Part I: RNA-seq analysis
Part II: scRNA analysis

RNA-seq applications
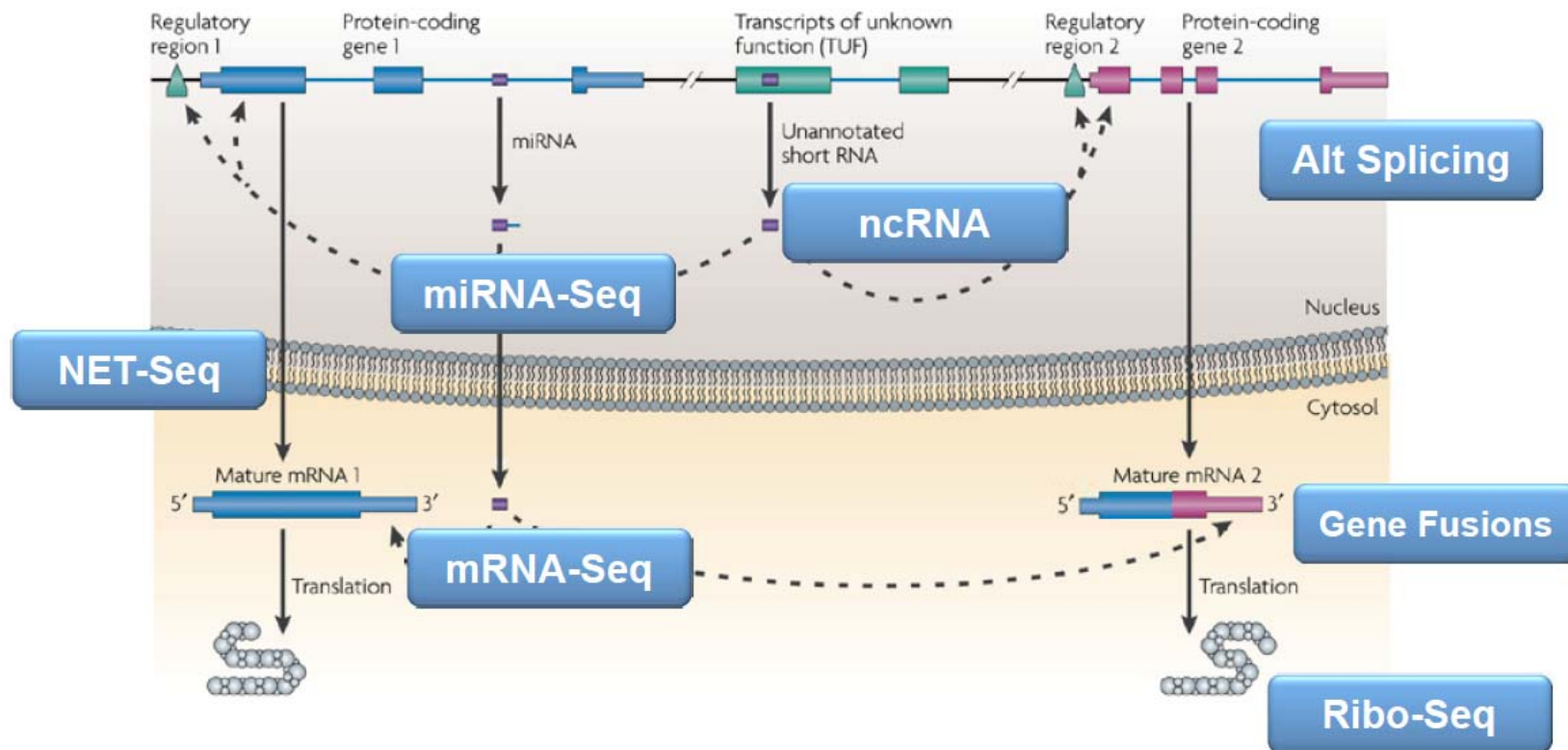
Scope of Transcriptome Sequenicng

Single-cell RNA-Seq
scRNA-Seq analyzes gene expression of individual cells. Single cells are isolated prior to RNA extraction.

1. RNA isolation

2. RNA selection

3. cDNA library preparation

4. Fragmentation, size selection

5. Adapter ligation

6. Validation, quantification

7. Sequencing, analysis

ATGCAGCGACGACGAGCACG
CGACTACGACTCGACTCACG
GCACGACGAGTGTTGAGCAG
GGCAGCGTGCAGCGACGGAC

Note: Details of the workflow can vary depending on the chosen sequencing platform.

Total RNA-Seq
Whole transcriptome analysis omits selection in order to sequence coding and non-coding RNA

mRNA-Seq
mRNA can be enriched via poly-A selection or ribosomal RNA depletion

Small RNA-Seq
Size selection methods can be used to enrich for noncoding RNA such as miRNA

Targeted RNA-Seq
Specific RNA sequences can be selected using hybridization probes

https://www.biocompare.com/Molecular-Biology/26572-RNA-Seq/

RNA integrity - BioAnalyzer

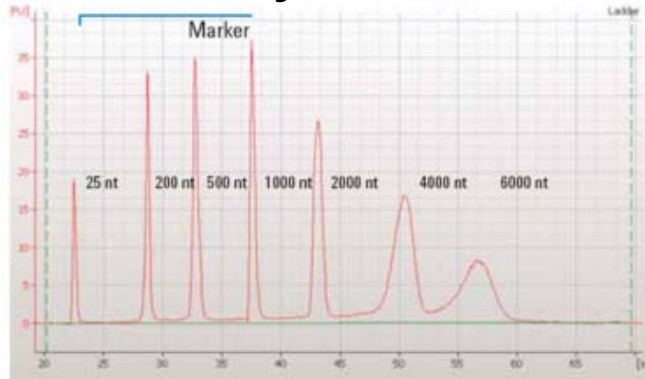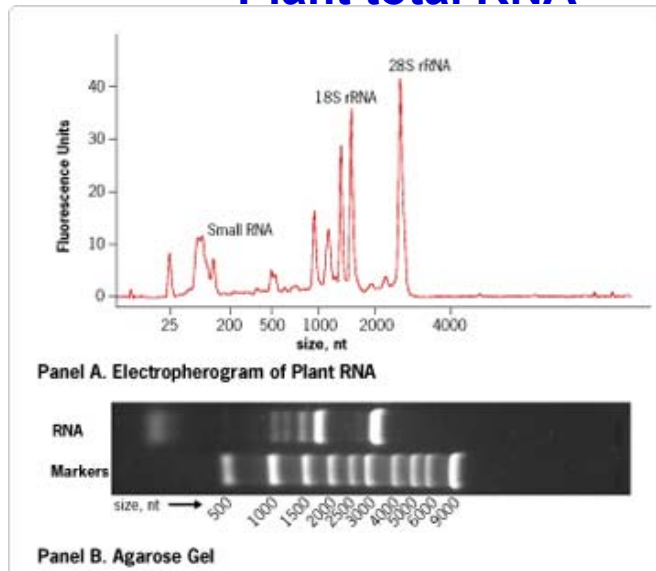## BioAnalyzer RNA ladder

