

# CAPÍTULO 1

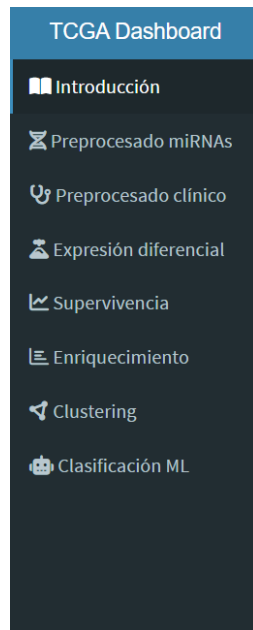
---

## Manual de usuario

---

Aquí se explicará y se mostrará con imágenes las distintas pestañas o apartados de la aplicación web desarrollada.

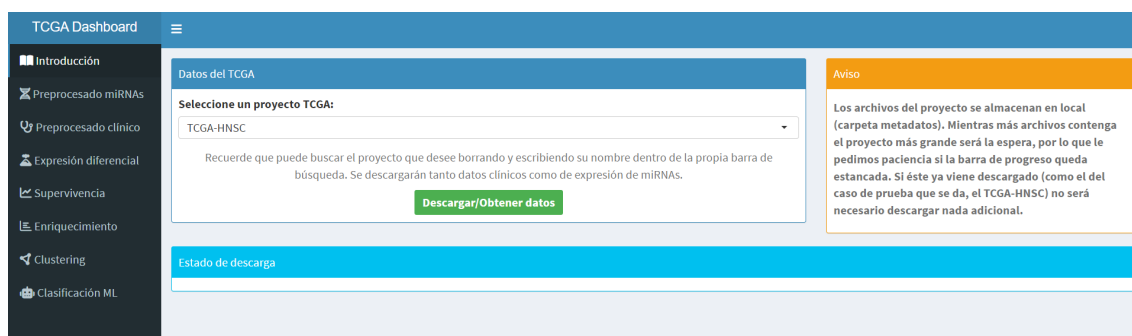
**Barra lateral.** Aparece en todas las vistas de la aplicación, aunque la podemos ocultar si pulsamos el icono de las tres barras horizontales que aparece justo al lado del nombre de la aplicación. Además del nombre, observamos una entrada para cada vista de la aplicación junto con un icono que ayuda a distinguirla.



**Figura 1.1:** Barra del menú lateral.

**Introducción.** Es la primera vista que obtenemos al entrar en la aplicación (Figura 1.2). En ella se pueden observar tres cajas principales:

- **Datos del TCGA:** donde el usuario puede seleccionar el proyecto TCGA que desee.
- **Aviso:** en color naranja, donde se explica cómo funciona el proceso de descarga y se advierte al usuario de que éste puede ser largo en ciertos casos.
- **Estado de descarga:** donde se indica si los datos están listos para su uso.

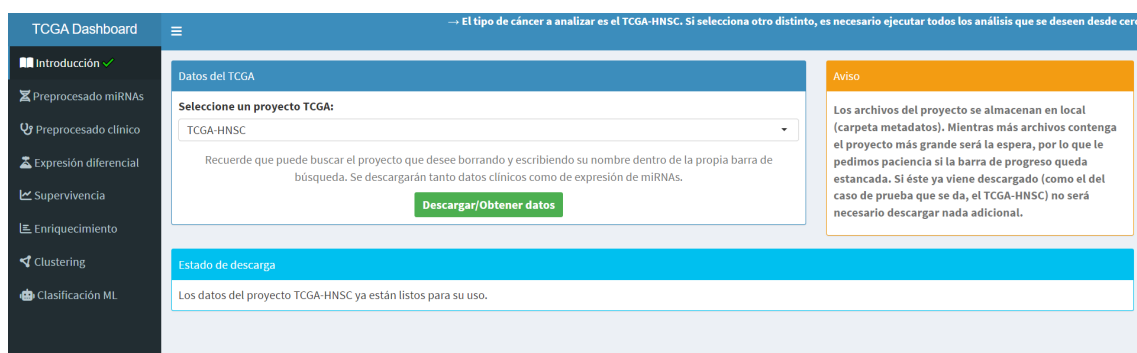


**Figura 1.2:** Pestaña de Introducción sin pulsar el botón verde.

Una vez que se pulsa el botón de descarga u obtención de datos, la vista general cambia (Figura 1.3) tal que:

- En la parte de arriba se nos muestra el tipo de cáncer o cohorte que se ha seleccionado para su uso.
- En el menú lateral se observa cómo el apartado de 'Introducción' aparece con un tick verde, indicándonos que ese paso ya se ha realizado.

y por último se observa el cambio en el estado de descarga.



**Figura 1.3:** Pestaña de Introducción una vez se pulsa el botón verde.

**Preprocesado miRNAs.** Esta vista permite al usuario ejecutar el preprocesado de los datos de expresión de miRNAs. Se distinguen dos cajas principales:

- **Parámetros a modificar en el preprocesado:** donde se permite establecer los valores deseados para los diferentes filtros.
- **Estado del preprocesado:** donde primero se le advierte al usuario de que debe descargar u obtener los datos de un proyecto antes de continuar, y después, una vez ejecutado el preprocesado, se indica el número de miRNAs que se han filtrado (ver Figura 1.5).

TCGA Dashboard

→ El tipo de cáncer a analizar es el TCGA-HNSC. Si selecciona otro distinto, es necesario ejecutar todos los análisis que se deseen desde cero.

**Parámetros a modificar en el preprocesado**

Una vez normalizados los miRNAs a lecturas mapeadas por millón (lo que quiere decir que los conteos de expresión de todos los microARNs para una sola muestra no suman uno, sino un millón), se le pide al usuario que:

**Establezca un porcentaje máximo de pacientes permitidos en los que la expresión de los miRNAs filtrados superen el valor 1 (un valor cercano al 0)**

50

Es decir, si se escoge el 50% (valor establecido por defecto) quiere decir que se rechazarán los miRNAs cuya expresión en pacientes es inferior a 1 en, al menos, el 50% del total.

