



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

**BÚSQUEDA DE GENES IMPLICADOS
A LA ALTA RESISTENCIA A
PIRAZINAMIDA BASADO EN EL
ANÁLISIS TRANSCRIPTÓMICO DE
*Mycobacterium tuberculosis***

**TESIS PARA OPTAR EL GRADO DE
MAESTRO EN MICROBIOLOGÍA**

FERNANDO MOISES MAMANI SANCA

LIMA – PERÚ

2022

ASESOR DE TESIS

Dra. Patricia Sheen Cortavarría

JURADO EVALUADOR

PRESIDENTE

MG. Dora Jesus Maurtua Torres

VOCAL

DR. Pablo Tsukayama Cisneros

SECRETARIO

MG. Juan Carlos Agapito Panta

AGRADECIMIENTOS

A mi padre Antonio Joaquín Mamani Hanco, mi madre Maruja Andrea Sanca Cho, hermano y hermanas que desde el comienzo de esta aventura nunca me dejaron de apoyar, y también indiscutiblemente a mis amigos.

DEDICATORIO

Dedicada a toda mi familia y a todo(a)s las personas que me ayudaron
con una sonrisa eterna.

FUENTES DE FINANCIAMIENTO

La realización de esta tesis fue llevada a cabo como gracias al
financiamiento por PROCIENCIA - CONCYTEC en el marco de la
convocatoria Proyecto Investigación Básica, 2018-01.

TABLA DE CONTENIDOS

RESUMEN

ABSTRACT

| | |
|------------------------------------------------------------------------------------------------------------------|----|
| 1. INTRODUCCIÓN..... | 1 |
| 2. MARCO TEÓRICO..... | 3 |
| 2.1. <i>Mycobacterium tuberculosis</i> : Agente causal..... | 3 |
| 2.1.1. Tratamiento..... | 4 |
| 2.2. La pirazinamida (PZA)..... | 4 |
| 2.2.1. Resistencia a pirazinamida..... | 5 |
| 2.2.2. Condiciones que afectan la susceptibilidad a PZA: Relación entre el pH y la concentración de PZA..... | 7 |
| 2.2.2.1. pH..... | 7 |
| 2.2.2.2. Tamaño del inóculo..... | 8 |
| 2.2.2.3. Fase de crecimiento y estado metabólico..... | 8 |
| 2.3. Transcriptómica..... | 8 |
| 2.3.1. Uso del análisis transcriptómico frente a la resistencia antibiótica..... | 9 |
| 2.3.2. Uso del análisis transcriptómico en <i>M. tuberculosis</i> | 10 |
| 3. HIPÓTESIS..... | 13 |
| 4. OBJETIVOS..... | 13 |
| 4.1. Objetivo general..... | 13 |
| 4.2. Objetivos específicos..... | 13 |
| 5. METODOLOGÍA..... | 14 |
| 5.1. Diseño del estudio..... | 14 |
| 5.2. Evaluación del crecimiento de <i>M. tuberculosis</i> H37Rv..... | 14 |
| 5.2.1. Estudio Preliminar y selección de las condiciones de cultivo: pH y concentración de PZA..... | 14 |
| 5.2.1.1. Condiciones de cultivo en medio líquido 7H9..... | 14 |
| 5.2.1.2. Medios de cultivo..... | 14 |
| 5.3. Cultivo y extracción de ARN de <i>M. tuberculosis</i> H37Rv a pH y concentraciones de PZA determinados..... | 16 |

| | |
|------------------------------------------------------------------------------------------------------------------|----|
| 5.3.1. Caracterización de las muestras de ARN de <i>M. tuberculosis</i> | 19 |
| 5.4. Preparación de librerías y secuenciación transcriptómica..... | 19 |
| 5.5. Análisis bioinformático..... | 20 |
| 5.5.1. Control de calidad..... | 20 |
| 5.5.2. Alineamiento y ensamblado de los reads..... | 21 |
| 5.5.3. Análisis de la expresión diferencial de genes y enriquecimiento de rutas metabólicas..... | 21 |
| 5.5.4. Selección de genes candidatos..... | 22 |
| 5.5.5. Gráfica Volcano..... | 22 |
| 6. RESULTADOS..... | 24 |
| 6.1. Evaluación y selección de las condiciones de cultivo: pH y concentración de PZA..... | 24 |
| 6.2. Cultivo y extracción de ARN de <i>M. tuberculosis</i> H37Rv a pH y concentraciones de PZA determinados..... | 27 |
| 6.2.1. Extracción de ARN..... | 27 |
| 6.3. Análisis de las librerías de cDNA..... | 29 |
| 6.4. Análisis bioinformático..... | 31 |
| 6.4.1. Control de calidad..... | 31 |
| 6.5. Genes de expresión diferencial..... | 33 |
| 6.5.1. Comparación de las condiciones de pH 6.3 a 0 vs 50 µg/ml de PZA..... | 33 |
| 6.5.2. Comparación entre pH 7.0 a 0 vs 50 µg/ml de PZA | 35 |
| 6.5.3. Comparación entre pH 7.0 y pH 6.3 ambos sin PZA..... | 37 |
| 6.5.4. Comparación entre pH 7.0 y pH 6.3 ambos con 50 µg/ml de PZA..... | 38 |
| 6.6. Selección de los genes diferencialmente expresados por la presencia de PZA y no por el cambio de pH..... | 39 |
| 6.7. Rutas metabólicas afectadas por PZA..... | 40 |
| 7. DISCUSIÓN..... | 44 |
| 8. CONCLUSIÓN..... | 53 |
| 8.1. Recomendaciones..... | 54 |

| | |
|--------------------------------|----|
| 9. ASPECTO ÉTICO..... | 54 |
| 10. BIBLIOGRAFÍA..... | 55 |
| 11. ANEXO | |
| 11.1. Comandos Bioinformáticos | |

RESUMEN

La tuberculosis (TB), enfermedad producida por la bacteria *Mycobacterium tuberculosis*, es tratada por un conjunto de medicamentos de primera línea incluyendo la pirazinamida (PZA). La PZA es altamente específica y elimina a las micobacterias en su estado semi-latente; sin embargo, aún se desconoce el mecanismo de acción. Las micobacterias resistentes a PZA por lo general presentan mutaciones en el gen *pncA* que codifica a la proteína pirazinamidasa (Pzasa), la cual transforma a la PZA en su estado activo, ácido pirazinoico (POA), pero se ha podido identificar cepas resistentes con Pzasas activas con ausencia de mutaciones en *pncA* o con mutaciones que no afectan regiones importantes de la enzima, sugiriendo la presencia de otros mecanismos. El objetivo en el presente trabajo fue identificar genes involucrados en la resistencia a PZA por medio del análisis transcriptómico de la cepa de *M. tuberculosis* H37Rv cultivada bajo diferentes estados de estrés (pH 6.3 y 7.0 y de 0 y 50 µg/mL de PZA). Las bibliotecas de cDNA generadas a partir del ARN fueron secuenciadas en pares terminales (paired-end). Las secuencias se alinearon con el genoma de referencia y las abundancias se cuantificaron utilizando Salmon. El análisis de expresión diferencial se realizó con el paquete edgeR, mientras que el análisis de enriquecimiento génico se realizó con GOseq. Los genes expresados diferencialmente (ED) se identificaron con un $\log_{2}FC \geq 2$ y ≤ -2 , y valor $p < 0,05$ umbral. De un total de 3979 genes analizados, se encontraron 13 genes de ED, los cuales están involucrados en actividad similar a las porinas, canales de poro ancho, transportadora transmembrana pasiva y acción antioxidante. El gen Rv0387c de la familia de proteínas PE/PPE, con función probable de porina o transportador

de moléculas, podría subexpresarse en presencia de PZA inhibiendo su entrada o salida del POA, aumentando la resistencia a PZA en *M. tuberculosis*.

Palabras clave: *Mycobacterium tuberculosis*, pirazinamida, transcriptoma, ARN, genes, expresión diferencial, secuenciación.

ABSTRACT

Tuberculosis (TB), a disease caused by the bacterium *Mycobacterium tuberculosis*, is treated by a set of first-line drugs including pyrazinamide (PZA). PZA is highly specific and kills mycobacteria in their semi-dormant state; however, the mechanism of action is still unknown. PZA-resistant mycobacteria usually have mutations in the *pncA* gene that encodes the protein pyrazinamidase (Pzase), which transforms PZA into its active state, pyrazinoic acid (POA), but it has been possible to identify resistant strains with Pzases active with the absence of mutations in *pncA* or with mutations that do not affect important regions of the enzyme, suggesting the presence of other mechanisms. The objective in this work was to identify genes involved in resistance to PZA by means of transcriptomic analysis of the *M. tuberculosis* strain H37Rv cultivated under different stress states (pH 6.3 and 7.0 and 0 and 50 $\mu\text{g/mL}$ of PZA). The cDNA libraries generated from the RNA were sequenced in paired-end pairs. Sequences were aligned to the reference genome and abundances were quantified using Salmon. Differential expression analysis was performed with the edgeR package, while gene enrichment analysis was performed with Goseq. Differentially expressed (DE) genes were identified with $\log\text{FC} \geq 2$ and ≤ -2 , and p-value < 0.05 threshold. From a total of 3979 genes analyzed, 13 ED genes were found, which are involved in porin-like activity, wide pore channels, passive transmembrane transporter and antioxidant action. The Rv0387c gene of the PE/PPE family of proteins, with a probable porin or molecular transporter function, could be underexpressed in the presence of PZA, inhibiting its entry or exit from the POA, increasing resistance to PZA in *M. tuberculosis*.

Keywords: Mycobacterium tuberculosis, pyrazinamide, transcriptome, RNA, genes, differential expression, sequencing.

1. INTRODUCCIÓN

La tuberculosis (TB) es una enfermedad infecciosa provocada por *Mycobacterium tuberculosis*, que con mayor frecuencia afecta a los pulmones. La Organización Mundial de la Salud (OMS) reporta que la TB es una de las 10 principales causas de muerte en todo el mundo (1) y por ende es altamente estudiada a nivel mundial. La presencia de cepas de *M. tuberculosis* que son resistentes a los medicamentos de isoniazida y rifampicina (MDR) incrementan su letalidad (2).

La pirazinamida (PZA), antibiótico, que forma parte del tratamiento de primera línea con acción única debido a que elimina a *M. tuberculosis* en estado semi-latente. Esta característica permitió la reducción del tratamiento de 12-9 meses a 6 meses, y su uso se mantiene en los tratamientos de TB MDR (4, 5), pero el mecanismo de acción concreta de la actividad de PZA todavía no es conocida.

La PZA es una pro-droga por lo que requiere de la actividad de la proteína pirazinamidasa de *M. tuberculosis* (PZA_{sa}) para transformarse a su estado activo ejerciendo así su actividad bactericida. (6). El mecanismo conocido hasta ahora de PZA comienza con su ingreso a la bacteria por medio de difusión pasiva en donde requiere de la enzima citoplasmática pirazinamidasa (PZ_{asa}) codificada por el gen *pncA*, para convertirse a su estado activo antituberculoso conocido como ácido pirazinoico (POA). El mecanismo de acción de POA está en investigación; se plantea que podría estar actuando como un protonóforo acidificando el medio intracelular, alterando el potencial de membrana que finalmente es letal para la bacteria (4, 7) así como afectando algunos posibles blancos como la proteína ribosomal S1 (RpsA), enzima L-aspartato descarboxilasa (PanD) y al enzima ácido graso sintasa (FasI).

El principal mecanismo de resistencia está asociado a la pérdida de la actividad de la PZAsa debido a la presencia de mutaciones en el gen *pncA* (72-97%); sin embargo, estudios recientes indicarían que la resistencia a PZA podría estar relacionada a la presencia de mutaciones en genes que codifican a RpsA, PanD y Fasl.

El uso de nuevas técnicas para la identificación de nuevos genes relacionados a la resistencia a PZA, es importante porque permitirían el desarrollo de nuevos medicamentos que puedan complementar el tratamiento actual contra la TB y evitar la generación de cepas de TB MDR. La transcriptómica es una opción altamente viable en la evaluación de los diferentes genes que puedan estar expresándose ante un tipo de estrés, en este caso por la PZA, permitiendo la identificación de los posibles mecanismos implicados con la resistencia mediada por la transcripción del ADN. De Welzen *et al.* encontró que la alta resistencia a etionamida está mediada por la combinación de mutaciones luego de comparar el transcriptoma de aislados clínicos. Esto demuestra que las modificaciones transcripcionales globales podrían contribuir a la resistencia o el mantenimiento de la aptitud bacteriana (fitness) (18). En este estudio se evaluó el efecto de la PZA en el transcriptoma de *M. tuberculosis* en pH neutro y ácido (pH 6.3) como un factor de estrés, permitiendo observar la expresión de los posibles genes implicados y deducir su importancia en la alta resistencia a la pirazinamida para entender los diferentes mecanismos que podrían estar participando en el desarrollo de la resistencia a PZA.

2. MARCO TEÓRICO

2.1. *Mycobacterium tuberculosis*: Agente causal

La especie bacteriana *M. tuberculosis*, es el agente causal que produce la enfermedad de la tuberculosis (TB). *M. tuberculosis* pertenece al género *Mycobacterium*, es un bacilo Gram positivo, ácido alcohol resistente, con dimensiones de entre 0.2-0.6 micras (μm) de ancho y 1-10 (μm) de largo (19), recto o ligeramente curvo, aerobio estricto, sin movimiento, de crecimiento lento y no formadora de esporas ni cápsulas (20, 21). La propiedad de ácido-resistencia es debido a la presencia de ácidos micólicos en su pared, y este también atribuye un factor importante en la identificación por tinción (22).

M. tuberculosis pertenece al denominado complejo tuberculosis, micobacterias capaces de producir enfermedades (*M. caprae*, *M. canetti*, *M. bovis*, *M. africanum* y *M. microti*) (21, 23). A pesar que el humano es su hospedero preferente, *M. tuberculosis* ha sido hallado en animales tanto domésticos como salvajes. Su lugar de residencia por lo general es en los pulmones, ya que en estos órganos presenta oxígeno necesario para su proliferación. *M. tuberculosis* es una bacteria de crecimiento lento con un periodo de división de 18-20 horas, capaz de resistir ambientes fríos, a la congelación y desecación (23).

Una de las características que la hacen diferencial a otras bacterias es su pared celular, la cual está constituida por diferentes moléculas como el complejo micolil-arabinogalactano-peptidoglicano (mAGP), ejerciendo un papel importante en el mantenimiento de una estructura de la membrana estable esencial para su viabilidad (24). También presenta una pared rica en ácidos micólicos encargada de la unión

del complejo mAGP con las membranas superiores, es altamente hidrofóbico y está relacionada a la virulencia (25).

La complejidad de la pared de *M. tuberculosis*, por los diferentes compuestos presentados además de las bombas proteicas y su baja permeabilidad, está involucrada a la resistencia a los diferentes fármacos y su supervivencia en ambientes fuera de su hospedero (26).

2.1.1. Tratamiento

El tratamiento de TB en pacientes sin riesgo de presentar TB-MDR se basa en una terapia combinada con medicamentos de primera línea aplicado por 6 meses que contiene isoniazida (INH), rifampicina (RIF), etambutol (ETM) y pirazinamida (PZA) (3).

2.2. La pirazinamida (PZA)

La PZA, análoga de la nicotinamida es un bactericida con acción única, elimina a *M. tuberculosis* en estado semi-latente, característica que permitió la reducción del tratamiento de 12-9 meses a 6 meses, y su uso se mantiene en los tratamientos de TB MDR (4, 5), pero el mecanismo de acción concreta de la actividad de PZA todavía no es conocida.

Así mismo, se sabe que la PZA actúa a pH ácido, similar al ambiente en donde se encuentra *M. tuberculosis* luego de ser fagocitado por los macrófagos (4, 6). Además, es altamente específica y es una pro-droga por lo que requiere de la actividad de la proteína PZAsa de *M. tuberculosis* para transformarse a su estado activo ejerciendo así su actividad bactericida. (6). El mecanismo conocido hasta

ahora de PZA comienza con su ingreso a la bacteria por medio de difusión pasiva, ya dentro la PZA necesita de la enzima citoplasmática PZasa codificada por el gen *pncA*, para convertirse a su estado activo antituberculoso conocido como ácido pirazinoico (POA). Producido el POA es expulsado por un sistema de bombas de flujos al medio extracelular en donde al encontrarse en un ambiente ácido, una porción del POA expulsado se protona convirtiéndose en una especie no cargada (HPOA) con mayor permeabilidad que permite su reingreso a la bacteria y ya dentro este liberaría el protón captado extracelularmente, acidificando el medio intracelular, alterando el potencial de membrana que finalmente es letal para la bacteria (4, 7).

2.2.1. Resistencia a pirazinamida

A raíz que la actividad de la PZA depende de su conversión a su estado activo por parte de la enzima PZasa, hay una gran variedad de estudios que reportan que la resistencia a la PZA es explicada por la mutación en el gen *pncA* (72-97%). Sin embargo, las mutaciones se distribuyen en forma diversa en todo el marco de lectura (ORF) del gen *pncA* que comprende 561 pares de bases (pb) modificando de manera variable la actividad de la PZasa (8, 9). Por lo tanto, diferentes trabajos se enfocan en el descubrimiento de nuevos mutantes “Hot spot” o mutaciones relacionadas para la identificación de susceptibilidad a PZA; también reportan que el uso de la secuenciación es la mejor opción en la detección de cepas resistentes a PZA con una sensibilidad de 91,3% y una especificidad de 96,8% (9, 10,11).

Pero la resistencia a la PZA no solo se basa en las mutaciones en el gen *pncA* de forma exclusiva, como se menciona anteriormente; el mecanismo de PZA aún no

es totalmente conocido y estudios recientes indicarían que PZA podría tener otras vías target las cuales estarían afectando en forma conjunta a la resistencia por dicho medicamento (27).

En los últimos años diferentes estudios han demostrado que la resistencia a PZA también se debe a mutaciones en posibles targets de unión del POA, como por ejemplo Yang *et al.* encontró mutaciones en el gen que codifica a la proteína RpsA (12, 13, 14). Además, Shuo Zhang *et al.* demostró que ciertas cepas de *M. tuberculosis* resistentes a PZA sin mutaciones en los genes *pncA* y *rpsA*, presentaban mutaciones en el gen *panD* que codifica al aspartato decarboxilasa (participa en la síntesis de β -alanina), por lo que también podrían estar ligado a la resistencia a PZA (15). Por otra parte, Bharati Pandey *et al.* identificó una doble mutación en histidina a arginina en el residuo 21 (H21R) e isoleucina a valina en el residuo 29 (I49V) del gen *panD* dando lugar a resistencia a PZA (24) y por todo lo anterior Anthony RM *et al.* reafirmó que el gen *panD* es el objetivo principal de POA (16). Por otro lado, Sheen *et al.* al realizar un análisis del genoma completo de 68 cepas de *M. tuberculosis* de pacientes con TB, encontró cepas resistentes a PZA sin presencia de mutaciones en el gen *pncA* y de estas cepas se pudo identificar nuevos genes asociados significativamente con resistencia PZA, estos genes están relacionados con el transporte de iones metálicos, regulación del pH y la evasión del sistema inmunológico; sugiriendo nuevos posibles mecanismo de resistencia a PZA como también el efecto de PZA a múltiples blancos (17).

2.2.2. Condiciones que afectan la susceptibilidad a PZA: Relación entre el pH y la concentración de PZA

PZA es activo contra *M. tuberculosis* en condiciones únicas, múltiples factores afectan su actividad como el estado metabólico del bacilo tuberculoso, el tamaño del inóculo, la fase de crecimiento, el pH, la concentración del fármaco, entre otros (28, 29, 30). Curiosamente, un común denominador entre todos ellos es que la susceptibilidad a PZA se favorece en condiciones de estrés en *M. tuberculosis*. Por todo esto, Anthony & et al plantearon la hipótesis de que los factores de estrés relacionados con el estado de latencia inducen el cambio hacia un fenotipo sensible a PZA, en donde, el o los targets de POA son esenciales para la supervivencia de *M. tuberculosis* (27).

2.2.2.1. pH

La característica notable de PZA es que actúa in vitro a pH ácidos; en medios alcalinos, el fármaco pierde su actividad. La PZA puede probarse cuando el rango de pH está entre 5,5 y 6,0 pero la concentración mínima inhibitoria (MIC) cambia de 50 a aproximadamente 100 µg/ml de PZA (30, 72). El cambio de MIC puede predecirse mediante la ecuación de Henderson-Hasselbach dado que PZA se hidroliza completamente a POA (30). Aunque se ha demostrado la actividad de la PZA a pH 5,5, la PZA mata sólo el 72% de la población incluso cuando se utiliza una concentración alta de PZA (1000 µg/ml) (31), concentración 20 veces mayor que las concentraciones encontradas en el suero humano (50 µg/ml) (32).

2.2.2.2. Tamaño del inóculo

Un tamaño de inóculo grande reduce en gran medida la actividad de PZA. El aumento del tamaño del inóculo de 10^7 a 10^8 células/ml provocó la inactivación de la PZA (33). Este efecto puede explicarse por el aumento del pH por el amonio, producto de la desaminación de la PZA por la PZasa bacteriana. El pH en cultivos con 10^7 - 10^8 células/ml puede aumentarse de 5,5 a casi neutro (30), mientras que se observó un pequeño aumento del pH con 10^6 células/ml, un inóculo comúnmente utilizado en las pruebas de sensibilidad a la PZA.

