



UNAM
POSGRADO
Ciencias Biológicas



¿Qué hago y cómo lo hago?

Bioinformática para el análisis de datos de expresión *RNAseq* en búsqueda de blancos terapéuticos en el cáncer de mama

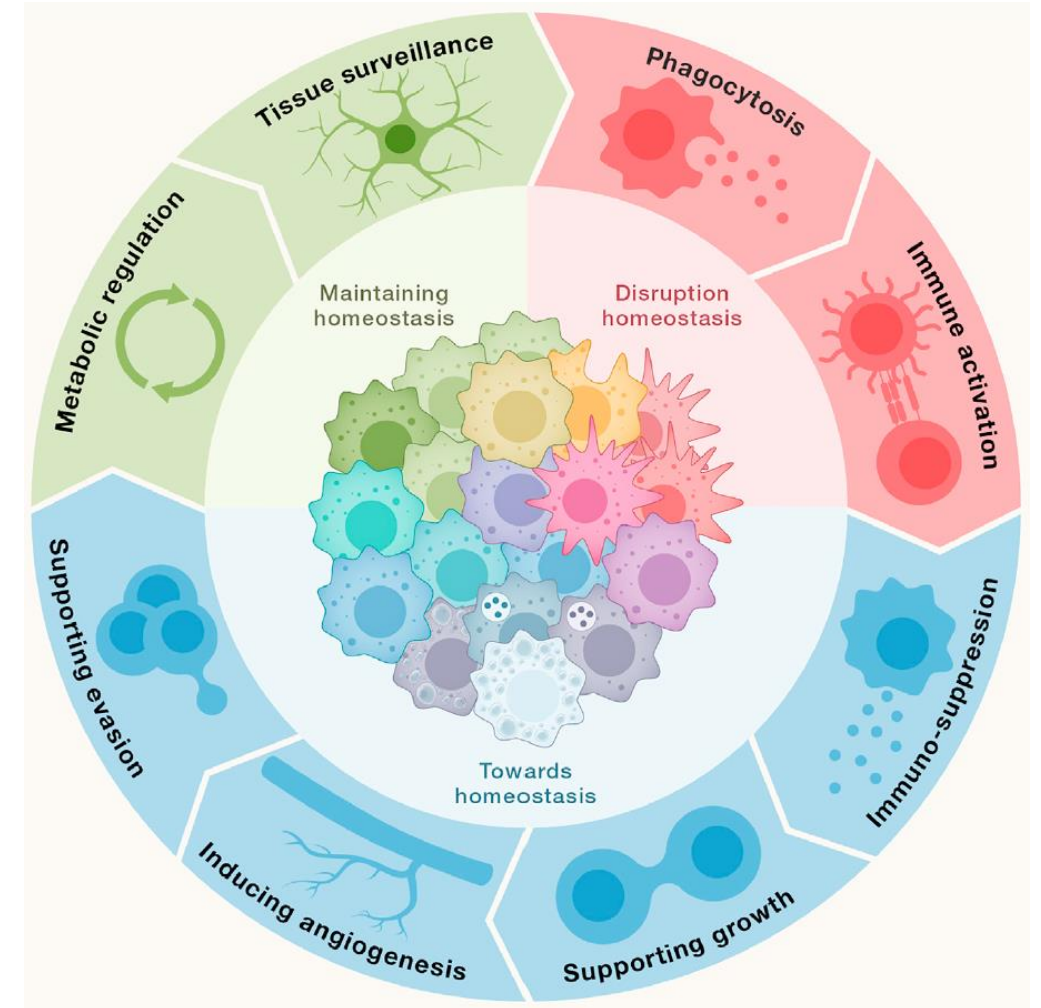
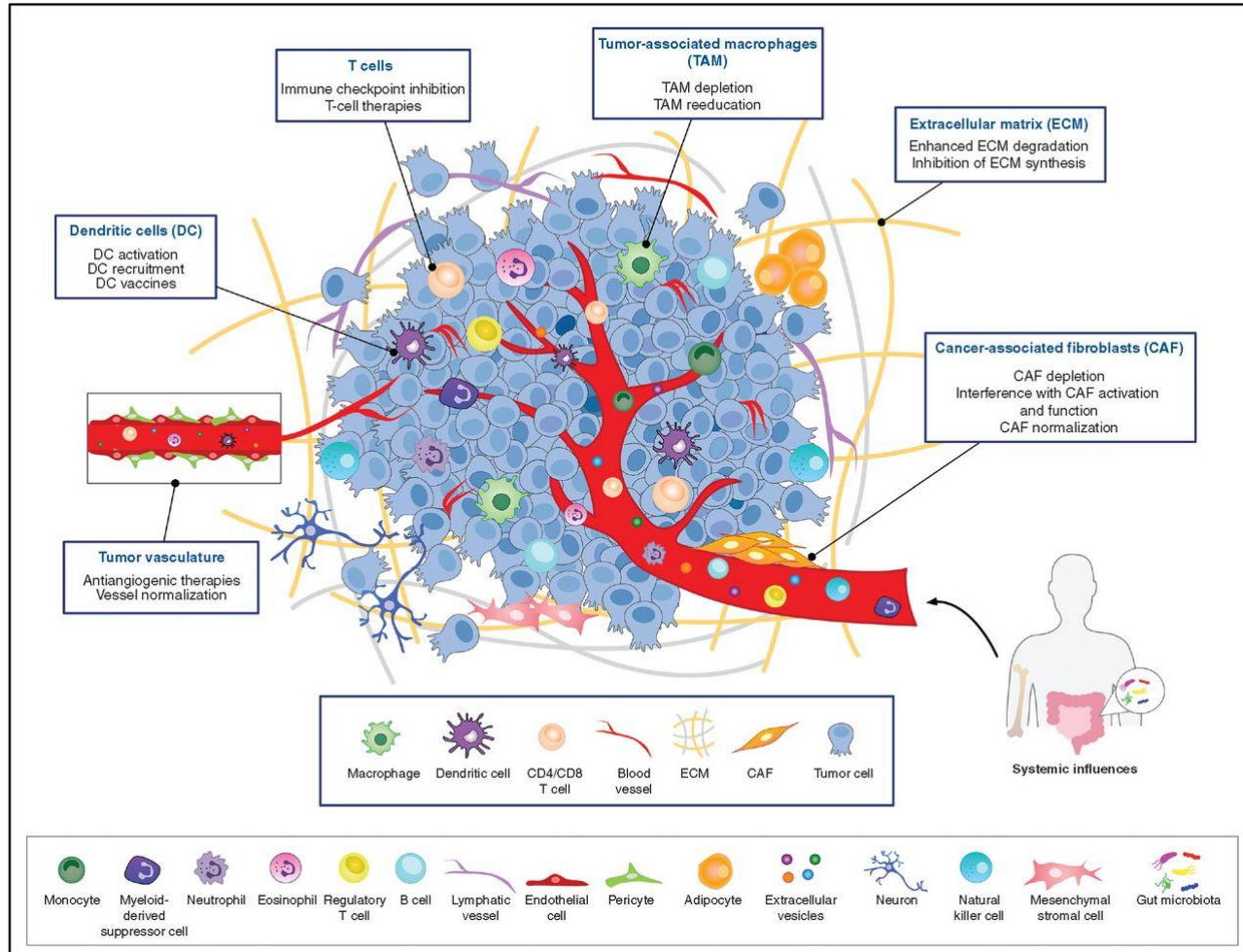
Biólogo Experimental: Jhonatan Raúl Martínez Valderrama

