Identificación de genes expresados diferencialmente en ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial utilizando herramientas bioinformáticas.

Jarre Moreano, Alanis Rafaela

Departamento de Ciencias de la Vida y de la Agricultura

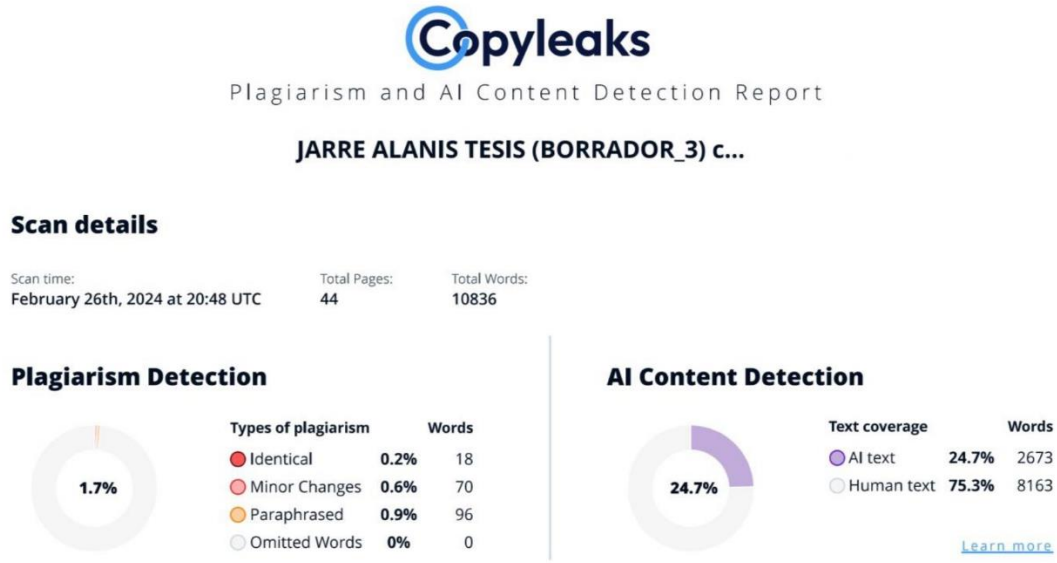
Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniera en Biotecnología

Flores Flor, Francisco Javier Ph. D.

07 de marzo del 2024

Resultados de la herramienta para verificación y/o análisis de similitud de contenidos



Copyleaks

Plagiarism and AI Content Detection Report

JARRE ALANIS TESIS (BORRADOR_3) c...

Scan details

Scan time:	Total Pages:	Total Words:
February 26th, 2024 at 20:48 UTC	44	10836

Plagiarism Detection

1.7%

Types of plagiarism		Words
Identical	0.2%	18
Minor Changes	0.6%	70
Paraphrased	0.9%	96
Omitted Words	0%	0

AI Content Detection

24.7%

Text coverage		Words
AI text	24.7%	2673
Human text	75.3%	8163

Learn more

Plagiarism Results: (6)

Logran revertir la fibrosis pulmonar en modelos murinos con nanopartíc... **0.7%**

https://www.ciberes.org/noticias/logran-revertir-la-fibrosis-pulmonar-en-modelos-murinos-con-nanoparticula...

...

Logran revertir la fibrosis pulmonar en modelos murinos con nanopartíc... **0.7%**

https://www.ciberer.es/noticias/logran-revertir-la-fibrosis-pulmonar-en-modelos-murinos-con-nanoparticulas...

...

Firmado electrónicamente por:
FRANCISCO JAVIER
FLORES FLOR

.....

Flores Flor, Francisco Javier

Director

Departamento de Ciencias de la Vida y de la Agricultura

Carrera De Ingeniería en Biotecnología

Certificación

Certifico que el trabajo de titulación: "**Identificación de genes expresados diferencialmente en ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial utilizando herramientas bioinformáticas**" fue realizado por la señorita **Jarre Moreano, Alanis Rafaela;** el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 27 de febrero del 2024

Firmado electrónicamente por:
FRANCISCO JAVIER
FLORES FLOR

.....

Flores Flor, Francisco Javier

C.C: 1713443479

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Responsabilidad de Autoría

Yo, **Jarre Moreano, Alanis Rafaela**, con cédula de ciudadanía n°1310507734, declaro que el contenido, ideas y criterios del trabajo de titulación**: "Identificación de genes expresados diferencialmente en ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial utilizando herramientas bioinformáticas"** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí 27 de febrero del 2024

……..

Jarre Moreano, Alanis Rafaela

C.C: 1310507734

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Autorización de Publicación

Yo, **Jarre Moreano, Alanis Rafaela**, con cédula de ciudadanía n°1310507734, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Título: "Identificación de genes expresados diferencialmente en ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial utilizando herramientas bioinformáticas"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.

Sangolquí, 27 de febrero del 2024

……..

Jarre Moreano, Alanis Rafaela

C.C: 1310507734

Dedicatoria

A mis padres, Danny & Katiuska, por siempre estar en primera fila de todos mis logros.

A mi hermanita, Tesla, porque me pone los pies sobre la tierra, es mi ancla y me motiva a seguir.

A mi amor, Emilio, porque me admira y siempre reconoce -y grita- mis triunfos, y, ha sido un apoyo fundamental cuando he tenido fracasos.

A mi mejor amiga, Diana, por hacerme entender que soy inteligente aun cuando me siento la persona más tonta del mundo.

A mis abuelos, Rafael y Oswaldo, que no están aquí conmigo ahora físicamente, pero, sepan que los pienso cada día, que quise ser ingeniera como ellos, y que, a pesar de todas las circunstancias, fueron la razón para seguir adelante con la carrera, y ahora con esta tesis. Sé que están en primera fila también.

A mis abuelas, Letty y Gardenia, porque son mujeres luchadoras y trabajadoras que siempre han sido mi ejemplo.

Al resto de mi familia y amigos, a los que de corazón les interesa mi felicidad y celebran mis triunfos. Gracias por siempre estar.

Alanis Rafaela Jarre Moreano

Agradecimientos

A mis padres, Danny & Katiuska, por siempre estar ahí para mí y por sobre todo por creer en mí y en mis capacidades desde siempre.

A mi hermanita, Tesla, porque a pesar de sus bromas, siempre está ahí para mí, incluso cuando siento que no puedo seguir.

A mi amor, Emilio, por prestarme su hombro para llorar cuando es necesario y decirme las palabras correctas para ayudarme a respirar profundo, centrarme y seguir. Gracias por ser mi luz.

A mi tutor, el Dr. Francisco Flores, por su paciencia, sus sugerencias y su ayuda incondicional. Por brindarme su confianza para poder avanzar con esta tesis.

A todos quienes fueron mis docentes, que han dejado una huella inigualable en mi corazón, por sus enseñanzas y lecciones de vida. Y también, agradezco a todos quienes fueron mis compañeros en cada una de sus clases, por ayudarme cuando lo necesité y ser buenos amigos.

A mis amigos que me dejó la ESPE, Massiel, Michelle, Diego, Joseph y Jean, porque a pesar de que ya no nos vemos, siguen pendientes de mí y son con quienes me puedo desahogar con temas que los demás no entenderían.

Alanis Rafaela Jarre Moreano

Índice de contenido

Listado de tablas

Listado de figuras

Listado de abreviaturas

EDG – Expresión diferencial de genes

FP – Fibrosis pulmonar

MSC – Células madre mesenquimales

ARN – Ácido Ribonucleico

ARNm – Ácido Ribonucleico mensajero

ADN – Ácido desoxirribonucleico

GO – Ontología génica

ENA - European Nucleotide Archive

FDR - Tasa de Descubrimiento Falso

Resumen

Los estudios bioinformáticos, enfocados en el análisis diferencial de genes en modelos de ratones *Mus musculus* con fibrosis pulmonar, ofrecen una vía innovadora para comprender los mecanismos moleculares, identificar patrones de expresión génica y evaluar tratamientos, con el potencial de impulsar avances significativos en el manejo clínico y el desarrollo terapéutico de la enfermedad. El control de calidad y evaluación de duplicados garantiza la integridad de los datos de RNA-seq, respaldando la alta calidad de las muestras y el éxito del proceso de alineación. El análisis de expresión diferencial utilizando muestras de control y dos tratamientos: el primero con células mesenquimales humanas (hMSC) con nanopartículas de óxido de hierro (Fe-hMSC) y el segundo con con células mesenquimales humanas (hMSC) con nanopartículas de óxido de hierro infusionadas con pioglitazona (PgFe-hMSC), los cuales revelan más de 12,000 genes diferencialmente expresados de alta calidad. El análisis de expresión génica con DESeq2 revela relaciones diferenciales entre tratamientos, destacando la efectividad potencial del tratamiento 2 con PgFe-hMSC. Además, el enriquecimiento funcional, especialmente con respecto a la función molecular dentro de la ontología génica, y la significancia en la vía de transducción olfatoria sugieren mecanismos relevantes en la fibrosis pulmonar, respaldando la utilidad de las herramientas bioinformáticas para comprender la patología y la respuesta a tratamientos específicos.

Palabras Clave: fibrosis pulmonar, control de calidad, análisis de expresión diferencial, enriquecimiento funcional

Abstract

Bioinformatics holds immense potential in deciphering the complexities of diseases like pulmonary fibrosis, particularly through gene expression analysis in *Mus musculus* mouse models. This approach allows for a deeper understanding of molecular mechanisms, gene expression patterns, and treatment efficacy, driving advancements in clinical management and therapeutic development. Central to this progress is rigorous quality control and duplicate evaluation, ensuring the integrity of RNA-seq data and enabling seamless alignment. Through meticulous scrutiny, researchers extract valuable insights from vast genomic information. Differential expression analysis compares distinct treatments with control samples. Notably, treatments involving human mesenchymal cells (hMSC) infused with iron oxide nanoparticles (Fe-hMSC) and hMSCs infused with iron oxide nanoparticles and pioglitazone (PgFe-hMSCs) reveal over 12,000 differentially expressed genes, illuminating molecular dynamics in pulmonary fibrosis. Further analysis with tools like DESeq2 reveals nuanced treatment relationships, emphasizing personalized therapeutic approaches tailored to individual genetic signatures. Functional enrichment analysis, particularly in gene ontology molecular functions, highlights critical pathways like olfactory transduction. These insights deepen our understanding of pulmonary fibrosis and underscore bioinformatics' crucial role in disease elucidation and targeted interventions. This synergy between bioinformatics and experimental research signals a new era in precision medicine. Comprehensive genomic analyses pave the way for more effective therapeutic strategies, offering hope to those battling pulmonary fibrosis and other challenging diseases.

Keywords: pulmonary fibrosis, quality control, differential expression analysis, functional enrichment

Capítulo I: Introducción

Formulación del problema

La fibrosis pulmonar, una enfermedad crónica y progresiva, se caracteriza por la formación excesiva de tejido cicatricial en los pulmones, afectando la función respiratoria y la calidad de vida de los pacientes (Kim et al., 2018). A pesar de los avances en la comprensión de los mecanismos, persisten desafíos importantes en la identificación precisa de factores genéticos y respuestas moleculares relacionadas con el desarrollo y progresión de la fibrosis pulmonar. En este contexto, la utilización de herramientas computacionales para realizar análisis diferencial de genes en modelos de ratones con fibrosis pulmonar emerge como una estrategia innovadora para abordar estas incógnitas (Wu et al., 2020).

Dentro de la literatura científica se ha demostrado el valor de las técnicas de análisis diferencial de genes para identificar patrones de expresión génica asociados con diversas enfermedades, incluida la fibrosis pulmonar (Torres-Soria et al., 2022). Estas herramientas permiten comparar los perfiles de expresión génica entre grupos de ratones con fibrosis pulmonar y aquellos sometidos a diferentes tratamientos, proporcionando así una visión detallada de las alteraciones moleculares. Sin embargo, la aplicación específica de estas herramientas en el contexto de tratamientos experimentales, y la comprensión detallada de cómo influyen en la expresión génica, aún se encuentran en una fase incipiente (Liu et al., 2020).

Existen estudios previos con base bioinformática entre los cuales, Wan et al. resumen sistemáticamente las causas y las vías de señalización de la senescencia celular en la FP, también analiza objetivamente el impacto de la senescencia en AEC y fibroblastos en el desarrollo del PF (Wan et al., 2023). Posteriormente se comprobó que la proteína de unión a la actina F (TRIOBP) modula la señalización de la beta catenina, mediante una regulación de miR-29b en la fibrosis pulmonar idiopática (Wang et al., 2024). En otro estudio, Gajjala et al.

utilizaron un modelo de ratón de fibrosis pulmonar inducida por bleomicina, mostró que la sobreexpresión de Sox9 específica de miofibroblastos aumenta la activación de los fibroblastos y la fibrosis pulmonar (Gajjala et al., 2021).

Identificar los cambios en la expresión génica en respuesta a diferentes intervenciones terapéuticas no solo arrojará luz sobre los mecanismos moleculares involucrados para fines de esta investigación, sino que también proporcionará información crucial sobre la eficacia de los tratamientos existentes y potenciales(Zeng et al., 2018).

Este estudio se propone explorar de manera detallada y sistemática la aplicación de herramientas computacionales para el análisis diferencial de genes en modelos de ratones con fibrosis pulmonar sometidos a tratamientos específicos. La formulación de este problema de investigación busca contribuir significativamente al conocimiento actual en el campo, ofreciendo perspectivas innovadoras y aplicables que podrían influir directamente en la atención clínica y el desarrollo de terapias más efectivas para la fibrosis pulmonar.

Justificación del problema

En el complejo entramado de enfermedades pulmonares, la fibrosis pulmonar emerge como un desafío médico significativo, afectando la calidad de vida de los pacientes. La necesidad de comprender los mecanismos de acción de esta patología se vuelve imperativa para mejorar las estrategias terapéuticas (Chen et al., 2018). La aplicación de la bioinformática para llevar a cabo un análisis diferencial de genes en modelos de ratones con fibrosis pulmonar y sometidos a tratamientos específicos se presenta como una vía prometedora. Esta aproximación puede arrojar luz sobre las complejas interacciones genéticas implicadas en la progresión de la fibrosis pulmonar, marcando un hito en la investigación biomédica actual (Leal et al., 2018).

La diversidad de tratamientos disponibles para la fibrosis pulmonar destaca la necesidad crítica de identificar marcadores genéticos específicos que respondan de manera diferencial a las intervenciones terapéuticas (Romero et al., 2022). El análisis computacional de la expresión génica en modelos murinos ofrece la posibilidad de discernir patrones moleculares sutiles que podrían no ser evidentes mediante métodos convencionales. Esta perspectiva no solo proporcionaría una comprensión más profunda de la eficacia de los tratamientos actuales, sino que también podría revelar nuevos objetivos terapéuticos emergentes (Melia & Waxman, 2020).

La complejidad de la fibrosis pulmonar y su variabilidad en la respuesta a tratamientos plantea interrogantes significativos en la práctica clínica. La implementación de la bioinformática para el análisis diferencial de genes busca abordar estas variabilidades, ofreciendo una visión más personalizada de las respuestas terapéuticas (Christov et al., 2022). En este sentido, este enfoque no solo podría contribuir a la estratificación de pacientes según su perfil genético, sino también a la identificación de subgrupos que se beneficiarían de tratamientos específicos (Chen et al., 2018).

La falta de estudios específicos que exploren a fondo la aplicación de instrumentos bioinformáticos en el análisis diferencial de genes en modelos de fibrosis pulmonar constituye una brecha evidente en la literatura biomédica actual (Ayoob & Kangas, 2020). La justificación de esta investigación radica en conocer las diferencias entre dos tipos de terapias de reposición mitocondrial con la finalidad de que a futuro se puedan utilizar a manera de terapias más precisas y personalizadas. La aplicación de herramientas computacionales en este contexto no solo ampliará nuestra comprensión de los mecanismos moleculares, sino que también contribuirá directamente a mejorar los resultados clínicos para los pacientes afectados por la fibrosis pulmonar (Blasco et al., 2019).