**Establezca un porcentaje máximo de pacientes a los que se les permite que la expresión de los miRNAs filtrados sean un valor nulo**

10

Es decir, si se escoge el 10% (valor establecido por defecto) quiere decir que se rechazarán los miRNAs cuya expresión o conteo en la totalidad de pacientes superan el 10% de valores nulos. Se recomienda no superar el 20% para evitar introducir sesgos en la posterior imputación de estos valores.

**Establezca un umbral de la varianza a superar por cada miRNA**

0,5

Por último, se imputan los valores de expresión nulos resultantes mediante la función `impute.knn` del proyecto Bioconductor.

**Ejecutar preprocesado con estos valores**

**Estado del preprocesado**

Por ahora nada. Recuerde descargar u obtener los datos del proyecto deseado para su análisis en la ventana de 'Introducción' antes de continuar

Figura 1.4: Vista general de la pestaña Preprocesado miRNAs.

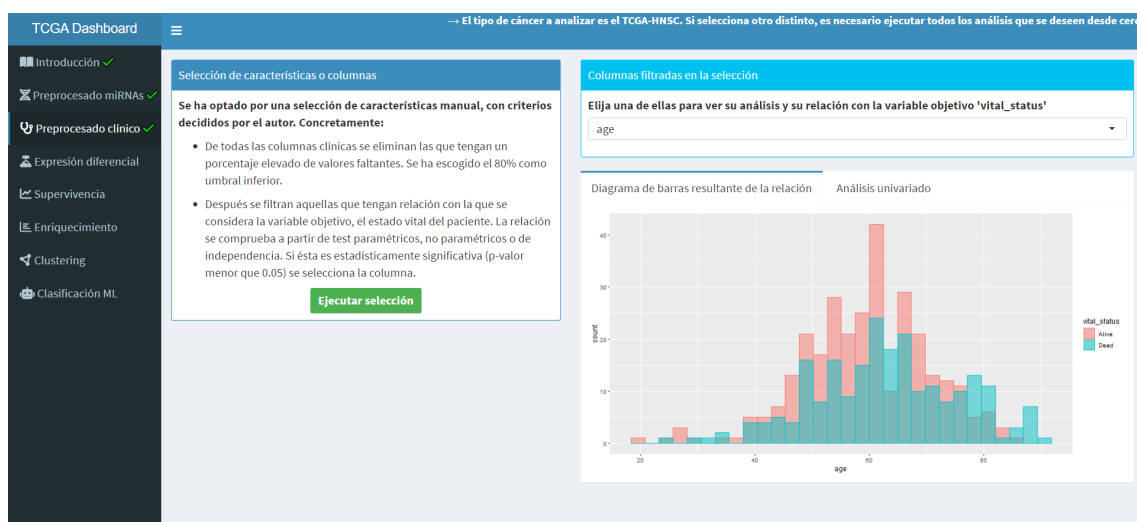
**Estado del preprocesado**

Datos preprocesados correctamente. Hay 341 miRNAs.

Figura 1.5: Caja de estado actualizada.

**Preprocesado clínico.** En esta pestaña se ejecuta el preprocesado de los datos clínicos. Se observan tres cajas principales:

- **Selección de características o columnas:** donde se comentan los criterios establecidos para ello. Aquí no es necesario que el usuario modifique nada; simplemente ejecutar el botón.
- **Columnas filtradas en la selección:** donde se muestran las columnas clínicas significativas que nos han quedado para nuestra cohorte seleccionada.
- **Diagrama de barras resultante de la relación / Análisis univariado.** En el primer apartado se observa la relación de la variable seleccionada con la variable objetivo `vital_status`. En el segundo se observa un análisis, que varía dependiendo del tipo de la variable (ver Figuras 1.6 y 1.7).



**Figura 1.6:** Vista general de la pestaña Preprocesado clínico, incluyendo el diagrama de barras correspondiente a la variable seleccionada.

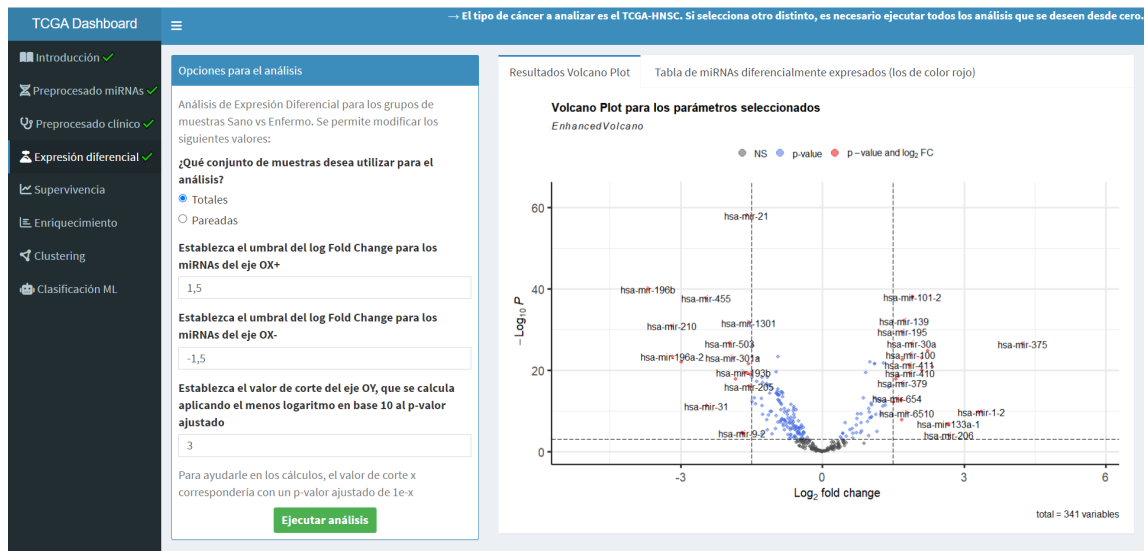


**Figura 1.7:** Vista general de la pestaña Preprocesado clínico, incluyendo el análisis univariado de la variable seleccionada.