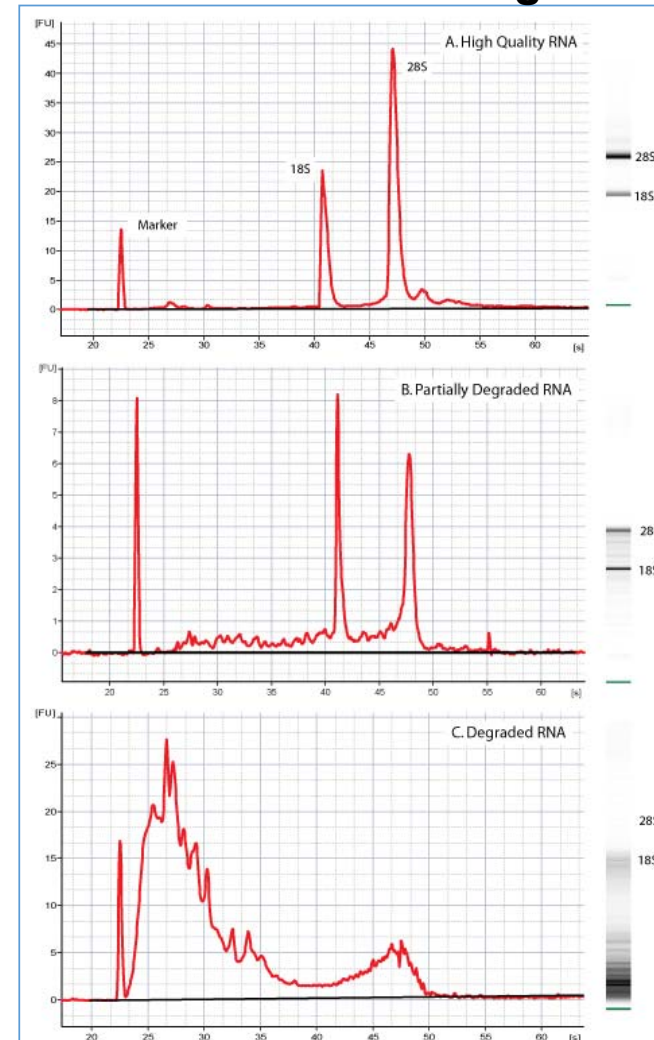Figure 1   RNA 6000 Nano ladder

Plant total RNA

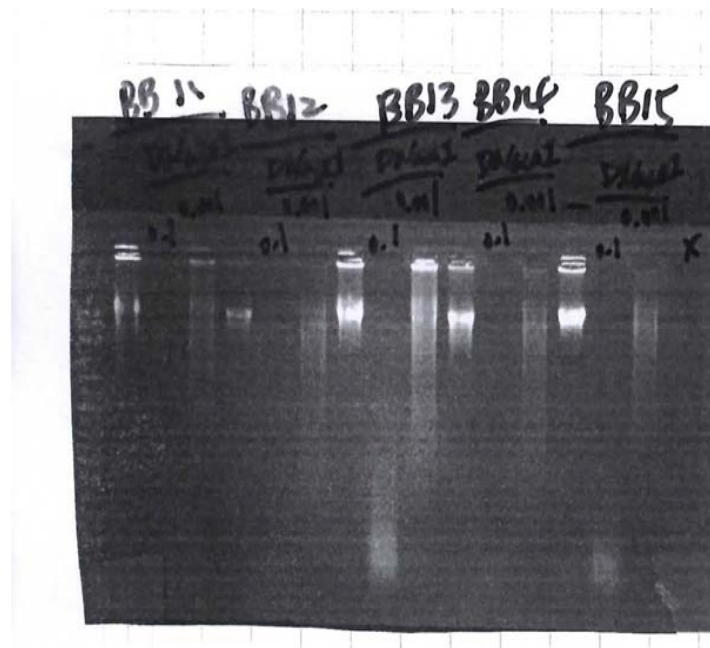Panel A. Electropherogram of Plant RNA
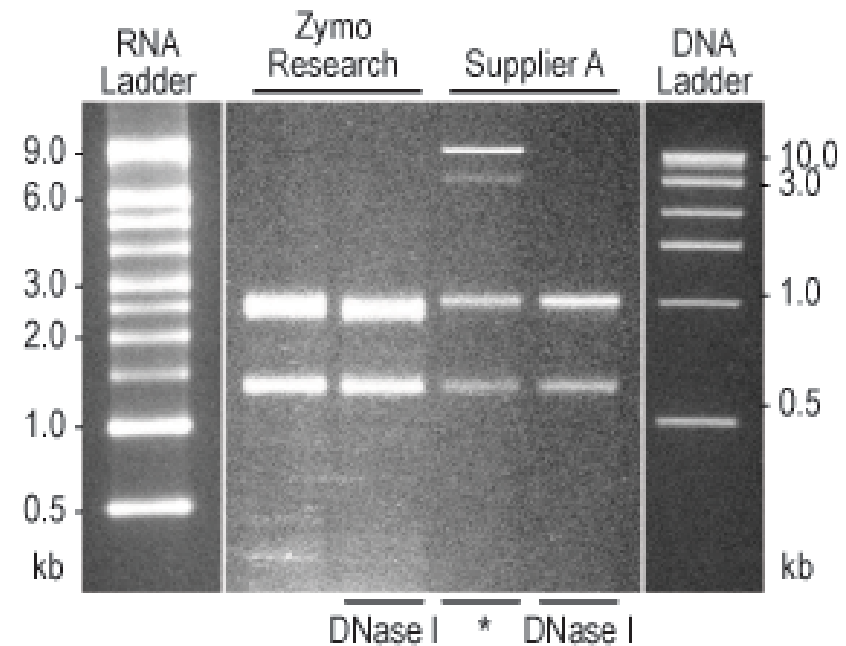
Panel B. Agarose Gel
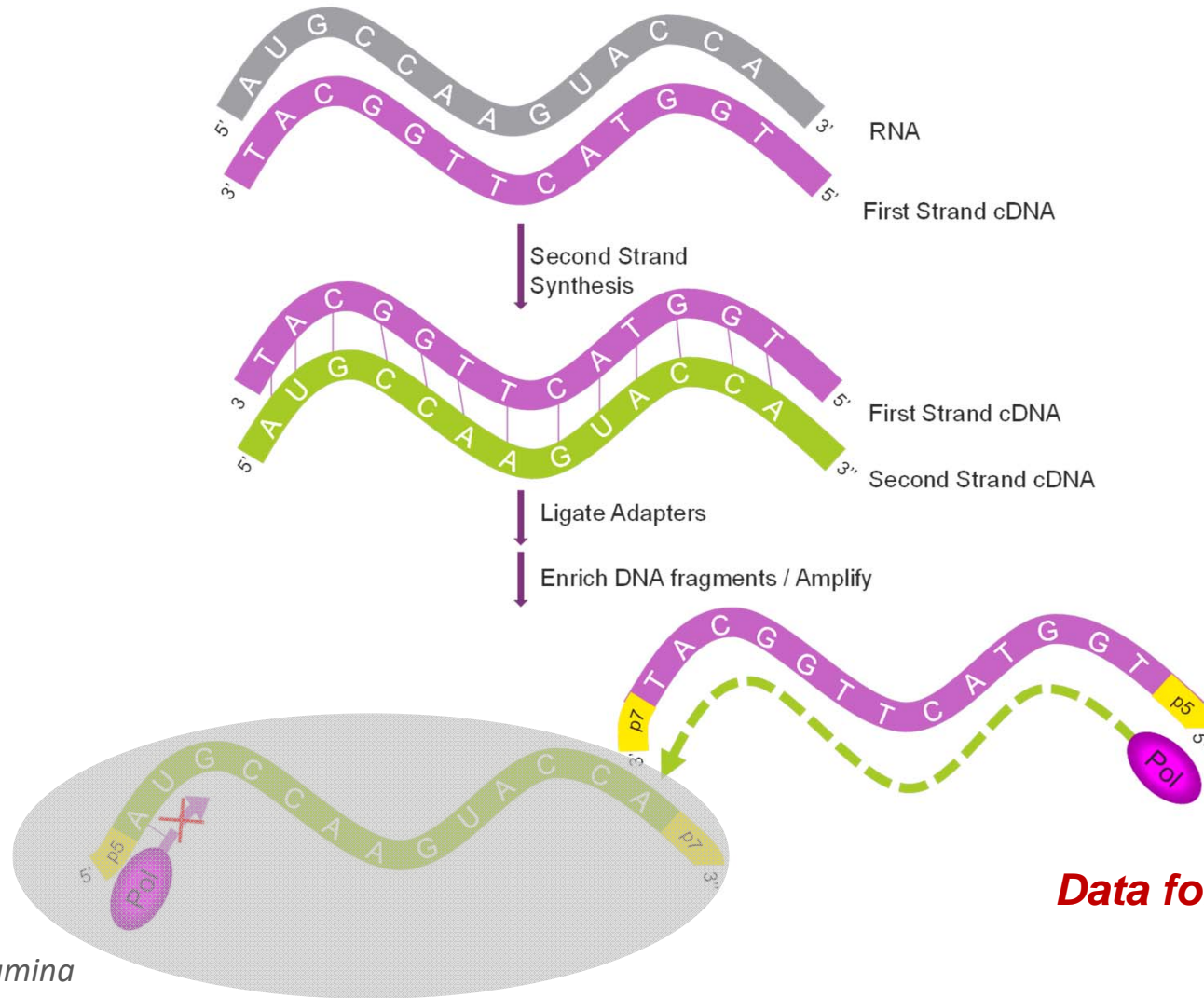
Human RNA – various degradation

