THESIS

IDENTIFICATION OF SMALL EXTRACELLULAR RNA FRAGMENTS OF *MYCOBACTERIUM TUBERCULOSIS*

Submitted by

Sarah Winter Sheldon

Department of Microbiology, Immunology, and Pathology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2014

Master's Committee:

Advisor: John Belisle

Mary Jackson
Lawrence Goodridge

ABSTRACT

IDENTIFICATION OF SMALL EXTRACELLULAR RNA FRAGMENTS OF *MYCOBACTERIUM TUBERCULOSIS*

In 2012, the World Health Organization reported 8.6 million estimated incident cases of tuberculosis, 1 million deaths among HIV-negative people, and 0.3 million deaths from HIV-associated tuberculosis. The Stop TB Partnership has a 2015 goal of reducing the 1990 prevalence rates by half. In order to accomplish this goal, there is a large effort to develop new vaccines, diagnostics, treatment, and therapeutics. Understanding how the pathogen, *Mycobacterium tuberculosis*, interacts with the host is critical to the development of these goals. An emerging area of interest is how host cells respond to bacterial nucleic acids; there are several bacteria that produce nucleic acids that impact pathogenesis through recognition by host pattern recognition receptors.

Previous work by Obregón-Henao et al. found that the culture filtrate (CF) of *M. tuberculosis* was able to induce apoptosis in monocytes, and the material was identified as small stable RNAs. Through cloning, the *M. tuberculosis* small RNA present in the CF was found to predominantly consist of tRNA and rRNA with lengths between 30 and 70 bases. The goal of this work was to further understand the composition of the small, stable, extracellular RNA of *M. tuberculosis*.

The first step in further elucidating the extracellular RNA population was to develop an RNA isolation method, allowing for the reliable purification of RNA from the CF of *M. tuberculosis* H37Rv. The method developed previously was not optimized for RNA purification,

and a more streamlined method was needed. Available commercial kits did not fit the specific needs of the project as a method was needed to isolate small RNAs from large volumes of CF. The method developed resulted primarily in small RNAs and allowed for isolation of extracellular RNA free of contaminants that could interfere with biological assays, including DNA, protein, LAM, and LPS. The kinetics of RNA release into the CF was examined, comparing the rate of RNA release to that of protein. The RNA and protein were found to have parallel release rates, which could indicate active release rather than passive release of the RNA.

Once a reliable RNA isolation method was developed, the composition of the extracellular RNA was interrogated utilizing Next Generation Sequencing as a high-throughput method. A pilot study was developed to determine the optimal concentration of extracellular RNA for sequencing. The Next Generation Sequencing provided a better understanding of the components of the secreted or released RNA. Ribosomal RNA and transfer RNA fragments were found to be present in the extracellular RNA, correlating to what was found by Obergón-Henao et al. A third group of small RNAs were also identified in this study, many of which corresponded to small RNAs previously reported in the literature, however novel small RNA sequences were also identified.

The possibility of bias in the sequencing technology was investigated using synthesized tRNA DNA oligonucleotides (stDNA oligos) added at specific concentrations. The quantitation bias study indicated that some bias occurs, although the cause is unknown. All of the stDNA oligos in the sample were identified, giving some confidence in the qualitative nature of this technology. However, based on the possibility of bias, it may be too generous to state that the technology is truly quantitative. Based on these studies, it is possible to say with confidence that what is identified is present, but not that things are not missed.

The long-term goals of this work are to fully understand how the extracellular RNA interacts with the host at a molecular level and to understand the mechanism of RNA release. In order to accomplish these goals, it will be necessary to evaluate more *M. tuberculosis* extracellular RNA using Next Generation Sequencing. A time course study with Next Generation Sequencing should also be done to see if the RNA composition changes over time, as well as for comparison to intracellular small RNAs. It would also be important to develop an assay to confirm fragments found using the Next Generation Sequencer, as well as to evaluate selective release from *M. tuberculosis.*

ACKNOWLEDGEMENTS

There are many people who have made this work possible and deserve acknowledgement. My time at Colorado State University has been a transformative, life changing experience, and I am truly grateful to everyone who has helped me on this journey.

I would like to thank my advisor, Dr. John Belisle for giving me the opportunity to work in his laboratory. He has pushed me beyond my comfort zone, teaching me to become a better critical thinker, to solve problems, and to challenge myself.

I would like to thank my committee members Dr. Jackson and Dr. Goodridge for their time and support throughout my graduate career.

I would like to thank past and present members of the Belisle and Slayden labs for all of their help, support, and friendship during my time at CSU. It has been an absolute pleasure to work with so many amazing people. Special thanks go to Darragh Heaslip and Luke Kingry. Thank you to Darragh for teaching me so much from day one; his patience, guidance, and friendship have been invaluable. Thank you to Luke for all of his help and patience teaching me about RNA and Next Generation Sequencing data analysis.

I would not be here without the support and love of my amazing family. A huge thanks goes to my parents Cathy and Peter Sheldon. Thank you for always being there for me and truly believing I could do anything I wanted. Thank you to my beautiful and talented sister Allison Sheldon. I am so lucky to have such great friend for a sister; you are a constant source of inspiration to me.

DEDICATION

This thesis is enthusiastically dedicated to my wonderful husband Tim. Thank you for supporting me every step of the way, through the highs and lows, joys and sorrows. Thank you for being a freakin trooper, I love you and I like you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

.csfasta	Colorspace data files
.gbk	GenBank reference genome
A	Adenine
Ag85A	Antigen 85A
APC	Antigen-presenting cell
BCA	Bicinchoninic acid
BCG	*Mycobacterium bovis* bacilli Camette-Guérin
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
cAMP	Cyclic adenosine monophosphate
cDC	Conventional dendritic cell
cDNA	Complementary DNA
CF	Culture filtrate
CFP10	Culture filtrate protein 10
CLR	C-type lectin receptor
ConA	Concanavalin A
CpG	2'-deoxyribo-cytidine-phosphate-guanoine
CR3	Complement receptor 3
CSP	Cytosolic surveillance pathway
CSU	Colorado State University
DC	Dendritic cell

DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3 grabbing non-integrin
DEAE	Diethylaminoethanol
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DOTS	Directly Observed Treatment, Short-Course
dsRNA	Double-stranded RNA
DST	Drug susceptibility testing
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMB	Ethambutol
ESAT-6	Early-secreted antigenic target 6
EtBr	Ethidium bromide
EU	Ethynyluridine
G	Guanine
GAS	Glycerol alanine salts
GBS	Group B *Streptococcus*
GtRNAdb	Genomic tRNA database
HIV	Human immundifficency virus
HPLC	High-performance liquid chromatography
IAA	Isoamyl alcohol extraction
ICD	Isocitrate dehydrogenase
IDT	Integrated DNA Technologies

IFN	Interferon
Ig	Immunoglobulin
IGRA	Interferon gamma release assays
IKK	IκB kinase
IL	Interleukin
INH	Isoniazid
IP RPC	Ion-pair reverse-phase chromatography
IPS-1	Interferon-β promoter stimulator 1
IRF3	Interferon regulatory factor 2
kDa	Kilodalton
LAL	Limulum Amebocyte Lysate
LAM	Lipoarabinomannan
Lgp2	Laboratory of genetics and physiology 2
LLO	Listeriolysin O
LM	Lipomannan
LPS	Lipopolysaccharide
MA	Mycolic acid
ManLAM	Mannose-capped LAM
MDA5	Melanoma differnetiation associated gene 5
MDR-TB	Multidrug-resistant Tuberculosis
MFE	Minimum free energy
MGC	Multinucleated giant cells
MHC	Major histocompatibility complex

miRNA	Micro RNA
mRNA	Messenger RNA
MTBC	*Mycobacterium tuberculosis* complex
MV	Membrane vesicle
MVA85A	Modified vaccinia virus Ankara 85 A
ncRNA	Non-coding RNA
NGS	Next Generation Sequencing
NK	Natural killer
NLR	Nucleotide-binding oligomerization domain-like receptors
NOD	Nucleotide-binding oligomerization domain
NT	Nucleotide
NTM	Nontuberculosis mycobacteria
ORN	Oligoribonucleotide
PAMP	Pathogen associated molecular patterns
PAS	p-aminosalicylic acid
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
pDC	Plasmacytoid dendritic cell
PDIM	Phthiocerol dimycocerosate
PEI	Polyethylenimine
PIMs	Phosphatidylinositol mannosides
piRNA	Pico RNA
PKR	Protein-kinase-RNA-regulated receptors

Pol	Polymerase
Poly I:C	Polyinosine-polycytidylic acid
PPD	Purified Protein Derivative
PRR	Pattern recognition receptor
PZA	Pyrazinamide
QA	Quality Assurance
QC	Quality Control
QFT	QuantiFERON-TB Gold In-Tube
qRT-PCR	Quantitative real time PCR
RD1	Region of difference 1
RIF	Rifampicin
RIG-I	Retinoic acid-inducible gene-I
RLH	Retinoic acid-inducible gene-I-like RNA helicases
RLR	Retinoic acid-inducible gene-I-like receptors
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
RT	Reverse transcription
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
siRNA	Small-interfering RNA
SOLiD	Support Oligonucleotide ligation detection

SR	Scavenger receptor
sRNA	Small RNA
ssRNA	Single-stranded RNA
stDNA	
oligo	Synthesized tRNA DNA Oligonucleotide
T	Thymidine
TAT	Twin-arginine transporter
TB	Tuberculosis
TBE	Tris Borate EDTA
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline and Tween 20
TDM	Trehalose dimycolate
TDR	Totally drug-resistant
Th1	T helper type 1
Th2	T helper type 2
TLR	Toll-like receptor
TMM	Trehalose monomycolate
TNF-α	Tumor necrosis factor alpha
Treg	T regulatory cell
tRNA	Transfer RNA
TSSS	Type VII Secretion System
TST	Tuberculin Skin Test
U	Uracil

V	Volts
VSV	Vesicular stomatitis virus
WHO	World Health Organization
XXDR	Extremely drug-resistant

Chapter I

Literature Review

1.1 General Introduction to Mycobacteria, *Mycobacterium tuberculosis* and Tuberculosis

1.1.1 *Mycobacterium* spp.

The genus *Mycobacterium* is made up of many different species, which are often broadly divided into two categories: *Mycobacterium tuberculosis* complex (MTBC) and nontuberculosis mycobacteria (NTM). The classification of these organisms is based on several defining characteristics including rate of growth and virulence [1]. While the topic of this thesis focuses on *M. tuberculosis*, it is important to acknowledge other important species in this genius. Some are obligate human pathogens, while others are considered opportunistic. Table 1.1 lays out a few of the *Mycobacterium* spp., dividing them into pathogenic and opportunistic, also identifying each organism as slow or rapid growers.

1.1.2 *Mycobacterium tuberculosis* general characteristics and physiology

Mycobacterium tuberculosis is the causative agent of tuberculosis (TB) and is a highly effective pathogen that has plagued humans for centuries. TB is a disease that has shaped our past and continues to shape our future. Despite the efforts of the WHO, it is still the second leading cause of death from an infectious disease worldwide [6].

In 1882, Robert Koch was the first to isolate *M. tuberculosis* and establish it as the causative agent of TB. Prior to this event, TB was believed to be an inherited disease [7]. *M. tuberculosis* is an acid-fast, Gram-positive bacillus with a unique and complex cell envelope along with a G/C rich genome [8]. As noted in Table 1.1, *M. tuberculosis* is a slow growing bacterium, which is due, in part, to the cell wall, reviewed in the next section [9].

Table 1.1 Prominent *Mycobacterium* species. Shows prominent *Mycobacterium* species separated by pathogenicity and rate of growth. Adapted from references [1-5].

