

# The role of RNase E in *Mycolicibacterium smegmatis*

by  
Ying Zhou



A Dissertation  
Submitted to the Faculty of the  
WORCESTER POLYTECHNIC INSTITUTE  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Biology and Biotechnology

Jan 25<sup>th</sup>, 2022

APPROVED BY:

Dr. Scarlet Shell, Advisor  
Biology and Biotechnology, WPI

Dr. Reeta Rao, Committee Member  
Biology and Biotechnology, WPI

Dr. Elizabeth Ryder, Committee Member  
Biology and Biotechnology, WPI

Dr. Patrick Flaherty, Committee Member  
Mathematics and Statistics, Umass-Amherst

# Abstract

TB is caused by *Mycobacterium tuberculosis* (Mtb) and is one of the leading causes of death worldwide. During infection, Mtb can survive and adapt within the host, and this is dependent on the tight regulation of gene expression. mRNA degradation is a major process in this regulation, and it is well characterized in *E. coli* and *B. subtilis*, but not in mycobacteria. Therefore, it is fundamental to have a better understanding of mRNA degradation in mycobacteria. We used *Mycobacterium smegmatis* which is non-pathogenic and fast-growing for all our studies. In Chapter 2, we showed a global analysis of the transcriptome organization and post-transcriptional mRNA cleavage landscape in *M. smegmatis* in log phase and hypoxia conditions. In addition to defining transcription start sites (TSSs), we identified over 3,000 RNA cleavage sites. Importantly, a novel sequence motif was found among these cleavage sites which is different from what has been reported in *E. coli*. Then we aimed to determine which ribonuclease was the major contributor to these cleavage events, and we hypothesized that it was RNase E. In chapter 3, we determined mRNA half-lives with high confidence when RNase E was normally expressed or repressed. As expected, we showed a global mRNA stabilization when RNase E was repressed. The degree of stabilization varied among transcripts and several potential causes of this variation were assessed, such as mRNA abundance and the 5' UTRs of transcripts. Importantly, we also mapped the RNase E cleavage sites *in vivo* and *in vitro* and found that RNase E cleaved at the sequence RN↓CNU, consistent with the sequence motif identified in chapter 2. These findings demonstrate an important role for RNase E in mRNA degradation. RNase E interacts with proteins to form RNA degradosome in *E. coli* and the interactions commonly happen in the scaffold domain of RNase E. To better understand mycobacterial RNase E, in chapter 4 we mapped the boundaries of the RNase E catalytic domain and scaffold domains, which are different from *E. coli* RNase E. The first 330 residues of the N-terminal scaffold domain had roles in the sub-cellular

localization of RNase J, cell size, gene expression, and mRNA stabilization. Overall, we have compiled evidence showing that RNase E is important in mRNA degradation in *M. smegmatis*.

# Table of Contents

Abstract.....	i
Acknowledgements.....	vii
Chapter 1 : An overview of the roles of ribonucleases in mRNA degradation .....	1
An overview of the roles of ribonucleases in mRNA degradation.....	2
Introduction.....	2
Endonucleases.....	2
Exoribonucleases .....	11
RNases in mycobacteria .....	16
References.....	17
Chapter 2 : Defining the Transcriptional and Post-transcriptional Landscapes of <i>Mycobacterium smegmatis</i> in Aerobic Growth and Hypoxia.....	28
Defining the Transcriptional and Post-transcriptional Landscapes of <i>Mycobacterium smegmatis</i> in Aerobic Growth and Hypoxia .....	29
Introduction.....	30
Results.....	33
Discussion .....	53
Materials and Methods.....	57
References .....	61
Supplementary Figures .....	69
Supplementary Tables .....	76
Chapter 3 : The dominant role of RNase E in shaping the <i>Mycobacterium smegmatis</i> transcriptome .....	77
The dominant role of RNase E in shaping the <i>Mycobacterium smegmatis</i> transcriptome .	78
Introduction.....	79
Results.....	81
Discussion .....	90
Materials and Methods.....	92
References .....	98
Tables.....	101
Supplementary Figures .....	102
Supplementary Tables .....	105
Chapter 4 : Defining the roles of the degradosome-scaffolding domains of RNase E in <i>Mycobacterium smegmatis</i> .....	106
Defining the roles of the degradosome-scaffolding domains of RNase E in <i>Mycobacterium smegmatis</i> .....	107
Introduction.....	108
Results.....	110
Materials and Methods.....	126
References .....	129
Tables.....	133

Supplementary Figures .....	140
Supplementary Tables .....	144
Chapter 5 : Conclusions and Future directions .....	149
Defining the Transcriptional and Post-transcriptional Landscapes of <i>Mycobacterium smegmatis</i> in Aerobic Growth and Hypoxia .....	150
The dominant role of RNase E in shaping the <i>Mycobacterium smegmatis</i> transcriptome	152
Defining the roles of the degradosome-scaffolding domains of RNase E in <i>Mycobacterium smegmatis</i> .....	154
References .....	157

## List of Figures

Figure 1-1: A combination of cleavage events by major ribonucleases and RNA helicases contribute to RNA degradation. ....	3
Figure 1-2: The catalytic domain of RNase E is conserved in <i>E. coli</i> , <i>M. tuberculosis</i> , and <i>M. smegmatis</i> .....	6
Figure 2-1: Mapping and categorization of transcription start sites in <i>M. smegmatis</i> .....	34
Figure 2-2: <i>M. smegmatis</i> promoter -10 regions are dominated by the ANNNT motif.....	40
Figure 2-3: Leader features are conserved in mycobacteria .....	43
Figure 2-4: Cleavage site positions are biased with respect to sequence context and genetic location.....	46
Figure 2-5: The transcriptional landscape substantially changes upon oxygen limitation .....	51
Figure 3-1: Knockdown of <i>rne</i> expression causes growth cessation and reduced mRNA degradation rates in <i>M. smegmatis</i> .....	83
Figure 3-2: Knockdown of <i>rne</i> causes substantial changes in <i>M. smegmatis</i> transcript abundance levels.....	84
Figure 3-3: Knockdown of <i>rne</i> causes transcriptome-wide mRNA stabilization, which explains some of the gene expression changes and appears to be partially compensated for by decreased transcription rates. ....	85
Figure 3-4: Leadered transcripts and highly abundant transcripts tend to be more strongly stabilized upon <i>rne</i> knockdown .....	86
Figure 3-5: RNase E is responsible for previously reported <i>M. smegmatis</i> mRNA cleavage sites and cleavage 5' of cytidines .....	89
Figure 4-1: Defining the essential domain of <i>M. smegmatis</i> RNase E .....	114
Figure 4-2: RNase E and RNase J appear to localize together and RNase J localization is more diffuse when both the N- and C-terminal scaffold domains of RNase E are deleted.....	116
Figure 4-3: RNase J localization is partially dependent on RNase E .....	117
Figure 4-4: Differential expression analysis of <i>rne</i> <sup>Δ2-330</sup> and <i>rne</i> <sup>Δ2-330, Δ825-1037</sup> compared to full-length <i>rne</i> .....	121
Figure 4-5: Deletion of the N-terminal 330 residues of RNase E leads to slower mRNA degradation .....	123

## List of Tables

Table 3-1: Strains and plasmids used in this study.....	101
Table 4-1: Proteins identified in greater abundance from RNase E immunoprecipitation followed by LC-MS/MS compared to a mock immunoprecipitation.....	133
Table 4-2: Expression level of ESX-1 genes in <i>rne</i> <sup>Δ2-330, Δ825-1037</sup> and <i>rne</i> <sup>Δ2-330</sup> strains relative to the full-length <i>rne</i> strain. ....	135
Table 4-3: Expression levels of mycofactocin genes and their regulator in <i>rne</i> <sup>Δ2-330, Δ825-1037</sup> and <i>rne</i> <sup>Δ2-330</sup> relative to full-length <i>rne</i> .....	136
Table 4-4: Expression levels of genes differentially expressed in <i>rne</i> <sup>Δ2-330</sup> but not in <i>rne</i> <sup>Δ2-330, Δ825-1037</sup> .....	137

Table 4-5: Relative expression levels of glycerol metabolism enzymes in <i>rne</i> <sup>Δ2-330, Δ825-1037</sup> and <i>rne</i> <sup>Δ2-330</sup> relative to full-length <i>rne</i> .....	140
---	-----

## List of Supplemental Figures

Figure S2-1: Hypoxia model similar to the Wayne model.....	69
Figure S2-2: Construction of 5'-end-directed libraries .....	70
Figure S2-3: Workflow for noise filtering and TSS prediction in the different datasets .....	71
Figure S2-4: TSSs identified in the different datasets.....	71
Figure S2-5: Workflow used for TSS classification .....	72
Figure S2-6: Distribution of antisense TSSs .....	72
Figure S2-7: Cleavage sites distribution within genes according to coding sequence context...	73
Figure S2-8: Validation of a medium confidence pTSS .....	74
Figure S2-9: Gene expression levels in RNAseq expression libraries in hypoxia.....	75
Figure S2-10: Correlation between expression data and 5' end-directed libraries data in hypoxia .....	75
Figure S2-11: Changes in RNA cleavage within coding sequences in hypoxic conditions .....	76
Figure S3-1: Filtering genes for half-life calculations.....	102
Figure S3-2: Half-lives of genes in control conditions are similar .....	103
Figure S3-3: Accessible 5' ends are not correlated with the differential stabilization when <i>rne</i> is repressed .....	104
Figure S3-4: <i>In vitro</i> -transcribed partial duplex RNA substrate used for RNase E cleavage assays .....	105
Figure S4-1: Construction of a chromosomally His-FLAG-tagged RNase E strain.....	140
Figure S4-2: Rifampicin killing assay to test the protein functionality of RNase J-Dendra .....	141
Figure S4-3: Derivations of mCherry-tagged RNase E.....	141
Figure S4-4: Expression of FLAG-RNase E and RNase J-Dendra in strains harboring both tagged RNases .....	142
Figure S4-5: Growth kinetics to test the protein function of tagged PNPase .....	143

## List of Supplemental Tables

Table S4-1: Percentage of positive transformants during the construction of FLAG-tagged RNase E mutants.....	144
Table S4-2: Percentage of positive transformants during the construction of RNase E-mCherry mutants .....	144
Table S4-3: Genes differentially expression in <i>rne</i> <sup>Δ825-1037</sup> relative to full-length <i>rne</i> .....	145
Table S4-4: Strains and plasmids.....	146
Table S4-5: Primers for qPCR.....	148

# Acknowledgements

I want to give full of my thanks to Dr. Scarlet Shell, my advisor. I cannot describe my gratitude and happiness in words. I have been so lucky to come to WPI and be the first PhD student in the Shell Lab. I will never regret to choose this program because I have a chance to work with an excellent supervisor for seven years. This is a wonderful journey, and it will become a best part in my life. I learned a lot from Scarlet. Your passion about research and hard-working impressed me a lot. I always told my family that my supervisor is a good instructor as she will never say no if I need help. You gave me unconditional support in my project and always came up with new ideas when I got stuck. You encouraged me to present posters and give talks in meetings and conferences. All of these made me confident about myself to be a good researcher. Thank you, Scarlet!!! You are always my best supervisor.

I also want to thank all my dissertation committee members for their critical roles in giving me suggestions and ideas to continue my research. Dr. Reeta Rao, Dr. Elizabeth Ryder, Dr. Patrick Flaherty, it is a good pleasure to have you to let me think about the bigger picture of my research and get instructions when I need help. And thanks for all the feedback on my dissertation.

I want to thank all the past and current lab members of the Shell lab. We shared happiness and offered help to each other. We also hold parties to celebrate the birthday and publications. Dr. Carla Martini, I want to thank you for your help and comments on my experiments. Dr. Diego Vargas Blanco, thank you for your kindness when I was in trouble. Huaming Sun, thank you for helping me in data analysis. Dr. Louis Roberts, thank you for your help in giving instructions to undergraduates who are involved in my RNase E protein purification project. I also want to give my thanks to Julia and Junpei, the time we spent together was good. My last thanks are for all

the undergraduates I met in the Shell lab, they were really doing an excellent job and offered a big help in my project. I had a joyful time in the Shell lab, thank you to all of you!

I want to give my heartfelt thanks to my parents. You are the strongest support behind me and made me brave enough to pursue my PhD degree. I will never forget the words that you told me when I was wondering about myself. Thank you for giving me an opportunity to come to US to get a better study and let me step further than others. This is a long trip, without you, I can never imagine that I can make it. I also want to thank my grandparents for their love and support.

Without a doubt, I want to thank my husband, Yuanzhou Yang. You are my best friend to share happiness and anxiety. Thanks for coming back to US, I really appreciated it. You have been a necessary part in my life and thanks for your love and support. I will also be a strong support for you and best wishes to your research.

I would like to thank Bo Yang, my best friend in the department. I am so glad to meet you in the department. I value your friendship and best wishes to your graduation. I also want to thank all my friends from "32 John St", the two years we spent together gave my PhD life a good start.

Lastly, I want to thank all the games that I played, and the TV shows I watched. PhD study is a struggle sometimes, the relax time makes me full of energy and ready to go again.

# Chapter 1 : An overview of the roles of ribonucleases in mRNA degradation

# An overview of the roles of ribonucleases in mRNA degradation

## Introduction

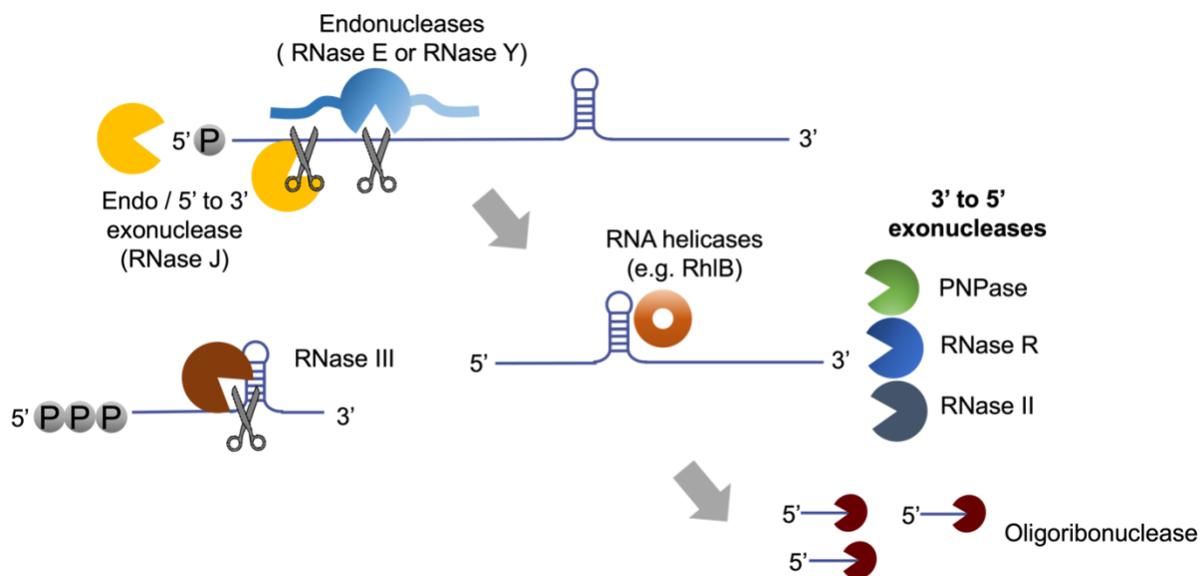
Gene expression regulation is a principal way that bacteria phenotypically adapt to diverse environments. Transcription and mRNA degradation are two major processes determining how much mRNA is available for protein synthesis. mRNA degradation is usually carried out by a combination of cleavage events by different ribonucleases (RNases), including endonucleases and exonucleases, with the assistance of RNA helicases (Figure 1-1). mRNA degradation makes an important contribution to ribonucleotide recycling by degrading useless transcripts. Rapid mRNA turnover also helps cells to quickly adapt to variable environments by permitting rapid changes in gene expression. The studies reported in this dissertation reflect our efforts to understand the mechanisms and consequences of mRNA processing and degradation in mycobacteria, a group of bacteria that includes important human pathogens such as *Mycobacterium tuberculosis* as well as the non-pathogenic model *Mycobacterium smegmatis*. To place this work into context, we will first review some of what is known about mRNA degradation in bacteria.

## Endonucleases

Principle endonuclease such as RNase E, RNase G, and RNase Y specifically work on single-stranded RNAs, and RNase III shows specificity for double-stranded RNAs. Different combinations of these contribute to mRNA degradation in different bacterial species.

## RNase E

RNase E is one of the most important enzymes in mRNA degradation in many bacterial species, first identified in *E. coli* and initially found to be responsible for rRNA processing (Apirion, 1978). Then, it was showed that there was an increased half-life of bulk mRNA when an RNase E mutant strain was shifted to a non-permissive temperature, and it preferred to cleave “AU” rich regions with a loose consensus cleavage site sequence “G/AN↓A/UUU” (Babitzke & Kushner, 1991; Chao et al., 2017; McDowall et al., 1994; Mudd et al., 1990; Mudd et al., 1988). RNase E also has a strong preference for single-stranded regions of RNA substrates (Mackie, 1992; Mackie & Genereaux, 1993; McDowall et al., 1995). Further studies also showed that in the absence of RNase E, around 60% of coding sequence showed a differential expression in *E. coli* (Stead et al., 2011), suggesting its global role in mRNA metabolism in *E. coli*.



**Figure 1-1: A combination of cleavage events by major ribonucleases and RNA helicases contribute to RNA degradation.** RNase III can specifically cleave double-stranded mRNA and RNase E in *E. coli* and mycobacteria or RNase Y in *B. subtilis* prefer to cleave single-stranded 5' monophosphorylated mRNA. RNase J has dual endonuclease and 5' to 3' exonuclease activities and exists in *B. subtilis* and mycobacteria, but not in *E. coli*. RNA helicases can unwind secondary structure to facilitate nonspecific 3' to 5' exoribonuclease activities by PNPase, RNase II, and RNase R. RNase R itself can degrade mRNAs with secondary structures. The remaining short products are degraded by oligoribonuclease to mononucleotides for recycling.

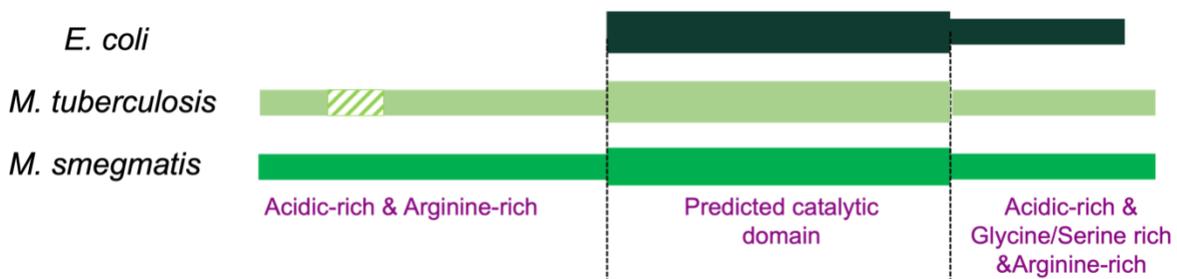
*E. coli* RNase E is encoded by *rne* (also known as *ams* or *hmp*) and the gene product is a 118 kDa protein consisting of 1061 amino acid residues (Babitzke & Kushner, 1991; Casaregola et al., 1994; Chauhan & Apirion, 1991; Cormack et al., 1993; Melefors & von Gabain, 1991; Mudd et al., 1990; Taraseviciene et al., 1991). The N-terminal domain of *E. coli* RNase E was found sufficient to perform a site-specific cleavage with a single-stranded oligoribonucleotide - BR13 - and was suggested to contain catalytic sites for this endonuclease activity (McDowall & Cohen, 1996; McDowall et al., 1995; Taraseviciene et al., 1995). The N-terminal catalytic domain (1-529 residues) of *E. coli* RNase E protein forms a homo-tetramer with each protomer composed of two globular domains, one 'large' and one 'small' (Callaghan et al., 2005). The 'large' domain contains small subdomains including RNase H endoribonuclease family, S1 domain, 5' sensor region, and DNase I. However, only the S1 domain and 5' sensor region appeared to be important in triggering a conformational change in the active site for RNA cleavage (Callaghan et al., 2005; Koslover et al., 2008).

Then, the C-terminal domain of *E. coli* RNase E, a region enriched with arginine, was shown to be an RNA-binding subdomain and not necessary for endonuclease activity (McDowall & Cohen, 1996; Taraseviciene et al., 1995). Moreover, a proline rich sequence in the C-terminal half was showed to be responsible for an observed slower migration of purified RNase E protein in SDS/polyacrylamide gels (Casaregola et al., 1994; McDowall & Cohen, 1996). Overall, The C-terminal domain of RNase E is characterized by regions containing highly charged and acidic amino acid compositions in *E. coli* (Casaregola et al., 1994). Compared to the N-terminal catalytic domain, the C-terminal half of *E. coli* RNase E is less structured and self-interacted due to the self-interaction domain (Callaghan et al., 2004). Importantly, the C-terminal domain contains distinct binding sites for exoribonuclease polynucleotide phosphorylase (PNPase), DEAD box RNA helicase (RhIB), and the glycolytic enzyme enolase (Vanzo et al., 1998). As there were no direct interactions found between PNPase, RhIB, and enolase, the C-terminal domain of RNase

E appeared to be the structural core of the assembled RNA degradosome (Vanzo et al., 1998). RNase E binds to other proteins in different species, and the degradosome components vary among species (Ait-Bara & Carpousis, 2010; Hardwick et al., 2011; Jager et al., 2001; Lee & Cohen, 2003; Miczak et al., 1996; Plocinski et al., 2019; Purusharth et al., 2005; Rosana et al., 2016; Stoppel et al., 2012; Van den Bossche et al., 2016; Zhang et al., 2014). In this cooperative fashion, RNase E is a key to initiate mRNA degradation.

*E. coli* RNase E shows a strong preference in cleaving 5'-monophosphorylated RNAs while the cleavage events are partially inhibited on 5'-triphosphorylated substrates (Mackie, 1998). The 5'-end dependent cleavage of RNase E was suggested to be a property of the N-terminal catalytic domain while the C-terminal half was not required (Jiang et al., 2000). Later, this 5'-end preference was explained by the structure of RNase E's catalytic domain containing a 5'-end binding pocket that can only accommodate 5'-monophosphorylated RNAs (Callaghan et al., 2005). On the other hand, 'direct entry' cleavage was proposed to be a major pathway for degrading RNA as RNase E was shown by other groups to be able to cleave RNAs rapidly without a 5' monophosphorylated end (Clarke et al., 2014; Kime et al., 2010). With this internal entry mechanism, RNA pyrophosphohydrolase (RppH)-like enzymes, which can convert 5' triphosphates to 5' monophosphates, are less required for RNase E-mediated RNA processing and degradation. Another example that supports RNase E having a major 'direct entry' pathway is the maturation of transfer RNA (tRNA) in *E. coli* (Kime et al., 2014). Mostly recently, a work showed that RNase E linearly scanned from the 5'-monophosphorylated end of single-stranded RNA for cleavage sites and this attack was hindered by structural obstacles upstream (Richards & Belasco, 2019). While both RNA substrate recognition pathways exist, a cooperation between them may be equally important in RNase E activity.

In *E. coli*, RNase E was found to be a membrane-binding protein, localizing to the inner cytoplasmic membrane of cells (Khemici et al., 2008). This localization feature was attributed to 'segment A' which is a 15-residues Membrane Targeting Sequence (MTS) at the beginning of C-terminal domain of RNase E. Deletion of the MTS induced a growth defect, suggesting the functional importance of RNase E's association with the membrane (Khemici et al., 2008). Further study also showed that RhlB was membrane-associated along with RNase E and this membrane binding was through the MTS of RNase E (Strahl et al., 2015). Also in this study, the formation of short-lived foci of RNase E was observed, which was proposed to be explained by a transient clustering of RNase E that would increase cooperation with other degradation bodies in mRNA degradation (Strahl et al., 2015). Another interesting study showed that mRNAs localized at the inner membrane were degraded faster than cytoplasmic mRNAs (Moffitt et al., 2016). The observed faster degradation of membrane-localized mRNAs was dependent upon the RNA degradosome being located at the membrane (Moffitt et al., 2016).



**Figure 1-2: The catalytic domain of RNase E is conserved in *E. coli*, *M. tuberculosis*, and *M. smegmatis*.** The catalytic domains in *M. tuberculosis* and *M. smegmatis* are located in the central region of RNase E, different from *E. coli*. Regions containing highly conserved sequences are named as shown, as well as other regions containing acidic-rich, arginine-rich or glycine/serine rich sequences. The region shown in stripes in *M. tuberculosis* RNase E is not present in *M. smegmatis* RNase E.

RNase E is also essential in mycobacteria (Sasseti et al., 2003; Sasseti & Rubin, 2003; Taverniti et al., 2011) and it plays a key role in rRNA processing in *M. tuberculosis* (Zeller et al., 2007). In

addition to rRNA processing, mRNA cleavage by RNase E was found in *Mycobacterium smegmatis* (Taverniti et al., 2011). There is a conserved catalytic domain in mycobacterial RNase E with substantial homology to that of *E. coli* RNase E. However, the catalytic domain in mycobacterial RNase E is centrally located between two predicted scaffold domains, in contrast to *E. coli* where there is only one, C-terminal scaffold domain (Figure 1-2) (Zeller et al., 2007). Key degradosome components were also defined in *M. tuberculosis*, and they include an RNA helicase, PNPase, and RNase J (Plocinski et al., 2019).

### **RNase G**

In *E. coli*, while RNase E is essential for cell viability, its ortholog RNase G is not (Apirion & Lassar, 1978; Li et al., 1999; Ono & Kuwano, 1979; Wachi et al., 1999). A genetic analysis showed that there was a functional relationship between RNase E and RNase G (initially called CafA) (Wachi et al., 1997; Wachi et al., 1999). RNase G had a high sequence similarity (31%) to the N-terminal catalytic domain of RNase E and both RNase E and RNase G appeared to be required for 5' maturation of 16S ribosomal RNA (Li et al., 1999; McDowall et al., 1993). Overexpression of wild-type RNase G did not fully compensate for the lethality of RNase E deletion; however, with the unnatural extension of several amino acid residues at the N-terminus or C-terminus of RNase G, the complementation by this modified RNase G was sufficient to make cells viable (Deana & Belasco, 2004; Lee et al., 2002). Compared to RNase E, the absence of RNase G had minor effects on *E. coli* transcriptome-wide mRNA abundance based on a microarray analysis. This may be explained by the lower cellular RNase G concentration but also by the different intrinsic properties of these two ribonucleases (Deana & Belasco, 2004; Lee et al., 2002).

## **RNase Y**

Distinct from *E. coli*, RNase E is absent in *B. subtilis*. Instead, RNase Y (previously called YmdA) was identified as an essential endoribonuclease. RNase Y was able to cleave the *yitJ*, a riboswitch with a SAM-binding aptamer domain upstream (Hunt et al., 2006; Shahbadian et al., 2009). Further, it was shown that the half-life of mRNA was increased in an RNase Y depleted strain and this impact was on a global scale, indicating RNase Y's key role in mRNA degradation in *B. subtilis* (Durand, Gilet, Bessieres, et al., 2012; Laalami et al., 2013; Lehnik-Habrink, Schaffer, et al., 2011; Shahbadian et al., 2009).

The protein structure of RNase Y is distinct from RNase E. RNase Y consists of a transmembrane domain, a coiled-coil disordered domain, two central KH and HD domains containing important catalytic residues, and a highly conserved C-terminal domain (Lehnik-Habrink, Newman, et al., 2011). Although the structures of RNase E and RNase Y are distinct, they appear to share similar functions as mediators of RNA degradation complexes. RNase Y was proposed to scaffold the formation of a degradosome-like protein complex in *B. subtilis* with RNase J1, RNase J2, PNPase, and two glycolytic enzymes, 6-phosphofructokinase and enolase, by bacterial two-hybrid analyses (Commichau et al., 2009). Another study showed that a DEAD box RNA helicase, CshA, interacted with RNase Y and PNPase, mediated by the C-terminal domain of CshA (Lehnik-Habrink, Newman, et al., 2011; Lehnik-Habrink et al., 2010). However, since the interaction between RNase Y and RNase J1 was weak and the complex was never purified *in vivo* (Commichau et al., 2009), some researchers suspect that the interactions in this complex may be transient in nature.

There are additional functional similarities between RNase E and RNase Y. They are both essential endoribonuclease and have a preference for 5'-monophosphorylated RNA substrates,

and they both have major effects on mRNA degradation (Shahbadian et al., 2009). Another important similarity is the localization of these two proteins, as they are both membrane-binding proteins and their localization is important for cell growth (Hunt et al., 2006; Khemici et al., 2008). Moreover, a recent work showed that full length *E. coli* RNase E could replace RNase Y in *B. subtilis* by restoring the growth of an RNase Y deletion strain, and that RNase E localized to the inner face of the cell membrane when ectopically expressed in *B. subtilis* (Laalami et al., 2021). Taken together, there might be an evolutionarily conserved mechanism governing the organization of bacterial mRNA degradation.

### **RNase III**

RNase III was found and purified from *E. coli* in the 1960s and shown to be a double-stranded specific endoribonuclease (Robertson et al., 1968). RNase III plays an important role in rRNA processing and the maturation of 16S and 23S rRNAs as well as a limited extent of mRNA processing (Afonyushkin et al., 2005; Babitzke et al., 1993; Gordon et al., 2017; Talkad et al., 1978; Young & Steitz, 1978). It was found that only 12% of the coding sequences were affected in an RNase III null mutant strain in *E. coli* (Stead et al., 2011). RNase III is highly conserved across bacteria and eukaryotes, mainly in its catalytic domain (Błaszczuk et al., 2004; Filippov et al., 2000). It was also found in mycobacteria (Taverniti et al., 2011). In *E. coli*, RNase III is encoded by the *rnc* gene and is the first gene located in *rnc-era-recO* operon (*rnc* operon). Interestingly, RNase III can cleave the stem-loop of its own 5' UTR to induce a rapid degradation of the *rnc* transcript to achieve autoregulation (Bardwell et al., 1989). A further study showed that this autoregulation was mediated by a portable RNA stability element (*rncO*) located in the *rnc* transcript (Matsunaga et al., 1996).

The RNase III family can be divided into different classes. The class containing *E. coli* RNase III is simpler, containing an endonuclease domain and a double-stranded RNA binding domain (Kharrat et al., 1995). RNase III functions as a homodimer with  $Mg^{2+}$  dependent active sites in each monomer and can act as a processing enzyme or a binding protein without processing (Dasgupta et al., 1998; Kharrat et al., 1995; Nashimoto & Uchida, 1985). Each strand of the dsRNA can be cleaved, yielding products with 5' monophosphates and 3' hydroxyls with a 2-bp overhang at the 3' end (Nicholson, 1999). The dsRNA substrates of RNase III are commonly around 20 base pairs, and a reduced cleavage rate was observed for shorter sequence (Pertzev & Nicholson, 2006; Robertson, 1982).

*E. coli*, *B. subtilis*, and mycobacteria all have RNase III, but RNase III is only essential in *B. subtilis* because the dsRNA cleavage of RNase III is involved in host defense mechanisms as it is required for the protection from toxin genes (Durand, Gilet, & Condon, 2012). Compared to RNase E and RNase Y, RNase III has a limited role in mRNA degradation. Studies in *E. coli* and *B. subtilis* both showed that there were a small number of transcripts affected by the absence of RNase III (Durand, Gilet, Bessieres, et al., 2012; Stead et al., 2011).

## **RNase J**

RNase J is a unique RNase as it harbors dual endonuclease and 5'-3' exoribonuclease activities and exists in *B. subtilis* and mycobacteria. For this RNase, more details will be described in the "Exoribonuclease" section.

## **Other Endonucleases**

RNase P is an unusual endonuclease as it contains an essential RNA component and a protein subunit, first characterized in tRNA maturation by removal of nucleotides from the 5' ends of tRNA

precursors (Kole et al., 1980; Robertson et al., 1972; Stark et al., 1978). In *E. coli*, the RNA component was termed M1 RNA, and the protein component was termed C5. These two portions could reconstitute a holoenzyme *in vitro* (Guerrier-Takada et al., 1983). It was found that M1 RNA itself could process tRNA precursors under certain conditions but C5 protein greatly increased RNase P cleavage activity *in vitro*. *In vivo*, the protein portion C5 was required for normal activity. RNase P has a limited role in mRNA metabolism. Examples showed that RNase P was able to cleave in the intergenic regions of polycistronic mRNAs and affected the abundance of downstream transcripts, in cooperation with RNase E (Alifano et al., 1994; Li & Altman, 2003). RNase P was also found to have a homolog in mycobacteria (Taverniti et al., 2011).

YbeY is a recently identified endonuclease in many bacteria including *E. coli* (Jacob et al., 2013), and it was shown to be involved in processing of all three rRNAs (Davies et al., 2010). Further, the YbeY ortholog in *Sinorhizobium meliloti* was linked to mRNA metabolism as it could affect the regulation of sRNA-mRNA pathways (Pandey et al., 2011). Another study showed that YbeY is a single-strand specific endonuclease and participated in rRNA and mRNA degradation in *E. coli* (Jacob et al., 2013). A genome-wide mRNA and sRNA expression profiling study in *E. coli* revealed that the deletion of YbeY had a global effect on sRNA-mediated gene expression, both Hfq-dependent and independent, which was consistent with the previous observation in Jacob's research (Pandey et al., 2014).

## **Exoribonucleases**

Exoribonucleases in bacteria commonly work after the degradation of mRNA has been initiated by endonucleolytic cleavage and perform subsequent degradation of initial cleavage products. Most of them act on single-stranded RNA in a 3' to 5' direction, including PNPase, RNase II, and RNase R, but there is an exception that is a 5' to 3' exoribonuclease - RNase J.

## 3' to 5' Exoribonucleases

### PNPase

PNPase is a phosphorolytic exoribonuclease that participates in mRNA degradation in *E. coli* (Donovan & Kushner, 1986; Har-El et al., 1979; Kinscherf & Apirion, 1975). PNPase can act as an exoribonuclease as well as a polymerase. PNPase prefers to degrade RNAs with single-stranded 3' ends and its exoribonuclease activities can be impeded by double-stranded stem-loop structures. In addition, at least 7-10 base pairs of 3' overhang were found to be required for its activity (Blum et al., 1999; Py et al., 1996). As a polymerase, it can add single-stranded poly(A) tails to the 3' ends of transcripts and this polyadenylation strongly facilitates its exoribonucleolytic activity on 3' ends that would otherwise be inaccessible due to secondary structure (Blum et al., 1999; Coburn & Mackie, 1996a; Mackie, 1989; Mohanty & Kushner, 2000; Xu & Cohen, 1995).

Each monomer of PNPase has two PH domains linked to an  $\alpha$ -helical domain, a KH domain, and a conserved S1 domain (Jarrige et al., 2002; Regnier et al., 1987; Symmons et al., 2000). The S1 and KH domains are not required for PNPase activity, but they are critical for the substrate binding and product release (Stickney et al., 2005). PNPase forms a trimer, and the six PH domains assemble into a ring structure containing a central channel (Shi et al., 2008). The ring-like structure can trap single-stranded RNAs in the central channel and then the RNAs bind the active sites (Buttner et al., 2005; Lorentzen & Conti, 2005).

PNPase cooperates with other RNases in mRNA degradation. The deletion of PNPase in *E. coli* did not affect cell viability, however, cells with PNPase and RNase II or PNPase and RNase R double deletions were not viable, indicating that there is some functional redundancy among the exoribonucleases in *E. coli* (Cheng & Deutscher, 2005; Cheng et al., 1998; Donovan & Kushner, 1986; Spickler & Mackie, 2000). Moreover, PNPase was found to directly interact with the RNA

helicase RhlB in *E. coli*, suggesting that PNPase and RhlB may work together in the removal of structured RNAs (Py et al., 1996). PNPase has been found as a component of the RNA degradosome, in many species, usually interacting with RNase E (Ait-Bara & Carpousis, 2010; Carpousis et al., 1994; Hardwick et al., 2011; Van den Bossche et al., 2016; Zhang et al., 2014), including *M. tuberculosis* (Plocinski et al., 2019).

In *E. coli*, PNPase auto-regulates its own expression at a post-transcriptional level. RNase III was found to cleave the double-stranded stem-loop in the 5' UTR upstream of the PNPase coding sequence, resulting in upstream and downstream fragments that remain base-paired together. PNPase then degrades the 3' end of the upstream product, removing the region that base-pairs with the downstream region, and exposing the downstream region to RNase E for degradation (Carzaniga et al., 2009; Jarrige et al., 2001; Portier et al., 1987; Robert-Le Meur & Portier, 1992).

## **RNase II**

RNase II is an important hydrolytic 3' to 5' exoribonuclease that can non-specifically degrade RNAs to 5'- mononucleotides in *E. coli* (Deutscher & Reuven, 1991). Compared to PNPase, the activity of RNase II was strongly inhibited by secondary structures at the 3' end, but it efficiently degrades unstructured RNAs like poly(A) tails (Guarneros & Portier, 1990; Marujo et al., 2000; Mott et al., 1985; Wu et al., 1981). Moreover, RNase II was shown to protect some structured 3' ends of mRNAs from PNPase activity (Coburn & Mackie, 1996b; Hajnsdorf et al., 1994).

Unlike PNPase, RNase II is a monomeric enzyme and contains two cold-shock domains, an RNB catalytic domain with active sites, and an S1 domain (Frazao et al., 2006). It was also shown that the narrow channel in the catalytic center only allowed single-stranded RNA to access the catalytic pocket, explaining the specificity of RNase II for ssRNA. Moreover, the cold-shock

domains and S1 domain act as RNA binding domains to provide an 'anchor' for the RNA substrates.

In *E. coli*, RNase II was found to associate with the cytoplasmic membrane by its N-terminal amphipathic helix, and this membrane association was suggested to maintain its normal function in cells (Lu & Taghbalout, 2013).

## **RNase R**

RNase R was identified in an *E. coli* RNase II deficient strain (Cheng et al., 1998; Gupta et al., 1977). RNase R and RNase II are both in the RNR superfamily of exoribonucleases and have similarities in catalytic activities, such as non-specific 3' to 5' degradation and release of 5'-mononucleotides (Cheng & Deutscher, 2002). However, differences exist. One major difference is the substrate specificity; RNase R can degrade rRNA while RNase II cannot (Zuo & Deutscher, 2001). Importantly, RNase R can efficiently degrade mRNA with extensive secondary structures, such as repetitive extragenic palindromic (REP) sequence, which is a highly conserved, 35–40 bp element in *E. coli* (Cheng & Deutscher, 2005; Gilson et al., 1987; Higgins et al., 1988). This difference can be mainly explained by the nuclease domain in RNase R that is sufficient to bind and degrade RNAs including structured RNAs (Matos et al., 2011; Vincent & Deutscher, 2009). RNase R was found to interact with RNase E in *Pseudomonas syringae* (Purusharth et al., 2005).

## **5' Exoribonuclease & Endonuclease**

In previous studies, 3' exoribonucleases were found and well characterized; however, it was a surprise when the first bacterial 5' to 3' exoribonuclease, RNase J1, was found in *B. subtilis* in 2007 (Mathy et al., 2007). Previously, 5' to 3' exoribonucleases had only been reported in eukaryotes. RNases J1 and J2 had been identified two years prior in *B. subtilis* as endonucleases

that appeared to have functional similarity to RNase E in *E. coli* (Even et al., 2005). Taken together, RNase J is a unique RNase that harbors endonuclease activity as well as 5' to 3' exonuclease activity. Like RNase E, RNase J also shows a preference for cleavage of 5' monophosphorylated substrates (Even et al., 2005; Richards et al., 2011). In a *B. subtilis* double deletion strain where RNase J1 was under the control of a IPTG-inducible Pspac promoter, a small increase of global mRNA half-life was observed comparing to WT, less than 2-fold change (Even et al., 2005). However, specific mRNAs were efficiently degraded by RNase J1/J2 (Mader et al., 2008). Studies also showed that RNase Y and RNase J1/J2 could form a complex, indicating a major role for RNase J in a cooperative mRNA degradation in *B. subtilis* (Durand, Gilet, Bessieres, et al., 2012).

Compared to RNase J1, the exoribonuclease activity of RNase J2 was found to be significantly weaker, around 100-fold less. However, the endonuclease activities were similar for RNase J1 and J2 (Mathy et al., 2010). The important role of RNase J1 in the exoribonuclease activities might be correlated with the essentiality of RNase J1 in *B. subtilis* (Britton et al., 2007). RNase J1 and J2 formed a complex in *B. subtilis*, and this complex appeared to be a primary form in the cells (Mathy et al., 2010). Further studies showed that RNase J1 and J2 associated to form homodimers respectively and then a heterotetramer (Newman et al., 2011). Structural analysis of the *Thermus thermophilus* RNase J protein revealed the interesting finding that the endonuclease and exonuclease activities are carried out at the same active site, indicating that the degradation mode might switch quickly between endo- and exoribonucleolytic on the same RNA (Commichau et al., 2009; Li de la Sierra-Gallay et al., 2008).

Distinct from *E. coli* and *B. subtilis*, mycobacteria have both RNase E and RNase J. RNase J is not essential in mycobacteria, but it also showed a dual activity and a role in rRNA processing

(Taverniti et al., 2011). Further, *M. tuberculosis* work found RNase J interacts with RNase E in mRNA degradosomes (Plocinski et al., 2019).

## **Oligoribonucleases**

Different from exoribonucleases, oligoribonucleases appeared to efficiently degrade very short RNA substrates of no more than five nucleotides (Datta & Niyogi, 1975). As exoribonucleases have difficulty in degrading mRNA substrates completely, oligoribonucleases are essential for the degradation of the end products into mononucleotides (Ghosh & Deutscher, 1999). The oligoribonuclease Orn is essential in *E. coli* and its homolog is not present in all bacterial species. However, there are oligoribonucleases with similar functions in other species, including mycobacteria (Fang et al., 2009; Ghosh & Deutscher, 1999; Mechold et al., 2007).

## **RNases in mycobacteria**

It is obvious that ribonucleases play important roles in mRNA degradation as well as processing of stable RNAs. The interactions between various RNases precisely govern mRNA decay rates, sometimes resulting in selective degradation of different sets of mRNA substrates. Previous studies have focused heavily on well-established organisms, but as the combination of RNases in bacteria varies, it would be helpful to further investigate the roles and interactions of RNases in different organisms to broaden our understanding of mRNA metabolism and regulation. Moreover, the emergence of new techniques in recent years has facilitated transcriptome-wide study of mRNA degradation in different growth conditions.

Mycobacteria include important human pathogens that continue to cause disease and death around the world. Knowledge of the basic biology of these bacteria is needed to understand why antibiotic treatment often fails, and to lay a knowledge foundation to facilitate development of

better antibiotics. We therefore sought to study mRNA degradation in mycobacteria and elucidate the roles of RNases in post-transcriptional regulation as well as their cooperation in mRNA metabolism.

Here we investigated the roles of mycobacterial RNases in mRNA degradation in several ways. In chapter 2, we characterized the transcriptional and post-transcriptional landscape in *M. smegmatis* and in the mapped cleavage sites, we found a novel consensus motif. Considering the important role of RNase E in other species, we hypothesized that RNase E was a major contributor to mRNA degradation. Then in chapter 3, we used an inducible knockdown system to repress the expression of RNase E and assess the impact on mRNA metabolism in *M. smegmatis*. We found a global mRNA stabilization when we repressed RNase E. Moreover, we confirmed that RNase E cleaved at the sequence RN ↓ CNU *in vivo* and *in vitro*, consistent with our results in chapter 1. These results demonstrated the importance of RNase E in mRNA metabolism. Lastly in chapter 4, we sought to investigate the boundary of scaffold domains in RNase E and used RNase E mutant strains to define the function of scaffold domains from different aspects. We showed that deletion of 330 residues of the N-terminal scaffold domain impacted sub-cellular location of RNase J, cell size, gene expression, and mRNA stability. Taken together, we demonstrated an important role of RNase E in mRNA degradation in *M. smegmatis*.

