

<https://medicinaprecisionandalucia.iavante.es> | #PANMEP

PANMEP

PROGRAMA ANDALUZ DE FORMACIÓN EN MEDICINA PERSONALIZADA Y DE PRECISIÓN

EXPERTO UNIVERSITARIO EN MEDICINA PERSONALIZADA Y DE PRECISIÓN

TÉCNICAS

FORMACIÓN
IIVANTE
Fundación
Progreso y Salud



LUIS JAVIER MARTÍNEZ GONZÁLEZ, PhD

Unidad de Genómica, Genyo
Centro de Genómica e Investigación Oncológica
Pfizer | Universidad de Granada | Junta de Andalucía
Departamento de Bioquímica y Biología Molecular III
e Inmunología. Universidad de Granada.

COLABORA



ORGANIZAN



INTRODUCCIÓN AL ANÁLISIS DE ARN MEDIANTE SECUENCIACIÓN POR NGS

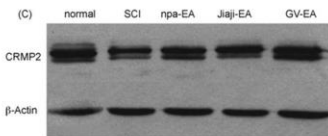
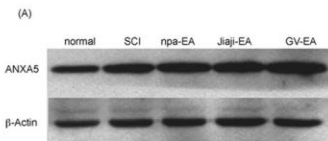
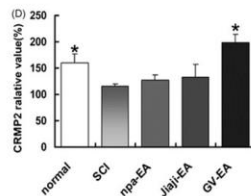
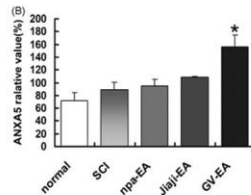
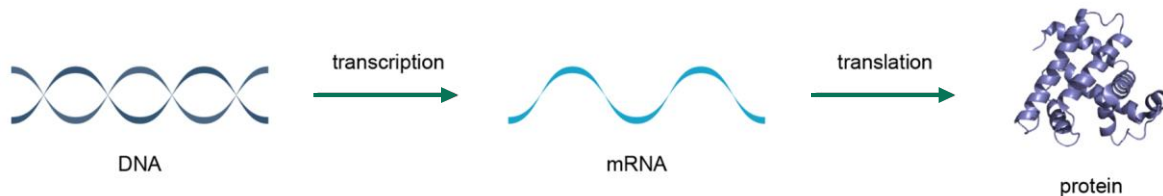
Estudios de Expresión



INTRODUCCIÓN AL ANÁLISIS DE ARN MEDIANTE SECUENCIACIÓN POR NGS

Dogma central de la biología molecular:

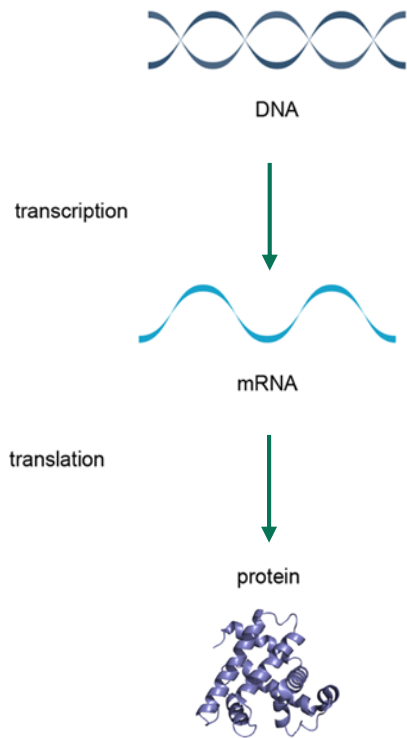
Un GEN -> Una PROTEINA (F. Crick 1958)



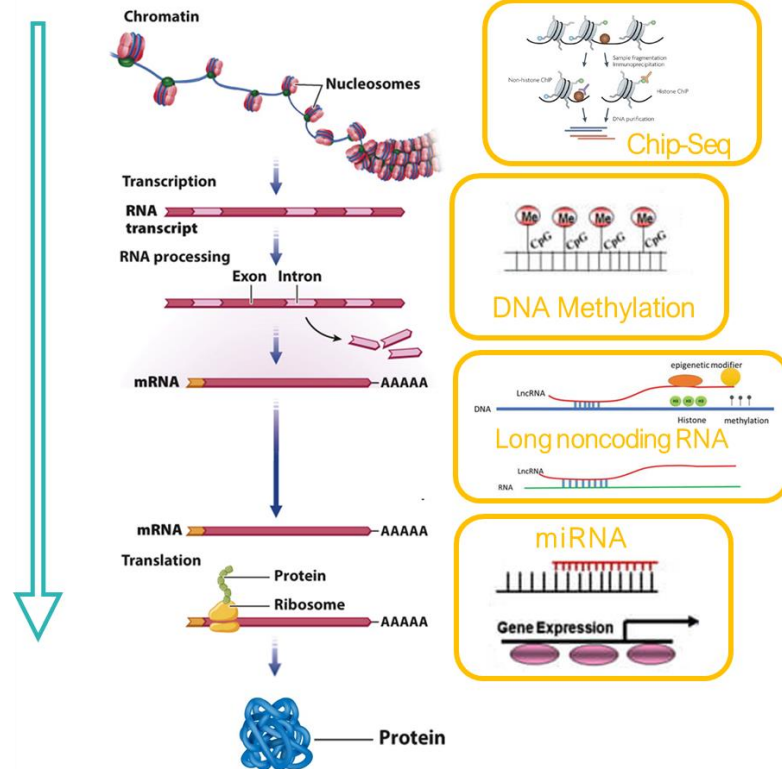
INTRODUCCIÓN AL ANÁLISIS DE ARN MEDIANTE SECUENCIACIÓN POR NGS

Dogma central de la biología molecular:

Un GEN -> Una PROTEINA (F. Crick 1958)



Levels of gene expression regulation

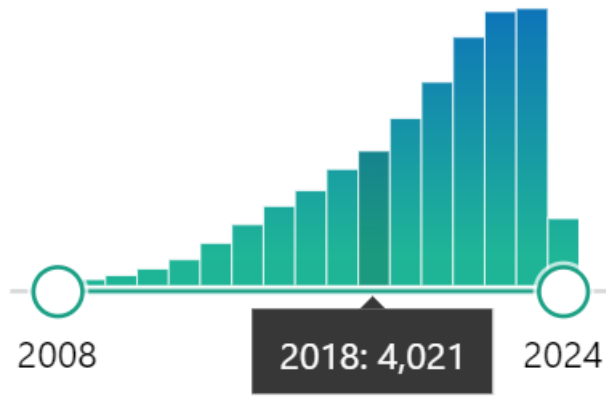


INTRODUCCIÓN AL ANÁLISIS DE ARN MEDIANTE SECUENCIACIÓN POR NGS

RNA-Seq es una técnica que permite conocer el ARN presente (todo o parte de este) en una muestra biológica y cuantificar cada una de las moléculas presentes.



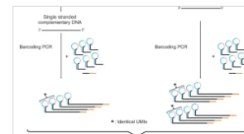
INTRODUCCIÓN AL ANÁLISIS DE ARN MEDIANTE SECUENCIACIÓN POR NGS



Digital RNA sequencing using unique molecular identifiers enables ultrasensitive RNA mutation analysis

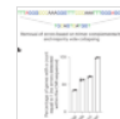
Digital RNA sequencing using unique molecular identifiers and well-performing reverse transcription allows detection of mutant allele frequencies < 0.1%.

Manuel Luna Santamaría, Daniel Andersson ... Anders Ståhlberg



Homotrimer barcodes enable accurate counting of RNA molecules during high-throughput RNA sequencing

We pinpoint PCR artifacts as the primary source of inaccurate quantification in both short- and long-read RNA sequencing, a problem that intensifies with an increase in PCR cycles in both bulk and single-cell sequencing contexts. To overcome this challenge, we engineered a novel unique molecular identifier (UMI) barcode composed of homotrimer nucleotide blocks. This design facilitates accurate quantification of RNA molecules, substantially improving molecular counting.



RNA structure profiling at single-cell resolution reveals new determinants of cell identity

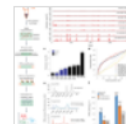
Mapping T cell landscapes in B cell lymphomas

As a major microenvironmental component in B cell lymphomas, T cells are highly relevant for current immunotherapeutic treatment strategies of such tumours. A study now provides an unprecedented multimodal insight into the composition and features of T cell subsets of the four main types of nodal B cell non-Hodgkin lymphoma.

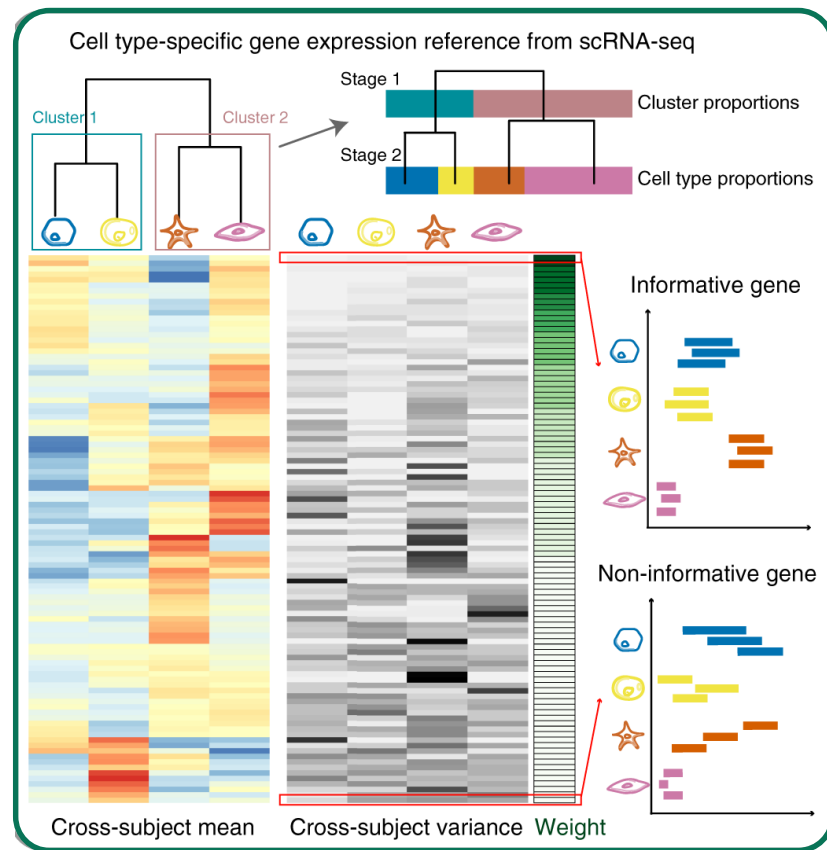
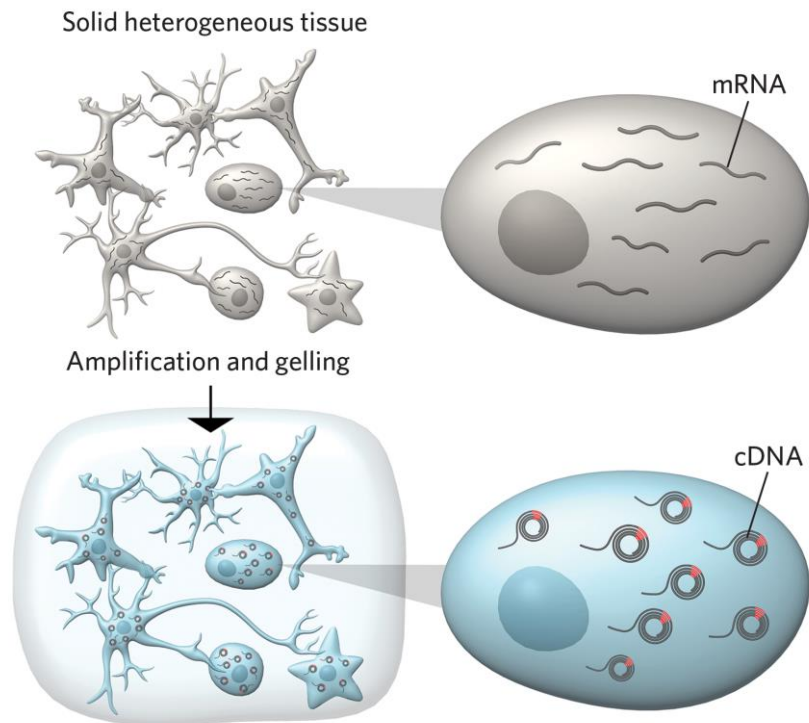
Ralf Küppers



aneous determination of
lowing new insights into

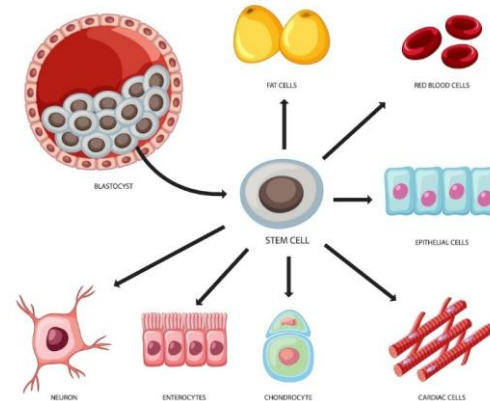
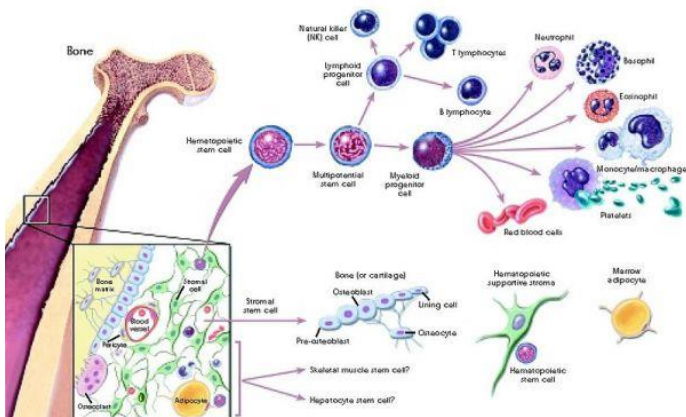
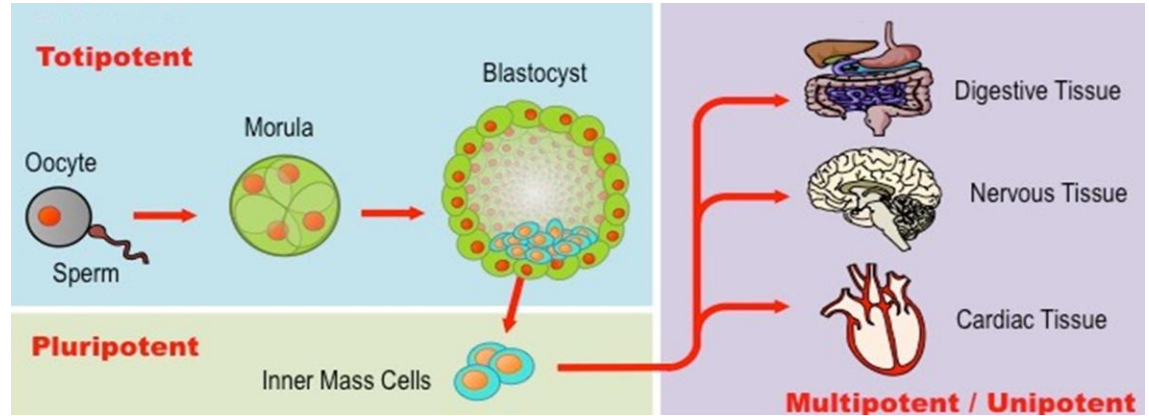