2.2.2.3. Fase de crecimiento y estado metabólico

La PZA es significativamente más activa contra los bacilos tuberculosos semi-latentes (34) y los bacilos viejos (7, 30) que tienen un metabolismo menos activo y menores reservas de energía que los bacilos más jóvenes y de rápido crecimiento. Las reservas de energía se ven fácilmente afectadas.

2.3. Transcriptómica

La transcriptómica es el estudio del total de ARN mensajero de un organismo en una circunstancia específica. La transcriptómica permite comparar expresiones diferenciales de genes entre poblaciones celulares en respuesta a diferentes condiciones u tratamientos.

Para su análisis existen diferentes métodos de uso de alto rendimiento como el microarrays y la tecnología de NSG (Next generation sequencing), cada uno de estos específicos y de gran utilidad en sus estudios respectivos. El microarrays útil para estudios de rutina de secuencias diana, mientras que el NSG llamado "RNA-

seq” permite comprender de forma detallada y masiva la diversidad de genes expresados en corto tiempo (35).

2.3.1. Uso del análisis transcriptómico frente a la resistencia antibiótica

La transcriptómica ha sido aplicada en el estudio del mecanismo de enfermedades, diagnóstico clínico y el desarrollo de nuevos fármacos de enfermedades humanas como animales (36). Denis *et al* reportó potenciales targets farmacológicos en enfermedades neurológicas y psiquiátricas mediante el análisis de células de tejido cerebral pudo identificar 5 genes. En el tratamiento del cáncer, los tumores que presentan cierto tipo de similitud de expresión pueden ser tratados en forma “dirigida” (38). La transcriptómica también es útil en la identificación de moléculas que se activan ante la presencia del fármaco y su interacción con su objetivo, permitiendo identificar nuevos objetivos y desarrollo de nuevos fármacos o uso de los ya existentes (39, 36).

La transcriptómica está siendo aplicada también en la identificación de los genes implicados en la resistencia y sensibilidad a los fármacos. Los microorganismos usan diferentes vías metabólicas para su supervivencia ante el uso indiscriminado de antibióticos, desencadenando la aparición de microorganismos MDR. Oluwasegun *et al.* analizó el transcriptoma de biopelículas de *Candida albicans*, levadura patogénica oportunista, en presencia de concentraciones subinhibidoras de ácido araquidónico y fluconazol encontrando expresión diferencial de genes involucrados en la síntesis de metionina y generación de ATP, la biosíntesis de AMP, el metabolismo del fumarato y la oxidación de ácidos grasos (40). Ariane *et*

al. mediante la secuenciación cualitativa de ARN de 135 *Pseudomonas aeruginosa*, identificó genes asociados a la resistencia de fluroquinolonas, aminoglucósidos y β -lactámicos, como también posibles biomarcadores con relación directa a la resistencia, permitiendo brindar información en la elaboración de futuros fármacos, como también herramientas de diagnóstico molecular para su rápido y dirigido tratamiento, mitigando potenciales evoluciones de resistencias futuras (41).

2.3.2. Uso del análisis transcriptómico en *M. tuberculosis*

En las últimas décadas, los esfuerzos por conocer mejor a *M. tuberculosis*, el uso del análisis transcriptómico a providenciado de información valiosa para entender cuáles podrían ser sus mecanismos de defensa o facultades alternas en su supervivencia ante sus diferentes factores de estrés como también ayudar a identificar el modo de acción y predecir los objetivos de los medicamentos actuales o potencialmente nuevos contra la tuberculosis. Dentro de estos análisis se ha podido identificar que uno de los mecanismos de *M. tuberculosis* frente a un ambiente adverso es la carencia de una secuencia Shine-Dalgarno (SD) en su gran mayoría de genes (cuarta parte del transcriptoma), reflejando una gran plasticidad de *M. tuberculosis* en su regulación transcripcional, a diversas condiciones de estrés (42, 43).

Wilson *et al.* en su estudio, monitorearon los cambios en la expresión génica en respuesta a la isoniazida, destacando la inducción de cinco genes organizados como un operón (*fabD-acpM-kasA-kasB-accD6*) que codifican las enzimas sintasa de ácidos grasos tipo II (FAS-II), involucradas en la producción de fosfolípidos de membrana (44). Boshoff *et al.* estudió la respuesta transcripcional a los antibióticos

y condiciones inhibitorias en *M. tuberculosis*, donde genes que estaban regulados se agrupaban en función de su mecanismo de acción conocida y logró predecir el mecanismo de acción de varios agentes desconocidos (45). Mientras tanto, Walter *et al.* estudió la expresión genética de tuberculosis pulmonar sensible a fármacos tratados con la combinación estándar de isoniazida, rifampicina, PZA y etambutol, encontrando una disminución en la abundancia de ARNm bacteriano que demostró la muerte de la mayoría de las bacterias 4 días después de comenzar el tratamiento, pero la abundancia de ARNm bacteriano después de eso indicó tolerancia al fármaco. Los genes implicados en el crecimiento, el metabolismo y la síntesis de lípidos estaban regulados a la baja en las bacterias tolerantes a los fármacos, mientras que los implicados en la respuesta al estrés, como los del módulo de toxina-antitoxina (TA), y las bombas de salida del fármaco estaban regulados al alza (46).

En la búsqueda de nuevos compuestos naturales antimicobacteriano contra *M. tuberculosis*, Liang *et al.* estudió el efecto de la queleritrina, demostrando que hay muchos genes que estaban regulados al alza o la baja e involucrados en varias vías por acción de este compuesto, lo que podría deberse a una respuesta adaptativa o la inhibición de la traducción (47). También ciertas mutaciones pueden ser compensatorias frente a prejuicios producidas por una primera mutación contra un medicamento específico, Zhihong Xu *et al.* analizó el perfil transcripcional total de 6 cepas resistentes con mutaciones *rpoB* y *rpoB/rpoC* donde se observó que los efectos de la mutación en *rpoC* regula positivamente los genes de las proteínas ribosómicas y en las vías de síntesis de aminoácidos. Estas mutaciones podían participar en la respuesta compensadora ante la interrupción del metabolismo

energético y la maquinaria de fosforilación oxidativa por la mutación en *rpoB* (48). Por otro lado, Gerjo J. *et al.* realizó un estudio donde analizó el transcriptoma de 2 cepas de *M. tuberculosis* una sensible y otra resistente a rifampicina, con una mutación H526Y en el gen *rpoB*. Durante un tiempo de exposición, encontraron que los genes Rv0559c y Rv0560c fueron inducidos fuertemente en la cepa resistente sugiriendo que estos 2 genes juegan un papel importante en la resistencia a rifampicina (49). Frente a TB extremadamente resistentes (XDR) Guohua Yu *et al.* realizó un estudio transcripcional de 2 cepas de TB XDR y 1 sensible contra 3 medicamentos antituberculosos, hallando en estas cepas XDR, tres genes reguladores significativamente enriquecidos (*regX3*, *mprA* y *mtrA*) que muestran sobreexpresión que pertenecen al sistema regulador de dos componentes, lo que sugiere vínculos entre estos sistemas de respuesta y resistencia.

3. HIPÓTESIS

- La presencia de PZA produce expresión genética diferencial de resistencia en *Mycobacterium tuberculosis*.

4. OBJETIVOS

4.1. Objetivo General

Determinar que genes podrían estar relacionados a la resistencia de PZA mediante el análisis transcriptómico (ARN-seq) de *M. tuberculosis* a diferentes concentraciones de PZA y pHs.

4.2. Objetivos específicos

- Determinar las concentraciones de PZA y pH que permita el crecimiento del 50% de la cepa de referencia H37Rv de *M.tuberculosis* en 4 semanas.
- Analizar el transcriptoma de la cepa de referencia H37Rv de *M. tuberculosis* bajo las condiciones seleccionadas de PZA y pH.
- Determinar los posibles genes implicados a una alta eficiencia de resistencia a PZA, mediante la diferencia cuantitativa de los cultivos estresados y control.
- Inferir la importancia biológica de los posibles genes implicados en el mecanismo de acción de PZA.

5. METODOLOGÍA

5.1. Diseño del estudio

El estudio experimental fue realizado con la cepa de referencia sensible a los antibióticos incluyendo PZA ATCC de *M. tuberculosis* H37Rv en el laboratorio de Mycobacterium de bioseguridad de nivel 3 de los Laboratorios de Investigación y Desarrollo (LID) de la Universidad Peruana Cayetano Heredia.

5.2. Evaluación del crecimiento de *M. tuberculosis* H37Rv

5.2.1. Estudio preliminar y selección de las condiciones de cultivo: pH y concentración de PZA.

5.2.1.1. Condiciones de cultivo en medio líquido 7H9

M. tuberculosis fue cultivado en medio líquido Middlebrook 7H9 conteniendo glicerol y OADC con 0, 5, 50, 100 y 400 µg/ml de PZA a los siguientes pHs: 6.0, 6.2, 6.5 y 7.0. Se evaluó la inhibición del 50% del crecimiento bacteriano a comparación del medio sin PZA.

. La concentración de 100 µg/ml, considerado el MIC a pH 6.0 fue incluido como control de inhibición).

5.2.1.2. Medios de cultivo

Caldo 7H9 Middlebrook (7H9-OADC). El caldo fue preparado de acuerdo a las indicaciones del proveedor. Brevemente, el medio 7H9 Middlebrook (BD Difco) fue suplementado con glicerol. El pH fue modificado con NaOH y HCl y autoclavado. El medio finalmente fue suplementado con OADC al 10% (ácido oléico, albúmina, dextrosa y catalasa). y tween 80 al 0.05%.

Agar 7H10 Middlebrook (7H10-OADC). El agar fue preparado de acuerdo a las indicaciones del proveedor. El medio 7H10 Middlebrook (BD Difco) fue suplementado con 5% de OADC.

Stock de PZA. A partir de un stock de 15 mg/ml diluido en agua miliQ de pyrazinecarboxamide (PZA) (Sigma-Aldrich) se prepararon stocks de 15000, 3700, 1850 y 185 µg/ml. Los stocks fueron usados para obtener las concentraciones finales de PZA en 10 ml de caldo de cultivo 7H9-OADC (Tabla 1).

| | Concentración Inicial de PZA | Volumen a usar de los stocks | Concentración Final de PZA por cultivo |
|---------|------------------------------|------------------------------|----------------------------------------|
| Stock 1 | 15000 µg/ml | 0.27 ml | 400 µg/ml |
| Stock 2 | 3700 µg/ml | 0.27 ml | 100 µg/ml |
| Stock 3 | 1850 µg/ml | 0.27 ml | 50 µg/ml |
| Stock 4 | 185 µg/ml | 0.27 ml | 5 µg/ml |

Tabla 1. Preparación de las 4 concentraciones de PZA finales en medio 7H9-OADC.

Inóculo de la cepa de referencia ATCC *M. tuberculosis* H37Rv. La cepa de *M. tuberculosis* H37Rv fue sembrado en 2 placas de agar 7H10-OADC e incubadas 2 semanas a 37°C. El cultivo bacteriano fue trasladado con un asa de siembra de 10 µl a un tubo de 15 ml con 6 perlas de vidrio y 1 ml de solución salina al 0.85%. El contenido bacteriano fue homogenizado usando un vortex durante 2 min seguido de 10 min de reposo para evitar los posibles aerosoles generados. Esta suspensión fue transferida por goteo con una pipeta de transferencia a un tubo conteniendo 20 ml de solución salina hasta llegar a un OD₆₀₀ de 1.2.

Preparación del medio de cultivo con su condición respectiva. Se preparó 10 ml del medio de cultivo 7H9-OADC para cada condición de pH y concentración de PZA con *M. tuberculosis* a un OD final de 0.07. Los componentes fueron: 9.23 ml del caldo 7H9-OADC al pH indicado, 0.5 ml de la cepa (OD de 1.2) y 0.27 ml de PZA de la concentración indicada para llegar a las concentraciones finales del acuerdo a la Tabla 1.

Crecimiento de *M. tuberculosis* H37Rv y registro. La evaluación del efecto de PZA y el pH se realizó mediante la determinación del OD cada 7 días durante 4 semanas. Los cultivos fueron incubados en el shaker MaxQ 4450 (Thermo Fisher Scientific) a 200 rpm a 37°C. Cada tubo fue homogenizado manualmente y 1ml fue transferido a una cubeta para la medición.

La contaminación por crecimiento de hongos u otros organismos fue verificado en los cultivos de caldo luego de 2 semanas de crecimiento. Una alícuota de 50 µl fue extruido de los cultivos en caldo 7H9-OADC para ser agregados en placas con agar 7H10-OADC e incubadas 5 días; luego del cual la placa fue revisada para detectar presencia de contaminantes.

5.3. Cultivo y extracción de ARN de *M. tuberculosis* H37Rv a pH y concentraciones de PZA determinados

Los resultados del estudio anterior se usaron para seleccionar las 4 condiciones finales para el inicio del estudio del transcriptoma. El criterio que se tuvo en cuenta fue que el porcentaje del crecimiento final del cultivo no fuera más del 50% del crecimiento en el medio sin PZA. Dónde las condiciones de pH y PZA seleccionadas fueron: 6.3 y 7.0 de pH y concentraciones de 0 y 50 µg/ml de PZA.

Baja carga bacteriana a pH 6.3 y alta carga bacteriana a pH 7.0 con actividad inhibitoria que no sobrepase el 50% del medio sin PZA (50 µg/ml).

Cultivo en medio 7H9. Los cultivos fueron realizados en volúmenes de 15 ml mezclando lo siguiente: 1.0 ml del cultivo de *M. tuberculosis* (OD₆₀₀: 1.3), 13.6 ml del caldo 7H9-OADC y 0.4 ml de PZA (Figura 1) y 50 µg/ml de PZA. 0.4 ml de caldo 7H9-OADC fue reemplazado en las condiciones en los medios sin PZA. El OD inicial fue de 0.07. Los cultivos fueron realizados por triplicado. La incubación de los cultivos se realizó por 3 semanas a 200 rpm a 37°C, obteniendo ODs de entre 0.4 a 1.0.

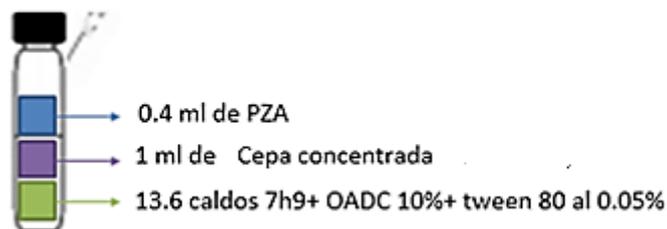


Figura 1. Volúmenes en el cultivo de *M. tuberculosis* H37Rv para extracción de ARN.

Extracción de ARN. Los cultivos fueron cosechados cuando alcanzaron un OD₆₀₀ entre 0.4 (cultivos con PZA) y 1.0 (cultivos sin PZA) mediante centrifugación a 3000 x g por 20 min a 17°C. El sobrenadante fue descartado y el precipitado fue lavado con 5 ml de buffer TE Tris EDTA pH 8.0 (preparado con agua DEPC), se homogenizó suavemente con una pipeta de transferencia y se realizó nuevamente una centrifugación a 3000 x g por 20 min a 17°C. Finalmente, al pellet se agregó

600 μ l o 1000 μ l de reactivo Trizol TRI Reagent (Zymoresearch) y se homogenizó vigorosamente con una pipeta de transferencia.

La variación de volúmenes del reactivo Trizol es respecto al OD obtenido en los cultivos, con ODs mayor a 0.7 se utilizó 1 ml de trizol y para ODs menores se utilizó 600 μ l. Luego se transfirió la suspensión homogenizada a tubos tapa rosca de 2 ml con beads de 0.1 mm de diámetro de Zirconia/Silica (BioSpec Products) y se almacenó en hielo para su posterior uso.

Las células fueron lisadas en el homogeneizador Fast Prep a 6.5 m/s durante 30 seg, realizando 2 ciclos y entre cada ciclo se colocó los tubos en hielo por 3 min. Los tubos fueron centrifugados a 12500 x g por 15 min a 4°C y mantenidos fríos en una caja tecnopor con hielo. Luego se transfirió la fase soluble (fase superior de Trizol) a un tubo de 1.5 ml libres de ARNs, para después agregar 200 μ l de cloroformo mezclándolo durante 15 seg. de forma manual agitando vigorosamente y se dejó durante 10 min en hielo. Después, se centrifugó a 12500 x g por 15 min a 4°C. Se transfirió la fase superior, sin tocar la interface (ADN de color blanco), a otro tubo de 1.5 ml libres de ARNs agregando 500 μ l de etanol absoluto frio y homogenizando por inversión unas 10 veces, terminado el homogenizado se almacenó a -20°C por 30 min. Después se centrifugó a 12500 x g por 15 min a 4°C, se decantó el etanol de un golpe y lo restante se retiró con un tip de 200 μ l con cuidado de no tocar el pellet. Luego se agregó 1 ml de etanol al 75% frio (diluido con agua DEPC), se invirtió 10 veces y se centrifugó nuevamente a 12500 x g por 5 min a 4°C. Se decantó el etanol nuevamente de un solo golpe y se centrifugó a 12500 x g por 1 min y 30 seg. a 4°C, se descartó el etanol sobrante con un tip de 200 μ l y se dejó secar por 5 min a temperatura ambiente o hasta que se evapore el

etanol sobrante. Para finalizar, el pellet fue re suspendido con 20 μ l de agua libre de DNasas y RNasas; y se almacenó a -70°C .

5.3.1. Caracterización de las muestras de ARN de *M. tuberculosis*

La cantidad de ARN se determinó usando el kit Qubit RNA HS Assay (Invitrogen), la integridad y calidad del ARN se determinó usando el kit Qubit™ RNA IQ Assay (Invitrogen). La concentración y pureza del ARN también fue determinado mediante el análisis de los ratios 260/280 y 260/230, respectivamente, obtenidos en el Nanodrop 2000 (Thermo Scientific). La integridad de las muestras fue analizada, también, por medio de corridas electrofónicas, donde se realizó un mix de la muestra con el reactivo 2X RNA Loading Dye (Thermo Scientific), para luego ser analizadas por electroforesis en gel de 1.5% de agarosa por 30 min a 95V en el PowerPac Basic (Bio-Rad) en buffer TBE 1X. Los geles fueron teñidos con bromuro de etidio por 15 min.

5.4. Preparación de librerías y secuenciación transcriptómica

Al menos 2 μ g (hasta 10 μ g) de ARN en un volumen de 20 μ l fue usado para obtener ARN mensajero (ARNm). El ARN ribosomal fue degradado con el kit Ribominus Bacteria Transcriptome Isolation (Invitrogen). La secuenciación realizada en los extremos 5' y 3' (Paired-End) con el objetivo de obtener mayor información de las librerías elaboradas evitando posibles errores de lecturas de los nucleótidos por sesgos de replicación y errores inducidos durante la amplificación de las secuencias de ADN. Los fragmentos menores de 300 pares de bases (pb) fueron generados

usando el kit TruSeq stranded mARN Library Prep (Illumina) y las lecturas fueron de 75 pares de bases tanto del extremo 5' y 3' (2 x 75pb).

Las librerías fueron cuantificadas mediante el kit Qubit™ 1X dsDNA HS Assay (Invitrogen) en el Qubit 4 y 6 de las 12 librerías de cDNA escogidas al azar fueron analizadas en una electroforesis en gel de agarosa. Las muestras fueron secuenciadas en el secuenciador masivo NextSeq 550 (Illumina) con una profundidad de 5 millones de reads por muestra para obtener datos de alta calidad (50).

5.5. Análisis bioinformático

5.5.1. Control de calidad

Tras la secuenciación, la calidad de las lecturas, o reads, fue evaluado mediante el programa FastQC v0.11.9 (51), permitiendo conocer los estándares de calidad: i) calidad QC (52) y ii) el porcentaje de guaninas y citosinas (%GC). El quality scores o QC es el promedio de puntaje que se obtiene de cada uno de los nucleótidos que conforman la secuencia, que se refiere a la probabilidad de dicho nucleótido haya sido asignado por error (52). El QC es calificado dentro de un rango de calidad de 0-34, donde los $QC \geq 28$ son considerados de muy buena calidad, indicando que la base está correctamente posicionada y con una precisión de 99.9 %. Por otro lado, conociendo el porcentaje de GC del genoma de *M. tuberculosis* (~60 %), se espera que el promedio de % GC de los reads sea cercano a este. Por lo tanto, se tomaron en cuenta los reads con un $QC > 28$ y $\% GC < 60$. Se realizaron las comparaciones

entre la posición (base pair) vs. su calidad (QC o Quality Control); y porcentaje total de guaninas y citosinas (%GC) vs. número de secuencias en la muestra.

5.5.2. Alineamiento y ensamblado de los reads

El proceso de alineamiento y ensamblado de los reads de cada una de las muestras se realizó usando el genoma de referencia *M. tuberculosis* H37Rv obtenida de la base de datos ENSEMBL Bacteria (53) y se utilizó el programa Salmon versión 1.6.0 (54) para cuantificar la abundancia de cada uno de los genes, expresada como *counts per gene*, o número de reads alineados para cada gen, generando una matriz de conteo de reads.

5.5.3. Análisis de la expresión diferencial de genes y enriquecimiento de rutas metabólicas

El análisis de la expresión diferencial se realizó a partir de la matriz de conteo de reads. Primero, las muestras fueron normalizadas usando el método TMM (55); luego, las medias de la expresión de cada gen fueron comparadas entre todas las muestras usando el paquete estadístico EdgeR v3.36.0 de Bioconductor v3.14 (56) el cual permite determinar la expresión diferencial de cada gen entre las muestras, al comparar la expresión o abundancia de cada transcrito. Además, se realizó el proceso de normalización de los reads a sus genes asignados por medio del CPM (Counts per million), permitiendo hacer la corrección diferencial de la profundidad de la secuenciación entre las muestras es decir una comparación de recuento de genes entre réplicas del mismo grupo de muestras.