En la actualidad, se ha destacado la relevancia de la transferencia mitocondrial como un elemento crucial para preservar la funcionalidad de las mitocondrias y prevenir la apoptosis en enfermedades. Este reconocimiento subraya su importancia prospectiva en los procesos patológicos (Clemente-Suárez et al., 2023). Con el progreso de las metodologías moleculares y bioquímicas, se ha logrado una comprensión más profunda de los mecanismos involucrados en los trastornos mitocondriales asociados con diversas enfermedades. Esto ha posicionado a las mitocondrias como un objetivo de gran relevancia tanto para instituciones de investigación como para la industria farmacéutica (Zhang & Miao, 2023). La disfunción mitocondrial, característica común en numerosas enfermedades, que abarcan trastornos neurodegenerativos, desórdenes metabólicos y cáncer, ha intensificado aún más el interés en estas estructuras celulares (Liu et al., 2022).

La utilización de la bioinformática para el análisis diferencial de genes en modelos murinos con fibrosis pulmonar y sometidos a tratamientos específicos representa un campo de investigación vital y prometedor (Matthews et al., 2021). En el Ecuador, existen investigaciones sobre la fibrosis pulmonar a nivel molecular, sin embargo, no existen investigaciones con la utilización de herramientas bioinformáticas. Esta estrategia no solo podría transformar nuestra comprensión de la enfermedad a nivel molecular, sino que también podría tener un impacto significativo en la personalización de las terapias, abriendo nuevas perspectivas para mejorar la calidad de vida de los pacientes con fibrosis pulmonar (Torres-Soria et al., 2022). En conjunto, estos pasos pueden impulsar la adaptación de la atención médica a la genética nacional y mejorar la salud y el bienestar de la población ecuatoriana.

Si nos enfocamos en lo epidemiológico, la incidencia de la fibrosis pulmonar es de 6.8-17.4 casos nuevos por cada 100 000 habitantes anualmente; sin embargo, nuevos estudios que han estimado una cifra de 14-42.7 casos por cada 100000 habitantes de manera anual, por

lo que se estima un promedio de 4760 nuevos casos cada año, con una incidencia más grande en mujeres (Remón et al., 2016).

Objetivos de la investigación

Objetivo general

Identificar genes expresados diferencialmente en ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial utilizando herramientas bioinformáticas.

Objetivos específicos

- Obtener secuencias crudas de RNAseq de ratones con fibrosis pulmonar expuestas a dos terapias de reposición mitocondrial a partir de bases de datos públicas.
- Limpiar y normalizar las secuencias de RNAseq obtenidas
- Identificar genes con expresión diferencial entre los tratamientos.
- Realizar un análisis funcional de los genes identificados

Capitulo II: Marco Teórico

Fibrosis pulmonar

La fibrosis pulmonar, una afección crónica del pulmón, puede ser potencialmente mortal debido a diversas causas (Chen et al., 2018). La patogénesis de esta enfermedad sigue siendo poco clara, y actualmente, no hay un tratamiento farmacológico efectivo, lo que resulta en un pronóstico clínico desfavorable (Pourgholamhossein et al., 2018; Yang et al., 2019).

Esta condición se caracteriza por la acumulación de miofibroblastos y matriz extracelular en el pulmón, desencadenando inflamación y provocando la pérdida de la estructura pulmonar original debido al depósito excesivo y desorganizado de colágeno y matriz extracelular. Este proceso conduce a la remodelación pulmonar y, en última instancia, a la insuficiencia respiratoria, manifestándose en el engrosamiento y la cicatrización del tejido conectivo pulmonar (Li et al., 2021; Wu et al., 2020; Xin et al., 2019).

Estudios extensos han establecido que un aumento en el número de fibroblastos, combinado con su excesivo depósito de matriz extracelular (ECM) en el pulmón, resulta en la destrucción de la estructura alveolar, la reducción de la distensibilidad pulmonar y la alteración de la función de intercambio de gases (Leal et al., 2018).

A pesar de estos conocimientos, la causa de esta enfermedad sigue siendo desconocida. Actualmente se cree que la formación de fibrosis pulmonar implica diversos factores, incluyendo la matriz extracelular, citocinas y quimiocinas inflamatorias(Tam et al., 2020). La actina alfa del músculo liso (α-SMA) se identifica como una marca de miofibroblastos, que se originan a partir de fibroblastos y desempeñan un papel crucial en la producción de matriz extracelular y citocinas relacionadas con la fibrosis(Yang et al., 2019).

Cuando la causa es desconocida, se denomina fibrosis pulmonar idiopática, siendo la forma más común de esta afección. Este trastorno afecta eventualmente a lóbulos pulmonares

completos, iniciando con cambios fibróticos microscópicos en las áreas periféricas que avanzan gradualmente hacia el interior, y su progresión puede conducir a la insuficiencia respiratoria (Leal et al., 2018; Simon et al., 2023).

A pesar de su sombrío pronóstico, la tasa de supervivencia se ha informado tan baja como de 2 a 5 años después del diagnóstico de fibrosis pulmonar. Además de esta baja supervivencia, los pacientes se enfrentan a la escasez de tratamientos eficaces(Andugulapati et al., 2020).

Los factores genéticos y epigenéticos siguen siendo fundamentales en el desarrollo del proceso fibrótico, aunque la contribución y la interacción de las variantes identificadas aún no están completamente esclarecidas(Torres-Soria et al., 2022). En este contexto de susceptibilidad genética, se reconoce que las microlesiones repetidas en el epitelio alveolar son el impulsor inicial de un proceso de reparación alterado, donde diversas células pulmonares desarrollan comportamientos anómalos, contribuyendo al desarrollo y mantenimiento del proceso fibrótico (Sgalla et al., 2018).

Tratamientos para la Fibrosis Pulmonar

A lo largo del tiempo, se han desarrollado diversas opciones terapéuticas y protocolos para abordar la fibrosis pulmonar, aunque los resultados en términos de curación o mejora en la calidad de vida han sido limitados. A pesar de la evidencia de la contribución de las citocinas en la respuesta inflamatoria y su relevancia en modelos experimentales de fibrosis pulmonar, la mayoría de los enfoques terapéuticos destinados a bloquear la inflamación pulmonar excesiva han mostrado fracasos en las etapas preclínicas (Simon et al., 2023).

Terapias de Reposición Mitocondrial

Las mitocondrias, orgánulos fundamentales con un papel no solo en la generación de energía, sino también en la señalización celular, desempeñan un papel crucial en el

mantenimiento de la homeostasis celular (Zhang & Miao, 2023). Existe una sólida base que respalda la importancia de las mitocondrias en este equilibrio, y numerosas pruebas indican que la disfunción mitocondrial conlleva consecuencias perjudiciales (Liu et al., 2022). La transferencia mitocondrial, que involucra la transmisión de componentes mitocondriales, como ADN, ARN y proteínas, entre células, ha demostrado mejorar la funcionalidad mitocondrial y reducir el estrés oxidativo en diversos tipos celulares. Diversos estudios han planteado la hipótesis de que la transferencia mitocondrial podría constituir un enfoque terapéutico prometedor para abordar enfermedades vinculadas con la disfunción mitocondrial (Clemente-Suárez et al., 2023).

Células Madre Mesenquimales. Las células madre mesenquimales (MSC) representan células progenitoras multipotentes con capacidad de proliferación y regeneración. Estas células mesenquimales, presentes en diversos tejidos adultos como médula ósea, grasa, piel, placenta y corazón, tienen la capacidad de migrar fácilmente a través de los vasos sanguíneos en respuesta a daños, mediados por factores inflamatorios y la invocación de células inflamatorias(Saleh et al., 2022).

Las MSC desempeñan un papel crucial en la modulación de la proliferación, activación y función efectora de las células inmunes, siendo relevantes en la patogénesis de enfermedades pulmonares inflamatorias, tanto agudas como crónicas (Wan et al., 2023b).

Estas células estromales, presentes en la matriz de varios tejidos, como cartílago, hueso, grasa, médula ósea y derivados del mesodermo, exhiben características de células madre, incluyendo capacidad de autorrenovación y plasticidad. Esta capacidad les permite proliferar y diferenciarse en diversos tipos celulares, contribuyendo a la formación de órganos en circunstancias específicas (Choi et al., 2023). Se ha demostrado que las MSC tienen beneficios terapéuticos en diversas enfermedades, como insuficiencia cardíaca isquémica,

hipertensión arterial pulmonar, accidentes cerebrovasculares, enfermedad renal crónica y sepsis (Li et al., 2021).

En este estudio, se emplearon dos tipos de MSC: las MSC humanas con nanopartículas de oxído de hierro Fe-hMSC, y las MSC humanas con nanopartículas de óxido de hierro infusionadas con pioglitazona, el cual es un medicamento para tratar la diabetes tipo 2, denominadas Pg-Fe-hMSC (Huang et al., 2023).

La Biología Computacional

En la era digital, la biología se ha vuelto dependiente de la computación y la colaboración. Los proyectos de investigación contemporáneos pueden abarcar múltiples sistemas de modelos, hacer uso de diversas tecnologías de ensayo, recopilar datos de diferentes tipos y necesitar estrategias computacionales complejas. Esto hace que el diseño y la ejecución efectivos sean desafiantes o incluso imposibles para un científico individual (Way et al., 2021).

La biología computacional desempeña un papel crucial en el análisis de redes biológicas, donde se construyen y analizan redes para comprender las interacciones entre genes, proteínas y metabolitos en procesos biológicos y enfermedades (Adams et al., 2022). Este enfoque se basa en un marco teórico llamado modelo gráfico, que representa la estructura biológica y el flujo funcional de información a través de ella. La comprensión y modelado de la estructura de la red proporcionan un mejor conocimiento de sus mecanismos evolutivos y de sus comportamientos dinámicos y funcionales (Liu et al., 2020).

En el descubrimiento de fármacos, la biología computacional contribuye al caracterizar los mecanismos moleculares de unión al ligando, identificar sitios activos/de unión y refinar la estructura de las posiciones de unión del ligando-objetivo. La mayoría de estos enfoques

destacan la importancia de determinar con precisión los sitios activos/de unión en la proteína diana (Zhang et al., 2022).

Dentro del ámbito de la biología computacional, la biomedicina destaca como un campo relevante que utiliza diversas herramientas y técnicas informáticas para abordar cuestiones médicas y biológicas de importancia, como el estudio del comportamiento de tratamientos para la fibrosis pulmonar a nivel genético (Hong et al., 2022).

La ciencia biomédica, que abarca disciplinas como epidemiología clínica, microbiología médica e ingeniería biomédica, tiene como objetivo desarrollar nuevas terapias, tratamientos y tecnologías para el manejo de enfermedades, dolencias y discapacidades (Blasco et al., 2019).

Secuenciación de ARN (RNA-Seq)

El ácido ribonucleico (ARN) ocupa una posición central en el dogma central de la biología molecular, que establece que el ADN se transcribe en ARN y este último se traduce en proteínas. No obstante, investigaciones recientes han revelado un papel en constante expansión del ARN en las células (Chen & Wong, 2019). El ARN, una molécula monocatenaria compuesta por cuatro nucleótidos (adenina, guanina, uracilo o citosina), no solo participa en la transmisión de información genética, sino que también regula la expresión génica (Jovic et al., 2022). Además de la información de la secuencia, las modificaciones químicas agregan complejidad al ARN, emergiendo como una nueva capa de regulación de la expresión génica (Zhang et al., 2022).

El ARN cumple diversas funciones, proporcionando un andamiaje físico para complejos proteicos, catalizando reacciones químicas, regulando la replicación del ADN, y controlando la transcripción y traducción de proteínas (Chen & Wong, 2019) . Para llevar a cabo sus funciones multifacéticas, el ARN se clasifica según su tamaño, estructura, modificaciones químicas y secuencia, constituyendo el transcriptoma de ARN en una célula (Young et al., 2010). Dada la

amplitud del transcriptoma de ARN, que incluye >20,000 genes y cientos de miles de transcripciones en células humanas, se requieren herramientas y plataformas específicas para analizar, procesar y evaluar estas moléculas cruciales (Tan et al., 2021).

El ARN desempeña diversas funciones biológicas no solo a través de secuencias de nucleósidos canónicos, sino también mediante múltiples modificaciones estructurales, tanto conocidas como desconocidas (Zhang et al., 2019). La secuenciación de ARN (RNA-Seq) es una poderosa técnica utilizada en la investigación molecular para analizar el perfil de expresión génica a nivel de ARN en una muestra biológica (Yépez et al., 2022).

El avance continuo de la secuenciación de ARN ha llevado el análisis del transcriptoma a una nueva era, con mayor eficiencia y menor costo (Hong et al., 2022; Jovic et al., 2022). La secuenciación de ARN ofrece valiosos conocimientos para la investigación y tratamiento de diversas enfermedades, y se anticipa que, con la medicina de precisión, será ampliamente utilizada para estudiar diferentes tipos de enfermedades (Hong et al., 2020).

RNA-Seq, al examinar directamente la abundancia y secuencia de la transcripción en todo el transcriptoma, facilita la identificación sistemática de eventos de transcripción aberrantes, como genes expresados en niveles anómalos, empalmes incorrectos de genes y variantes raras expresadas monoalélicamente (Yépez et al., 2022).Además de mejorar el diagnóstico, RNA-Seq puede aportar comprensión sobre los mecanismos patogénicos moleculares y los fundamentos genéticos de las variantes(Tan et al., 2021). Aunque estos estudios iniciales son prometedores, la implementación clínica rutinaria de RNA-Seq requiere flujos de trabajo computacionales sólidos, controles de calidad establecidos y adecuada cantidad y profundidad de secuenciación del material de ARN(Begik et al., 2022).

La secuenciación de ARN de próxima generación (RNA-Seq) es una técnica establecida y versátil utilizada en investigación molecular para detectar secuencias enriquecidas en tejidos

y momentos específicos, permitiendo caracterizar la expresión génica diferencial en respuestas a diversos estímulos (Kumar & Kirti, 2023).

Control de Calidad

Los datos representan el fundamento esencial de cualquier investigación, ya que la calidad de los resultados está directamente vinculada a la calidad de los datos recopilados (Gliklich et al., 2014). Galaxy emerge como una plataforma que facilita, mediante pasos simples, la utilización de sistemas y métodos de datos no identificados para lograr, evaluar o controlar la calidad de los datos de investigación (AbuHalimeh, 2022).

La alta calidad de los datos es crucial, requiriendo precisión y idoneidad para el análisis estadístico (Bhatt, 2023). Más importante aún, los datos de alta calidad deben limitar su "nivel de variación aceptable" de manera arbitraria para no comprometer las conclusiones derivadas del análisis estadístico. Además, deben cumplir con los requisitos reglamentarios pertinentes especificados para la calidad de los datos (Krishnankutty et al., 2012).

Durante el proceso de secuenciación, es común que se introduzcan errores, como la incorporación de nucleótidos incorrectos a la secuencia. Estos errores, derivados de las limitaciones técnicas de cada plataforma de secuenciación, pueden distorsionar el análisis y llevar a una interpretación errónea de los datos (Hiltemann et al., 2023). La presencia de adaptadores también es posible si las lecturas son más extensas que los fragmentos secuenciados, y la eliminación de estos puede mejorar el número de lecturas alineadas (Begik et al., 2022).

Mapeo

Para dar sentido a las lecturas, es esencial identificar el origen de las secuencias dentro del genoma y determinar a qué genes pertenecen. Este proceso, conocido como mapeo de las lecturas con la referencia, se lleva a cabo al tener un genoma de referencia para el organismo

(Hiltemann et al., 2023). En este proyecto de investigación, se emplearon datos de RNA-seq de *Mus musculus*, y a pesar de que el grupo control presenta fibrosis pulmonar, las secuencias de alta calidad se asignaron al genoma de referencia proporcionado por Galaxy para *Mus musculus.*

El mapeo representa un paso crítico en la interpretación de los datos de RNA-seq, ya que implica asignar lecturas a características genómicas (Koch et al., 2018). Este proceso nos permite evaluar la expresión de una característica, como un gen, midiendo el número de lecturas asignadas a dicho gen. En muchos casos, no es posible asignar cada lectura de manera única a un gen específico, ya que algunas lecturas pueden tener "mapeo múltiple", es decir, ser compatibles con varios genes (Rotwein, 2019).

Análisis Diferencial de expresión de genes

La expresión genética es el proceso mediante el cual la información contenida en el ADN se convierte en instrucciones para la producción de proteínas u otras moléculas, implicando la transcripción del ADN en ARN mensajero (ARNm) seguida de su traducción en proteínas (Coker et al., 2019). El análisis de la expresión génica se utiliza para evaluar el patrón de alteraciones genéticas que se producen en condiciones específicas, en un tejido o en una célula individual. Esto implica medir la cantidad de transcripciones de ADN presentes en una muestra de tejido o células para obtener información sobre qué genes se expresan y en qué niveles (Weger et al., 2021).