	Mycobacterium species	**Growth**
Pathogenic	*Mycobacterium tuberculosis*	Slow
	Mycobacterium leprae	Slow
	Mycobacterium africanum	Slow
	Mycobacterium bovis	Slow
	Mycobacterium abscessus	Rapid
Opportunistic	*Mycobacterium avium*	Slow
	Mycobacterium ulcerans	Slow
	Mycobacterium kansasii	Slow
	Mycobacterium intracellulare	Slow
	Mycobacterium scrofulaceum	Slow
	Mycobacterium marinum	Slow
	Mycobacterium malmoense	Slow
	Mycobacterium simiae	Slow
	Mycobacterium szulgai	Slow
	Mycobacterium xenopi	Slow
	Mycobacterium haemophilum	Slow
	Mycobacterium smegmatis	Rapid
	Mycobacterium chelonae	Rapid
	Mycobacterium fortuitum	Rapid

1.1.2.1 Cell wall

M. tuberculosis has a unique and complex cell wall, which plays a role in pathogenesis, especially because it is the first thing that comes into contact with an antigen-presenting cell (APC). The extensive cell wall structure is important in protecting *M. tuberculosis* from the host immune response, and adds increased resistance to degradation by lysosomal enzymes in the phagosome [9]. The basic features of the *M. tuberculosis* cell wall will be briefly reviewed in this section, focusing on pathogenic aspects.

Because of the unique cell wall, *M. tuberculosis* is not a true Gram-positive bacterium, possessing Gram-negative attributes, including an outer membrane-like structure [10]. *M. tuberculosis* has an outer layer in addition to the typical Gram-positive peptidoglycan that

contains lipids, glycolipids, and polysaccharides. Figure 1.1 shows a simplified schematic of the *M. tuberculosis* cell wall, which includes some of the most biologically important structures.

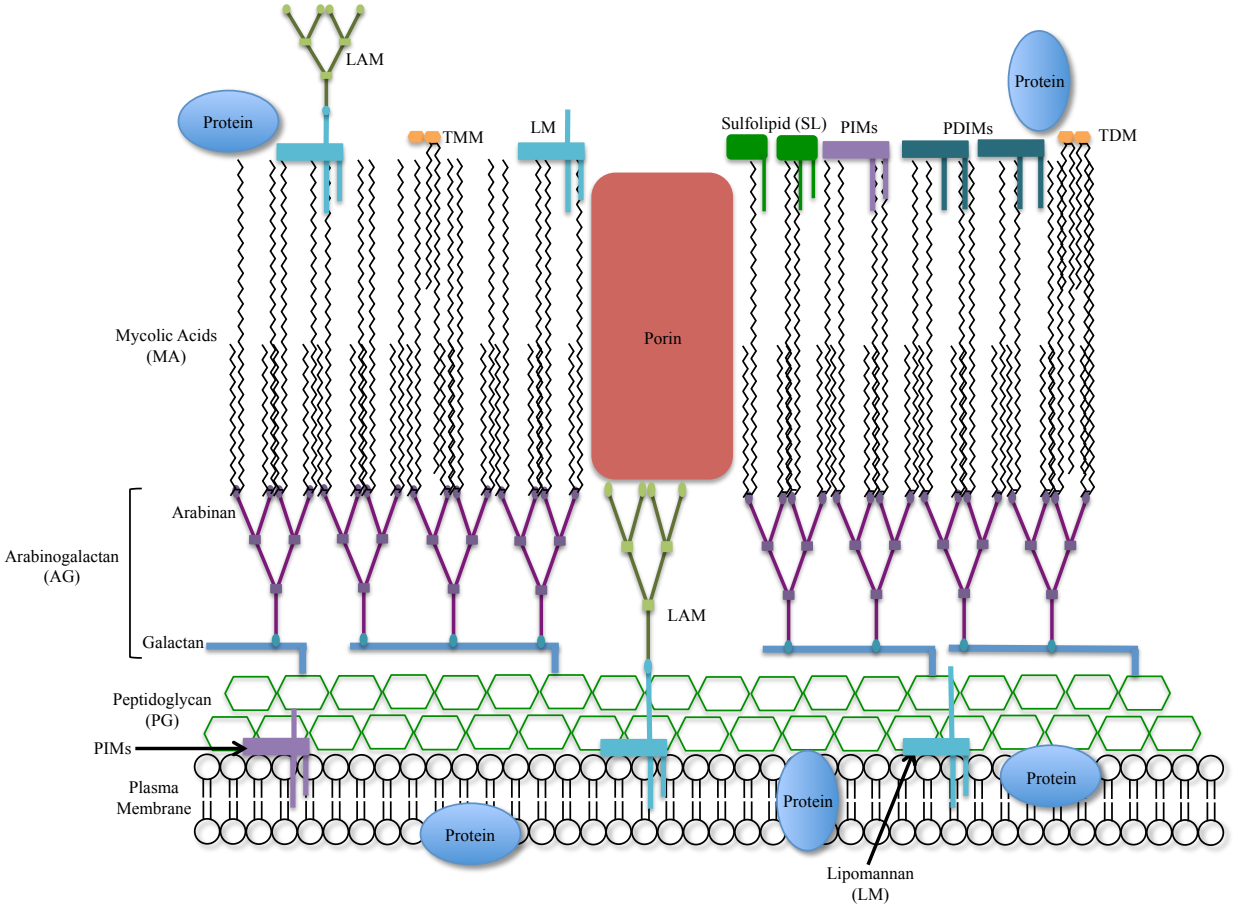


Figure 1.1 Mycobacterial cell wall. Figure shows some of the most biologically important structures within the plasma membrane and the outer membrane (Adapted from Chatterjee et al., 1998 [11] and Minnikin et al., 2002 [12]; permissions obtained).

The plasma membrane is surrounded by peptidoglycan covalently linked to arabinogalactan (AG) and mycolic acids (MA), both of which are unique to the actinomycete line of bacteria including the *Cornebacterium-Mycobacteria-Nocardia* (CMN) branch [9, 11, 13, 14]. The hydrophobic mycobacterial MAs are long-chain fatty acids that are mainly oriented towards the outside of the cell wall [9, 15-18]. Phosphatidylinositol mannosides (PIMs),

lipomannan (LM), and lipoarabinomannan (LAM) are all known to exist in a plasma membrane-bound form as well as in the outer layer [9, 14, 19-21].

In order to determine the pathogenic implications of the cell wall components of *M. tuberculosis*, researchers performed early studies by separating the various molecules and testing them individually. Using purified components, LAM has been implicated in multiple immunologic functions including abrogation of T cell activation [22], the inhibition of IFNγ-induced functions [23], scavenging of cytotoxic oxygen free radicals [24], and the production of macrophage associated cytokines including TNF-α [25-29]. Unlike *M. smegmatis* and *M. chelonae*, LAM in *M. tuberculosis* is mannose-capped, referred to as ManLAM, and has been considered an important virulence factor [30, 31]. Mannose caps are thought to be involved in both the attenuation of the host immune response and mycobacterial entry into macrophages [11]. Studies on the importance of ManLAM in virulence were preformed using purified product and studies often yielded conflicting results [20, 23, 32-37]. Mannose capping is not restricted to virulent strains of *M. tuberculosis*, indicating that ManLAM may not be an important virulence factor on its own [38-40].

Sulfolipids and trehalose dimycolates (TDM) are located in the outer layer of the cell wall, and have been found to be important in the survival of *M. tuberculosis* in the host cell (Section 1.2) although like ManLAM, early studies were done using purified products [41-51]. The outer layer also contains the lipid phthiocerol dimycocerosate (PDIM), which has been shown to be necessary for virulence in mouse models [52-55]. *M. tuberculosis* mutants rendered unable to synthesize or translocate PDIM are attenuated *in vivo*, although the exact role has not yet been identified [54, 56-59]. The mode of action and importance in virulence is not known

for many of these cell wall components and the development of mutants will be important in determining the true biological importance of these components.

1.1.2.2 Protein Secretion and translocation systems of *M. tuberculosis*

Secreted mycobacterial proteins make up a critical component of the host-pathogen interaction that occurs in TB, and the systems responsible for export or secretion are important mechanisms of virulence. This section briefly reviews four methods *M.tuberculosis* utilizes for export or secretion; the Sec pathway, the twin-arginine transporter (TAT) pathway, Type VII secretion system (T7SS), and production of membrane vesicles (MV).

The Sec pathway is a translocation system that is highly conserved in bacteria, and serves to translocate proteins containing N-terminal signal sequences in an unfolded state across the cytoplasmic membrane. One of the major components of the Sec secretion pathway is SecA, and *M. tuberculosis* is one of the few pathogenic bacteria that has two copies of the SecA protein, *secA1*, and *secA2* [8]. Deletion mutants were used to evaluate the roles of these two components, and *secA2* was found to be nonessential, as the *ΔsecA2* mutant is not able to prevent phagosome maturation, which is so critical to *M. tuberculosis* survival in the host cell [60-62]. SecA2 does contribute to the pathogenesis of *M. tuberculosis*, as it has been found to be the secretion mechanism for the protein superoxide dismutase A (SodA), which helps the pathogen elude oxidative attack by macrophages [61].

The TAT pathway is also a translocation system, which serves to translocate folded proteins across the cytoplasmic membrane. The TAT pathway is essential to *M. tuberculosis*, although the role in pathogenesis is not fully understood [63].

M. tuberculosis has five T7SS, also termed Esx 1-5, which are responsible for transporting proteins without a typical signal sequence. The Esx-1 secretion system is the best understood of the T7SS, and is involved in virulence. Esx-1 is encoded on the region of difference 1 (RD1), which is deleted in the attenuated vaccine strain *Mycobacterium bovis* Calmette-Guérin (BCG). The absence of RD1 and Esx-1 in BCG is associated with the decreased virulence of this vaccine strain [64].

There are several effectors secreted by the Esx-1 system also encoded on the RD1, including early-secreted antigenic target 6 (ESAT-6) (Esx-A) and culture filtrate protein 10 (CFP10) (Esx-B) [65]. ESAT-6 and CFP10 have effects on the innate as well as the adaptive immune response. ESAT-6 is important in the recruitment of uninfected macrophages to the site of infection, and both *M. tuberculosis* antigens are recognized by $CD4^+$ and $CD8^+$ T cells [66, 67]. ESAT-6 can also make pores in host membranes, which are responsible for three main consequences in the host cell. The first result of pore formation is host cell death, which is imperative for inducing formulation of necrotic granulomas [68]. The second is mycobacterial escape from the phagosome, which allows access to the cytoplasm after prolonged infection [69]. The third is secretion of bacterial effector molecules into the host cytoplasm like mycobacterial DNA, discussed in section 1.3.3.6 [69, 70].

In addition to the other methods of secretion and translocation, *M. tuberculosis* is also able to release products through active MVs [71]. MVs are released from bacteria through a process of budding from the outer membrane can contain genetic material, lipids, proteins, DNA, and RNA [72-76], which are used as a way for the bacteria to interact with it's environment [75, 77]. Many different mycobacterial species have been found to form and release MVs including *M. ulcerans*, *M. tuberculosis*, and BCG, although only the virulent species were found to possess

lipoprotein agonists of TLR2 [71]. When the MVs were given to mice before infection with *M. tuberculosis*, there was an increase in the bacterial replication in the lungs and spleens after infection [71]. MVs from BCG and *M. tuberculosis* were also found to induce a strong humoral immune response in TB patients [78]. Pardos-Rosales et al. later found that iron limitation, which occurs in the host cell environment, affects MV production, stimulating release and changing the lipid content, further implicating MVs in *M. tuberculosis* cell communication, modulated by the host environment [77]. Membrane vesicles provide a way for *M. tuberculosis* to release immunologically active factors into the host, which could help this pathogen survive and persist [77, 78].

