Roles of regulation of mRNA cleavage in *Mycobacterium smegmatis*

By

Paula de Camargo Bertuso

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the degree

Master of Science

In

Biology & Biotechnology

May 2016

Approved:

Dr Scarlet Shell, major advisor:_____

Dr David Adams, committee member:_____

Dr Reeta Rao, committee member:_____

ABSTRACT

One third of the world's population is infected with *Mycobacterium tuberculosis*, the bacterium that causes TB. During an infection, bacteria often survive host immune system attacks, which include oxidative stress conditions for bacteria growing inside macrophages. This makes treatment difficult and time-consuming. We hypothesize bacteria can adapt to environmental conditions by changing their mRNA maturation and degradation profiles. Using a model system, *Mycobacteruim smegmatis,* we focus on how mRNA expression is affected by oxidative stress. After construction and sequencing of RNA expression libraries, preliminary analysis showed that after three hours of H_2O_2 exposure most upregulated genes were related to DNA repair, while downregulated genes included transport proteins. After six hours of exposure, upregulated genes were similar to three hours and downregulated genes included tRNAs. 5' end mapping libraries were also constructed to access differential cleavage site abundance under oxidative stress conditions. We also investigated the roles RNase J may have in stress response and mRNA processing in Mycobacteria. RNase J and RNase E are thought to be the major RNases in bacteria. While most bacteria only have one of them, mycobacteria encode both in their genome, with RNase J being non-essential. We constructed a set of 4 strains (WT, RNase J overexpression, RNase J deletion, and complemented RNase J deletion) and tested their drug resistance and stress tolerance. Results suggests that RNase J deletion and overexpression alter drug sensitivity. Stress tolerance assays showed that WT is more tolerant to oxidative stress, followed by RNase J deletion strain and overexpression and complemented RNase J deletion strains, with the last two showing no growth when cultured with H_2O_2. Analysis of the expression profile of these strains was performed to help understand if gene expression differences are responsible for the phenotypes observed. For the complemented RNase J deletion, one operon had almost all its genes upregulated. This operon encodes a hydrogenase (Hyd3), suggesting that redox balance in the strain is perturbed.

ACKOWLEDGMENTS

I would like to give my great thanks to:

My parents, Plinio and Cilmara, my sister Camila and my boyfriend Pedro for all the love and support throughout these last 2 years even from distance.

My dearest friends Thalita and Ying. Thalita for being such an amazing person and a great listener and Ying for being the best lab-mate I could ask for.

My advisor Dr. Scarlet Shell for all the guidance and knowledge she shared with me that made grow so much as a scientist and as a person.

To my committee members Dr. David Adams and Dr. Reeta Rao for all the insightful suggestions to my project to make it more conclusive.

To all Shell lab members, who helped create a great and enjoyable environment in the lab.

To Dr José M. Argüello, Dr Jagan Srinivasan, Dr Joseph B. Duffy and Dr Robert E. Dempski, for the use of their equipment which helped support my research project.

To all faculty, staff and specially the graduate students from the Department of Biology and Biotechnology from WPI.

For all the support and encouragement from all the amazing people I had the pleasure to meet and became friends with.

The financial support from my sponsor, CAPES foundation, Ministry of Education of Brazil under grant 88888.075595/2013-00.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TABLE OF ABBREVIATIONS

Abbreviations	Full Name
µg	Micrograms
µg/mL	Micrograms per milliliter
µg/µL	Micrograms per microliter
µL	Microliter
ADC	Albumin Dextrose Catalase
ADS	Albumin Dextrose Saline
BCG	Bacillus Calmette Guerin
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CDC	Centers for Disease Control and Prevention
DMSO	Dimethyl Sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
g	Grams
hyg	Hygromicin
INH	Isoniazid
L	Liter
LB	Luria-Bertani
M. smegmatis	*Mycobacterium smegmatis*
MDR	Multidrug-resistant
mg	Milligrams
MIC	Minimal Inhibitory Concentration
mL	Milliliters
mM	Millimolar
mm	Millimeters
MOP	Mycobacterium optimal promoter
Mtb	*Mycobacterium tuberculosis*
N	Normal
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
nt	Nucleotide
OD	Optical Density

PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
pg	Picograms
pg/µL	Picograms per microliter
qPCR	Quantitative Polymerase Chain Reaction
RIF	Rifampicin
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript per million mapped reads
rpm	Rotations per minute
SNP	Single nucleotide polymorphism
TB	Tuberculosis
Tm	Melting temperature
Tris	Trisaminomethane
w/v	weight/volume
WT	Wild type

1. INTRODUCTION TO TUBERCULOSIS AS A GLOBAL PUBLIC HEALTH PROBLEM

Mycobacterium tuberculosis is the causative agent of tuberculosis (TB), which normally affects the lungs and can be fatal without proper treatment. According to the World Health Organization (WHO), latent TB is estimated to infect 2 to 3 billion people worldwide while around 5% to 15% of this group will develop active TB disease during life. WHO recommends that patients with drug-susceptible TB are treated with a combination of isoniazid (INH), rifampicin (RIF), ethambuthol and pyrazinamide for 2 months followed by a treatment with only INH and RIF for 4 months (WHO, 2015). For being quite a long treatment, some patients do not complete the full 6 months of treatment, which facilitates the development of drug resistant TB. Multidrug-resistant strains (MDR) are already a problem worldwide (Figure 1). These strains are resistant to both INH and RIF and the recomended treatment includes 20 months of more expensive and toxic drugs (WHO, 2015).

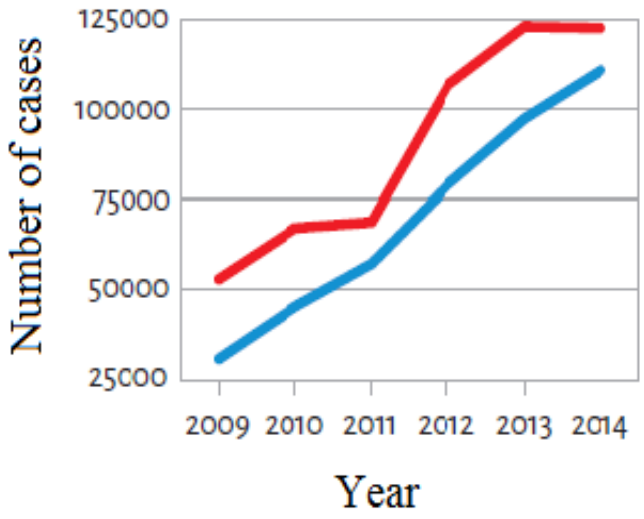

Figure 1: Global trend of multi-drug resistant (MDR-TB) cases and additional RIF-resistant TB cases detected (red) from 2009 to 2014 compared to number of TB patients treated with MDR-TB specific drug from 2009 to 2014 (blue). Modified from WHO, 2015.

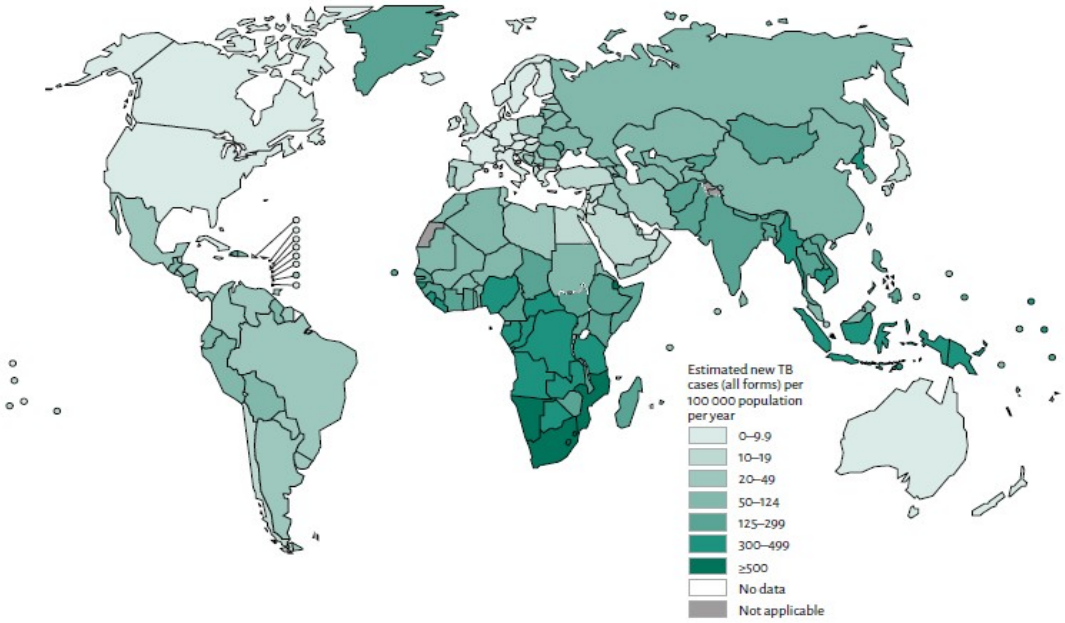

Figure 2: Estimated new global TB cases (all forms) per 100,000 population in 2014. Retrieved from WHO, 2015.

Figure 2 shows that the majority of new cases are concentrated in Africa and Asia. When we look at the high-burden countries, places where the incidence, prevalence and mortality of TB are high (WHO, 2015), as shown in the following table, are mainly in developing countries.

Table 1: 22 high-burden countries according to WHO, 2015.

22 countries with high TB burden			
Afghanistan	Ethiopia	Nigeria	Uganda
Bangladesh	India	Pakistan	United Republic of Tanzania
Brazil	Indonesia	Philippines	Viet Nam
Cambodia	Kenya	Russian Federation	Zimbabwe
China	Mozambique	South Africa	
Democratic Republic of Congo	Myanmar	Thailand	

The majority of these countries do not have sufficient resources to offer people proper health care, and that is part of the reason why the number of TB cases remains alarming. It is reported that the 22 high-burden countries represent 80% of the TB cases worldwide (WHO, 2015).

In 2014, 1.5 million deaths were attributed to TB around the globe (WHO, 2015). Considering that this disease has an extremely high cure rate when promptly diagnosed and correctly treated, the number of deaths is surprising (WHO, 2015). As stated before, the long treatments used for TB can be a problem for most people that might stop taking the medicine once they start to feel better even if the 6 months of treatment are not over. It is therefore important to study the biology of TB in order to find a better and faster way to fight this disease.

Starting in 1921, the BCG vaccine was administrated in humans to protect against the effects of TB. The vaccine consists in an attenuated strain of *Mycobacterium bovis* developed at the Pasteur Institute in France (CDC, 1996). Between 1945 and 1975, several studies were conducted to address BCG vaccine efficacy with results ranging between zero to 80% (CDC, 1996). In young children BCG is considered effective against tuberculous meningitis and miliary tuberculosis, while the protection effectiveness for pulmonary TB varies from 2 to 80% (CDC, 1996).

BCG vaccination can commonly lead to false-positive results in the tuberculin skin test. The test is not able to differentiate between those currently infected with TB and who have an immune response to the BCG and react to the tuberculin (CDC, 1996). Is reported that people previously vaccinated with BCG showed reactivity between 0 to 19 mm when tested, and the recommendations from CDC are that TB diagnosis can be made if for a BGC-vaccinated person the induration on the test is $\geq$ 10 mm (CDC, 1996).

Taken together this information led us to conclude that BCG vaccination does not necessarily protect people from getting TB infection, with its efficacy being uncertain, and testing can lead to false-positives from the BCG vaccine (CDC, 1996). For these reasons some countries where the disease is rare, like the U.S., limit the vaccination to only groups at risk (CDC, 1996).

These findings about BCG efficacy and possible problems help corroborate that, besides finding a better and faster way to fight this disease, we also need a better strategy to prevent new cases of TB.

1.1. CURRENT TREATMENTS FOR TUBERCULOSIS

As stated in section 1, it is recommended that patients with drug-susceptible TB are treated with a combination of isoniazid (INH), rifampicin (RIF), ethambuthol and pyrazinamide for 2 months followed by a treatment with only INH and RIF for 4 months (WHO, 2015).

Isoniazid is a prodrug that, once inside the mycobacterial cell, is converted to the active drug by the catalase-peroxidase KatG (Timmins and Deretic, 2006). This conversion generates reactive species that inhibit mycolic acid synthesis (Timmins and Deretic, 2006). Rifampicin stops transcription initiation in bacteria by targeting the β subunit of RNA polymerase (gene rpoB) (Agrawal *et al*, 2015).

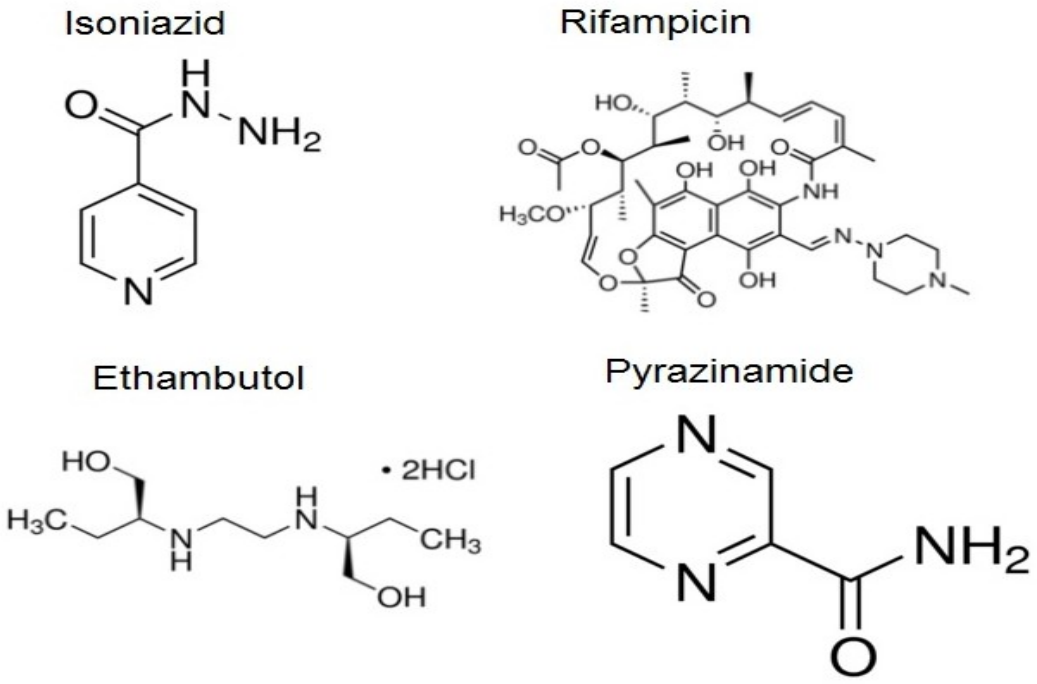

Figure 3: Drugs used for treatment of drug-susceptible TB. Reproduced from http://www.sigmaaldrich.com/united-states.html

Ethambutol is capable of arresting the multiplication of mycobacterial cells making cells non-viable after several hours of application (Forbes *et al*., 1962). Its effect is due to interference of ethambutol in the synthesis of key metabolites for cell multiplication, that will then impair cell metabolism and eventually make then lose viability (Forbes *et al*., 1962). The drug inhibits the transference of mycolic acids to the cell wall of *M. smegmatis* (Takayama *et al.*, 1979)

Pyrazinamide, on the other hand, is converted to pyrazinoic acid by PZase once inside the mycobacterial cell and inhibits membrane transport function in *M. tuberculosis* by disrupting membrane energetics (Zhang *et al.*, 2003). The drug will target more efficiently the old non-replicating bacteria as their membrane is fragile and normally present a lower membrane potential with less active metabolism when compared with young replicating bacteria (Zhang *et al.*, 2003).

These four drugs (figure 3) are the first line of treatment against *M. tuberculosis* and are effective in most of the cases when administrated to eliminate drug-susceptible TB. In the case of a patient with drug-resistant TB, second line drugs, like fluoroquinolones, can be used.

Fluoroquinolones are a class of antibiotics that interacts with DNA gyrase and topoisomerase IV (Hooper, 2000). The drug forms a complex with those two enzymes and prevents DNA replication in bacteria (Hooper, 2000).

1.2. ROLE OF POST-TRANSCRIPTIONAL REGULATION IN MYCOBACTERIAL STRESS RESPONSE

In order to address these issues it is necessary to understand how mycobacteria can adapt to hostile environments and thrive there, like for example the human body. When infecting a host *Mycobacterium tuberculosis* (Mtb) has to find ways to avoid the immune response and, when it is not able to avoid it, resist the host "attacks". When infecting a host, Mtb resides in the macrophages (Pieters, 2008). The metabolism of a macrophage in the lung liberates reactive oxygen species (ROS) that affect the bacteria growing inside these cells (Dussurget *et al*, 1998).

This process will only happen if Mtb becomes phagocytosed by an activated macrophage where the phagosome and the lysosome fuse in an attempt to kill the bacteria (Pieters, 2008). In the case that the bacteria get phagocytosed by a non-activated macrophage, it can prevent the fusion of the phagosome with the lysosome by a cholesterol mediated mechanism, and this allows Mtb to avoid facing high concentrations of ROS that can damage the cells (Pieters, 2008). Tolerance to

ROS and oxidative stress seems to play a role in Mtb infection and survival within a host but the mechanisms remain to be fully elucidated. ROS and oxidative stress effects on mycobacteria will be covered more in depth in section 1.3.

Transcript degradation is known to help bacteria to adapt to different environments by performing RNA cleavage and changing their expression profile (Arraiano *et al*, 2010). The enzymes that play a major role in RNA cleavage are the RNases (ribonucleases). RNases are known to function as a clean up mechanism for the cell, as they take part in degradation of transcripts and protect the cell from producing aberrant proteins from defective RNAs (Arraiano *et al*, 2010).

The ability of a particular transcript to evade the action of RNases can dictate its stability. Most mRNA can escape from the RNases by using secondary structures to "hide" the cleavage sites or, in the case of a highly translated mRNA, the ribosomes will help to protect the transcript (Arraiano *et al*, 2010).

RNA processing can also make a transcript more stable in response to environmental conditions, and this will also help to determine the amount of that specific protein in the cell (Rochat et al, 2013). Some RNAs, for example tRNA and rRNA, are only completely functional following transcript cleavages during maturation (Arraiano *et al*, 2010).

Once cleaved by an RNase, a can have the following possible fates: the entire transcript can be (1) destabilized and then degraded, (2) selectively destabilized or stabilized, where one fragment may go under degradation and the other will get more stable and presumably more protein will be translated, and (3) entirely stabilized (Figure 4).

It is also worth noting that the 5' end of the primary transcript contains a triphosphate while the 5' end after the cleavage site contains a monophosphate, and this is a key feature to differentiate between the 5' end origin and identify mRNA cleavage sites that may lead to transcript stabilization genome-wide (Güell *et al*, 2011).

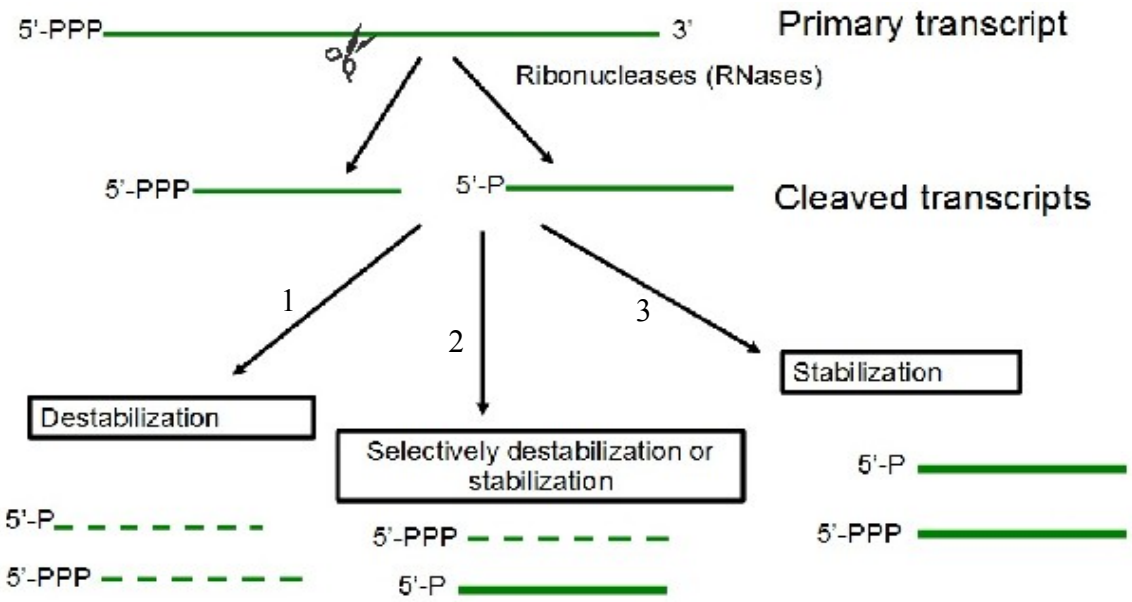

Figure 4: Possible fates a transcript may have after cleavage by an RNase.

Since RNases play a critical role in mRNA fate in the bacterial cell, and mRNA profiles can be changed due to environmental conditions, understanding how and in which conditions RNases work can help us to understand how Mtb adapt to the stress conditions it faces when infecting a host.

Table 2: Occurrence of RNases in different bacteria. Modified from Traverniti *et al*, 2011.

	E. coli	*B. subtilis*	*M. smegmatis*	*M. tuberculosis*
Endoribonucleases				
RNase E	+	−	+	+
RNase G	+	−	−	−
RNase III	+	+	+	+
Mini III	−	+	−	−
RNase M5	−	+	−	−
RNase P	+	+	+	+
RNase Z	+	+	+	+
RNase I	+	−	−	−
RNase Y	−	+	−	−
RNase J	−	+	+	+
Exoribonucleases				
RNase J	−	+	+	+
PNPase	+	+	+	+
RNase T	+	−	−	−
RNase R	+	+	−	−
RNase II	+	−	−	−
RNase PH	−	+	+	+
YhaM	−	+	−	−
Orn	+	−	+	+
NrnA	−	+	+?	+?

Table 2 shows the occurrence of RNases in various bacteria. *E. coli* has RNase E but not RNase J, while *B. subtilis* has RNase J but not RNase E. Mycobacteria, on the other hand, have both RNase E and RNase J, and this combination was not found in any other bacteria analyzed (Traverniti *et al*, 2011). RNase E is known to be an essential gene in mycobacteria, while RNase J is non-essential (Traverniti *et al*, 2011). The fact that it is a non-essential gene makes it possible to knock out RNase J more easily, and this knock out strain in *M. smegmatis* is known to have no growth defect (Traverniti *et al*, 2011). RNase J is also known to have exo and endoribonucleotic activity, which is unlike any other of the RNases presented on table 2, and also plays a role in ribosomal RNA maturation and stability (Traverniti et al, 2011). For all these reasons we decided to look at the roles that RNase J may have in mRNA cleavage and adaptation to stress. More information about RNase J can be found in section 1.3.

To perform the work described in this thesis, *Mycobacterium smegmatis* was used as a model system for *Mycobacterium tuberculosis*. *M. smegmatis* is a non-pathogenic fast growing mycobacteria which, as can be seen in table 2, has the same RNases as Mtb (Traverniti *et al*, 2011). The fact they have similar RNases suggests that the RNA degradation and maturation patterns are the same in *M. smegmatis* and Mtb (Traverniti *et al*, 2011). Genetic manipulation can be done in *M. smegmatis* relatively easily using molecular biology tools (Altaf *et al*, 2010), which corroborates its use as a model system for Mtb.

1.3. OXIDATIVE STRESS IN MYCOBACTERIA

ROS (reactive oxygen species) are known to be involved in several biological processes, for example: mutation, degenerative diseases, aging, development, inflammation and cell signaling (Kohen and Nyska, 2002).

The catalytic process of some enzymes and the dismutation of superoxide can produce H_2O_2, which can damage the cell in concentrations as low as 10 μM and can diffuse across biological

membranes (Kohen and Nyska, 2002). The damage caused by the presence of H_2O_2 includes degradation and/or inactivation of proteins, release of iron inside the cell, and oxidation of structures such as DNA and lipids (Kohen and Nyska, 2002). Figure 4 shows the reaction of iron with hydrogen peroxide (H_2O_2). Iron ions are the most likely to take part in the Fenton reaction because of the high amount present in the cells when compared with other metals (Kohen and Nyska, 2002). When those ions are present on the surface of proteins and DNA it can facilitate the H_2O_2 related damage to those structures due to the proximity between the molecule and the radicals generated after the H_2O_2 decomposition (Kohen and Nyska, 2002).

$$Fe^{+2} + H_2O_2 \longrightarrow Fe^{+3} + OH^{\cdot} + OH^{-}$$

Figure 5: Fenton reaction between iron and hydrogen peroxide. This reaction may occur with any transition metal that contains un-paired electrons. Reproduced from (Kohen and Nyska, 2002).

Most ROS have short half-lives since they react with other molecules very rapidly; $OH^{\cdot}$ is an example of such a ROS (Kohen and Nyska, 2002). H_2O_2 on the other hand is somewhat more stable and can reach other sites in the cells far from where it was originally produced (Kohen and Nyska, 2002).

If an organism is growing aerobically it will generate ROS due to the oxidation of redox enzymes (Broden *et al*, 2016). This process in which bacteria experience ROS as a byproduct of their metabolism can be called endogenous oxidative stress (Hassett and Cohen, 1986). On the other hand, the source of the ROS can also be exogenous, as when pathogenic bacteria are attacked by the host immune system and inflammation is induced (Hassett and Cohen, 1989). This is the scenario that Mtb is likely to face when it gets phagocytosed by an activated macrophage as discussed in section 1.

In order to fight oxidative stress bacteria have evolved some defenses, such as for example DNA and protein repair mechanisms and detoxifying enzymes (Hassett and Cohen, 1989). An efficient DNA repair mechanism, for example, can promptly identify a damaged site, remove it and

incorporate undamaged bases in its place (Kohen and Nyska, 2002). Table 3 shows examples of detoxifying enzymes. It is not clear why bacteria have multiple enzymes that are likely to be used to resolve excess peroxides in the cell, for example, *B. subtilis* appear to have 9 potential peroxide scavenging enzymes in its genome (Broden *et al*, 2016).

Table 3: Mycobacterial defenses against Oxidative stress. Retrieved from Dussurget and Smith, 1998.

Enzyme	Gene	Notes[b]
Catalase/peroxidase	*katG*	Found in most mycobacteria; inactive in *Mycobacterium leprae*; found in *Mycobacterium tuberculosis*; associated with resistance to INH and virulence in guinea pigs
Catalase	*katE*	Found in a limited number of mycobacteria
Mn-SOD	*sodA*	Most mycobacteria have an Mn-SOD
Fe-SOD	*sodA*	*M. tuberculosis* has an Fe-SOD[c]
Cambialistic SOD	*sodA*	*Mycobacterium smegmatis* SOD is active with either iron or manganese as a metal ion ligand
Thioredoxin	*trxA/trxB*	Found in many mycobacteria, including *M. leprae* and *M. tuberculosis*
Alkyl hydroperoxide reductase	*ahpC*	Found in many mycobacteria, including *M. leprae*, *M. tuberculosis*, *M. smegmatis* and *Mycobacterium avium*

[a]Abbreviations: INH, isoniazid; SOD, superoxide dismutase.
[b]In *Escherichia coli*, KatG and AhpC are positively regulated by OxyR, the Mn-SOD is positively regulated by SoxR/SoxS, and KatE is positively regulated by KatF (RpoS).
[c]An open reading frame coding for a putative Cu/Zn-SOD is found in the *M. tuberculosis* genome.

Is thought that iron deficiency may lead to elevation of oxidative stress levels in the cell due to decreased function of some of the protective enzymes, and bacteria may face iron deficiency when infecting a host (Dussurget and Smith, 1998). In order to try to resolve the high levels of oxidative stress, the mycobacterial genome encodes enzymes like the superoxide dismutases (SODs) that are responsible for the conversion of superoxides to H_2O_2; the H_2O_2 is then converted to simpler organic peroxides by catalases and peroxidases like KatG (Dussurget and Smith, 1998).

Isoniazid can also be linked to increased oxidative stress in mycobacteria when we look at KatG function: (1) KatG is a catalase/peroxidase that will help resolve the high concentrations of H_2O_2 by converting it to simpler peroxides such as hydrogen peroxide, and (2) KatG will convert the INH prodrug to its active form which inhibits the mycolic acids synthesis (Dussurget and Smith, 1998). It is believed that the INH conversion also generates ROS, which can contribute to the

degree of drug toxicity (Dussurget and Smith, 1998). It is postulated that INH will diffuse passively into the mycobacteria cell and KatG is not involved in the progdrug transport (Bardou *et al*., 1998). Once INH is inside the cell, KatG will promote a peroxidation reaction and "activate" INH, and this reaction will result in the production of oxygen radical species (Timmins and Daretic, 2006) which can then damage different components of the cell.

The mycolic acids of some pathogenic mycobacteria were reported to have decreased sensitivity to lipid peroxidation due to cyclopropanation and this may be considered another defense against oxidative stress (Dussurget and Smith, 1998). When oxidized in the presence of NADH, INH will form the complex INH-NADH that inhibits the enzyme InhA, one of the major players in mycolic acids synthesis (Timmins and Daretic, 2006). INH is also capable of forming the complex INH-NADP when $NADP^+$ is available and then inhibits the enzyme MabA *in vitro*, also known to be a key player in mycolic acid synthesis (Timmins and Daretic, 2006).

The presence of ROS are also reported to change the activation of different sigma factors, for example in *Streptomyces coelicolor*, where SigR activates the expression of a whole set of genes in response to oxidative stress (Touzain *et al*, 2008). In Mtb, SigH is considered the ortholog of SigR (Touzain *et al*, 2008). This suggests that SigH may also influence gene expression under oxidative stress conditions.

Li *et al*, 2015 analyzed the transcriptional response to H_2O_2 in *M. smegmatis* by using low and high concentrations of hydrogen peroxide in a similar manner as was done in the present study. Li *et al*, 2015 data also suggests that different sigma factors will regulate the response to different levels of oxidative stress inside the mycobacterial cell. It was also shown that the protective enzymes mentioned earlier, such as *katG* and *trx* have increased expression upon exposure to oxidative stress (Li *et al*, 2015).

1.4. ROLES OF RNASE J IN MYCOBACTERIA

It was suggested that the major RNases responsible for RNA metabolism profile changes in bacteria are RNases E and J (Traverniti *et al*, 2011). And as discussed previously, RNase J is a non-essential gene in Mtb and *M. smegmatis* (Traverniti *et al*, 2011). RNase J is also unique due to its 5' to 3' exoribonucleolytic activity as well as endoribonucleotic activity (Traverniti *et al*, 2011). Here the roles of RNase J in mycobacteria will be addressed in more depth.

In *M.smegmatis* RNase J is 558 amino acids in length and is encoded by the gene MSMEG_2685 (Traverniti *et al*, 2011). *M. smegmatis* RNase J can be classified as a β-CASP zinc-dependent metallo-β-lactamase (Traverniti *et al*, 2011), and in this family of proteins the zinc is used to perform the cleavage reaction (Callebaut *et al*, 2002). The structure of RNase J from *M. smegmatis*, containing the β-CASP domain can be seen in Figure 6.

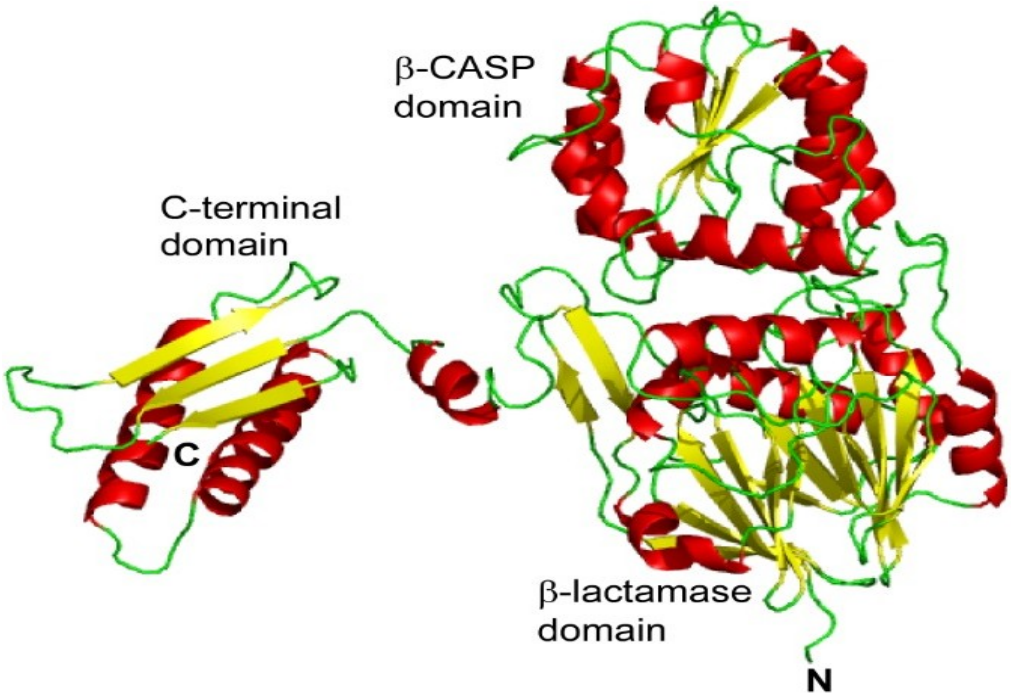

Figure 6: RNase J 3D structure from *M. smegmatis* (MSMEG_2685). Modified from (Traverniti *et al*, 2011).

The β-CASP domain stands for metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2 and can be found in bacteria, eukaryotes and archaea (Callebaut *et al*, 2002). Some metallo-β-

lactamase proteins can use nucleic acids as substrates, including a subunit of the cleavage and polyadenylation specificity factor (CPSF) which is thought to take part in RNA cleavage, and other enzymes such as SNM1 from mice and PSO2 from yeast which play roles in DNA repair (Callebaut *et al*, 2002). Finally, Artemis which is also part of the metallo-β-lactamase group has been described as a recombination/DNA repair enzyme and can have exo and endonuclease activity depending on the complex it is linked to (Callebaut *et al*, 2002).