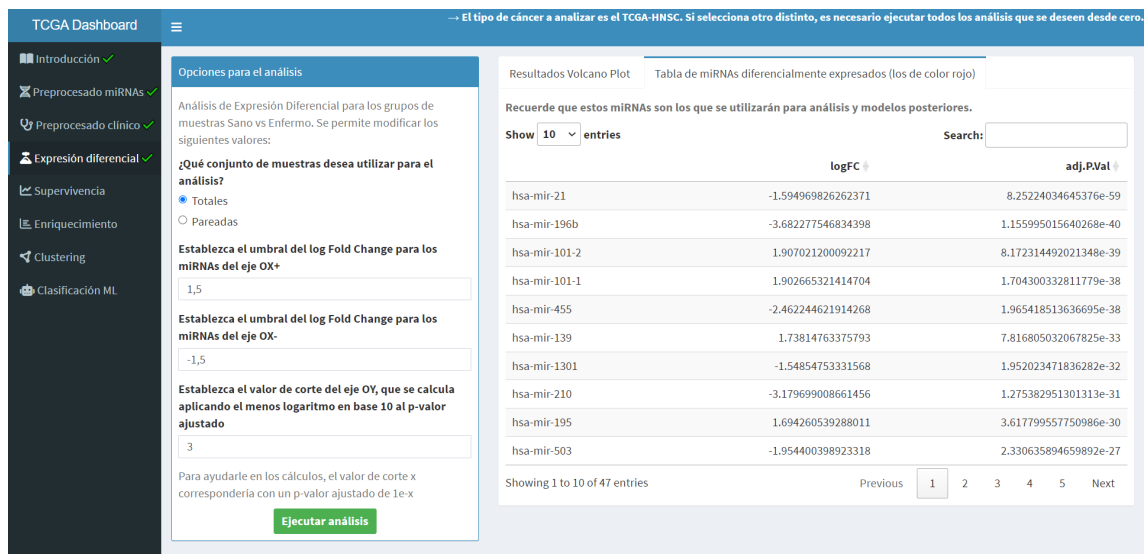
**Expresión diferencial.** Esta vista permite ejecutar el análisis del mismo nombre para los grupos de muestras sanas vs enfermas o tumorales. Se distinguen dos cajas principales:

- **Opciones para el análisis:** donde se permiten seleccionar los valores de los parámetros que actúan en él.
- **Resultados Volcano Plot / Tabla de miRNAs diferencialmente expresados.** En el primer apartado se visualiza el Volcano Plot. En el segundo se observa una tabla con los datos (logFC y p-valor ajustado) de los

miRNAs que aparecían de color rojo en el gráfico, pudiendo ordenar por columnas cómodamente (ver Figuras 1.8 y 1.9).



**Figura 1.8:** Vista general de la pestaña Expresión diferencial con los resultados del Volcano Plot.

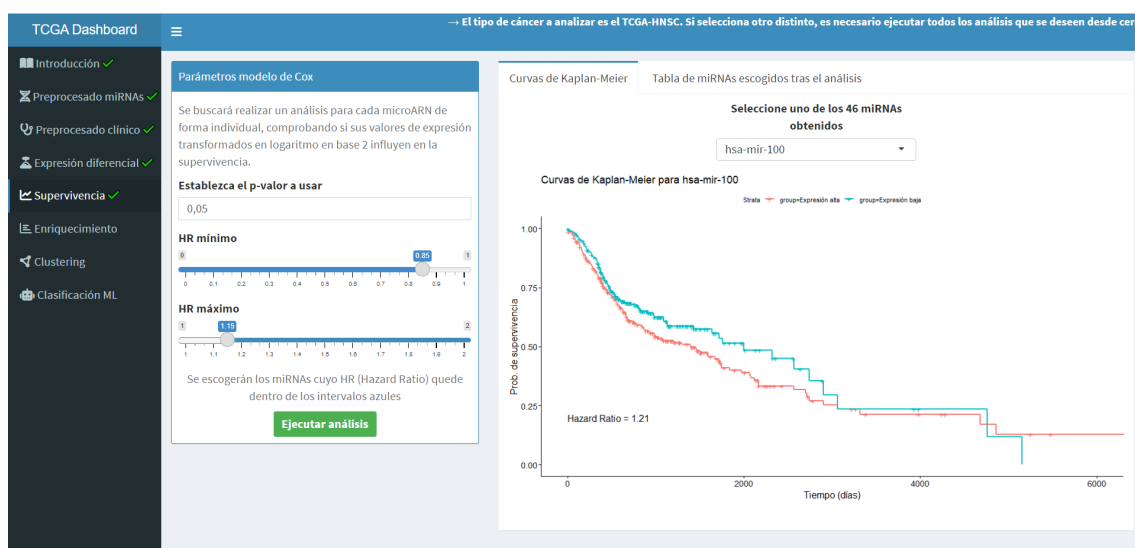


**Figura 1.9:** Vista general de la pestaña Expresión diferencial con la tabla de miRNAs diferencialmente expresados.

**Supervivencia.** En esta pestaña se permite ejecutar el análisis de Supervivencia para cada miRNA de manera individual. Se distinguen dos cajas principales:

- **Parámetros modelo de Cox:** donde se seleccionan los valores deseados.