1.1.3 Pathology, Epidemiology, and Control of *M. tuberculosis*

M. tuberculosis is spread through aerosol, and generally infects the lungs, termed pulmonary TB. Infection is not limited to the lungs however, and can affect other parts of the body, termed extrapulmonary TB. Within *M. tuberculosis* infection, there are two TB-related conditions: latent TB infection and active TB disease. In latent TB infection, the patient is infected with the bacterium, although they are not infectious and show no outward symptoms. In TB disease the bacteria are actively growing, causing the patient to become infectious and develop symptoms [79]. In general, a relatively small proportion of people infected with *M. tuberculosis* will develop active TB disease, although for people with weak immune systems or human immunodeficiency virus (HIV), the risk is much higher [80, 81].

In 2012, the World Health Organization (WHO) reported 8.6 million estimated incident cases of tuberculosis, 1 million deaths among HIV-negative people, and 0.3 million deaths from HIV-associated tuberculosis [6]. Geographically, 58% of the world's TB cases in 2012 were in

South-East Asia and the Western Pacific Regions [6]. The African region accounts for one fourth of the world's cases, India 26%, China 12%, and the Americas with only 3% [6]. There are multiple factors to this including areas with high rates of HIV, poverty, and population density.

In 1995, the WHO introduced and expanded a strategy for tuberculosis control termed DOTS (Directly Observed Treatment, Short-Course). DOTS operates using five key technical as well as managerial elements involving diagnosis, administration of anti-tuberculosis drugs, securing sufficient resources and political commitment, and a system for data monitoring [82]. In 1998, the Stop TB Partnership was formed in part to help accelerate DOTS implementation [82]. Since 1995, 56 million people have been successfully treated for TB in countries that have adopted DOTS [6]. The Stop TB Partnership has a 2015 goal of reducing the 1990 prevalence rates by half. Since 1990 the TB mortality rate has fallen by 45% globally, and incidence rates are falling in most parts of the world [6]. Because of the large number of people affected by tuberculosis, research to develop new drugs, diagnostics, and vaccines is extremely important.

1.1.3.1 Treatment

The thick cell wall, described in section 1.1.2.1, allows for decreased permeability contributing to inherent drug resistance [31]. TB can be cured by standard drug therapy, although the treatment times are long, due in part, to the slow rate of *M. tuberculosis* growth and the complex cell wall. The current regimen involves two months of treatment with first-line drugs: rifampicin (RIF), isoniazid (INH), ethambutol (EMB), and pyrazinamide (PZA); followed by four months of RIF and INH [6]. When first line drugs are ineffective, a group of second-line drugs are available; this includes thiamides, cycloserine, aminoglycosides, p-aminosalicylic acid

(PAS), fluorquinolones, and cyclic peptides [83]. Second-line drugs are more harmful to patients, often having severe side effects and long treatment times.

There are many reasons why new TB drugs are needed, including the rise of MDR-TB (discussed in the next section). Current drug regimens require long treatment times and often have toxicity associated with use, resulting in a need for drugs that are able to shorten the treatment time as well as improving efficacy and tolerability [6]. Because patients with HIV are more likely to develop TB disease [6] it is also important to consider HIV patients when developing new drugs. Any anti-TB treatments used for these patients need to be used in conjunction with antiretroviral therapy (ART) for HIV [6, 84]. The drugs at a physician's disposal for treating TB are around 50 years old, although recently, bedaquiline, a new anti-TB drug was approved by the FDA for use as part of combination drug therapy in adults [6, 85]. According to the 2013 WHO Global Tuberculosis Report, there a currently ten new or repurposed drugs in the late phases of clinical development [6].

In addition to drug toxicity and resistance, there are many socioeconomic factors that impact the treatment of TB. Socioeconomic factors include immigration, poverty, malnutrition, war, and limited access to medication/medical care [86]. The cost of drugs is a concern for treating impoverished patients and, second-line drugs are even more expensive than first-line drugs [87]. Getting the cost of current drugs or new drugs down would have an impact on the number of patients that are treated for TB. Many other obstacles in treatment exist that would require social change, and Stop TB Strategy aims to find resolutions for these issues [6]. One such obstacle is the need close follow-up and patient adherence to therapy [86], and that it can be difficult for individuals in rural and poor communities to get formal health care [87]. The Stop

TB Partnership aims to help accomplish it's goal by finding ways to improve health policies and finding ways to deliver necessary supplies and information [6].

1.1.3.2 Drug resistance

The development of drug resistant *M. tuberculosis* strains can be attributed in part, to poor TB control, a lack or misuse of drugs, and lack of patient follow-up [88]. In 2012 the WHO reported 450,000 incident cases and 170,000 deaths due to Multi-drug resistant TB (MDR-TB), with the highest levels in eastern Europe and central Asia [6]. Only 81 total confirmed cases of MDR-TB were reported in the US in 2012, comprising only 0.018% of the MDR-TB cases world wide [6]. Of the documented MDR-TB cases, 9.6% were recorded as Extensively drug-resistant TB (XDR-TB) [6]. MDR-TB is resistant to at least INH and RIF [6]. XDR-TB is resistant to at least INH and RIF, along with fluroquinolones, and more than one second-line drug [89].

A new category of resistant TB has been identified in Iran, India, and Italy, termed either extremely drug-resistant (XXDR), or totally drug-resistant (TDR); strains in this category are resistant to all first- and second-line drugs, along with additional drugs [89-91]. The rise in MDR/XDR-TB provides a pressing need for drugs that are capable of combating these difficult pathogens.

1.1.3.3 Vaccines

There is also a need for an effective TB vaccine to help prevent spread of the disease. There is only one vaccine that is currently available to protect against TB, *Mycobacterium bovis* bacillus Calmette-Guérin (BCG). BCG was found to have varied levels of protection in adults ranging from 0-80% [92]. While not entirely effective, BCG is able to protect against

disseminated TB in infants, so it is still widely used [93, 94]. There are two major strategies being used in the development of a new vaccine for TB. The first strategy is to produce a vaccine that is more efficient than BCG, and the second is to utilize a 'prime-boost' strategy [6]. The 'prime-boost' strategy for vaccine development is attractive, as many of the populations most in need of a vaccine already immunize with BCG [95]. There are currently 12 anti-TB vaccine candidates in Phase I, Phase II, or Phase IIb trials [6]. New vaccines fall into two major categories: live mycobacterial vaccines, and subunit vaccines [92].

A prominent example of one of these candidates is the recombinant modified vaccinia virus Ankara, which expresses the *M. tuberculosis* antigen 85A (Ag85A), named MVA85A [96]. MVA85A follows the 'prime-boost' strategy, and BCG vaccination followed by a boost with MVA85A was shown to improve protection in animals [97-100]. MVA85A was also found to be immunogenic and safe in South African adults [95]. Proof of concept Phase IIb trials to show safety and evaluate efficacy in infants for this candidate began in South Africa in 2009, and study results were recently published [6]. Despite encouraging early studies, MVA85A was not successful in providing additional protection from TB disease or *M. tuberculosis* infection in infants, although no concern of safety was raised [96, 101]. Further studies are necessary to evaluate MVA85A, as it may be more effective in a different age group such as adults, adolescents, or older children [6, 96]. Researchers also wish to investigate whether the vaccine would work better as a booster for adolescents and adults, further studying the interaction with BCG [101]. MVA85A is currently being tested on HIV positive adults in Senegal and South Africa to determine the efficacy with HIV patients [101]. Although the results of the MVA85A trial were not as positive as many had hoped, there may still be a use for this vaccine once other questions have been investigated [101, 102].

1.1.3.4 Diagnostics

Early detection and treatment are important in TB control, and can reduce the rate of transmission [103]. Current diagnostics for TB are often time consuming, and it is especially difficult to identify patients with latent TB [6]. This section discusses the available methods for diagnosis including the pros and cons to each method.

The conventional method of TB diagnosis is sputum smear microscopy and bacterial culture [6]. Unfortunately, this method takes a considerable amount of time, making the development of rapid diagnostic tests important for TB control. Sputum smear microscopy also relies on the acquisition of a good sample, which can be difficult to obtain, especially in children and HIV positive patients, as they do not produce sputum [104, 105]. Sputum acquisition also introduces a risk to healthcare workers because of the possibility of aerosolization during sample collection [106]. Sputum can have a low concentration of TB bacilli, rendering the sample smear negative [104]. The detection limit for microscopy is approximately $5x10^3$ bacilli per ml of specimen [107, 108]. Sample quality is greatly improved when patients are given specific instructions on sputum production, limiting the collection of saliva rather than sputum and increasing the volume of sample obtained, especially in women [109-112].

Chest radiography is another method used to look for TB, although it is not specific to TB, identifying abnormalities that could be attributed to other diseases. There are still uses for chest radiography as it can be a cost-effective solution and more readily available than other methods of diagnosis [113].

The Tuberculin skin test (TST) is a widely used diagnostic test, first developed by Robert Koch in the 1880s [114]. The TST involves intradermal injection of purified protein derivative (PPD), which contains a mixture of *M. tuberculosis* antigens [115]. When a patient has been

exposed to *M. tuberculosis*, they develop cell-mediated immunity to tuberculin antigens in PPD. The patient will have a delayed-type hypersensitivity skin reaction to PPD at the injection site, testing for the presence of a cellular immune response within 48 to 72 hours [115]. The TST is far from an ideal diagnostic, as it can only give an indication of TB disease or latent disease, and should not be the only method used. A positive result does not diagnose, as it could be a result of exposure to environmental mycobacteria, or previous BCG immunization. The impact of BCG vaccination on a TST depends on multiple factors including when the vaccination was given, and how many doses were administered [115]. A negative result does not rule out TB, as immunosuppressed patients may not react, leading to an incorrect interpretation of results [115]. The TST is still the preferable method for the serial testing of health care workers in the United States [116].

Interferon gamma (IFN-γ) release assays (IGRA) are another diagnostic platform used to test for active TB disease or latent TB infection, although it cannot distinguish between the two. Like the TST, IGRAs also test for the presence of a cellular immune response. One of the downsides to IGRAs is that we don't know how long adaptive immune responses to mycobacterial antigens last when mycobacteria are no longer present [117]. These assays use antigens that are specific to *M. tuberculosis* like ESAT-6 and CFP10 to stimulate T cells, and measure the release of IFN-γ [115]. These assays have lower sensitivity in immunocompromised patients, including patients with HIV and children. IGRAs are not affected by BCG vaccination status, and results can be available within 24 hours, relatively quickly when compared to the TST [115]. There are two main tests, the QuantiFERON-TB Gold In-Tube (QFT) (Cellestis/Qiagen, Carnegie, Australia) and the T-SPOT.TB (Oxford Immunotec, Abingdon, United Kingdom) [115]. QFT is an enzyme-linked immunosorbent assay (ELISA) that uses ESAT-6, CFP-10, and

TB7.7 to test whole blood [115]. TSPOT.TB is an enzyme-linked immunosorbent spot (ELISPOT) assay that uses PBMCs rather than whole blood [115]. PBMCs are incubated with ESAT-6 and CFP-10 to look for the number of IFN-γ producing T cells, or spot-forming cells [115].

Drug susceptibility testing (DST) is an important step once a diagnosis has been made in order to determine if the bacterium is resistant to first- or second-line anti-TB drugs. This determination is especially important in MDR-TB cases [6]. In 2010, the WHO recommended the rapid molecular test Xpert® MTB/RIF, which tests for pulmonary TB as well as rifampicin resistance [6]. New diagnostics are necessary for TB, especially tests that can be done quickly, inexpensively, and instrument free [6].

1.2 The host-pathogen interactions of *M. tuberculosis*

M. tuberculosis is a complex organism that is notoriously difficult to study for many reasons, including its ability to live undetected in its host for prolonged periods of time. *M. tuberculosis* has evolved with humans, and as such the bacterium has devised mechanisms to ensure its survival. Because of the need for new drugs and vaccines, there is an emphasis on studying the initial host-pathogen interactions and the initial stages of infection leading to disease. This section aims to describe some notable interactions between *M. tuberculosis* and the host immune system, including both the innate and adaptive immune response.