Among the proteins containing the β-CASP domain identified by Callebaut *et al*, 2002, it is worth noting the presence of Rv2752c from *Mycobacterium tuberculosis* and ML1512 from *Mycobacterium leprae*. Upon searching Rv2752c in the Tuberculist website (http://tuberculist.epfl.ch/index.html) it is possible to see that the ortholog for this gene in *M. leprae* is ML1512 and in *M. smegmatis* the ortholog is MSMEG_2685, which is RNase J. This helps corroborate that Mtb and *M. smegmatis* have similar RNase J domain structures, suggesting the *M. smegmatis* may be a good model for studying the function of RNase J in Mtb.

Much of the work done to understand the roles of RNase J was done in *Bacillus subtilis*, which actually has RNase J1 and RNase J2 (Arraiano *et al*, 2010). RNase J1 was shown to be non-essential in *B. subtilis* (Figaro *et al,* 2013) and prefers single-stranded 5' monophosphate ends as substrates for its exoribonuclease activity (Arraiano *et al*, 2010). RNase J1 and J2 are reported to from a complex but overall activity of RNase J2 is around 100 fold lower than RNase J1 activity (Figaro *et al,* 2013). The exoribonucleatic activity of RNase J1 was also reported to be responsible for the 16S rRNA maturation and regulation of some specific mRNA transcripts in *Bacillus thuringiensis*, including the cryIIIA transcript which encodes for a insecticidal protein and the tryptophan operon (trp) (Arraiano *et al*, 2010).

When there is enough tryptophan in the cell, the TRAP (trp RNA-binding attenuation protein) complex binds to the transcript (Arraiano *et al*, 2010) and RNase J1 then cleaves this transcript between the TRAP complex and the transcription terminator (ter) (Condon, 2010). The

fragment upstream is then degraded by the PNPase enzyme while the fragment downstream of the cleavage site is degraded by RNase J1 exoribonucleatic activity (Condon, 2010). Figure 7 shows a graphical representation of the fate of the trp transcript in *B. subtilis* after endoribonucleotic activity of RNase J.

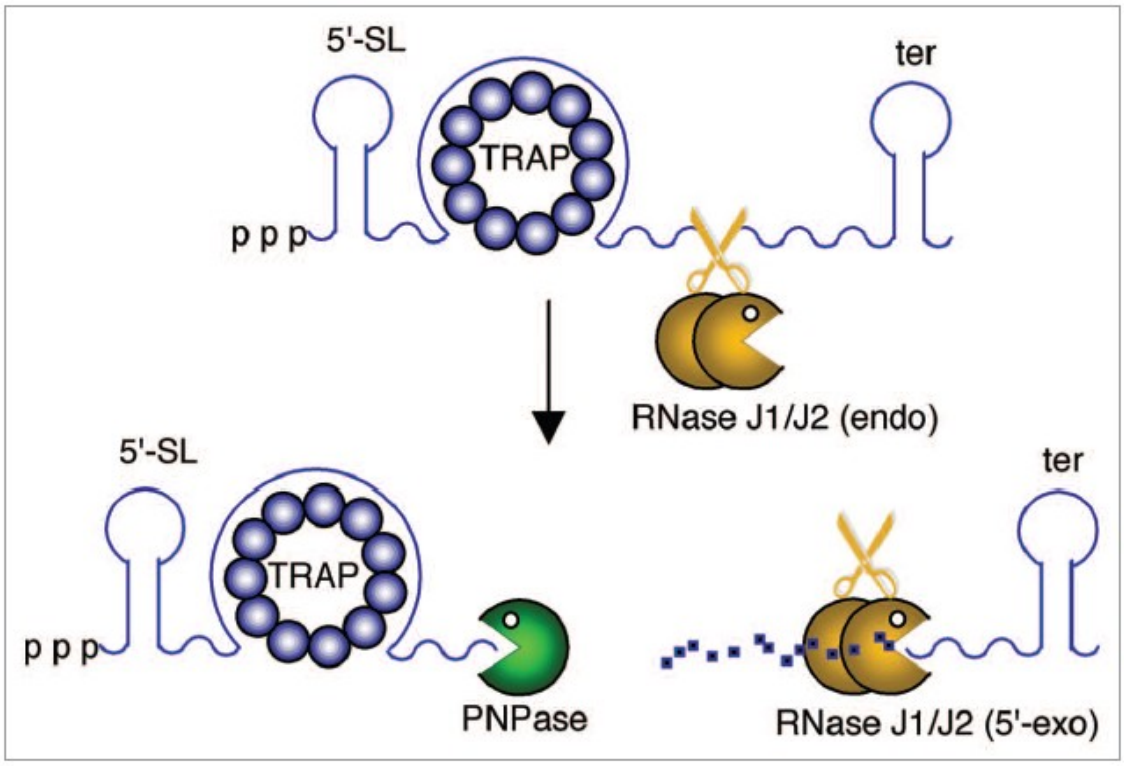

Figure 7: Fate of the trp transcript in *B. subtilis* after endoribonucleotic activity of RNase J. Reproduced from Condon, 2010.

Overall, in the case of *B. subtilis,* mRNA stability shows a small increase in the RNAse J1/J2 double deletion strain and these enzymes are considered to be functionally homologous to *E. coli* RNase E (Even *et al*, 2005). In Even *et al*., 2005 mRNA decay was measured by culturing early exponential phase cultures with ^{3}H-uridine for a few minutes to label the mRNA. Transcription was then stopped and total RNA extracted in several time intervals; the amount of mRNA present in each time point sample was then measured by the amount of radiolabelled uridine and the reported results showed a small increase in mRNA stability (Even *et al*, 2005). In *B. subtilis*, lack of RNase J1 was also shown to cause growth defects, hypersensitivity to antibiotics and alterations in cell morphology that suggest defects in actin-like proteins (Figaro *et al*, 2013). In *Staphylococcus*

aureus, an RNase J deletion strain was reported to accumulate monophosphorylated 5' ends when compared to a WT strain (Linder *et al*, 2014). RNase J was also reported to act in 16S rRNA maturation and takes part in both cleavage of the transcript from RNase P in *S. aureus* and degradation of the 5' end of the initial transcript after cleavage (Linder *et al*, 2014).

In *M. smegmatis*, RNase J was shown to be responsible for degrading 16S rRNA while RNase E is the enzyme responsible for the maturation of this transcript (Traverniti *et al*, 2011). For the 23S rRNA, on the other hand, RNase J plays a major role in its maturation (Traverniti *et al*, 2011) since an RNase J deletion strain had no detectable mature 23S rRNA in the cell (Traverniti *et al*, 2011). Finally, both RNase J and RNAse E were reported to be involved in maturation of the 5S rRNA (Traverniti *et al*, 2011). Since rRNA maturation in *M. smegmatis* was shown to be different from both *E. coli*, which does not have RNase J, and *B. subtilis*, where RNase E is not present (Traverniti *et al*, 2011), understanding the role of RNase J in mycobacteria might help reveal if and how RNase E and RNAse J functionally interact inside the cell and how this might influence stress survival in Mtb.

2. PROJECT PURPOSE

This project has two main goals: (1) understand how oxidative stress conditions change RNA expression profiles in mycobacteria, and whether RNA cleavage contributes to these changes by influencing transcript stability and fate, and (2) understand the roles of RNase J in *M. smegmatis* and if this enzyme has any involvement in mycobacteria response to stress.

To achieve the first goal, RNA-seq expression libraries and 5' end mapping libraries were constructed under normal and oxidative stress conditions to investigate genes expression and transcript cleavage in *M. smegmatis* transcriptome wide.

In order to achieve the second goal, an RNase J knock out strain was created using recombineering strategy with homologous recombination, and compared with WT, overexpression and complemented strains. RNase J knockout and overexpression strains were phenotypically analyzed by measuring drug resistance and stress tolerance. RNA-seq expression libraries were also constructed to analyze the effects of RNase J on gene expression transcriptome-wide.

3. MATERIAL AND METHODS

3.1. STRAINS USED

This work was developed using the *Mycobacteruim smegmatis* strain mc^2 155, referred to herein as WT strain, and *Escherichia coli* DH5-alpha competent cells.

3.2. MEDIA PREPARATION

3.2.1. *M. smegmatis* MEDIA

3.2.1.1. MIDDLEBROOK 10x ADC SUPPLEMENT:

50g/L of BSA fraction V were dissolved in Milli Q water and agitated for at least 2 hours with low heat. Next 20g/L dextrose, 8.5 g/L NaCl and 30 mg/L catalase were added to the media and mixed with agitation for a few minutes. Media was filter sterilized and stored at 4 °C until use.

3.2.1.2. MIDDLEBROOK 10X ADS SUPPLEMENT (FOR OXIDATIVE STRESS EXPERIMENTS):

50g/L of BSA fraction V were dissolved in Milli Q water and agitated for at least 2 hours with low heat. Next 20g/L dextrose and 8.5 g/L NaCl were added to the media and mixed with agitation for a few minutes. Media was filter sterilized and stored at 4 °C until use.

3.2.1.3. MIDDLEBROOK 7H9 BROTH

4.7 g/L of 7H9 Broth powder, 0.25% of the final solution volume of 20% Tween 80, 0.4 % of the final solution volume of 50% glycerol and 10% of the final solution volume of 10x ADC or ADS were dissolved in Milli Q water with agitation for a few minutes. Media was then filter sterilized and stored at 4 °C until use.

3.2.1.4. MIDDLEBROOK 7H10 AGAR

19 g/L of 7H10 Agar powder was mixed with Milli Q water and 1% of the final solution volume of 50% glycerol and mixed for a few minutes. Media was then autoclaved 20 minutes at 121 °C and when medium cooled down 100 mL ADC was added and antibiotic if needed. 25 mL of media was then poured in each petri dish and when solid stored at 4 °C until use. In the case of square plate 35 mL per plate is used.

3.2.2. *E. coli* MEDIA

3.2.2.1. LB BROTH

25 g/L of LB Broth powder was dissolved on Milli Q water and mixed until complete dissolution of the powder. Media was autoclaved 20 minutes at 121 °C and stored in room temperature until use.

3.2.2.2. LB AGAR

40 g/L of LB Agar powder was dissolved on Milli Q water and mixed until complete dissolution of the powder. Media was autoclaved 20 minutes at 121 °C and stored in room temperature until use.

3.3 DNA AND RNA ISOLATION FROM MYCOBACTERIA

3.3.1. DNA ISOLATION FROM MYCOBACTERIA

Unless noted, all centrifugation steps were done at room temperature.

5 mL to 10 mL of liquid culture in log phase were centrifuged at 4000 rpm for 5 minutes and cells were ressuspended in 200 µL of TE Buffer (10 mM Tris pH 7.5 and 1 mM EDTA) and transferred to a beat-beating tube (MP Biochemicals lysing matrix B, product #6911-500). 400 µL phenol:chloroform:isoamyl alcohol (25:24:1) pH 8 was added to each tube and the tubes were

vortexed for a couple minutes in order to break the cell envelopes. Samples were then centrifuged at 14000 rpm for 5 minutes and the upper aqueous layer was transferred to a new tube.

1 mL of ice-cold ethanol (95%-100%) was added to each sample and mixed gently. Samples were then centrifuged at 14000 rpm for 5 minutes at 4 °C, the supernatant was aspirated off and 1 mL of 70% ethanol added to wash the samples.

After the wash samples were centrifuged at 14000 for 5 minutes at room temperature and ethanol was aspirated off. Tubes were then centrifuged one more time at 14000 for 30 seconds and any residual ethanol was removed with a pipet.

The pellet was dryed with the tubes open for a few minutes and 100 µL of ultra pure water added to ressuspend the samples. Finally, DNA concentration was measured using NanoDrop® and samples stored in -20 °C until use.

3.3.2. TOTAL RNA EXTRACTION FROM MYCOBACTERIA

5 mL to 10 mL of liquid culture with OD between 0.5 to 1 were centrifuged at 4000 rpm for 5 minutes at 20 °C. Supernatant was then removed without disturbing the pellet and resuspend in 1 mL TRIzol (Invitrogen). Samples were transferred to bead-beating tubes (MP Biochemicals lysing matrix B, product #6911-500) and bead-beat in the FastPrep 5G twice for 40 seconds at 9 msec, placing samples on ice for 2 minutes in between.

300 µL chloroform was added to each sample and mixed and samples were centrifuged at 15000 rpm for 15 minutes at 4 ºC. Aqueous layer was then pipeted and added to a new tube containing 600 µL isopropanol and mixed. Sample tubes were left in -20 °C freezer for 45 minutes to overnight.

After the -20 °C incubation, samples were centrifuged at 15000 rpm for 10 minutes at 4 ºC and the supernatant was poured off. 1 mL 75% ethanol was added and the tubes were gently inverted several times to wash the pellet. Samples were centrifuged again for 1 minute at 15000 rpm

at 4 ºC and supernatant was poured off.

Samples were then centrifuged again briefly to bring liquid to bottom of tube and the remaining supernatant was pipetted off. Tubes were allowed to stay at room temperature with lids closed for 10 minutes and any remaining supernatant was removed.

Pellets were ressuspended in 100 µL of ultra pure RNase-free water, RNA concentration was then measured using NanoDrop® and samples were stored at -80 ºC until use.

3.3.3. DNASE TREATMENT

In order to minimize DNA contamination in RNA extraction samples, a maximum of 20 µg of RNA was treated with 2.5 µL Ambion DNase Turbo enzyme, 10 µL of 10x DNase buffer and ultra pure RNase-free water to a final volume of 100 µL. Samples were then incubated for 1 hour at 37 °C with gentle agitation. After DNase treatment proceed to RNA clean up.

3.3.4. RNA CLEAN UP

All centrifugation steps were done in room temperature and samples were stored in -80 °C until use.

To start the procedure, RNA samples had their volume adjusted to 100 µL with RNase-free water if needed then 350 µL RTL buffer was added to each sample. Following, 250 µL of ethanol (96-100%) was added to each sample and mixed well by pipeting. Samples where then transferred to an RNeasy® column and centrifuged for 15 seconds at 15000 rpm. The flow-through was discarded and 500 µL RPE buffer to each sample to wash the membrane. After 15 seconds centrifugation at 15000 rpm, the flow-through was once again discarded and the wash was repeated.

A third wash step was performed by adding 500 µL RPE buffer and samples were centrifuged for 2 minutes at 15000 rpm. Samples were then transferred to a 2 mL new collection tube provided by the manufacture and centrifuged at 15000 rpm for 1 minute to collect any

remaining buffer. Samples were placed in a clean 1.5 mL collection tube and between 50 µL to 100 µL of RNase-free water was added to the columns, depending on the expected yield. Samples were centrifuged at 15000 rpm for 1 minute to elute the RNA.

3.4. OXIDATIVE STRESS EXPRESSION AND 5' END MAPPING LIBRARIES PREPARATION

Using the oxidative stress media containing ADS as supplement, WT strain was thawed from the -80 °C frozen stock and cultured for 1 day. On the next day cells were diluted and were grown overnight to reach OD_{600} of approximately 0.8. Cultures were diluted to OD 0.15 and incubated in triplicate with different mM concentrations of hydrogen peroxide for 9h with OD_{600} measurements taken at every 3h.

Once the hydrogen peroxide concentration for library construction (3.2 mM) was chosen, cells were again grown using the same method as described above and harvested at 3h and 6h for library construction. After total RNA extraction of the samples as previously stated, an RNA clean up procedure were performed in the cells followed by a DNase treatment and another RNA clean up.

To start with the libraries procedure, 6 out of the 9 harvested samples were chosen. The samples used for this assay can be seen on the following table:

Table 4: Oxidative Stress experiment samples used for library construction

Condition	Replicates
Control – no H_2O_2 3h	1 and 2
3.2 mM H_2O_2 3h	1 and 2
3.2 mM H_2O_2 6h	1 and 3

3.4.1. OXIDATIVE STRESS EXPRESSION LIBRARIES

To construct the expression libraries the Epicentre ScriptSeq™ Complete kit for bacteria was used following the manufacturer's recommendations as shown in the next sections:

3.4.1.1. rRNA DEPLETION

For this procedure we used the batch washing protocol to prepare the beads for use. For each of the 6 samples 225 µL of magnetic beads were used.

The total of 1350 µL of magnetic beads were placed into a clean microcentrifuge tube and then placed into a magnetic stand for 1 minute so the solution turned clear. The supernatant was discarded and 1350 µL of RNA-free water was added to the tube to wash the beads. Samples were mixed at medium speed and placed on the magnetic stand. The supernatant was removed and the wash step repeated on more time.

The tube was then removed from the magnetic stand and 360 µL of Bead Resuspension Solution was added. Samples were mixed at medium speed and divided into 65 µL aliquots. 1 µL of Riboguard RNase Inhibitor was added to each aliquot and tubes were stored in room temperature until use.

As the washed magnetic beads were waiting at room temperature the volumes containing 5 µg of RNA for each sample were calculated and combined with RNase-free water for a total of 26 µL per sample. To each sample 4 µL of Ribo-Zero Reaction Buffer and 10 µL of Ribo-Zero rRNA Removal Solution were added for a final volume of 40 µL per sample. The reaction were mixed by pipeting and incubated at 68 °C for 10 minutes followed by a 5 minute incubation at room temperature.

The treated RNA samples were then mixed with the magnetic beads by pipeting and then mixed by vortexing for 10 seconds. The tubes were incubated as room temperature for 5 minutes and then mixed by vortexing for 10 seconds one more time. Samples were incubated at 50 °C for 5

minutes and immediately placed in a magnetic stand for at least 1 minute until the solution become clear. The supernatant, containing the rRNA-depleted samples were transferred to new micro centrifuge tubes and ut through a purification procedure.

Before starting the purification, sample volumes were adjusted to 180 µL using RNase-free water. Then 18 µL of 3 M sodium acetate and 2 µL glycogen were added to each sample and gently mixed. 600 µL ice-cold 100% ethanol was added to each sample and mixed gently and the samples were placed in -20 °C for at least 1 hour.

After the 1 hour incubation samples were centrifuged at 15000 rpm for 30 minutes at 4 °C and the supernatant was discarded. The pellet was washed twice with ice-cold 70% ethanol and centrifuged at 15000 rpm at 4 °C. After carefully removing the supernatant from the second wash, samples were centrifuged for an extra minute to collect any residual ethanol and any remaining supernatant was discarded. The samples were allowed to stay at room temperature for 5 minutes and the pellet was dissolved in 10 µL RNase-free water. If not used immediately, samples were stored in -80 °C.

3.4.1.2. RNA FRAGMENTATION, cDNA SYNTHESIS AND TERMINAL-TAGGING

After rRNA depletion, 2 µL of the treated sample was mixed with 7 µL RNase-free water, 1 µL RNA fragmentation solution and 2 µL of cDNA synthesis primer to fragment the RNA and anneal the cDNA synthesis primer. The samples were incubated at 85 °C for 5 minutes and the fragmentation was stopped by placing the tubes in an ice-water bath.

Next, a cDNA synthesis reaction was prepared by mixing with each sample 3 µL cDNA synthesis PreMix, 0.5 µL 100mM DTT and 0.5 µL StartScript reverse transcriptase. Samples were then incubated at 25 °C for 54 minutes followed by an incubation at 42 °C for 20 minutes using a thermocycler. The reaction was equilibrated to 37 °C and 1 µL of finishing solution was added to each sample and mixed by pipeting. Samples were then incubated at 37 °C for 10 minutes followed

by 3 minutes at 95 °C. Once the incubation was finished, samples were equilibrated to 25 °C and had 7.5 µL terminal tagging premix and 0.5 µL DNA polymerase added. Samples were mixed by pipeting and incubated at 25 °C for 15 minutes followed by 3 minutes at 95 °C. Samples were equilibrated at 4 °C until the purification step was initiated.

For this purification step the columns from the MinElute kit from Quiagen were used. 125 µL of buffer PB were added to each sample and the final solution was transferred to a labeled MinElute column. To bind the cDNA, the columns were centrifuged for 1 minute at 14000 rpm. After the flow-through was discarded, 750 µL buffer PE was added to wash the samples and they were centrifuged for 1 minute at 14000 rpm. This wash step was repeated on more time and after the flow-through was discarded, the samples were centrifuged for an extra minute at 14000 rpm to remove any residual ethanol from buffer PE.

Columns were placed in clean microcentrifuge tubes and 25 µL buffer EB was added to elute the cDNA and samples were centrifuged for 1 minute at 14000 rpm. Samples were stored at -20 °C until use.

3.4.1.3. ADDING THE INDEXES AND FINAL LIBRARY PURIFICATION

22.5 µL of the purified di-tagged cDNA from the previous steps was mixed to 25 µL FailSafe PCR PreMix E, 1 µL forward PCR primer, 1 µL unique Script Seq Index PCR primer (reverse) and 0.5 µL FailSafe PCR enzyme. The PCR was done in two batches of 3 samples each where each batch contained one replicate of each condition. For the first batch we used 12 PCR cycles and for the second 14 cycles were performed. The conditions are as it follows: 3 minutes at 95 °C to denature the DNA, 12 or 14 cycles with 95 °C for 40 seconds, 55 °C for 30 seconds and 68 °C for 3 minutes and a final incubation using 68 °C for 7 minutes. The following table shows which unique index was used for each sample, and the sequences for the respective indexes can be found in section 6.1. of the Appendix.

Table 5: Unique Indexes used for Oxidative stress expression library construction

Sample	Unique Index
Control 3h replicate 1	Epicentre Kit Index 7
Control 3h replicate 2	Epicentre Kit Index 8
3.2mM H_2O_2 3h replicate 1	Epicentre Kit Index 9
3.2mM H_2O_2 3h replicate 2	Epicentre Kit Index 10
3.2mM H_2O_2 6h replicate 1	Epicentre Kit Index 11
3.2mM H_2O_2 6h replicate 3	Epicentre Kit Index 12

After the PCR samples were purified using the Agencourt AMPure XP Magnetic Beads purification method. Initially, AMPure XP beads were allowed to warm to room temperature and 50 μL of the beads solution were added to each sample and mixed thoroughly by pipeting. The total volume of sample (100 μL) was transferred to a clean microcentrifuge tube and incubated at room temperature for 15 minutes. Samples were then placed in a magnetic stand and incubated at room temperature for 5 minutes until the liquid appeared clear.

The supernatant was removed and discarded and the tubes which were still located in the magnetic stand were washed twice with 200 μL of 80% ethanol for 30 seconds. The beads were allowed to air dry on the magnetic stand for 15 minutes at room temperature and the tubes were removed from the stand after 20 μL nuclease-free water was added to the samples. The beads were resuspended by pipeting and incubated in room temperature for 2 minutes and placed again in the magnetic stand for at least 5 minutes until the liquid appeared completely clear. The supernatant, which contained the finalized library, was transferred to a new tube, DNA concentration was measured using QuBit and samples stored in -20 °C until the start of the sequencing procedure.

The following figure shows an overview of the library construction procedure used.

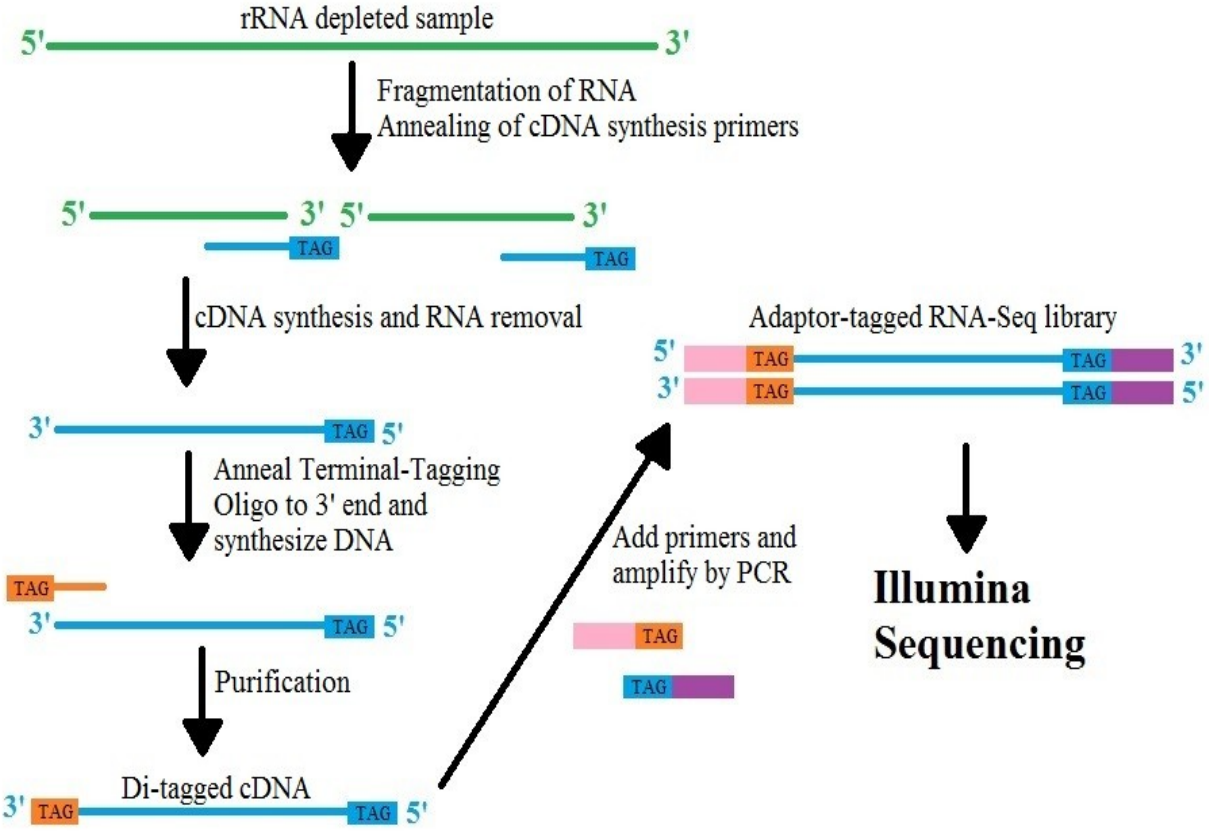

Figure 8: Overview of the oxidative stress expression library preparation.

3.4.2. OXIDATIVE STRESS 5' END MAPPING LIBRARIES

rRNA depleted samples from section 3.4.1.1 were also used for construction of 5' end mapping libraries. Libraries were constructed essentially as described in Shell *et al.*, 2015, with details below.

3.4.2.1. DEPHOSPHORYLATION AND DNA-RNA HYBRID LIGATION

For each one of the rRNA-depleted samples we used a total of 6 µL: 3 µL for the converted library and 3 µL for the non-converted library. For the converted library, which was the one where the dephosphorylation occurred, we combined 3 µL of rRNA-depleted sample, 2 µL 10X Buffer, 1 µL 5' Polyphosphatase and 14 µL RNase-free water. For the non-converted library, 1 µL water was substituted for the 1 µL of 5' Polyphosphatase enzyme. All 12 samples were incubated at 37 °C for

1 h and then purified using the RNA clean up as described in section 3.3.4. with only changing the elution step. Instead of using 50 µL RNase-free water, for this step samples were eluted twice using 30 µL of water containing RNase OUT (10 µL in 1 mL of water) and DTT (10 µL in 1 mL of water) for a final volume of 60 µL. After purification, 1 µL 100 mM Tris pH 7.5 was added to each sample and they were concentrated to 8 µL using a speedvac centrifuge.

1 µL of oligo SSS392 1 µg/µL (DNA-RNA hybrid adapter) was added to each sample and they were incubated for 10 minutes at 65 °C and snap-cooled in an ice-water bath once the incubation period was over. In sequence, the following was added to each sample: 1 µL RNase OUT, 10 µL 50% PEG 8000, 3 µL 10X T4 RNA ligase I buffer, 3 µL 10 mM ATP, 3 µL DMSO and 1 µL T4 RNA ligase I. Samples were then incubated at 20 °C overnight. On the next day, samples were purified using the RNA clean up as described in section 3.3.4 and concentrated to 5 µL using a speedvac centrifuge.

The following figure shows how the dephosphorylation happens in the 5' end and which structures (transcription start sites and cleavage sites) each library construction will capture since the enzyme T4 RNA ligase used in this procedure is only able to act in a 5' monophosphate.

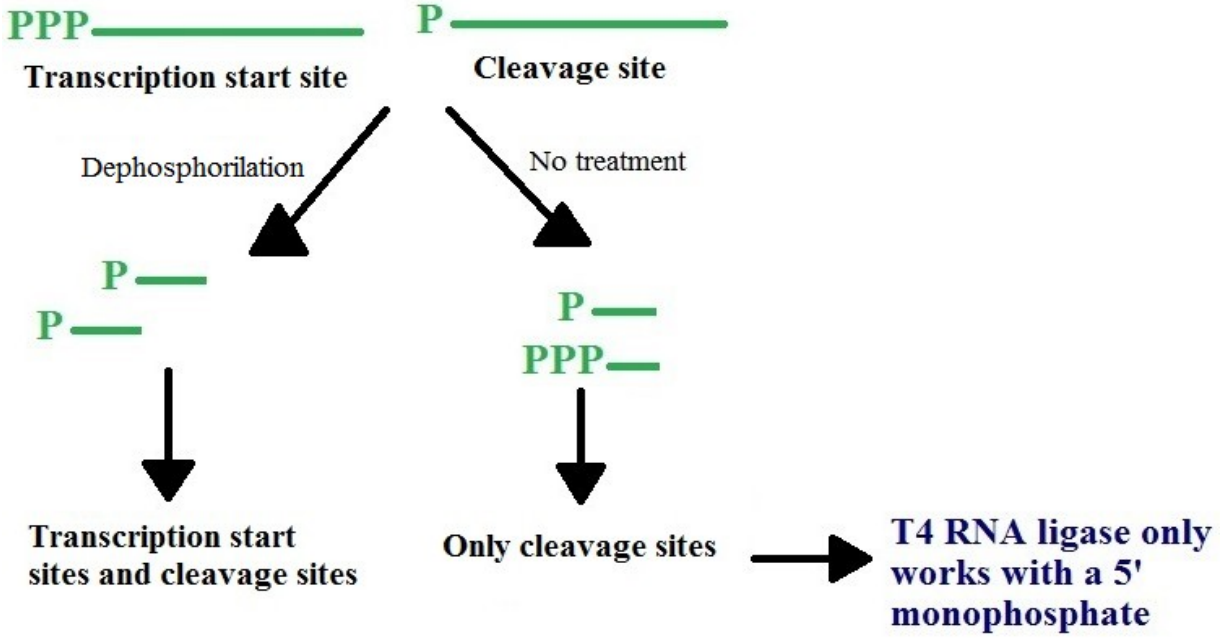

Figure 9: How to differentiate between a transcription start site and a cleavage site based on their 5' end structures and set ups for the converted (dephosphorylation) and non-converted (no treatment) libraries.

3.4.2.2. RNA FRAGMENTATION AND FIRST STRAND cDNA SYNTHESIS

The 5 µL samples with DNA-RNA adapter ligated had 4 µL 5X NEBNext first strand synthesis reaction buffer and 1 µL random primers SSS397 (1 µg/µL) added and samples were heated at 94 °C for 11 minutes in a pre heated thermocycler in order to fragment the RNA. After the incubation samples were snap-cooled in an ice-water bath.

To perform the first strand cDNA synthesis, the following was added to each sample: 0.5 µL murine RNase inhibitor, 0.5 µL Actinomycin D (0.1 µg/µL), 1 µL II ProtoScript Reverse Transcriptase and 3.5 µL RNase-free water. Next, the samples were incubated at 25 °C for 10 minutes, followed by a 42 °C incubation for 50 minutes and 15 minutes at 70 °C. 10 µL 1 N NaOH and 10 µL 500 mM EDTA were added to each of the 12 samples and they were incubated at 65 °C for 15 minutes to degrade any remaining RNA present in the samples.

25 µL 1 M Tris pH 7.5 was added to each sample and samples were purified using MinElute columns according to the following protocol: 600 µL PB buffer were added to each reaction and mixed by pipeting. Samples were applied to a MinElute column and centrifuged for 1 minute at 14000 rpm to bind the cDNA. The flow-through was loaded in the columns one more time and the centrifugation step was repeated to increase the final yield of cDNA that was binding the column. Flow-through was then discarded and 750 µL PE buffer was added to wash the columns and samples were centrifuged for 1 minute at 14000 rpm. This wash step was repeated for all samples and the flow-through was discarded. Samples were then centrifuged for another minute at 14000 rpm to eliminate any residual PE buffer. Columns were then placed in clean microcentrifuge tubes and cDNA eluted in 30 µL RNase-free water by 1 minute centrifugation at 14000 rpm. The elution step was then repeated with another 30 µL water for a final volume of 60 µL per sample.

After purification samples were concentrated to 14.1 µL using a speedvac centrifuge.

3.4.2.3. PCR TO ADD INDEXES AND FULL-LENGTH ILLUMINA ADAPTERS

Using the 14.1 µL samples from the previous step, PCR reactions were performed to add the unique index to each sample. In a clean tube the following were combined with the 14.4 µL cDNA sample: 10 µL 5X GC Phusion Buffer, 2.5 µL 10 mM Forward primer (SSS398), 20 µL 5M Betaine, 0.5 µL Phusion polymerase, 0.4 µL dNTP mix (25 mM each) and 2.5 µL10 mM Reverse unique index primer. The samples were then moved to a thermocycler using the following conditions: 98 °C for 3 minutes, 12 cycles with 98 °C for 80 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds and a final extension incubation at 72 °C for 5 minutes. The following table shows which unique reverse index primer was used for each sample:

Table 6: Unique indexes used for 5' end mapping library construction and sample differentiation.

Sample	Unique index	Index sequence
Control 1 - 3 h - non-converted library	SSS 399	CGTGAT
Control 1 - 3 h - converted library	SSS 403	GCCTAA
Control 2 - 3 h - non-converted library	SSS 651	ATTGGC
Control 2 - 3 h - converted library	SSS 653	GGACGG
3.2 mM 1 - 3 h - non-converted library	SSS 654	CTCTAC
3.2 mM 1 - 3 h - converted library	SSS 655	GCGGAC
3.2 mM 2 - 3 h - non-converted library	SSS 656	TTTCAC
3.2 mM 2 - 3 h - converted library	SSS 657	GGCCAC
3.2 mM 1 - 6 h - non-converted library	SSS 658	CGAAAC
3.2 mM 1 - 6 h - converted library	SSS 659	CGTACG
3.2 mM 3 - 6 h - non-converted library	SSS 660	CCACTC
3.2 mM 3 - 6 h - converted library	SSS 661	GCTACC

After the PCR, samples were purified using the Agencourt AMPure XP Magnetic Beads purification following the guidelines as described: the magnetic AMPure Beads were allowed to equilibrate to room temperature and mix 1.2 volumes (60 µL) beads per volume DNA (50 µL) by pipeting at least 10 times. Each sample were then transferred to a clean microcentrifuge tube and

incubated for 15 minutes at room temperature. Tubes were then places in a magnetic stand and incubated for at least 5 minutes until the liquid became clear and the supernatant was discarded. With the tubes still on the magnetic stand, beads were washed 2 times with 200 µL of 80% ethanol for 30 seconds and the supernatant was discarded. Tubes were allowed to air dry with the caps opened for 10 minutes on the magnetic stand. Beads were removed from the magnetic stand and resuspended in 50 µL RNase-free water. After incubate the cell at room temperature for 2 minutes, tubes were place again in the magnetic stand for 5 minutes. The clean supernatant, which contains the libraries, were transferred to clean tubes and DNA concentration was measured using QuBit.