- **Curvas de Kaplan-Meier / Tabla de miRNAs escogidos tras el análisis.** En el primer apartado se puede seleccionar cualquier miRNA de los filtrados para visualizar sus curvas del modelo según si su expresión es Alta o Baja filtrando por la mediana. En el segundo apartado se pueden consultar tanto el p-valor como el Hazard Ratio de los miRNAs filtrados, ordenando por cualquiera de las columnas (ver Figuras 1.10 y 1.11).



**Figura 1.10:** Vista general de la pestaña Supervivencia con las curvas de Kaplan-Meier para un miRNA de los filtrados.

miRNA	p-valor	hazard ratio
hsa-mir-100	0.0006134143322621976	1.213167304044155
hsa-mir-127	0.0001468218250968937	1.236010182597949
hsa-mir-134	0.0002406007525699813	1.234429336169751
hsa-mir-142	0.0004716011401437065	0.8491110251444978
hsa-mir-148b	0.02838836867267327	0.7741399861615271
hsa-mir-152	0.00786436251076758	1.219616253035368
hsa-mir-154	0.00210844398700863	1.224148509730024
hsa-mir-15a	0.04447626510485249	0.8446453664625552
hsa-mir-181c	0.008274133393288705	0.7821705739672064
hsa-mir-299	0.001838708381783802	1.170565002025008

**Figura 1.11:** Vista general de la pestaña Supervivencia con la tabla de miRNAs filtrados.

**Enriquecimiento.** Esta vista permite al usuario ejecutar el análisis de Enriquecimiento. Se distinguen dos cajas principales:

- **Genes diana / Análisis de enriquecimiento.** En el primer apartado se seleccionan los parámetros para obtener los genes diana. En el segundo apartado, se seleccionan aquellos correspondientes al análisis. En este último es donde podremos ejecutar todo el proceso tras pulsar el botón (ver Figuras 1.12 y 1.13).
- **Resultados:** donde aparece un diagrama de barras indicando las categorías más significativas, estadísticamente hablando.

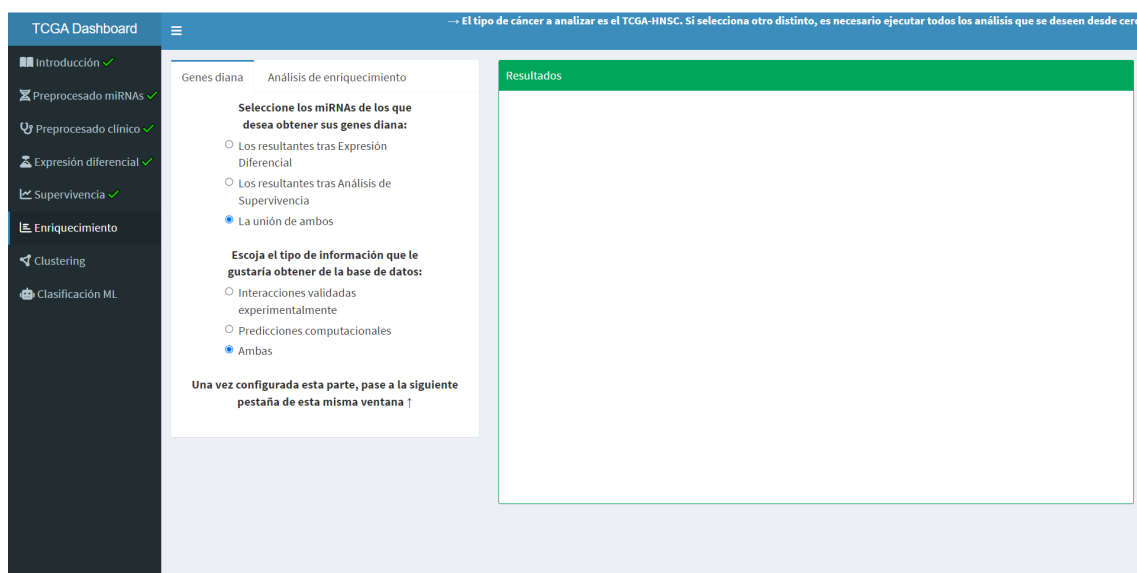


Figura 1.12: Vista general de la pestaña Enriquecimiento con las opciones para seleccionar genes diana.