Following infection with a bacterial pathogen, the type of immune response is determined in part by the initial interaction between the pathogen and the host innate immune system. Pathogen associated molecular patterns (PAMPs), are molecules on microbial pathogens that bind to specific host pattern recognition receptors (PRRs), resulting in the activation of the innate immune response [118]. The binding of PAMPs to specific PRRs initiates signaling pathways

leading to the phosphorylation and activation of transcription factors. These transcription factors are capable of inducing production of many cytokines that are important in the innate immune response of the host. The type of innate immune response triggered by the interaction between PAMPs and PRRs tailors the type of adaptive immune response in the event that the innate immune response is unsuccessful on its own [118]. A few bacterial components that are recognized by PRRs and the innate immune system include: bacterial lipopolysaccharide (LPS), lipoproteins, peptidoglycan, flagellin, and teichoic acid [119]. These PAMPs are highly conserved among microbial pathogens and are not readily made by mammalian cells; over time, various receptors of the innate immune system have evolved to recognize these microbial products [118].

Initial interactions between *M. tuberculosis* and the host cell primarily occur with alveolar macrophages and dendritic cells (DCs), although some *in vitro* model systems use monocytes to study immunological interactions. The bacilli enter mononuclear phagocytes through receptor mediated phagocytosis and there are several *M. tuberculosis* products that can be recognized by PRRs during phagocytosis [120, 121].

M. tuberculosis can be opsonized by host molecules including immunoglobulins (Ig), complement, and lectins [122, 123]. Opsonization allows for enhanced phagocytosis, host cell activation, and microbicidal activity. Non-opsonized *M. tuberculosis* can be recognized by a number of different PRRs including: C-type lectin receptors (CLRs), scavenger receptors (SRs), Toll-like receptors (TLRs), nucleotide-binding oligomerization domain (NOD)-like receptors (NLR), and complement receptor 3 (CR3) [124-130].

C-type lectin receptors (CLRs) are a class of receptors on phagocytic cells and four different types are involved in the binding and uptake of *M. tuberculosis*: the mannose receptor,

dendritic cell-specific intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN), Mincle, and Dectin-I [126, 131]. The mannose residues on ManLAM allow for *M. tuberculosis* detection by the mannose receptor on macrophages and dendritic cells (DCs) [132-137]. DC-SIGN is a transmembrane receptor located on the surface of immature DCs, which is also able to detect ManLAM, mediating ManLAM induced production of IL-10 [130, 138, 139]. Mincle is expressed on the surface of macrophages and is an essential receptor for TDM, although there is some redundancy with other PRRs also able to recognize TDM including the one of the SRs [140-144]. Dectin-I has been shown to work in concert with TLR2 in mediating production of TNF-α [145, 146]. Additionally, Dectin-I has a role in the production of IL-12p40, which is important in the formation of granulomas [147].

NOD1 and NOD2 are PRRs that are able to detect peptidoglycan [148]. NOD2 has been found to be required for the optimal production of proinflammatory cytokines [149, 150], as well as play a role in the recognition and control of infection found by using knockout mice [151]. Polymorphisms associated with NLRs are more susceptible to *M. tuberculosis* infection [152-154].

TLRs 1, 2, 4, and 9 have been implicated in the host response and recognition to *M. tuberculosis* and will be discussed in greater depth in section 1.3.2 [155-159]. Lipoproteins of *M. tuberculosis* are TLR2 agonists and TLR2 is important for macrophage production of proinflammatory cytokines TNF-α, IL-1, IL-6, and IL-12 [160, 161].

In the course of many bacterial infections, phagocytosis by a macrophage causes death of the pathogen as a result of phagosome and lysosome fusion, or phagosomal maturation, leading to an acidic, hydrolytically active environment. *M. tuberculosis* and other intracellular pathogens are able to survive inside the macrophage by inhibiting this event. Inhibition of

phago-lysosome fusion requires live *M. tuberculosis* and only occurs with pathogenic bacteria. Mycobacteria-containing phagosomes stall at an early stage of maturation, retaining a pH of 6.4 [162]. Molecules that are proposed to modulate phagosome maturation are the cell wall glycolipids ManLAM, TDM, and sulfolipids, along with the bacterial phosphatase SapM and serine/threonine kinase PknG [162]. Modulation of the phagosome by *M. tuberculosis* is critical to the success of this pathogen. The early endosome provides a favorable environment that protects *M. tuberculosis* from serum immunoglobulins and complement, also giving the bacteria access to nutrients including iron through recycling endosomes [163].

Once *M. tuberculosis* becomes established inside the host cell, a localized proinflammatory response attracts immune cells including mononuclear cells and T cells resulting in a granuloma [164]. The formation of a granuloma becomes important for both the host and the bacteria, keeping the infection at bay, while providing the bacteria with protection from the immune system along with nutrient sources [164]. The granuloma is made up of macrophages, multinucleated giant cells (MGCs), foamy macrophages, T cells, B cells, and fibroblasts [131, 165]. When T cells fail to maintain the granuloma, as occurs with HIV, reactivation of the bacilli is promoted and active TB disease can progress [164]. The granuloma is also a place where cell-to-cell spread of *M. tuberculosis* occurs via release of bacilli from host cells through apoptosis and phagocytosis by uninfected cells, allowing for further spread of infection [131].

Despite the ability of *M. tuberculosis* to live inside macrophages, host cells possess methods of eliminating the bacilli. T cells and a T helper 1 (Th1)-type response are important for the host adaptive immune response and are critical for controlling *M. tuberculosis* during latent, or containment stages of disease. Cell mediated immunity is important for control of *M.*

tuberculosis largely because it is an intracellular organism, making antibodies less effective [165]. Although the adaptive immune response is important for the control of *M. tuberculosis* growth, it is not sufficient to abolish the infection [166]. T cells that secrete IFN-γ are important to granuloma formation and maintenance [117]. Exposure to IFN-γ before phagocytosis of *M. tuberculosis* allows the macrophage to overcome mycobacterial inhibition of phagosome maturation [162]. IFN-γ can activate autophagy, or delivery of the bacilli to the lysosome [131]. TNF-α works with IFN-γ to activate macrophages and modulate apoptosis [167]. NK cells are an important source of IFN-γ [168] and express granulysin, which has direct antimycobacterial activity.

There are several subsets of $CD4^+$ T cells: Th1 cells, Th17 cells, and T regulatory cells (Tregs); and all have a role in the host response to *M. tuberculosis*. Tregs secrete the anti-inflammatory cytokine IL-10 and TGFβ [165]. Th17 cells initiate the recruitment of Th1 cells to the lung through up regulation of chemokines [95].

T helper type 1 (Th1) cells are essential for primary resistance to infection and protective immunity in humans [131]. TNF-α is important early on in granuloma formation and the protective host response [34], and acts as a mediator of macrophage activation [165].

Th1 cells recognize peptide antigens that are bound to major histocompatibility complex (MHC) class II molecules. *M. tuberculosis* is able to diminish the ability of macrophages to present antigens to $CD4^+$ T cells, which makes it more difficult for the host to eliminate infection, down regulate MHC class II on the surface [165]. *M. tuberculosis* is also able to inhibit MHC II peptide loading and trafficking to the cell surface [169]. Resting macrophages express low levels of MHC class II molecules on the surface, once a macrophage is activated, antigen presentation is up regulated, increasing MHC II on the surface [162]. Antigen presenting

cells (APCs) are able to produce TGFβ, IL-6, and IL-10, which can directly and indirectly affect T cell function and proliferation [165].

The role of $CD8^+$ T cells in *M. tuberculosis* infection has been debated through the years and the details have not been fully elucidated. Many studies show that $CD8^+$ T cells are an important component in protection and control of TB infection [170-178]. $CD8^+$ T cells contribute to protection through secretion of IFN-γ [179, 180] as well as their ability to initiate cytotoxic lysis of host cells [181-183]. There are also studies suggesting $CD8^+$ T cells are not necessary during infection and have no protective role [184-187]. It is possible for $CD8^+$ T cells to arise, which can secrete the immunosuppressive IL-10 during a chronic *M. tuberculosis* infection, although the significance of this particular cell type is unclear [188]. $CD8^+$ T cells do not optimally produce IFN-γ without the presence of $CD4^+$ T cells, and this interaction between the two types requires further study [173, 189-192].

$CD8^+$ T cells recognize peptide epitopes bound to MHC class I, cytoplasmic antigens. MHC class I restricted and nonclassically restricted $CD8^+$ T cells (CD1). CD1b presents mycolic acid, LAM, PDIM, glucose monomycolate, and CD1c presents isoprenoid glycolipids. $CD8^+$ T cells are able to lyse infected cells by producing perfoin, granzymes, and granulysin, and produce cytokines like IFN-γ to recruit cellular effector cells [165].

Apoptosis and autophagy are two additional host cell mechanisms for controlling *M. tuberculosis* infection [164]. Apoptosis can also benefit *M. tuberculosis* depending on what stage of infection it occurs in [166]. Apoptosis promotes cross-presentation of antigens by $CD8^+$ T cells [131].

Type I Interferon Response and *M. tuberculosis*

Type I IFNs (IFN-α/β) were originally described and thought of as a part of the antiviral type response, although it is found more and more in response to bacterial infection as well. Type I IFNs induce the transcription of a large group of genes, and the products of these genes enhance the host's ability to resist viral infections. This class of IFNs are produced by many different cell types in response to stimulation by transmembrane and cytosolic receptors which will be discussed in more detail later [193].

Depending on stage of infection, type I IFNs could be either detrimental or beneficial to the host. They are important in activating components of the innate and adaptive immune system including APC maturation, activation of T cells, B cells, and NK cells [194, 195]. Because of their role in regulation, type I IFNs can suppress the innate immune response to bacterial infections and also play a role in apoptosis, increasing the susceptibility of different cell types [70, 193].

M. tuberculosis has been shown to induce a type I IFN response in macrophages although its role in susceptibility to disease has not been fully elucidated [70, 196]. The ability to induce a type I IFN response by specific *M. tuberculosis* clinical strains has been correlated with virulence, and The Esx-1 secretion pathway is necessary for this response to occur [197, 198]. The induction of the type I IFN response limits the host's ability to clear *M. tuberculosis* from the spleen and lungs [199, 200]. Mice deficient in the type I IFN receptor were found to have decreased mortality when compared to wild type mice, although the bacterial load in the lungs was similar [198]. IFN-α/β have been found to modulate IL-12-dependent IFN-γ production and impair Th1 responses in mice [198, 201].

Individually, IFN-α is able to prime macrophages to produce IL-10 and impair the mycobacteriostatic activity, and has been shown to enhance mycobacterial replication in human macrophages infected with BCG that could normally control infection [201]. IFN-α can be produced by macrophages; DCs, promoting both Th1 and DC maturation [34]; and lymphocytes [202]. IFN-α has also been shown to increase apoptosis in monocytes [201]. Figure 1.2 shows an overview of the basic mechanisms involved in the interaction between *M. tuberculosis* and the host cell.

