# Introducción

### ¿Qué están intentando hacer los investigadores?

Obtener información biológica relevante sobre el cáncer de colón, como por ejemplo, genes y vías que están afectadas metabólicamente, para luego probar su hipótesis in vitro.

### ¿Cómo describen su trabajo?

Logran su objetivo mediante un análisis de datos de microarrays disponibles en línea, utilizando software como Excel, manualmente descubren relevancias biológicas observando los diferentes valores de expresión.

### ¿Dónde trabajan?

En un laboratorio del centro de investigación, con notebooks y computadoras de gama media-baja.

### ¿Qué retos enfrentan?

Resulta imposible analizar grandes cantidades de datos por restricciones de memoria y capacidad de procesamiento de software y hardware, además es complejo realizar este análisis de manera manual y lleva mucho tiempo.

### ¿Qué alternativas usan?

Ocasionalmente, optan por utilizar herramientas de software web, pero que no siempre devuelven resultados útiles o resultan muy complejas de entender y utilizar, otra alternativa que realizan es hacer el análisis de manera manual, pero incluyendo pequeñas cantidades de muestras para no sobrecargar la memoria, lo cual puede no proveer resultados satisfactorios y abarcativos.

# Requisitos funcionales

## Procesar muestras

* El sistema debe cargar archivos de metadatos de muestras en formato MAGE-TAB.
* El sistema debe permitir al usuario seleccionar muestras en particular de las cargadas desde el archivo.
* El sistema debe permitir al usuario elegir a que grupo designar cada muestra.
* El sistema debe agrupar las muestras que pertenezcan al mismo grupo y promediar los valores de intensidad de las sondas.
* El sistema debe permitir al usuario establecer un valor de error estándar máximo para el cual quitar un grupo de sondas que posean un valor superior.
* El sistema debe avisar al usuario en caso de que haya sondas en alguna muestra de un grupo que presenten valores con un error estándar mayor a la establecida y guardar los datos en un archivo.
* El sistema debe guardar el archivo con los nombres de sonda e intensidades promedio de cada grupo que hayan pasado el filtro.

## Control de calidad

* El sistema debe realizar gráficos en ciertos puntos de control del procesamiento, a fin de verificar la calidad de los datos.

## Procesar sondas

* El sistema debe cargar archivos con los nombres de sonda e intensidades de cada grupo.
* El sistema debe permitir filtrar sondas basándonos en sí la sonda mapea con múltiples genes.
* El sistema debe permitir filtrar sondas que mapean al mismo gen si su error estándar supera un umbral.
* El sistema debe permitir al usuario elegir el organismo y el umbral de error estándar.
* El sistema debe avisar al usuario en caso de que haya sondas en alguna muestra de un grupo que no sean específicos al organismo y guardar los datos en un archivo.
* El sistema debe avisar al usuario en caso de que haya genes de un grupo que presenten valores de error estándar mayores a la establecida y guardar los datos en un archivo.
* El sistema debe guardar el archivo con los nombres de genes e intensidades de cada grupo, también debe incluir los nombres de los que fueron filtrados por alto error.

## Procesar genes

* El sistema debe cargar archivos mapeados que contengan el nombre del gen y las intensidades por grupo.
* El sistema debe permitir al usuario seleccionar el grupo de control para realizar los modelos para el cálculo de expresión diferencial.
* El sistema debe permitir al usuario establecer un valor de p-value máximo para el cual se filtran los genes que posean un valor superior.
* El sistema debe calcular p-value y log2foldchange de los genes.
* El sistema debe permitir filtrar genes si su p-value supera un umbral.
* El sistema debe avisar al usuario en caso de que haya genes de un grupo que presenten valores de p-value mayores al establecido y guardar los datos en un archivo.
* El sistema debe enriquecer los genes con términos GO, vías que afectan, enfermedades asociadas y otra información complementaria, desde bases de datos curadas.
* El sistema debe guardar el archivo con los nombres de genes, intensidades y enriquecimientos, también debe incluir los nombres de los que fueron filtrados.

## Generar resultados

* El sistema debe permitir cargar múltiples resultados para permitir un metanálisis entre distintos estudios.
* El sistema debe poder generar los siguientes resultados de expresión:
  + Cantidad genes y sondas sobre expresadas por nivel.
  + Gráfico de barras de cada gen con valores de expresión de la muestra control y de las de tratamiento.
  + Tabla de sobre expresión de genes mostrando los genes y agrupado por nivel de sobre expresión.
  + Tabla de cantidad de sondas por gen.
  + Volcano plot con los valores log2foldchange.
* El sistema debe generar los siguientes resultados de vías:
  + Gráfico de vías (Fireworks), similar al que provee Reactome.
  + Gráfico de barras que muestre vías sobreexpresadas con p-value.
* El sistema debe generar los siguientes resultados de metanálisis:
  + Un diagrama de Venn que muestre que compare que genes se encuentran diferenciados en los estudios.
  + Heatmap de valores de intensidad de cada gen por grupo.
* El sistema debe generar los siguientes resultados de un gen:
  + Su página en Expression atlas.
  + El pathway asociado a ese gen.
* El sistema debe generar los siguientes resultados basándose en los términos GO:
  + Mosaic plot de las funciones de los genes.