DNase I treatment

gDNA sample

RNA sample

Strand-specific RNA-seq prep

Source: Illumina

Data format: RF

Types and Characteristics of NGS Reads

- **Read length:**   **Short**   **Long**

    50-300bp   *500-15,000bp*

- **Read types:**

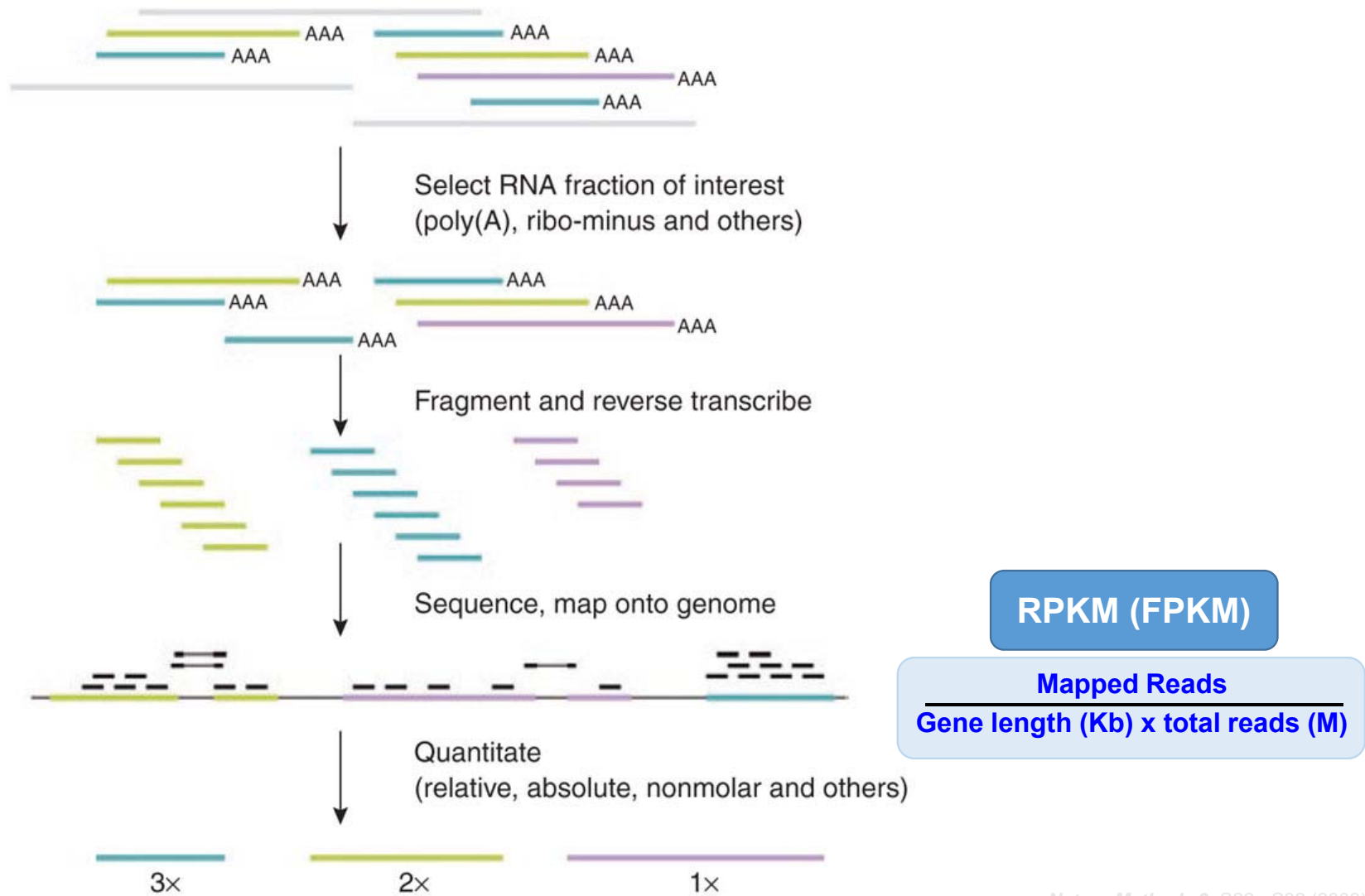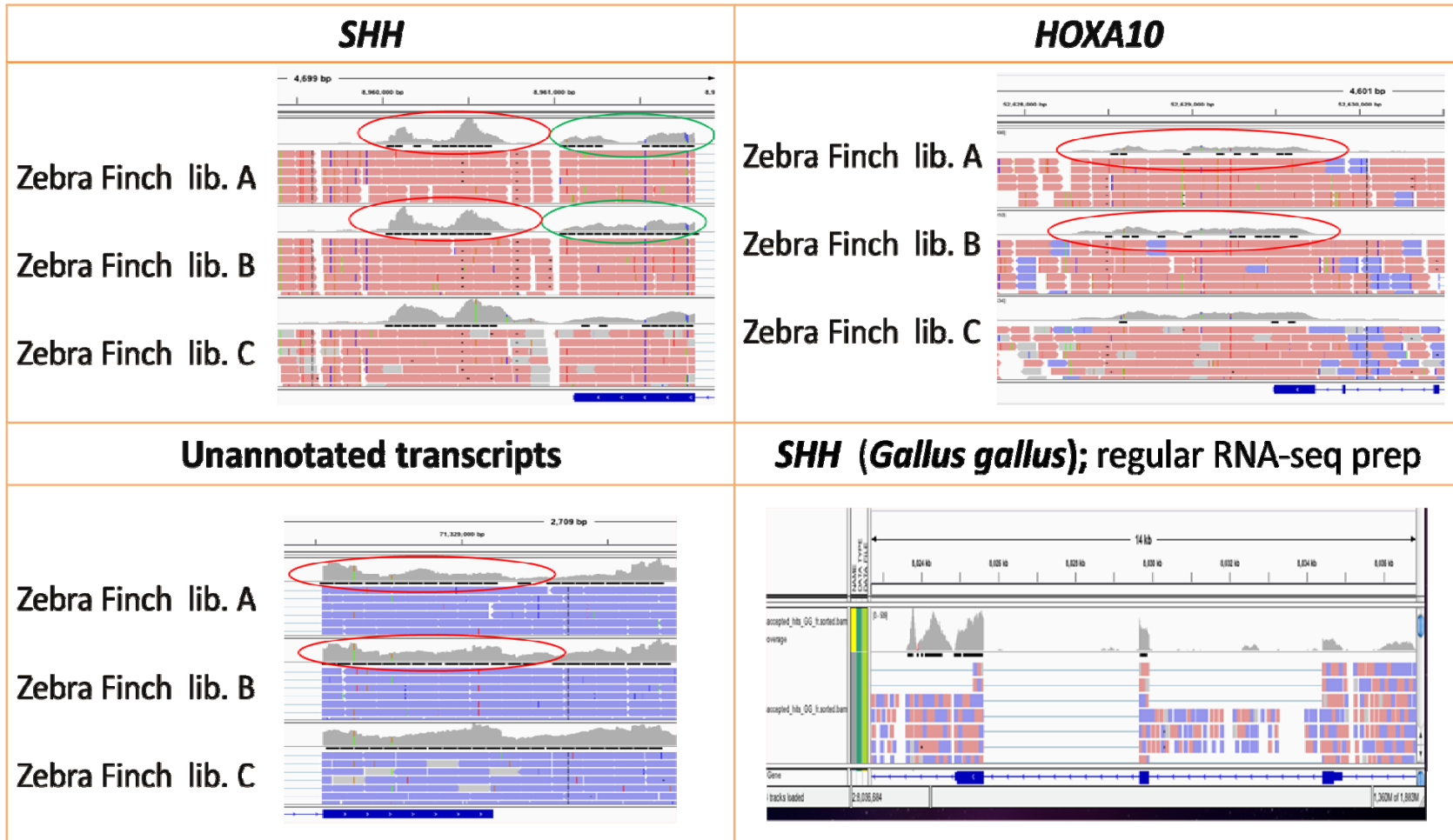   **SR** (single end)   50bp-20kb

    PE (paired-end)   50-300 bp;
                                                    1~1.5 kb jump

Transcriptome profiling: RNA-seq

Select RNA fraction of interest
(poly(A), ribo-minus and others)

Fragment and reverse transcribe

Sequence, map onto genome

Quantitate
(relative, absolute, nonmolar and others)

3×   2×   1×

RPKM (FPKM)

$$\frac{\text{Mapped Reads}}{\text{Gene length (Kb) x total reads (M)}}$$

8

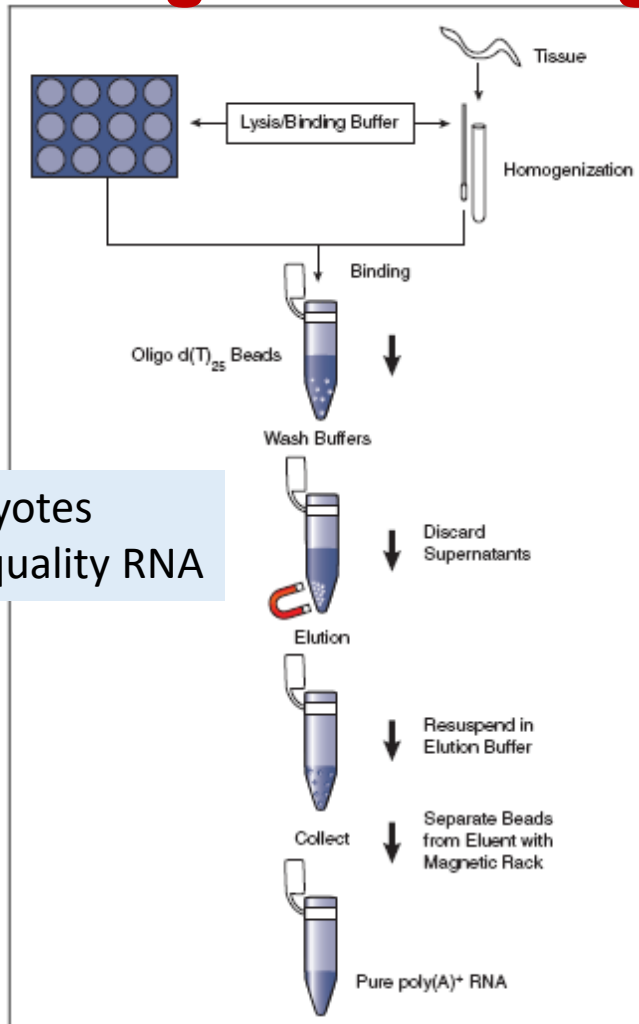# Stranded RNA-seq: >90% strand-specificity

*Genome viewer: IGV genome viewer (version 2.2), plus strand(Blue), minus strand(Red), no paired end info (could be plus or minus strand, Gray). In non-stranded protocol, the amount of blue alignments are almost equal to red alignment
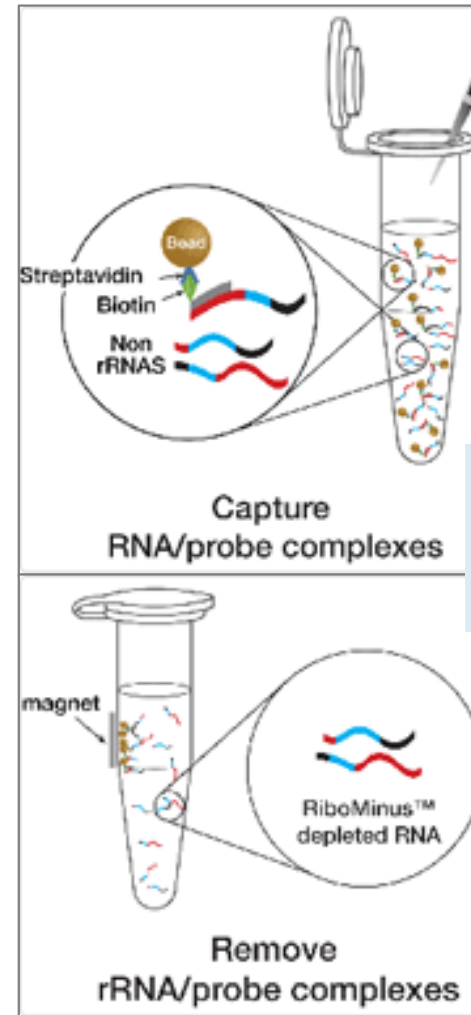
Courtesy of Chih-kuan Chen