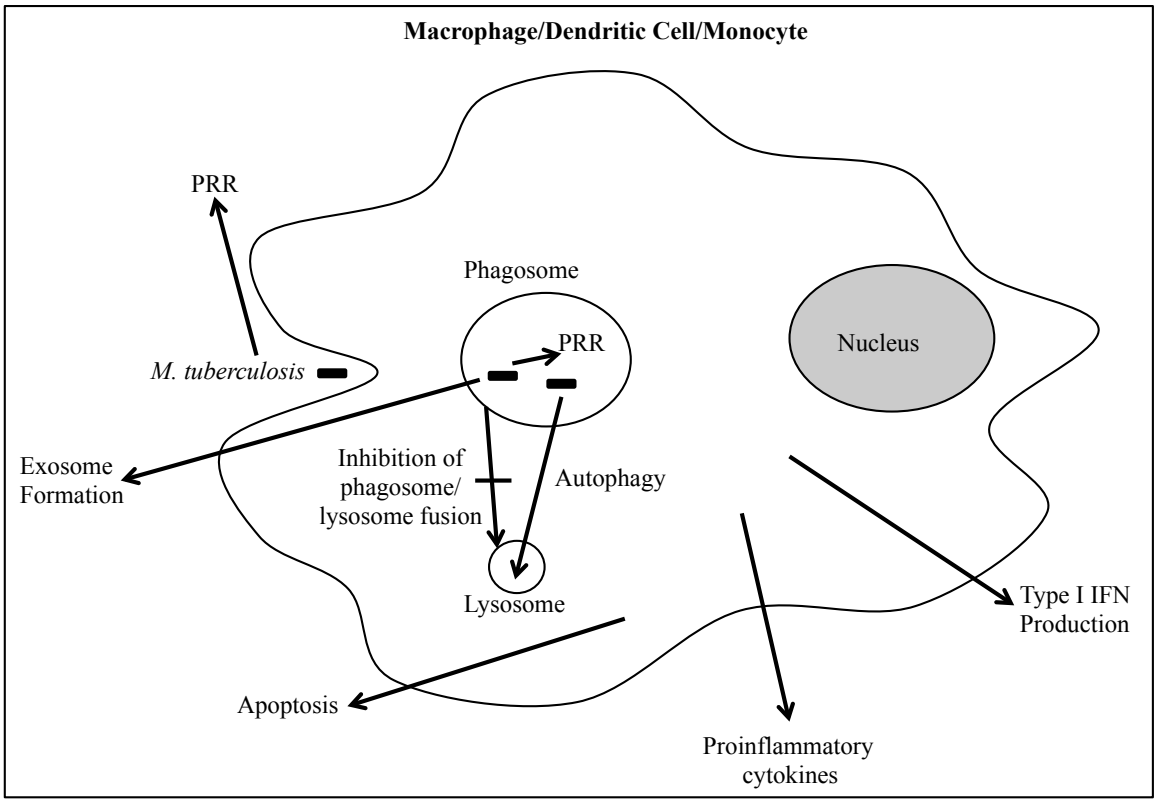


Figure 1.2 Basic molecular pathogenesis of *M. tuberculosis*. Overview of the basic molecular mechanisms induced by *M. tuberculosis* in the host cell. The host cell has pattern recognition receptors on the outside of the cell and inside the phagosome/early endosome. Inhibition of phagolysosomal fusion allows *M. tuberculosis* to survive intracellularly. Mycobacterial antigens can be released from the cells through exosome formation. After infection, the host cell is induced to produce proinflammatory cytokines like TNF-α, IL-1, IL-6, IL-12, as well as a type I IFN response. The host can attempt to eliminate *M. tuberculosis* using autophagy and apoptosis.

1.3 Host-microbe interactions via pathogen derived nucleic acid

1.3.1 RNA

Ribonucleic acid, or RNA, is an essential building block of life, and is likely the earliest genetic material in evolutionary history [203-207]. It is critical for the function of prokaryotic cells, eukaryotic cells, and viruses. RNA has a wide range of functions including storing and transmitting genetic information, and aiding in the synthesis of essential proteins for cell survival [206, 208-210].

RNA is composed of molecules containing a ribose sugar, a phosphate group, and a ribonucleotide base [209, 210]. Ribonucleotides are similar to the dioxynucleotides of DNA; both have adenine (A), guanine (G), and cytosine (C), but in place of the thymidine (T) of DNA, RNA contains uracil (U) [211]. The RNA of eukaryotic and prokaryotic cells is generally single-stranded (ssRNA). Some viruses are unique in that they contain double-stranded RNA (dsRNA) [212-214]. Typical ssRNA is able to fold or coil back on itself to form complex secondary and tertiary structures; the most common of which are hairpins with complementary base pairing and helical organization [209, 210, 215].

One of the most important roles of RNA in a cell is its involvement in transcription and translation [208]. In order to accomplish this, there are three main types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). These differ in their functions, location in the cell, and structure [209]. rRNA, mRNA, tRNA are considered “classic RNAs” [216].

The template for protein synthesis is mRNA, the only coding RNA, transcribed from the template strand of DNA [210]. The primary function of mRNA is to carry the information from

the nuclear DNA to the ribosome, where mRNA will guide amino acid assembly, forming a functional protein [217].

rRNA and tRNA are part of the protein-synthesizing machinery [210, 218, 219], and are both non-coding RNA [220]. rRNA has three major roles: it contributes to ribosome structure, it helps position the mRNA on the ribosome, and it also plays a catalytic role in protein synthesis [209, 218]. rRNA is a major component of the ribosome, along with polypeptides, and is the most abundant type of RNA [221]. The translation process occurs on the ribosome with complex assemblies of rRNAs [210, 218]. The bacterial ribosome is made up of a 30S subunit and a 50S subunit. The 30S subunit contains 16S rRNA, and the 50S subunit contains 5S and 23S rRNA, each of which contains a translational domain [209].

tRNA forms a cloverleaf-like structure generally consisting of a D arm and loop, a T arm and loop, an acceptor stem, a variable loop, and an anticodon loop (Figure 1.3) [222, 223]. This structure makes tRNA more stable than some of its other RNA counterparts, especially mRNA which degrades easily [224]. tRNA functions as the adaptor in protein synthesis, carrying activated amino acids to the ribosome to form proteins [219, 220]. The anticodon region recognizes the codon on the mRNA, and attaches the correct amino acid [225].

RNA is notoriously difficult to study. It is generally not very stable in solution due in part to the activity of ribonucleases (RNases), which are enzymes that process and degrade RNA [227]. Because of these difficulties, there are many aspects of how bacterial RNA affects biological systems that are not fully understood.

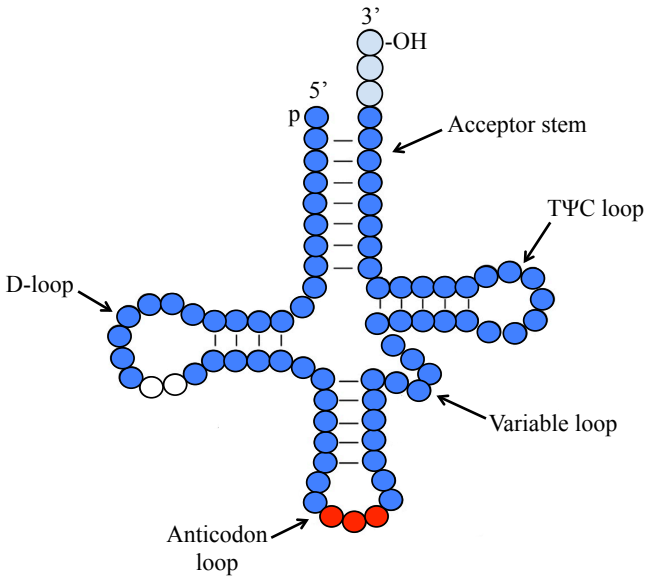


Figure 1.3 tRNA Secondary Structure. Circles represent nucleotides and lines represent bases. White circles in the D-loop represent residues present in some, but not all tRNAs. Red circles represent anticodon residues, and light blue circles represent the 3' CCA end (adapted from Phizicky et al., [226], permissions obtained).

1.3.1.1 Small RNA

Small RNAs (sRNA) have become a popular topic of research across multiple organisms. sRNA research was originally focused on cancer biology, but has more recently become a growing topic in the study of bacterial pathogens. Many bacterial sRNAs have been found through bioinformatic predictions. When bioinformatics screens are preformed several key elements are examined including promoter and terminator motifs in intergenic regions (IGR), phylogenetic conservation, and RNA structure similarities [228].

There are many types of sRNA including micro RNA (miRNA), pico RNA (piRNA), small-interfering RNA (siRNA), and non-coding (ncRNAs), ranging from 50 to 400 nucleotides in size [216]. These RNAs are typically not translated into proteins and have been shown to regulate a large number of functions related to bacterial viability and growth. sRNAs have been

shown to modulate numerous stress responses in bacteria as well as important components in microbial physiology including: catabolism, intracellular iron homeostasis, cell cycle, quorum sensing, toxicity, secretion system, cellular stress, and persistence [216]. sRNAs often act by base-pairing with specific target mRNAs altering their translation or stability, while others associate with specific proteins and modulate their function; allowing bacteria to quickly adjust to new environments [216]. sRNAs interact with regulatory proteins; possess specialized housekeeping functions; and interact with specific mRNAs in a regulatory manner affecting transcription, translation, and stability [216]. sRNAs can be further divided into two categories that are based on the mode of base pairing. *Cis*-encoded and *trans*-encoded. *Cis*-encoded sRNAs are within the target-encoding mRNA and are transcribed in an antisense orientation to the target mRNA [229]. *Trans*-encoded sRNA is typically in the IGR and often has multiple targets [230]. Many sRNAs have been identified although only a few have been studied in detail.

In 2009 Arnvig and Young published the first experimental evidence for sRNAs in *M. tuberculosis* [231]. The sRNAs were found by making cDNA libraries from low-molecular weight *M. tuberculosis* RNA. Nine putative sRNA molecules were found and differential expression was observed between exponential and stationary phases of growth, as well as in stress conditions [231]. This initial study was followed by a next generation sequencing analysis of the *M. tuberculosis* transcriptome [232]. The sequencing approach revealed a large number of sRNAs. Over one fourth of the sequence reads were ncRNA, with more intergenic regions than annotated coding sequences. The sRNAs were again monitored in exponential and stationary growth, with an increase in the stationary phase cultures. They also preformed a mouse study,

finding large amounts of the sRNAs in the lungs of infected mice, indicating a possible role in pathogenesis [232].

Using both a cloning technique and computational techniques, DiChiara et al. identified 34 novel sRNAs in BCG [233]. For the sRNAs found in BCG, homologues were identified in *M. tuberculosis* and *M. smegmatis*. Eight of the sRNAs were found in *M. tuberculosis* and BCG only, three were found in BCG and *M. smegmatis* only, 12 were found in all three species, and 14 were found in BCG alone [233]. The differential expression between the different species could be an indicator that some participate in pathogenicity.

Pelly et al. preformed a screen for ncRNAs in *M. tuberculosis* by cloning size-selected sRNAs from total RNA of *M. tuberculosis* [234]. Using this method, 12 clones were identified within intergenic regions. Of the 12 clones, four were novel sRNAs, and eight colones contained sequences from a region of the genome between *MT1302* and *MT1303*, termed ncrMT1302, which had previously been found by DiChiara et al [233, 234]. Further study showed that *ncrMT1302* is expressed at high levels in *M. tuberculosis.* Low pH, the addition of rifampicin, and supplementation with cAMP can modulate the levels of ncrMT1302 y [234]. Additionally, this was the first report to identify an expression of an *M. tuberculosis* sRNA *in vivo* [234].

The host has many miRNAs that are important in the immune response to *M. tuberculosis.* Host miRNAs have been found to be important in the regulation of T cell differentiation, the function of macrophages, DCs, and NK cells. There is research being conducted to use specific miRNAs as a biomarker for tuberculosis [235]. The following sections attempt to demonstrate what is known about bacterial RNA and how it affects the host immune system.

1.3.2 Receptors that recognize nucleic acid

Nucleic acids are conserved throughout all life forms, and the recognition of viral nucleic acid by mammalian cells is well described [236]. There are several types of PRRs involved in the recognition of foreign nucleic acids. These PRRs include: toll-like receptors (TLRs), protein-kinase-RNA-regulated receptors (PKRs), retinoic acid-inducible gene-1-like RNA helicases (RLH), and the nucleotide-binding oligomerization domain (NOD)-like receptors (NLR) family. Each of the previously mentioned types of PRRs plays a different role in various types of nucleic acid detection.