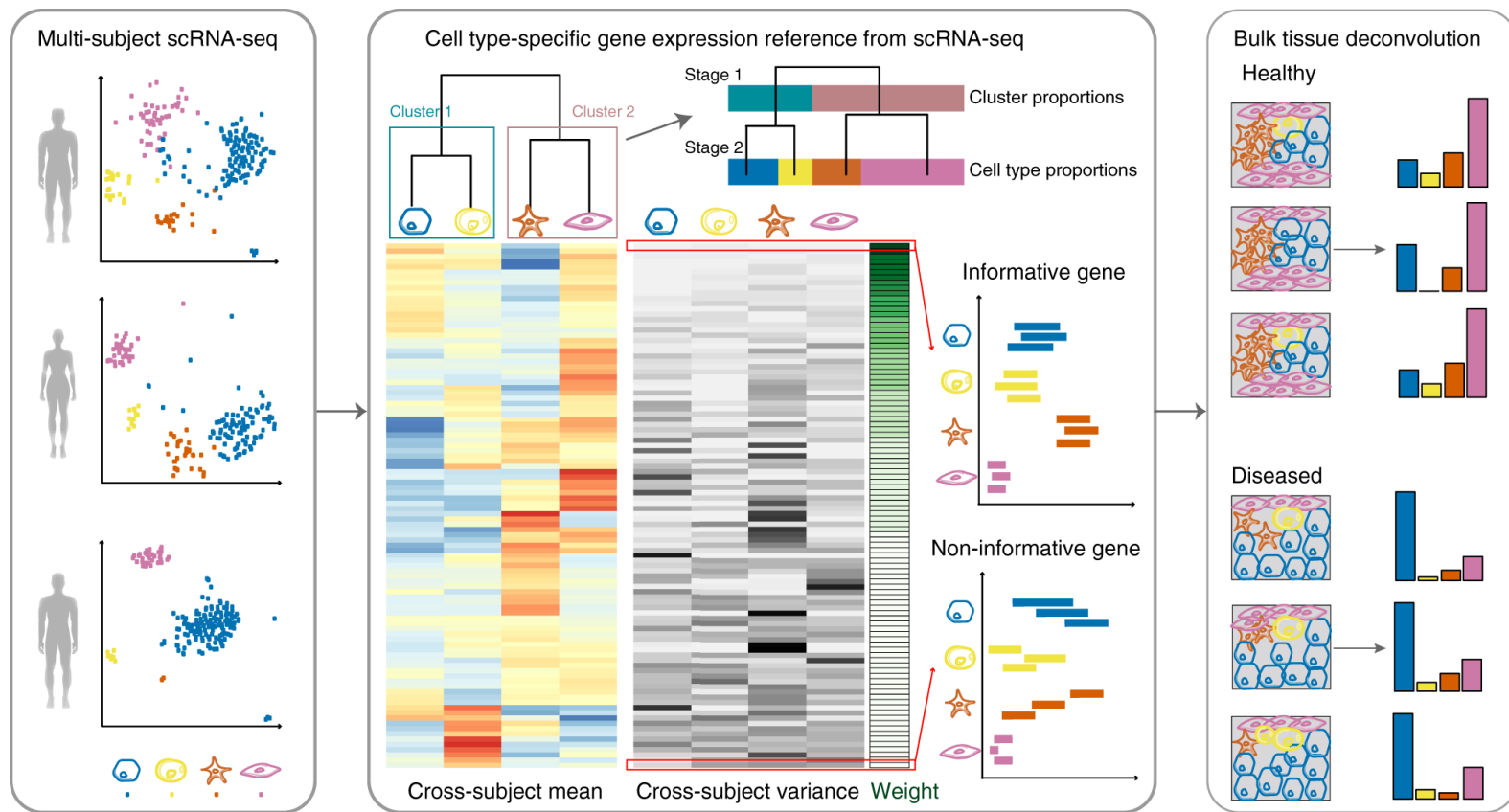
EL ARN EN EL TEJIDO



EL ARN EN EL TEJIDO

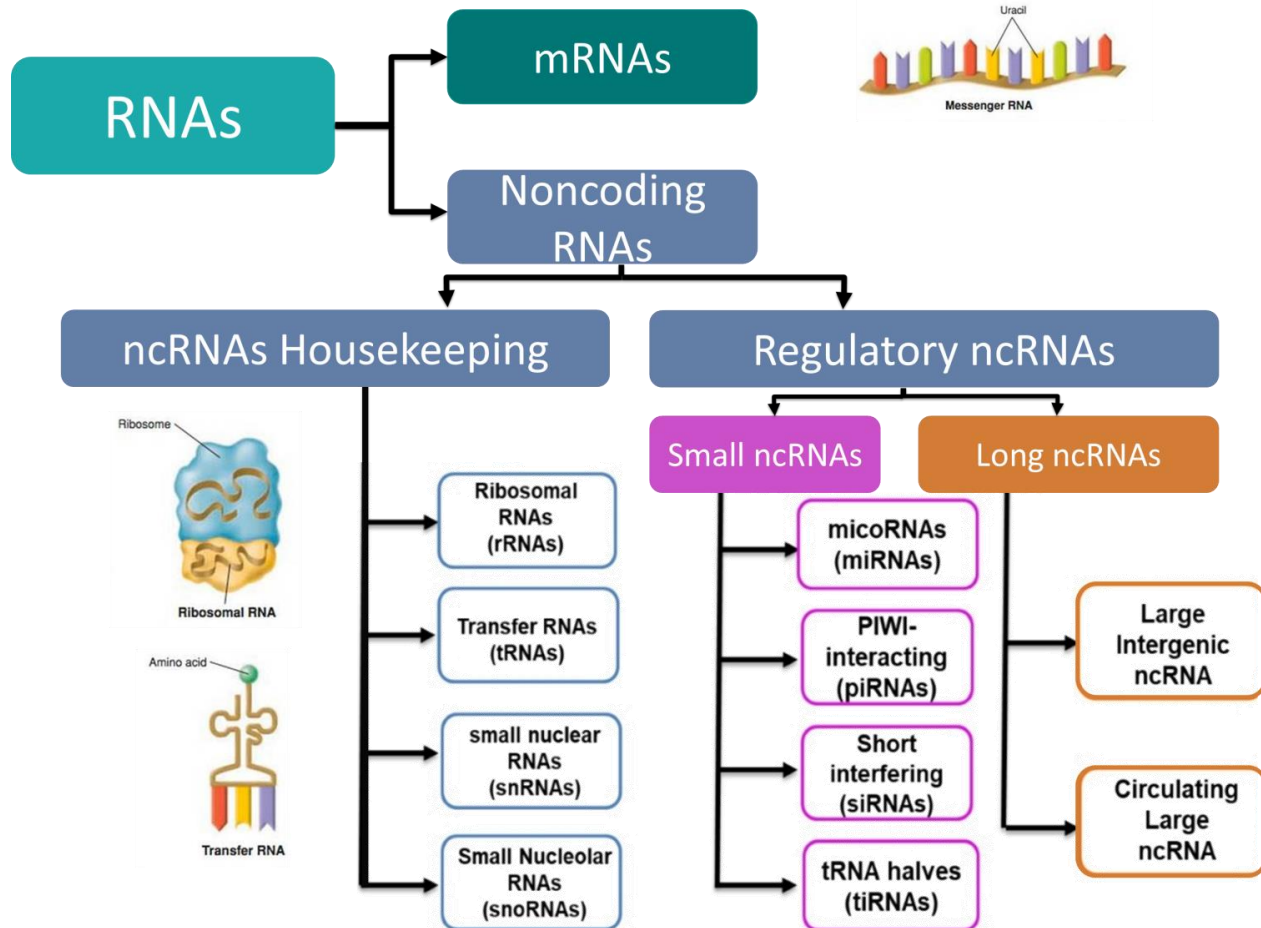
Patrones de expresión génica específicos de tejidos y células





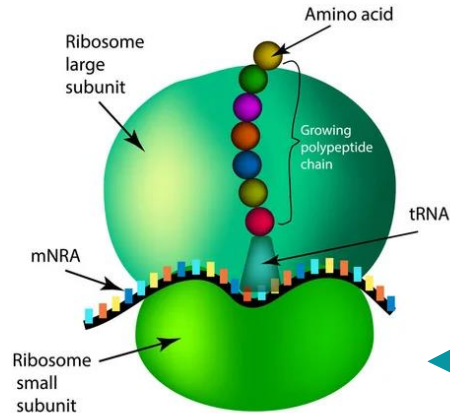
TIPOS DE ARN

¿Qué moléculas leemos cuando hacemos RNA-Seq?

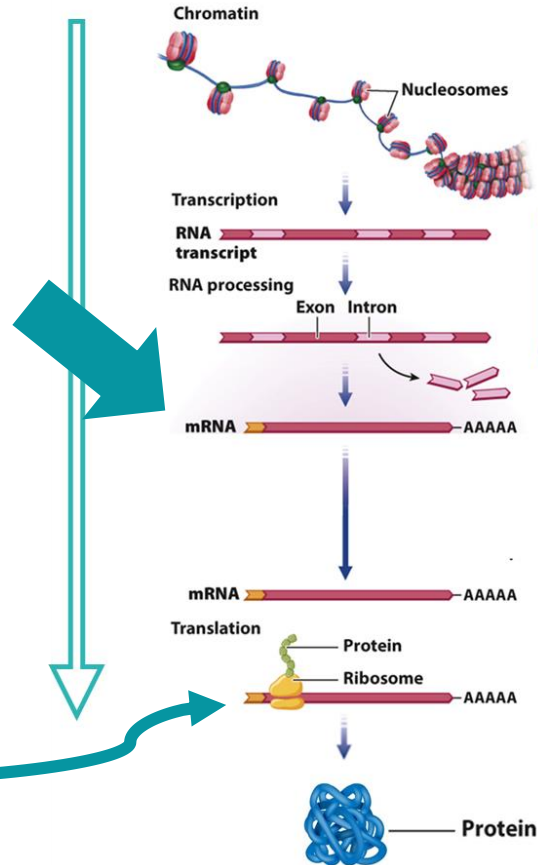


TIPOS DE ARN

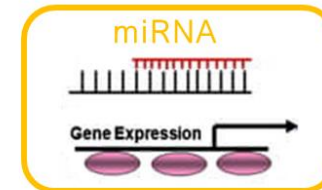
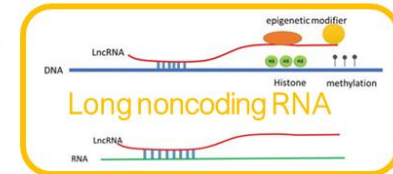
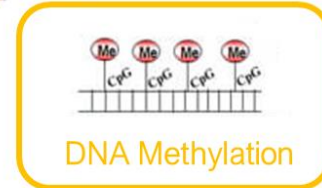
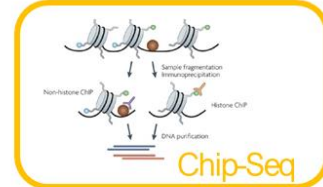
¿Cuál es su actividad?