## References

- Afonyushkin, T., Vecerek, B., Moll, I., Blasi, U., & Kaberdin, V. R. (2005). Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res*, 33(5), 1678-1689. <https://doi.org/10.1093/nar/gki313>
- Ait-Bara, S., & Carpousis, A. J. (2010). Characterization of the RNA degradosome of *Pseudoalteromonas haloplanktis*: conservation of the RNase E-RhlB interaction in the gammaproteobacteria. *J Bacteriol*, 192(20), 5413-5423. <https://doi.org/10.1128/JB.00592-10>

- Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C. M., Bruni, C. B., & Carlomagno, M. S. (1994). Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA. *Genes Dev*, *8*(24), 3021-3031. <https://doi.org/10.1101/gad.8.24.3021>
- Apirion, D. (1978). Isolation, genetic mapping and some characterization of a mutation in *Escherichia coli* that affects the processing of ribonucleic acid. *Genetics*, *90*(4), 659-671. <https://doi.org/10.1093/genetics/90.4.659>
- Apirion, D., & Lassar, A. B. (1978). A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *J Biol Chem*, *253*(5), 1738-1742. <https://www.ncbi.nlm.nih.gov/pubmed/342528>
- Babitzke, P., Granger, L., Olszewski, J., & Kushner, S. R. (1993). Analysis of mRNA decay and rRNA processing in *Escherichia coli* multiple mutants carrying a deletion in RNase III. *J Bacteriol*, *175*(1), 229-239. <https://doi.org/10.1128/jb.175.1.229-239.1993>
- Babitzke, P., & Kushner, S. R. (1991). The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of *Escherichia coli*. *Proc Natl Acad Sci U S A*, *88*(1), 1-5. <https://doi.org/10.1073/pnas.88.1.1>
- Bardwell, J. C., Regnier, P., Chen, S. M., Nakamura, Y., Grunberg-Manago, M., & Court, D. L. (1989). Autoregulation of RNase III operon by mRNA processing. *EMBO J*, *8*(11), 3401-3407. <https://www.ncbi.nlm.nih.gov/pubmed/2583104>
- Blaszczak, J., Gan, J., Tropea, J. E., Court, D. L., Waugh, D. S., & Ji, X. (2004). Noncatalytic assembly of ribonuclease III with double-stranded RNA. *Structure*, *12*(3), 457-466. <https://doi.org/10.1016/j.str.2004.02.004>
- Blum, E., Carpousis, A. J., & Higgins, C. F. (1999). Polyadenylation promotes degradation of 3'-structured RNA by the *Escherichia coli* mRNA degradosome in vitro. *J Biol Chem*, *274*(7), 4009-4016. <https://doi.org/10.1074/jbc.274.7.4009>
- Britton, R. A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J., & Condon, C. (2007). Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol*, *63*(1), 127-138. <https://doi.org/10.1111/j.1365-2958.2006.05499.x>
- Buttner, K., Wenig, K., & Hopfner, K. P. (2005). Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol Cell*, *20*(3), 461-471. <https://doi.org/10.1016/j.molcel.2005.10.018>
- Callaghan, A. J., Aurikko, J. P., Ilag, L. L., Gunter Grossmann, J., Chandran, V., Kuhnel, K., Poljak, L., Carpousis, A. J., Robinson, C. V., Symmons, M. F., & Luisi, B. F. (2004). Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *J Mol Biol*, *340*(5), 965-979. <https://doi.org/10.1016/j.jmb.2004.05.046>
- Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., & Luisi, B. F. (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, *437*(7062), 1187-1191. <https://doi.org/10.1038/nature04084>
- Carpousis, A. J., Van Houwe, G., Ehretsmann, C., & Krisch, H. M. (1994). Copurification of *E. coli* RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, *76*(5), 889-900. [https://doi.org/10.1016/0092-8674\(94\)90363-8](https://doi.org/10.1016/0092-8674(94)90363-8)
- Carzaniga, T., Briani, F., Zangrossi, S., Merlino, G., Marchi, P., & Deho, G. (2009). Autogenous regulation of *Escherichia coli* polynucleotide phosphorylase expression revisited. *J Bacteriol*, *191*(6), 1738-1748. <https://doi.org/10.1128/JB.01524-08>
- Casaregola, S., Jacq, A., Laoudj, D., McGurk, G., Margaron, S., Tempete, M., Norris, V., & Holland, I. B. (1994). Cloning and analysis of the entire *Escherichia coli* ams gene. ams is identical to hmp1 and encodes a 114

- kDa protein that migrates as a 180 kDa protein. *J Mol Biol*, 238(5), 867.  
<https://doi.org/10.1006/jmbi.1994.1344>
- Chao, Y., Li, L., Girodat, D., Forstner, K. U., Said, N., Corcoran, C., Smiga, M., Papenfort, K., Reinhardt, R., Wieden, H. J., Luisi, B. F., & Vogel, J. (2017). In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Mol Cell*, 65(1), 39-51. <https://doi.org/10.1016/j.molcel.2016.11.002>
- Chauhan, A. K., & Apirion, D. (1991). The rne gene is the structural gene for the processing endoribonuclease RNase E of Escherichia coli. *Mol Gen Genet*, 228(1-2), 49-54. <https://doi.org/10.1007/BF00282446>
- Cheng, Z. F., & Deutscher, M. P. (2002). Purification and characterization of the Escherichia coli exoribonuclease RNase R. Comparison with RNase II. *J Biol Chem*, 277(24), 21624-21629.  
<https://doi.org/10.1074/jbc.M202942200>
- Cheng, Z. F., & Deutscher, M. P. (2005). An important role for RNase R in mRNA decay. *Mol Cell*, 17(2), 313-318.  
<https://doi.org/10.1016/j.molcel.2004.11.048>
- Cheng, Z. F., Zuo, Y., Li, Z., Rudd, K. E., & Deutscher, M. P. (1998). The vacB gene required for virulence in Shigella flexneri and Escherichia coli encodes the exoribonuclease RNase R. *J Biol Chem*, 273(23), 14077-14080.  
<https://doi.org/10.1074/jbc.273.23.14077>
- Clarke, J. E., Kime, L., Romero, A. D., & McDowall, K. J. (2014). Direct entry by RNase E is a major pathway for the degradation and processing of RNA in Escherichia coli. *Nucleic Acids Res*, 42(18), 11733-11751.  
<https://doi.org/10.1093/nar/gku808>
- Coburn, G. A., & Mackie, G. A. (1996a). Differential sensitivities of portions of the mRNA for ribosomal protein S20 to 3'-exonucleases dependent on oligoadenylation and RNA secondary structure. *J Biol Chem*, 271(26), 15776-15781. <https://doi.org/10.1074/jbc.271.26.15776>
- Coburn, G. A., & Mackie, G. A. (1996b). Overexpression, purification, and properties of Escherichia coli ribonuclease II. *J Biol Chem*, 271(2), 1048-1053. <https://doi.org/10.1074/jbc.271.2.1048>
- Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Volker, U., & Stulke, J. (2009). Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics*, 8(6), 1350-1360.  
<https://doi.org/10.1074/mcp.M800546-MCP200>
- Cormack, R. S., Genereaux, J. L., & Mackie, G. A. (1993). RNase E activity is conferred by a single polypeptide: overexpression, purification, and properties of the ams/rne/hmp1 gene product. *Proc Natl Acad Sci U S A*, 90(19), 9006-9010. <https://doi.org/10.1073/pnas.90.19.9006>
- Dasgupta, S., Fernandez, L., Kameyama, L., Inada, T., Nakamura, Y., Pappas, A., & Court, D. L. (1998). Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the Escherichia coli endoribonuclease RNase III--the effect of dsRNA binding on gene expression. *Mol Microbiol*, 28(3), 629-640.  
<https://doi.org/10.1046/j.1365-2958.1998.00828.x>
- Datta, A. K., & Niyogi, K. (1975). A novel oligoribonuclease of Escherichia coli. II. Mechanism of action. *J Biol Chem*, 250(18), 7313-7319. <https://www.ncbi.nlm.nih.gov/pubmed/170260>
- Davies, B. W., Kohrer, C., Jacob, A. I., Simmons, L. A., Zhu, J., Aleman, L. M., Rajbhandary, U. L., & Walker, G. C. (2010). Role of Escherichia coli YbeY, a highly conserved protein, in rRNA processing. *Mol Microbiol*, 78(2), 506-518. <https://doi.org/10.1111/j.1365-2958.2010.07351.x>
- Deana, A., & Belasco, J. G. (2004). The function of RNase G in Escherichia coli is constrained by its amino and carboxyl termini. *Mol Microbiol*, 51(4), 1205-1217. <https://doi.org/10.1046/j.1365-2958.2003.03905.x>

- Deutscher, M. P., & Reuven, N. B. (1991). Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, *88*(8), 3277-3280. <https://doi.org/10.1073/pnas.88.8.3277>
- Donovan, W. P., & Kushner, S. R. (1986). Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proc Natl Acad Sci U S A*, *83*(1), 120-124. <https://doi.org/10.1073/pnas.83.1.120>
- Durand, S., Gilet, L., Bessieres, P., Nicolas, P., & Condon, C. (2012). Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of *Bacillus subtilis* mRNAs. *PLoS Genet*, *8*(3), e1002520. <https://doi.org/10.1371/journal.pgen.1002520>
- Durand, S., Gilet, L., & Condon, C. (2012). The essential function of *B. subtilis* RNase III is to silence foreign toxin genes. *PLoS Genet*, *8*(12), e1003181. <https://doi.org/10.1371/journal.pgen.1003181>
- Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D., & Putzer, H. (2005). Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Res*, *33*(7), 2141-2152. <https://doi.org/10.1093/nar/gki505>
- Fang, M., Zeisberg, W. M., Condon, C., Ogryzko, V., Danchin, A., & Mechold, U. (2009). Degradation of nanoRNA is performed by multiple redundant RNases in *Bacillus subtilis*. *Nucleic Acids Res*, *37*(15), 5114-5125. <https://doi.org/10.1093/nar/gkp527>
- Filippov, V., Solovyev, V., Filippova, M., & Gill, S. S. (2000). A novel type of RNase III family proteins in eukaryotes. *Gene*, *245*(1), 213-221. [https://doi.org/10.1016/s0378-1119\(99\)00571-5](https://doi.org/10.1016/s0378-1119(99)00571-5)
- Fraza, C., McVey, C. E., Amblar, M., Barbas, A., Vornheim, C., Arraiano, C. M., & Carrondo, M. A. (2006). Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature*, *443*(7107), 110-114. <https://doi.org/10.1038/nature05080>
- Ghosh, S., & Deutscher, M. P. (1999). Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc Natl Acad Sci U S A*, *96*(8), 4372-4377. <https://doi.org/10.1073/pnas.96.8.4372>
- Gilson, E., Perrin, D., Saurin, W., & Hofnung, M. (1987). Species specificity of bacterial palindromic units. *J Mol Evol*, *25*(4), 371-373. <https://doi.org/10.1007/BF02603122>
- Gordon, G. C., Cameron, J. C., & Pflieger, B. F. (2017). RNA Sequencing Identifies New RNase III Cleavage Sites in *Escherichia coli* and Reveals Increased Regulation of mRNA. *mBio*, *8*(2). <https://doi.org/10.1128/mBio.00128-17>
- Guarneros, G., & Portier, C. (1990). Different specificities of ribonuclease II and polynucleotide phosphorylase in 3'mRNA decay. *Biochimie*, *72*(11), 771-777. [https://doi.org/10.1016/0300-9084\(90\)90186-k](https://doi.org/10.1016/0300-9084(90)90186-k)
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, *35*(3 Pt 2), 849-857. [https://doi.org/10.1016/0092-8674\(83\)90117-4](https://doi.org/10.1016/0092-8674(83)90117-4)
- Gupta, R. S., Kasai, T., & Schlessinger, D. (1977). Purification and some novel properties of *Escherichia coli* RNase II. *J Biol Chem*, *252*(24), 8945-8949. <https://www.ncbi.nlm.nih.gov/pubmed/336625>
- Hajnsdorf, E., Steier, O., Coscoy, L., Teyssset, L., & Regnier, P. (1994). Roles of RNase E, RNase II and PNPase in the degradation of the rpsO transcripts of *Escherichia coli*: stabilizing function of RNase II and evidence for efficient degradation in an *ams pnp rnb* mutant. *EMBO J*, *13*(14), 3368-3377. <https://www.ncbi.nlm.nih.gov/pubmed/7519147>
- Har-El, R., Silberstein, A., Kuhn, J., & Tal, M. (1979). Synthesis and degradation of lac mRNA in *E. coli* depleted of 30S ribosomal subunits. *Mol Gen Genet*, *173*(2), 135-144. <https://doi.org/10.1007/BF00330303>

- Hardwick, S. W., Chan, V. S., Broadhurst, R. W., & Luisi, B. F. (2011). An RNA degradosome assembly in *Caulobacter crescentus*. *Nucleic Acids Res*, *39*(4), 1449-1459. <https://doi.org/10.1093/nar/gkq928>
- Higgins, C. F., McLaren, R. S., & Newbury, S. F. (1988). Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene*, *72*(1-2), 3-14. [https://doi.org/10.1016/0378-1119\(88\)90122-9](https://doi.org/10.1016/0378-1119(88)90122-9)
- Hunt, A., Rawlins, J. P., Thomaidis, H. B., & Errington, J. (2006). Functional analysis of 11 putative essential genes in *Bacillus subtilis*. *Microbiology (Reading)*, *152*(Pt 10), 2895-2907. <https://doi.org/10.1099/mic.0.29152-0>
- Jacob, A. I., Kohrer, C., Davies, B. W., RajBhandary, U. L., & Walker, G. C. (2013). Conserved bacterial RNase YbeY plays key roles in 70S ribosome quality control and 16S rRNA maturation. *Mol Cell*, *49*(3), 427-438. <https://doi.org/10.1016/j.molcel.2012.11.025>
- Jager, S., Fuhrmann, O., Heck, C., Hebermehl, M., Schiltz, E., Rauhut, R., & Klug, G. (2001). An mRNA degrading complex in *Rhodobacter capsulatus*. *Nucleic Acids Res*, *29*(22), 4581-4588. <https://doi.org/10.1093/nar/29.22.4581>
- Jarrige, A., Brechemier-Baey, D., Mathy, N., Duche, O., & Portier, C. (2002). Mutational analysis of polynucleotide phosphorylase from *Escherichia coli*. *J Mol Biol*, *321*(3), 397-409. [https://doi.org/10.1016/s0022-2836\(02\)00645-9](https://doi.org/10.1016/s0022-2836(02)00645-9)
- Jarrige, A. C., Mathy, N., & Portier, C. (2001). PNPase autocontrols its expression by degrading a double-stranded structure in the pnp mRNA leader. *EMBO J*, *20*(23), 6845-6855. <https://doi.org/10.1093/emboj/20.23.6845>
- Jiang, X., Diwa, A., & Belasco, J. G. (2000). Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. *J Bacteriol*, *182*(9), 2468-2475. <https://doi.org/10.1128/JB.182.9.2468-2475.2000>
- Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M., & Pastore, A. (1995). Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J*, *14*(14), 3572-3584. <https://www.ncbi.nlm.nih.gov/pubmed/7628457>
- Khemici, V., Poljak, L., Luisi, B. F., & Carpousis, A. J. (2008). The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol Microbiol*, *70*(4), 799-813. <https://doi.org/10.1111/j.1365-2958.2008.06454.x>
- Kime, L., Clarke, J. E., Romero, A. D., Grasby, J. A., & McDowall, K. J. (2014). Adjacent single-stranded regions mediate processing of tRNA precursors by RNase E direct entry. *Nucleic Acids Res*, *42*(7), 4577-4589. <https://doi.org/10.1093/nar/gkt1403>
- Kime, L., Jourdan, S. S., Stead, J. A., Hidalgo-Sastre, A., & McDowall, K. J. (2010). Rapid cleavage of RNA by RNase E in the absence of 5' monophosphate stimulation. *Mol Microbiol*, *76*(3), 590-604. <https://doi.org/10.1111/j.1365-2958.2009.06935.x>
- Kinscherf, T. G., & Apirion, D. (1975). Polynucleotide phosphorylase can participate in decay of mRNA in *Escherichia coli* in the absence of ribonuclease II. *Mol Gen Genet*, *139*(4), 357-362. <https://doi.org/10.1007/BF00267975>
- Kole, R., Baer, M. F., Stark, B. C., & Altman, S. (1980). *E. coli* RNAase P has a required RNA component. *Cell*, *19*(4), 881-887. [https://doi.org/10.1016/0092-8674\(80\)90079-3](https://doi.org/10.1016/0092-8674(80)90079-3)
- Koslover, D. J., Callaghan, A. J., Marcaida, M. J., Garman, E. F., Martick, M., Scott, W. G., & Luisi, B. F. (2008). The crystal structure of the *Escherichia coli* RNase E apoprotein and a mechanism for RNA degradation. *Structure*, *16*(8), 1238-1244. <https://doi.org/10.1016/j.str.2008.04.017>
- Laalami, S., Bessieres, P., Rocca, A., Zig, L., Nicolas, P., & Putzer, H. (2013). *Bacillus subtilis* RNase Y activity in vivo analysed by tiling microarrays. *PLoS One*, *8*(1), e54062. <https://doi.org/10.1371/journal.pone.0054062>
- Laalami, S., Cavaiuolo, M., Roque, S., Chagneau, C., & Putzer, H. (2021). *Escherichia coli* RNase E can efficiently replace RNase Y in *Bacillus subtilis*. *Nucleic Acids Res*, *49*(8), 4643-4654. <https://doi.org/10.1093/nar/gkab216>

- Lee, K., Bernstein, J. A., & Cohen, S. N. (2002). RNase G complementation of rne null mutation identifies functional interrelationships with RNase E in Escherichia coli. *Mol Microbiol*, *43*(6), 1445-1456. <https://doi.org/10.1046/j.1365-2958.2002.02848.x>
- Lee, K., & Cohen, S. N. (2003). A Streptomyces coelicolor functional orthologue of Escherichia coli RNase E shows shuffling of catalytic and PNPase-binding domains. *Mol Microbiol*, *48*(2), 349-360. <https://doi.org/10.1046/j.1365-2958.2003.03435.x>
- Lehnik-Habrink, M., Newman, J., Rothe, F. M., Solovyova, A. S., Rodrigues, C., Herzberg, C., Commichau, F. M., Lewis, R. J., & Stulke, J. (2011). RNase Y in Bacillus subtilis: a Natively disordered protein that is the functional equivalent of RNase E from Escherichia coli. *J Bacteriol*, *193*(19), 5431-5441. <https://doi.org/10.1128/JB.05500-11>
- Lehnik-Habrink, M., Pfortner, H., Rempeters, L., Pietack, N., Herzberg, C., & Stulke, J. (2010). The RNA degradosome in Bacillus subtilis: identification of CshA as the major RNA helicase in the multiprotein complex. *Mol Microbiol*, *77*(4), 958-971. <https://doi.org/10.1111/j.1365-2958.2010.07264.x>
- Lehnik-Habrink, M., Schaffer, M., Mader, U., Diethmaier, C., Herzberg, C., & Stulke, J. (2011). RNA processing in Bacillus subtilis: identification of targets of the essential RNase Y. *Mol Microbiol*, *81*(6), 1459-1473. <https://doi.org/10.1111/j.1365-2958.2011.07777.x>
- Li de la Sierra-Gallay, I., Zig, L., Jamalli, A., & Putzer, H. (2008). Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol*, *15*(2), 206-212. <https://doi.org/10.1038/nsmb.1376>
- Li, Y., & Altman, S. (2003). A specific endoribonuclease, RNase P, affects gene expression of polycistronic operon mRNAs. *Proc Natl Acad Sci U S A*, *100*(23), 13213-13218. <https://doi.org/10.1073/pnas.2235589100>
- Li, Z., Pandit, S., & Deutscher, M. P. (1999). RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO J*, *18*(10), 2878-2885. <https://doi.org/10.1093/emboj/18.10.2878>
- Lorentzen, E., & Conti, E. (2005). Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. *Mol Cell*, *20*(3), 473-481. <https://doi.org/10.1016/j.molcel.2005.10.020>
- Lu, F., & Taghbalout, A. (2013). Membrane association via an amino-terminal amphipathic helix is required for the cellular organization and function of RNase II. *J Biol Chem*, *288*(10), 7241-7251. <https://doi.org/10.1074/jbc.M112.408674>
- Mackie, G. A. (1989). Stabilization of the 3' one-third of Escherichia coli ribosomal protein S20 mRNA in mutants lacking polynucleotide phosphorylase. *J Bacteriol*, *171*(8), 4112-4120. <https://doi.org/10.1128/jb.171.8.4112-4120.1989>
- Mackie, G. A. (1992). Secondary structure of the mRNA for ribosomal protein S20. Implications for cleavage by ribonuclease E. *J Biol Chem*, *267*(2), 1054-1061. <https://www.ncbi.nlm.nih.gov/pubmed/1370457>
- Mackie, G. A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, *395*(6703), 720-723. <https://doi.org/10.1038/27246>
- Mackie, G. A., & Genereaux, J. L. (1993). The role of RNA structure in determining RNase E-dependent cleavage sites in the mRNA for ribosomal protein S20 in vitro. *J Mol Biol*, *234*(4), 998-1012. <https://doi.org/10.1006/jmbi.1993.1654>
- Mader, U., Zig, L., Kretschmer, J., Homuth, G., & Putzer, H. (2008). mRNA processing by RNases J1 and J2 affects Bacillus subtilis gene expression on a global scale. *Mol Microbiol*, *70*(1), 183-196. <https://doi.org/10.1111/j.1365-2958.2008.06400.x>

- Marujo, P. E., Hajnsdorf, E., Le Derout, J., Andrade, R., Arraiano, C. M., & Regnier, P. (2000). RNase II removes the oligo(A) tails that destabilize the rpsO mRNA of Escherichia coli. *RNA*, 6(8), 1185-1193. <https://doi.org/10.1017/s135583820000073x>
- Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T., & Condon, C. (2007). 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell*, 129(4), 681-692. <https://doi.org/10.1016/j.cell.2007.02.051>
- Mathy, N., Hebert, A., Mervelet, P., Benard, L., Dorleans, A., Li de la Sierra-Gallay, I., Noirot, P., Putzer, H., & Condon, C. (2010). Bacillus subtilis ribonucleases J1 and J2 form a complex with altered enzyme behaviour. *Mol Microbiol*, 75(2), 489-498. <https://doi.org/10.1111/j.1365-2958.2009.07004.x>
- Matos, R. G., Barbas, A., Gomez-Puertas, P., & Arraiano, C. M. (2011). Swapping the domains of exoribonucleases RNase II and RNase R: conferring upon RNase II the ability to degrade ds RNA. *Proteins*, 79(6), 1853-1867. <https://doi.org/10.1002/prot.23010>
- Matsunaga, J., Dyer, M., Simons, E. L., & Simons, R. W. (1996). Expression and regulation of the rnc and pdxJ operons of Escherichia coli. *Mol Microbiol*, 22(5), 977-989. <https://doi.org/10.1046/j.1365-2958.1996.01529.x>
- McDowall, K. J., & Cohen, S. N. (1996). The N-terminal domain of the rne gene product has RNase E activity and is non-overlapping with the arginine-rich RNA-binding site. *J Mol Biol*, 255(3), 349-355. <https://doi.org/10.1006/jmbi.1996.0027>
- McDowall, K. J., Hernandez, R. G., Lin-Chao, S., & Cohen, S. N. (1993). The ams-1 and rne-3071 temperature-sensitive mutations in the ams gene are in close proximity to each other and cause substitutions within a domain that resembles a product of the Escherichia coli mre locus. *J Bacteriol*, 175(13), 4245-4249. <https://doi.org/10.1128/jb.175.13.4245-4249.1993>
- McDowall, K. J., Kaberdin, V. R., Wu, S. W., Cohen, S. N., & Lin-Chao, S. (1995). Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. *Nature*, 374(6519), 287-290. <https://doi.org/10.1038/374287a0>
- McDowall, K. J., Lin-Chao, S., & Cohen, S. N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J Biol Chem*, 269(14), 10790-10796. <https://www.ncbi.nlm.nih.gov/pubmed/7511606>
- Mechold, U., Fang, G., Ngo, S., Ogryzko, V., & Danchin, A. (2007). YtqI from Bacillus subtilis has both oligoribonuclease and pAp-phosphatase activity. *Nucleic Acids Res*, 35(13), 4552-4561. <https://doi.org/10.1093/nar/gkm462>
- Melefors, O., & von Gabain, A. (1991). Genetic studies of cleavage-initiated mRNA decay and processing of ribosomal 9S RNA show that the Escherichia coli ams and rne loci are the same. *Mol Microbiol*, 5(4), 857-864. <https://doi.org/10.1111/j.1365-2958.1991.tb00759.x>
- Miczak, A., Kaberdin, V. R., Wei, C. L., & Lin-Chao, S. (1996). Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proc Natl Acad Sci U S A*, 93(9), 3865-3869. <https://doi.org/10.1073/pnas.93.9.3865>
- Moffitt, J. R., Pandey, S., Boettiger, A. N., Wang, S., & Zhuang, X. (2016). Spatial organization shapes the turnover of a bacterial transcriptome. *Elife*, 5. <https://doi.org/10.7554/eLife.13065>
- Mohanty, B. K., & Kushner, S. R. (2000). Polynucleotide phosphorylase functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in Escherichia coli. *Proc Natl Acad Sci U S A*, 97(22), 11966-11971. <https://doi.org/10.1073/pnas.220295997>

- Mott, J. E., Galloway, J. L., & Platt, T. (1985). Maturation of *Escherichia coli* tryptophan operon mRNA: evidence for 3' exonucleolytic processing after rho-dependent termination. *EMBO J*, *4*(7), 1887-1891. <https://www.ncbi.nlm.nih.gov/pubmed/2992951>
- Mudd, E. A., Krisch, H. M., & Higgins, C. F. (1990). RNase E, an endoribonuclease, has a general role in the chemical decay of *Escherichia coli* mRNA: evidence that rne and ams are the same genetic locus. *Mol Microbiol*, *4*(12), 2127-2135. <https://doi.org/10.1111/j.1365-2958.1990.tb00574.x>
- Mudd, E. A., Prentki, P., Belin, D., & Krisch, H. M. (1988). Processing of unstable bacteriophage T4 gene 32 mRNAs into a stable species requires *Escherichia coli* ribonuclease E. *EMBO J*, *7*(11), 3601-3607. <https://www.ncbi.nlm.nih.gov/pubmed/3061803>
- Nashimoto, H., & Uchida, H. (1985). DNA sequencing of the *Escherichia coli* ribonuclease III gene and its mutations. *Mol Gen Genet*, *201*(1), 25-29. <https://doi.org/10.1007/BF00397981>
- Newman, J. A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C. R., & Lewis, R. J. (2011). Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure*, *19*(9), 1241-1251. <https://doi.org/10.1016/j.str.2011.06.017>
- Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev*, *23*(3), 371-390. <https://doi.org/10.1111/j.1574-6976.1999.tb00405.x>
- Ono, M., & Kuwano, M. (1979). A conditional lethal mutation in an *Escherichia coli* strain with a longer chemical lifetime of messenger RNA. *J Mol Biol*, *129*(3), 343-357. [https://doi.org/10.1016/0022-2836\(79\)90500-x](https://doi.org/10.1016/0022-2836(79)90500-x)
- Pandey, S. P., Minesinger, B. K., Kumar, J., & Walker, G. C. (2011). A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. *Nucleic Acids Res*, *39*(11), 4691-4708. <https://doi.org/10.1093/nar/gkr060>
- Pandey, S. P., Winkler, J. A., Li, H., Camacho, D. M., Collins, J. J., & Walker, G. C. (2014). Central role for RNase YbeY in Hfq-dependent and Hfq-independent small-RNA regulation in bacteria. *BMC Genomics*, *15*, 121. <https://doi.org/10.1186/1471-2164-15-121>
- Pertzev, A. V., & Nicholson, A. W. (2006). Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res*, *34*(13), 3708-3721. <https://doi.org/10.1093/nar/gkl459>
- Plocinski, P., Macios, M., Houghton, J., Niemiec, E., Plocinska, R., Brzostek, A., Slomka, M., Dziadek, J., Young, D., & Dziembowski, A. (2019). Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of *Mycobacterium tuberculosis*. *Nucleic Acids Res*, *47*(11), 5892-5905. <https://doi.org/10.1093/nar/gkz251>
- Portier, C., Dondon, L., Grunberg-Manago, M., & Regnier, P. (1987). The first step in the functional inactivation of the *Escherichia coli* polynucleotide phosphorylase messenger is a ribonuclease III processing at the 5' end. *EMBO J*, *6*(7), 2165-2170. <https://www.ncbi.nlm.nih.gov/pubmed/3308454>
- Purusharth, R. I., Klein, F., Sulthana, S., Jager, S., Jagannadham, M. V., Evguenieva-Hackenberg, E., Ray, M. K., & Klug, G. (2005). Exoribonuclease R interacts with endoribonuclease E and an RNA helicase in the psychrotrophic bacterium *Pseudomonas syringae* Lz4W. *J Biol Chem*, *280*(15), 14572-14578. <https://doi.org/10.1074/jbc.M413507200>
- Py, B., Higgins, C. F., Krisch, H. M., & Carpousis, A. J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature*, *381*(6578), 169-172. <https://doi.org/10.1038/381169a0>
- Regnier, P., Grunberg-Manago, M., & Portier, C. (1987). Nucleotide sequence of the pnp gene of *Escherichia coli* encoding polynucleotide phosphorylase. Homology of the primary structure of the protein with the RNA-

- binding domain of ribosomal protein S1. *J Biol Chem*, 262(1), 63-68.  
<https://www.ncbi.nlm.nih.gov/pubmed/2432069>
- Richards, J., & Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Mol Cell*, 74(2), 284-295 e285.  
<https://doi.org/10.1016/j.molcel.2019.01.044>
- Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., Condon, C., & Belasco, J. G. (2011). An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol Cell*, 43(6), 940-949. <https://doi.org/10.1016/j.molcel.2011.07.023>
- Robert-Le Meur, M., & Portier, C. (1992). E.coli polynucleotide phosphorylase expression is autoregulated through an RNase III-dependent mechanism. *EMBO J*, 11(7), 2633-2641.  
<https://www.ncbi.nlm.nih.gov/pubmed/1628624>
- Robertson, H. D. (1982). Escherichia coli ribonuclease III cleavage sites. *Cell*, 30(3), 669-672.  
[https://doi.org/10.1016/0092-8674\(82\)90270-7](https://doi.org/10.1016/0092-8674(82)90270-7)
- Robertson, H. D., Altman, S., & Smith, J. D. (1972). Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid precursor. *J Biol Chem*, 247(16), 5243-5251.  
<https://www.ncbi.nlm.nih.gov/pubmed/4560501>
- Robertson, H. D., Webster, R. E., & Zinder, N. D. (1968). Purification and properties of ribonuclease III from Escherichia coli. *J Biol Chem*, 243(1), 82-91. <https://www.ncbi.nlm.nih.gov/pubmed/4865702>
- Rosana, A. R., Whitford, D. S., Fahlman, R. P., & Owtrim, G. W. (2016). Cyanobacterial RNA Helicase CrhR Localizes to the Thylakoid Membrane Region and Cosediments with Degradosome and Polysome Complexes in *Synechocystis* sp. Strain PCC 6803. *J Bacteriol*, 198(15), 2089-2099.  
<https://doi.org/10.1128/JB.00267-16>
- Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, 48(1), 77-84. <https://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A*, 100(22), 12989-12994. <https://doi.org/10.1073/pnas.2134250100>
- Shahbabian, K., Jamalli, A., Zig, L., & Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *EMBO J*, 28(22), 3523-3533. <https://doi.org/10.1038/emboj.2009.283>
- Shi, Z., Yang, W. Z., Lin-Chao, S., Chak, K. F., & Yuan, H. S. (2008). Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA*, 14(11), 2361-2371.  
<https://doi.org/10.1261/rna.1244308>
- Spickler, C., & Mackie, G. A. (2000). Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *J Bacteriol*, 182(9), 2422-2427. <https://doi.org/10.1128/JB.182.9.2422-2427.2000>
- Stark, B. C., Kole, R., Bowman, E. J., & Altman, S. (1978). Ribonuclease P: an enzyme with an essential RNA component. *Proc Natl Acad Sci U S A*, 75(8), 3717-3721. <https://doi.org/10.1073/pnas.75.8.3717>
- Stead, M. B., Marshburn, S., Mohanty, B. K., Mitra, J., Pena Castillo, L., Ray, D., van Bakel, H., Hughes, T. R., & Kushner, S. R. (2011). Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res*, 39(8), 3188-3203. <https://doi.org/10.1093/nar/gkq1242>
- Stickney, L. M., Hankins, J. S., Miao, X., & Mackie, G. A. (2005). Function of the conserved S1 and KH domains in polynucleotide phosphorylase. *J Bacteriol*, 187(21), 7214-7221. <https://doi.org/10.1128/JB.187.21.7214-7221.2005>

- Stoppel, R., Manavski, N., Schein, A., Schuster, G., Teubner, M., Schmitz-Linneweber, C., & Meurer, J. (2012). RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast. *Nucleic Acids Res*, *40*(17), 8593-8606. <https://doi.org/10.1093/nar/gks613>
- Strahl, H., Turlan, C., Khalid, S., Bond, P. J., Kebalo, J. M., Peyron, P., Poljak, L., Bouvier, M., Hamoen, L., Luisi, B. F., & Carpousis, A. J. (2015). Membrane recognition and dynamics of the RNA degradosome. *PLoS Genet*, *11*(2), e1004961. <https://doi.org/10.1371/journal.pgen.1004961>
- Symmons, M. F., Jones, G. H., & Luisi, B. F. (2000). A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, *8*(11), 1215-1226. [https://doi.org/10.1016/s0969-2126\(00\)00521-9](https://doi.org/10.1016/s0969-2126(00)00521-9)
- Talkad, V., Achord, D., & Kennell, D. (1978). Altered mRNA metabolism in ribonuclease III-deficient strains of Escherichia coli. *J Bacteriol*, *135*(2), 528-541. <https://doi.org/10.1128/jb.135.2.528-541.1978>
- Taraseviciene, L., Bjork, G. R., & Uhlin, B. E. (1995). Evidence for an RNA binding region in the Escherichia coli processing endoribonuclease RNase E. *J Biol Chem*, *270*(44), 26391-26398. <https://doi.org/10.1074/jbc.270.44.26391>
- Taraseviciene, L., Miczak, A., & Apirion, D. (1991). The gene specifying RNase E (rne) and a gene affecting mRNA stability (ams) are the same gene. *Mol Microbiol*, *5*(4), 851-855. <https://doi.org/10.1111/j.1365-2958.1991.tb00758.x>
- Taverniti, V., Forti, F., Ghisotti, D., & Putzer, H. (2011). Mycobacterium smegmatis RNase J is a 5'-3' exo-endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol*, *82*(5), 1260-1276. <https://doi.org/10.1111/j.1365-2958.2011.07888.x>
- Van den Bossche, A., Hardwick, S. W., Ceyssens, P. J., Hendrix, H., Voet, M., Dendooven, T., Bandyra, K. J., De Maeyer, M., Aertsen, A., Noben, J. P., Luisi, B. F., & Lavigne, R. (2016). Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome. *Elife*, *5*. <https://doi.org/10.7554/eLife.16413>
- Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M., & Carpousis, A. J. (1998). Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome. *Genes Dev*, *12*(17), 2770-2781. <https://doi.org/10.1101/gad.12.17.2770>
- Vincent, H. A., & Deutscher, M. P. (2009). The roles of individual domains of RNase R in substrate binding and exoribonuclease activity. The nuclease domain is sufficient for digestion of structured RNA. *J Biol Chem*, *284*(1), 486-494. <https://doi.org/10.1074/jbc.M806468200>
- Wachi, M., Umitsuki, G., & Nagai, K. (1997). Functional relationship between Escherichia coli RNase E and the CafA protein. *Mol Gen Genet*, *253*(4), 515-519. <https://doi.org/10.1007/s004380050352>
- Wachi, M., Umitsuki, G., Shimizu, M., Takada, A., & Nagai, K. (1999). Escherichia coli cafA gene encodes a novel RNase, designated as RNase G, involved in processing of the 5' end of 16S rRNA. *Biochem Biophys Res Commun*, *259*(2), 483-488. <https://doi.org/10.1006/bbrc.1999.0806>
- Wu, A. M., Christie, G. E., & Platt, T. (1981). Tandem termination sites in the tryptophan operon of Escherichia coli. *Proc Natl Acad Sci U S A*, *78*(5), 2913-2917. <https://doi.org/10.1073/pnas.78.5.2913>
- Xu, F., & Cohen, S. N. (1995). RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation. *Nature*, *374*(6518), 180-183. <https://doi.org/10.1038/374180a0>
- Young, R. A., & Steitz, J. A. (1978). Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in Escherichia coli ribosomal precursor RNA. *Proc Natl Acad Sci U S A*, *75*(8), 3593-3597. <https://doi.org/10.1073/pnas.75.8.3593>

- Zeller, M. E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T., & Kaberdin, V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem J*, *403*(1), 207-215. <https://doi.org/10.1042/BJ20061530>
- Zhang, J. Y., Deng, X. M., Li, F. P., Wang, L., Huang, Q. Y., Zhang, C. C., & Chen, W. L. (2014). RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region. *RNA*, *20*(4), 568-579. <https://doi.org/10.1261/rna.043513.113>
- Zuo, Y., & Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res*, *29*(5), 1017-1026. <https://doi.org/10.1093/nar/29.5.1017>

## Chapter 2 : Defining the Transcriptional and Post-transcriptional Landscapes of *Mycobacterium smegmatis* in Aerobic Growth and Hypoxia

# Defining the Transcriptional and Post-transcriptional Landscapes of *Mycobacterium smegmatis* in Aerobic Growth and Hypoxia

**M. Carla Martini<sup>1</sup>, Ying Zhou<sup>1</sup>, Huaming Sun<sup>2</sup> and Scarlet S. Shell<sup>1,2\*</sup>**

<sup>1</sup> Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA, United States

<sup>2</sup> Program in Bioinformatics and Computational Biology, Worcester Polytechnic Institute, Worcester, MA, United States

Edited by: Pere-Joan Cardona, Germans Trias i Pujol Health Sciences Research Institute (IGTP), Spain

This chapter is a reproduction of a manuscript that was published as:

Martini MC, Zhou Y, Sun H and Shell SS (2019) Defining the Transcriptional and Post-transcriptional Landscapes of *Mycobacterium smegmatis* in Aerobic Growth and Hypoxia. *Front. Microbiol.* 10:591. doi: 10.3389/fmicb.2019.00591

## Author Contributions

MM, YZ, and SS conceived and designed the experiments. MM and YZ performed the experiments. MM, HS, and SS analyzed the data. MM and SS wrote the manuscript.

---

## Abstract

The ability of *Mycobacterium tuberculosis* to infect, proliferate, and survive during long periods in the human lungs largely depends on the rigorous control of gene expression. Transcriptome-wide analyses are key to understanding gene regulation on a global scale. Here, we combine 5'-end-directed libraries with RNAseq expression libraries to gain insight into the transcriptome organization and post-transcriptional mRNA cleavage landscape in mycobacteria during log phase growth and under hypoxia, a physiologically relevant stress condition. Using the model organism *Mycobacterium smegmatis*, we identified 6,090 transcription start sites (TSSs) with high confidence during log phase growth, of which 67% were categorized as primary TSSs for

annotated genes, and the remaining were classified as internal, antisense, or orphan, according to their genomic context. Interestingly, over 25% of the RNA transcripts lack a leader sequence, and of the coding sequences that do have leaders, 53% lack a strong consensus Shine-Dalgarno site. This indicates that like *M. tuberculosis*, *M. smegmatis* can initiate translation through multiple mechanisms. Our approach also allowed us to identify over 3,000 RNA cleavage sites, which occur at a novel sequence motif. To our knowledge, this represents the first report of a transcriptome-wide RNA cleavage site map in mycobacteria. The cleavage sites show a positional bias toward mRNA regulatory regions, highlighting the importance of post-transcriptional regulation in gene expression. We show that in low oxygen, a condition associated with the host environment during infection, mycobacteria change their transcriptomic profiles and endonucleolytic RNA cleavage is markedly reduced, suggesting a mechanistic explanation for previous reports of increased mRNA half-lives in response to stress. In addition, a number of TSSs were triggered in hypoxia, 56 of which contain the binding motif for the sigma factor SigF in their promoter regions. This suggests that SigF makes direct contributions to transcriptomic remodeling in hypoxia-challenged mycobacteria. Taken together, our data provide a foundation for further study of both transcriptional and posttranscriptional regulation in mycobacteria.

---

## Introduction

Tuberculosis is a disease of global concern caused by *Mycobacterium tuberculosis* (Mtb). This pathogen has the ability to infect the human lungs and survive there for long periods, often by entering into non-growing states. During infection, Mtb must overcome a variety of stressful conditions, including nutrient starvation, low pH, oxygen deprivation and the presence of reactive oxygen species. Consequently, the association of Mtb with its host and the adaptation to the surrounding environment requires rigorous control of gene expression.

As the slow growth rate and pathogenicity of Mtb present logistical challenges in the laboratory, many aspects of its biology have been studied in other mycobacterial species. One of the most widely used models is *Mycobacterium smegmatis*, a non-pathogenic fast-growing bacterium. While there are marked differences between the genomes of Mtb and *M. smegmatis*, such as the highly represented PE/PPE-like gene category and other virulence factors present in Mtb and poorly represented or absent in *M. smegmatis*, these organisms have at least 2,117 orthologous genes (Prasanna and Mehra, 2013) making *M. smegmatis* a viable model to address certain questions about the fundamental biology of mycobacteria. Indeed, studies using *M. smegmatis* have revealed key insights into relevant aspects of Mtb biology including the Sec and ESX secretion systems involved in transport of virulence factors (Coros et al., 2008; Rigel et al., 2009), bacterial survival during anaerobic dormancy (Dick et al., 1998; Bagchi et al., 2002; Trauner et al., 2012; Pecsí et al., 2014) and the changes induced during nutrient starvation (Elharar et al., 2014; Wu et al., 2016; Hayashi et al., 2018). However, the *M. smegmatis* transcriptome has been less extensively studied than that of Mtb.

Identification of transcription start sites (TSSs) is an essential step toward understanding how bacteria organize their transcriptomes and respond to changing environments. Genome-wide TSS mapping studies have been used to elucidate the general transcriptomic features in many bacterial species, leading to the identification of promoters, characterization of 5' untranslated regions (5' UTRs), identification of RNA regulatory elements and transcriptional changes in different environmental conditions (examples include (Albrecht et al., 2009; Mitschke et al., 2011; Cortes et al., 2013; Schlüter et al., 2013; Dinan et al., 2014; Ramachandran et al., 2014; Sass et al., 2015; Shell et al., 2015b; Thomason et al., 2015; Berger et al., 2016; Cuklina et al., 2016; D'arrigo et al., 2016; Heidrich et al., 2017; Li et al., 2017)). To date, two main studies have reported the transcriptomic landscape in Mtb during exponential growth and carbon starvation (Cortes et

al., 2013; Shell et al., 2015b). These complementary studies revealed that, unlike most bacteria, a substantial percentage (~25%) of the transcripts are leaderless, lacking a 5' UTR and consequently a Shine-Dalgarno ribosome-binding site. In addition, a number of previously unannotated ORFs encoding putative small proteins were found (Shell et al., 2015b), showing that the transcriptional landscape can be more complex than predicted by automated genome annotation pipelines. Thus, TSS mapping is a powerful tool to gain insight into transcriptomic organization and identify novel genes. Less is known about the characteristics of the *M. smegmatis* transcriptome. A recent study reported a number of *M. smegmatis* TSSs in normal growth conditions (Li et al., 2017). However, this work was limited to identification of primary gene-associated TSSs and lacked an analysis of internal and antisense TSSs, as well as characterization of promoter regions and other relevant transcriptomic features. In addition, Potgieter et al. (2016) validated a large number of annotated ORFs using proteomics and were able to identify 63 previously unannotated leaderless ORFs.

To achieve a deeper characterization of the *M. smegmatis* transcriptional landscape, we combined 5' -end-mapping and RNAseq expression profiling under two different growth conditions. Here we present an exhaustive analysis of the *M. smegmatis* transcriptome during exponential growth and hypoxia. Unlike most transcriptome-wide TSS analyses, our approach allowed us to study not only the transcriptome organization in different conditions, but also the frequency and distribution of RNA cleavage sites on a genome wide scale. Whereas regulation at the transcriptional level is assumed to be the main mechanism that modulates gene expression in bacteria, post-transcriptional regulation is a key step in the control of gene expression and has been implicated in the response to host conditions and virulence in various bacterial pathogens (Kulesekara et al., 2006; Mraheil et al., 2011; Heroven et al., 2012; Schifano et al., 2013; Holmqvist et al., 2016). Here we show that the predominant RNA cleavage sequence motif in *M.*

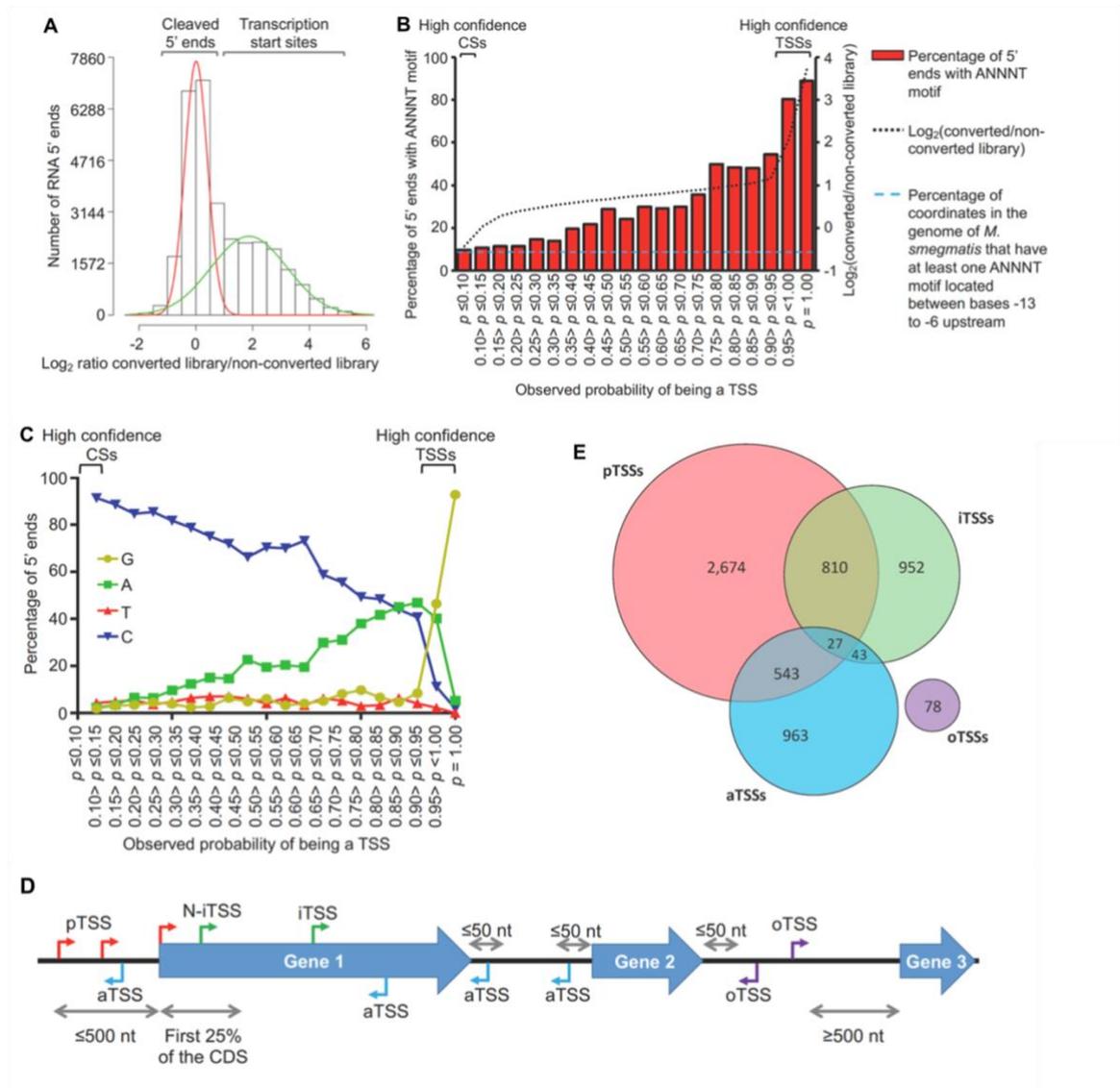
*smegmatis* is distinct from what has been reported for other bacteria. We also show that RNA cleavage decreases during adaptation to hypoxia, suggesting that RNA cleavage may be a refinement mechanism contributing to the regulation of gene expression in harsh conditions.

## Results

### Mapping, Annotation, and Categorization of Transcription Start Sites

In order to study the transcriptome structure of *M. smegmatis*, RNAs from triplicate cultures in exponential phase were used to construct 5' end mapping libraries (Dataset 1) according to our previously published methodologies (Shell et al., 2015a, b) with minor modifications. Briefly, our approach relies on comparison of adapter ligation frequency in a dephosphorylated (converted) library and an untreated (non-converted) library for each sample. The converted libraries capture both 5' triphosphate and native 5' monophosphate-bearing transcripts, while the non-converted libraries capture only native 5' monophosphate-bearing transcripts (Supplementary Figure S2-2). Thus, assessing the ratios of read counts in the converted/non-converted libraries permits discrimination between 5' triphosphate ends (primary transcripts from transcription start sites) and 5' monophosphate ends (cleavage sites). By employing a Gaussian mixture modeling analysis (Figure 2-1A) we were able to identify 5,552 TSSs in *M. smegmatis* with an observed probability of being a TSS  $\geq 0.95$  (high confidence TSSs, Supplementary Table S2-2). A second filtering method allowed us to obtain 222 additional TSSs from Dataset 1 (Supplementary Figure S2-3). A total of 5,774 TSSs were therefore obtained from Dataset 1. In addition, data from separate libraries constructed as controls for the hypoxia experiment (Dataset 2) in 'The Transcriptional Landscape Changes in Response to Oxygen Limitation' were also included in this analysis to obtain TSSs. After noise filtering (Supplementary Figure S2-3), 4,736 TSSs from Dataset 2 were identified. The union of the two datasets yielded a total of 6,090 non-redundant high confidence

TSSs, of which 4,420 were detected in both datasets (Supplementary Figure S2-4 and Supplementary Table S2-2).



**Figure 2-1: Mapping and categorization of transcription start sites in *M. smegmatis*.** (A) Diagram showing the ratios of coverage in the converted/non-converted libraries for each coordinate. Gaussian mixture modeling was used to discriminate between TSSs and CSs. For this analysis, the 15,720 coordinates from Dataset 1 were used. (B) Abundance of the ANNNT promoter motif located between bases -13 to -6 upstream of the 15,720 coordinates. The light blue dashed line indicates the percentage of coordinates in the genome of *M. smegmatis* that have at least one ANNNT motif located between bases -13 to -6 upstream (9.7%). (C) Base frequency at the +1 position among the 15,720 5' ends from Dataset 1. (D) Categories for TSS annotation based on the genomic context. TSSs were classified according to their relative position to genes as primary (pTSSs, red), internal (iTSSs, green), antisense (aTSSs, light blue) and orphan (oTSSs, violet). (E) Distribution of TSSs among the different categories.

Although not all 5' ends could be classified with the Gaussian mixture modeling, we were able to assign 57% of the 5' ends in Dataset 1 to one of the two 5' end populations with high confidence (5,552 TSSs and 3,344 CSs). To validate the reliability of the Gaussian mixture modeling used to classify 5' ends, we performed two additional analyses. First, according to previous findings in *Mtb* (Cortes et al., 2013) and other well studied bacteria (Sassetal.,2015; Bergeretal.,2016; C~uklinaetal.,2016; D'arrigo et al., 2016), we anticipated that TSSs should be enriched for the presence of the ANNNT -10 promoter consensus motif in the region upstream. Evaluation of the presence of appropriately- spaced ANNNT sequences revealed that 5' ends with higher probabilities of being TSSs are enriched for this motif, whereas for those 5' ends with low probabilities of being TSSs (and thus high probabilities of being CSs) have ANNNT frequencies similar to that of the *M. smegmatis* genome as a whole (Figure 2-1B). Secondly, we predicted that TSSs should show enrichment for A and G nts at the +1 position, given the reported preference for bacterial RNA polymerases to initiate transcription with these nts (Lewis and Adhya, 2004; Mendoza-Vargas et al., 2009; Mitschke et al., 2011; Cortes et al., 2013; Shell et al., 2015b; Thomason et al., 2015; Berger et al., 2016). Thus, we analyzed the base enrichment in the +1 position for the 5' ends according to the *p*-value in the Gaussian mixture modeling (Figure 2-1C). These results show a clear increase in the percentage of G and A bases in the position +1 as the probability of being a TSS increases, while the percentage of sequences having a C at +1 increases as the probability of being a TSS decreases. These two analyses show marked differences in the sequence contexts of TSSs and CSs and further validate the method used for categorization of 5' ends.

To study the genome architecture of *M. smegmatis*, the 6,090 TSSs were categorized according to their genomic context (Figures 2-1D, E and Supplementary Table S2-2). TSSs located  $\leq 500$

nt upstream of an annotated gene start codon in the *M. smegmatis* str. mc<sup>2</sup>155 (accession NC\_008596) reference genome was classified as primary TSSs (pTSS). TSSs within annotated genes on the sense strand were denoted as internal (iTSS). When an iTSS was located in the first quarter of an annotated gene, it was sub-classified as N-terminal associated TSS (N-iTSS) and was further examined to determine if it should be considered a primary TSS (see below). TSSs located on the antisense strand either within a gene or within a 5' UTR or 3' UTR were grouped as antisense TSSs (aTSSs). Finally, TSSs located in non-coding regions that did not meet the criteria for any of the above categories were classified as orphan (oTSSs). When a pTSS also met the criteria for classification in another category, it was considered to be pTSS for the purposes of downstream analyses. A total of 4,054 distinct TSSs met the criteria to be classified as pTSSs for genes transcribed in exponential phase. These pTSSs were assigned to 3,043 downstream genes, representing 44% of the total annotated genes (Supplementary Table S2-3). This number is lower than the total number of genes expressed in exponential phase, in large part due to the existence of polycistronic transcripts (see operon prediction below). Interestingly, 706 (23%) of these genes have at least two pTSSs and 209 (7%) have three or more, indicating that transcription initiation from multiple promoters is common in *M. smegmatis*. We used 5' RACE to confirm seven selected pTSSs (Supplementary Table S2-1), all of which mapped to the same position by both methods. Four of these were novel TSSs not reported by Li et al. (2017).

A total of 995 iTSSs (excluding the iTSSs that were also classified as a pTSS of a downstream gene, see Supplementary Figure S2-5 for classification workflow) were identified in 804 (12%) of the annotated genes, indicating that transcription initiation within coding sequences is common in *M. smegmatis*. iTSSs are often considered to be pTSSs of downstream genes, to be spurious events yielding truncated transcripts, or to be consequences of incorrect gene start annotations. However, there is evidence supporting the hypothesis that iTSSs are functional and highly

conserved among closely related bacteria (Shao et al., 2014), highlighting their potential importance in gene expression.

We were also able to detect antisense transcription in 12.5% of the *M. smegmatis* genes. Antisense transcription plays a role in modulation of gene expression by controlling transcription, RNA stability, and translation (Morita et al., 2005; Kawano et al., 2007; Andre et al., 2008; Fozo et al., 2008; Giangrossi et al., 2010) and has been found to occur at different rates across bacterial genera, ranging from 1.3% of genes in *Staphylococcus aureus* to up to 46% of genes in *Helicobacter pylori* (Beaume et al., 2010; Sharma et al., 2010). Of the 1,006 aTSSs identified here (excluding those that were primarily classified as pTSSs), 881 are within coding sequences, 120 are within 5' UTRs and 72 are located within 3' UTRs (note that some aTSS are simultaneously classified in more than one of these three subcategories, Supplementary Figure S2-6). While we expect that many of the detected antisense transcripts have biological functions, it is difficult to differentiate antisense RNAs with regulatory functions from transcriptional noise. In this regard, Lloréns-Rico et al. (2016) reported that most of the antisense transcripts detected using transcriptomic approaches are a consequence of transcriptional noise, arising at spurious promoters throughout the genome. To investigate the potential significance of the *M. smegmatis* aTSSs, we assessed the relative impact of each aTSS on local antisense expression levels by comparing the read depth upstream and downstream of each aTSS in our RNAseq expression libraries. We found 318 aTSSs for which expression coverage was  $\geq 10$ -fold higher in the 100 nt window downstream of the TSS compared to the 100 nt window upstream (Supplementary Table S2-4). Based on the magnitude of the expression occurring at these aTSS, we postulate that they could represent the 5' ends of candidate functional antisense transcripts rather than simply products of spurious transcription. However, further work is needed to test this hypothesis. Finally,

78 oTSSs were detected across the *M. smegmatis* genome. These TSSs may be the 5' ends of non-coding RNAs or mRNAs encoding previously unannotated ORFs.