El modelo fue corregido por medio del batch effects, permitiendo evitar los sesgos introducidos por los 3 procesos de secuenciación realizadas.

Los resultados obtenidos fueron expresados como tasa de cambio o fold change (FC) y se filtraron en base al p-value (menor a 0.05) asociado a cada comparación entre condiciones con una significancia de 95%.

Para determinar las rutas metabólicas o procesos biológicos afectados por la presencia de PZA de los genes seleccionados, se usó el análisis de enriquecimiento genético (GEA) siendo esta una técnica que proporciona anotaciones funcionales de un conjunto de genes, implementando a la par el paquete GOSep v1.46, herramienta de análisis ontológica genética (GO) que permite la categorización de genes para datos RNA-seq (57).

5.5.4. Selección de genes candidatos

La selección de genes fue realizada teniendo en cuenta los resultados del logaritmo del fold change (logFC): Se seleccionaron los genes con mayor o iguales a 2 logFC (up-regulated), sobreexpresión y menores o iguales de -2 logFC (down-regulated), subexpresión. Los logaritmos de fold change p-value menores de 0.05 fueron seleccionados. La identificación de los genes seleccionados se realizó en la base de datos abierta de la página del NCBI.

5.5.5. Gráfica Volcano

La gráfica Volcano es un diagrama de dispersión que permite identificar rápidamente cambios significativos de una gran número de datos replicados, integrando los valores del fold change y significancia estadística por medio de dos

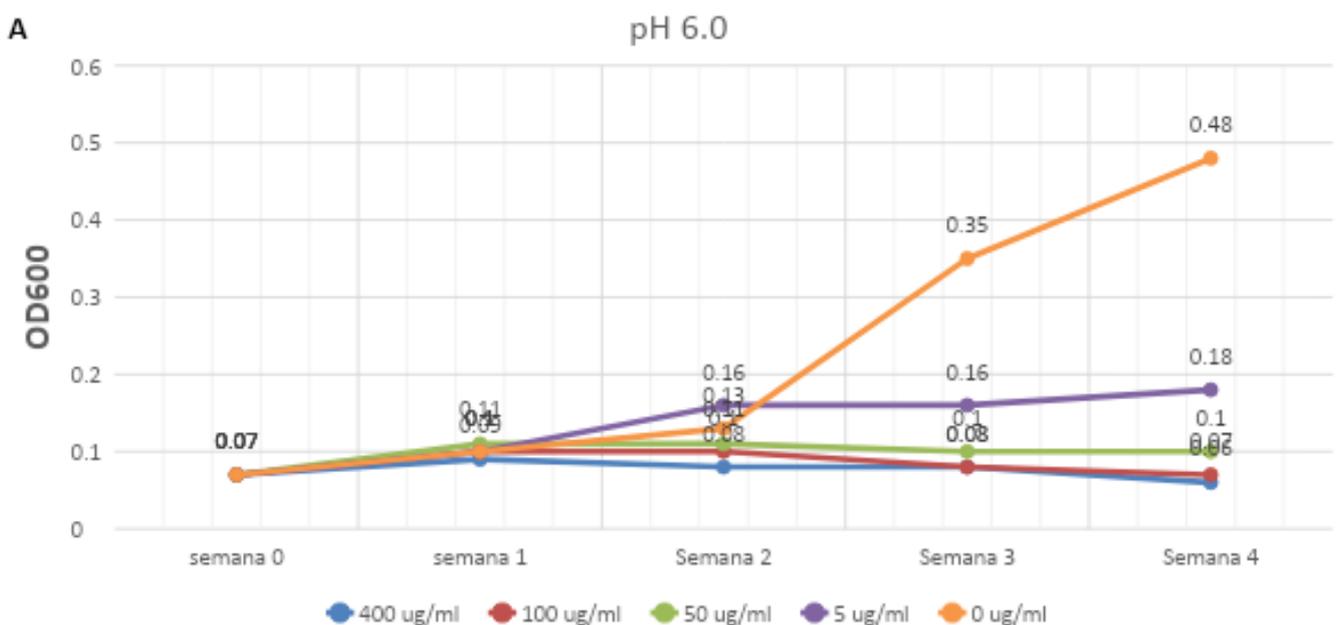
ejes: eje Y, representado por el $-\log$ en base 10 del p-value y eje X, representado por el \log en base 2 del fold change de las 2 condiciones evaluadas, permitiendo identificar valores de expresión diferencial (58), en el caso del presente estudio de aquellos genes que tuvieron una expresión alterada significativa frente a los tratamientos de estrés aplicados (pH y PZA) a *M. tuberculosis*.

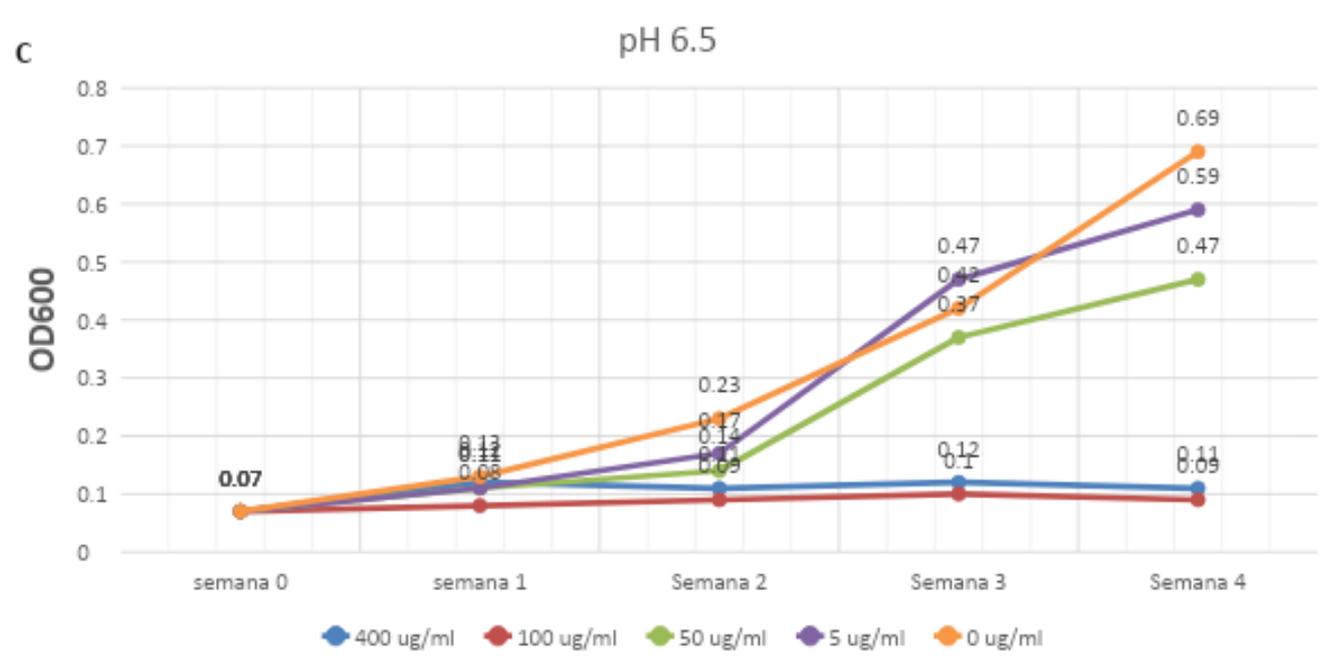
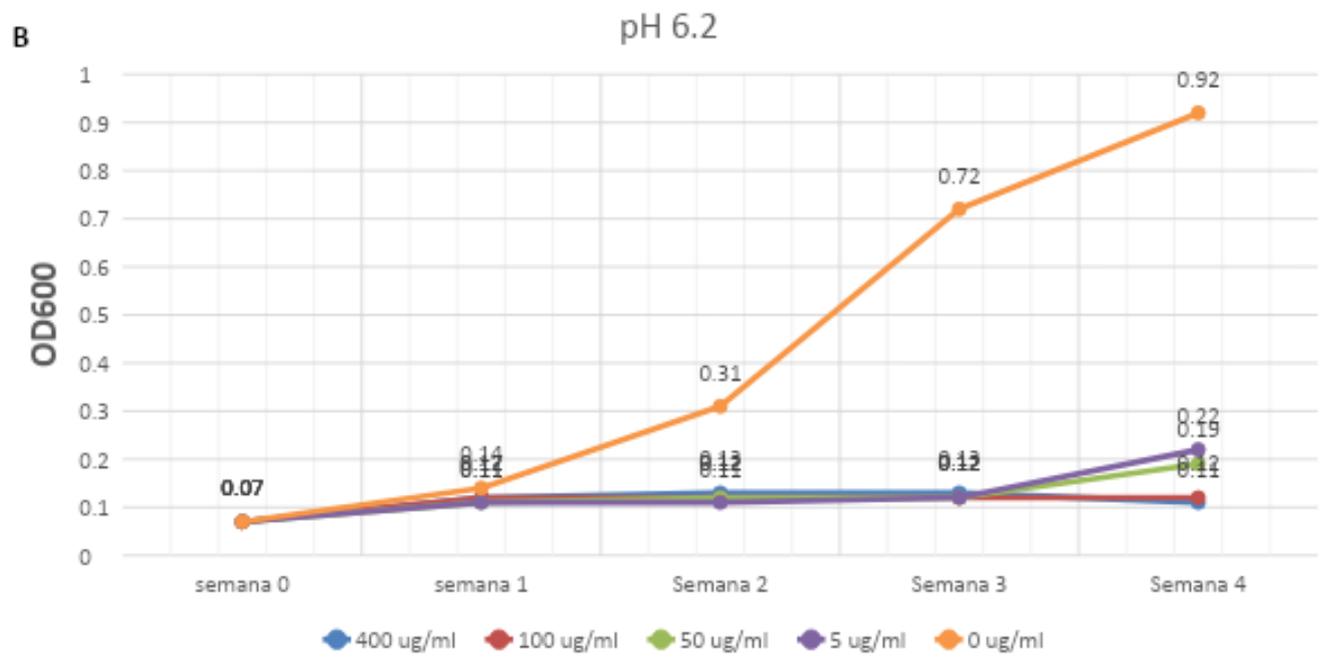
Dentro de la selección de datos en el gráfico, para una mejor visualización de los genes de expresión diferencial (ED) significativos, se seleccionaron dos puntos de corte en el eje X (Log_2 FC); -2 para genes subexpresados y +2 para genes sobre expresados.

6. RESULTADOS

6.1. Evaluación y selección de las condiciones de cultivo: pH y concentración de PZA

Crecimiento de *M. tuberculosis* a diferentes condiciones. En la Figura 2 se muestra el crecimiento en todas las concentraciones de PZA y pHs. El crecimiento fue bajo a pH 6, el OD en el medio control (sin PZA) fue de 0.48 y los OD máximo en los medios con PZA fue de 0.18 en la concentración más baja de PZA (5 $\mu\text{g/ml}$). En pH 6.2 el crecimiento en el medio control alcanzó un OD de 0.92 y a la concentración de 50 $\mu\text{g/ml}$ se observó un OD de 0.19; es decir, menos del 50% del crecimiento del medio sin PZA. En pH 6.5, el porcentaje de crecimiento en 50 $\mu\text{g/ml}$ sobrepasa el 50% del crecimiento del medio sin PZA. En pH 7.0, el medio control pasó el OD de 1 y a la concentración de 50 $\mu\text{g/ml}$ no pasó el 50% del crecimiento sin PZA.





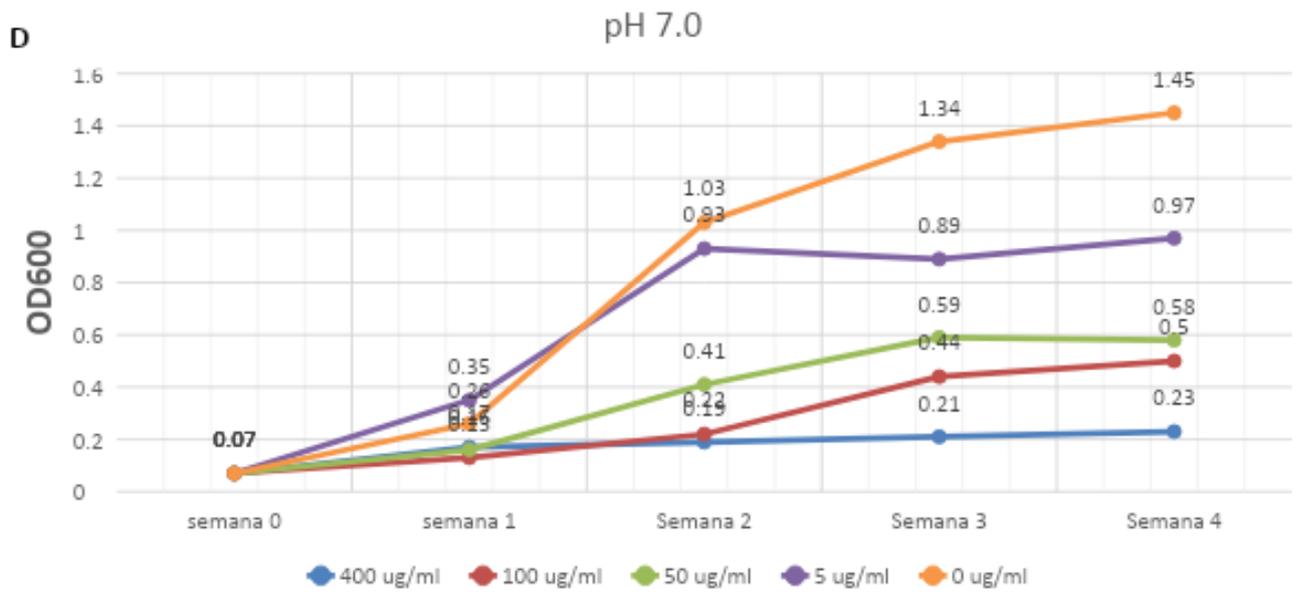


Figura 2. Crecimiento de *M. tuberculosis* H37Rv (OD₆₀₀) durante 4 semanas en diferentes pH y concentraciones de PZA. A) pH 6.0, B) pH 6.2, C) pH 6.5 y D) pH 7.0.

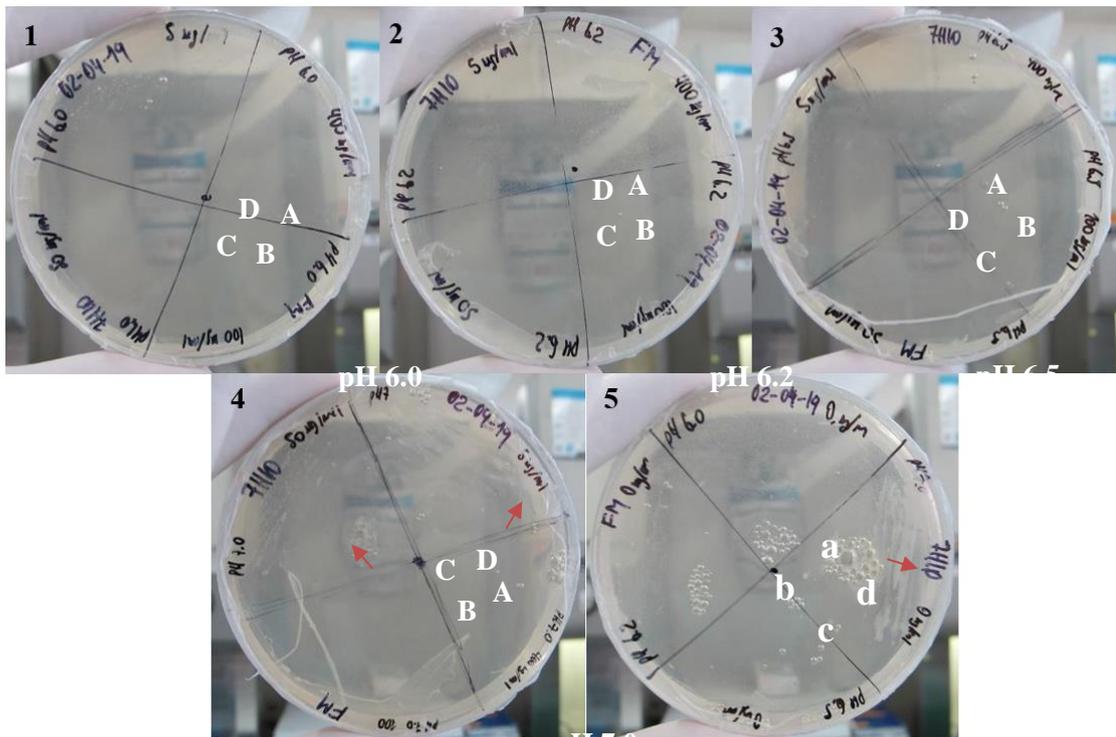


Figura 3. Cultivo negativo de contaminación en medio 7H10 de los cultivos de *M. tuberculosis* H37Rv. Cultivos a pHs 1) 6.0, 2) 6.2, 3) 6.5, 4) 7.0 con sus diferentes concentraciones de A) 400; B) 100; C) 50; D) 5 μ g/ml de PZA. La placa 5 comprende

los cultivos a una concentración de 0 µg/ml en pHs: 5.a) pH 6.0; 5.b) pH 6.2; 5.c) 6.5 y 5.d) 7.0. (→) crecimiento leve de la cepa *M. tuberculosis* H37Rv.

No se observó contaminación de los caldos 7H9-OADC luego de la siembra en el medio de cultivo 7H10-OADC para *M. tuberculosis* (Figura 3). Además, se observó un ligero crecimiento anticipado en los cultivos con pH 7.0(neutro) con concentraciones bajas de PZA (Figura 3.4 y 3.5)

6.2. Cultivo y extracción de ARN de *M.tuberculosis* H37Rv a pH y concentraciones de PZA determinados

Las condiciones seleccionadas para realizar los experimentos fueron: pH 6.3 y 7.0 con 0 y 50 µg/ml (4 condiciones). Se escogió el medio con pH 6.3 con el objetivo de aumentar ligeramente la proporción del crecimiento entre el medio sin PZA y en el medio con PZA sin sobrepasar el 50% del crecimiento en el medio sin PZA. El medio a pH 7.0 fue seleccionado como control en el que se verifica el efecto de PZA, pero en altas concentraciones bacterianas. Usando estas condiciones, *M. tuberculosis* fue cultivado durante 3 semanas.

6.2.1. Extracción de ARN

Los resultados del análisis del ARN extraído de los cultivos en las 4 condiciones realizadas en triplicado (12 muestras) son mostrados en la Tabla 2. En resumen, en cada muestra se obtuvo más de 2 µg de ARN con alto grado de pureza (IQ mayor a 7) e integridad evidenciada con la ausencia de degradación y ADN y presencia de las bandas que corresponden al ARN ribosomal: 23S, 16S y 5S en la Figura 4. Los parámetros de las muestras de ARN fueron óptimos para realizar el agotamiento y preparación de librerías.

| N° muestra | Repetición | pH | PZA µg/ml | ARN total ng/µl (Nanodrop) | Ratio 260/280 | Ratio 260/230 | ARN total ng/µl (Qubit 4) | IQ (Qubit 4) |
|------------|------------|-----|--------------|----------------------------------|------------------|------------------|---------------------------------|-----------------|
| 1 | 1 | 7 | 0 | 898 | 1.97 | 2.06 | 1120 | 8.5 |
| 2 | 1 | 7 | 50 | 640 | 1.94 | 2.08 | 610 | 9.1 |
| 3 | 1 | 6.3 | 0 | 747 | 1.94 | 1.87 | 680 | 7.7 |
| 4 | 1 | 6.3 | 50 | 218 | 1.74 | 1.85 | 207 | 7.1 |
| 5 | 2 | 7 | 0 | 864 | 1.91 | 2.16 | 810 | 8.2 |
| 6 | 2 | 7 | 50 | 2109 | 1.99 | 1.89 | 2130 | 7.4 |
| 7 | 2 | 6.3 | 0 | 578 | 1.92 | 1.93 | 540 | 7.3 |
| 8 | 2 | 6.3 | 50 | 280 | 1.8 | 1.81 | 196 | 6.9 |
| 9 | 3 | 7 | 0 | 415 | 1.93 | 1.8 | 360 | 7.9 |
| 10 | 3 | 7 | 50 | 356 | 1.88 | 1.93 | 332 | 6.9 |
| 11 | 3 | 6.3 | 0 | 1204 | 1.97 | 1.91 | 1200 | 7.9 |
| 12 | 3 | 6.3 | 50 | 1001 | 1.93 | 1.80 | 886 | 8.5 |

Tabla 2. Caracterización de la extracción de ARN de las muestras de *M. tuberculosis* H37Rv. Realizadas por Nanodrop 2000 y Qubit 4.

