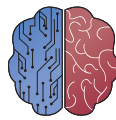




UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería de la Salud



INGENIERÍA  
DE LA SALUD

**TFG del Grado en Ingeniería de la  
Salud**

**Análisis bioinformático de los  
genes implicados en el cáncer  
hepático  
Documentación Técnica**

Presentado por Ana Paula Cuesta Asín  
en Universidad de Burgos

9 de julio de 2024

Tutores: Rubén Ruiz González – Antonia Maiara  
Marques do Nascimento



## Índice general

# I

|  |           |
|--|-----------|
| E.1. Cuaderno de trabajo . . . . .                             | 25        |
| E.2. Configuración y parametrización de las técnicas . . . . . | 26        |
| E.3. Detalle de resultados . . . . .                           | 27        |
| <b>Apéndice F Anexo de sostenibilización curricular</b>        | <b>29</b> |
| F.1. Introducción . . . . .                                    | 29        |
| F.2. Reflexión Personal sobre la Sostenibilidad . . . . .      | 29        |
| F.3. Competencias de Sostenibilidad Adquiridas . . . . .       | 30        |
| F.4. Conclusión . . . . .                                      | 31        |
| <b>Bibliografía</b>  | <b>33</b> |

---

# Índice de figuras

---

|  |    |
|--|----|
| C.1. Estructuración de los directorios del proyecto. |    |
| Fuente: Elaboración propia . . . . .                 | 10 |

---

# Índice de tablas

---

|   |    |
|---|----|
| D.1. CU-1 Consultar Experimentos. . . . . | 21 |
| D.2. CU-2 Añadir Nuevos Datos. . . . .    | 22 |
| D.3. CU-3 Generar Reportes. . . . .       | 23 |

## Apéndice A

---

# Plan de Proyecto Software

---

### A.1. Planificación temporal

La planificación temporal del proyecto incluye diversas fases, desde la recopilación de datos hasta el análisis y la presentación de resultados. Cada fase tiene un conjunto de tareas específicas con fechas de inicio y fin estimadas.

**- Recopilación de Datos y Búsqueda de Bibliografía (1 mes):**

1. Descarga de datos de expresión génica y variantes genéticas.
2. Preparación de archivos de datos para el análisis.

Fecha de inicio: 1 de enero de 2024 Fecha de fin: 31 de enero de 2024

**- Preprocesamiento de Datos (1 mes):**

1. Normalización y limpieza de datos.
2. Filtrado de datos relevantes para el análisis.

Fecha de inicio: 1 de febrero de 2024 Fecha de fin: 29 de febrero de 2024

**- Análisis de Expresión Diferencial (1 mes):**

1. Implementación de análisis utilizando el paquete limma.
2. Identificación de genes diferencialmente expresados.

Fecha de inicio: 1 de marzo de 2024 Fecha de fin: 30 de marzo de 2024

**- Análisis de Variantes Genéticas (1 mes):**

1. Análisis de variantes en genes clave (CTNNB1 y TP53).

2. Interpretación de resultados y validación cruzada.

Fecha de inicio: 1 de abril de 2024 Fecha de fin: 30 de abril de 2024

**- Visualización de Resultados (1 mes):**

1. Generación de gráficos y diagramas para la representación de datos.

2. Preparación de figuras para la presentación de resultados.

Fecha de inicio: 1 de mayo de 2024 Fecha de fin: 31 de mayo de 2024

**- Redacción del Informe Final (1 mes):**

1. Redacción del informe final del proyecto.

2. Revisión y edición del documento.

Fecha de inicio: 1 de junio de 2024 Fecha de fin: 5 de julio de 2024

## Planificación económica

La planificación económica del proyecto incluye los costos asociados a cada fase del proyecto, desde la adquisición de datos hasta la presentación final de los resultados.

**Costos de Hardware:**

Adquisición de equipos de computación y almacenamiento.

Costo estimado: 2,000 euros

**Costos de Software:**

Licencias para herramientas bioinformáticas (si es necesario).

Costo estimado: 500 euros

**Costos de Personal:**

Salarios del equipo de investigación (6 meses).

Costo estimado: 15,000 euros

**Costos de Publicación y Presentación:**

Publicación en revistas científicas y presentación en conferencias.

Costo estimado: 2,000 euros

**Gastos Misceláneos:**

Otros gastos relacionados con el proyecto (transporte, suministros, etc.).



Costo estimado: 1,000 euros

**Total estimado del proyecto: 20,500 euros**

## Viabilidad legal

La viabilidad legal del proyecto se asegura mediante el cumplimiento de todas las regulaciones y normativas aplicables en cada etapa del proyecto.

### **Cumplimiento de Normativas Éticas:**

Asegurar que todas las prácticas de investigación cumplan con los estándares éticos establecidos, para ello se debe obtener aprobación ética para el uso de datos genéticos y biológicos.

### **Protección de Datos Personales:**

Garantizar la privacidad y protección de los datos personales utilizados en el proyecto. Para llevar a cabo esto es necesario aplicar medidas de anonimización y cumplir con la normativa GDPR.