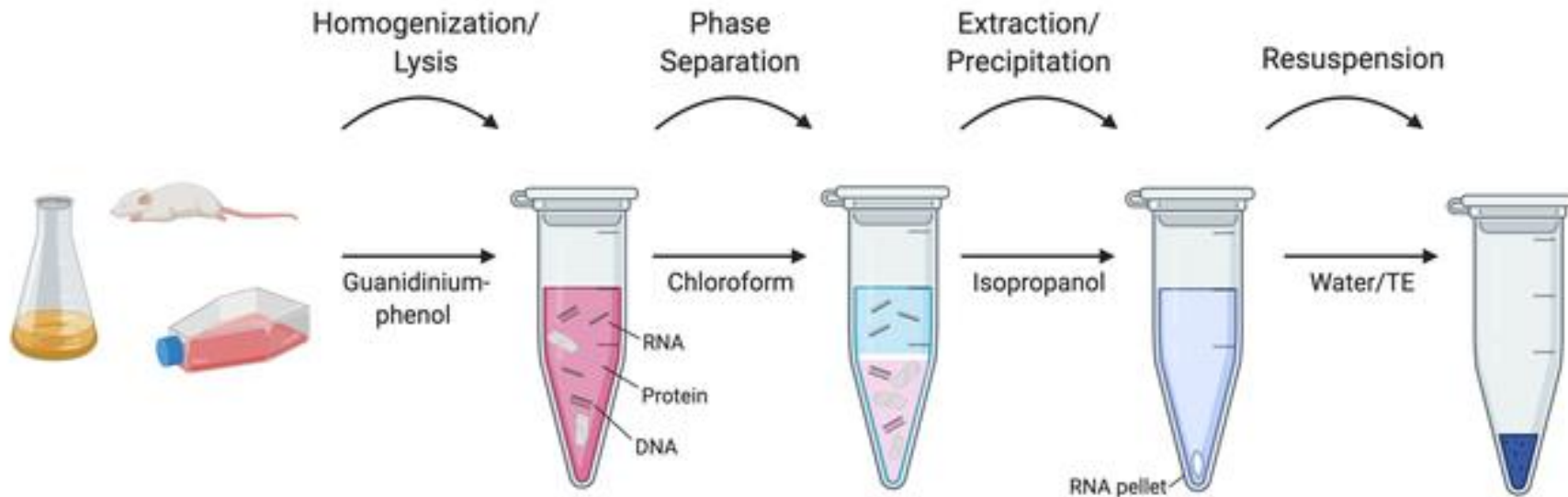
Levels of gene expression regulation



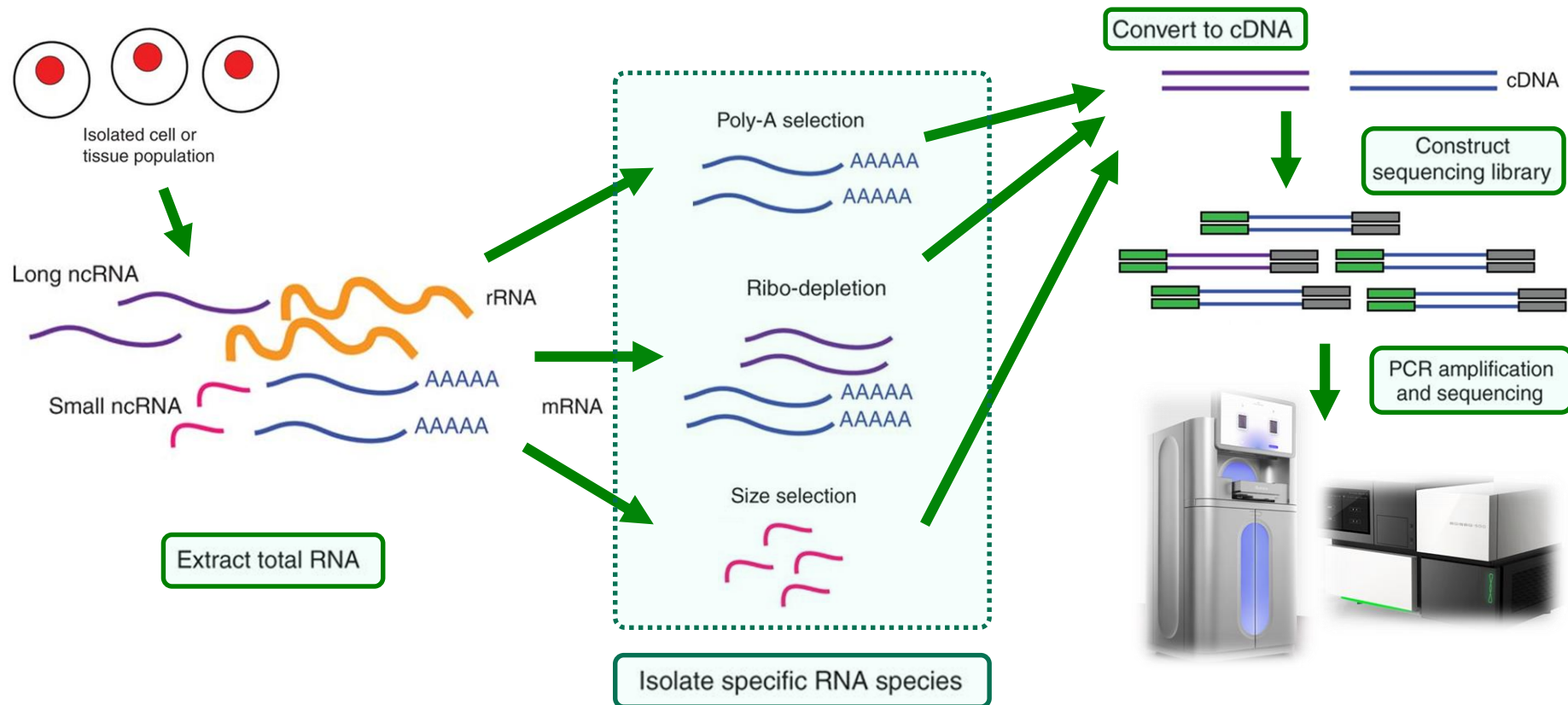
Epigenetics regulation



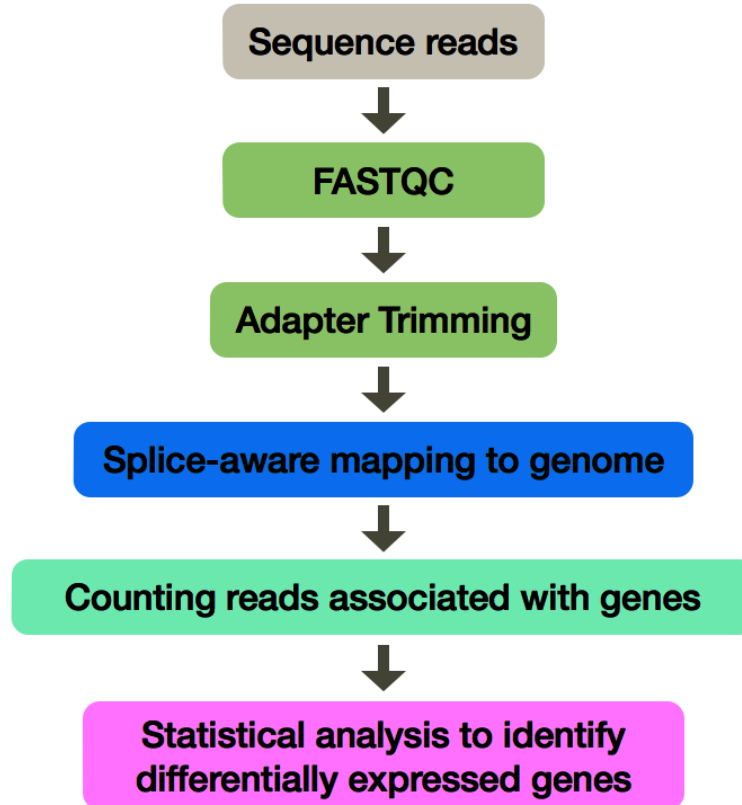
PROTOCOLO:



PROTOCOLO:



RESULTADOS RNA-Seq



RESULTADOS RNA-Seq

Generación de FastQ



Per Read Metrics

READ	CYCLES	YIELD	PROJECTED YIELD	ALIGNED (%)	ERROR RATE (%)	INTENSITY CYCLE 1	%>Q30
Read 1	75	24.28 Gbp	24.28 Gbp	0.00	0.00	8639.33	94.85
Read 2 (I)	8	2.30 Gbp	2.30 Gbp	0.00	0.00	5287.41	91.92
Read 3 (I)	8	2.29 Gbp	2.29 Gbp	0.00	0.00	4749.41	93.15
Read 4	75	24.20 Gbp	24.20 Gbp	0.00	0.00	7706.69	90.85
Non-index Reads Total	150	48.48 Gbp	48.48 Gbp	0.00	0.00	8173.01	92.85
Total	166	53.07 Gbp	53.07 Gbp	0.00	0.00	6595.71	92.82

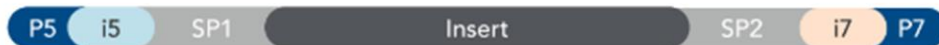
RESULTADOS RNA-Seq

Interpretación de los resultados

Single index



Unique dual index



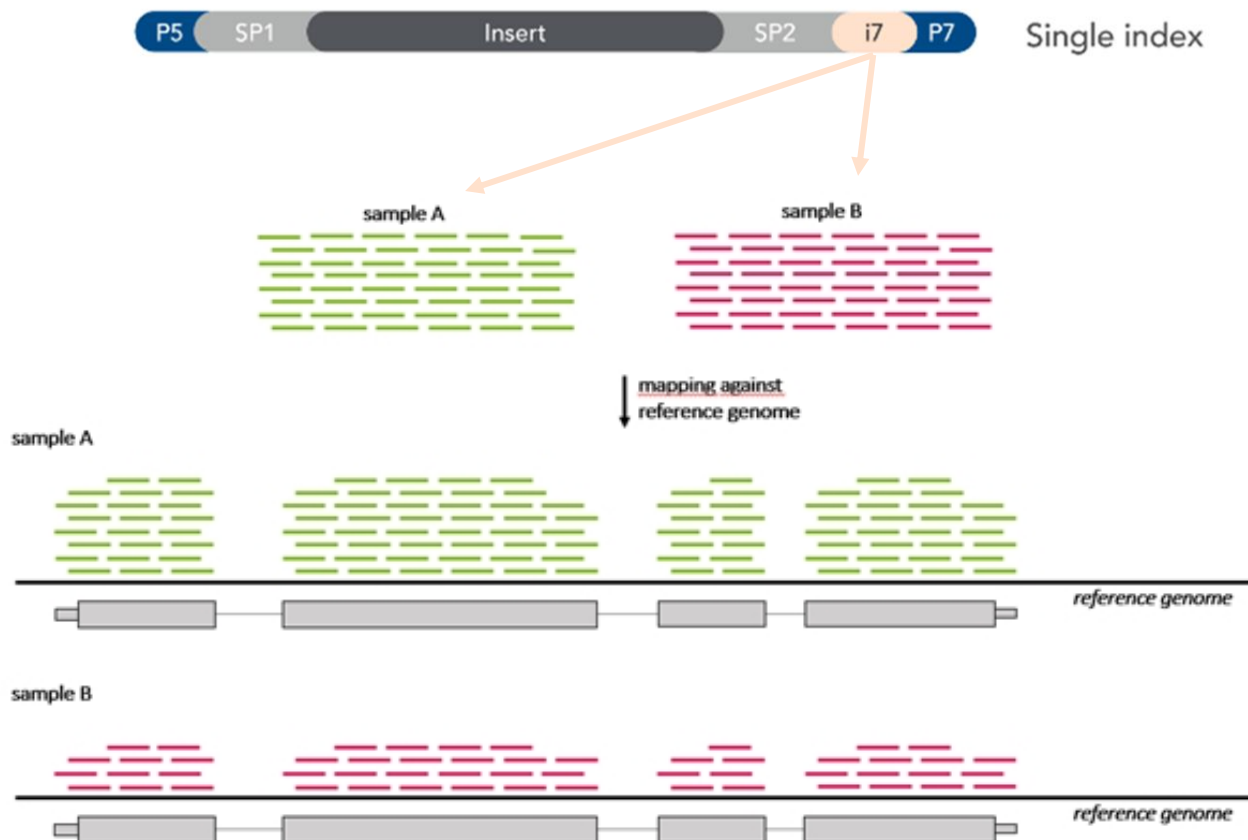
xGen UDI-UMI adapter



-  **Flow cell binding sequence:** Platform-specific sequences for library binding to instrument
-  **Sequencing primer sites:** Binding sites for general sequencing primers
-   **Sample indexes:** Short sequences specific to a given sample library
-  **Molecular index/barcode:** Short sequence used to uniquely tag each molecule in a given sample library
-  **Insert:** Target DNA or RNA fragment from a given sample library

