

Análisis transcriptómicos de la expresión génica

Máster Universitario en Bioinformática

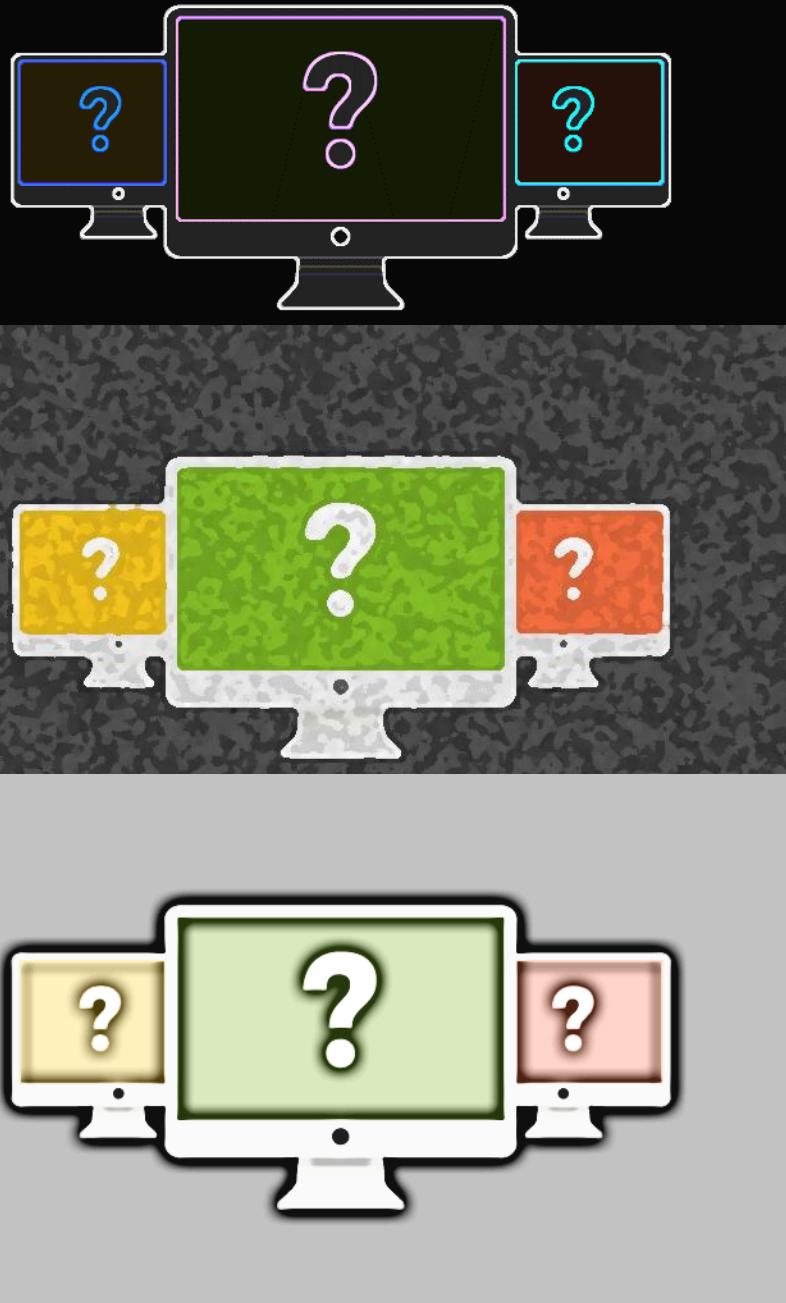
Sesión 13



**Universidad
Internacional
de Valencia**

Dra. Paula Soler Vila
paula.solerv@professor.universidadviu.com

De:
 Planeta Formación y Universidades



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

PRACTIQUEMOS



Ontología genética *anotación/análisis*

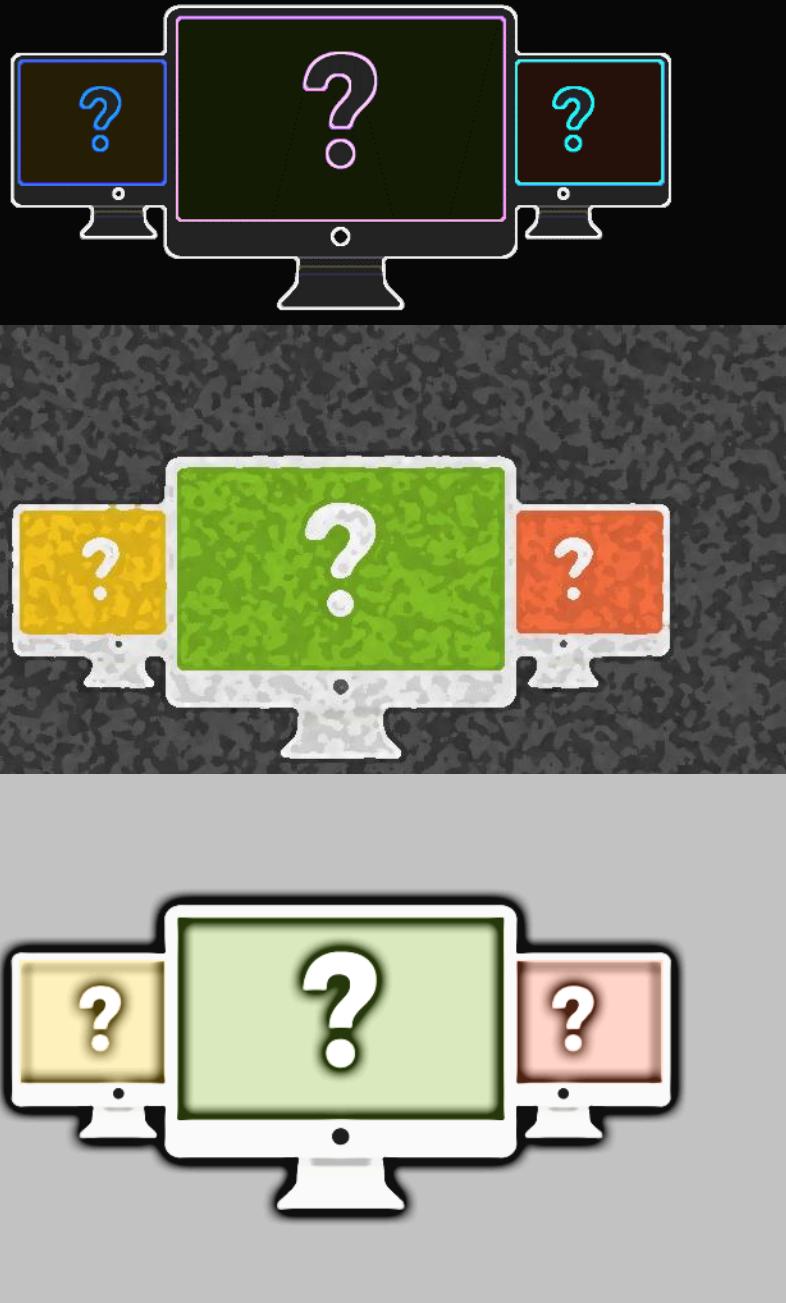
```
library(org.Mm.eg.db)
library(limma)
library(GO.db)

# 1. Gene ontology(GO) analysis
go <- goana(res,species = "Mm")
topGO(go, number = 2, ontology = "BP")

Term Ont N Up Down P.Up P.Down
GO:0022613 ribonucleoprotein complex biogenesis BP 399 25 194 1 4.782448e-48
GO:0042254 ribosome biogenesis BP 291 12 157 1 2.175947e-4

go <- goana(DEG$ENTREZID,species = "Mm")
topGO(go, number = 2, ontology = "BP")
```