### **Propiedad Intelectual:**

Respetar y proteger los derechos de propiedad intelectual asociados al uso de datos y herramientas bioinformáticas, para asegurar el uso adecuado de licencias y reconocer las fuentes de datos y software.

### **Publicación de Resultados:**

Cumplir con las políticas de acceso abierto y las regulaciones de publicación científica. Para llevar a cabo esto es necesario publicar los resultados en revistas de acceso abierto y proporcionar acceso a los datos subyacentes.



## *Apéndice B*

---

# Documentación de usuario

---

### B.1. Requisitos software y hardware para ejecutar el proyecto.

A continuación vamos a detallar los requisitos específicos que son necesarios para la realización del presente proyecto.

#### Hardware

: A continuación vamos a enumerar los elementos de Hardware que necesitamos para nuestro proyecto.

- **Procesador:** El procesador debe ser de tipo Intel Core i5 o superior, ya que este tipo de procesadores proporcionan un rendimiento adecuado para manejar grandes volúmenes de datos y realizar cálculos intensivos como los del proyecto.

- **Memoria RAM:** Esta tarjeta de memoria debe contener 8 GB como mínimo, aunque se recomiendan tarjetas con 16 GB de memoria, ya que los datos pueden llegar a ocupar gran parte de la memoria de estas. Esto se debe a que una cantidad adecuada de memoria es crucial para procesar datos de expresión génica sin problemas.

- **Espacio en Disco:** Se necesitan al menos 50 GB de espacio libre para almacenar los datos descargados y los resultados del análisis.

- **Sistema Operativo:** Este Sistema Operativo (S.O.) debe ser Windows 10, macOS 10.15, o Linux (Ubuntu 20.04 LTS) para asegurar la compatibili-

dad con el software requerido, ya que si utilizamos versiones anteriores se puede ver comprometida la compatibilidad de estas con el software utilizado.

## Software

: Entre los requisitos de Software que necesitamos para el desarrollo del proyecto se encuentran los siguientes:

- **Lenguaje R:** Este lenguaje de programación debe presentar una versión 4.0.0 o superior, ya que esta versión es crucial para la ejecución de scripts de análisis de expresión génica.

- **Python:** Este lenguaje debe presentar una versión 3.8 o superior, la cual es requerida y utilizada para el procesamiento adicional de datos y visualización. En versiones anteriores puede que se limiten algunas de las funcionalidades.

- **RStudio:** Una versión 1.4 o superior es estrictamente necesaria, ya que en caso contrario, este entorno de desarrollo integrado (IDE) para R puede verse comprometido a la hora de mostrar todas las funcionalidades del paquete BioConductor.

- **Jupyter Notebook:** Este editor de código se encuentra incluido en la distribución de Anaconda, para ejecutar y documentar el código Python en un entorno interactivo.

- **Paquetes de R:** Se requieren los paquetes GEOquery, limma, hgu133a2.db, y hgu133a.db para la descarga y análisis de datos de expresión génica, los cuales se encuentran almacenados en el proyecto BioConductor.

- **Bibliotecas de Python:** Se necesitan Biopython, requests, pandas y numpy, entre otras para el procesamiento adicional y análisis de datos, así como para la obtención de información acerca de los genes desde la API de Ensembl.

## B.2. Instalación / Puesta en marcha

### Instalación de R y RStudio

:

## Descargar e Instalar R

: Para realizar la descarga e instalación de este lenguaje de programación se deben seguir los siguientes pasos: 1. Visitar el sitio web de [CRAN](<https://cran.r-project.org/>). 2. Seleccionar la versión adecuada para su sistema operativo y seguir las instrucciones de instalación.

## Descargar e Instalar RStudio

: 1. Acceder a [RStudio](<https://rstudio.com/products/rstudio/download/>). 2. Descargar el instalador para su sistema operativo y seguir las instrucciones de instalación que aparecen por pantalla.

## Instalación de Python y Jupyter Notebook:

### Descargar e Instalar Anaconda:

1. Visite la página de [Anaconda](<https://www.anaconda.com/products/individual>). 2. Descargue el instalador correspondiente a su sistema operativo. 3. Siga las instrucciones del asistente de instalación.

### Instalar Jupyter Notebook:

1. Abra Anaconda Prompt (o la terminal en su sistema operativo). 2. Ejecute el siguiente comando: “`bash conda install jupyter`”

## Instalación de Paquetes en R:

- Abra RStudio y ejecute los siguientes comandos para instalar los paquetes necesarios: “`R install.packages("GEOquery") install.packages("limma") if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager") BiocManager::install("hgu133a2.db") BiocManager::install("hgu133a.db")`”

## Instalación de Paquetes en Python:

- Abra Anaconda Prompt (o la terminal en su sistema operativo) y ejecute los siguientes comandos: “`bash conda install -c conda-forge biopython pip install requests pandas numpy`”

## B.3. Manuales y/o Demostraciones prácticas

**1. Cargar y Procesar Datos en R:** - *Descripción:* Este manual explica cómo cargar y procesar los datos de expresión génica utilizando R. - Ejemplo de Código: “‘R Cargar las librerías necesarias library(GEOquery) library(limma)