1.3.2.1 Toll-like receptors

TLRs are transmembrane receptors that play a major role in pathogen nucleic acid recognition in the host, and are among the most well described PRRs. TLRs are expressed in DCs, macrophages, B cells, and non-immune cells like fibroblasts and epithelial cells [237]. The TLR family is able to recognize both bacterial and viral products, and are involved in targeting and eliminating intracellular pathogens in the innate immune response [119, 238]. When a ligand is recognized by a TLR, a formation of homodimers or heterodimers occurs, and the cytoplasmic domain of TLRs interact with adaptor molecules and transmit signals initiating a signaling cascade. These signaling cascade can have different outcomes depending on the TLR, including leading to the transcription of NF-κB and MyD88 [239]. The interaction of a TLR with its specific ligand results in the activation of transcription factors.

TLRs 3, 7, 8, and 9 recognize nucleic acids and reside in intracellular compartments including the endosome, lysosome, and endoplasmic reticulum (ER) of host cells [240, 241]. When an intracellular TLR detects foreign nucleic acid, it triggers what is historically known as an anti-viral type response through these transcription factors. This response produces type I

IFNs and inflammatory cytokines such as interleukin-12 (IL-12), IL-1, and IL-6 [236, 241]. TLR polymorphisms can alter host susceptibility to human pathogens [242, 243].

1.3.2.1.1 TLR 3

TLR3 recognizes polyinosine-polycytidylic acid (poly (I:C)), a synthetic analogue of dsRNA, and natural dsRNA [236]. dsRNA is produced by many viruses during their replication [236]. The lysis of virally infected cells can lead to the release of dsRNA, which can be taken up by a transmembrane receptor and delivered to the intracellular compartment where TLR3 resides. Viruses inside cells can also release dsRNA during their replication which can then be detected by TLR3 [236].

TLR3 recognizes dsRNA partially by the ionic and hydrogen bonds associated with the sugar-phosphate backbone of the dsRNA [241]. Host RNA has a low affinity for TLR3 due to RNA modifications which include polyadenylation (poly-A tails) and nucleotide methylation [238]. Selective recognition of dsRNA by TLR3 leads to the activation of NF-κB and the production of IFN-α/β [236], which exert antiviral and immunostimulatory activities [119]. Activation of TLR3 also leads to production of immunoregulatory and antiviral cytokines IL-6 and IL-12, which help elicit cytotoxic responses needed for the elimination of intracellular pathogens [236, 244]. These cytokines can help control the replication of viruses and contribute to the initiation of the adaptive immune response [244].

1.3.2.1.2 TLR7 and TLR8

TLRs 7 and 8 are phylogenetically similar to one another, and for this reason, they are often studied together [241]. They were originally found to recognize synthetic anti-viral and anti-tumor compounds such as imidazoquinolines and guanine nucleoside analogues, as well as

whole RNA from viruses, and certain synthetic singe-stranded oligoribonucleotides (ORNs) [241, 245]. TLR7 and 8 have also been shown to sense short dsRNAs [246]. They were also originally thought to function for detection of viral RNA only, although more recent research has shown that they can also detect bacterial RNA [246]. Bacterial RNA contains less frequently modified nucleosides than human RNA, which limits TLR7 and 8 sensing of phagocytosed self-RNA [246]. Despite their similarities, TLR7 and 8 are not identical, and more research has led to discoveries of their differences. TLR7 is highly expressed in plasmacytoid dendritic cells (pDCs) and B cells [241, 246].

TLR8 is primarily, although not exclusively expressed in monocytes [246]. TLR8 preferentially recognizes bacterial RNA and ssRNA derived from HIV, vesicular stomatitis virus (VSV), and influenza A virus [245]. It is possible that ssRNA makes its way to the endosomes of cells through receptor-mediated endocytosis, where viral particles are degraded [241]. When RNA is detected by TLR8 in monocytes it leads to the production of IL-12p70, which then increases the expression of TLR8 in Natural Killer (NK) cells [246].

1.3.2.1.3 TLR9

TLR9 recognizes unmethylated 2'-deoxyribo-cytidine-phosphate-guanosine (CpG) DNA motifs and synthetic CpG oligodeoxynucleotides (ODNs) [239, 241]. CpG DNA is present in prokaryotes, but is rare in eukaryotes [247]. In prokaryotic DNA, the cytosines are rarely methylated, unlike those in eukaryotic DNA. This difference helps to prevent the innate immune system from recognizing host DNA [248]. Recognition of DNA occurs independently of the base sequence, as the sugar backbone 2'-deoxyribose is what allows the DNA to be recognized by TLR9 [241].

TLR9 is generally localized in the ER, and then migrates to the endosome where it recognizes its ligand [249]. The intracellular location of TLR9 is likely another factor that helps prevent recognition of self DNA and subsequent autoimmunity [239]. TLR9 is expressed in various cell types including plasmacytoid dendritic cells (pDCs), conventional DCs (cDCs), and macrophages [241]. When TLR9 is stimulated, different cytokines or transcription factors are produced depending on the type of CpG-DNA stimulation [239]. When TLR9 molecules in pDCs encounter their ligands, large amounts of type I IFNs are produced, which can directly activate T cells, linking the innate and the adaptive immune systems [239].

1.3.2.2 Protein kinase-RNA-regulated receptors

While TLRs are an important part of the pathogen recognition system, there are many TLR-independent mechanisms for recognizing foreign nucleic acids including the PKRs. PKRs are serine/threonine protein kinases, located in the cytosol, which are induced by IFNs and activated by dsRNA [250, 251]. PKRs contain two conserved dsRNA binding domains in their N-terminal region [194]. The recognition of dsRNA can induce PKR to activate the IκB kinase (IKK) complex, leading to the release of NF-κB, resulting in an inflammatory response and the production of type I IFNs [250]. Induction of these pathways helps to kill virally infected cells aiding the host in its defense against viral infections.

1.3.2.3 Retinoic acid-inducible gene-1-like RNA helicases

The retinoic acid-inducible gene-I (RIG-I)-like RNA helicase (RLH) family is also able to recognize nucleic acid in a TLR-independent manner, and includes: RIG-I, melanoma differentiation associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (Lgp2) [241, 252, 253]. These receptors are found in the cytoplasm of both immune and non-immune

cells, where they survey for the presence of viral RNA [241]. When members of the RLH family recognize the appropriate ligand, dsDNA for example, a typical antiviral-like response ensues, with induction of type I IFN. The dsDNA can be derived from viruses, bacteria, or even self.

RIG-I is a viral and bacterial PRR able to recognize short sequences of cytoplasmic dsRNA as well as various ssRNA viruses [194, 241, 254]. MDA5 is required for the recognition of some RNA viruses as well as poly (I:C) and binds long sequences of dsRNA [241]. Lgp2 is also able to bind dsRNA, and was recently found to be required for antiviral responses mediated by RIG-I and MDA5 by Satoh et al. [255].

1.3.2.4 Nucleotide-binding oligomerization domain (NOD)-like receptors

The Nucleotide-binding oligomerization domain (NOD)-like receptors (NLR) family is divided into five subfamilies, each with a different function. NOD2 binds viral RNA, which leads to the production of type I IFNs [237]. Some of the NLRs are capable of forming multi-protein complexes called inflammasomes that can secrete IL-1 family cytokines [241]. One of these inflammasomes is called NALP3, which is required for caspase-1 activation in response to dsRNA and bacterial RNA as well as other ligands [256]. The interaction between PAMPs and NALP3 is thought to be indirect, as no direction between the two has been shown [241]. The NALP3 inflammasome is important in the secretion of pro-inflammatory cytokines to help combat infection [241].

1.3.3 Specific microbes and the role of nucleic acids in the innate immune response

The previous section outlines multiple nucleic acid receptors and begins to show how pathogen RNA can modulate the immune response and possibly help a pathogen establish infection. The role of nucleic acids and their impact on the immune system has been more broadly studied in viruses, and there are many examples of this interaction. One nice example is from Moon et al., where they found that a noncoding RNA from a flavivirus is able to disrupt cellular mRNA decay in the host cell [257]. This mRNA decay causes interference with regulation of cellular gene expression that may influence the cell's ability to mount an effective immune response [257]. This section will outline specific microbes and the role of nucleic acids in pathogenesis.

1.3.3.1 *Listeria monocytogenes*

Listeria monocytogenes is the causative agent of listeriosis, a food-borne disease that can cause sepsis and meningitis in immunocompromised patients. This disease can also cause severe fetal infections or abortion in pregnant women [258, 259]. *L. monocytogenes* is a Gram-positive bacterium that infects and resides in macrophages [194, 260]. It gains access to the cytoplasm of macrophages by disrupting the phagosomal membrane using a pore-forming toxin called listeriolysin O (LLO), allowing the bacteria to replicate in the cytosol [260]. Once in the cytosol, the bacteria induce a type I IFN response, although the mechanism by which *L. monocytogenes* induces the response is still largely unknown [261]. The induction of type I IFN is independent of TLR and cytoplasmic RNA-sensing pathways [260], but seems to be related to multidrug resistance pathways [262]. These multidrug resistance pathways cannot transport large RNA or DNA polymers, but may be able to transport small RNA molecules [260]. Five novel ncRNAs were found in *L. monocytogenes* that are only found in pathogenic *L. monocytogenes,* and are likely involved in virulence [263]. Like *M. tuberculosis*, *L. monocytogenes* has two copies of the

SecA protein, as well as an Esx-1-like secretion system [264, 265]. *L. monocytogenes* was recently found to actively secrete small RNAs through the Sec secretion system, relying on SecA2 [65]. The secretion of small RNAs resulted in strong activation of RIG-I [65]. Hagmahn et al. were able to visualize the bacterial RNA being secreted into the cytosol of human cells [266].

L. monocytogenes is a good example of a pathogen that uses the type I IFN response to its benefit. The role of type I IFN in apoptosis could be of great importance in the pathogenesis of *L. monocytogenes* [70]. Studies conducted with mice lacking the type I IFN receptor found that these mice were more resistant to *L. monocytogenes* infection than wild-type mice [259, 267, 268]. The decreased survival of wild-type mice is likely a result of IFN-mediated apoptosis of splenocytes [259]. In this case, the type I IFN is not acting against the bacteria, but is instead acting against the host, allowing the bacteria to survive the innate immune response.

1.3.3.2 *Legionella pneumophila*

Legionella pneumophila is a Gram-negative bacterium that is ubiquitous in freshwater and soil amoebae. When humans inhale contaminated water, especially from air-conditioning units, the bacterium can cause a type of pneumonia called Legionnaires' disease [269, 270]. *L. pneunophila* replicates in the vesicles of alveolar macrophages, which helps it elude the immune response [270]. This bacterium is one of the only bacteria known to induce a type I IFN response through the MDA5 cytosolic RNA-detection pathway [271].

The interferon regulatory factor 3 (IRF3) and interferon-β promoter stimulator 1 (IPS-1) are critical for the IFNβ response, and are also important for the control of intracellular replication by *L. pneumophila* in lung epithelial cells [271]. In an experiment done by Opitz et al., mice lacking the IPS-1 gene had reduced induction of type I IFN by *L. pneumophila* [271].

In this experiment, the replication of *L. pneumophila* in mouse alveolar macrophages decreased with increasing IFNα/β [271]. The exact mechanism of how this occurs is unclear; one hypothesis is that *L. pneumophila* DNA is translocated into the host cells, which leads to the production of RNA ligands through DNA-dependent RNA polymerase III (Pol III) transcription [272]. There is however, no direct evidence that DNA translocation occurs during infection [260]. Another study showed that transfection of macrophages with *L. pneumophila* RNA, but not DNA, is able to induce a RIG-I-dependent type I IFN response [260]. This study also found a that MDA5 made a contribution to RIG-I in induction of type I IFNs. Unlike *L. monocytogenes*, the induction of the type I IFN response seems to have a protective effect against *L. pneumophila*.

