

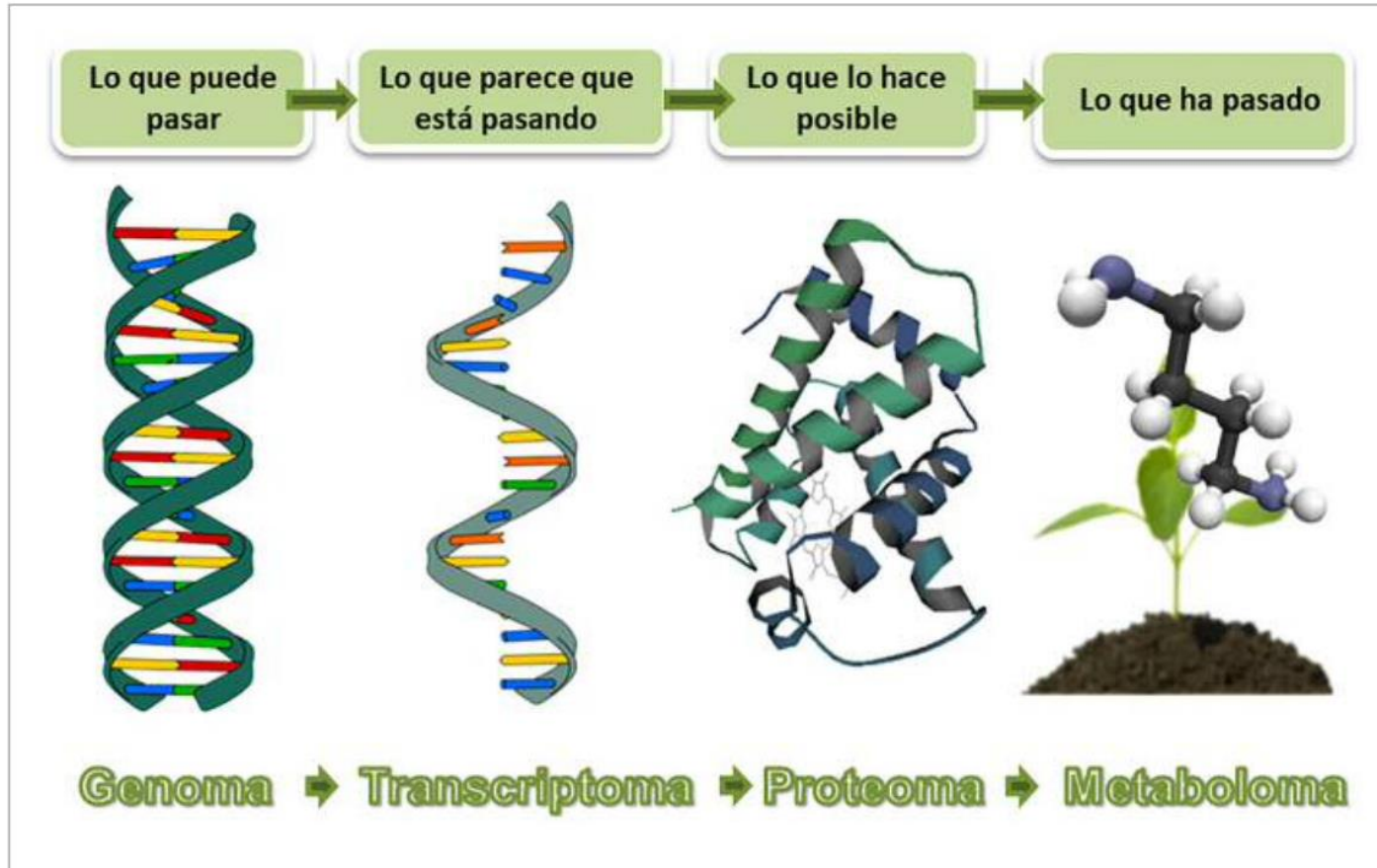


Introducción a R CON DATOS DE EXPRESIÓN DIFERENCIAL

M.C. Anahí Canedo Téxon
anahi.canedo@ecosur.mx



Ciencias ómicas