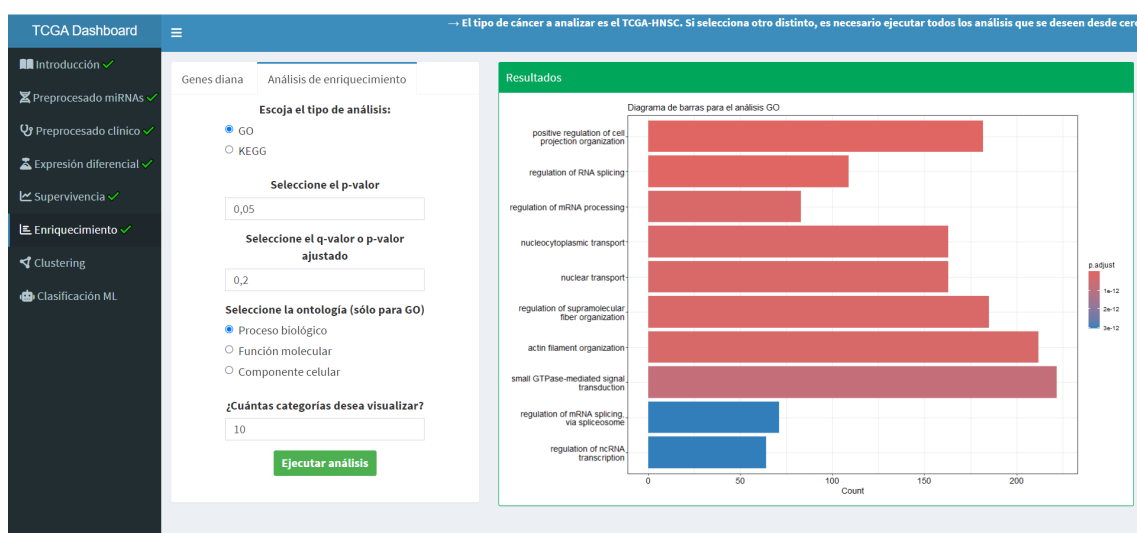
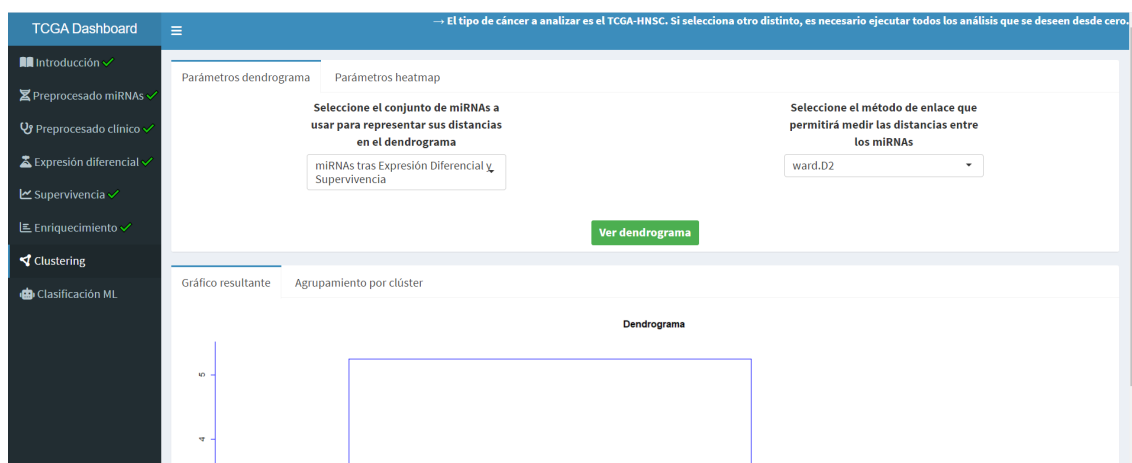


Figura 1.13: Vista general de la pestaña Enriquecimiento con las opciones del propio análisis que le da nombre, visualizando un resultado.

**Clustering.** En esta pestaña se ejecuta el modelo de clustering jerárquico. Se distinguen dos cajas principales, una arriba y otra abajo:

- En la caja de arriba encontramos dos apartados:
  - **Parámetros dendrograma:** donde se pueden seleccionar los diferentes parámetros y visualizar el dendrograma (ver Figura 1.14).
  - **Parámetros heatmap:** donde se puede seleccionar los diferentes parámetros y visualizar el mapa de calor. Una vez se ejecute este último botón, aparecerá un tick verde para este apartado de clustering del menú (ver Figura 1.15).
- En la caja de abajo se encuentran otros dos apartados:
  - **Gráfico resultante:** donde se pueden visualizar tanto el dendrograma como el mapa de calor.
  - **Agrupamiento por clúster:** aquí volvemos a distinguir varios apartados que permiten ejecutar los resultados finales del modelo.
    - \* Primero se permiten seleccionar dos parámetros: el clúster y el número de muestras.
    - \* Después nos vuelve a aparecer otra ventana con tres sub-apartados. El primero (Tabla miRNAs) permite visualizar una tabla con los miRNAs del clúster, el segundo (Tabla con las variables clínicas) permite ver estas variables para las muestras seleccionadas, y el tercero y último (Análisis univariado de variables clínicas) permite seleccionar una de estas variables y observar su distribución para el conjunto de muestras seleccionado (ver Figuras 1.16, 1.17 y 1.18).



**Figura 1.14:** Vista general de la pestaña Clustering con las opciones para el dendrograma.





Figura 1.15: Vista general de la pestaña Clustering con las opciones para el mapa de calor.

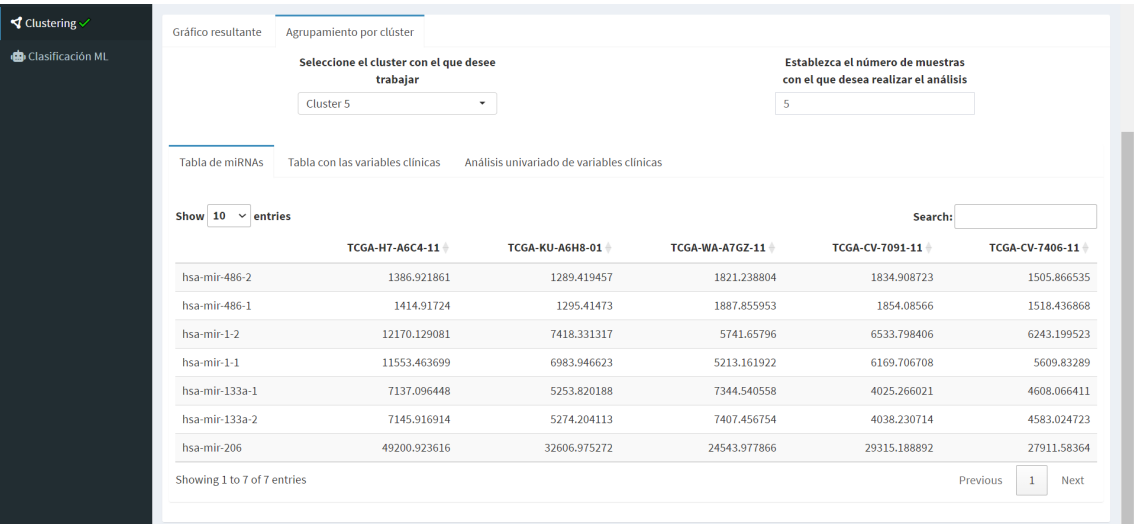


Figura 1.16: Tabla de miRNAs del mismo clúster.