Dr. Ezequiel M. Fuentes Pananá
Dr. Diego Prada Gracia
Dra. Carla Daniela Robles Espinoza
Dra. Cecilia Suárez Arriaga

Noviembre, 2024

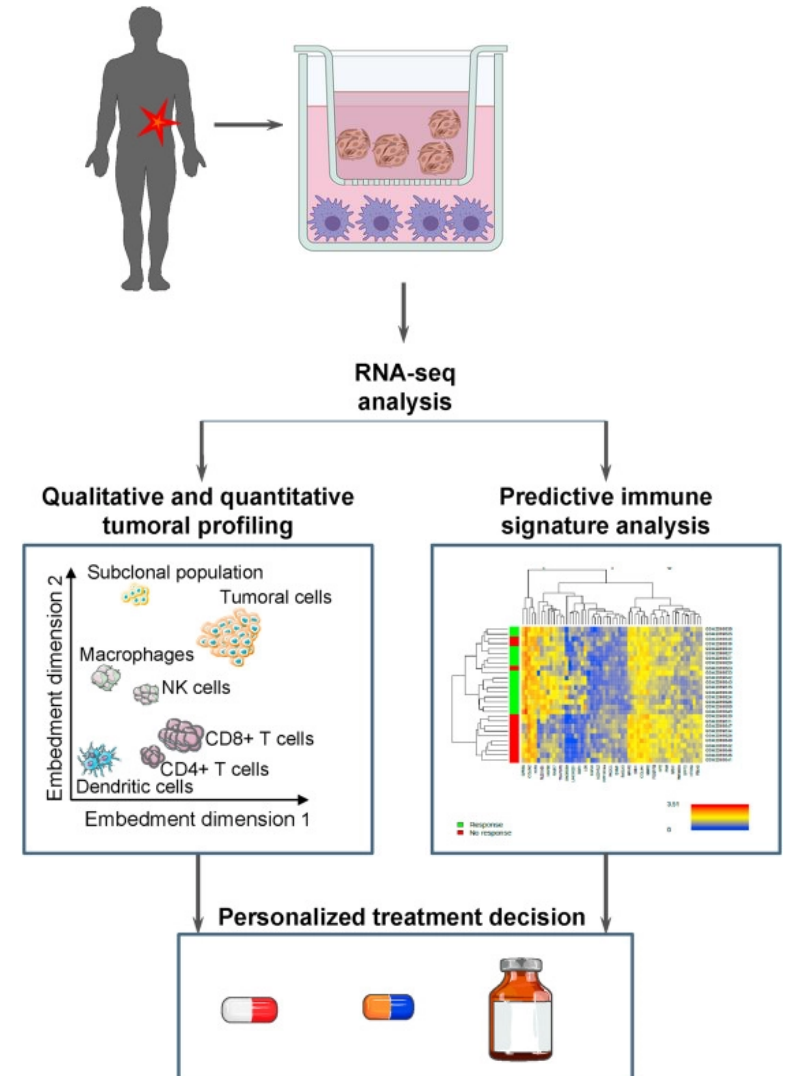
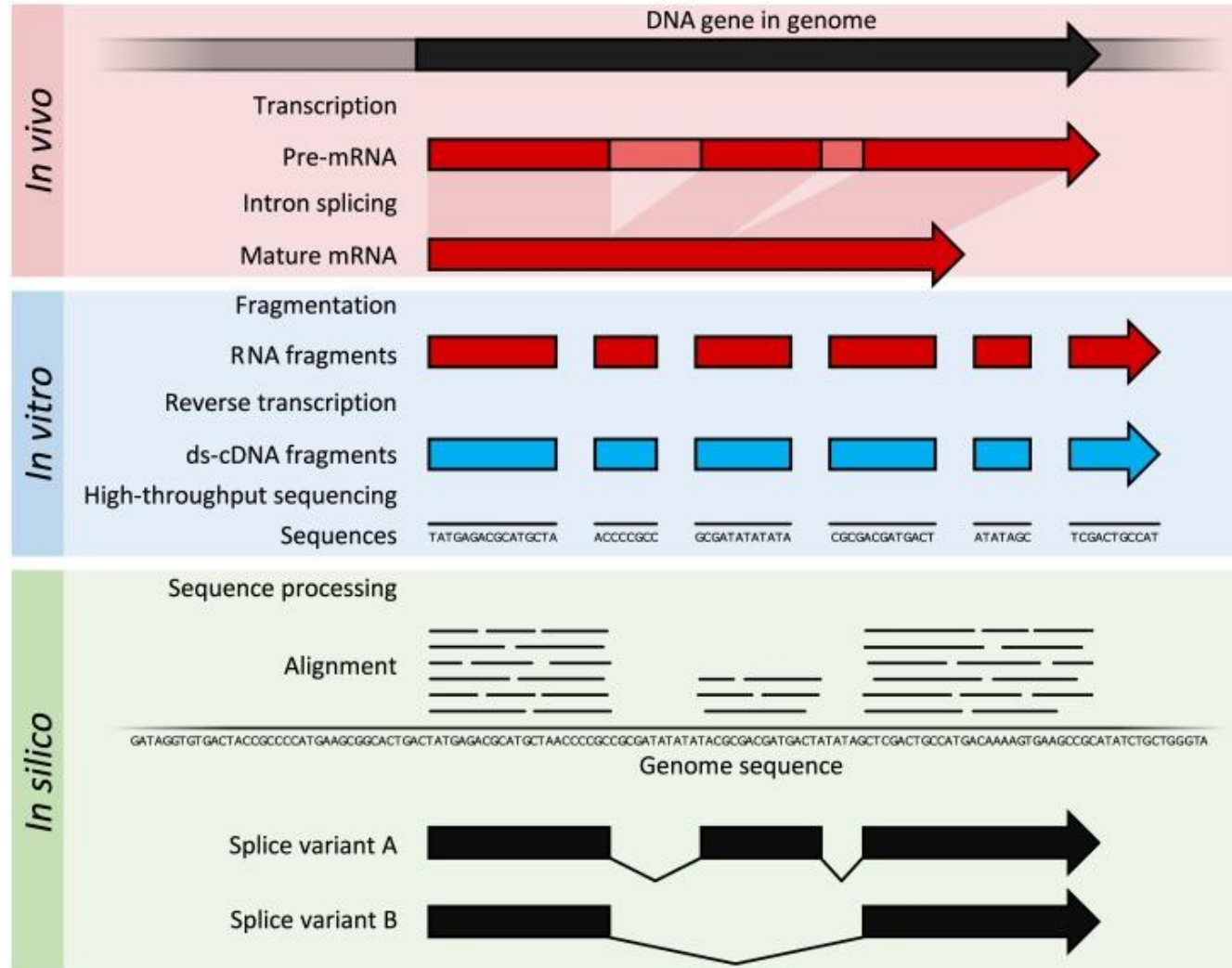
¿Qué es lo que hago?