Term	Ont	N	DE	P.DE
GO:0006956	complement activation	BP	187	5 5.983414e-05
GO:0006955	immune response	BP	1752	13 9.285810e-05



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

DUDAS ACTIVIDAD 1

- Compartir todos los comandos necesarios para asegurar la **reproducibilidad** del análisis y facilitar mi **comprensión**.
- Indicar en las capturas de pantalla todo aquello que sea **informativo**; si la salida de la ejecución de un comando es muy extensa, visualizar las primeras líneas que verifican su correcta ejecución.

Pregunta 4

¿Cuántas letras del alfabeto distintas podemos encontrar en la columna **CIGAR** en nuestro archivo SRR1552444_hisat2.sam?



La letra M indica una coincidencia de alineación entre las secuencias, y aparece 158.000 veces.

DUDAS ACTIVIDAD 1

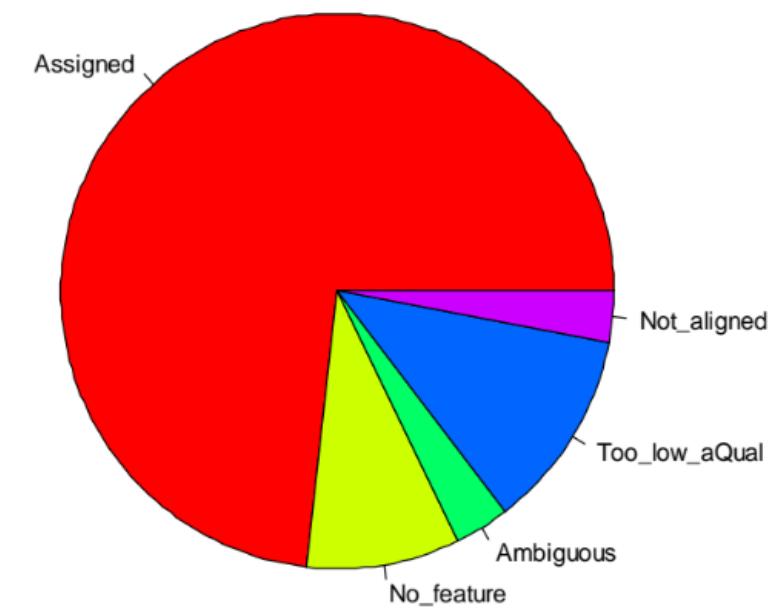
Pregunta 6

ENSMUSG00000106667	0
ENSMUSG00000106668	0
ENSMUSG00000106669	0
ENSMUSG00000106670	0
ENSMUSG00000106671	1
no_feature	2483249
ambiguous	908221
too_low_aQual	3261453
not_aligned	817651
alignment_not_unique	0

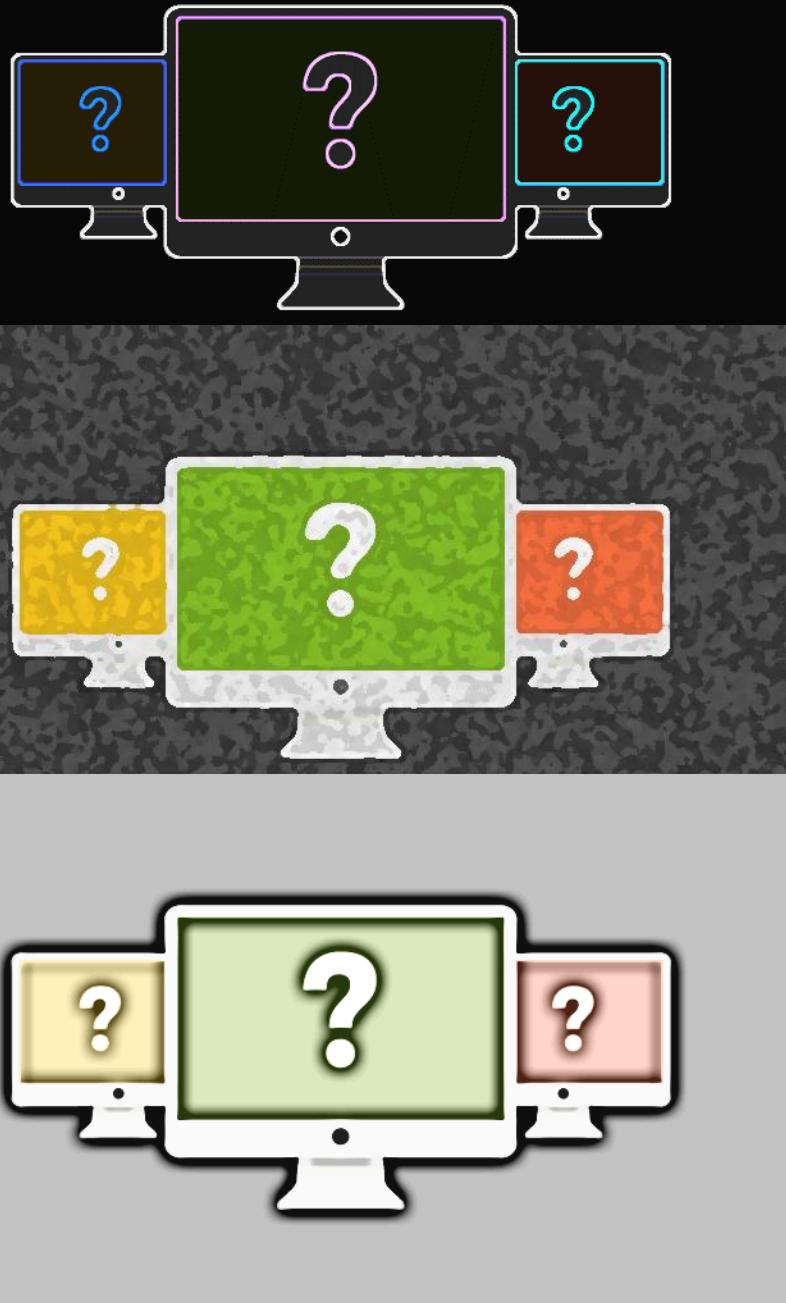
1) Obtener el número total de lecturas asignadas (**SUM**)

2) Obtener el número de lecturas no asignadas

3) Computar el porcentaje de cada valor teniendo en cuenta el número total de lecturas (asignadas como no asignadas)



4) Realizar un gráfico y comentar (adicinar el código empleado)



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

ACTIVIDAD 2



Actividad 2-1C

Archivos adjuntos: [Actividad 2.docx](#) (17,382 KB)
[Actividad 2.pdf](#) (136,137 KB)

From gene counts to DEG and pathways



El propósito de esta actividad es que el estudiante demuestre la adquisición de habilidades y competencias necesarias para llevar a cabo un **tratamiento estadístico** adecuado sobre los datos de conteo de un experimento de ARN-seq. Esto implica determinar los genes expresados de manera diferencial entre diferentes escenarios y extraer los principales términos ontológicos enriquecidos.