1.3.3.3 Group B *Streptococcus*

Group B *Streptococcus* (GBS) is a group of Gram-positive extracellular bacteria that causes sepsis, meningitis, or pneumonia in human neonates [273, 274]. Transmission occurs prior to, or during birth, when the mother is colonized by GBS. The bacterium can also cause meningitis when acquired perinatally, although this sequale is less common [274].

GBS is another example of the induction of a type I IFN response that is beneficial to the host. When GBS is phagocytosed by conventional DCs (cDCs) and macrophages, the bacteria undergo degradation and lysis in the phagolysosome. The degraded bacterial products remain in the phagolysosme and it is there that nucleic acids from GBS act as ligands for TLR7 and TLR9 [273, 275]. RNA from GBS can be sensed by TLR7 in the endosome of cDCs [237]. In macrophages however, GBS activates a different pathway that is independent of TLR signaling [275]. In this pathway, liberated bacterial genomic DNA activates an unknown cytosolic DNA sensor [273].

1.3.3.4 *Helicobacter pylori*

Helicobacter pylori is a Gram-negative bacterium that can cause peptic ulcer disease, non-cardia adenocarcinoma of the stomach, as well as gastro-esophageal reflux disease. *H. pylori* is able to colonize the stomach despite harsh environmental conditions, and can maintain persistence throughout the life of host [276]. There is evidence suggesting *H. pylori* is a human-specific pathogen [277], and that co-evolution of the microbe and the host has likely occurred [278].

In order to survive for a prolonged period of time in the stomach of infected patients, *H. pylori* needs to evade the host immune system, and some of the methods are similar to those of *M. tuberculosis*. *H. pylori* is also able to inhibit phagosome maturation and persist in the early endosome [279]. *H. pylori* has a high rate of mutation as well as a high frequency of recombination, leading to a versatile population in a single host, which contributes to its ability to evade the host adaptive immune system [276]. *H. pylori* is also able to interfere with the uptake and processing of antigens as well as the activation of immune cells [280].

The role of TLR8 polymorphisms in *M. tuberculosis,* which will be discussed later, prompted Gantier et al. to investigate TLR8 signaling in *H. pylori* [279]. Infection with *H. pylori* promotes the induction of IFN-γ, which induces expression of TLR8 [279]. This study was the first evidence that TLR8 is able to function after phagocytosis of bacteria [279]. Additionally, Rad et al. used DCs to look at nucleic acid activation by *H. pylori* and found that TLR9 recognizes *H. pylori* DNA [281]. They also showed that *H. pylori* RNA induces proinflammatory cytokines in a TLR-independent manner, identifying RIG-I as the receptor for the bacterial RNA. Recognition of *H. pylori* RNA by RIG-I induces type I IFN production and

DC activation. These studies were the first evidence that bacterial RNA could be recognized by RIG-I to produce type I IFNs [281].

1.3.3.5 *Borrelia burgdorferi*

Borrelia burgdorferi is the spirochetal causative agent of Lyme disease, which is a public health problem in the United States [282] as well as Europe [283]. The primary route of transmission is by *Ixodes scapularis* ticks [284]. Lyme disease is a complex illness ranging from a local skin rash, called erythema migrans, to systemic problems including meningitis, encephalitis, arthritis, and carditis [285]. Activation of the innate immune system is an important component to the pathogenesis of *B. burgdorferi*, and a great deal of research has been done to better understand this response.

Phagocytosis of live *B. burgdorferi* leads to the transcription of IFN-β in human monocytes, which was previously solely attributed to TLR1/2 stimulation by *B. burgdorferi* lipoproteins [242]. It was later found that the type I interferon response could not be from TLR1/2 alone and that the induction of IFN-β was a result of TLR8-mediated signaling [286, 287]. Cervantes et al. recently discovered that *B. burgdorferi* RNA serves as the ligand for TLR8 [242]. Their studies show that activation occurs in the phagosome and that bacterial nucleic acids are not transferred into the cytosol. This is the first known example of bacterial RNA as an activator for TLR8 in human cells [242].

1.3.3.6 *Mycobacterium tuberculosis*

The previous sections help to illustrate the biological role nucleic acid can play in the pathogenesis of an organism. Although much of this research is fairly recent and needs to be

elucidated, it provides us with possibilities of how this may occur in other organisms. As discussed in section 1.3.1.1, sRNAs have been found in *M. tuberculosis*, and the following section discusses the possibility of nucleic acid playing a role in its pathogenesis as well.

Studies by Stanley et al. and Pandey et al. indicate that TLR signaling may not be involved in the induction of the type I IFN response against *M. tuberculosis* [70, 288]. Instead, these studies show that the presence of nucleic acids, like RNA and/or cell fragments in the cytosol may be of greatest importance. The type-VII secretion system may be involved in directly translocating these molecules or permeabilizing the phagosomal membrane, which would allow bacterial molecules to enter the cytosol [260]. This is a benefit for the organism, as it is able to elude the host immune response inside the macrophage.

Manzanillo et al. found extracellular *M. tuberculosis* DNA is able to trigger the cytosolic surveillance pathway (CSP) in macrophages and that this cytosolic signaling was required for virulence [69]. This response was found to be independent of TLR/NOD, and RNA-sensing molecules. They propose that the Esx-1 system is utilized to deliver ESAT-6 to cause conduits in the phagosomal membrane, allowing the DNA to come in to contact with cytosolic receptors. There are a few ideas as to how the DNA becomes extracellular, which include bacterial lysis, naturally occurring DNA on the surface of the bacterium, and active or passive release through outer membrane vesicles [69].

The role of *M. tuberculosis* nucleic acid and its interaction with the immune system has been explored recently. Obregón-Henao et al. investigated the finding that RNA in the culture filtrate (CF) of *M. tuberculosis* was able to induce early apoptosis in human monocytes [289]. Additional work showed that the apoptosis was a caspase-8 dependent, TNF-α and caspase-1

independent mechanism. The RNA affected the ability of human monocytes to control mycobacterial infection [289].

Although it has been shown that *M. tuberculosis* RNA detection in the host cell is linked to caspase-8, the exact mechanism is not known. Based on the information discussed above about receptors that recognize RNA, there are several possible ways in which the *M. tuberculosis* RNA could be interacting with the host cell. TLRs 3, 7, and 8 are located in the early endosome and are able to detect RNA, making them possible candidates. TLR8 can also lead to apoptosis, and is therefore a likely candidate based on the previous studies. TLR8 polymorphisms in humans have been studied and are associated with increased susceptibility to TB [243], although the mechanism used for TLR8 to recognize *M. tuberculosis* is unknown [290].

It is also possible that the RNA is able to move out of the early endosome into the cytosol to activate receptors like RIG-I, MDA5, and Lgp2. As discussed earlier, *M. tuberculosis* has also been shown to form exosomes, and the RNA could be packaged in these, where it could then be released and recognized by other cells [71]. Because of there was a lack of caspase-1 involvement in the apoptotic pathway studied by Obregón-Henao et al., it is not likely that the CF RNA is inducing an inflammasome response, such as the NALP3 inflammasome [289]. Figure 1.4 shows the various ways described in this section in which *M. tuberculosis* RNA could be interacting with the host cell.

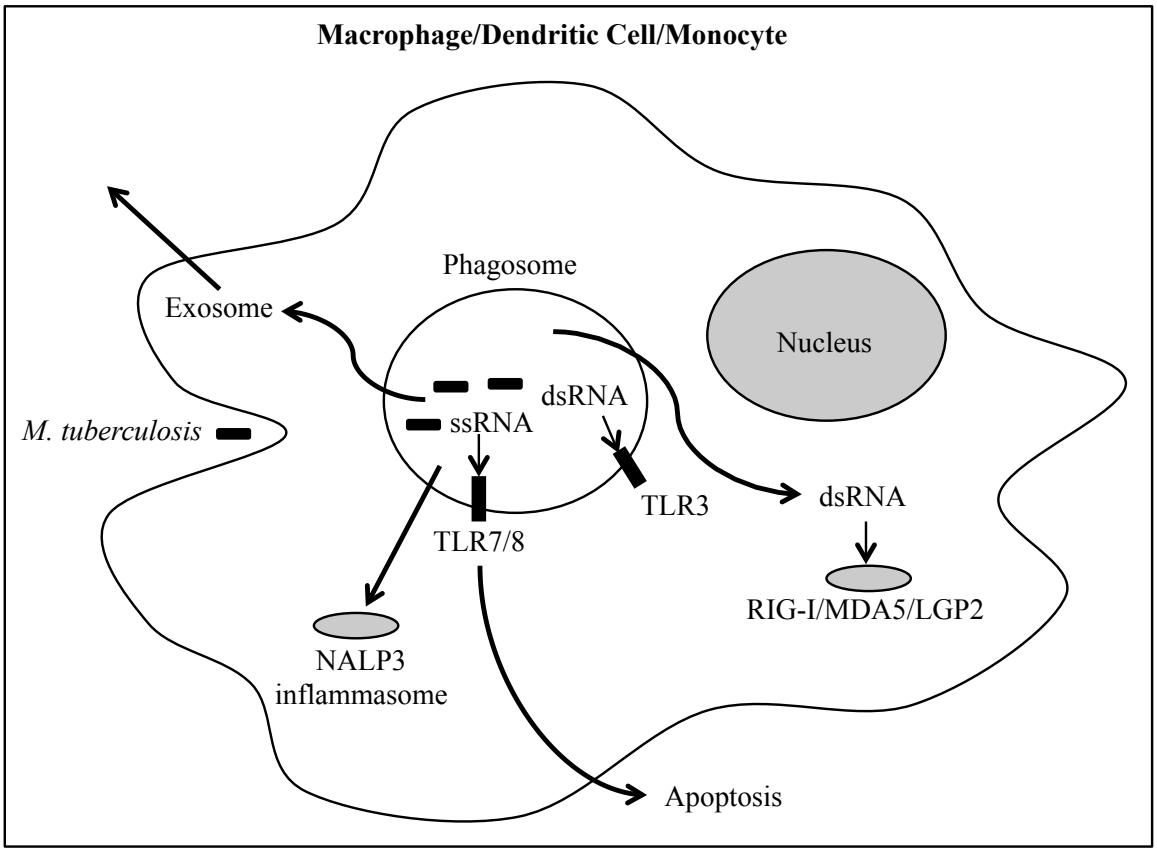


Figure 1.4 Ways in which *M. tuberculosis* RNA could be interacting with the host cell. TLR7/8, 3 are located in the early endosome and are able to detect RNA. Activation of TLR7/8 can also lead to apoptosis, which is important based on previous studies. Bacterial RNA could get out into the cytosol and activate a receptor like RIG-I. RNA could also be packaged in exosomes where it could then be released and recognized by other cells.

1.3.4 Specific RNA sequences recognized

Specific RNA sequences have been identified that bind to TLR7 and 8. These sequences are natural viral and eukaryotic RNA with U-rich sequences [291, 292]. Studies have coupled these sequences with synthetic ORNs to determine the immune response triggered by each one [291]. The GU-rich ORNs induce substantially stronger IFN-α production from human peripheral blood mononuclear cells (PBMCs) than other combinations of nucleotides tested [291]. At least four nucleotides of the specific TLR7 and 8 stimulating RNA sequences are

required to stimulate cytokine responses. The major sequences found to activate TLR7 and 8 are: UUGU, GUUC, GUUU, UUUC, UGUU, and UCUC. Some sequences are specific to TLR8, and are generally AU-rich, including: AUGU, UAUA, AUAU, AUAC, UAUU, UUAC, CUAC, GUAC, or UAUC [291]. It is now thought that immune stimulation could be due to the specific motifs impact on secondary structure, rather than the sequence itself [246]. All of the *M. tuberculosis* tRNAs have sequence homology to sequences found to activate both TLR7 and 8, and many have sequence homology to TLR8 specific reactive sequences [291].