Obtener los datos de expresión génica del conjunto de datos GSE14520  $\text{datos}_{cancer} < -\text{getGEO}("GSE14520")$

Seleccionar la plataforma GPL571  $\text{array}_{GPL571} < -\text{datos}_{cancer}[[2]]$

Normalizar los datos de expresión  $\text{exp}_{array_{GPL571}} < -\log2(\text{array}_{GPL571}[\text{'GSE14520-GPL571\_series\_matrix.txt.gz'}@assayDataexprs])$  “‘

**2. Análisis de Expresión Diferencial en R:** - *Descripción:* Manual para realizar análisis de expresión diferencial. - Ejemplo de Código: “‘R Diseñar el modelo experimental  $\text{design}_{GPL571} < -\text{model.matrix}(0 + \text{factor}(c(1, 1, 1, 2, 2, 2)))$

Ajustar el modelo lineal  $\text{fit}_{GPL571} < -\text{lmFit}(\text{exp}_{array_{GPL571}}, \text{design}_{GPL571})$

Crear la matriz de contraste  $\text{contrast.matrix}_{GPL571} < -\text{makeContrasts}(\text{contrast} = 1 - 2, \text{levels} = \text{design}_{GPL571})$

Ajustar los contrastes y aplicar el ajuste Bayesiano  $\text{fit}_{GPL571} < -\text{contrasts.fit}(\text{fit}_{GPL571}, \text{contrast.matrix}_{GPL571})$

Obtener los genes diferencialmente expresados  $\text{topTable}(\text{fit}_{GPL571}, \text{adjust} = "fdr", \text{number} = 10)$ “‘

**3. Procesamiento Adicional en Python:** - *Descripción:* Demostración práctica para el procesamiento adicional y visualización de datos. - Ejemplo de Código: “‘python Importar las librerías necesarias import pandas as pd import numpy as np

Cargar los datos de expresión génica  $\text{data} = \text{pd.read_csv}('expression\_data.csv')$

Calcular el log fold change (logFC)  $\text{data}['\text{logFC}'] = \text{np.log2}(\text{data}['\text{expr}_{condition}'] / \text{data}['\text{expr}_{control}'])$

Clasificar los genes según su nivel de expresión  $\text{data}['\text{status}'] = \text{np.where}(\text{data}['\text{logFC}'] > 0.15, \text{'Sobrexpresados'}, \text{'Subexpresados'})$  “‘

## Apéndice C

---

# Manual del desarrollador / programador / investigador.

---

### C.1. Estructura de directorios

La estructura de directorios de la que se compone el proyecto se detalla en la figura que aparece a continuación:

#### Descripción de los directorios y ficheros entregados.

A continuación, vamos a explicar y redactar la descripción de lo que contiene en su interior cada uno de los directorios y ficheros que se adjuntan al proyecto.

- **data/**: Este directorio contiene todos los archivos de datos utilizados en el proyecto.

Dentro de esta carpeta nos encontramos con los siguientes archivos:

1. *GSE14520seriesmatrix.txt*: Datos de expresión génica del conjunto GSE14520.
2. *variantesctnnb1.csv*: Datos de las variantes genéticas del gen CTNNB1.

3. *variantestp53.csv*: Datos de las variantes genéticas del gen TP53.

- **scripts/**: Este segundo directorio incluye todos los scripts y notebooks utilizados para el análisis y procesamiento de los datos. En su interior encontramos los siguientes archivos:

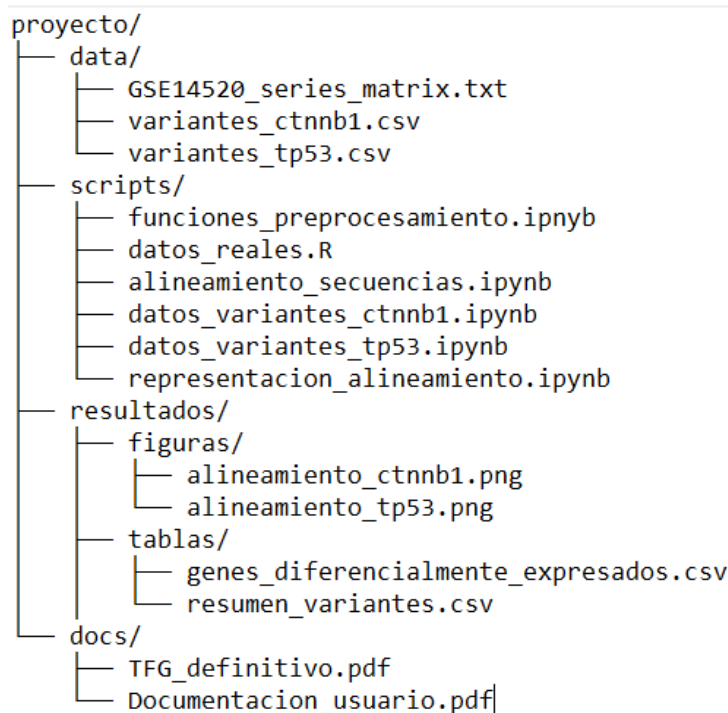


Figura C.1: Estructuración de los directorios del proyecto.