Since the samples concentrations were lower than expected and the QuBit was not able to measure then, samples were concentrated to 14.1 µL using the speedvac centrifuge and more PCR cycles performed to enrich for full-length products. For this PCR reaction, the 14.1 µL samples were mixed with 10 µL 5X GC Phusion Buffer, 2.5 µL 10 µM primer SSS 401, 2.5 µL 10 µM primer SSS 402, 20 µL 5 M Betaine, 0.4 µL 25 mM each dNTP mix and 0.5 µL Phusion polymerase by pipeting. Followed the same condition as the previous PCR, a further 8 cycles were performed in the samples by using a thermocycler. Samples were then purified again using the AMPure purification procedure as previously stated with the difference that instead of 1.2 volumes of beads, it was used 1.8 volumes (90 µL) of beads.

DNA concentrations were measured by QuBit and libraries were stored at -20 °C until the start of the sequencing procedure. The following figure shows an overall procedure for the 5' end mapping library construction.

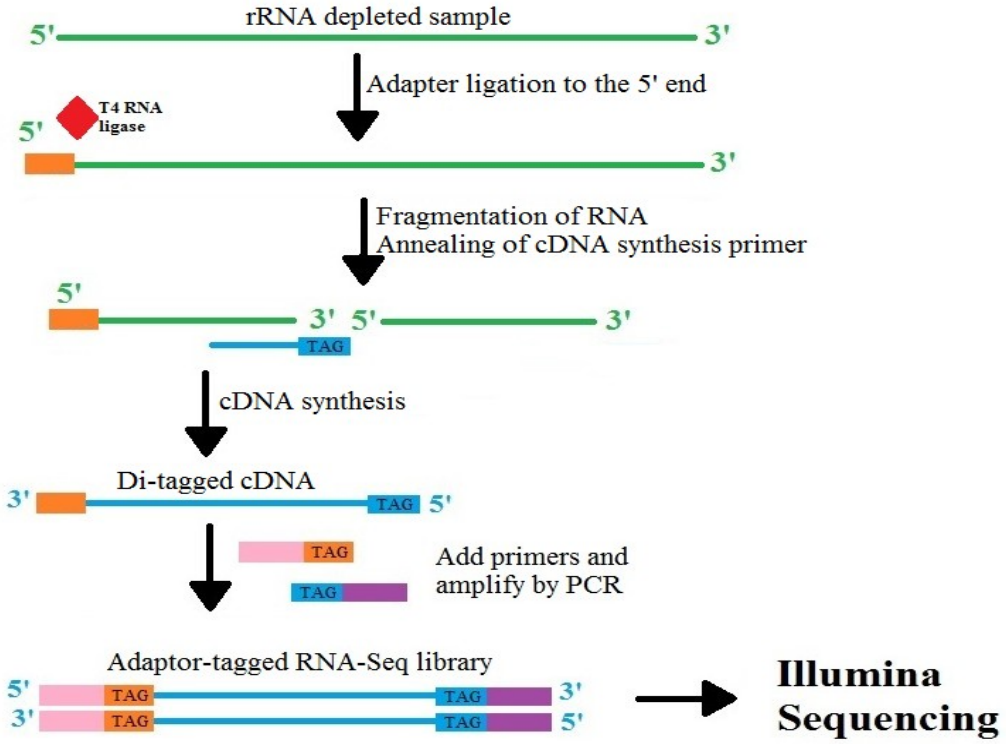

Figure 10: Overview of the oxidative stress 5' end mapping library preparation.

3.5. RNASE J KNOCKOUT CONSTRUCTION AND COMPLEMENTATION

Figure 11 shows a graphic representation of the procedures that discussed in the next sections.

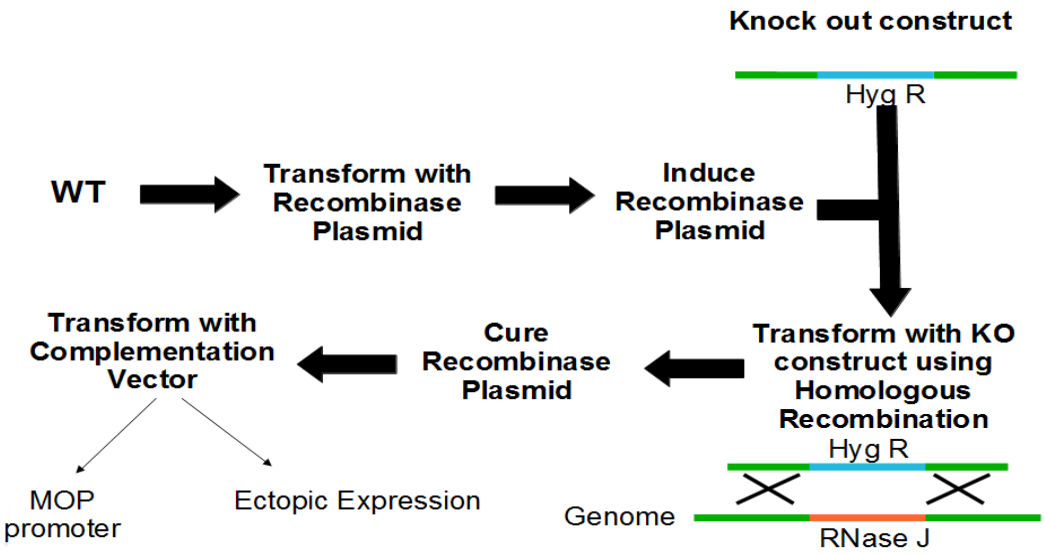

Figure 11: Overview of KO construction and complementation.

3.5.1. KNOCKOUT CONSTRUCTION

In order to knockout RNase J homologous recombination was used to replace the RNase J coding sequence with a hygromicin resistence gene. The RNase J sequence and also its neighbor upstream and downstream sequences were retrieved from the Smegmalist website.

Using SeqBuider™ software for DNA sequence visualization primer sets were constructed to amplify a region with 500 nucleotides upstream (primer SSS 589 and SSS 590) and 500 nucleotides downstream (primers SSS 591 and SSS 592) of the RNase J coding sequence. Genomic DNA extracted from *M. smegmatis* strain mc^2 155 was used as template for these PCRs.

To amplify the hygromicin resistance gene, the plasmid pMV762 was used a template with primers SSS 584 and SSS 583. A PCR using primers SSS 598 and SSS 583 was then performed to connect the upstream flank with the hyg resistance gene and, last, a PCR using primers SSS 589 and SSS 592 was performed to connect together upstream flank, hyg resistance gene and downstream flank to form the final construct. For these PCRs, Q5 DNA polymerase was used due to its low error rate and the primers Tm and annealing temperatures were calculated using the NEB Tm calculator available at http://tmcalculator.neb.com/#!/.

Mycobacteria in general have a low DNA uptake and a low homologous recombination rate, which makes challenging to manipulate their genomes (Van Kessel and Hatfull, 2008). In order to increase recombination rates a strategy known as recombineering can be used, which consists in use a plasmid containing mycobacteriophage recombination genes (Van Kessel and Hatfull, 2008).

Once enough construct (around 2 µg DNA) was made, we used the following protocol to prepare *M. smegmatis* competent cells using a WT strain which was previously tranformed with a recombinase plasmid pJV53 (Van Kessel and Hatfull, 2008).

50 mL of liquid cultures were grown overnight with OD_{600} between 0.6 to 1 with 50 µL the recombinase plasmid inducer (inducer solution was made by adding 0.8 µL Isovaleronitrile in % mL DMSO). Culture was transferred to conical tubes and centrifuged for 5 minutes at room

temperature using 4000 rpm. The supernatant was discarded and the pellet washed with 50mL 10% glycerol and centrifuged for 10 minutes at room temperature using 4000 rpm. Supernatant was discarded and the wash step repeated one more time.

The pellet was then ressuspended in 5ml of 10% glycerol and centrifuged again using the same conditions as before. The supernatant was once again discarded and the pellet was ressuspended in 300 µL 10% glycerol. Ressuspended cells were divided in 50 µL aliquots and used for transformation. Aliquots not used immediately were frozen by contact with a solution containing dry ice and 95% ethanol and stored at -80 °C.

Next, following the protocol, an aliquot of competent cells was mixed with the knockout construct (around 2 µg) of interest and transferred to a 2mm electroporation cuvette. Cells were electroporated using *E. coli* settings for the 2 mm cuvette and then transferred to a clean microcentrifuge tube containing 200 µL of 7H9 media. Cells were incubated with shaking at 37 °C for 4 hours to recover from transformation and then plated in 7H10 plates containing the selective antibiotic.

After a few days of incubation, 12 colonies were visible in the plate. They were each picked and cultured in liquid media without any antibiotic in order to facilitate the strains to lose the recombinase plasmid. 100 µL of the cell culture and 1/10 and 1/100 dilutions were then plated on 7H10 agar containing 15% sucrose to select for the cultures that lost the plasmid. The recombinase plasmid has the SacB gene which makes sucrose toxic to the cell and then ideally they will not grown in media containing sucrose.

Using the colonies from the sucrose plates a second selection was done in which the colonies were streaked in plates containing kanamycin (the selection drug for the plasmid), then plates containing sucrose and 7H10 plates with no drug. Cells which lost the plasmid should grow on the sucrose and no drug plates only. The knockout strains were then validated by PCR with primers SSS 593 and SSS 594 with anneal "outside" of the knockout construct and the colonies with

the correct size were sent to sequencing to further check the nucleotide sequence surrounding the integration site.

3.5.2. COMPLEMENTING THE KO STRAIN

The next step was to construct a complementation vector to restore the RNase J activity in the KO strain. A complementation construct was built using the RNase J coding sequence, a synthetic ribosome binding site and the MOP promoter. The construct was amplified using the primers SSS 644 and SSS 645.

Plasmid pJEB402 was cut with Puv II and Hind III restriction enzymes and a Gibson Assembly reaction was performed to combine the vector backbone with the complementation construct as stated in the following protocol:

1.5 µL of vector backbone and 3.5 µL of insert were mixed together in a 0.2 mL tube. Five µL of Gibson assembly 1 step master mix (2x) was added to the sample and it was incubated for 1 hour at 50 °C. 2 µL were used for the *E. coli* transformation and the remaining solution was stored at -20 °C.

A tube of NEB 5- α competent cells was thawed on ice for 10 minutes and had 2 µL of the Gibson Assembly solution added to it. The cells were then placed on ice for 30 minutes without mixing. Next, cells were submitted to a heat shock at 42 °C for 30 seconds and then placed on ice for another 5 minutes. Cells were then resuspended in 200 µL of SOC rich media and incubated with 250 rpm shaking at 37 °C for 1 hour. Last, 20 µL and 200 µL of cells were plated in LB containing kanamycin and incubated overnight at 37 °C.

A few colonies were chosen and grown in liquid LB with kanamycin for a day and then a Miniprep (Qiagen®) reaction was performed following the instructions from the manufacturer in order to isolate the plasmid.

Next competent cells from the *M. smegmatis* KO strain were made and transformation proceeded as stated in section 3.5.1. A transformation with the WT was done following the same

protocol in order to obtain the overexpression strain (WT+J).

To obtain the strains JKO+E and WT+E, RNase J KO and WT strains respectively were transformed using the empty backbone vector before the insertion of RNase J coding sequence (pJEB402).

This particular vector has a promoter called MOP which is a strong synthetic constitutive promoter designed specifically for mycobacterial cells (Fortune *et al*, 2005).

3.6. RNASE J STRAINS PHENOTYPIC ASSAYS

3.6.1. qPCR TO DETERMINE THE RELATIVE EXPRESSION LEVEL

To determine the relative expression level for the WT+E, WT+J, JKO+E and JKO+J strains cells were cultured and had their total RNA extracted as stated in section 3.3.2. followed by a DNase treatment and RNA clean up as stated in sections 3.3.3. and 3.3.4.

After concentrations for the samples were measured, the volume necessary for 1 µg RNA for each sample was calculated and transferred to a clean tube. RNase-free water was used to complete a final volume of 5.25 µL per reaction. This samples were also tested for genomic DNA contamination using a no-RT (reverse transcriptase) control.

0.5 µL 100 mM Tris and 0.5 µL 1 mg/mL random hexamers were added to each sample and mixed by pipeting. Samples were incubated for 10 minutes at 70 °C and snapped-cool in an ice water bath. Next, 2 µL 5X First strand buffer, 0.5 µL 10 mM each dNTP, 0.5 µL 100 mM DTT, 0.25 µL RNase OUT and 0.5 µL Superscriptt III reverse transcriptase was added to each test sample. For the no-RT samples, 0.5 µL RNase-free water was added instead of Superscriptt III reverse transcriptase. All samples were then incubated at 25 °C for 5 minutes, 50 °C for 2 hours and 70 °C for 15 minutes.

After the incubation, samples had 5 µL 0.5 M EDTA and 5 µL NaOH added and were incubated for 15 minutes at 65 °C. Next, 12.5 µL 1 M Tris was added to each sample and then were

purified using Quiagen MinElute columns following the procedure as stated:

500 µL PB buffer was added to each sample and mixed by pipeting. Samples were transferred to labeled MinELute columns and centrifuged at 14000 rpm for 1 minute to bind the cDNA to the columns. Flow-through was discarded and 750 µL PE buffer was added to wash the samples. Samples were centrifuged at 14000 rpm for 1 minute and flow-through discarded. Samples centrifuged again at 14000 rpm for 1 minute to eliminate any residual buffer from the columns. Columns were transferred to clean microcentrifuge tubes and 50 µL EB buffer added to elute the cDNA. Columns were centrifuged at 14000 rpm for 1 minute and cDNA concentrations were measured using a Nanodrop. The samples without RT (no-RT) should have significant lower concentration than test samples to corroborate that there is low to none genomic DNA contamination. Samples were stored on -20 °C until use.

Meanwhile, IDT's primer quest tool was used to design three set of primers to amplify RNase J coding sequence retrieved from Smegmalist. Primers sets were then tested using WT strain cDNA and comparing with validation primers JR273 and JR274. WT cDNA was diluted to 200 pg/µL, 40 pg/µL, 8 pg/µL and 1.6 pg/µL using RNase-free water. Primer sets, on the other hand, were diluted together with their pairs using RNase-free water to a final concentration of 2.5 mM each primer. Each cDNA concentration were tested in technical triplicates by adding to each well of the 96 well PCR plate: 2 µL primer set mix, 10 µL Taq universal SybrGreen Supermix, 6 µL water and 2 µL cDNA. Water controls were also performed to address amplicon contamination in the water stock. The plate with the final samples was then ran in an Applied Biosystems 7500 qPCR machine using the following cycle conditions: 50 °C for 2 minutes, 95 °C for 10 minutes and 40 cycles using 95 °C for 15 seconds and 61 °C for 1 minute. The machine was set to standard curve assay type with 20 µL per reaction and the melt curve was included to address quality of the primers annealing. Data for the primers tested and their efficiency values can be found in Appendix 7.2. For analysis proposes a C_T threshold of 0.2 was used and efficiency was

calculated based on the equation 10^(-1/slope) – 1. Based on the data analysis, the primer set SSS706-707 was chosen to perform the relative expression qPCR.

To address the relative expression for the WT+E, WT+J, JKO+E and JKO+J strains, 40 pg cDNA per reaction was used. For these experiments were used 3 biological replicates per strains.

Samples cDNA were diluted to 20 pg/µL using RNase-free water and for each well of the 96 well PCR plate was added 2 µL 2.5 mM each primer mix, 10 µL Taq universal SybrGreen Supermix, 6 µL water and 2 µL cDNA. For the water control samples the 2 µL cDNA were substituted by RNase-free water. Cycling conditions and machine settings were the same as used for testing the primers, with the exception that the melt curve was not included for this reaction. For data analysis the Δ_{CT} method was used to compare the C_T obtained with the SSS706-707 primers and the C_T obtained with the validation primers (JR274-274). Relative expression level was then determined by the formula: $2^{-\Delta CT}$.

3.6.2. MINIMAL INHIBITORY CONCENTRATION

Late log OD_{600} cultures (between 0.6 and 1) were diluted to OD_{600} 0.002 and set aside until use. Meanwhile the antibiotic of interest was dissolved in 7H9 in a concentration 4 times higher than the higher concentration to be tested on the plate.

A 96 well plate was obtained and 100 µL of drug-free 7H9 liquid media was added to each well. Next, 100 µL of the previously dissolved drug was added to the top row of the plate and mixed well by pipeting. A 2 fold serial dilution was then performed so the next row would always have half as much drug concentration than the previous row. For the last row when the serial dilution took place, 100 µL of media was discarded.

100 µL of the bacterial stock previously diluted was then added to each well and mixed by pipeting. Plate was covered with a breathable film and incubated for 24 h at 37 °C with 125 rpm shaking. Once the first 24 h incubation was done, a resazurin solution with 0.02% (w/v) in destilled

water was prepared and filter sterilized and 20 µL of this solution was added to each well of the test plate. After resazurin was added, the plate was covered with breathable film and incubated for 20 to 24 h at 37 °C with 125 rpm shaking.

The resazurin when its oxidized form is a blue dye and turns pink when it is reduced to resorufin, which is a pink dye, when viable cells are grown aerobically (Taneja and Tyagi, 2007).

The minimal inhibitory concentration (MIC) is defined as the lowest drug concentration were no change in resazurin color occurred. The following figure shows an overview of the MIC assay procedure:

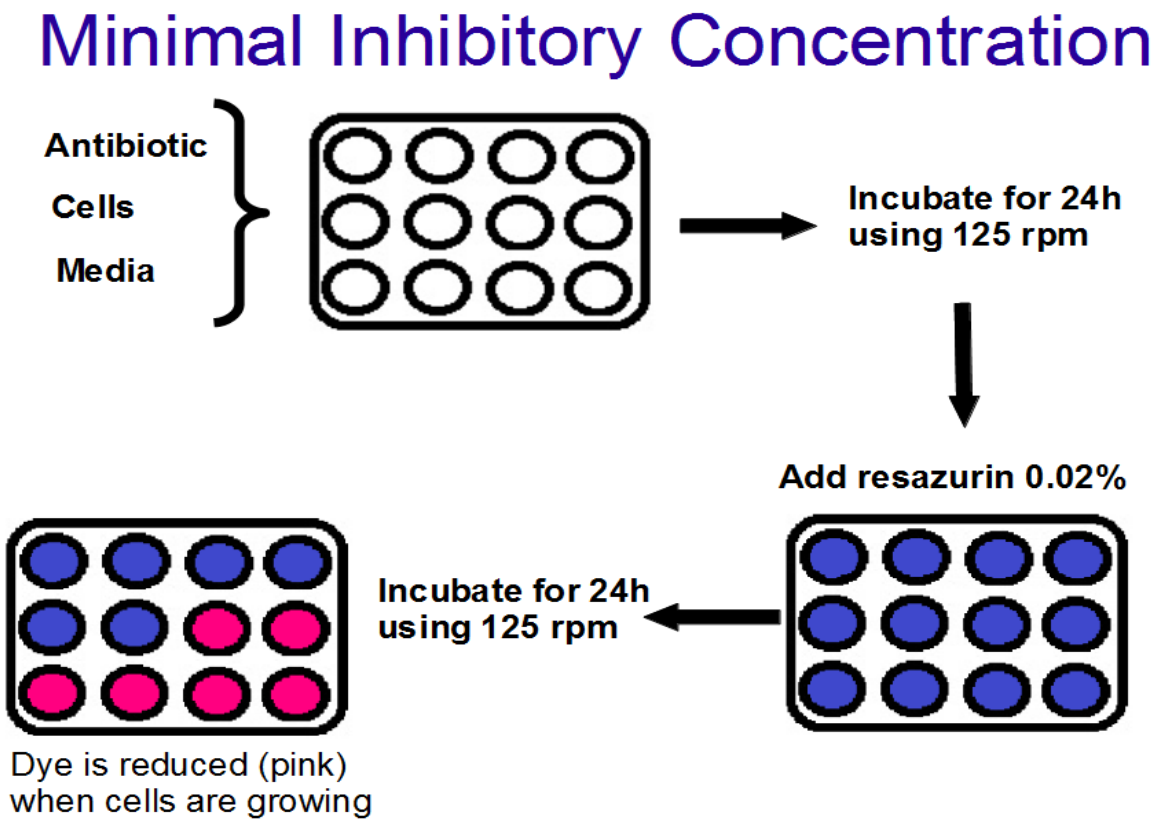

Figure 12: Overall procedure to run a MIC (Minimal Inhibitory Concentration) assay using a well plate.

3.6.3. ANTIBIOTIC KILLING ASSAYS

Cells where grown in 7H9 liquid media until reach late log phase (around OD_{600} 0.6 - 1) and then adjusted to OD_{600} 0.1. 3 mL of media were added to each 15 mL conical tube and the drug in the tested concentration was added to half of then. Tubes were then incubated for 24h at 37 °C with 250 rpm shaking. In parallel, 200 µL of the OD_{600} 0.1 samples were used to performed a 10 fold serial dilution and 5 µL of all dilutions were spotted in squared shaped 7H10 plates in triplicates and incubated at 37 °C for 3 days. These plates are addressed as input CFU.

After the 24 h incubation the tubes are removed from the incubator and centrifuged at 4000 rpm for 5 minutes at room temperature. The supernatant is discarded and pellet is ressuspended in 3 mL of drug-free media to wash out the drug from the cells. Using 200 µL of the ressuspended cells, a 10 fold serial dilution was performed and 5 µL of all dilutions were spotted in squared shaped 7H10 plates in duplicate and incubated at 37 °C for 3 days. This plates are addressed as output CFU.

After the 3 days of incubation, colonies were counted and duplicates averaged. To address the amount of killing the ratio output CFU over input CFU was calculated and the 3 biological replicates averaged.

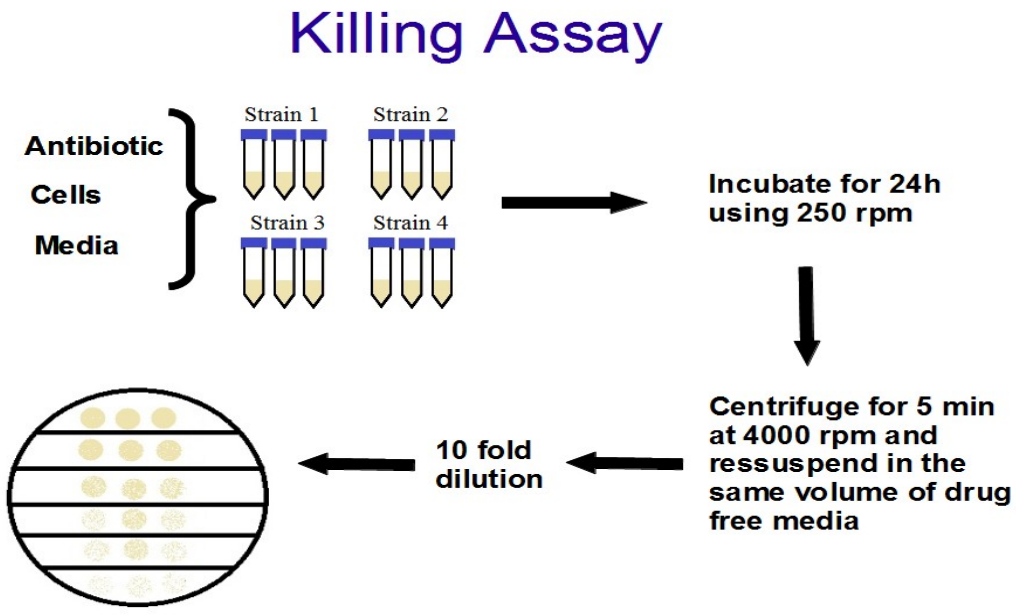

Figure 13: Overall procedure to perform a antibiotic killing assay using 15 ml conical tubes.

3.6.4. OXIDATIVE STRESS GROWTH CURVE

Cells were thawed from -80 °C and grown for a day in 7H9 liquid media containing ADS supplement and 20 µg/mL Kanamicin. Cultures were then centrifuged at 4000 rpm for 5 minutes at room temperature and the supernatant discarded twice to wash off the drug and cells were grown in 7H9+ADS overnight until they reach late log OD_{600} (around 0.6 - 1).

The next morning cultures were diluted to OD_{600} 0.15 and cultured in biological triplicates with 0 and 3.2 mM H_2O_2. OD_{600} measurements were taken every 3 h for the first 9 h and at 29 or 30 h.

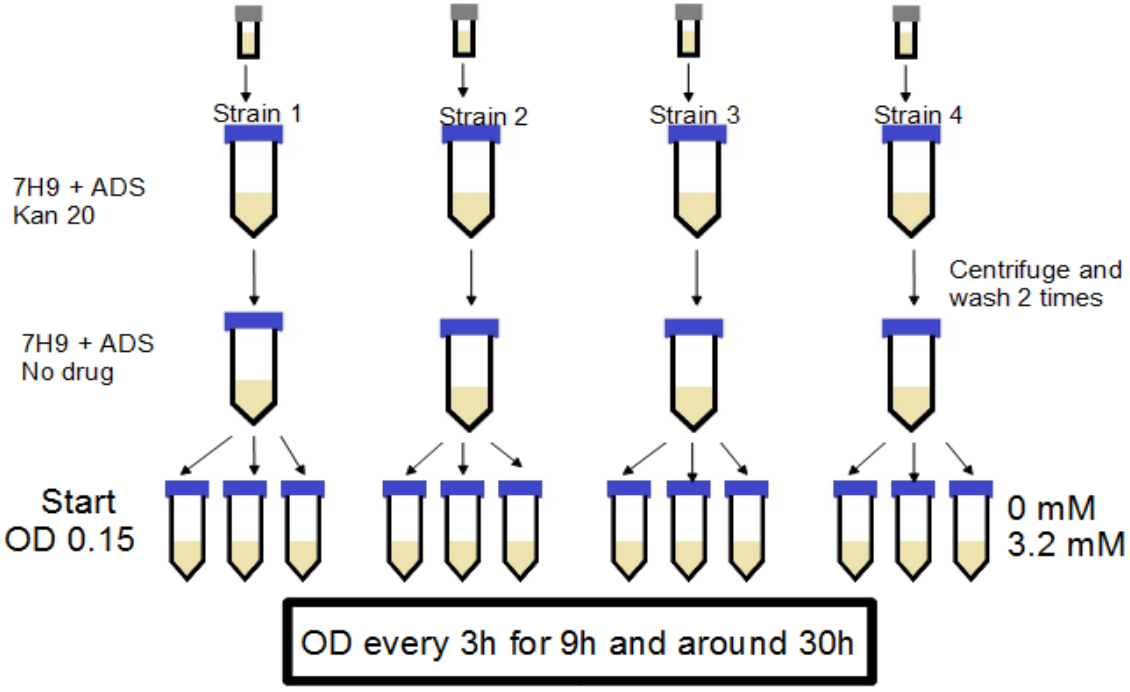

Figure 14: Overview of an Oxidative stress growth curve experiment.

3.7. RNASE J OVEREXPRESSION STRAIN EXPRESSION LIBRARY PREPARATION

Using the 7H9 media, WT+E and WT+J strains were thawed from the -80 °C frozen stock and cultured for 1 day with 20 µg/mL Kananicin. On the next day cells were centrifuged 2 times at 4000 rpm for 5 minutes and ressuspended in 5 mL drug free media and diluted to grown overnight

in triplicate at 37 °C until reach OD of approximate 0.8.

After total RNA extraction of the samples as previously stated on section 3.3.2., an RNA clean up procedure were performed in the cells followed by a DNase treatment and another RNA clean up (complete procedure is shown in sections 3.3.3. and 3.3.4.

For the libraries construction only 2 of the 3 replicates of each strain was used. The samples used for this assay can be seen on the following table:

Table 7: RNase J strains samples used for library construction

Condition	Replicates
WT+E	1 and 2
WT+J	1 and 2
JKO+J	1 and 2

3.7.1. rRNA REMOVAL

For this procedure the non-magnetic Ribo-Zero™ rRNA Removal Kit for gram-positive bacteria from Epicentre was used following the manufacture recommendations with some minor modifications as stated in the sequence:

The microspheres were allowed to to equilibrate to room temperature and then mixed by vortexing for 20 seconds. 65 µL of microspheres were transferred into a clean 2 mL wash tube and centrifuged at 15000 rpm for 3 minutes at room temperature and the supernatant discarded. The microspheres were then washed with 130 µL of wash solution and mixed by vortexing to ressuspend the spheres. Tubes were centrifuged again at 15000 rpm for 3 minutes at room temperature and the supernatant discarded.

65 µL of Ressuspension solution was added and the spheres were ressuspended by vortexing. 1 µL of RiboGuard RNase Inhibitor was added to each tube and the tubes were stored at room temperature until use.

Meanwhile, the RNA samples were treated by adding in a 0.2 mL tube 4 µL Ribo-zero reaction buffer, 5 µg total RNA sample, 9 µL Ribo-zero rRNA removal solution and RNase-free

water to complete 40 µL. Samples were incubated in a thermocycler at 68 °C for 10 minutes and at room temperature for 15 minutes.

The RNA samples were then added to the ressupended microspheres solution and mixed by pipetting at least 15 times and vortexing at medium speed. The tubes were then incubated at room temperature for 10 minutes with vortex mixing for 5 seconds every 3 to 4 minutes. After the incubation period was over the samples were mixed by vortexing for 5 seconds and placed on a heat block at 50 °C for 10 minutes. The RNA-microspheres solutions were transferred to a microsphere removal unit and centrifuged at 15000 rpm for 1 minute at room temperature. The eluate containing the rRNA depleted samples was then purificated using a ethanol precipitation protocol as shown.

Samples had their volumes adjusted to 180 µL with RNase-free water and 18 µL of 3 M sodium acetate and 2 µL of glycogen (10 mg/mL) were added to each sample and mixed by vortexing. Following, 600 µL ice-cold 100% ethanol was added to each sample and mixed gently and the samples were placed in -20 °C for at least 1 hour.

After the 1 hour incubation samples were centrifuged at 15000 rpm for 30 minutes at 4 °C and the supernatant was discarded. The pellet was washed twice with ice-cold 70% ethanol and centrifuged at 15000 rpm at 4 °C for 5 minutes. After carefully remove the supernatant of the second wash, samples were centrifuged for extra 30 seconds to collect any residual ethanol and any remaining supernatant was discarded. The samples were allowed to stay at room temperature for 5 minutes and the pellet was dissolved in 20 µL RNase-free water. If not used immediately, samples were stored in -80 °C.

2 µL of each rRNA depleted sample were sent to Bioanalyzer analysis to address the quality and the rRNA content of the samples.

3.7.2. RNA FRAGMENTATION AND cDNA SYNTHESIS

The library construction for these samples as described in this section and in section 3.7.3. and 3.7.4. were performed using the NEXTflex™ Rapid Directional qRNA-Seq™ Kit from Bioo Scientific Corp. with some modifications as following:

4 µL from each of the rRNA samples were combined with 10 µL RNase-free water and 5 µL NEXTFlex RNA fragmentation buffer and mixed thoroughly by pipetting. Samples were then incubated at 85 °C for 6 minutes to fragment the RNA and then placed in a ice-water bath to stop the reaction.

For the first cDNA strand synthesis 1 µL NEXTFlex first strand synthesis primer was added to the fragmented RNA and incubated for 5 minutes at 65 °C. After the tubes were placed in a ice-water bath to stop the reaction, 4 µL NEXTFlex directional first strand buffer mix and 1 µL NEXTFlex rapid reverse transcriptase was added to each sample and after mixed by pipetting, samples were incubated for 10 minute at 25 °C followed by 50 minutes at 50 °C and 15 minutes at 70 °C.

For the second cDNA strand synthesis 25 µL NEXTFlex directional second strand synthesis mix were added to each samples and after mixed by pipetting samples were incubated for 1 hour at 16 °C.

The samples were then purified using the Agencourt AMPure XP Magnetic Beads following the guidelines as stated in the sequence: the magnetic AMPure Beads were allowed to equilibrate to room temperature and 90 µL beads were mixed to each cDNA (50 µL) sample by pipeting. Each sample were then transferred to a clean microcentrifuge tube and incubated for 5 minutes at room temperature. Tubes were then places in a magnetic stand and incubated for at least 5 minutes until the liquid became clear and the supernatant was discarded. With the tubes still on the magnetic stand, beads were washed 2 times with 200 µL 80% ethanol for 30 seconds and the supernatant was discarded. Tubes were removed from stand and allowed to air dry with the caps opened for 2

minutes. Beads were removed from the magnetic stand and ressuspended in 17 µL RNase-free water by pipetting. After incubate the tubes at room temperature for another 2 minutes, tubes were placed again in the magnetic stand for 5 minutes. 16 µL of the clean supernatant were transferred to clean tubes stored at -20 °C until use.

3.7.3. ADENYLATION AND ADAPTER LIGATION

16 µL of each of the purified second strand synthesis product were mixed with 4.5 µL NEXTFlex Adenylation mix by thoroughly pipetting and incubated at 37 °C for 30 minutes followed by 5 minutes at 70 °C. Samples were then placed on ice and had 27.5 µL NEXTFlex ligation mix and 2 µL NEXTFlex directional qRNA molecular labels added to each sample. Samples were then mixed by pipetting and incubated for 10 minutes at 30 °C to perform the adapter ligation to the samples. After the incubation period, samples were purified twice using the Agencourt AMPure XP Magnetic Bead method.

The magnetic AMPure Beads were allowed to equilibrate to room temperature and 40 µL beads were mixed to each cDNA (50 µL) sample by pipeting. Each sample were then transferred to a clean microcentrifuge tube and incubated for 5 minutes at room temperature. Tubes were then places in a magnetic stand and incubated for at least 5 minutes until the liquid became clear and the supernatant was discarded. With the tubes still on the magnetic stand, beads were washed 2 times with 200 µL of 80% ethanol for 30 seconds and the supernatant was discarded. Tubes were removed from stand and allowed to air dry with the caps opened for 2 minutes. Beads were removed from the magnetic stand and ressuspended in 51 µL RNase-free water by pipetting. After incubate the tubes at room temperature for another 2 minutes, tubes were placed again in the magnetic stand for 5 minutes and 50 µL of the clean supernatant were transferred to clean tubes.