Análisis transcriptómico de macrófagos en el microambiente tumoral del cáncer de mama.



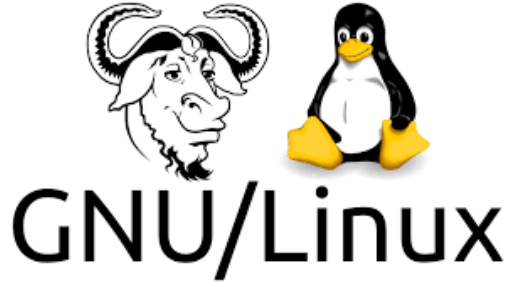
¿Qué es lo que hago?

Análisis transcriptómico de macrófagos en el microambiente tumoral del cáncer de mama.

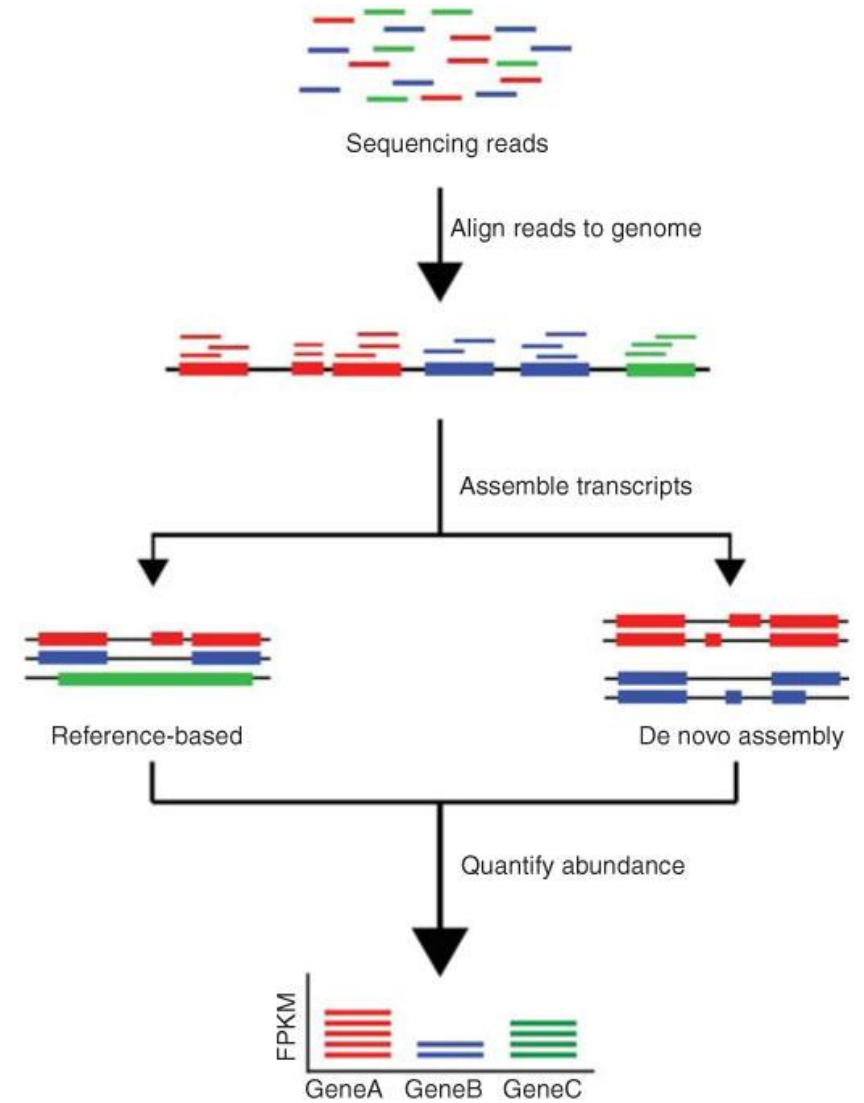


¿Cómo lo hago?

Aplicando métodos, recursos y herramientas computacionales.