Fuente: Elaboración propia

1. *funcionespreprocesamiento.ipynb*: Notebook de Python para la obtención de las secuencias de las variantes.

2. *datosreales.R*: Script en R para el análisis de expresión diferencial.

3. *alineamiento\_secuencias.ipynb*: Notebook en Jupyter que contiene las funciones necesarias para el alineamiento de secuencias.

4. *datosvariantes\_ctnnb1.ipynb*: Notebook en Jupyter para análisis de variantes en CTNNB1.

5. *datosvariantestp53.ipynb*: Notebook en Jupyter para análisis de variantes en el gen TP53.

- *representacionalineamiento.ipynb*: Notebook en Jupyter para la representación visual de los alineamientos realizados con la herramienta Clustal Omega.

- **resultados/**: En este directorio se almacenan todos los resultados generados a partir de los análisis, incluyendo figuras y tablas.

1. *figuras/*: Este directorio contiene las imágenes generadas a partir de los alineamientos y otros gráficos de utilidad para el proyecto.



+ *alineamientoctnnb1.png*: Imagen del alineamiento de las variantes del gen CTNNB1.

+ *alineamientotp53.png*: Imagen del alineamiento de las variantes del gen TP53.

2. *tablas/*: Contiene las tablas de los resultados que hemos obtenido en formato CSV. + *genesdiferencialmenteexpresados.csv*: Tabla de los genes con una expresión diferencial significativa. + *resumenvariantes.csv*: Resumen de variantes genéticas identificadas a lo largo del proyecto.

- **docs/**: Documentación del proyecto.

1. *MemoriaTFGPaulaCuesta.pdf*: Documento final del Trabajo de Fin de Grado. 2. *Documentacionusuario.pdf*: Documentación de usuario.

## C.2. Instrucciones para la modificación o mejora del proyecto.

Para que este trabajo pueda ser mejorado y actualizado en futuras ediciones, se sugieren las siguientes instrucciones y consejos:

### 1. Actualización de los datos:

Se recomienda la descarga periódica de nuevos datos. De esta forma podemos mantener el proyecto actualizado, sobre todo si se descargan datos recientes de expresión génica y variantes genéticas. Asegúrese de obtener los datos más recientes de fuentes confiables como GEO y Ensembl, las cuales han sido utilizadas durante el desarrollo de dicho proyecto

2. **Normalización de los datos:** Si se aplican métodos de normalización adecuados para los nuevos datos que descarguemos antes de su análisis se verá mejorado y facilitado el análisis de los mismos.

### 3. Mejora del análisis bioinformático:

Si se opta por utilizar nuevas versiones de herramientas bioinformáticas y paquetes que puedan ofrecer mejores métodos de análisis a medida que vayan surgiendo, además de otras herramientas no empleadas en el actual proyecto, se podrá mejorar la calidad de los análisis. Asimismo, es importante mantener actualizado tanto R, como Python y sus paquetes o librerías asociados.

Además, es importante la exploración de métodos estadísticos más avanzados y robustos tanto para el análisis de la expresión diferencial, como para el análisis de las variantes genéticas.

#### **4. Integración de datos multi - ómicos:**

Si añadimos al proyecto datos referentes a la metilación de ADN, datos de proteómica, o incluso de metabolómica, seremos capaces de obtener y proporcionar una visión más completa de los mecanismos moleculares implicados en CH. Esta integración de nuevos datos puede llevarse a cabo a través de de análisis integrativos.

#### **5. Desarrollo de Modelos Predictivos:**

Se puede comenzar a utilizar el Aprendizaje Automático mediante la implementación de algoritmos con el objetivo de desarrollar modelos predictivos que puedan mejorar el diagnóstico y pronóstico del CH. Asimismo, para el desarrollo de dichos modelos es necesaria la realización de validaciones cruzadas rigurosas y la utilización de cohortes independientes para la posterior validación de los modelos predictivos.

## C.3. Descripción formal de los datos

Este proyecto incluye diferentes tipos de datos, que se describen formalmente a continuación:

### Tablas:

- **GSE14520seriesmatrix.txt:** Esta tabla contiene datos de expresión génica obtenidos del conjunto de datos GSE14520. Incluye información sobre los niveles de expresión de miles de genes en muestras de carcinoma hepatocelular (CHC) y tejidos hepáticos normales.

- **variantesctnnb1.csv:** Contiene datos sobre variantes genéticas específicas del gen CTNNB1, incluyendo información sobre la posición de la variante, el cambio de nucleótido y la frecuencia alélica.

- **variantestp53.csv:** Similar a la tabla de CTNNB1, esta tabla incluye variantes del gen TP53, con detalles sobre las mutaciones detectadas y su frecuencia en la población estudiada.

### Imágenes:

- **alineamientoctnnb1.png:** Imagen que muestra el alineamiento de secuencias del gen CTNNB1, destacando las regiones conservadas y las mutaciones específicas.