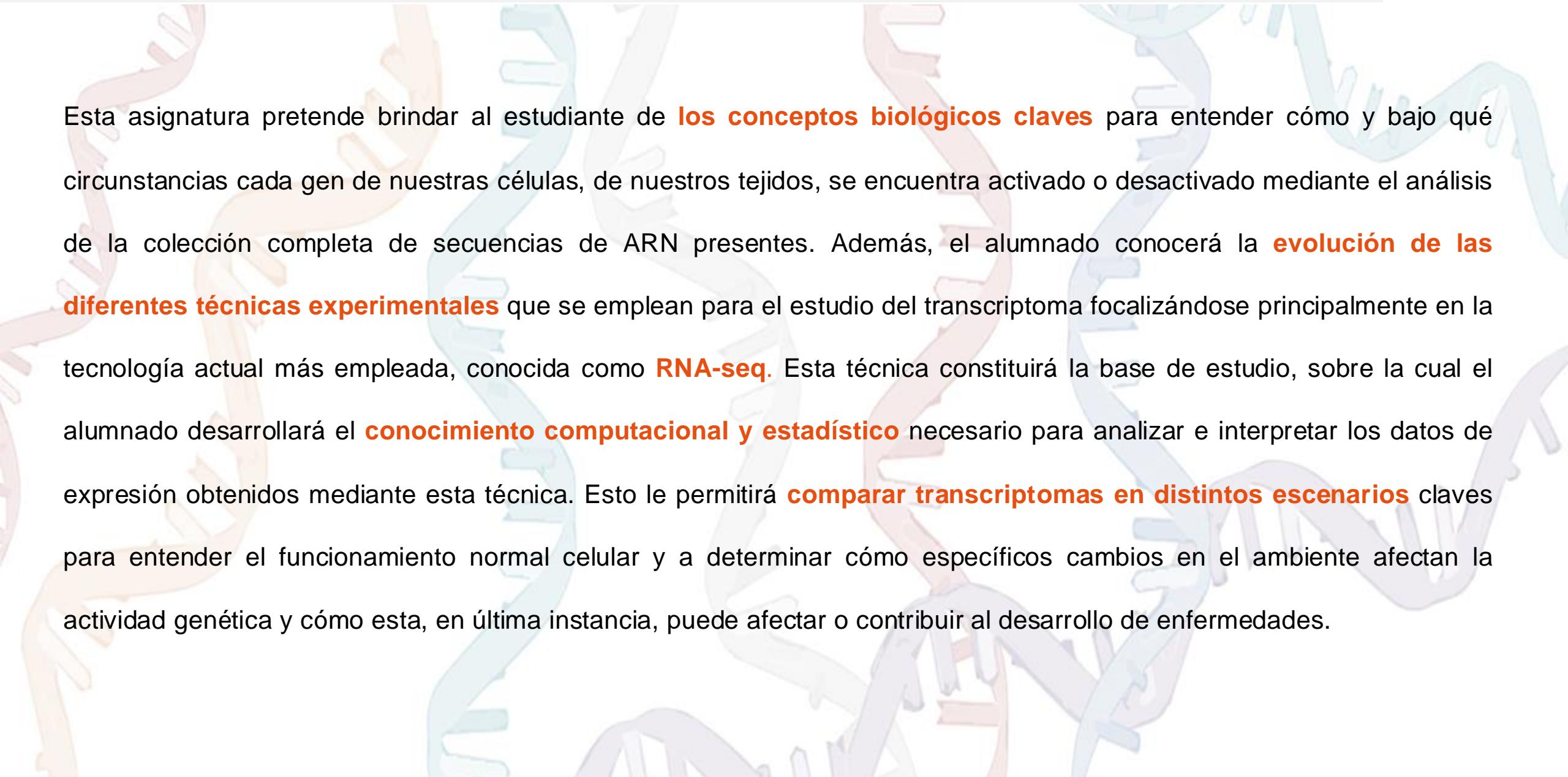
Para conocer todos los detalles de la actividad, se deberá acceder al documento denominado **Actividad 2 disponible en formato docx y en pdf.



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

Objetivos: Presentación de la asignatura



Esta asignatura pretende brindar al estudiante de **los conceptos biológicos claves** para entender cómo y bajo qué circunstancias cada gen de nuestras células, de nuestros tejidos, se encuentra activado o desactivado mediante el análisis de la colección completa de secuencias de ARN presentes. Además, el alumnado conocerá la **evolución de las diferentes técnicas experimentales** que se emplean para el estudio del transcriptoma focalizándose principalmente en la tecnología actual más empleada, conocida como **RNA-seq**. Esta técnica constituirá la base de estudio, sobre la cual el alumnado desarrollará el **conocimiento computacional y estadístico** necesario para analizar e interpretar los datos de expresión obtenidos mediante esta técnica. Esto le permitirá **comparar transcriptomas en distintos escenarios** claves para entender el funcionamiento normal celular y a determinar cómo específicos cambios en el ambiente afectan la actividad genética y cómo esta, en última instancia, puede afectar o contribuir al desarrollo de enfermedades.

Encuesta de satisfacción profesor/asignatura

Fecha de inicio: 12-09-2024 (03:00 hora peninsular española)

Fecha de fin: 19-09-2024 (23:59 hora peninsular española)



Si algún alumno/a no lo ha hecho: completad la encuesta de satisfacción docente ahora