The second purification step was then repeated the same way as stated above and the samples were eluted in a final volume of 34 µL of RNase-free water to recover 33 µL in the last

step. Samples were stored in -20 until use.

3.7.4. PCR AMPLIFICATION

For each sample from section 3.7.3., 1 µL NEXTFlex Uracil DNA glycosylase was added and samples were incubated at 37 °C for 30 minutes followed by an incubation at 98 °C for 2 minutes. Samples were then transferred to ice to stop the reaction and 12 µL NEXTFlex PCR master mix, 2 µL NEXTFlex qRNA-Seq Universal forward primer and 2 µL NEXTFlex Index reverse primer were added to each tube. Samples WT+E 1 and WT+J 1 were mixed by pipetting then incubated in a thermocycler using the stated conditions: 98 °C for 2 minutes followed by 12 cycles with 98 °C for 30 seconds, 65 °C for 30 seconds and 72 °C for 60 seconds and a final extension of 4 minutes at 72 °C. For samples WT+E 2 and WT+J 2 the same PCR reaction was performed but instead 12 cycles, it was used 16 cycles. Finally, samples were purified using the Agencourt AMPure XP Magnetic Bead method as showed in section 3.7.3. with the difference that the final samples were eluted in 17 µL of RNase-free water to recover 16 µL in the last step of the second purification. The libraries concentrations were measured using QuBit and they were stored in -20 °C until sent for sequencing. The following table shows which of the unique indexes used for each sample:

Table 8: Unique indexes used for WT+E, WT+J and JKO+J libraries

Sample	Unique index
WT+E rep 1	Primer Index 1 from kit
WT+J rep 1	Primer Index 2 from kit
JKO+J rep 1	Primer Index 4 from kit
WT+E rep 2	SSS 786
WT+J rep 2	SSS 787
JKO+J rep 2	SSS789

Since sample WT+J 1 had a low final concentration, 4 extras cycles were performed using

the following protocol: sample was concentrated using a speedvac to a final volume of 14.1 µL and then mixed by pipetting with 10 µL 5X GC Phusion Buffer, 2.5 µL 10 mM primer SSS 401, 2.5 µL 10 mM primer SSS 402, 20 µL 5M Betaine, 0.5 µL Phusion polymerase and 0.4 µL dNTP mix (25 mM each). The sample was then moved to a thermocycler using the following conditions: 98 °C for 3 minutes, 4 cycles with 98 °C for 80 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds and a final extension incubation at 72 °C for 5 minutes.

WT+J 1 sample was then purified using the Agencourt AMPure XP Magnetic Beads following the guidelines as stated in section 3.7.2. Finally, concentrations were measured using QuBit and sample was stored in -20 °C until sent for sequencing.

The following figure shows an overview of the library construction procedure:

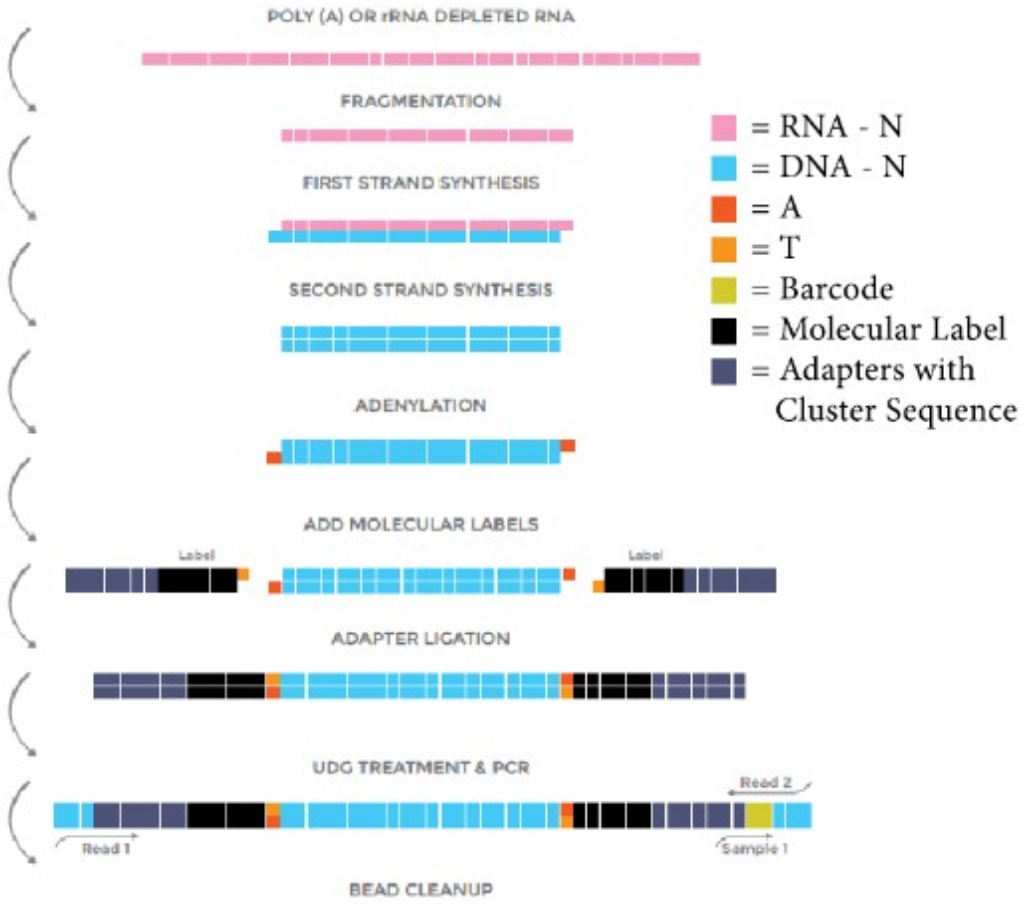

Figure 15: Overview of WT+E and WT+J strains expression library construction. Modified from: ©Bioo Scientific Corp. NEXTflex™ Rapid Directional qRNA-Seq™ Kit Manual. 2014. V14.09

3.8. DATA ANALYSIS FOR LIBRARIES

3.8.1. EXPRESSION LIBRARIES

After next generation sequencing using the Illumina® platform the reads were aligned back to *M. smegmatis* genome using the software Rockhopper (Tjaden, 2015; McClure *et al,* 2013) available at: http://cs.wellesley.edu/~btjaden/Rockhopper/.

After alignment the conditions were the RPKM values were 0 were substituted by value 1 and ratio between 2 different conditions was calculated. Finally the $\log_2$ RPKM ratios for each comparison were calculated and the genes were identified where the change was equal or greater than 2 fold ($\log_2$ values equal or greater than 1 for upregulated genes and $\log_2$ values equal or smaller than -1 for downregulated genes). In this list of genes the ones considered statistically differentially expressed were the ones with adjusted p-value below 0.05.

After the alignment several genes showed as "hypothetical proteins" so in order to find the function of those genes, their amino acid sequences were entered in BLAST (Altschul *et al*, 1990) and Phyre 2 (Kelley *et al*, 2015) software. This way most of the differentially expressed genes functions could be addressed.

Next, in order to find the pathways that were enriched under the tested conditions, the software DAVID (Huang *et al*, 2009) was used.

3.8.2. 5' END MAPPING LIBRARIES

After the sequencing procedure the output for the samples was aligned with *M. smegmatis* genomes. After the alignment only the first nucleotide of each read 1 was analyzed for the libraries. Since there might be a little discrepancy between which is actually the true first nt of a particular transcript since RNA polymerase may not always start the transcription on the same exact nucleotide, normalizations are needed to get rid of the background. Samples were then run into a script to compare each site with its 10 coordinates upstream and downstream. This fist procedure

had the objective to get rid of the background and find the coverage ratio for each position. Following each coordinate were compared with the 4 sites upstream and downstream to find the one with the highest coverage. If this particular site is the one with the highest coverage it was considered the significant one, otherwise the adjacent positions were analyzed until the highest coverage value among the 9 coordinates (the 4 sites upstream, the site in question and the 4 sites downstream) was found.

An arbitrary cutoff of 100 coverage on the converted libraries was used in order to get rid of the background. If a sample had coverage on the converted library lower than 100 for all 3 conditions (no H_2O_2, H_2O_2 3h and H_2O_2 6h) it was excluded from the analysis. The cutoff chosen for this analysis is a little high due to the limitations of the library construction results as will be discussed in the results section.

Next, the $\log_2$ mean of the coverage of the 2 replicates from the converted libraries of one condition were plotted against each other in order to find the differential 5' ends when comparing the 2 conditions in question. The differential 5' ends were considered the outliers of those graphs.

The coordinates found for the outliers were then analyzed using the software Artemis (Rutherford *et al*, 2000) and compared with the expression libraries in order to address if a specific coordinate was a transcription start site or a cleavage site.

4. RESULTS FOR OXIDATIVE STRESS PROFILES

4.1. CULTURES FOR OXIDATIVE STRESS EXPERIMENTS

We sought to identify an experimental condition in which *M. smegmatis* experienced oxidative stress but was not killed. We therefore cultured a WT strain (*Mycobacterium smegmatis* strain mc^2 155) with different concentrations of hydrogen peroxide to address how much stressor the bacteria is able to face and still continue growing.

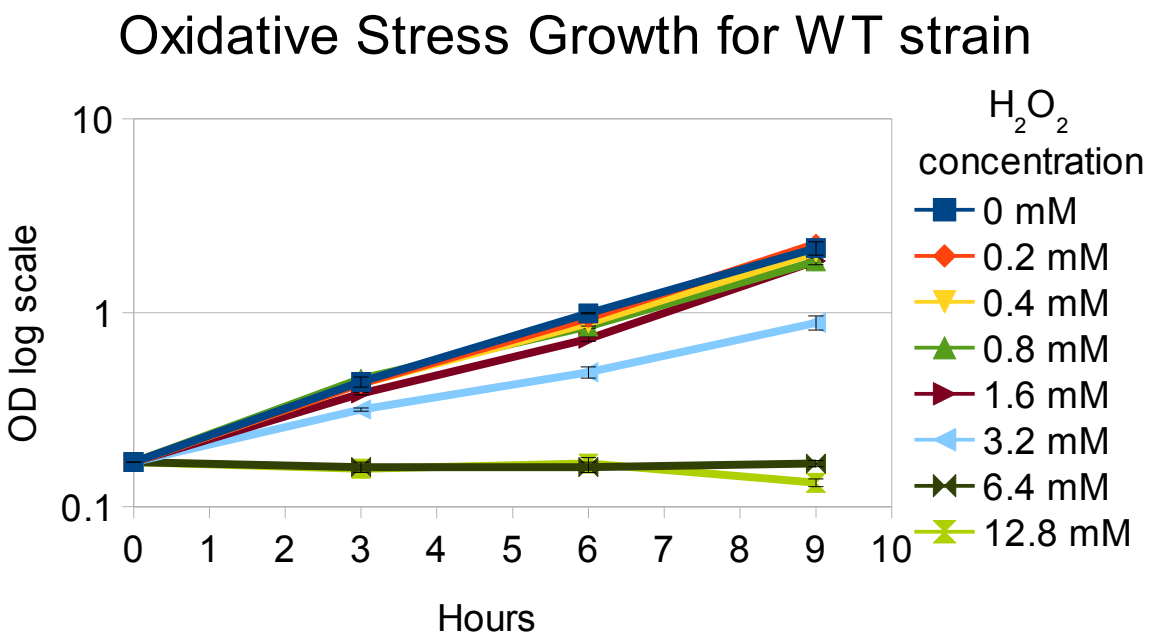

Figure 16: Oxidative stress growth for WT strain. Triplicates cultures with initial OD_{600} 0.15 were cultured at 37 °C with different concentrations of hydrogen peroxide (H_2O_2) and their growth was followed for 9h. Error bars show standard deviation.

Figure 16 shows that different concentrations of hydrogen peroxide differentially impact the overall growth rate. While lower concentrations such as 0.2 and 0.4 mM have almost no apparent effect on cell growth, which were comparable to no exposure of H_2O_2, high concentrations such as 6.4 or 12.8 mM showed no apparent overall growth although is not clear if the cells in those cultures are dead. Li *et al*, 2015 reported that 7 mM hydrogen peroxide for 30 minutes was lethal to *M. smegmatis*, suggesting that, at least for the 12.8 mM concentration, cells are not viable.

Since 3.2 mM hydrogen peroxide impaired cell growth but did not halt it completely, we chose this concentration for preparation of samples for RNA-seq libraries construction.

4.2. OXIDATIVE STRESS EXPRESSION LIBRARIES

To understand how mycobacteria is able to tolerate ROS, we decided to study *M. smegmatis* transcriptome under oxidative stress by constructing RNA-seq libraries. Cells were cultured in liquid media with 3.2 mM hydrogen peroxide or without hydrogen peroxide and harvested after 3 and 6h to perform RNA extraction. Samples then had their rRNA depleted and the remaining mRNA was fragmented. Adapters were connected to both ends of the fragments and a PCR to enrich for fragments containing the full-length adapters was performed in order to provide enough sample for Illumina sequencing.

After sequencing, the software Rockhopper (Tjaden, 2015) was used to align reads to the *M. smegmatis*, normalize reads counts bewteen samples, and perform differential expression analysis. We then calculated the $\log_2$ ratios of RPKM values between pairs of conditions to determine the fold change between those specific conditions. After applying a cutoff of 2 fold expression difference between test sample and control sample and an adjusted p-value of 0.05 or less, lists of upregulated and downregulated genes were generated (tables 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20).

Table 9: Differentially expressed genes under oxidative stress conditions. DNA binding and repair. Numbers in italics are not significant.

DNA binding and repair				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
Rv0629c	MSMEG_1325	exodeoxyribonuclease V subunit alpha RecD	1,58	*1,27*
Rv0630c	MSMEG_1327	exodeoxyribonuclease V subunit beta RecB	1,64	*1,08*
Rv0631c	MSMEG_1328	exodeoxyribonuclease V subunit gamma RecC	1,86	*1,39*
Rv3395c	MSMEG_1620	Thought to be regulated by Rv2720|LEXA.	5,29	4,25
Rv3394c	MSMEG_1622	DNA repair polymerase	5,33	4,06
Rv3370c	MSMEG_1633	error-prone DNA polymerase DnaE2	5,13	3,94
Rv3297	MSMEG_1756	endonuclease VIII	2,87	2,12
Rv3202c	MSMEG_1941	helicase, UvrD/Rep family protein	2,81	1,91
Rv3201c	MSMEG_1943	ATP-dependent DNA helicase	2,12	*1,17*
Rv3062	MSMEG_2277	ATP-dependent DNA ligase LigB	1,51	*1,19*
	MSMEG_2663	HNH endonuclease signature domain	3,94	2,58
Rv2737c	MSMEG_2723	recombinase A RecA	2,45	*1,59*
	MSMEG_2724	recombination regulator RecX	2,73	1,58
Rv2720	MSMEG_2740	LexA repressor	2,85	1,75
	MSMEG_2742	DNA-damage-inducible protein	4,21	2,94
Rv2594c	MSMEG_2943	Holliday junction resolvase RuvC	1,83	*1,26*
Rv2593c	MSMEG_2944	Holliday junction DNA helicase RuvA	1,96	*1,21*
Rv2592c	MSMEG_2945	Holliday junction DNA helicase RuvB	2,04	1,58
	MSMEG_3164	DNA-binding protein	*-0,90*	-1,11
	MSMEG_3292	DNA-binding protein	2,22	*1,58*
Rv1765c / Rv2015	MSMEG_3404	HNH endonuclease domain-containing protein	2,63	2,18
Rv1633	MSMEG_3816	excinuclease ABC subunit B UvrB	2,10	1,64
	MSMEG_3935	Possible integrase	*-1,25*	-1,05
	MSMEG_3957	HNH endonuclease	3,91	*2,58*
Rv2478c	MSMEG_4701	single-stranded DNA-binding protein	*1,00*	1,58
Rv1277	MSMEG_5004	DNA repair exonuclease	2,30	1,79
Rv2100	MSMEG_5026	HNH endonuclease	4,75	3,58
Rv1210	MSMEG_5082	DNA-3-methyladenine glycosylase I (Tag)	1,89	*1,23*
Rv0336 / Rv051	MSMEG_5106	HNH endonuclease	3,46	*2,32*
	MSMEG_5451	alkylated DNA repair protein	2,97	*1,50*
Rv2100/ Rv3776	MSMEG_5583	HNH endonuclease	5,39	4,80
	MSMEG_5876	HNH endonuclease	3,05	*1,72*
	MSMEG_5935	ATP-dependent DNA helicase	1,58	*1,11*
Rv1378c / Rv3074	MSMEG_6025	HNH endonuclease	6,17	5,13
Rv3585	MSMEG_6079	DNA repair protein RadA	1,54	*1,30*
Rv3129	MSMEG_6368	DNA-binding protein	1,49	*0,64*
	MSMEG_6443	DNA polymerase IV DinP	3,46	*2,32*
Rv1128c / Rv1148c / Rv1702c	MSMEG_6573	HNH endonuclease signature domain	4,95	*3,46*

Table 10: Differentially expressed genes under oxidative stress conditions. Membrane proteins. Genes in italics are not significant.

Membrane proteins				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
	MSMEG_0185	MmpL6 protein	2,58	*0,62*
	MSMEG_0520	porin	*-1,46*	-1,14
Rv0309	MSMEG_0635	transpeptidase	*-1,74*	-1,22
Rv0463	MSMEG_0904	membrane protein	*-1,02*	-1,14
	MSMEG_1266	ankyrin	*2,17*	2,58
Rv3482c	MSMEG_1312	chromosome segregation ATPase	2,12	*1,42*
	MSMEG_1315	MscS family protein	*-0,28*	-1,01
	MSMEG_1530	integral membrane protein	1,52	*0,64*
	MSMEG_1680	major pilin protein fimA	*-0,81*	-1,09
	MSMEG_1777	UsfY protein	*-0,29*	-1,12
	MSMEG_1789	Possible membrane protein	*-0,63*	-1,15
	MSMEG_1791	UsfY protein	*-0,63*	-1,25
	MSMEG_1792	Possible membrane protein	*0,08*	-1,35
	MSMEG_2004	arabinose-proton symporter	*-0,37*	-1,48
	MSMEG_2542	Membrane protein	3,58	*2,81*
	MSMEG_2925	permease membrane component	*-0,50*	-1,32
	MSMEG_2946	transmembrane protein	1,58	*0,79*
	MSMEG_3022	transglycosylase associated protein	*-0,44*	-1,05
Rv1477 / Rv1478	MSMEG_3145	secreted cell wall-associated hydrolase	-2,19	*-0,55*
	MSMEG_3255	DoxX subfamily protein	*-0,86*	-1,50
	MSMEG_3439	putative membrane protein	*-0,79*	-2,01
	MSMEG_3541	cytochrome C biogenesis protein transmembrane region	*-0,35*	-1,18
	MSMEG_3593	sensor histidine kinase	*-0,14*	-1,08
	MSMEG_3815	drug efflux membrane protein	1,58	*1,14*
	MSMEG_4030	membrane spanning protein	1,58	*0,26*
	MSMEG_4536	TQXA domain-containing protein	*-1,42*	-1,46
	MSMEG_4689	LprJ	-2,41	*-0,41*
	MSMEG_5011	membrane protein	1,68	*0,61*
	MSMEG_5051	major facilitator superfamily protein	*-1,45*	-1,38
	MSMEG_5187	tetracycline-resistance determinant TetV	1,48	*1,15*
Rv1069c	MSMEG_5278	transmembrane protein	*-1,35*	-1,17
	MSMEG_5540	membrane protein	*-0,64*	-1,58
	MSMEG_5581	integral membrane protein	1,91	*0,88*
	MSMEG_6006	membrane protein	2,27	*1,12*
	MSMEG_6500	membrane protein	*-0,43*	-1,50
	MSMEG_6758	transport integral membrane protein	-2,05	-1,17
	MSMEG_6783	integral membrane protein	3,07	2,23

Table 11: Differentially expressed genes under oxidative stress conditions. Detoxification. Genes in italics are not significant.

Detoxification				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
Rv0187	MSMEG_0224	O-methyltransferase MdmC	1,26	*0,77*
	MSMEG_0447	Peroxiredoxin	2,14	*0,37*
	MSMEG_3560	glyoxalase	*-0,28*	-1,21
	MSMEG_5027	glyoxalase	2,17	*1,81*
	MSMEG_5216	glyoxalase	3,16	2,54
Rv0887c	MSMEG_5680	glyoxalase	2,54	1,95
Rv0801	MSMEG_5827	glyoxalase	1,57	*0,76*
	MSMEG_6232	catalase KatA	*-0,35*	-1,22

Table 12: Differentially expressed genes under oxidative stress conditions. Electron transport chain and some metabolism genes. Genes in italics are not significant.

Electron transport chain				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
	MSMEG_0451	oxidoreductase	*-0,48*	-1,28
	MSMEG_0685	oxidoreductase, molybdopterin-binding subunit	*-0,66*	-1,50
	MSMEG_0686	oxidoreductase	*-0,26*	-1,53
Rv3054c	MSMEG_1016	NAD(P)H-dependent FMN reductase	2,81	2,05
	MSMEG_1782	oxidoreductase, short chain dehydrogenase/reductase	*-0,35*	-1,15
Rv3054c	MSMEG_2296	NAD(P)H-dependent FMN reductase	2,81	2,05
	MSMEG_2651	alkanesulfonate monooxygenase	-1,64	*-0,28*
	MSMEG_3137	oxidoreductase	1,83	*0,93*
	MSMEG_4519	NrtC protein (oxidoreductase activity)	-3,70	-2,70
	MSMEG_5189	oxidoreductase	*-0,36*	-1,30
	MSMEG_6703	N5,N10- methylenetetrahydromethanopterin reductase-related protein	3,32	*2,00*
	MSMEG_6747	oxidoreductase	2,06	*1,32*
	MSMEG_6859	oxidoreductase	2,52	2,17
	MSMEG_6884	NADP oxidoreductase	2,32	1,25
Metabolism				
Rv2991	MSMEG_0048	pyridoxamine 5'-phosphate oxidase	*-1,03*	-1,13
	MSMEG_0147	C-5 sterol desaturase	*0,00*	-1,00
	MSMEG_0208	ribonuclease	*-0,34*	-1,29
	MSMEG_0218	3-demethylubiquinone-9 3-methyltransferase	1,93	*0,68*
	MSMEG_0239	O-acetylhomoserine/O-acetylserine sulfhydrylase	*-1,24*	-1,08
Rv3714c	MSMEG_0256	subunit of E3 ubiquitin ligase	2,55	*1,36*
	MSMEG_0303	alcohol dehydrogenase	1,78	*0,42*
	MSMEG_0338	acyl-CoA dehydrogenase	*-0,74*	-1,08
	MSMEG_0543	pyridine nucleotide-disulfide oxidoreductase	5,73	*1,58*
-	MSMEG_0546	methyltransferase	2,70	2,17
	MSMEG_0669	ADP-glucose pyrophosphorylase	*-0,32*	-1,29
	MSMEG_0670	FAD dependent oxidoreductase	*-0,29*	-1,40
	MSMEG_0684	aldehyde oxidase	*-0,41*	-1,07
	MSMEG_0773	acetyltransferase	2,68	1,77
Rv3714c	MSMEG_0850	subunit of E3 ubiquitin ligase	2,91	*1,32*
Rv0068 Rv0439c	MSMEG_0876	short chain dehydrogenase	1,22	*-0,26*
	MSMEG_1026	acetyltransferase	1,87	1,42
	MSMEG_1031	Ketosteroid isomerase-related protein	*0,80*	1,34
Rv0548c	MSMEG_1075	naphthoate synthase	*-1,25*	-1,17
	MSMEG_1114	short chain dehydrogenase	2,08	1,70
	MSMEG_1225	Possible aromatic acid decarboxylase	3,47	2,11
	MSMEG_1236	Mpr protein	3,00	1,66
	MSMEG_1269	Ser/Thr protein phosphatase	2,77	2,25
	MSMEG_1272	ribosylglycohydrolase	2,42	2,58
	MSMEG_1280	deacetylase	1,92	*1,18*

Table 13: Differentially expressed genes under oxidative stress conditions. Metabolism genes part 2. Genes is italics are not significant.

Metabolism

Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
Rv0751c	MSMEG_1496	3-hydroxyisobutyrate dehydrogenase MmsB	-1,83	*-0,98*
Rv0752c	MSMEG_1497	acyl-CoA dehydrogenase	-2,22	-1,37
Rv0753c	MSMEG_1498	methylmalonate-semialdehyde dehydrogenase MmsA	-2,03	*-0,88*
Rv0458	MSMEG_1543	eptc-inducible aldehyde dehydrogenase	-1,82	*-0,97*
	MSMEG_1546	coenzyme B12-dependent glycerol dehydrogenase small subunit	-2,14	*-0,24*
	MSMEG_1548	propanediol utilization: dehydratase, medium subunit	-2,06	*-0,64*
	MSMEG_1768	flavodoxin	*-0,36*	-1,03
	MSMEG_1771	methylase	*-0,60*	-1,38
Rv3281	MSMEG_1812	acetyl-/propionyl-coenzyme A carboxylase AccE5	*-0,58*	-1,22
	MSMEG_1916	lipoprotein	*-0,57*	-1,70
	MSMEG_1975	amidohydrolase	1,81	1,32
	MSMEG_2079	alcohol dehydrogenase	*-0,90*	-1,30
	MSMEG_2177	fmnh2-utilizing oxygenase	2,22	*1,87*
	MSMEG_2262	hydrogenase-2, small subunit HybA	-1,84	*-0,74*
	MSMEG_2264	peptidase M52, hydrogen uptake protein	*-1,26*	-1,00
	MSMEG_2306	acetyltransferase	1,87	1,42
	MSMEG_2311	Ketosteroid isomerase-related protein	*0,80*	1,34
	MSMEG_2346	phytoene synthase	*-0,39*	-1,54
	MSMEG_2347	phytoene dehydrogenase	*-0,62*	-1,65
Rv2858c	MSMEG_2597	aldehyde dehydrogenase	*-0,58*	-1,00
	MSMEG_2650	acyl-CoA dehydrogenase	-2,81	-1,00
	MSMEG_2791	pyridoxamine 5'-phosphate oxidase	1,97	*1,05*
	MSMEG_2850	cell entry (mce) related family protein	1,49	*0,95*
	MSMEG_2906	LLM class F420-dependent oxidoreductase	1,81	*0,94*
	MSMEG_2920	aldo/keto reductase	1,58	*1,00*
Rv1493	MSMEG_3159	methylmalonyl-CoA mutase	-1,86	*-0,90*
Rv1596	MSMEG_3201	nicotinate-nucleotide pyrophosphorylase	*0,11*	-1,08
	MSMEG_3311	acyl carrier protein	*0,07*	-1,00
	MSMEG_3325	carboxylate-amine ligase	*1,11*	1,32
	MSMEG_3336	hydrolase	1,17	*0,11*
	MSMEG_3384	isoflavone reductase	2,53	1,58
	MSMEG_3385	ketosteroid isomerase	2,65	1,78
	MSMEG_3418	polyketide cyclase	*-0,15*	-1,06
	MSMEG_3555	hydrolase	*0,89*	1,10
	MSMEG_3583	monooxygenase	1,64	*0,70*
Rv3733c	MSMEG_3679	phosphohydrolase	*1,38*	1,38
Rv1728c	MSMEG_3714	glycoside hydrolase	*0,53*	1,58
	MSMEG_3812	acyl-CoA thioesterase	*-0,92*	-1,22
Rv2004c	MSMEG_3942	Possible adenylate kinase	*-1,29*	-1,17
	MSMEG_3962	lactate 2-monooxygenase	1,58	*0,26*
	MSMEG_4027	zinc-containing alcohol dehydrogenase superfamily protein	1,77	*0,49*
	MSMEG_4031	Possible keratin associated protein	1,61	*0,91*
	MSMEG_4032	zinc-binding alcohol dehydrogenase	2,75	2,61
	MSMEG_4063	amidohydrolase	2,36	1,90
	MSMEG_4064	zinc-binding alcohol dehydrogenase	3,12	2,93
Rv3463	MSMEG_4141	LLM class F420-dependent oxidoreductase	2,27	1,28
Rv2122c	MSMEG_4181	phosphoribosyl-ATP pyrophosphatase	*-0,76*	-1,36
	MSMEG_4389	monooxygenase	-2,22	*-0,81*
	MSMEG_4399	Possible alpha-ketoglutarate decarboxylase	3,16	2,02
	MSMEG_4465	cutinase	2,00	*0,49*
	MSMEG_4538	cysteine desulfurase	-2,00	-1,11
	MSMEG_4618	isochorismatase	*-0,21*	-2,04

Table 14: Differentially expressed genes under oxidative stress conditions. Metabolism genes part 3. Genes in italics are not significant.

Metabolism

Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
Rv2495c	MSMEG_4710	branched-chain alpha-keto acid dehydrogenase subunit E2	*-1,58*	-1,00
Rv2500c	MSMEG_4715	acyl-CoA dehydrogenase	*-1,58*	-1,03
Rv2501c	MSMEG_4716	acetyl-/propionyl-coenzyme A carboxylase subunit alpha	*-1,31*	-1,24
Rv0974c Rv2502c	MSMEG_4717	carboxyl transferase domain-containing protein	*-1,46*	-1,17
	MSMEG_4736	polysaccharide pyruvyl transferase	1,29	1,47
	MSMEG_4737	polysaccharide pyruvyl transferase	1,38	1,57
	MSMEG_4742	clavaldehyde dehydrogenase	*-1,32*	-1,12
	MSMEG_5002	hydrolase	1,87	1,22
Rv1140	MSMEG_5188	caax amino protease	*-0,23*	-1,02
Rv3130c	MSMEG_5242	acyltransferase	*-1,38*	-1,09
Rv3129	MSMEG_5243	pyridoxamine 5'-phosphate oxidase	*-0,86*	-1,26
Rv1070c	MSMEG_5277	enoyl-CoA hydratase	*-1,35*	-1,30
	MSMEG_5343	acetyltransferase family protein	*-0,67*	-1,46
	MSMEG_5400	dehydrogenase	*-0,50*	-1,44
	MSMEG_5401	polyketide cyclase	*-0,36*	-1,39
	MSMEG_5404	propionate--CoA ligase	-2,06	-0,94
	MSMEG_5582	glucose-6-phosphate dehydrogenase	3,80	2,66
Rv0927c	MSMEG_5584	short chain dehydrogenase	2,70	*2,00*
Rv0926c	MSMEG_5586	diacylglycerol kinase	1,44	*0,65*
	MSMEG_5590	carboxylate-amine ligase	*-0,15*	-1,26
Rv0905	MSMEG_5639	enoyl-CoA hydratase	*-0,58*	-1,04
Rv1482c / Rv3517	MSMEG_5643	subunit of E3 ubiquitin ligase	2,77	1,87
	MSMEG_5732	monooxygenase	1,17	*0,40*
Rv2590	MSMEG_5739	long-chain fatty-acid--CoA ligase	2,20	*1,38*
	MSMEG_5753	serine/threonine specific protein phosphatase	2,29	1,39
	MSMEG_5799	nucleoside-diphosphate-sugar epimerase	*-0,21*	-1,42
	MSMEG_5826	pyruvate decarboxylase	*-0,77*	-1,56
Rv0761c	MSMEG_5866	alcohol dehydrogenase	*-1,41*	-1,17
Rv1482c / Rv351	MSMEG_5917	subunit of E3 ubiquitin ligase	3,70	*2,32*
	MSMEG_5956	NAD dependent epimerase/dehydratase	*0,76*	1,47
Rv3607c	MSMEG_6102	dihydroneopterin aldolase	*-0,76*	-1,02
Rv1254	MSMEG_6230	acyltransferase	*-0,38*	-1,24
Rv3700c	MSMEG_6246	pyridoxal-phosphate-dependent transferase	2,16	*1,45*
Rv3702c	MSMEG_6248	class II glutamine amidotransferase	2,28	*1,51*
Rv3703c	MSMEG_6249	iron(II)-dependent oxidoreductase EgtB	2,27	1,70
Rv3704c	MSMEG_6250	glutamate--cysteine ligase	2,09	*1,32*
Rv3734c Rv3740c	MSMEG_6322	bifunctional wax ester synthase/acyl-CoA diacylglycerol acyltransferase	*-0,51*	-1,32
Rv1758	MSMEG_6354	serine esterase, cutinase	*-0,56*	-1,45
	MSMEG_6454	LLM class F420-dependent oxidoreductase	1,51	*0,90*
	MSMEG_6482	putative O-methyltransferase	*0,00*	-1,14
	MSMEG_6512	acyl-CoA dehydrogenase	*-0,90*	-1,39
	MSMEG_6610	lipoprotein	*-0,17*	-1,10
	MSMEG_6616	S-(hydroxymethyl)glutathione dehydrogenase	*-0,54*	-1,10
Rv1130	MSMEG_6645	2-methylcitrate dehydratase	4,39	*3,32*
	MSMEG_6646	methylisocitrate lyase PrpB	3,74	2,42
Rv1131	MSMEG_6647	citrate synthase	3,74	2,06
Rv2967c	MSMEG_6648	pyruvate carboxylase Pyc	3,17	*1,81*
	MSMEG_6664	methylenetetrahydrofolate reductase	*-0,38*	-1,04
Rv3696c	MSMEG_6759	glycerol kinase	*-2,02*	-1,19
Rv3302c	MSMEG_6761	glycerol-3-phosphate dehydrogenase	*-1,68*	-1,07
	MSMEG_6768	halogenase	*-0,29*	-1,06
	MSMEG_6781	lipase	3,39	*2,17*
	MSMEG_6858	epoxide hydrolase 1	2,49	1,93
	MSMEG_6885	MmcI protein	1,77	*1,26*

Table 15: Differentially expressed genes under oxidative stress conditions. ATP, metal and RNA binding proteins; stress response and secreted proteins. Values in italics are not significant.

Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
ATP binding				
	MSMEG_1013	ATP-binding protein	1,72	*0,65*
	MSMEG_2293	ATP-binding protein	1,72	*0,65*
	MSMEG_6612	ATPase, MoxR family protein	*-0,26*	-1,14
Metal binding				
	MSMEG_0637	iron-sulfur binding oxidoreductase	*-0,34*	-1,32
	MSMEG_1976	metal-sulfur cluster biosynthetic enzyme	*1,00*	1,49
	MSMEG_2415	hemerythrin HHE cation binding protein	*-0,44*	-1,32
Rv2578c	MSMEG_2990	Fe-S protein, radical SAM family protein	5,19	3,58
Rv1909c	MSMEG_3460	ferric uptake regulation protein	1,35	*0,52*
Rv1284	MSMEG_4985	carbonic anhydrase	*-1,14*	-1,10
	MSMEG_5936	hemophore-related protein	*-0,06*	-1,17
Rv3677c	MSMEG_6190	metallo-beta-lactamase	*-0,60*	-1,46
RNA binding				
Rv3296	MSMEG_1757	DEAD/DEAH box helicase	1,64	*1,17*
	MSMEG_1790	Rho termination factor	*-0,67*	-1,42
	MSMEG_3170	RNA-binding protein	*-0,58*	-1,18
	MSMEG_6005	RNA polymerase subunit sigma-70	2,30	1,76
Stress response				
	MSMEG_1770	general stress protein CsbD	*0,29*	-1,58
	MSMEG_3242	starvation-inducible DNA-binding protein	3,74	2,32
	MSMEG_3582	ATP-dependent protease La (lon)	1,93	1,49
Rv2031c	MSMEG_3932	heat shock protein hspX	*-1,70*	-2,12
	MSMEG_5733	universal stress protein family protein	*-0,91*	-1,07
Secreted protein				
	MSMEG_5355	PPE family protein	1,95	2,13
	MSMEG_6144	PE family protein	*-0,42*	-1,22

Table 16: Differentially expressed genes under oxidative stress conditions. Transcription regulators.

Transcription regulator				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
	MSMEG_1303	transcriptional regulator	1,69	*1,05*
Rv0691c	MSMEG_1420	mycofactocin system transcriptional regulator	-2,58	-1,26
	MSMEG_1953	transcription factor WhiB	1,77	*1,51*
	MSMEG_2125	glycerol operon regulatory protein	*-0,41*	-1,44
	MSMEG_2905	transcriptional regulator	1,83	*0,76*
	MSMEG_3180	MerR family transcriptional regulator	*0,63*	1,35
	MSMEG_3447	two-component system response regulator	2,79	2,08
	MSMEG_4300	transcription regulator AmtR	*-0,17*	-1,24
	MSMEG_5186	HicB family protein	3,17	*1,81*
	MSMEG_5542	transcriptional regulator	*-0,62*	-1,44
	MSMEG_6677	LysR family transcriptional regulator	1,38	1,38
	MSMEG_0188	MarR family transcriptional regulator	1,18	*-0,22*
	MSMEG_0268	GntR family transcriptional regulator	*-1,08*	-1,26
	MSMEG_0448	MarR family transcriptional regulator	*-0,82*	-1,14
	MSMEG_2017	MerR family transcriptional regulator	-6,38	*-0,52*
Rv3744	MSMEG_5405	ArsR family transcriptional regulator	1,39	1,22
	MSMEG_5731	GntR family transcriptional regulator	1,62	*0,87*
Rv0081	MSMEG_6451	ArsR family transcriptional regulator	*1,00*	3,58

Table 17: Differentially expressed genes under oxidative stress conditions. Sigma factors and transporters proteins. Values in italics are not significant.

Sigma factor				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
Rv0182c	MSMEG_0219	RNA polymerase factor sigma-70	1,72	*0,91*
Rv3223c	MSMEG_0573	ECF sigma factor RpoE1	1,58	*-0,26*
Rv3223c	MSMEG_0574	ECF sigma factor RpoE1	1,58	1,58
	MSMEG_1418	RNA polymerase ECF-type sigma factor	*-1,32*	1,58
Transporter proteins				
	MSMEG_0113	taurine transporter permease TauC	-2,17	-1,26
	MSMEG_0116	taurine import ATP-binding protein TauB	-1,83	*-0,70*
	MSMEG_0549	ABC transporter permease	*-0,67*	-1,00
	MSMEG_0551	nitrate/sulfonate/bicarbonate ABC transporter ATPase	*-1,58*	-1,37
	MSMEG_1169	integral membrane transporter	1,77	*1,00*
Rv0821c Rv3301c	MSMEG_1605	phosphate transporter regulatory protein PhoU	*-0,37*	-1,34
	MSMEG_1683	cytosine/purines uracil thiamine allantoin permease	*-0,72*	-1,01
Rv1859	MSMEG_2014	molybdenum import ATP-binding protein ModC	-3,70	*-0,38*
Rv1858	MSMEG_2015	molybdate ABC transporter permease ModB	-4,91	*-0,20*
Rv1857	MSMEG_2016	molybdate ABC transporter periplasmic molybdate-binding protein ModA	-6,13	*-0,64*
	MSMEG_2680	amino acid transporter	1,58	*0,49*
	MSMEG_2796	C4-dicarboxylate ABC transporter	3,25	*1,00*
	MSMEG_2924	permease binding-protein component	*-0,31*	-1,40
	MSMEG_2926	glycine betaine/carnitine/choline transport ATP-binding protein opuCA	*-0,32*	-1,15
	MSMEG_2927	ABC transporter permease	*-0,48*	-1,06
	MSMEG_3314	transporter	*-0,14*	-1,46
	MSMEG_3386	shikimate transporter	2,04	*1,17*
	MSMEG_3536	sugar transporter	*0,06*	-1,66
	MSMEG_4318	Formate/nitrite family of transporters-like protein	*-0,93*	-1,13
	MSMEG_4355	peptide ABC transporter permease	*-1,12*	-1,32
	MSMEG_4385	peptide ABC transporter	-2,12	*-0,70*
Rv1244	MSMEG_5054	ABC transporter substrate-binding protein	*-0,87*	-1,22
	MSMEG_5102	ABC transporter ATP-binding protein	1,81	2,17
	MSMEG_5559	metabolite/sugar transporter	*0,18*	-1,36
	MSMEG_6307	glutamine transporter permease	*-1,66*	-1,04
	MSMEG_6309	ABC transporter ATP-binding protein	*-1,37*	-1,47
	MSMEG_6544	transporter	*-0,57*	-1,29
	MSMEG_6769	transporter monovalent cation:proton antiporter-2 (CPA2) family protein	*-0,42*	-1,16

Table 18: Differentially expressed genes under oxidative stress conditions. Protein regulation and tRNA. Values in italics are not significant.

Protein regulation and tRNA				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
	MSMEG_3755	5S ribosomal RNA rrfA	*-0,24*	-3,93
	MSMEG_4929	5S ribosomal RNA rrfB	*-0,37*	-4,03
	MSMEG_5489	50S ribosomal protein L32	*0,04*	-1,87
Rv0881	MSMEG_5687	23s ribosomal RNA methyltransferase	*-0,58*	-1,00
Rv2057c	MSMEG_6067	50S ribosomal protein L33	*0,42*	1,22
	MSMEG_6068	50S ribosomal protein L28	*0,65*	1,39
Rv3701c	MSMEG_6247	histidine-tRNA ligase	1,81	*1,00*
	MSMEG_6615	dithiol-disulfide isomerase	*-0,59*	-1,09
	MSMEG_0008	Ile tRNA	*-1,08*	-2,83
	MSMEG_0037	Leu tRNA	*-0,19*	-4,97
	MSMEG_0677	Gly tRNA	*-0,42*	-1,06
	MSMEG_1166	Tyr tRNA	*-0,19*	-2,86
	MSMEG_1337	Thr tRNA	*-0,31*	-2,95
	MSMEG_1965	Met tRNA	*0,05*	-2,71
	MSMEG_2138	Ala tRNA	*0,76*	-3,19
	MSMEG_2384	Gln tRNA	*0,16*	-2,73
	MSMEG_2385	Glu tRNA	*0,27*	-3,28
	MSMEG_2833	Val tRNA	*-0,51*	-1,91
	MSMEG_2836	Val tRNA	*-0,38*	-3,23
	MSMEG_3734	Pro tRNA	*-1,17*	-1,00
	MSMEG_4201	Leu tRNA	*0,32*	-2,65
	MSMEG_4478	Asn tRNA	*-0,64*	-3,36
	MSMEG_4725	His tRNA	*-0,51*	-1,54
	MSMEG_4746	Lys tRNA	*1,13*	-2,46
	MSMEG_5598	Arg tRNA	*-0,25*	-3,83
	MSMEG_5756	Asp tRNA	*-0,62*	-2,40
	MSMEG_5758	Lys tRNA	*-0,48*	-4,98
	MSMEG_6152	Thr tRNA	*-0,41*	-3,33
	MSMEG_6204	Pro tRNA	*0,10*	-3,86
	MSMEG_6287	Ser tRNA	*-0,44*	-2,23
	MSMEG_6326	Ser tRNA	*0,07*	-5,43
	MSMEG_6350	Ser tRNA	*-0,46*	-2,15

Table 19: Differentially expressed genes under oxidative stress conditions. Proteins with unknown function. Values in italics are not significant.

Unknown function				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/noH_2O_2
	MSMEG_0183	hypothetical protein	2,63	*0,93*
Rv0184	MSMEG_0222	hypothetical protein	1,13	*0,39*
Rv0185	MSMEG_0223	hypothetical protein	1,58	*1,00*
	MSMEG_0493	hypothetical protein	2,97	2,26
	MSMEG_0494	hypothetical protein	3,46	*2,00*
	MSMEG_0558	hypothetical protein	*0,47*	1,28
	MSMEG_0757	hypothetical protein	1,91	*1,54*
	MSMEG_0910	hypothetical protein	*2,58*	3,58
	MSMEG_0963	hypothetical protein	*0,21*	-1,42
	MSMEG_1223	hypothetical protein	2,00	*1,14*
	MSMEG_1237	hypothetical protein	3,75	3,38
	MSMEG_1271	hypothetical protein	3,25	*2,32*
	MSMEG_1273	hypothetical protein	2,91	2,75
	MSMEG_1279	hypothetical protein	1,98	*1,16*
	MSMEG_1282	hypothetical protein	1,96	*1,14*
	MSMEG_1619	hypothetical protein	4,15	3,09
	MSMEG_1627	hypothetical protein	*-1,15*	-2,21
	MSMEG_1766	hypothetical protein	*-0,58*	-1,56
	MSMEG_1781	hypothetical protein	*-0,54*	-2,17
	MSMEG_1783	hypothetical protein	*-0,51*	-1,30
	MSMEG_1788	hypothetical protein	*-0,53*	-1,13
	MSMEG_2115	hypothetical protein	*-0,89*	-1,27
	MSMEG_2199	hypothetical protein	*-0,65*	-1,11
	MSMEG_2258	hypothetical protein	*-1,32*	-1,00
	MSMEG_2261	hypothetical protein	-2,04	*-0,55*
	MSMEG_2376	hypothetical protein	*-0,39*	-1,12
	MSMEG_2679	hypothetical protein	1,40	*1,21*
	MSMEG_2958	hypothetical protein	*-1,00*	-2,58
	MSMEG_3295	hypothetical protein	*3,00*	3,70
	MSMEG_3352	hypothetical protein	1,46	*0,81*
	MSMEG_3443	hypothetical protein	*-0,40*	-1,02
	MSMEG_3537	hypothetical protein	*0,13*	-1,12
	MSMEG_3680	hypothetical protein	2,19	1,85
	MSMEG_3713	hypothetical protein	1,95	2,85
	MSMEG_4026	hypothetical protein	1,67	*-0,30*
	MSMEG_4028	hypothetical protein	1,87	*0,87*
Rv2137c	MSMEG_4195	hypothetical protein	*-0,28*	-1,44
	MSMEG_4257	hypothetical protein	1,99	*1,36*
Rv2342	MSMEG_4479	hypothetical protein	*-0,82*	-1,55
	MSMEG_4499	hypothetical protein	2,57	1,59
	MSMEG_4500	hypothetical protein	2,58	*0,00*
	MSMEG_4518	hypothetical protein	-3,22	-3,49
	MSMEG_4749	hypothetical protein	*1,81*	1,58
	MSMEG_4961	hypothetical protein	2,20	0,75
	MSMEG_4993	hypothetical protein	*-0,22*	-1,62
	MSMEG_5157	hypothetical protein	1,54	*1,13*
	MSMEG_5406	hypothetical protein	*-0,84*	-1,95
	MSMEG_5461	hypothetical protein	-2,54	*-0,27*
	MSMEG_5543	hypothetical protein	*0,42*	-1,10
	MSMEG_5558	hypothetical protein	2,67	1,45
	MSMEG_5803	hypothetical protein	*-0,65*	-1,00
	MSMEG_5875	hypothetical protein	*-0,86*	-1,24
	MSMEG_6211	hypothetical protein	*-0,31*	-1,58
	MSMEG_6355	hypothetical protein	*-0,49*	-1,08
	MSMEG_6456	hypothetical protein	*0,02*	-1,43
	MSMEG_6498	hypothetical protein	2,56	1,30
	MSMEG_6501	hypothetical protein	*-0,03*	-1,28
	MSMEG_6566	hypothetical protein	2,17	*1,06*
	MSMEG_6609	hypothetical protein	*-0,47*	-1,41

Table 20: Differentially expressed genes under oxidative stress conditions. Proteins with undeterminated function and antigen and cell division. Values in italics are not significant.

Undeterminated function				
Orthologs in TB	Synonym	Product	log2 H_2O_2 3h/no H_2O_2	log2 H_2O_2 6h/no H_2O_2
	MSMEG_0114	extracellular solute-binding protein	*1,17*	-1,00
	MSMEG_0186	MmpS2 protein	2,77	*1,30*
	MSMEG_0550	sulfonate binding protein	*-0,98*	-1,33
Rv0288 Rv3017c Rv3019c	MSMEG_0621	low molecular weight protein antigen 7	*-0,65*	-1,00
	MSMEG_0696	alanine-rich protein	*-0,38*	-1,44
	MSMEG_1193	TROVE domain-containing protein	2,20	2,20
	MSMEG_1251	gp58 [Mycobacterium phage Spartacus]	2,00	1,64
	MSMEG_1264	prophage Lp1 protein 5	3,32	2,47
	MSMEG_1270	gp75 [Mycobacterium phage Marvin]	3,30	3,51
	MSMEG_1281	putative cytosolic protein	1,98	*1,32*
	MSMEG_1356	DinB superfamily protein	1,96	1,51
	MSMEG_1357	Azi37 / DinB superfamily protein	3,58	2,81
	MSMEG_1384	cupin	*-0,81*	-2,22
	MSMEG_1421	mycofactocin precursor	*-1,29*	-1,95
Rv0692	MSMEG_1422	mycofactocin system protein MftB	*-1,40*	-1,33
	MSMEG_1544	PduO protein	-2,12	*-0,12*
	MSMEG_1755	Anti-sigma factor, ChrR	1,96	*0,97*
	MSMEG_1921	SecC motif-containing protein	*-0,47*	-1,00
Rv1754c	MSMEG_2107	conserved secreted protein	-1,78	*-0,83*
	MSMEG_3289	gp61 protein	*0,10*	-1,02
	MSMEG_3298	response regulator receiver domain-containing protein	1,69	2,15
	MSMEG_3377	gp15 protein	1,62	1,93
Rv2207	MSMEG_4275	phosphoribosyltransferase	*-0,65*	-1,02
	MSMEG_5502	ATPase AAA	-1,81	-1,25
	MSMEG_5761	cupin	*0,79*	1,75
	MSMEG_6243	response regulator receiver domain-containing protein	*-0,86*	-1,00
	MSMEG_6359	trypsin domain-containing protein	*-0,24*	-1,06
Antigen and Cell division				
Rv1677 Rv2878c	MSMEG_3543	Soluble secreted antigen MPT53	*-0,42*	-1,50
Rv1209	MSMEG_5083	cell division protein DivIVA	1,94	1,80

Among the upregulated genes, there was substantial overlap between the two time points while the overlap between the downregulated genes was quite small (figure 17). It is also worth noting that for the 3h condition, the majority of the differentially expressed genes (186 out of 217) were upregulated, while for the 6h samples the majority of the differentially expressed genes (201 out of 298) were downregulated.

Previous studies on yeast have shown that H_2O_2 inhibits translation elongation and can increase the time of ribosomal transit in mRNA by 50% in eucaryotes (Grant, 2011). The finding of some tRNAs on the downregulated list of genes for the 6h exposure to hydrogen peroxide is also consistent with the idea that translation is reduced. This may be also explained by the observation that oxidative stress induces reduction in protein production by inhibiting translation presumably to save energy for the essential processes in the cell and prevent synthesis of defective proteins (Grant, 2011).

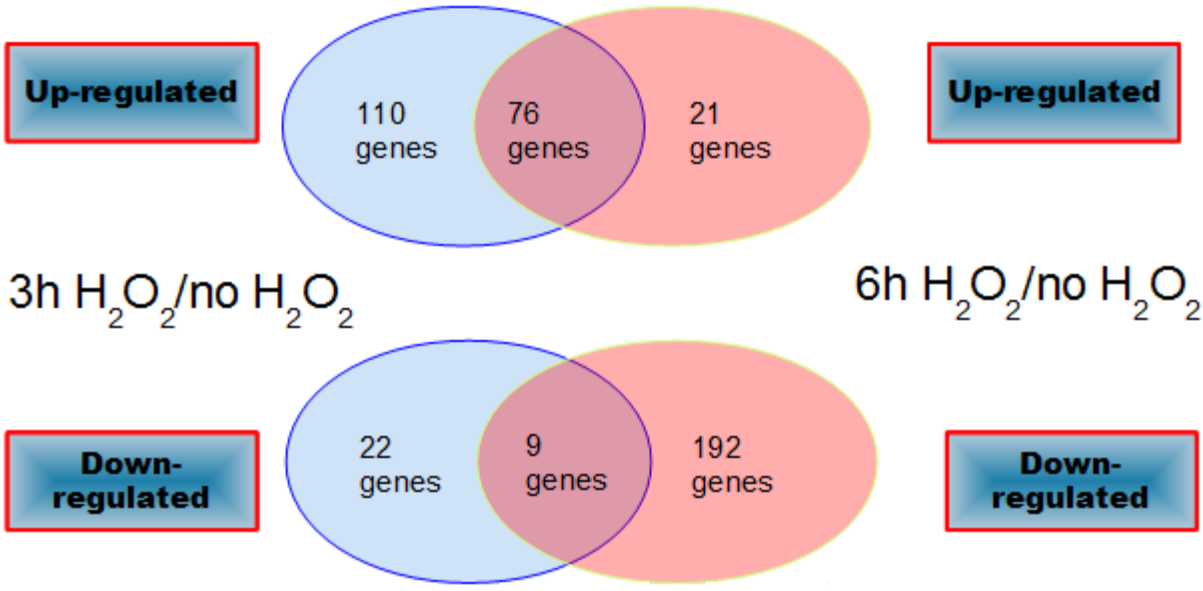

Figure 17: Overall analysis of number of genes up and downregulated under oxidative stress conditions and overlap between the 2 sets of genes.

The next step was to run DAVID analysis (Huang *et al*, 2009) in order to identify the enriched pathways for the up and downregulated genes under oxidative stress conditions (table 13). As the table shows, DAVID (Huang *et al*, 2009) could not identify any statistically significant (0.05

or smaller value for the Bonferroni correction) clusters within the downregulated genes lists. For the upregulated gene lists on the other hand, both time-points showed enrichment in HNH nucleases, with enrichment scores of 5.87 and 4.22 for 3h and 6h of exposure to oxidative stress. The enrichment scores tell us which gene groups are more significant in the tested experiment and is measured by the software using the geometric mean of the p-values of the terms in a particular cluster followed by the application of the minus log of the mean (Huang *et al.*, 2007). Therefore a high enrichment score suggests that this particular group of members is involved in more relevant roles (Huang *et al.*, 2007). For the three hour time point, 168 genes of a total of 186 genes were included in the analysis, and of those 168 genes, 11 were classified as belonging to the HNH nuclease cluster. For the six hour time-point, 88 genes out of a total of 97 genes were included in the analysis and of those 88 genes, 5 genes were classified as belonging to the HNH nuclease cluster. DAVID was not able to include some if the genes in the analysis because some of then were not able able to be mapped to the database. Even using a custom background, the gene list had to be converted to DAVID format in order to run the software and, also for this step, not all genes were able to be mapped to *M. smegmatis* genome.

Table 21: DAVID result for the enrichment pathway analysis for both up and downregulated genes under 3h and 6h of hydrogen peroxide treatment. The table is only showing the statistically significant clusters when using a cutoff of 0.05 for the Bonferroni correction.

H_2O_2 3h / no H_2O_2 upregulated genes		
Cluster name	Enrichment score	Bonferroni correction
HNH Nuclease	5.87	≤ 0.036
H_2O_2 3h / no H_2O_2 downregulated genes		
Cluster name	Enrichment score	Bonferroni correction
No statistically significant clusters found		
H_2O_2 6h / no H_2O_2 upregulated genes		
Cluster name	Enrichment score	Bonferroni correction
HNH Nuclease	4.22	≤ 0.0098
H_2O_2 6h / no H_2O_2 downregulated genes		

Cluster name	Enrichment score	Bonferroni correction
No statistically significant clusters found		

Having HNH nucleases enriched within the upregulated genes was an expected result as this enzyme domain is found in proteins shown to take part in many cellular processes, including DNA repair and recombination (Pediaditakis et al, 2012). Also, HNH nuclease family genes were previously shown to to be induced in *M. smegmatis* in response to oxidative stress (Li *et al*, 2015) and, as stated before, oxidative stress may cause DNA damage (Dussurget and Smith, 1998) which explains why this class of proteins is upregulated in response to ROS.

A previous study (Li *et al,* 2015) exposed *M. smegmatis* early log phase cultures to 0.2 mM and 7 mM H_2O_2 for 30 minutes and then performed a RNA-seq experiment to try to understand mycobacterial response to oxidative stress under low and high concentrations of hydrogen peroxide. We constructed heat maps to compare the expression of differentially expressed genes identified in that study and in ours (Figure 18).

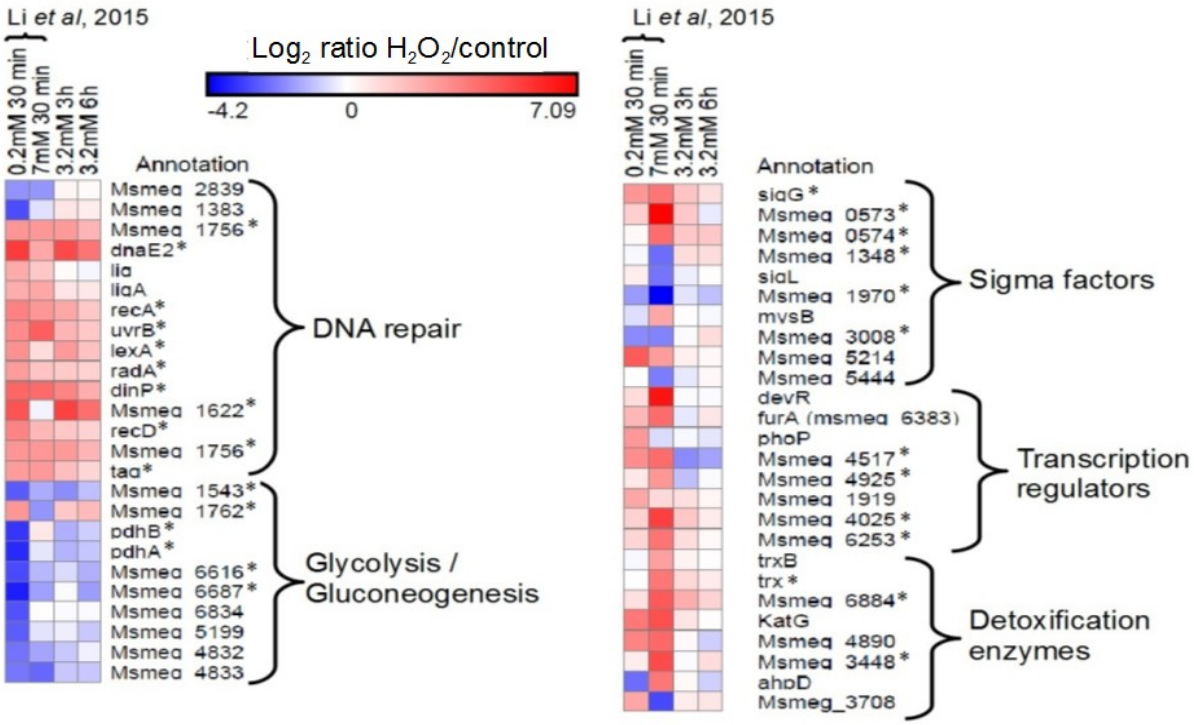

Figure 18: Heat map showing the comparison between data from Li *et al,* 2015 and the data gathered in this project. The figure shows the log_2 ratios of FPKM (Li *et al,* 2015) or log_2 ratios of RPKM (data from this project) for the included genes when compared 2 different conditions. Genes were grouped by function according

to Li *et al,* 2015. Test samples are the ones containing hydrogen peroxide in the specific concentrations shown in the figure and control samples consists in samples were the cells were not exposed to hydrogen peroxide during the experiment. This figure only includes the genes found on both data sets and the * indicates genes that differential expression (2 fold or higher difference) in at least one comparison in our data. This heat map was constructed using the software GENE-E from Broad Institute available at: http://www.broadinstitute.org/cancer/software/GENE-E/index.html.

We saw that many of the sigma factors, transcription regulators and detoxification enzymes that were included in the Li *et al*, 2015 study were also differentially expressed 3h or 6h after addition of 3.2 mM of hydrogen peroxide in our study.

As expected, DNA repair enzymes are upregulated for both data sets, with special attention to MSMEG_1622 and dnaE2, which were the most highly upregulated genes in our study. dnaE2 (MSMEG_1633) encodes a DNA polymerase III α chain which is known to be involved in DNA repair and emergence of RIF resistance in Mtb (Boshoff *et al*, 2003). MSMEG_1622 sequence was retrieved from Smegmalist website and analyzed using BLAST, which showed the gene is likely to encode a DNA repair polymerase that facilitate translesion synthesis.

When looking at the transcription factors, we noted that our data does not seem to have any strong differential expression. Li *et al.*, 2015 data, on the other hand, present a trend of upregulated transcription factor, specially for the 7 mM condition. It is interesting to notice that in some cases, for example MSMEG_4517 and MSMEG_4925, we do see an opposite trend in the data, where our data for those particular genes shows downregulation and Li's group show upregulation.

4.3. OXIDATIVE STRESS 5' END MAPPING LIBRARIES

With the objective of exploring the RNA cleavage sites present in *M. smegmatis* to infer if changes in RNA cleavage contribute to the oxidative stress response, we constructed 5' end mapping libraries from the same RNA samples used for expression library construction above. After rRNA depletion, each sample were divided into two and one of then went through an polyphosphatase treatment while a mock treatment was performed with the other half. The polyphosphatase treatment degraded the triphosphate present in the 5' end of the transcription start

sites to monophosphate so the enzyme T4 RNA ligase could act in both transcription start sites and cleavage sites, that naturally have a monophosphate in their 5' end. On the mock treatment sample on the other hand, only the cleavage sites will have the adapters ligated once they were the only ones with a monophosphate in the 5' end. Next, an adapter was ligated to the 5' end monophosphate of all the available mRNA in the sample. RNA was fragmented, and cDNA was synthesized using a random primer linked to a 3' adapter sequence. PCR was performed to enrich for the fragments with both adapters, and samples were sequenced using Illumina platform. The first nucleotide of each Read 1 corresponds to the precise 5' end of a transcript.

Reads were aligned to *M. smegmatis* genome and, after some normalization steps, the first nucleotide of each transcript (henceforth called "5' end coverage") was analyzed in order to (1) infer if that particular coordinate was a transcription start site or a cleavage site and (2) determine if abundance of each 5' end was affected by hydrogen peroxide.

Reads with 5' end coverage below an arbitrary cutoff of 100 normalized reads were filtered out to remove low-abundance 5' ends and limit further analysis to reproducible data. To address if the arbitrary cutoff for 5' end coverage of the converted libraries was appropriate to eliminate the background from the samples, graphs showing the correlation between the 2 biological replicates used in this experiment are shown in figure 19. A appropriate library should have little to no background and good correlation between the biological replicates of the same condition.

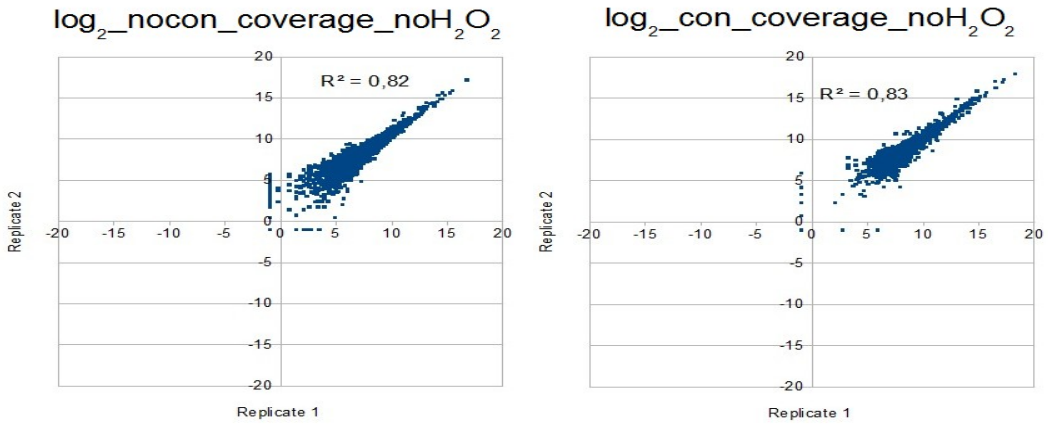

Figure 19: Comparison between replicate 1 and 2 for the control libraries with no hydrogen peroxide. The left panel shows the log_2 values for 5' end coverage of the non-converted libraries when no H_2O_2 is present in the culture. The right panel shows the log_2 values for coverage of the converted libraries when no H_2O_2 is present

in the culture. R^2 values show how correlated the samples are with each other.

The graphs above show that the 100 cutoff of the converted libraries, although conservative, is good option due to the limitations of this data set. When analyzing the results without applying the arbitrary background, the biological replicates were less well correlated (figure 20), which makes difficult to compare conditions and to compare converted and non-converted libraries within each condition.

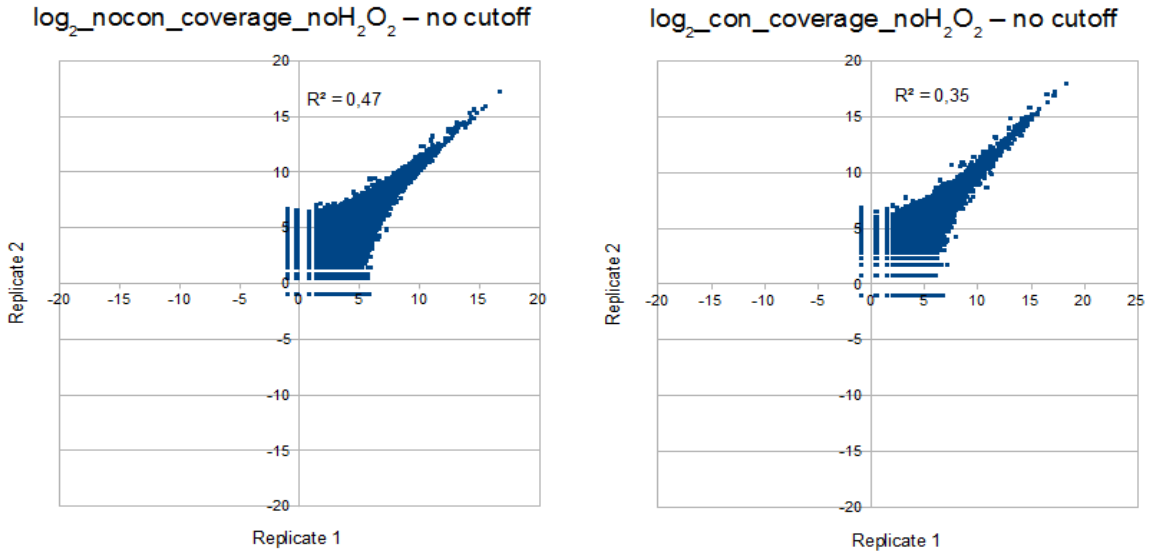

Figure 20: Comparison between replicate 1 and 2 for the control libraries with no hydrogen peroxide before applying the cutoff of 100 coverage for the converted libraries.

To assess our ability to distinguish transcription start sites from cleavage sites, the ratio of coverage for each 5' ens in the converted library relative to the non-converted library was determined and the distributions of the $\log_2$ of these ratios were assess. Log_2 converted/non-converted ratio distributions for the 2 no H_2O_2 replicates are shown (figure 22). We expected to see binomial distributions with cleavage sites equally represented in the two libraries and transcriptions start sites more highly represented in the converted library (figure 21). However, the distributions were not clearly binomial, indicating that our ability to distinguish between transcription start sites and cleavage sites is limited.

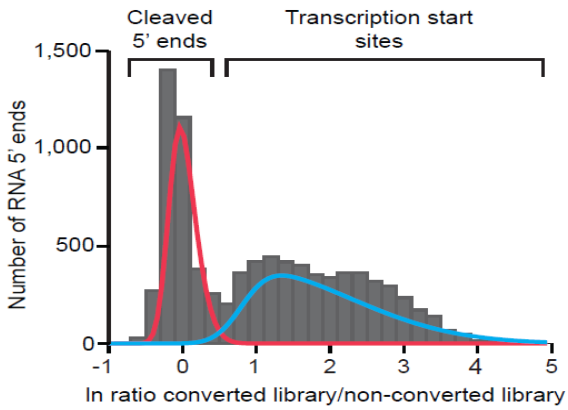

Figure 21: Expected histogram plot for the 5' end mapping data showing a binomial distribution. Retrieved from Shell *et al*., unpublished data.