Out of the 995 iTSSs identified, 457 were located within the first quarter of an annotated gene (N-iTSSs). In cases where we could not predict a pTSS with high confidence, we considered the possibility that the start codon of the gene was misannotated and the N-iTSS was in fact the primary TSS. Although we do not discount the possibility that functional proteins can be produced when internal transcription initiation occurs far downstream of the annotated start codon, we only considered N-iTSSs candidates for gene start reannotation when there was a start codon (ATG, GTG, or TTG) in-frame with the annotated gene in the first 30% of the annotated sequence. In this way, we suggest re-annotations of the start codons of 213 coding sequences (see Supplementary Figure S2-5 and Supplementary Table S2-5). These N-iTSSs were considered to be pTSSs (N-iTSSs → pTSSs) for all further analyses described in this work.

## **Operon Prediction**

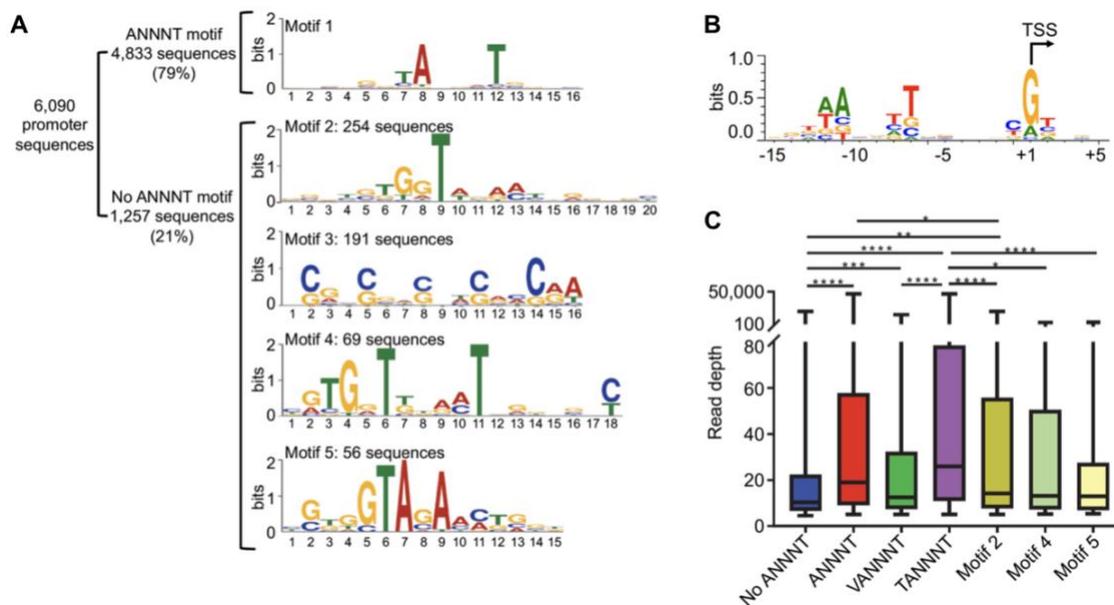
To predict operon structure, we combined 5' end libraries and RNAseq expression data. We considered two or more genes to be co-transcribed if (1) they had spanning reads that overlapped both the upstream and downstream gene in the expression libraries, (2) at least one TSS was detected in the 5' end- directed libraries for the first gene of the operon, and (3) the downstream gene(s) lacked pTSSs and iTSSs (for more detail, see section "Materials and Methods"). Thus, we were able to identify and annotate 294 operons with high confidence across the *M. smegmatis* genome (Supplementary Table S2-6). These operons are between 2 and 4 genes in length and comprise a total of 638 genes. Our operon prediction methodology has some limitations. For example, operons not expressed in exponential growth phase could not be detected in our study. Furthermore, internal promoters within operons can exist, leading to either monocistronic

transcripts or suboperons (Guell et al., 2009; Paletta and Ohman, 2012; Skliarova et al., 2012). We limited our operon predictions to genes that appear to be exclusively co-transcribed, excluding those cases in which an internal gene in an operon can be alternatively transcribed from an assigned pTSS. Finally, our analysis did not capture operons in which the first gene lacked a high-confidence pTSS. Despite these limitations, our approach allowed us to successfully identify new operons as well as previously described operons. Previously reported operons that were captured by our predictions included the *furA-katG* (MSMEG\_6383-MSMEG\_6384) operon involved in oxidative stress response (Milano et al., 2001), the *vapB-vapC* (MSMEG\_1283-MSMEG\_1284) Toxin–Antitoxin module (Robson et al., 2009) operon, and the *ClpP1-ClpP2* (MSMEG\_4672-MSMEG\_4673) operon involved in protein degradation (Raju et al., 2012).

### **Characterization of *M. smegmatis* Promoters Reveals Features Conserved in *M. tuberculosis***

Most bacterial promoters have two highly conserved regions, the –10 and the –35, that interact with RNA polymerase via sigma factors. However, it was reported that the –10 region is necessary and sufficient for transcription initiation by the housekeeping sigma factor SigA in mycobacteria, and no SigA –35 consensus motifs were identified in previous studies (Cortes et al., 2013; Newton-Foot and Gey van Pittius, 2013; Li et al., 2017; Zhu et al., 2017). To characterize the core promoter motifs in *M. smegmatis* on a global scale we analyzed the 50 bp upstream of the TSSs. We found that 4,833 of 6,090 promoters analyzed (79%) have an ANNNT motif located between positions –6 to –13 upstream the TSSs (Figure 2-2A). In addition, 63% of the promoters with ANNNT motifs have a thymidine preceding this sequence (TANNNT). This motif is similar to that previously described in a transcriptome–wide analysis for Mtb (Cortes et al., 2013) and for most bacterial promoters that are recognized by the  $\sigma_{70}$  housekeeping sigma factor (Ramachandran et al., 2014; Sassetal.,2015; Bergeretal.,2016; C`uklinaetal.,2016; D`arrigo et al., 2016). However,

no apparent bias toward specific bases in the NNN region was detected in our study or in *Mtb*, while in other bacteria such as *Escherichia coli*, *Salmonella enterica*, *Burkholderia cenocepacia*, *Pseudomonas putida*, and *Bacillus subtilis* an A/T preference was observed in this region (Jarmer et al., 2001; Ramachandran et al., 2014; Sass et al., 2015; Berger et al., 2016; D'arrigo et al., 2016). We were unable to detect a consensus motif in the -35 region either using MEME server (Bailey et al., 2015) or manually assessing the possible base- enrichment in the -35 region. Analysis of the sequences in the immediate vicinity of TSSs revealed that G and A are the most frequent bases at the +1 position, and C is considerably more abundant at -1 (Figure 2-2B).



**Figure 2-2: *M. smegmatis* promoter -10 regions are dominated by the ANNNT motif.** (A) Identification of promoter motifs. Consensus motifs were identified by using MEME. The 20 nt upstream the 6,090 TSSs were used for the initial analysis. Those sequences lacking an ANNNT -10 motif between positions -13 and -6 (1,257) were used to identify other conserved promoter sequences. Motif 2 (20 nt length) and Motif 4 (18 nt length) are located immediately upstream of the TSS (at the -1 position), while the spacing of Motif 5 varies from -4 to -1 relative to the TSS, with -3 being the dominant position (75% of the motifs). (B) The sequences flanking 3,500 randomly chosen TSSs were used to create a sequence logo by WebLogo 3 (Crooks et al., 2004), revealing the two dominant spacings for the ANNNT motif and base preferences in the immediate vicinity of the TSS. (C) Comparison of apparent promoter activity for different motifs. Mean normalized read depth in the converted libraries from Dataset 1 was compared for TSSs having or lacking the ANNNT motif in the -10 region, and ANNNT-associated TSSs were further subdivided into those

(Continued on next page)

the extended TANNNT motif or conversely the VANNNT sequence (where V = A, G or C). Motifs 2, 4, and 5 in Figure 2-2A are also included. \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$  (Kruskal–Wallis test with post-test for multiple comparisons).

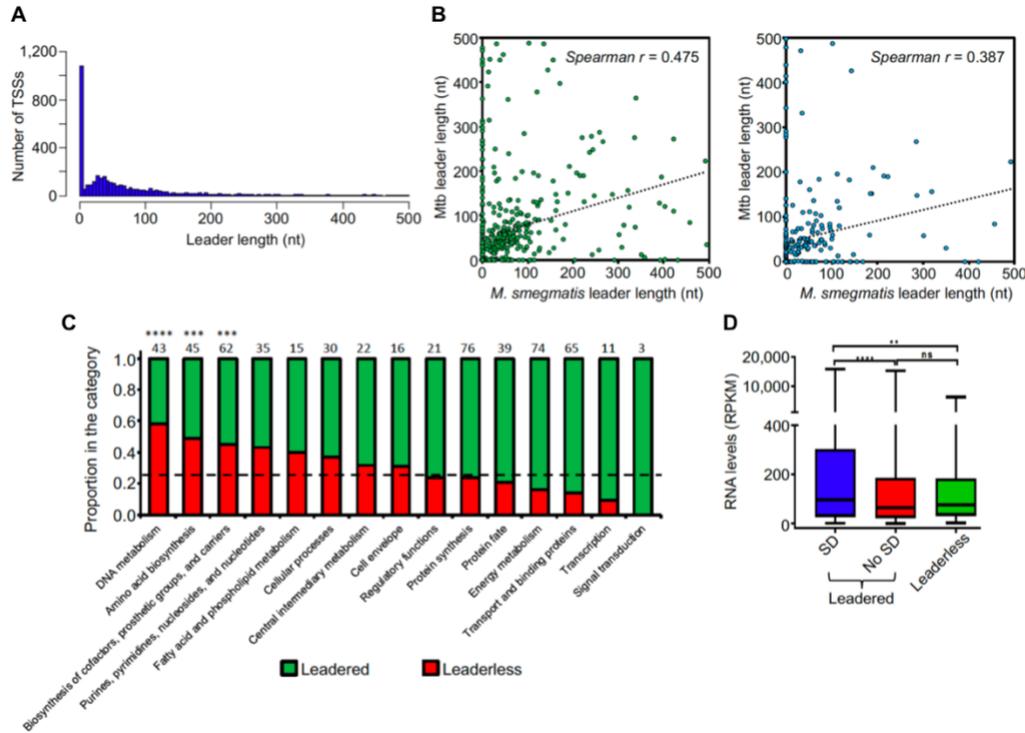
Interestingly, we identified several alternative motifs in the –10 promoter regions of transcripts lacking the ANNNT motif (Figure 2-2A). One of these, (G/C)NN(G/C)NN(G/C), is likely the signature of *M. smegmatis*' codon bias in the regions upstream of iTSSs. The other three sequences are candidate binding sites for alternative sigma factors, which are known to be important in regulation of transcription under diverse environmental conditions. However, the identified consensus sequences differ substantially from those previously described in mycobacteria (Raman et al., 2001, 2004; Sun et al., 2004; Lee et al., 2008a, b; Song et al., 2008; Veyrier et al., 2008; Humpel et al., 2010; Gaudion et al., 2013). The TSSs having these sigma factor motifs and the associated genes are listed in Supplementary Table S2-7. We next examined the relationship between promoter sequence and promoter strength, as estimated by the read depths in the 5' end converted libraries. As shown in Figure 2-2C, the expression levels of transcripts with ANNNT –10 motifs are on average substantially higher than those lacking this sequence. In addition, promoters with the full TANNNT motif are associated with more highly abundant transcripts compared to those having a VANNNT sequence, where V is G, A or C. These results implicate TANNNT as the preferred –10 sequence for the housekeeping sigma factor, SigA, in *M. smegmatis*. As shown in Figure 2-2C, expression levels of transcripts having the motif 2 in Figure 2-2A were significantly increased when compared to the total pool of transcripts lacking the ANNNT motif.

## **Leaderless Transcription Is a Prominent Feature of the *M. smegmatis* Transcriptome**

5' UTRs play important roles in post-transcriptional regulation and translation, as they may contain regulatory sequences that can affect mRNA stability and/or translation efficiency. Whereas in

most bacteria 5' UTR-bearing ("leadered") transcripts predominate, this is not the case for Mtb, in which near one quarter of the transcripts have been reported to be leaderless (Cortes et al., 2013; Shell et al., 2015b). To investigate this feature in *M. smegmatis*, we analyzed the 5' UTR lengths of all genes that had at least one pTSS. We found that for 24% of the transcripts the TSS coincides with the translation start site or produces a leader length  $\leq 5$  nt, resulting in leaderless transcripts (Figure 2-3A). This is less than the 40% reported for *M. smegmatis* in a smaller TSS-mapping study (Li et al., 2017), and suggests that the proportions of leaderless transcripts are in fact similar for *M. smegmatis* and Mtb. A total of 1,099 genes (including those re-annotated in section "Mapping, Annotation, and Categorization of Transcription Start Sites") have leaderless transcripts, and 155 of those (14%) are also transcribed as leadered mRNAs from separate promoters. Two of the pTSSs we validated by 5' RACE (Supplementary Table S2-1) belong to leaderless transcripts. For leadered transcripts, the median 5' UTR length was 69 nt. Interestingly, 15% of the leaders are  $>200$  nt, suggesting that these sequences may contain potential regulatory elements. We then sought to compare the leader lengths of *M. smegmatis* genes with the leader lengths of their homologs in Mtb. For this analysis we used two independent pTSS mapping Mtb datasets obtained from Cortes et al. (2013) and Shell et al. (2015b) (Figure 2-3B). To avoid ambiguities, we used only genes that had a single pTSS in both species. Our results show a statistically significant correlation of leader lengths between species, suggesting that similar genes conserve their transcript features and consequently may have related regulatory mechanisms. Additionally, comparison of leaderless transcription in *M. smegmatis* and Mtb revealed that 62% or 73% of the genes that are only transcribed as leaderless in *M. smegmatis* also lack a 5' UTR in MTB, according to Cortes et al. (2013) or Shell et al. (2015b), respectively (Supplementary Table S2-8). We next assessed if leaderless transcripts are associated with particular gene categories and found the distribution across categories was uneven (Figure 2-3C). The three categories "DNA metabolism," "Amino acid biosynthesis," and "Biosynthesis of

cofactors, prosthetic groups and carriers” were significantly enriched in leaderless transcripts ( $p$ -value  $< 0.05$ , hypergeometric test), while “Signal transduction,” “Transcription,” and “Transport and binding proteins” appear to have fewer leaderless transcripts.



**Figure 2-3: Leader features are conserved in mycobacteria.** (A) Leader length distribution. The 4,054 pTSSs and the pTSSs of the 213 reannotated genes (N-iTSSs  $\rightarrow$  pTSSs) were used. (B) Leader length correlation between *M. smegmatis* and Mtb genes. The leader sequences of genes having a single unique pTSS in both species (leader length  $\geq 0$  and  $\leq 500$  nt) were used. 508 homologous genes in Cortes et al. (2013) (left figure) and 251 homologous genes in Shell et al. (2015b) (right figure) were used. When a gene in *M. smegmatis* had more than one homolog in Mtb, that with the highest identity was considered. Spearman  $r$   $p$ -value  $< 0.00001$  in both cases. (C) Distribution of leaderless transcripts among different functional TIGRfam functional categories (Haft et al., 2001). 557 genes having TIGRfam categories were used for this analysis. Genes having both leadered and leaderless transcripts were excluded. The black dashed line indicates the expected proportion of leaderless genes (25%) according to the global analysis performed in this study. The numbers above each bar indicate the total number of genes used for this analysis in each category (leaderless + leadered). \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$  (Chi-Square test with Bonferroni correction for multiple comparisons). (D) RNA levels vary according to leader status. Mean expression levels were compared for genes expressed with leaders containing a canonical SD sequence (SD) or not (No SD) or lacking leaders (leaderless). Gene expression was quantified by RNAseq. Genes were classified as containing an SD sequence if at least one of the three tetramers AGGA, GGAG, or GAGG (core sequence AGGAGG) were present in the region  $-6$  to  $-17$  nt relative to the start codon. rRNAs, tRNAs, sRNAs, and genes expressed as both leadered and leaderless transcripts were excluded. \*\*\*\* $p < 0.0001$ , \*\* $p < 0.005$ ; ns: not significant. (Kruskal–Wallis test with post-test for multiple comparisons).

We next evaluated the presence of the Shine-Dalgarno ribosome-binding site (SD) upstream of leadered coding sequences. For this analysis, we considered those leaders containing at least one of the three tetramers AGGA, GGAG or GAGG (core sequence AGGAGG) in the region -6 to -17 relative to the start codon to possess canonical SD motifs. We found that only 47% of leadered coding sequences had these canonical SD sequences. Thus, considering also the leaderless RNAs, a large number of transcripts lack canonical SD sequences, suggesting that translation initiation can occur through multiple mechanisms in *M. smegmatis*. We further compared the relative expression levels of leaderless and leadered coding sequences subdivided by SD status. Genes expressed as both leadered and leaderless transcripts were excluded from this analysis. We found that on average, expression levels were significantly higher for those genes with canonical SD sequences than for those with leaders but lacking this motif and for those that were leaderless (Figure 2-3D). Together, these data suggest that genes that are more efficiently translated have also higher transcript levels. Similar findings were made in Mtb, where proteomic analyses showed increased protein levels for genes with SD sequences compared to those lacking this motif (Cortes et al., 2013).

### **Identification of Novel Leaderless ORFs in the *M. smegmatis* Genome**

As GTG or ATG codons are sufficient to initiate leaderless translation in mycobacteria (Shell et al., 2015b; Potgieter et al., 2016), we used this feature to look for unannotated ORFs in the *M. smegmatis* NC\_008596 reference genome. Using 1,579 TSSs that remained after pTSS assignment and gene reannotation using N-iTSSs (see Supplementary Figure S2-5) we identified a total of 66 leaderless ORFs encoding putative proteins longer than 30 amino acids, 5 of which were previously identified (Shell et al., 2015b). 83% of these ORFs were predicted in other annotations of the *M. smegmatis* mc<sup>2</sup>155 or MKD8 genome [NC\_018289.1, (Gray et al., 2013)], while 10 of the remaining ORFs showed homology to genes annotated in other mycobacterial species and *Helobdella robusta* and two ORFs did not show homology to any known protein. The

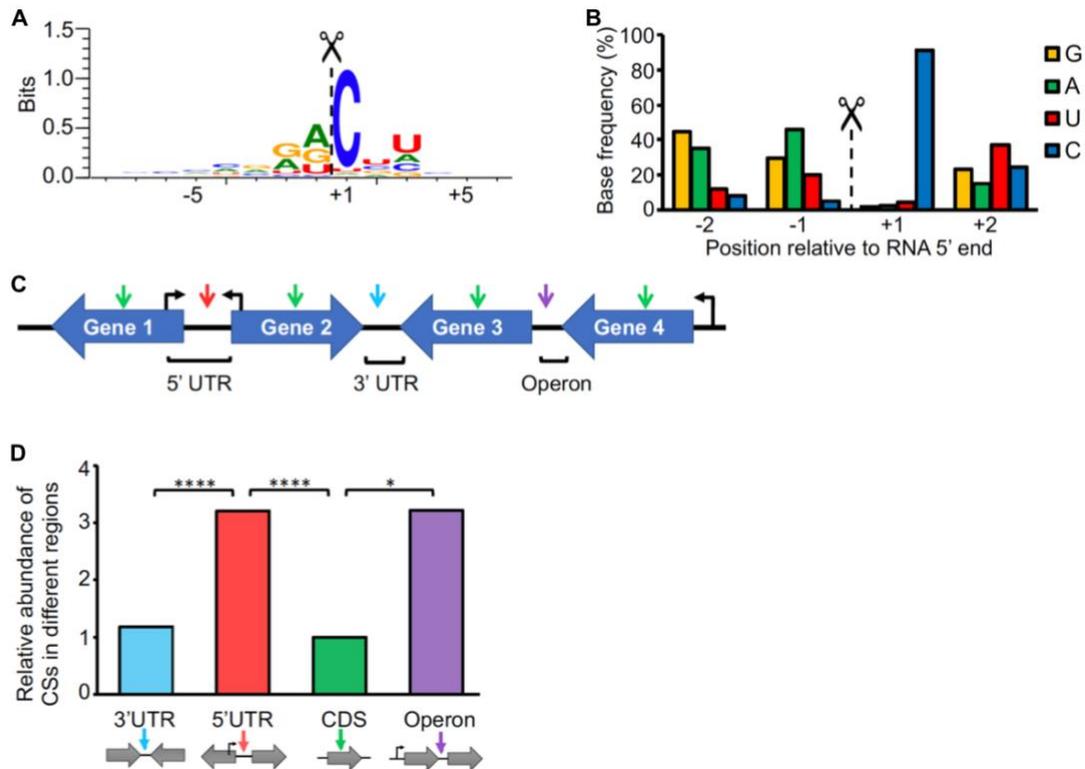
TSS of ORF15 was validated by 5'RACE. These results show that automatic annotation of genomes can be incomplete and highlight the utility of transcriptomic analysis for genome (re)annotation. Detailed information on these novel putative ORFs is provided in Supplementary Table S2-9.

### **Endonucleolytic RNA Cleavage Occurs at a Distinct Sequence Motif and Is Common in mRNA Regulatory Regions**

As our methodology allows us to precisely map RNA cleavage sites in addition to TSSs, we sought to analyze the presence and distribution of cleavage sites in the *M. smegmatis* transcriptome. mRNA processing plays a crucial role in regulation of gene expression, as it is involved in mRNA maturation, stability and degradation (Arraiano et al., 2010). Mixture modeling identified 3,344 CSs with a posterior probability  $\geq 0.9$  (high confidence CSs) (Figure 2-1A and Supplementary Table S2-10). To determine the sequence context of the CSs, we used the regions flanking the 5' ends to generate a sequence logo (Figure 2-4A). There was a strong preference for a cytosine in the +1 position (present in more than the 90% of the CSs) (Figure 2-4B), suggesting that it may be structurally important for RNase recognition and/or catalysis.

Cleaved 5' ends can represent either degradation intermediates or transcripts that undergo functional processing/maturation. In an attempt to investigate CS function, we classified them according to their locations within mRNA transcripts (Figure 2-4C and Supplementary Table S2-10). We found that, after normalizing to the proportion of the expressed transcriptome that is comprised by each location category, cleaved 5' ends are more abundant within 5' UTRs and intergenic regions of operons than within coding sequences and 3' UTRs (Figure 2-4D). Stringent criteria were used in these analyses to avoid undesired bias (Figure 2-4C and see section

“Materials and Methods”). While one would expect the CSs associated with mRNA turnover to be evenly distributed throughout the transcript, enrichment of CSs within the 5' UTRs as well as between two co-transcribed genes may be indicative of cleavages associated with processing and maturation. Alternatively, these regions may be more susceptible to RNases due to lack of associated ribosomes. Here we predicted with high confidence that at least 101 genes have one or more CSs in their 5' UTRs (Supplementary Table S2-11).



**Figure 2-4: Cleavage site positions are biased with respect to sequence context and genetic location.**

(A) Sequence context of cleavage sites. The sequences flanking the 3,344 high-confidence CSs were used to create the sequence logo with WebLogo 3 (Crooks et al., 2004). (B) Base preference for RNA cleavage. The base frequencies for the -2 to +2 positions were determined. (C) Cleavage site categories based on the genetic context. CSs are denoted with arrows. 5' UTR: the CS is within the leader of a gene, and the genes upstream and downstream of the CS are divergent (Gene 1 and Gene 2, red arrow). CDS: The CS is within a coding sequence (green arrow). 3' UTR: the genes upstream and downstream of the CS are convergent (Gene 2 and Gene 3, light blue arrow). Operon: The CS is between two genes with the same orientation and the first gene in the operon has a pTSS according to Supplementary Table S6 (violet arrow). (D) Distribution of cleavage sites. The frequency of CSs in each location was normalized to the proportion of the genome that the location category comprised. The proportions were then normalized to the CDS category, which was set as 1. \*\*\*\* $p < 0001$ , \* $p < 0.01$  (Chi-square test).

We detected cleaved 5' ends within the coding sequences of 18% of *M. smegmatis* genes, ranging from 1 to over 40 sites per gene. We analyzed the distribution of CSs within coding sequences (Supplementary Figure S2-7), taking into consideration the genomic context of the genes. When analyzing the distribution of CSs within the coding sequences of genes whose downstream gene has the same orientation, we observed an increase in CS frequency in the region near the stop codon (Supplementary Figure S2-7A). However, when only coding sequences having a downstream gene on the opposite strand (convergent) were considered, the distribution of CSs through the coding sequences was significantly different ( $p$ -value < 0.0001, Kolmogorov-Smirnov D test) with the CSs more evenly distributed throughout the coding sequence (Supplementary Figure S2-7B). This suggests that the cleavage bias toward the end of the genes observed in Supplementary Figure S2-7A may be due to the fact that many of these CSs are actually occurring in the 5' UTRs of the downstream genes. In cases where the TSS of a given gene occurs within the coding sequence of the preceding gene, a CS may map to both the coding sequence of the upstream gene and the 5' UTR of the downstream gene. In these cases, we cannot determine in which of the two transcripts the cleavage occurred. However, cleavages may also occur in polycistronic transcripts. We therefore assessed the distributions of CSs in the operons predicted above. The distribution of CSs in genes co-transcribed with a downstream gene showed a slight increase toward the last part of the gene (Supplementary Figure S2-7C). This may reflect cases in which polycistronic transcripts are cleaved near the 3' end of an upstream gene, as has been reported for the *furA-katG* operon, in which a cleavage near the stop codon of *furA* was described (Milano et al., 2001; Sala et al., 2008; Taverniti et al., 2011). The *furA-katG* cleavage was identified in our dataset, located 1 nt downstream of the previously reported position. A similar enrichment of CSs toward stop codons was also observed in a recent genome-wide RNA cleavage analysis in *S. enterica* (Chao et al., 2017), although in

this case the high frequency of cleavage may be also attributed to the U preference of RNase E in this organism, which is highly abundant in these regions.

### **Prediction of Additional TSSs and CSs Based on Sequence Context**

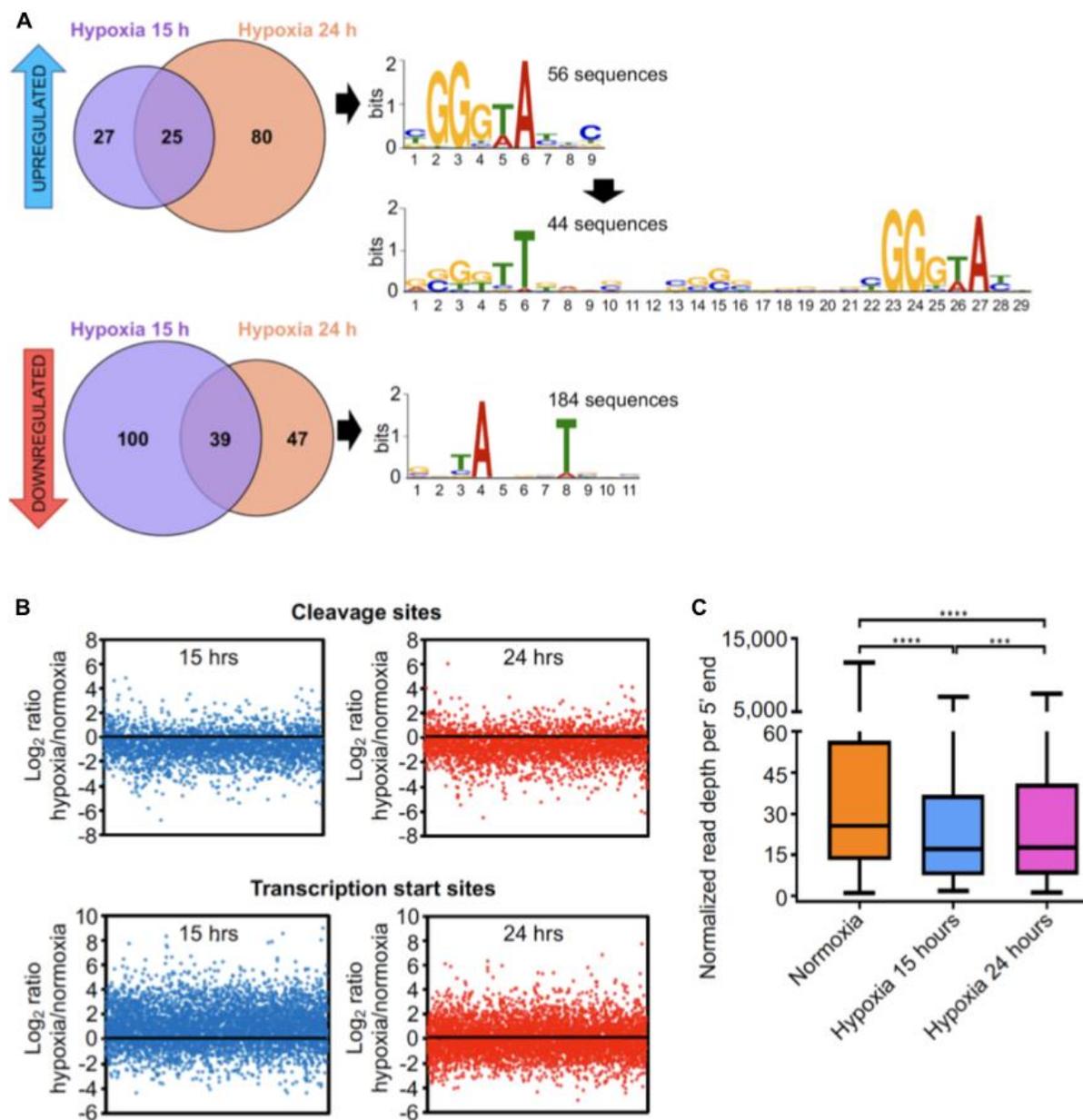
The sequence contexts of TSSs (Figure 2-2B) and CSs (Figure 2-4A) were markedly different, as G and A were highly preferred in the TSS +1 position whereas C was highly preferred in the CS +1 position, and TSSs were associated with a strong overrepresentation of ANNNT –10 sites while CSs were not. These sequence-context differences not only provide validation of our methodology for distinguishing TSSs from CSs, as discussed above, but also provide a means for making improved predictions of the nature of 5' ends that could not be categorized with high confidence based on their converted/non-converted library coverage alone. Taking advantage of these differences, we sought to obtain a list of additional putative TSSs and CSs. Thus, of the 5' ends that were not classified with high confidence by mixture modeling, we selected those that had an appropriately positioned ANNNT motif upstream and a G or an A in the +1 position and classified them as TSSs with medium confidence (Supplementary Table S2-12). In the same way, 5' ends with a C in the +1 position and lacking the ANNNT motif in the region upstream were designated as medium confidence CSs (Supplementary Table S2-13). In this way, we were able to obtain 576 and 4,838 medium confidence TSSs and CSs, respectively. Additional validation of a medium confidence TSS was performed for gene MSMEG\_0063 using 5' RACE. We were able to corroborate that, as predicted, transcription of this gene is initiated 139 bp upstream the coding sequence and that either deletion or mutation of the predicted –10 promoter region dramatically decreased transcription initiation (Supplementary Figure S2-8). These results support the value of TSS prediction based on –10 promoter region motif and base composition at +1 position and highlight the importance of the –10 ANNNT promoter motif for mycobacterial transcription. Three medium confidence CSs (86927+, 87293+, and 5038902–) were also validated using 5' RACE.

Although we are aware of the limitations of these predictions, these lists of medium confidence 5' ends provide a resource that may be useful for guiding further studies. 5' ends that did not meet the criteria for high or medium confidence TSSs or CSs are reported in Supplementary Table S2-14.

## **The Transcriptional Landscape Changes in Response to Oxygen Limitation**

We sought to study the global changes occurring at the transcriptomic level in oxygen limitation employing a system similar to the Wayne model (Wayne and Hayes, 1996) (see section “Materials and Methods”). Two timepoints were experimentally determined in order to evaluate transcriptomic changes during the transition into hypoxia (Supplementary Figure S2-1). A different enzyme was used for conversion of 5' triphosphates to 5' monophosphates in these 5'-end libraries, and it appeared to be less effective than the enzyme used for the 5' end libraries in Dataset 1. As a consequence, our ability to distinguish TSSs from CSs de novo in these datasets was limited. However, we were able to assess changes in abundance of the 5' ends classified as high-confidence TSSs or CSs in Dataset 1, as well as identify a limited number of additional TSSs and CSs with high confidence (Supplementary Figure S2-4 and Supplementary Table S2-3). Corresponding RNAseq expression libraries revealed that, as expected, a large number of genes were up and downregulated in response to oxygen limitation (Supplementary Figure S2-9 and Supplementary Table S2-15). We next investigated the transcriptional changes in hypoxia by assessing the relative abundance of TSSs in these conditions. We found 318 high- confidence TSSs whose abundance varied substantially between exponential phase and hypoxia (Supplementary Table S2-16). A robust correlation was observed between the pTSS peak height in the 5'-end-directed libraries and RNA levels in the expression libraries for hypoxia (Supplementary Figure S2-10). In an attempt to identify promoter motifs induced in hypoxia, we analyzed the upstream regions of those TSSs whose abundance increased (fold change  $\geq 2$ ,

adjusted  $p$ -value  $\leq 0.05$ ). Interestingly, we detected a conserved GGGTA motif in the  $-10$  region of 56 promoters induced in hypoxia using MEME (Figure 2-5A and Supplementary Table S2-16). This motif was reported as the binding site for alternative sigma factor SigF (Rodrigue et al., 2007; Hartkoorn et al., 2010; Humpel et al., 2010). Additionally, the extended  $-35$  and  $-10$  SigF motif was found in 44 of the 56 promoter sequences (Figure 2-5A and Supplementary Table S2-16). SigF was shown to be induced in hypoxia at the transcript level in Mtb (Iona et al., 2016) and highly induced at the protein level under anaerobic conditions using the Wayne model in *Mycobacterium bovis* BCG strain and Mtb (Michele et al., 1999; Galagan et al., 2013). In *M. smegmatis*, SigF was shown to play a role under oxidative stress, heat shock, low pH and stationary phase (Gebhard et al., 2008; Humpel et al., 2010; Singh et al., 2015) and *sigF* RNA levels were detected in exponential phase at a nearly comparable level to *sigA* (Singh and Singh, 2008). Here, we did not detect significant changes in expression of the *sigF* gene in hypoxia at the transcript level. However, this is consistent with reported data showing that *sigF* transcript levels remain unchanged under stress conditions in *M. smegmatis* (Gebhard et al., 2008), as it was postulated that SigF is post-transcriptionally modulated via an anti-sigma factor rather than through *sigF* transcription activation (Beaucher et al., 2002). We noted that, in the case of TSSs whose abundance was reduced in hypoxia, almost the totality of the promoters contains the  $-10$  ANNNT  $\sigma 70$  binding motif. We then examined the presence of SigF motif in the regions upstream of 5' ends that were not classified as high confidence TSSs. We speculate that 5' ends associated with this motif may be potential TSSs triggered by hypoxia. We found 96 additional putative TSSs that were (1) overrepresented in hypoxia and (2) associated with appropriately-spaced SigF motifs (Supplementary Table S2-17). Three of the hypoxia- induced genes with SigF motifs (MSMEG\_3460, MSMEG\_4195 and MSMEG\_5329) have homologous genes induced in hypoxia in Mtb (Park et al., 2003; Rustad et al., 2008).



**Figure 2-5: The transcriptional landscape substantially changes upon oxygen limitation.** (A) TSSs significantly increased or decreased in hypoxia. 132 TSSs were overrepresented (upper panel) and 186 were underrepresented (lower panel) in different hypoxia stages. The upstream regions of these TSSs were used to search for promoter motifs using MEME. (B) The mean normalized read depths for each 5' end in the non-converted libraries were compared between hypoxia and normoxia. Graphics show the Log<sub>2</sub> of the ratios of read depth for each CSs at 15 h (upper left) and 24 h (upper right), and the Log<sub>2</sub> of the ratios of the read depth for each TSSs at 15 h (lower left) and 24 h (lower right) compared to normoxia. (C) Normalized read depth at high-confidence cleavage sites under normoxia and the transition into hypoxia. \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ ; ns, not significant (non-parametric Wilcoxon matched-pairs signed rank test).

It is well known that under anaerobic conditions mycobacteria induce the DosR regulon, a set of genes implicated in stress tolerance (Rosenkrands et al., 2002; O'Toole et al., 2003; Park et al., 2003; Roberts et al., 2004; Rustad et al., 2008; Honaker et al., 2009; Leistikow et al., 2010). The DosR transcriptional regulator was highly upregulated at both hypoxic timepoints in the expression libraries (13 and 18-fold at 15 and 24 h, respectively, Supplementary Figure 2-S9) and 30 out of the 49 DosR-activated genes (Berney et al., 2014) were upregulated in our dataset. Thus, we hypothesized that the DosR binding motif should be present in a number of regions upstream the TSSs that were upregulated in hypoxia. Analysis of the 200 bp upstream the TSSs using the CentriMo tool for local motif enrichment analysis (Bailey and Machanick, 2012) allowed us to detect putative DosR motifs in 13 or 53 promoters, depending on whether a stringent (GGGACTTNGNCCCT) or a weak (RRGNCYWNNNGNMM) consensus sequence was used as input (Lun et al., 2009; Berney et al., 2014; Gomes et al., 2014) (Supplementary Table S2-16). At least two of the 13 genes downstream of these TSSs were previously reported to have DosR motifs by Berney et al. (2014) and RegPrecise Database (Novichkov et al., 2013) and two others are homologs of genes in the Mtb DosR regulon that were not previously described in *M. smegmatis* as regulated by DosR (Supplementary Table S2-16).

We then used CentriMo to search for DosR motifs in the regions upstream of 5' ends that were not classified as high confidence TSSs, given that TSSs derived from hypoxia-specific promoters may have been absent from Dataset 1. We found 36 putative TSSs associated with 20 different genes (Supplementary Table S2-18), of which 11 have been shown to have DosR binding motifs (Berney et al., 2014). Five of these are homologs of genes in the Mtb DosR regulon.

## ***M. smegmatis* Decreases RNA Cleavage Under Oxygen Limitation**

There is evidence that mycobacterial mRNA is broadly stabilized under hypoxia and other stress conditions (Rustad et al., 2013; Ignatov et al., 2015). Thus, we anticipated that RNA cleavage should be reduced under hypoxia as a strategy to stabilize transcripts. We compared the relative abundance of each high confidence CS in stress and in exponential phase (Figure 2-5B) and found that RNA cleavage is significantly reduced in both hypoxia 15 and 24h on a global scale (Figure 2-5C). In contrast, relative abundance of TSSs did not decrease in these conditions, indicating that the reduction in CSs is not an artifact of improper normalization (Figure 2-5B). When the ratios of CSs abundance in hypoxia/normal growth of individual genes were analyzed, we observed the same behavior (Supplementary Figure S2-11). These results indicate that the number of cleavage events per gene decreases during adaptation to hypoxia, which could contribute to the reported increases in half-life (Rustad et al., 2013).

## **Discussion**

In recent years, genome-wide transcriptome studies have been widely used to elucidate the genome architecture and modulation of transcription in different bacterial species (Albrecht et al., 2009; Mendoza-Vargas et al., 2009; Mitschke et al., 2011; Cortes et al., 2013; Schlüter et al., 2013; Dinan et al., 2014; Ramachandran et al., 2014; Innocenti et al., 2015; Sass et al., 2015; Thomason et al., 2015; Berger et al., 2016; Cuklina et al., 2016; D'arrigo et al., 2016; Heidrich et al., 2017; Li et al., 2017; Zhukova et al., 2017). Here we combined 5'-end-directed libraries and RNAseq expression libraries to shed light on the transcriptional and post-transcriptional landscape of *M. smegmatis* in different physiological conditions.

The implementation of two differentially treated 5'-end libraries followed by Gaussian mixture modeling analysis allowed us to simultaneously map and classify 5' ends resulting from nucleolytic

cleavage and those resulting from primary transcription with high confidence. We were able to classify 57% of the 5' ends in Dataset 1 with high confidence. In addition, we elaborated a list of medium confidence TSSs and CSs (Supplementary Tables S2-12, S2-13). These lists constitute a valuable resource for the research community.

Analysis of TSS mapping data allowed us to identify over 4,000 primary TSSs and to study the transcript features in *M. smegmatis*. The high proportion of leaderless transcripts, the lack of a consensus SD sequence in half of the leadered transcripts, and the absence of a conserved -35 consensus sequence indicate that the transcription-translation machineries are relatively robust in *M. smegmatis*. These findings are consistent with a recent study that mapped a 2,139 TSSs in *M. smegmatis* (Li et al., 2017). The apparent robustness of translation is shared with *Mtb*, where 25% of the transcripts lack a leader sequence (Cortes et al., 2013; Shell et al., 2015b). In addition, high abundances of transcripts lacking 5' UTRs have been reported in other bacteria including *Corynebacterium diphtheria*, *Leptospira interrogans*, *Borrelia burgdorferi*, and *Deinococcus deserti*, the latter having 60% leaderless transcripts (de Groot et al., 2014; Adams et al., 2017; Zhukova et al., 2017; Wittchen et al., 2018). Considering the high proportion of leaderless transcripts and the large number of leadered transcripts that lack a SD sequence (53%), it follows that an important number of transcripts are translated without canonical interactions between the mRNA and anti-Shine-Dalgarno sequence, suggesting that *M. smegmatis* has versatile mechanisms to address translation. A computational prediction showed that the presence of SD can be very variable between prokaryotes, ranging from 11% in *Mycoplasma* to 91% in *Firmicutes* (Chang et al., 2006). Cortes et al. (2013) reported that the 55% of the genes transcribed with a 5' UTR lack the SD motif. The correlation of leader lengths for homologous genes in *M. smegmatis* and *M. tuberculosis* (Figure 2-3B) suggests that some genes may share additional UTR-

associated regulatory features, although further work is required to investigate the possible regulatory roles of 5' UTRs in both species.

To begin to understand the role of RNA cleavage in mycobacteria, we identified and classified over 3,000 CSs throughout the *M. smegmatis* transcriptome, presenting the first report of an RNA cleavage map in mycobacteria. The most striking feature of the CSs was a cytidine in the +1 position, which was true in over 90% of the cases. While the RNases involved in global RNA decay in mycobacteria have not been yet elucidated, some studies have implicated RNase E as a major player in RNA processing and decay (Kovacs et al., 2005; Zeller et al., 2007; Csanadi et al., 2009; Taverniti et al., 2011), given its central role in other bacteria such as *E. coli* and its essentiality for survival in both *M. smegmatis* and Mtb (Sasseti et al., 2003; Sasseti and Rubin, 2003; Griffin et al., 2011; Taverniti et al., 2011; DeJesus et al., 2017). It is therefore possible that mycobacterial RNase E, or other endonucleases with dominant roles, favor cytidine in the +1 position. Interestingly, the sequence context of cleavage found here is different from that described for *E. coli*, for which the consensus sequence is (A/G)N↓AU (Mackie, 2013) or *S. enterica*, in which a marked preference for uridine at the +2 position and AU-rich sequences are important for RNase E cleavage (Chao et al., 2017).

RNA cleavage is required for maturation of some mRNAs (Li and Deutscher, 1996; Condon et al., 2001; Gutsell and Jain, 2010; Moores et al., 2017). Therefore, the observation that CSs are enriched in 5' UTRs and intergenic regions suggests that processing may play roles in RNA maturation, stability, and translation for some transcripts in *M. smegmatis*. A high abundance of processing sites around the translation start site was also observed in *P. aeruginosa* and *S. enterica* in transcriptome-wide studies (Chao et al., 2017; Gill et al., 2018), suggesting that 5'

UTR cleavage may be a widespread post-transcriptional mechanism for modulating gene expression in bacteria.

Regulation of RNA decay and processing plays a crucial role in adaptation to environmental changes. We present evidence showing that RNA cleavage is markedly reduced in conditions that result in growth cessation. It was previously demonstrated that in low oxygen concentrations mycobacteria reduce their RNA levels (Ignatov et al., 2015) and mRNA half-life is strikingly increased (Rustad et al., 2013), likely as a mechanism to maintain adequate transcript levels in the cell without the energy expenditures that continuous transcription would require. While several traits are involved in the regulation of transcript abundance and stability, the observation that cleavage events are pronouncedly reduced in these conditions pinpoint this mechanism as a potential way to control RNA stability under stress. In agreement with this hypothesis, RNase E was modestly but significantly decreased at the transcript level in early and late hypoxia (fold change = 0.63 and 0.56, respectively,  $p$ -value adjusted  $<0.05$ ), suggesting that reducing the RNase E abundance in the cell may be a strategy to increase transcript half-life. Further study is needed to better understand the relationship between transcript processing and RNA decay in normoxic growth as well as stress conditions.

Hypoxic stress conditions were also characterized by major changes in the TSSs. 5'-end-mapping libraries revealed that over 300 TSSs varied substantially when cultures were limited in oxygen. We found that 56 transcripts triggered in hypoxia contain the SigF promoter binding motif, indicating that this sigma factor plays a substantial role in the *M. smegmatis* hypoxia response. While previous work revealed increased expression of SigF itself in hypoxia in *Mtb* (Galagan et al., 2013; Iona et al., 2016; Yang et al., 2018), this is the first report demonstrating the direct impact of SigF on specific promoters in hypoxic conditions in mycobacteria. Further work is

needed to better understand the functional consequences of SigF activation in both organisms in response to hypoxia.

The work reported here represents the most complete *M. smegmatis* transcriptome map to date. We have almost doubled the number of mapped TSSs and report the presence and locations of internal and antisense TSSs as well as primary TSSs. Comparison of TSSs used in log phase and hypoxia revealed a signature of SigF activity in hypoxia, which has not been previously reported. We report the presence of locations of thousands of RNA cleavage sites, which reveals for the first time the consensus sequence recognized by the major mycobacterial RNase(s) that produces monophosphorylated 5' ends. Cleavage sites are enriched in 5' UTRs and intergenic regions, suggesting that these locations are more accessible to RNases and/or subject to regulation by RNA processing. Cleaved RNAs are relatively less abundant in hypoxic *M. smegmatis* cultures, suggesting that RNase activity is reduced as part of the phenotypic transition into hypoxia-induced growth arrest.

## **Materials and Methods**

### **Strains and Growth Conditions Used in This Study**

*Mycobacterium smegmatis* strain mc<sup>2</sup>155 was grown in Middlebrook 7H9 supplemented with ADC (Albumin Dextrose Catalase, final concentrations 5 g/L bovine serum albumin fraction V, 2 g/L dextrose, 0.85 g/L sodium chloride, and 3 mg/L catalase), 0.2% glycerol and 0.05% Tween 80. For the exponential phase experiment (Dataset 1), 50 ml conical tubes containing 5 ml of 7H9 were inoculated with *M. smegmatis* to have an initial OD = 0.01. Cultures were grown at 37 °C and 200 rpm. Once cultures reached an OD of 0.7–0.8, they were frozen in liquid nitrogen and stored at –80°C until RNA purification. For hypoxia experiments (Dataset 2), a protocol similar to the Wayne model (Wayne and Hayes, 1996) was implemented. Briefly, 60 ml serum bottles

(Wheaton, product number 223746, actual volume to top of rim 73 ml) were inoculated with 36.5 ml of *M. smegmatis* culture with an initial OD = 0.01. The bottles were sealed with rubber caps (Wheaton, W224100-181 Stopper, 20 mm) and aluminum caps (Wheaton, 20 mm aluminum seal) and cultures were grown at 37°C and 125 rpm to generate hypoxic conditions. Samples were taken at an early stage of oxygen depletion when growth had slowed but not completely stopped (15 h) and at a later stage when a methylene blue indicator dye was fully decolorized and growth had ceased (24 h). These time points were experimentally determined according to growth curve experiments (see Supplementary Figure S2-1). 15 ml of each culture were sampled and frozen immediately in liquid nitrogen until RNA extraction.

### **RNA Extraction**

RNA was extracted as follows: frozen cultures stored at -80°C were thawed on ice and centrifuged at 4,000 rpm for 5 min at 4°C. The pellets were resuspended in 1 ml Trizol (Life Technologies) and placed in tubes containing Lysing Matrix B (MP Bio). Cells were lysed by bead-beating twice for 40 s at 9 m/sec in a FastPrep 5G instrument (MP Bio). 300 µl chloroform was added and samples were centrifuged for 15 min at 4,000 rpm at 4°C. The aqueous phase was collected, and RNA was purified using Direct-Zol RNA miniprep kit (Zymo) according to the manufacturer's instructions. Samples were then treated with DNase Turbo (Ambion) for 1 h and purified with an RNA Clean & Concentrator-25 kit (Zymo) according to the manufacturer's instructions. RNA integrity was checked on 1% agarose gels and concentrations were determined using a Nanodrop instrument. Prior to library construction, 5 µg RNA was used for rRNA depletion using Ribo-Zero rRNA Removal Kit (Illumina) according to the manufacturer's instructions.

## Construction of 5'-End-Mapping Libraries

For CS categorization in Figure 2-4D, we established stringent criteria in order to determine the frequency of CSs in each location category relative to the amount of the genome comprising that location category. For 3' UTR regions, we considered only CSs that were located between 2 convergent genes. To assess frequency relative to the whole genome, we considered the sum of all regions located between two convergent genes. For 5' UTRs we considered all CSs located between 2 divergent genes, and the sum of all leader lengths for genes having a pTSS whose upstream gene is in the opposite strand (divergent) determined in this study was used for assessing relative frequency. For 5' ends corresponding to cleavages between co-transcribed genes we used the operon structures determined in this study, and the sum of all their intergenic regions was used for assessing relative frequency. Finally, for CSs located within coding sequences all genes were considered, as all of them produced reads in the expression libraries. The sum of all coding sequences in NC\_008596 genome was used for assessing relative frequency, after subtracting overlapping regions to avoid redundancy.