Clustering ✓

Clasificación ML

Gráfico resultante Agrupamiento por clúster

Seleccione el cluster con el que desea trabajar

Cluster 5

Establezca el número de muestras con el que desea realizar el análisis

5

Tabla de miRNAs Tabla con las variables clínicas Análisis univariado de variables clínicas

Show 4 entries

Search:

	tipo_muestra	vital_status	time	age	number_of_lymphnodes_positive_by_he	gender	pathologic_stage	pathologic_T	pathologic_M
TCGA-H7-A6C4-11	sana	Dead	414	35	5	FEMALE	Stage IVA	T2	M0
TCGA-KU-A6H8-01	enferma	Dead	327	41	2	MALE	Stage IVA	T2	M0
TCGA-WA-A7GZ-11	sana	Dead	625	58	0	MALE	is_missing	T2	is_missing
TCGA-CV-7091-11	sana	Alive	3381	54	0	MALE	Stage I	T1	is_missing

Figura 1.17: Tabla con los valores clínicos de las muestras seleccionadas.

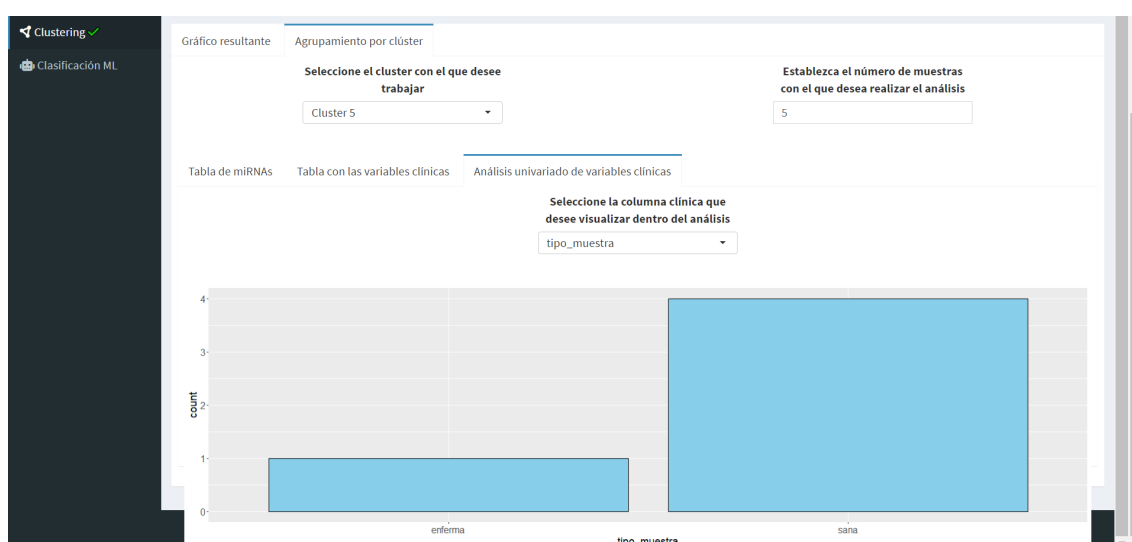
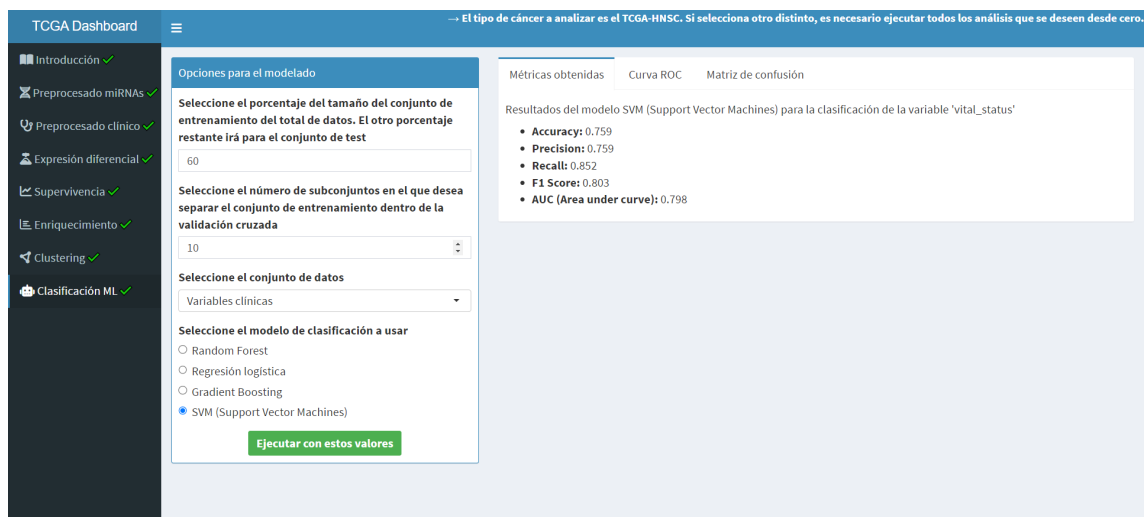