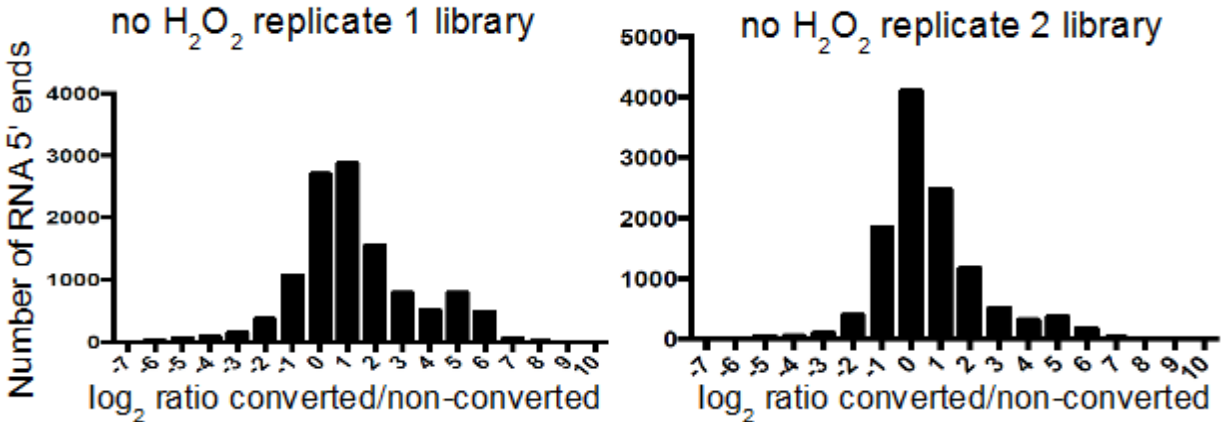

Figure 22: Histograms showing the number of 5' ends by log_2 ratio of converted over non-converted coverages for the 2 replicates of the no H_2O_2 condition.

Next, we looked into the plots comparing the log_2 ratio between the converted and non-converted libraries for the 2 biological replicates from each condition (figure 23). The R^2 for these graphs are below expected. R^2 represent the correlation between the samples and while it was expected that biological replicates would have some variability, values around 0.25 to 0.47 are low and limit our ability to distinguish transcription start sites from cleavage sites with high confidence. This further shows the limitations of the data set used.

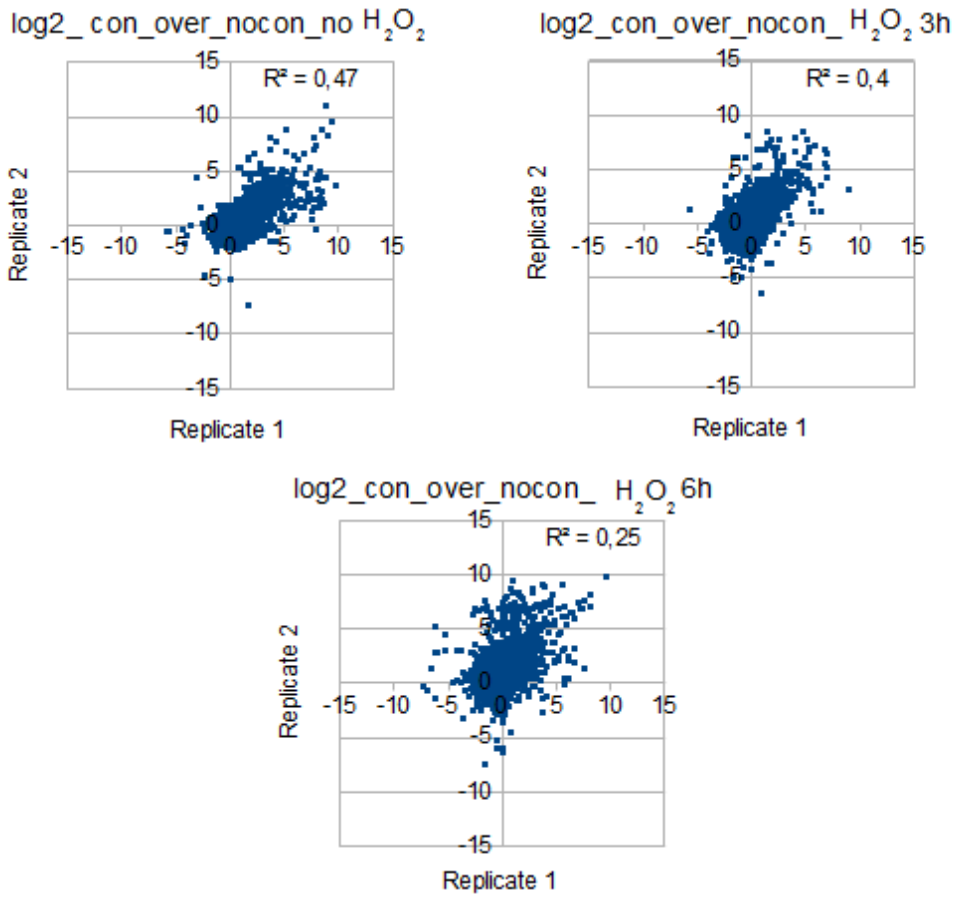

Figure 23: Plots of log_2 ratios between the converted and non-converted libraries comparing the 2 biological replicates for no H_2O_2, H_2O_2 3h and H_2O_2 6h. R^2 values show the correlation between the 2 biological replicates.

The dephosphorylation for the converted libraries may not have worked as well as planned, so a way to increase this reaction efficiency would be to increase the incubation time from 1 hour to 2 hours for example. In the case of the fragmentation, the samples appeared to be over fragmented, (figure 24) which lead to smaller fragments that could be easily washed away during the purification steps. Over-fragmentation may also lead to excess variability between replicates due to less available starting material for the enrichment PCR step, and therefore more PCR cycles are needed. More PCR cycles leads to more variability between samples because of "jackpotting" (Shell, personal communication). So instead of using 94 °C for 11 minutes, for next libraries constructions could be a nice idea to decrease time and temperature for the fragmentation step.

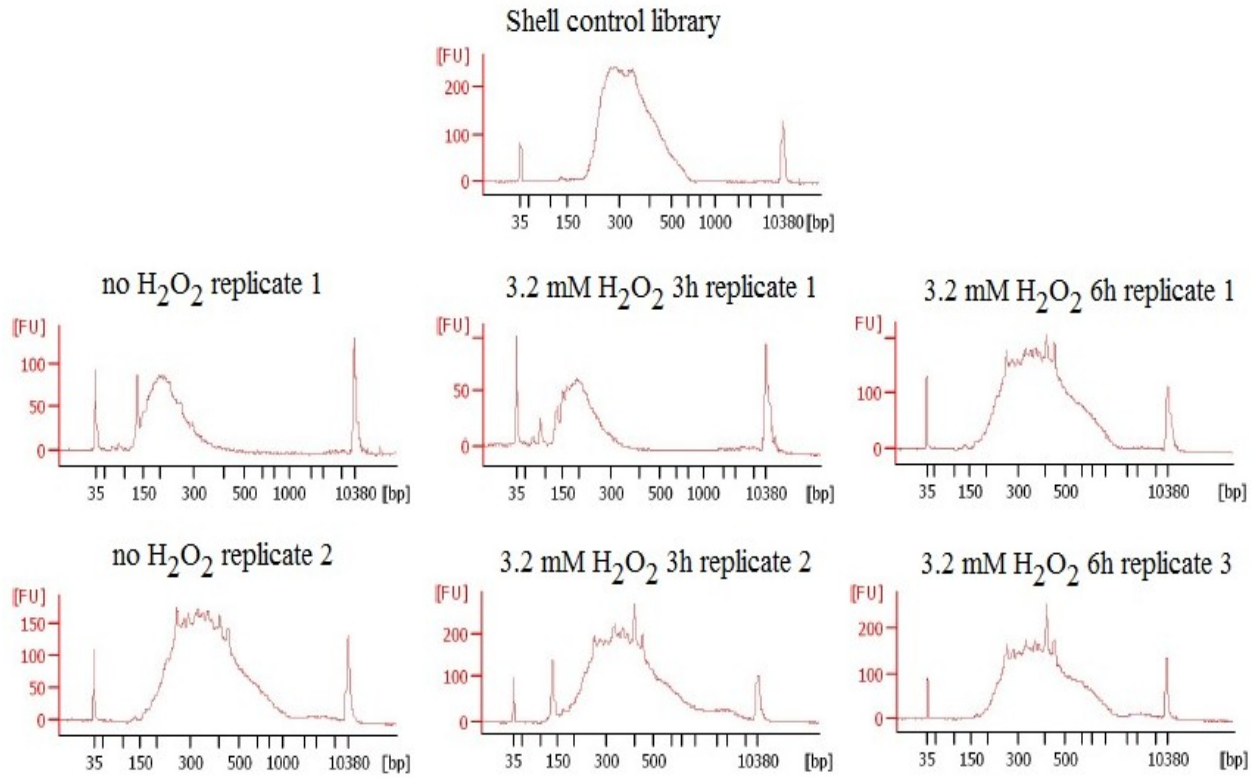

Figure 24: Results from Bioanalyzer using oxidative stress libraries samples. Samples were compared with a library from Mtb constructed by Professor Shell in an independent experiment. The no H_2O_2 replicate 1 and 3.2 mM H_2O_2 3h replicate 1 libraries appear to have an smaller overall fragment distribuition size than the other samples, which suggests that those samples had their mRNA overfragmented. We can also notice a sharp peak around 150 bp for some of the samples, for example no H_2O_2 replicate 1 and 3.2 mM H_2O_2 3h replicate 2. This peak is possible adapter dimer that was not efficiently removed from the samples by the magnetic bead purification.

Even with some limitations as stated above, the highest and lowest ratios are likely to represent transcription start sites and cleavage sites, respectively. In order to verify that prediction, sequence logos were created with Web Logos (Crooks *et al*, 2004) using the 500 highest $\log_2$ ratios between the converted and non-converted libraries and the 500 lowest $\log_2$ ratios between the converted and non-converted libraries for each condition (figures 25, 26, 27).

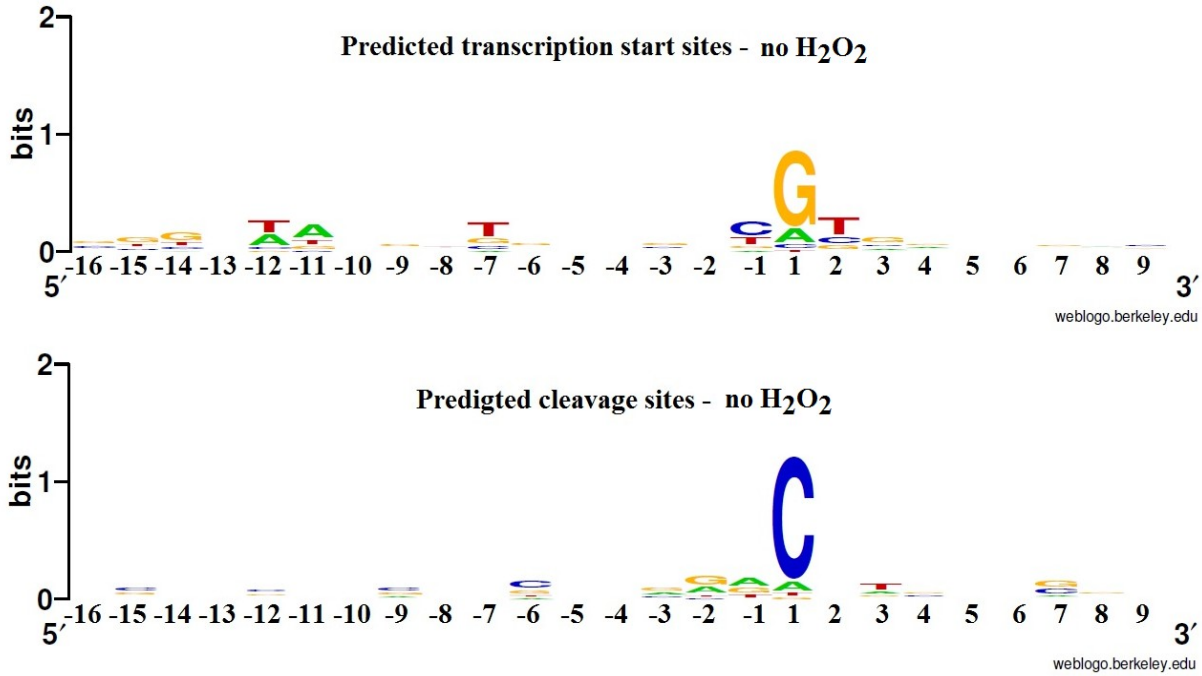

Figure 25: Web logos for the predicted transcription start sites (500 highest ratios) and cleavages sites (500 lowest ratios) under no H_2O_2 conditions. Figures made using the website http://weblogo.berkeley.edu/logo.cgi

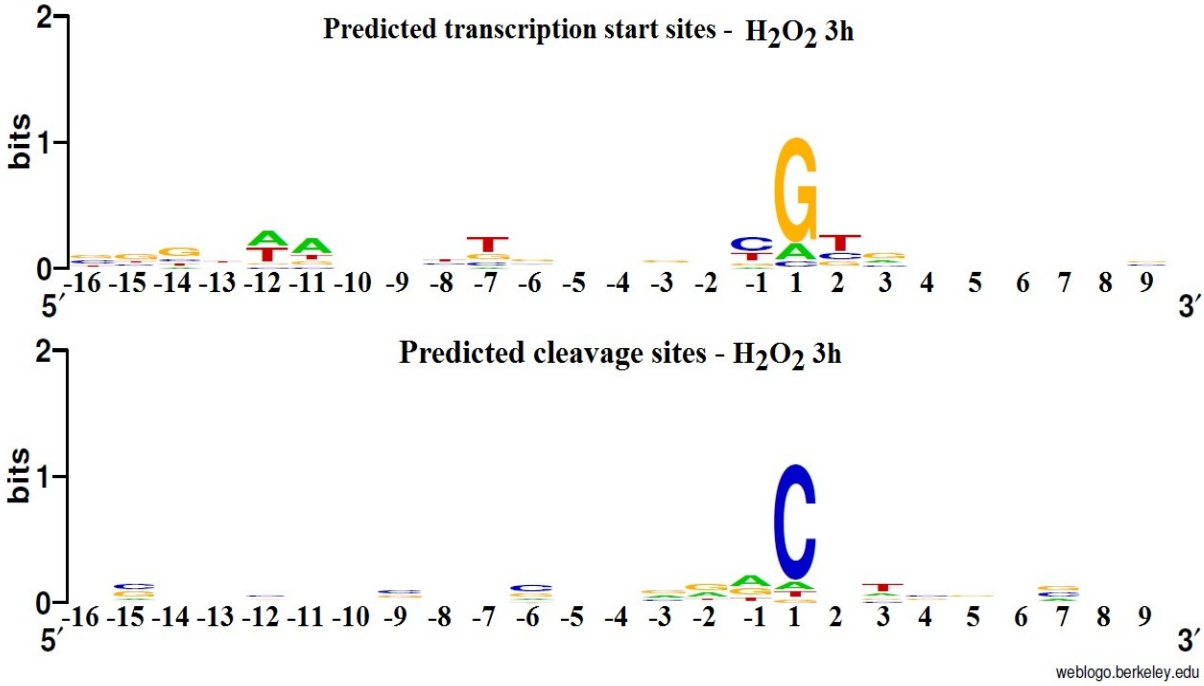

Figure 26: Web logos for the predicted transcription start sites (500 highest ratios) and cleavages sites (500 lowest ratios) under H_2O_2 3h conditions. Figures made using the website http://weblogo.berkeley.edu/logo.cgi

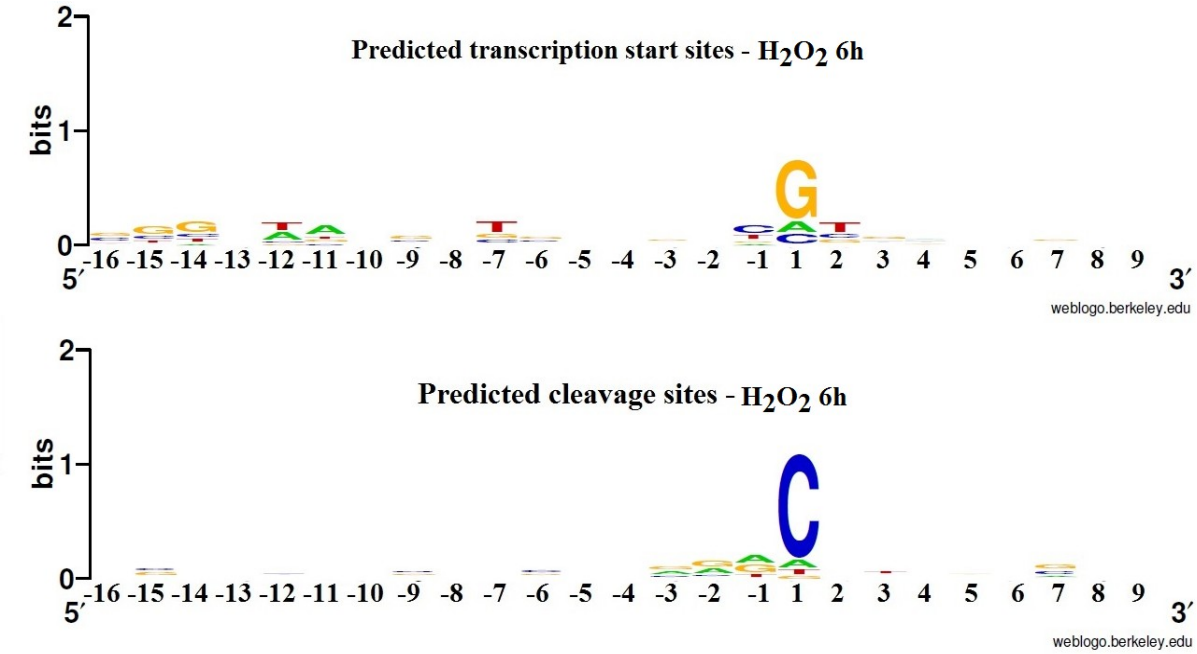

Figure 27: Web logos for the predicted transcription start sites (500 highest ratios) and cleavages sites (500 lowest ratios) under H_2O_2 6h conditions. Figures made using the website http://weblogo.berkeley.edu/logo.cgi

The sequence logos indicate that the highest 500 ratios in each condition are indeed enriched for transcription start sites, as we observed an adenine and thymine-rich region around the -10 site, which resembles a housekeeping sigma factor binding site from *E. coli* (Lodish *et al*, 2000) and *M. tuberculosis* (Shell *et al*, 2015). In *M. tuberculosis*, a purine at the position +1 is reported to be characteristic of transcription start sites (Shell *et al*, 2015), consistent with our data and supporting the idea that the 500 highest ratios represent transcription start sites in *M. smegmatis*. For the predict cleavage site (500 lowest ratios) is possible to see that before the cleavage site every 3 nucleotides we usually see either a cytosine or a guanine. This pattern is characteristic of mycobacterial coding sequences, where the last nt of the codon will be more likely to be a cytosine or guanine as a results of GC rich genomes (Scapoli *et al*, 2009). Also, as described and mapped in Mtb, most cleavage sites occurred immediately between cytosines (Shell *et al*, unpublished data), so observing enrichment for this residue suggests that the 500 lowest ratios are indeed cleavage sites.

In order to understand the basis of the gene expression differences observed in section 4.2.,

where we discussed the results for the expression libraries we constructed, we wanted to identify both transcription starts sites and cleavage sites that were differentially abundant in oxidative stress. We therefore compared different conditions to see which were the differentially expressed 5' ends. This was done by plotting the mean of the log_2 converted library coverages from the 2 replicates for one condition against the mean of the log_2 converted library coverage from the 2 replicates for the other condition (figure 28).

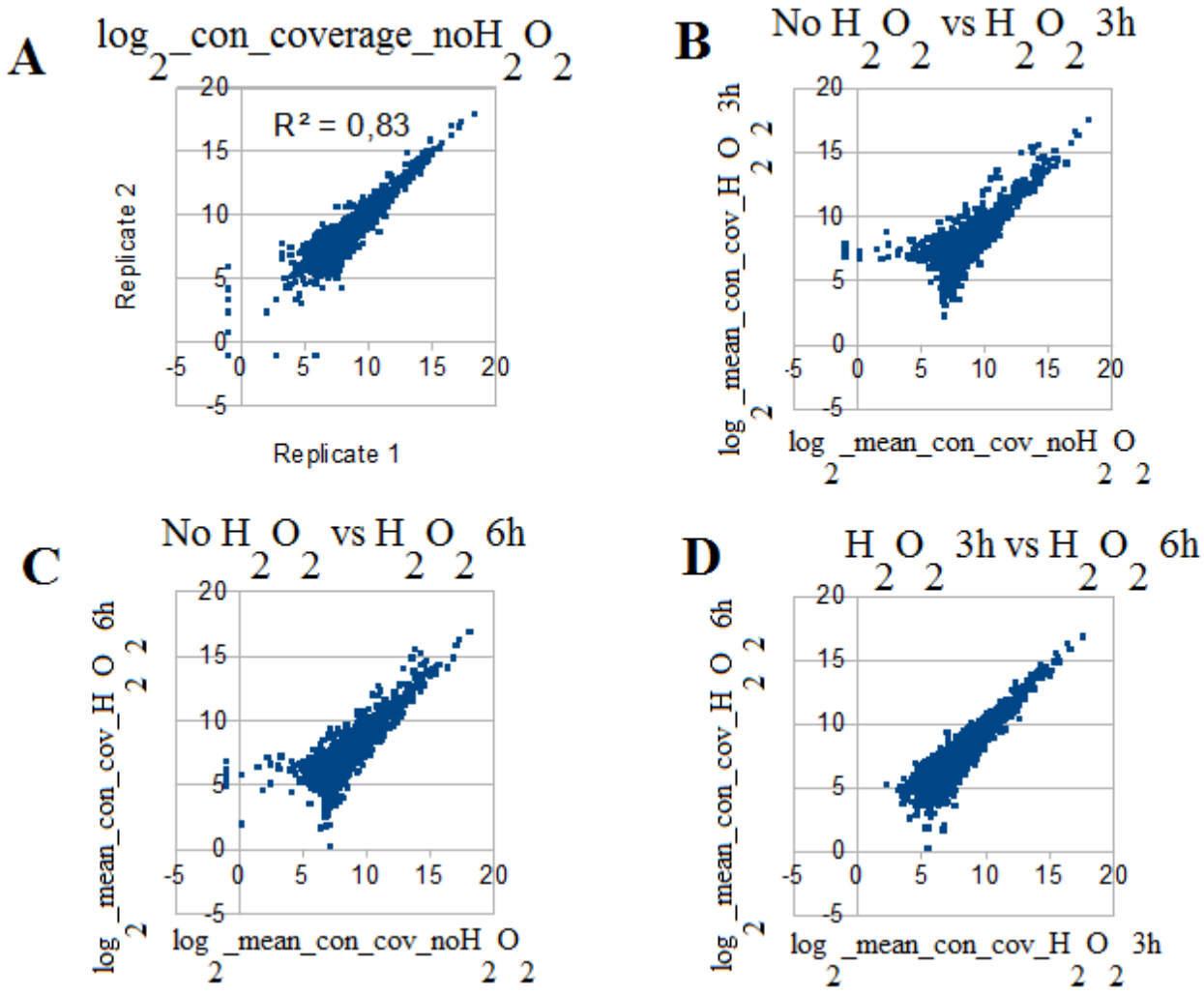

Figure 28: Comparison between the biological replicates converted libraries under no H_2O_2 conditions with no H_2O_2 vs H_2O_2 3h, no H_2O_2 vs H_2O_2 6h and H_2O_2 3h vs H_2O_2 6h. For all plots only the converted libraries were used. log_2_mean_con_cov strands for log_2 mean of the coverage of the 2 replicates from the converted libraries.

Figure 28 shows in panel A the comparison between the 2 replicates under no H_2O_2 conditions. This plot serves as a "calibrator" to analyze the other plots as we expect no biological

meaningful differential 5' end abundance between the two biological replicates. Panel B shows the differential 5' ends when comparing no H_2O_2 with H_2O_2 3h conditions, the differential 5' ends are the outliers of the plot. The same is true for panel C which compares no H_2O_2 with H_2O_2 6h; again the outliers correspond to the differential 5' ends. Finally, panel D shows the comparison between H_2O_2 3h with H_2O_2 6h. This plot has fewer outliers, suggesting that the number of differential 5' ends is minimal. In order to establish a meaningful cutoff for identifying differentially abundant 5' end we determined the $\log_2$ of the mean 5'end converted library coverage ratio between conditions, and calculated Z-scores for these ratios. A 5'end was considered to be differentially abundant if it had a Z-score of at least 3, representing $\log_2$ ratios at least 3 standard deviations from the mean. Data analysis will continue be performed using the differentially abundant 5' ends to understand the mechanisms mycobacteria uses under oxidative stress conditions.

5. PHENOTYPIC ASSAYS FOR RNASE J

5.1 CONSTRUCTION OF RNASE J KNOCKOUT, OVEREXPRESSION AND COMPLEMENTED STRAINS.

In order to construct an RNAse J deletion strain, we used a recombineering strategy (Van Kessel and Hatfull, 2008) as described in the methods section 3.5.1. A plasmid encoding phage recombinases behind an inducible promoter was transformed into the WT strain and induced prior to transformation with a cassette to replace the RNase J coding sequence with the hygromicin resistance gene. In order to evaluate the effects of overexpression of RNase J, the RNase J coding sequence was cloned into plasmid PJEB402, which integrates in the L5 attB site (Lee *et al.*, 1991), to create the plasmid pSS138. pSS138 was transformed into both the WT strain and the RNase J deletion (JKO) strain to create overexpression strain and complemented strains, respectively. pSS138 contains a synthetic promoter called MOP, which stands for mycobacterial optimal promoter, and it is a strong promoter with constitutive expression (Fortune *et al*, 2005). For this

complementation plasmid the synthetic promoter was preferred over the native promote as we were not able to identify with confidence the exact place in *M. smegmatis* genome where the RNase J native promoter was.

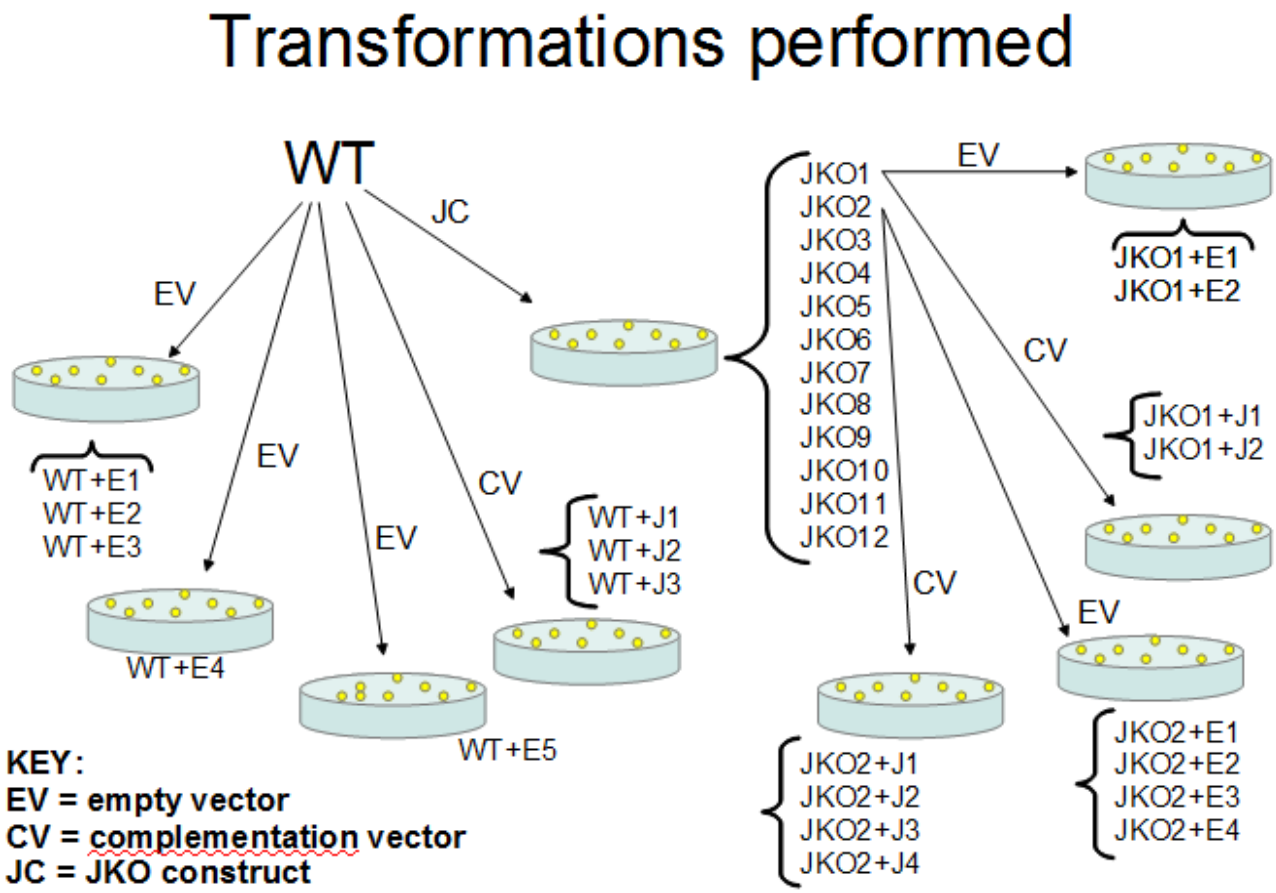

Figure 29: Transformations performed during the course of this project to obtain the WT+E, WT+J, JKO+E, JKO+J strains.

To validate our approach, the expression level of RNase J was assayed by qRT-PCR in the following strains: WT+E (WT strain transformed with "empty" plasmid pJEB402), WT+J (WT strain transformed with RNase J complementation plasmid pSS138), JKO+E (RNase J KO strain transformed with "empty" plasmid pJEB402) and JKO+J (RNase J KO strain transformed with plasmid pSS138). The RNase J expression relative to the housekeeping gene *sigA* was determined (Figure 30).

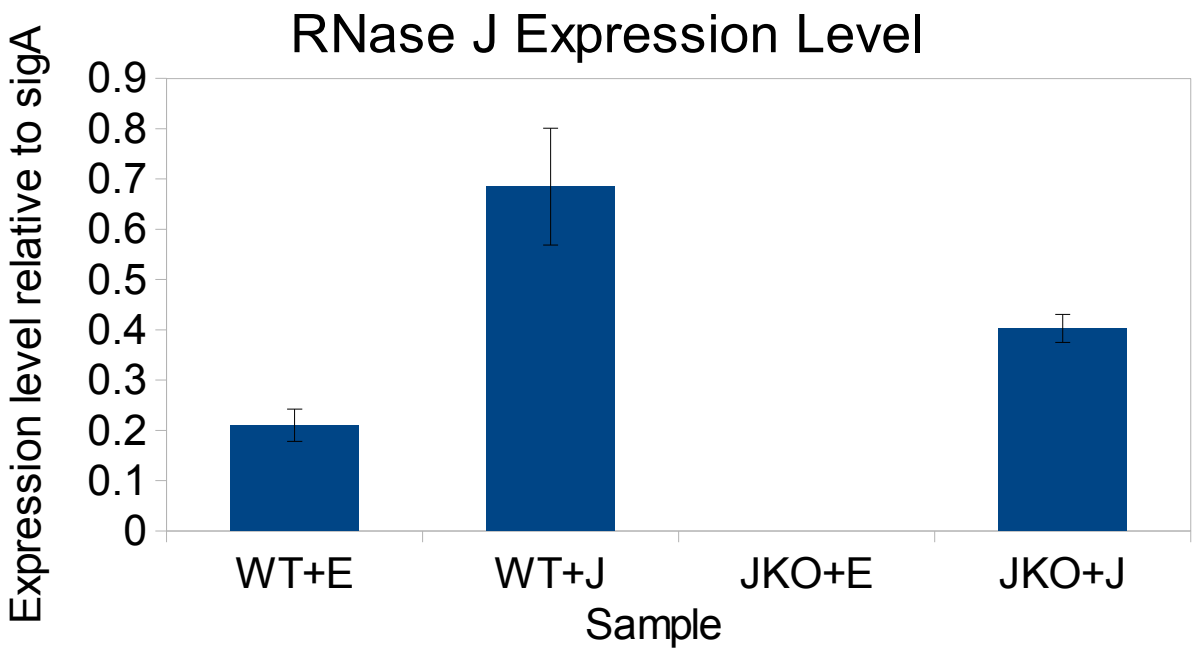

Figure 30: RNase J expression level relative to the housekeeping gene sigA for strains WT+E, WT+J, JKO+E and JKO+J. Experiment was done in biological triplicates for each strain. Error bars show standard deviation.

We observed that the WT+J strain has 3.5 fold higher RNase J expression than the WT+E strain. This was expected as the WT+J strain contains 2 copies of the RNase J coding sequence. JKO+E on the other hand had no detectable expression of RNase J consistent with the successful deletion of the gene from the genome. The JKO+J strain showed about twice the expression level of the WT+E strain. For the KO complemented strain, we had no expectations regarding its expression level as we did not know beforehand how the MOP promoter would compare with the native promoter. These results suggests that MOP is a stronger promoter than the RNase J native promoter.

5.2. MINIMAL INHIBITORY CONCENTRATION

Once the set of 4 strains were constructed and validated for RNase J mRNA level, we decided to challenge them using the first line of defense against Mtb antibiotics, INH and RIF, to determinate if RNase J plays any role in drug tolerance. In order to do that, drug tolerance was measured by calculating the Minimal Inhibitory Concentration (MIC) which is defined as the lower

drug concentration where there is no color change in the resazurin dye. Briefly, cells were grown with different concentrations of antibiotic for 48h, and after the first 24h a resazurin dye was added to determine which cultures were still growing under the tested conditions. Unfortunately our results had substantial variability between isolates than was expected for the same genotype (Table 22).

Table 22: Summary of MIC results for isoniazid (INH) and rifampicin (rif) for strains and isolates tested. JKO1 and JKO2 are 2 different RNase J isolates that were independently transformed with the empty vector and complementation vector. WT+E 4* is a strain originated from a different transformation than the other WT+E isolates used for this assay.