Methods of mRNA enrichment

Oligo-dT binding



Eukaryotes
High quality RNA

rRNA removal

Prokaryotes
Non-A-tailed RNA
Degraded RNA

Iso-Seq sample QC requirements
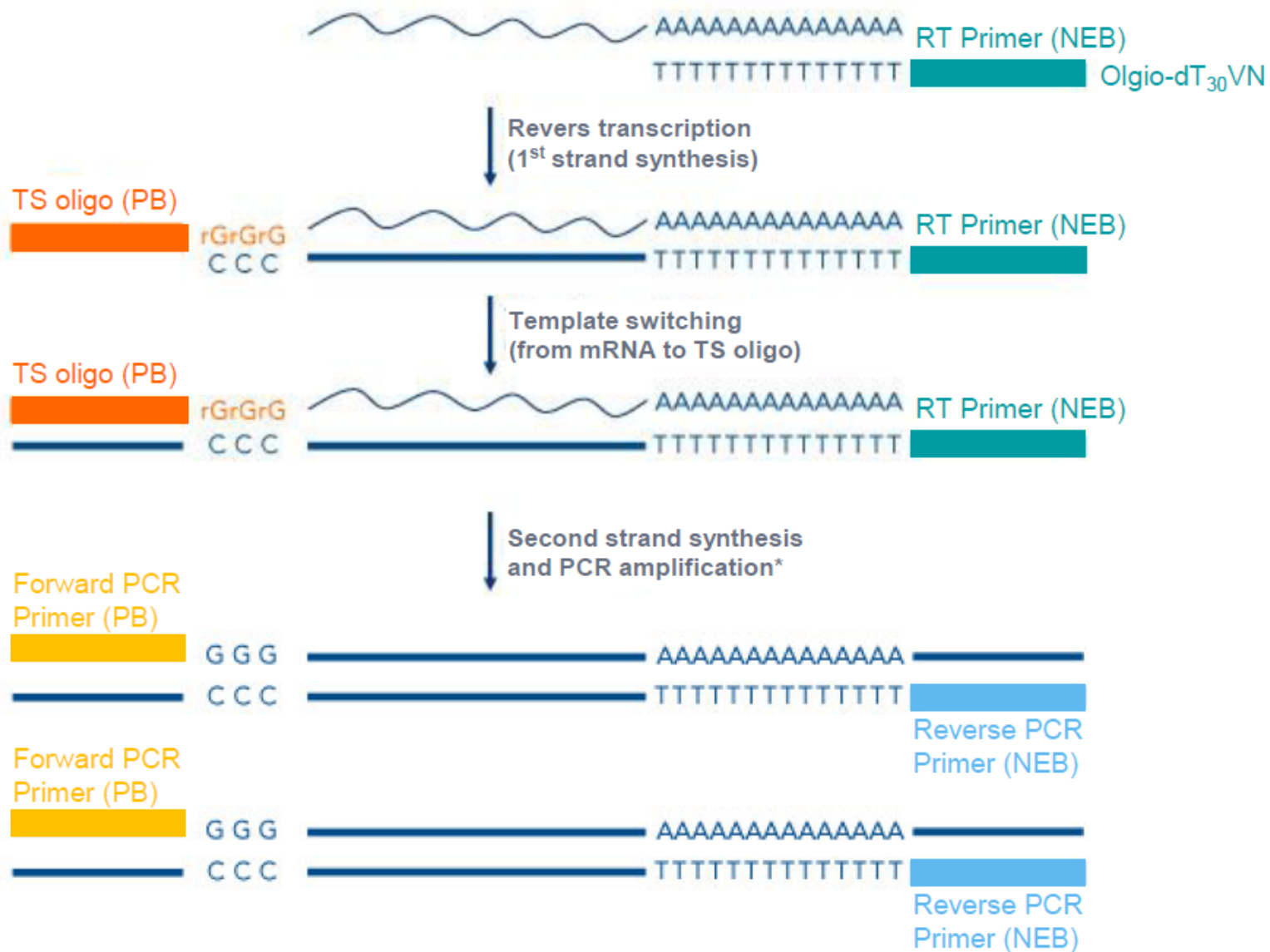
Evaluation of input total RNA sample integrity

- Sample QC of input total RNA samples should be assessed by measuring the RNA Integrity Number (RIN) using a Bioanalyzer 2100 instrument (Agilent Technology)

- RIN score (1 to 10) is related to the ratio of the area under the 28s and 18s fragment peaks and also takes into account the signal intensity above the baseline in the Inter-Region and Fast Region since this is where degradation products appear

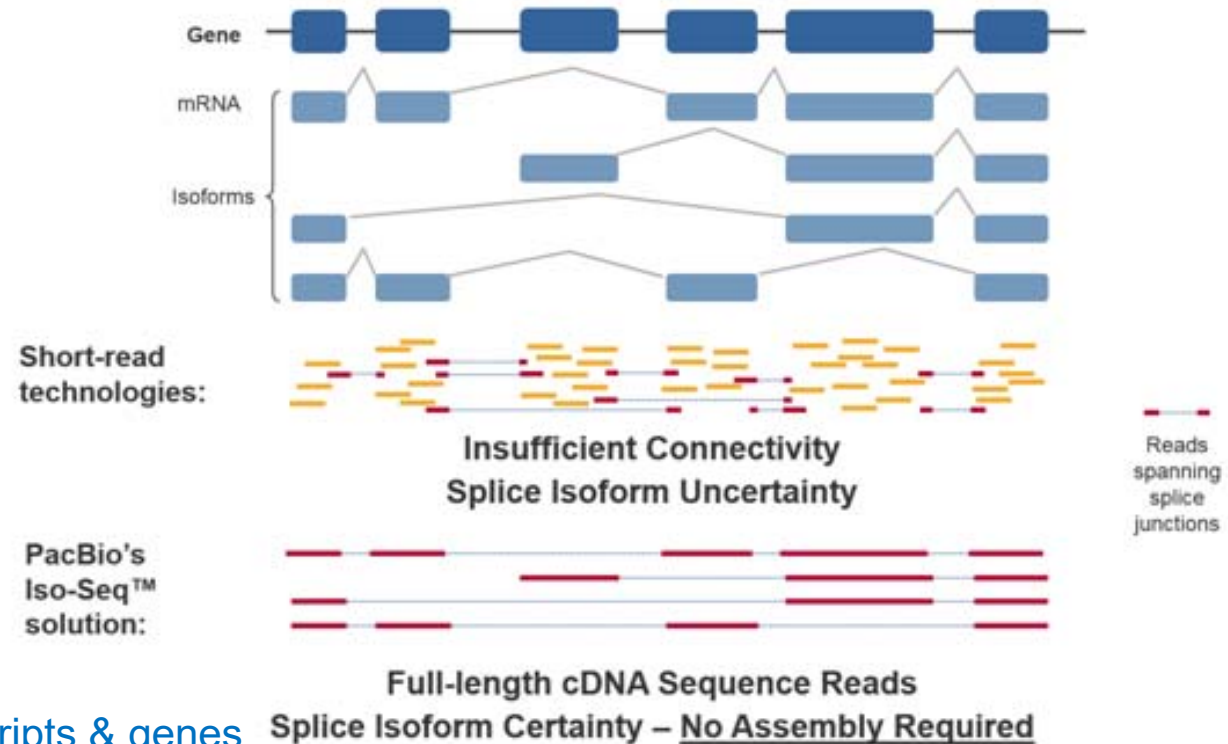- Higher RIN numbers are correlated with better overall sample quality and lower degradation
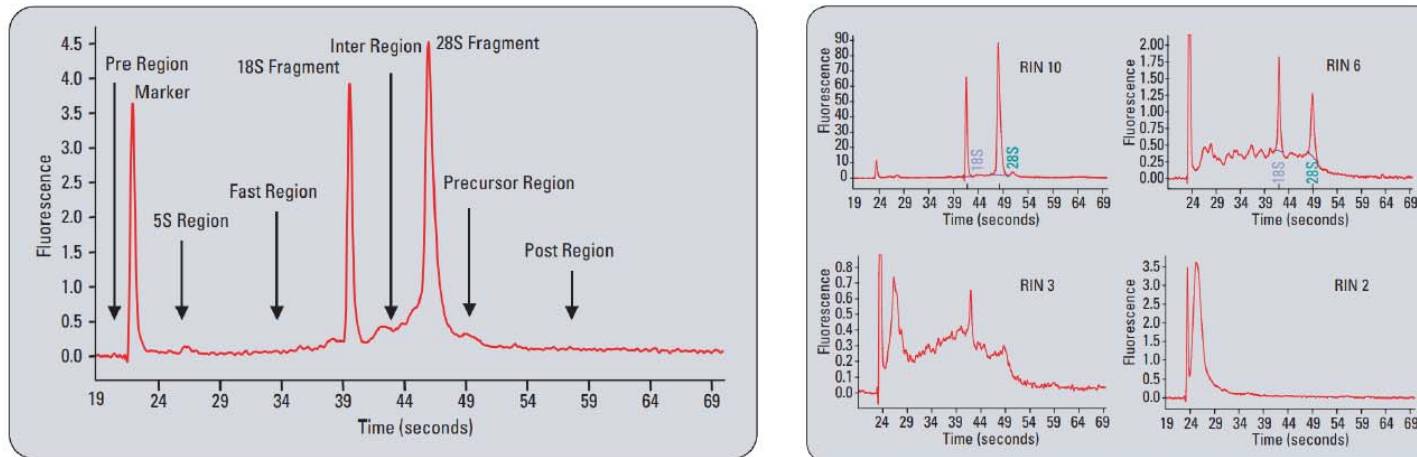
Left: Bioanalyzer electropherogram detailing the regions that are indicative of RNA quality. Right: Sample electropherograms corresponding to different RNA Integrity Number (RIN) scores. Samples range from intact (RIN 10), to degraded (RIN 2). Images from Agilent Application Note: RNA Integrity Number (RIN) – Standardization of RNA Quality Control (5989-1165EN)

A RIN ≥7.0 (ideally ≥8.0) is sufficient for the Iso-Seq protocol. Samples with a RIN <7.0 can be processed, but the risk of significant underperformance or even failure is greatly increased.