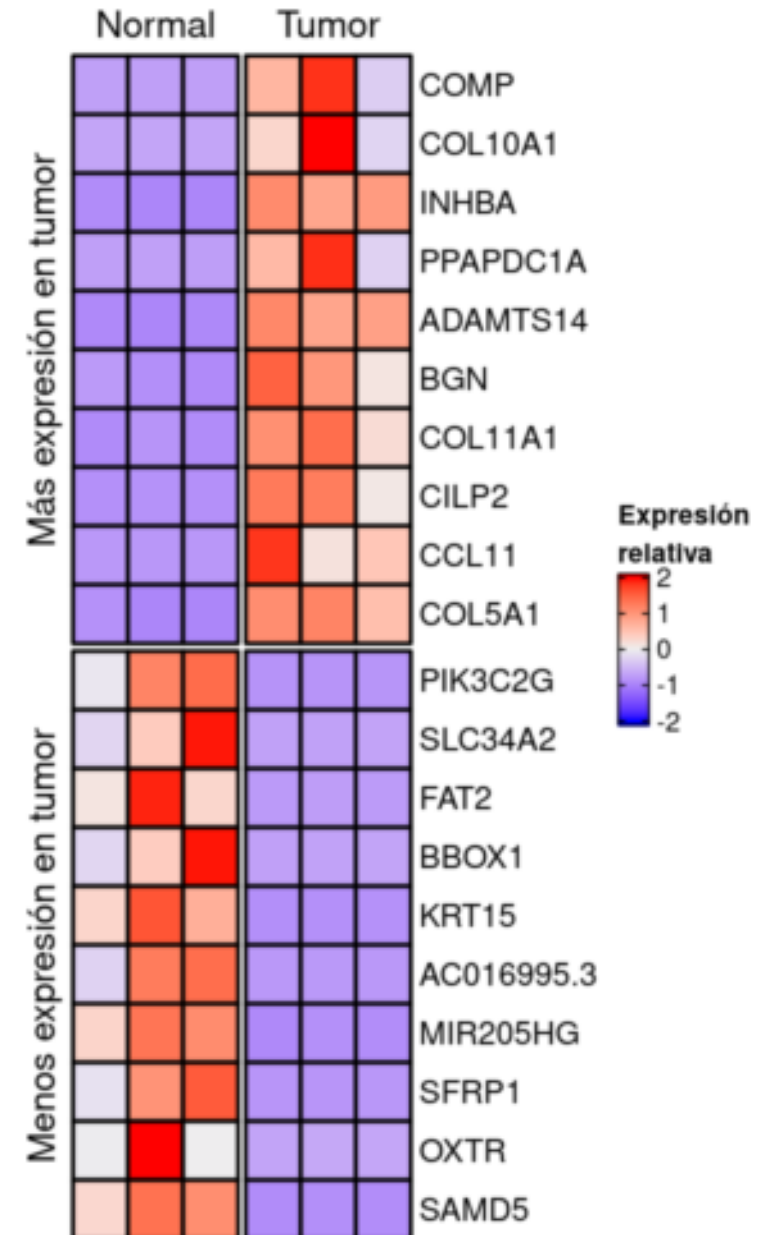


- Estudio holístico de los sistemas biológicos
- Se caracterizan por emplear tecnologías de alto rendimiento, es decir, que generan grandes cantidades.