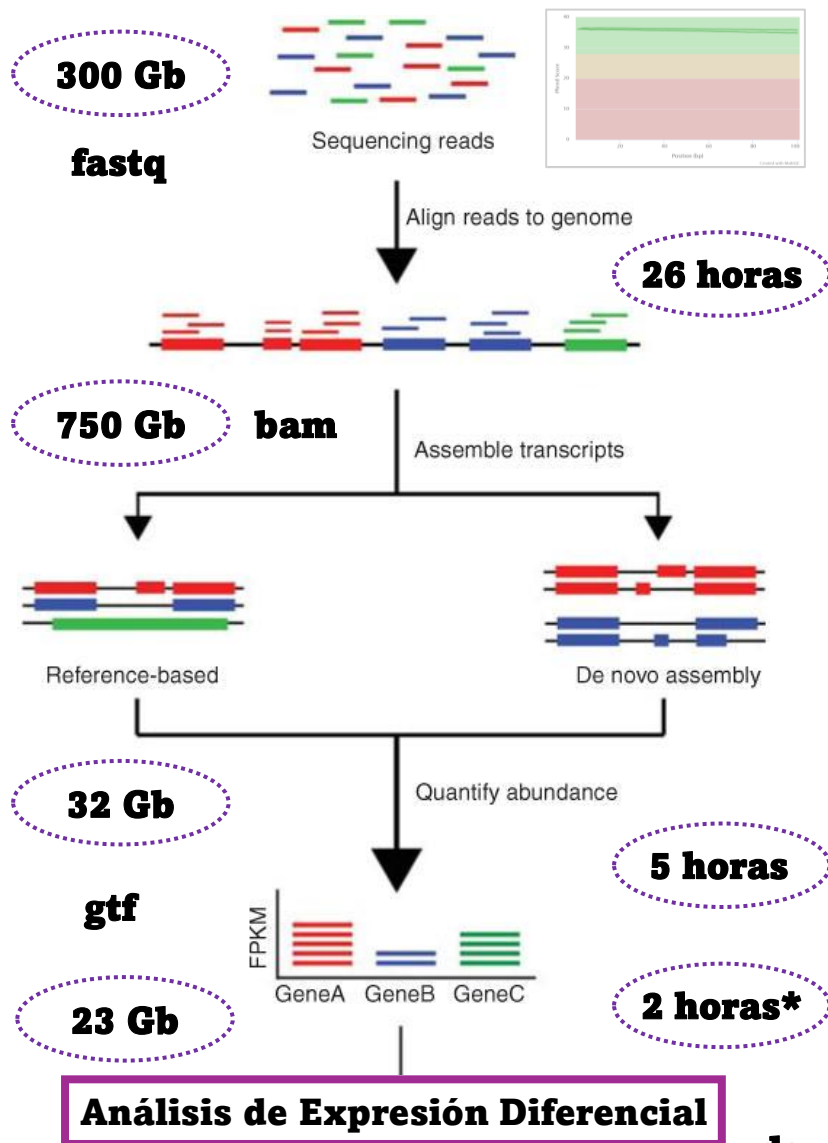
Precision 7920 Tower: Intel Xeon Gold 5218R; 20 cores; 40 threads;
NVIDIA T400 4Gb, 64 Gb RAM 4 Tb HDD & 1Tb SSD.



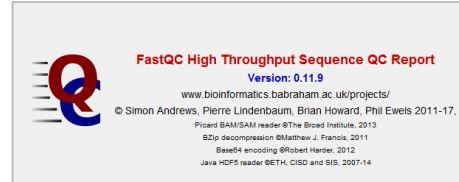
Análisis de Expresión Diferencial

¿Cómo lo hago?

Aplicando métodos, recursos y herramientas computacionales.



156 archivos FASTQ: 13 condiciones experimentales con triplicado biológico y replica una técnica.



Trimmomatic: A flexible read trimming tool for Illumina NGS data

Citations

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

[Bioinformatics](#). 2013 Jan; 29(1): 15–21.

Published online 2012 Oct 25. doi: [10.1093/bioinformatics/bts635](#)

PMCID: PMC3530905

PMID: [23104886](#)

STAR: ultrafast universal RNA-seq aligner

[Alexander Dobin](#),^{1,*} [Carrie A. Davis](#),¹ [Felix Schlesinger](#),¹ [Jorg Drenkow](#),¹ [Chris Zaleski](#),¹ [Sonali Jha](#),¹ [Philippe Batut](#),¹ [Mark Chaisson](#),² and [Thomas R. Gingeras](#)¹

Letter | Published: 18 February 2015

StringTie enables improved reconstruction of a transcriptome from RNA-seq reads

[Mihaela Pertea](#), [Geo M Pertea](#), [Corina M Antonescu](#), [Tsung-Cheng Chang](#), [Joshua T Mendell](#) & [Steven L Salzberg](#) ✉

[Nature Biotechnology](#). 33, 290–295 (2015) | [Cite this article](#)

Protocol | Published: 11 August 2016

Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown

[Mihaela Pertea](#), [Daehwan Kim](#), [Geo M Pertea](#), [Jeffrey T Leek](#) & [Steven L Salzberg](#) ✉