En la cuantificación de la expresión génica, se compara el número de lecturas secuenciadas con el número de pares de bases secuenciadas de un fragmento de ADN en relación con una fuente genómica o de transcriptoma reconocida (Tasker et al., 2017). La precisión de esta cuantificación depende de que las lecturas secuenciadas proporcionen suficiente información distintiva para permitir la aplicación de algoritmos bioinformáticos que

correlacionen las lecturas con los genes apropiados según el número de pares de bases secuenciadas (Alharbi & Vakanski, 2023).

La expresión diferencial de genes se refiere al análisis e interpretación de las variaciones en la abundancia de transcripciones génicas dentro de un transcriptoma. Las listas de genes que difieren entre dos conjuntos de muestras generalmente se obtienen mediante herramientas de análisis de datos de RNA-seq o mediante pruebas estadísticas de conjuntos de datos (Chen & Wong, 2019).

Aunque los métodos moleculares actuales para el análisis de la expresión génica son herramientas probadas y poderosas, presentan desafíos como la sensibilidad inadecuada, la dependencia crítica de la calidad del ARN, limitaciones en la capacidad multiplex, altos costos y tiempos prolongados para obtener resultados (Farhang-Ghahremani et al., 2023).

La comparación de patrones diferenciales de expresión genética ha permitido identificar elementos comunes significativamente enriquecidos en clases de genes con funciones específicas, como la síntesis de proteínas, la administración de hormonas y la plasticidad morfológica (Kwak et al., 2020).

El análisis de expresión diferencial significa tomar los datos del recuento de lecturas normalizados y realizar un análisis estadístico para descubrir cambios cuantitativos en los niveles de expresión entre grupos experimentales (Ritchie et al., 2015).El objetivo de las pruebas de expresión diferencial es determinar qué genes se expresan en diferentes niveles entre condiciones. Estos genes pueden ofrecer información biológica sobre los procesos afectados por las condiciones de interés (Kim et al., 2019).

El análisis de expresión génica diferencial, derivado de datos de RNA-seq, generalmente incluye tres etapas: normalización de recuentos, estimación de parámetros del

modelo estadístico y pruebas de expresión diferencial (Rapaport et al., 2013). En esta tesis, se utilizó la plataforma Galaxy para realizar el análisis de expresión genética.

Galaxy Project

Galaxy Project es una herramienta integral en el análisis diferencial de genes, proporcionando una interfaz web que incorpora las herramientas necesarias y facilita la reproducibilidad (Hiltemann et al., 2023).En el ámbito de la investigación genómica, Galaxy Project se destaca como una plataforma colaborativa que revoluciona el análisis genómico diferencial, ofreciendo una interfaz accesible y poderosa para realizar análisis genómicos de manera eficiente (Spoor et al., 2020).

Al abordar el análisis diferencial de genes, Galaxy es una herramienta esencial para la identificación de patrones de expresión génica que pueden ser cruciales en el estudio de diversas condiciones y enfermedades (Damiani et al., 2020). Con un enfoque modular y una interfaz gráfica, Galaxy simplifica la creación de flujos de trabajo específicos para cada investigación, desde la preparación de datos hasta la interpretación de resultados, permitiendo la reproducción y compartición transparente de análisis (Mirela-Bota et al., 2021).

Análisis de Enriquecimiento Funcional u Ontología Génica

Los genes codifican productos génicos, que a menudo son proteínas, pero también pueden ser moléculas de ARN no codificantes (ARNnc), desempeñando funciones a nivel molecular, celular y orgánico (Chen et al., 2021). La Ontología Génica (GO) es una base de conocimientos que proporciona una representación completa, estructurada y computacionalmente accesible de las funciones genéticas de genes en cualquier organismo celular o virus (Carbon et al., 2021). En la investigación en ciencias biológicas, el recurso GO se ha vuelto fundamental, respaldando el análisis de experimentos y sistemas biológicos a gran escala (Aleksander et al., 2023).

La GO es una ontología biomédica que utiliza un vocabulario controlado de términos para describir funciones fisiológicas normales de entidades biológicas, como proteínas y ARNnc, en diversas especies y campos biológicos, de manera consistente y computacionalmente accesible (Kramarz et al., 2020). Los términos GO se asocian manualmente con entidades biológicas por biocuradores científicos, basándose en información experimental publicada, y automáticamente por conductos electrónicos, utilizando criterios de similitud cuidadosamente diseñados (Aleksander et al., 2023).

Los enlaces resultantes entre los términos GO y las entidades biológicas se denominan "anotaciones". La GO comprende tres categorías de términos que describen funciones moleculares, procesos y componentes celulares (Wood et al., 2020).

- Función molecular (MF): Actividades realizadas por un producto genético a nivel molecular.
- Componente celular (CC): Ubicaciones relacionadas con las estructuras celulares donde se realizan las funciones moleculares.
- Proceso biológico (BP): Programas biológicos que comprenden actividades moleculares que actúan en conjunto para lograr un resultado particular a nivel celular u organismo de organismos multicelulares (Jezernik et al., 2022).

Cada gen puede estar involucrado en diversos procesos y funciones, representados en GO como múltiples anotaciones para un solo gen (Wood et al., 2020). Sin embargo, debido a limitaciones espaciales, funcionales o temporales, es poco probable que los mismos genes lleven a cabo ciertas combinaciones de funciones o procesos. Identificar pares de términos GO que es poco probable que estén anotados correctamente en los mismos genes puede indicar posibles anotaciones erróneas (Li et al., 2022).

ShinyGO

ShinyGO es una aplicación Shiny basada en paquetes de R/Bioconductor y una extensa base de datos de anotaciones y rutas recopiladas de diversas fuentes (Ge et al., 2020). Aprovechando el marco Shiny, que permite el acceso a varios paquetes R para visualización y análisis estadísticos, ShinyGO presenta características únicas, como mostrar genes de consulta en diagramas de rutas y redes PPI basadas en el acceso a la interfaz del programa de aplicación (API) a KEGG y STRING. También permite visualizar superposiciones entre rutas enriquecidas utilizando agrupaciones jerárquicas y redes interactivas, e identificar diferencias estadísticamente significativas en el tipo de gen, longitud, contenido de GC y distribución cromosómica entre los genes de consulta y el fondo

Hipótesis

Al tratar la fibrosis pulmonar con dos tipos de células madre mesenquimales (MSC) modificadas: una con nanopartículas de óxido de hierro (Fe-hMSC) y la otra con nanopartículas de óxido de hierro infusionadas con pioglitazona (Pg-Fe-hMSC), se observa una expresión génica diferencial de los mismos genes. La expresión génica diferencial implica que los niveles de expresión de ciertos genes varían entre los tratamientos con Fe-hMSC y Pg-Fe-hMSC, lo que sugiere que estos tratamientos pueden tener efectos distintos en los mecanismos moleculares implicados en la fibrosis pulmonar.

Capítulo III: Metodología

El presente trabajo de investigación se estableció en dos fases: Como primera fase se tuvo el pre-análisis que consiste en los pasos que llevan hasta el control de calidad, y; como segunda fase, la fase de análisis, que incluye todos los pasos para tanto el análisis de EDG como el análisis de enriquecimiento funcional.

Pre-análisis

Participantes

El presente trabajo de investigación estuvo elaborado por Alanis Rafaela Jarre Moreano, egresada de la Carrera de Ingeniería en Biotecnología, con ayuda de la tutoría del Ing. Francisco Flores Ph. D., quien es docente investigador del departamento de Ciencias de la Vida y la Agricultura en la Universidad de las Fuerzas Armadas – ESPE. En cuanto al financiamiento de la investigación, este mismo no fue necesario debido al trabajo bioinformático, sin embargo, cualquier gasto mínimo estuvo a cargo de la estudiante.

Zona de estudio

Este trabajo investigativo fue realizado en la ciudad de Portoviejo, mediante el uso de plataformas bioinformáticas gratuitas, disponibles en la web.

Duración de la investigación

Este proyecto de investigación tuvo una duración de aproximadamente 4 meses y medio, iniciando a finales de septiembre del 2023 y, estuvo culminado para enero del 2024.

Elección de las plataformas de estudio

El proyecto Galaxy (https://galaxyproject.org/) inició en el 2005, con el objetivo de permitir la ciencia biomédica basada en datos para análisis accesible y reproducibles independientemente, y también una comunicación transparente de los análisis(Afgan et al., 2018). Esta plataforma fue seleccionada debido a su interfaz sencilla de manejar, y porque no

necesita codificación. Una vez cargados los datos seleccionados para secuenciarse, se utilizó este servidor gratuito para realizar el control de calidad, mapeo y análisis diferencial de genes.

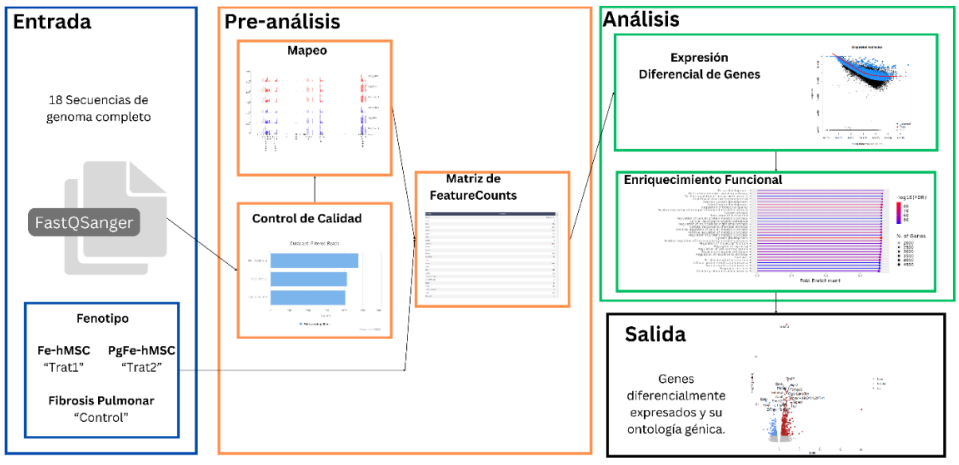
La plataforma Shiny, permite el acceso a muchos paquetes R potentes para visualización y análisis estadísticos, desarrollamos una nueva herramienta basada en la base de datos de anotaciones en Ensembl y bases de datos de rutas de muchas otras fuentes (Ge et al., 2020). Una vez obtenidos los genes diferencialmente expresados, se procedió a utilizar la plataforma ShinyGO, para el proceso de enriquecimiento funcional u Ontología de Genes, lo cual nos permitió realizar un análisis en profundidad de listas de genes, con visualización gráfica de enriquecimiento, vías, características de genes e interacciones de proteínas (Ge & Jung, 2021).

Esquema de desarrollo

Este esquema de desarrollo fue realizado posterior a un análisis de todos los pasos a seguir para este proyecto y se realizó un resumen para el mismo, se realizó en Canva con la autoría de Alanis Jarre para poder esquematizar de manera gráfica y sencilla los pasos para este proyecto.

Figura 1

Esquema de desarrollo



Set de datos

Los datos se encontraron en las bases de datos del NCBI, los cuales fueron publicados en dicha plataforma para un artículo de investigación científica con el tema: “Intervención eficiente para la fibrosis pulmonar mediante transferencia mitocondrial promovida por la biogénesis mitocondrial” por su traducción a español. Realizada por Huang et al., 2023; con DOI: 10.1038/s41467-023-41529-7. Se extrajeron del European Nucleotide Archive (ENA), donde estaban bajo el código: PRJNA948365.

Estos datos abarcan 3 grupos de muestras emparejadas tomadas de ratones *Mus musculus* con fibrosis pulmonar (FP) como control, y 3 grupos de muestras tomadas de ratones en las mismas condiciones, pero cada uno tratado con uno de los tratamientos: Fe-hMSC y Pg-Fe-hMSC. Esto resultó en un total de 9 pares de muestras que posteriormente fueron sometidos a los distintos tratamientos dentro del servidor Galaxy.

Control de calidad

Por lo previamente mencionado, el control de calidad de secuencias es un paso esencial en cualquier análisis. En Galaxy se utilizaron distintas herramientas para poder lograrlo: FastQC para crear los reportes de la calidad de la secuencia; MultiQC para combinar los reportes generados; y, Cutadapt para mejorar la calidad de las secuencias por medio de cortes y filtros (Tekman et al., 2021).

Luego del cargado de los datos, se procede a crear una colección, para que aquí puedan estar como lista de datos pareados. Sin embargo, con la versión de MultiQC que existe actualmente, no se pueden utilizar colecciones pareadas, por lo que se transformaron a una lista con datos forward y reverse, con la opción flatten collection.

Luego de tener la colección como pares de un alista simple, se vuelve a utilizar el programa FastQC con el output que tuvimos, que finalmente se vuelve a utilizar MultiQC, utilizando el output Raw data de FastQC.

Posteriormente se utiliza Cutadapt para poder filtrar por tamaño y, lo cual se seleccionó un tamaño mínimo de 20 bp, y se seleccionó para que se pueda utilizar el archivo con MultiQC para que se generen los gráficos.

Mapeo

Para este proceso en la plataforma Galaxy, se utiliza RNA STAR para el mapeo con el genoma de referencia del *Mus musculus* (GRCm39), y, MultiQC para el reporte final. Después del mapeo, ahora tenemos información sobre dónde se encuentran las lecturas en el genoma de referencia y qué tan bien se mapearon. El siguiente paso en el análisis de datos de RNA-Seq es la cuantificación del número de lecturas asignadas a características genómicas (Afgan et al., 2018).

En RNA Star, se selecciona la colección que ha sido previamente cortada con Cutadapt, y se obtuvo el genoma de referencia en este caso de *Mus musculus* poniendo que como output nos de una lectura de gen por lectura contada como GeneCounts. Posteriormente, con los resultados de STAR se utiliza MultiQC, se utiliza específicamente el output Log obtenido.

Coteo de Número de Lecturas por gen anotado

Nuevamente utilizando MultiQC con el output Gene counts de STAR, y utilizando el archivo generado por RNA Star de lecturas por gen.

Para hacer la comparación de la expresión de cada gen en distintas condiciones, se requiere hacer previamente una cuantificación de número de lecturas por gen, o para poder ser más específicos, el número de lecturas mapeadas en los exones de cada gen (Hiltemann et al., 2023). Para esto en Galaxy se tienen dos opciones, de las cuales se utilizó FeatureCounts, ya

que esta opción ofrece una manera de poder adaptar el conteo de lecturas a lo que se necesita.

En featureCounts se utilizó el archivo generado con RNA Star mapped.bam, se seleccionó el archivo de anotación de genes de la historia, el cual sería el del genoma GRCm39*;* como filtro se seleccionó exones*,* como output se seleccionó Gene-ID, se seleccionna que se cree un archivo de tamaño de los genes. Se seleccionó también que si hay pares se cuenten como si fueran un solo fragmento, y una calidad mínima de mapeo por lectura de 10 bp. Posteriormente se vuelve a utilizar el MultiQC con el archivo de resumen que se ha generado por feature counts.

Análisis

Análisis de expresión diferencial de genes

Lo que nos ofrece FeatureCounts es una tabla con el número de lecturas asignadas a cada gen en la anotación proporcionada (Damiani et al., 2020). Esta tabla es lo que necesitamos como punto de partida para realizar el análisis de expresión diferencial de genes. Este análisis sirve para permitir identificar la expresión diferencial de genes inducida por los distintos tratamientos.

Para poder realizar el análisis de expresión de genes se utiliza la herramienta DESeq2, la cual sirve perfecto para datos de RNA-seq y para realizar el análisis de EDG. Primero, todos los archivos obtenidos en conteo de lecturas se hacen un solo archivo en una sola tabla y aplica la normalización a profundidad de secuenciación y composición de biblioteca.

Lo que hace el DESeq2 cumple con dos pasos fundamentales que son: 1) Se estima la varianza biológica utilizando las réplicas para cada condición, y; 2) Se estima la importancia de la expresión de genes entre dos condiciones, cualquiera que estas sean (Afgan et al., 2018).