PacBio

**Starting total RNA input (~300 ng)**
(Contains poly(a)-tailed mRNA)

AAAAAAAAAAAAA RT Primer (NEB)

TTTTTTTTTTTTT Olgio-dT$_{30}$VN

Revers transcription
(1st strand synthesis)

TS oligo (PB)

rGrGrG AAAAAAAAAAAAA RT Primer (NEB)

C C C TTTTTTTTTTTTT

Template switching
(from mRNA to TS oligo)

TS oligo (PB)

rGrGrG AAAAAAAAAAAAA RT Primer (NEB)

C C C TTTTTTTTTTTTT

Second strand synthesis
and PCR amplification*

Forward PCR
Primer (PB)

G G G AAAAAAAAAAAAA

C C C TTTTTTTTTTTTT

Reverse PCR
Primer (NEB)

Forward PCR
Primer (PB)

G G G AAAAAAAAAAAAA

C C C TTTTTTTTTTTTT

Reverse PCR
Primer (NEB)

Isoform Sequencing by PacBio

- Discover novel transcripts & genes
- Identify fusion genes
- Resolve alternative polyadenylation
- Identify retained introns
- Find anti-sense transcription
- Annotate gene isoforms & alternative splicing events
- Recover missing exons
- Improve isoform-abundance quantification accuracy

NGS High Throughput Genomics Core at BRCAS
新世代基因體定序核心實驗室
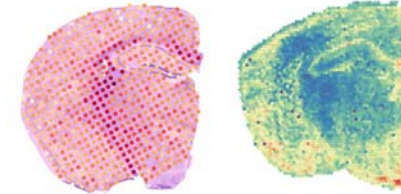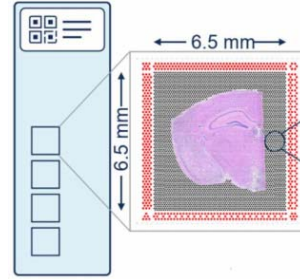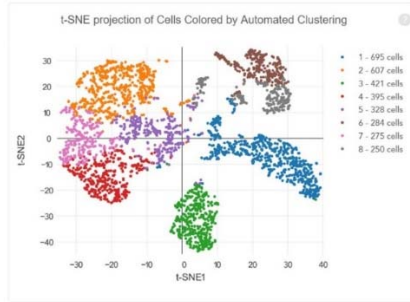
Single-Cell

Spatial seq

10x Genomics

Countess *Chromium X*

CytAssist *EVOS*

HiSeq2500 **MiSeq** **NextSeq2000** **PacBio Sequel & SQIIe** **Nanopore GridION** **Promethion P2 Solo**

14

New Tech on transcriptome sequencing

- **10x Genomics: Single-cell & Spatial analyses**

- **PacBio MAS/Kinnex: HT-HiFi**

Project considerations

Single cell

Sample:
- Source: repeatability
- Cell count, viability, purity
- Complexity
- Target cell%
- Cell vs nuclei

Experimental design
- Contrasts, time course
- Replicates
- Sequencing scale
- Bioinformatics
- Expected cell types for clustering

Spatial

Sample:
- Source: repeatability
- Tissue property & Complexity
- Embedding quality
- Tissue size & target region%
- Sample optimization

Experimental design
- Contrasts, time course
- Replicates
- Sequencing scale
- Bioinformatics
- Domain annotation: histology vs immuno-staining

10x Genomics: Single-cell applications

Countess III   *ChromiumX*
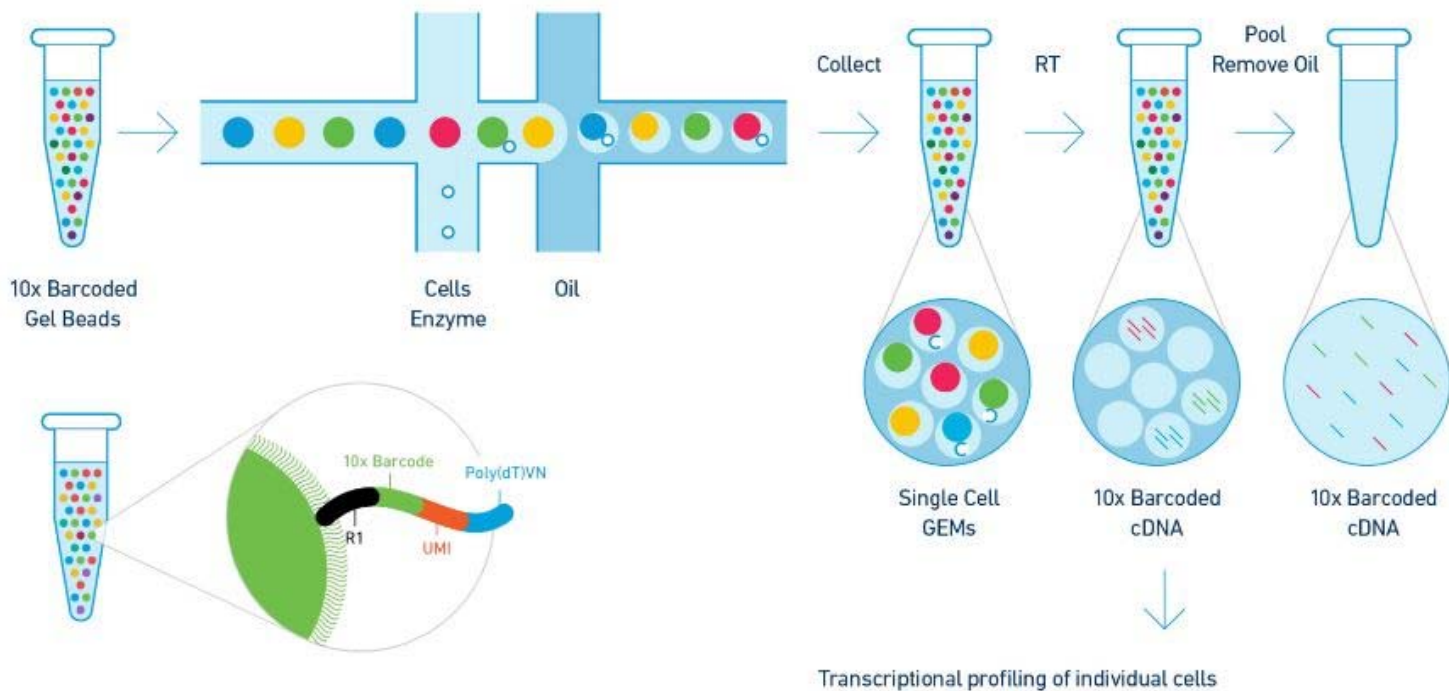
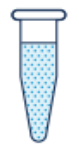| **10x Single Cell** |
| --- |
| 10x Chromium 3′ Single Cell RNA Prep |
| Single Cell Chip G |
| 10x Chromium VDJ 5′ Single Cell RNA Prep |
| Single Cell Chip K |
| 10x ATAC-seq prep (Chromatin accessibility) |