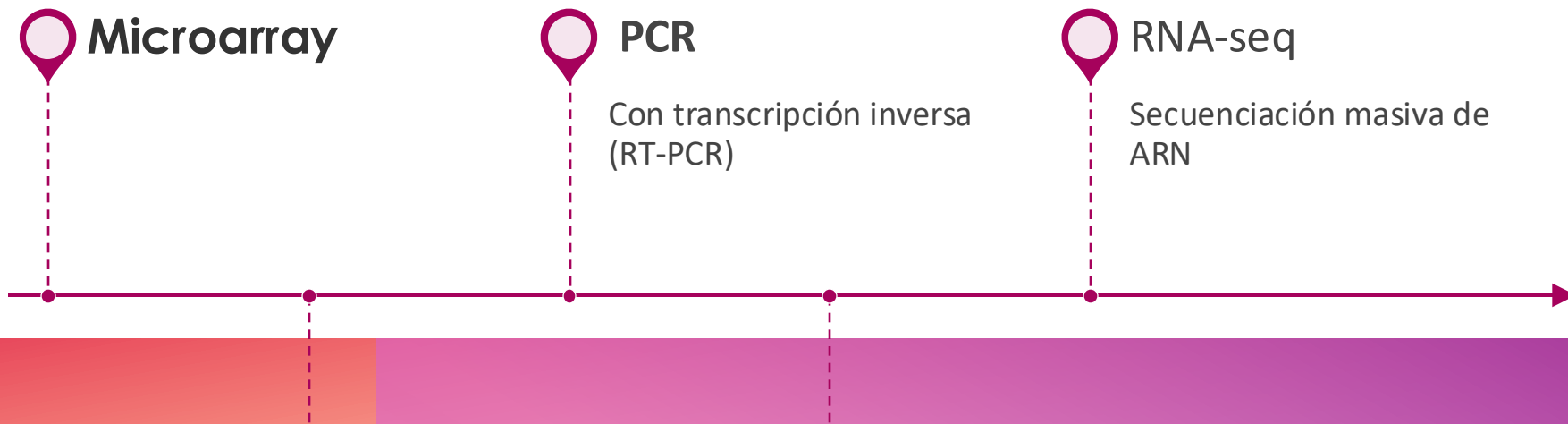
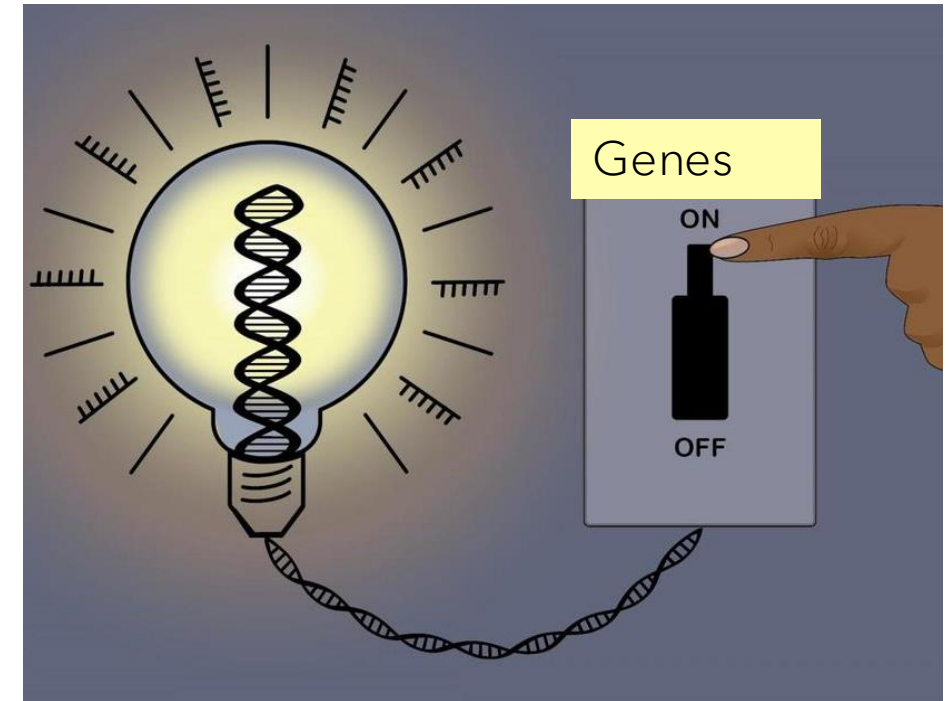
Saito, 2013 ; Ziegler et al., 2009 ; Caudai *et al.*, 2021