## 5' RACE (Rapid Amplification of cDNA Ends)

For validation of TSSs and CSs, RNA samples from *M. smegmatis* were split in two and treated with or without RPPH (NEB) in order to remove the native 5' triphosphates of primary transcripts or not, respectively. Then, an adapter oligo SSS1016 (CTGGAGCACGAGGACACTGACATGGACTGAAGGAGTrArGrArArA, where nts preceded by "r" are ribonucleotides and the rest of the oligo is composed of deoxyribonucleotides) was ligated to the RNA 5' ends using T4 RNA ligase (NEB). Prior to ligation, 8 µl of RNA sample were combined with 1 µl of 1 µg/µl adapter oligo and incubated at 65°C for 10 min. For ligation, the 9 µl of RNA-oligo mix were combined with: 10 µl 50% PEG8000, 3 µl 10X ligase buffer, 3 µl 10 mM ATP, 3 µl DMSO, 1 µl Murine RNase inhibitor (NEB), and 1 µl T4 ligase (NEB). Ligation reactions were

incubated at 20°C overnight and then cleaned using RNA Clean and Concentrator 25 kit (Zymo). Both RPPH- treated and mock-treated samples were used for cDNA synthesis. Reactions in absence of reverse transcriptase were performed to control for genomic DNA contamination. For amplification of specific 5' ends, PCR was done using a forward primer SSS1017 binding to the adapter oligo (CTGGAGCACGAGGACACTGA) and a reverse (specific) primer binding near the predicted 5' end (see Supplementary Table S2-1). For PCRs, a touchdown protocol in which the annealing temperature was reduced 1°C every cycle was performed as follows: (i) initial step of DNA denaturation at 95°C for 5 min, (ii) 17 cycles of 95°C for 30 s, 72–55°C (touchdown) for 20 s and 68°C for 25 s, (iii) 20 cycles of 95°C for 30 s, 55°C for 20 s and 68°C for 25 s and (iv) a final elongation step at 68°C for 5 min. Each amplified fragment was sequenced using the specific primer. A TSS or CS was validated if (i) the 5' end position coincided with that mapped the 5' end libraries and (ii) the PCR product was more abundant in the RPPH than in the no RPPH treatment (TSS) or the PCR product was equally abundant in the RPPH and in the no RPPH treatment (CS). For validation of the MSMEG\_0063 promoter, an *M. smegmatis* mutant strain lacking the region comprising the genes MSMEG\_0062-MSMEG\_0066 was transformed with either of the 3 following constructs: (i) Wt promoter, which has the gene MSMEG\_0063 with the native predicted promoter region and the downstream genes MSMEG\_0064-MSMEG\_0066, (ii)  $\Delta$ promoter, in which the predicted promoter region for MSMEG\_0063 was deleted, and (iii) Mutated promoter, in which two point mutations were introduced in the predicted -10 region of the MSMEG\_0063 promoter. These constructs were inserted in the L5 site of the *M. smegmatis* genome.

## **DATA AVAILABILITY**

All next-generation sequencing data are available in raw and processed forms on the GEO site, accession number GSE128412.

## FUNDING

This work was supported by NSF CAREER award 1652756 to SS and by institutional startup funds from WPI.

## ACKNOWLEDGMENTS

We are grateful to Zheyang Wu and Thomas Ierger for helpful advice on the mixture modeling and Michael Chase for helpful advice on other aspects of our data analysis pipeline. We thank Todd Gray and Keith Derbyshire for the kind gift of the *msmeg\_0062*-*msmeg\_0066* deletion strain used for promoter validation. We thank members of the Shell lab for technical assistance and helpful discussions. All Illumina sequencing was performed by the UMass Medical School Deep Sequencing Core.

## References

- Adams P. P., Flores Avile C., Popitsch N., Bilusic I., Schroeder R., Lybecker M., et al. (2017). In vivo expression technology and 5' end mapping of the *Borrelia burgdorferi* transcriptome identify novel RNAs expressed during mammalian infection. *Nucleic Acids Res.* 45 775–792. 10.1093/nar/gkw1180
- Albrecht M., Sharma C. M., Reinhardt R., Vogel J., Rudel T. (2009). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res.* 38 868–877. 10.1093/nar/gkp1032
- Andre G., Even S., Putzer H., Burguiere P., Croux C., Danchin A., et al. (2008). S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* 36 5955–5969. 10.1093/nar/gkn601
- Arraiano C. M., Andrade J. M., Domingues S., Guinote I. B., Malecki M., Matos R. G., et al. (2010). The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol. Rev.* 34 883–923. 10.1111/j.1574-6976.2010.00242.x
- Bagchi G., Das T. K., Tyagi J. S. (2002). Molecular analysis of the dormancy response in *Mycobacterium smegmatis*: expression analysis of genes encoding the DevR–DevS two-component system, Rv3134c and chaperone  $\alpha$ -crystallin homologues. *FEMS Microbiol. Lett.* 211 231–237.
- Bailey T. L., Johnson J., Grant C. E., Noble W. S. (2015). The MEME Suite. *Nucleic Acids Res.* 43 W39–W49. 10.1093/nar/gkv416
- Bailey T. L., Machanick P. (2012). Inferring direct DNA binding from CHIP-seq. *Nucleic Acids Res.* 40:e128. 10.1093/nar/gks433

- Beaucher J., Rodrigue S., Jacques P. E., Smith I., Brzezinski R., Gaudreau L. (2002). Novel *Mycobacterium tuberculosis* anti-sigma factor antagonists control sigmaF activity by distinct mechanisms. *Mol. Microbiol.* 45 1527–1540. 10.1046/j.1365-2958.2002.03135.x
- Beaume M., Hernandez D., Farinelli L., Deluen C., Linder P., Gaspin C., et al. (2010). Cartography of methicillin-resistant *S. aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *PLoS One* 5:e10725. 10.1371/journal.pone.0010725
- Benaglia T., Chauveau D., Hunter D. R., Young D. (2009). mixtools: an R package for analyzing finite mixture models. *J. Stat. Softw.* 32 1–29. 10.18637/jss.v032.i06
- Berger P., Knödler M., Förstner K. U., Berger M., Bertling C., Sharma C. M., et al. (2016). The primary transcriptome of the *Escherichia coli* O104: H4 pAA plasmid and novel insights into its virulence gene expression and regulation. *Sci. Rep.* 6:35307. 10.1038/srep35307
- Berney M., Greening C., Conrad R., Jacobs W. R., Jr., Cook G. M. (2014). An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc. Natl. Acad. Sci. U.S.A.* 111 11479–11484. 10.1073/pnas.1407034111
- Chang B., Halgamuge S., Tang S. L. (2006). Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene* 373 90–99. 10.1016/j.gene.2006.01.033
- Chao Y., Li L., Girodat D., Förstner K. U., Said N., Corcoran C., et al. (2017). In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA Pathways. *Mol. Cell* 65 39–51. 10.1016/j.molcel.2016.11.002
- Condon C., Brechemier-Baey D., Beltchev B., Grunberg-Manago M., Putzer H. (2001). Identification of the gene encoding the 5S ribosomal RNA maturase in *Bacillus subtilis*: mature 5S rRNA is dispensable for ribosome function. *RNA* 7 242–253. 10.1017/S1355838201002163
- Coros A., Callahan B., Battaglioli E., Derbyshire K. M. (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol. Microbiol.* 69 794–808. 10.1111/j.1365-2958.2008.06299.x
- Cortes T., Schubert O. T., Rose G., Arnvig K. B., Comas I., Aebersold R., et al. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* 5 1121–1131. 10.1016/j.celrep.2013.10.031
- Crooks G. E., Hon G., Chandonia J. M., Brenner S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14 1188–1190. 10.1101/gr.849004
- Csanadi A., Faludi I., Miczak A. (2009). MSMEG\_4626 ribonuclease from *Mycobacterium smegmatis*. *Mol. Biol. Rep.* 36 2341–2344. 10.1007/s11033-009-9454-1
- Čuklina J., Hahn J., Imakaev M., Omasits U., Förstner K. U., Ljubimov N., et al. (2016). Genome-wide transcription start site mapping of *Bradyrhizobium japonicum* grown free-living or in symbiosis—a rich resource to identify new transcripts, proteins and to study gene regulation. *BMC Genomics* 17:302. 10.1186/s12864-016-2602-9
- D'arrigo I., Bojanovič K., Yang X., Holm Rau M., Long K. S. (2016). Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*. *Environ. Microbiol.* 18 3466–3481. 10.1111/1462-2920.13326
- de Groot A., Roche D., Fernandez B., Ludanyi M., Cruveiller S., Pignol D., et al. (2014). RNA sequencing and proteogenomics reveal the importance of leaderless mRNAs in the radiation-tolerant bacterium *Deinococcus deserti*. *Genome Biol. Evol.* 6 932–948. 10.1093/gbe/evu069

- DeJesus M. A., Gerrick E. R., Xu W., Park S. W., Long J. E., Boutte C. C., et al. (2017). Comprehensive essentiality analysis of the mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8:e2133-16. 10.1128/mBio.02133-16
- Dick T., Lee B. H., Murugasu-Oei B. (1998). Oxygen depletion induced dormancy in *Mycobacterium smegmatis*. *FEMS Microbiol. Lett.* 163 159–164. 10.1111/j.1574-6968.1998.tb13040.x
- Dinan A. M., Tong P., Lohan A. J., Conlon K. M., Miranda-CasoLuengo A. A., Malone K. M., et al. (2014). Relaxed selection drives a noisy noncoding transcriptome in members of the *Mycobacterium tuberculosis* complex. *MBio* 5:e1169-14. 10.1128/mBio.01169-14
- Elharar Y., Roth Z., Hermelin I., Moon A., Peretz G., Shenkerman Y., et al. (2014). Survival of mycobacteria depends on proteasome-mediated amino acid recycling under nutrient limitation. *EMBO J.* 33 1802–1814. 10.15252/embj.201387076
- Fozo E. M., Kawano M., Fontaine F., Kaya Y., Mendieta K. S., Jones K. L., et al. (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* 70 1076–1093. 10.1111/j.1365-2958.2008.06394.x
- Galagan J. E., Minch K., Peterson M., Lyubetskaya A., Azizi E., Sweet L., et al. (2013). The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* 499 178–183. 10.1038/nature12337
- Gaudion A., Dawson L., Davis E., Smollett K. (2013). Characterisation of the *Mycobacterium tuberculosis* alternative sigma factor SigG: its operon and regulon. *Tuberculosis* 93 482–491. 10.1016/j.tube.2013.05.005
- Gebhard S., Humpel A., McLellan A. D., Cook G. M. (2008). The alternative sigma factor SigF of *Mycobacterium smegmatis* is required for survival of heat shock, acidic pH and oxidative stress. *Microbiology* 154 2786–2795. 10.1099/mic.0.2008/018044-0
- Giangrossi M., Prosseda G., Tran C. N., Brandi A., Colonna B., Falconi M. (2010). A novel antisense RNA regulates at transcriptional level the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Res.* 38 3362–3375. 10.1093/nar/gkq025
- Gill E. E., Chan L. S., Winsor G. L., Dobson N., Lo R., Ho Sui S. J., et al. (2018). High-throughput detection of RNA processing in bacteria. *BMC Genomics* 19:223. 10.1186/s12864-018-4538-8
- Gomes A. L., Abeel T., Peterson M., Azizi E., Lyubetskaya A., Carvalho L., et al. (2014). Decoding ChIP-seq with a double-binding signal refines binding peaks to single-nucleotides and predicts cooperative interaction. *Genome Res.* 24 1686–1697. 10.1101/gr.161711.113
- Gray T. A., Palumbo M. J., Derbyshire K. M. (2013). Draft genome sequence of MKD8, a conjugal recipient *Mycobacterium smegmatis* strain. *Genome Announc.* 1:e0014813. 10.1128/genomeA.00148-13
- Griffin J. E., Gawronski J. D., Dejesus M. A., Ioerger T. R., Akerley B. J., Sasseti C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* 7:e1002251. 10.1371/journal.ppat.1002251
- Guell M., van Noort V., Yus E., Chen W. H., Leigh-Bell J., Michalodimitrakis K., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science* 326 1268–1271. 10.1126/science.1176951
- Gutgsell N. S., Jain C. (2010). Coordinated regulation of 23S rRNA maturation in *Escherichia coli*. *J. Bacteriol.* 192 1405–1409. 10.1128/JB.01314-09
- Haft D. H., Loftus B. J., Richardson D. L., Yang F., Eisen J. A., Paulsen I. T., et al. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* 29 41–43. 10.1093/nar/29.1.41

- Hartkoorn R. C., Sala C., Magnet S. J., Chen J. M., Pojer F., Cole S. T. (2010). Sigma factor F does not prevent rifampin inhibition of RNA polymerase or cause rifampin tolerance in *Mycobacterium tuberculosis*. *J. Bacteriol.* 192 5472–5479. 10.1128/JB.00687-10
- Hayashi J. M., Richardson K., Melzer E. S., Sandler S. J., Aldridge B. B., Siegrist M. S., et al. (2018). Stress-induced reorganization of the mycobacterial membrane domain. *MBio* 9:e1823-17. 10.1128/mBio.01823-17
- Heidrich N., Bauriedl S., Barquist L., Li L., Schoen C., Vogel J. (2017). The primary transcriptome of *Neisseria meningitidis* and its interaction with the RNA chaperone Hfq. *Nucleic Acids Res.* 45 6147–6167. 10.1093/nar/gkx168
- Heroven A., Sest M., Pisano F., Scheb-Wetzel M., Böhme K., Klein J., et al. (2012). Crp induces switching of the CsrB and CsrC RNAs in *Yersinia pseudotuberculosis* and links nutritional status to virulence. *Front. Cell. Infect. Microbiol.* 2:158. 10.3389/fcimb.2012.00158
- Holmqvist E., Wright P. R., Li L., Bischler T., Barquist L., Reinhardt R., et al. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* 35 991–1011. 10.15252/embj.201593360
- Honaker R. W., Leistikow R. L., Bartek I. L., Voskuil M. I. (2009). Unique roles of DosT and DosS in DosR regulon induction and *Mycobacterium tuberculosis* dormancy. *Infect. Immun.* 77 3258–3263. 10.1128/IAI.01449-08
- Humpel A., Gebhard S., Cook G. M., Berney M. (2010). The SigF regulon in *Mycobacterium smegmatis* reveals roles in adaptation to stationary phase, heat, and oxidative stress. *J. Bacteriol.* 192 2491–2502. 10.1128/JB.00035-10
- Ignatov D. V., Salina E. G., Fursov M. V., Skvortsov T. A., Azhikina T. L., Kaprelyants A. S. (2015). Dormant non-culturable *Mycobacterium tuberculosis* retains stable low-abundant mRNA. *BMC Genomics* 16:954. 10.1186/s12864-015-2197-6
- Innocenti N., Golumbeanu M., Fouquier d'Herouel A., Lacoux C., Bonnin R. A., Kennedy S. P., et al. (2015). Whole-genome mapping of 5' RNA ends in bacteria by tagged sequencing: a comprehensive view in *Enterococcus faecalis*. *RNA* 21 1018–1030. 10.1261/rna.048470.114
- Iona E., Pardini M., Mustazzolu A., Piccaro G., Nisini R., Fattorini L., et al. (2016). *Mycobacterium tuberculosis* gene expression at different stages of hypoxia-induced dormancy and upon resuscitation. *J. Microbiol.* 54 565–572. 10.1007/s12275-016-6150-4
- Jarmer H., Larsen T. S., Krogh A., Saxild H. H., Brunak S., Knudsen S. (2001). Sigma A recognition sites in the *Bacillus subtilis* genome. *Microbiology* 147 2417–2424. 10.1099/00221287-147-9-2417
- Kawano M., Aravind L., Storz G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* 64 738–754. 10.1111/j.1365-2958.2007.05688.x
- Kovacs L., Csanadi A., Megyeri K., Kaberdin V. R., Miczak A. (2005). Mycobacterial RNase E-associated proteins. *Microbiol. Immunol.* 49 1003–1007. 10.1111/j.1348-0421.2005.tb03697.x
- Kulesekara H., Lee V., Brenic A., Liberati N., Urbach J., Miyata S., et al. (2006). Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. *Proc. Natl. Acad. Sci. U.S.A.* 103 2839–2844. 10.1073/pnas.0511090103
- Lee J. H., Geiman D. E., Bishai W. R. (2008a). Role of stress response sigma factor SigG in *Mycobacterium tuberculosis*. *J. Bacteriol.* 190 1128–1133.

- Lee J. H., Karakousis P. C., Bishai W. R. (2008b). Roles of SigB and SigF in the *Mycobacterium tuberculosis* sigma factor network. *J. Bacteriol.* 190 699–707.
- Leistikow R. L., Morton R. A., Bartek I. L., Frimpong I., Wagner K., Voskuil M. I. (2010). The *Mycobacterium tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *J. Bacteriol.* 192 1662–1670. 10.1128/JB.00926-09
- Lewis D. E., Adhya S. (2004). Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol. Microbiol.* 54 692–701. 10.1111/j.1365-2958.2004.04318.x
- Li H., Durbin R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25 1754–1760. 10.1093/bioinformatics/btp324
- Li X., Mei H., Chen F., Tang Q., Yu Z., Cao X., et al. (2017). Transcriptome landscape of *Mycobacterium smegmatis*. *Front. Microbiol.* 8:2505. 10.3389/fmicb.2017.02505
- Li Z., Deutscher M. P. (1996). Maturation pathways for *E. coli* tRNA precursors: a random multienzyme process in vivo. *Cell* 86 503–512. 10.1016/S0092-8674(00)80123-3
- Lloréns-Rico V., Cano J., Kamminga T., Gil R., Latorre A., Chen W. H. (2016). Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv.* 2:e1501363. 10.1126/sciadv.1501363
- Love M. I., Huber W., Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. 10.1186/s13059-014-0550-8
- Lun D. S., Sherrid A., Weiner B., Sherman D. R., Galagan J. E. (2009). A blind deconvolution approach to high-resolution mapping of transcription factor binding sites from ChIP-seq data. *Genome Biol.* 10:R142. 10.1186/gb-2009-10-12-r142
- Mackie G. A. (2013). RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* 11 45–57. 10.1038/nrmicro2930
- Mendoza-Vargas A., Olvera L., Olvera M., Grande R., Vega-Alvarado L., Taboada B., et al. (2009). Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* 4:e7526. 10.1371/journal.pone.0007526
- Michele T. M., Ko C., Bishai W. R. (1999). Exposure to antibiotics induces expression of the *Mycobacterium tuberculosis* sigF gene: implications for chemotherapy against mycobacterial persistors. *Antimicrob. Agents Chemother.* 43 218–225. 10.1128/AAC.43.2.218
- Milano A., Forti F., Sala C., Riccardi G., Ghisotti D. (2001). Transcriptional regulation of furA and katG upon oxidative stress in *Mycobacterium smegmatis*. *J. Bacteriol.* 183 6801–6806. 10.1128/JB.183.23.6801-6806.2001
- Mitschke J., Georg J., Scholz I., Sharma C. M., Dienst D., Bantscheff J., et al. (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* 108 2124–2129. 10.1073/pnas.1015154108
- Moores A., Riesco A. B., Schwenk S., Arnvig K. B. (2017). Expression, maturation and turnover of DrrS, an unusually stable, DosR regulated small RNA in *Mycobacterium tuberculosis*. *PLoS One* 12:e0174079. 10.1371/journal.pone.0174079
- Morita T., Maki K., Aiba H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* 19 2176–2186. 10.1101/gad.1330405

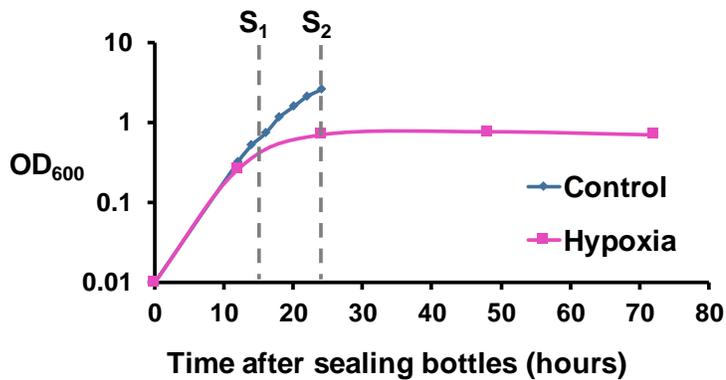
- Mraheil M. A., Billion A., Mohamed W., Mukherjee K., Kuenne C., Pischmarov J., et al. (2011). The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. *Nucleic Acids Res.* 39 4235–4248. 10.1093/nar/gkr033
- Newton-Foot M., Gey van Pittius N. C. (2013). The complex architecture of mycobacterial promoters. *Tuberculosis* 93 60–74. 10.1016/j.tube.2012.08.003
- Novichkov P. S., Kazakov A. E., Ravcheev D. A., Leyn S. A., Kovaleva G. Y., Sutormin R. A., et al. (2013). RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 14:745. 10.1186/1471-2164-14-745
- O'Toole R., Smeulders M. J., Blokpoel M. C., Kay E. J., Loughheed K., Williams H. D. (2003). A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in *Mycobacterium smegmatis*. *J. Bacteriol.* 185 1543–1554. 10.1128/JB.185.5.1543-1554.2003
- Paletta J. L., Ohman D. E. (2012). Evidence for two promoters internal to the alginate biosynthesis operon in *Pseudomonas aeruginosa*. *Curr. Microbiol.* 65 770–775. 10.1007/s00284-012-0228-y
- Park H. D., Guinn K. M., Harrell M. I., Liao R., Voskuil M. I., Tompa M., et al. (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol. Microbiol.* 48 833–843. 10.1046/j.1365-2958.2003.03474.x
- Pecsi I., Hards K., Ekanayaka N., Berney M., Hartman T., Jacobs W. R., et al. (2014). Essentiality of succinate dehydrogenase in *Mycobacterium smegmatis* and its role in the generation of the membrane potential under hypoxia. *MBio* 5:e1093-14. 10.1128/mBio.01093-14
- Potgieter M. G., Nakedi K. C., Ambler J. M., Nel A. J., Garnett S., Soares N. C., et al. (2016). Proteogenomic analysis of *Mycobacterium smegmatis* using high resolution mass spectrometry. *Front. Microbiol.* 7:427. 10.3389/fmicb.2016.00427
- Prasanna A. N., Mehra S. (2013). Comparative phylogenomics of pathogenic and non-pathogenic mycobacterium. *PLoS One* 8:e71248. 10.1371/journal.pone.0071248
- Raju R. M., Unnikrishnan M., Rubin D. H., Krishnamoorthy V., Kandror O., Akopian T. N., et al. (2012). *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathog.* 8:e1002511. 10.1371/journal.ppat.1002511
- Ramachandran V. K., Shearer N., Thompson A. (2014). The primary transcriptome of *Salmonella enterica* serovar Typhimurium and its dependence on ppGpp during late stationary phase. *PLoS One* 9:e92690. 10.1371/journal.pone.0092690
- Raman S., Hazra R., Dascher C. C., Husson R. N. (2004). Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. *J. Bacteriol.* 186 6605–6616. 10.1128/JB.186.19.6605-6616.2004
- Raman S., Song T., Puyang X., Bardarov S., Jacobs W. R., Jr., Husson R. N. (2001). The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J. Bacteriol.* 183 6119–6125. 10.1128/JB.183.20.6119-6125.2001
- Rigel N. W., Gibbons H. S., McCann J. R., McDonough J. A., Kurtz S., Braunstein M. (2009). The accessory SecA2 system of mycobacteria requires ATP binding and the canonical SecA1. *J. Biol. Chem.* 284 9927–9936. 10.1074/jbc.M900325200
- Roberts D. M., Liao R. P., Wisedchaisri G., Hol W. G., Sherman D. R. (2004). Two sensor kinases contribute to the hypoxic response of *Mycobacterium tuberculosis*. *J. Biol. Chem.* 279 23082–23087. 10.1074/jbc.M401230200

- Robson J., McKenzie J. L., Cursons R., Cook G. M., Arcus V. L. (2009). The vapBC operon from *Mycobacterium smegmatis* is an autoregulated toxin-antitoxin module that controls growth via inhibition of translation. *J. Mol. Biol.* 390 353–367. 10.1016/j.jmb.2009.05.006
- Rodrigue S., Brodeur J., Jacques P. E., Gervais A. L., Brzezinski R., Gaudreau L. (2007). Identification of mycobacterial sigma factor binding sites by chromatin immunoprecipitation assays. *J. Bacteriol.* 189 1505–1513. 10.1128/JB.01371-06
- Rosenkrands I., Slayden R. A., Crawford J., Aagaard C., Clifton III E., Andersen P. (2002). Hypoxic response of *Mycobacterium tuberculosis* studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J. Bacteriol.* 184 3485–3491. 10.1128/JB.184.13.3485-3491.2002
- Rustad T. R., Harrell M. I., Liao R., Sherman D. R. (2008). The enduring hypoxic response of *Mycobacterium tuberculosis*. *PLoS One* 3:e1502. 10.1371/journal.pone.0001502
- Rustad T. R., Minch K. J., Brabant W., Winkler J. K., Reiss D. J., Baliga N. S., et al. (2013). Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Res.* 41 509–517. 10.1093/nar/gks1019
- Sala C., Forti F., Magnoni F., Ghisotti D. (2008). The katG mRNA of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* is processed at its 5' end and is stabilized by both a polypurine sequence and translation initiation. *BMC Mol. Biol.* 9:33. 10.1186/1471-2199-9-33
- Sass A. M., Van Acker H., Förstner K. U., Van Nieuwerburgh F., Deforce D., Vogel J., et al. (2015). Genome-wide transcription start site profiling in biofilm-grown *Burkholderia cenocepacia* J2315. *BMC Genomics* 16:775. 10.1186/s12864-015-1993-3
- Sasseti C. M., Boyd D. H., Rubin E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48 77–84. 10.1046/j.1365-2958.2003.03425.x
- Sasseti C. M., Rubin E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. U.S.A.* 100 12989–12994. 10.1073/pnas.2134250100
- Schifano J. M., Edifor R., Sharp J. D., Ouyang M., Konkimalla A., Husson R. N., et al. (2013). Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc. Natl. Acad. Sci. U.S.A.* 110 8501–8506. 10.1073/pnas.1222031110
- Schlüter J.-P., Reinkensmeier J., Barnett M. J., Lang C., Krol E., Giegerich R., et al. (2013). Global mapping of transcription start sites and promoter motifs in the symbiotic  $\alpha$ -proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics* 14:156. 10.1186/1471-2164-14-156
- Shao W., Price M. N., Deutschbauer A. M., Romine M. F., Arkin A. P. (2014). Conservation of transcription start sites within genes across a bacterial genus. *MBio* 5:e1398-14. 10.1128/mBio.01398-14
- Sharma C. M., Hoffmann S., Darfeuille F., Reignier J., Findeiss S., Sittka A., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464 250–255. 10.1038/nature08756
- Shell S. S., Chase M. R., loerger T. R., Fortune S. M. (2015a). RNA sequencing for transcript 5'-end mapping in mycobacteria. *Methods Mol. Biol.* 1285 31–45. 10.1007/978-1-4939-2450-9\_3
- Shell S. S., Wang J., Lapierre P., Mir M., Chase M. R., Pyle M. M., et al. (2015b). Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genet.* 11:e1005641. 10.1371/journal.pgen.1005641

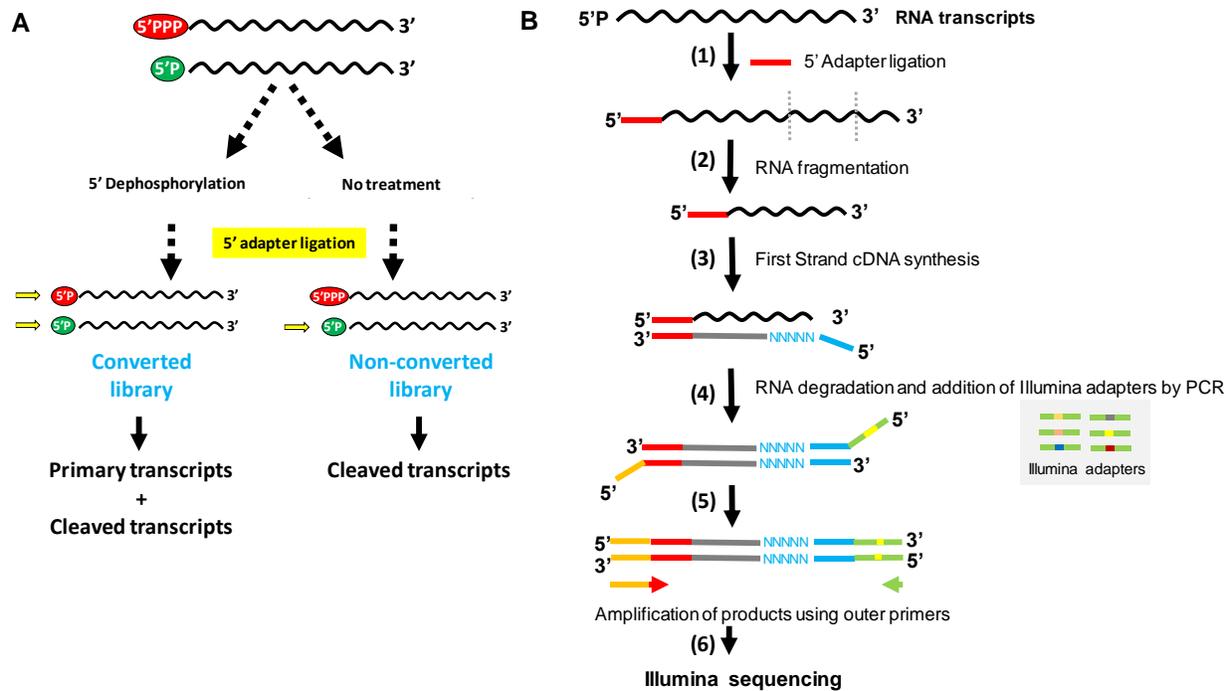
- Singh A. K., Dutta D., Singh V., Srivastava V., Biswas R. K., Singh B. N. (2015). Characterization of *Mycobacterium smegmatis* sigF mutant and its regulon: overexpression of SigF antagonist (MSMEG\_1803) in *M. smegmatis* mimics sigF mutant phenotype, loss of pigmentation, and sensitivity to oxidative stress. *Microbiologyopen* 4 896–916. 10.1002/mbo3.288
- Singh A. K., Singh B. N. (2008). Conservation of sigma F in mycobacteria and its expression in *Mycobacterium smegmatis*. *Curr. Microbiol.* 56 574–580. 10.1007/s00284-008-9126-8
- Skliarova S. A., Kreneva R. A., Perumov D. A., Mironov A. S. (2012). The characterization of internal promoters in the *Bacillus subtilis* riboflavin biosynthesis operon. *Genetika* 48 1133–1141.
- Song T., Song S. E., Raman S., Anaya M., Husson R. N. (2008). Critical role of a single position in the -35 element for promoter recognition by *Mycobacterium tuberculosis* SigE and SigH. *J. Bacteriol.* 190 2227–2230. 10.1128/JB.01642-07
- Sun R., Converse P. J., Ko C., Tyagi S., Morrison N. E., Bishai W. R. (2004). *Mycobacterium tuberculosis* ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol. Microbiol.* 52 25–38. 10.1111/j.1365-2958.2003.03958.x
- Taverniti V., Forti F., Ghisotti D., Putzer H. (2011). *Mycobacterium smegmatis* RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol. Microbiol.* 82 1260–1276. 10.1111/j.1365-2958.2011.07888.x
- Thomason M. K., Bischler T., Eisenbart S. K., Förstner K. U., Zhang A., Herbig A., et al. (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* 197 18–28. 10.1128/JB.02096-14
- Trauner A., Loughheed K. E., Bennett M. H., Hingley-Wilson S. M., Williams H. D. (2012). The dormancy regulator DosR controls ribosome stability in hypoxic mycobacteria. *J. Biol. Chem.* 287:24053. 10.1074/jbc.M112.364851
- Veyrier F., Said-Salim B., Behr M. A. (2008). Evolution of the mycobacterial SigK regulon. *J. Bacteriol.* 190 1891–1899. 10.1128/JB.01452-07
- Wayne L. G., Hayes L. G. (1996). An in vitro model for sequential study of shutdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect. Immun.* 64 2062–2069.
- Wittchen M., Busche T., Gaspar A. H., Lee J. H., Ton-That H., Kalinowski J., et al. (2018). Transcriptome sequencing of the human pathogen *Corynebacterium diphtheriae* NCTC 13129 provides detailed insights into its transcriptional landscape and into DtxR-mediated transcriptional regulation. *BMC Genomics* 19:82. 10.1186/s12864-018-4481-8
- Wu M.-L., Gengenbacher M., Dick T. (2016). Mild nutrient starvation triggers the development of a small-cell survival morphotype in mycobacteria. *Front. Microbiol.* 7:947. 10.3389/fmicb.2016.00947
- Yang H., Sha W., Liu Z., Tang T., Liu H., Qin L., et al. (2018). Lysine acetylation of DosR regulates the hypoxia response of *Mycobacterium tuberculosis*. *Emerg. Microbes Infect.* 7:34. 10.1038/s41426-018-0032-2
- Zeller M. E., Csanadi A., Miczak A., Rose T., Bizebard T., Kaberdin V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem. J.* 403 207–215. 10.1042/BJ20061530
- Zhu Y., Mao C., Ge X., Wang Z., Lu P., Zhang Y., et al. (2017). Characterization of a minimal type of promoter containing the -10 element and a guanine at the -14 or -13 position in mycobacteria. *J. Bacteriol.* 199:e385-17. 10.1128/JB.00385-17

Zhukova A., Fernandes L. G., Hugon P., Pappas C. J., Sismeiro O., Coppée J. Y., et al. (2017). Genome-wide transcriptional start site mapping and sRNA Identification in the pathogen *Leptospira interrogans*. *Front. Cell Infect. Microbiol.* 7:10. 10.3389/fcimb.2017.00010

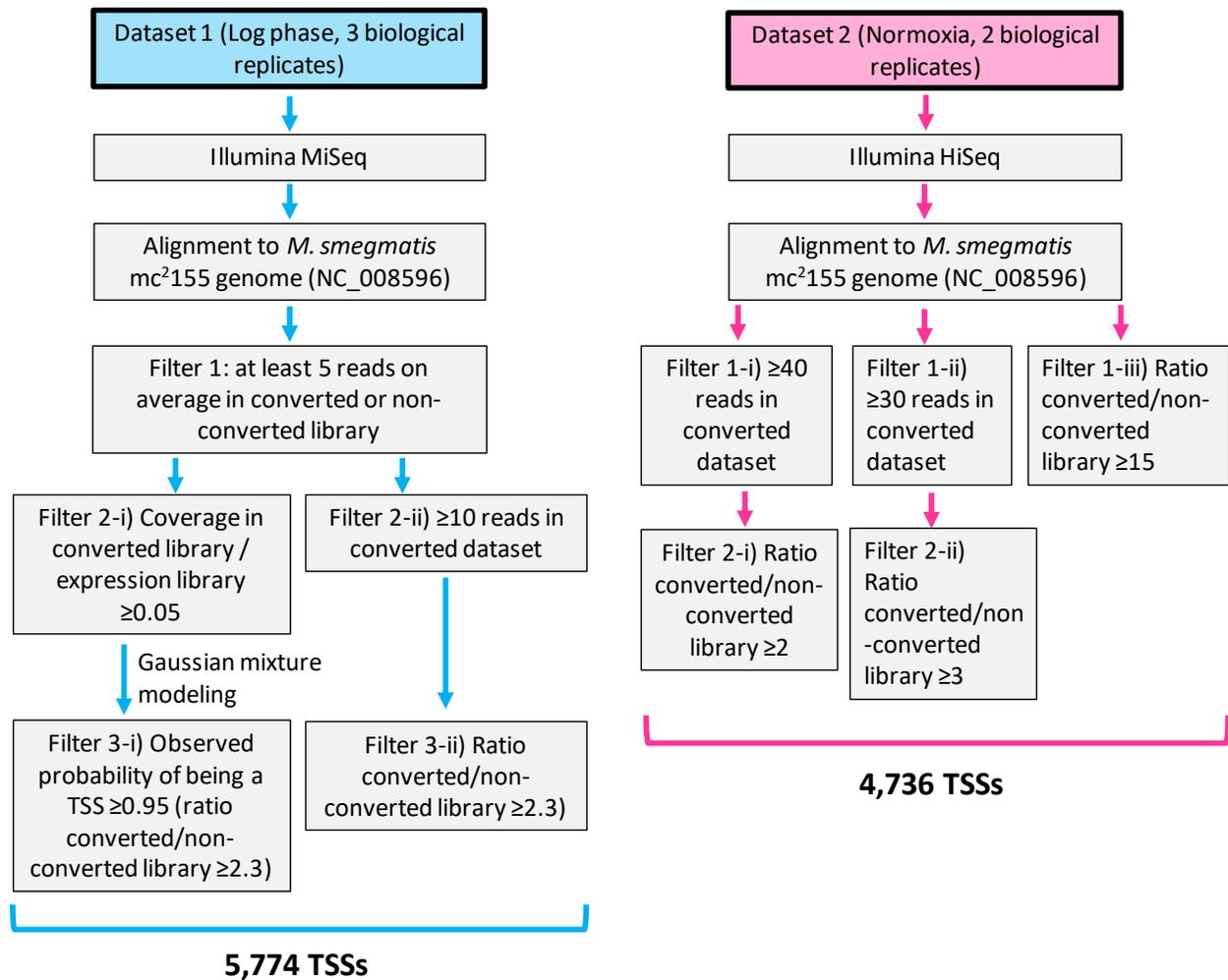
## Supplementary Figures



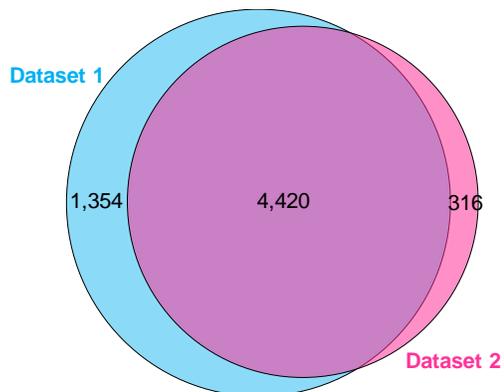
**Figure S2-1: Hypoxia model similar to the Wayne model.** Cultures were grown in sealed flasks to produce a gradual reduction in oxygen. Samples were taken at 15 (S<sub>1</sub>) and 24 (S<sub>2</sub>) hours after bottles were sealed. For control, cultures were sampled at an OD = 0.8.



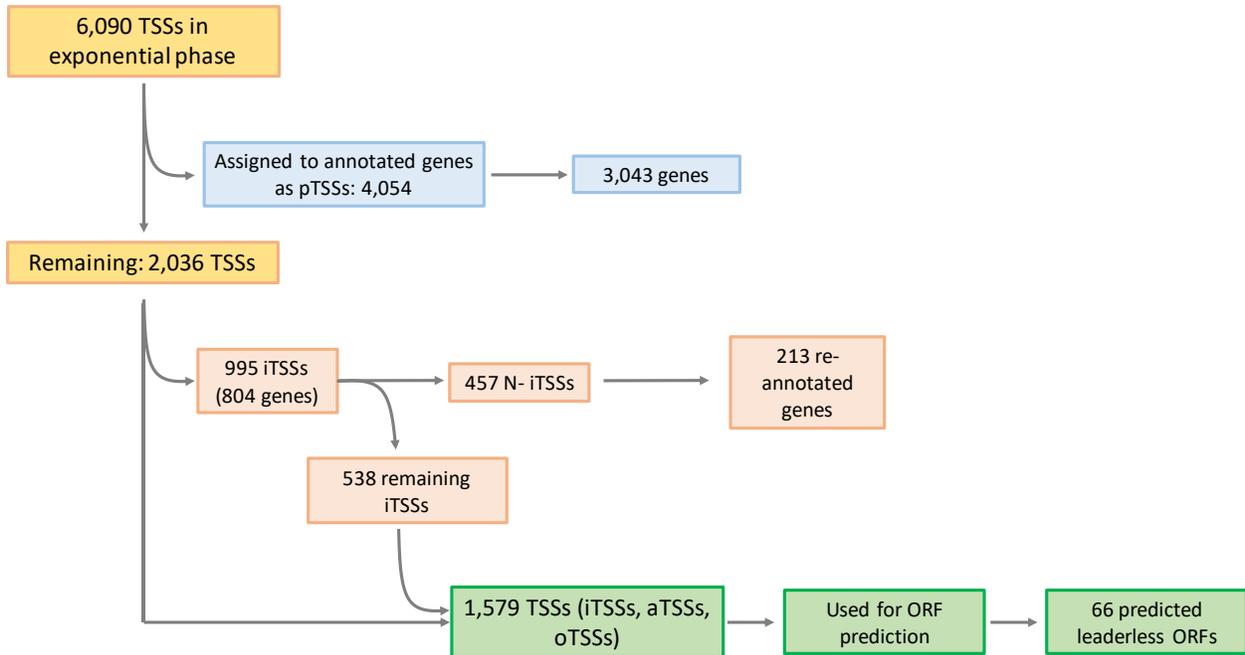
**Figure S2-2: Construction of 5'-end-directed libraries.** A) RNA samples were split in two parts and treated differentially. RNA for Library 1 (converted) was treated with RPPH to convert triphosphates in monophosphates, allowing the capture of 5' end that are primary transcripts or cleaved RNAs. RNA for Library 2 (non-converted) was mock-treated, allowing the capture of cleaved transcripts. B) Workflow of 5'-end-directed libraries. After RPPH or 5' polyphosphatase treatment, adapter SSS392 (TCCCTACACGACGCTCTTCCGAUCU) was ligated to the 5' monophosphate ends (1). Then, RNA was fragmented by heating at 85°C for 6 min (log phase experiment) or at 94°C for 11 min (hypoxia experiment) (2) and first strand cDNA synthesis was carried out using the degenerate primer SSS397 (CTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN) (3). RNA was then degraded and DNA was amplified using universal adapter sequence SSS398 (AATGATACGGCGACCACCGAGATCTACTCTTCCCTACACGACGCTCTTC) and primers bearing Illumina indexes (4). Adapter-bearing products were PCR-amplified using outer primers SSS401 (AATGATACGGCGACCACCGAGATC) and SSS402 (CAAGCAGAAGACGGCATAACGAGAT) to enrich for full-length fragments. 4 (log phase experiment) or 16 (hypoxia experiment) PCR cycles were performed (5). Finally, libraries were sequenced using Illumina technology (6).



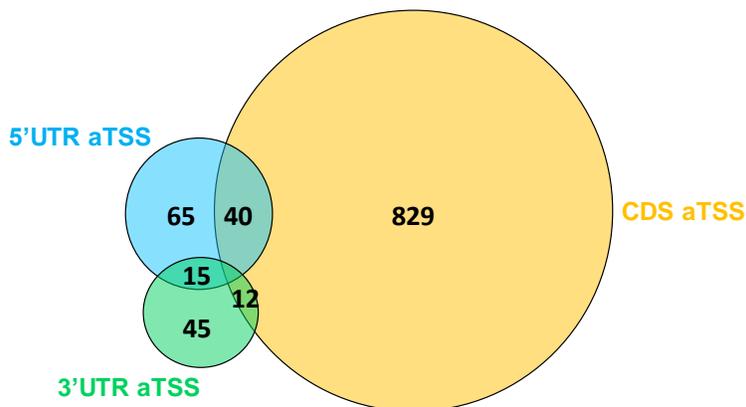
**Figure S2-3: Workflow for noise filtering and TSS prediction in the different datasets.** Normoxia refers to the control (log phase) used in the hypoxia experiment.



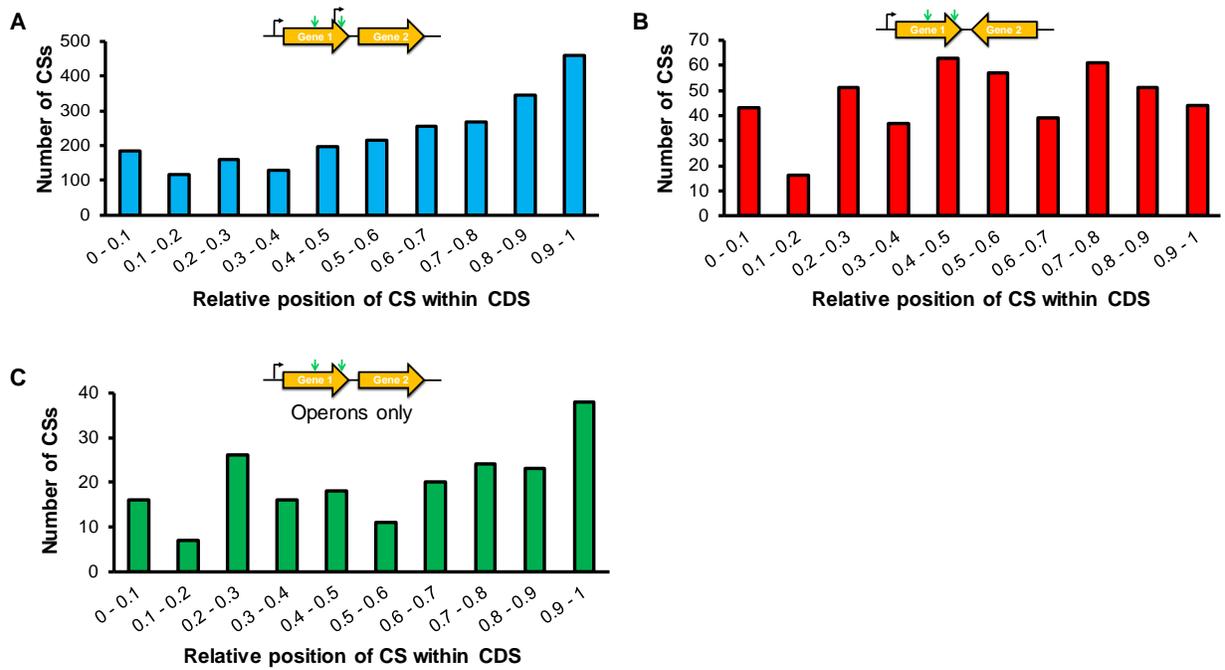
**Figure S2-4: TSSs identified in the different datasets.** Dataset 1: exponential phase (5,774 TSSs), Dataset 2: Normoxia (4,736 TSSs).



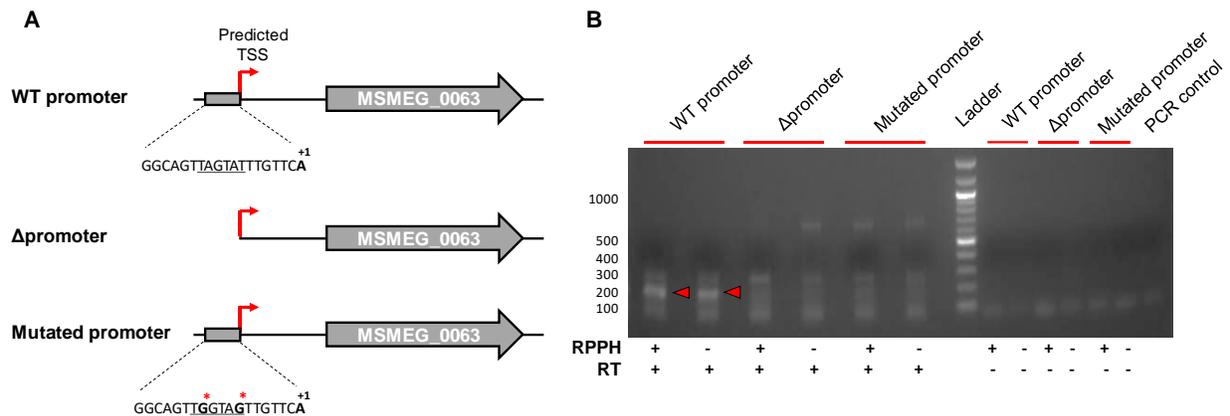
**Figure S2-5: Workflow used for TSS classification.** A complete scheme of the procedure used to classify TSSs is shown. TSSs located within 0-500 nt upstream of an annotated coding sequence were classified as pTSSs. TSSs located within annotated coding sequences were classified as iTSSs. iTSSs located within the first 25% of an annotated coding sequence were subclassified as N-iTSSs. When a gene lacked a pTSS, had an N-iTSS, and had an in-frame start codon downstream of the N-iTSS and within the first 30% of the coding sequence, the start codon of the gene was re-annotated. aTSSs (TSSs located on the antisense strand of a coding sequence, 5' UTR, or 3' UTR) and oTSSs (TSSs not belonging to any of the above-mentioned categories) were assigned as described in Figure 1D and Materials and Methods.



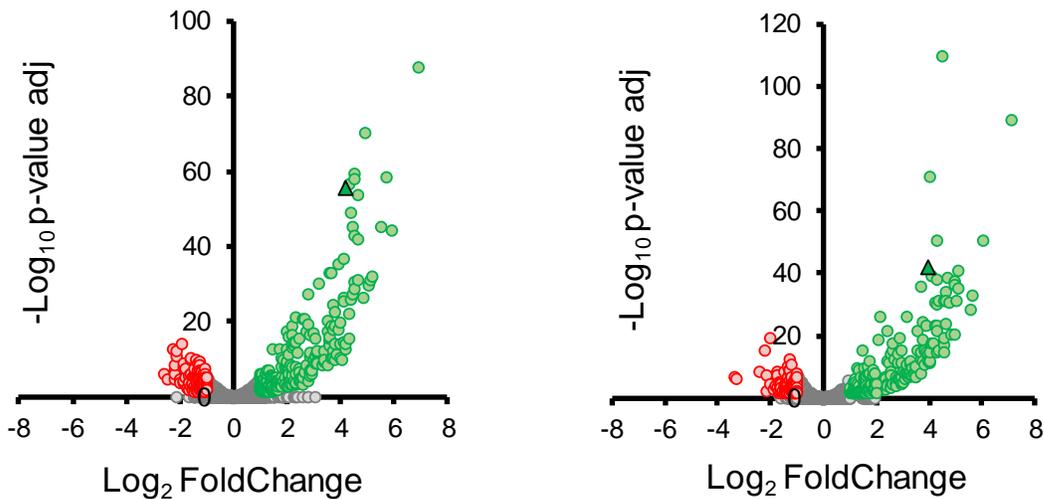
**Figure S2-6: Distribution of antisense TSSs.** The 1,006 aTSSs were classified according to their positions in 5' UTRs, 3' UTRs, and CDSs (coding sequences).