El análisis EDG se resuelve a partir de recuentos de lecturas, y su objetivo es corregir la variabilidad en las mediciones utilizando las réplicas, las cuales son de mucha importancia para resultados más precisos (Ritchie et al., 2015).

Análisis de enriquecimiento funcional

Luego de obtener todos los genes expresados diferencialmente, se prosiguió a extraer la lista de genes, luego se procedió a procesar los genes en la plataforma ShinyGO (Ge & Jung, 2021). En donde una vez procesados los genes se tienen distintas ventanas para observar los resultados entre las cuales destacan: Enriquecimiento, cuadros estadísticos, árbol de relación de los genes, la cadena de genes, las vías a las cuales pertenecen los genes, gráficos estadísticos, etc (Ge et al., 2020). Aquí se pueden tomar como referencia principal: El componente celular, la función molecular, o, procesos biológicos; para los distintos análisis.

.

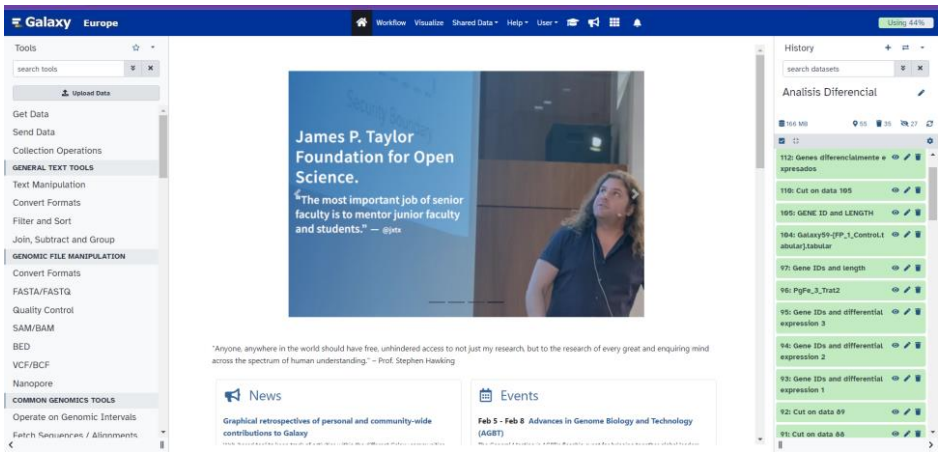
Capítulo III: Resultados

Plataforma utilizada

En la interfaz de Galaxy (Figura 2) se puede observar la barra de tareas realizadas al lado derecho, las herramientas del lado izquierdo en donde solo tipeando la herramienta que se desea utilizar aparecen las opciones.

Figura 2

Interfaz de la plataforma Galaxy



Set de datos

El set de datos se extrajo de la plataforma del ENA, como el proyecto PRJNA948365, en donde se descargaron los 18 archivos disponibles en formato fasta. Dentro de estos archivos se tenían 3 repeticiones de cada control, tratamiento 1 (Fe-hMSC) y tratamiento 2 (PgFe-hMSC); y todos los datos estaban en pares con una longitud de 8572864428 pares de bases cada una.

A cada uno de los datos se le puso su propio nombre como indicativo, los de grupo control con el nombre FP_#_Control, en donde el # identificaba si era la muestra de la repetición 1, 2 o 3. Los del grupo tratado con Fe-hMSC, se escogió como nombre Fe_#_Trat1, para señalar que era el tratamiento 1 y el número de repetición; mientras que el grupo tratado

con PgFe-hMSC, se denominó PgFe_#_Trat2. Estas denominaciones serán importantes posteriormente para usar DESeq2, ya que se necesita etiquetar las muestras.

Análisis de control de calidad

En la primera parte del análisis de control de calidad, como se mencionó en la metodología se pasa por FastQC, lo cual nos da un resumen de los datos (tabla 1), en donde se puede observar los porcentajes de duplicados, porcentaje de GC en las secuencias, y el tamaño de las secuencias en millones.

Tabla 1

Datos estadísticos generales obtenidos de Galaxy utilizando MultiQC

Nombre de la muestra	% Duplicados	% GC	Seq en Millones
FP_1_Control_forward	52.2%	51%	23.9
FP_1_Control_reverse	51.6%	51%	23.9
Fe_1_Trat1_forward	51.5%	51%	20.8
Fe_1_Trat1_reverse	50.3%	51%	20.8
PgFe_1_Trat2_forward	60.6%	49%	20.3
PgFe_1_Trat2_reverse	59.5%	49%	20.3
FP_2_Control_forward	51.4%	51%	21.5
FP_2_Control_reverse	50.7%	51%	21.5
Fe_2_Trat1_forward	52.1%	50%	22.3
Fe_2_Trat1_reverse	49.5%	50%	22.3
PgFe_2_Trat2_forward	57.3%	48%	21.9
PgFe_2_Trat2_reverse	53.4%	48%	21.9
FP_3_Control_forward	50.3%	51%	22.2
FP_3_Control_reverse	48.8%	51%	22.2
Fe_3_Trat1_forward	49.7%	50%	22.3
Fe_3_Trat1_reverse	47.9%	50%	22.3
PgFe_3_Trat2_forward	54.9%	50%	23.6
PgFe_3_Trat2_reverse	52.6%	50%	23.6

Recorte de datos

Dentro de los resultados obtenidos por Cutadapt, en donde se filtraron las secuencias con una longitud de mínimo 20 bp. Sin embargo, se puede notar que todas las secuencias pasaron este filtro, por lo cual no se cortaron datos dentro de este control de calidad, pero posteriormente, en el mapeo, se hace una revisión a profundidad de las características de estas secuencias.

Mapeo

RNA-Star es un alineador universal ultrarrápido de RNA-Seq, y nos permite observar el alineamiento de las secuencias. Siguiendo por la línea de los Scores alineados (Figura 3), se puede observar que las secuencias que se encuentran únicamente mapeadas (azul), son las que nos brindan el porcentaje mayor a 89.6%.

Figura 3

STAR: Scores alineados

Nota: a) primera repetición, b) segunda repetición y, c) tercera repetición.

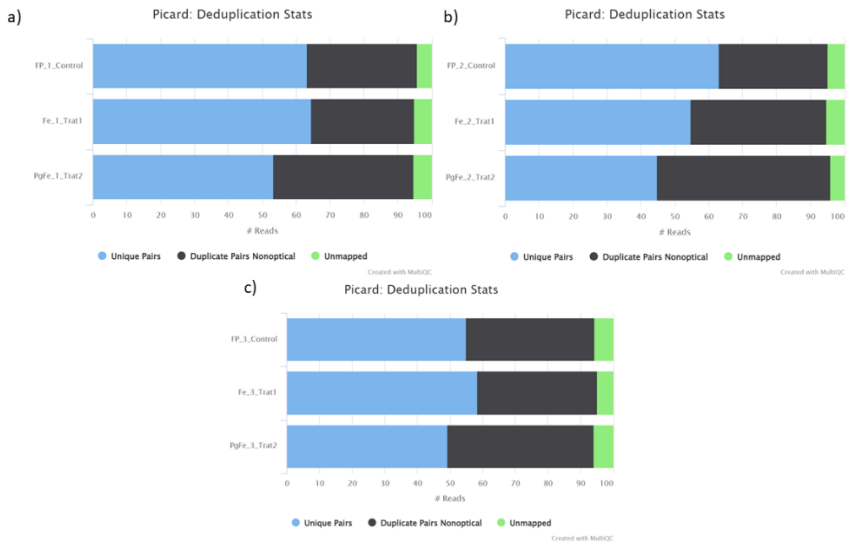
También se pueden constatar (Figura 3) las secuencias que han sido mapeados con loci múltiple (celeste) van del 3 al 4.5%, mientras que las que han mapeado muchos loci (amarillo) casi no se pueden diferenciar porque solo son el 0.1% de cada secuencia. En lo que respecta a las secuencias muy cortas como para mapearse, estas representan un intervalo de entre 4 a 5.7% entre todas las secuencias.

Inspección de resultados de mapeo

Duplicados. Para un chequeo más a profundidad de la calidad de los datos, se utiliza MarkDuplicates para poder obtener la cantidad de lecturas duplicadas, ya que examina registros alineados en el conjunto de datos SAM o BAM suministrado para localizar moléculas duplicadas. Luego, todos los registros se escriben en el archivo de salida con los registros duplicados marcados.

Figura 4

Estadísticas de duplicación



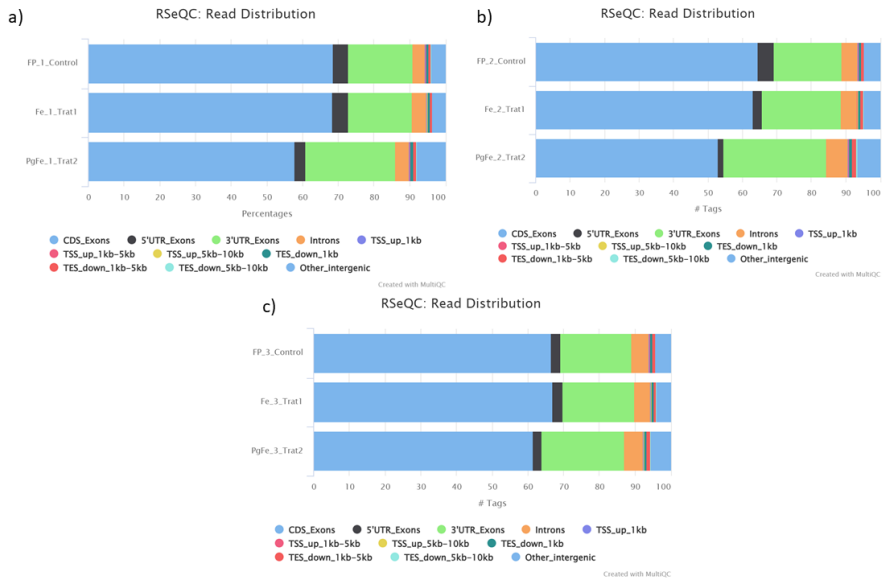
Nota: a) primera repetición, b) segunda repetición y, c) tercera repetición.

Todos estos resultados se pueden constatar con los gráficos obtenidos del mismo programa, en donde tenemos las estadísticas de duplicación (Figura 4), donde se resume que: las secuencias pareadas únicas (azul) van en un intervalo de entre 45 al 64% los duplicados (negro) van de 32.1 al 51% alrededor de las secuencias; mientras que las secuencias sin mapear (verde) van desde el 2 al 5%.

Distribución de lecturas entre funciones. Al realizar esto, se espera que el resultado sea que la mayoría de las lecturas se asignen a exones en lugar de intrones o regiones intergénicas. Por lo cual como se puede observar en la Figura 5, se observa que más del 85% en todas las secuencias son exones, mientras que el resto se dividen entre intrones (naranja), y regiones intergénicas.

Figura 5

Distribución de lecturas con RSeQC



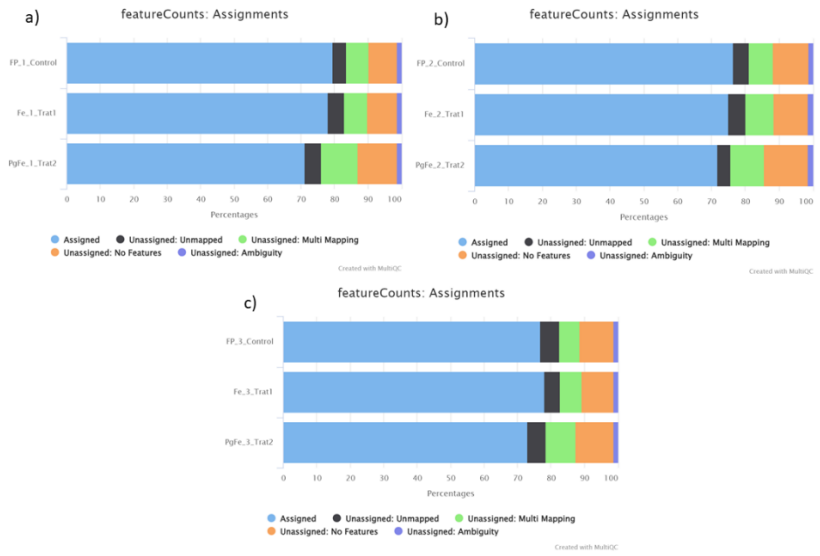
Nota: a) primera repetición, b) segunda repetición y, c) tercera repetición.

Conteo de número de lecturas por gen anotado

Conteo de lecturas por gen. Una vez utilizando el feature counts para el conteo de lecturas por gen, se obtienen las asignaturas de los genes, en donde podemos observar que sobre el 70% de los genes se encuentran asignados (azul), alrededor de un 5% se encuentra sin asignar y sin mapear (negro), alrededor de un 8% se encuentra sin asignar, pero con multimapeo (verde), y alrededor de 10% se encuentra sin asignar y sin características (naranja) y el resto sin asignar, porque tienen ambigüedad. (celeste)

Figura 6

Porcentaje de genes asignados con Feature Counts



Nota: a) primera repetición, b) segunda repetición y, c) tercera repetición.

Análisis de expresión diferencial

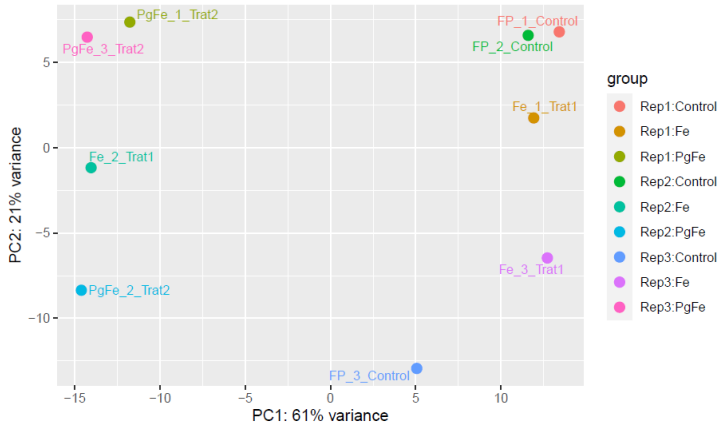
Identificación de las características diferencialmente expresadas

Una vez obtenidas las tablas de feature counts: counts, con esta información se procede a utilizar el DESeq2.

Dentro de los resultados, se obtiene un gráfico PCA (Figura 7), en donde se puede observar la distancia entre un plano y el otro, en donde el control y el tratamiento 1 con Fe-hMSC se encuentran dentro del mismo; mientras que todas las muestras del tratamiento 2, y, solo la réplica 2 del tratamiento 1 se encuentran dentro del otro plano.

Figura 7

Gráfico PCA de las dos primeras dimensiones de un análisis de componentes principales, realizado con los recuentos normalizados de las muestras.



Primero se agrupan por tratamiento (el primer factor) y en segundo lugar por número de repetición (el segundo factor), como en el gráfico PCA (Figura 7). Nuevamente se observa como el grupo control con el grupo de tratamiento 1 se encuentran más cercanos entre ellos que con el grupo de tratamiento 2.

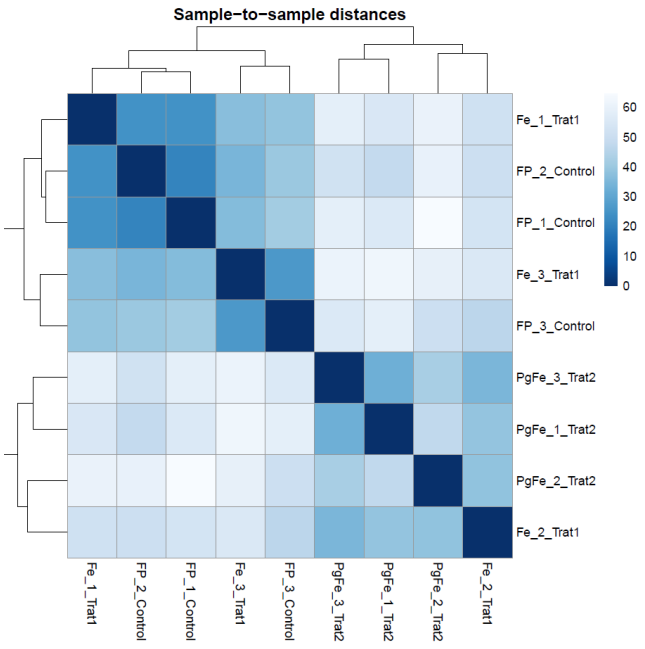
En este caso específico, se menciona que el grupo control y el grupo de tratamiento 1 están más cercanos entre sí que con el grupo de tratamiento 2 en el gráfico PCA. Esto sugiere que hay una mayor similitud en la expresión génica entre el grupo control y el grupo de tratamiento 1 en comparación con el grupo de tratamiento 2. Es importante destacar que esta

interpretación se basa en la variación observada en los datos y puede indicar posibles diferencias en la respuesta al tratamiento entre los grupos.