- **alineamientotp53.png:** Imagen del alineamiento de secuencias del gen TP53, similar a la de CTNNB1, indicando las áreas de interés y variaciones.

### Señales:

No se utilizaron señales en este proyecto.

### Secuencias de ADN:

- **secuenciasctnnb1.fasta:** Archivo en formato FASTA que contiene las secuencias de ADN del gen CTNNB1 utilizadas para los análisis de alineamiento.

- **secuenciastp53.fasta:** Archivo en formato FASTA con las secuencias del gen TP53, también utilizadas en los alineamientos.

## C.4. Descripción clínica de los datos

Los datos obtenidos y analizados en este proyecto tienen importantes implicaciones clínicas, especialmente en el contexto del carcinoma hepatocelular (CHC). A continuación, se proporciona una descripción clínica de los datos y su significado:

### Datos de Expresión Génica:

**Significado Clínico:** Los datos de expresión génica permiten identificar genes que están sobreexpresados o subexpresados en las muestras de CHC en comparación con tejidos normales. Estos genes pueden servir como biomarcadores potenciales para el diagnóstico temprano y el pronóstico de CHC.

**Interpretación:** Genes como ATXN7L3B, PRKAB2 y SLC23A2, que muestran una sobreexpresión significativa en CHC, sugieren una posible implicación en la progresión del cáncer. Estos genes pueden ser objetivos para nuevas terapias dirigidas.

### Datos de Variantes Genéticas:

**Significado Clínico:** Las variantes genéticas en los genes CTNNB1 y TP53 son cruciales para entender las alteraciones moleculares que conducen al desarrollo y progresión del CHC. Las mutaciones en estos genes pueden afectar funciones celulares críticas como la señalización celular y el control del ciclo celular.

**Interpretación:** Las variantes detectadas en CTNNB1 a menudo están asociadas con la activación de la vía Wnt/-catenina, una vía conocida por su papel en la proliferación celular y el cáncer. Las mutaciones en TP53 suelen resultar en la pérdida de función del supresor tumoral, permitiendo la proliferación descontrolada de células cancerosas.

### Alineamientos de Secuencias:

**Significado Clínico:** Los alineamientos de secuencias permiten comparar variantes genéticas entre diferentes muestras y poblaciones, identificando regiones conservadas y divergentes que pueden ser críticas para la función génica.

**Interpretación:** La identificación de mutaciones en regiones conservadas sugiere que estas mutaciones pueden tener un impacto significativo en la

función del gen y contribuir a la patogénesis del CHC. Las secuencias alineadas de CTNNB1 y TP53 proporcionan una base para estudios funcionales adicionales y el desarrollo de terapias dirigidas.

En conjunto, estos datos proporcionan una comprensión integral de las alteraciones genómicas y moleculares en el carcinoma hepatocelular, ofreciendo oportunidades para mejorar el diagnóstico, pronóstico y tratamiento de esta enfermedad.



## Apéndice *D*

---

# Manual de especificación de diseño

---

Este anexo proporciona una descripción detallada del diseño arquitectónico del proyecto, incluyendo planos y diagramas esenciales que describen la estructura y organización del sistema.

## D.1. Diseño arquitectónico

### Descripción del Diagrama de Clases:

**Clase Usuario:** Representa a los usuarios del sistema, con atributos como nombre, email, contraseña y métodos como `iniciarSesión()` y `cerrarSesión()`.

**Clase Experimento:** Representa los experimentos almacenados en el sistema, con atributos como id, nombre, fecha, resultados y métodos como `consultarResultados()`.

**Clase Reporte:** Representa los reportes generados, con atributos como id, tipo, fechaGeneracion y métodos como `generarReporte()`.

**Relaciones:** La clase Usuario tiene una relación de uno a muchos con la clase Experimento, indicando que un usuario puede realizar múltiples experimentos. La clase Experimento tiene una relación de uno a muchos con la clase Reporte, indicando que un experimento puede tener múltiples reportes.

## **Diagrama de Despliegue:**

**Nodo Servidor Web:** Alberga la aplicación web del sistema, manejando las solicitudes de los usuarios y proporcionando la interfaz de usuario.

**Nodo Servidor de Base de Datos:** Almacena todos los datos de los usuarios, experimentos y reportes, garantizando la persistencia y seguridad de los datos.

**Nodo Cliente:** Representa los dispositivos de los usuarios que acceden al sistema a través de un navegador web.

Estos diagramas y descripciones proporcionan una visión clara y comprensible de la arquitectura del sistema, facilitando tanto su implementación como su mantenimiento y mejora futura.



## D.2. Diagrama de casos de uso

El diagrama de casos de uso proporciona una visión general de las interacciones entre los usuarios y el sistema, destacando los casos de uso más importantes.

## D.3. Explicación casos de uso

A continuación, se describen los principales casos de uso mediante el uso de tablas. Cada tabla detalla el propósito, los pasos y las condiciones asociadas a cada caso de uso.