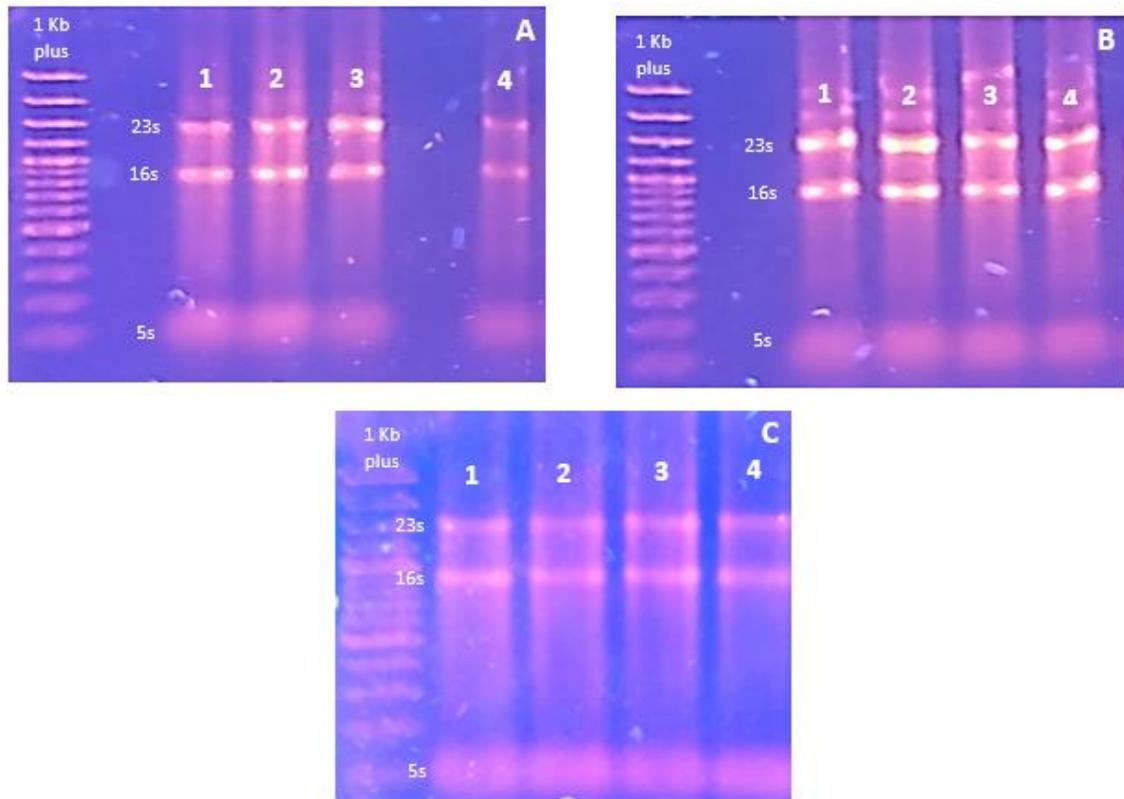


Figura 4. Observación de ARN total de *M. tuberculosis* H37R. A) Primera repetición, B) segunda repetición y C) tercera repetición. Las condiciones de cultivo: 1) pH 6.3 sin PZA; 2) pH 6.3 con 50 µg/ml de PZA; 3) pH 7.0 sin PZA y 4) pH con 50 µg/ml de PZA. Presencia de las subunidades ribosomales 23S, 16S y 5S, y ausencia de ADN.

6.3. Análisis de las librerías de cDNA

Los resultados de la electroforesis en gel de agarosa (Figura 5) confirman la óptima elaboración de librerías, presentando éstas un tamaño mayor a 260 pb y concentraciones de más de 10 ng/µl (Tabla 3).

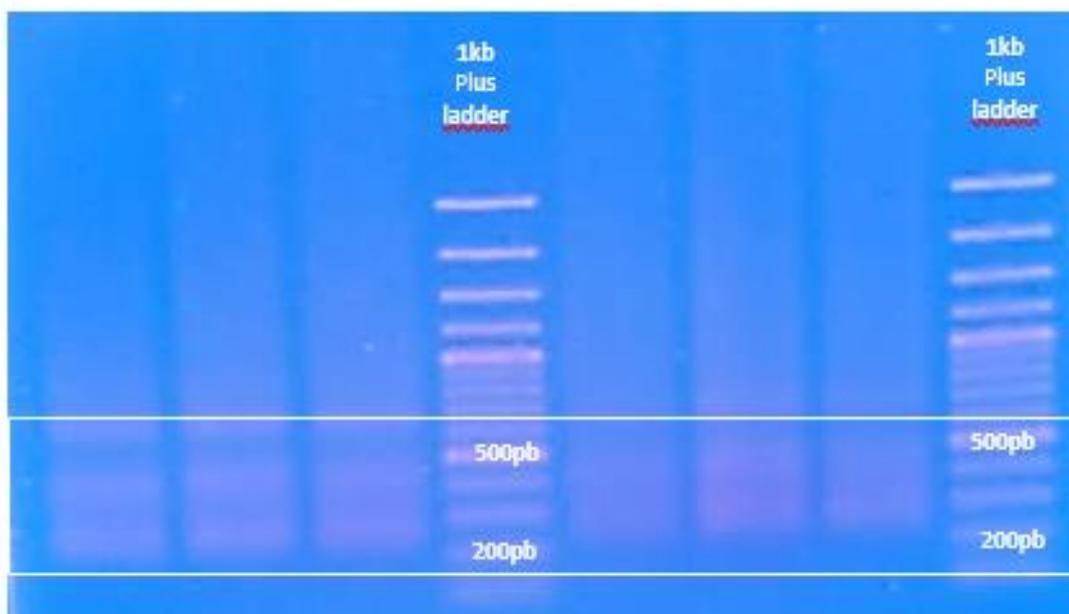


Figura 5. Electroforesis de las librerías de cDNA con bandas de tamaños mayores a 260 pb.

| Muestra | pH | Pza($\mu\text{g/ml}$) | cDNA total $\text{ng}/\mu\text{l}$ |
|---------|-----|-------------------------|------------------------------------|
| 1 | 7 | 0 | 30.0 |
| 2 | 7 | 50 | 30.8 |
| 3 | 6.3 | 0 | 30.8 |
| 4 | 6.3 | 50 | 33.0 |
| 5 | 7 | 0 | 45.0 |
| 6 | 7 | 50 | 46.8 |
| 7 | 6.3 | 0 | 43.0 |
| 8 | 6.3 | 50 | 45.6 |
| 9 | 7 | 0 | 35.2 |
| 10 | 7 | 50 | 28.6 |
| 11 | 6.3 | 0 | 30.4 |
| 12 | 6.3 | 50 | 41.2 |

Tabla 3. Cuantificación de las librerías de *M. tuberculosis* H37Rv usando Qubit 4.

6.4. Análisis bioinformático

6.4.1. Control de calidad

La calidad obtenida fue mayor a $QC \geq 28$ (Figura 8, sector verde) lo cual significa que 1 base de 1000 bases generadas fue incorrecta obteniendo un porcentaje de precisión del 99.9% para cada muestra. Por cada muestra secuenciada mediante “Paired-end” se obtuvieron 2 archivos FastQC, tanto de su extremo 5´y 3´ (Figura 6).

La línea azul demarca la calidad de la media, y ésta se mantiene en el sector verde donde están los $QC \geq 28$, indicando que la secuencia es de óptima calidad. Cabe destacar que para las últimas bases hay un pequeño declive (Figura 8A) indicando un comportamiento normal en el transcurso del secuenciamiento (a mayor longitud de pares de bases pueden aparecer pequeños errores de lectura) procediendo a eliminar y corregir para su posterior trabajo en el alineamiento y análisis de expresión diferencial genética.

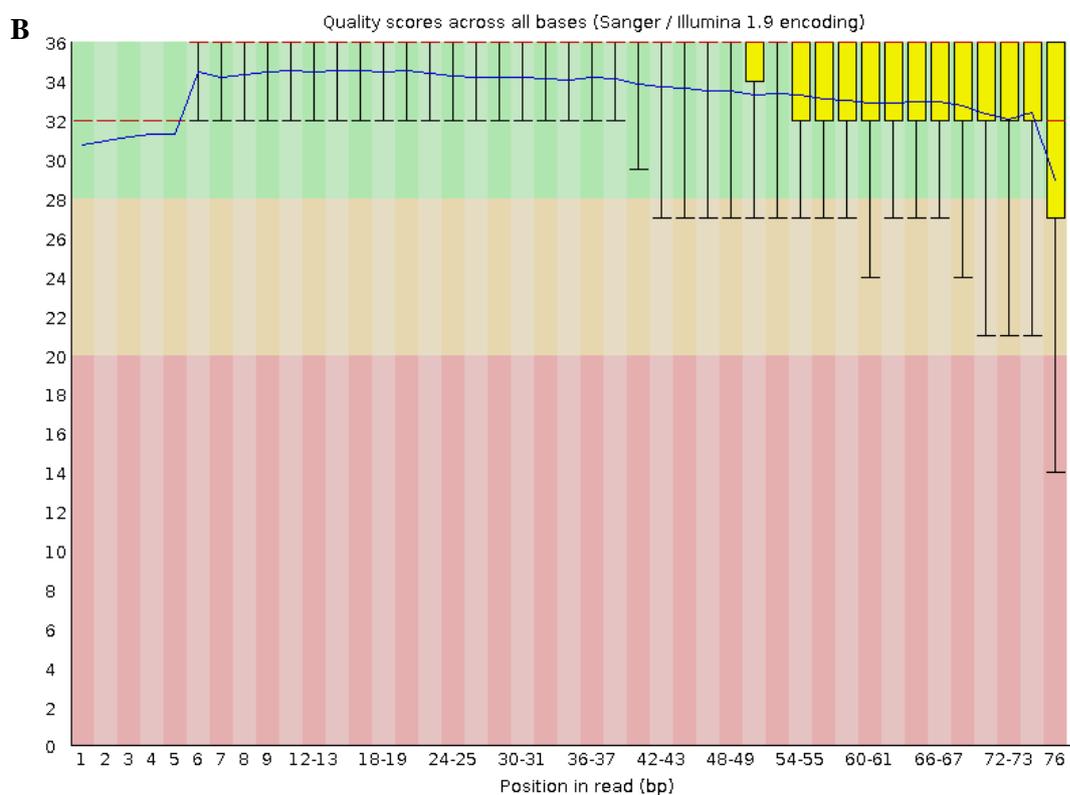
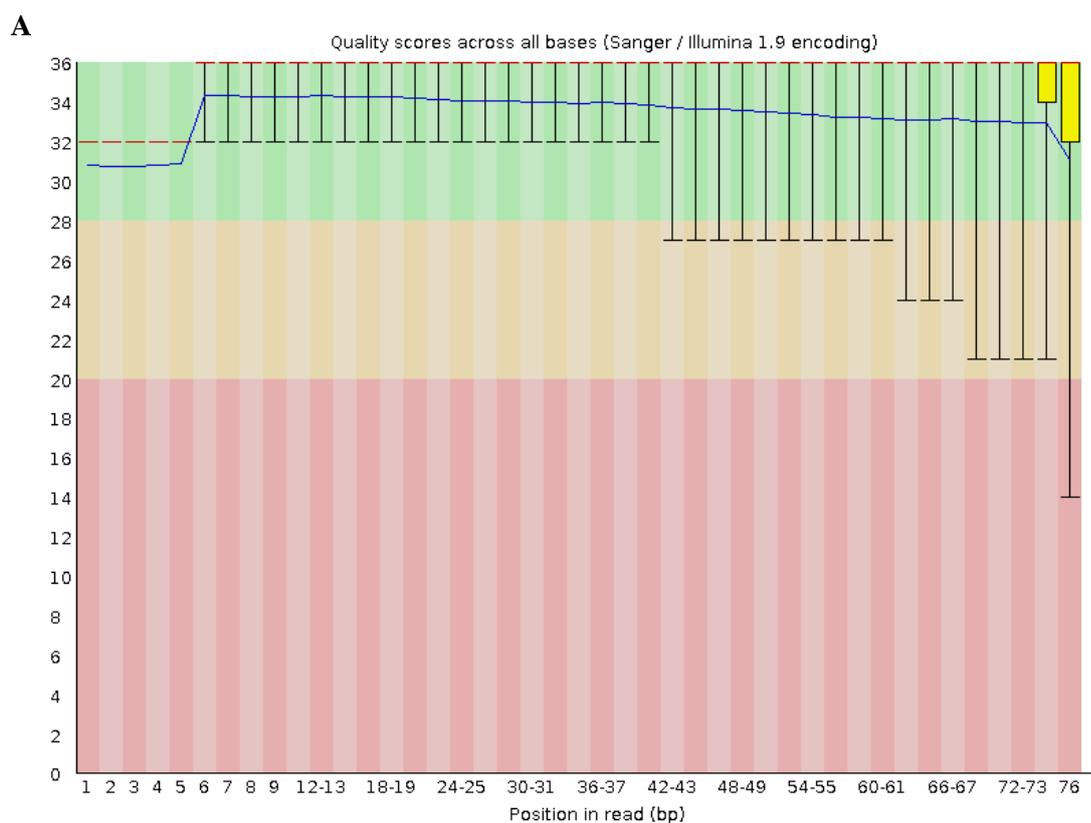


Figura 6. Grafica de quality control (QC) vs position in read (Bp) de la muestra

1(pH 7- PZA 0ug/ml). A: lecturas 5', B: lecturas 3'.

6.5. Expresión genética diferencial

Se secuenciaron 3979 genes, los cuales se utilizaron para realizar 2 comparaciones: pH 7 con 0 $\mu\text{g/ml}$ PZA contra pH 7.0 con 50 $\mu\text{g/ml}$ de PZA (pH7, 0 vs 50 $\mu\text{g/ml}$ de PZA) y pH 6.3 con 0 $\mu\text{g/ml}$ de PZA contra pH 6.3 con 50 $\mu\text{g/ml}$ (pH 6.3, 0 vs 50 $\mu\text{g/ml}$ de PZA). También para poder descartar algún gen de expresión diferencial que esté relacionado al efecto del pH y no por PZA, se realizó la comparación de la expresión genética entre pH 7.0 y pH 6.3 en ausencia y presencia de 50 $\mu\text{g/ml}$ PZA.

6.5.1. Comparación de las condiciones de pH 6.3 a 0 vs 50 $\mu\text{g/ml}$ de PZA

Once genes se expresaron diferencialmente, 8 de éstos se subexpresaron (Down-regulated) Rv1714, Rv2635, Rv0387c, Rv3113, Rv0601c, Rv1574, Rv2913c y Rv2651c; y 3 se sobreexpresaron (up-regulated) Rv2626c, Rv2275 y Rv2274A (Figura 7).

H37Rv pH 6.3, pza 0 vs. pH 6.3, pza 50

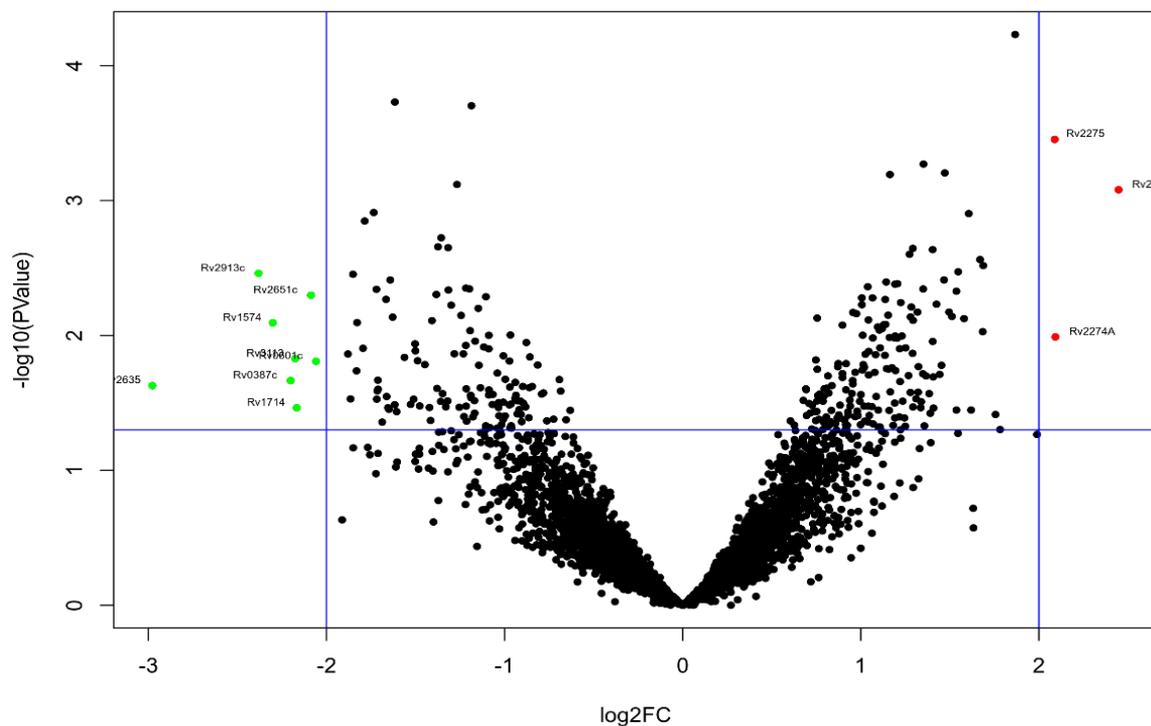


Figura 7. Volcano-plot Log₂FC vs -Log₁₀ (p-value) de la comparación de condición de pH 6.3 con 0 vs 50 µg/ml de PZA en *M. tuberculosis* H37Rv. Los puntos verdes indican subexpresión (-2) y los rojos sobreexpresión (+2).

| Genes | Descripción | Expresión | Log2FC | FC | P-value |
|---------|---------------------------------|-----------|--------|-------|---------|
| Rv1714 | Oxidoreductase | DR | -2.152 | 0.225 | 0.036 |
| Rv2635 | Hypothetical protein | DR | -2.941 | 0.130 | 0.026 |
| Rv0387c | Pseudogen | DR | -2.193 | 0.219 | 0.025 |
| Rv3113 | Phosphatase | DR | -2.178 | 0.221 | 0.012 |
| Rv0601c | Two component sensor kinase HK2 | DR | -2.044 | 0.242 | 0.012 |
| Rv1574 | Phage protein | DR | -2.297 | 0.204 | 0.007 |
| Rv2913c | D-amino acid aminohydrolase | DR | -2.381 | 0.192 | 0.004 |
| Rv2651c | Prophage protease | DR | -2.086 | 0.235 | 0.003 |
| Rv2626c | Hypoxic response protein | UR | 2.448 | 5.455 | 0.000 |

| | | | | | |
|---------|-------------------------------------|----|-------|-------|-------|
| Rv2275 | Cyclo(L-tyrosyl-L-tyrosyl) synthase | UR | 2.074 | 4.211 | 0.000 |
| Rv2274A | Antitoxin MazE8 | UR | 2.086 | 4.245 | 0.009 |

Tabla 4. Genes de expresión diferencial candidatos con p-value, logFC significativos y descripción biológica de la cepa de *M. tuberculosis* H37Rv en la comparación de sin PZA (0 µg/ml) contra la presencia de PZA (50 µg/ml) a un pH 6.3. FC: Fold chang; DR: Down-regulated; UR: Up-regulated.

El gen Rv2626c fue el de mayor expresión diferencial en el grupo de genes de sobreexpresión (up-regulated), con un logFC de 2.4 y un p-value de <0.01, siendo identificado como una “Proteína de respuesta hipóxica”, mientras el gen Rv2635 tubo la mayor subexpresión (Down-regulated) con un logFC de -2.9 y un p-value de 0.026, identificada como una “Proteína hipotética” (Tabla 4).

6.5.2. Comparación entre pH 7.0 a 0 vs 50 µg/ml de PZA

Cuatro genes se sobreexpresaron (Figura 8) (Rv3599c, Rv2307D, Rv2654c y Rv2816c), donde Rv3599c tuvo la mayor expresión en comparación de las restantes (Tabla 5).