DESeq2 también nos brinda como uno de sus resultados un mapa de calor (Figura 8), el cual ofrece una descripción general de las similitudes y diferencias entre muestras: el color representa la distancia entre las muestras. Azul oscuro significa distancia más corta, es decir, muestras más cercanas dados los recuentos normalizados.

Figura 8

Mapa de calor de la matriz de distancias de muestra a muestra (con agrupamiento) basado en los recuentos normalizados.



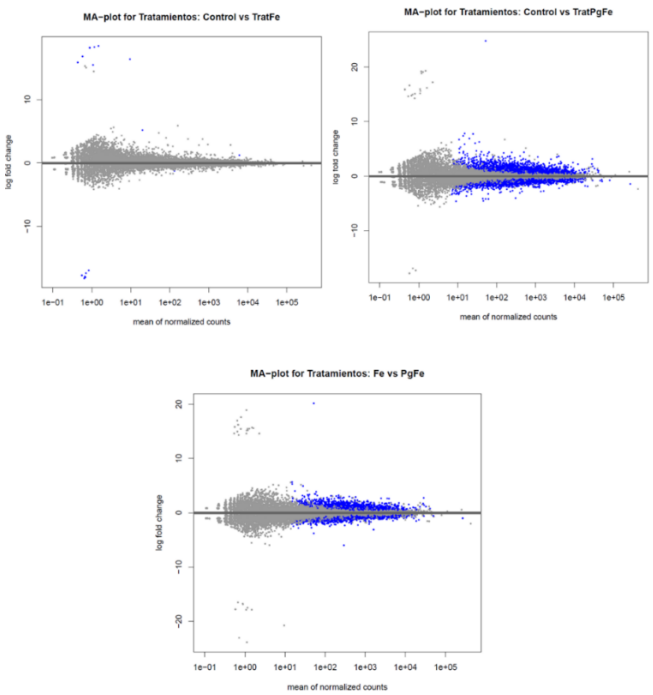
Como el último de los gráficos obtenidos de la herramienta DESeq2 se obtuvo un gráfico MA de cada una de las muestras comparadas entre sí (Figura 9), el cual compara cada una de las muestras entre sí, dando una vista global de la relación entre el cambio de expresión de condiciones, la fuerza de expresión promedio de los genes y la capacidad del algoritmo para detectar la expresión genética diferencial.

Un gráfico MA, muestra la representación visual de los datos genómicos, aplicando una gráfico Bland-Altman, demuestra la diferencia entre medidas de varias muestras, transformando los datos a escalas M (porción logarítmica) y A (promedio de la media). Los genes que superaron el umbral de significancia (valor p ajustado <0,1) están coloreados en azul.

En este gráfico (Figura 9) nuevamente notamos la diferencia entre el comportamiento del tratamiento 1 (TratFe en la imagen), con el tratamiento 2 (TratPgFe en la imagen) cuando se comparan con el control y entre ellos. Existe un mayor número de genes que superaron el umbral de significancia cuando se compara el tratamiento 2, con el control como con el tratamiento 1; mientras que al comparar el grupo de control con el tratamiento 1, no hay diferencia.

Figura 9

Gráfico MA de cada una de las muestras comparadas entre sí.



Extracción y anotación de los genes diferencialmente expresados

DESeq2 también deja un archivo tabular normalizado en donde se puede observar los genes que se han expresado diferencialmente, en este caso, existen más de 30000 genes que componen a *Mus musculus,* prácticamente igual que en los humanos y de estos aproximadamente el 40% se expresan de manera diferencial, lo cual se observa claramente en el gráfico MA, observando la distribución.

DESeq2 es una herramienta bioinformática utilizada para el análisis de datos de expresión génica a partir de secuenciación de ARN. Esta herramienta permite identificar genes que se expresan de manera diferencial entre diferentes condiciones experimentales.

Además de realizar el análisis estadístico para identificar los genes diferencialmente expresados, DESeq2 también genera un archivo tabular normalizado. Este archivo contiene información sobre la expresión de cada gen en las diferentes muestras, después de aplicar correcciones para tener en cuenta la composición de las bibliotecas y otras variaciones experimentales.

En el caso mencionado, se indica que de los aproximadamente 30000 genes que posee el *Mus musculus*, cerca de 15000 se expresan de manera diferencial. Esto significa que, en el experimento analizado, se han identificado alrededor 15000 genes cuya expresión varía entre las condiciones experimentales estudiadas.

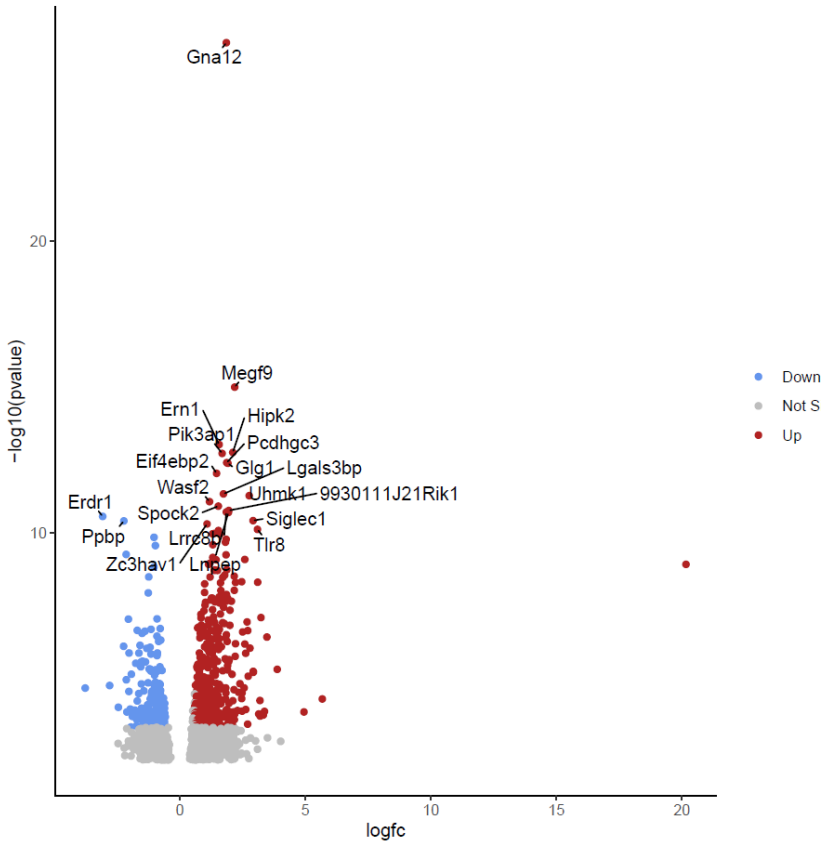
Visualización de la expresión de los genes diferencialmente expresados

De la misma manera y aprovechando ya los datos normalizados, se prosiguió a obtener un gráfico de volcán (Figura 10), en donde se puedan observar los genes que se encuentran diferencialmente expresados en una mayor cantidad. Se pueden observar los 20 genes con mayor expresión diferencial entre la lista de más de 15000 genes, los de color rojo con una media positiva y los de color celeste, una negativa.

En este gráfico, se interpreta en el eje “Y” lo significativas, a nivel estadístico, que son las diferencias en la expresión génica, ya que lo genes que sean más estadísticamente significativos estarán hacia la parte superior, ya que poseen valores p más bajos. Mientras que en el eje “X” se muestra lo grande de la diferencia de expresión génica, que quiere decir que si se encuentra en valores mayores a 0, el gen se encuentra regulado positivamente (color rojo). Mientras que si se encuentra en valores menores que 0, el gen se encuentra regulado negativamente (color celeste). Mientras que los valores cercanos a 0, significa que no existen grandes diferencias en la expresión de genes.

Figura 10

Gráfico de volcán en donde se observan los genes con mayor expresión diferencial entre las muestras.



Tomando en cuenta las interpretaciones del gráfico de volcán, se puede observar que, de los aproximadamente 15000 genes expresados diferencialmente, existen alrededor de 50 que son más estadísticamente significativos, que los otros; y que existen más genes regulados positivamente que los genes regulados negativamente. Mientras que el otro 50% de los no tenían grandes diferencias de expresión genética.

Análisis de enriquecimiento funcional

Dentro del análisis de enriquecimiento funcional realizado utilizando la plataforma online de ShinyGO, en donde se debe rescatar que los análisis que se realizan de manera estadística en esta plataforma virtual tienen como base el método Chi-cuadrado para sacar sus distintos valores de p, teniendo en cuenta las características de los genes. Para este análisis solo se tomaron en cuenta los primeros 5000 genes expresados diferencialmente.

Dentro de esta plataforma de enriquecimiento funcional, los resultados más relevantes son los FDRs. El FDR, o Tasa de Descubrimiento Falso, representa una corrección aplicada a los p-valores derivados del análisis de enriquecimiento funcional. Su función principal es controlar la proporción de descubrimientos falsos, ofreciendo así una evaluación más precisa de la relevancia biológica de los resultados.

GO Procesos Biológicos

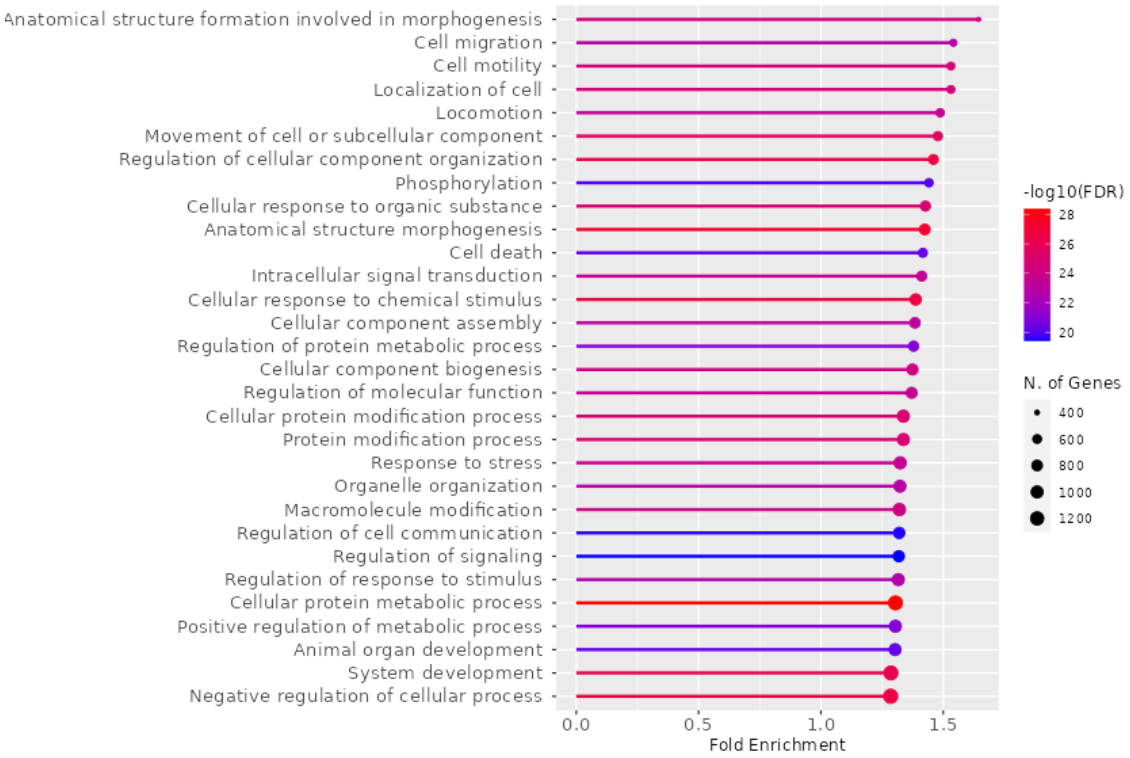
Al obtener los resultados del enriquecimiento funcional en base a los procesos biológicos de los distintos genes que han sido diferencialmente expresados (Figura 11), se puede observar que los genes que se obtuvo una mayor significancia (color azul) en procesos biológicos como lo son la fosforilación, la muerte celular, la regulación de señales y de comunicación celular, y también de desarrollo de órganos.

Al lado izquierdo tenemos el listado de los distintos procesos biológicos dentro de los cuales estos genes forman parte, mientras del lado derecho, podemos observar las

observaciones, como el tamaño del punto se correlaciona al número de genes, y el color de la línea es en relación al FDR.

Figura 11

Distribución de enriquecimiento funcional basado en procesos biológicos de los diferentes genes.



Si se observa el FDR, cuando el color es azul, quiere decir que el FDR es bajo o significativo, lo cual indica que la asociación que existe entre los genes y la vía, en este caso procesos biológicos, es altamente significativa. El morado indicaría asociación media y el rojo una asociación que no es estadísticamente significativa con la vía.

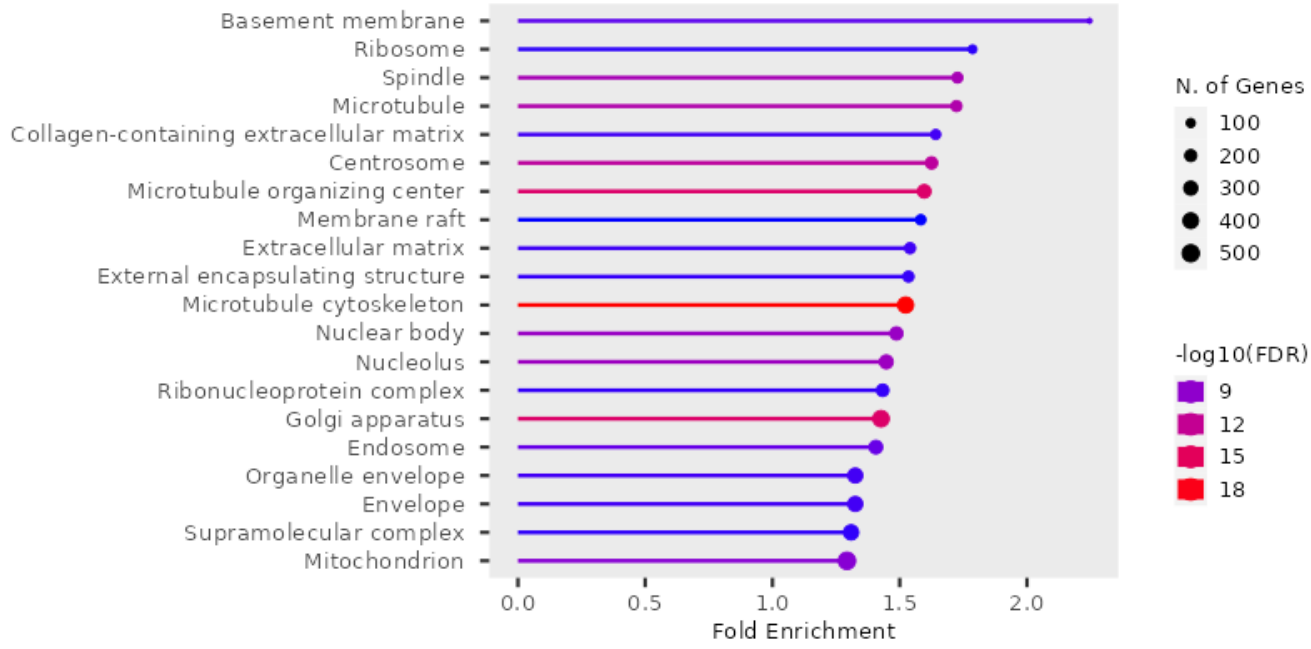
GO Componente Celular

Los resultados del enriquecimiento funcional en base al componente celular en los genes que obtuvimos del análisis de expresión diferencial (Figura 12), podemos observar que

existen muchos más componentes celulares que están estrechamente relacionados a los genes procesados. Son 9 componentes celulares dentro de los que se encuentran: Robosoma, matriz extracelular que contiene colágeno, balsa de membrana, matriz extracelular, estructura encapsuladora externa, complejo ribonucleo-proteína, envoltura de organelo, envoltura y complejo supramolecular.