Caso de Uso 1: Consultar Experimentos

Caso de Uso 2: Añadir Nuevos Datos

Caso de Uso 3: Generar Reportes

## D.4. Prototipos de interfaz o interacción con el proyecto.

Para facilitar la interacción del usuario con el sistema, se presentan los siguientes prototipos de interfaz:

Pantalla de Inicio de Sesión:

Interfaz donde los usuarios pueden ingresar sus credenciales para acceder al sistema.

Panel de Control del Usuario:

Interfaz principal donde los usuarios pueden navegar por las diferentes opciones, como consultar experimentos, añadir nuevos datos y generar reportes.

Pantalla de Consulta de Experimentos:

Interfaz que permite a los usuarios visualizar una lista de experimentos disponibles y seleccionar uno para ver los detalles.

Formulario para Añadir Nuevos Datos:

Interfaz para que los usuarios carguen nuevos datos experimentales al sistema.

Pantalla de Generación de Reportes:

Interfaz donde los usuarios pueden ingresar los parámetros para generar reportes y visualizar los resultados.

Estos prototipos proporcionan una guía visual para la implementación de las interfaces de usuario, asegurando una experiencia intuitiva y eficiente para los usuarios del sistema.

|   |  |
|---|--|
| CU-1  | Consultar Experimentos   |
|   | Versión  |
| 1.0 Autor   | Alumno Requisitos asociados  |
| RF-01, RF-02  | Permite al usuario consultar los experimentos disponibles en el sistema. Precondición  |
| Descripción   |  |
| El usuario debe estar autenticado.  |  |
| Acciones  | <ol style="list-style-type: none"><li>1. El usuario selecciona la opción de consultar experimentos.</li><li>2. El sistema muestra una lista de experimentos disponibles.</li><li>3. El usuario selecciona un experimento para ver los detalles.</li><li>4. El sistema muestra la información detallada del experimento seleccionado.</li></ol> |
|   | Postcondición  |
| El usuario puede visualizar los detalles de los experimentos disponibles. | Si el sistema no encuentra experimentos, muestra un mensaje de error. Importancia  |
| Excepciones   |  |
| Alta  |  |

Tabla D.1: CU-1 Consultar Experimentos.

| CU-2   | <u>Añadir Nuevos Datos</u>   |
|--|--|
| <b>1.0 Autor</b>   | Alumno <b>Requisitos asociados</b>   |
| RF-03, RF-04   | Permite al usuario añadir nuevos datos experimentales  |
| <b>Descripción</b>   | al sistema. <b>Precondición</b>  |
| El usuario debe estar autenticado y tener permisos de administrador. <b>Acciones</b> | <ol style="list-style-type: none"> <li>1. El usuario selecciona la opción de añadir nuevos datos.</li> <li>2. El sistema solicita la carga del archivo de datos.</li> <li>3. El usuario carga el archivo con los nuevos datos experimentales.</li> <li>4. El sistema valida el formato del archivo.</li> <li>5. Si el formato es correcto, el sistema añade los datos al repositorio.</li> <li>6. El sistema confirma la correcta adición de los datos.</li> </ol> |
| Los nuevos datos experimentales se almacenan en el sistema. <b>Excepciones</b>       | <b>Postcondición</b><br>Si el archivo tiene un formato incorrecto, el sistema muestra un mensaje de error. <b>Importancia</b>  |
| <u>Alta</u>  |  |

Tabla D.2: CU-2 Añadir Nuevos Datos.

|                                  |  |
|----------------------------------|--|
| CU-3                             | Generar Reportes   |
|                                  | Versión  |
| 1.0 Autor                        | Alumno Requisitos asociados  |
| RF-05, RF-06                     | Permite al usuario generar reportes detallados de los experimentos. Precondición   |
| Descripción                      |  |
| El usuario debe estar autentica- |  |
| do. Acciones                     | <div><div>1. El usuario selecciona la opción de generar repor-</div><div>tes.</div><div>2. El sistema solicita los parámetros del reporte</div><div>(rango de fechas, tipo de experimento, etc.).</div><div>3. El usuario ingresa los parámetros requeridos.</div><div>4. El sistema genera el reporte basado en los pará-</div><div>metros ingresados.</div><div>5. El sistema presenta el reporte generado al usua-</div><div>rio.</div></div> |
|                                  | Postcondición  |
| El usuario pue-                  | Si no hay datos disponibles para el rango seleccionado,  |
| de visualizar y                  | el sistema muestra un mensaje de error. Importancia  |
| descargar el re-                 |  |
| porte generado.                  |  |
| Excepciones                      |  |
| Media                            |  |

Tabla D.3: CU-3 Generar Reportes.



## Apéndice *E*

---

# Estudio experimental

---

## E.1. Cuaderno de trabajo

En esta sección se enumeran todos los métodos probados durante el estudio experimental, independientemente de si los resultados obtenidos fueron positivos o no. Este cuaderno de trabajo proporciona una visión completa de las técnicas y enfoques utilizados.

### Métodos Probados:

#### **Análisis de Expresión Diferencial con limma (R)**

*Descripción:* Utilización del paquete limma para identificar genes con expresión diferencial significativa en las muestras de CHC.