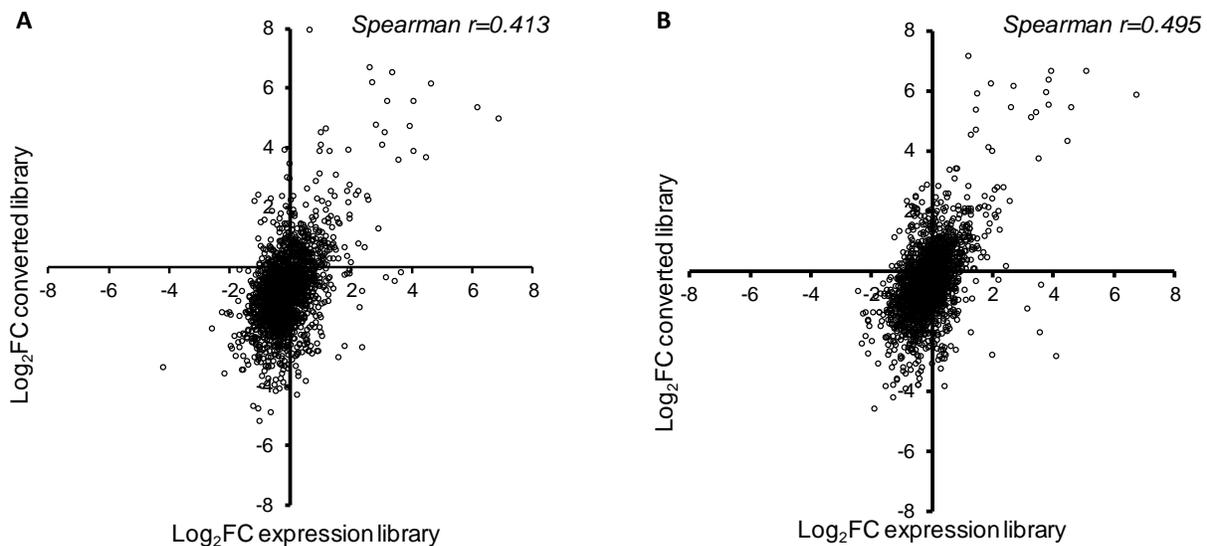
**Figure S2-7: Cleavage sites distribution within genes according to coding sequence context.** The number of cleavage sites according to the relative position in the coding sequence is represented considering A) only coding sequences whose downstream gene is in the same strand, B) only coding sequences whose downstream gene is in the opposite strand (convergent), and C) only genes having a downstream gene transcribed as an operon. The CS distribution is significantly different between graphics A and B (p-value <0.0001, Kolmogorov Smirnov D test).



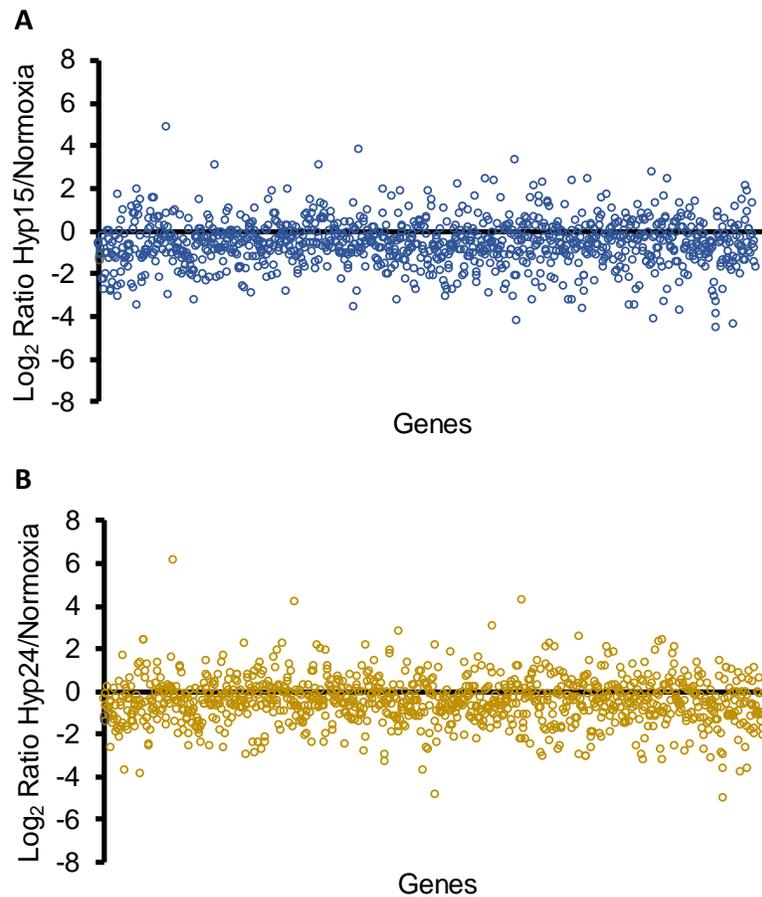
**Figure S2-8: Validation of a medium confidence pTSS.** A) Constructs used to validate the medium confidence pTSS of MSMEG\_0063 were cloned into pJEB402 plasmid and integrated in the L5 site in the genome of an *M. smegmatis* strain lacking *msmeg\_0062*-*msmeg\_0066*. The WT promoter construct has the wildtype promoter region;  $\Delta$ promoter has a deletion of the region upstream of the predicted pTSS; and mutated promoter has a replacement of two bases (red asterisks) in the -10 promoter region (underlined sequence). B) 1% agarose gel showing the 5' RACE amplification products. The red arrows indicate the band corresponding to the predicted pTSS. At the bottom is indicated whether the RNA samples were treated with pyrophosphohydrolase (RPPH) prior to adapter ligation and whether cDNA synthesis with reverse transcriptase (RT) was performed. PCR control: water.



**Figure S2-9: Gene expression levels in RNAseq expression libraries in hypoxia.** Changes in transcript levels were obtained by DEseq analysis, comparing each indicated condition to the control experiment. Genes upregulated (245 or 266 at 15 or 24 h, respectively) and downregulated 106 or 158 at 15 or 24 h, respectively) with a fold change  $\geq 2$  and a corrected p value  $\leq 0.05$  are highlighted in green and red, respectively. The triangle indicates expression of MSMEG\_5244 (dosR) gene.



**Figure S2-10: Correlation between expression data and 5' end-directed libraries data in hypoxia.** The X axis represents the Log2 of the fold change in the expression libraries from hypoxia/normoxia datasets and the Y axis represents the Log2 of the fold change in the peak height in hypoxia/normoxia 5' end-directed libraries. The analysis was done for hypoxia at 15 hours (A) and 24 hours (B). Genes having only one pTSS were used. The correlation is significant in both cases, with a p-value  $< 0.00001$ .



**Figure S2-11: Changes in RNA cleavage within coding sequences in hypoxic conditions.** The number of cleavage events within each coding sequence was compared through the different conditions. The Log<sub>2</sub> of the ratio of the number of cleavages in hypoxia/control are shown. Each dot represents a specific gene. A) Hypoxia 15 hours, B) Hypoxia 24 hours.

## Supplementary Tables

All tables can be accessed online:

<https://www.frontiersin.org/articles/10.3389/fmicb.2019.00591/full#supplementary-material>

Chapter 3 : The dominant role of RNase E in shaping the  
*Mycobacterium smegmatis* transcriptome

# The dominant role of RNase E in shaping the *Mycobacterium smegmatis* transcriptome

**Ying Zhou, Huaming Sun, Samantha R. Joubran, Alexa Davis, Joseph Dainis, Louis A. Roberts, Scarlet S. Shell**

---

## **Abstract**

*Mycobacterium tuberculosis* is the one of the leading causes of death worldwide and it can survive and adapt within the host. During this adaptation, the success of the bacteria relies on regulation of gene expression as well as macromolecule synthesis and degradation. mRNA degradation is one of these regulated processes. However, the mechanisms of such regulation remain poorly understood. In *E. coli*, RNase E is major player in mRNA degradation and has been well studied. However, much less is known about the role of RNase E in mycobacteria. Here, we used *Mycobacterium smegmatis*, a non-pathogenic model, to study the role of RNase E in mRNA degradation. As RNase E is essential in mycobacteria, we used an inducible knockdown system combined with RNA-seq to determine mRNA half-lives transcriptome-wide and investigate the effects on mRNA metabolism when RNase E is repressed. We were able to determine the mRNA half-lives in *rne* repression and a control condition with high/medium confidence for more than 4,000 transcripts. In the half-life analysis, we found a 2-fold stabilization or more for 3,624 mRNAs when we repressed *rne*, indicating a global role of RNase E in mRNA degradation in mycobacteria. A varied stabilization of mRNA among genes was observed, and we assessed potential factors that may contribute to this differential sensitivity to RNase E. mRNA abundance and 5' UTRs appear to be key factors; however, they do not fully explain the variability, and more factors need to be considered. Lastly, we mapped the RNase E cleavage sites with 5' RACE and 3' RACE *in vivo* and *in vitro* and found that RNase E cleaved at the sequence RN↓CNU, consistent with the major cleavage site motif that we previously identified in the *M. smegmatis* transcriptome. Most

of the in vivo-mappable mRNA cleavage products are therefore likely products of RNase E cleavage. Taken together, RNase E has a dominant role in mRNA metabolism in *M. smegmatis*.

---

## Introduction

Mycobacteria are a globally important group of bacteria including the pathogen *Mycobacterium tuberculosis* which kills over a million people each year (WHO, 2021) as well as numerous environmental bacteria and opportunistic pathogens. Mycobacteria are phylogenetically divergent from better-studied models such as *Escherichia coli*, and consequently, numerous aspects of their fundamental biology remain poorly understood. mRNA metabolism is clearly a critical aspect of mycobacterial biology, as regulation of gene expression facilitates adaptation to stressors both during infection and in the environment, and regulation of mRNA degradation permits energy conservation during severe stress. However, the roles and regulation of mRNA degradation enzymes remain largely undefined.

The endoribonuclease RNase E is a critical component of the bulk mRNA degradation machinery in gram-negative bacteria. In *E. coli*, RNase E cleaves single-stranded mRNAs in A/U-rich regions and interacts with other RNA degradation proteins to increase the efficiency of mRNA degradation (Babitzke & Kushner, 1991; Carpousis et al., 1994; McDowall et al., 1994; Mudd et al., 1988; Py et al., 1996; Vanzo et al., 1998). In contrast, the better-studied gram-positive bacteria such as *Bacillus subtilis* and *Staphylococcus aureus* lack RNase E completely and rely on other RNases such as RNase J and RNase Y. Mycobacteria are phylogenetically gram-positive, despite having cell envelopes that prevent gram staining. However, they encode orthologs of RNase E, and these genes are essential in both *M. tuberculosis* and the non-pathogenic model *Mycobacterium smegmatis* (Sasseti et al., 2003; Sasseti & Rubin, 2003; Taverniti et al., 2011). The essentiality of RNase E suggests it may be a critical component of the bulk mRNA degradation machinery in mycobacteria. Consistent with this, mycobacterial RNase E was shown to interact with other

RNases such as RNase J and PNPase (Plocinski et al., 2019). It was also shown to contribute to rRNA maturation (Taverniti et al., 2011).

We previously showed that the *M. smegmatis* transcriptome is shaped by endonucleolytic cleavage events that produce mRNA fragments with monophosphorylated 5' ends (Martini et al., 2019). RNase E is known to produce cleavage products with monophosphorylated 5' ends in other organisms (Mackie, 1998). Taken together with the observation that the mycobacterial cleavage sites appeared to be present preferentially in single-stranded regions, and the paucity of other candidate RNases predicted to cleave with those properties, we hypothesized that RNase E was responsible for the majority of the cleavage sites we mapped in *M. smegmatis*. However, the mycobacterial cleavage sites occurred primarily in a sequence context distinct from that reported to be cleaved by *E. coli* RNase E. Most mycobacterial mRNA cleavages occurred immediately upstream of a cytidine, with a preference for 1-2 purines immediately upstream and uridine three nts downstream of the cleavage site (RR↓CNU). A previous report tested the cleavage specificity of mycobacterial RNase E *in vitro*; however, the substrates used in that study did not include "RRCNU" (Zeller et al., 2007).

Given the clear importance of RNase E in mycobacteria and lack of information on its role, we sought to define its function in mycobacterial mRNA metabolism. We used an inducible system to interrogate the effects of knockdown of *rne*, the gene encoding RNase E, in *M. smegmatis*. We found that RNase E has a rate-limiting role in degradation of most mRNAs, and that its cleavage signature is ubiquitous across the transcriptome. We then used purified RNase E to confirm its cleavage specificity *in vitro*. Together, our results implicate RNase E as the predominant source of cleaved mRNAs in the transcriptomes of both *M. smegmatis* and *M. tuberculosis* as well as a critical mediator of bulk mRNA degradation in these organisms.

## Results

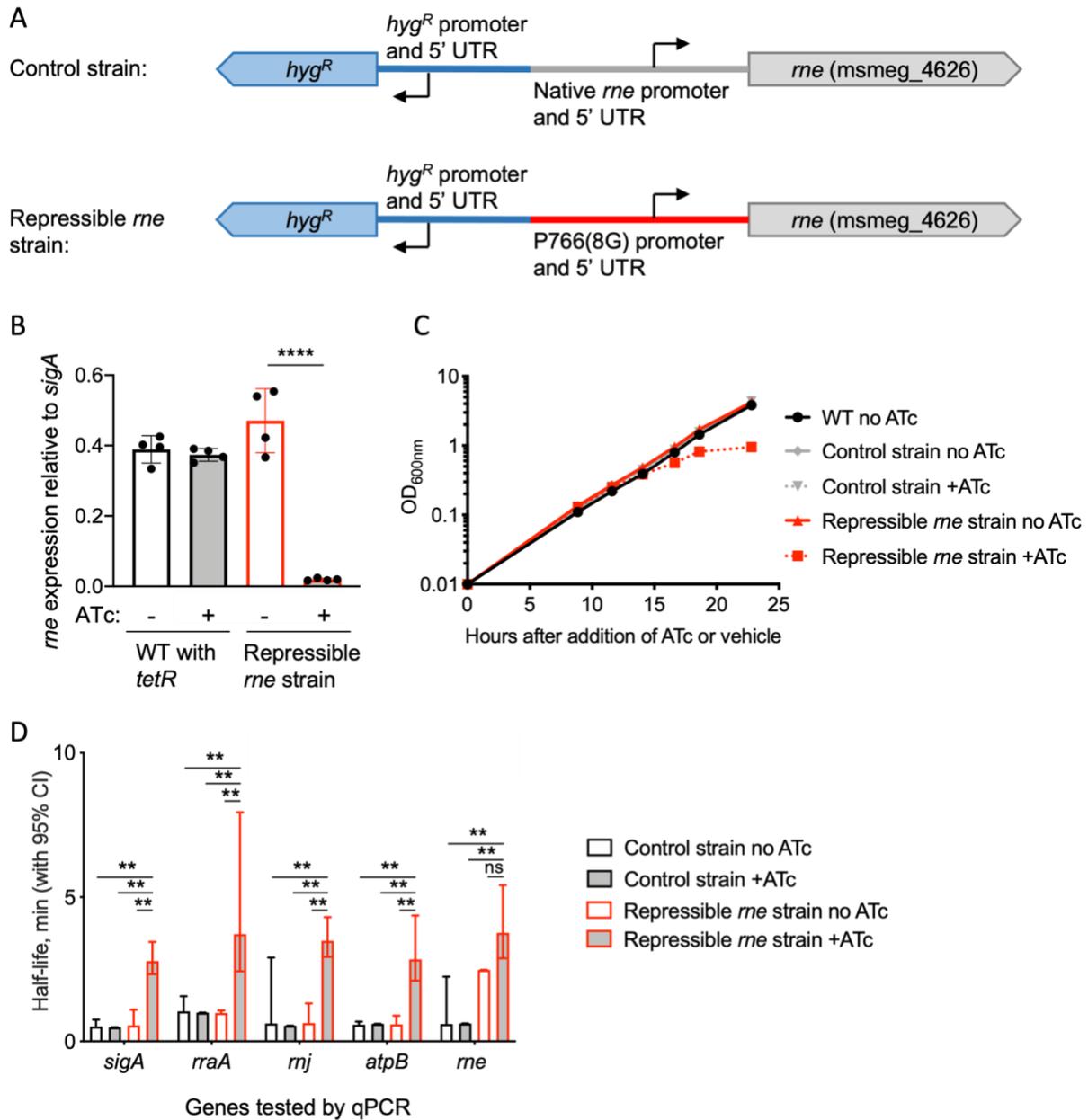
### **RNase E has a global role in *M. smegmatis* mRNA degradation**

RNase E is encoded by the gene *rne*, which was predicted to be essential in *M. tuberculosis* and shown to be essential in *M. smegmatis* (Sasseti et al., 2003; Sasseti & Rubin, 2003; Taverniti et al., 2011). To investigate the function of this enzyme in *M. smegmatis*, we therefore constructed a strain in which transcription of *rne* (msmeg\_4626) could be repressed by addition of anhydrotetracycline (ATc) (Ehrt et al., 2005) (Figure 3-1A, Table 3-1). Replacement of the native *rne* promoter and 5' UTR (Martini et al., 2019) with the P766(8G) promoter and associated 5' UTR (Johnson et al., 2020) produced a strain in which ATc caused a constitutively expressed reverse TetR to bind the promoter and repress *rne* expression (Figure 3-1B). We hereafter refer to this as the repressible *rne* strain. Consistent with the known essentiality of *rne*, growth slowed approximately 14 hours after addition of ATc and later ceased (Figure 3-1C). Construction of the repressible strain resulted in insertion of a hygromycin resistance gene upstream of *rne*. We therefore constructed an isogenic strain in which the hygromycin resistance gene was inserted upstream of the native copy of *rne* (hereafter referred to as the control strain; Figure 3-1A).

While the essentiality of *rne* could be due to its role in mRNA degradation, rRNA maturation, or both, we were specifically interested in determining the role of RNase E in mRNA metabolism. We therefore evaluated the impact of *rne* knockdown on mRNA degradation rates prior to the slowing of growth. We measured the half-lives of several mRNAs by using quantitative PCR (qPCR) to determine transcript abundance at timepoints following addition of rifampicin to block transcription initiation. The half-lives of all tested genes were lengthened upon *rne* knockdown (Figure 3-1D). To determine the generalizability of this observation, we used RNAseq to measure mRNA half-lives transcriptome-wide. RNAseq libraries were constructed from RNA extracted from triplicate cultures of each strain and condition at various timepoints after the addition of rifampicin.

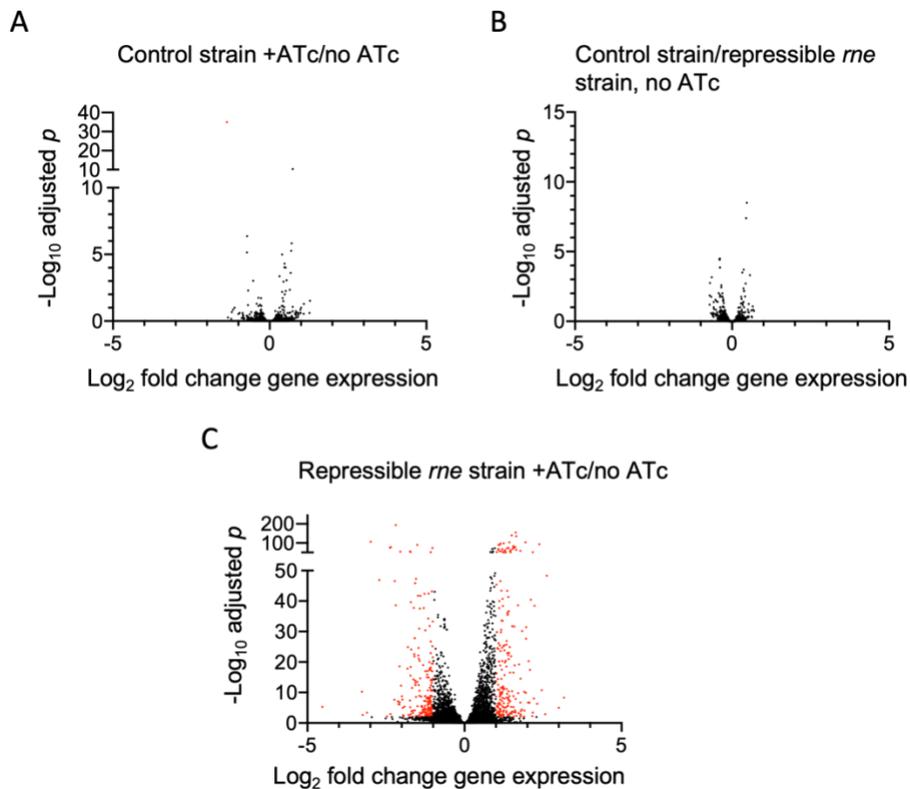
qPCR was used to establish relative values for a set of calibrator genes, and these were used to normalize the coverage values obtained from the RNAseq libraries as described in detail in the methods section. Libraries were made from the repressible *rne* strain following 8 hours of treatment with ATc (*rne* knockdown condition), the repressible *rne* strain in the absence of ATc, and the control strain harboring the native *rne* promoter in the presence and absence of ATc. The timepoint for analysis of the *rne* knockdown condition was carefully chosen to maximize our power to detect relevant phenotypes, but prior to the slowing of growth, as we expect growth changes would themselves affect mRNA stability as has been reported by us and many others (Esquerre et al., 2014; Esquerre et al., 2015; Lin et al., 2012; Nilsson et al., 1984).

To confirm that cells in the three control conditions were physiologically similar, we performed differential expression analysis on the libraries constructed from RNA harvested immediately after the addition of rifampicin to capture steady-state gene expression levels (Figure 3-2 and Table S3-1). The presence of ATc affected a small number of genes regardless of the *rne* promoter (Figure 3-2A), and *rne* itself had significantly but modestly reduced expression in the promoter replacement strain in the absence of ATc compared to the control strain (~1.4-fold, Figure 3-2B). In contrast, 428 genes were expressed at substantially and significantly different levels (fold change  $\geq 2$ , adjusted  $p < 0.01$ ) in the *rne* knockdown condition compared to the repressible *rne* strain in the absence of ATc (Figure 3-2C).

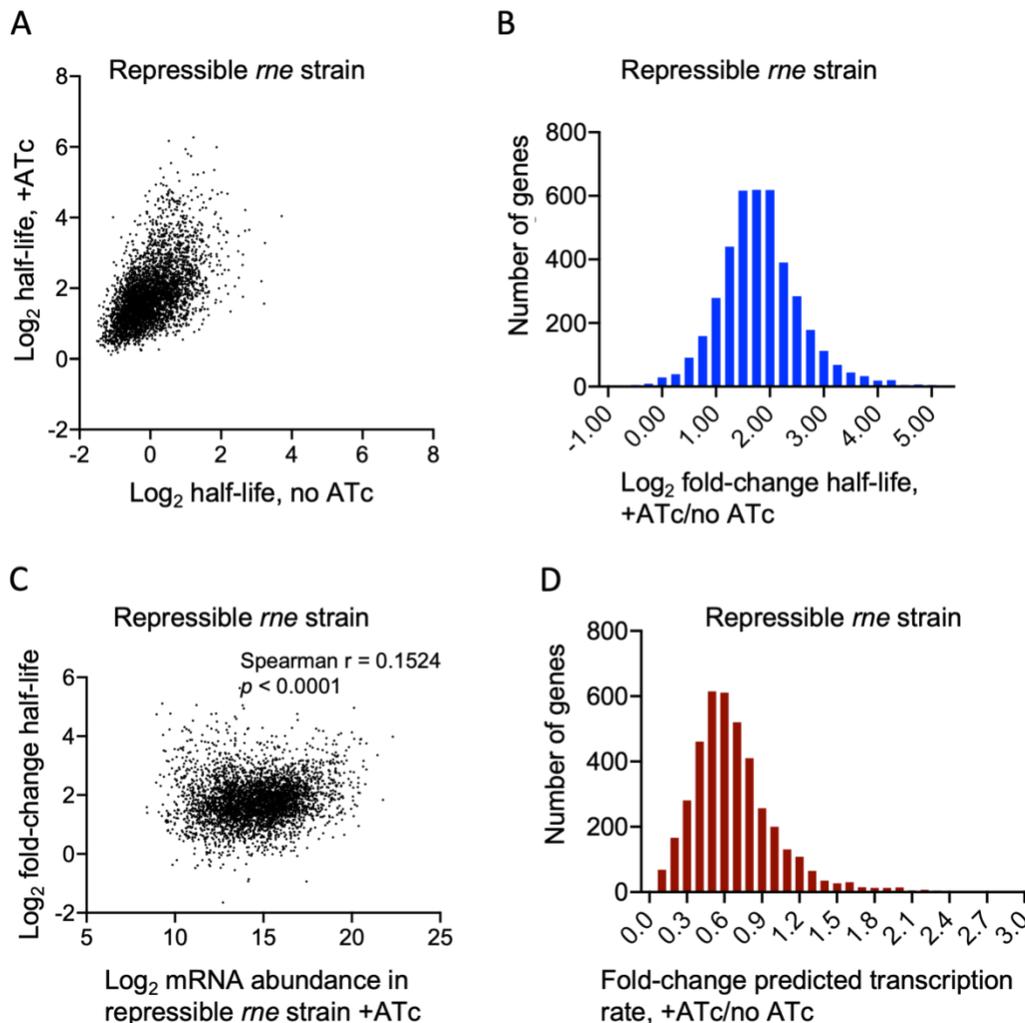


**Figure 3-1: Knockdown of *rne* expression causes growth cessation and reduced mRNA degradation rates in *M. smegmatis*.** A. Promoter replacement strategy to construct a strain in which *rne* expression is repressed by addition of ATc. B. *rne* transcript levels were reduced in the repressible *rne* strain following 3 hrs of exposure to ATc. \*\*\*\*=  $p < 0.001$ , two-tailed t test. C. Growth of the repressible *rne* strain slowed approximately 15 hours after addition of ATc. D. The half-lives of five transcripts were determined by using qPCR to measure mRNA levels at several time-points following addition of RIF to block new transcription. \*\*=  $p < 0.01$ , pair-wise comparisons by linear regression.

To identify transcripts that were direct targets of RNase E, we calculated half-lives for each gene in each condition as described in the methods section and Figure S3-1. We determined high-confidence half-lives for 1662 genes as well as medium-confidence half-lives for an additional 3964 genes in the *rne* knockdown condition. We were able to calculate high-confidence half-lives for 4,073 of these genes in the repressible *rne* strain in the absence of ATc as well. Half-lives were similar in comparisons between control conditions (Figure S3-2). In contrast, half-lives of most genes were increased in the *rne* knockdown (Figure 3-3A, B). The half-lives of 3,624 genes increased by 2-fold or more, and an additional 78 genes had no measurable degradation in the *rne* knockdown. Together, these data are consistent with RNase E playing a rate-limiting step in the degradation of at least 89% of the transcriptome.



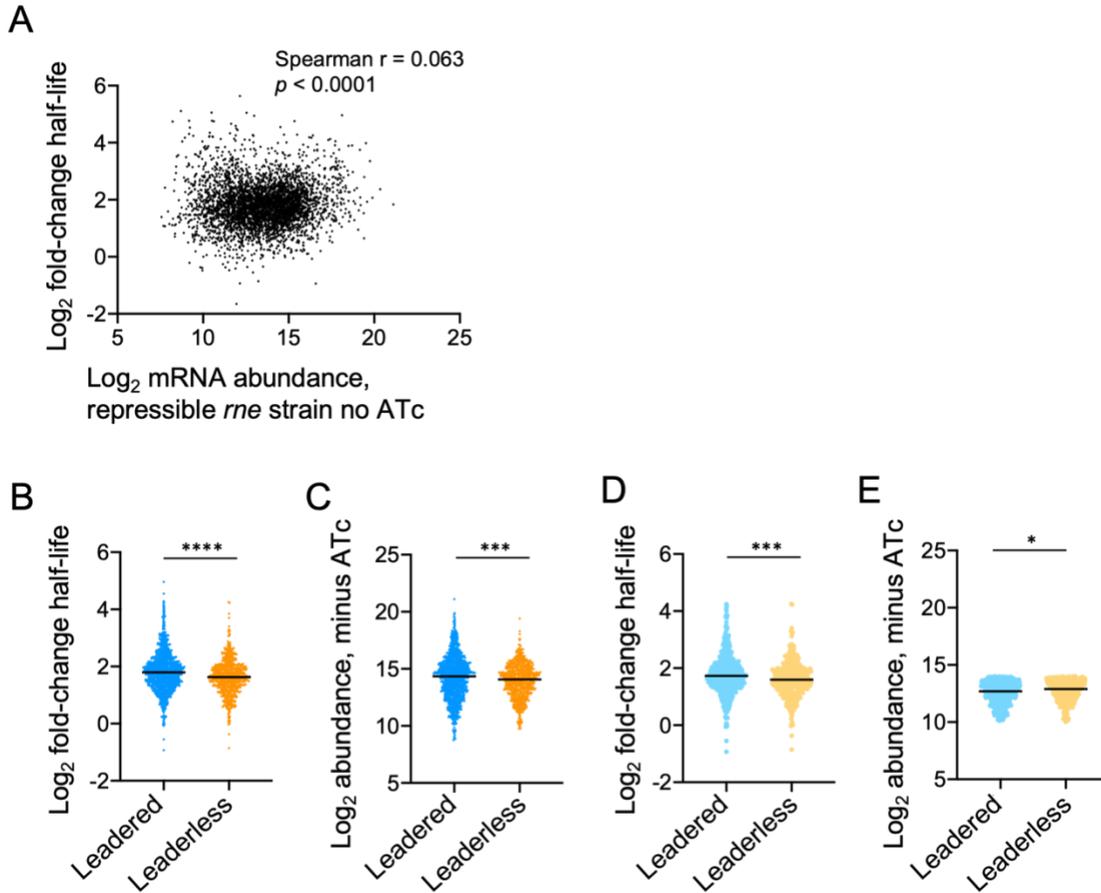
**Figure 3-2: Knockdown of *rne* causes substantial changes in *M. smegmatis* transcript abundance levels.** Differential expression analysis was performed on the indicated strains immediately after addition of rifampicin and eight hours after addition of ATc when indicated. Red dots indicate genes with significantly different abundance ( $\log_2$  fold change  $\geq 2$ , adjusted  $p < 0.01$ ) for the indicated comparisons.



**Figure 3-3: Knockdown of *rne* causes transcriptome-wide mRNA stabilization, which explains some of the gene expression changes and appears to be partially compensated for by decreased transcription rates.** A-B. Most genes had longer half-lives after eight hours of exposure to ATc to knock down *rne* expression. C. A weak but statistically significant positive correlation between mRNA abundance and degree of stabilization upon *rne* knockdown suggests that some of the gene expression changes upon knockdown are due to slower transcript degradation. D. Transcription rates in the *rne* knockdown condition were predicted from the known steady-state abundances and degradation rates.

Many of the changes in transcript abundance upon *rne* knockdown appeared to be attributable to increased mRNA half-lives. There was a modest but statistically significant positive correlation between transcript abundance in the *rne* knockdown condition and the fold-change half-life

between the control and knockdown conditions (Figure 3-3C). We used the measured mRNA abundances and half-lives to calculate predicted transcription rates. A majority of genes had decreased transcription rates, consistent with physiological compensation for the decreased mRNA degradation rates (Figure 3-3D, Table S3-2).



**Figure 3-4: Leadered transcripts and highly abundant transcripts tend to be more strongly stabilized upon *rne* knockdown.**

**A.** There was a weak but statistically significant positive correlation between the abundance of transcripts prior to *rne* knockdown and the degree of stabilization upon knockdown. **B.** Leadered transcripts were on average stabilized more than leaderless transcripts upon *rne* knockdown. **C.** Leadered transcripts had a higher median abundance prior to *rne* knockdown. **D.** The trend toward greater stabilization of leadered transcripts was still evident when examining subsets of genes for which the leadered group did not have higher median abundance than the leaderless group. **E.** The abundances of the gene subsets used in panel D. For B-E, the horizontal bars indicate the median and asterisks indicate comparisons by the Mann-Whitney test.

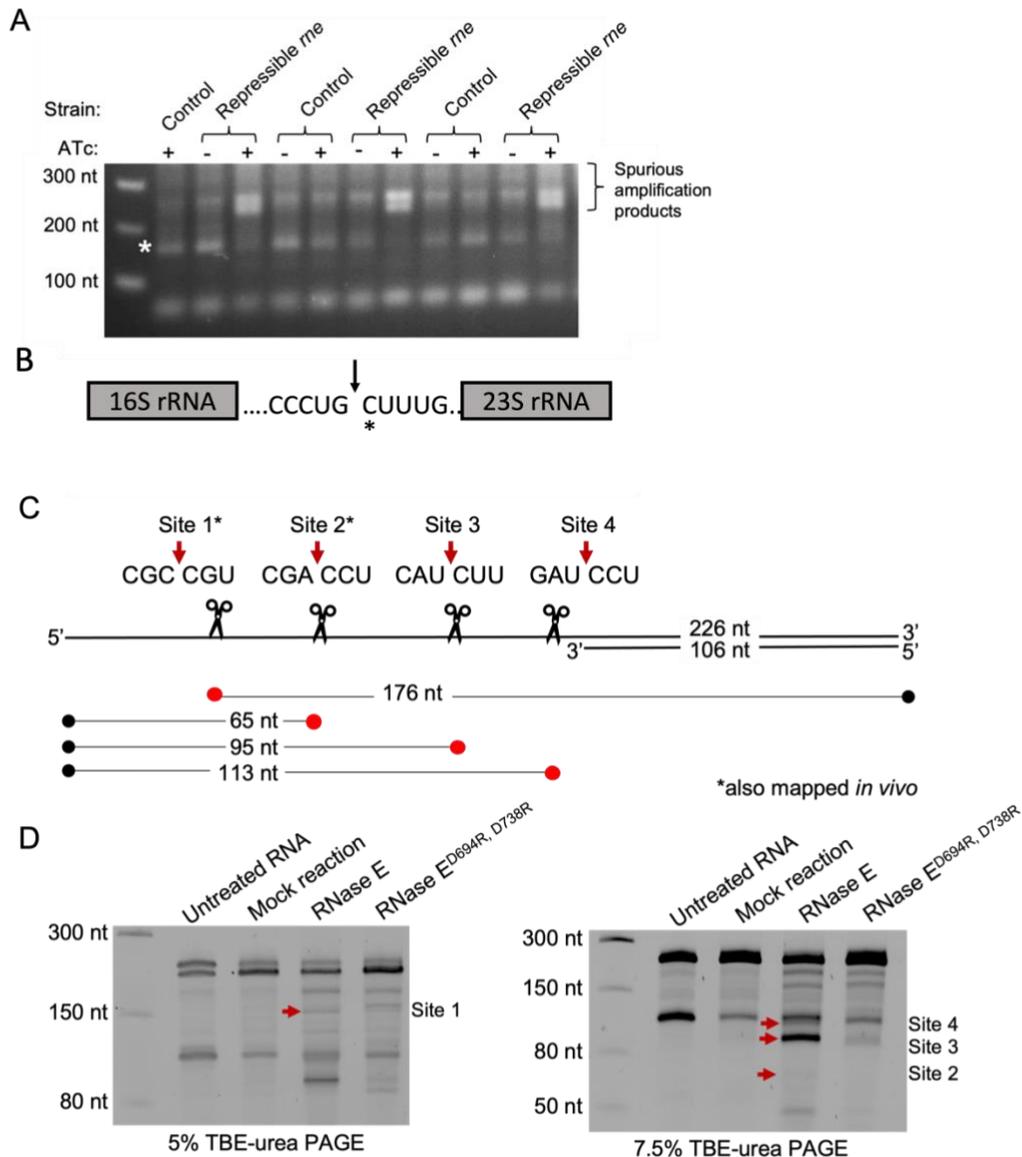
While most of the transcriptome was stabilized in the *rne* knockdown condition, the degree of stabilization varied substantially among genes (Figure 3-3B). To investigate the factors that influence transcript sensitivity to RNase E levels, we examined fold-change half-life in the *rne* knockdown as a function of other potentially relevant characteristics. There was a weak but statistically significant correlation between mRNA abundance in the control condition and fold-change in half-life upon *rne* knockdown (Figure 3-4A), suggesting that more abundant transcripts may be more sensitive to RNase E. On average, leaderless genes were less affected by *rne* knockdown than leadered genes (Figure 3-4B). Leaderless genes also had a lower median abundance than leadered genes in the control condition (Figure 3-4C); however, this did not appear to explain the difference in response to *rne* knockdown because that difference persisted when examining a subset of genes for which leadered transcript abundance was not higher (Figure 3-4D, E). Leadered genes may therefore be more sensitive to RNase E than leaderless genes. However, both gene type categories contained genes that were unaffected by *rne* knockdown as well as genes that were strongly affected, indicating that additional factors affect RNase E sensitivity. Given that RNase E is strongly stimulated by engagement of transcript 5' ends in *E. coli* (Mackie, 1998), we considered that accessible 5' ends might make transcripts more sensitive to RNase E. However, we did not find correlations between fold-change in half-life upon *rne* knockdown and predicted secondary structure near the 5' ends of transcripts (Figure S3-3).

### **RNase E is responsible for previously reported *M. smegmatis* mRNA cleavage sites and cleavage 5' of cytidines**

Given the global role for RNase E implied by these data, we wondered if RNase E was the enzyme responsible for mRNA cleavage events that we previously mapped (Martini et al., 2019). Those cleavage events occurred across the transcriptome at a sequence motif not previously associated with any RNase in any organism. The dominant feature of the cleavage site sequence context

was a cytidine immediately downstream of the cleavage site. We confirmed the role of RNase E in N↓C site cleavages in two ways. First, we used 5' RACE to qualitatively compare the abundance of 5' ends arising from a putative RNase E cleavage event in the rRNA precursor (Figure 3-5A). We mapped a 5' end in the spacer region between the 16S and 23S rRNAs resulting from cleavage at the sequence UG↓CU (Figure 3-5B). Consistent with the idea that RNase E is responsible for cleaving this site, the band corresponding to the 5' end produced by the cleavage event is fainter in the *rne* knockdown (Figure 3-5A). This is consistent with a previously reported role for RNase E in cleaving near this location (Taverniti et al., 2011), although the method used in that report did not permit precise identification of the 5' end as we did here.

Second, we overexpressed and purified the catalytic domain of *M. smegmatis* RNase E in *E. coli* to test its cleavage specificity *in vitro*. This recombinant RNase E catalytic domain lacked the predicted scaffold domains encoded by residues 2-145 and 825-1037, similar to the RNase E variants used for *in vitro* work in many reports (Zeller et al., 2007). It also had N-terminal 6x-His and FLAG epitope tags to facilitate purification. A variant containing the predicted catalytic site mutations D694R and D738R was purified to use as a control. The purified proteins were incubated with an *in vitro*-transcribed substrate that contained a duplex region and a single-stranded region (Figure 3-5C & Figure S3-4). Bands that appeared only in the digest with the catalytically active enzyme were subject to 5' and 3' RACE to map the cleavage site locations (Figure 3-5D). We mapped four distinct cleavage sites, all in the single-stranded portion of the substrate (Figure 3-5C). Two were at the same positions as cleavage sites that we previously mapped *in vivo*, and all four occurred at the sequence motif RN↓CNU. These data confirm the propensity of RNase E to cleave single-stranded RNAs at the phosphodiester bond 5' to cytidines.



**Figure 3-5: RNase E is responsible for previously reported *M. smegmatis* mRNA cleavage sites and cleavage 5' of cytidines.** A. RNA samples extracted from triplicate cultures of control strain and repressible *rne* strain were run on a TBE-agarose gel and the abundance of 5' end (band with \*) products arising from a putative RNase E cleavage event was compared. B. Schematic of the mapped 5' end in the spacer region between the 16S and 23S rRNAs. C. The locations of four mapped cleavage sites from 5' RACE and 3' RACE are shown schematically. The schematic is not to scale. D. *In vitro*-transcribed substrates (Figure S3-4) were incubated with purified RNase E (residues 146-824) and its variant with catalytic site mutations (D694R and D738R), then cleaved products were divided into two parts and run on 5% and 7.5% TBE-urea PAGE gels. Products for which a 5' or 3' end was mapped in panel C are indicated with red arrows.

## Discussion

Here we used a combination of approaches to define the role of RNase E in mycobacterial mRNA degradation and identify its targets. The dramatic effect of RNase E knockdown on mRNA degradation rates in *M. smegmatis* is consistent with the essentiality of this enzyme in mycobacteria. Variability in the extent of stabilization among genes suggests that while RNase E likely contributes to degradation of most mRNAs, other RNases may contribute differentially across the transcriptome. For example, the essential exoribonuclease PNPase could conceivably be the major degradation factor for the genes that are minimally affected by RNase E knockdown. An alternative explanation is that some mRNAs may be exquisitely susceptible to degradation, such that they are still efficiently degraded by the small amounts of RNase E present in the knockdown condition. Leadered transcripts appeared to be more sensitive to RNase E levels than leaderless transcripts, suggesting that 5' UTRs may serve as platforms for engagement with RNase E. However, the lack of correlation between fold stabilization and predicted secondary structure near the 5' ends of transcripts suggests that the effects of 5' UTRs on RNase E engagement cannot be explained simply by exposure of transcript 5' ends.

Taken together, our data implicate RNase E as the enzyme responsible for mRNA cleavage events 5' of cytidines, which are widespread *in vivo* in both *M. smegmatis* (Martini et al., 2019) and *M. tuberculosis* (our unpublished data). This cleavage sequence preference was not found in a previous publication reporting the *in vitro* activity of *M. tuberculosis* RNase E (Zeller et al., 2007). However, the test substrates used in that study did not contain the recognition motif RR↓CNU that we found to be most prominent *in vivo*. Our results are therefore not inconsistent with that study, but rather expand upon it. Both our *in vivo* and *in vitro* data indicate that RNase E has a strong preference for cleaving 5' of cytidines, but that the impact of the surrounding sequence is weak. This could mean that the identities of the surrounding nts are unimportant for RNase E binding

and cleavage, or that the identities of those nts are important but act in combinatorial ways that are not obvious from the data currently available. Interpretation of the *in vivo* cleavage patterns is complicated because (1) cleavage is likely affected by ribosomes and RNA-binding proteins that protect and expose particular regions and (2) cleavage products that are rapidly degraded are not detected and our methods therefore are biased towards identification of cleavage events that produce stable products. *In vitro*, there was a clear preference for cleavage 5' of cytidines, but there were many cytidines that did not produce detectable cleavage products, indicating that RNase E prefers certain positions within the test substrate. We examined secondary structure predictions of the substrate and found that the cleaved positions did not correspond to the positions most likely to be in single-stranded loops. The *in vitro* cleavage pattern therefore cannot be easily explained by the predicted secondary structure. Stem-loops near cleavage sites have been shown to stimulate or direct cleavage by *E. coli* RNase E in some contexts (Bandyra et al., 2018; Schuck et al., 2009; Updegrave et al., 2019), and therefore the sites cleaved in our study could be dictated in part by such cis-acting elements. *Cis*-acting unpaired regions have also been shown to affect cleavage by *E. coli* RNase E (Kime et al., 2014). The potential impact of the scaffold domains (which were partially deleted in our purified RNase E) should also be considered, as the *E. coli* RNase E scaffold domains were recently shown to affect catalytic activity (Ali & Gowrishankar, 2020).

The strong preference of mycobacterial RNase E to cleave 5' of cytidines contrasts with the lack of strong base specificity by *E. coli* RNase E at this position (McDowall et al., 1994). Residue F67 in *E. coli* RNase E is highly conserved among the Proteobacteria and was proposed to play a key role in the catalytic mechanisms by forming a binding pocket for the base one or two nt downstream of the cleavage site (Callaghan et al., 2005). Mutating this residue to Ala in *E. coli* abolished activity *in vitro* (Callaghan et al., 2005). However, the equivalent position in both *M. smegmatis* and *M. tuberculosis* encodes Val. It is tempting to speculate that differences in the key

residues that position the RNA substrate in the active site are responsible for the differences in cleavage sequence preference for mycobacterial vs *E. coli* RNase E; further work is needed to investigate this question.

## Materials and Methods

### Bacterial strains and culture conditions

*Mycobacterium smegmatis* strain mc2155 and derivatives (Table 3-1) were grown in Middlebrook 7H9 liquid medium supplemented with glycerol, Tween-80, catalase, glucose, and sodium chloride as described (Vargas-Blanco et al., 2019) or on Middlebrook 7H10 with the same supplements except for Tween-80. *Escherichia coli* NEB-5-alpha (New England Biolabs) was used for cloning and BL21(DE3) was used for protein overexpression. *E. coli* was grown on LB. Liquid cultures were grown at 37°C with a shaker speed of 200 RPM. When indicated, anhydrotetracycline was used at 200 ng/mL. Antibiotic concentrations used were 25 µg/mL for kanamycin and 150 µg/mL for hygromycin.

### *M. smegmatis* strain construction

#### SS-M\_0418:

The “inducible RNase E KD” was built using the “Mycobacterial recombineering” approach for gene replacement (van Kessel & Hatfull, 2008). 2 µg target DNA fragment containing 500 bp upstream of *rne* (*msmeg\_4626*) native promoter, Hygromycin cassette, plasmid promoter and its 5' UTR containing tet operator (*tetO*) and the first 500 bp of *rne* coding sequence was PCR amplified from pSS187 and DNA dialysis was performed before transformation into SS-M\_0078 (Wildtype *M. smegmatis* with recombinase plasmid pNit-recET-Kan). pSS187 was built using NEBuilder HiFi assembly. Colonies were confirmed by sequencing. Counterselection with 15%

sucrose was followed by PCR screening to identify an isolate (SS-M\_0151) which lost the recombinase plasmid after 24-48 hours incubation in 7H9 medium without drug. SS-M\_0151 was further transformed with plasmid pSS291 encoding a Tet repressor (TetR) into the L5 phage integration site.

#### **Construction of SS-M\_0424:**

A hygromycin-resistant control strain was built using similar protocol as described for SS-M\_0418, the difference being that the target DNA fragment that was transformed into SS-M\_0078 only contained the hygromycin resistance cassette to be placed upstream of the native promoter of *rne*.

#### **RNA extraction, RNAseq library construction, and sequencing**

Cultures were grown to an OD of 0.8-0.9 and divided into a series of 14 ml conical tubes. Rifampicin was added to a final concentration of 150 µg/mL and cultures were harvested at (0, 1, 2, 4, 8, 16, 32 min) by freezing in liquid nitrogen. Frozen cultures were stored at -80°C and thawed on ice for RNA extraction. RNA was extracted as in (Vargas-Blanco et al., 2019). Illumina libraries were constructed and sequenced by the Broad Institute Microbial 'Omics Core using the library construction procedure described in (Shishkin et al., 2015).

#### **cDNA synthesis and quantitative PCR**

cDNA was synthesized as described in (Vargas-Blanco et al., 2019) and qPCR was performed using the conditions described in (Vargas-Blanco et al., 2019).

#### **RNAseq data analysis for expression and half-life determination**

Samples harvested immediately after addition of rifampicin were used to determine steady-state gene expression levels. Reads were aligned with Bowtie v1.2.2 (Langmead et al., 2009), read

alignment processed by SAMtools v1.9 (Danecek et al., 2021), counts determined by HTSeq v0.10.1 (Anders et al., 2015), and differential expression analysis performed with DESeq2 v1.10.1 (Love et al., 2014).

To calculate mRNA half-lives, data from all of the timepoints following rifampicin treatment were processed. First, reads were aligned using BWA-MEM v0.7.17 (Li, 2013). Next, the resulting alignments were processed for each strand by SAMtools v1.10 (Danecek et al., 2021). The raw coverage of each coordinate was calculated through BEDTools v2.29.1 (Quinlan & Hall, 2010). Then we conducted a two-step normalization of the raw coverage. First, coverage was normalized by the total amount of reads of each sample. Then we used the averaged qPCR estimated expression level relative to time zero of Control strain (no ATc) over 8 genes (*sigA*, *rraA*, *esxB*, *atpE*, *msmeg\_4626*, *msmeg\_4665*, *msmeg\_5691*, *msmeg\_6941*) to calculate the normalization factors for all the samples. Specifically, for a given time point  $T_n$ , we calculated the normalization factor  $F_{T_n}$  with qPCR targeted gene expression measurement as indicated below:

$$T_{n,i} \{Gene_i \text{ target expected coverage}\} = \left( \frac{T_{n,i} \{Gene_i \text{ qPCR target}\}}{T_{0,i} \{Gene_i \text{ qPCR target}\}} \right) * T_{0,i} \{Gene_i \text{ target coverage}\}$$

$$F_{T_n} = \frac{1}{8} \sum_{i=1}^8 \frac{T_{n,i} \{Gene_i \text{ target expected coverage}\}}{T_{n,i} \{Gene_i \text{ target coverage}\}}$$

Then the final normalized coverage for each coordinate is calculated by multiplying first step normalized coverage with the normalization factor for each sample. The coverage for each gene is then represented by the summation of the normalized coverage of its coordinates.

## **5' RACE to map a putative RNase E cleavage site in the rRNA transcript**

Enzymes were obtained from New England Biolabs unless otherwise specified. Five hundred ng of each RNA sample were mixed with 1 µg of oligo SSS1016 in a total volume of 9 µl, incubated at 65°C for 10 minutes, and cooled on ice for 5 minutes. Each sample was combined with 21 µl of ligation mix containing 10 µl of 50% PEG8000, 3 µl of 10X T4 RNA ligase buffer, 3 µl of 10 mM ATP, 3 µl of DMSO, 1 µl of murine RNase inhibitor, and 1 µl of T4 RNA ligase. Samples were incubated at 20°C overnight and purified with a Zymo RNA Clean & Concentrator-5 kit according to the manufacturer's instructions with the following modifications: samples were first diluted by addition of 20 µl of RNase-free water, and samples were eluted in 8 µl of RNase-free water. Three µl of each purified ligation were then subject to cDNA synthesis or mock (no-RT) cDNA synthesis. Samples were combined with 1 µl of a mix containing 50 mM Tris pH 7.5 and 500 ng/µl random primers (Invitrogen), incubated at 70°C for 10 minutes, and snap-cooled in an ice-water bath. cDNA was then synthesized as in (Vargas-Blanco et al., 2019). 35 ng of cDNA or the equivalent volume of the corresponding no-RT sample were mixed with 2.5 µl 10X Taq buffer, 1.25 µl each 10 µM primers SSS1017 and SSS2210, 1.25 µl DMSO, 0.5 µl of 10 mM each dNTP mix, 0.167 µl Taq polymerase, and water to a final volume of 25 µl. Cycling conditions were 5 minutes at 95°C, 35 cycles of 30 seconds at 95°C, 20 seconds at 52°C, and 25 seconds at 68°C, and a final 5 minute incubation at 68°C. PCRs were run on 1.5% agarose gels and bands that appeared in cDNA samples but not in no-RT samples were excised and sequenced with SSS2210 to identify the adapter/RNA junctions.