Figura 12

Distribución de enriquecimiento funcional basado en los componentes celulares de los diferentes genes.

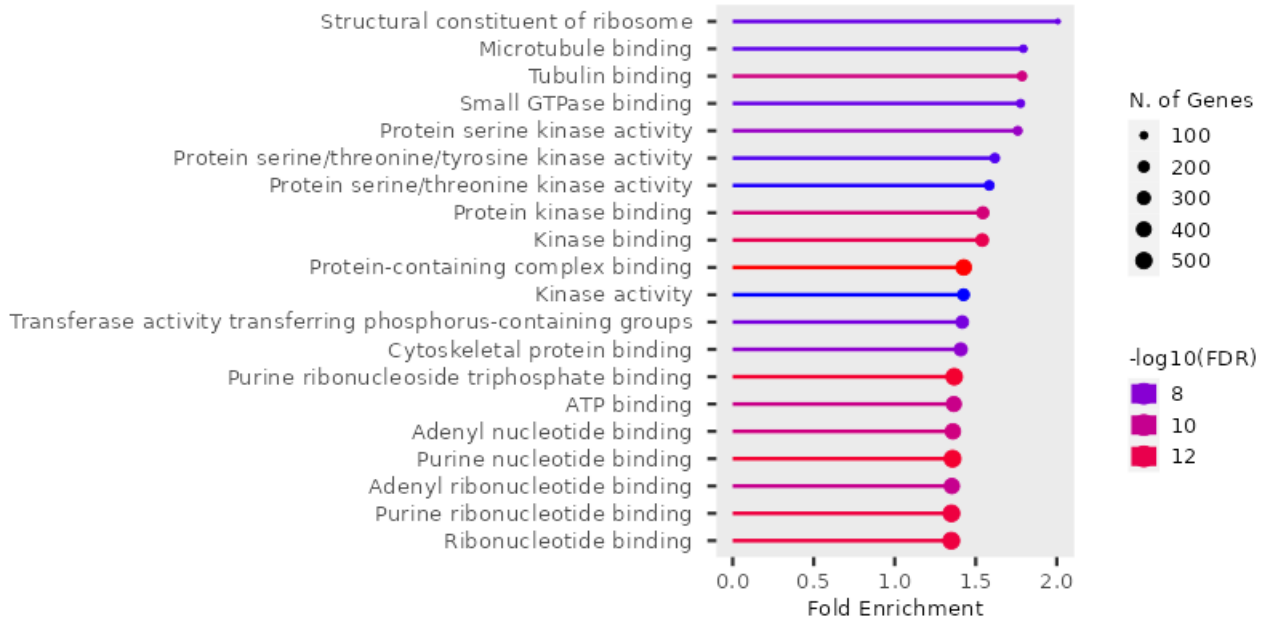


GO Función Molecular

En los que respecta al enriquecimiento funcional en base a la función molecular (Figura 13), se pude observar que, destacan las funciones moleculares relacionadas con la proteína quinasa: actividad de quinasa, actividad quinasa de las proteínas serina-treonina y actividad quinasa de las proteínas serina-treonin-tirosina. Estas siendo altamente significativas y relacionadas a los genes.

Figura 13

Distribución de enriquecimiento funcional basado en la función molecular de los diferentes genes.

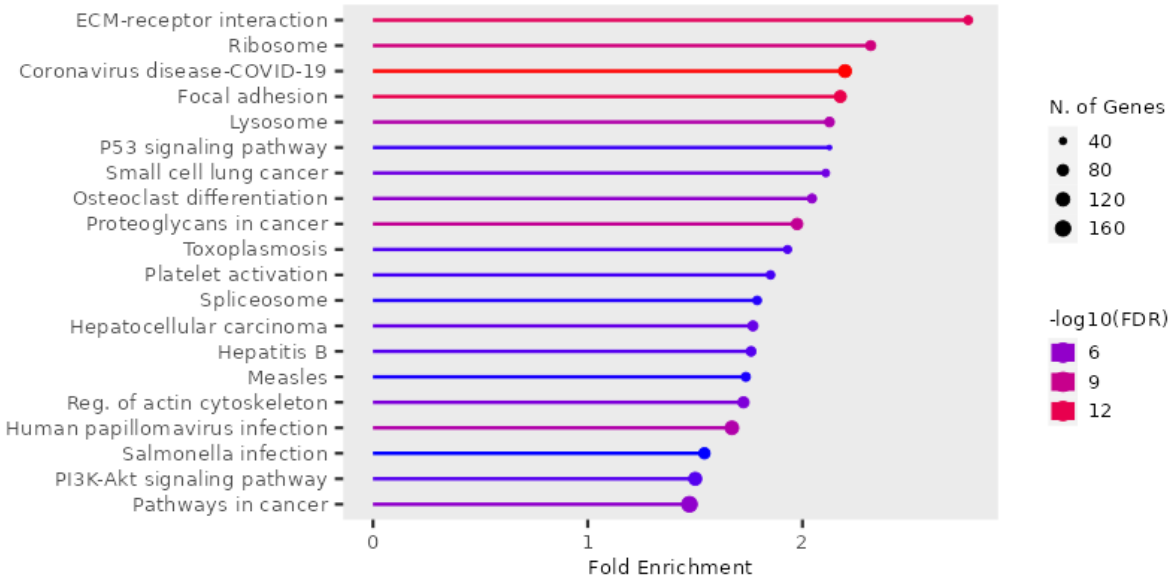


Enciclopedia de genes y genomas de Kioto (KEGG)

Tomando en cuenta los resultados de enriquecimiento funcional basado en la enciclopedia de genes (Figura 14), se pude observar que dentro de las vías donde los genes se encuentran estrechamente relacionados se tienen principalmente a: Vía de señalización de P53, toxoplasmosis, activación de plaquetas, espliceosoma, carcinoma hepatocelular, hepatitis B, sarampión, infección de salmonela, y, vías de señalización de PI3K-Akt.

Figura 14

Distribución de enriquecimiento funcional basado en KEGG

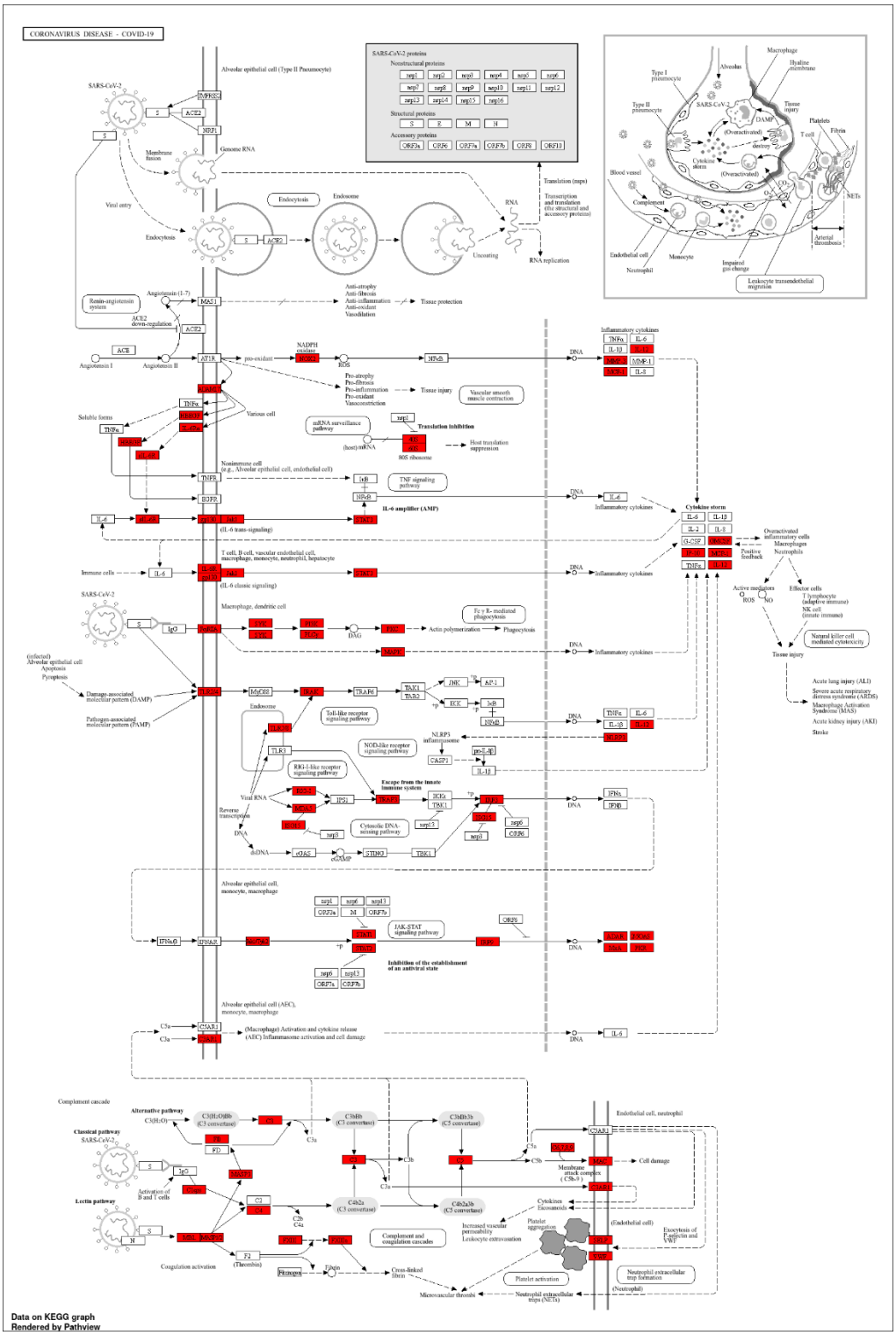


La evaluación de enriquecimiento funcional a través de KEGG involucra la identificación de las vías biológicas y procesos celulares en los cuales los genes con expresión diferencial se encuentran sobre o subexpresados de manera significativa.

Dentro de los resultados de ShinyGO, en la sección de KEGG, se presentan diversas rutas metabólicas que indican las enfermedades en las que se observaron implicaciones. Dentro de las rutas que destacan como estrechamente relacionadas a los genes no se encuentra ninguna que afecte los pulmones o sea parte de procesos de los pulmones. Sin embargo, destacamos que existe la relación de los genes con dos vías de señalización de cáncer como lo son la P53 y la PI3k/Akt. Dentro de resultados de falsos positivos tenemos genes que se encuentran estrechamente relacionados a la enfermedad de COVID-19 (Figura 15), en donde los marcados con rojo son los genes relacionados con dicha enfermedad.

Figura 15

Ruta metabólica de la enfermedad del COVID-19



Capítulo IV: Discusión

Cuando hablamos del control de calidad, luego de la revisión de cada una de las secuencias, el mapeo permite que se cerciore que, los datos han sido tratados con cuidado y no existe ninguna clase de contaminación. A pesar de que cuando se realizó el corte de las secuencias menores a 20 bp, no hubo secuencias de menores a lo mismo, cuando se realizaba el mapeo, existieron varios puntos que, gracias a bibliografía, constatar que las muestras se encontraban de alta calidad.

Luego de utilizar STAR, el cual es un alineador universal, al obtener los resultados de los Scores alineados (Figura 3), todas las repeticiones tienen más de 90% de mapeo único, y se sostiene que las muestras no tienen contaminación debido a que tiene más de un 70% de mapeo único que es lo mínimo aceptable dentro de la herramienta STAR (Dobin et al., 2013).

Esto permitió avanzar con el control de duplicados (Figura 4), donde se observa que todas las secuencias tienen duplicados que representan menos del 50%, excepto una muestra del tratamiento 2, réplica 2, cuyos duplicados alcanzan el 51%. Es importante destacar que una cobertura de lectura extremadamente alta para transcripciones particularmente expresadas en datos de RNA-seq puede resultar en niveles de duplicación de lectura del 70% o más (Koch et al., 2018). Sin embargo, dado que los duplicados detectados se encuentran dentro de un rango adecuado, se considera apropiado continuar con el mapeo de las secuencias

Considerando otro punto de control de calidad de las secuencias, se analizó la distribución de las lecturas con RSeQC (Figura 5), y se observó que más del 85% de las lecturas abarcaban exones en cada muestra. En un contexto normal de secuencias alineadas, se espera que más del 80% de las lecturas correspondan a exones, alrededor del 2% a intrones y aproximadamente un 5% a regiones intergénicas (Wang et al., 2012). Estos resultados indican que la alineación y el mapeo han sido exitosos.

Como última prueba de mapeo y control de calidad, se realizó el conteo de asignaciones con Feature Counts (Figura 6). En este análisis, se observó que más del 70% de las lecturas de genes han sido asignados correctamente. Es importante destacar que, si este porcentaje hubiera sido inferior al 50%, se habría requerido investigar si las lecturas fueron mapeadas adecuadamente y verificar las anotaciones correspondientes a la versión del genoma de referencia (Liao et al., 2014).

DESeq2 se utiliza para estimar la dependencia de la varianza-media en los datos de recuento de ensayos de secuenciación de alto rendimiento y probar la expresión diferencial basándose en un modelo que utiliza la distribución binomial negativa (Love et al., 2014). En cuanto al análisis de componentes principales (PCA, por sus siglas en inglés) (Figura 7), este procedimiento estadístico permite resumir la información contenida en grandes tablas de datos en un conjunto más pequeño de "índices resumidos" que son más fáciles de visualizar y analizar (Camargo, 2022.). Se observa que, entre los dos tratamientos, el que más se asemeja al control es el tratamiento 1, mientras que el tratamiento 2 se separa en el plano. Solo la réplica 2 del tratamiento 1 con Fe-hMSC se asemeja al tratamiento 2 con PgFe-hMSC. Esto podría indicar que el tratamiento 2 con PgFe-hMSC podría ser más efectivo debido a la dispersión de los tratamientos con el control dentro de los planos, en donde las muestras del tratamiento 1 se observa más relacionado con el control, mientras que las muestras con el tratamiento dos se alejan de ellos. Estas relaciones se reflejan en el mapa de calor (Figura 8), donde el control se relaciona estrechamente con el tratamiento uno, mientras que las tres réplicas del tratamiento 2 se agrupan en el otro lado del gráfico.

En el gráfico de media móvil (MA) (Figura 9), una cantidad considerable de puntos de datos que superan un límite en el eje y sugiere una mayor presencia de genes bajo regulación positiva. Por otro lado, una mayor proporción de puntos por debajo de -1 indicaría una intensa regulación negativa en los genes (Monier et al., 2018). Al comparar los tres gráficos de la

Figura 9, se observa que tanto el control como el tratamiento 1 con Fe-hMSC son muy similares entre sí, lo que no muestra significancia (puntos azules) al compararlos entre ellos. Sin embargo, al comparar tanto las muestras del grupo control como las muestras del tratamiento 1 con Fe-hMSC con el tratamiento 2 con PgFe-hMSC, se pueden observar más genes expresados diferencialmente con regulación positiva, lo que sugiere una mayor diferencia en la expresión génica entre estos tratamientos.

Esto previamente mencionado, se puede constatar en la investigación de Huang et al., 2023; donde utilizaron los mismos datos y llegaron a la conclusión que el tratamiento 2 con Pg-Fe-hMSC logró mejores resultados tanto para la capacidad de transferencia mitocondrial altamente eficiente y sostenida, lo cual permitió los mejores resultados terapéuticos, en comparación al tratamiento 1 con Fe-hMSC.

Posteriormente, tomando en cuenta el gráfico de volcán que nos permite observar claramente la regulación tanto positiva, como negativa, de los genes y su significancia, se puede observar una clara distinción entre los genes que tienen una mayor significancia entre todos, que son alrededor de 50, y una mayor cantidad de genes con regulación positiva, de los 15000 que se pudo observar previamente en el gráfico MA que se expresaron diferencialmente.