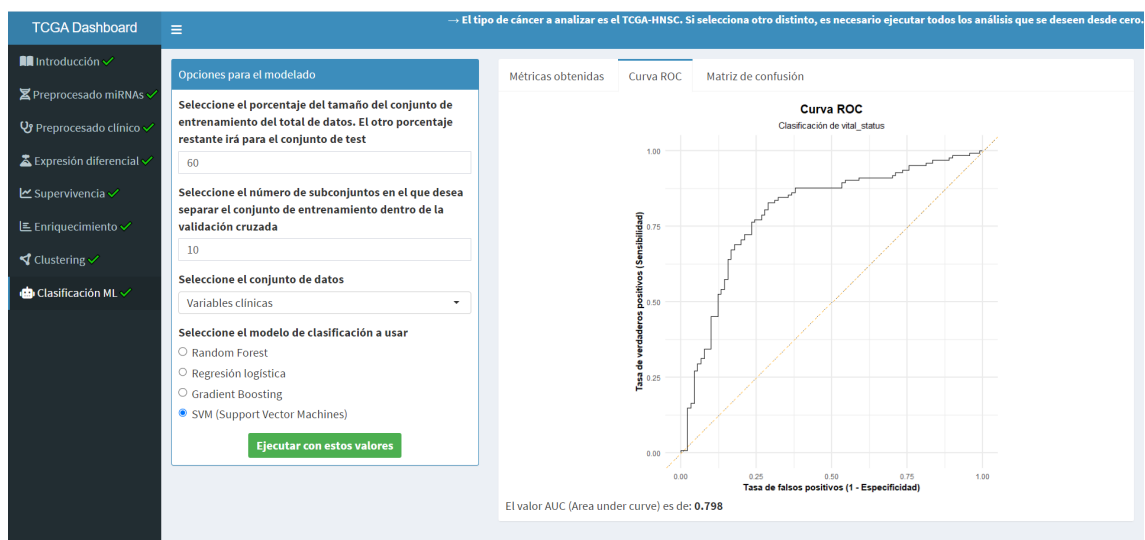
Figura 1.18: Análisis univariado para la variable clínica seleccionada.

**Clasificación ML.** En esta vista se encuentra todo lo relacionado con los modelos de clasificación de la variable `vital_status`. Se aprecian dos cajas principales:

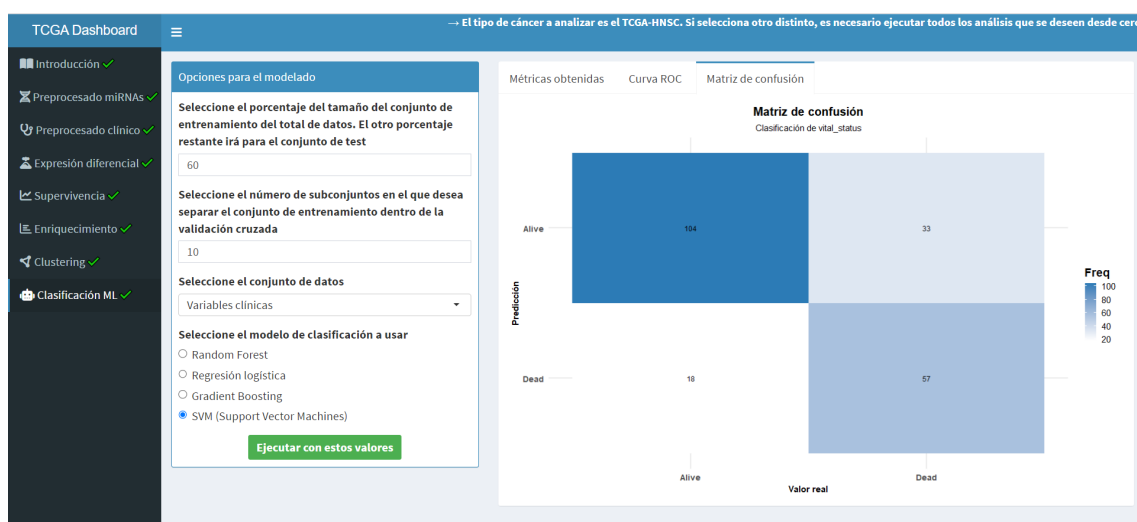
- **Opciones para el modelado:** aquí podemos seleccionar los distintos parámetros para el modelado, incluido el modelo principal que se desea aplicar.
- **Métricas obtenidas / Curva ROC / Matriz de confusión.** El primer apartado permite visualizar las métricas del modelo elegido junto con los parámetros seleccionados. El segundo apartado da como resultado la Curva ROC del modelo, y el tercer apartado muestra la matriz de confusión correspondiente (ver Figuras 1.19, 1.20 y 1.21).



**Figura 1.19:** Vista general de la pestaña Clasificación ML, donde se muestran las distintas métricas para los parámetros seleccionados.



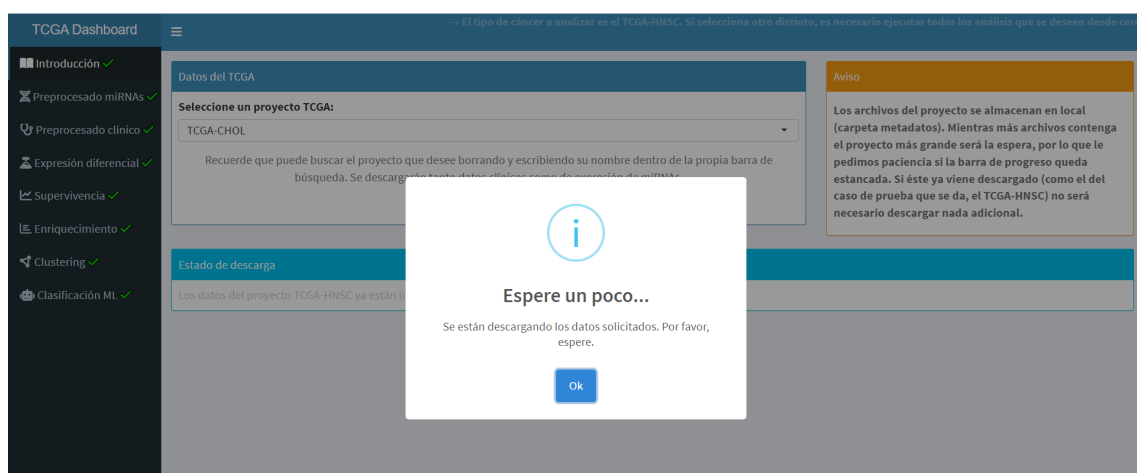
**Figura 1.20:** Vista general de la pestaña Clasificación ML, donde se muestran la curva ROC del modelo.