Chromium X
10X Chromium X Single Cell Multiome
10X Chromium X Single Cell ATAC

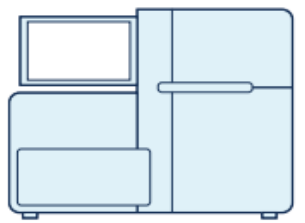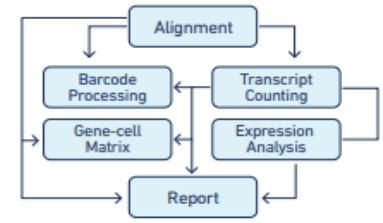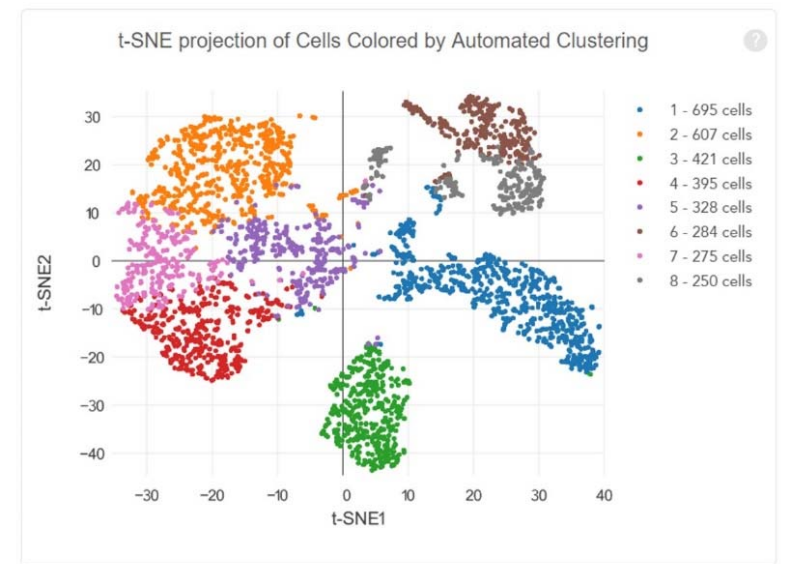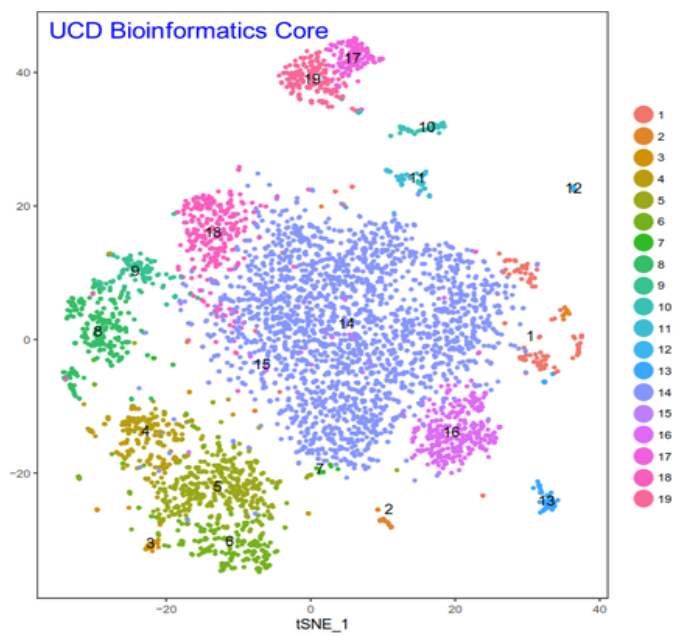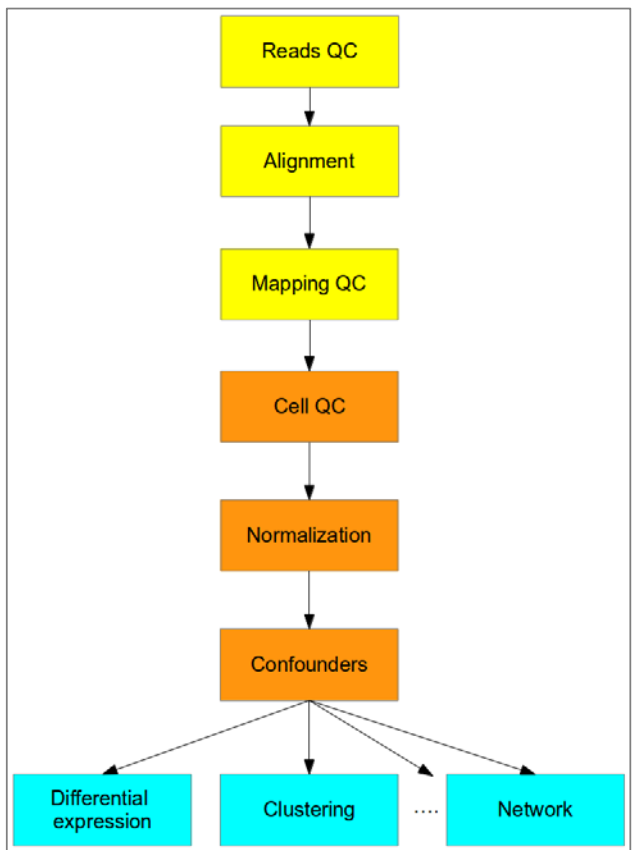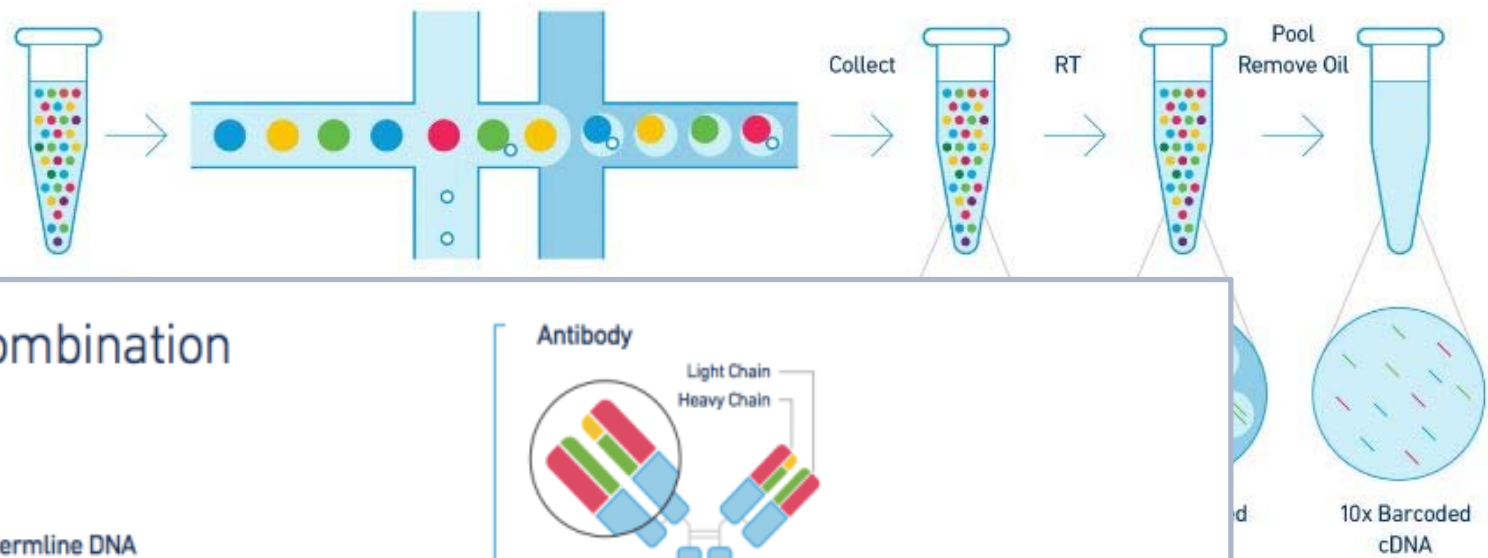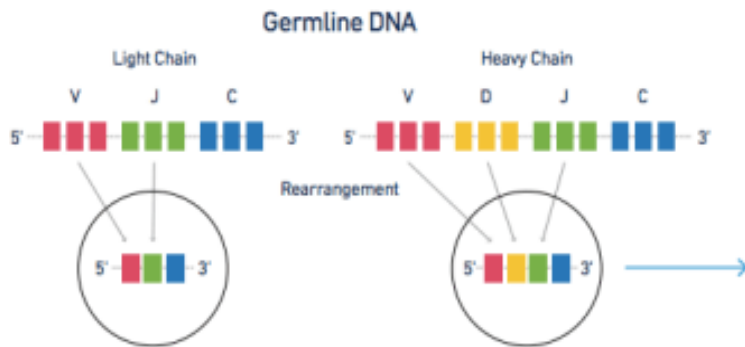| Input	Library Construction	Sequencing	Data Analysis	Data Visualization
Cell Suspension	Barcoding & Library Construction	Sequence Transcriptome	Pipelines	Report & Visualization
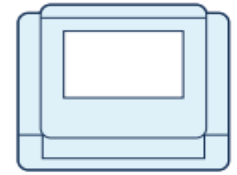
Pipelines: Alignment → Barcode Processing / Transcript Counting → Gene-cell Matrix / Expression Analysis → Report

Pipeline flow:
- Reads QC
- Alignment
- Mapping QC
- Cell QC
- Normalization
- Confounders
- Differential expression / Clustering / .... / Network

UCD Bioinformatics Core (tSNE_1 vs tSNE, clusters 1–19)

t-SNE projection of Cells Colored by Automated Clustering
- 1 - 695 cells
- 2 - 607 cells
- 3 - 421 cells
- 4 - 395 cells
- 5 - 328 cells
- 6 - 284 cells
- 7 - 275 cells
- 8 - 250 cells

https://dnatech.genomecenter.ucdavis.edu/single-cell-analyses/

10x Genomics: Single-cell applications

V(D)J Recombination

Germline DNA

Light Chain
V   J   C

Heavy Chain
V   D   J   C

Rearrangement

Antibody
Light Chain
Heavy Chain

B-Cell Receptor
Light Chain
Heavy Chain

T-Cell Receptor
Variable Regions
Constant Regions
Transmembrane Region
β Chain
α Chain

Collect   RT   Pool Remove Oil

10x Barcoded cDNA

Single-cell VDJ
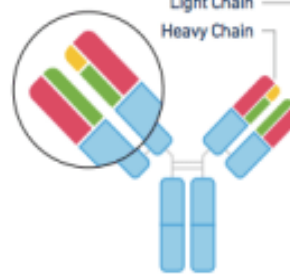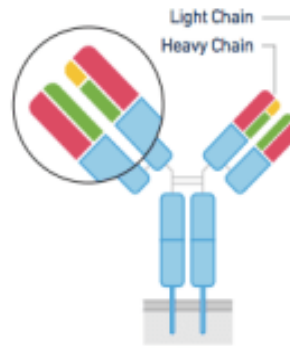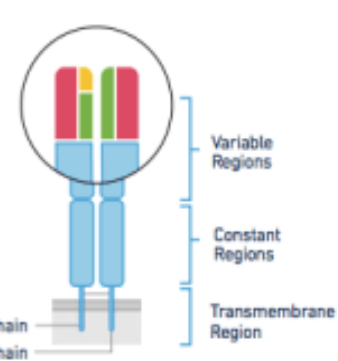
10x Multiome: Sc-ATAC + snRNA
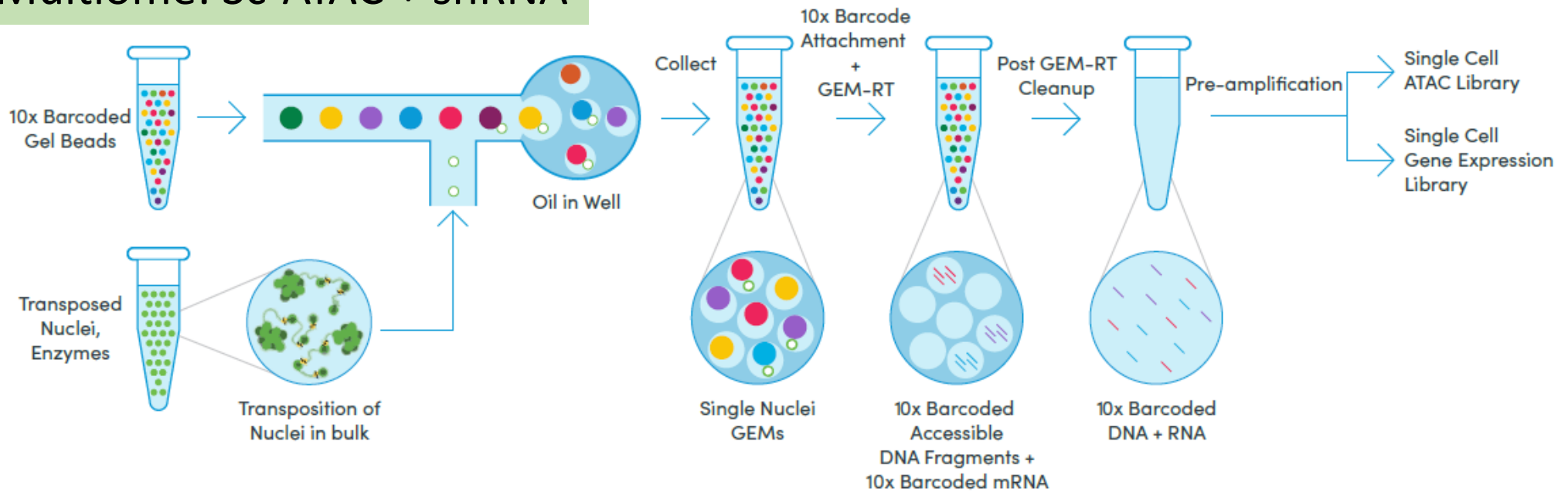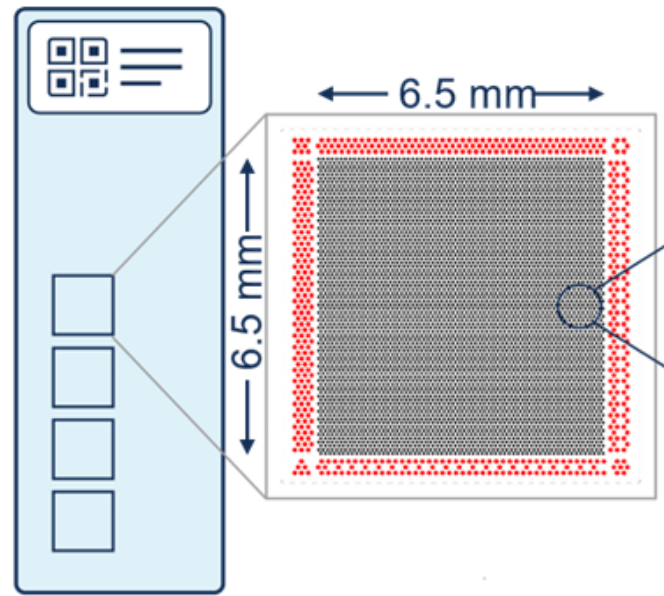