## **Overexpression and purification of recombinant RNase E variants**

An RNase E variant encoding residues 146-824 was cloned into pET-42 for overexpression in *E. coli* as follows. pSS348, carrying the *M. smegmatis rne* coding sequence with a  $\Delta$ 2-145 partial N-terminal deletion,  $\Delta$ 825-1037 full C-terminal deletion, and an N-terminal addition of 6XHis,

3XFLAG, TEV protease cleavage site, and a 4xGly linker sequence was used as a template for creation of pSS420. pSS420 was then used as a template for creation of pSS421, which has the mutations D694R and D738R to make it catalytically dead.

These two RNase E variants were purified as described in (Davis & Dainis, 2020).

### ***In vitro* RNA synthesis and *in vitro* RNase E cleavage assay**

For *in vitro* transcription, template DNA for A-initiated primary *atpB-E* sense transcript and anti-sense transcript was PCR amplified using gene specific oligonucleotides with a T7  $\phi$ 2.5 promoter and “AGG” transcription start sites (TAATACGACTCACTATTAGG) at the 5' end of the forward oligonucleotide. Monophosphorylated RNA was synthesized from DNA templates in the presence of a 50-fold molar excess of AMP over ATP (Luciano et al., 2017) with T7 RNA polymerase (NEB M0251). Each 50  $\mu$ L reaction mixture contained 1X reaction buffer, 5 mM DTT, 1 mM UTP, 1 mM CTP, 1 mM GTP, 0.5 mM ATP, 25 mM AMP, 5 units/ $\mu$ L T7 RNA polymerase, 1 unit/ $\mu$ L Murine RNase inhibitor and 2  $\mu$ g DNA template. The *in vitro* transcription reaction was incubated at 37°C for 16 hours. The resulting transcripts were treated with TURBO DNase at 37°C for 30 minutes before further purification with a Zymo RNA Clean & Concentrator-5 kit.

Before the cleavage assay, the primary *atpB-E* sense transcript and anti-sense transcript were combined at a 1:1 molar ratio for RNA annealing and the mixtures were incubated in the presence of 5X annealing buffer (50 mM Tris-HCl, pH 7.9, 0.5 mM EDTA, pH 8.0, 100 mM NaCl) in a 10  $\mu$ L reaction, and incubated for 1 min at 90°C, then slowly cooled down to room temperature (No less than 30 min). The resulting annealed RNA mix was immediately stored at -80°C.

*In vitro* RNase E cleavage reactions were heated at 65°C for 3 min prior to adding the enzyme, then cooled and incubated at 37°C for 1 hour following addition of the enzyme. The reaction buffer

was composed of 20 mM Tris-HCl, pH7.9, 100 mM NaCl, 5% Glycerol, 0.01% IGEPAL, 0.1 mM DTT, 10 mM MgCl<sub>2</sub>, and each reaction containing 300 ng annealed RNA mix and 80 ng of purified RNase E. For mock reactions, water was used instead of enzyme. Reactions were stopped by adding equal amount of 2X Invitrogen™ Gel loading buffer II and then subjected to electrophoresis on a 7.5% or 5% polyacrylamide-8 M urea gel and visualized after 15 min staining with SYBR Gold Nucleic Acid gel stain. Bands of interest were excised and RNA was recovered using Zymo small-RNA PAGE recovery kit, followed by identification of cleavage sites by 5' RACE or 3' RACE.

### **5' RACE and 3' RACE to map cleavage sites from *in vitro* RNase E cleavage assays**

For 5' RACE, RNA extracted from bands as described above was mixed with 1 μL of 1 μg/μL RNA oligo SSS1016 at 65°C for 5 min, chilled on ice and then incubated with T4 RNA ligase (NEB M0437M), Murine RNA inhibitor, 100% DMSO, 10 mM ATP, 50% PEG 8000 at 20°C for 18 hours. Ligated RNA was column-purified and followed by cDNA synthesis using gene specific reverse oligo SSS916 located close to 3' end oligo SSS916. cDNA was purified and then was used as template to perform Taq PCR with SSS1018 (short version DNA primer of SSS1016) and SSS916. PCR amplified product was sent for DNA sequencing using SSS916.

For 3'RACE, RNA extracted from bands as described above was mixed with 1 μL of 1 μg/μL RNA oligo SSS2433 at 65°C for 5 min, chilled on ice and then incubated with T4 RNA ligase (NEB M0437M), Murine RNA inhibitor, 100% DMSO, 10 mM ATP, and 50% PEG 8000 at 17°C for 18 hours. Ligated RNA was purified and followed by cDNA synthesis using reverse oligo SSS2434 (short version DNA primer of SSS2433). cDNA was purified and then was used as template to perform Taq PCR with gene specific forward oligo SSS917 located at 5' end and SSS2434. PCR amplified product was sent for DNA sequencing using SSS917.

## Funding

This work was supported by NSF CAREER award 1652756 to SS and by institutional startup funds from WPI.

## Acknowledgement

We thank members of the Shell lab for technical assistance and helpful discussions.

## References

- Ali, N., & Gowrishankar, J. (2020). Cross-subunit catalysis and a new phenomenon of recessive resurrection in *Escherichia coli* RNase E. *Nucleic Acids Res*, *48*(2), 847-861. <https://doi.org/10.1093/nar/gkz1152>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169. <https://doi.org/10.1093/bioinformatics/btu638>
- Babitzke, P., & Kushner, S. R. (1991). The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of *Escherichia coli*. *Proc Natl Acad Sci U S A*, *88*(1), 1-5. <https://doi.org/10.1073/pnas.88.1.1>
- Bandyra, K. J., Wandzik, J. M., & Luisi, B. F. (2018). Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E. *Mol Cell*, *72*(2), 275-285 e274. <https://doi.org/10.1016/j.molcel.2018.08.039>
- Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., & Luisi, B. F. (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, *437*(7062), 1187-1191. <https://doi.org/10.1038/nature04084>
- Carpousis, A. J., Van Houwe, G., Ehretsmann, C., & Krisch, H. M. (1994). Copurification of *E. coli* RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, *76*(5), 889-900. [https://doi.org/10.1016/0092-8674\(94\)90363-8](https://doi.org/10.1016/0092-8674(94)90363-8)
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2). <https://doi.org/10.1093/gigascience/giab008>
- Davis, A. B., & Dainis, J. P. (2020). *Investigating the Cleavage Specificity of Mycobacterium smegmatis RNase E in vitro*. <https://digital.wpi.edu/show/w95052875>
- Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., & Schnappinger, D. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res*, *33*(2), e21. <https://doi.org/10.1093/nar/gni013>
- Esquerre, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & Coccagn-Bousquet, M. (2014). Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res*, *42*(4), 2460-2472. <https://doi.org/10.1093/nar/gkt1150>

- Esquerre, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C., Coccagn-Bousquet, M., & Girbal, L. (2015). Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genomics*, *16*, 275. <https://doi.org/10.1186/s12864-015-1482-8>
- Johnson, E. O., Office, E., Kawate, T., Orzechowski, M., & Hung, D. T. (2020). Large-Scale Chemical-Genetic Strategy Enables the Design of Antimicrobial Combination Chemotherapy in *Mycobacteria*. *ACS Infect Dis*, *6*(1), 56-63. <https://doi.org/10.1021/acsinfecdis.9b00373>
- Kime, L., Clarke, J. E., Romero, A. D., Grasby, J. A., & McDowall, K. J. (2014). Adjacent single-stranded regions mediate processing of tRNA precursors by RNase E direct entry. *Nucleic Acids Res*, *42*(7), 4577-4589. <https://doi.org/10.1093/nar/gkt1403>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, *10*(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- Lin, H. H., Lin, C. H., Hwang, S. M., & Tseng, C. P. (2012). High growth rate downregulates *fumA* mRNA transcription but is dramatically compensated by its mRNA stability in *Escherichia coli*. *Curr Microbiol*, *64*(5), 412-417. <https://doi.org/10.1007/s00284-012-0087-6>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Luciano, D. J., Vasilyev, N., Richards, J., Serganov, A., & Belasco, J. G. (2017). A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in *Escherichia coli*. *Mol Cell*, *67*(1), 44-54 e46. <https://doi.org/10.1016/j.molcel.2017.05.035>
- Mackie, G. A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, *395*(6703), 720-723. <https://doi.org/10.1038/27246>
- Martini, M. C., Zhou, Y., Sun, H., & Shell, S. S. (2019). Defining the Transcriptional and Post-transcriptional Landscapes of *Mycobacterium smegmatis* in Aerobic Growth and Hypoxia. *Front Microbiol*, *10*, 591. <https://doi.org/10.3389/fmicb.2019.00591>
- McDowall, K. J., Lin-Chao, S., & Cohen, S. N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J Biol Chem*, *269*(14), 10790-10796. <https://www.ncbi.nlm.nih.gov/pubmed/7511606>
- Mudd, E. A., Prentki, P., Belin, D., & Krisch, H. M. (1988). Processing of unstable bacteriophage T4 gene 32 mRNAs into a stable species requires *Escherichia coli* ribonuclease E. *EMBO J*, *7*(11), 3601-3607. <https://www.ncbi.nlm.nih.gov/pubmed/3061803>
- Nilsson, G., Belasco, J. G., Cohen, S. N., & von Gabain, A. (1984). Growth-rate dependent regulation of mRNA stability in *Escherichia coli*. *Nature*, *312*(5989), 75-77. <https://doi.org/10.1038/312075a0>
- Plocinski, P., Macios, M., Houghton, J., Niemiec, E., Plocinska, R., Brzostek, A., Slomka, M., Dziadek, J., Young, D., & Dziembowski, A. (2019). Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of *Mycobacterium tuberculosis*. *Nucleic Acids Res*, *47*(11), 5892-5905. <https://doi.org/10.1093/nar/gkz251>
- Py, B., Higgins, C. F., Krisch, H. M., & Carpousis, A. J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature*, *381*(6578), 169-172. <https://doi.org/10.1038/381169a0>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841-842. <https://doi.org/10.1093/bioinformatics/btq033>

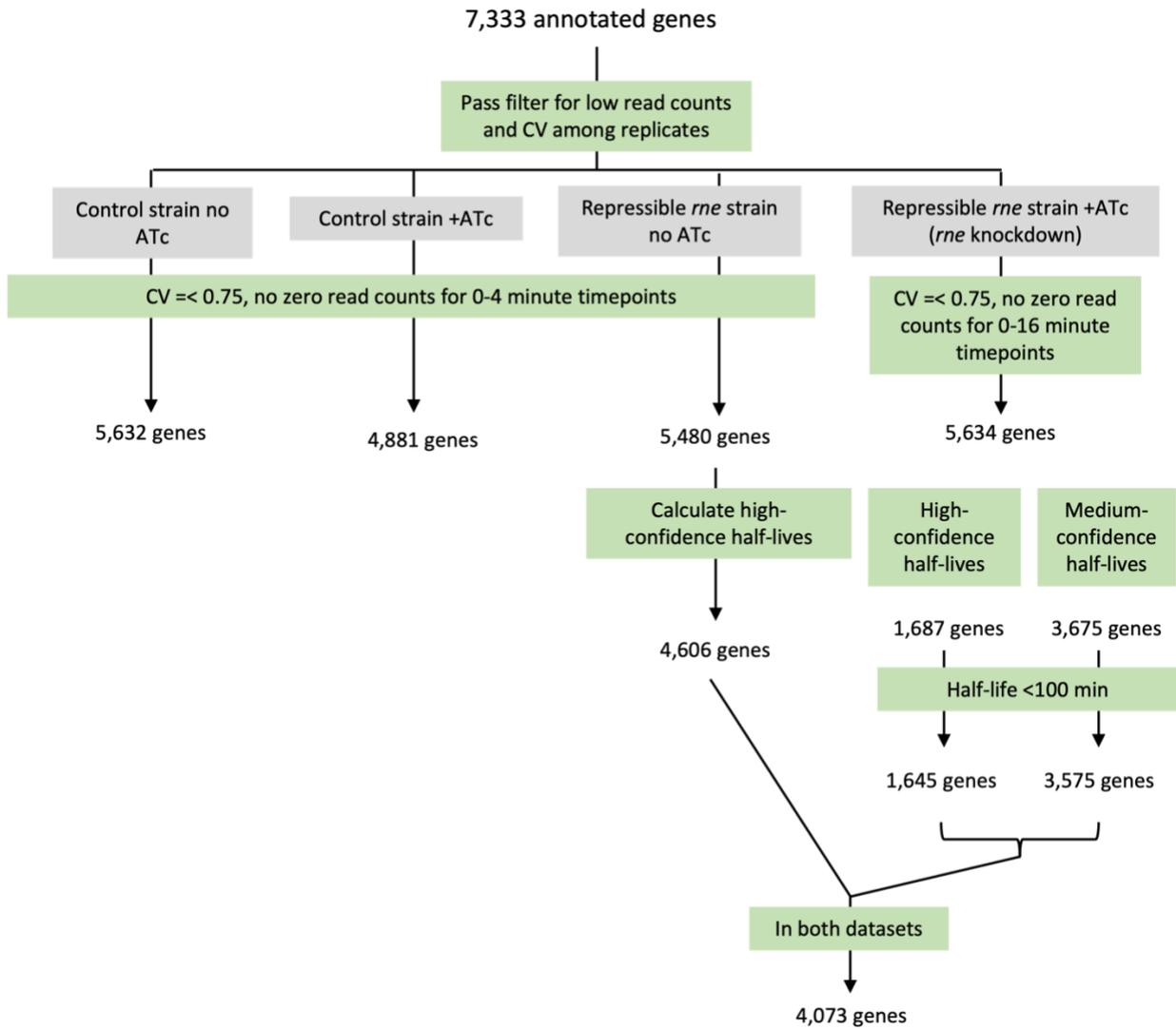
- Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, *48*(1), 77-84. <https://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A*, *100*(22), 12989-12994. <https://doi.org/10.1073/pnas.2134250100>
- Schuck, A., Diwa, A., & Belasco, J. G. (2009). RNase E autoregulates its synthesis in Escherichia coli by binding directly to a stem-loop in the rne 5' untranslated region. *Mol Microbiol*, *72*(2), 470-478. <https://doi.org/10.1111/j.1365-2958.2009.06662.x>
- Shishkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R. P., Rudy, R. F., Patel, M. M., Novod, N., Hung, D. T., Gnirke, A., Garber, M., Guttman, M., & Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat Methods*, *12*(4), 323-325. <https://doi.org/10.1038/nmeth.3313>
- Taverniti, V., Forti, F., Ghisotti, D., & Putzer, H. (2011). Mycobacterium smegmatis RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol*, *82*(5), 1260-1276. <https://doi.org/10.1111/j.1365-2958.2011.07888.x>
- Updegrave, T. B., Kouse, A. B., Bandyra, K. J., & Storz, G. (2019). Stem-loops direct precise processing of 3' UTR-derived small RNA MicL. *Nucleic Acids Res*, *47*(3), 1482-1492. <https://doi.org/10.1093/nar/gky1175>
- van Kessel, J. C., & Hatfull, G. F. (2008). Mycobacterial recombineering. *Methods Mol Biol*, *435*, 203-215. [https://doi.org/10.1007/978-1-59745-232-8\\_15](https://doi.org/10.1007/978-1-59745-232-8_15)
- Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M., & Carpousis, A. J. (1998). Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome. *Genes Dev*, *12*(17), 2770-2781. <https://doi.org/10.1101/gad.12.17.2770>
- Vargas-Blanco, D. A., Zhou, Y., Zamalloa, L. G., Antonelli, T., & Shell, S. S. (2019). mRNA Degradation Rates Are Coupled to Metabolic Status in Mycobacterium smegmatis. *mBio*, *10*(4). <https://doi.org/10.1128/mBio.00957-19>
- WHO. (2021). WHO Global Tuberculosis Report. *World Health Organization*.
- Zeller, M. E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T., & Kaberdin, V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem J*, *403*(1), 207-215. <https://doi.org/10.1042/BJ20061530>

## Tables

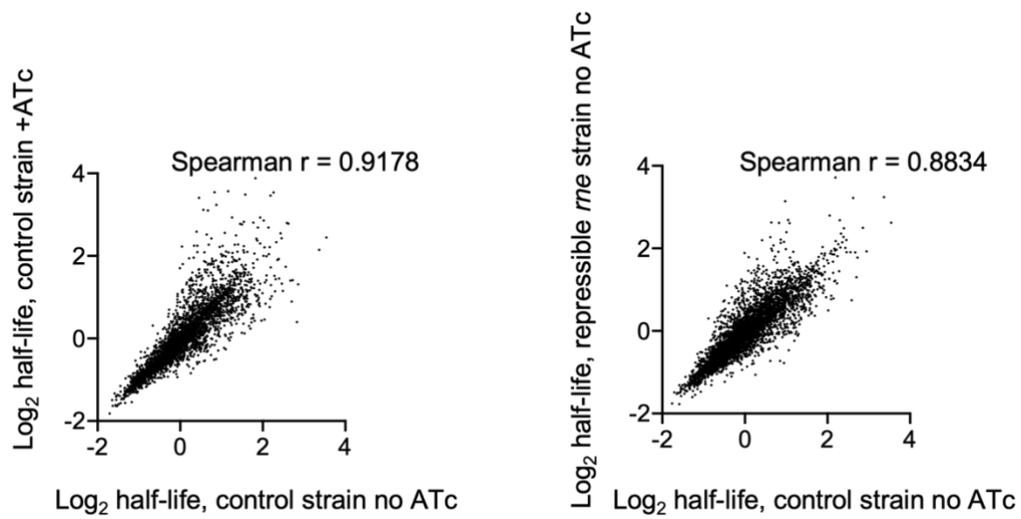
**Table 3-1: Strains and plasmids used in this study.**

Species	Strain	Plasmid	Description	Source
<i>M. smegmatis</i>	mc <sup>2</sup> 155	None	Widely-used lab strain	ATCC
<i>M. smegmatis</i>	SS-M_0424	pSS291: tetR38 driven by promoter ptb38, L5 integrating, kan <sup>R</sup>	mc <sup>2</sup> 155 with the <i>hyg</i> <sup>R</sup> gene inserted with its own promoter upstream of, and divergent from, <i>rne</i> .	This study
<i>M. smegmatis</i>	SS-M_0418	pSS291: tetR38 driven by promoter ptb38, L5 integrating, kan <sup>R</sup>	mc <sup>2</sup> 155 in which the <i>rne</i> ( <i>msmeg_4626</i> ) promoter and UTR (436nt upstream relative to the <i>rne</i> start codon) are replaced by the P766(8G) promoter and associated 5' UTR (Johnson et al., 2020). Additionally, the <i>hyg</i> <sup>R</sup> gene inserted with its own promoter upstream of, and divergent from, <i>rne</i> .	This study
-	-	pET42	Plasmid harbors inducible T7 promoter	-
<i>E. coli</i>	SS-E_0345	pSS420, Kan <sup>R</sup>	pET42 containing #146-824aa of <i>rne</i> tagged with 6xHis-3xFLAG between start codon and rest <i>rne</i> coding sequence.	This study
<i>E. coli</i>	No strain name assigned	pSS421, Kan <sup>R</sup>	pET42 containing #146-824aa of <i>rne</i> tagged with 6xHis-3xFLAG between start codon and rest <i>rne</i> coding sequence, and active sites mutations D694R and D738R	This study

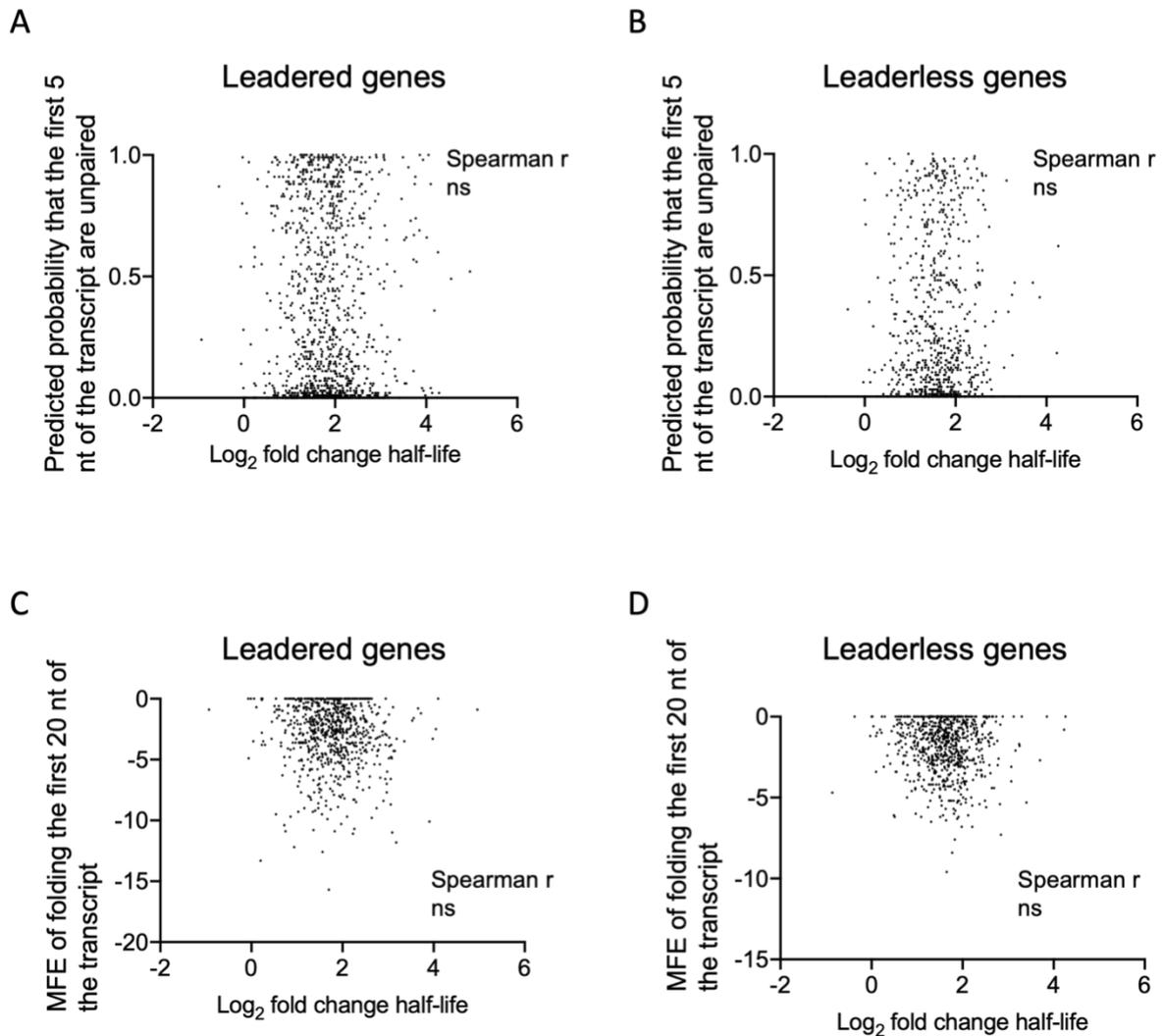
## Supplementary Figures



**Figure S3-1: Filtering genes for half-life calculations.** For 7,333 annotated genes in different conditions: control strain with or without ATc and repressible *rne* strain with or without ATc, we performed different levels of filter for half-life calculations. For *rne* knockdown, we pooled genes with high-confidence half-lives and medium-confident half-lives together for further comparison and analysis.



**Figure S3-2: Half-lives of genes in control conditions are similar.** Left panel: Log<sub>2</sub> half-life comparison was performed between control strain with ATc or no ATc. Right panel: Log<sub>2</sub> half-life comparison was performed between control strain no ATc and repressible *rne* strain no ATc.



**Figure S3-3: Predicted secondary structure at transcript 5' ends is not correlated with the differential stabilization when *rne* is repressed.** For leadered (A) or Leaderless (B) genes, no correlation was found between fold-change in half-life upon *rne* knockdown and unpaired probability of first 5 nt of transcripts. Also, no correlation was found between fold-change in half-life upon *rne* knockdown and minimum free energy of folding for the first 20 nt of transcripts for leadered (C) or leaderless (D) genes.

5' -AGGGCGCUGAUCGCCAUGUUCCCCUGGUACAUCCAGUGGUUCCCCAACGC<sup>^</sup>CGUGUGGAAG-3'

5' -ACCUUCGA<sup>^</sup>CCUGUUCGUCGGCCUCAUCCAGGCCUUCAU<sup>^</sup>CUUCUCGCUGCUGACGAU<sup>^</sup>CC-3'

5' -UGUACUUCAGCCAGUCGAUGGAACUGGACCACGAGGACCACUGACGAGCAACCCUGCUGGA-3'

3' -AUGAAGUCGGUCAGCUACCUUGACCUGGUGCUCCUGGUGACUGCUCGUUGGGACGACCU-5'

5' -CCGAACAAAUCCCUACGACCCGAUCGACACGAACUCUGACGGCAACA-3'

3' -GGCUUGUUUAGGGGAUGCUGGGCUAGCUGUGCUUGAGACUGCCGUUGU-5'

**Figure S3-4: *In vitro*-transcribed partial duplex RNA substrate used for RNase E cleavage assays.**

Black font indicates the sense strand corresponding to the 3' 159 nt of the *M. smegmatis atpB* coding sequence (underlined) and 64 nt of the intergenic region between *atpB* and *atpE*. Blue font indicates an antisense strand used to block RNase E cleavage. Cleavage sites mapped in Figure 3-5 are shown by red carets. The first 3 nt of the transcript were added to facilitate transcription by T7 polymerase and are not part of the *M. smegmatis* gene sequence.

## Supplementary Tables

All tables (Table S3-1 and S3-2) can be accessed online via WPI database.

Chapter 4 : Defining the roles of the degradosome-scaffolding domains of RNase E in *Mycolicibacterium smegmatis*

# Defining the roles of the degradosome-scaffolding domains of RNase E in *Mycolicibacterium smegmatis*

**Ying Zhou, M. Carla Martini, Junpei Xiao, Scarlet S. Shell**

---

## **Abstract**

Regulation of mRNA degradation contributes to control of gene expression in mycobacteria, and allows for saving of energy under stress conditions. In *E. coli*, RNase E plays a key role in mRNA degradation and acts a central recruiter with a scaffold domain that interacts with other proteins to form the RNA degradosome and facilitate mRNA degradation. We aimed to define the scaffold domains of RNase E in *Mycolicibacterium smegmatis* and investigate the roles of these domains. First, we performed an RNase E immunoprecipitation and LC-MS/MS to identify enriched proteins; these identified proteins showed a high degree of similarity to key degradosome-like proteins found in previous *M. tuberculosis* work. We made predictions about the boundaries of the catalytic domain of RNase E and made RNase E scaffold domain deletion strains to assess the effects of scaffold deletions to RNase localization, cell size, and mRNA degradation. The catalytic domain of RNase E was found to be centrally localized with two scaffold domains at the N-terminus and C-terminus. We also observed a co-localization between mCherry-tagged RNase E and Dendra-tagged RNase J by fluorescence microscopy and found a diffuse pattern in the RNase J-Dendra signal when we deleted 330 residues of the RNase E N-terminal scaffold domain in combination with deletion of the C-terminal scaffold domain. Deletion of 330 residues of the N-terminal scaffold domain alone induced slower growth, longer cells, and mRNA stabilization while this deletion combined with C-terminal scaffold domain deletion reversed the growth rate defect and produced shorter cells and more mRNA stabilization. The importance of 330 residues at the N-terminal scaffold domain was also supported by our RNA-seq differential gene expression analysis. Taken

together, our results indicate an important role of the first 330 residues of the RNase E N-terminal scaffold domain in mRNA degradation, growth, gene expression, and RNase J localization.

---

## Introduction

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* (Mtb) and is one of the leading causes of death worldwide from an infectious agent (WHO, 2021). Mtb can adapt and survive in various microenvironments within the host. Changes in gene expression and macromolecule turnover are prominent components of bacterial stress responses. However, the mechanisms that regulate mRNA degradation in mycobacteria are poorly understood.

Much of the fundamental knowledge of bacterial mRNA degradation comes from studies of *Escherichia coli*. In *E. coli*, RNase E is an important endoribonuclease with a well-characterized role in the initiation of mRNA degradation (Laalami et al., 2014). *E. coli* RNase E is composed of a catalytic N-terminal domain and an unstructured C-terminal scaffold domain that directly interacts with a 3' to 5' exoribonuclease polynucleotide phosphorylase (PNPase), a DEAD box RNA helicase (RhIB), and the glycolytic enzyme enolase to form a multiple-proteins complex termed the RNA degradosome (Carpousis et al., 1994; Miczak et al., 1996; Py et al., 1996). RNase E orthologs are found in many bacteria and can also form degradosome-like protein complexes; however, the complex components are varied (Hardwick et al., 2011; Lee & Cohen, 2003; Plocinski et al., 2019; Purusharth et al., 2005). In addition, the sRNA chaperone Hfq can also be involved in small RNA-mediated mRNA degradation when interacting with RNase E (Ikeda et al., 2011). In *E. coli*, the catalytic domain of RNase E forms a homo-tetramer and contains a 5' sensing pocket for substrates (Callaghan et al., 2003; Callaghan et al., 2005). While the N-terminal domain is essential for cell viability, the C-terminal domain is not, and the sequence of

the C-terminal domain is also not well conserved in other RNase E homologs (Apirion & Lassar, 1978; Kaberdin et al., 1998; Ono & Kuwano, 1979).

RNase E is absent from *Bacillus subtilis*, which instead uses the endonuclease RNase Y to initiate mRNA decay (Shahbadian et al., 2009). In addition to RNase Y, RNase J1 and its paralogue RNase J2 have specific endoribonuclease activities similar to that of RNase E (Even et al., 2005). Compared to RNase J2, RNase J1 is essential in *B. subtilis* and harbors a stronger activity. Both RNase J paralogues possess a 5' to 3' exonucleolytic activity and strong preferences for substrates with 5' monophosphates (Mathy et al., 2007; Mathy et al., 2010). A degradosome-like protein complex is present in *B. subtilis*, containing RNase Y, RNase J1/J2, PNPase and glycolytic enzymes (Commichau et al., 2009). While RNase J is not found in *E. coli*, mycobacteria encode both RNase J and RNase E. RNase J in *Mycolicibacterium* (formerly *Mycobacterium*) *smegmatis* has dual endonuclease and 5' to 3' exonuclease activities similar to RNase J1 in *B. subtilis* (Taverniti et al., 2011). As the best-studied bacteria have either RNase E or RNase J but not both, how these two RNases work together in mycobacterial mRNA metabolism is an open question.

Previous studies showed that RNase E is also essential in mycobacteria (Sasseti et al., 2003; Sasseti & Rubin, 2003; Taverniti et al., 2011), and the catalytic domain of *M. tuberculosis* RNase E is similar to that of *E. coli* RNase E, existing as tetramers in solution (Zeller et al., 2007). In addition, Mtb RNase E is a 5' end dependent endonuclease and able to cleave the precursor of 5S rRNA (Zeller et al., 2007). The key role of mycobacterial RNase E in rRNA processing was further supported by another study in *M. smegmatis* (Taverniti et al., 2011). RNase E was also showed to cleave the *furA-katG* transcript, indicating its potential role in mRNA processing and degradation in mycobacteria (Taverniti et al., 2011). Most recently, a work defined the core degradosome components in *M. tuberculosis* including RNase E, RNase J and PNPase and also showed transcriptome-wide expression changes when RNase E was repressed by CRISPRi

(Plocinski et al., 2019). Our work described in Chapter 3 of this dissertation confirm the major role played by RNase E in bulk mRNA degradation in *M. smegmatis*.

To better understand RNase E, here we used *M. smegmatis* as a model organism to investigate the boundaries and roles of its scaffold domains. We constructed RNase E mutant strains with deletions in the predicted scaffold domains and assessed the impacts on growth, sub-cellular localization, gene expression, and mRNA degradation rates. Our results show that the catalytic region of RNase E in *M. smegmatis* is centrally located and a region of 330 residues in the N-terminal scaffold domain is important for mRNA degradation, growth, gene expression, and localization of RNase J.

## Results

### LC-MS/MS of FLAG-tagged RNase E reveals RNase E-associated proteins

Previous studies have revealed that RNase E interacts with different proteins in various species to form variations of the RNA degradosome (Ait-Bara & Carpousis, 2010; Hardwick et al., 2011; Jager et al., 2001; Lee & Cohen, 2003; Miczak et al., 1996; Plocinski et al., 2019; Purusharth et al., 2005; Rosana et al., 2016; Stoppel et al., 2012; Van den Bossche et al., 2016; Zhang et al., 2014). We therefore sought to identify the proteins that interact with RNase E in *M. smegmatis*. To achieve this, we constructed a His-3xFLAG-tagged RNase E strain by placing a His-3xFLAG tag between the start codon and the coding sequence of RNase E on the *M. smegmatis* chromosome (Figure S4-1A). We then performed anti-FLAG immunoprecipitations with the tagged strain to enrich for RNase E and associated proteins. We performed immunoprecipitations with a WT strain in parallel as a control. LC-MS/MS was used to identify proteins in the eluates, and intensity Based Absolute Quantitation (iBAQ) was used as a proxy to assess the relative protein abundance. We used the ratio of normalized iBAQ values in the tagged and WT strains to

identify proteins enriched by RNase E immunoprecipitation (Table 4-1). The enriched proteins included orthologs of several proteins previously identified as components of the *M. tuberculosis* degradosome, such as RNase J, PNPase, and the RNA helicase RhIE (Plocinski et al., 2019). PNPase was commonly found to interact with RNase E to form degradosomes in other bacteria (Ait-Bara & Carpousis, 2010; Carpousis et al., 1994; Hardwick et al., 2011; Plocinski et al., 2019; Van den Bossche et al., 2016; Zhang et al., 2014). In addition, two putative RNA-binding proteins, msmeg\_2436 and msmeg\_6941, were also co-purified with RNase E in our experiment, as was reported for *M. tuberculosis* (Plocinski et al., 2019). Hence, our data suggested that the core degradosome components in *M. smegmatis* and *mtb* have a high degree of similarity.

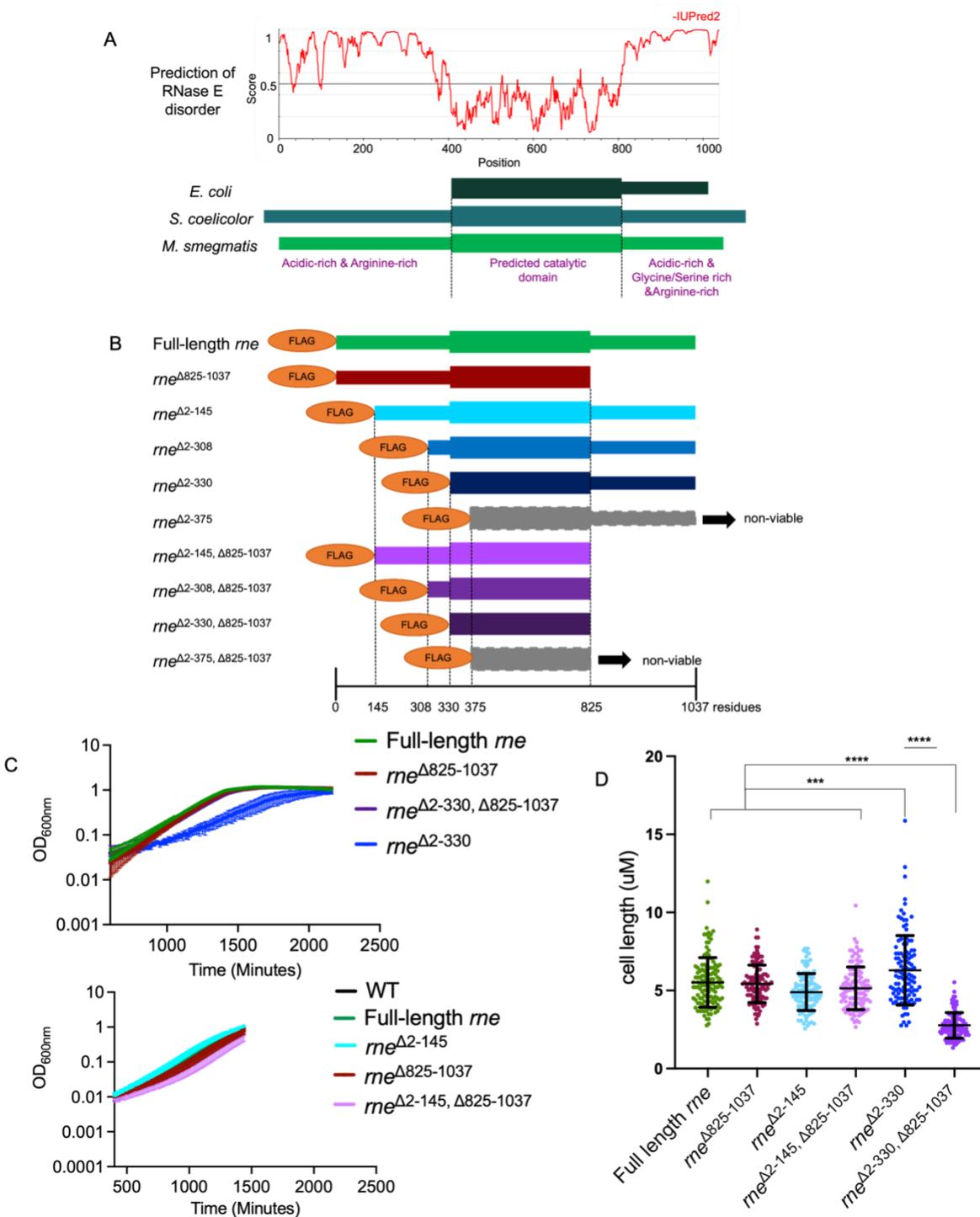
### **Defining the essential domain of *M. smegmatis* RNase E**

Previous studies showed that the N-terminal, catalytic domain of *E. coli* RNase E is evolutionarily conserved in *M. tuberculosis* and many other bacteria, while the C-terminal scaffold domain is poorly conserved (Kaberdin et al., 1998). The scaffold domain interacts with other proteins to form RNA degradosomes (Khemici & Carpousis, 2004; Lopez et al., 1999). In mycobacteria and other actinobacteria, RNase E was predicted to have two scaffold domains with a central catalytic domain (Lee & Cohen, 2003; Zeller et al., 2007) (Figure 4-1A). To predict the boundaries between the scaffold and catalytic domains of RNase E in *M. smegmatis*, we assessed disorder prediction and sequence conservation. We used IUPred2 to predict regions of disorder (Dosztanyi, 2018) and found that the central region of RNase E was predicted to be ordered and the N- and C-terminal regions were predicted to be disordered, as expected (Figure 4-1A). We then did a multiple sequence alignment with RNase E from several other species (not shown) in order to identify the places on either side of the catalytic domain where the amino acid sequence became poorly conserved, as expected for the disordered scaffold domains. Although the unstructured scaffold domains of *M. smegmatis* did not show appreciable sequence similarity to RNase E from other species, these regions contained acidic and arginine-rich regions, as seen in the scaffold

domains of RNase E from other species (Al-Husini et al., 2018; Casaregola et al., 1994; McDowall & Cohen, 1996; Taraseviciene et al., 1995).

To experimentally define the catalytic domain of RNase E in *M. smegmatis* and get deeper insights into the roles of the scaffold domains, we used the disorder prediction and sequence alignments to choose positions for making scaffold domain deletion mutants. To efficiently construct and test variants of this essential protein, we placed a copy of the full-length *rne* gene with its native promoter and 5' UTR and an N-terminal His-3xFLAG tag in a plasmid (pSS267) that integrated into the L5 phage integration site in the *M. smegmatis* chromosome. The coding sequence of *rne* was then deleted from the native locus using a two-step recombination approach to create a strain in which the only copy of *rne* was at the L5 site. Other *rne* variants were then constructed using the same plasmid backbone with a different drug resistance marker and transformed into this strain to see if they could replace pSS267. Plasmids used in this study are listed in Table S4-4 and the *rne* variants are diagramed in Figure 4-1B. Successful replacement of pSS267 with an incoming plasmid indicates that the *rne* variant encoded on that plasmid can support *M. smegmatis* viability. False positives occur when a strain acquires the incoming plasmid without loss of the resident plasmid and can be identified by having resistance to both drugs. We found that an *rne* variant encoding a deletion of residues 2-375 could not support viability, as we only obtained false positive transformants (0/140 transformants were replacements). The same was true for a deletion of residues 2-375 combined with deletion of residues 825-1037 (0/142 transformants were replacements). However, deletions of residues 2-330 and/or 825-1037 were

able to efficiently replace full-length *rme* (Table S4-1). Thus, our working model is that the catalytic domain of *M. smegmatis* RNase E lies between residues 331 and 824.



(Continued on next page)

**Figure 4-1: Defining the essential domain of *M. smegmatis* RNase E.** (A) Prediction of protein structure of *M. smegmatis* RNase E by IUPred2 (Dosztanyi, 2018) (top) and schematic of the domain architecture of RNase E from the indicated organisms. (B) Schematic of FLAG-RNase E derivations placed in single copy at the L5 phage integration site. (C) Growth curves of FLAG-tagged RNase E mutant strains. Except for the WT strain, the indicated *rne* alleles were present at the L5 site and the native copy was deleted. Three biological replicates were grown for each strain. The optical densities at 600 nm ( $OD_{600nm}$ ) were measured and initial OD for each strain was normalized to 0.005. For the bottom panel, the minor differences in growth curves were not reproducible. (D) Cell length of FLAG-tagged RNase E mutant strains. For each strain,  $n > 100$ , and the mean and SD are shown. Strains were compared with ANOVA with Tukey's HSD. \*\*\*,  $p < 0.001$ ; \*\*\*\*,  $p < 0.0001$ .

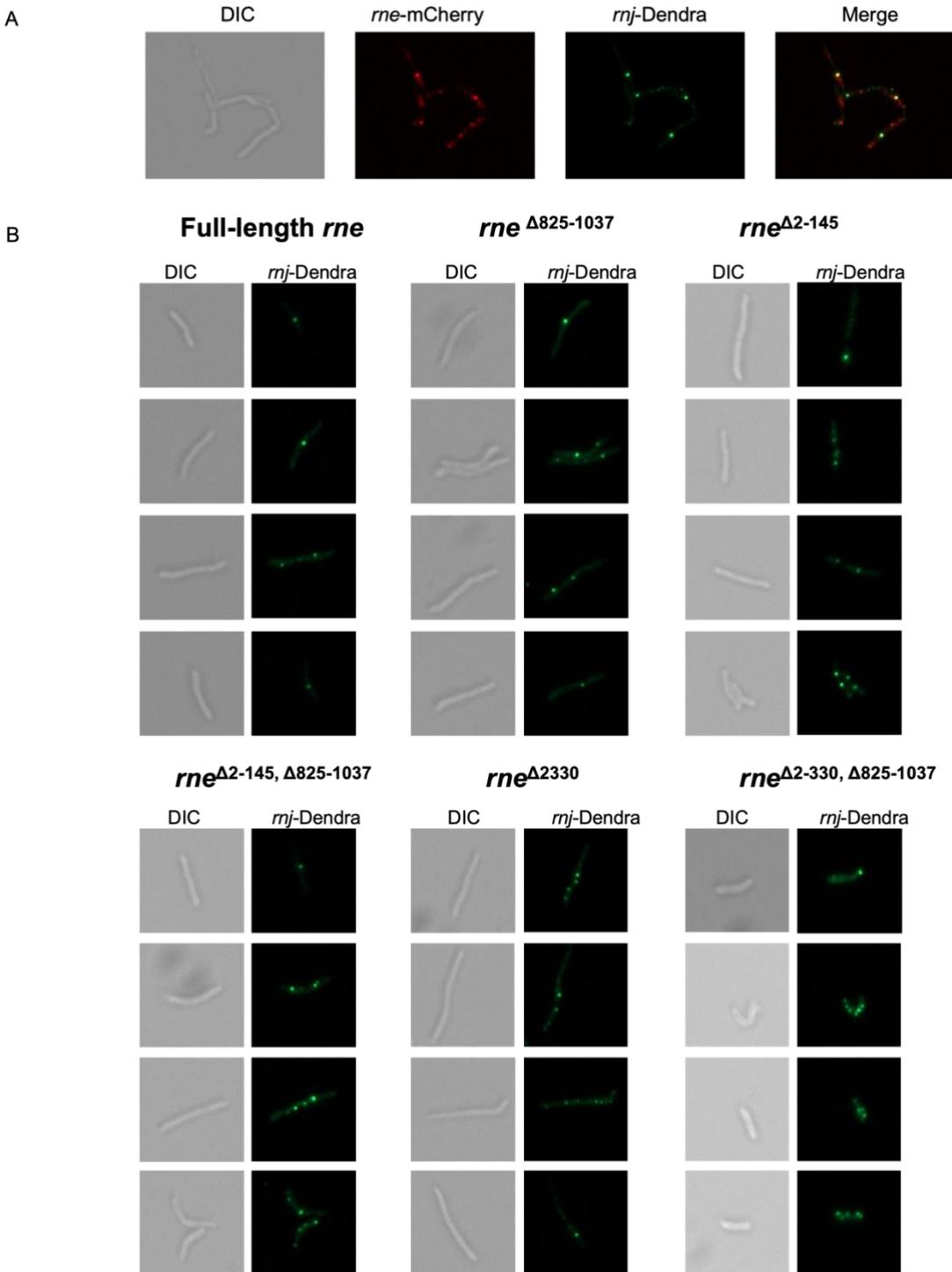
We then performed *in vitro* growth assays to compare the growth kinetics of all FLAG-*rne* tagged strains to the wildtype *M. smegmatis*. The FLAG-*rne* <sup>$\Delta 2-330$</sup>  mutant strain grew slower than WT; however, all other FLAG-*rne* strains grew similarly to WT, including the FLAG-*rne* <sup>$\Delta 2-330, \Delta 825-1037$</sup>  double deletion strain (Figure 4-1C). Moreover, we checked the cell length of the FLAG-*rne* tagged strains by microscopy and analyzed cell-length distributions ( $n > 100$ ) for each strain. The mean cell length was similar for the full-length FLAG-*rne* strain, the FLAG-*rne* <sup>$\Delta 2-145$</sup>  strain, the FLAG-*rne* <sup>$\Delta 825-1037$</sup>  strain, and the FLAG-*rne* <sup>$\Delta 2-145, \Delta 825-1037$</sup>  double deletion. However, the FLAG-*rne* <sup>$\Delta 2-330$</sup>  cells were longer than full-length FLAG-*rne* while those of the FLAG-*rne* <sup>$\Delta 2-330, \Delta 825-1037$</sup>  double deletion strain were shorter (Figure 4-1D). Deletion of the full N-terminal scaffold domain alone therefore induced an elongated cell phenotype as well as a slower growth rate, while the additional deletion of the C-terminal scaffold domain restored normal growth while having an opposite effect on cell size.

### **RNase E and RNase J co-localize, and RNase J localization is partially dependent on RNase E**

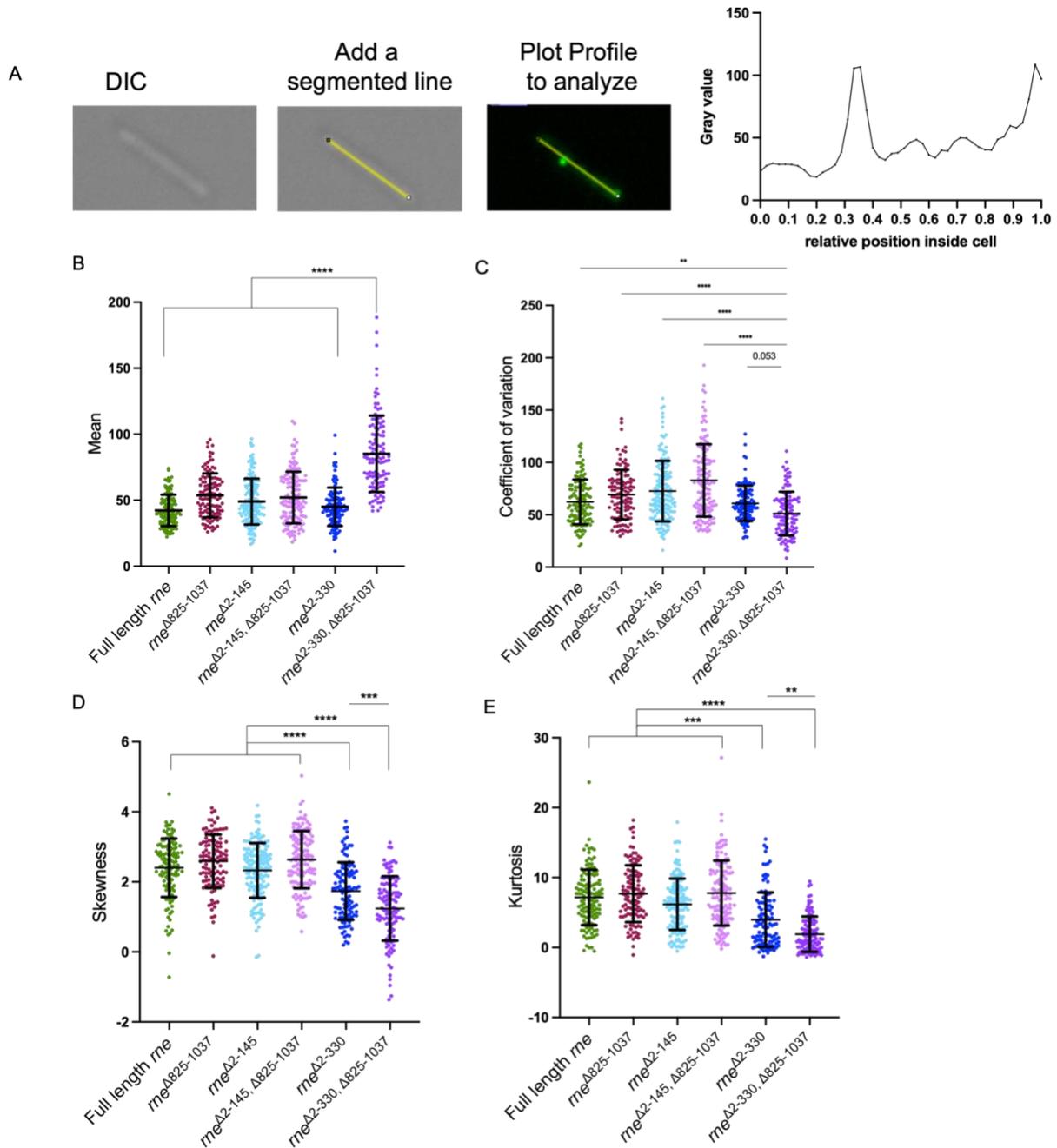
RNase J is a degradosome component in *Bacillus subtilis* and was found to be a core degradosome-associated enzyme in Mtb (Commichau et al., 2009; Plocinski et al., 2019). We therefore hypothesized that RNase J would interact with RNase E in *M. smegmatis* as it does in Mtb (Plocinski et al., 2019). To investigate the effects of this interaction on subcellular localization

of RNases, we constructed an *M. smegmatis* strain with mCherry-tagged RNase E (*rne*-mCherry) and Dendra-tagged RNase J (*rnj*-dendra; plasmid obtained from Judd et al.,2021). The strain lacked the native copy of *rne*, which confirmed the functionality of the *rne*-mCherry fusion. To confirm that the Dendra-tagged RNase J was functional, we took advantage of a previous observation that *M. smegmatis* *rnj* deletion mutants are hypersensitive to rifampicin. We conducted a rifampicin (12 µg/mL) killing assay on the WT parent, an *rnj* deletion strain, and the *rnj* deletion strain complemented with plasmids containing *rnj* or *rnj*-dendra. We observed that killing of the *rnj*-dendra complemented strain was similar to that of WT, confirming that this fusion protein retains at least some of its normal functions (Figure S4-2). When we assessed RNase localization patterns by fluorescence microscopy, we found that *rne*-mCherry showed both diffuse cytoplasmic and punctate localization and *rnj*-dendra largely showed a punctate pattern (Figure 4-2A). As expected, the RNase E and RNase J puncta appeared to localize together, confirming that they interact, either directly or indirectly interaction, in live cells.