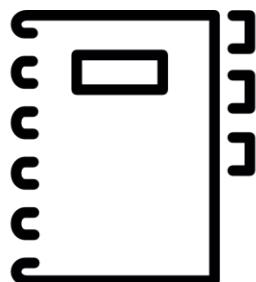
Unos minutos para completar la encuesta



Sistema de evaluación

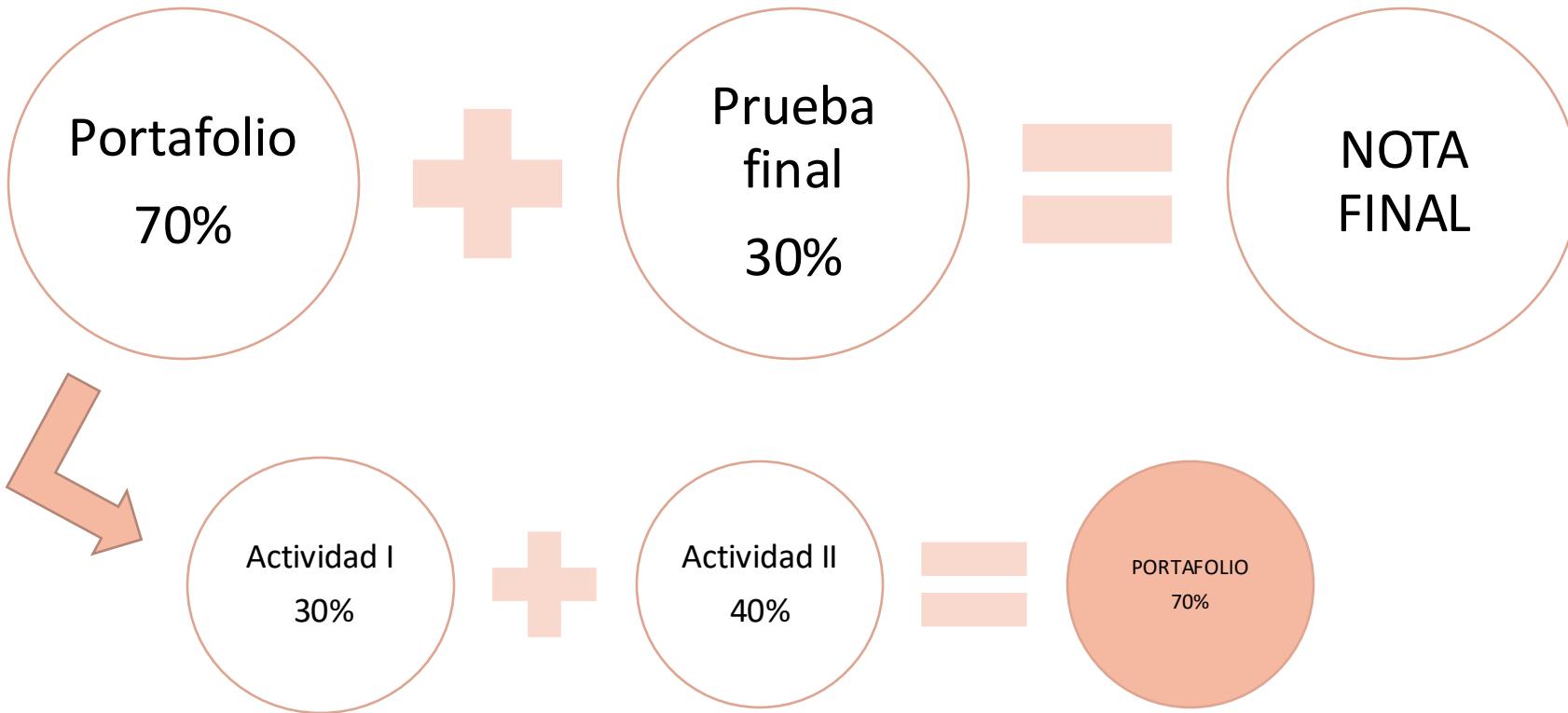


Fechas de realización de la prueba	
1ª Convocatoria	Viernes 25 de octubre del 2024 Franja A: 11:00 – 13:00 (hora peninsular española) Franja B: 19:00 – 22:00 (hora peninsular española)
2ª Convocatoria	Lunes 16 de diciembre del 2024 Franja A: 11:00 – 13:00 (hora peninsular española) Franja B: 19:00 – 22:00 (hora peninsular española)



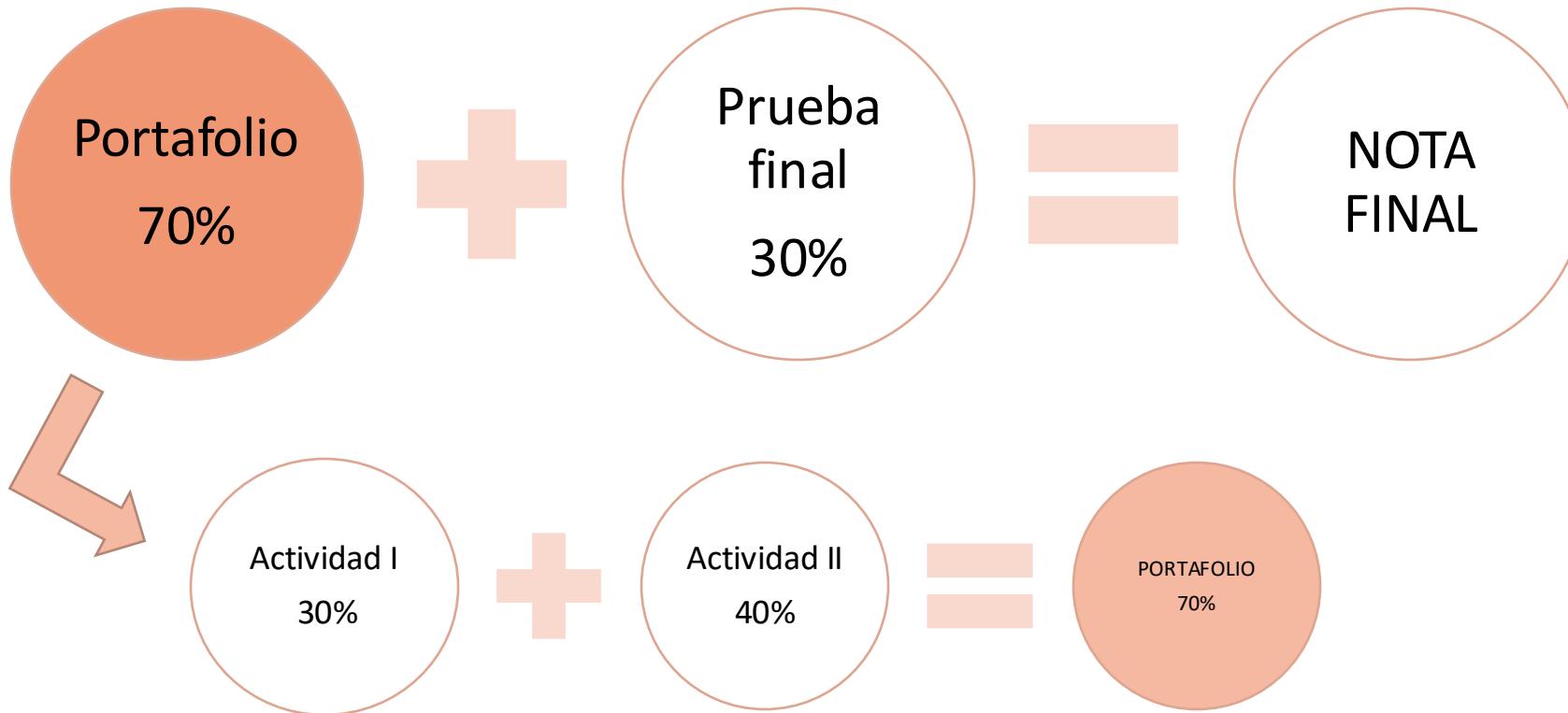
Fechas límite de entrega del portafolio	
1ª Convocatoria	Viernes 25 de octubre del 2024 a las 23:59 (hora peninsular española)
2ª Convocatoria	Lunes 16 de diciembre del 2024 a las 23:59 (hora peninsular española)

Se guarda la nota del portafolio o de la prueba final superada para segunda convocatoria



Para superar la asignatura, será necesario obtener como mínimo un **5** en cada apartado.