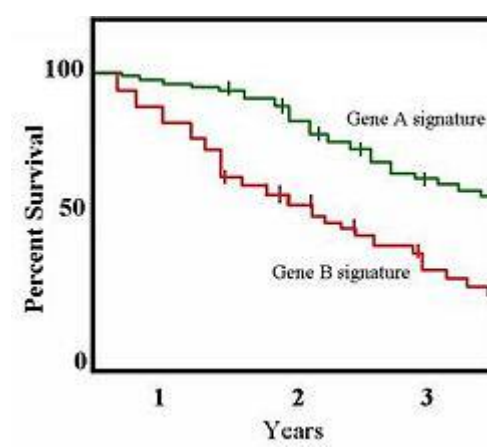
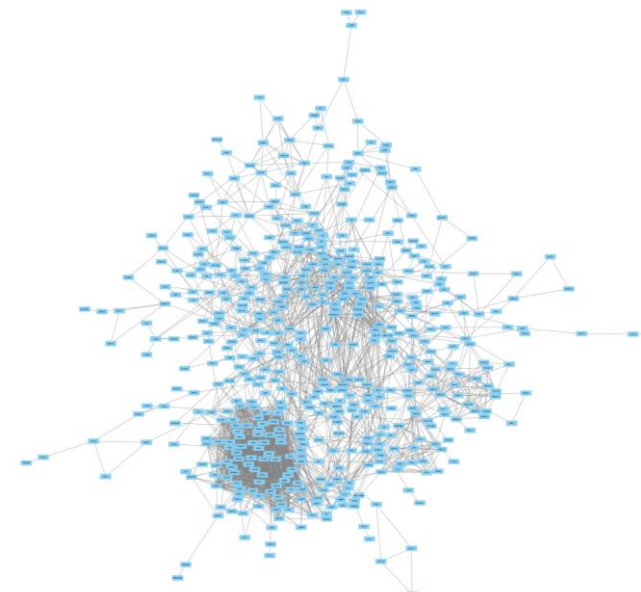
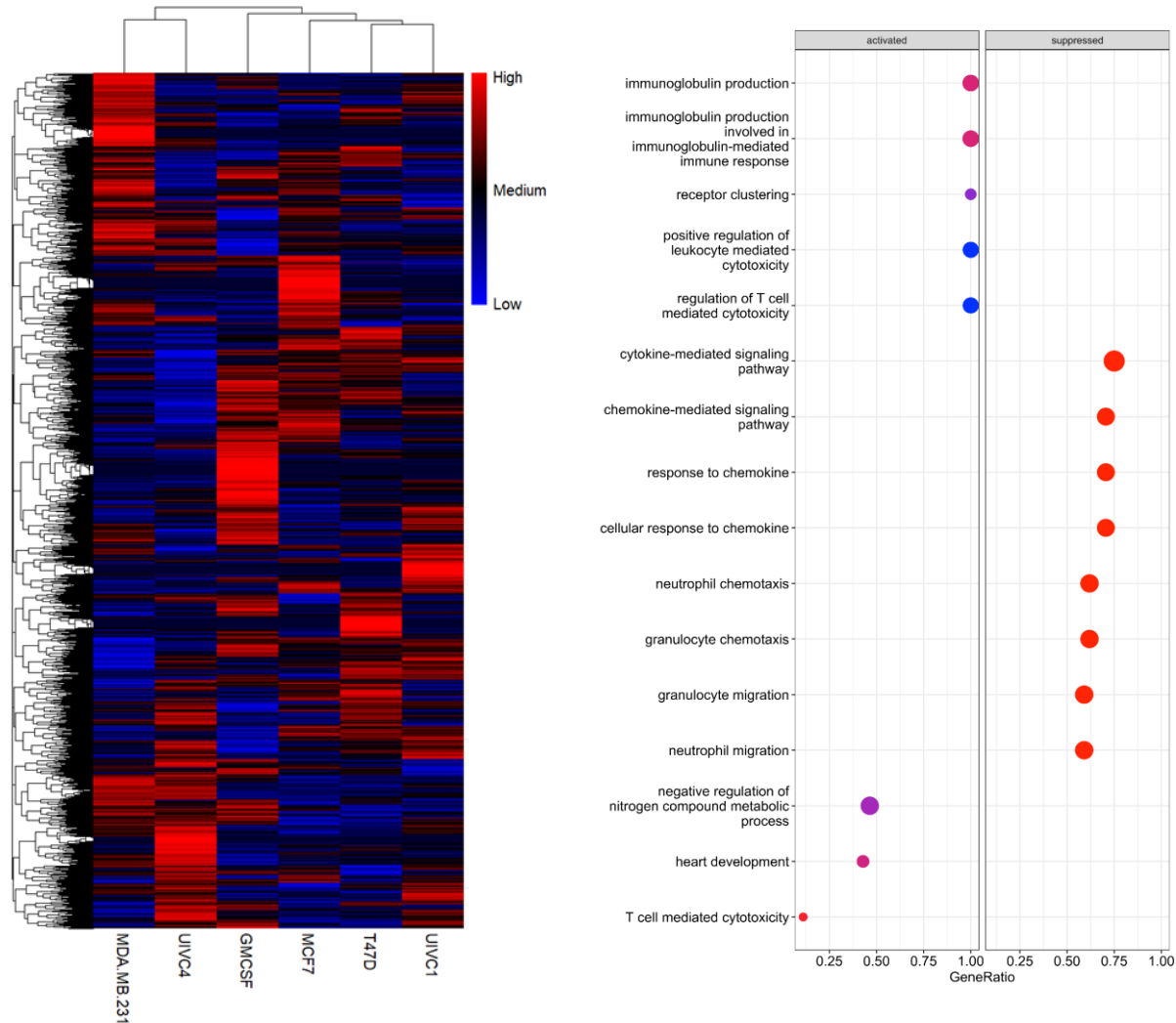
[Nature Protocols](#). 11, 1650–1667 (2016) | [Cite this article](#)