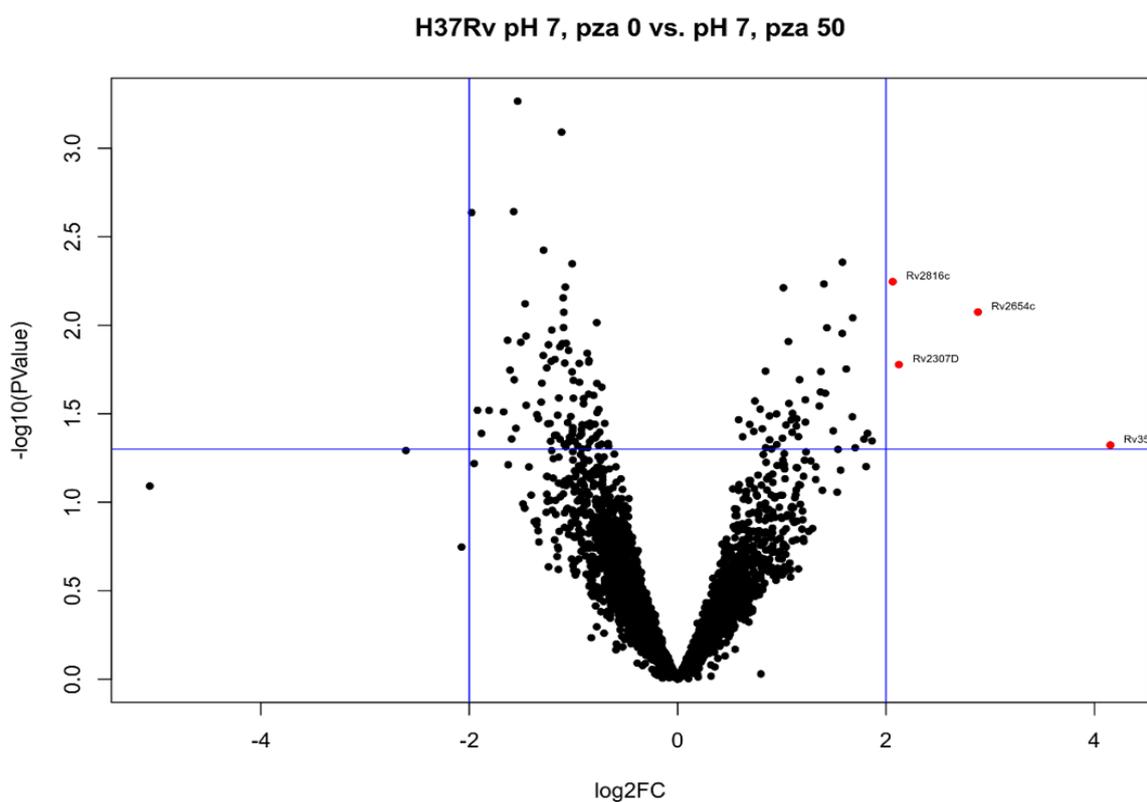


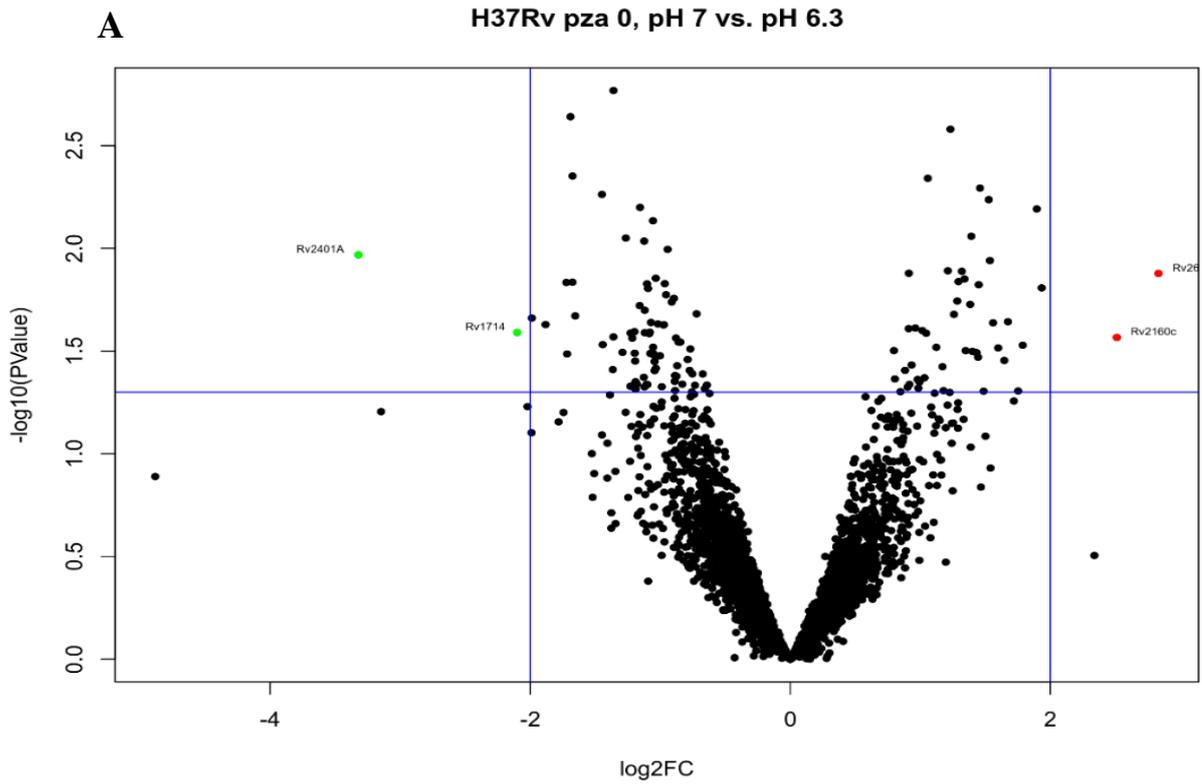
Figura 8. Volcano-plot LogFc vs $-\text{Log}_{10}(\text{p-value})$ de la comparación de condición de pH 7.0 con 0 vs 50 $\mu\text{g/ml}$ de PZA en *M. tuberculosis* H37Rv. Los puntos rojos indican sobreexpresión (+2).

| Genes | Descripción | Expresión | LogFC | FC | P-Value |
|---------|-----------------------------------------|-----------|-------|--------|---------|
| Rv3599c | Hypothetical short protein | UR | 3.985 | 15.835 | 0.047 |
| Rv2307D | Hypothetical protein | UR | 2.105 | 4.302 | 0.018 |
| Rv2654c | Antitoxin | UR | 2.871 | 7.320 | 0.004 |
| Rv2816c | CRISPR-associated endoribonuclease Cas2 | UR | 2.074 | 4.212 | 0.002 |

Tabla 5. Genes con expresión diferencial con p-value, logFC significativos y descripción biológica de la cepa de *M. tuberculosis* H37Rv en la comparación de sin PZA contra la presencia de PZA (50 $\mu\text{g/ml}$) a un pH 7. FC: Fold change, CPM: counts per million. UR: Up-regulated.

6.5.3. Comparación entre pH 7.0 y pH 6.3 ambos sin PZA

Ocho genes se expresaron diferencialmente; 2 de los cuales se subexpresaron (Rv1714 v 2401A) y 2 se sobreexpresaron (Rv2160c y Rv2654c) (Figura 9).



B

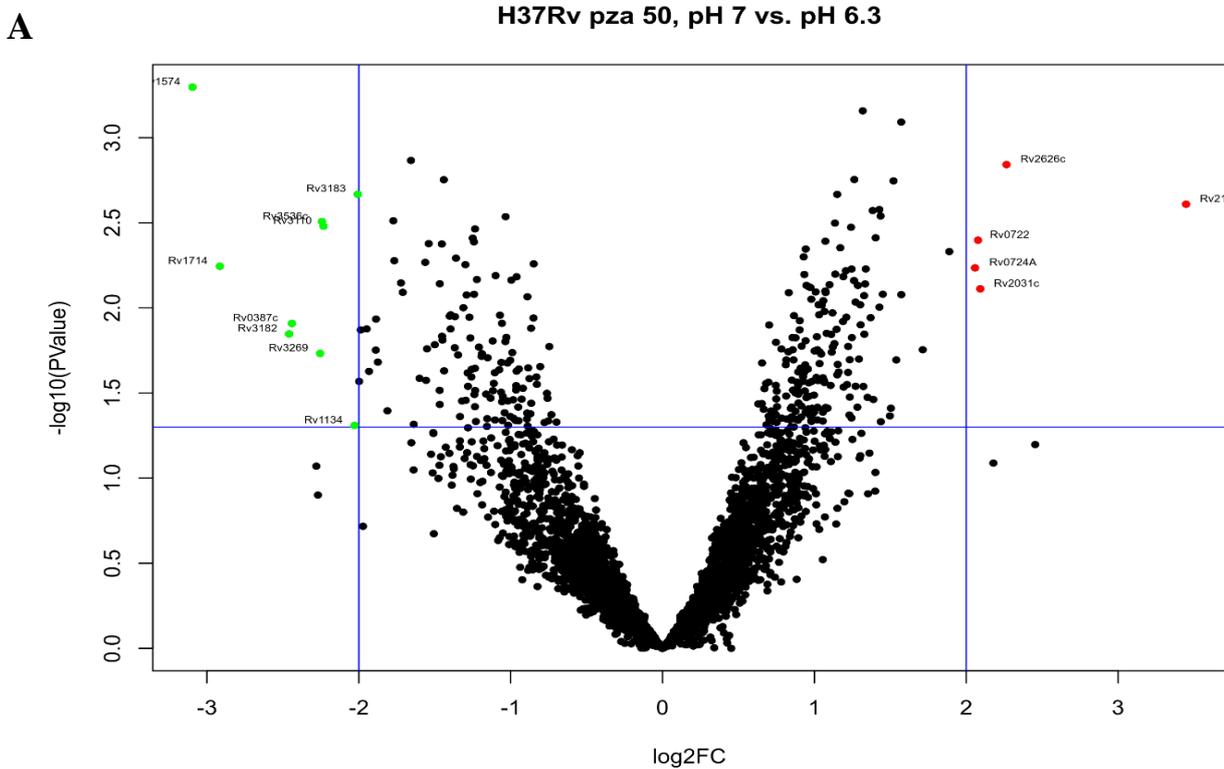
| Genes | Descripción | Expresión | LogFC | FC | P-Value |
|---------|--------------------------------|-----------|--------|-------|---------|
| Rv1714 | Uncharacterized oxidoreductase | DR | -2.094 | 0.234 | 0.025 |
| Rv2160c | Conserved hypothetical protein | UR | 2.544 | 5.833 | 0.020 |
| Rv2401A | Conserved membrane protein | DR | -3.296 | 0.101 | 0.011 |
| Rv2654c | Antitoxin | UR | 2.802 | 6.976 | 0.007 |

Figura 11. A) Volcano-plot LogFc vs $-\text{Log}_{10}(\text{p-value})$ de la comparación de la condición de 0 $\mu\text{g/ml}$ de PZA a pH 7.0 vs pH 6.3. en *M. tuberculosis* H37Rv. Los puntos verdes indican subexpresión y los rojos sobreexpresión. B) Cuadro de Genes de Expresión diferencial con p-value, logFC significativos de la cepa de *M.*

tuberculosis H37Rv relacionados al pH. FC: Fold change, CPM: counts per million, DR: Down-regulated; UR: Up-regulated.

6.5.4. Comparación entre pH 7.0 y pH 6.3 ambos con 50 µg/ml de PZA

Catorce genes se expresaron diferencialmente, 5 de ellos se sobreexpresaron (Rv2160c, Rv2626c, Rv2031c, Rv0722, Rv0724A) y 9 subexpresados (Rv3183, Rv1134, Rv3110, Rv3536c, Rv3269, Rv0387c, Rv3182, Rv1714, Rv1574) (Figura 10).



B

| Genes | Descripción | Expresión | LogFC | FC | P-Value |
|---------|---------------------------------------------|-----------|--------|--------|---------|
| Rv2160c | Conserved <i>hypothetical protein</i> | UR | 3.447 | 10.908 | 0.002 |
| Rv2626c | Hypoxic response protein 1 | UR | 2.265 | 4.806 | 0.001 |
| Rv2031c | Alpha-crystallin | UR | 2.092 | 4.265 | 0.008 |
| Rv0722 | 50S ribosomal protein L30 | UR | 2.077 | 4.220 | 0.004 |
| Rv0724A | Uncharacterized protein | UR | 2.058 | 4.163 | 0.006 |
| Rv3183 | Possible transcriptional regulatory protein | DR | -2.006 | 0.249 | 0.002 |
| Rv1134 | Hypothetical protein | DR | -2.028 | 0.245 | 0.049 |

| | | | | | |
|---------|------------------------------------------------|----|--------|-------|-------|
| Rv3110 | Pterin-4-alpha-carbinolamine dehydratase moaB1 | DR | -2.232 | 0.213 | 0.003 |
| Rv3536c | Probable hydratase | DR | -2.242 | 0.211 | 0.003 |
| Rv3269 | Conserved protein | DR | -2.254 | 0.210 | 0.019 |
| Rv0387c | PPE-SVP domain-containing protein | DR | -2.440 | 0.184 | 0.012 |
| Rv3182 | Conserved hypothetical protein | DR | -2.459 | 0.182 | 0.014 |
| Rv1714 | Uncharacterized oxidoreductase | DR | -2.915 | 0.133 | 0.006 |
| Rv1574 | Probable PhiRv1 phage related protein | DR | -3.095 | 0.117 | 0.001 |

Figura 10. A) Volcano-plot LogFc vs $-\text{Log}_{10}(\text{p-value})$ de la comparación de la condición de 50 $\mu\text{g/ml}$ de PZA a pH 7.0 vs pH 6.3. en *M. tuberculosis* H37Rv. Los puntos verdes indican subexpresión y los rojos sobreexpresión. B) Cuadro de Genes de Expresión diferencial con p-value, logFC significativos de la cepa de *M. tuberculosis* H37Rv relacionados al pH. FC: fold change, CPM: counts per million, DR: Down-regulated; UR: Up-regulated.

6.6. Selección de los genes diferencialmente expresados por la presencia de PZA y no por el cambio de pH

En la tabla 7 se observa que los genes diferencialmente expresados por efecto de PZA a pH de 6.3 (11 genes) y 7.0 (4 genes) son distintos. 1 gen expresado diferencialmente de cada una de estas comparaciones fue retirado debido a que éstos se expresaron como resultado del cambio exclusivo del pH (pH 6.3 vs pH 7 sin PZA y pH 6.3 vs pH 7 con PZA): Rv1714 y Rv2654c. Finalmente se seleccionaron 10 genes para el grupo de la comparación a pH 6.3 y 3 genes para el grupo de la comparación a pH 7.0.

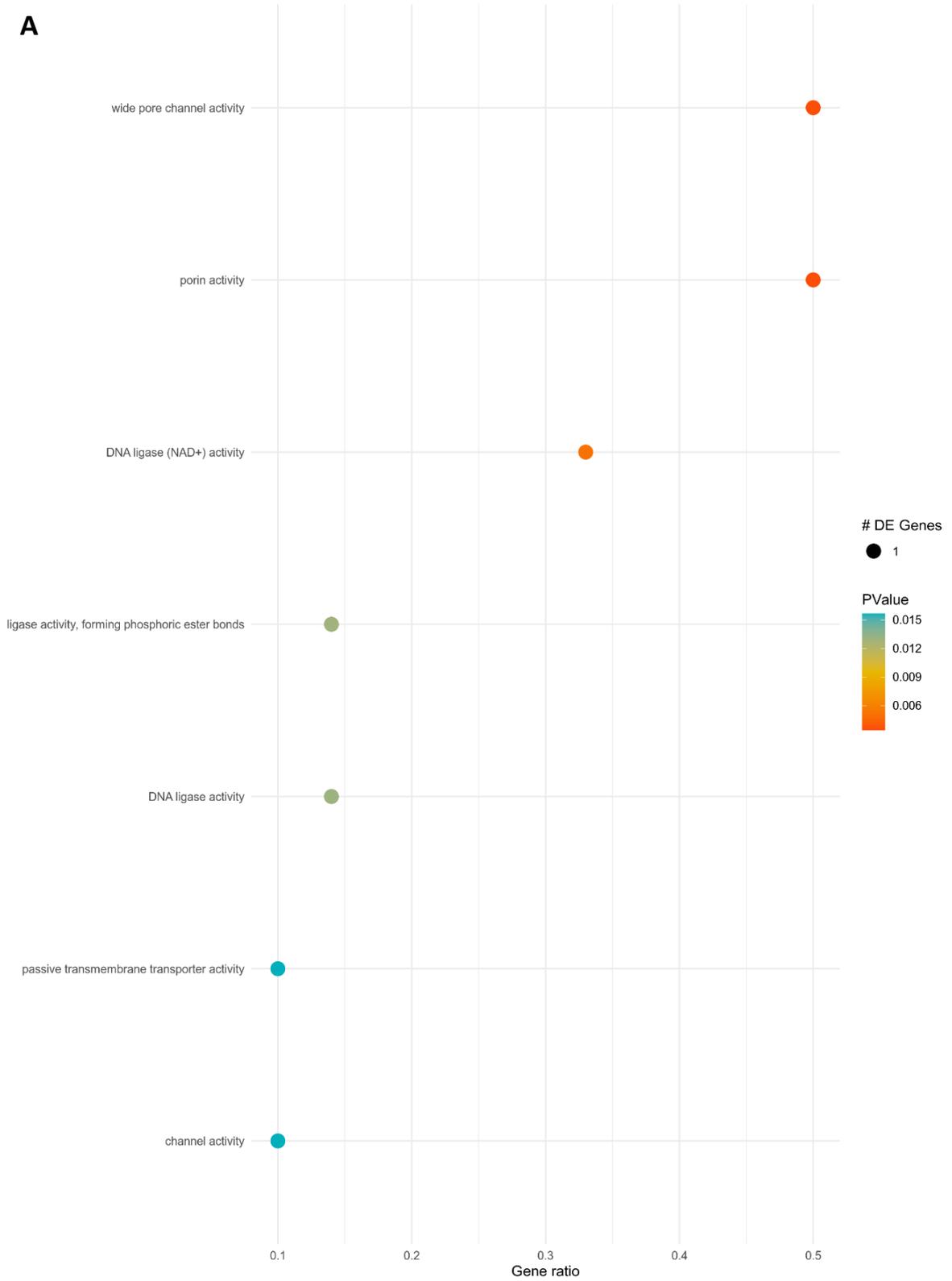
| A. Comparación pH 6.3, 0 vs 50 µg/ml PZA | B. Comparación pH 7.0, 0 vs 50 µg/ml PZA | C. Expresión de genes por pH |
|-----------------------------------------------------|-----------------------------------------------------|-----------------------------------------|
| Rv1574 | | - |
| Rv0387c | | - |
| Rv0601c | | - |
| Rv2274A | | - |
| Rv2275 | | - |
| Rv2626c | | - |
| Rv2635 | | - |
| Rv2651c | | - |
| Rv2913c | | - |
| Rv3113 | | - |
| | Rv2307D | - |
| | Rv3599c | - |
| | Rv2816c | - |
| | Rv2654c | Rv2654c |
| Rv1714 | | Rv1714 |
| | | Rv2160c |
| | | Rv2401A |

Tabla 7. Genes con expresión diferencial en las comparaciones entre 0 vs 50 µg/ml PZA a pH 6.3 (A) y pH 7.0 (B). La región sombreada representa los genes ED debido al pH (C), la posición de genes una misma fila representa la coincidencia entre comparaciones.

6.7. Rutas metabólicas afectadas por PZA

El análisis de las rutas metabólicas asociadas a los 10 genes seleccionados en la comparación a pH 6.3 entre 0 PZA vs 50 PZA mostró 7 funciones moleculares afectadas (Figura 11A) y a pH 7.0 entre 0 PZA vs 50 PZA mostró 12 funciones moleculares afectadas (Figura 11B) (Tabla 8). Dentro de los cuales un gen, forma independiente, afecta significativamente de una de las funciones.

A



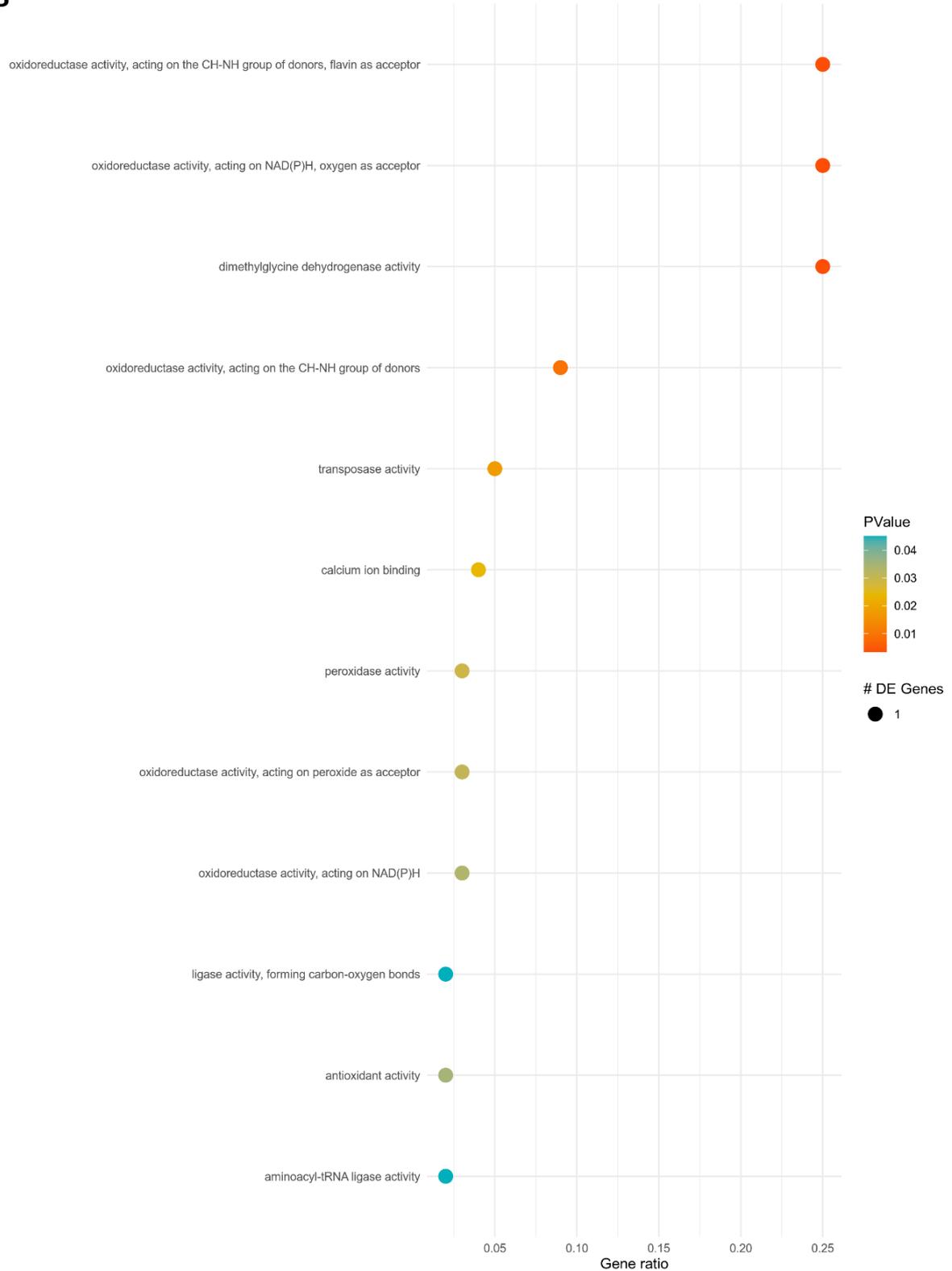
B

Figura 11. Funciones moleculares afectadas de los genes de expresión diferencial (ED) Análisis de enriquecimiento genético (GEA) por efecto de PZA a pH 6.3 (A) y pH 7.0 (B).

| Condiciones evaluadas | Número de genes ED | Rutas metabólicas afectadas |
|-----------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| pH 6.3 0 vs 50 µg/ml PZA | UR: 3, DR: 7 | Actividad porina. Actividad del canal de poro ancho. Actividad de ADN ligasa (NAD+). Actividad ligasa, formando enlaces éster fosfórico. Actividad de ADN ligasa. Actividad del canal. Actividad transportadora transmembrana pasiva. |
| pH 7.0 0 vs 50 µg/ml PZA | UR: 4 | Actividad oxidorreductasa, que actúa sobre el grupo de donantes CH-NH, flavina como aceptor. Actividad dimetilglicina deshidrogenasa. Actividad oxidorreductasa, que actúa sobre NAD (P) H, oxígeno como aceptor. Actividad oxidorreductasa, que actúa sobre el grupo de donantes CH-NH. Actividad transposasa. Unión de iones de calcio. Actividad peroxidasa. Actividad oxidorreductasa, que actúa sobre el peróxido como aceptor. Actividad oxidorreductasa, que actúa sobre NAD (P) H Actividad antioxidante. Actividad de aminoacil-tRNA ligasa. Actividad ligasa, formando enlaces carbono-oxígeno. |

Tabla 8. Principales rutas metabólicas afectadas por la presencia de PZA a pH 6.3

y pH 7.0. DR: Subexpresados; UR: Sobreexpresados; ED: Expresión diferencial.