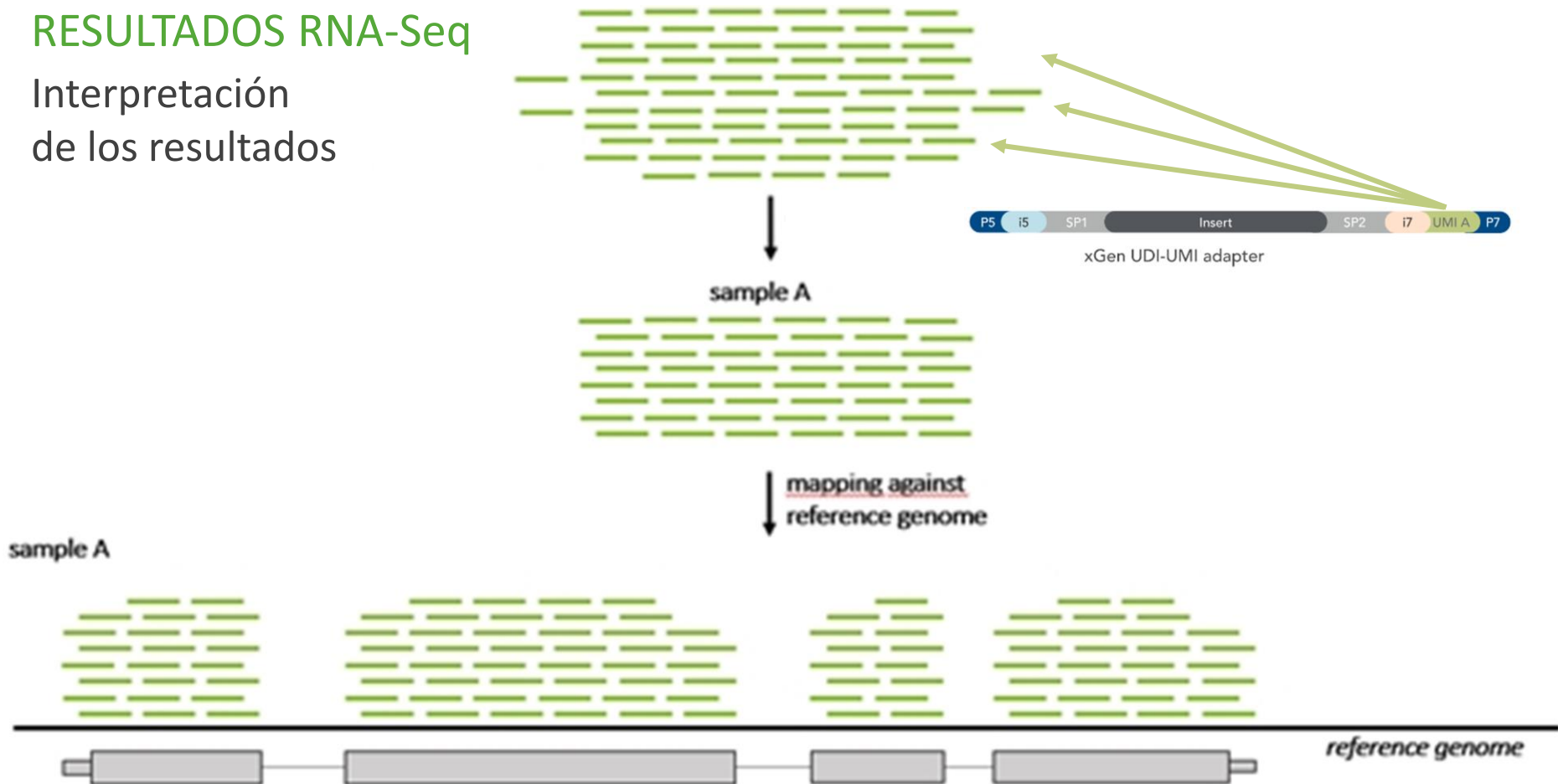
RESULTADOS RNA-Seq

Interpretación de los resultados



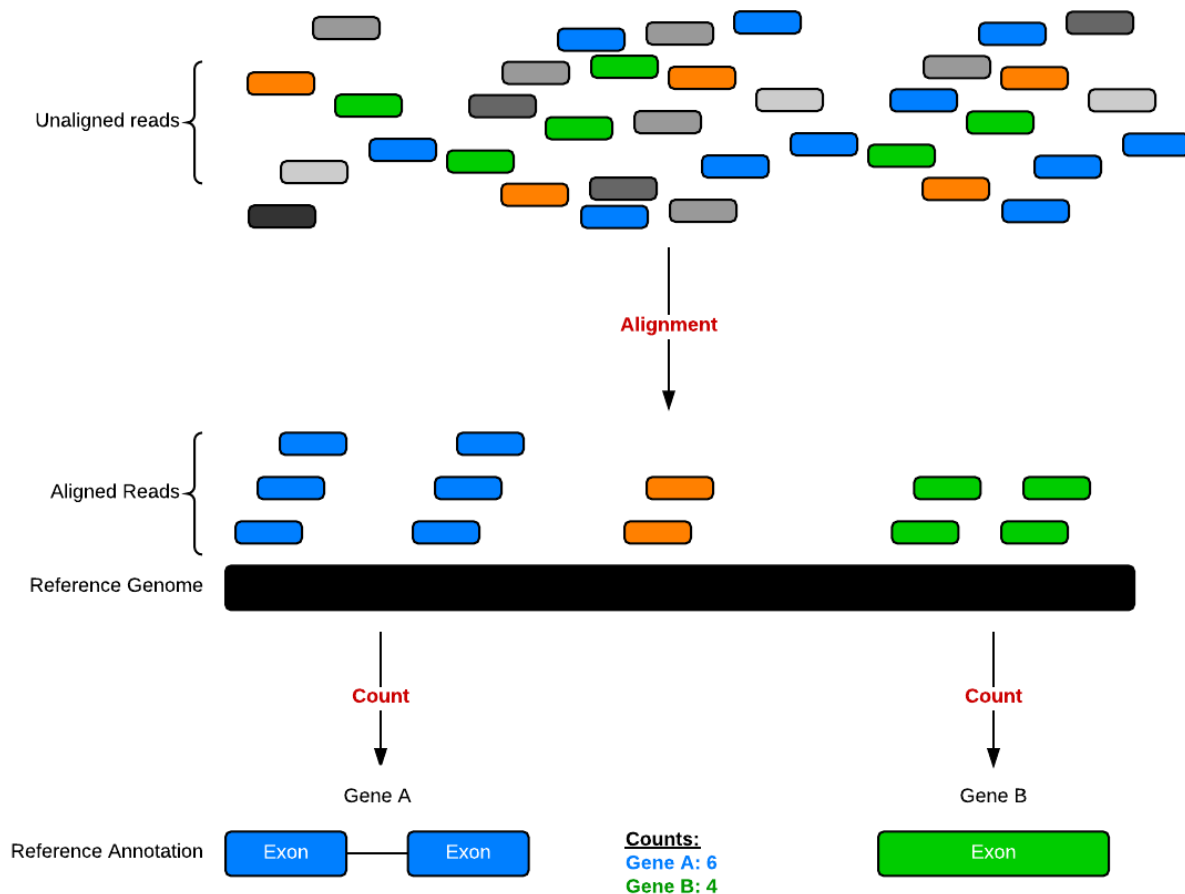
RESULTADOS RNA-Seq

Interpretación de los resultados



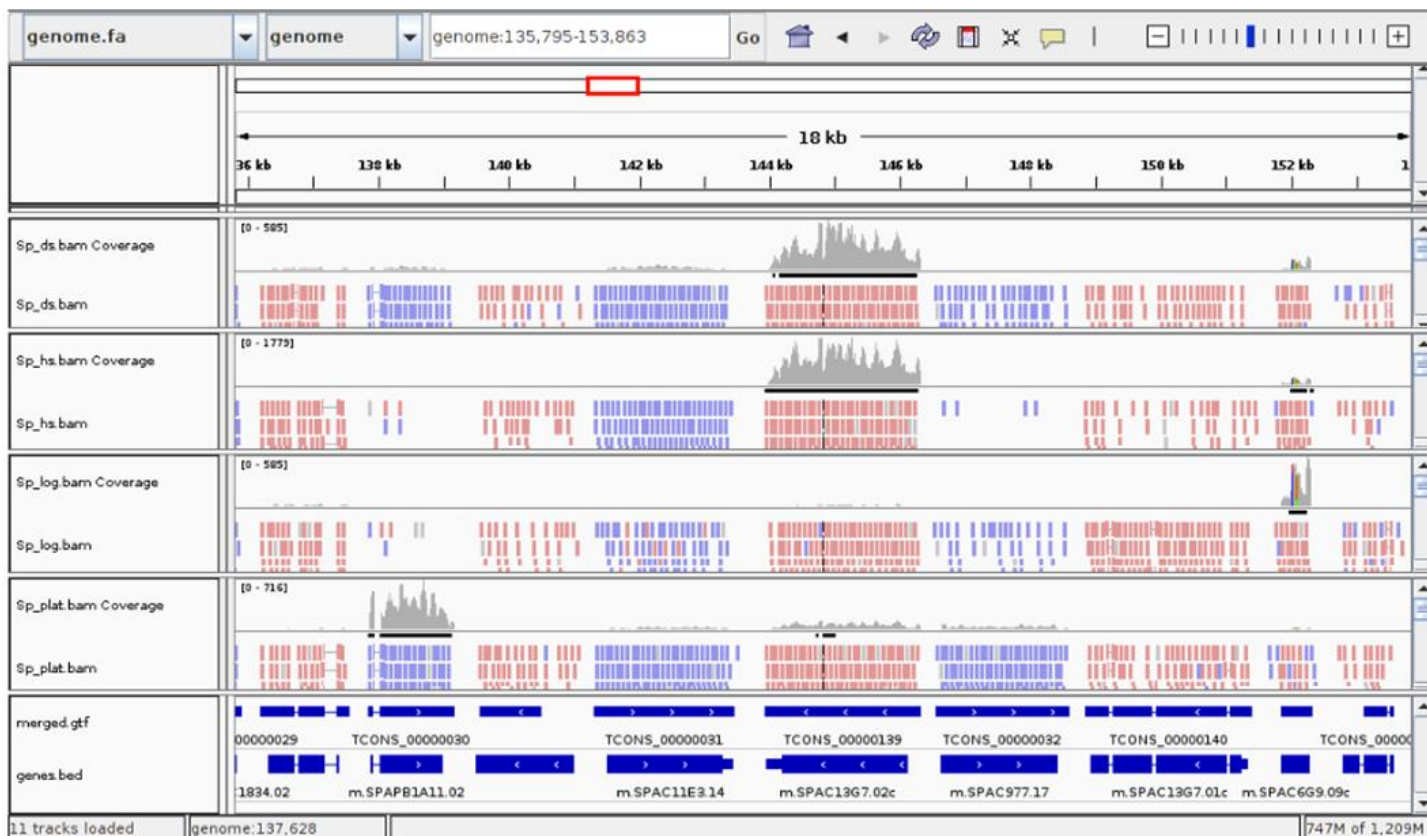
RESULTADOS RNA-Seq

Interpretación de los resultados



RESULTADOS RNA-Seq

Interpretación de los resultados



RESULTADOS RNA-Seq

Interpretación de los resultados

Gene A 600 bases

Gene B 1100 bases

Gene C 1400 bases

$$\text{RPKM} = 12 / (0.6 * 6) = 3.33$$

$$\text{RPKM} = 24 / (1.1 * 6) = 3.64$$

$$\text{RPKM} = 11 / (1.4 * 6) = 1.31$$

Sample 1



N = 6M

Sample 2



N = 8M

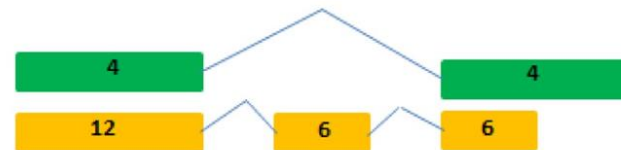
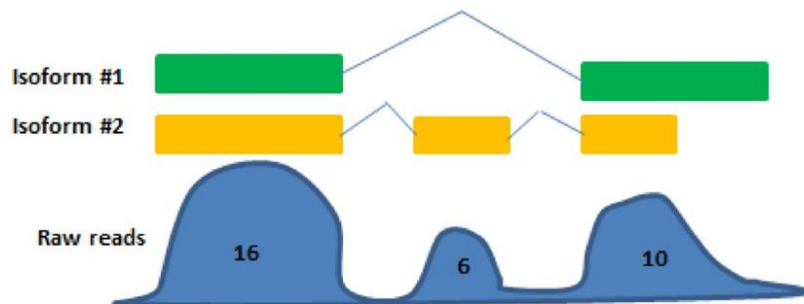
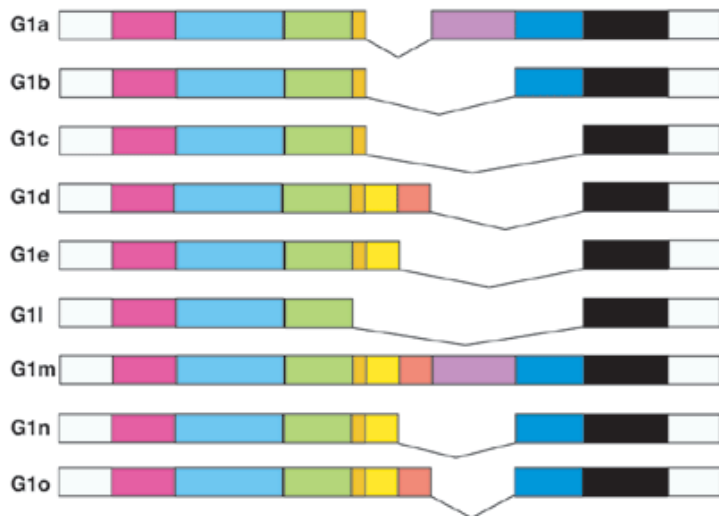
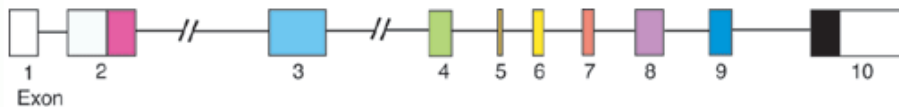
$$\text{RPKM} = 19 / (0.6 * 8) = 3.96$$

$$\text{RPKM} = 28 / (1.1 * 8) = 1.94$$

$$\text{RPKM} = 16 / (1.4 * 8) = 1.43$$

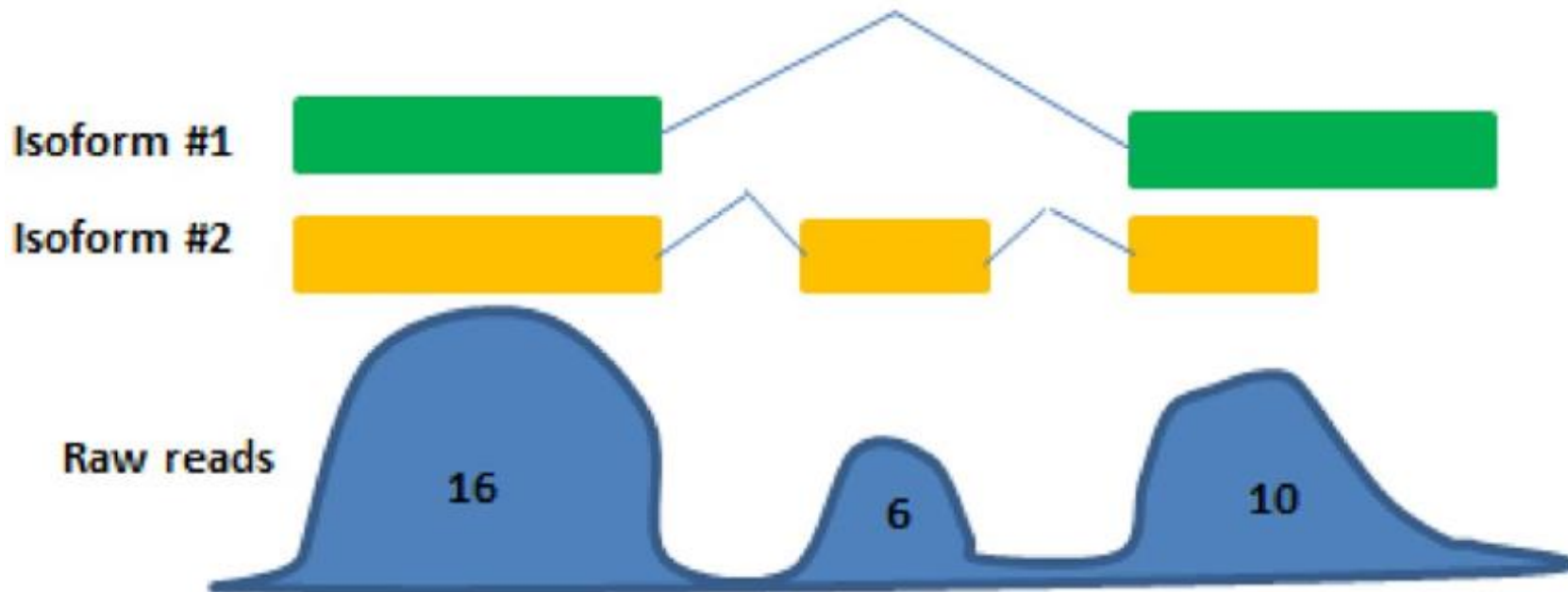
RESULTADOS RNA-Seq

Interpretación de los resultados

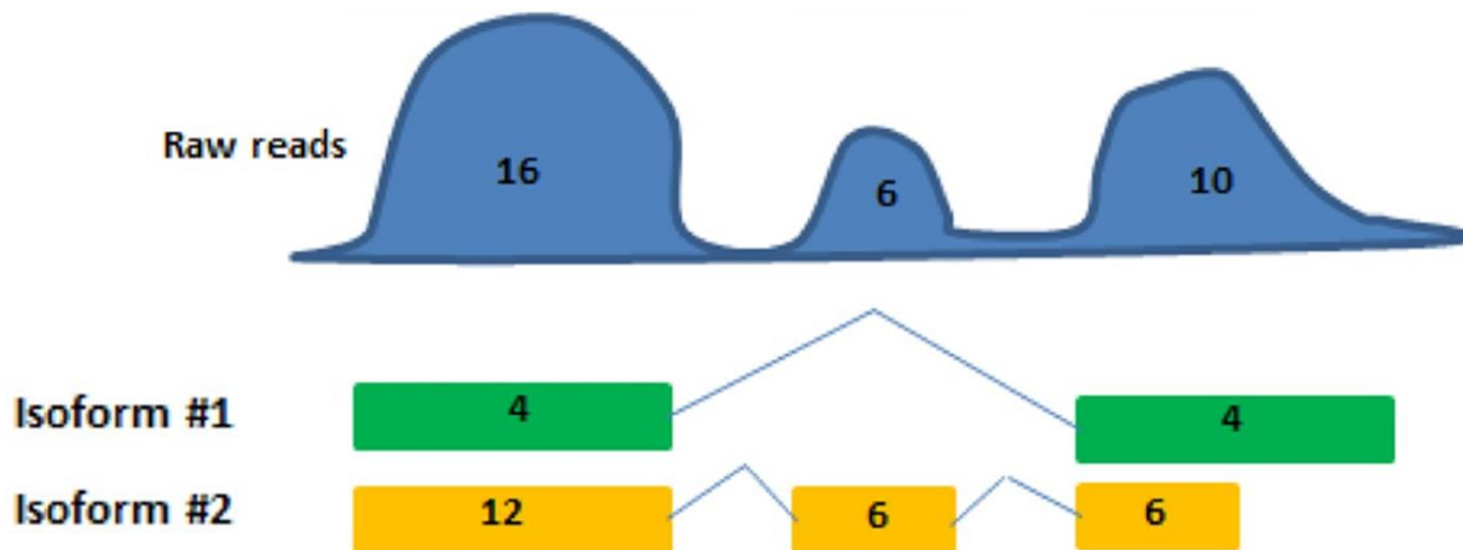


Relative isoform abundance (#1/#2): 25% / 75%
 RPKM for isoform #1 and #2: 2 and 6
 RPKM for gene (=sum of isoforms): 8 (=2+6)