We next wanted to investigate the roles of the RNase E scaffold domains in mediating interactions with RNase J, so we constructed truncated versions of mCherry-tagged RNase E. As described for the FLAG-tagged constructs, we made plasmids containing truncated mCherry-tagged RNase E and transformed them into the Giles phage integration site to replace a resident plasmid encoding full-length *rne*. Dendra-tagged RNase J was present at the L5 phage integration site in all the strains. The initial plan was to construct various mCherry-tagged RNase E derivatives (Figure S4-3). However, for most of the mCherry-tagged RNase E mutants we were unable to obtain colonies in which the plasmid encoding full-length *rne* was replaced by the plasmid encoding the truncations. This indicated that the N-terminal deleted RNase E mutants tagged with mCherry could not fulfill the essential function of RNase E (Table S4-2).



**Figure 4-2: RNase E and RNase J appear to localize together and RNase J has more diffuse localization when both the N- and C-terminal scaffold domains of RNase E are deleted. (A)** Representative live cells imaging showing the colocalization of RNase E-mCherry and RNase J-Dendra, merge fluorescence is shown in right column. **(B)** Representative images of live cells showing the localization of RNase J-Dendra from RNase E mutant strains. All images were taken with a Zeiss Axio Imager fluorescence microscope with an apotome using a 63x/1.4 Oil DIC objective.



**Figure 4-3: RNase J localization is partially dependent on RNase E.** (A) An example showing the analysis method “Plot Profile” from image J and corresponding intensity distribution plot from a single cell. Quantification of (B) Mean, (C) coefficient of variation, (D) Skewness, and (E) Kurtosis of the intensity plots acquired from each RNase E mutant strain. Each dot represents the indicated value for a single cell. For each strain,  $n > 100$  cells were analyzed, and the mean and SD are shown. Strains were compared with ANOVA with Tukey’s HSD. \*\*,  $P < 0.05$ , \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ .

As an alternative, we chose to analyze the localization pattern of Dendra-tagged RNase J with the previous FLAG-tagged RNase E mutants we made. To achieve this, we put Dendra-tagged RNase J into the Giles site of the mutant strains and imaged live cells (Figure 4-2B). *mj*-Dendra still showed a punctate pattern in the strains harboring the truncated versions of RNase E. However, there appeared to be more diffuse cytoplasmic fluorescence in addition to puncta in the strains in which residues 2-330 of RNase E were deleted. To quantify this phenotype, we measured the variability in fluorescence signal intensity along the lengths of cells. We used the analysis method “Plot Profile” from Image J to measure the intensities of pixels along the cell (Figure 4-3A). For each cell, a segmented line can be drawn from one end of the cell to the other end, and a plot of single pixel values along the line will be reported, in our case representing the intensity distribution along the cell. With this analytical tool, we were able to get the intensity distributions of hundreds of single cells from each mutant strain. The mean value of intensities collected from the strain harboring *mne*<sup>Δ2-330, Δ825-1037</sup> were significantly higher than those of the strain harboring full-length *mne* (Figure 4-3B). We confirmed by western blot that this strain expresses RNase J-Dendra approximately 2-fold higher than the other strains (Figure S4-4). We also confirmed that the RNase E variants were expressed at levels similar to, or slightly higher than, the full-length RNase E (Figure S4-4).

We reasoned that if a strain had diffuse signal in addition to puncta, the cytoplasmic signal intensities would be less variable than in strains with puncta only. We therefore analyzed the coefficient of variation (CV) to determine the variation of intensity along the lengths of cells. As predicted, the CVs of the intensity distributions collected from *mne*<sup>Δ2-330</sup> and *mne*<sup>Δ2-330, Δ825-1037</sup> were smaller than those of other strains (Figure 4-3C), demonstrating that signal intensities are less punctate and more diffuse. Moreover, we compared the skewness and kurtosis of the distribution patterns with as they are two commonly used metrics of distribution shape. We found significant differences between *mne*<sup>Δ2-330</sup>, *mne*<sup>Δ2-330, Δ825-1037</sup>, and other strains regarding skewness (Figure 4-

3D) and kurtosis (Figure 4-3E), respectively. Compared to the other strains, the skewness of  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  were closer to 0, reflecting the relative lack of pixels with extremely low values. Kurtosis is an indicator of the heaviness of the tails of a distribution, with lower kurtosis meaning fewer extreme values. The kurtosis of  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  was lower than that of the other strains, again consistent with there being fewer pixels with extremely low values. Taken together, these metrics confirm that the *rnj*-dendra intensity in  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  was more diffuse.

### **Investigation of interaction between RNase E and PNPase**

PNPase is a 3' to 5' exonuclease and has been defined in mRNA degradosome of *E. coli*, *S. coelicolor* and *mtb* (Lee & Cohen, 2003; Miczak et al., 1996; Plocinski et al., 2019). To learn whether RNase E interacts with PNPase in *M. smegmatis*, we tagged PNPase with Dendra fluorescent protein to visualize the localization pattern of PNPase by microscope. However, we found that fusing Dendra to either end of PNPase (N terminal or C terminal) caused PNPase to be non-functional, as assessed by the inability of the fusions to support viability when the native copy was repressed by CRISPRi (Figure S4-5). Thus, we are not able to investigate the interaction between RNase E and PNPase due to technical limitations.

### **Differential expression analysis of RNase E mutant strains**

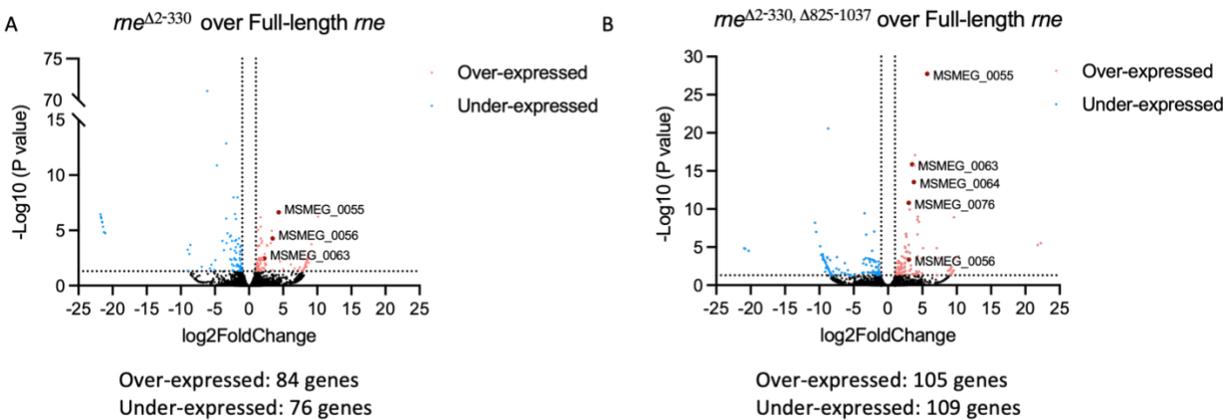
Given the phenotypes observed from deletion of residues 2-330 of the RNase E N terminus, and the impact of the C-terminal domain deletion on these phenotypes, we decided to investigate the global transcription changes in these strains by RNA-seq. We performed RNA-seq on strains harboring the following FLAG-tagged versions of *rne* expressed from the L5 site: full-length,  $\Delta 2-330$ ,  $\Delta 825-1037$ , and the double deletion of 2-330 and 825-1037. Genes with  $\geq 2$ -fold expression difference compared to the full-length strain with adjusted *P* values  $\leq 0.05$  were considered to be differentially expressed.

Comparing  $rne^{\Delta 825-1037}$  to full-length  $rne$ , 22 genes were found to be differentially expressed (8 genes over-expressed and 14 under-expressed) (Supplementary table 3). The small number of differentially expressed genes caused by the C-terminal scaffold domain deletion is consistent with the similar cell size and growth rate of that strain compared to full-length  $rne$  (Figure 1C and 1D).

We next focused on the differential gene expression of  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  compared to full-length  $rne$ . For the former comparison, 160 genes were differentially expressed (84 genes over-expressed and 76 genes under-expressed) (Figure 4-4A) and for the latter, 214 genes were differentially expressed (105 genes over-expressed and 109 genes under-expressed) (Figure 4-4B). Among these genes, we found 58 genes overlapped in both comparison groups. Pathway analysis revealed that in both  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  there was overexpression of several genes belonging to the ESX-1 secretion system locus, which is conserved in many mycobacteria and plays a role in conjugation in *M. smegmatis* (Coros et al., 2008; Gey Van Pittius et al., 2001). Remarkably, 12 genes from the ESX-1 locus in  $rne^{\Delta 2-330, \Delta 825-1037}$  were significantly over-expressed and most of these showed the same trend in  $rne^{\Delta 2-330}$  (Table 4-2). Another set of genes showed an under-expression trend in both  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  (Table 4-3). These were mycofactocin (MFT) biosynthesis genes, which appear to have roles in cholesterol and ethanol utilization, and MSMEG\_6242, which is also important for ethanol utilization (Krishnamoorthy et al., 2019).

Next, to identify the specific genes that may contribute to different cell size in  $rne^{\Delta 2-330, \Delta 825-1037}$  and  $rne^{\Delta 2-330}$ , we focused on the genes within the “cell wall and cell division” category. The gene encoding the porin MspA (MSMEG\_0965), which is important for nutrient uptake and involved in cell wall permeability, was under-expressed ( $\log_2$ FoldChange: -2.83,  $p < 0.001$ ) specifically in  $rne^{\Delta 2-330, \Delta 825-1037}$  but not in  $rne^{\Delta 2-330}$  ( $\log_2$ FoldChange: -0.083) (Hillmann et al., 2007; Stahl et al.,

2001). Further, we found that *whmD* (MSMEG\_1831), encoding the transcriptional regulator WhiB2, was over-expressed ( $\log_2\text{FoldChange}$ : 1.38,  $p = 0.0266$ ) in the  $rne^{\Delta 2-330, \Delta 825-1037}$  strain. A previous study showed that overexpression of *whmD* in *M. smegmatis* induced small cells with multiple septa and significant lysis and only slightly slowed growth kinetics in liquid medium (Gomez & Bishai, 2000), and could therefore potentially explain the short cell phenotype we observed specifically for  $rne^{\Delta 2-330, \Delta 825-1037}$  while the *whmD* in  $rne^{\Delta 2-330}$  had similar expression as WT ( $\log_2\text{FoldChange}$ : 0.037) (Figure 1D). We also examined genes differentially expressed specifically in the  $rne^{\Delta 2-330}$  strain that might be related to the long cell phenotype and slower growth rate. Genes that were differentially expressed in this strain but not in the  $rne^{\Delta 2-330, \Delta 825-1037}$  double deletion strain are listed in Table 4-4. However, we did not identify obvious candidates among these 90 genes to explain the growth defect in the  $rne^{\Delta 2-330}$  strain, and further work will be needed to address the mechanistic basis of that phenotype.



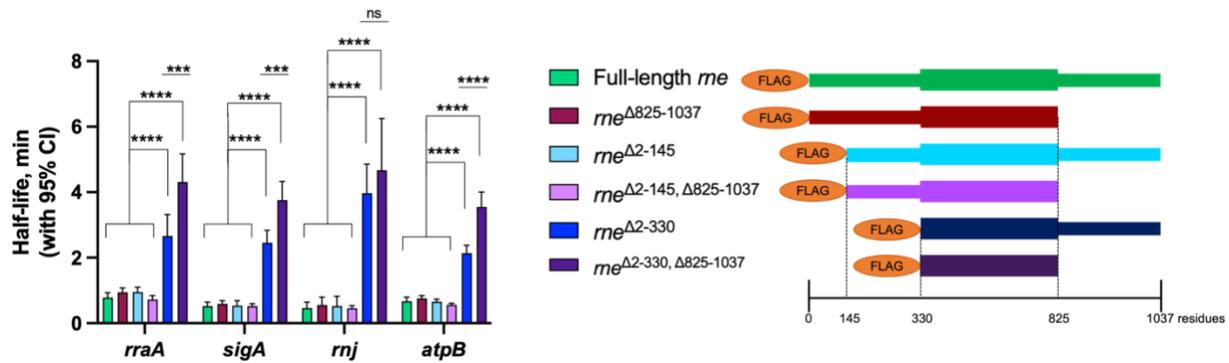
**Figure 4-4: Differential expression analysis of strains expressing  $rne^{\Delta 2-330}$  and  $rne^{\Delta 2-330, \Delta 825-1037}$  comparing to full-length *rne*.** Volcano plots highlighting differentially expressed genes between (A)  $rne^{\Delta 2-330}$  and full-length *rne* and (B)  $rne^{\Delta 2-330, \Delta 825-1037}$  and full-length *rne*. The genes are colored if they passed the thresholds for statistical significance (adjusted  $P$  value  $\leq 0.05$ ) and  $\log_2$  fold change  $\geq 1$  or  $\leq -1$ . Red indicates overexpression and blue indicates underexpression.

Finally, we found the *glpFKD* operon was underexpressed in  $rne^{\Delta 2-330}$  and two of these three genes were also underexpressed in  $rne^{\Delta 2-330, \Delta 825-1037}$  (Table 4-5). In *M. smegmatis*, the *glpFKD* operon is involved in glycerol catabolism (Bong et al., 2019). We also checked two cyclic AMP

receptor protein genes (Crp; MSMEG\_6189 and MSMEG\_0539) (Sharma et al., 2014) as MSMEG\_6189 was shown to negatively regulate *glpFKD* (Bong et al., 2019). MSMEG\_0539 was underexpressed in both mutant strains (-1.58 log<sub>2</sub>FoldChange in *rne*<sup>Δ2-330, Δ825-1037</sup> and -1.53 log<sub>2</sub>FoldChange in *rne*<sup>Δ2-330</sup>) while MSMEG\_6189 was not (-0.73 log<sub>2</sub>FoldChange in *rne*<sup>Δ2-330, Δ825-1037</sup> and -2.0 log<sub>2</sub>FoldChange in *rne*<sup>Δ2-330</sup>, but not meeting the p value cut-off). The decreased expression of the *glpFKD* operon therefore cannot be explained by overexpression of its negative regulator MSMEG\_6189. However, our data would be consistent with a scenario in which MSMEG\_0539 positively regulates the *glpFKD* operon or is co-regulated with the *glpFKD* operon by another regulator. Previous studies showed that MSMEG\_0539 and MSMEG\_6189 are two functional paralogs Crp and have different biochemical properties in *M. smegmatis* (Aung et al., 2015; Sharma et al., 2014). Our data suggest that the roles of these two Crp orthologs might be different. Moreover, considering the important role of RNase E in mRNA degradation and previous findings about the tight correlation between mRNA stability and energy metabolism (Vargas-Blanco et al., 2019), the scaffold domains of RNase E might be involved in coordinating the regulation of RNA metabolism and energy metabolism in *M. smegmatis*.

### **Deletion of the N-terminal 330 residues of RNase E leads slower mRNA degradation**

To determine if deletion of RNase E scaffold domains affected mRNA degradation rates, we measured the half-lives of four genes (*rraA*, *sigA*, *rnj*, and *atpB*) in all of our scaffold domain mutants. While deletion of residues 2-145, residues 825-1037, or both in combination did not affect mRNA degradation rates, deletion of residues 2-330 residues caused slower mRNA degradation and the combination of this deletion with deletion of the C-terminal scaffold domain slowed mRNA degradation even more (Figure 5-4). Taken together, this indicates that the N-terminal scaffold domain is required for efficient mRNA degradation and that the N-terminal and C-terminal domains may have partially redundant functions.



**Figure 4-5: Deletion of the N-terminal 330 residues of RNase E leads to slower mRNA degradation.**

Transcript half-lives for the indicated genes were measured for RNase E mutant strains by blocking transcription with 150  $\mu$ g/mL rifampicin and measuring RNA abundance at several timepoints by quantitative PCR. Constructs of RNase E mutants are shown in the right. Error bars denote 95% CI. Half-lives were compared using linear regression analysis,  $n=3$ . \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ; ns,  $P > 0.05$ .

## Discussion

While RNase E is well characterized in *E. coli*, less is known about this essential enzyme in mycobacteria. Here we sought to increase understanding of the roles of the scaffold domains in *M. smegmatis* RNase E. Our LC-MS/MS results revealed that the proteins associated with RNase E in *Mtb* and *M. smegmatis* have a high similarity (Table 4-1 and (Plocinski et al., 2019)), indicating that the key components of RNA degradosomes in these two species may be the same, as well as the potential RNA binding proteins. The co-localization of RNase E-mCherry and RNase J-Dendra from the live cell microscopy also confirms the interaction between RNase E and RNase J in live cells (Figure 4-2A). Taken together, our data support the view that RNase J is a key degradosome component interacting with RNase E in mycobacteria. Our findings also provide insights into a role of RNase E scaffold domains in the interaction with RNase J (Figure 4-2B). The localization of RNase J-Dendra was significantly diffused when the full N-terminal and C-terminal scaffold domains of RNase E were deleted, indicating that RNase J localization is partially dependent on the RNase E scaffold domains and that residues 146-330 of the N-terminal scaffold are more important for this function than residues 2-145 (Figure 4-3A-E). However, we

do not know the exact binding sites for RNase J within the necessary RNase E scaffold domain regions. Also, we cannot exclude the existence of indirect interactions between RNase E and RNase J.

Another interesting finding about RNase E in *M. smegmatis* is its localization. In *E. coli*, RNase E binds to the inner face of the cell membrane (Khemici et al., 2008) and its subcellular localization is important for the efficiency of global mRNA degradation (Hadjeras et al., 2019). In contrast, our data show that RNase E in *M. smegmatis* is localized throughout the cell, consistent with the fact that a 15-residue membrane targeting sequence (MTS) located in the non-catalytic of RNase E in *E. coli* is absent from *M. smegmatis* RNase E. A future study of the correlation between RNase E protein localization and mRNA localization would shed more light on how and where RNase E recruits mRNA for degradation.

We have shown here that the catalytic domain of RNase E is centrally located, between N- and C-terminal scaffold domains (Figure 1A). This is consistent with previous work on RNase E showing that the Mtb RNase E catalytic domain was in the central region which starts from residues 332 (corresponding to *M. smegmatis* RNase E residue 384) (Zeller et al., 2007). Previous studies showed that RNase E is essential in mycobacteria, as it is in *E. coli* (Sasseti et al., 2003; Sasseti & Rubin, 2003; Taverniti et al., 2011). However, the scaffold domains are not evolutionarily conserved and not necessary for cell growth. The single deletion of the C-terminal scaffold domain in *M. smegmatis* RNase E did not induce any distinct changes in cell phenotype and only affected expression of a few genes, suggesting that it does not make a unique, important contribution to RNase E function in the growth conditions we used as long as the N-terminal domain is present. In contrast, the deletion of the N-terminal 330 residues, comprising most of the N-terminal scaffold domain, caused a substantial cell growth defect, cell elongation, slowed mRNA degradation, and significant gene expression changes for many genes. Interestingly,

deletion of the C-terminal domain largely restored the growth defect of the *rne*<sup>Δ2-330</sup> strain, while causing the opposite cell size defect and exacerbating the mRNA degradation defect. The gene expression changes responsible for these phenotypes are unclear. And, we do not know how the combination of deletion of C-terminal scaffold domain with *rne*<sup>Δ2-330</sup> restores cell growth. A further biological function analysis will be helpful to explain the severe phenotype of the *rne*<sup>Δ2-330</sup> strain. Nevertheless, as WhiB2 overexpression from our RNA-seq analysis is only found in *rne*<sup>Δ2-330, Δ825-1037</sup> and considering its important role in mycobacterial septum formation and cell division (Gomez & Bishai, 2000), the shorter cell phenotype could potentially be explained by this expression change. Taken together, all our phenotypic observations from RNase E mutants suggest an important role of RNase E in cell growth and division.

Our RNA-seq analysis identified a link between RNase E and glycerol metabolism, as the *glpFkD* operon and one of its regulators were under-expressed in *rne*<sup>Δ2-330</sup> and *rne*<sup>Δ2-330, Δ825-1037</sup> (Table 3). This is not the first time that RNase E was identified in relation to energy metabolism. Enolase and aconitase, enzymes in glycolysis and the TCA cycle, respectively, have been identified as members of the RNase E degradosomes of *E. coli* and *Caulobacter crescentus*, respectively (Hardwick et al., 2011; Miczak et al., 1996). While we did not find any glycolytic enzymes in our RNase E pull-down, it is possible that such interactions between RNase E and energy metabolism enzymes do occur in mycobacteria and did not appear in our dataset for technical reasons. Future studies should investigate the interaction of degradosomes with components of energy pathways in mycobacteria.

## Materials and Methods

### Strains and Growth Conditions Used in This Study

The *M. smegmatis* strains used in this work are derived from strain mc<sup>2</sup>155. For *M. smegmatis* and its derivatives (Table S4-4), strains were grown in Middlebrook 7H9 medium with albumin-dextrose-catalase (ADC) supplementation (final concentration is 5g/liter bovine serum albumin fraction V, 2 g/liter dextrose, 0.85 g/liter NaCl, and 3 mg/liter catalase), 0.2% glycerol, and 0.05% Tween 80. Cultures were shaken at 200 rpm and grown at 37°C to log phase at the time of harvest.

The chromosomal His-3xFLAG tagged RNase E strain (SS-M\_0296) was built using a two-step strategy (See (Vargas-Blanco et al., 2019)).

His-3xFLAG tagged RNase E mutant strains were built in steps. First, we integrated the plasmid pSS267 containing the 436 nt upstream of the *rne* start codon (the TSS is at -236 relative to the start codon and the 200 nt upstream of the TSS were assumed to contain the promoter), with the His-3xFLAG coding sequence inserted after the start codon of *rne* into L5 integration site of wildtype mc<sup>2</sup>155, resulting in a strain (SS-M\_0367) with a second FLAG-tagged *rne* copy in the L5 site. Then plasmid pSS298 was built based on pJM1 and contained 1 kb of the sequence upstream of *rne*, the last 150bp of the *rne* coding sequence and 500 bp of the sequence downstream of *rne*. pSS298 was transformed into SS-M\_0367 and the colonies were selected with hygromycin and confirmed by sequencing. Counterselection with 15% sucrose was used to identify colonies that had successfully undergone a second crossover. We confirmed with PCR screening and DNA sequencing to get the strain (SS-M\_0461) in which pSS298 was lost and most of the *rne* coding sequence was deleted from the native locus. With SS-M\_0461, we were able to construct His-3xFLAG tagged RNase E mutant strains by transforming plasmids

containing *rne* variants (Table S4-4) into it and selected the positive colonies that were Nat<sup>R</sup> but not Kan<sup>R</sup>.

mCherry-tagged RNase E mutant strains were built as generally described for the FLAG-tagged constructs. We constructed a plasmid pSS418 containing the 436 nt upstream of *rne* start codon (TSS is at -236 relative to start codon and 200nt upstream of TSS were assumed to contain the promoter), *rne* coding sequence and mCherry coding sequence inserted before the stop codon of *rne* and integrated this plasmid into SS-M\_0461 Gile integration site. Then pSS322 (Dendra fused at the C-terminus of RNase J) was transformed into this strain to achieve a replacement of pSS267 with pSS322 at the L5 integration site, resulting SS-M\_0805. With SS-M\_0805, we have a hyg<sup>R</sup> plasmid containing mCherry tagged RNase E in Giles integration site and Dendra tagged RNase J in L5 integration site. We were able to transform plasmids containing mCherry tagged *rne* variants (Table S4-4) into it and selected the positive colonies that were Nat<sup>R</sup> but not Hyg<sup>R</sup>.

### **RNA extraction and determination of mRNA stability**

RNA extraction, cDNA synthesis, and quantitative PCR to determine mRNA half-lives were done as described in (Vargas-Blanco et al., 2019)

### **RNA-seq analysis**

Illumina-compatible RNAseq libraries were constructed using a protocol developed by the lab of Sarah Fortune that will be published elsewhere. Libraries were sequenced at the UMass Medical School Deep Sequencing core facility on an Illumina HiSeq instrument to obtain 100 nt paired end reads. Reads were aligned to the NC\_008596 reference genome using Burrows-Wheeler Aligner (Li & Durbin, 2009). The FeatureCounts tool was used to assign mapped reads to genomic features. DESeq2 (Love et al., 2014) was used to assess changes in gene expression in RNA

expression libraries. The DESeq2 package internally corrects for library size. The input count matrix for differential expression analysis were un-normalized counts, which allow assessing the measurement precision correctly. Genes with an adjusted  $p$ -value less than 0.05 and fold change  $\geq 2$  were regarded as differentially expressed.

### **Anti-FLAG immunoprecipitation and LC-MS/MS**

We performed anti-FLAG immunoprecipitation with SS-M\_0296 and instructions was followed as description in ANTI-FLAG M2 Magnetic Beads (Sigma-Aldrich M8823) and eluted with 3xFLAG peptide (Sigma-Aldrich F4799). A mock anti-FLAG immunoprecipitation control was performed with WT  $mc^2155$ . LC-MS/MS was performed on the eluates by the UMass Medical School Mass Spectrometry Facility. We used normalized intensity Based Absolute Quantitation (iBAQ) as a proxy to assess the relative difference between tagged RNase E immunoprecipitation and mock immunoprecipitation. Two independent immunoprecipitations were performed and sent for LC-MS/MS, and the results were pooled together for further analysis. Based on the ratio of normalized iBAQ between tagged strain and WT, we sorted out all the enriched proteins ( $> 3.0$  ratio) and listed in table 4-1.

### **Live cell microscopy and FIJI/ImageJ analysis**

Live cell images were taken with Zeiss Axio Imager fluorescence microscope with an apotome using a 63x/1.4 Oil DIC objective. For each channel, settings for image acquisition were identical among RNase E mutant strains. In the FIJI/ImageJ, a segmented line was manually added to each cell from one end to the other in DIC image. Then this segmented line was copied to the same position in the corresponding fluorescent channel image using Analyze > Tools > ROI manager. Intensity distributions from one end of the cell to other end were obtained using Analyze > Plot Profile.

## Funding

This work was supported by NSF CAREER award 1652756 to SS and by institutional startup funds from WPI.

## Acknowledgements

We thank members of the Shell lab and Louis Roberts for technical assistance and helpful discussions. LC-MS/MS was performed by the Mass Spectrometry Facility in UMass Chan Medical School. All Illumina sequencing was performed by the UMass Chan Medical School Deep Sequencing Core.

## References

- Ait-Bara, S., & Carpousis, A. J. (2010). Characterization of the RNA degradosome of *Pseudoalteromonas haloplanktis*: conservation of the RNase E-RhlB interaction in the gammaproteobacteria. *J Bacteriol*, *192*(20), 5413-5423. <https://doi.org/10.1128/JB.00592-10>
- Al-Husini, N., Tomares, D. T., Bitar, O., Childers, W. S., & Schrader, J. M. (2018). alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies. *Mol Cell*, *71*(6), 1027-1039 e1014. <https://doi.org/10.1016/j.molcel.2018.08.003>
- Apirion, D., & Lassar, A. B. (1978). A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *J Biol Chem*, *253*(5), 1738-1742. <https://www.ncbi.nlm.nih.gov/pubmed/342528>
- Aung, H. L., Dixon, L. L., Smith, L. J., Sweeney, N. P., Robson, J. R., Berney, M., Buxton, R. S., Green, J., & Cook, G. M. (2015). Novel regulatory roles of cAMP receptor proteins in fast-growing environmental mycobacteria. *Microbiology (Reading)*, *161*(Pt 3), 648-661. <https://doi.org/10.1099/mic.0.000015>
- Bong, H. J., Ko, E. M., Song, S. Y., Ko, I. J., & Oh, J. I. (2019). Tripartite Regulation of the glpFKD Operon Involved in Glycerol Catabolism by GylR, Crp, and SigF in *Mycobacterium smegmatis*. *J Bacteriol*, *201*(24). <https://doi.org/10.1128/JB.00511-19>
- Callaghan, A. J., Grossmann, J. G., Redko, Y. U., Ilag, L. L., Moncrieffe, M. C., Symmons, M. F., Robinson, C. V., McDowall, K. J., & Luisi, B. F. (2003). Quaternary structure and catalytic activity of the *Escherichia coli* ribonuclease E amino-terminal catalytic domain. *Biochemistry*, *42*(47), 13848-13855. <https://doi.org/10.1021/bi0351099>
- Callaghan, A. J., Redko, Y., Murphy, L. M., Grossmann, J. G., Yates, D., Garman, E., Ilag, L. L., Robinson, C. V., Symmons, M. F., McDowall, K. J., & Luisi, B. F. (2005). "Zn-link": a metal-sharing interface that organizes the quaternary structure and catalytic site of the endoribonuclease, RNase E. *Biochemistry*, *44*(12), 4667-4675. <https://doi.org/10.1021/bi0478244>

- Carpousis, A. J., Van Houwe, G., Ehretsmann, C., & Krisch, H. M. (1994). Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, 76(5), 889-900. [https://doi.org/10.1016/0092-8674\(94\)90363-8](https://doi.org/10.1016/0092-8674(94)90363-8)
- Casaregola, S., Jacq, A., Laoudj, D., McGurk, G., Margaron, S., Tempete, M., Norris, V., & Holland, I. B. (1994). Cloning and analysis of the entire Escherichia coli ams gene. ams is identical to hmp1 and encodes a 114 kDa protein that migrates as a 180 kDa protein. *J Mol Biol*, 238(5), 867. <https://doi.org/10.1006/jmbi.1994.1344>
- Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Volker, U., & Stulke, J. (2009). Novel activities of glycolytic enzymes in Bacillus subtilis: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics*, 8(6), 1350-1360. <https://doi.org/10.1074/mcp.M800546-MCP200>
- Coros, A., Callahan, B., Battaglioli, E., & Derbyshire, K. M. (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. *Mol Microbiol*, 69(4), 794-808. <https://doi.org/10.1111/j.1365-2958.2008.06299.x>
- Dosztanyi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Sci*, 27(1), 331-340. <https://doi.org/10.1002/pro.3334>
- Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D., & Putzer, H. (2005). Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic Acids Res*, 33(7), 2141-2152. <https://doi.org/10.1093/nar/gki505>
- Gey Van Pittius, N. C., Gamielien, J., Hide, W., Brown, G. D., Siezen, R. J., & Beyers, A. D. (2001). The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria. *Genome Biol*, 2(10), RESEARCH0044. <https://doi.org/10.1186/gb-2001-2-10-research0044>
- Gomez, J. E., & Bishai, W. R. (2000). whmD is an essential mycobacterial gene required for proper septation and cell division. *Proc Natl Acad Sci U S A*, 97(15), 8554-8559. <https://doi.org/10.1073/pnas.140225297>
- Hadjeras, L., Poljak, L., Bouvier, M., Morin-Ogier, Q., Canal, I., Coccain-Bousquet, M., Girbal, L., & Carpousis, A. J. (2019). Detachment of the RNA degradosome from the inner membrane of Escherichia coli results in a global slowdown of mRNA degradation, proteolysis of RNase E and increased turnover of ribosome-free transcripts. *Mol Microbiol*, 111(6), 1715-1731. <https://doi.org/10.1111/mmi.14248>
- Hardwick, S. W., Chan, V. S., Broadhurst, R. W., & Luisi, B. F. (2011). An RNA degradosome assembly in Caulobacter crescentus. *Nucleic Acids Res*, 39(4), 1449-1459. <https://doi.org/10.1093/nar/gkq928>
- Hillmann, D., Eschenbacher, I., Thiel, A., & Niederweis, M. (2007). Expression of the major porin gene mspA is regulated in Mycobacterium smegmatis. *J Bacteriol*, 189(3), 958-967. <https://doi.org/10.1128/JB.01474-06>
- Ikeda, Y., Yagi, M., Morita, T., & Aiba, H. (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Mol Microbiol*, 79(2), 419-432. <https://doi.org/10.1111/j.1365-2958.2010.07454.x>
- Jager, S., Fuhrmann, O., Heck, C., Hebermehl, M., Schiltz, E., Rauhut, R., & Klug, G. (2001). An mRNA degrading complex in Rhodobacter capsulatus. *Nucleic Acids Res*, 29(22), 4581-4588. <https://doi.org/10.1093/nar/29.22.4581>
- Judd, J. A., Canestrari, J., Clark, R., Joseph, A., Lapierre, P., Lasek-Nesselquist, E., Mir, M., Palumbo, M., Smith, C., Stone, M., Upadhyay, A., Wirth, S. E., Dedrick, R. M., Meier, C. G., Russell, D. A., Dills, A., Dove, E., Kester, J., Wolf, I. D., . . . Derbyshire, K. M. (2021). A Mycobacterial Systems Resource for the Research Community. *mBio*, 12(2). <https://doi.org/10.1128/mBio.02401-20>

- Kaberdin, V. R., Miczak, A., Jakobsen, J. S., Lin-Chao, S., McDowall, K. J., & von Gabain, A. (1998). The endoribonucleolytic N-terminal half of Escherichia coli RNase E is evolutionarily conserved in Synechocystis sp. and other bacteria but not the C-terminal half, which is sufficient for degradosome assembly. *Proc Natl Acad Sci U S A*, *95*(20), 11637-11642. <https://doi.org/10.1073/pnas.95.20.11637>
- Khemici, V., & Carpousis, A. J. (2004). The RNA degradosome and poly(A) polymerase of Escherichia coli are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol*, *51*(3), 777-790. <https://doi.org/10.1046/j.1365-2958.2003.03862.x>
- Khemici, V., Poljak, L., Luisi, B. F., & Carpousis, A. J. (2008). The RNase E of Escherichia coli is a membrane-binding protein. *Mol Microbiol*, *70*(4), 799-813. <https://doi.org/10.1111/j.1365-2958.2008.06454.x>
- Krishnamoorthy, G., Kaiser, P., Lozza, L., Hahnke, K., Mollenkopf, H. J., & Kaufmann, S. H. E. (2019). Mycofactocin Is Associated with Ethanol Metabolism in Mycobacteria. *mBio*, *10*(3). <https://doi.org/10.1128/mBio.00190-19>
- Laalami, S., Zig, L., & Putzer, H. (2014). Initiation of mRNA decay in bacteria. *Cell Mol Life Sci*, *71*(10), 1799-1828. <https://doi.org/10.1007/s00018-013-1472-4>
- Lee, K., & Cohen, S. N. (2003). A Streptomyces coelicolor functional orthologue of Escherichia coli RNase E shows shuffling of catalytic and PNPase-binding domains. *Mol Microbiol*, *48*(2), 349-360. <https://doi.org/10.1046/j.1365-2958.2003.03435.x>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lopez, P. J., Marchand, I., Joyce, S. A., & Dreyfus, M. (1999). The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol Microbiol*, *33*(1), 188-199. <https://doi.org/10.1046/j.1365-2958.1999.01465.x>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T., & Condon, C. (2007). 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell*, *129*(4), 681-692. <https://doi.org/10.1016/j.cell.2007.02.051>
- Mathy, N., Hebert, A., Mervelet, P., Benard, L., Dorleans, A., Li de la Sierra-Gallay, I., Noirot, P., Putzer, H., & Condon, C. (2010). Bacillus subtilis ribonucleases J1 and J2 form a complex with altered enzyme behaviour. *Mol Microbiol*, *75*(2), 489-498. <https://doi.org/10.1111/j.1365-2958.2009.07004.x>
- McDowall, K. J., & Cohen, S. N. (1996). The N-terminal domain of the rne gene product has RNase E activity and is non-overlapping with the arginine-rich RNA-binding site. *J Mol Biol*, *255*(3), 349-355. <https://doi.org/10.1006/jmbi.1996.0027>
- Miczak, A., Kaberdin, V. R., Wei, C. L., & Lin-Chao, S. (1996). Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proc Natl Acad Sci U S A*, *93*(9), 3865-3869. <https://doi.org/10.1073/pnas.93.9.3865>
- Ono, M., & Kuwano, M. (1979). A conditional lethal mutation in an Escherichia coli strain with a longer chemical lifetime of messenger RNA. *J Mol Biol*, *129*(3), 343-357. [https://doi.org/10.1016/0022-2836\(79\)90500-x](https://doi.org/10.1016/0022-2836(79)90500-x)
- Plocinski, P., Macios, M., Houghton, J., Niemiec, E., Plocinska, R., Brzostek, A., Slomka, M., Dziadek, J., Young, D., & Dziembowski, A. (2019). Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of Mycobacterium tuberculosis. *Nucleic Acids Res*, *47*(11), 5892-5905. <https://doi.org/10.1093/nar/gkz251>

- Purusharth, R. I., Klein, F., Sulthana, S., Jager, S., Jagannadham, M. V., Evgenieva-Hackenberg, E., Ray, M. K., & Klug, G. (2005). Exoribonuclease R interacts with endoribonuclease E and an RNA helicase in the psychrotrophic bacterium *Pseudomonas syringae* Lz4W. *J Biol Chem*, *280*(15), 14572-14578. <https://doi.org/10.1074/jbc.M413507200>
- Py, B., Higgins, C. F., Krisch, H. M., & Carpousis, A. J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature*, *381*(6578), 169-172. <https://doi.org/10.1038/381169a0>
- Rosana, A. R., Whitford, D. S., Fahlman, R. P., & Owtrim, G. W. (2016). Cyanobacterial RNA Helicase CrhR Localizes to the Thylakoid Membrane Region and Cosediments with Degradosome and Polysome Complexes in *Synechocystis* sp. Strain PCC 6803. *J Bacteriol*, *198*(15), 2089-2099. <https://doi.org/10.1128/JB.00267-16>
- Sasseti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, *48*(1), 77-84. <https://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Sasseti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A*, *100*(22), 12989-12994. <https://doi.org/10.1073/pnas.2134250100>
- Shahbabian, K., Jamali, A., Zig, L., & Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *EMBO J*, *28*(22), 3523-3533. <https://doi.org/10.1038/emboj.2009.283>
- Sharma, R., Zaveri, A., Gopalakrishnapai, J., Srinath, T., Varshney, U., & Visweswariah, S. S. (2014). Paralogous cAMP receptor proteins in *Mycobacterium smegmatis* show biochemical and functional divergence. *Biochemistry*, *53*(49), 7765-7776. <https://doi.org/10.1021/bi500924v>
- Stahl, C., Kubetzko, S., Kaps, I., Seeber, S., Engelhardt, H., & Niederweis, M. (2001). MspA provides the main hydrophilic pathway through the cell wall of *Mycobacterium smegmatis*. *Mol Microbiol*, *40*(2), 451-464. <https://doi.org/10.1046/j.1365-2958.2001.02394.x>
- Stoppel, R., Manavski, N., Schein, A., Schuster, G., Teubner, M., Schmitz-Linneweber, C., & Meurer, J. (2012). RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast. *Nucleic Acids Res*, *40*(17), 8593-8606. <https://doi.org/10.1093/nar/gks613>
- Taraseviciene, L., Bjork, G. R., & Uhlin, B. E. (1995). Evidence for an RNA binding region in the *Escherichia coli* processing endoribonuclease RNase E. *J Biol Chem*, *270*(44), 26391-26398. <https://doi.org/10.1074/jbc.270.44.26391>
- Taverniti, V., Forti, F., Ghisotti, D., & Putzer, H. (2011). *Mycobacterium smegmatis* RNase J is a 5'-3' exo-endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol*, *82*(5), 1260-1276. <https://doi.org/10.1111/j.1365-2958.2011.07888.x>
- Van den Bossche, A., Hardwick, S. W., Ceysens, P. J., Hendrix, H., Voet, M., Dendooven, T., Bandyra, K. J., De Maeyer, M., Aertsens, A., Noben, J. P., Luisi, B. F., & Lavigne, R. (2016). Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome. *Elife*, *5*. <https://doi.org/10.7554/eLife.16413>
- Vargas-Blanco, D. A., Zhou, Y., Zamalloa, L. G., Antonelli, T., & Shell, S. S. (2019). mRNA Degradation Rates Are Coupled to Metabolic Status in *Mycobacterium smegmatis*. *mBio*, *10*(4). <https://doi.org/10.1128/mBio.00957-19>
- WHO. (2021). WHO Global Tuberculosis Report. *World Health Organization*.
- Zeller, M. E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T., & Kaberdin, V. R. (2007). Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem J*, *403*(1), 207-215. <https://doi.org/10.1042/BJ20061530>

Zhang, J. Y., Deng, X. M., Li, F. P., Wang, L., Huang, Q. Y., Zhang, C. C., & Chen, W. L. (2014). RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region. *RNA*, 20(4), 568-579. <https://doi.org/10.1261/rna.043513.113>

## Tables

**Table 4-1: Proteins identified in greater abundance from RNase E immunoprecipitation followed by LC-MS/MS compared to a mock immunoprecipitation.**

Abundance ratio from tagged RNase E compared to WT	Protein	<i>M. smegmatis</i> gene ID	Mtb gene ID
37941000	Conserved hypothetical protein	MSMEG_5691	Rv0877
31037002	Bacterial proteasome-activating AAA-ATPase	MSMEG_3902	Rv2115c
6927016	Transcription termination factor Rho	MSEMG_4954	Rv1297
4347232	RNA binding protein Jag	MSMEG_6941	Rv3920c
3414000	KH domain RNA binding protein YlqC	MSMEG_2436	Rv2908c
3364402	Ssu ribosomal protein S1p	MSMEG_3833	Rv1630
1640750	DNA-binding protein HU	MSMEG_2389	Rv2986c
1352250	Integration host factor MihF	MSMEG_3050	Rv1388
911935	DNA-directed RNA polymerase, beta subunit, RpoB/RpoC	MSMEG_1367	Rv0667
901055	Putative cystathionine gamma-synthase	MSMEG_5265	Rv1079
770435	Lipoprotein	MSMEG_6064	/
677524	Aspartate-semialdehyde dehydrogenase	MSMEG_6256	Rv3708c
552203	DNA-directed RNA polymerase, beta subunit, RpoB/RpoC	MSMEG_1368	Rv0668
399280	Conserved hypothetical protein	MSMEG_5715	/
376327	Negative regulator of genetic competence ClpC/mecB	MSMEG_6091	Rv3596c
325135	RNase J	MSMEG_2685	Rv2752c
213033	Methylcrotonyl-CoA carboxylase, carboxyl transferase subunit	MSMEG_4717	Rv2502c
168116	Pyridine nucleotide-disulphide oxidoreductase	MSMEG_1416	Rv0688
155902	DNA Gyrase, A subunit	MSMEG_0006	Rv0006

140532	Sigma factor MysA	MSMEG_2758	Rv2703
106064	DEAD/DEAH box helicase	MSMEG_1930	Rv3211
87569	Acyl-CoA dehydrogenase domain protein	MSMEG_6511	/
72618	Alpha oxoglutarate ferredoxin oxidoreductase, beta subunit	MSMEG_4645	Rv2454c
51502	Inosine-5'-monophosphate dehydrogenase	MSMEG_1602	Rv3411c
38	PNPase	MSMEG_2656	Rv2783c
10	Acyl-CoA dehydrogenase domain protein	MSMEG_6512	/
6	Acyl carrier protein	MSMEG_4326	Rv2244
5	NAD(P) transhydrogenase, beta subunit	MSMEG_0109	/
4	Thiamine biosynthesis protein ThiC	MSMEG_0826	Rv0423c
4	Chaperone protein DnaK	MSMEG_0709	Rv0350
4	Putative non-ribosomal peptide synthase	MSMEG_0401	/

**Table 4-2: Expression level of ESX-1 genes in *rne*<sup>Δ2-330, Δ825-1037</sup> and *rne*<sup>Δ2-330</sup> strains relative to the full-length *rne* strain.**

	<i>Rne</i> <sup>Δ2-330, Δ825-1037</sup>		<i>rne</i> <sup>Δ2-330</sup>		Gene_Product
	log2FoldChange	adjusted <i>p</i> value	log2FoldChange	adjusted <i>p</i> value	
MSMEG_0055	5.71	1.88E-28	4.34	2.28E-07	EspE
MSMEG_0056	3.05	4.39E-04	3.45	5.44E-05	Hypothetical protein espF
MSMEG_0057	2.11	0.042	1.69	0.045	ESX-1 secretion-associated protein espG1
MSMEG_0059	1.78	0.001	1.34®	0.138®	ATPase, AAA family protein eccA1
MSMEG_0063	3.52	1.35E-16	2.25	3.48E-03	PE family protein PE35
MSMEG_0064	3.76	2.93E-14	1.86®	0.058®	PPE family protein
MSMEG_0065	2.11	1.75E-04	0.85®	0.499®	ESAT-6-like protein EsxB
MSMEG_0069	1.69	0.009	-0.17®	0.975®	Translation initiation factor IF-2 protein
MSMEG_0076	3.01	1.54E-11	1.83	0.004	Antigen MTB48
MSMEG_0078	1.82	0.006	0.37®	0.930®	EspA_EspE domain-containing protein
MSMEG_0082	2.40	2.93E-04	0.93®	0.393®	ESX conserved component eccE1
MSMEG_0083	2.85	3.96E-05	2.63	2.13E-04	Membrane-anchored mycosin mycp1

®Either log2FoldChange or adjusted *p* value did not meet the significance criteria in our analysis.

**Table 4-3: Expression levels of mycofactocin genes and their regulator in *rne*<sup>Δ2-330, Δ825-1037</sup> and *rne*<sup>Δ2-330</sup> relative to full-length *rne*.**

	<i>Rne</i> <sup>Δ2-330, Δ825-1037</sup>		<i>rne</i> <sup>Δ2-330</sup>		Gene_Product
	log2FoldChange	adjusted <i>p</i> value	log2FoldChange	adjusted <i>p</i> value	
MSMEG_6242	-7.87	0.046	-6.13	7.6174E-72	Alcohol dehydrogenase, iron-containing
MSMEG_1421	-2.43	0.812®	-3.87	0.310®	Mycofactocin precursor peptide MftA
MSMEG_1422	-2.98	0.130®	-2.56	0.005	Peptide chaperone MftB
MSMEG_1423	-1.67	0.033	-2.07	3.2604E-07	Mycofactocin maturase MftC
MSMEG_1424	-1.76	0.001	-2.30	1.02E-08	Pre-mycofactocin synthase MftD
MSMEG_1425	-1.73	0.026	-1.81	0.018	Mycofactocin precursor peptide peptidase MftE
MSMEG_1426	0.36®	0.991®	-0.58®	0.895®	Pre-mycofactocin glycosyltransferase MftF

® Either log2FoldChange or adjusted *p* value does not meet the significance criteria in our analysis.