7. DISCUSIÓN

En el presente estudio se identificaron los potenciales genes y/o rutas metabólicas involucrados en el mecanismo de acción de PZA mediante la comparación de los transcriptomas de *M. tuberculosis* obtenidos a través de RNA-seq bajo el efecto de PZA a pH ácido (pH 6.3) y neutro (pH 7). Dentro de ellos se identificó 10 genes en pH 6.3 y 3 genes en pH 7.0.

Uno de ellos, el gen Rv0387c, fue identificado como un pseudogen el cual tuvo una subexpresión 2 veces menor ($\text{LogFC} = -2.193$) a pH 6.3 por efecto de PZA. Este gen pertenece a la familia de proteínas PE/PPE caracterizadas por presentar cerca del extremo N-terminal regiones conservadas con secuencias de Pro-Glu (PE) y Pro-Pro-Glu (PPE) respetivamente, conformada por 107 proteínas PE y 69 PPE. Además, están clasificadas en diferentes familias y subfamilias debido a la presencia de sus diferentes dominios conservados (59). Dentro de estas familias se encuentra la familia multigénica conocida como PE/PPE, llamada así por su organización de formato bicistrónico con un gen PE seguido de un gen PPE (60). Por su conformación, estas varían mucho en tamaño, encontrándose desde péptidos pequeños con 28 aminoácidos (aa) a grandes proteínas de 588 aa, Rv0387c está conformada por 244 aa perteneciente a las proteínas PEE, también conocida como PPE9.

En la última década se ha visto el gran interés en estudiar esta familia de proteínas PE/PPE, ya que representa el 10% del total de genes en el genoma de *M. tuberculosis*. Además, son exclusivas de la familia de las micobacterias y producidas en mayor cantidad en aquellas micobacterias patogénas (10%) (61). La mayoría de estas proteínas se localizan en la superficie de la membrana externa

celular o se secretan. También desempeñan acción en la modulación de la patogenicidad y virulencia, regulación en el sistema inmune innato del huésped, inhibición de la formación de los fagolisosomas en los macrófagos y fuente principal de variación antigénica (62, 60).

El traslado de estas proteínas para su secreción y/o ubicación en la membrana externa de las micobacterias se realiza mediante un sistema especializado de secreción de tipo VII o también llamada ESX, el cual comprende 5 tipos de proteínas: ESX-1, ESX-2, ESX-3, ESX-4 y ESX-5. ESX5 se encuentra exclusivamente en la membrana interna de las micobacterias de crecimiento lento como *M. tuberculosis* (61, 63). La interacción entre las proteínas PE/PPE y las proteínas de secreción es esencial en el crecimiento y supervivencia de *M. tuberculosis* (64) porque permite regular los efectos adversos o favorables en el estado latente o activa; captando de forma eficiente nutrientes esenciales, liberando factores de virulencia o ejerciendo resistencia a medicamentos.

También se sabe que algunas proteínas de la familia de PE/PPE actúan de forma independiente y otras en conjunto. PPE38, en forma independiente, participa en la secreción mediada por el sistema ESX-5 (65). LipY, participa en la degradación de triacilglicéridos (TAG), siendo la lipasa principal activada en ausencia de nutrientes. LipY es traslocada a la membrana externa de la bacteria mediante el sistema de secreción ESX-5; su dominio PE actúa como una señal permitiendo su transporte y luego de su remoción (61).

Dentro de nuestros resultados no se observó ningún tipo de variación en la expresión de los genes que participan o están involucrados en la formación de este tipo de sistema de secreción ESX-5, por lo que la resistencia y/o supervivencia

estaría relacionada a Rv0387c y no de algún otro gen intermediario que participaría en su transporte. La subexpresión de Rv0387c bajo estrés (presencia de PZA) podría estar relacionada con la disminución de la liberación de moléculas al medio externo, la disminución del número de células bacterianas y la disminución de las funciones metabólicas. Estas condiciones le permitirían entrar a *M. tuberculosis* al estado de latencia y sobrevivir en situaciones extremas.

Otra función importante de las proteínas PPEs, es la relacionada a la resistencia a fármacos. PPE51 estaría involucrada en la resistencia a un compuesto bactericida candidato a fármaco, 3,3-bis-di (metilsulfonil) propionamida (3bMP1) debido a que una mutación simple en PPE51 generó resistencia a este fármaco. PPE51 es translocada por el sistema de secreción ESX-5 y una proteína de la familia PE19 (66) a través de una señal de secreción conservada YxxD/E en extremo C-terminal, la que estaría involucrada en la interacción con otra región conservada de PPE llamada WxG (67). Además, PE19+PPE51 se ensamblarían heterodiméricamente, formando una porina permitiendo el pase de moléculas tanto de exportación e importación (66). Por lo tanto, cabe la posibilidad de que Rv0387c podría comportarse de la misma forma respecto a la regulación de ingreso y salida de PZA o de su especie activa POA en *M. tuberculosis*, por medio de la formación de una porina, donde su subexpresión evitaría la formación de esta supuesta porina disminuyendo considerablemente la acción de PZA. Además, dentro de nuestros resultados de rutas metabólicas afectadas, observamos alteraciones en las rutas como actividad porina, actividad de canal de poro ancho y actividad transportadora transmembrana pasiva (Tabla 8), reforzando nuestra hipótesis de la posible función de Rv0387c en *M. tuberculosis* bajo efectos de PZA.

Una de las características de la familia de las micobacterias es la presencia de una capa en la membrana externa lipídica gruesa conformada por ácido micólicos; glicolípidos como lipooligosacáridos (LOS), micolatos de trehalosa y dimicocerosatos de tiocerol (PDIMs), generando una barrera altamente impermeable e hidrófoba (63), donde PDIM estaría constituyendo el 46% de la barrera lipídica en *M. tuberculosis*. En el mismo estudio de PP51 pudieron observar que PDIM estaría también participando en la resistencia a este fármaco 3bMP1, ya que una mutación en un solo nucleótido del gen *fadD26*, que codifica una coenzima-A implicada en la biosíntesis de PDIM, en una cepa con el gen *ppe51* en *knockout*, se observó una menor resistencia en comparación la cepa con el gen *PPE51* mutante y *PDMI* con síntesis normalizada. Esta alteración en PDIM estaría permitiendo el pase a moléculas hidrofóbicas e hidrofílicas (63; 64 ;66).

Considerando estos estudios podríamos reforzar que la acción de la resistencia a PZA no estaría solo involucrada a una mutación específica a un solo gen, sino que las alteraciones fisiológicas en *M. tuberculosis* también estarían jugando un papel importante en la total resistencia. Además, otra proteína de la familia PE, PE11, regulada positivamente bajo estrés: pH ácido, inanición y latencia dentro del macrófago altera la composición lipídica de la membrana externa de *M. tuberculosis*. La hidrofobicidad de *M. smegmatis* con la proteína recombinante PE11 aumentó debido al incremento en la resistencia a múltiples antibióticos: Rifampicina, isoniazida, etambutol, vancomicina y ampicilina (68). También proteínas de la familia PE, PE_PGRS, fueron reguladas positivamente frente a kanamicina (69).

En un estudio de las alteraciones de expresión y regulación de las proteínas PE/PPE, se observó que 128 de los 169 genes PE/PEE variaron considerablemente frente a diferentes condiciones: En el macrófago, en presencia de DETA/NO, H₂O₂, hipoxia, diamida, etambutol, lisosoma, a 45°C y entre otros. Dentro de los 41 genes que no se observó algún tipo de variación, se encontró a Rv0387c (62). Esto nos estaría dando posibles indicios, conforme a nuestros resultados obtenidos, que la disminución de expresión de Rv0387c podría deberse a la presencia exclusiva de PZA, reforzando su importancia en la resistencia y supervivencia de *M. tuberculosis* frente a este prodroga antituberculosa

Teniendo en cuenta que la familia de proteínas PE/PPE juegan un rol importante en a la supervivencia de *M. tuberculosis* en múltiples condiciones, siendo regulada de diferentes formas, cabe atribuirle un factor más, el cual contribuiría en su virulencia. Dentro de miles de genomas analizados de *M. tuberculosis* de aislados de pacientes clínicos, se obtuvo polimorfismos homoplásticos de nucleótido único (SNP) presentes en todas las categorías funcionales de genes y en mayor proporción dentro de la familia PE/PEE, donde estas mutaciones se asociaron fuertemente con la resistencia a antibióticos y estos SNP estarían atribuyéndole la alta diversidad funcional de *M. tuberculosis* que tiene con su respectivo huésped evidenciando un mecanismo de coevolución (70). Por lo tanto, se necesitaría realizar otros estudios que permitiría caracterizar cuál sería su función específica de Rv0387c y su posible implicancia en la resistencia de PZA y posiblemente extrapolar para los diferentes linajes de complejo *M. tuberculosis* (MTBC), como L5 y L6 de *M. africanum* las que causan la tuberculosis humana en África occidental (71).

Otro gen de ED obtenido en esta comparación a pH 6.3 fue Rv2275, que es un ciclodipéptido sintasa, que actúa como sustrato del citocromo P450 tipo 121A1(CYP121A1) para la formación del dicetopiperazina ciclo(L-tyr-L-tyr) permitiendo la formación de las micociclosina, que es un metabolito secundario de *M. tuberculosis* (76). Se reporta que CYP121A1 cumple la función esencial en el crecimiento de *M. tuberculosis* y posiblemente participaría en las funciones como catabolizar el colesterol como fuente de carbono durante la infección crónica, el establecimiento de la infección, la susceptibilidad a los fármacos y la persistencia (77). Además, se ha podido demostrar que múltiples azoles con actividad antimicrobiana y antituberculosa se unirían con gran afinidad a CYP121A1 (78). Por lo tanto, la sobreexpresión del gen Rv2275 hallado dentro de nuestros resultados, podría estar indicando una posible competencia de unión a CYP121A1 contra el PZA presente. En donde una mayor cantidad de ciclodipéptido sintasa posiblemente interaccionaría primero con CYP121A1, a comparación de las concentraciones circundantes de PZA, mitigando sus efectos antituberculosos y manteniendo la viabilidad de la cepa micobacteriana.

Otro gen sobreexpresado fue Rv2274A, (ligeramente superior a Rv2275) quien codifica la antitoxina MazE8, antagonista de la toxina MazF8, pertenece a la familia MazEF de Toxina-antitoxina (TA) tipo II en *M. tuberculosis*. La relevancia de esta familia de proteínas recae en su participación en la activación de diferentes fenotipos de patogenicidad, tolerancia y persistencia, y por lo tanto supervivencia de *M. tuberculosis* ante diferentes factores de estrés en su etapa de desarrollo (79). No obstante, no hay información específica de la posible función de MazE8 y su target MazF8 ante un estrés específico en *M. tuberculosis*, pero sí de sus familiares.

Tiwari P. *et al.* reporta que MazF3, MazF6 y MazF9, inducirían el estado bacteriostático de sus cepas micobacterianas bajo efecto estrés de diferentes medicamentos, donde estas ribonucleasas se sobreexpresaron permitiendo una signante tolerancia a las drogas usadas (80). Mismos resultados obtuvo Wei L. *et al* evaluando las mismas toxinas, bajo drogas antituberculosos mediante qRT-PCR (81). Pero respecto a MazF8, en sus resultados no se obtuvo inhibición del crecimiento de sus cepas, además que sus similares de MazE8 se subexpresaron, que resulta contrario a lo obtenido en nuestro estudio; esto puede deberse a que los factores de estrés son completamente diferentes a lo reportado en nuestro trabajo, deduciendo que esta sobreexpresión de Rv2274A, que conlleva a la degradación de MazF8, estaría actuando como una nueva vía de resistencia a PZA, permitiendo la supervivencia de *M. tuberculosis*, pero que es necesario estudios complementario en la identificación funcional de este complejo TA para la deducción precisa de un posible resistencia.

A pesar de la obtención de otros genes de ED hallados en esta condición acida (pH6.3), la falta de información previa para cada uno de ellos es carente, pero a raíz de los datos obtenidos en nuestro trabajo permitiría a futuras investigaciones estudiar estos genes en relación al efecto de resistencia a PZA en *M. tuberculosis* y posiblemente su importancia como posibles targets para la elaboración de futuros medicamentos.

Dentro de la comparación sin PZA (0 µg/ml) vs con PZA (50 µg/ml) a un pH 7.0, se obtuvo 3 genes de ED. Sabemos que el efecto de PZA, en cualquier concentración, a condiciones alcalinas se ve inhibida y que en condiciones menos acidas el efecto bactericida de PZA disminuye. Pero a pesar del uso en este trabajo

de un pH neutro, donde se observó que el crecimiento de *M. tuberculosis* fue afectado en presencia de PZA, evitando un crecimiento no mayor al 50% en comparación a su condición de cultivo sin PZA (Figura 4. D), se observó una diferenciación genética, no tan numerosa comparado a la condición de cultivo a pH ácido. De los 3 genes, Rv3599c y Rv2307D son proteínas Hipotéticas y Rv2816c es endorribonucleasa Cas2 asociada a CRISPR, donde los 3 se sobreexpresaron.

Rv3599c no presenta mayor información además de ser solo una proteína hipotética a diferencia del gen, Rv2307B. Esta proteína, también hipotética, está relacionada a una posible isla genómica en *M. tuberculosis* (73). Estas islas genómicas son fragmentos de ADN exógeno, captadas por transferencia horizontal de genes de otros organismos que pueden providenciar ciertos beneficios asociados a síntesis de metabolitos secundarios, resistencia a antibióticos y mayor patogenicidad (74). Esto indicaría que habría una posible relación de resistencia a PZA por parte del gen Rv2307B, pero es necesario la identificar certera del gen y realizar otras pruebas para la determinación de su funcionabilidad y posible participación.

Por otro lado, Rv2816c, gen que codifica la proteína Cas2, cumple la función de inmunidad contra fagos y plásmidos. Pero además Qinqin H. *et al.* descubrieron que la sobreexpresión de Cas2 bajo diferentes factores de estrés como condiciones ácidas, SDS y estrés oxidativo, cambio la morfología y capacidad de supervivencia de una cepa de *M. smegmatis* recombinada con Cas2 (75). Esta información constataría con la sobreexpresión de Rv2816c obtenida en nuestros resultados bajo el estrés de PZA, indicando una posible relación de resistencia bajo modificaciones estructurales de *M. tuberculosis*, disminuyendo un posible ingreso del medicamento u otros factores que tendrían que ser evaluados en futuros estudios.

En este estudio, el pH fue un factor de estrés usado; sin embargo, PZA es activa ante varios factores de estrés. Para conocer el efecto de PZA en *M. tuberculosis* bajo otros factores estrés principalmente los que se encuentran en el cuerpo humano es importante realizar más estudios, permitiendo ampliar la posibilidad de encontrar otros genes que podrían estar participando en a la alta resistencia a PZA o constatar los datos obtenidos hasta ahora por medio de nuestros resultados.

Además, se consiguió ARN de *M. tuberculosis* de alta calidad minimizando los posibles errores que se puedan obtener al manipular este material genético tan delicado, de igual forma se consiguió estandarizar el uso del secuenciador Next seq 550 para análisis transcriptómico en *M. tuberculosis*, siendo unos de los primeros a en todo el Perú, permitiendo abrir paso a futuras investigaciones siguiendo la base de nuestro estudio en el descubrimiento de genes implicados a la resistencia de medicamentos no solo de *M. tuberculosis* sino de otros microorganismos de importancia de la salud pública.

8. CONCLUSIONES

- Las condiciones de crecimiento de *M. tuberculosis* seleccionadas de PZA a pH 6.3 y 7.0 con 50 ug/ml logró obtener aproximadamente el 50% del crecimiento comparado al medio sin PZA con ODs menores de 1 (pH 6.3) y mayores de 1 (pH 7.0).
- La metodología desarrollada de extracción ARN permitió obtener ARN de buena calidad (integridad y pureza) para la secuenciación y análisis transcriptómico bajo las condiciones de crecimiento en estrés seleccionadas.
- Trece genes con expresión diferencial relacionadas con las siguientes rutas metabólicas: Actividad porina, transporte transmembrana pasivo y de oxidorreductasa, donde Rv2275, Rv2274A, Rv2307B, Rv2816c y Rv0387c son los genes con mayor información científica al presente, que podrían tener relación con la alta resistencia a PZA.
- Se identificó el gen Rv0387c, perteneciente a la familia de proteínas llamadas PE/PPE, el cual posiblemente tenga una implicancia significativa en la evasión del efecto de la PZA actuando como una proteína transportadora u tipo de porina.
- Rv2275, Rv2274A, Rv2307B, Rv2816c posiblemente actuarían como potenciales genes relacionados a la resistencia de PZA bajo las funciones de competidor de PZA, regulador de toxina, isla genómica y modificador fenotípico, respectivamente, en *M. tuberculosis*.

8.1. Recomendaciones

- Se recomienda un análisis específico de los genes Rv0387, Rv2275, Rv2274A, Rv2307B, Rv2816c, por medio de su bloqueo o silenciamiento usando las técnicas de knockout o Crispr Cas9, que permita determinar su implicancia en la resistí a PZA.
- Se recomienda el uso de PCR cuantitativa (qPCR) para explicar de forma cuantitativa la expresión diferencial de cada uno de os genes hallados.
- Observar cuales serían sus alteraciones de expresión del gen Rv0387 con PZA y bajo otras condiciones de estrés que permita simular más su ambiente *in vivo*.

9. ASPECTOS ÉTICOS

El presente estudio fue llevado a cabo como parte del proyecto “Efecto de la pirazinamida en el transcriptoma de *Mycobacterium tuberculosis*: Búsqueda de nuevo blanco terapéuticos” financiado por PROCENCIA - CONCYTEC en el marco de la convocatoria Proyecto Investigación Básica, 2018-01 (Contrato: 113-2018). El presente proyecto está registrado con el Código SIDISI 104497 con constancia de aprobación del Comité de Ética CAREG-ORVEI-125-19.

10. BIBLIOGRAFÍA

1. World Health Organization. (2018). Global tuberculosis Report.
2. World Health Organization .(2018). Tuberculosis. Recuperado <http://www.who.int/en/news-room/fact-sheets/detail/tuberculosis>
3. Ministerio de Salud. ACTUALIZACIÓN DEL SUB NUMERAL 7. TRATAMIENTO DE LA TUBERCULOSIS DE LA NTS N° 041-MINSA/DGSP-V.01 “NORMA TÉCNICA DE SALUD PARA EL CONTROL DE LA TUBERCULOSIS”. R.M. N° 383-2006/MINSA.
4. Zhang, Y., Shi, W., Zhang, W., & Mitchison, D. (2013). Mechanisms of Pyrazinamide Action and Resistance. *Microbiol Spectrum*. Vol 2(4), pág 1–12.
5. BRITISH THORACIC ASSOCIATION. (1982). A Controlled Trial of Six Months Chemotherapy in Pulmonary Tuberculosis. Second Report: Results during the 24 Months after the End of Chemotherapy. *Am Rev Respir Dis.*; Vol. 126(3), pág. 460-462.
6. Guarrant, R., Walker, D., Weller, P . (2011). Tuberculosis and Atypical Mycobacterial Infections. *Tropical Infectious Diseases: Principles, Pathogens and Practice*, Vol. 3, pág. 228-247.
7. Zhang, Y., Wade, M. M., Scorpio, A., Zhang, H., & Sun, Z . (2003). Mode of action of pyrazinamide: disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid. *Journal of Antimicrobial Chemotherapy*. Vol 53(5), pág. 790-795.
8. Zhang, Y., & Yew, W. (2009). Mechanisms of drug resistance in Mycobacterium tuberculosis. *The international journal of tuberculosis and*

lung disease : the official journal of the International Union against Tuberculosis and Lung Disease, Vol. 13(11), pág.1320–1330.