Transcriptómica-RNAseq

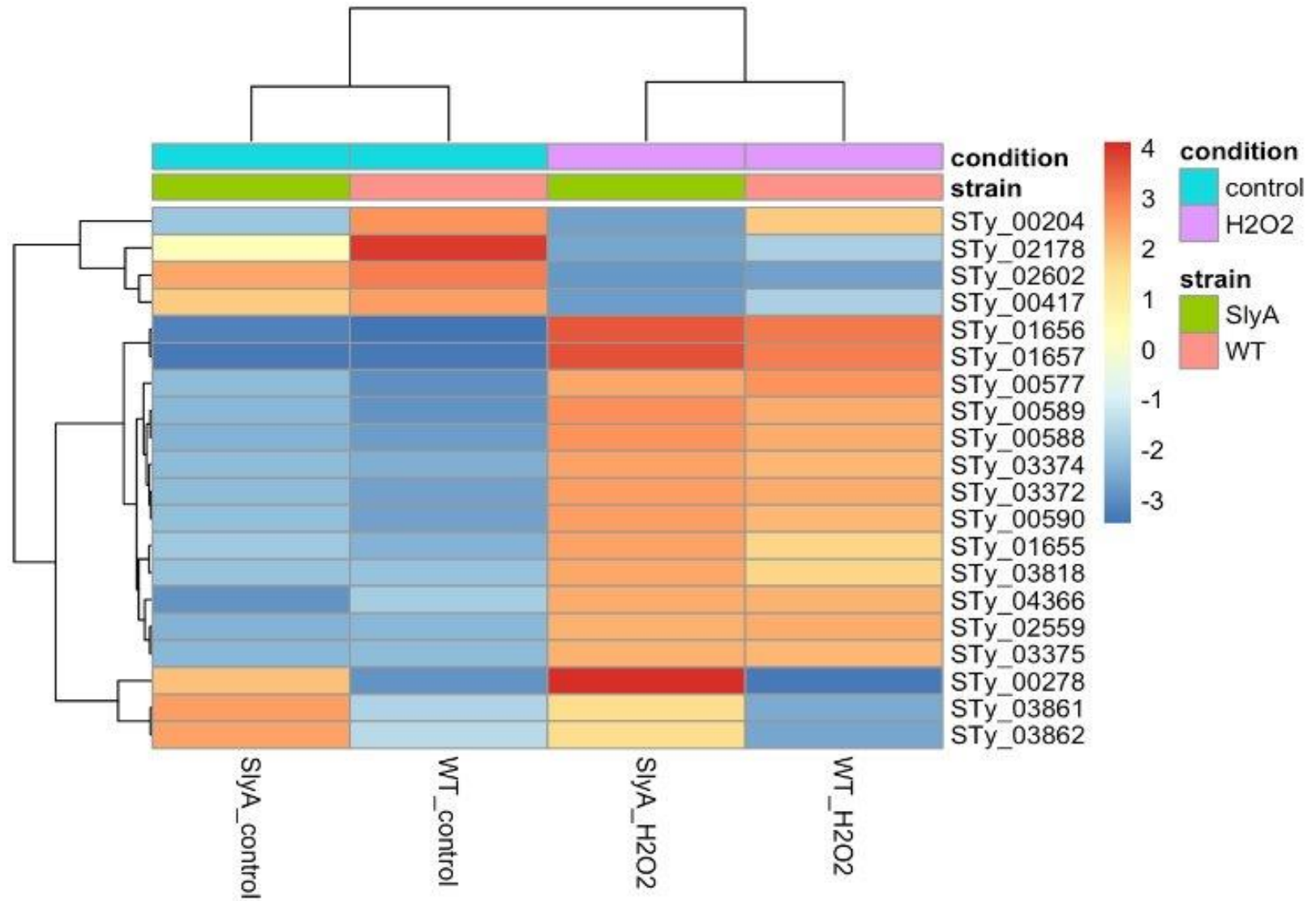
Es la ciencia que estudia el **patrón de expresión génica** en un organismo o en células específicas bajo circunstancias



- El análisis de la expresión génica es el estudio de cómo se activan y desactivan los genes en diferentes células y tejidos, y cómo esto afecta su función y comportamiento.
- Procesos biológicos como el desarrollo, la diferenciación, las enfermedades y la respuesta a los estímulos.



La tecnología más utilizada es RNASeq a partir de la cual se pueden realizar análisis de expresión génica diferencial para detectar aquellos **genes que se expresan diferencialmente respecto a una condición control.**