Nota final >= 5: no se permite ir a segunda convocatoria para “subir nota”



Para observar las **CORRECCIONES** debes acceder de nuevo a las actividades a través del apartado de “Tus Calificaciones”



CONTENIDO DE LAS CLASES + manual de la asignatura

Instrucciones generales del examen:

- Duración: 60 minutos

Características de las preguntas:

- 20 preguntas tipo test (0.4 puntos/pregunta) cada una con cuatro opciones de respuesta de las cuales sólo una es correcta + NS/NC
- 2 preguntas de respuesta breve (1 punto/pregunta)

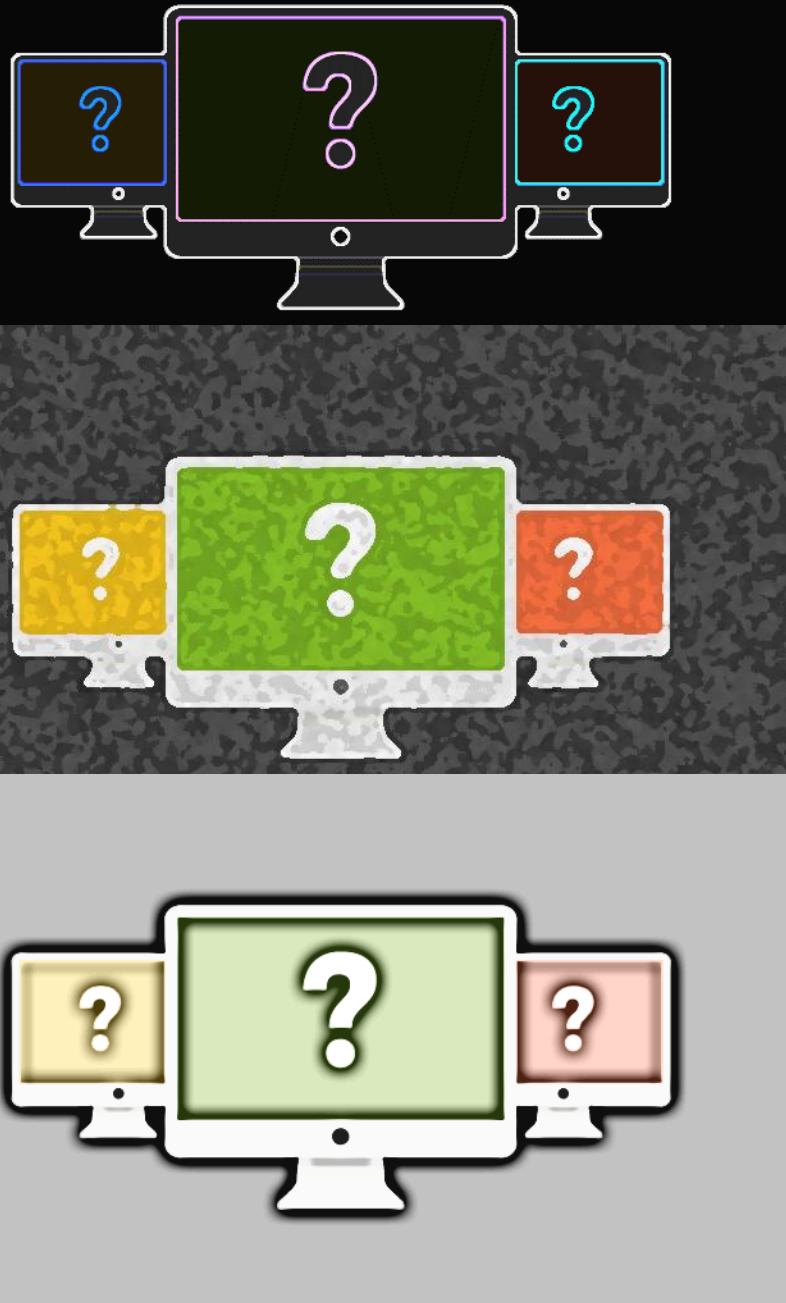
⚠ Recordad:

- Las respuestas **incorrectas** restan un 33% de la puntuación de la pregunta.
- Las respuestas en blanco o NS/NC, ni suman, ni restan.



Errores más comunes detectados

1. Mal posicionamiento de la cámara externa
2. Empleo de dos monitores
3. Comandos -> *Copy/Paste*
4. Aplicaciones abiertas:
 1. Word
 2. PowerPoint
 3. Mail
 4. WhatsApp
 5. Skype
 6. M.Teams
 7. Discord
 8. Adobe Acrobat
 9. Jerarquía de archivos/directorios



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

PRIMER CONTACTO



PSP

PlayStation®Portable

Preguntas y respuestas

P1. Acerca de los datos de recuentos de RNA-seq, ¿cuál de las siguientes afirmaciones es falsa?

1. Siguen una distribución binomial negativa.
2. Sufren una sobredispersión de la varianza.
3. Sufren una sobredispersión de la media.
4. Necesitan ser escalados para hacer comparativas.

P2. La función glmQLFTTest la aplicamos para:

1. Estimar la dispersión común en el pipeline de edgeR.
2. Estimar la dispersión de cada gen en el pipeline de edgeR.
3. Calcular la prueba de significancia o T-test.
4. Para corregir los p-values mediante el método de Benjamini-Hochberg

P3. El filtraje de genes de baja expresión:

1. Reduce el número de test estadísticos en la prueba múltiple.
2. Puede aumentar el número de genes finalmente declarados como significativos.
3. Reduce el número falsos positivos.
4. Todas son verdaderas.