rds, csv, tsv, txt, png, pdf...

¿Para qué un análisis de expresión diferencial?

Para identificar los genes que se expresan preferentemente en determinada condición, establecer los procesos biológicos en los que se ven implicados, saber cuáles de ellos aparentemente tienen una mayor importancia para abordarlos como marcadores diagnósticos y blancos terapéuticos.



¿Cómo se ve el proceso de lo que hago?

```
File Edit Selection View Go Run Terminal Help
$ star_alignment.sh X
home > jrmaval > marval > test_rnaseq > $ star_alignment.sh
1 #!/bin/bash
2
3 mkdir -p alignment
4
5 # Variables:
6 input_dir="/mnt/Guardado/data_mac2/trimming_data/trimmed_reads"
7
8 # Iniciar el temporizador
9 start_time=$(date +%s.%N)
10
11 # Alineamiento: Loop para alinear cada par de archivos R1 y R2 con STAR
12
13 for muestra in {37..54}; do
14     # Agregar un cero antes del número de muestra si es menor que 10
15     muestra_formatted=$(printf "%03d" $muestra)
16
17     # Nombre base de los archivos R1 y R2 para esta muestra
18     base_name="EF1CM3SS25_${muestra_formatted}_S${muestra}"
19
20     # Loop para cada lectura (L001, L002, etc.)
21     for lectura in {1..2}; do
22         # Rutas a los archivos FASTQ para esta muestra y lectura
23         r1_file="${base_name}_L00${lectura}_R1_001_paired.fastq.gz"
24         r2_file="${base_name}_L00${lectura}_R2_001_paired.fastq.gz"
25         r1_path="${input_dir}/${r1_file}"
26         r2_path="${input_dir}/${r2_file}"
27
28         # Usando STAR
29         STAR --genomeDir /mnt/Guardado/data_mac2/trimming_data/trimmed_reads/Genome \
30             --sjdbGTFfile /mnt/Guardado/data_mac2/trimming_data/trimmed_reads/Genome/genes.gtf \
31             --sjdbOverhang 100 \
32             --runThreadN 32 \
33             --readFilesIn "${r1_path}" "${r2_path}" \
34             --readFilesCommand zcat \
35             --outFileNamePrefix "alignment/${base_name}_L00${lectura}" \
36             --outSAMtype BAM SortedByCoordinate \
37             --quantMode TranscriptomeSAM
38     done
39 done
40
41 # Analisis de Calidad del Alineamiento
42 mkdir -p alignment/quality_alignment
```

```
jrmaval@LAPTOP-8SI0DC1R: X + v
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$ ls
bitacora.md data extraer_rna.sh qc.sh qc_reads reads saludo.sh star_alignment.sh trimm.sh
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$ nohup ./saludo.sh &
[1] 265016
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$ nohup: ignoring input and appending output to 'nohup.out'

[1]+  Done                  nohup ./saludo.sh
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$ tail -f nohup.out
Hola, Mundo

^C
(QualityControl) jrmaval@LAPTOP-8SI0DC1R:~/marval/test_rnaseq$ top
top - 21:42:16 up 2 days, 16 min, 1 user, load average: 0.00, 0.01, 0.00
Tasks: 58 total, 1 running, 57 sleeping, 0 stopped, 0 zombie
%Cpu(s):  0.2 us,  0.3 sy,  0.0 ni, 99.5 id,  0.0 wa,  0.0 hi,  0.1 si,  0.0 st
MiB Mem : 7616.4 total, 5952.7 free, 741.7 used, 922.0 buff/cache
MiB Swap: 2048.0 total, 2048.0 free,  0.0 used, 6559.0 avail Mem

  PID USER      PR  NI   VIRT   RES    SHR   S  %CPU  %MEM    TIME+  COMMAND
258429 jrmaval   20   0 968268 96908 39604 S   5.3   1.2   1:25.41 node
258377 jrmaval   20   0 956644 91100 41136 S   0.3   1.2   0:04.64 node
   1 root      20   0 166056 11244  8000 S   0.0   0.1   8:31.77 systemd
   2 root      20   0 2280    1308  1188 S   0.0   0.0   0:00.04 init-systemd(Ub
   7 root      20   0 2344    136   132 S   0.0   0.0   0:04.36 init
  46 root     19  -1 47832 15648 14588 S   0.0   0.2   0:02.94 systemd-journal
  71 root     20   0 22096  6024  4552 S   0.0   0.1   0:18.16 systemd-udev
  82 root     20   0  4496   180   32 S   0.0   0.0   0:00.02 snapfuse
  84 root     20   0  4496   168   20 S   0.0   0.0   0:00.00 snapfuse
  86 root     20   0  4784  1804  1368 S   0.0   0.0   0:00.86 snapfuse
  91 root     20   0  4496   200   48 S   0.0   0.0   0:00.00 snapfuse
  93 root     20   0  4676  1044   816 S   0.0   0.0   0:00.05 snapfuse
  96 root     20   0  4496   176   28 S   0.0   0.0   0:00.00 snapfuse
 102 root     20   0  4628   196   52 S   0.0   0.0   0:00.03 snapfuse
 105 root     20   0  4628   160   12 S   0.0   0.0   0:00.03 snapfuse
 109 root     20   0  4496   196   44 S   0.0   0.0   0:00.00 snapfuse
 115 root     20   0  4740  1644  1132 S   0.0   0.0   0:02.20 snapfuse
 117 root     20   0  4496   160   16 S   0.0   0.0   0:00.00 snapfuse
 119 root     20   0  4796  1856  1428 S   0.0   0.0   0:01.32 snapfuse
 138 systemd+ 20   0 25540 12648  8452 S   0.0   0.2   0:01.00 systemd-resolve
 177 root     20   0  4308   2776  2524 S   0.0   0.0   0:00.57 cron
 179 message+ 20   0  8592  4656  4120 S   0.0   0.1   0:00.49 dbus-daemon
 185 root     20   0 30096 19492 10304 S   0.0   0.2   0:00.12 networkd-dispat
```