Control



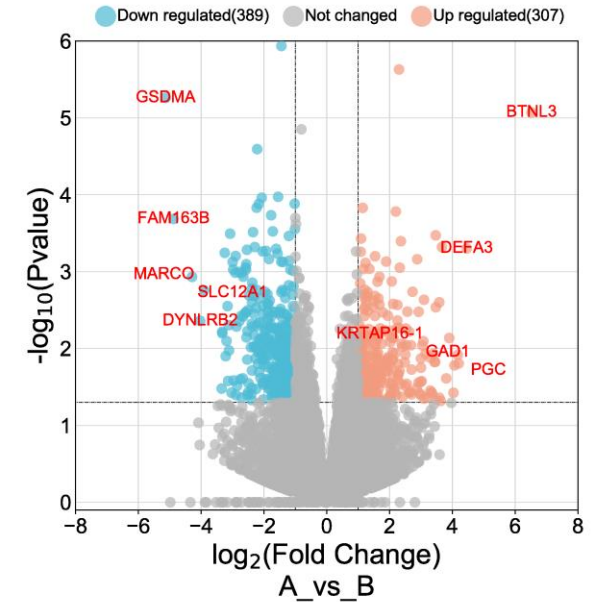
Tratamiento



Fotografía
instantánea

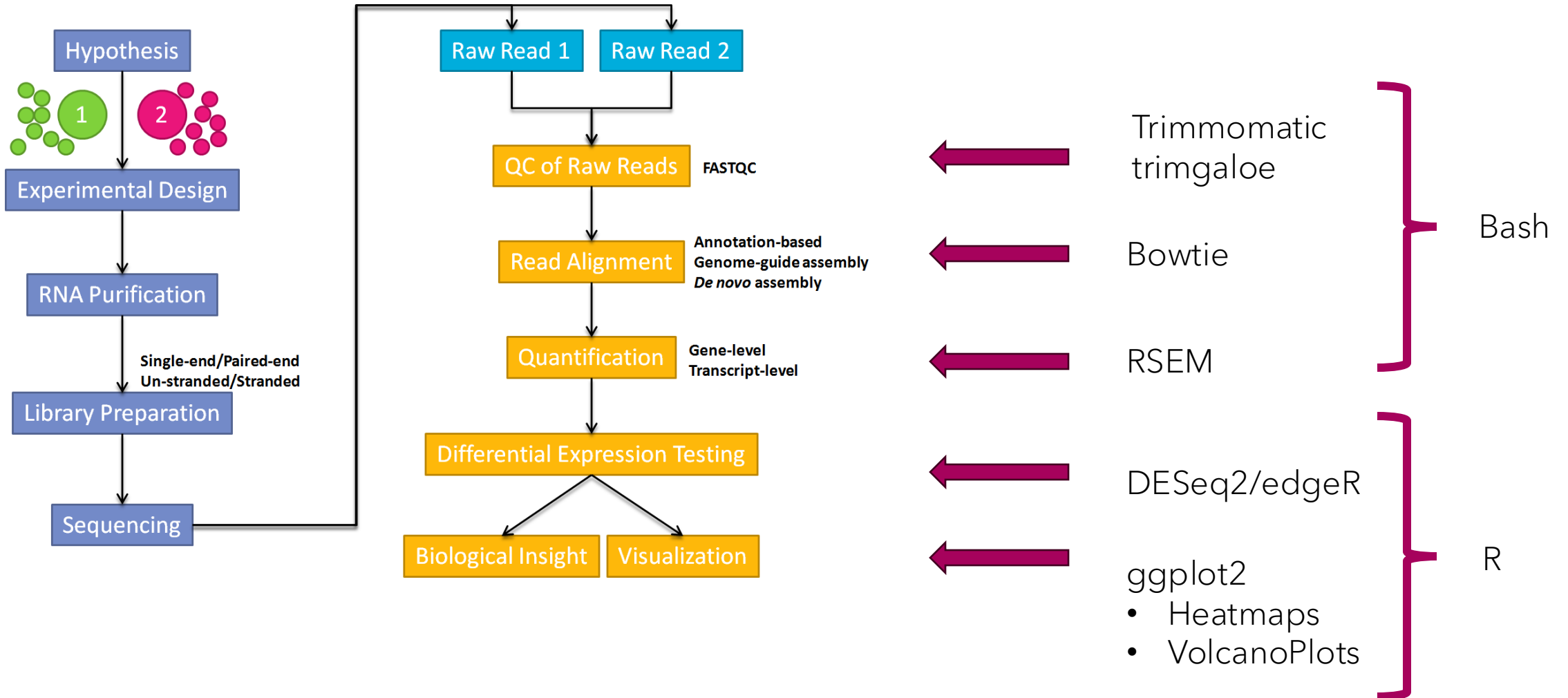


Secuenciación



Expresión
Diferencial

PIPELINE



Bowtie2

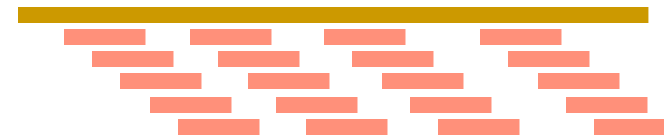
Mapeador de lecturas



RSEM (RNA-Seq by Expectation-Maximization)

Cuantificación del número de lecturas asignadas a cada unigen

$$\text{Abundancia de un gen} = \frac{\text{No. de lecturas}}{\text{transcrito}}$$



Normalización

$$FPKM = \frac{10^9 \times \text{Número de lecturas mapeadas al unigene}}{\text{Número total de lecturas por experimento} \times \text{longitud del transcrito (kb)}}$$

