**RESOLUCIÓN PARA LA MEJORA Y APLICACIONES DE LA INTERFAZ OMICS INTEGRATOR**

TESIS DE MAESTRÍA

PRESENTADO POR:

**ESTEBAN VARGAS PARRA**

DIRECTOR:

**ALEX SANCHEZ PLA**

**ÍNDICE GENERAL**

[INTRODUCCIÓN 3](#_Toc99834292)

[Contexto y justificación del trabajo 3](#_Toc99834293)

[Descripción general 3](#_Toc99834294)

[Justificación del Trabajo Final de Máster 3](#_Toc99834295)

[Objetivos 4](#_Toc99834296)

[Objetivos generales 4](#_Toc99834297)

[Objetivos específicos 4](#_Toc99834298)

[Enfoque y método a seguir 4](#_Toc99834299)

[Planificación con hitos y temporización 5](#_Toc99834300)

[Tareas 5](#_Toc99834301)

[Calendario 5](#_Toc99834302)

[Hitos 10](#_Toc99834303)

[Análisis de riesgo 10](#_Toc99834304)

[Plan de trabajo 10](#_Toc99834305)

[Resultados esperados 10](#_Toc99834306)

[Plan de trabajo 11](#_Toc99834307)

[ESTADO DEL ARTE 11](#_Toc99834308)

[Bibliografía 12](#_Toc99834309)

**ÍNDICE DE FIGURAS**

[**Figura 1.** Primera fase 7](#_Toc100008797)

[**Figura 2.** Segunda fase 7](#_Toc100008798)

[**Figura 3.** Tercera fase 8](#_Toc100008799)

[**Figura 4.** Cuarta fase 9](#_Toc100008800)

[**Figura 5.** Fase final 10](#_Toc100008801)

# INTRODUCCIÓN

# Contexto y justificación del trabajo

## Descripción general

El presente trabajo Final de Máster (TFM) se basará en un proyecto aplicativo donde se tomará una interfaz denominada Omics Integrator, la cual fue creada en un trabajo previo, con el fin de procesar datos ómicos; a esta aplicación se le añadirán nuevas herramientas y técnicas que sean suficientemente eficientes, con el fin de manejar este tipo de datos de una manera más óptima en un tiempo razonable.

# Justificación del Trabajo Final de Máster

La genómica, proteómica y metabolómicas se han visto envueltas durante los últimos años, en el área de las ciencias biológicas en el desarrollo de nuevas técnicas, con el fin de estudiar conjuntamente procesos de expresión de los genes y sus productos [1].

Antiguamente se creía que llegar a estudiar un sistema biológico sería algo inalcanzable, sin embargo, con el auge de la bioinformática, la mejora de nuevas técnicas y el avance de la tecnología, se ha logrado interpretar procesos o sistemas enteros pertenecientes a las células, generando cantidades robustas de datos [2].

La aplicación Omics Integrator, es una aplicación que se encarga de integrar diferentes aplicaciones de análisis de datos ómicos, para llegar a resultados como lo son el VolcanoPlot, Análisis de Componentes Principales (PCA), Filtraciones y Mapas de Calor.

A pesar de ser una aplicación con las funcionalidades mencionadas previamente, ésta requiere de algunas mejoras, donde se permita relacionar de manera más sencilla al usuario que la pondrá en práctica con la interfaz, así como se incluyan nuevas funcionalidades con ejemplos y vídeos interactivos con el fin de ser más práctica al momento de emplearse.

Por ende, se empleará el paquete Shiny perteneciente a R, con el objetivo de desarrollar nuevas técnicas de análisis y funciones interactivas en la aplicación Omics Integrator. Con este TFM se buscará dejar un aplicativo que no sólo sea más potente, sino que además tenga la opción de tener mejoras concordes a las futuras nuevas técnicas de análisis vayan surgiendo en esta área de estudio.

# Objetivos

A continuación, serán enunciados los objetivos que se desean alcanzar con el Trabajo Final de Máster.