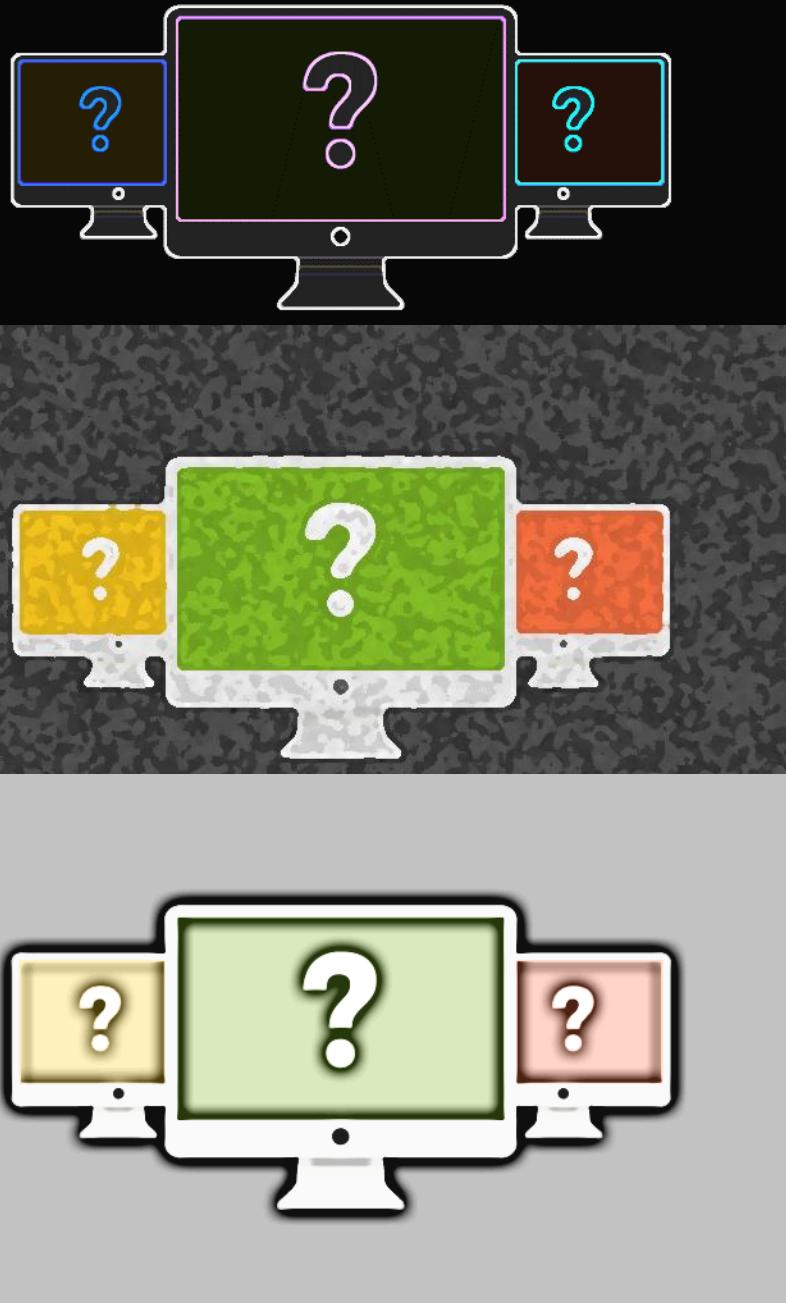
Preguntas y respuestas

P4. Cada punto en el gráfico de ...

1. Upset plot representa la dispersión de los genes respecto a la media de expresión.
2. MD representa el nivel de cambio de cada gen con respecto a la media de expresión.
3. Volcán representa el nivel de cambio de cada gen con respecto a la media de expresión.
4. Venn representa la significancia estadística de cada gen respecto a la magnitud del cambio.

P5. La Ontología de Genes se agrupa en tres ontologías que se corresponden con diversos aspectos de la biología celular y son:

1. Función molecular, proceso biológico y localización celular.
2. Función molecular, metabolismo biológico y localización celular.
3. Proceso molecular, función biológica y localización celular.
4. Proceso molecular, función biológica y alternaciones relacionadas con patologías descritas



Aspectos a tratar

- 1 Finalizar el análisis de ontología génica (*goana*)
- 2 Resolución dudas de la **Actividad 1** del portafolio
- 3 Explicación de la **Actividad 2** del portafolio
- 4 **Tutoría general:**
 - Objetivos de la asignatura
 - Encuesta de satisfacción del alumnado
 - Sistema de evaluación
 - Sistema biométrico
- 5 Preguntas y Respuestas
- 6 Curiosidades sobre *single cell RNA-seq*

Análisis de única célula

EMBL-EBI Training > On-demand training > Course materials > Single-cell RNA-seq analysis using R

Login Register



[Enter materials](#)

[Mark as favourite](#)

These materials include:

[Videos](#)

[Practicals](#)

[Slides](#)

[Published](#)
04 November 2022

[English](#)

[Contact](mailto:courses@ebi.ac.uk)
courses@ebi.ac.uk

This set of materials includes those from the 2022 course [Single-Cell RNA-seq Analysis using R](#); it covers the analysis of scRNA-seq data using R and command line tools. Lectures and practical sessions will cover both droplet-based and plate-based scRNA-seq analysis pipelines from raw reads to cell clusters. They will explore and interpret data using R as well as the Human Cell & Single Cell Expression Atlases.

Using these materials

These course materials provide a mixture of pre-recorded lectures, presentations and practicals to help advance your skills in the analysis of single-cell RNA-seq data. You may select your topic of interest from the [Course content](#) page to view the relevant materials or work your way through all the course materials.

To find out more about the trainers who created these materials, follow the links from the Course content page or go directly to the [Trainer biographies](#) page. You can also find the software requirements for the practicals in the [Technical help sheet](#).

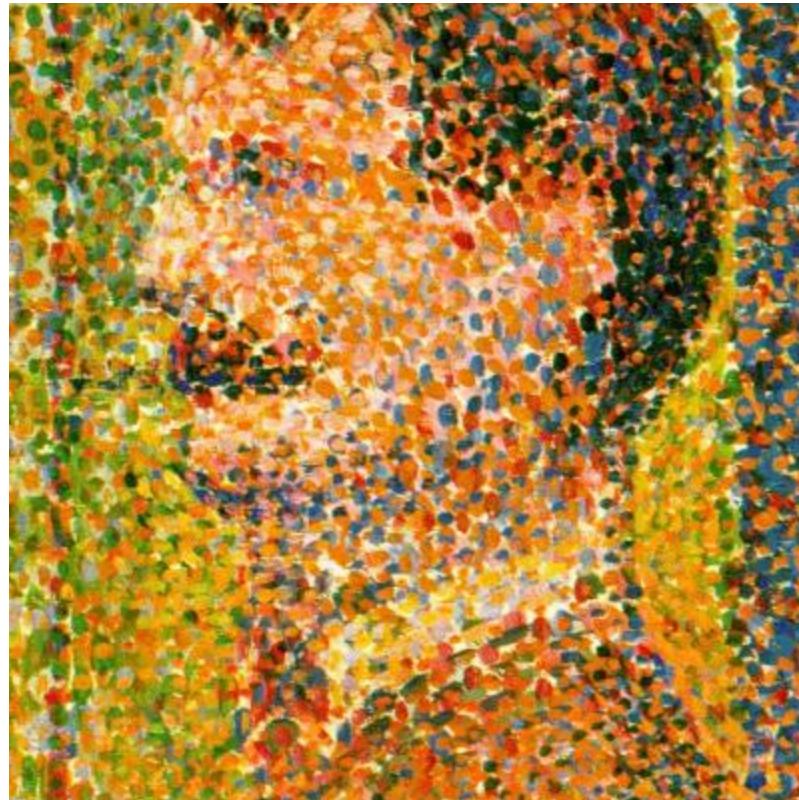
In the [Further learning](#) section you may explore the details about the EMBL-EBI's free access online tutorials and webinars on a variety of life sciences topics.

If you would like to provide feedback on this set of course materials, please use the form on the [Your feedback](#) page.