RESULTADOS RNA-Seq



RESULTADOS RNA-Seq

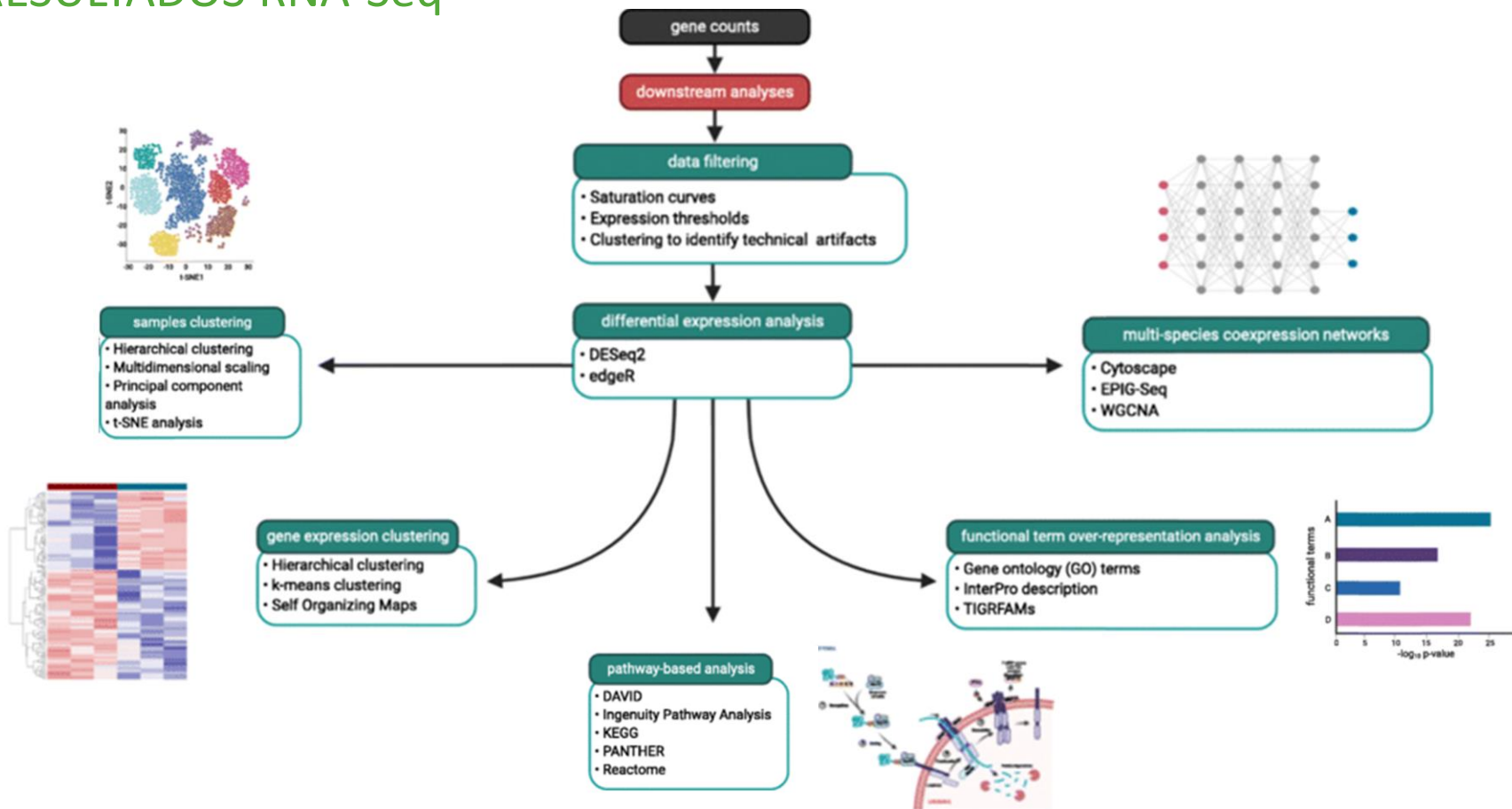


Relative isoform abundance (#1/#2): 25% / 75%

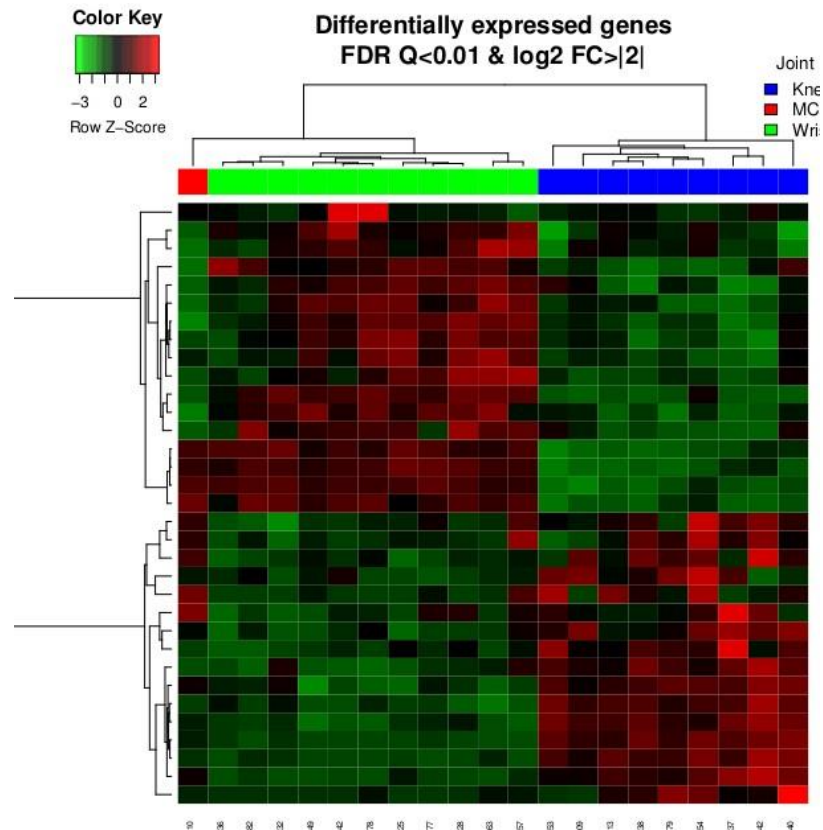
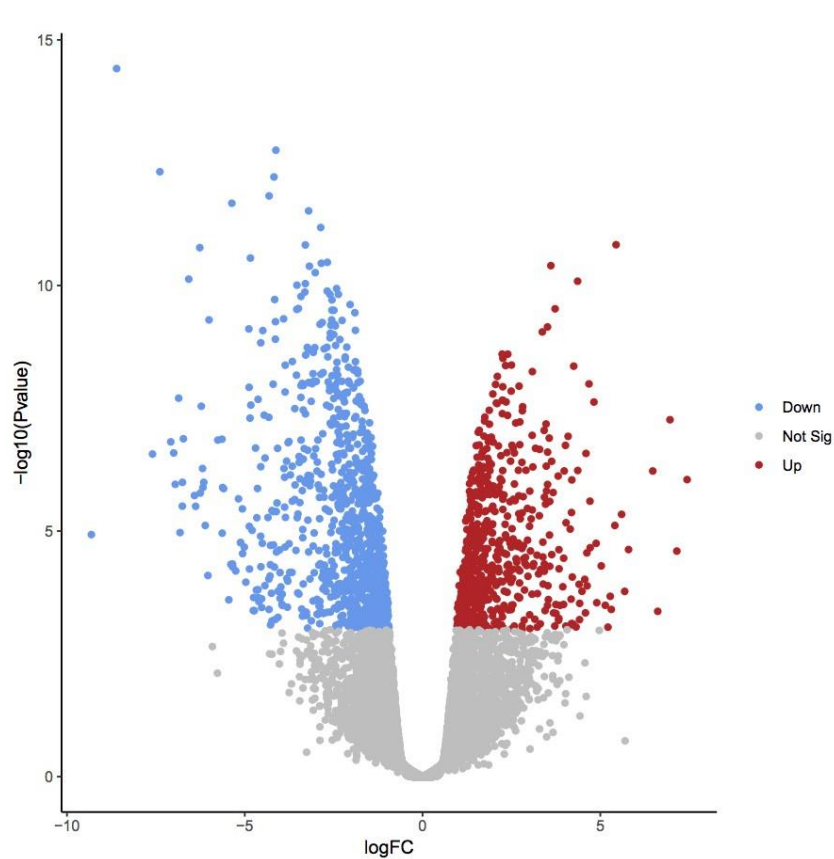
RPKM for isoform #1 and #2: 2 and 6

RPKM for gene (=sum of isoforms): 8 (=2+6)

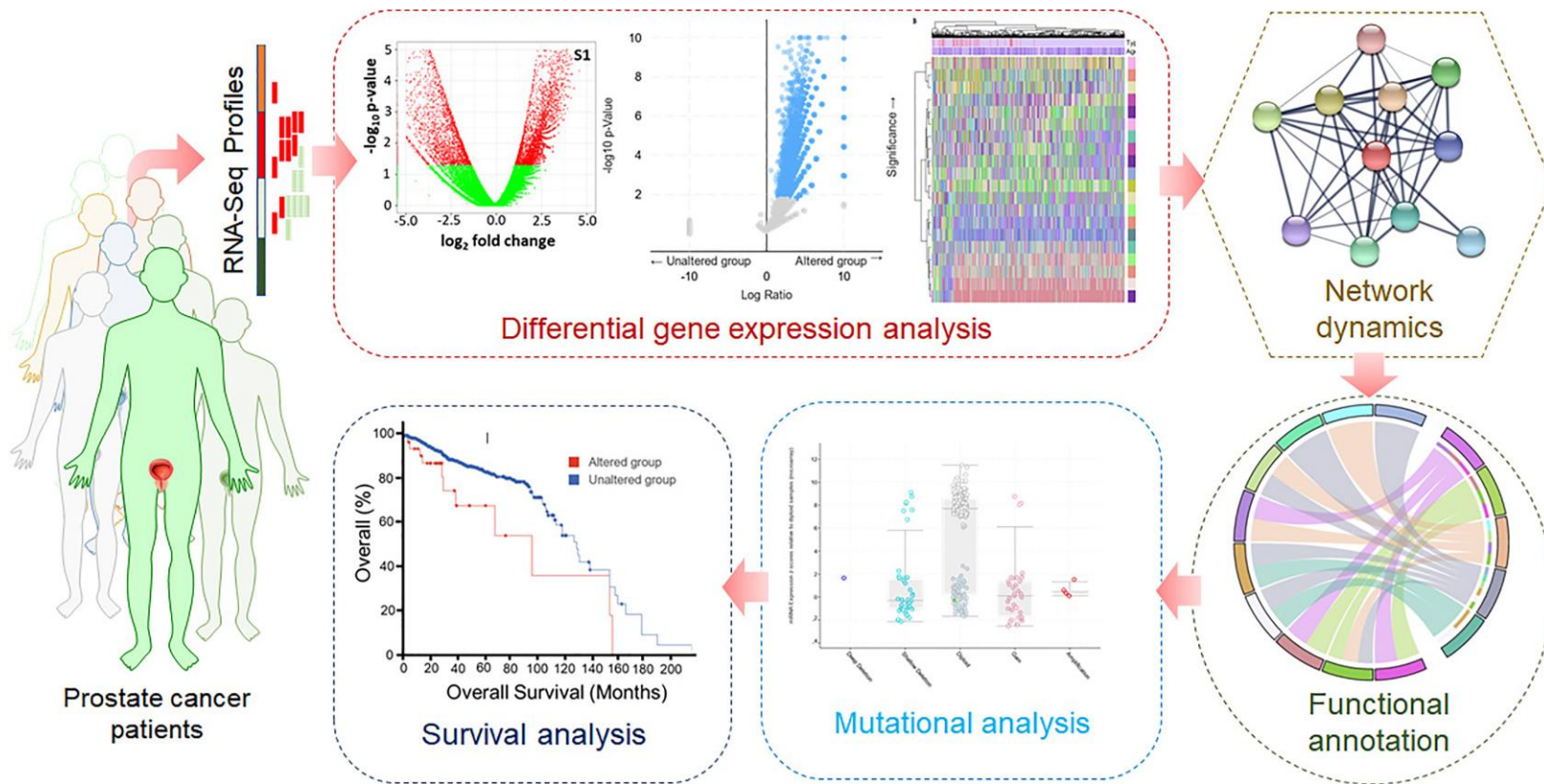
RESULTADOS RNA-Seq



RESULTADOS RNA-Seq

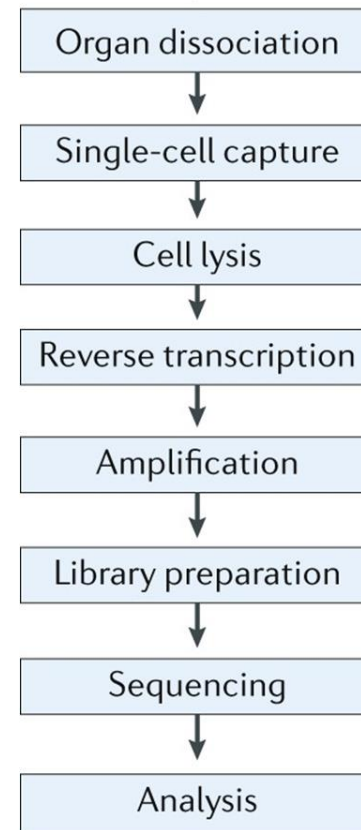
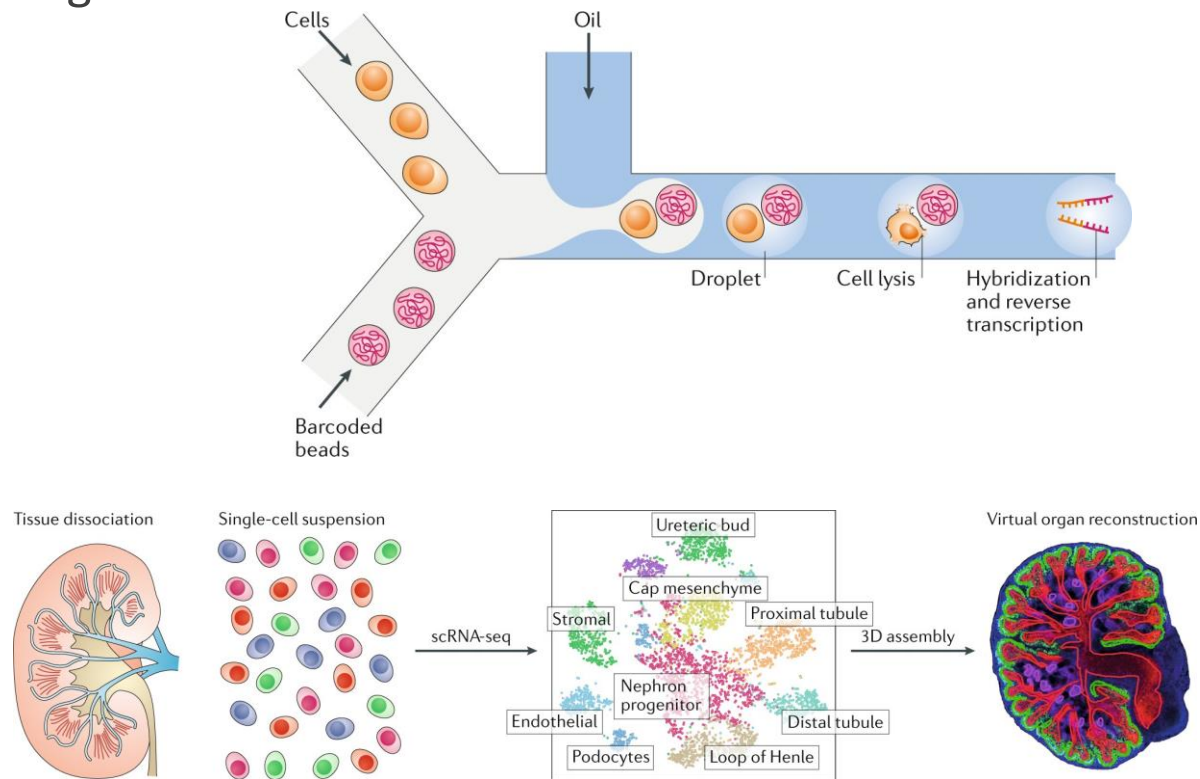