## Objetivos generales

1. Integrar a la aplicación Omics Integrator, nuevas funciones y mejoras de análisis ómicos.
2. Incluir en la aplicación Omics Integrator, ejemplos y documentación interactiva para el usuario

## Objetivos específicos

1. Funciones y mejoras:
2. Realizar una búsqueda rigurosa de métodos de análisis de datos ómicos
3. Indicar las ventajas y desventajas de los distintos métodos de análisis
4. Seleccionar el o los más completos e incluirlos en el aplicativo.
5. Validar los datos con una plataforma de análisis
6. Ejemplos y documentación:
7. Realizar una documentación en con el instructivo del uso del aplicativo.
8. Incluir vídeos interactivos con ejemplos de cada función del aplicativo.
9. Disponer el software con funciones actualizada a los usuarios de interés.

# Enfoque y método a seguir

Como primer paso para llevar a cabo respectivo proyecto, es buscar las diferentes fuentes bibliográficas investigativas a cerca de análisis ómicos, con el fin de saber y comprender el tratamiento de los datos y la funcionalidad de ello. A pesar de que esta fase no es muy detallada, se realiza la el primer contacto con el tema de interés.

Basados en ello, se van planteando las formas de trabajo:

* Primero se desarrolla la parte teórica del proyecto (búsqueda de bibliografía, antecedentes, comparaciones, etc…) y después de ello se procede a la parte práctica (funciones disponibles en R, paquetes, ejemplos), se hace reconocimiento y análisis del Omics Integrator, se evalúan las ventajas y desventajas para identificar mejoras del aplicativo.
* Empezar con las dos partes del proyecto a la vez, lo práctico y lo teórico.

Se procede a emplear la primera opción, ya que es un tema robusto y se requiere una búsqueda minuciosa por parte del investigador, para definir cuáles son las aplicaciones que el Omics Integrator no posee y además aquellas que están disponibles cómo se podrían mejorar. Además, que, con esto realizado, se puede proceder a crear un paquete en R y poder implementarlo directamente en la aplicación web.

Así mismo, aquellas aplicaciones que se desean incluir o mejorar en el software, al contar con una base teórica definida, permitirá tener en cuenta el tiempo en el cual se procede a desarrollar el código en R ya sea que se cuente con unas bases o que deba empezarse desde cero. Además, una vez se empiece con la creación del código, es importante que el adiestramiento del paquete Shiny para la modificación del aplicativo y la inclusión de nuevas funciones.

# Planificación con hitos y temporización

En esta parte será especificada las tareas que se llevarán a cabo en el Trabajo Fin de Máster, asimismo, se empleará un calendario que detallará las distintas labores y lapsos de entrega. Finalmente, se evaluarán y analizarán los riesgos que puedan alterar el proyecto.

## Tareas

A partir de lo mencionado previamente se permite desglosar en las siguientes actividades:

* Realizar una búsqueda rigurosa de métodos de análisis de datos ómicos (OE 1, con una duración de 30 horas).
* Para cada método de análisis se asignarán los siguientes trabajos:
* Identificar la funcionalidad y uso de los métodos (O.E. 1.I)
* A partir de literatura y trabajos, establecer las ventajas y desventajas de los análisis a emplear (O.E. 1.II)
* Realizar un cuadro comparativo que permita determinar cuáles son los análisis más completos y establecer cuál o cuáles serán incluidos a partir de las ventajas y desventajas (O.E. 1.III)
* Buscar paquetes estadísticos en R que permitan el cálculo de los análisis deseados, si no hay uno programarlo (O.E. 1.III)
* Comparar resultados con plataformas previas, como Galaxy (O.E. 1.IV)
* Adiestramiento en el manejo de Shiny y Git.hub (9 horas O.E. 1 y 2)
* Crear el código en R con el respectivo análisis (60 horas dependiendo de la complejidad O.E. 1)
* Crear una documentación para el uso del aplicativo con todas sus funciones con un paso a paso (10 horas O.E. 2.I)
* Realizar vídeos interactivos con ejemplos para el uso de la aplicación, una opción viable para la creación del vídeo es OBS Studio (20 horas O.E. II)
* Subir los vídeos al código con la ayuda del paquete *vembedr* (2 horas O.E. II)
* Interactuar con el software que quede a disposición del usuario de interés (Tiempo:N.A. O.E. III)

# Calendario

En la siguiente gráfica se presentará el calendario por meses del Trabajo Final de Máster. Para ello se tomó en cuenta la disponibilidad del estudiante la cual está basada principalmente los fines de semana y festivos, mientras que los días entre semana la dedicación será máxima de unas 3 horas por día.