Solution Features

- Integrate gene expression and epigenomic landscape through direct measurement in the same cell, eliminating the need for inferring relationships in silico

- Identify linkages between putative regulatory elements and their target genes

- Simple and robust workflow

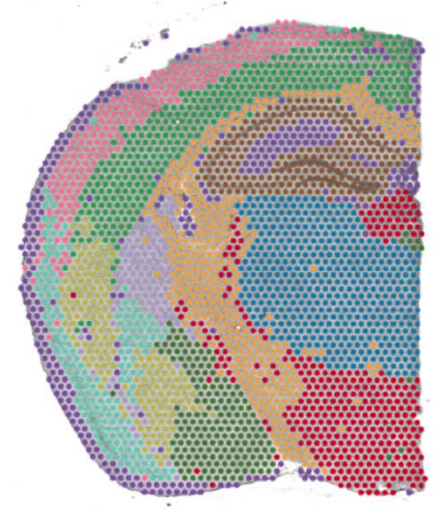- Easy-to-use software for data analysis and visualization

System Features

- Efficiently partition 500–10,000 nuclei per channel, for up to 80,000 nuclei per run

- Scalable; run up to 8 samples in parallel

- Recover up to 65% of loaded nuclei

- High sensitivity

- Low microfluidic multiplet rate (<1% per 1000 nuclei)

- Demonstrated with cell lines, primary cells, cryopreserved samples, and fresh and flash-frozen tissue

10x Genomics: Spatial Transcriptome

Annotation directly on H&E image
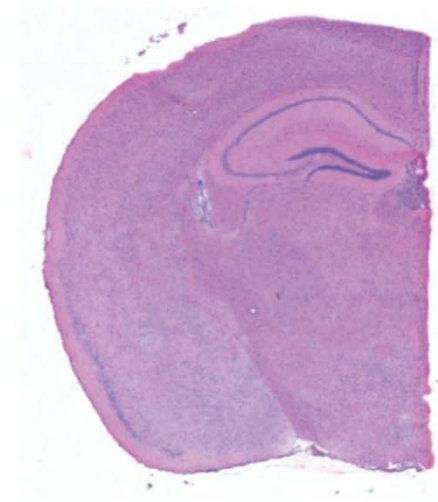
Captured H&E image from the microscope

Thalamus
Lateral forebrain bundle
Isocortex (deep)
Hypothalamus
Cluster
Hippocampus
Isocortex (sup.)
Amygdalar nuclei
Cortical subplate
Piriform area
Caudoputamen

Lct
Dsp
C1ql2
Lefty1
Spink8
Crlf1
Fibcd1
Cabp7
Klk8
Gm2115
Lrrc10b
Cebpd
Iqgap2
Wipf3
Homer3
Neurod6
Nr3c2
Hpca
Shisa6
Pcdh20

Visium Spatial Gene Expression Slide
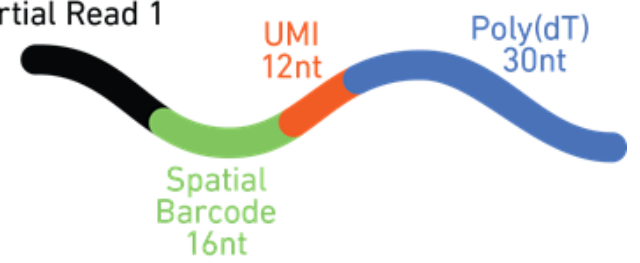
Captur
~5000
S

6.5 mm

6.5 mm

100 μm

55 μm

Partial Read 1

UMI 12nt

Poly(dT) 30nt

Spatial Barcode 16nt

10x Genomics: Spatial transcriptome

OCT

FFEP

CytAssist

EVOS

10x Visium

(X-VTO)	10X Visium Tissue Optimization
(X-VGE)	10X Visium Gene Expression
(X-VFP)	10X Visium Gene Expression for FFPE
(X-VFP-Cyt)	10X Visium CytAssist Gene Expression for FFPE

Visium Projects Processed

Organisms
- Mouse
- Human
- Maize

Tissue types
- Brain
- Cerebellum
- Liver, spleen
- Gum/Cheek
- Lung
- Leaf

Complete setup for spatial transcriptome workflow
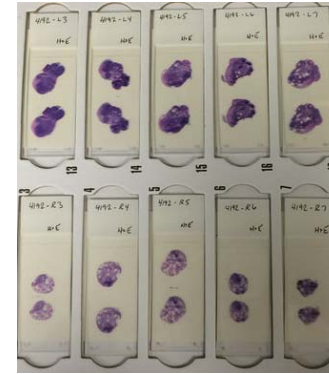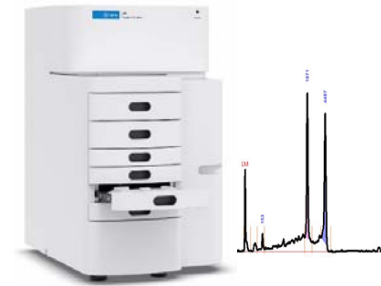
1. OCT embedding or FFPE blocks

2. Cryosectioning

3. H&E staining

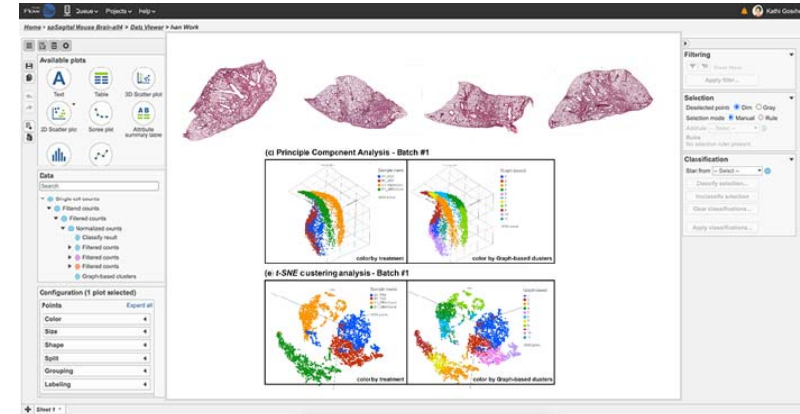## 4. Tissue catalog

5. RNA QC

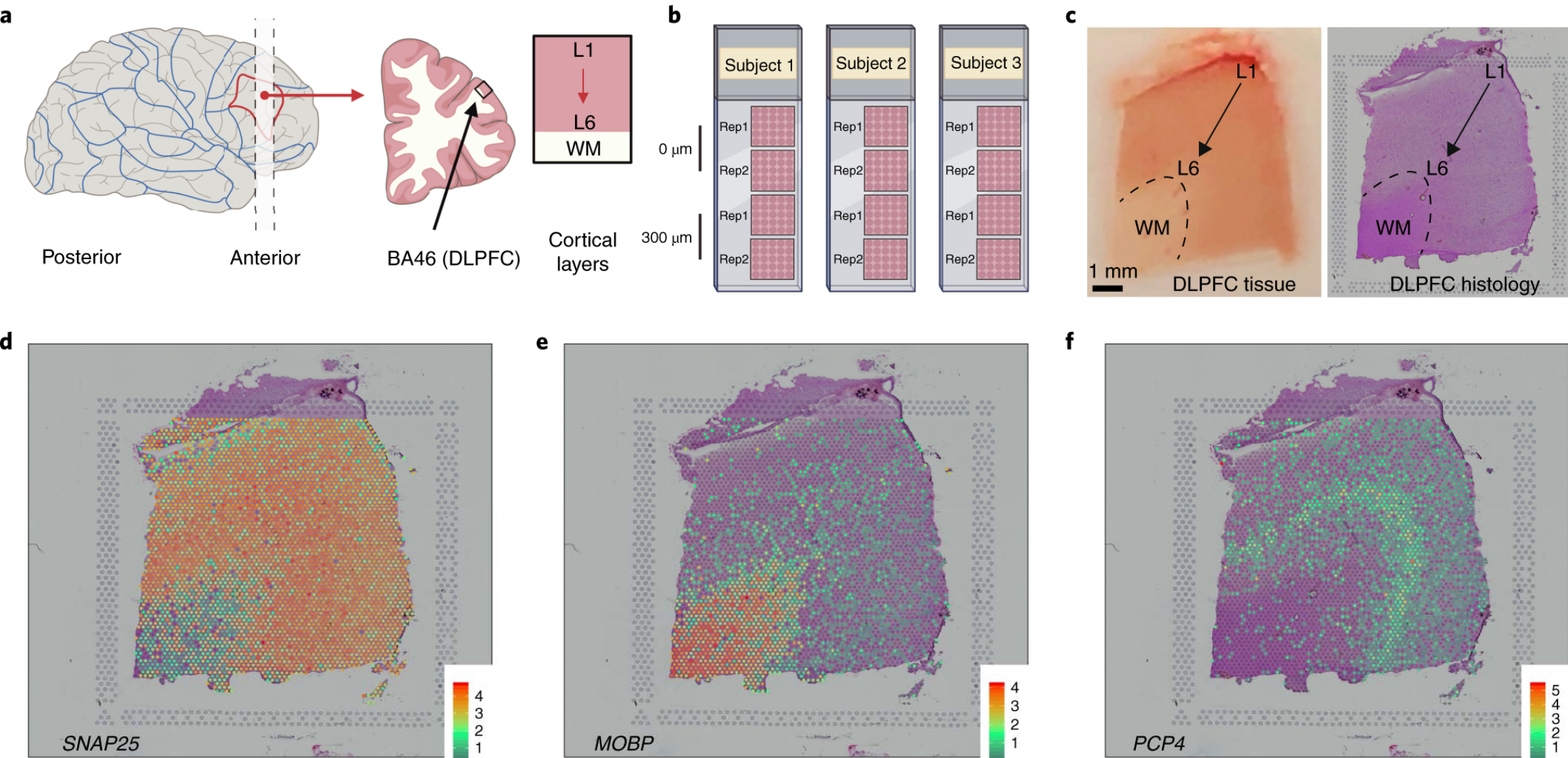## 6. Visium prep & Sequencing

7. Fluorescent imaging

8. Data analysis with GUI interface

Maynard, K.R., Collado-Torres, L., Weber, L.M. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci 24, 425–436 (2021). https://doi.org/10.1038/s41593-020-00787-0