RESULTADOS RNA-Seq

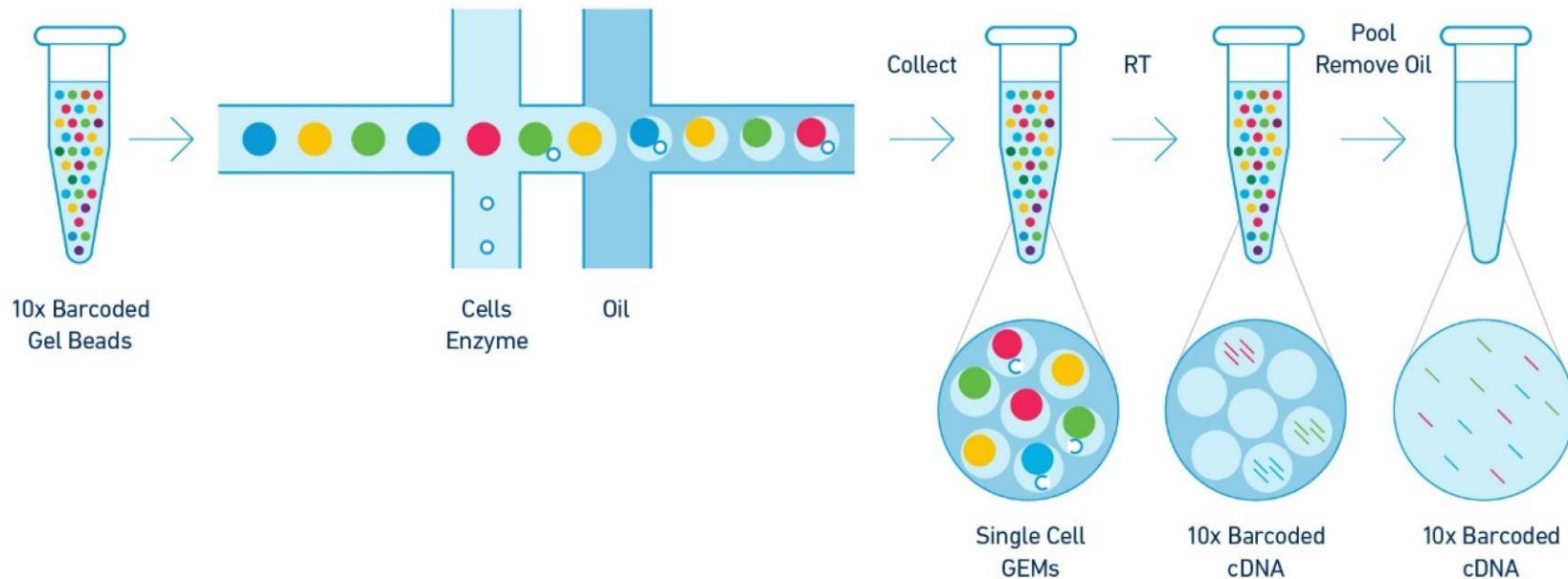


INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

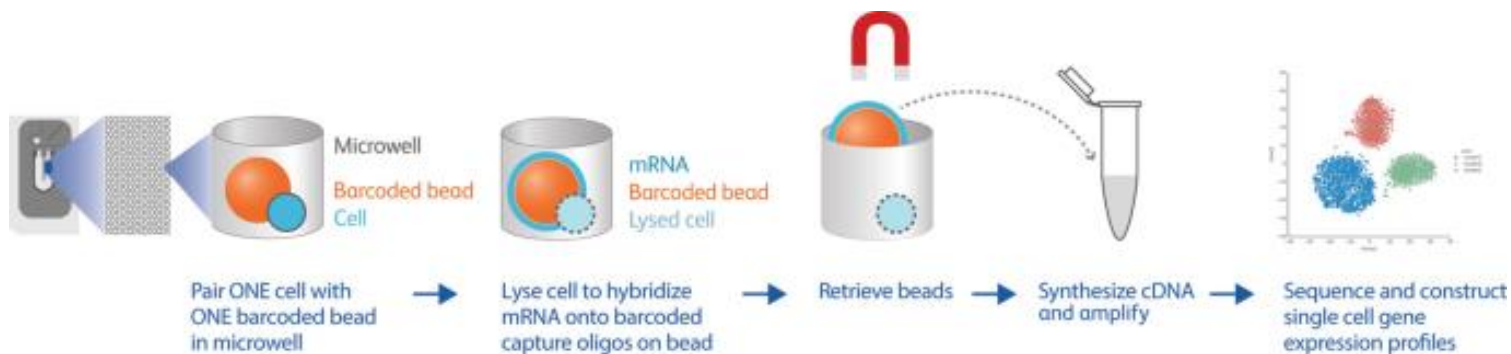
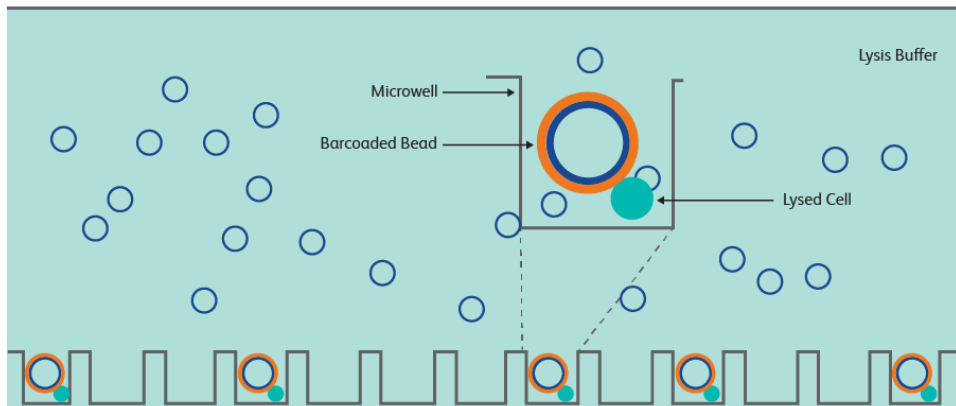
Single Cell



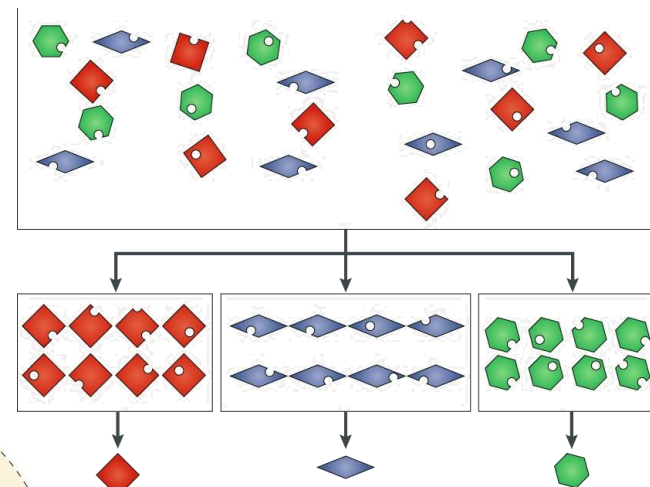
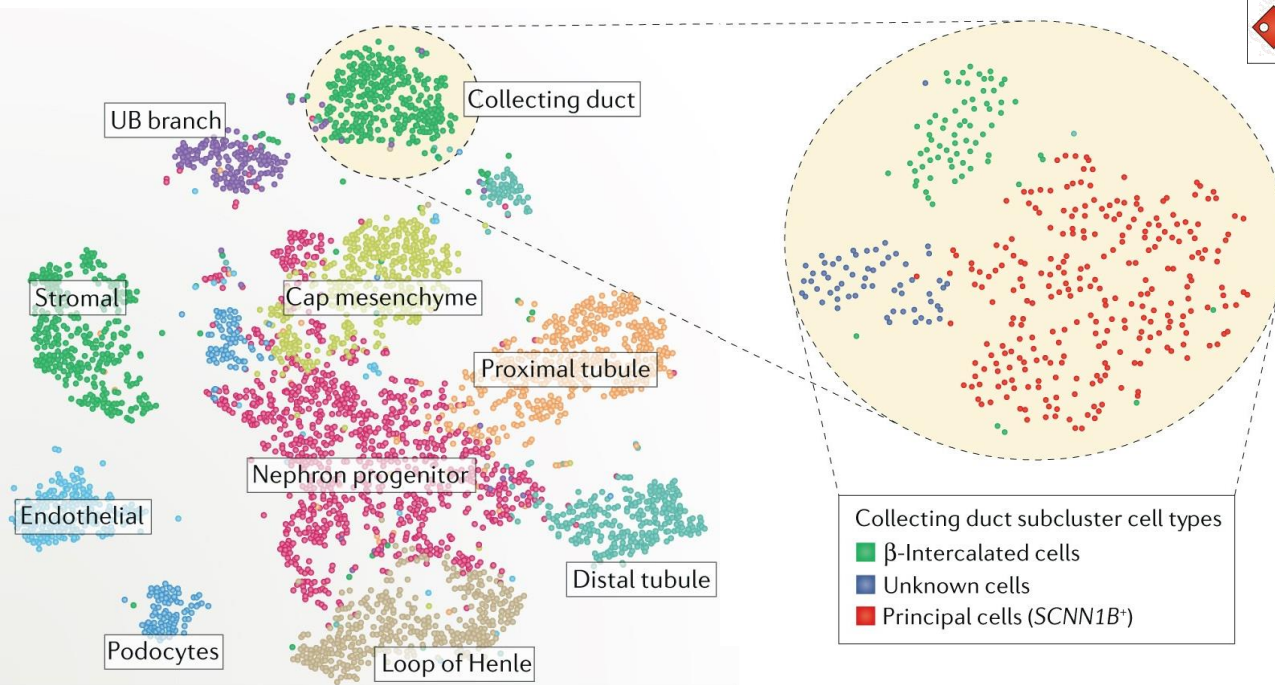
INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



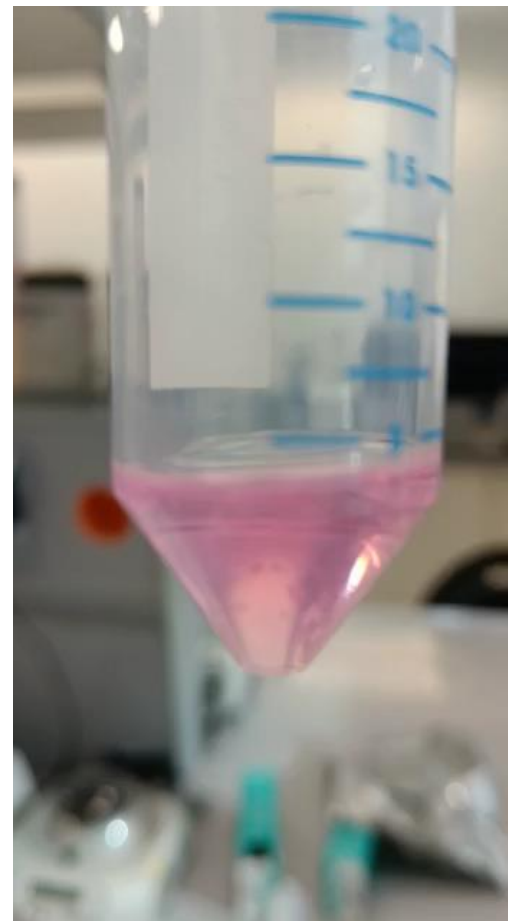
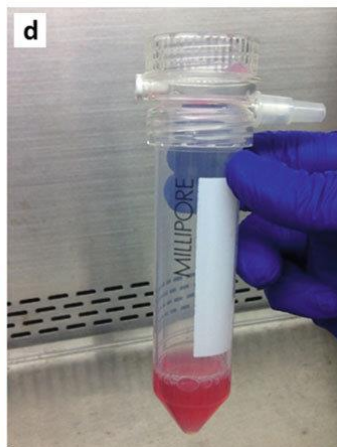
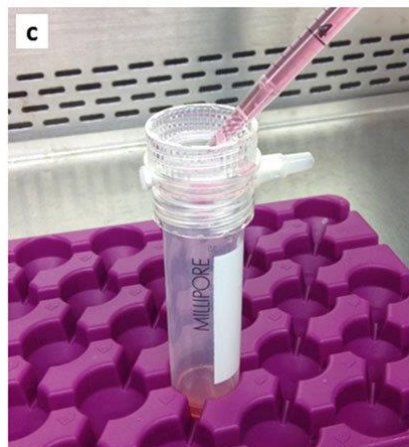
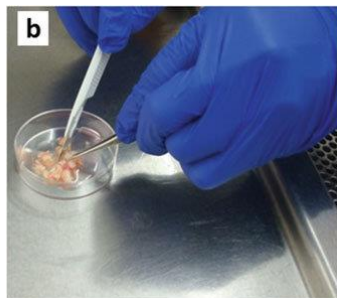
INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



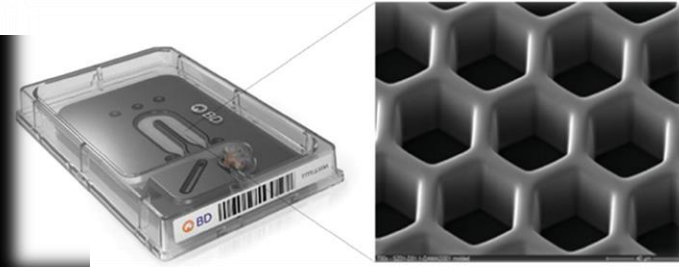
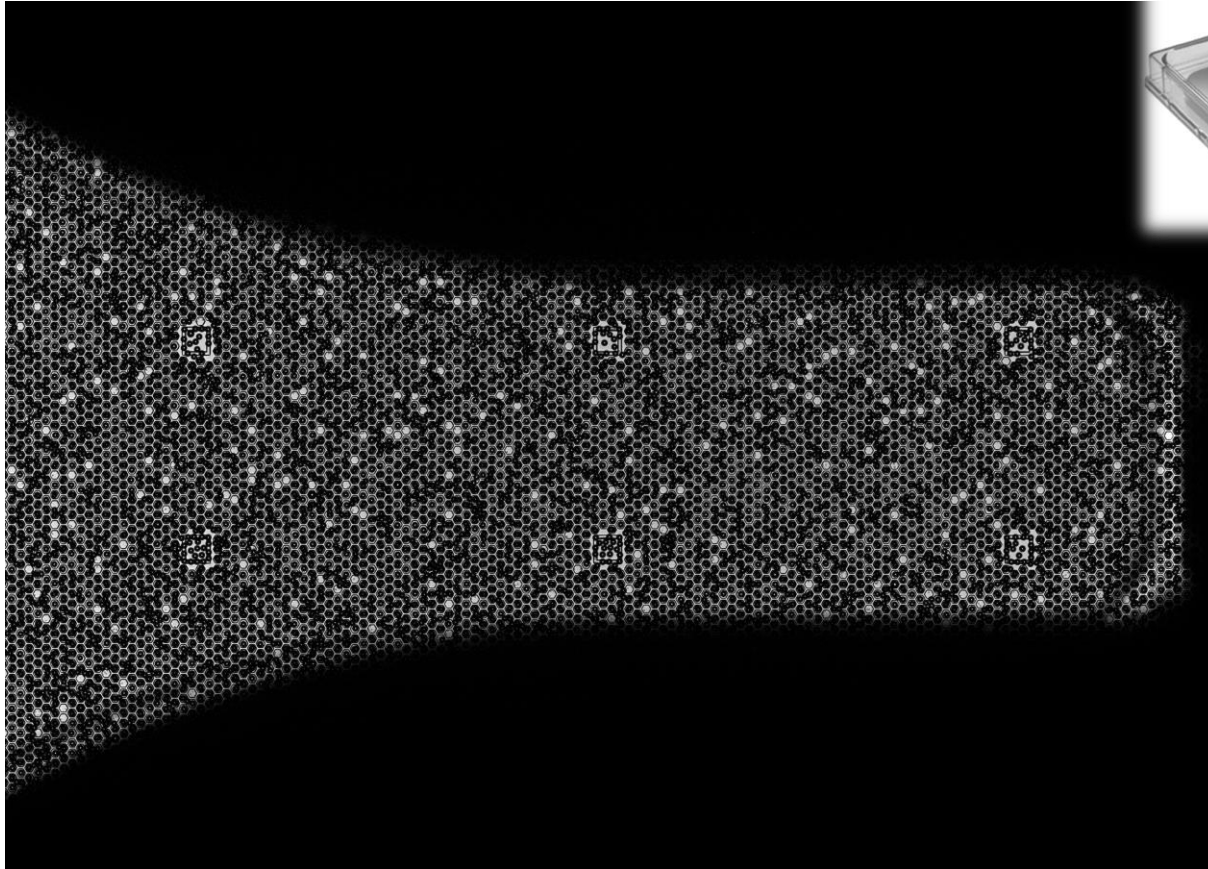
INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