**Figura 1.** Primera fase

Como se puede ver en el mes de febrero, el día 23 empieza la definición de los puntos tenidos en cuenta para el plan de trabajo. Como primera fase se definieron los objetivos, cómo se llegarán a cumplir esos objetivos y, además, cuánto tiempo tomará la realización de cada tarea.



**Figura 2.** Segunda fase



**Figura 3.** Tercera fase

Para los meses de marzo y abril se trabajará en toda la parte robusta del presente proyecto, trabajando directamente sobre cada objetivo y tarea, además durante estos meses se hará el trabajo directo en la búsqueda completa de la bibliografía y el trabajo que llevará más tiempo, que es la creación de los códigos y la validación de los resultados.



**Figura 4.** Cuarta fase

Para el mes de mayo en sus días finales, se tiene prevista la entrega del documento con finalizado.



**Figura 5.** Fase final

Finalmente, en el mes de junio se llevará a cabo la realización de la presentación y la defensa pública.

# Hitos

Analizando las actividades planteadas en el proyecto, se enmarcan a continuación los hitos, que son aquellos percances donde si se presenta algún retraso, influyen negativamente en los plazos de las demás tareas.

**Tabla 1**

|  |  |  |
| --- | --- | --- |
| **Hito** | **PEC** | **Fecha** |
| Plan de trabajo | PEC 1 | 23/02/2022 |
| Desfase de trabajo | PEC 1 | 08/03/2022 |
| Búsqueda de bibliografía | PEC2 | 07/03/2022 |
| Creación del Paquete en R y aplicación Shiny | PEC2 y PEC3 | 11/04/2022 |
| Creación de documento y vídeo | PEC3 | 16/05/2022 |
| Entrega documento | PEC4 | 02/06/2022 |
| Elaboración presentación | PEC5a | 06/06/2022 |
| Defensa pública | PEC5b | 24/06/2022 |

# Análisis de riesgo

En todos los proyectos se presentan externalidades que afectan negativamente el avance del mismo, provocando que la TFM no se lleve a cabo como se esperaba.

Enseguida, se señalan algunos de los factores que logre afectar lo que se planeó con aterioridad:

1. El tiempo en el que se realiza el proyecto es muy corto.
2. El trabajo proyecta varios temas abiertos, por lo cual definir uno específico es riguroso
3. No llevar a cabo el plan de trabajo como se había estructurado
4. Que el plan de trabajo no haya sido el correcto
5. Poco conocimiento en la creación de paquetes en R
6. No poder crear las modificaciones y aplicaciones del sitio web, debido a que el autor no ha tenido contacto con el paquete Shiny.
7. Hubo un desfase de trabajo en la entrega de la PEC 1, lo cual provocó un retraso en el avance del trabajo final como se esperaba.

# Plan de trabajo

El del documento presente, en el que se indica y propone el desarrollo del trabajo, detallando como se va a proceder en la elaboración del TMF. Además, se señalan los objetivos, los hitos, el tiempo estimado de cada actividad. Finalmente, se definen los posibles riesgos que conlleva el proyecto.

# Resultados esperados

Finalizado el proyecto se obtendrán las siguientes salidas que se especifican a continuación:

## Plan de trabajo

Es el documento actual, sobre el que se detalla qué se va a realizar, demarcando el trabajo y señalando la manera en la que se procederá. Además, los objetivos del TFM se especifican, así como se planifican los hitos y el tiempo llevado a cabo en cada tarea. Finalmente se establecen y distinguen los posibles riesgos que pueden aparecer en la realización del proyecto.