Al comparar los distintos enfoques del enriquecimiento funcional, es importante considerar la relación de los genes en cada uno de los procesos biológicos, componentes celulares y funciones moleculares. En los gráficos destaca el FDR, el cual ajusta los p-valores para minimizar los falsos positivos, mejorando la confiabilidad de los hallazgos en análisis de enriquecimiento funcional. Dentro de los resultados se puede observar que el enriquecimiento funcional basado en la componente celular es el que presenta más genes estrechamente relacionados en comparación con todas las relaciones de ontología de genes. Por lo tanto, se presume que la mayoría de los genes en esta categoría forman parte de los componentes celulares (Ge et al., 2020).

Al analizar las vías KEGG a las que podrían pertenecer la mayoría de los genes previamente diferencialmente expresados, se observó que las vías donde los genes se encuentran estrechamente relacionados se tienen principalmente a: Vía de señalización de P53, toxoplasmosis, activación de plaquetas, espliceosoma, carcinoma hepatocelular, hepatitis B, sarampión, infección de salmonela, y, vías de señalización de PI3K-Akt.

Dentro de estas vías se puede destacar las vías de señalización P53 y PI3k/Akt, debido a que son vías de señalización de cáncer, también tenemos el espliceosoma que genera RNAm maduro, las cuales se encuentran estrechamente relacionados a procesos cancerígenos, lo cual sugiere que una fibrosis pulmonar a largo plazo puede transformarse en cáncer. Las vías de infección de salmonela y toxoplasmosis son vías donde la infección se adquiere por comida, lo cual no representan vías de interés para este tema. Mientras que la hepatitis B y el carcinoma hepatocelular son dos enfermedades que afectan el hígado, lo cual puede sugerir que la fibrosis pulmonar, o en su defecto, la medicación con la que se está tratando a la larga puede afectar el hígado. Observando más a fondeo la vía del sarampión, el único síntoma considerable y distintivos de las otras infecciones mencionadas sería el posible caso de pneumonía, ya que afectaría al pulmón.

Sin embargo, y con muchos genes que estarían relacionándose estrechamente a esta ruta, tenemos un resultado con FDR que dice ser falso positivo, pero enfocado en genes que pertenecen a la enfermedad de COVID-19, la cual afecta las vías respiratorias, por lo cual se puede investigar la relación de los genes de esta enfermedad con los de la fibrosis pulmonar en el futuro. (Persuy et al., 2015).

Capítulo V: Conclusiones

- Se obtuvieron tres réplicas de cada muestra de: control, tratamiento 1 con Fe-hMSC y tratamiento 2 con PgFe-hMSC, de bases de datos públicas, las cuales se utilizaron para realizar un análisis de expresión diferencial de genes y se obtuvieron alrededor de 15000 genes diferencialmente expresados.

- El mapeo permitió constatar que el procesamiento de las lecturas obtenidas de la base de datos generó secuencias de alta calidad y sin signos de contaminación.

- Dentro de los tratamientos, se observa una clara diferencia entre el tratamiento 1 (Fe-hMSC) y el tratamiento 2 (PgFe-hMSC). Mientras que el tratamiento 1 no mostró resultados significativos en comparación con el grupo de control, en el grupo de tratamiento 2 se observó una diferencia en la expresión de genes y a su vez demostró una alta capacidad de transferencia mitocondrial y excelentes resultados terapéuticos entre los dos tratamientos. Lo cual se pude utilizar de punto de partida para más investigaciones involucrando las células madre mesenquimales con nanopartículas de óxido de hierro y pioglitazona como terapia para la fibrosis pulmonar.

- En el análisis del enriquecimiento funcional se puede concluir que la mayoría de los genes se encuentran estrechamente relacionados a los componentes celulares, donde tuvieron más de 9 componentes celulares con mayor significancia. Por el lado de las rutas KEGG, la mayoría de genes forman parte de rutas cancerígenas, sin embargo, tomando en cuanta los probables falsos positivos se encontraron que muchos de los genes pertenecen también a la ruta metabólica de la enfermedad del COVID-19, lo cual como son dos enfermedades que afectan las vías respiratorias, se puede usar de punto de partida para posteriores investigaciones.

Capítulo VI: Recomendaciones

- Al tomar datos de bibliotecas públicas, revisar y tomar en cuenta si vienen en pares o son únicas, identificar con qué nombres se les va a diferenciar a cada una de las muestras para evitar confusiones al momento de procesar los datos, ya que estos deben ser claros para quien maneje las muestras.
- Es necesario hacer varias pruebas que permitan identificar si las muestras son de buena calidad o si se contaminaron, sobre todo si no las estamos tomando nosotros, sino que son tomadas de una base de datos. Es por esto que es indispensable hacer un control de calidad y mapeo.
- Al comparar dos tratamientos para realizar un análisis de expresión diferencial de genes, es importante observar cómo se comporta cada tratamiento en comparación con el control y entre sí. Por lo tanto, es necesario obtener gráficos separados para cada comparación. Aunque DESeq2 proporciona un único gráfico, se puede ejecutar el análisis dos veces más para obtener las comparaciones deseadas. Además, es aconsejable generar un mapa de volcán para visualizar el análisis de expresión diferencial de genes e identificar los genes con mayor grado de expresión.
- Cuando se hace un análisis de enriquecimiento funcional, es necesario observar cada uno de los campos y la clasificación de los genes con sus respectivos valor-p antes de tomar la decisión de cuál va a ser la vía final de los genes. ShinyGO es una herramienta sencilla para poder observar toda la estadística y sus gráficos de manera eficiente y con una interfaz amigable.

Capítulo VII: Bibliografía

AbuHalimeh, A. (2022). Improving Data Quality in Clinical Research Informatics Tools. *Frontiers in Big Data*, *5*, 871897. https://doi.org/10.3389/FDATA.2022.871897

Adams, J. I., Ferebee, T., Minto, M., Pennerman, K. K., & Chambwe, N. (2022). *Ten simple rules for creating a global network in computational biology*. https://doi.org/10.1371/journal.pcbi.1010528

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(Web Server issue), W537. https://doi.org/10.1093/NAR/GKY379

Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1). https://doi.org/10.1093/GENETICS/IYAD031

Alharbi, F., & Vakanski, A. (2023). Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. *Bioengineering*, *10*(2). https://doi.org/10.3390/BIOENGINEERING10020173

Andugulapati, S. B., Gourishetti, K., Tirunavalli, S. K., Shaikh, T. B., & Sistla, R. (2020). Biochanin-A ameliorates pulmonary fibrosis by suppressing the TGF-β mediated EMT, myofibroblasts differentiation and collagen deposition in in vitro and in vivo systems. *Phytomedicine*, *78*, 153298. https://doi.org/10.1016/J.PHYMED.2020.153298

Ayoob Id, J. C., & Kangas, J. D. (2020). *10 simple rules for teaching wet-lab experimentation to computational biology students, i.e., turning computer mice into lab rats*. https://doi.org/10.1371/journal.pcbi.1007911

Begik, O., Mattick, J. S., & Novoa, E. M. (2022). Exploring the epitranscriptome by native RNA sequencing. *RNA*, *28*(11), 1430. https://doi.org/10.1261/RNA.079404.122

Bhatt, A. (2023). Data quality – The foundation of real-world studies. *Perspectives in Clinical Research*, *14*(2), 92. https://doi.org/10.4103/PICR.PICR_12_23

Blasco, A. I., Endres, M. G., Sergeev, R. A., Jonchhe, A., Maximilian Macaluso, N. J., Narayan, R., NatoliID, T., Paik, J. H., Briney, B., Wu, C., Su, A. I., Subramanian, A., & Lakhani, K. R. (2019a). *Advancing computational biology and bioinformatics research through open innovation competitions*. https://doi.org/10.1371/journal.pone.0222165

Blasco, A. I., Endres, M. G., Sergeev, R. A., Jonchhe, A., Maximilian Macaluso, N. J., Narayan, R., NatoliID, T., Paik, J. H., Briney, B., Wu, C., Su, A. I., Subramanian, A., & Lakhani, K. R. (2019b). *Advancing computational biology and bioinformatics research through open innovation competitions*. https://doi.org/10.1371/journal.pone.0222165

Camargo, A. (s/f). *PCAtest: testing the statistical significance of Principal Component Analysis in R*. https://doi.org/10.7717/peerj.12967

Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., … Elser, J. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325. https://doi.org/10.1093/NAR/GKAA1113

Chen, C., Wang, Y. yan, Wang, Y. xia, Cheng, M. qun, Yin, J. bing, Zhang, X., & Hong, Z. peng. (2018a). Gentiopicroside ameliorates bleomycin-induced pulmonary fibrosis in mice via inhibiting inflammatory and fibrotic process. *Biochemical and Biophysical Research Communications*, *495*(4), 2396–2403. https://doi.org/10.1016/J.BBRC.2017.12.112

Chen, C., Wang, Y. yan, Wang, Y. xia, Cheng, M. qun, Yin, J. bing, Zhang, X., & Hong, Z. peng. (2018b). Gentiopicroside ameliorates bleomycin-induced pulmonary fibrosis in mice via inhibiting inflammatory and fibrotic process. *Biochemical and Biophysical Research Communications*, *495*(4), 2396–2403. https://doi.org/10.1016/J.BBRC.2017.12.112

Chen, L., & Wong, G. (2019). Transcriptome Informatics. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*, 324–340. https://doi.org/10.1016/B978-0-12-809633-8.20204-5

Chen, Y., Verbeek, F. J., & Wolstencroft, K. (2021). Establishing a consensus for the hallmarks of cancer based on gene ontology and pathway annotations. *BMC Bioinformatics*, *22*(1). https://doi.org/10.1186/S12859-021-04105-8

Chloe Li, K. Y., Cook, A. C., & Lovering, R. C. (2022). GOing Forward With the Cardiac Conduction System Using Gene Ontology. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/FGENE.2022.802393/FULL

Choi, S. M., Mo, Y., Bang, J. Y., Ko, Y. G., Ahn, Y. H., Kim, H. Y., Koh, J., Yim, J. J., & Kang, H. R. (2023). Classical monocyte-derived macrophages as therapeutic targets of umbilical cord mesenchymal stem cells: comparison of intratracheal and intravenous administration in a mouse model of pulmonary fibrosis. *Respiratory Research*, *24*(1), 68. https://doi.org/10.1186/S12931-023-02357-X

Christov, Z., Karabencheva-Christova, T., Zhang, Y., Luo, M., Wu, P., Wu, S., Lee, T.-Y., & Bai, C. (2022). International Journal of Molecular Sciences Application of Computational

Biology and Artificial Intelligence in Drug Design. *Int. J. Mol. Sci*, *2022*, 13568. https://doi.org/10.3390/ijms232113568

Clemente-Suárez, V. J., Martín-Rodríguez, A., Yáñez-Sepúlveda, R., & Tornero-Aguilera, J. F. (2023). Mitochondrial Transfer as a Novel Therapeutic Approach in Disease Diagnosis and Treatment. *International Journal of Molecular Sciences*, *24*(10). https://doi.org/10.3390/IJMS24108848

Coker, H., Wei, G., & Brockdorff, N. (2019). m6A modification of non-coding RNA and the control of mammalian gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1862*(3), 310–318. https://doi.org/10.1016/J.BBAGRM.2018.12.002

Damiani, C., Rovida, L., Maspero, D., Sala, I., Rosato, L., Di Filippo, M., Pescini, D., Graudenzi, A., Antoniotti, M., & Mauri, G. (2020). MaREA4Galaxy: Metabolic reaction enrichment analysis and visualization of RNA-seq data within Galaxy. *Computational and Structural Biotechnology Journal*, *18*, 993–999. https://doi.org/10.1016/J.CSBJ.2020.04.008

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Farhang Ghahremani, M., Kai, K., Seto, Y., Cho, W., Miller, M. C., Smith, P., & Englert, D. F. (2023). RESEARCH Open Access Novel method for highly multiplexed gene expression profiling of circulating tumor cells (CTCs) captured from the blood of women with metastatic breast cancer. *Journal of Translational Medicine*, *21*, 414. https://doi.org/10.1186/s12967-023-04242-z

Gajjala, P. R., Kasam, R. K., Soundararajan, D., Sinner, D., Huang, S. K., Jegga, A. G., & Madala, S. K. (2021). Dysregulated overexpression of Sox9 induces fibroblast activation in pulmonary fibrosis. *JCI Insight*, *6*(20). https://doi.org/10.1172/JCI.INSIGHT.152503

Ge, S. X., Jung, D., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628. https://doi.org/10.1093/BIOINFORMATICS/BTZ931

Gliklich, R. E., Dreyer, N. A., & Leavy, M. B. (2014). *Data Collection and Quality Assurance*. https://www.ncbi.nlm.nih.gov/books/NBK208601/

Hiltemann, S., Rasche, H., Gladman, S., Hotz, H. R., Larivière, D., Blankenberg, D., Jagtap, P. D., Wollmann, T., Bretaudeau, A., Goué, N., Griffin, T. J., Royaux, C., Bras, Y. Le, Mehta, S., Syme, A., Coppens, F., Droesbeke, B., Soranzo, N., Bacon, W., … Batut, B. (2023). Galaxy Training: A powerful framework for teaching! *PLoS Computational Biology*, *19*(1). https://doi.org/10.1371/journal.pcbi.1010752

Hong, M., Tao, S., Zhang, L., Diao, L. T., Huang, X., Huang, S., Xie, S. J., Xiao, Z. D., & Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology & Oncology*, *13*(1), 166. https://doi.org/10.1186/S13045-020-01005-X

Hong, Y., Lin, Z., Yang, Y., Jiang, T., Shang, J., & Luo, Z. (2022). Biocompatible Conductive Hydrogels: Applications in the Field of Biomedicine. *International journal of molecular sciences*, *23*(9). https://doi.org/10.3390/IJMS23094578

Huang, T., Lin, R., Su, Y., Sun, H., Zheng, X., Zhang, J., Lu, X., Zhao, B., Jiang, X., Huang, L., Li, N., Shi, J., Fan, X., Xu, D., Zhang, T., & Gao, J. (2023). Efficient intervention for pulmonary fibrosis via mitochondrial transfer promoted by mitochondrial biogenesis. *Nature Communications*, *14*(1). https://doi.org/10.1038/S41467-023-41529-7

Jezernik, G., Gorenjak, M., & Potočnik, U. (2022). Gene Ontology Analysis Highlights Biological Processes Influencing Non-Response to Anti-TNF Therapy in Rheumatoid Arthritis. *Biomedicines*, *10*(8). https://doi.org/10.3390/BIOMEDICINES10081808/S1

Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, *12*(3). https://doi.org/10.1002/CTM2.694

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, M. S., Kim, S. H., Jeon, D., Kim, H. Y., & Lee, K. (2018). Changes in expression of cytokines in polyhexamethylene guanidine-induced lung fibrosis in mice: Comparison of bleomycin-induced lung fibrosis. *Toxicology*, *393*, 185–192. https://doi.org/10.1016/J.TOX.2017.11.017

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A beginner's guide to analysis of RNA sequencing data. *American Journal of Respiratory Cell and Molecular Biology*, *59*(2), 145–157. https://doi.org/10.1165/RCMB.2017-0430TR/SUPPL_FILE/DISCLOSURES.PDF

Kramarz, B., Huntley, R. P., Rodríguez-López, M., Roncaglia, P., Saverimuttu, S. C. C., Parkinson, H., Bandopadhyay, R., Martin, M. J., Orchard, S., Hooper, N. M., Brough, D., & Lovering, R. C. (2020). Gene Ontology Curation of Neuroinflammation Biology Improves the Interpretation of Alzheimer's Disease Gene Expression Data. *Journal of Alzheimer's Disease*, *75*(4), 1417. https://doi.org/10.3233/JAD-200207

Krishnankutty, B., Bellary, S., Naveen Kumar, B. R., & Moodahadu, L. S. (2012). Data management in clinical research: An overview. *Indian Journal of Pharmacology*, *44*(2), 168. https://doi.org/10.4103/0253-7613.93842

Kumar, D., & Kirti, P. B. (2023). The genus Arachis: an excellent resource for studies on differential gene expression for stress tolerance. *Frontiers in Plant Science*, *14*, 1275854. https://doi.org/10.3389/FPLS.2023.1275854

Kwak, M., Erdag, G., & Slingluff, C. L. (2020). Gene expression analysis in formalin fixed paraffin embedded melanomas is associated with density of corresponding immune cells in those tissues. *Scientific Reports*, *10*(1). https://doi.org/10.1038/S41598-020-74996-9

