# 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

## Convertir datos

* El sistema debe convertir un archivo de muestra con datos de microarrays de diferentes compañías a un formato estándar que contenga el nombre de la sonda, intensidad y otros metadatos de la muestra a la que pertenecen.
* El sistema debe guardar el archivo con formato estándar de manera local.

## Procesar muestras

* El sistema debe cargar archivos de muestras con el formato estándar.
* El sistema debe permitir al usuario seleccionar muestras en particular de las cargadas desde los archivos.
* 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 varianza mínimo de aviso entre valores de sonda de las muestras de un mismo grupo.
* El sistema debe avisar al usuario en caso de que haya sondas en alguna muestra de un grupo que presenten valores con una varianza mayor a la establecida.
* El sistema debe guardar el archivo con los nombres de sonda e intensidades promedio de cada grupo.

## Procesar sondas

* El sistema debe cargar archivos con los nombres de sonda e intensidades de cada grupo.
* El sistema debe mapear cada sonda con su gen, vía y términos GO.
* El sistema debe permitir filtrar sondas.
* El sistema debe guardar el archivo con los nombres de sonda, gen, vía e intensidades de cada grupo.

## Filtrar sondas

* El sistema debe permitir al usuario filtrar las sondas por organismo.
* El sistema debe permitir al usuario filtrar las sondas por p-value.
* El sistema debe permitir al usuario filtrar las sondas por un umbral de error estándar.
* El sistema debe avisar al usuario que sondas no pasaron el filtro y guardarlas en un archivo separado.

## Pre procesar resultados

* El sistema debe cargar archivos mapeados que contengan el nombre de la sonda, gen, vía, términos GO y las intensidades por grupo.
* El sistema debe permitir al usuario seleccionar un gen o vía en particular para los resultados.
* El sistema debe permitir al usuario seleccionar que muestra se utilizará como control y que muestra o muestras se utilizarán como tratamiento para los resultados.
* El sistema debe permitir al usuario elegir el tipo de gráfico o tabla a realizar.

## Generar resultados

* El sistema debe poder generar los siguientes resultados de expresión:
  + Cantidad genes y sondas sobre expresadas por nivel.
  + Heatmap de valores de intensidad de cada sonda por muestra.
  + Heatmap de valores de intensidad de cada gen por muestra.
  + 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 y cantidad de sondas totales con valores en los datos.
* El sistema debe generar los siguientes resultados de vías:
  + Tabla de sobre expresión de vías general, con el nivel de sobre activación, el nombre y p-value.
  + Gráfico de vías, similar al que provee Reactome.
  + Gráfico de barras que muestre cantidad de sondas sobre expresadas por vía.
* El sistema debe generar los siguientes resultados de expresión y vías:
  + Tabla de genes y vías que incluya genes, cantidad de total por vía y el nivel de sobre expresión.

## 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 mostrará al usuario en formato tabla los datos en cada pestaña de procesamiento.
* El sistema permitirá al usuario organizar de manera ascendente o descendente por cada columna de la tabla las filas de los datos.
* Los resultados se abrirán en una ventana nueva con un tamaño de píxeles igual a la pantalla completa.
* En todas las pestañas el log con los avisos se encontrará debajo indicando el estado actual de los procedimientos.
* La interfaz del sistema debe mostrar clara y fácilmente los metadatos de las muestras.

## Interfaces de software

Se contempla la posibilidad de que el sistema se comunique mediante API a otros sistemas para realizar algunos de sus procedimientos, pero no hay ningún requisito de este tipo.