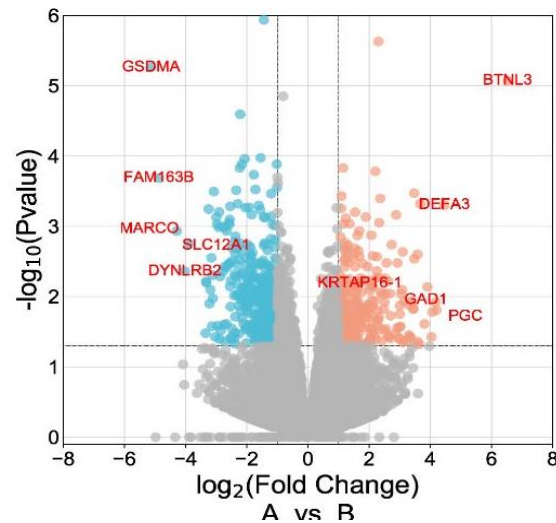
FPKM (Fragments Per Kilobase Million).

TPM (Transcripts Per Kilobase Million)

deseq2/edgeR

Identificación de los genes diferencialmente expresados entre cada uno de los tejidos, tratamientos, células

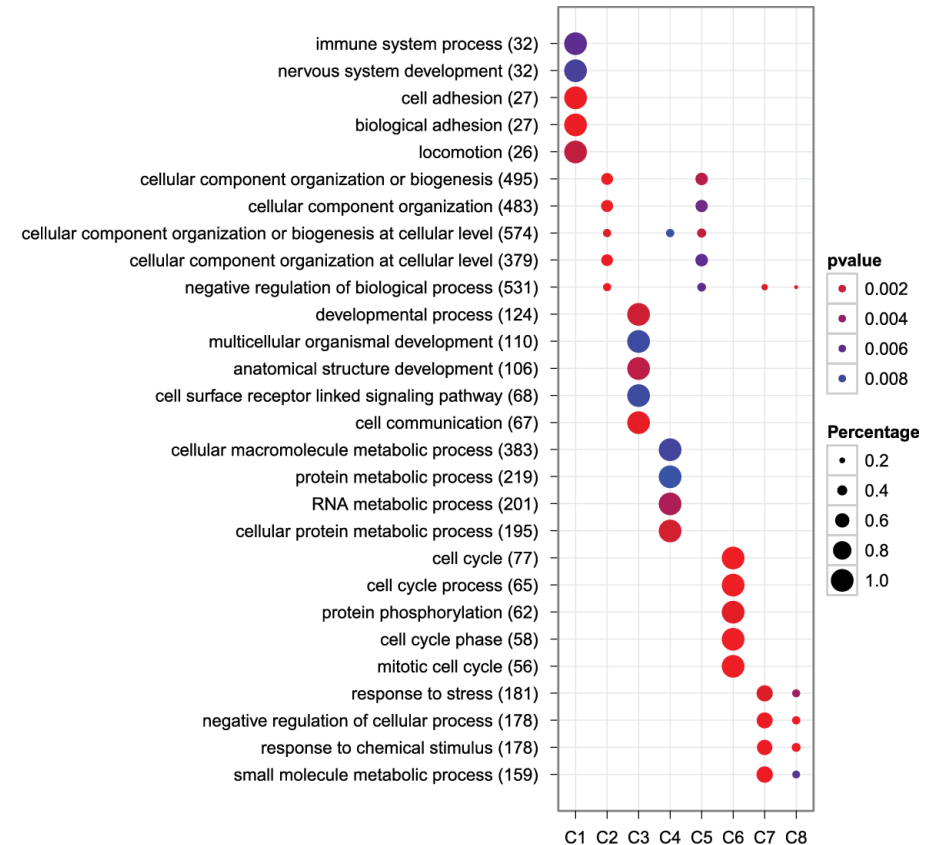
1. Comparativos pareados entre muestras
2. Establecer punto de corte basado en FoldChange (tasa de cambio)
3. Test estadístico de corrección FDR (Falso descubrimiento)



$$\text{FoldChange} = \frac{\text{Abundancia del gen 1 en el \acute{o}rgano A}}{\text{Abundancia del gen 1 en el \acute{o}rgano B}}$$