Leal, M. P., Brochetti, R. A., Ignácio, A., Câmara, N. O. S., da Palma, R. K., de Oliveira, L. V. F., de Fátima Teixeira da Silva, D., & Lino-dos-Santos-Franco, A. (2018a). Effects of formaldehyde exposure on the development of pulmonary fibrosis induced by bleomycin in mice. *Toxicology Reports*. https://doi.org/10.1016/J.TOXREP.2018.03.016

Leal, M. P., Brochetti, R. A., Ignácio, A., Câmara, N. O. S., da Palma, R. K., de Oliveira, L. V. F., de Fátima Teixeira da Silva, D., & Lino-dos-Santos-Franco, A. (2018b). Effects of formaldehyde exposure on the development of pulmonary fibrosis induced by bleomycin in mice. *Toxicology Reports*, *5*, 512–520. https://doi.org/10.1016/J.TOXREP.2018.03.016

Li, D. Y., Li, R. F., Sun, D. X., Pu, D. D., & Zhang, Y. H. (2021). Mesenchymal stem cell therapy in pulmonary fibrosis: a meta-analysis of preclinical studies. *Stem Cell Research & Therapy*, *12*(1). https://doi.org/10.1186/S13287-021-02496-2

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Liu, C., Ma, Y., Zhao, J., Nussinov, R., Zhang, Y. C., Cheng, F., & Zhang, Z. K. (2020a). Computational network biology: Data, models, and applications. *Physics Reports*, *846*, 1–66. https://doi.org/10.1016/J.PHYSREP.2019.12.004

Liu, C., Ma, Y., Zhao, J., Nussinov, R., Zhang, Y. C., Cheng, F., & Zhang, Z. K. (2020b). Computational network biology: Data, models, and applications. *Physics Reports*, *846*, 1–66. https://doi.org/10.1016/J.PHYSREP.2019.12.004

Liu, Z., Sun, Y., Qi, Z., Cao, L., & Ding, S. (2022). Mitochondrial transfer/transplantation: an emerging therapeutic approach for multiple diseases. *Cell & Bioscience 2022 12:1*, *12*(1), 1–29. https://doi.org/10.1186/S13578-022-00805-7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Matthews, B. J., Melia, T., & Waxman, D. J. (2021). Harnessing natural variation to identify cis regulators of sex-biased gene expression in a multi-strain mouse liver model. *PLoS Genetics*, *17*(11). https://doi.org/10.1371/journal.pgen.1009588

Melia, T., & Waxman, D. J. (2020). Genetic factors contributing to extensive variability of sex-specific hepatic gene expression in Diversity Outbred mice. *PLoS ONE*, *15*(12 December). https://doi.org/10.1371/journal.pone.0242665

Mirela-Bota, P., Aguirre-Plans, J., Meseguer, A., Galletti, C., Segura, J., Planas-Iglesias, J., Garcia-Garcia, J., Guney, E., Oliva, B., & Fernandez-Fuentes, N. (2021). Galaxy InteractoMIX: An Integrated Computational Platform for the Study of Protein–Protein Interaction Data. *Journal of Molecular Biology*, *433*(11), 166656. https://doi.org/10.1016/J.JMB.2020.09.015

Monier, B., Mcdermaid, A., Zhao, J., & Ma, Q. (s/f). *ViDGER Supplementary Material*. Recuperado el 31 de enero de 2024, de https://academic.oup.com/bib/article/20/6/2044/5066173

Persuy, M. A., Sanz, G., Tromelin, A., Thomas-Danguin, T., Gibrat, J. F., & Pajot-Augy, E. (2015). Mammalian olfactory receptors: Molecular mechanisms of odorant detection, 3D-modeling, and structure-activity relationships. *Progress in Molecular Biology and Translational Science*, *130*, 1–36. https://doi.org/10.1016/bs.pmbts.2014.11.001

Pourgholamhossein, F., Rasooli, R., Pournamdari, M., Pourgholi, L., Samareh-Fekri, M., Ghazi-Khansari, M., Iranpour, M., Poursalehi, H. R., Heidari, M. R., & Mandegary, A. (2018). Pirfenidone protects against paraquat-induced lung injury and fibrosis in mice by modulation of inflammation, oxidative stress, and gene expression. *Food and Chemical Toxicology*, *112*, 39–46. https://doi.org/10.1016/J.FCT.2017.12.034

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D., & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*(9), 1–13. https://doi.org/10.1186/GB-2013-14-9-R95/COMMENTS

Remón, L., Uvidia, G., & Castro, O. (2016). *CASO CLÍNICO*.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015a). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/NAR/GKV007

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015b). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/NAR/GKV007

Romero, Y., Balderas-Martínez, Y. I., Vargas-Morales, M. A., Castillejos-López, M., Vázquez-Pérez, J. A., Calyeca, J., Torres-Espíndola, L. M., Patiño, N., Camarena, A., Carlos-Reyes, Á., Flores-Soto, E., León-Reyes, G., Sierra-Vargas, M. P., Herrera, I., Luis-García, E. R., Ruiz, V., Velázquez-Cruz, R., & Aquino-Gálvez, A. (2022). Effect of Hypoxia in the Transcriptomic Profile of Lung Fibroblasts from Idiopathic Pulmonary Fibrosis. *Cells*, *11*(19). https://doi.org/10.3390/cells11193014

Rotwein, P. (2019). Gene Mapping by RNA-sequencing: A Direct Way to Characterize Genes and Gene Expression through Targeted Queries of Large Public Databases. *Bio-protocol*, *9*(1). https://doi.org/10.21769/BIOPROTOC.3129

Saleh, M., Fotook Kiaei, S. Z., & Kavianpour, M. (2022). Application of Wharton jelly-derived mesenchymal stem cells in patients with pulmonary fibrosis. *Stem Cell Research & Therapy*, *13*(1). https://doi.org/10.1186/S13287-022-02746-X

Sgalla, G., Iovene, B., Calvello, M., Ori, M., Varone, F., & Richeldi, L. (2018). Idiopathic pulmonary fibrosis: pathogenesis and management. *Respiratory Research 2018 19:1*, *19*(1), 1–18. https://doi.org/10.1186/S12931-018-0730-2

Simon, K. S., Coelho, L. C., Veloso, P. H. de H., Melo-Silva, C. A., Morais, J. A. V., Longo, J. P. F., Figueiredo, F., Viana, L., Silva Pereira, I., Amado, V. M., Mortari, M. R., & Bocca, A. L. (2023a). Innovative Pre-Clinical Data Using Peptides to Intervene in the Evolution of Pulmonary Fibrosis. *International Journal of Molecular Sciences*, *24*(13). https://doi.org/10.3390/ijms241311049

Simon, K. S., Coelho, L. C., Veloso, P. H. de H., Melo-Silva, C. A., Morais, J. A. V., Longo, J. P. F., Figueiredo, F., Viana, L., Silva Pereira, I., Amado, V. M., Mortari, M. R., & Bocca, A. L. (2023b). Innovative Pre-Clinical Data Using Peptides to Intervene in the Evolution of

Pulmonary Fibrosis. *International Journal of Molecular Sciences*, *24*(13). https://doi.org/10.3390/ijms241311049

Spoor, S., Wytko, C., Soto, B., Chen, M., Almsaeed, A., Condon, B., Herndon, N., Hough, H., Jung, S., Staton, M., Wegrzyn, J., Main, D., Feltus, F. A., & Ficklin, S. P. (2020). Tripal and Galaxy: Supporting reproducible scientific workflows for community biological databases. *Database*, *2020*. https://doi.org/10.1093/database/baaa032

Tam, B. Y., Chiu, K., Chung, H., Bossard, C., Nguyen, J. D., Creger, E., Eastman, B. W., Mak, C. C., Ibanez, M., Ghias, A., Cahiwat, J., Do, L., Cho, S., Nguyen, J., Deshmukh, V., Stewart, J., Chen, C. W., Barroga, C., Dellamary, L., … Yazici, Y. (2020). The CLK inhibitor SM08502 induces anti-tumor activity and reduces Wnt pathway gene expression in gastrointestinal cancer models. *Cancer Letters*, *473*, 186–197. https://doi.org/10.1016/J.CANLET.2019.09.009

Tan, K. T., Ding, L. W., Wu, C. S., Tenen, D. G., & Yang, H. (2021). Repurposing RNA sequencing for discovery of RNA modifications in clinical cohorts. *Science Advances*, *7*(32). https://doi.org/10.1126/SCIADV.ABD2605

Tasker, J. G., Voisin, D. L., & Armstrong, W. E. (2017). The Cell Biology of Oxytocin and Vasopressin Cells. *Hormones, Brain and Behavior*, 305–336. https://doi.org/10.1016/B978-0-12-803592-4.00058-4

Tekman, M., Batut, B., Ostrovsky, A., Antoniewski, C., Clements, D., Ramirez, F., Etherington, G. J., Hotz, H. R., Scholtalbers, J., Manning, J. R., Bellenger, L., Doyle, M. A., Heydarian, M., Huang, N., Soranzo, N., Moreno, P., Mautner, S., Papatheodorou, I., Nekrutenko, A., … Gruning, B. (2021). A single-cell RNA-sequencing training and analysis suite using the Galaxy framework. *GigaScience*, *9*(10). https://doi.org/10.1093/GIGASCIENCE/GIAA102

Torres-Soria, A. K., Romero, Y., Balderas-Martínez, Y. I., Velázquez-Cruz, R., Torres-Espíndola, L. M., Camarena, A., Flores-Soto, E., Solís-Chagoyán, H., Ruiz, V., Carlos-Reyes, Á., Salinas-Lara, C., Luis-García, E. R., Chávez, J., Castillejos-López, M., & Aquino-Gálvez, A. (2022a). Functional Repercussions of Hypoxia-Inducible Factor-2α in Idiopathic Pulmonary Fibrosis. En *Cells* (Vol. 11, Número 19). MDPI. https://doi.org/10.3390/cells11192938

Torres-Soria, A. K., Romero, Y., Balderas-Martínez, Y. I., Velázquez-Cruz, R., Torres-Espíndola, L. M., Camarena, A., Flores-Soto, E., Solís-Chagoyán, H., Ruiz, V., Carlos-Reyes, Á., Salinas-Lara, C., Luis-García, E. R., Chávez, J., Castillejos-López, M., & Aquino-Gálvez, A. (2022b). Functional Repercussions of Hypoxia-Inducible Factor-2α in Idiopathic Pulmonary Fibrosis. En *Cells* (Vol. 11, Número 19). MDPI. https://doi.org/10.3390/cells11192938

Wan, R., Wang, L., Zhu, M., Li, W., Duan, Y., & Yu, G. (2023a). Cellular Senescence: A Troy Horse in Pulmonary Fibrosis. *International Journal of Molecular Sciences*, *24*(22). https://doi.org/10.3390/IJMS242216410

Wan, R., Wang, L., Zhu, M., Li, W., Duan, Y., & Yu, G. (2023b). Cellular Senescence: A Troy Horse in Pulmonary Fibrosis. *International Journal of Molecular Sciences*, *24*(22). https://doi.org/10.3390/IJMS242216410

Wang, L., Wang, S., & Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184–2185. https://doi.org/10.1093/BIOINFORMATICS/BTS356

Wang, L., Zhao, W., Xia, C., Ma, S., Li, Z., Wang, N., Ding, L., Wang, Y., Cheng, L., Liu, H., Yang, J., Li, Y., Rosas, I., & Yu, G. (2024). TRIOBP modulates β-catenin signaling by regulation of miR-29b in idiopathic pulmonary fibrosis. *Cellular and Molecular Life Sciences*, *81*(1). https://doi.org/10.1007/S00018-023-05080-4

Way, G. P., Carninci, P., Carvalho, B. S., de Hoon, M., Finley, S. D., C Gosline, S. J., Lê Cao, K.-A., H Lee, J. S., Marchionni, L., Robine, N., Sindi, S. S., Theis, F. J., H Yang, J. Y., Carpenter, A. E., & Fertig, E. J. (2021). *A field guide to cultivating computational biology*. https://doi.org/10.1371/journal.pbio.3001419

Weger, B. D., Gobet, C., David, F. P. A., Atger, F., Martin, E., Phillips, N. E., Charpagne, A., Weger, M., Naef, F., & Gachon, F. (2021). Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3). https://doi.org/10.1073/PNAS.2015803118/-/DCSUPPLEMENTAL

Wood, V., Carbon, S., Harris, M. A., Lock, A., Engel, S. R., Hill, D. P., Van Auken, K., Attrill, H., Feuermann, M., Gaudet, P., Lovering, R. C., Poux, S., Rutherford, K. M., & Mungall, C. J. (2020). Term Matrix: a novel Gene Ontology annotation quality control system based on ontology term co-annotation patterns. *Open Biology*, *10*(9). https://doi.org/10.1098/RSOB.200149

Wu, H., Yu, Y., Huang, H., Hu, Y., Fu, S., Wang, Z., Shi, M., Zhao, X., Yuan, J., Li, J., Yang, X., Bin, E., Wei, D., Zhang, H., Zhang, J., Yang, C., Cai, T., Dai, H., Chen, J., & Tang, N. (2020). Progressive Pulmonary Fibrosis Is Caused by Elevated Mechanical Tension on Alveolar Stem Cells. *Cell*, *180*(1), 107-121.e17. https://doi.org/10.1016/j.cell.2019.11.027

Xijin Ge, S., & Jung, D. (s/f). *ShinyGO: a web application for in-depth analysis of gene sets*. Recuperado el 17 de enero de 2024, de http://ge-lab.org/go/

Xin, X., Yao, D., Zhang, K., Han, S., Liu, D., Wang, H., Liu, X., Li, G., Huang, J., & Wang, J. (2019). Protective effects of Rosavin on bleomycin-induced pulmonary fibrosis via suppressing fibrotic and inflammatory signaling pathways in mice. *Biomedicine & Pharmacotherapy*, *115*, 108870. https://doi.org/10.1016/J.BIOPHA.2019.108870

Yang, L., Chen, P. pan, Luo, M., Shi, W. lan, Hou, D. shun, Gao, Y., Xu, S. fu, & Deng, J. (2019). Inhibitory effects of total ginsenoside on bleomycin-induced pulmonary fibrosis in mice. *Biomedicine & Pharmacotherapy*, *114*, 108851. https://doi.org/10.1016/J.BIOPHA.2019.108851

Yépez, V. A., Gusic, M., Kopajtich, R., Mertes, C., Smith, N. H., Alston, C. L., Ban, R., Beblo, S., Berutti, R., Blessing, H., Ciara, E., Distelmaier, F., Freisinger, P., Häberle, J., Hayflick, S. J., Hempel, M., Itkis, Y. S., Kishita, Y., Klopstock, T., … Prokisch, H. (2022). Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Medicine*, *14*(1), 22. https://doi.org/10.1186/S13073-022-01019-9

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2). https://doi.org/10.1186/GB-2010-11-2-R14

Zeng, H. jin, Liu, Z., Wang, Y. ping, Yang, D., Yang, R., & Qu, L. bo. (2018). Studies on the anti-aging activity of a glycoprotein isolated from Fupenzi (Rubus chingii Hu.) and its regulation on klotho gene expression in mice kidney. *International Journal of Biological Macromolecules*, *119*, 470–476. https://doi.org/10.1016/J.IJBIOMAC.2018.07.157

Zhang, N., Shi, S., Jia, T. Z., Ziegler, A., Yoo, B., Yuan, X., Li, W., & Zhang, S. (2019). A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. *Nucleic Acids Research*, *47*(20), e125. https://doi.org/10.1093/NAR/GKZ731

Zhang, T. guang, & Miao, C. yu. (2023). Mitochondrial transplantation as a promising therapy for mitochondrial diseases. *Acta Pharmaceutica Sinica B*, *13*(3), 1028–1035. https://doi.org/10.1016/J.APSB.2022.10.008

Zhang, Y., Lu, L., & Li, X. (2022). Detection technologies for RNA modifications. *Experimental & Molecular Medicine*, *54*(10), 1601. https://doi.org/10.1038/S12276-022-00821-0

Zhang, Y., Luo, M., Wu, P., Wu, S., Lee, T. Y., & Bai, C. (2022). Application of Computational Biology and Artificial Intelligence in Drug Design. *International Journal of Molecular Sciences 2022, Vol. 23, Page 13568*, *23*(21), 13568. https://doi.org/10.3390/IJMS232113568