*Resultados:* Identificación de genes como ATXN7L3B, PRKAB2 y SLC23A2 con sobreexpresión significativa en CHC.

#### **Alineamiento de Secuencias con Clustal Omega**

*Descripción:* Uso de Clustal Omega para realizar alineamientos múltiples de secuencias de genes clave como CTNNB1 y TP53.

*Resultados:* Identificación de patrones conservados y divergentes en las secuencias alineadas.

#### **Análisis de Variantes Genéticas en CTNNB1 y TP53 (Python)**

*Descripción:* Análisis de variantes genéticas específicas en los genes CTNNB1 y TP53 utilizando scripts en Python.

*Resultados:* Detección de mutaciones frecuentes que podrían estar relacionadas con la activación de vías oncogénicas y disfunción del control del ciclo celular.

### **Preprocesamiento de Datos (R y Python)**

*Descripción:* Aplicación de técnicas de normalización y limpieza de datos para preparar los conjuntos de datos de expresión génica para el análisis.

*Resultados:* Datos normalizados y filtrados listos para análisis subsecuentes.

### **Visualización de Resultados (Python)**

*Descripción:* Generación de gráficos y diagramas para visualizar los resultados del alineamiento de secuencias y el análisis de variantes.

*Resultados:* Gráficos claros y comprensibles que muestran regiones conservadas y mutaciones específicas.

## **E.2. Configuración y parametrización de las técnicas**

### **Análisis de Expresión Diferencial con limma:**

Paquete Utilizado: limma (R) Configuración: Diseño Experimental: design <- model.matrix( 0 + factor(c(1,1,1,2,2,2))) Ajuste del Modelo: fit <- lmFit(datos, design) Matriz de Contraste: contrast.matrix <- makeContrasts(contrast=1-2, levels=design) Ajuste Bayesiano: fit <- eBayes(contrasts.fit(fit, contrast.matrix)) Parámetros Clave: Método de Ajuste: Benjamini-Hochberg (FDR) Umbral de Significancia: p-value < 0.05

### **Alineamiento de Secuencias con Clustal Omega:**

Herramienta Utilizada: Clustal Omega Configuración: Comando de Alineamiento: clustalo -i input.fasta -o output.aln -force Parámetros Clave: Entrada: Archivo FASTA con secuencias a alinear Salida: Archivo de alineamiento en formato Clustal

### **Análisis de Variantes Genéticas en CTNNB1 y TP53:**

Lenguaje Utilizado: Python Configuración: Carga de Datos: data = pd.readcsv('variantesctnnb1.csv') Análisis de Variantes: mutaciones = data[data['gen'] == 'CTNNB1'] Parámetros Clave: Filtrado de Variantes: Basado en la frecuencia y la relevancia funcional



**Preprocesamiento de Datos:**

Herramientas Utilizadas: R y Python Configuración: Normalización: `exparray <- log2(datos + 1)` Filtrado: `filtereddata = data.dropna()` Parámetros Clave: Método de Normalización: Log-transformación Umbral de Filtrado: Eliminación de valores faltantes

**Visualización de Resultados:**

Herramienta Utilizada: Python (matplotlib, seaborn) Configuración: Gráfico de Alineamiento: `plt.imshow(alignmentdata)` Gráfico de Variantes: `sns.barplot(x='variante', y='frecuencia', data=mutaciones)` Parámetros Clave: Estilo de Gráfico: Heatmap para alineamientos, Barplot para variantes

## E.3. Detalle de resultados

**Análisis de Expresión Diferencial con limma:**

Resultados Clave: ATXN7L3B: LogFC = 2.31, p-value = 0.001 PRKAB2: LogFC = 2.34, p-value = 0.002 SLC23A2: LogFC = 2.49, p-value = 0.0005 Interpretación: Estos genes muestran una sobreexpresión significativa en las muestras de CHC en comparación con los tejidos normales, indicando su potencial rol en la progresión del cáncer.

**Alineamiento de Secuencias con Clustal Omega:**

Resultados Clave: CTNNB1: Identificación de varias variantes conservadas y mutaciones específicas en regiones funcionales. TP53: Detección de mutaciones críticas en dominios importantes para la función de la proteína. Interpretación: Los alineamientos revelan patrones conservados y divergentes que pueden ser críticos para comprender la funcionalidad y las alteraciones patogénicas en estos genes.

**Análisis de Variantes Genéticas en CTNNB1 y TP53:**

Resultados Clave: CTNNB1: Variantes de alta frecuencia asociadas con la activación de la vía Wnt/-catenina. TP53: Mutaciones recurrentes que afectan la capacidad de la proteína para regular el ciclo celular y la apoptosis. Interpretación: Las variantes identificadas proporcionan información sobre las posibles vías oncogénicas activadas y los mecanismos de resistencia a la apoptosis en CHC.