Enriquecimiento Funcional

Significado Biol\u00f3gico



1. Preparing Reference Sequences

RSEM can extract reference transcripts from a genome if you provide it with gene annotations in a GTF/GFF3 file. Alternatively, you can provide RSEM with transcript sequences directly.

```
rsem-prepare-referenc --bowtie2 Transcriptoma.fasta TrinityAssembly
```

2. Calculating Abundance Matrix

Estimar la abundancia de transcripciones a partir de datos de RNA-Seq. Estas herramientas consideran factores como la profundidad de lectura, la longitud de la transcripción y los sesgos

```
abundance_estimates_to_matrix.pl --est_method RSEM --cross_sample_norm TMM  
--out_prefix AbundanceMatrix
```

3. Calculating Expression Values

La normalización garantiza que los valores de expresión sean comparables entre muestras. La normalización tiene en cuenta el tamaño de la biblioteca, la longitud de los genes y otros factores de confusión.

```
rsem-calculate-expression --bowtie2 --paired-end HQ_EILTFCF01_R1.fastq  
HQ_EILTFCF01_R2.fastq TrinityAssembly Sample.genes.results
```

Counts

Número de lecturas asignadas a cada unigen

TPM

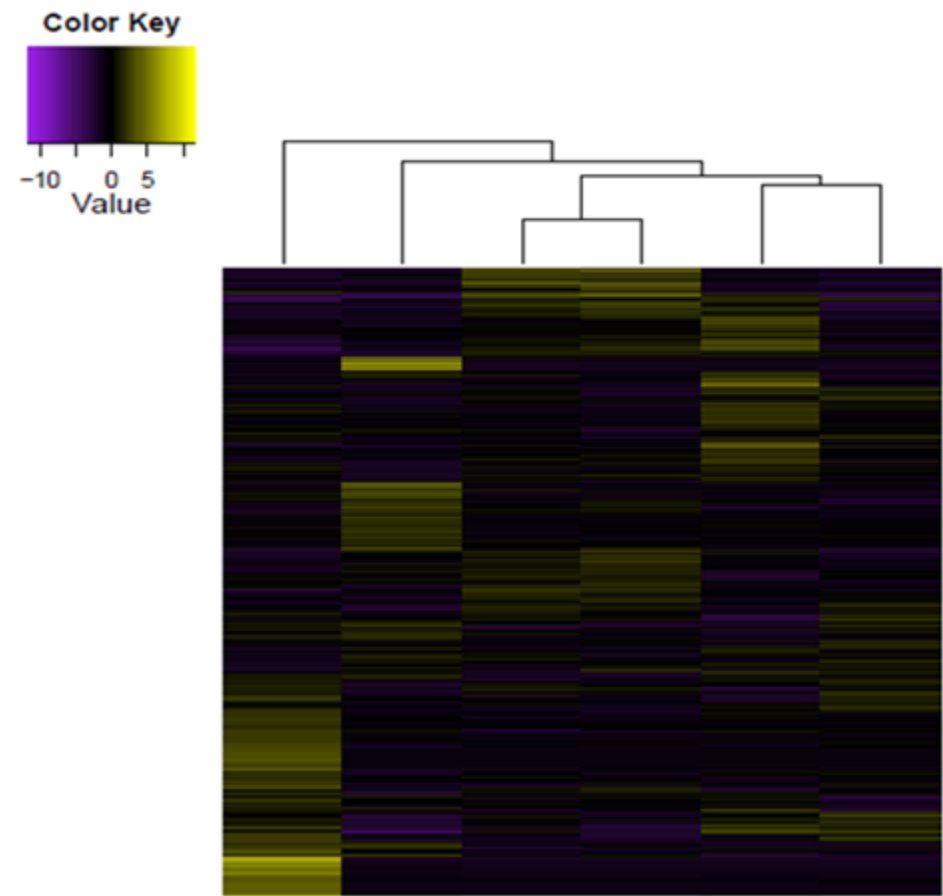
Permite hacer comparativos entre muestras

FPKM

Permite hacer comparativos de genes dentro de una misma muestra.