9. Yadon, A. N., Maharaj, K., Adamson, J. H., Lai, Y. P., Sacchettini, J. C., Ioerger, T. R., Rubin, E. J., & Pym, A. S . (2017). A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. *Nature communications*, Vol.8 (1), pág. 588.
10. Kahbazi, M., Sarmadian, H., Ahmadi, A., Didgar, F., Sadrnia, M., Poolad, T., & Arjomandzadegan, M . (2018). Novel Mutations in pncA Gene of Pyrazinamide Resistant Clinical Isolates of *Mycobacterium tuberculosis*. *Scientia pharmaceutica*. Vol.86 (2), pág. 15.
11. CRyPTIC Consortium and the 100,000 Genomes Project, Allix-Béguet, C., Arandjelovic, I., Bi, L., Beckert, P., Bonnet, M., Bradley, P., Cabibbe, A. M., Cancino-Muñoz, I., Caulfield, M. J., Chaiprasert, A., Cirillo, D. M., Clifton, D. A., Comas, I., Crook, D. W., De Filippo, M. R., de Neeling, H., Diel, R., Drobniewski, F. A., Faksri, K., ... Zhu, B. (2018). Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *The New England journal of medicine*, Vol. 379(15), pág. 1403-1415.
12. Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J. S., Barry, C. E., 3rd, Wang, H., Zhang, W., & Zhang, Y. (2011). Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science (New York, N.Y.)*, Vol.333 (6049), pág. 1630–1632.
13. Yang, J., Liu, Y., Bi, J., Cai, Q., Liao, X., Li, W., Guo, C., Zhang, Q., Lin, T., Zhao, Y., Wang, H., Liu, J., Zhang, X., & Lin, D . (2015). Structural basis for

- targeting the ribosomal protein S1 of *Mycobacterium tuberculosis* by pyrazinamide. *Molecular Microbiology*, Vol. 95(5), pág. 791–803.
14. Chakraborty, S., & Rhee, K. Y . (2015). Tuberculosis Drug Development: History and Evolution of the Mechanism-Based Paradigm. *Cold Spring Harbor perspectives in medicine*, Vol 5(8): a021147.
 15. Zhang, S., Chen, J., Shi, W., Liu, W., Zhang, W., & Zhang, Y . (2013). Mutations in *panD* encoding aspartate decarboxylase are associated with pyrazinamide resistance in *Mycobacterium tuberculosis*. *Emerging microbes & infections*, Vol.2 (6), pág. 34.
 16. Anthony, R. M., den Hertog, A. L., & van Soolingen, D. (2018). 'Happy the man, who, studying nature's laws, Thro' known effects can trace the secret cause.' Do we have enough pieces to solve the pyrazinamide puzzle?, *Journal of Antimicrobial Chemotherapy*, Vol 73(7), pág. 1750–1754.
 17. Sheen, P., Requena, D., Gushiken, E., Gilman, R. H., Antiparra, R., Lucero, B., Lizárraga, P., Cieza, B., Roncal, E., Grandjean, L., Pain, A., McNerney, R., Clark, T. G., Moore, D., & Zimic, M . (2017). A multiple genome analysis of *Mycobacterium tuberculosis* reveals specific novel genes and mutations associated with pyrazinamide resistance. *BMC Genomics*, Vol.18(1), pág. 769.
 18. de Welzen, L., Eldholm, V., Maharaj, K., Manson, A. L., Earl, A. M., & Pym, A. S . (2017). Whole-transcriptome and -genome analysis of extensively drug-resistant *Mycobacterium tuberculosis* clinical isolates identifies downregulation of *ethA* as a mechanism of ethionamide resistance. *Antimicrob Agents Chemother*. Vol. 61(12), pág. 1461-17.

19. BDATA BIOI. *Mycobacterium tuberculosis*, Instituto Nacional de Seguridad e Higiene en el Trabajo (INSHT), Available online at: <https://www.insst.es/documents/94886/353495/Mycobacterium+tuberculosis.pdf/9d648564-f9f0-41ff-9ae9-6c7af638bb52?version=1.0&t=1528734433284>
20. Joklik, Wolfgang K., Willet, Hilda P., Amos, Bernard, Wilfert, Catherine M. (1994). ZINSSER MICROBIOLOGIA. 20ª EDICION.
21. González, M., J. (2014). Microbiología de la tuberculosis. Seminarios de la Fundación Española de Reumatología, Vol 15(1), pág. 25-33.
22. Madigan, T., Martinko, J., M, Parker, J. (2004). BROCK- BIOLOGIA DE LOS MICROORGANISMOS. 10ª EDICION.
23. Túniz Bastidaa, MR. García Ramosa, ML. Pérez del Molinoa, FL. Lado Lado. (2002). Epidemiología de la tuberculosis. Medicia Integral. Vol. 39(5), pág. 171-239.
24. Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. Cold Spring Harbor perspectives in medicine, Vol. 5(8), a021113.
25. Nataraj, V., Varela, C., Javid, A., Singh, A., Besra, G. S., & Bhatt, A. (2015). Mycolic acids: deciphering and targeting the Achilles' heel of the tubercle bacillus. Molecular microbiology, Vol 98(1), pág. 7–16.
26. Maitra, A., Munshi, T., Healy, J., Martin, L. T., Vollmer, W., Keep, N. H., & Bhakta, S. (2019). Cell wall peptidoglycan in *Mycobacterium tuberculosis*: An Achilles' heel for the TB-causing pathogen. FEMS microbiology reviews, Vol 43(5), pág. 548–575.

27. Ates, L. S., Dippenaar, A., Sayes, F., Pawlik, A., Bouchier, C., Ma, L., Warren, R. M., Sougakoff, W., Majlessi, L., van Heijst, J., Brossier, F., & Brosch, R. (2018). Unexpected Genomic and Phenotypic Diversity of *Mycobacterium africanum* Lineage 5 Affects Drug Resistance, Protein Secretion, and Immunogenicity. *Genome biology and evolution*, Vol. 10(8), pág.1858–1874.
28. Hewlett, D., Jr, Horn, D. L., & Alfalla, C. (1995). Drug-resistant tuberculosis: inconsistent results of pyrazinamide susceptibility testing. *JAMA*, 273(12), 916–917.
29. Zhang, Y., & Mitchison, D. (2003). The curious characteristics of pyrazinamide: a review. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease*, Vol.7(1), pág.6–21.
30. Zhang, Y., Permar, S., & Sun, Z. (2002). Conditions that may affect the results of susceptibility testing of *Mycobacterium tuberculosis* to pyrazinamide. *Journal of medical microbiology*, Vol. 51(1), pág.42–49.
31. Heifets, L. B., & Lindholm-Levy, P. J. (1990). Is pyrazinamide bactericidal against *Mycobacterium tuberculosis*?. *The American review of respiratory disease*, Vol.141(1), pág.250–252.
32. Ellard G. A. (1969). Absorption, metabolism and excretion of pyrazinamide in man. *Tubercle*, Vol. 50(2), pág.144–158.
33. McDERMOTT, W., & TOMPSETT, R. (1954). Activation of pyrazinamide and nicotinamide in acidic environments in vitro. *American review of tuberculosis*, Vol. 70(4), pág.748–754.

34. Heifets, L., & Lindholm-Levy, P. (1992). Pyrazinamide sterilizing activity in vitro against semidormant *Mycobacterium tuberculosis* bacterial populations. *The American review of respiratory disease*, Vol.145(5), pág.1223–1225.
35. Seong Woon Roh, Guy C.J. Abell, Kyoung-Ho Kim, Young-Do Nam, Jin-Woo Bae. (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in Biotechnology*, Vol.28(6), pág.291-299.
36. Yang, X., Kui, L., Tang, M., Li, D., Wei, K., Chen, W., Miao, J., & Dong, Y. (2020). High-Throughput Transcriptome Profiling in Drug and Biomarker Discovery. *Frontiers in genetics*, Vol.11(19).
37. Baird DA, Liu JZ, Zheng J, Sieberts SK, Perumal T, et al. (2021). Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLOS Genetics*, Vol 17(1).
38. NCI. (2020). Analyzing Tumor RNA May Help Match Patients with Most Effective Cancer Treatments. Available online at: <https://www.cancer.gov/news-events/cancer-currents-blog/2020/tumor-rna-cancer-precision-medicine>
39. Pabon, N. A., Xia, Y., Estabrooks, S. K., Ye, Z., Herbrand, A. K., Süß, E., Biondi, R. M., Assimon, V. A., Gestwicki, J. E., Brodsky, J. L., Camacho, C. J., & Bar-Joseph, Z. (2018). Predicting protein targets for drug-like compounds using transcriptomics. *PLoS computational biology*, Vol.14(12), e1006651.
40. Kuloyo, O., Fourie, R., Cason, E., Albertyn, J., & Pohl, C. H. (2020). Transcriptome Analyses of *Candida albicans* Biofilms, Exposed to

Arachidonic Acid and Fluconazole, Indicates Potential Drug Targets. G3 (Bethesda, Md.), Vol.10(9), pág.3099–3108.

41. Khaledi, A., Schniederjans, M., Pohl, S., Rainer, R., Bodenhofer, U., Xia, B., Klawonn, F., Bruchmann, S., Preusse, M., Eckweiler, D., Dötsch, A., & Häussler, S. (2016). Transcriptome Profiling of Antimicrobial Resistance in *Pseudomonas aeruginosa*. *Antimicrobial agents and chemotherapy*, Vol. 60(8), pág.4722–4733.
42. Sawyer, E. B., Grabowska, A. D., & Cortes, T. (2018). Translational regulation in mycobacteria and its implications for pathogenicity. *Nucleic acids research*, Vol 46(14), pág 6950–6961.
43. Briffotiaux, J., Liu, S., & Gicquel, B. (2019). Genome-Wide Transcriptional Responses of *Mycobacterium* to Antibiotics. *Frontiers in microbiology*, Vol.10, pág.249.
44. Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., & Schoolnik, G. K. (1999). Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 96(22), pág.12833–12838.
45. Boshoff, H. I., Myers, T. G., Copp, B. R., McNeil, M. R., Wilson, M. A., & Barry, C. E., 3rd (2004). The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *The Journal of biological chemistry*, Vol. 279(38), pág 40174–40184.

46. Walter, N. D., Dolganov, G. M., Garcia, B. J., Worodria, W., Andama, A., Musisi, E., Ayakaka, I., Van, T. T., Voskuil, M. I., de Jong, B. C., Davidson, R. M., Fingerlin, T. E., Kechris, K., Palmer, C., Nahid, P., Daley, C. L., Geraci, M., Huang, L., Cattamanchi, A., Strong, M., Davis, J. L. (2015). Transcriptional Adaptation of Drug-tolerant *Mycobacterium tuberculosis* During Treatment of Human Tuberculosis. *The Journal of infectious diseases*, Vol. 212(6), pág.990–998.
47. Liang, J., Zeng, F., Guo, A., Liu, L., Guo, N., Li, L., Jin, J., Wu, X., Liu, M., Zhao, D., Li, Y., Jin, Q., & Yu, L. (2011). Microarray analysis of the chelerythrine-induced transcriptome of *Mycobacterium tuberculosis*. *Current microbiology*, Vol.62(4), pág. 1200–1208.
48. Xu, Z., Zhou, A., Wu, J., Zhou, A., Li, J., Zhang, S., Wu, W., Karakousis, P. C., & Yao, Y. F. (2018). Transcriptional Approach for Decoding the Mechanism of *rpoC* Compensatory Mutations for the Fitness Cost in Rifampicin-Resistant *Mycobacterium tuberculosis*. *Frontiers in microbiology*, Vol. 9, pág.2895.
49. De Knecht GJ, Bruning O, ten Kate MT, de Jong M, van Belkum A, Endtz HP, Breit TM, Bakker-Woudenberg IA, de Steenwinkel JE. (2013). Rifampicin-induced transcriptome response in rifampicin-resistant *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*. Vol 93(1), pág.96-101.
50. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W., & Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?. *BMC genomics*, Vol.13, pág.734.

51. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
52. Illumina.2011. Quality Scores for Next-Generation Sequencing. Available online at: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf
53. Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D. M., Cambell, L., Carbajo, M., Chakiachvili, M., Christensen, M., Cummins, C., Cuzick, A., Davis, P., Fexova, S., Gall, A., George, N., Gil, L., Flicek, P. (2020). Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic acids research*, 48(D1), D689–D695.
54. Roberts, Adam, Lior Pachter. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*, Vol 10(1), pag. 71-73.
55. Robinson, M.D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, Vol 11(25).
56. Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. Vol. 26(1), pág:139-40.
57. Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, Vol 11(2), pág.R14.

58. Li, Wentian.(2012). Volcano plots in analyzing differential expressions with mRNA microarrays. *Journal of bioinformatics and computational biology*, Vol 10(06).
59. Smith I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews*, Vol 16(3), pág 463–496.
60. Sakshi K., Yadvir S., Khushbu S., Aditya M., Nasreen Z., E., Seyed E., H . (2012).Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. *Nucleic Acids Research*, Vol. 40(15), pág. 7113–7122.
61. Daleke, M. H., Cascioferro, A., de Punder, K., Ummels, R., Abdallah, A. M., van der Wel, N., Peters, P. J., Luirink, J., Manganelli, R., & Bitter, W. (2011). Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. *The Journal of biological chemistry*, Vol. 286(21), pág.19024–19034.
62. Li, W., Deng, W., & Xie, J. (2019). Expression and regulatory networks of *Mycobacterium tuberculosis* PE/PPE family antigens. *Journal of cellular physiology*, Vol. 234(6), pág.7742–7751.
63. Ehtram, A., Shariq, M., Ali, S., Quadir, N., Sheikh, J. A., Ahmad, F., Sharma, T., Ehtesham, N. Z., & Hasnain, S. E. (2021). Teleological cooption of *Mycobacterium tuberculosis* PE/PPE proteins as porins: Role in molecular immigration and emigration. *International journal of medical microbiology : IJMM*, Vol. 311(3), 151495.

64. Ates, L. S., Ummels, R., Commandeur, S., van de Weerd, R., Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S. R., Abdallah, A. M., Abd El Ghany, M., Abdel-Haleem, A. M., Pain, A., Jiménez, C. R., Bitter, W., & Houben, E. N. (2015). Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLoS genetics*, Vol. 11(5), e1005190.
65. Ates, L. S., Dippenaar, A., Ummels, R., Piersma, S. R., van der Woude, A. D., van der Kuij, K., Le Chevalier, F., Mata-Espinosa, D., Barrios-Payán, J., Marquina-Castillo, B., Guapillo, C., Jiménez, C. R., Pain, A., Houben, E., Warren, R. M., Brosch, R., Hernández-Pando, R., & Bitter, W. (2018). Mutations in *ppe38* block PE_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nature microbiology*, Vol.3(2), pág.181–188.
66. Wang, Q., Boshoff, H., Harrison, J. R., Ray, P. C., Green, S. R., Wyatt, P. G., & Barry, C. E., 3rd (2020). PE/PPE proteins mediate nutrient transport across the outer membrane of *Mycobacterium tuberculosis*. *Science (New York, N.Y.)*, Vol.367(6482), pág1147–1151.
67. Louis, S., A. (2020). New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Molecular Microbiology*, Vol. 113(1), pág.4-21.
68. Singh, P., Rao, R.N., Reddy, J.R., Prasad, R., Kotturu, S.K., Ghosh, S., & Mukhopadhyay, S. (2016). PE11, a PE/PPE family protein of *Mycobacterium tuberculosis* is involved in cell wall remodeling and virulence. *Scientific Reports*, Vol. 6(21624).

69. Srivastava, V., Jain, A., Srivastava, B. S., & Srivastava, R. (2008). Selection of genes of *Mycobacterium tuberculosis* upregulated during residence in lungs of infected mice. *Tuberculosis (Edinburgh, Scotland)*, Vol. 88(3), pág.171–177.
70. Tantivitayakul, P., Ruangchai, W., Juthayothin, T., Smittipat, N., Disratthakit, A., Mahasirimongkol, S., Viratyosin, W., Tokunaga, K., & Palittapongarnpim, P. (2020). Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in *Mycobacterium tuberculosis*. *Scientific reports*, Vol. 10(1), pág. 8024.
71. Ates, L. S., Dippenaar, A., Sayes, F., Pawlik, A., Bouchier, C., Ma, L., Warren, R. M., Sougakoff, W., Majlessi, L., van Heijst, J., Brossier, F., & Brosch, R. (2018). Unexpected Genomic and Phenotypic Diversity of *Mycobacterium africanum* Lineage 5 Affects Drug Resistance, Protein Secretion, and Immunogenicity. *Genome biology and evolution*, Vol. 10(8), pág.1858–1874.
72. Salfinger, M., & Heifets, L. B. (1988). Determination of pyrazinamide MICs for *Mycobacterium tuberculosis* at different pHs by the radiometric method. *Antimicrobial agents and chemotherapy*, Vol.32(7), pág.1002–1004.
73. Becq, J., Gutierrez, M. C., Rosas-Magallanes, V., Rauzier, J., Gicquel, B., Neyrolles, O., & Deschavanne, P. (2007). Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Molecular biology and evolution*, Vol. 24(8), pág. 1861–1871.
74. Daniel, M. (2017). Uma nova abordagem para a identificação de ilhas genômicas em bactérias com base no método de agrupamento mean shift. Programa de Pós-Graduação em Informática, Universidade Federal da Paraíba.
<https://repositorio.ufpb.br/jspui/bitstream/tede/9040/2/arquivototal.pdf>

75. Huang, Q., Luo, H., Liu, M., Zeng, J., Abdalla, A. E., Duan, X., Li, Q., & Xie, J. (2016). The effect of Mycobacterium tuberculosis CRISPR-associated Cas2 (Rv2816c) on stress response genes expression, morphology and macrophage survival of Mycobacterium smegmatis. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, Vol. 40, pág. 295–301.
76. Zhu, X.; McAtee, C. C.; Schindler, C. S. (2018). Scalable Synthesis of Mycocyclosin. *Org. Lett.*, Vol. 20, pág. 2862–2866.
77. Ortega Ugalde, S., Wallraven, K., Speer, A., Bitter, W., Grossmann, T. N., & Commandeur, J. (2020). Acetylene containing cyclo(L-Tyr-L-Tyr)-analogs as mechanism-based inhibitors of CYP121A1 from Mycobacterium tuberculosis. *Biochemical pharmacology*, Vol. 177, pág. 113938.
78. Pascal B., Marie Le D., Alistair F., Olivier L., Mickaël J., Jean-Baptiste C., Alain L., Robert T., Marie C., Cédric M., Christophe D., Roger G., Jean-Luc P., Muriel G. (2009). Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences*, Vol. 106 (18) pág. 7426-7431.
79. Zaychikova, M. V., Zakharevich, N. V., Sagaidak, M. O., Bogolubova, N. A., Smirnova, T. G., Andreevskaya, S. N., Larionova, E. E., Alekseeva, M. G., Chernousova, L. N., & Danilenko, V. N. (2015). Mycobacterium tuberculosis Type II Toxin-Antitoxin Systems: Genetic Polymorphisms and Functional Properties and the Possibility of Their Use for Genotyping. *PLoS one*, 10(12), e0143682
80. Tiwari, P., Arora, G., Singh, M. et al. (2015). MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs. *Nat Commun*, Vol. 6, pág. 6059 (2015).
81. Liu Wei, Zhao Jili, Qu Yanlin, Xie Wanying; Yuan Li. (2017). Detection on expression levels of mazE F toxin-antitoxin system in Mycobacterium tuberculosis by qRT-PCR. *Chinese Journal of Zoological Diseases*, Vol. 33(2), pág. 143-147.