The 10x projects we have processed

Organisms

- Mouse
- Human
- Maize
- Bird
- Marine

Samples
- Flash frozen (OCT)
- FFPE

Tissue types

- Brain
- Cerebellum
- Liver
- Gum/cheek
- Spleen
- Lung
- Embryonic leaves
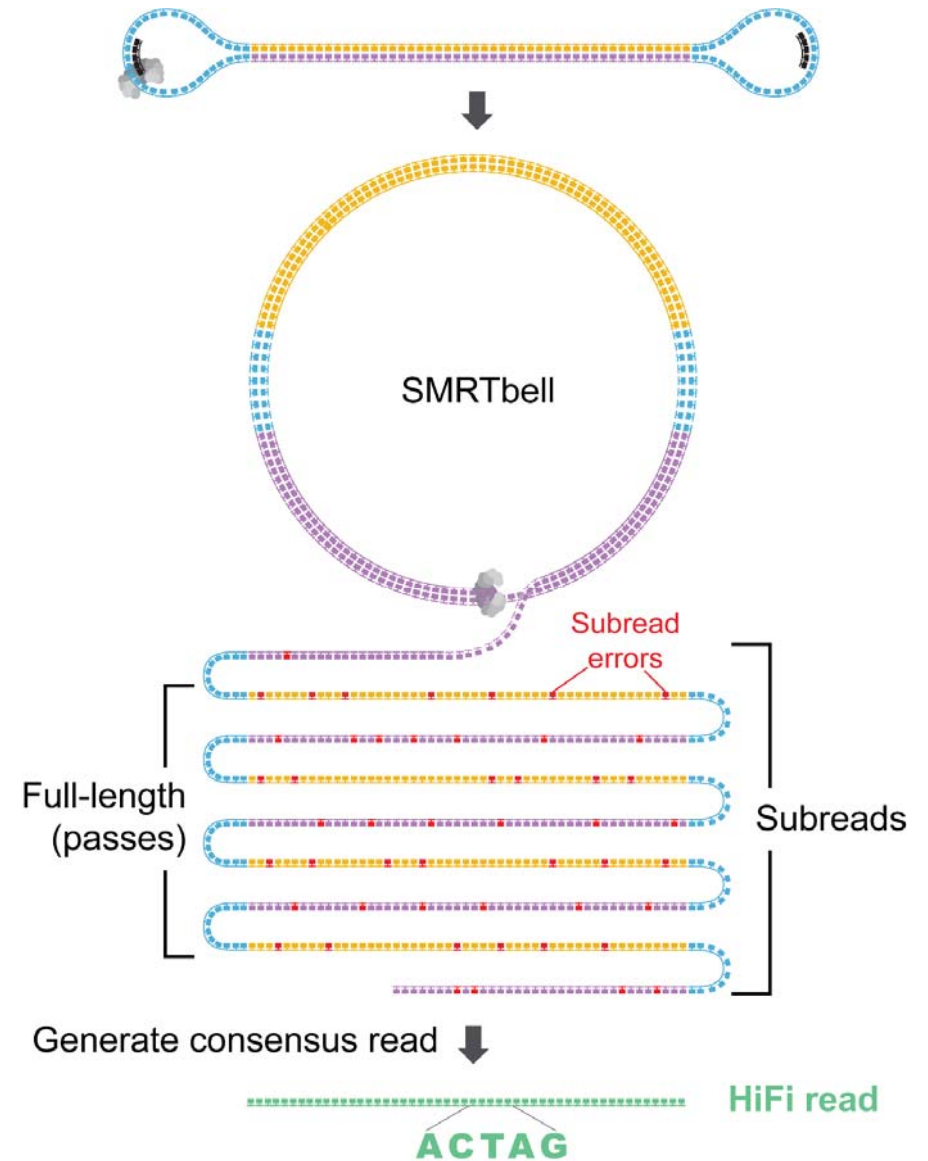- Skin
- More plants (ongoing)

Biological Qs

- Cardiac hypertrophy
- Chronic pain mechanism
- Neuronal degeneration
- Cancer/tumor
- Stress/injury
- Covid immuno-drug develop.
- Plant embryogenesis
- C4 photosynthesis evolution
- Skin repair / hair regeneration

Highly accurate PacBio HiFi
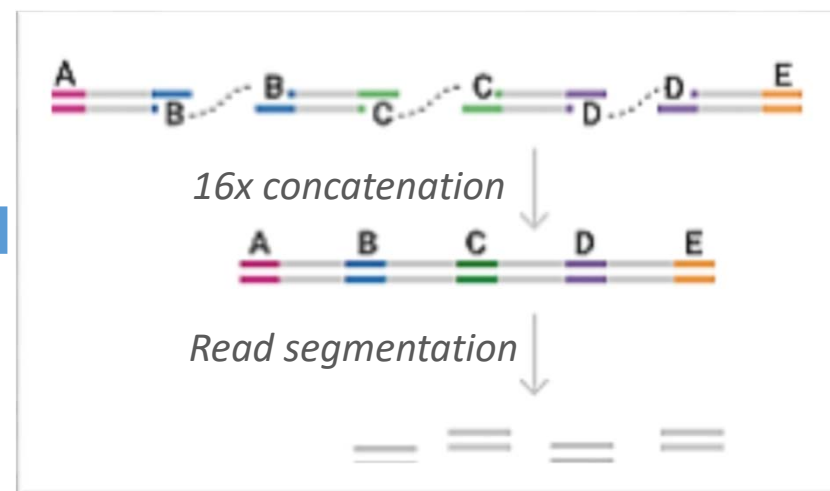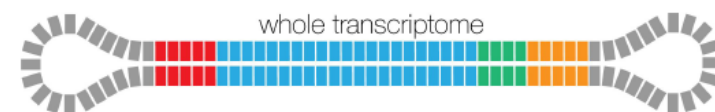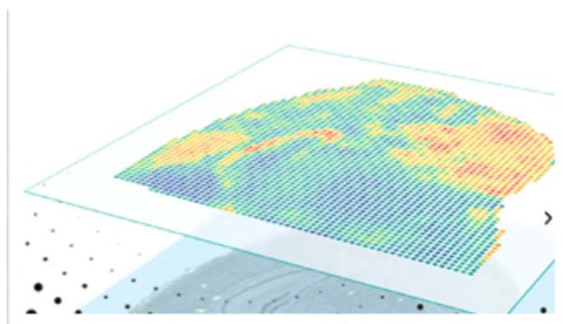
Now we have CCS / HiFi data...

What can we use it for?

Technology merge: 10x + MAS-Isoform Sequencing

PacBio Multiplexed Array Sequencing:
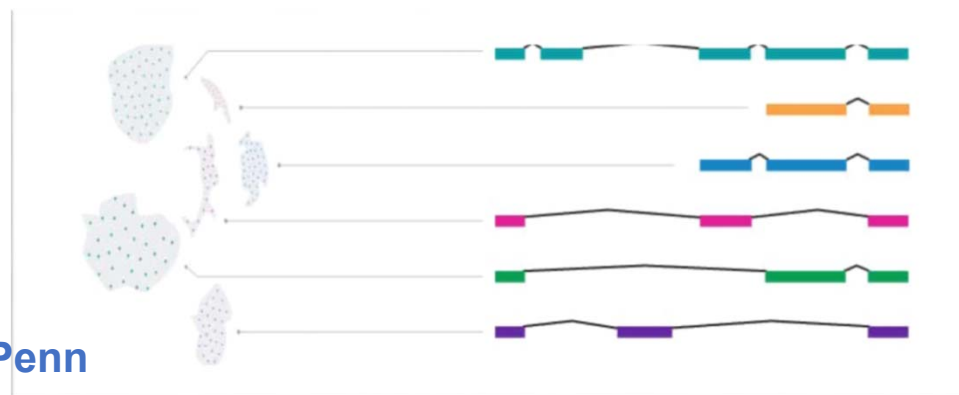
- **10X scRNA**
- **10X Spatial-Visium**

"...Detection of 214,516 unique isoforms covering 22,391 genes, 72.6% of the isoforms are novel."

Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders.
doi: https://doi.org/10.1101/2023.03.25.534016

UCLA+UPenn

*Kinnex Promotion:
8-12 SMRTcells
(while supply last!)*

MAS/Kinnex Applications

	MAS-seq	Academy	Industry
(P-M)	MAS-Seq SMRTbell Prep for FL16S	1,200	1,500
(P-Nh)	MAS-Seq SMRTbell Prep for Bulk Iso-Seq (homebrew)	22,200	27,750
(P-Nk)	MAS-Seq SMRTbell Prep for Bulk Iso-Seq (Kinnex)	10,800	13,500
(P-MSC)	MAS-Seq SMRTbell Prep for 10X Single Cell	26,100	32,630
(P-MVG)	MAS-Seq SMRTbell Prep for 10X Visium	26,100	32,630

#More about MAS-Seq: Kinnex full-length RNA kit, Kinnex 16S rRNA kit, Kinnex single-cell RNA kit.

PacBio Sequel IIe Sequencing Services

MAS/Kinnex Sequnecing

	Service Item	Academy	Industry
(PB-IIe)	Sequel IIe Seq Run	106,400	133,000
(PB-IIe-EB)	Sequel IIe Seq Run (early bird)	85,100	106,380
(PB-IIe-K)	Sequel IIe Seq Run (kinnex promotion)	65,000	81,250

Explore the power of full-length isoform profiling at Sc/Sp dimensions!