# ESTADO DEL ARTE

Con el avance de las plataformas modernas para mediciones ómicas de alto rendimiento, éstas se han vuelto necesarias, con el fin obtener avances de estudios biológicos, biomédicos, entre otros; buscando adoptar un enfoque que otorgue información sobre los sistemas biológicos. Datos de distintas fuentes ómicas, como la genómica, proteómica y metabolómica, se han integrado para descubrir el complejo funcionamiento de la biología de sistemas, empleando algoritmos predictivos basados en un lenguaje de aprendizaje automático. A partir de estos métodos de aprendizajes estadísticos, se brindan nuevas técnicas para integrar y analizar datos ómicos, permitiendo el hallazgo de biomarcadores, con el potencial de predecir diferenciación de expresión génica [3].

## Ómicas

Esta palabra proveniente del inglés, la cual se refiere al estudio de un conjunto en total [2], en el área de las ciencias biológicas se emplea a delimitar el estudio de los distintos sistemas biológicos que contiene el funcionamiento de la célula [4]. La primera disciplina ómica en aparecer fue la genómica, enfocándose en el estudio de genomas completos, como oposición a lo que se realizaba tradicionalmente en la genética que examinaba variantes de manera individual o genes únicos; estos estudios genómicos otorgaron un cerco útil para mapear y estudiar variantes genéticas específicas que favorecen tanto a enfermedades mendelianas como complejas [5].

Las tecnologías ómicas que aporta la biología de sistemas, son herramientas importantes para análisis integrales, debido a que, aquellas tecnologías tradicionales que se basan desde un enfoque reduccionista, han sido insuficientes para resolver problemas en un sistema biológico. En lugar de esto, la biología de sistemas posee un enfoque holístico, con el fin de comprender mejor todo el proceso tanto a nivel cuantitativo como cualitativo, proporcionando información sobre enfermedades, terapias, toxicidades, entre otros [6].

Antiguamente el estudio en su totalidad de un sistema biológico era una labor improbable, además la bioinformática era una herramienta de acceso limitado, donde los estudios se restringían a un único elemento o interacción específica, como saber si un gen estaba expresándose o no; pero en la actualidad, las técnicas modernas son capaces de examinar procesos o sistemas completos de las células, creando cantidades robustas de datos [7].

# Bibliografía

[1] P. Hieter and M. Boguski, “Functional genomics: It’s all how you read it,” *Science (80-. ).*, vol. 278, no. 5338, pp. 601–602, 1997, doi: 10.1126/science.278.5338.601.

[2] D. Tabas Madrid, “Herramientas eficientes para el análisis masivo de datos ómicos,” p. 167, 2018, [Online]. Available: https://eprints.ucm.es/id/eprint/49798/1/T40488.pdf.

[3] P. S. Reel, S. Reel, E. Pearson, E. Trucco, and E. Jefferson, “Using machine learning approaches for multi-omics data analysis: A review,” *Biotechnol. Adv.*, vol. 49, 2021, doi: 10.1016/j.biotechadv.2021.107739.

[4] Q. Pan, O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, “Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing,” *Nat. Genet.*, vol. 40, no. 12, pp. 1413–1415, 2008, doi: 10.1038/ng.259.

[5] Y. Hasin, M. Seldin, and A. Lusis, “Multi-omics approaches to disease,” *Genome Biol.*, vol. 18, no. 1, 2017, doi: 10.1186/s13059-017-1215-1.

[6] B. Karahalil, “Overview of Systems Biology and Omics Technologies,” *Curr. Med. Chem.*, vol. 23, no. 37, pp. 4221–4230, 2016, doi: 10.2174/0929867323666160926150617.

[7] V. Marx, “Biology: The big challenges of big data TL - 498,” *Nature*, vol. 498 VN-, no. 7453, pp. 255–260, 2013, [Online]. Available: http://dx.doi.org/10.1038/498255a.