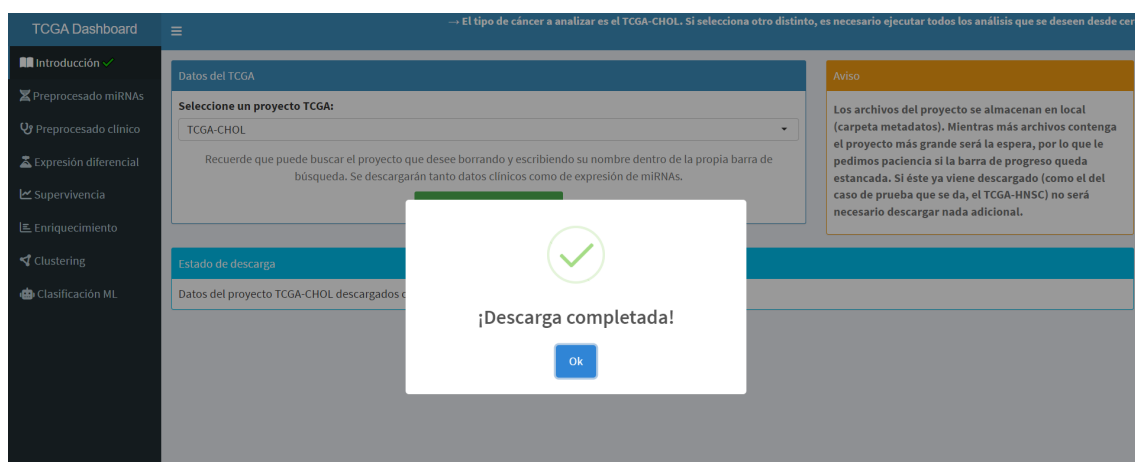
**Figura 1.21:** Vista general de la pestaña Clasificación ML, donde se muestran la matriz de confusión del modelo.

**Selección de una nueva cohorte.** Una vez terminado de realizar un análisis completo para el tipo de cáncer escogido, se da la opción de seleccionar otro tipo. Para ello, volvemos a la ventana de Introducción, seleccionamos la nueva cohorte que se desee estudiar y pulsamos el botón verde. Si no tenemos los datos descargados, aparecerá primero una alerta que nos dirá que esperemos un poco, y una vez se complete la descarga aparecerá otra que nos notificará sobre la descarga satisfactoria de los datos.

Seguidamente, observaremos que en el menú lateral no aparecen los distintos ticks verdes. Esto nos indica que para la nueva cohorte no se ha realizado nada todavía y por tanto es un nuevo análisis desde cero, algo a lo que también hace referencia el mensaje de arriba del todo.



**Figura 1.22:** Descarga de nuevos datos en curso.



**Figura 1.23:** Descarga de nuevos datos finalizada.



**Figura 1.24:** Vista general de la ventana Introducción tras la descarga de los nuevos datos.

Por último, comentar tres cosas:

1. En las pestañas que tardan más en realizar sus procesos (concretamente, los dos preprocesados y la de clasificación) aparecen alertas que indican al usuario que debe esperar un poco (como la de la Figura 1.22).
2. Cada vez que se ejecuta un nuevo análisis o técnica (sea preprocesado, análisis estadístico o modelo de ML), se vuelve a la pestaña de Introducción al tener que renderizar el menú lateral para que aparezca el tick verde al lado de la técnica seleccionada. Esto solo ocurre la primera vez que se ejecuta.
3. La aplicación está pensada para que no se puedan adelantar pasos; es decir, no se puede ejecutar el análisis de Expresión Diferencial sin ejecutar un preprocesado previo de los miRNAs (ver 1.25). Esto ocurre de igual con demás casos de la misma índole.

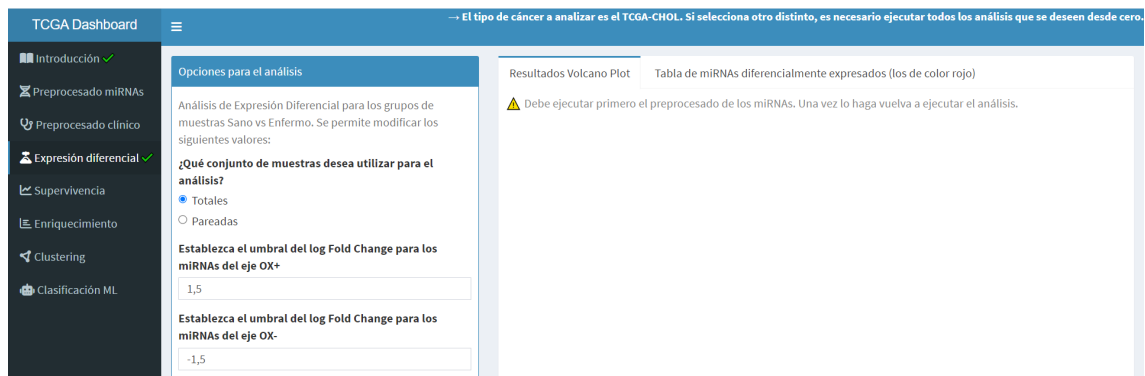


Figura 1.25: Aviso al intentar ejecutar el análisis de Expresión Diferencial.