## Avisar (loguear)

* El sistema debe avisar al usuario cada vez que una operación se completa con éxito, generó un aviso o falló.

# 

# Requisitos no funcionales

| **Requerimiento** | **Definición** | **Valoración (1-3)** |
| --- | --- | --- |
| Eficiencia de desempeño | |  |
| Comportamiento temporal | Los tiempos de respuesta y procesamiento y los ratios de throughput de un sistema cuando lleva a cabo sus funciones bajo condiciones determinadas en relación con un banco de pruebas (benchmark) establecido. | 2 |
| Utilización de recursos | Las cantidades y tipos de recursos utilizados cuando el software lleva a cabo su función bajo condiciones determinadas. | 2 |
| Capacidad | Grado en que los límites máximos de un parámetro de un producto o sistema software cumplen con los requisitos. | 3 |
| Compatibilidad | |  |
| Coexistencia | Capacidad del producto para coexistir con otro software independiente, en un entorno común, compartiendo recursos comunes sin detrimento. | 1 |
| Interoperabilidad | Capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada. | 3 |
| Usabilidad | |  |
| Aprendizabilidad | Capacidad del producto que permite al usuario aprender su aplicación. | 3 |
| Operabilidad | Capacidad del producto que permite al usuario operarlo y controlarlo con facilidad. | 3 |
| Protección de errores | Capacidad del sistema para proteger a los usuarios de hacer errores. | 2 |
| Estética | Capacidad de la interfaz de usuario de agradar y satisfacer la interacción con el usuario. | 1 |
| Fiabilidad | |  |
| Recuperabilidad | Capacidad del producto software para recuperar los datos directamente afectados y restablecer el estado deseado del sistema en caso de interrupción o fallo. | 2 |
| Mantenibilidad | |  |
| Modularidad | Capacidad de un sistema o programa de ordenador (compuesto de componentes discretos) que permite que un cambio en un componente tenga un impacto mínimo en los demás. | 3 |
| Reusabilidad | Capacidad de un activo que permite que sea utilizado en más de un sistema software o en la construcción de otros activos. | 2 |
| Analizabilidad | Facilidad con la que se puede evaluar el impacto de un determinado cambio sobre el resto del software, diagnosticar las deficiencias o causas de fallos en el software, o identificar las partes a modificar. | 3 |
| Modificabilidad | Capacidad del producto que permite que sea modificado de forma efectiva y eficiente sin introducir defectos o degradar el desempeño. | 3 |
| Testeabilidad | Facilidad con la que se pueden establecer criterios de prueba para un sistema o componente y con la que se pueden llevar a cabo las pruebas para determinar si se cumplen dichos criterios | 2 |
| Baratura de mantenimiento | Facilidad y bajo costo de recursos económicos para tener al software en funcionamiento | 3 |
| Portabilidad | |  |
| Adaptabilidad | Capacidad del producto que le permite ser adaptado de forma efectiva y eficiente a diferentes entornos determinados de hardware, software, operacionales o de uso. | 3 |
| Instalabilidad | Facilidad con la que el producto se puede instalar y/o desinstalar de forma exitosa en un determinado entorno | 3 |
| Reemplazabilidad | Capacidad del producto para ser utilizado en lugar de otro producto software determinado con el mismo propósito y en el mismo entorno. | 3 |

# 

# Requisitos de la interfaz externa

## Interfaces de usuarios

* El sistema permitirá al usuario organizar de manera ascendente o descendente, por cada columna de la tabla, las filas de los datos.
* El sistema permitirá al usuario manualmente suprimir filas de las tablas de datos
* El sistema permitirá al usuario reordenar las columnas de las tablas de datos.
* El sistema permitirá filtrar los datos mediante un filtro desplegable en la cabecera de cada columna.
* El sistema preguntará siempre al usuario directorio de guardado y archivos a cargar mediante una ventana emergente.
* Los resultados se abrirán en una ventana nueva con un tamaño de píxeles igual a la pantalla completa.
* Las ventanas que solicitan información adicional al usuario, como un directorio, deben bloquear la ventana principal hasta que la información haya sido emitida.

## Interfaces de software

Se utilizarán múltiples API, tanto de manera directa, como indirecta mediante librerías de Python. Ellas son:

g:Profiler: Mediante la librería de Python gprofiler.

Rectome: Mediante la librería rectome2py de Python, y luego con PyQtWebEngine para mostrar su página en un navegador web de ventana.