Ejecutando StringTie para EF1CM3SS25_S25_L001Aligned.sortedByCoord.out.bam...

Procesado: EF1CM3SS25_S25_L001Aligned.sortedByCoord.out.bam -> ballgown/EF1CM3SS25_S25_L001/genes.gtf

Ejecutando StringTie para EF1CM3SS25_S25_L002Aligned.sortedByCoord.out.bam...

Procesado: EF1CM3SS25_S25_L002Aligned.sortedByCoord.out.bam -> ballgown/EF1CM3SS25_S25_L002/genes.gtf

Ejecutando StringTie para EF1CM3SS26_S26_L001Aligned.sortedByCoord.out.bam...

Procesado: EF1CM3SS26_S26_L001Aligned.sortedByCoord.out.bam -> ballgown/EF1CM3SS26_S26_L001/genes.gtf

Ejecutando StringTie para EF1CM3SS26_S26_L002Aligned.sortedByCoord.out.bam...

Procesado: EF1CM3SS26_S26_L002Aligned.sortedByCoord.out.bam -> ballgown/EF1CM3SS26_S26_L002/genes.gtf

Done, Step 4

Tiempo de ejecución: 303 minutos 50.849019744 segundos.

Listo pinche desesperado, el trabajo se hizo en 304 min

(END)

Reportes: Trimming

Recibidos x

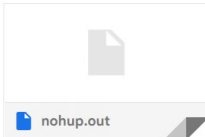
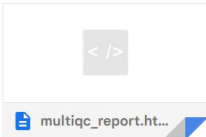


adminiuvic

para mí ▾

El **alineamiento** y análisis de calidad de las muestras se hizo en: 5 min

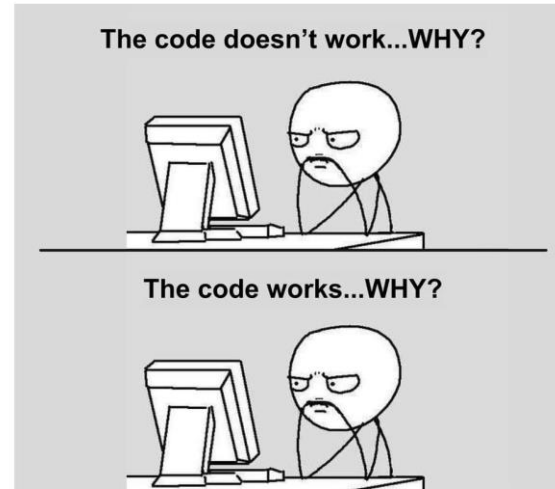
2 archivos adjuntos • Analizado por Gmail ⓘ



¿Cómo es el proceso?



Me:



Dudas y comentarios:



***“Ciencia es creer en la
ignorancia de los científicos.”***

Richard Phillips Feynman (1918-1988)

***“Todos los modelos son erróneos,
pero algunos son útiles”***

George Edward Pelham Box (1976)