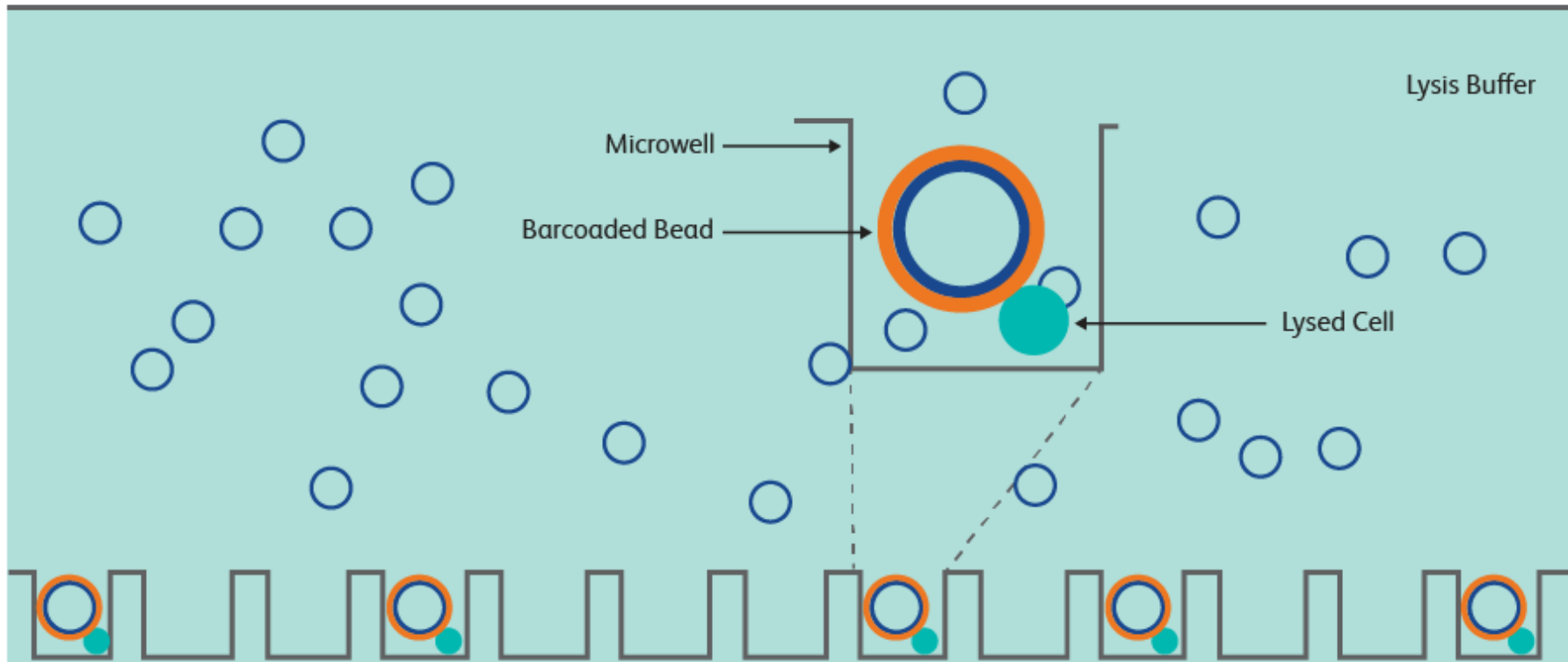
INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

RNA extraction



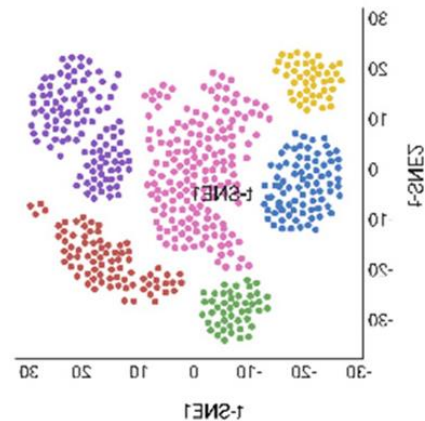
cDNA
senthesis



Single-cell
sequencing

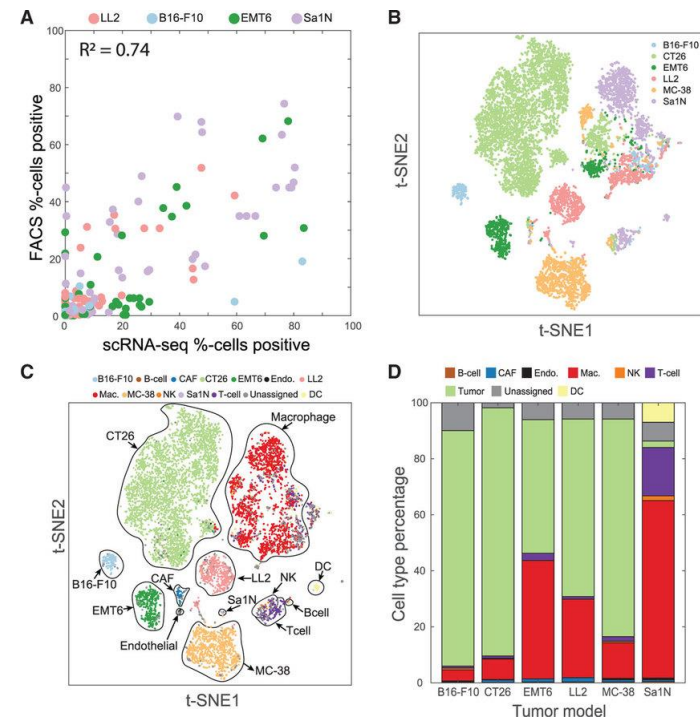
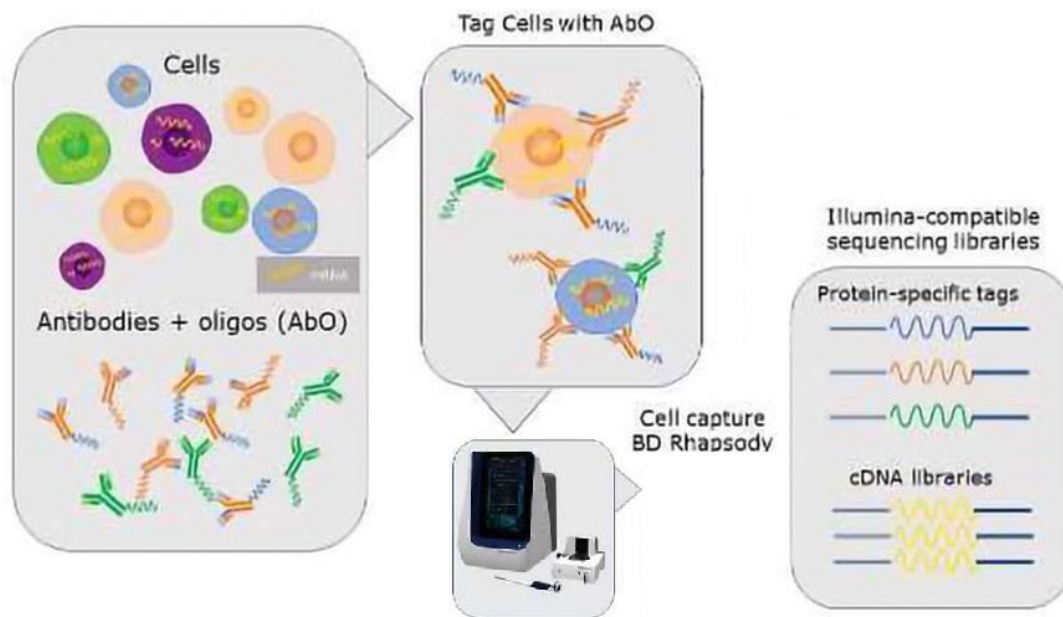


Expression
profile



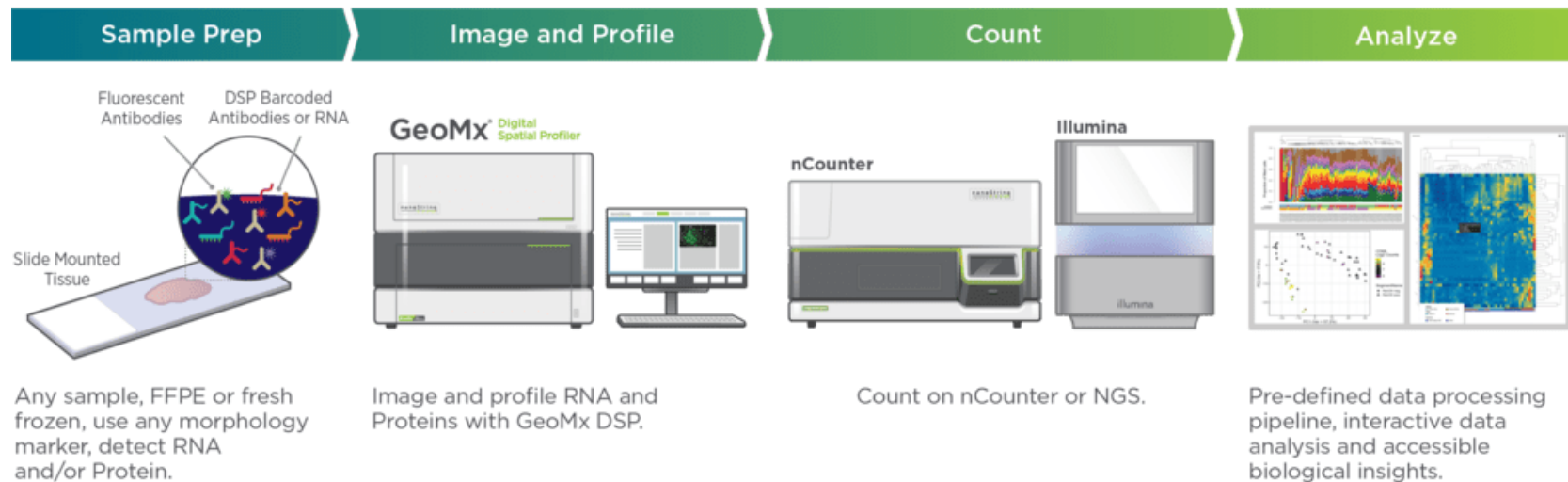
Cell type identification

INCORPORACIÓN DE NUEVAS TECNOLOGÍAS



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

Transcriptómica espacial



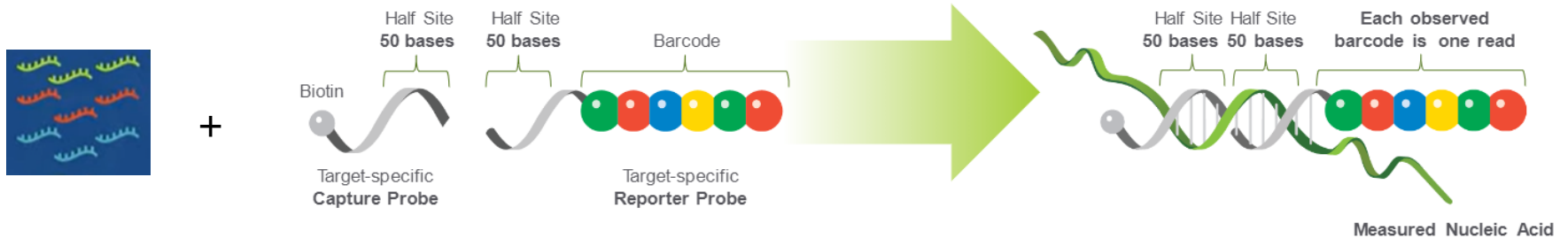
INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

Transcriptómica espacial

nCounter® Technology - Faster than qPCR. Simpler than NGS

Highly Multiplexed, Single Molecule, Digital Counting

A simple and cost-effective solution for multiplex analysis of up to 800 RNA or protein targets from precious samples.



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

Transcriptómica espacial

DAY 1: Slide staining

Manual method or automated
with Leica BOND RX/RX^m

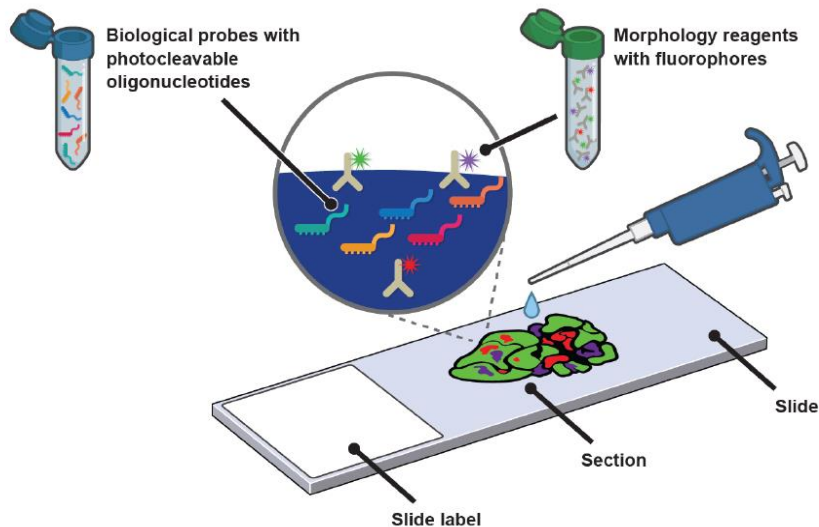


**Deparaffinize
& prepare tissue**



Probe incubation overnight

- With morphology markers
for protein samples



INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

Transcriptómica espacial

DAY 2: Process slides on GeoMx® DSP



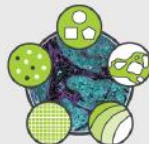
Stain & wash slides

- DNA stain, with morphology markers for RNA samples



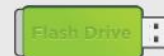
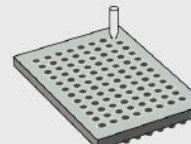
Start GeoMx run

- Load instrument
- Identify slides



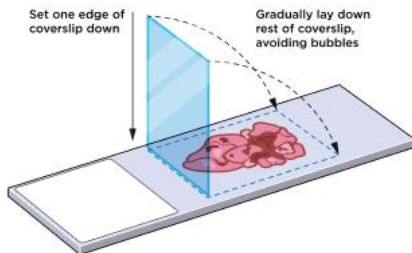
Scan & select ROIs

- Segment ROIs (optional)