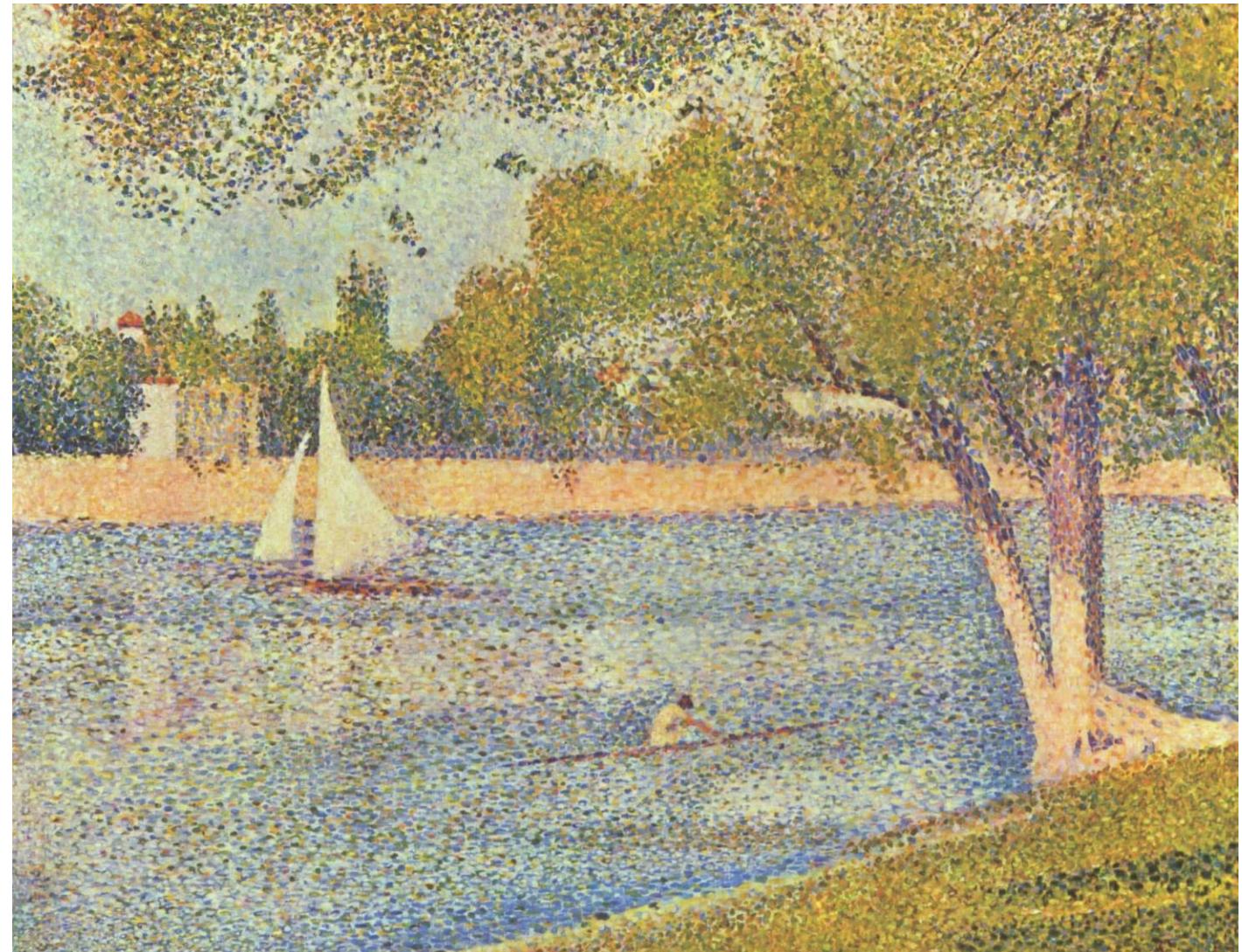
<https://www.ebi.ac.uk/training/materials/single-cell-rna-seq-analysis-using-r/>

12/09/2024

Análisis de única célula



Georges Seurat



Análisis de única célula



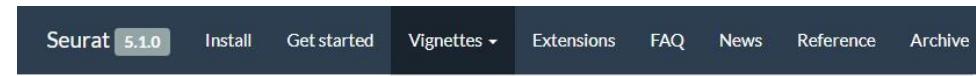
Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our [install page](#). This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

- Paper: [Dictionary learning for integrative, multimodal, and scalable single-cell analysis](#)
- Vignette: [Streamlined integration of scRNA-seq data](#)
- Vignette: [Cross-modality bridge integration](#)
- Website: [Azimuth-ATAC, reference-mapping for scATAC-seq datasets](#)



Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our [install page](#). This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

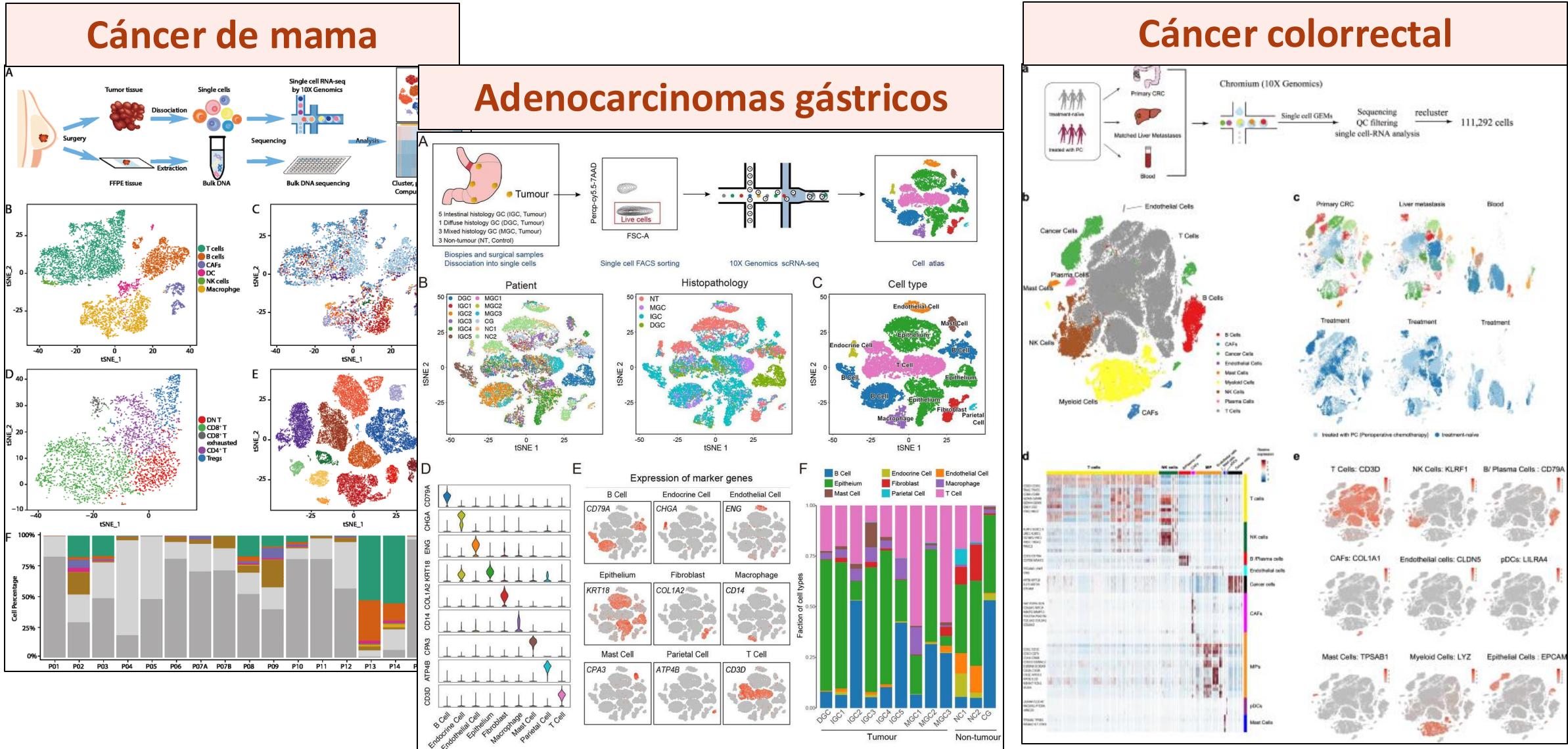
- Paper: [Dictionary learning for integrative, multimodal, and scalable single-cell analysis](#)
- Vignette: [Streamlined integration of scRNA-seq data](#)
- Vignette: [Cross-modality bridge integration](#)
- Website: [Azimuth-ATAC, reference-mapping for scATAC-seq datasets](#)

- **Flexible, interactive, and highly scalable analysis:** The speed of modern sequencing technologies is outpacing even Moore's law. In Seurat v5, we introduce 'sketch-based analysis', where representations are updated incrementally via iterative analysis - while the full dataset is being processed.