Strain	MIC - INH (µg/mL)	Times measured	MIC - RIF (µg/mL)	Times measured
WT	3	2	1.56	2
WT+E 1	6	2	3.125	2
WT+E 2	-	-	1.56 / 3.125	2
WT+E 3	12 / 24	1	-	-
WT+E 4*	-	-	1.56	1
WT+J 1	12	2	6.25	2
WT+J 2	6	1	3.125 / >3.125	2
WT+J 3	6	1	3.125	1
JKO1	6	2	<0.39	2
JKO1+E 1	6	2	<0.39	2
JKO1+E 2	-	-	1.56	1
JKO2+E 1	3	1	0.78	1
JKO2+E 2	3	1	1.56	2
JKO2+E 3	-	-	0.78	1
JKO2+E 4	-	-	0.78	1
JKO1+J 1	24	2	0.78	2
JKO1+J 2	> 24	2	1.56	1
JKO2+J 1	12	2	3.125	1
JKO2+J 2	12	1	1.56	1
JKO2+J 3	-	-	1.56	1
JKO2+J 4	-	-	1.56	1

Even there is no data for some presented isolates, we notice there is some variability between isolates of the same strain. Isolate WT+E 3 shows extreme resistance to INH when compared to other WT+E isolates, so we believe it is not representative of the real phenotype for this stain. Rather, it may have acquired a spurious mutation that confers INH resistance. The MIC values that more likely represent the real phenotype for WT+E is 6 μg/mL for INH and 1.56 μg/mL for RIF since they match more closely the untransformed WT strain. Taneja and Tyagi, 2007 reported, using a similar assay, that the MIC for INH is 4 μg/mL for WT *M. smegmatis* mc^2155 strain, which is similar to the results obtained by our untransformed WT (3 μg/mL). For RIF, on the other hand, the reported MIC values are between 0.5 and 1 μg/mL (Taneja and Tyagi, 2007) which are a little lower than the 1.56 μg/mL but still can be considered similar.

When considering the WT+J strain, the INH MIC is most likely 6 μg/mL which is comparable to the WT+E isolates, and the RIF phenotype is likely between 3.125 μg/mL and 6.25 μg/mL. The WT+J strains appear to be slightly more resistant to RIF than WT+E, suggesting that having more RNase J in the cell gives the mycobacteria an advantage in the presence of RIF.

For the JKO isolates the MIC seems to be somewhere between 3 μg/mL and 6 μg/mL for INH and between 0.78 μg/mL and 1.56 μg/mL for RIF. Lack of RNase J therefore seems to have little impact on the ability of these bacteria to grow in the presence of INH and RIF, although there may be a subtle increase in sensitivity to RIF.

For the complementation strain isolates (JKO+J), the INH MIC appears to be between 12 μg/mL and 24 μg/mL, while for the RIF MIC the value is more likely to be 1.56 μg/mL. When looking at the INH phenotype, the JKO+J isolates show more tolerance to INH than WT+E and WT+J. Comparing these results with RNase J expression levels makes it difficult to draw any obvious conclusion, as the strain with more RNase J transcript levels (WT+J) did not show the higher MIC value, while the strain with no detectable RNase J expression (JKO) had the lower MIC value.

When looking at the RIF results, on the other hand, we see the WT phenotype restored in the JKO+J isolates. Is also possible to observe a trend in these results: more expression of RNase J will modestly increase tolerance to RIF stress, while no RNase J at all will modestly increase sensitivity to this particular drug. This trend correlates with RNase J relative expression values from section 5.1.

5.3. ANTIBIOTIC KILLING ASSAYS

The MIC assays shown in the previous section are only able to address if bacteria are growing or not under the tested conditions and, in this case, a lack of growth does not necessarily mean death. Bacteria may persist in a viable but non-growing state, and resume growth when conditions improve. For this reason MIC assays are often complemented with a killing assay.

The following figure shows the results for the INH killing assay comparing strains with the empty vector and without the empty vector.

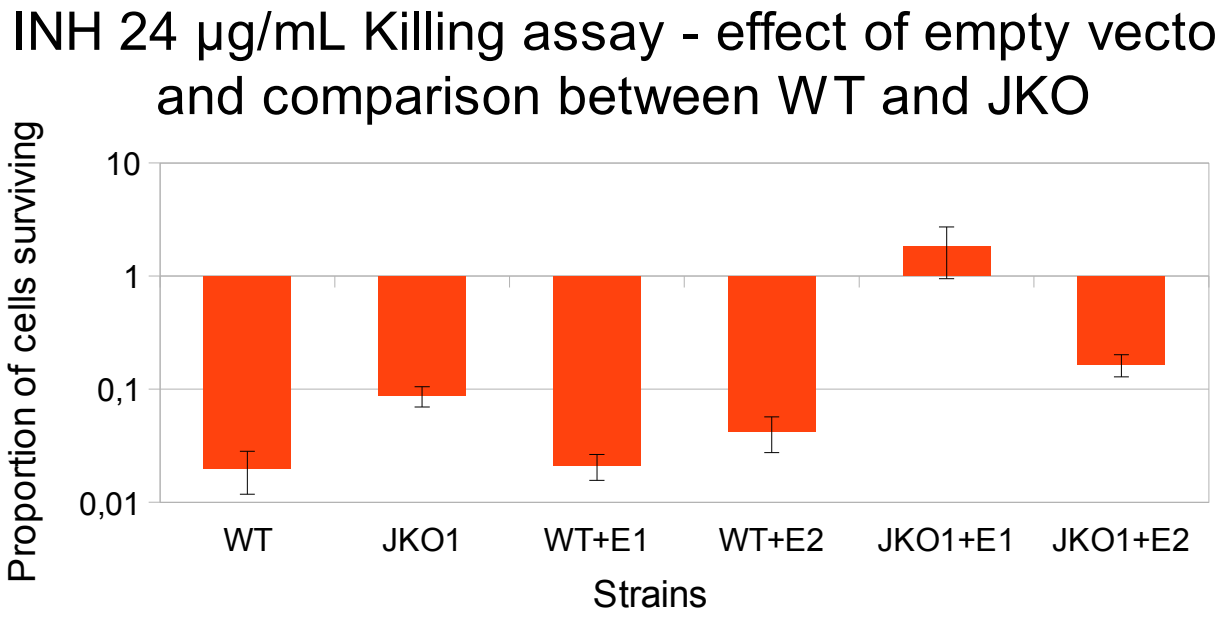

Figure 31: INH 24 µg/mL killing assay to compare strains containing the empty vector (WT+E1, WT+E2, JKO1+E1 and JKO1+E2) and without the empty vector (WT and JKO1). Survival rate is calculated by dividing CFU on drug-free plates after 24h of exposure to drug in liquid media by input CFU. Error bars show one standard deviation.

Figure 31 shows that there is an apparent difference between survival of WT and JKO1, where JKO1 strain seems able to survive better when challenged with INH. It also shows that while transforming the empty vector in the WT strain appeared to have no effect on survival, it did affect the survival when transforming the RNase J KO strain (JKO1) as JKO1+E1 looks different from JKO1 although is unclear if JKO1+E2 is different from JKO1. It is worth noticing that JKO1+E1 and JKO1+E2 are colonies derived from the same transformation plate and they both came from the JKO1 strain.

JKO1+E1 presents an extreme phenotype when challenged with 24 µg/mL INH, where it is actually able to grow while the other strains are killied. Since this is not seen in the other RNase J KO strains, it suggests that for this particular isolate some other mutation may have happened when the transformation was performed.

This graph also suggests the existence of variability between strains originated from the same common ancestor. For this reason, we decided to test the other available isolates from the initial RNase J KO transformation plate using the same drug concentration of 24 µg/mL INH (figures 32, 33, 34).

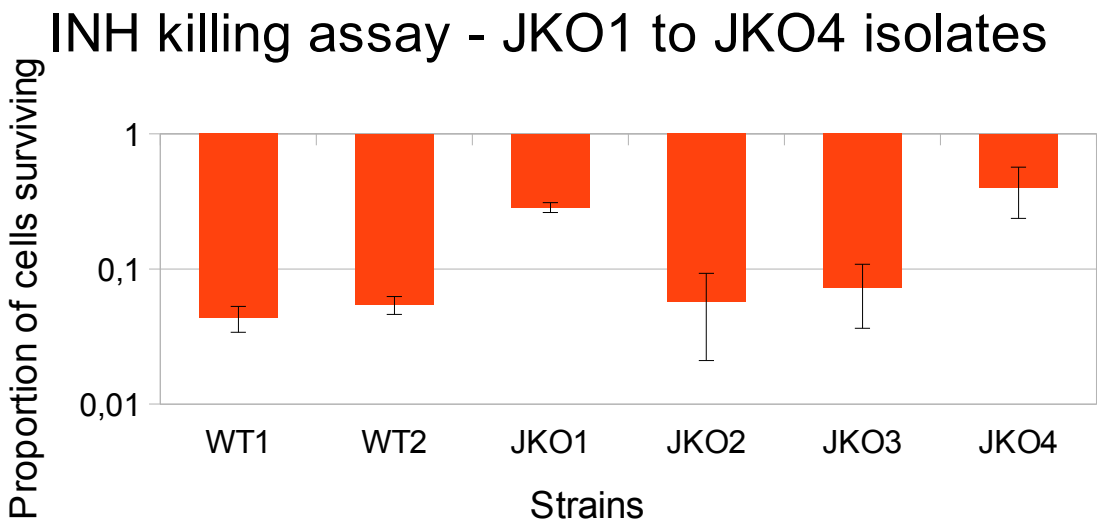

Figure 32: INH 24 µg/mL killing assay using isolates JKO1 to JKO4. Survival rate is calculated by dividing output CFU by input CFU. Error bars show one standard deviation.

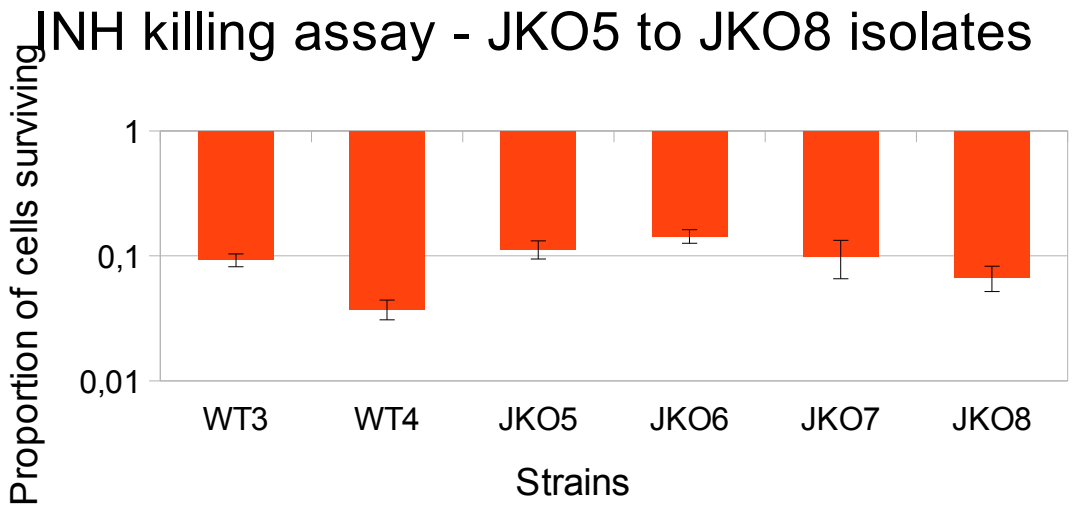

Figure 33: INH 24 µg/mL killing assay using isolates JKO5 to JKO8. Survival rate is calculated by dividing output CFU by input CFU. Error bars show standard deviation.

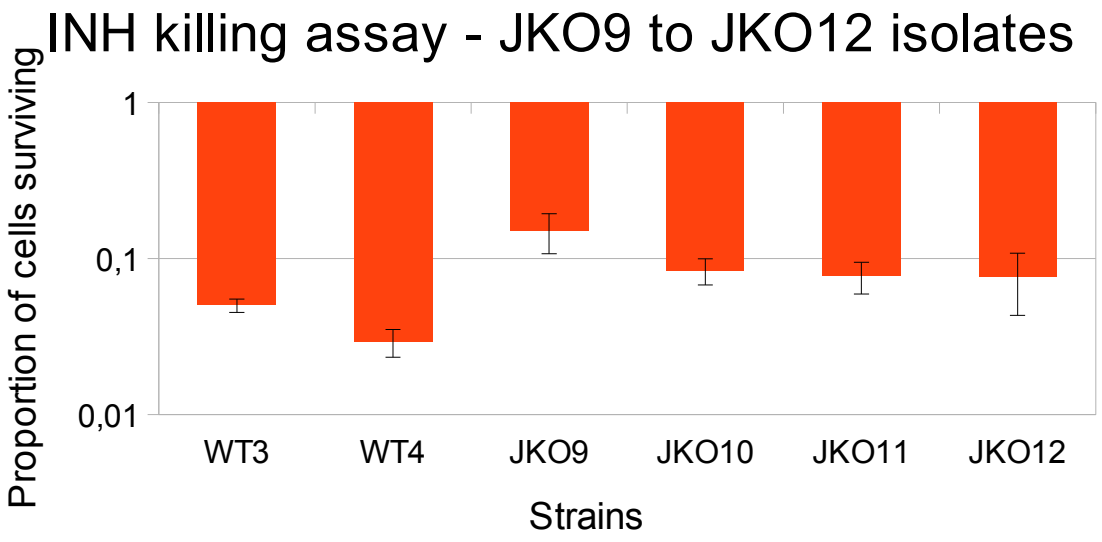

Figure 34: INH 24 µg/mL killing assay using isolates JKO9 to JKO12. Survival rate is calculated by dividing output CFU by input CFU. Error bars show one standard deviation.

The 3 figures above show the phenotypes for colonies JKO1 to JKO12 in order to address the RNase J KO resistance to INH. For colonies JKO1 and JKO4 we observed less killing of the bacteria than for the WT colonies. On the other hand, colonies JKO2, JKO3, JKO5, JKO6, JKO7, JKO8, JKO9, JKO10, JKO11 and JKO12 did not seem to have any difference when compared to WT colonies. This leads us to conclude that the effect on RNase J KO survival under INH induced

stress conditions is not large enough, if any, to be detected by this assay.

5.4. OXIDATIVE STRESS GROWTH CURVE

In order to test if RNase J has a role in oxidative stress tolerance in mycobacteria, cells were grown in liquid media with 3.2 mM H_2O_2 for around 30 hours (figure 35).

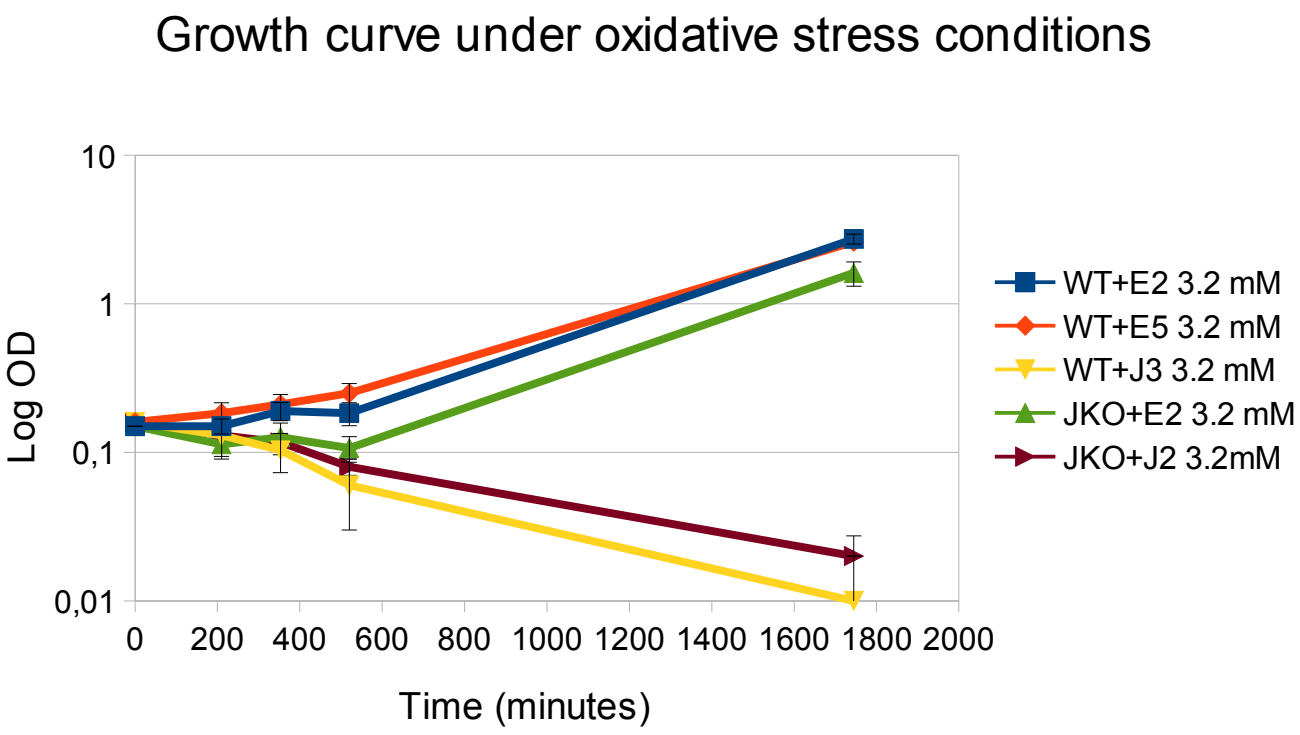

Figure 35: Graph showing the phenotype of some isolates of the RNase J related strains when challenged with H_2O_2. Error bars are one standard deviation.

The graph shows that the strains WT+E2, WT+E5 and JKO1+E2 are able to grow when 3.2 mM H_2O_2 was added to the media at time zero. JKO+E2 is also shown to have an overall slower growth when compared with the WT+E2 and WT+E5 strains in the presence of 3.2 mM H_2O_2. On the other hand, strains WT+J3 and JKO1+J2 did not show any growth after 30 h of culture. We were surprised that the JKO+J2 strain did not grow, as we had expected it would have a similar phenotype to WT+E. This result suggests that excess of RNase J transcript, as shown is section 5.1

that WT+J and JKO+J have more relative expression of RNase J than WT+E, confers a disadvantage when facing oxidative stress conditions.

5.5. EXPRESSION LIBRARIES FOR RNASE J OVEREXPRESSION AND COMPLEMENTATION STRAINS

Expression libraries were constructed using the 4 sets of strains (WT+E1, WT+J1, JKO1+E1 and JKO1+J1) to address how RNase J could change the cell transcriptome. Briefly, RNA was extracted from late log phase cultures and rRNA was depleted. After fragmentation of the RNA, the transcripts were used as template for cDNA synthesis and adapters were attached to both ends of the cDNA fragments by PCR reactions. Given the fact that the JKO1+E1 isolate had drug assay phenotypes that were not seen in other JKO isolates, we suspect that this particular isolate has an additional mutation. For this reason the results obtained for this strain expression library will not be discussed in this section.

Results obtained from the Bioanalyzer (Figure 33) for the rRNA depleted samples for strains WT+E, WT+J and JKO+J indicate that the WT+E1 profile is quite different from WT+E2 and the WT+J and JKO+J isolates, showing a greater proportion of high molecular weight RNAs. We concluded from that the profile from WT+E1 is not representative of this strain and suspected that something aberrant occurred during the rRNA depletion step.

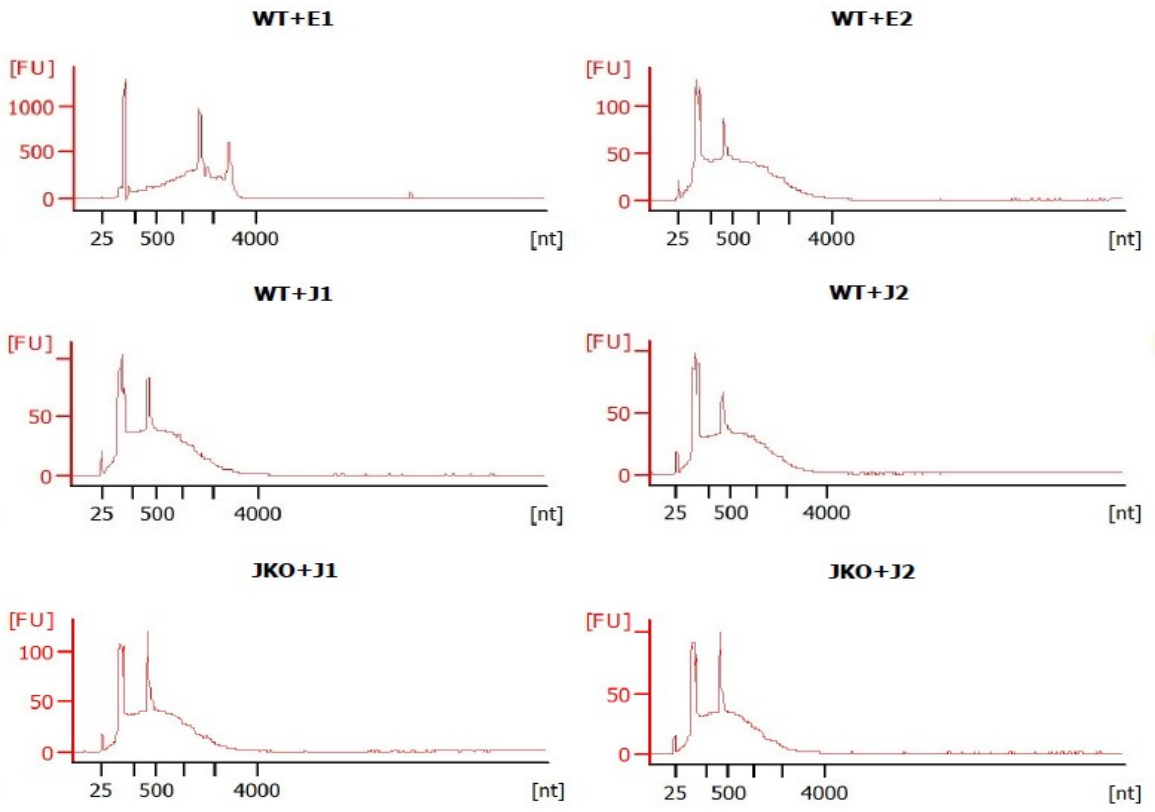

Figure 36: Bioanalyzer results for WT+E, WT+J and JKO+J samples used for library construction.

The fragment size of the library WT+E1 replicate was high, so we decided not to sequence it since it would not be multiplexed properly with the other libraries. In order to maintain 2 replicates per sample, we used RNA from the the original WT+E1 in order to construct another library (WT+E1B). This time, the rRNA depletion was performed using a different rRNA depletion kit and when compared with WT+E2 the sequence coverage from the two libraries revealed differences in some genes. This observation, combined with the observation of near complete rRNA depletion, led us to the hypothesis that some mRNAs could be depleted along with the rRNA and that using more input material might alleviate this.

Coming back to the profiles from figure 36, it is possible to conclude that the rRNA depletion reaction depleted similar material from samples WT+E2, WT+J1, WT+J2 JKO+J1 and JKO+J2, but not for WT+E1. For this reason sample WT+E1 was excluded from the next analysis.

When analyzing the $\log_2$ ratios of the WT+J and WT+E and applying the cutoff of 2 fold

difference and an adjusted p-value smaller than 0.05 is was possible to obtain 25 upregulated genes and 1 downregulated gene. The low number of differentially expressed gene may be explained by the lack of statistical power due to the fact only one replicate of WT+E strain is being used.

Libraries for the JKO+J were also constructed, and since the complementation vector is using a stronger promoter than RNase J native promoter, this strain can also be considered an overexpression strain although it has less RNase J expression then the WT+J strain. The JKO+J strain, after applied the cutoff mentioned earlier, had 45 upregulated genes and 4 downregulated genes. The following figure show the comparison between the differentially expressed genes from WT+J and JKO+J when compared to WT+E:

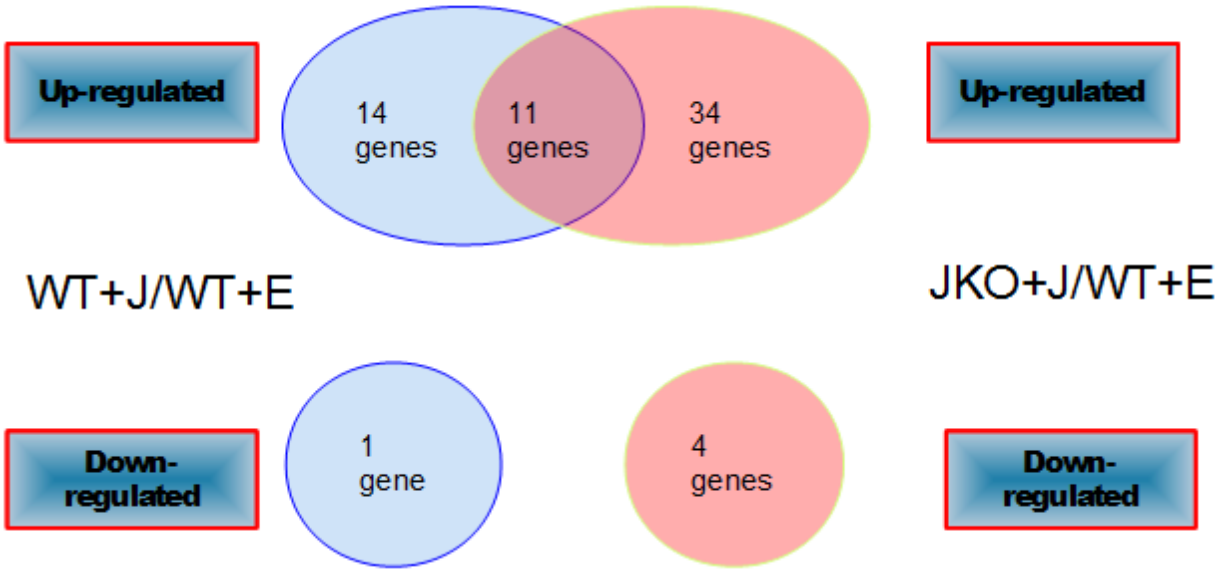

Figure 37: Overall analysis of number of genes up and downregulated for the RNase J overexpression and complemented strains, and the overlap between the 2 sets of genes.

The figure shows an overlap of 11 genes for the upregulated lists, but no overlap when looking at the downregulated genes. By looking a little closer to the upregulated gene list for the JKO+J/WT+E condition, we noted see the upregulation of the genes MSMEG_3928, MSMEG_3929 and MSMEG_3930 (Figure 38); which encodes for [NiFe] hydrogenase subunits alpha, delta and beta, respectively.

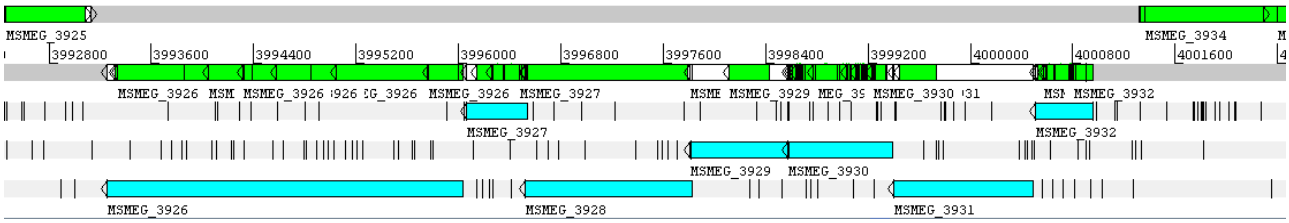

Figure 38: Location of the operon which contains genes MSMEG_3928, MSMEG_3929 and MSMEG_3930. Screen shot of the view from Artemis (Rutherford *et al*, 2000).

The previous picture shows the location of the operon where the genes of interest are encountered in the minus strand of *M. smegmatis* genome. It is worth noting that the gene MSMEG_3927 which encodes the protein peptidase M52, a hydrogen uptake protein, and gene MSMEG_3932 which encodes the heat shock protein hspX, are also upregulated in the JKO+J strain.

Previous studies have shown that genes MSMEG_3931-MSMEG_3928 encode a enzyme known as Hyd3 which is regulated by the DosR regulon (Berney *et al,* 2014). DosR is shown to act together with DosS and DosT, and regulates response under hypoxic conditions by maintaining the redox balance inside the cells (Berney *et al,* 2014). This balance is then maintained by activation of hydrogen production by 3 different hydrogenases: Hyd1, Hyd2 and Hyd3 (Berney *et al,* 2014). The $\log_2$ values obtained for those genes are around 6.34. This suggests that the JKO+J strain can not maintain the redox balance in the cells which may explain the absence of growth under oxidative stress conditions.

Table 23: Differentially expressed genes according with the 2 fold and p-value 0.05 cutoff for strains WT+J and JKO+J when compared with the WT+E. Values in italic are not significant.

Orthologs in TB	Synonym	Product	log2 ratio WJ/WE	log2 ratio JJ/WE	Function
	MSMEG_0462	MmpS4 protein	*-0,77*	-3,09	Membrane protein
	MSMEG_3937	membrane protein	2,81	5,56	Membrane protein
	MSMEG_3946	transmembrane protein	*0,77*	2,86	Membrane protein
	MSMEG_5602	Possible exportin-1 protein	4,00	*3,70*	Membrane protein
Rv0242c	MSMEG_0372	3-ketoacyl-ACP reductase FabG	*0,31*	2,69	Metabolism
	MSMEG_2262	hydrogenase-2, small subunit HybA	1,17	1,74	Metabolism
	MSMEG_2604	4-carboxymuconolactone decarboxylase PacC	-3,81	*-3,81*	Metabolism
Rv1594	MSMEG_3199	quinolinate synthetase NadA	*0,61*	2,53	Metabolism
Rv1622c	MSMEG_3232	cytochrome D ubiquinol oxidase subunit II Cyd B	*1,20*	2,84	Metabolism
	MSMEG_3465	fatty-acid—CoA ligase	1,19	2,58	Metabolism
	MSMEG_3927	peptidase M52, hydrogen uptake protein	*1,00*	5,86	Metabolism
	MSMEG_3928	[NiFe] hydrogenase subunit alpha	*2,32*	6,02	Metabolism
	MSMEG_3929	[NiFe] hydrogenase subunit delta	*2,00*	6,61	Metabolism

Table 23: continued

Orthologs in TB	Synonym	Product	log2 ratio WJ/WE	log2 ratio JJ/WE	Function
	MSMEG_3931	[NiFe] hydrogenase subunit beta	2,17	6,34	Metabolism
	MSMEG_3933	diacylglycerol O-acyltransferase	0,89	2,70	Metabolism
	MSMEG_3935	integrase	0,83	2,16	Metabolism
Rv2004c	MSMEG_3942	Possible adenylate kinase	0,64	2,67	Metabolism
Rv3130c	MSMEG_3948	acyltransferase	1,58	4,55	Metabolism
Rv2032 / Rv312	MSMEG_3952	NAD(P)H nitroreductase	0,79	2,33	Metabolism
Rv3131	MSMEG_3955	NAD(P)H nitroreductase	0,38	4,79	Metabolism
Rv2379c	MSMEG_4510	peptide synthetase mbtf	0,32	3,95	Metabolism
Rv2380c	MSMEG_4511	linear gramicidin synthetase subunit B	0,42	3,17	Metabolism
	MSMEG_4512	polyketide synthase	0,00	4,07	Metabolism
Rv2383c	MSMEG_4515	non-ribosomal peptide synthetase	0,54	3,72	Metabolism
Rv3130c	MSMEG_5242	acyltransferase	-0,13	3,15	Metabolism
Rv2032	MSMEG_5246	NAD(P)H nitroreductase	0,33	3,02	Metabolism
Rv0983	MSMEG_5486	peptidase S1 and S6, chymotrypsin/Hap	-0,84	2,63	Metabolism
	MSMEG_5506	Possible oxidoreductase	3,70	3,46	Metabolism
Rv3588c	MSMEG_6082	carbonic anhydrase	-0,97	2,52	Metabolism
	MSMEG_6511	acyl-CoA dehydrogenase	1,03	0,93	Metabolism
	MSMEG_6512	acyl-CoA dehydrogenase	1,08	1,64	Metabolism
	MSMEG_3926	cation-transporting ATPase Pma1	0,65	3,28	Metal binding
	MSMEG_6422	ferritin family protein	1,03	1,32	Metal binding
Rv2470	MSMEG_4695	globin	0,82	2,08	Oxygen binding
	MSMEG_3755	5S ribosomal RNA rrfA	4,68	5,50	Protein regulation
	MSMEG_3756	23S ribosomal RNA rrlA	2,91	3,00	Protein regulation
	MSMEG_4929	5S ribosomal RNA rrfB	4,43	3,97	Protein regulation
	MSMEG_4930	23S ribosomal RNA rrlB	2,91	3,00	Protein regulation
Rv2752c	MSMEG_2685	metallo-beta-lactamase superfamily protein (Rnase J)	2,32	2,85	Ribonuclease
Rv2031c	MSMEG_3932	Predicted to be in the DosR|Rv3133c regulon	3,08	5,22	Stress response
Rv2026c	MSMEG_3940	universal stress protein family protein	1,14	4,35	Stress response
Rv1996 / Rv2005c / Rv2623	MSMEG_3945	universal stress protein family protein	3,17	7,41	Stress response
Rv2026c	MSMEG_3950	universal stress protein family protein	0,10	2,80	Stress response
Rv2027c / Rv3132c	MSMEG_5241	GAF family protein	0,78	3,73	Stress response
	MSMEG_5245	universal stress protein family protein	0,47	3,24	Stress response
	MSMEG_5733	universal stress protein family protein	-0,09	2,95	Stress response
	MSMEG_2682	transcriptional regulator	0,00	4,09	Transcriptional regulator
	MSMEG_4825	TetR family transcriptional regulator	-1,46	-3,46	Transcriptional regulator
Rv3133c	MSMEG_5244	DevR family transcriptional regulator	1,31	2,84	Transcriptional regulator
	MSMEG_1786	STAS domain-containing protein	0,35	1,54	Transporter
	MSMEG_2834	Cys tRNA	1,38	0,46	tRNA
	MSMEG_5757	Glu tRNA	2,47	2,28	tRNA
	MSMEG_6326	Ser tRNA	1,66	1,11	tRNA
	MSMEG_6350	Ser tRNA	2,20	1,13	tRNA
	MSMEG_0009	hypothetical protein	1,33	1,27	unknown
	MSMEG_0987	hypothetical protein	-0,30	2,92	unknown
	MSMEG_1212	hypothetical protein	-1,00	-3,32	unknown
	MSMEG_1421	hypothetical protein	1,14	1,42	unknown
	MSMEG_2826	hypothetical protein	1,29	1,33	unknown
	MSMEG_3439	hypothetical protein	0,97	1,45	unknown
	MSMEG_5154	hypothetical protein	1,39	1,97	unknown
	MSMEG_5406	hypothetical protein	1,06	1,45	unknown
	MSMEG_5461	hypothetical protein	-1,35	-4,81	unknown
	MSMEG_6209	hypothetical protein	1,47	1,89	unknown

6. DISCUSSION

6.1. OXIDATIVE STRESS PROFILES

Mtb will produce oxidative stress as it grows aerobically (Broden *et al*, 2016), so it evolved to have a array of protective enzymes against this stressor (Dussurget and Smith, 1998). Oxidative stress has also been reported to be one of the types of stress Mtb will face when infecting a host (Dussurget *et al*, 1998). With the objective of understand the mechanisms by which mycobacteria can avoid the host immune "attacks", RNA-seq libraries ans 5' end mapping libraries were constructed with *M.smegmatis* cells cultured with and without hydrogen peroxide. With this libraries was possible to look into how mRNA profile changes can help the mycobacterial cell to improve their response against oxidative stress conditions.