ROI collection

- Collect ROI aspirates
- Finalize plate – download readout package to USB

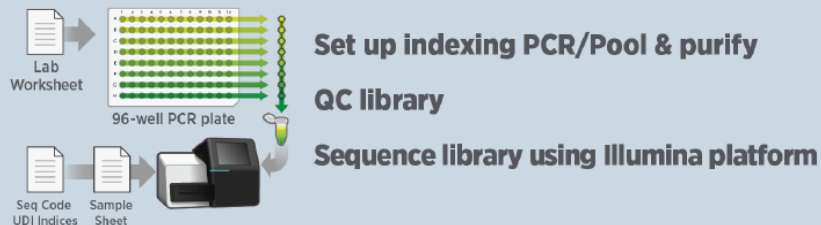


INCORPORACIÓN DE NUEVAS TECNOLOGÍAS

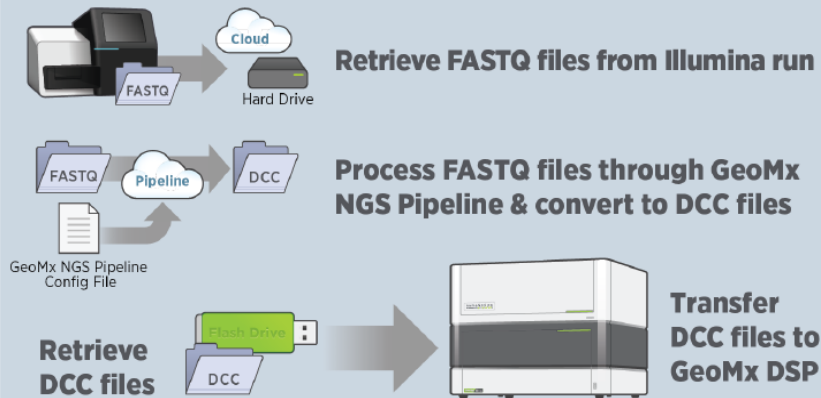
Transcriptómica espacial

NGS Readout

DAY 3: Library prep & sequencing



DAY 4: Process sequencing files to digital counts

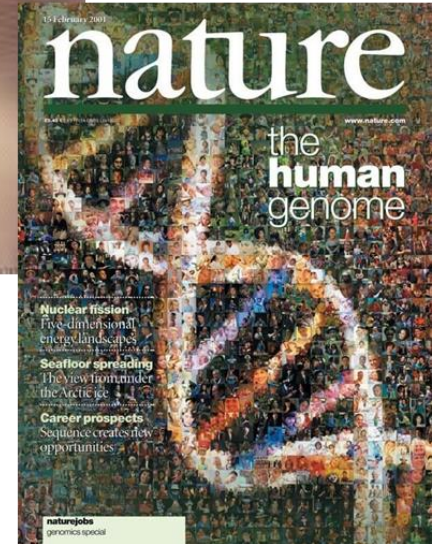


DAY 4-5: Data analysis on GeoMx DSP

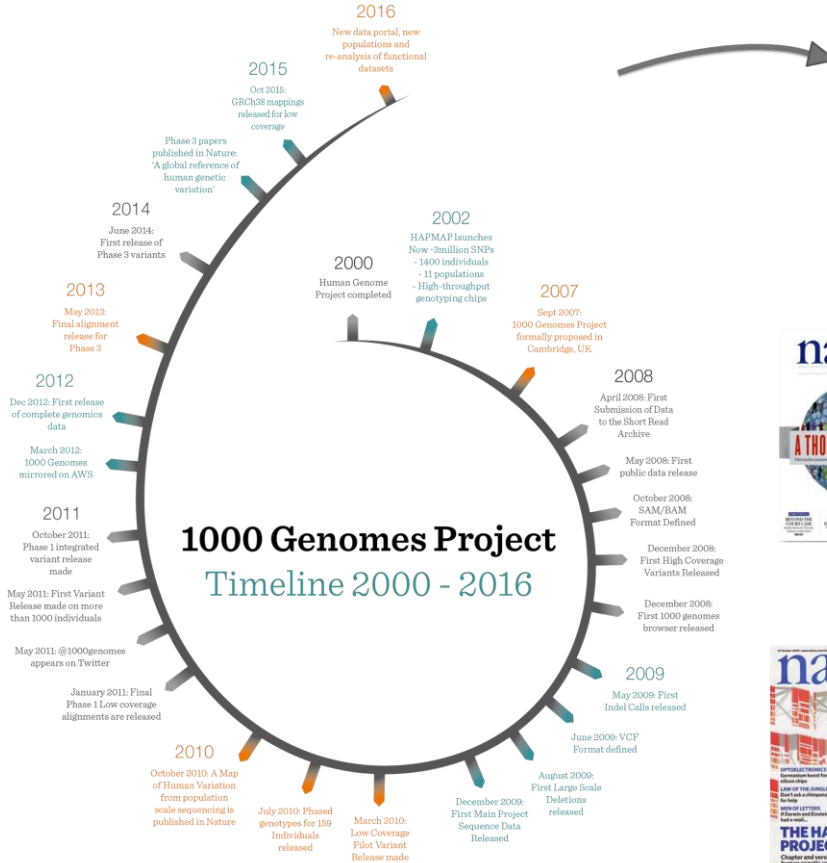


- Start new analysis**
- Run QC**
- Manage annotations**
- Data scaling & normalization**
- Statistical tests**
- Save visualizations & export (external analysis & publications)**

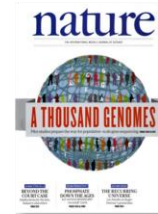
Proyectos de secuenciación ARN



Proyectos de secuenciación ARN



1000
WHOLE GENOMES SEQUENCED
AT BROAD INSTITUTE
APRIL 25, 2018

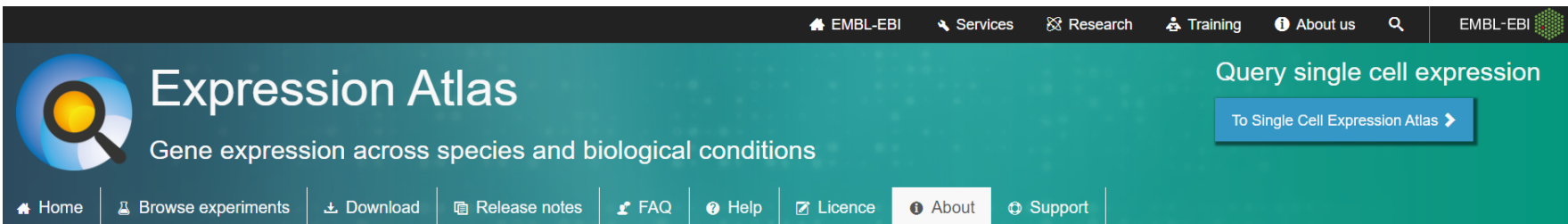


- **Piloto (20010)**
159 Individuos de 4 pop.
- **Fase I (2012):**
1000 Individuos de 14 pop.
- **Fase II (2015):**
+2500 Individuos de 26 pop.



- **Fase I (2005):**
1 SNP cada 5 kb. > 1M SNPs.
- **Fase II (2007):**
2M SNPs + 1M la fase I. 3.1 M.
- **Fase III (2009):**
+1.6 M (4.7M) 5 → 11 Pop.

Proyectos de secuenciación ARN



The screenshot shows the top section of the Expression Atlas website. It features a dark teal header with navigation links: Home, Browse experiments, Download, Release notes, FAQ, Help, Licence, About, and Support. The main banner area has a large magnifying glass icon and the text 'Expression Atlas' followed by 'Gene expression across species and biological conditions'. On the right, there is a search bar with the text 'Query single cell expression' and a button 'To Single Cell Expression Atlas >'. The EMBL-EBI logo is visible in the top right corner.

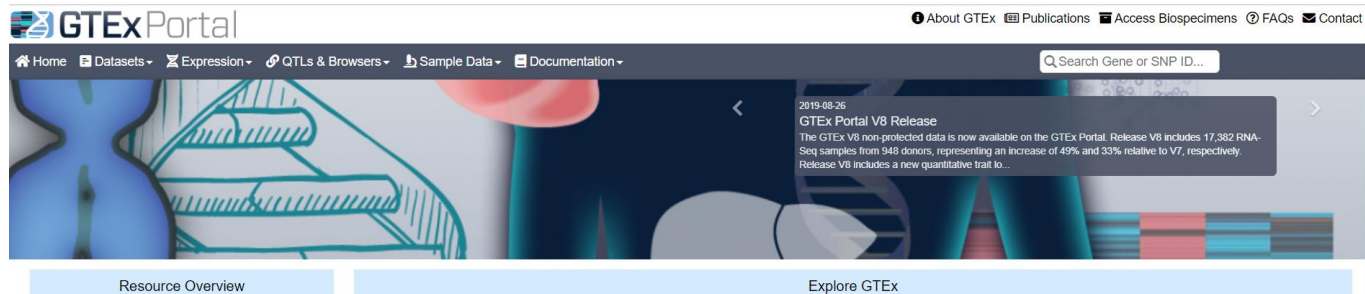
What is Expression Atlas?

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression. Our mission is to provide the scientific community with freely available information on the abundance and localisation of RNA (and proteins) across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others.

Expression Atlas aims to help answering questions such as 'where is my favourite gene expressed?' or 'how its expression changes in a disease?'



<https://gtexportal.org/>



The screenshot shows the GTEx Portal website. The header includes the GTEx logo and navigation links: Home, Datasets, Expression, QTLs & Browser, Sample Data, and Documentation. A search bar is present with the text 'Search Gene or SNP ID...'. The main content area features a large image of a DNA double helix and a news banner for '2019-08-26 GTEx Portal V8 Release'. The banner text states: 'The GTEx V8 non-protected data is now available on the GTEx Portal. Release V8 includes 17,382 RNA-Seq samples from 948 donors, representing an increase of 49% and 33% relative to V7, respectively. Release V8 includes a new quantitative trait lo...'. At the bottom, there are two buttons: 'Resource Overview' and 'Explore GTEx'.

Explore GTEx



Browse

Browse and search all data by gene

Browse and search all data by variant

[By Tissue](#)

Browse and search all data by tissue

[Histology Viewer](#)

Browse and search GTEx histology images



Single Cell

[Data Overview](#)

Learn more about available single cell data

[Multi-Gene Single Cell Query](#)

Browse and search single cell expression by gene and tissue



Expression

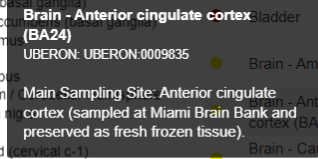
[Multi-Gene Query](#)

Browse and search expression by gene and tissue

[Transcript Browser](#)

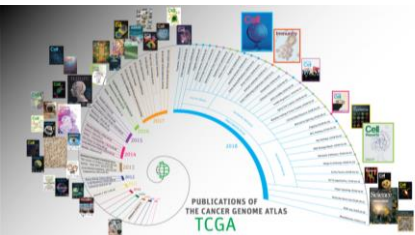
Visualize transcript expression and isoform structures

This page provides a visual representation of the biospecimen source sites (BSSs) for the collection of tissue from postmortem/organ procurement cases for the Genotype-Tissue Expression (GTEx) project.



Tissue	UBERON	Main Sampling Site
Adipose - Subcutaneous	UBERON:0002190	Subcutaneous tissue beneath the leg's skin sample.
Adipose - Visceral (Omentum)	UBERON:0010414	Adipose tissue on the large fold of parietal peritoneum that hangs down from the greater curvature of the stomach, passing in front of the small intestines.
Adrenal Gland	UBERON:0002369	Left, followed by the right if necessary for sufficient aliquots.
Artery - Aorta	UBERON:0001496	Ascending aorta or other thoracic regions (nonatherosclerotic).
Artery - Coronary	UBERON:0001621	Left and right, noncalcific regions only.
Artery - Tibial	UBERON:0007610	Left Tibial. Peripheral tibial artery from gastrocnemius region
Urinary bladder	UBERON:0001255	Central posterior urinary bladder, trimming from the outer wall if necessary.
Brain - Amygdala	UBERON:0001876	Amygdala (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Anterior cingulate cortex (BA24)	UBERON:0009835	Anterior cingulate cortex (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Caudate (basal ganglia)	UBERON:0001873	Caudate (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Cerebellar Hemisphere	UBERON:0002037	Cerebellar hemisphere (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Cerebellum	UBERON:0002037	Right cerebellum (sampled at donor collection site and preserved in PAXgene fixative).
Brain - Cortex	UBERON:0001870	Right cerebral frontal pole cortex (sampled at donor collection site and preserved in PAXgene fixative).
Brain - Frontal Cortex (BA9)	UBERON:0009834	Right cerebral frontal pole cortex (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Hippocampus	UBERON:0001954	Hippocampus (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Hypothalamus	UBERON:0001898	Hypothalamus (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Nucleus accumbens (basal ganglia)	UBERON:0001882	Nucleus accumbens basal ganglia (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Putamen (basal ganglia)	UBERON:0001874	Putamen basal ganglia (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Spinal cord (cervical c-1)	UBERON:0006469	Spinal cord (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
		Substantia nigra (sampled at Miami Brain Bank and preserved as

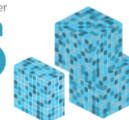
Proyectos de secuenciación ARN



NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over
2.5
PETABYTES
of data



TCGA data describes
33
DIFFERENT
TUMOR TYPES

...including
10
RARE
CANCERS

To put this into perspective, 1 petabyte of data is equal to

212,000
DVDs



...based on paired tumor and normal tissue sets collected from

11,000
PATIENTS

...using
7
DIFFERENT
DATA TYPES



Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary

Data Release 28.0 - February 02, 2021

PROJECTS



68

PRIMARY SITES



67

CASES



84.591

FILES



615.761

GENES

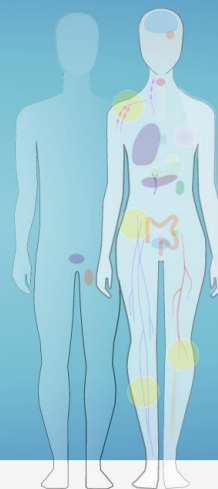


23.535

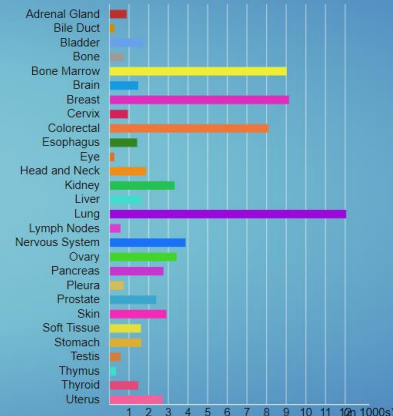
MUTATIONS



3.461.256



Cases by Major Primary Site



<https://portal.gdc.cancer.gov/>

Proyectos de secuenciación ARN

<http://www.tcgaportal.org/>

Proyectos de secuenciación ARN

<https://gtexportal.org/home/>

rs4880, SOD2, Liver

INTELIGENCIA ARTIFICIAL y RNA-Seq

<https://ualcan.path.uab.edu/>

FORMACIÓN **IAVANTE** Fundación Progreso y Salud

CENTRO DE SIMULACIÓN CLÍNICA AVANZADA

@IAVANTE_FPS | #IAVANTEformación | www.IAVANTE.es



Gracias por su atención

www.IAVANTE.es

IAVANTE