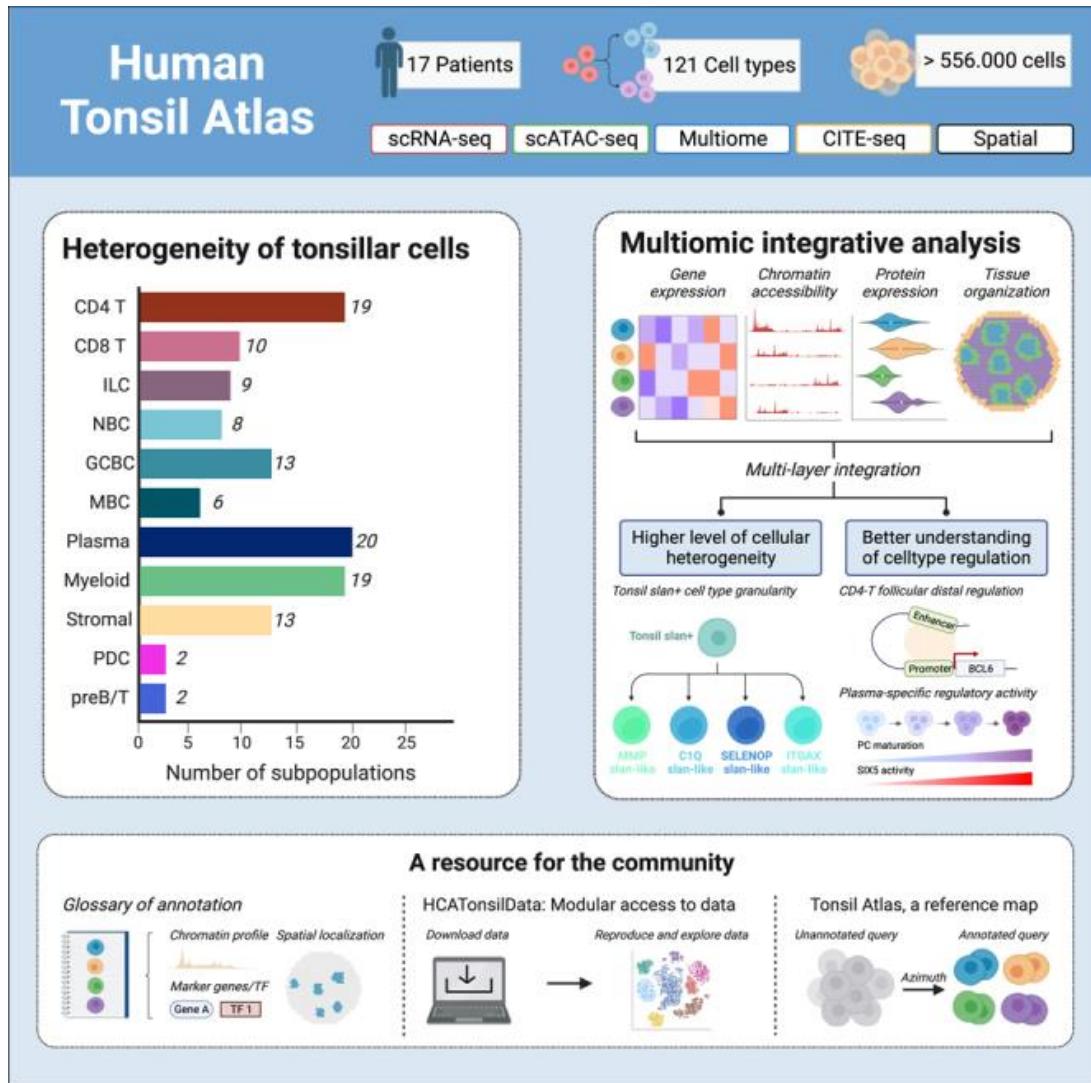
We enable high performance via the R package `Seurat`.

<https://satijalab.org/seurat/>

El atlas del cáncer



An atlas of cells in the human tonsil



ArrayExpress	updated reads spatial transcriptomics	9 months ago
CITE-seq	harmonize CITE-seq objects	2 years ago
MCL	MCL full analysis revision	10 months ago
Zenodo	corrected	4 months ago
data	fastq paths ArrayExpress	9 months ago
multome	CR multome 2	10 months ago
scATAC-seq	remove BCL2-T, not included in final dataset	10 months ago
scRNA-seq	sequencing slan cells	10 months ago
spatial_transcriptomics	fastq paths spatial	10 months ago
LICENSE	code release preprint	2 years ago
README.md	hashed libraries	3 months ago
current.Rproj	code release preprint	2 years ago
README	MIT license	

An Atlas of Cells in the Human Tonsil

Palatine tonsils are secondary lymphoid organs (SLOs) representing the first line of immunological defense against inhaled or ingested pathogens. We generated an atlas of the human tonsil composed of >556,000 cells profiled across five different data modalities, including single-cell transcriptome, epigenome, proteome, and immune repertoire sequencing, as well as spatial transcriptomics. This census identified 121 cell types and states, defined developmental trajectories, and enabled an understanding of the functional units of the tonsil. Exemplarily, we stratified myeloid slan-like subtypes, established a BCL6 enhancer as locally active in follicle-associated T and B cells, and identified SIX5 as putative transcriptional regulator of plasma cell maturation. Analyses of a validation cohort confirmed the presence, annotation, and markers of tonsillar cell types and provided evidence of age-related compositional shifts. We demonstrate the value of this resource by annotating cells from B cell-derived mantle cell lymphomas, linking transcriptional heterogeneity to normal B cell differentiation states of the human tonsil.

This repository contains all the scripts, notebooks and reports to reproduce all analysis from our manuscript entitled [An Atlas of Cells in the Human Tonsil](#), published in *Immunity* in 2024. Here, we describe how to access the data, document the most important packages and versions used, and explain how to navigate the directories and files in this repository.

<https://github.com/Single-Cell-Genomics-Group-CNAG-CRG/TonsilAtlas>



viu

Universidad
Internacional
de Valencia

universidadviu.com

De:
 Planeta Formación y Universidades