First, the ideal concentration for the assays has found to be 3.2 mM H_2O_2. For this particular concentration, cell growth was impaired but not completely stopped. This way we believe we were able to get the strongest level of response by the cells. Cells were chosen to be harvested at three and six hours time points because *M.smegmatis* generation time is around 3 hours.

When both test conditions (3h 3.2 mM H_2O_2 and 6h 3.2 mM H_2O_2) where compared with no H_2O_2 a cutoff of 2 fold and p-value of 0.05 was applied to get the list of the differentially expressed genes under these test conditions. The first thing we can notice is that that is a great overlap among the upregulated samples (76 genes) which is not seen in the downregulated genes where only 9 genes were present in both samples. For the 3 hour time point, the majority of the differentially expressed genes were upregulated while for the 6 hour time point the majority of the differentially expressed genes were downregulated. This may be due to the fact that oxidative stress will induces reduction in protein production (Li *et al*, 2015) to save energy for the cell. Some tRNAs were found on the downregulated genes for the 6 hour time point which corroborate the idea that translation if reduced. An analysis using the software DAVID was performed using these 2 genes lists and it did not shown any statistically significant results for the downregulated genes. The upregulated genes

on the other hand had enrichment of clusters, for both 3 hour and 6 hour time points, containing HNH nuclease genes, which was expected since they are know to be involved in DNA repair and were also previously shown to be upregulated in M. smegmatis under oxidative stress conditions (Pediaditakis et al, 2012 and Li *et al*, 2015). Another interesting set of genes found upregulated in our samples were MSMEG_6645, MSMEG_6646 and MSMEG_6647, which are involved in the methylcitrate cycle. This pathway was also reported to be upregulated in TB in response to oxidative stress (Voskuil *et al.*, 2011). We can speculate about what is happening in the cell under oxidative stress conditions and why it upregulates this pathway. One possibility would be that, after some period of exposure to excess hydrogen peroxide, cells start to see lipid damage and in order to restore balance will "discard" the damage structures and create new ones (Shell, personal communication). Once the old lipids, which were damaged, are no longer useful, the cell might use the β-oxidation pathway and the methylcitrate cycle to "recycle" its components (Shell, personal communication).

Our data was also compared to Li *et al*, 2015 which tested exposure of culture of *M. smegmatis* in early log phase to 30 minutes of 0.2 mM H_2O_2 and 7 mM H_2O_2. A visual representation of this comparison can be seen on figure 17. Some clusters of genes showed similar trends for the 2 data sets, like for example the upregulation of DNA repair genes as discussed previously on section 4.2. For some others clusters on the other hand, like sigma factors, transcription regulation and protective enzymes, there was not a clear trend that could be visualized in this data. It is interesting to notice that KatG was not found upregulated in any of the conditions tested in this study, which might indicate that the response seen to 3.2 mM $H2O_2$ after 3 and 6 hours are capturing the profile of RNA expression of "late" responses to $H2O_2$. Also is important to noticed that the libraries constructed in this study where using cells in late log phase and as previously reported, the phase which a culture is can influence in its RNA expression profile (Shemesh et al., 2006).

When looking at the 5' end mapping libraries we can see that due to technical problems during the library preparation, the reproducibility of the biological replicates was lower than expected (figure 23). This impaired our ability to distinguish between transcription start sites and cleavage sites and for this reason a relatively high cutoff for the coverage of the converted libraries was used, in order to compensate for those technical issues. In order to differentiate the transcription start sites from the cleavage sites, the $\log_2$ ratio between the converted and non-converted libraries from the same condition is taken. If a 5' end is a transcription start site, we expect it will have a $\log_2$ ratio greater than 0. The higher the ratio, the more likely that this specific 5' end is a transcription start site. On the other hand, if the $\log_2$ ratio between the converted and non-converted libraries is around 0, then this 5' end is likely to be a cleavage site.

We then took the 500 highest ratios for each condition tested, which should be representative of transcription start sites, and found their sequence logo. Similarly, the 500 lowest ratios for each condition, which should represent the cleavage sites, were run through a sequence logo software. As expected, the higher ratios did show and A and T rich region around -10 site and a purine (A or G) for the +1 nt position, which are characteristic of *E. coli* transcript start sites (Lodish *et al*, 2000) and for the lowest ratios a C was the most common nt found in position +1 which resembles Mtb cleavage sites (Shell *et al*, unpublished data).

Z-scores were then calculated to find the outliers in figures 26B to 26D in order to address which 5' ends were differently expressed in which comparison between 2 conditions. The cutoff of z-score of 3 was used in order to get the most significant outliers. Among the 3 conditions analyzed (no$H2O_2$ vs $H2O_2$ 3h, noH_2O_2 vs H_2O_2 6h and H_2O_2 3h vs H_2O_2 6h) 80 differentially expressed 5' ends were found (table 25).

Future work for this project consists in analyze the remaining genes in the differentially expressed list and look into more detail to the differentially expressed 5' ends in order to find specific transcripts that get cleaved under oxidative stress conditions. All this knowledge will help

understand how mycobacteria is able to tolerate and avoid the effect of high ROS in its environment. The genes that have enhanced expression or transcript stability when under oxidative stress can be used as target for new drugs, for example, once blocking normal oxidative stress response may be critical in killing bacteria and avoid bacterial dormancy when infecting a host.

6.2. RNASE J KNOCKOUT AND OVEREXPRESSION PHENOTYPES

In order to understand the roles of RNase J and how this enzyme can change RNA profiles under stress response conditions, a strain with deletion of RNase J was created. Since mycobacteria is known by its low homologous recombination rates (Van Kessel and Hatfull, 2008) a strategy know as recombineering was used where a plasmid in transformed into WT strain to facilitate the exchange of genetic material with linear DNA fragments (Van Kessel and Hatfull, 2008). The recombineering strategy helped indeed with the success hate of the transformation and the KO strain could be constructed. As control proposes, the complementation of the deletion strain was performed by adding RNase J coding sequence back to the bacterial genome in order to restore this gene activity.

Besides testing the effects of the lack of RNase J in the cell, we decided to construct an overexpression RNase J strain to address any possible vantage or advantage excess of this enzyme could bring to those cells and have a more complete idea of the possible mechanism by which RNase J work in the cell.

Initially, those set of four strains (WT with empty vector control, overexpression, deletion strain with empty vector control and complemented KO) had their RNase J relative expression determinated (figure 27) and results came as expected, with the overexpression strain having the highest relative expression, followed by the complemented strain, WT and RNase J deletion, which relative expression was undetectable. The complementation strain, as discussed in section 5.1., had more expression than WT because the promoter used for the vector construction was a synthetic

promoter which showed to be stronger than the native promoter. For these particular set of 4 isolates growth rates, antibiotic tolerance, colony morphology assays were conducted (data not shown). The strongest phenotypes seeing were a higher tolerance to INH on killing assays for the deletion strain, as can be seen for strain JKO1+E1 on figure 31, which was not complemented and a defect in colony morphology for the JKO1+E1 which was restored in the complemented strain (data not shown). Data from our collaborators from Fortune's lab using clinical Mtb isolates also indicates that deletion of RNase J confers cells a higher tolerance to INH and Figaro *et al*, 2013 showed previously that deletion of RNase J in will cause a cell morphology defect in *B. subtilis*.

With all these information in hand we then decided to test the reproducibility of the phenotypes we were seeing with those strains by using different isolates from the same strains as shown in sections 5.2. and 5.3. both MIC assays for INH and RIF and the killing assay for INH were showing variability in RNase J deletion strains phenotypes. Even with the variability reported for the isolates from the same strains, some trends about the antibiotic tolerance assays can be draw. The RNase J deletion strain appear to have a MIC smaller than WT for both tested drugs (INH and RIF) and the overexpression strain seems to have equal or higher MIC values than WT. For the complemented strain observed values is more difficult to suggest an overall trend since they don't correlate with the relative RNase J expression levels once for INH the MIC were higher than WT and overexpression strains and for RIF the MIC value were between the WT and RNase J deletion values (table 15). Since the MIC assay is based in a reduction reaction, it was hypothesized that, due the fact the JKO+J showed higher upregulated values for hydrogenases subunits, the high values seen for strain JKO+J on the INH MIC assay would be not real. On the other hand this is not true when we look into the RIF MIC assay results, which suggests a similar trend as observed in the qPCR reaction for the relative expression level. For the INH killing assays performed for this project is also possible to see some variability among the isolates. While JKO1, JKO4, JKO6 and JKO9 isolates seems to have more resistance to INH challenge, the majority of the isolates (8 out of

12) suggest that RNase J there is little to no effect in improving mycobacteria survival of INH. We plan to perform a whole genome sequencing and southern blots assays in order to identify the possible mutations that these isolates might have that make then a little more tolerant to INH than WT strain, specially on isolate JKO1+E1, which were showing an extreme phenotype in the killing assays performed during this project.

Next, we decided to challenge the RNase J deletion strain with hydrogen peroxide. Oxidative stress has being shown to contribute with INH toxicity (Dussurget and Smith, 1998) so we wanted to investigate if the selection for RNase J deletion strains saw by our collaborators could be actually due to the tolerance to oxidative stress and not so much by the drug itself. To address that cells were cultured with 3.2 mM H_2O_2 for about 30 hours and growth was measured by taking OD_{600}'s periodically. Surprisingly, our preliminary data shows that RNase J is capable of growing under oxidative stress conditions, although not in the same rate as WT (Figure 35). The overexpression and the complemented strains did not shown any growth on the conditions tested. These results suggests that excess of RNase J is detrimental to the cell when challenged with H_2O_2. It is important to notice that after 30h of incubation, the levels of H_2O_2 in the samples are not significant anymore which can lead us to conclude that the cells of the strains WT+J and JKO+J are likely non-viable anymore. It would be also interesting to look at other ROS generator, which might be more stable than hydrogen peroxide but also induce a different from what we have seen from this work.

Finally, the expression libraries constructed for the complemented strain showed one operon where almost all genes were upregulated. MSMEG_3928, MSMEG_3929 and MSMEG_3930 (figure 35) encodes for [NiFe] hydrogenase subunits alpha, delta and beta respectively are part of these operon and together form the protein Hyd3, which is regulated by the DosR stress response regulon (Berney *et al,* 2014). Although this result suggests that more expression of RNase J triggers stress reponse, those genes are not found to be upregulated in the overexpression strain. The only

gene that belong to this specif operon that is upregulated in both complemented and overexpression strains is MSMEG_3932, which encodes the heat shock protein hspX. More investigation at the differentially expressed genes needs to be done in order to completely understand RNase J function in mycobacteria.

As future directions for this project, we can cite the need to find the real phenotype for the RNase J deletion strain in order to understand the roles this enzyme may have in mRNA profiles changes in mycobacteria and possible oxidative stress tolerance. It would also be interesting to perform whole genome sequencing on the JKO1+E1 strain to verify what kind of additional mutation happened during the transformation procedure and how this had such a great influence in the strains phenotype. The data collected from the expression library for this strain can also help to find this possible mutation by looking at the possible SNPs in its genome when compared to WT genome.

7. APPENDIX

7.1. PRIMERS SEQUENCES:

Table 24: Sequences of the primes used during this project in alphabetical order. N means any of the 4 nucleotides chosen randomly and the red sequences shows the unique sequences of each index used for library construction.

Primer Name	Sequence 5' to 3'
Bioo kit Index 1	CAAGCAGAAGACGGCATACGAGATGTGTTCTAGTGACTGGAGTTCAGACGTG
Bioo kit Index 2	CAAGCAGAAGACGGCATACGAGATAAACATCGGTGACTGGAGTTCAGACGTG
Bioo kit Index 4	CAAGCAGAAGACGGCATACGAGATGTCTGTCAGTGACTGGAGTTCAGACGTG
Epicentre Kit Index 7	CAAGCAGAAGACGGCATACGAGATGTCTAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Epicentre Kit Index 8	CAAGCAGAAGACGGCATACGAGATTGAACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Epicentre Kit Index 9	CAAGCAGAAGACGGCATACGAGATCTAGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Epicentre Kit Index 10	CAAGCAGAAGACGGCATACGAGATATCGAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Epicentre Kit Index 11	CAAGCAGAAGACGGCATACGAGATCCGATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Epicentre Kit Index 12	CAAGCAGAAGACGGCATACGAGATGAACATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
JR 273	GACTACACCAAGGGCTACAAG
JR 274	TTGATCACCTCGACCATGTG
SSS 392	TCCCTACACGACGCTCTTCCGAUCU
SSS 397	CTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN*
SSS 398	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
SSS 399	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCT
SSS 401	AATGATACGGCGACCACCGAGATC
SSS 402	CAAGCAGAAGACGGCATACGAGAT
SSS 403	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCT
SSS 583	ACTAGCGTACGATCGACTGC
SSS 584	ACCAGCCCGTCATCGTCAAC
SSS 589	ATCTGCCGGTCGTGCTGTAC

SSS 590	GTTGACGATGACGGGCTGGTGAATCCCCTACCTGAGGACC
SSS 591	GCAGTCGATCGTACGCTAGTGCGATCTGAGCTCAACCCG
SSS 592	TCGACATGGGCAACGCGCT
SSS 593	GGTCTGCTCGCACACTTCAC
SSS 594	GCACAGCATCCGGGTGAAC
SSS 644	GTGAGCGCTCACAATTCGGATCCAGGAAGGAGATATACATATGAGCGCCGAACTCGC
SSS 645	TCAGTTAACTACGTCGACATCGATATCAGATCTCTATGACGGTCG
SSS 651	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCT
SSS 653	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGACGTGTGCT
SSS 654	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTTCAGACGTGTGCT
SSS 655	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTGCT
SSS 656	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGACGTGTGCT
SSS 657	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGACGTGTGCT
SSS 658	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTGCT
SSS 659	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTGCT
SSS 660	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCT
SSS 661	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTTCAGACGTGTGCT
SSS 706	TCATCCTCTCATCGGGTTTC
SSS 707	TTCGCGCTCAACCTTCT
SSS 708	ACGGGCGACATCAAACT
SSS 709	TGGAATCGCACAGGAACA
SSS 710	GTGACGGTCGGCAAGAT
SSS 711	AAACCCGATGAGAGGATGAG
SSS 786	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTG
SSS 787	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTG
SSS 789	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGAC

	GTG

7.2. PRIMERS EFFICIENCY FOR qPCR

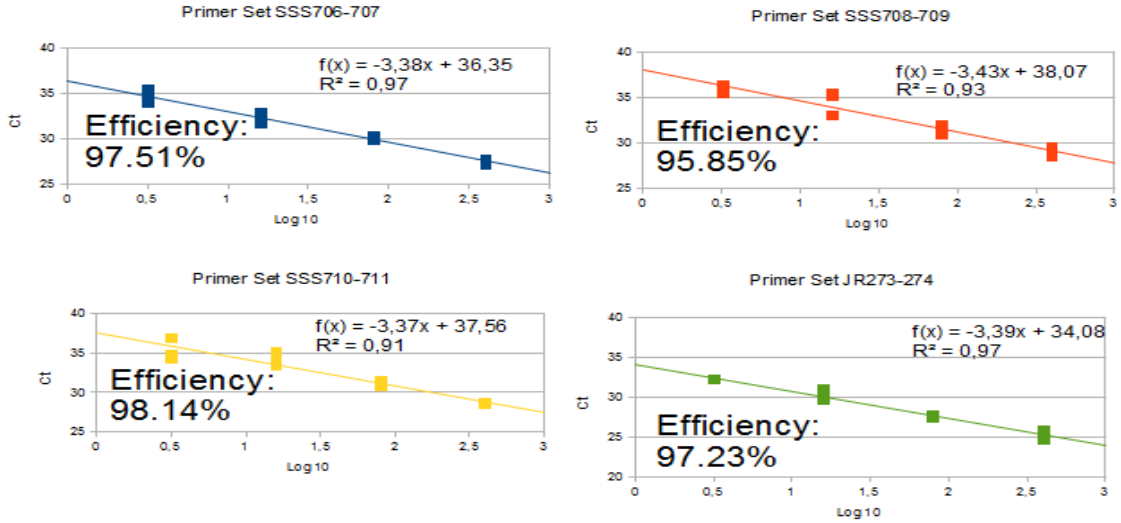

Figure 39: Primers tested prior to run the relative expression qPCR for RNase J. Primer set SSS706-707 was the one chosen for the following experiments and primer set JR273-274 are the validation primers that amplify a region of the housekeeping gene sigA.

7.3. Z-SCORES VALUES

Table 25: Z-scores for the 5'end mapping libraries. The green values show the significant values with the cutoff of a Z-score of more than 3. The unique identifier correspond to the nucleotide position at the genome with the strand were the gene is located (plus or minus strand).

Unique identifier	Z-score log2 ratio noHP vs HP3h	Z-score log2 ratio noHP vs HP6h	Z-score log2 ratio HP3h vs HP6h	Unique identifier	Z-score log2 ratio noHP vs HP3h	Z-score log2 ratio noHP vs HP6h	Z-score log2 ratio HP3h vs HP6h	Unique identifier	Z-score log2 ratio noHP vs HP3h	Z-score log2 ratio noHP vs HP6h	Z-score log2 ratio HP3h vs HP6h
1032818+	-1,1061912	3,87807589	6,29473264	3477089+	-5,10009419	-5,3561436	-0,1247135	5024564-	-0,79753499	1,76084192	3,23987644
1038813+	2,07964751	3,3293146	1,48743319	3477165+	-4,78178344	-4,4226386	0,63470621	5026601-	-0,88919933	2,09636891	3,77926393
1294340-	-3,4511799	-1,4714694	2,61623856	3477205+	-3,87016666	-3,5549321	0,54450877	5029425-	-0,61809172	-3,44692675	-3,52459801
1294380-	-3,387478	-2,0451089	1,81432212	3515532+	-3,20745883	-2,4098122	1,12410347	5147013+	1,65368329	-0,80046186	-3,14211196
1357857-	-2,6311756	-3,1904423	-0,6001689	3820231-	-0,58485859	2,34325624	3,69547213	5671331-	-7,66401881	-3,08362506	6,04074835
1562175+	-0,2554792	2,23802442	3,13762002	3825092-	-0,64865626	-3,4822854	-3,5294342	5671762-	-8,22803291	-7,18677984	1,62308041
1588629-	3,75405687	3,30456817	-0,7084405	3885819-	-3,01130535	-2,5158699	0,73746226	5671797-	-8,89357662	-7,52485412	2,05948682
1709181+	-7,6124244	-6,366399	1,85622622	4056388-	-2,8757672	0,98207339	4,94994417	5672104-	-5,79509959	-3,99553575	2,48056115
1711452+	-9,2144274	-8,1575684	1,68065424	4136260-	-4,13840395	-3,2813142	1,23452936	5672511-	-5,32356441	-4,00632458	1,85738334
1711460+	-5,0587045	-4,0508216	1,45913208	4137678-	-3,13370957	-2,703434	0,66044284	5672518-	-3,68112844	-3,4042805	0,48907558
1775874-	3,89533335	0,47546746	-4,4398375	429893+	0,55933701	3,41502077	3,56053965	5672527-	-3,85834533	-2,63374948	1,68472978
1871477+	0,90118423	6,16947198	6,57367113	432120-	1,30204211	3,28497505	2,4371757	5672921-	-6,35685006	-5,57774017	1,22217335
2000523-	-2,4164163	-3,1599948	-0,8396372	4341187-	-3,51814263	-2,9636584	0,83105589	5673067-	-9,90431931	-8,82076582	1,74071527
2025892+	1,26028907	3,22624142	2,41748417	447458-	1,32952258	-1,584312	-3,7062456	5673136-	-6,92195415	-6,37511461	0,9525893
2081398-	3,12064116	2,31722366	-1,1279978	4489608+	-3,16908667	-1,6520335	2,02502792	5775565+	0,84824452	4,42925258	4,45923597
211961-	-3,817055	-2,3423506	1,99686798	4578686-	-3,75804089	-2,9923173	1,10526988	6230194-	-0,14250746	2,82925466	3,73318163
2155354+	0,98871505	3,26376283	2,81566614	4578695-	-3,68574398	-3,1449044	0,82039677	6316855-	-3,10410452	-2,61055513	0,73867128
2157543-	1,75924012	3,02324415	1,51775987	4578702-	-3,01029527	-2,1843708	1,15197922	6626813-	-4,08970496	-3,02259394	1,4960979
2239870+	3,05957839	1,08119562	-2,5994878	4578708-	-3,13289105	-2,3059966	1,15791849	6626821-	-4,79308468	-3,17343112	2,21628699
2480080-	0,91080093	3,02453141	2,61631585	4578714-	-3,34682369	-2,7666799	0,85664282	6626855-	-3,79116213	-2,95591582	1,193753
248214+	-3,4983613	-2,5747713	1,2932894	4606182+	0,76261295	4,1610777	4,23355766	6785794+	-3,07654607	-3,10247624	0,08598962
2792272+	-3,013094	-2,4583334	0,81194698	4658177-	-9,52105699	-8,3191874	1,87436338	6785824+	-3,07691926	-2,56365583	0,76235275
2793196+	-3,6534255	-2,8481957	1,15079517	4658631-	-9,55428926	-9,5242706	0,40570786	6831079-	-3,63027563	-3,49401705	0,31076541
2793220+	-3,1506596	-3,0504072	0,24712432	4672121-	0,08589116	2,90394558	3,53157673	6832182-	-3,55394082	-2,68044787	1,23259016
2813368-	-3,3114071	-2,4704236	1,18246829	4703074+	0,02720319	2,75560965	3,42138574	6926+	2,38743297	-0,14749265	-3,27170636
2813600+	-5,0659814	-4,8259028	0,4963009	4807572-	1,76024273	3,34907086	1,92517115	6942400-	0,83568219	4,23552504	4,23247152
3004282+	-3,2137327	-1,6952384	2,02855542	5017068+	-0,97963862	1,39369668	3,01478068				

7.4. MEDIA RECIPES

7.4.1. *M. smegmatis* MEDIA recipes

Recipe for 1 L of Middlebrook 10x ADC supplement:

- 1 L Milli Q water
- 50 g BSA fraction V
- 20 g Dextrose
- 8.5 g NaCl
- 30 mg Catalase

Media was filter sterilized and stored at 4 °C until use.

Recipe for 1 L of Middlebrook 10x ADS supplement (for oxidative stress experiments):

- 1 L Milli Q water
- 50 g BSA fraction V
- 20 g Dextrose
- 8.5 g NaCl

Media was filter sterilized and stored at 4 °C until use.

Recipe for 1 L of Middlebrook 7H9 Broth

- 900 mL Milli Q water
- 4.7 g 7H9 Broth powder
- 2.5 mL 20% Tween 80
- 4 mL 50% glycerol
- 100 mL 10x ADC or ADS

Media was filter sterilized and stored at 4 °C until use.

Recipe for 1 L of Middlebrook 7H10 Agar

- 900 mL Milli Q water
- 19 g 7H10 Agar powder
- 10 mL 50% glycerol

Media was autoclaved 20 minutes at 121 °C and when medium cooled down 100 mL ADC was added and antibiotic if needed. 25 mL of media was then poured in each petri dish and when solid stored at 4 °C until use. In the case of square plate 35 mL per plate is used.

7.4.2. *E. coli* MEDIA recipes

Recipe for 1 L of LB Broth

- 1 L Milli Q water
- 25 g LB Broth powder

Media was autoclaved 20 minutes at 121 °C and stored in room temperature until use.

Recipe for 1 L of LB Agar

- 1 L Milli Q water
- 40 g LB Agar powder

Media was autoclaved 20 minutes at 121 °C and stored in room temperature until use.

8. REFERENCES

Agrawal P, Miryala S, Varshney U. Use of Mycobacterium smegmatis Deficient in ADP-Ribosyltransferase as Surrogate for Mycobacterium tuberculosis in Drug Testing and Mutation Analysis: e0122076. *PLoS One*. 2015;10.

Altaf M, Miller CH, Bellows DS, O'Toole R. Evaluation of the Mycobacterium smegmatis and BCG models for the discovery of Mycobacterium tuberculosis inhibitors. *Tuberculosis*. 2010;90:333-337.

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ. Basic Local Alignment Search Tool. *Journal of Molecular Biology.* 1990;215:403-410. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Arraiano CM, Andrade JM, Domingues S, et al. The critical role of RNA processing and degradation in the control of gene expression. *FEMS microbiology reviews*. 2010;34:883-923.

Bardou F, Raynaud C, Ramos C, Laneelle MA, Laneelle G. Mechanism of isoniazid uotake in *Mycobacterium tuberculosis*. *Microbiology* 144 (Pt 9);2539-2544. 1998.

Berney M, Greening C, Conrad R, William R. Jacobs J, Cook GM. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proceedings of the National Academy of Sciences*. 2014;111:11479-11484.

Bioo Scientific Corp. NEXTflex™ Rapid Directional qRNA-Seq™ Kit Manual. 2014. V14.09

Boshoff HIM, Reed MB, Barry CE, Mizrahi V. DnaE2 Polymerase Contributes to In Vivo Survival and the Emergence of Drug Resistance in Mycobacterium tuberculosis. *Cell*. 2003;113:183-193.

Broden, N. J. Flury, S. King, A. N. Schroeder, B. W. Coe, G. D. Faulkner, M. J. Insights into the function of a second, non-classical Ahp peroxidase, AhpA, in oxidative stress resistance in

Bacillus subtilis. *Journal of Bacteriology.* American Society for Microbiology. 2016.

Callebaut I, Moshous D, Mornon J, De Villartay J. Metallo-β-lactamase fold within nucleic acids processing enzymes: The β-CASP family. *Nucleic Acids Research*. 2002;30:3592-3601.

Centers for disease control and prevention - http://www.cdc.gov/tb/statistics/default.htm

Centers for disease control and prevention. The role of BCG vaccine in the prevention and control of Tuberculosis in the United States. A joint statement by the advisory council for the elimination of tuberculosis and the advisory committee on immunization practices. April 26, 1996 / Vol. 45 / No. RR-4. Available at: http://www.cdc.gov/mmwr/pdf/rr/rr4504.pdf

Condon C. What is the role of RNase J in mRNA turnover? *RNA Biology*. 2010;7:316-321.

Crooks GE, Hon G, Chandonia J, Brenner SE. WebLogo: A sequence logo generator. *Genome Research*. 2004;14:1188-1190.

Dussurget O, Smith I. Interdependence of mycobacterial iron regulation, oxidative-stress response and isoniazid resistance. *Trends in Microbiology*. 1998;6:354-358.

Epicentre. Ribo-zero™rRNA Removal Kit. Gram-positive bacteria manual. Lit.# 314.4/2012.

Epicentre. ScriptSeq™ Complete kit (bacteria) manual. Lit.#345. 4/2013. EPILIT345 Rev.A. Available at: http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/scriptseq-complete/scriptseq-complete-kit-bacteria-library-prep-guide.pdf

Even S, Pellegrini O, Zig L, et al. Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic acids research*. 2005;33:2141-2152.

Figaro S, Durand S, Gilet L, Cayet N, Sachse M, Condon C. Bacillus subtilis Mutants with Knockouts of the Genes Encoding Ribonucleases RNase Y and RNase J1 Are Viable, with Major Defects in Cell Morphology, Sporulation, and Competence. *Journal of Bacteriology*.

2013;195:2340-2348.

Forbes M, Kuck NA, Peets EA. MODE OF ACTION OF ETHAMBUTOL. *Journal of Bacteriology*. 1962;84:1099-1103.

Fortune SM, Jaeger A, Sarracino DA, et al. Mutually Dependent Secretion of Proteins Required for Mycobacterial Virulence. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:10676-10681.

GENE-E. Matrix visualization and analysis platform. Broad Institute. Software available at: http://www.broadinstitute.org/cancer/software/GENE-E/index.html.

Grant CM. Regulation of Translation by Hydrogen Peroxide. *Antioxidants & redox signaling*. 2011;15:191-203.

Güell M, Yus E, Lluch-Senar M, Serrano L. Bacterial transcriptomics: what is beyond the RNA horiz-ome? *Nature Reviews Microbiology*. 2011;9:658-669.

Hassett DJ, Cohen MS. Bacterial adaptation to oxidative stress: implications for pathogenesis and interaction with phagocytic cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1989;3:2574.

Hooper DC. Mechanisms of Action and Resistance of Older and Newer Fluoroquinolones. *Clinical Infectious Diseases*. 2000;31:S24-S28.

Huang DW, Lempicki RA, Sherman BT. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 2008;2009;4:44-57.

Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*. 2009;2008;37:1-13.

Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*.

2007;8:R183-R183.

Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*. 2011;91:8-13.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*. 2015;10:845-858. Available at: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

Kohen, R. Nyska, A.Oxidation of biological systems: Oxidative stress phenomena, antioxidants, redox reactions, and methods for their quantification. *Toxicologic Pathology.* 2002;30:620-650.

Lee MH, Pascopella L, Jacobs WR, Hatfull GF. Site-Specific Integration of Mycobacteriophage L5: Integration-Proficient Vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and Bacille Calmette-Guerin. *Proceedings of the National Academy of Sciences of the United States of America*. 1991;88:3111-3115.

Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList – 10 years after. *Tuberculosis*. 2011;91:1-7.

Li X, Wu J, Han J, Hu Y, Mi K. Distinct Responses of Mycobacterium smegmatis to Exposure to Low and High Levels of Hydrogen Peroxide. *PLOS ONE*. 2015;10:e0134595.

Linder P, Lemeille S, Redder P. Transcriptome-wide analyses of 5'-ends in RNase J mutants of a gram-positive pathogen reveal a role in RNA maturation, regulation and degradation. *PLoS genetics*. 2014;10:e1004207.

Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 10.2, Bacterial Transcription Initiation. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21612/

McClure R, Balasubramanian D, Sun Y, et al. Computational analysis of bacterial RNA-Seq

data. *Nucleic acids research*. 2013;41:e140-e140.

NEB Tm calculator - http://tmcalculator.neb.com/#!/

Pediaditakis M, Kaufenstein M, Graumann PL. Bacillus subtilis hlpB Encodes a Conserved Stand-Alone HNH Nuclease-Like Protein That Is Essential for Viability Unless the hlpB Deletion Is Accompanied by the Deletion of Genes Encoding the AddAB DNA Repair Complex. *Journal of Bacteriology*. 2012;194:6184-6194.

Pieters J. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. *Cell Host & Microbe*. 2008;3:399-407

Qiaprep® Miniprep Handbook. For purification of molecular grade DNA. Qiagen®. Fourth edition. June 2015. Available at: https://www.qiagen.com/resources/download.aspx?id=22df6325-9579-4aa0-819c-788f73d81a09&lang=en.

Rochat T, Bouloc P, Repoila F. Gene expression control by selective RNA processing and stabilization in bacteria. *FEMS Microbiology Letters*. 2013;344:104-113.

Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000;16:944-945.

Scapoli C, Bartolomei E, De Lorenzi S, et al. Codon and Aminoacid Usage Patterns in Mycobacteria. *Journal of Molecular Microbiology and Biotechnology*. 2009;17:53.

SeqBuilder™ (©2004-2016) Version 12.0.0 (233). DNASTAR. Madison, WI.

Shell SS, Wang J, Lapierre P, Mir M, Chase MR, Pyle MM, Gawande R, Ahmad R, Sarracino DA, Ioerger TR, Fortune SM, Derbyshire KM, Wade JT, Gray TA. Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape: e1005641. *PLoS Genetics*. 2015;11.

Shemesh M, Tam A, Feldman M, Steinberg D. Differential expression profiles of

Streptococcus mutans ftf, gtf and vicR genes in the presence of dietary carbohydrates at early and late exponential growth phases. *Carbohydrate Research*. 2006;341:2090-2097.

Takayama K, Armstrong EL, Kunugi KA, Kilburn JO. Inhibition by Ethambutol of Mycolic Acid Transfer into the Cell Wall of Mycobacterium smegmatis. *Antimicrobial Agents and Chemotherapy*. 1979;16:240-242.

Taneja NK, Tyagi JS. Resazurin reduction assays for screening of anti-tubercular compounds against dormant and actively growing Mycobacterium tuberculosis, Mycobacterium bovis BCG and Mycobacterium smegmatis. *Journal of Antimicrobial Chemotherapy*. 2007;60:288-293.

Taverniti V, Forti F, Ghisotti D, Putzer H. Mycobacterium smegmatis RNase J is a 5′-3′ exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Molecular Microbiology*. 2011;82:1260-1276.

Timmins GS, Deretic V. Mechanisms of action of isoniazid. *Molecular Microbiology*. 2006;62:1220-1227.

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

Timothy L. Bailey, Mikael Bodén, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, William S. Noble, "MEME SUITE: tools for motif discovery and searching", *Nucleic Acids Research*, 37:W202-W208, 2009.

Tjaden B. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome biology*. 2015;16:1-1.

Touzain F, Schbath S, Debled-Rennesson I, Aigle B, Kucherov G, Leblond P. SIGffRid: a tool to search for sigma factor binding sites in bacterial genomes using comparative approach and

biologically driven statistics. *BMC bioinformatics*. 2008;9:73-73.

Van Kessel, J. C. Hatfull, G. F. Mycobacterial Recombineering. In: Davis G. and Kayser K. J., eds. *Methods in Molecular Biology, vol. 435: Chromosomal Mutagenesis.* © Humana Press Inc., Totowa, NJ

Voskuil MI, Bartek IL, Visconti K, Schoolnik GK. The response of mycobacterium tuberculosis to reactive oxygen and nitrogen species. *Frontiers in microbiology*. 2011;2:105.

World Health Organization. Global Tuberculosis Report 2015. Available at: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1

Zhang Y, Wade MM, Scorpio A, Zhang H, Sun Z. Mode of action of pyrazinamide: disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid. *The Journal of antimicrobial chemotherapy*. 2003;52:790-795.