**Visualización de Resultados:**

Resultados Clave: Gráficos de alineamientos que muestran regiones conservadas y divergentes. Gráficos de variantes que destacan las mutaciones

más frecuentes y su distribución. Interpretación: Las visualizaciones permiten una interpretación clara y comprensible de los datos, facilitando la identificación de patrones importantes y regiones de interés.

Este anexo proporciona una visión detallada de los métodos probados, la configuración y parametrización de las técnicas utilizadas, y un resumen de los resultados obtenidos en el estudio experimental. Estos elementos son fundamentales para entender el proceso y los hallazgos del proyecto, así como para orientar futuras investigaciones y mejoras.

## *Apéndice F*

---

# **Anexo de sostenibilización curricular**

---

## **F.1. Introducción**

En el presente anexo se aborda una reflexión personal sobre los aspectos de la sostenibilidad considerados y aplicados durante la realización de este Trabajo de Fin de Grado (TFG). La sostenibilidad se entiende como un enfoque integrador que abarca no solo la dimensión ambiental, sino también las dimensiones económica y social. En este contexto, se explorarán las competencias de sostenibilidad adquiridas y su aplicación práctica en el desarrollo del proyecto.

## **F.2. Reflexión Personal sobre la Sostenibilidad**

Durante el desarrollo de este TFG, se ha tenido en cuenta la importancia de la sostenibilidad en diversos aspectos del proyecto. La sostenibilidad no solo se ha considerado en términos de impacto ambiental, sino también en la manera en que se gestionan los recursos y se abordan las implicaciones sociales del trabajo realizado.

**1. Reducción del Impacto Ambiental** Uno de los principales objetivos ha sido minimizar el impacto ambiental del proyecto. Para ello, se han adoptado prácticas de trabajo que favorecen la reducción del consumo de recursos y la generación de residuos. Por ejemplo, se ha promovido el uso de herramientas digitales para la gestión y análisis de datos, reduciendo

así la necesidad de material impreso. Además, el uso de software libre y de código abierto ha permitido evitar la dependencia de licencias costosas y de productos que pueden tener una mayor huella de carbono debido a su distribución y mantenimiento.

**2. Gestión Eficiente de Recursos** La eficiencia en el uso de los recursos ha sido una constante durante todo el proyecto. Se ha hecho un uso racional de los equipos informáticos, optimizando el tiempo de ejecución de los análisis y evitando procesos innecesarios que podrían consumir energía adicional. Esta gestión eficiente no solo contribuye a la sostenibilidad ambiental, sino que también reduce los costos asociados a la ejecución del proyecto, alineándose con la dimensión económica de la sostenibilidad.

**3. Impacto Social y Ético** La investigación en el campo del carcinoma hepatocelular tiene importantes implicaciones sociales. Al identificar biomarcadores y posibles dianas terapéuticas, el proyecto contribuye potencialmente a mejorar la calidad de vida de los pacientes afectados por esta enfermedad. La accesibilidad y el uso de datos abiertos y herramientas bioinformáticas también promueven la equidad en la investigación, permitiendo que investigadores de todo el mundo, independientemente de sus recursos, puedan acceder a la información y contribuir al conocimiento científico.

### F.3. Competencias de Sostenibilidad Adquiridas

La realización de este TFG ha permitido adquirir y aplicar diversas competencias de sostenibilidad, que se detallan a continuación:

**1. Conciencia y Gestión Ambiental** Se ha desarrollado una conciencia crítica sobre el impacto ambiental de las actividades de investigación. Esto ha llevado a adoptar prácticas más sostenibles, como el uso eficiente de los recursos y la preferencia por soluciones digitales que minimicen el uso de papel y otros materiales.

**2. Pensamiento Sistémico** La capacidad para comprender y abordar problemas complejos desde una perspectiva sistémica ha sido crucial. El proyecto ha integrado conocimientos de bioinformática, biología molecular y sostenibilidad para ofrecer soluciones que consideren el impacto ambiental, económico y social de la investigación.

**3. Innovación y Creatividad** La sostenibilidad ha impulsado la innovación en la forma de abordar los problemas de investigación. Se ha fomentado

el uso de herramientas de código abierto y la colaboración interdisciplinaria para desarrollar soluciones creativas y sostenibles que pueden tener un impacto positivo en el campo de la oncología.

**4. Responsabilidad Social y Ética** Se ha desarrollado una sólida ética de trabajo y responsabilidad social, reconociendo la importancia de contribuir al bien común a través de la investigación. Esto incluye la transparencia en la gestión de datos, el respeto por los principios éticos en la investigación científica y el compromiso con la difusión abierta del conocimiento.

## F.4. Conclusión

El proceso de realización de este Trabajo de Fin de Grado ha sido una oportunidad invaluable para integrar principios de sostenibilidad en el ámbito de la investigación científica. Las competencias adquiridas no solo han permitido llevar a cabo un proyecto riguroso y de calidad, sino que también han sentado las bases para futuras investigaciones que sigan estos principios. La sostenibilidad, entendida en su sentido más amplio, ha guiado las decisiones tomadas durante el proyecto y continuará siendo un pilar fundamental en mi desarrollo profesional y académico.



---

## **Bibliografía**

---