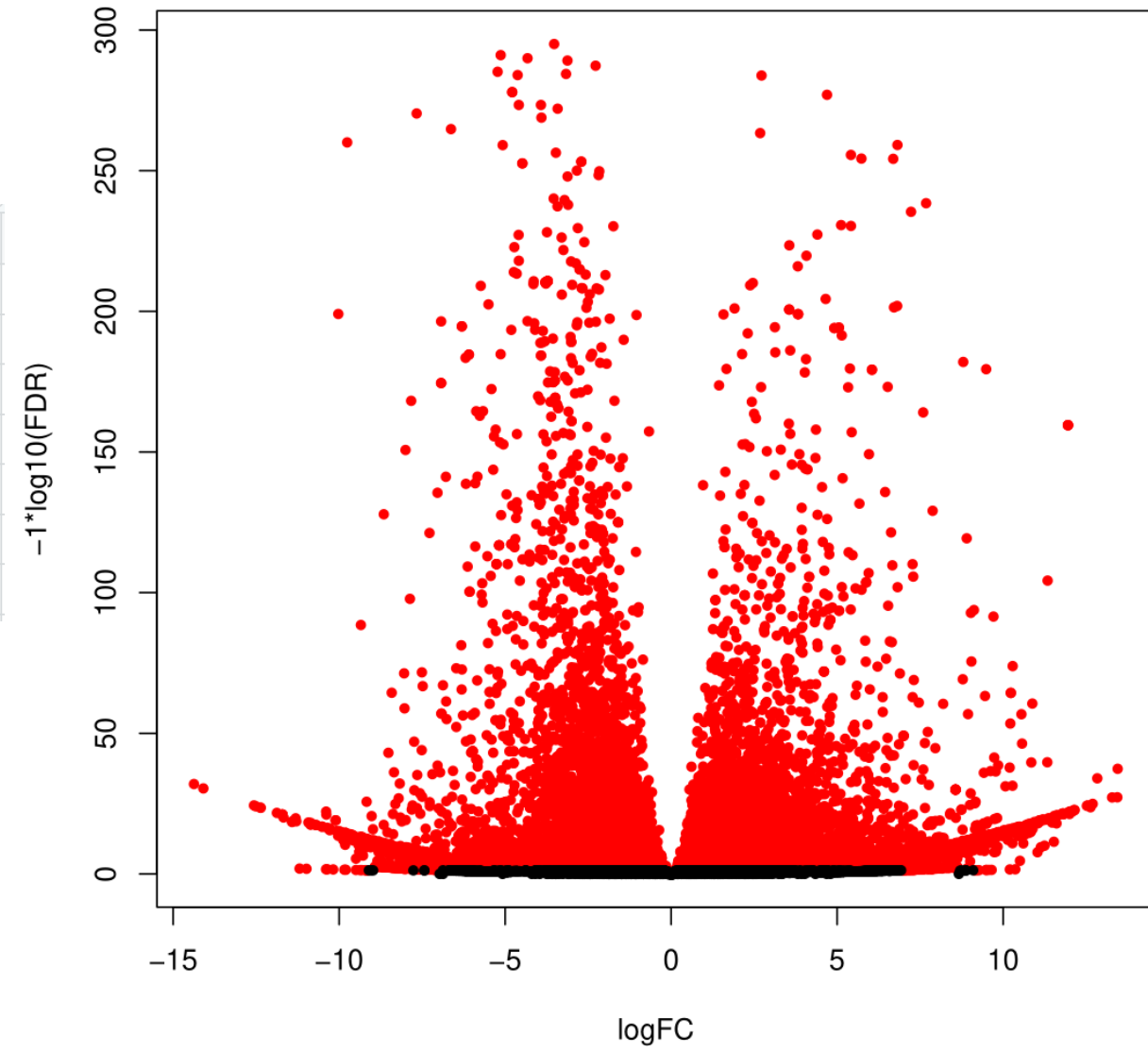
AbundanceMatrix.TMM.fpkm.matrix

GeneID	Control	Tratamiento.1	Tratamiento.2
Unigene169637	52.81	107.20	27.78
Unigene82021	7.00	6.00	4.00
Unigene184668	0.00	0.00	0.00
Unigene186330	0.00	18.43	3.68
Unigene2654	0.00	0.00	0.00
Unigene8580	0.00	0.00	0.00
Unigene99253	11.34	1.47	12.94



Sample.genes.results

gene_id	length	effective_length	expected_count	TPM	FPKM
Unigene100001	441	296.03	31.98	8.32	15.51
Unigene100005	231	88.33	0.80	0.70	1.30
Unigene100006	222	79.93	15.48	14.92	27.80
Unigene100007	156	25.15	2.31	7.08	13.20
Unigene100008	456	311.03	0.31	0.08	0.14
Unigene100010	447	302.03	384.85	98.16	182.89
Unigene100011	858	713.03	0.00	0.00	0.00





anahi.canedo@ecosur.mx