**Table 4-4: Expression levels of genes differentially expressed in *rne*<sup>Δ2-330</sup> but not in *rne*<sup>Δ2-</sup>**

330, Δ825-1037

	log2FoldChange	adjusted <i>p</i> value	Gene_product
MSMEG_0019	2.304004557	0.00134	Amino acid adenylation
MSMEG_0131	-1.862538402	0.00107	AMP-binding enzyme, putative
MSMEG_0172	2.344591974	0.00928	Probable conserved transmembrane protein, putative
MSMEG_0393	-1.773730902	0.01248	Fmt protein
MSMEG_0399	-1.624660355	0.00018	Conserved domain protein
MSMEG_0457	1.700825169	0.04924	DNA topoisomerase IV subunit B
MSMEG_0523	5.019675117	0.03363	DNA-binding protein
MSMEG_0549	-1.460956378	0.04533	ABC transporter, permease protein
MSMEG_0641	-1.256773996	0.02563	Binding-protein-dependent transport systems inner membrane component
MSMEG_0672	3.20388218	0.00014	Conserved hypothetical protein
MSMEG_0674	1.392906045	0.01941	ErfK/YbiS/YcfS/YnhG family protein
MSMEG_0694	1.505669633	0.03187	N/A
MSMEG_0757	4.543446594	0.00659	Hypothetical protein
MSMEG_0774	7.90441765	0.04944	N/A
MSMEG_0833	-1.323416871	0.00081	Conserved hypothetical protein
MSMEG_0871	8.531622378	0.00608	Putative aldehyde or xanthine dehydrogenase, molybdopterin binding subunit protein
MSMEG_0929	1.147922041	0.01019	ErfK/YbiS/YcfS/YnhG family protein
MSMEG_0945	-2.039702224	8.9E-05	Conserved domain protein
MSMEG_1287	7.965649221	0.03703	Cyclase/dehydrase superfamily protein
MSMEG_1324	-8.993029739	0.04481	Hypothetical protein
MSMEG_1343	-1.64326353	0.01248	tRNA-Trp
MSMEG_1397	-1.275306269	0.01603	Transcriptional regulator, TetR family protein
MSMEG_1422	-2.564790391	0.00468	Conserved hypothetical protein
MSMEG_1438	-1.461049781	5.2E-05	Ribosomal protein L23
MSMEG_1440	-1.049976511	0.03757	Ribosomal protein S19
MSMEG_1519	-1.112263715	0.02178	Translation initiation factor IF-1
MSMEG_1525	1.252947387	0.00027	50S ribosomal protein L17
MSMEG_1544	-3.683803101	0.00061	PduO protein
MSMEG_1546	-2.75636361	0.00414	Coenzyme B12-dependent glycerol dehydrogenase small subunit
MSMEG_1547	-3.336642203	1.4E-13	Glycerol dehydratase large subunit
MSMEG_1582	-1.899299778	9.3E-07	Chaperonin GroS

MSMEG_1669	1.236432415	0.01603	Succinate dehydrogenase, iron-sulfur protein
MSMEG_1679	1.333884979	0.00464	AmiB
MSMEG_1748	-21.26051033	1.5E-05	Conserved hypothetical protein
MSMEG_1812	-1.302087856	0.00526	Conserved hypothetical protein
MSMEG_1885	-1.359080316	0.00323	2Fe-2S iron-sulfur cluster binding domain protein
MSMEG_1950	1.962524605	0.02382	Conserved hypothetical protein
MSMEG_1951	2.274085532	0.03135	Conserved domain protein
MSMEG_1971	6.693641948	0.02736	Propane monooxygenase hydroxylase large subunit
MSMEG_2083	-2.366446176	0.0465	Inositol monophosphatase
MSMEG_2128	7.251696747	0.03537	Malonyl CoA decarboxylase
MSMEG_2236	8.029897042	0.03068	Putative thiolase
MSMEG_2347	6.125094176	0.01248	Phytoene dehydrogenase
MSMEG_2438	-1.286389374	0.03542	tRNA (guanine-N1)-methyltransferase
MSMEG_2531	8.242503355	0.01328	GntR family protein transcriptional regulator
MSMEG_2532	-2.226842319	0.01597	Dehydroquinase dehydratase, type II
MSMEG_2662	-21.77231619	3.4E-07	Hypothetical protein
MSMEG_2913	7.426325879	0.02244	Hydrolase
MSMEG_2974	1.463760199	0.00323	Peptidyl-prolyl cis-trans isomerase, cyclophilin-type
MSMEG_2986	8.021855327	0.03154	Amidohydrolase, AtzE family protein
MSMEG_3034	-2.090835532	0.00069	Metallopeptidase, M24 family protein
MSMEG_3086	1.503869735	0.00025	Triosephosphate isomerase
MSMEG_3241	-6.927427744	0.01972	Conserved hypothetical protein
MSMEG_3250	-1.271488886	0.03703	ABC transporter, ATP-binding protein
MSMEG_3316	8.241281544	0.01866	Transporter, major facilitator family protein
MSMEG_3318	8.500673328	0.00464	Oxidoreductase
MSMEG_3477	1.714449039	0.01024	Possible inv protein
MSMEG_3586	-21.06742639	1.8E-05	Major facilitator family protein transporter
MSMEG_3610	-21.16864807	1.6E-05	Conserved hypothetical protein
MSMEG_3701	-21.51508748	1.8E-06	Conserved hypothetical protein
MSMEG_3771	-2.483584378	9E-05	Arginine repressor
MSMEG_3792	-1.333661878	0.00348	Ribosomal protein L35
MSMEG_3811	1.412172205	0.04052	Universal stress protein family protein, putative
MSMEG_4095	8.76446061	0.00884	Putative monooxygenase
MSMEG_4125	-3.282834988	0.00014	Oxidoreductase, short chain dehydrogenase/reductase family protein
MSMEG_4137	-21.53596759	1.7E-06	Hypothetical protein
MSMEG_4145	-2.088537912	0.00421	Cupin 2 protein
MSMEG_4458	8.417077849	0.0209	Transmembrane transport protein
MSMEG_4597	-1.549172373	0.01732	Putative conserved lipoprotein lpph

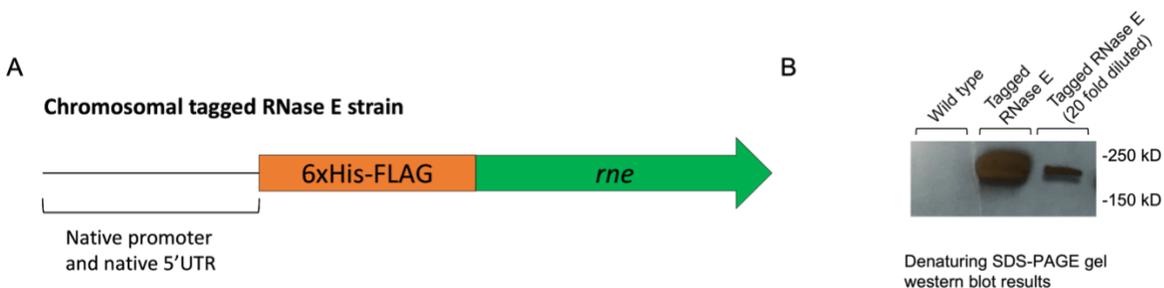
MSMEG_4715	1.709842813	0.03481	Acyl-CoA dehydrogenase
MSMEG_4920	1.372510384	0.03728	Acetyl-CoA acetyltransferase
MSMEG_5003	-8.732962719	0.00134	O-methyltransferase, family protein 3
MSMEG_5130	1.296453826	0.01248	conserved lipoprotein lpqW
MSMEG_5455	8.385813442	0.00753	PE family protein
MSMEG_5467	-4.937365446	0.005	tRNA-Ala
MSMEG_5763	8.759750621	0.00151	Conserved hypothetical protein
MSMEG_5822	7.896113681	0.00386	Conserved hypothetical protein
MSMEG_5948	8.238333469	0.01358	Glycosyl transferase
MSMEG_5969	-1.286420324	0.03435	Conserved hypothetical protein
MSMEG_6239	-2.147749706	0.00322	1,3-propanediol dehydrogenase
MSMEG_6253	1.904586038	9E-05	Fur family protein transcriptional regulator
MSMEG_6405	1.872484395	0.00464	Erp protein
MSMEG_6517	-21.51508748	1.8E-06	Hypothetical protein
MSMEG_6540	1.176711283	0.00753	Virulence factor Mce family protein
MSMEG_6646	8.311729552	0.01443	Methylisocitrate lyase
MSMEG_6732	-21.40654299	4.7E-06	Integral membrane protein
MSMEG_6758	-1.877438053	7.4E-05	Transport integral membrane protein
MSMEG_6768	9.156463015	0.00018	Halogenase
MSMEG_6804	-2.681516534	0.00051	Sugar ABC transporter substrate-binding protein
MSMEG_6850	7.981935131	0.03728	Alpha/beta hydrolase

**Table 4-5: Relative expression levels of glycerol metabolism enzymes in *rne*<sup>Δ2-330, Δ825-1037</sup> and *rne*<sup>Δ2-330</sup> relative to full-length *rne*.**

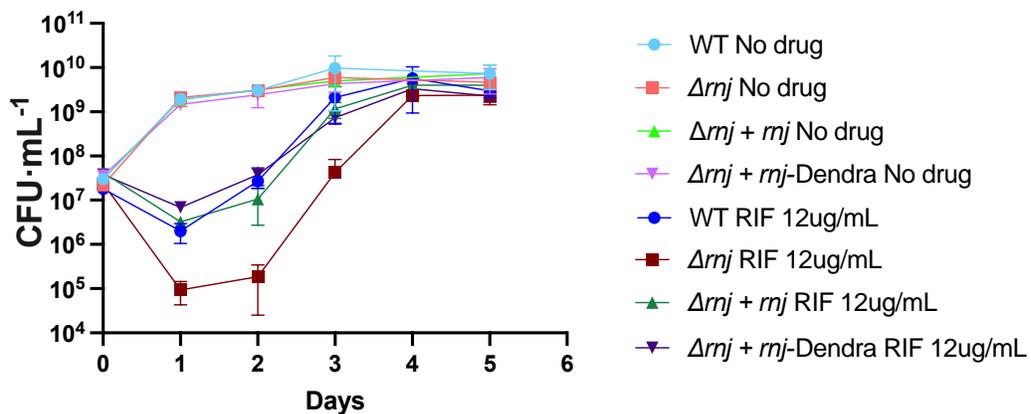
	<i>rne</i> <sup>Δ2-330, Δ825-1037</sup>		<i>rne</i> <sup>Δ2-330</sup>		Gene_Product
	log2FoldChange	adjusted <i>p</i> value	log2FoldChange	adjusted <i>p</i> value	
MSMEG_6758	-1.58®	0.852®	1.30	0.012	Glycerol kinase, glpF
MSMEG_6759	-1.44	0.017	-1.18	8.53E-04	Glycerol kinase, glpK
MSMEG_6761	-1.37	7.74E-04	-1.29	3.23E-04	Glycerol-3-phosphate dehydrogenase glpD
MSMEG_6189	-0.72®	0.666®	-0.20®	0.951®	Transcriptional regulator, Crp/Fnr family protein
MSMEG_0539	-1.58	0.045	-1.53®	0.051®	Transcriptional regulator, Crp/Fnr family protein

® Either log2FoldChange or adjusted *p* value does not meet the significance criteria in our analysis.

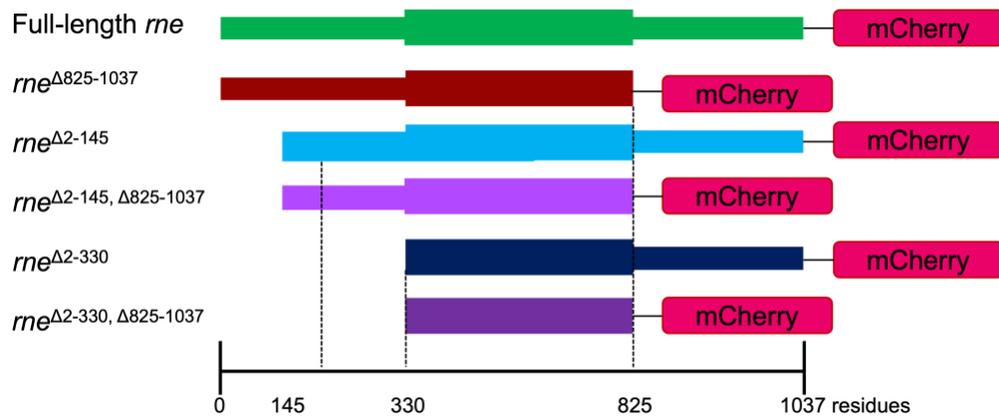
## Supplementary Figures



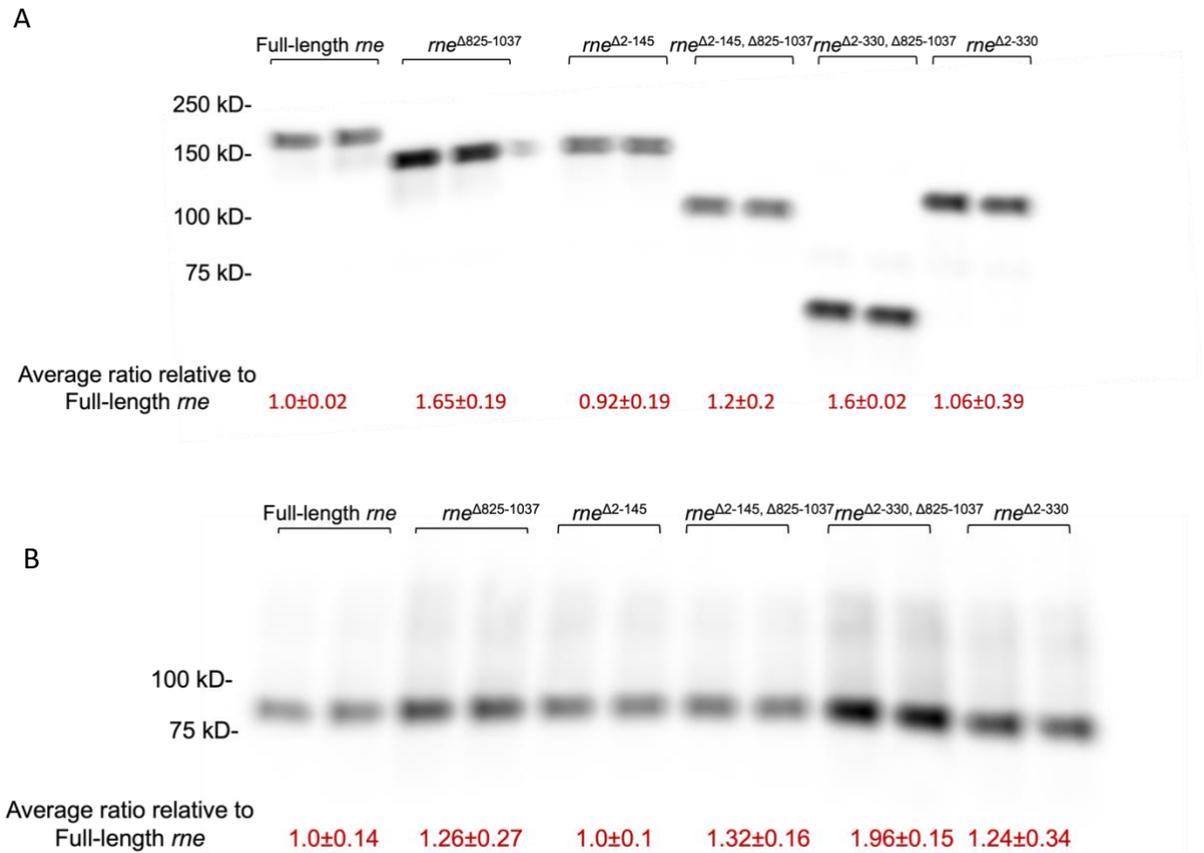
**Figure S4-1: Construction of a chromosomally His-FLAG tagged RNase E strain.** (A) *Rne* chromosomal locus in His-FLAG tagged RNase E strain. (B) Western blot (anti-FLAG) of tagged RNase E.



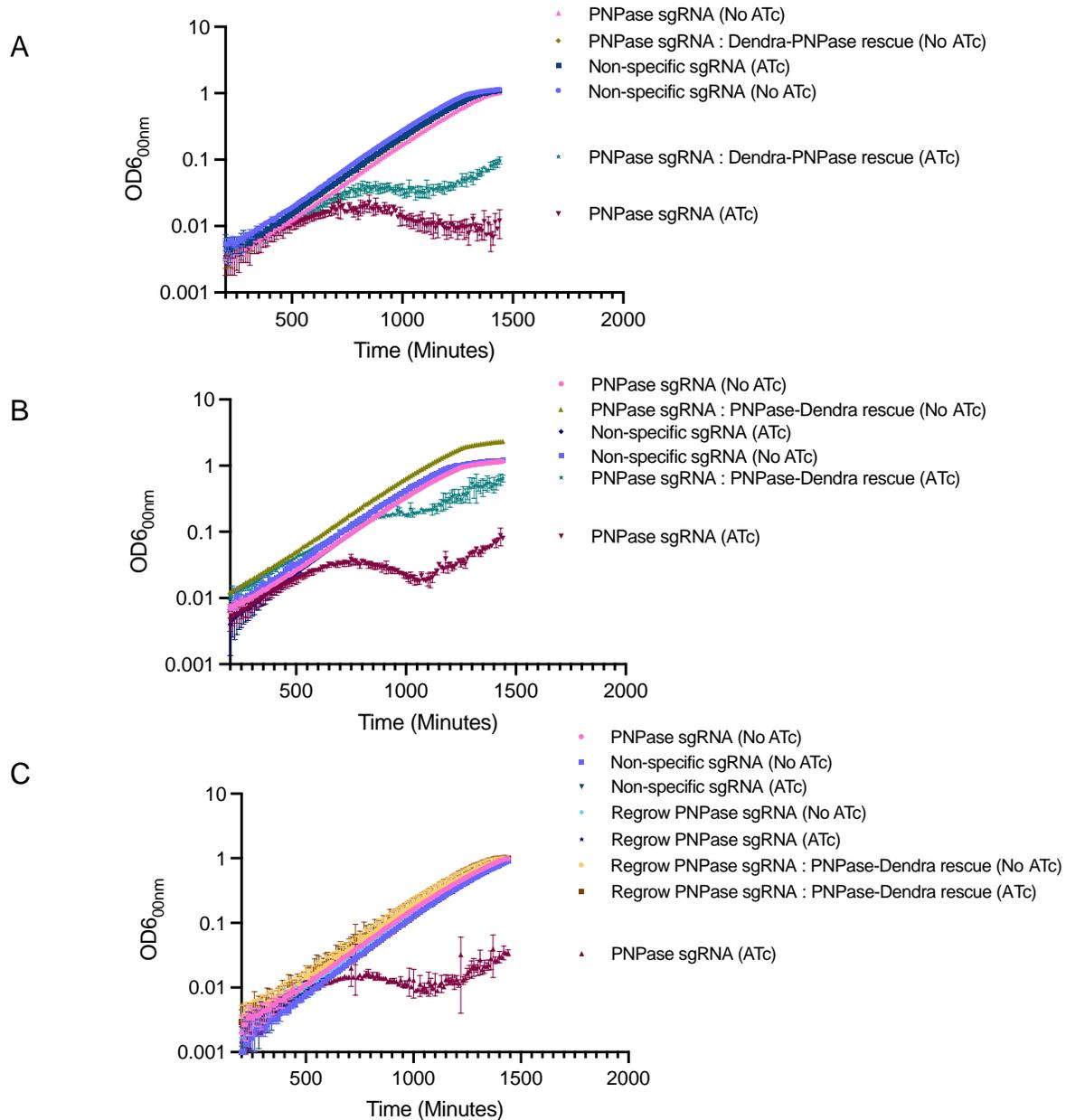
**Figure S4-2: Rifampicin killing assay to test the protein functionality of RNase J-Dendra.** WT,  $\Delta rnj$ ,  $\Delta rnj + rnj$ ,  $\Delta rnj + rnj$ -Dendra strains were grown in triplicate cultures and normalized to OD 0.1, then 12  $\mu$ g/mL rifampicin was added to each culture. No drug cultures were grown in parallel.



**Figure S4-3: Derivations of mCherry tagged RNase E.** mCherry was fused to the C-terminal end of RNase E mutant constructs.



**Figure S4-4: Expression of FLAG-RNase E and RNase J-Dendra in strains harboring both tagged RNases.** Western blot of (A) FLAG-RNase E and (B) RNase J-Dendra loaded with duplicate lysates from each RNase E mutant strain. The ratio of intensity for each strain was calculated by image J.



**Figure S4-5: Growth kinetics to test the protein function of tagged PNPase.** CRISPRi was used to knockdown PNPase expression as indicated. For PNPase-dendra rescue strain, the PAM in the dendra tagged PNPase coding sequence was mutated to avoid the specific binding of sgRNA for CRISPRi knockdown. Growth curves of CRISPRi PNPase, CRISPRi control and (A) Dendra tagged PNPase at the N-terminal in the Giles integration site in the CRISPRi PNPase strain to rescue PNPase, (B) Dendra tagged PNPase at the C-terminal in the Giles integration site in the CRISPRi PNPase strain to rescue PNPase. 200 ng/mL ATc was added to induce CRISPRi knock-down. (C) Cultures from (B) were regrown with or without Atc to confirm that the outgrowth of the PNPase-Dendra rescue was in fact due to loss of knockdown of the native PNPase. Three biological replicates of each strain were used for each experiment. The optical densities at 600nm ( $OD_{600nm}$ ) were measured and initial OD for each strain was normalized to 0.005.

## Supplementary Tables

**Table S4-1: Percentage of positive transformants during the construction of FLAG-tagged RNase E mutants**

Name	Total colonies #	Positive colonies #	Percentage positive
Full-length <i>rne</i>	32	22	68.75%
<i>rne</i> <sup>Δ825-1037</sup>	36	9	25.00%
<i>rne</i> <sup>Δ2-145</sup>	16	7	43.75%
<i>rne</i> <sup>Δ2-308</sup>	112	22	19.64%
<i>rne</i> <sup>Δ2-330</sup>	114	11	9.65%
<i>rne</i> <sup>Δ2-375</sup>	142	0	0
<i>rne</i> <sup>Δ2-145, Δ825-1037</sup>	10	3	30.00%
<i>rne</i> <sup>Δ2-308, Δ825-1037</sup>	115	10	8.70%
<i>rne</i> <sup>Δ2-330, Δ825-1037</sup>	114	20	17.54%
<i>rne</i> <sup>Δ2-375, Δ825-1037</sup>	140	0	0

**Table S4-2: Percentage of positive transformants during the construction of RNase E-mCherry mutants**

Name	Total colonies #	Positive colonies #	Percentage positive
Full-length <i>rne</i>	20	2	10.00%
<i>rne</i> <sup>Δ825-1037</sup>	56	2	3.57%
<i>rne</i> <sup>Δ2-145</sup>	228	0	0
<i>rne</i> <sup>Δ2-330</sup>	228	0	0
<i>rne</i> <sup>Δ2-145, Δ825-1037</sup>	228	0	0
<i>rne</i> <sup>Δ2-330, Δ825-1037</sup>	228	0	0

**Table S4-3: Genes differentially expression in *rne*<sup>Δ825-1037</sup> relative to full-length *rne***

	log2FoldChange	adjusted <i>p</i> value	Gene_Product
MSMEG_5380	-21.41	1.67E-05	tRNA-Leu
MSMEG_2842	-21.27	3.70E-05	Nitrilotriacetate monooxygenase component A
MSMEG_3522	-21.19	3.70E-05	Dopamine receptor D4
MSMEG_4137	-21.11	3.70E-05	Hypothetical protein
MSMEG_5299	-21.09	3.70E-05	Hypothetical protein
MSMEG_5531	-21.09	3.70E-05	ABC transporter ATP-binding protein
MSMEG_6517	-21.09	3.70E-05	Hypothetical protein
MSMEG_1412	-20.59	6.39E-05	Amino acid permease
MSMEG_6732	-20.59	6.39E-05	Integral membrane protein
MSMEG_2186	-19.43	2.87E-04	Conserved hypothetical protein
MSMEG_0413	-8.92	1.50E-02	Conserved hypothetical protein
MSMEG_5190	-8.87	1.81E-02	TetR-family protein transcriptional regulator
MSMEG_3304	-8.82	1.42E-02	Succinate semialdehyde dehydrogenase
MSMEG_5397	-8.71	3.64E-02	ATP-dependent DNA helicase RecQ
MSMEG_0757	4.78	4.86E-02	Hypothetical protein
MSMEG_5743	8.98	1.50E-02	Patatin
MSMEG_4424	9.40	4.86E-02	Endoribonuclease L-PSP
MSMEG_0774	9.92	2.73E-05	N/A
MSMEG_5033	20.65	3.35E-06	N/A
MSMEG_0718	22.26	1.61E-06	Acetyl-CoA synthetase
MSMEG_1551	22.55	1.10E-07	Hypothetical protein
MSMEG_4603	22.77	2.12E-08	N/A

**Table S4-4: Strains and plasmids**

#1 Plasmid	#1 Plasmid location	Resistance drug	Strain	Characteristics	#2 plasmid	#2 Plasmid location
-	-	-	SS-M_0296	His-3xFLAG tag between the start codon and coding sequence of chromosomal <i>rne</i>	-	-
pSS250	-	Hyg <sup>R</sup> , SacB	-	pJM1 plasmid containing 1kb sequence upstream and 1kb sequence downstream of the <i>rne</i> start codon, with the sequence encoding His-3xFLAG tag inserted after the <i>rne</i> start codon	-	-
pSS267	L5	Kan <sup>R</sup>	SS-M_0367	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and <i>rne</i> coding sequence in L5 site, full-length chromosomal <i>rne</i> copy is retained	-	-
pSS298	-	Hyg <sup>R</sup> , SacB	-	pJM1 plasmid containing 1kb sequence upstream of <i>rne</i> , last 150nt of <i>rne</i> coding sequence and 350bp sequence downstream of <i>rne</i>	-	-
pSS267	L5	Kan <sup>R</sup>	SS-M_0461	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS337	L5	Nat <sup>R</sup>	SS-M_0594	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and full-length <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS343	L5	Nat <sup>R</sup>	SS-M_0559	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #146-1037 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS351	L5	Nat <sup>R</sup>	SS-M_0562	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #2-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS348	L5	Nat <sup>R</sup>	SS-M_0565	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #146-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS371	L5	Nat <sup>R</sup>	SS-M_0682	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #331-1037 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS380	L5	Nat <sup>R</sup>	SS-M_0643	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #331-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS370	L5	Nat <sup>R</sup>	SS-M_0664	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #309-1037 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS379	L5	Nat <sup>R</sup>	SS-M_0643	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #309-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	-	-
pSS346	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #376-1037 <i>rne</i> coding sequence	-	-
pSS340	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #376-824 <i>rne</i> coding sequence	-	-
pSS322	-	Apramycin <sup>R</sup>	-	Plasmid containing UV15 promoter, <i>rnj</i> coding sequence tagged with Dendra	-	-

pSS418	Gile	Hyg <sup>R</sup> , Kan <sup>R</sup>	SS-M_0800	Plasmid containing native promoter, native 5'UTR, <i>rne</i> coding sequence tagged with mCherry in Giles site, another copy of <i>rne</i> is in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS267	L5
pSS418	Gile	Hyg <sup>R</sup> , Apramycin <sup>R</sup>	SS-M_0805	Plasmid containing native promoter, native 5'UTR, <i>rne</i> coding sequence tagged with mCherry in Giles site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS322	L5
pSS460	Gile	Nat <sup>R</sup> , Apramycin <sup>R</sup>	SS-M_0808	Plasmid containing native promoter, native 5'UTR, <i>rne</i> coding sequence tagged with mCherry in Giles site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS322	L5
pSS462	Gile	Nat <sup>R</sup> , Apramycin <sup>R</sup>	SS-M_0810	Plasmid containing native promoter, native 5'UTR, start codon, #146-824 <i>rne</i> coding sequence tagged with mCherry in Giles site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS322	L5
pSS463	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, #331-1037 <i>rne</i> coding sequence tagged with mCherry	-	-
pSS465	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, #331-824 <i>rne</i> coding sequence tagged with mCherry	-	-
pSS467	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, #2-824 <i>rne</i> coding sequence tagged with mCherry	-	-
pSS469	-	Nat <sup>R</sup>	-	Plasmid containing native promoter, native 5'UTR, start codon, #146-1037 <i>rne</i> coding sequence tagged with mCherry	-	-
pSS483	-	Hyg <sup>R</sup>	-	Plasmid containing UV15 promoter, <i>rnj</i> coding sequence tagged with Dendra	-	-
pSS337	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0854	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and full-length <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pSS343	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0860	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #146-1037 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pSS351	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0857	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #2-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pSS348	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0863	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #146-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pSS371	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0866	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #331-1037 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pSS380	L5	Nat <sup>R</sup> , Hyg <sup>R</sup>	SS-M_0869	Plasmid containing native promoter, native 5'UTR, start codon, His-3xFLAGtag and #331-824 <i>rne</i> coding sequence in L5 site, only 150 bp of chromosomal <i>rne</i> copy is retained	pSS483	Gile
pJR962	L5	Kan <sup>R</sup>	SS-M_0203	Plasmid containing non-specific sgRNA for CRISPRi knockdown control	-	-
pSS421	L5	Kan <sup>R</sup>	SS-M_0721	Plasmid containing PNPase-specific sgRNA for CRISPRi PNPase knockdown	-	-
pSS511	Gile	Hyg <sup>R</sup>	-	Plasmid containing Dendra tagged at the N-terminal of PNPase and the specific sgRNA binding site was mutated	pSS421	L5
pSS560	Gile	Hyg <sup>R</sup>	-	Plasmid containing Dendra tagged at the C-terminal of PNPase and the specific sgRNA binding site was mutated	pSS421	L5

**Table S4-5: Primers for qPCR**

Primer name	Gene	Directionality	Sequence 5'-3'
SSS697	<i>rraA</i> (msmeg_6439)	Forward	AACTACGGCGGCAAGAT
SSS698	<i>rraA</i> (msmeg_6439)	Reverse	GTCGAGAGGATCGACTTCAG
JR273*	<i>sigA</i> (msmeg_2758)	Forward	GACTACACCAAGGGCTACAAG
JR273*	<i>sigA</i> (msmeg_2758)	Reverse	TTGATCACCTCGACCATGTG
SSS706	<i>rnj</i> (msmeg_2685)	Forward	TCATCCTCTCATCGGGTTTC
SSS707	<i>rnj</i> (msmeg_2685)	Reverse	TTCGCGCTCAACCTTCT
SSS903	<i>atpB</i> (msmeg_4942)	Forward	TGTTTCGTGTTTCGTCTGCTAC
SSS904	<i>atpB</i> (msmeg_4942)	Reverse	CGGCTTGGCGAGTTCTT

\*Source: Rock et al., 2017.

## Chapter 5 : Conclusions and Future directions

## Defining the Transcriptional and Post-transcriptional Landscapes of *Mycobacterium smegmatis* in Aerobic Growth and Hypoxia

In *Mycobacterium smegmatis*, less was known about the transcriptional landscape. Thus, we aimed to characterize the transcriptome in *M. smegmatis* under exponential phase and hypoxia. We combined 5'-end-mapping and RNAseq expression profiling and used Gaussian mixture modeling analysis to map and classify high-confidence and medium-confidence transcription start sites (TSSs) and cleavage sites (CSs) in two different physiological conditions. We were able to identify a total of 6,090 high confidence TSSs and 4,054 of those were classified as primary TSSs (pTSSs) (Figure 2-1). Within these TSSs, we found an ANNNT motif located upstream as well as several other alternative motifs when ANNNT was missing (Figure 2-2). Interestingly, a large portion of pTSSs were leaderless transcripts and more than half of leadered transcripts lacked a consensus Shine-Dalgarno ribosome-binding site (SD) (Figure 2-3). This observation was similar to findings in *Mtb* (Cortes, Schubert et al. 2013, Shell, Wang et al. 2015), indicating they are prominent features of mycobacterial transcriptomes. In addition, the expression levels of those genes with canonical SD sequences were on average higher than leadered transcripts lacking SD motifs and leaderless transcripts. We also examined the differential abundance of TSSs under hypoxia stress conditions and found over 300 TSSs varied significantly (Figure 2-5). Further, a SigF binding motif was found in the promoter regions of 56 transcripts that had increased expression in hypoxia. Together with published work reporting increased expression of SigF in hypoxia in *Mtb* (Galagan, Minch et al. 2013, Iona, Pardini et al. 2016, Yang, Sha et al. 2018), these data suggest that SigF may play a key role on specific promoters in hypoxia responses in mycobacteria.

In addition to TSSs, we also precisely mapped over 3,000 RNA cleavage sites and performed an analysis about their presence and distribution in the transcriptome in *M. smegmatis*. We found a

strong preference for a cytosine in the +1 position in the sequence context of the CS, which was present in over 90% of mapped CSs (Figure 2-4). An enrichment of cleavage sites was found in 5' UTRs as well as the intergenic regions between two co-transcribed genes, indicating the importance of RNA cleavages in processing and maturation of mRNAs. Since RNase E is a major, essential player in mRNA processing and decay in *E. coli* (Sasseti, Boyd et al. 2003, Sasseti and Rubin 2003, Zeller, Csanadi et al. 2007, Taverniti, Forti et al. 2011), we hypothesized that RNase E was a major contributor in these mapped cleavage sites. Indeed, later work presented in chapter 3 confirmed the role of RNase E in generating these cleavage sites. In hypoxia, we showed that the RNA cleavage was markedly reduced (Figure 2-5). In mycobacteria, studies reported that mRNA levels were reduced, and mRNA half-lives were significantly increased in low oxygen conditions (Rustad, Minch et al. 2013, Ignatov, Salina et al. 2015), demonstrating a likely energy-saving mechanism to regulate transcript levels in stress conditions. Taken together, reduced RNA cleavage might be a potential way to control RNA stability during stress conditions.

Our work well characterized the transcriptomic landscape and doubled the number of mapped TSSs in *M. smegmatis*. In addition, our study was the first report showing the direct impact of SigF in mycobacteria under hypoxia conditions, and future work is needed to understand the mechanism underlying the activation of SigF as well as the significance of the SigF-related expression changes. We also found a novel feature of CSs with cytidine in the +1 position in *M. smegmatis*. When comparing to a previously reported RNase E cleavage site in the *furA-katG* transcript (Taverniti, Forti et al. 2011), we found a consistency in the sequence context in the cleavage sites. As the RNA cleavages were markedly reduced in hypoxia conditions, we also hypothesized that reducing the RNase E cleavages may be a strategy to stabilize the transcriptome in mycobacteria. To test this hypothesis, future experiments may need to focus on RNase E activities in different stress conditions.

## The dominant role of RNase E in shaping the *Mycolicibacterium smegmatis* transcriptome

As described above, we hypothesized that RNase E was the major ribonuclease producing the 5' monophosphorylated cleavage products that we mapped transcriptome-wide in *Mycolicibacterium smegmatis*. Considering the important role of RNase E reported in other organisms, we aimed to investigate the function of this enzyme in *M. smegmatis* in mRNA cleavage and degradation. We constructed a repressible *rne* strain in which the transcription of *rne* could be repressed by addition of ATc (Figure 3-1A). We also constructed a corresponding control strain (Figure 3-1A). We confirmed the essentiality of *rne* in *M. smegmatis* with the comparison of growth kinetics between two strains with or without ATc (Figure 3-1C). We used qPCR to measure mRNA half-lives of transcripts when we repressed *rne* expression and found longer half-lives for all tested genes (Figure 3-1D), indicating a potential global role of RNase E in mRNA degradation. Therefore, we constructed RNAseq libraries to measure mRNA half-lives transcriptome-wide. We tried to precisely define RNase E-specific effects on mRNA degradation by measuring half-lives on knockdown and control cultures that were in the same growth phase. (Figure 3-2). Taken together, our transcriptome-wide mRNA half-life experiment provides a powerful dataset for defining the role of RNase E in mRNA metabolism.

For the transcriptome-wide mRNA half-life analysis, we were able to determine high-confidence and medium-confidence half-lives when *rne* was normally expressed or repressed. Strikingly, the half-lives in the *rne* knockdown were increased by 2-fold or more for 3,624 genes, and some genes showed no measurable degradation (Figure 3-3A, B). Taken together, we concluded that RNase E played an important role in mRNA degradation for around 90% of the transcriptome.

During the analysis, we found a varied stabilization among genes in the *rne* knockdown condition. We wanted to investigate the factors that make some transcripts more sensitive than others. One of the factors we found was mRNA abundance, implying that more abundant transcripts may be more sensitive to RNase E (Figure 3-4A). We also observed a smaller effect of *rne* knockdown on leaderless genes comparing to leadered genes (Figure 3-4B). This might be explained by the intrinsic sensitivity of leadered genes to RNase E comparing to leaderless genes, but it is not clear yet. In both the leadered and leaderless categories, we found genes that were not affected when *rne* was repressed, indicating additional factors contribute to differential sensitivity to RNase E activity.

Lastly, we wanted to assess the hypothesis that RNase E is the major RNase producing cleavage sites we mapped previously (Chapter 2 and (Martini, Zhou et al. 2019)). We successfully mapped the 5' ends from processed rRNA products and cleaved products from an *in vitro* RNase E cleavage assay (Figure 3-5). With 5' RACE and 3' RACE, we found the RNase E cleaved at the sequence RN↓CNU, consistent with our previous findings described in Chapter 2 (Martini, Zhou et al. 2019). Taken together, we confirmed that RNase E is an important player in mRNA degradation, and it strongly preferred to cleave 5' of cytidines.

In this study we showed a dramatic effect of RNase E knockdown on mRNA degradation rates in *M. smegmatis*, however, we observed a variability in the extent of stabilization in the analysis. One of the potential explanations is the contributions from other RNases to mRNA degradation, such as the essential exoribonuclease PNPase.

From our *in vivo* (Martini, Zhou et al. 2019) and *in vitro* RNase E cleavage sites mapped in this study, we confirmed that RNase E has a strong preference for cleaving 5' of cytidines. We tried to interpret the results carefully, considering the effect from RNA-binding proteins and ribosomes

*in vivo* as well as the predicted secondary structure in the sequence context. RNase E direct-entry cleavage needs to be considered as it could perform *cis*-acting cleavage while stem-loops are found close to cleavage sites, which have been shown before (Schuck, Diwa et al. 2009, Kime, Clarke et al. 2014, Bandyra, Wandzik et al. 2018, Updegrave, Kouse et al. 2019). We also observed a different key residue in the active sites for RNA substrates recognition in RNase E in mycobacteria and *E. coli*, and suspected this is a potential reason underlying the observed cleavage preference difference between mycobacterial and *E. coli* RNase E.

## Defining the roles of the degradosome-scaffolding domains of RNase E in *Mycolicibacterium smegmatis*

Previous studies have reported RNase E-centered multi-protein complexes in bacteria, and proteins commonly bind to the scaffold domain of RNase E (Miczak, Kaberdin et al. 1996, Jager, Fuhrmann et al. 2001, Lee and Cohen 2003, Purusharth, Klein et al. 2005, Ait-Bara and Carpousis 2010, Hardwick, Chan et al. 2011, Stoppel, Manavski et al. 2012, Zhang, Deng et al. 2014, Rosana, Whitford et al. 2016, Van den Bossche, Hardwick et al. 2016, Plocinski, Macios et al. 2019). In mycobacteria, less was known about the boundaries of the RNase E scaffold domains and the roles of these scaffold domains. In this study, by using *M. smegmatis* as a model, we constructed His-3xFLAG tagged RNase E mutant strains with deletions in the predicted scaffold domains and investigated the impacts of scaffold domains in various categories, including cellular RNase E localization, gene expression and mRNA degradation. We constructed a chromosomal His-3xFLAG tagged RNase E strain to perform anti-FLAG immunoprecipitation assays, combined with LC-MS/MS, and we were able to enrich for RNase E-associated proteins. We found similar key degradosome components to those defined in *Mtb* ((Plocinski, Macios et al. 2019) and Table 4-1). We then aimed to define the boundaries of the scaffold domains of RNase E. We used IUPred2 to perform disorder prediction and multiple sequence alignment among several species

to assess the conservation of *M. smegmatis* RNase E. As predicted, the catalytic domain was located in the central region of RNase E and the N- and C- terminal regions were predicted to be disordered, similar to previous findings in Mtb (Zeller, Csanadi et al. 2007) (Figure 4-1A, B). With the further growth kinetics and cell length analysis, we found that the deletion of the N-terminal 330 residues of RNase E caused a severe growth defect and longer cell phenotype while the deletion of both N-terminal and C-terminal domain produced a shorter cell phenotype, but a similar growth rate as full-length *rne* (Figure 4-1C, D). We were also able to narrow the boundaries of the catalytic domain and provide a guidance for future research about RNase E protein structure.

Next, we were able to see co-localization of mCherry-tagged RNase E and Dendra-tagged RNase J (Figure 4-2A), confirming the association of RNase J with RNase E in *M. smegmatis*, similar to that in Mtb (Plocinski, Macios et al. 2019). Then we did investigation of the role of the RNase E scaffold domains in this association with RNase J. We did live cell imaging and observed that the intensity of RNase J-Dendra appeared to be more diffuse in addition to puncta when we deleted most of the scaffold domains (Figure 4-2B). To quantify the observation, we performed a statistical analysis of the variability in fluorescence signal intensity of RNase J-Dendra within cells. The function 'Plot Profile' from image J helped us to measure the intensity distribution along the cell (Figure 4-3A). The CVs of the intensity distributions of *rne*<sup>Δ2-330, Δ825-1037</sup> and *rne*<sup>Δ2-330</sup> were smaller than those of the other strains (Figure 4-3B), confirming the previously observed more diffuse pattern. Analyses of skewness (Figure 4-3D) and kurtosis (Figure 4-3E) also were consistent with this conclusion.

The importance of the N-terminal scaffold domain was further supported by our RNAseq analysis. Moreover, we observed that RNase E was not membrane-tethered in mycobacteria, unlike RNase E in *E. coli* (Khemici, Poljak et al. 2008), indicating an evolutionary divergence among RNase E homologs in bacteria.

In our RNAseq experiment, we aimed to characterize the transcripts that were affected by deletion of the scaffold domains of RNase E. The C-terminal scaffold domain was less important, as less transcripts were significantly affected (Table S4-3), and this conclusion was also supported by previous phenotypic observations (Figure 4-1). Compared to the C-terminal scaffold domain, the deletion of N-terminal 330 residues caused more transcripts to be over- and under-expressed (Figure 4-4A). Even more transcripts were affected when we deleted both the N-terminal and C-terminal domains (Figure 4-4B). We found overlapped genes between *rne*<sup>Δ2-330, Δ825-1037</sup> and *rne*<sup>Δ2-330</sup>, but also genes specifically affected in each strain. Overlapped genes included those from the ESX-1 secretion system (Table 4-2), Mycofactocins (Table 4-3) and glycerol metabolism (Table 4-4), indicating that these gene were affected by deletion of the N-terminal domain regardless of the presence or absence of the C-terminal domain. Other genes were found that could potentially explain the observed cell size phenotypes. Therefore, given the identified proteins in Table 4-1, we suspected that N-terminal scaffold domain is the major domain interacting with other proteins, but that the C-terminal domain clearly has a role as well. The cooperation between N-terminal and C-terminal scaffold domains was also supported by our half-life experiment, where deletion of both scaffold domain caused further mRNA stabilization beyond that of the N-terminal deletion alone, despite the C-terminal deletion alone not having a detectable effect (Figure 4-5).

RNase E was hypothesized to be a major player in mRNA degradation in mycobacteria, so a better understanding of the RNase E protein would help researchers to know its role in mRNA metabolism and interactions with other proteins as has been reported previously in *E. coli* (Lopez, Marchand et al. 1999). In our study, we were able to show the role of RNase E in affecting the localization pattern of RNase J, and we also tried to study the interaction between RNase E and PNPase. However, Dendra tagged PNPase was not fully functional (Figure S4-4) and we therefore did not wish to draw conclusions regarding its localization. Future studies may need to focus on other proteins that we pulled down in Table 4-1 to discover how RNase E function is

regulated. In addition, characterization of the RNA degradosome is a future direction to explore the mRNA degradation mechanisms in mycobacteria. Our FLAG-RNase E mutant strains will be good tools to perform immunoprecipitation to fulfill this purpose. Mapping the interacting sites of proteins with RNase E may help us to characterize the binding motifs that may exist in mycobacteria. In our RNAseq analysis, we found transcripts from the ESX-1 system were dramatically affected when RNase E was largely deleted, suggesting the possibility that the scaffold domains have a role in regulating these genes. Further experiments are needed to determine if the scaffold domains have binding sites for RNA binding proteins, which have been shown to facilitate the RNase E-mediated sRNA/mRNA pair degradation in post-transcriptional regulation in other bacteria (Masse, Escorcia et al. 2003, Van Assche, Van Puyvelde et al. 2015, Holmqvist and Vogel 2018). In addition, we also found transcripts from glycerol metabolism. This is an intriguing finding as future experiment could focus on the potential link between RNase E mediated post-transcriptional regulation and metabolic pathways in mycobacteria.

## References

- Ait-Bara, S. and A. J. Carpousis (2010). "Characterization of the RNA degradosome of *Pseudoalteromonas haloplanktis*: conservation of the RNase E-RhlB interaction in the gammaproteobacteria." *J Bacteriol* **192**(20): 5413-5423.
- Bandyra, K. J., J. M. Wandzik and B. F. Luisi (2018). "Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E." *Mol Cell* **72**(2): 275-285 e274.
- Cortes, T., O. T. Schubert, G. Rose, K. B. Arnvig, I. Comas, R. Aebersold and D. B. Young (2013). "Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*." *Cell Rep* **5**(4): 1121-1131.
- Galagan, J. E., K. Minch, M. Peterson, A. Lyubetskaya, E. Azizi, L. Sweet, A. Gomes, T. Rustad, G. Dolganov, I. Glotova, T. Abeel, C. Mahwinney, A. D. Kennedy, R. Allard, W. Brabant, A. Krueger, S. Jaini, B. Honda, W. H. Yu, M. J. Hickey, J. Zucker, C. Garay, B. Weiner, P. Sisk, C. Stolte, J. K. Winkler, Y. Van de Peer, P. Iazzetti, D. Camacho, J. Dreyfuss, Y. Liu, A. Dorhoi, H. J. Mollenkopf, P. Drogaris, J. Lamontagne, Y. Zhou, J. Piquenot, S. T. Park, S. Raman, S. H. Kaufmann, R. P. Mohn, D. Chelsky, D. B. Moody, D. R. Sherman and G. K. Schoolnik (2013). "The *Mycobacterium tuberculosis* regulatory network and hypoxia." *Nature* **499**(7457): 178-183.
- Hardwick, S. W., V. S. Chan, R. W. Broadhurst and B. F. Luisi (2011). "An RNA degradosome assembly in *Caulobacter crescentus*." *Nucleic Acids Res* **39**(4): 1449-1459.

Holmqvist, E. and J. Vogel (2018). "RNA-binding proteins in bacteria." Nat Rev Microbiol **16**(10): 601-615.

Ignatov, D. V., E. G. Salina, M. V. Fursov, T. A. Skvortsov, T. L. Azhikina and A. S. Kaprelyants (2015). "Dormant non-culturable Mycobacterium tuberculosis retains stable low-abundant mRNA." BMC Genomics **16**: 954.

Iona, E., M. Pardini, A. Mustazzolu, G. Piccaro, R. Nisini, L. Fattorini and F. Giannoni (2016). "Mycobacterium tuberculosis gene expression at different stages of hypoxia-induced dormancy and upon resuscitation." J Microbiol **54**(8): 565-572.

Jager, S., O. Fuhrmann, C. Heck, M. Hebermehl, E. Schiltz, R. Rauhut and G. Klug (2001). "An mRNA degrading complex in Rhodobacter capsulatus." Nucleic Acids Res **29**(22): 4581-4588.

Khemici, V., L. Poljak, B. F. Luisi and A. J. Carpousis (2008). "The RNase E of Escherichia coli is a membrane-binding protein." Mol Microbiol **70**(4): 799-813.

Kime, L., J. E. Clarke, A. D. Romero, J. A. Grasby and K. J. McDowall (2014). "Adjacent single-stranded regions mediate processing of tRNA precursors by RNase E direct entry." Nucleic Acids Res **42**(7): 4577-4589.

Lee, K. and S. N. Cohen (2003). "A Streptomyces coelicolor functional orthologue of Escherichia coli RNase E shows shuffling of catalytic and PNPase-binding domains." Mol Microbiol **48**(2): 349-360.

Lopez, P. J., I. Marchand, S. A. Joyce and M. Dreyfus (1999). "The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo." Mol Microbiol **33**(1): 188-199.

Martini, M. C., Y. Zhou, H. Sun and S. S. Shell (2019). "Defining the Transcriptional and Post-transcriptional Landscapes of Mycobacterium smegmatis in Aerobic Growth and Hypoxia." Front Microbiol **10**: 591.

Masse, E., F. E. Escorcía and S. Gottesman (2003). "Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli." Genes Dev **17**(19): 2374-2383.

Miczak, A., V. R. Kaberdin, C. L. Wei and S. Lin-Chao (1996). "Proteins associated with RNase E in a multicomponent ribonucleolytic complex." Proc Natl Acad Sci U S A **93**(9): 3865-3869.

Plocinski, P., M. Macios, J. Houghton, E. Niemiec, R. Plocinska, A. Brzostek, M. Slomka, J. Dziadek, D. Young and A. Dziembowski (2019). "Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of Mycobacterium tuberculosis." Nucleic Acids Res **47**(11): 5892-5905.

Purusharth, R. I., F. Klein, S. Sulthana, S. Jager, M. V. Jagannadham, E. Evgenieva-Hackenberg, M. K. Ray and G. Klug (2005). "Exoribonuclease R interacts with endoribonuclease E and an RNA helicase in the psychrotrophic bacterium Pseudomonas syringae Lz4W." J Biol Chem **280**(15): 14572-14578.

Rosana, A. R., D. S. Whitford, R. P. Fahlman and G. W. Owttrim (2016). "Cyanobacterial RNA Helicase CrhR Localizes to the Thylakoid Membrane Region and Cosediments with Degradosome and Polysome Complexes in Synechocystis sp. Strain PCC 6803." J Bacteriol **198**(15): 2089-2099.

Rustad, T. R., K. J. Minch, W. Brabant, J. K. Winkler, D. J. Reiss, N. S. Baliga and D. R. Sherman (2013). "Global analysis of mRNA stability in Mycobacterium tuberculosis." Nucleic Acids Res **41**(1): 509-517.

Sasseti, C. M., D. H. Boyd and E. J. Rubin (2003). "Genes required for mycobacterial growth defined by high density mutagenesis." Mol Microbiol **48**(1): 77-84.

Sasseti, C. M. and E. J. Rubin (2003). "Genetic requirements for mycobacterial survival during infection." Proc Natl Acad Sci U S A **100**(22): 12989-12994.

Schuck, A., A. Diwa and J. G. Belasco (2009). "RNase E autoregulates its synthesis in Escherichia coli by binding directly to a stem-loop in the rne 5' untranslated region." Mol Microbiol **72**(2): 470-478.

Shell, S. S., J. Wang, P. Lapierre, M. Mir, M. R. Chase, M. M. Pyle, R. Gawande, R. Ahmad, D. A. Sarracino, T. R. Ioerger, S. M. Fortune, K. M. Derbyshire, J. T. Wade and T. A. Gray (2015). "Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape." PLoS Genet **11**(11): e1005641.

Stoppel, R., N. Manavski, A. Schein, G. Schuster, M. Teubner, C. Schmitz-Linneweber and J. Meurer (2012). "RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast." Nucleic Acids Res **40**(17): 8593-8606.

Taverniti, V., F. Forti, D. Ghisotti and H. Putzer (2011). "Mycobacterium smegmatis RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation." Mol Microbiol **82**(5): 1260-1276.

Updegrave, T. B., A. B. Kouse, K. J. Bandyra and G. Storz (2019). "Stem-loops direct precise processing of 3' UTR-derived small RNA MicL." Nucleic Acids Res **47**(3): 1482-1492.

Van Assche, E., S. Van Puyvelde, J. Vanderleyden and H. P. Steenackers (2015). "RNA-binding proteins involved in post-transcriptional regulation in bacteria." Front Microbiol **6**: 141.

Van den Bossche, A., S. W. Hardwick, P. J. Ceysens, H. Hendrix, M. Voet, T. Dendooven, K. J. Bandyra, M. De Maeyer, A. Aertsen, J. P. Noben, B. F. Luisi and R. Lavigne (2016). "Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome." Elife **5**.

Yang, H., W. Sha, Z. Liu, T. Tang, H. Liu, L. Qin, Z. Cui, J. Chen, F. Liu, R. Zheng, X. Huang, J. Wang, Y. Feng and B. Ge (2018). "Lysine acetylation of DosR regulates the hypoxia response of Mycobacterium tuberculosis." Emerg Microbes Infect **7**(1): 34.

Zeller, M. E., A. Csanadi, A. Miczak, T. Rose, T. Bizebard and V. R. Kaberdin (2007). "Quaternary structure and biochemical properties of mycobacterial RNase E/G." Biochem J **403**(1): 207-215.

Zhang, J. Y., X. M. Deng, F. P. Li, L. Wang, Q. Y. Huang, C. C. Zhang and W. L. Chen (2014). "RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region." RNA **20**(4): 568-579.