11. Anexo:

11.1 Comandos Bioinformáticos:

```
import subprocess
```

```
#Create table with list of genes from transcriptome. The gene count will list  
genes in the same order,
```

```
#including genes with count = 0
```

```
#First value (row) is header. Will contain name of sample
```

```
genes = []
```

```
gen_c = ["Genes"]
```

```
infile = open("h37rv_cds.fa","r")
```

```
for line in infile:
```

```
    if line[0] == ">":
```

```
        genes.append(line.split(" ")[0].replace(">",""))
```

```
        gen_c.append(line.split(" ")[3].replace("gene:", ""))
```

```
infile.close()
```

```
final_count = []
```

```
#Index transcriptome
```

```
subprocess.call("salmon index -t h37rv_cds.fa -i h37rv_index -p 8",  
shell=True)
```

```
#List of R1 fastq
```

```
R1 = []
```

```
infile = open("R1j.txt","r")
```

```
for line in infile:
```

```
    R1.append(line.rstrip())
```

```
infile.close()
```

```
#List of R2 fastq
```

```
R2 = []
```

```
infile = open("R2j.txt","r")
```

```

for line in infile:
    R2.append(line.rstrip())
infile.close()
#For each couple of reads (R1 and R2) execute salom
for i in range(len(R1)):
    read = R1[i].split("/")[1].split("_")[0] #Name of output folder
    subprocess.call("salmon quant -i h37rv_index/ -l IU -1 " + R1[i] + "
-2 " + R2[i] + " -p 8 --validateMappings -o quants/" + read, shell=True)

    gen_id = []
    counts = []

    #Open quantification file
    infile = open("quants/" + read + "/quant.sf", "r")
    infile.readline() #skip first line (header)
    for line in infile:
        line = line.split('\t')
        gen_id.append(line[0])
        counts.append(str(line[-1].rstrip()))
    infile.close()

    gen_c[0] = gen_c[0] + '\t' + read
    c = 1
    for gen in genes:
        try:
            position = gen_id.index(gen)
            gen_c[c] = gen_c[c] + '\t' + counts[position]

        except:
            gen_c[c] = gen_c[c] + '\t' + "0"
        c += 1

```

```

outfile = open("counts_F.tsv","w")
for gen in gen_c:
    outfile.write(gen+"\n")
outfile.close()

# Load data and set conditions
samplesinfo <- c("A-1;pH-7_PZA-0", "A-2;pH-7_PZA-0", "A-3;pH-
7_PZA-0",
                "B-1;pH-7_PZA-50", "B-2;pH-7_PZA-50", "B-3;pH-7_PZA-
50",
                "C-1;pH-6_PZA-0", "C-2;pH-6_PZA-0", "C-3;pH-6_PZA-0",
                "D-1;pH-6_PZA-50", "D-2;pH-6_PZA-50", "D-3;pH-6_PZA-
50",
                "alpha-1;pH-7_PZA-0", "alpha-2;pH-7_PZA-0", "alpha-3;pH-
7_PZA-0",
                "beta-1;pH-7_PZA-50", "beta-2;pH-7_PZA-50", "beta-3;pH-
7_PZA-50",
                "teta-1;pH-7_PZA-50", "teta-2;pH-6_PZA-0", "teta-3;pH-
6_PZA-0",
                "gama-1;pH-6_PZA-50", "gama-2;pH-6_PZA-50", "gama-
3;pH-6_PZA-50",
                "I-1;pH-7_PZA-0", "I-2;pH-7_PZA-0", "I-3;pH-7_PZA-0",
                "II-1;pH-7_PZA-50", "II-2;pH-7_PZA-50", "II-3;pH-7_PZA-
50",
                "III-1;pH-7_PZA-50", "III-2;pH-7_PZA-50", "III-3;pH-
7_PZA-50",
                "IV-1;pH-6_PZA-50", "IV-2;pH-6_PZA-50", "IV-3;pH-
6_PZA-50")

# Try to group by treat + sample

```

```
Treat2 <- factor(c("H37Rv_pH-7_PZA-0", "H37Rv_pH-7_PZA-0",
"H37Rv_pH- 7_PZA-0","H37Rv_pH-7_PZA-50", "H37Rv_pH-7_PZA-
50", "H37Rv_pH-7_PZA-50", "H37Rv_pH-6_PZA-0", "H37Rv_pH-
6_PZA-0", "H37Rv_pH-6_PZA-0", "H37Rv_pH-6_PZA-50",
"H37Rv_pH-6_PZA-50", "H37Rv_pH-6_PZA-50",
"CSV10399_pH-7_PZA-0","CSV10399_pH-7_PZA-0", "CSV10399_pH-
7_PZA-0", "CSV10399_pH-7_PZA-50","CSV10399_pH-7_PZA-50",
"CSV10399_pH-7_PZA-50", "CSV10399_pH-7_PZA-
50","CSV10399_pH-6_PZA-0", "CSV10399_pH-6_PZA-0",
"CSV10399_pH-6_PZA-50", "CSV10399_pH-6_PZA-50",
"CSV10399_pH-6_PZA-50", "Ln2552_pH-7_PZA-0", "Ln2552_pH-
7_PZA-0", "Ln2552_pH-7_PZA-0", "Ln2552_pH-7_PZA-50",
"Ln2552_pH-7_PZA-50", "Ln2552_pH-7_PZA-50", "Ln2552_pH-
6_PZA-0", "Ln2552_pH-6_PZA-0", "Ln2552_pH-6_PZA-0",
"Ln2552_pH-6_PZA-50", "Ln2552_pH-6_PZA-50", "Ln2552_pH-
6_PZA-50"))
```

```
samplesgroup <-c("H1", "H1", "H1",
"H2", "H2", "H2",
"H3", "H3", "H3",
"H4", "H4", "H4",
"C1", "C1", "C1",
"C2", "C2", "C2",
"C3", "C3", "C3",
"C4", "C4", "C4",
"L1","L1", "L1",
"L2","L2", "L2",
"L3","L3", "L3",
"L4","L4", "L4")
```

```

# Counts
counts_salmon = read.table("F_counts.tsv", sep="\t", header = TRUE,
row.names = 1, stringsAsFactors = F)

#fc.dge <- DGEList(counts=fc$counts, genes = fc$annotation, samples =
samplesinfo, group = samplesgroup)
fc.dge <- DGEList(counts = counts_salmon, genes =
row.names(counts_salmon),
                 samples = colnames(counts_salmon), group = samplesgroup)
# Filter for genes with low counts across conditions
keep <- rowSums(cpm(fc.dge) > 1) >= 2
fc.dge <- fc.dge[keep, , keep.lib.sizes=FALSE]

# Normalize for library compositional bias
# calcNormFactors normalizes for RNA composition by finding a set of
scaling factors for the library sizes
# that minimize the log-fold changes between the samples for most genes.
fc.dge.norm <- calcNormFactors(fc.dge)

# Compute log2 counts-per-million without an offset:
logCPM <- cpm(fc.dge.norm, log = TRUE)
logCPMwonorm <- cpm(fc.dge, log = TRUE)

# Correct batch effect by including batch in model
Batch <- factor(c("run01", "run01", "run02",
                 "run01", "run01", "run02",
                 "run01", "run01", "run02",
                 "run02", "run02", "run01",
                 "run02", "run02", "run01",
                 "run02", "run02", "run01",
                 "run02", "run02", "run01",

```

```
"run02", "run02", "run03",  
"run02", "run02", "run03",  
"run02", "run02", "run03",  
"run02", "run02", "run03"))
```

```
Design <- model.matrix(~0+Batch+Treat2)  
colnames(Design) <- c("batch01", "batch02", "batch03",  
"CSV10399_pH-6_PZA-50", "CSV10399_pH-7_PZA-0",  
"CSV10399_pH-7_PZA-50",  
"H37Rv_pH-6_PZA-0", "H37Rv_pH-6_PZA-50",  
"H37Rv_pH-7_PZA-0", "H37Rv_pH-7_PZA-50",  
"Ln2552_pH-6_PZA-0", "Ln2552_pH-6_PZA-50",  
"Ln2552_pH-7_PZA-0", "Ln2552_pH-7_PZA-50")
```

```
# Estimate dispersion
```

```
fc.dge.disp <- estimateDisp(fc.dge.norm, design = Design, robust = TRUE)  
fc.dge.disp$common.dispersion
```

```
# Fit model / test for DE genes
```

```
fit <- glmQLFit(fc.dge.disp, Design, robust = TRUE)  
png(filename = "Pictures/QLplot.png", width = 1600, height = 1200, res =  
300)  
plotQLDisp(fit)  
dev.off()
```

```
#Prepare objects for GOseq analysis
```

```
table_length <- read.table("genes_length.tsv", sep="\t",  
col.names=c("Gene", "Length"), stringsAsFactors = F)  
Size_genes <- table_length$Length  
names(Size_genes) <- table_length$Gene  
toremove <- c("Rv1714", "Rv2654c", "Rv2654c", "Rv1714", "Rv2160c",  
"Rv2401A")
```

```

# Variable GOTerm containing GO terms
GOT <- read.table("allGoTermsForMycobacterium_tuberculosis_H37Rv.txt", sep
= "\t", stringsAsFactors = F)
GOTnames <- read.table("names.tsv", sep = "\t", stringsAsFactors = F)
GOTnames <- GOTnames$V1
GOT <- GOT[-1,]
i <- 1
X <- c()
Y <- c()
Name <- c()
GOTerm <- list()
while (i < length(GOT$V1)) {
  X <- c(X, GOT$V1[i])
  Y <- c(Y, paste0("GO:", GOT$V2[i]))

  if (GOT$V1[i] != GOT$V1[i+1] & GOT$V4[i] == "biological_process")
  {
    # if (GOT$V1[i] != GOT$V1[i+1]) {
    W <- list("X" = Y)
    Name <- c(Name, GOT$V1[i+1])
    X <- c()
    Y <- c()
    GOTerm <- c(GOTerm, W)
    }
  i <- i + 1
}
names(GOTerm) <- Name
#DE Analysis
# Check if there was a genuine need to adjust for batches by testing for DE
between batches

```

```

# by testing for DE between all treatments.
qlf <- glmQLFTest(fit, coef = 2:14)
tt <- topTags(qlf)
FDR <- p.adjust(qlf$table$PValue, method="BH")
sFDR <- sum(FDR < 0.05)
sdecidetest <- summary(decideTests(qlf, adjust.method = "none", p.value =
0.05, lfc = 2))

# Print to txt useful info#
writeLines(c("Common Dispersion:", fc.dge.disp$common.dispersion,
"\n", "Number of genes with pval < 0.05 and |logFC| > 2 (test between all
treatments)",
          sFDR, "\n", "Significative genes:"), "salmon_DGE_results.txt")
write.table(sdecidetest, file = "salmon_DGE_results.txt", append = T)

# DE Testing by conditions (comparisons)#
conditions <- read_excel("comparisons_new.xlsx")

for (i in (1:16)){

  file_name <- paste0(i, "_", conditions[i,1])
  title_name <- as.character(conditions[i,2])

  n <- length(conditions[4,])
  x <- as.numeric(unlist(c(conditions[i,3:n])))

  ###DE Test
  if (conditions[i,3] != 1 & conditions[i,3] != 0){
    qlf <- glmQLFTest(fit, coef = as.numeric(conditions[i,3]))
  } else {
    qlf <- glmQLFTest(fit, contrast = x)
  }
}

```

```

qlfgenes <- rownames(qlf$table)
qlfx <- qlf$table$logFC
tt <- topTags(qlf, sort.by = "PValue", n = 10000)
res <- tt$table
res <- cbind(res, "FoldChange" = 2^(res$logFC))
res <- subset(res, select = c(1,2,7,3,4,5,6))
write.xlsx(res, paste0("Tables/",file_name, ".xlsx"),
           col.names=TRUE, row.names=FALSE, append=FALSE)
if (nrow(res[res$FDR < 0.05,]) > 0) {

write.xlsx(res[res$FDR<0.05,],paste0("Tables/FDR/",file_name, ".xlsx"))
}

cat("\n", file = "salmon_DGE_results.txt", append = TRUE)
write.table(summary(decideTests(qlf, adjust.method = "none", p.value =
0.05, lfc = 2)), file = "salmon_DGE_results.txt", append = T)

```

###Gene enrichment analysis (GEA) - GOseq (based on pValue)

```

#Table with genes (column 1: genes; column 2: 1(significant) or 0 (non-
significant))

```

```

# Old, doesn't work:

```

```

#

```

```

genes=as.integer(p.adjust(tested$table$PValue[tested$table$logFC!=0],me
thod="BH")<.05)

```

```

#names(genes)=row.names(tested$table[tested$table$logFC!=0,])

```

```

#table(genes)

```

```

#selected <- c(rep(1,nrow(res[res$PValue<=0.05,])),
rep(0,nrow(res[res$PValue>0.05,])))

```

```

#DE <- matrix(selected, ncol=1)

```

```

#row.names(DE) <- res$genes

```

```

#Size <- Size_genes[names(Size_genes)%in%rownames(DE)]

```

```

#DE <- DE[rownames(DE)%in%names(Size_genes)]

```

```

#Size_genes <- Size_genes[names(Size_genes)%in%names(GOterm)]
#GOterm <- GOterm[names(GOterm)%in%names(Size_genes)]
#Size <- Size_genes[names(Size_genes)%in%res$genes]
#DE <- DE[res$genes%in%names(Size_genes)]
#names(DE) <- res$genes[res$genes%in%names(Size_genes)]
#DE[is.na(DE)] = 0
#pwf = nullp(DE, bias.data=Size)

#GO.wall=goseq(pwf, gene2cat=GOterm, use_genes_without_cat=FALSE)
#dev.off()

Selected <- rbind(res[res$PValue <= 0.05 & (res$logFC >= 2 | res$logFC
<= -2),],
                 res[res$PValue <= 0.05 & (res$logFC < 2 & res$logFC > -2),],
                 res[res$PValue > 0.05,])
binarysel <- c(rep(1, nrow(res[res$PValue <= 0.05 & (res$logFC >= 2 |
res$logFC <= -2),])),
              rep(0, nrow(res[res$PValue <= 0.05 & (res$logFC < 2 &
res$logFC > -2),])),
              rep(0, nrow(res[res$PValue > 0.05,])))
names(binarysel) <- Selected$genes

binarysel <- replace(binarysel, which(names(binarysel) %in% toremove),
0)
tokeep <- names(binarysel) %in% names(Size_genes)
binarysel <- binarysel[tokeep]

pwf <- nullp(binarysel, bias.data = Size_genes)
GO.wall <- goseq(pwf, gene2cat = GOterm, use_genes_without_cat = F)
dev.off()

```

#Analysis by Biological proc. (BP), Molecular funct. (MF) or Cellular compartment (CC)

```
#Over-represented pathways (Biological proc. bp)
GO.wall.or.bp      <-  GO.wall[GO.wall$ontology == "BP" &
GO.wall$over_represented_pvalue < 0.05,]
GO.wall.or.bp     <-  na.omit(GO.wall.or.bp)
GO.wall.or.bp.df  <-  data.frame("Description"           =
as.character(GO.wall.or.bp$term),
                                "Number of DE Genes"     =
as.integer(GO.wall.or.bp$numDEInCat),
                                "Number of Genes in Cat"  =
as.integer(GO.wall.or.bp$numInCat))

GO.wall.or.bp.df <- data.frame(cbind(GO.wall.or.bp.df,
                                "Gene ratio"           =
round(GO.wall.or.bp.df$Number.of.DE.Genes/GO.wall.or.bp.df$Number.
of.Genes.in.Cat, digits = 2)),
                                "PValue"              =
GO.wall.or.bp$over_represented_pvalue)
GO.wall.or.bp.df <- na.omit(GO.wall.or.bp.df)

if (nrow(GO.wall.or.bp.df) > 0) {
  write.xlsx(GO.wall.or.bp.df,
paste0("Tables/GOseq/bp/",file_name,".xlsx"), row.names=FALSE)

##Plot GEA
#Dot plot
tiff(filename      =      paste0("Pictures/GEA/bp/",file_name,".tiff"),
width=3024, height= 4048, res=300)
print (ggplot(GO.wall.or.bp.df) +
```

```

    geom_point(mapping = aes(x=Gene.ratio, y = reorder(Description,
Gene.ratio), color = PValue, size = Number.of.DE.Genes)) +
    scale_color_gradientn(colors=c("#FC4E07",          "#E7B800",
"#00AFBB")) +
    labs(x = "Gene ratio", y = "", size = "# DE Genes") +
    ggtitle("Over-representated GO terms") +
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.5)))
dev.off()

#Pie chart
pal<-c(randomColor(count=nrow(GO.wall.or.bp.df)))
tiff(filename = paste0("Pictures/GEA/bp/piechart/",file_name,".tiff"),
width=6024, height= 2048, res=300)
print (ggplot(GO.wall.or.bp.df, aes (x="", y=Number.of.DE.Genes, fill =
factor(Description)))+
    geom_col(position = 'stack', width = 0.1) +
    #          scale_color_gradientn(colors=c("#FC4E07",  "#E7B800",
"#00AFBB")) +
    scale_fill_manual(values=pal) +
    coord_polar("y") +
    theme_classic() +
    theme(plot.title = element_text(hjust = 0.5),
    axis.line = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    legend.position = "right") +
    labs(fill="", x = NULL, y = NULL))
dev.off()
}

```

```

#Over-represented pathways (Molecular func. mf)
GO.wall.or.mf      <-  GO.wall[GO.wall$ontology == "MF" &
GO.wall$over_represented_pvalue < 0.05,]
GO.wall.or.mf     <-  na.omit(GO.wall.or.mf)
GO.wall.or.mf.df  <-  data.frame("Description"
as.character(GO.wall.or.mf$term),
                                "Number of DE Genes"
as.integer(GO.wall.or.mf$numDEInCat),
                                "Number of Genes in Cat"
as.integer(GO.wall.or.mf$numInCat))

GO.wall.or.mf.df <- data.frame(cbind(GO.wall.or.mf.df,
                                "Gene ratio"
round(GO.wall.or.mf.df$Number.of.DE.Genes/GO.wall.or.mf.df$Number.
of.Genes.in.Cat, digits = 2)),
                                "PValue"
GO.wall.or.mf$over_represented_pvalue)
GO.wall.or.mf.df <- na.omit(GO.wall.or.mf.df)

if (nrow(GO.wall.or.mf.df)) {
write.xlsx(GO.wall.or.mf.df,
paste0("Tables/GOseq/mf/",file_name,"_bp.xlsx"), row.names=FALSE)

##Plot GEA
#Dot plot
tiff(filename      =      paste0("Pictures/GEA/mf/",file_name,".tiff"),
width=3024, height= 4048, res=300)
print (ggplot(GO.wall.or.mf.df) +
      geom_point(mapping = aes(x=Gene.ratio, y = reorder(Description,
Gene.ratio), color = PValue, size = Number.of.DE.Genes)) +

```

```

        scale_color_gradientn(colors=c("#FC4E07", "#E7B800",
"#00AFBB")) +
        labs(x = "Gene ratio", y = "", size = "# DE Genes") +
        ggtitle("Over-representated GO terms") +
        theme_minimal() +
        theme(plot.title = element_text(hjust = 0.5)))
dev.off()

```

```

#Pie chart
pal<-c(randomColor(count=nrow(GO.wall.or.mf.df)))
tiff(filename = paste0("Pictures/GEA/mf/piechart/",file_name,".tiff"),
width=6024, height= 2048, res=300)
print (ggplot(GO.wall.or.mf.df, aes (x="", y=Number.of.DE.Genes, fill =
factor(Description)))+
        geom_col(position = 'stack', width = 0.1) +
        #        scale_color_gradientn(colors=c("#FC4E07", "#E7B800",
"#00AFBB")) +
        scale_fill_manual(values=pal) +
        coord_polar("y") +
        theme_classic() +
        theme(plot.title = element_text(hjust = 0.5),
                axis.line = element_blank(),
                axis.text = element_blank(),
                axis.ticks = element_blank(),
                legend.position = "right") +
        labs(fill="", x = NULL, y = NULL))
dev.off()
}

```

```

#Over-represented pathways (Molecular func. mf)
GO.wall.or.cc <- GO.wall[GO.wall$ontology == "CC" &
GO.wall$over_represented_pvalue < 0.05,]

```

```

GO.wall.or.cc <- na.omit(GO.wall.or.cc)
GO.wall.or.cc.df <- data.frame("Description" =
as.character(GO.wall.or.cc$term),
"Number of DE Genes" =
as.integer(GO.wall.or.cc$numDEInCat),
"Number of Genes in Cat" =
as.integer(GO.wall.or.cc$numInCat))

```

```

GO.wall.or.cc.df <- data.frame(cbind(GO.wall.or.cc.df,
"Gene ratio" =
round(GO.wall.or.cc.df$Number.of.DE.Genes/GO.wall.or.cc.df$Number.
of.Genes.in.Cat, digits = 2)),
"PValue" = GO.wall.or.cc$over_represented_pvalue)
GO.wall.or.cc.df <- na.omit(GO.wall.or.cc.df)

```

```

if (nrow(GO.wall.or.cc.df)){
tryCatch(write.xlsx(GO.wall.or.cc.df,
paste0("Tables/GOseq/cc/",file_name,"_bp.xlsx"), row.names=FALSE))

```

```

##Plot GEA
#Dot plot
tiff(filename = paste0("Pictures/GEA/cc/",file_name,".tiff"),
width=3024, height= 4048, res=300)
print (ggplot(GO.wall.or.cc.df) +
geom_point(mapping = aes(x=Gene.ratio, y = reorder(Description,
Gene.ratio), color = PValue, size = Number.of.DE.Genes)) +
scale_color_gradientn(colors=c("#FC4E07", "#E7B800",
"#00AFBB")) +
labs(x = "Gene ratio", y = "", size = "# DE Genes") +
ggtitle("Over-representated GO terms") +

```

```

        theme_minimal() +
        theme(plot.title = element_text(hjust = 0.5)))
dev.off()

#Pie chart
pal<-c(randomColor(count=nrow(GO.wall.or.cc.df)))
tiff(filename = paste0("Pictures/GEA/cc/piechart/",file_name,".tiff"),
width=6024, height= 2048, res=300)
print (ggplot(GO.wall.or.cc.df, aes (x="", y=Number.of.DE.Genes, fill =
factor(Description)))+
        geom_col(position = 'stack', width = 0.1) +
        #       scale_color_gradientn(colors=c("#FC4E07", "#E7B800",
"#00AFBB")) +
        scale_fill_manual(values=pal) +
        coord_polar("y") +
        theme_classic() +
        theme(plot.title = element_text(hjust = 0.5),
                axis.line = element_blank(),
                axis.text = element_blank(),
                axis.ticks = element_blank(),
                legend.position = "right") +
        labs(fill="", x = NULL, y = NULL))
dev.off()
}

####Plots
##MD
png(filename = paste0("Pictures/",file_name,".png"), width = 2732, height
= 2048, res = 300)
plotMD(qlf, main = title_name, adjust.method = "none", p.value = 0.05)
abline(h = c(-2,2), col = "blue")
dev.off()

```