1.4 Project Rationale

After discovering that the CF of *M. tuberculosis* was able to induce early apoptosis in human monocytes, Obregón-Henao et al. set out to find what material was responsible. Because the CF of *M. tuberculosis* is largely protein based, it was hypothesized that the apoptosis was the result of released or secreted proteins. The CF was column fractionated and the active fraction was tested with Proteinase K, DNase, and RNase. Activity was lost when the fraction was treated with RNase, and unfractionated CF behaved in the same manner. To determine what type of RNA was present in the CF, a classical siRNA cloning technique was adopted. The *M. tuberculosis* RNA present in the CF was found to predominantly consist of tRNA and rRNA with lengths between 30 and 70 bases. Because of the possibility that the RNA could be degraded to 30-70 bases, the material in the CF was also immediately precipitated with acetone, indicating that the RNA is fairly stable in the CF [289].

In order to address the question of if the RNA was released as a result of active or passive release, the kinetic accumulation of RNA in the CF over time was also assessed. RNA was detected in the CF as early as two days of growth, and the accumulation paralleled that of protein secretion [289]. This is important since the presence of protein in the CF of early log phase

culture has been shown to be a result of active release [293]. The presence of RNA in early culture indicates that *M. tuberculosis* could be actively releasing these RNAs.

These previous findings lead to the hypothesis that this RNA population will be of biological importance. Before the biological importance can be evaluated, it is necessary to further understand the composition of the small, stable, extracellular RNA of *M. tuberculosis*. The study of small RNAs in *M. tuberculosis* will facilitate long-term goals of understanding how these products are released and how they interact with the host at a molecular level.

References

1. Shinnick, T. and R. Good, *Mycobacterial taxonomy.* European Journal of Clinical Microbiology and Infectious Diseases, 1994. **13**(11): p. 884-901.
2. Wagner, D. and L. Young, *Nontuberculous mycobacterial infections: a clinical review.* Infection, 2004. **32**(5): p. 257-270.
3. Primm, T.P., C.A. Lucero, and J.O. Falkinham, *Health impacts of environmental mycobacteria.* Clinical microbiology reviews, 2004. **17**(1): p. 98-106.
4. Katoch, V., *Infections due to non-tuberculous mycobacteria (NTM).* Indian Journal of Medical Research, 2004. **120**: p. 290-304.
5. Stahl, D.A. and J. Urbance, *The division between fast-and slow-growing species corresponds to natural relationships among the mycobacteria.* Journal of bacteriology, 1990. **172**(1): p. 116-124.
6. Organization, W.H., *Global tuberculosis report 2013.* 2013.
7. Daniel, T.M., *The history of tuberculosis.* Respiratory Medicine, 2006. **100**(11): p. 1862-1870.
8. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-544.
9. Brennan, P.J. and H. Nikaido, *The Envelope of Mycobacteria.* Annual Review of Biochemistry, 1995. **64**: p. 29-63.
10. Rastogi, N., *Recent observations concerning structure and function relationships in the mycobacterial cell envelope: elaboration of a model in terms of mycobacterial pathogenicity, virulence and drug-resistance.* Research in microbiology, 1991. **142**(4): p. 464-476.
11. Chatterjee, D. and K.-H. Khoo, *Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects.* Glycobiology, 1998. **8**(2): p. 113-120.
12. Minnikin, D.E., et al., *The Methyl-Branched Fortifications of Mycobacterium tuberculosis.* Chemistry & biology, 2002. **9**(5): p. 545-553.
13. Woese, C.R., *Bacterial evolution.* Microbiological reviews, 1987. **51**(2): p. 221.
14. Brennan, P., *Mycobacterium and other actinomycetes.* Microbial lipids, 1988. **1**: p. 203-298.
15. Minnikin, D. and M. Goodfellow. *Lipid composition in the classification and identification of acid-fast bacteria*. in *Society for Applied Bacteriology symposium series*. 1979.
16. Minnikin, D.E., *Lipids: complex lipids, their chemistry, biosynthesis and roles.* The biology of the mycobacteria, 1982. **1**: p. 95-184.
17. Dobson, G., et al., *Systematic analysis of complex mycobacterial lipids.* Society for Applied Bacteriology. Technical Series, 1985(20): p. 237-265.
18. McNeil, M., M. Daffe, and P.J. Brennan, *Location of the mycolyl ester substitutents in the cell-walls of Mycobacteria.* Journal of Biological Chemistry, 1991. **266**(20): p. 13217-13223.
19. Goldman, D.S., *Subcellular localization of individual mannose-containing phospholipids in Mycobacterium tuberculosis.* American Review of Respiratory Disease, 1970. **102**(4): p. 543-&.

20. Chatterjee, D., et al., *Lipoarabinomannan - multiglycosylated form of the mycobacterial mannosylphosphatidylinositols.* Journal of Biological Chemistry, 1992. **267**(9): p. 6228-6233.
21. Hunter, S.W. and P.J. Brennan, *Evidence for the presence of a phosphatidylinositol anchor on the lipoarabinomannan and lipomannan of Mycobacterium tuberculosis.* Journal of Biological Chemistry, 1990. **265**(16): p. 9272-9279.
22. Kaplan, G., et al., *Mycobacterium leprae Antigen-Induced Suppression of T-cell Proliferation in vitro.* Journal of Immunology, 1987. **138**(9): p. 3028-3034.
23. Sibley, L.D., et al., *Mycobacterial lipoarabinomannan inhibits gamma-interferon-mediated activation of macrophages.* Infection and Immunity, 1988. **56**(5): p. 1232-1236.
24. Chan, J., et al., *Lipoarabinomannan, a possible virulence factor involved in persistence of Mycobacterium tuberculosis within macrophages.* Infection and Immunity, 1991. **59**(5): p. 1755-1761.
25. Moreno, C., A. Mehlert, and J. Lamb, *The inhibitory effects of mycobacterial lipoarabinomannan and polysaccharides upon polyclonal and monoclonal human t-cell proliferation.* Clinical and Experimental Immunology, 1988. **74**(2): p. 206-210.
26. Moreno, C., et al., *Lipoarabinomannan from Mycobacterium tuberculosis induces the production of tumor necrosis factor from human and murine macrophages.* Clinical and Experimental Immunology, 1989. **76**(2): p. 240-245.
27. Barnes, P.F., et al., *Cytokine production induced by Mycobacterium tuberculosis lipoarabinomannan - relationship to chemical-structure.* Journal of Immunology, 1992. **149**(2): p. 541-547.
28. Chatterjee, D., et al., *Structural basis of capacity of lipoarabinomannan to induce secretion of tumor-necrosis-factor.* Infection and Immunity, 1992. **60**(3): p. 1249-1253.
29. Adams, L.B., Y. Fukutomi, and J.L. Krahenbuhl, *Regulation of murine macrophage effector functions by lipoarabinomannan from mycobacterial strains with different degrees of virulence.* Infection and Immunity, 1993. **61**(10): p. 4173-4181.
30. Strohmeier, G.R. and M.J. Fenton, *Roles of lipoarabinomannan in the pathogenesis of tuberculosis.* Microbes and Infection, 1999. **1**(9): p. 709-717.
31. Jarlier, V. and H. Nikaido, *Mycobacterial cell wall: structure and role in natural resistance to antibiotics.* FEMS microbiology letters, 1994. **123**(1): p. 11-18.
32. Kaur, D., et al., *Biosynthesis of mycobacterial lipoarabinomannan: role of a branching mannosyltransferase.* Proceedings of the National Academy of Sciences, 2006. **103**(37): p. 13664-13669.
33. Briken, V., et al., *Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response.* Molecular Microbiology, 2004. **53**(2): p. 391-403.
34. Giacomini, E., et al., *Infection of human macrophages and dendritic cells with Mycobacterium tuberculosis induces a differential cytokine gene expression that modulates T cell response.* The Journal of Immunology, 2001. **166**(12): p. 7033-7041.
35. Johansson, U., J. Ivanyi, and M. Londei, *Inhibition of IL-12 production in human dendritic cells matured in the presence of Bacillus Calmette-Guerin or lipoarabinomannan.* Immunology Letters, 2001. **77**(1): p. 63-66.
36. Demangel, C., P. Bertolino, and W.J. Britton, *Autocrine IL-10 impairs dendritic cell (DC)-derived immune responses to mycobacterial infection by suppressing DC*

trafficking to draining lymph nodes and local IL-12 production. European Journal of Immunology, 2002. **32**(4): p. 994-1002.
37. Rojas, M., et al., *Differential induction of apoptosis by virulent Mycobacterium tuberculosis in resistant and susceptible murine macrophages: role of nitric oxide and mycobacterial products.* J Immunol, 1997. **159**(3): p. 1352-1361.
38. Prinzis, S., D. Chatterjee, and P.J. Brennan, *Structure and antigenicity of lipoarabinomannan from Mycobacterium bovis BCG.* Journal of General Microbiology, 1993. **139**: p. 2649-2658.
39. Venisse, A., et al., *Structural features of lipoarabinomannan from Mycobacterium bovis BCG - determination of molecular-mass by laser-desorption mass-spectrometry.* Journal of Biological Chemistry, 1993. **268**(17): p. 12401-12411.
40. Khoo, K.H., et al., *Inositol phosphate capping of the nonreducing termini of lipoarabinomannan from rapidly growing strains of Mycobacterium.* Journal of Biological Chemistry, 1995. **270**(21): p. 12380-12389.
41. Bloch, H., *Studies on the virulence of tubercle bacilli - isolation and biological properties of a constituent of virulent organisms.* Journal of Experimental Medicine, 1950. **91**(2): p. 197-&.
42. Noll, H., et al., *The chemical structure of the cord factor of Mycobacterium tuberculosis.* Biochimica Et Biophysica Acta, 1956. **20**(2): p. 299-309.
43. Noll, H. and H. Bloch, *Studies on the chemistry of the cord factor of Mycobacterium tuberculosis.* Journal of biological chemistry, 1955. **214**(1): p. 251-65.
44. Goren, M.B., *Mycobacterial lipids - selected topics.* Bacteriological Reviews, 1972. **36**(1): p. 33-&.
45. Goren, M.B., et al., *Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(7): p. 2510-2514.
46. Rao, V., et al., *Mycobacterium tuberculosis controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule.* Journal of Experimental Medicine, 2005. **201**(4): p. 535-543.
47. Rao, V., et al., *Trans-cyclopropanation of mycolic acids on trehalose dimycolate suppresses Mycobacterium tuberculosis-induced inflammation and virulence.* Journal of Clinical Investigation, 2006. **116**(6): p. 1660-1667.
48. Rousseau, C., et al., *Sulfolipid deficiency does not affect the virulence of Mycobacterium tuberculosis H37Rv in mice and guinea pigs.* Infection and Immunity, 2003. **71**(8): p. 4684-4690.
49. Kumar, P., et al., *PapA1and PapA2 are acyltransferases essential for the biosynthesis of the Mycobacterium tuberculosis virulence factor Sulfolipid-1.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(27): p. 11221-11226.
50. Converse, S.E., et al., *MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(10): p. 6121-6126.
51. Brodin, P., et al., *High Content Phenotypic Cell-Based Visual Screen Identifies Mycobacterium tuberculosis Acyltrehalose-Containing Glycolipids Involved in Phagosome Remodeling.* Plos Pathogens, 2010. **6**(9).

52. Domenech, P. and M.B. Reed, *Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from Mycobacterium tuberculosis grown in vitro: implications for virulence studies.* Microbiology, 2009. **155**(11): p. 3532-3543.
53. Camacho, L.R., et al., *Analysis of the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis - Evidence that this lipid is involved in the cell wall permeability barrier.* Journal of Biological Chemistry, 2001. **276**(23): p. 19845-19854.
54. Cox, J.S., et al., *Complex lipid determine tissue specific replication of Mycobacterium tuberculosis in mice.* Nature, 1999. **402**(6757): p. 79-83.
55. Rousseau, C., et al., *Virulence attenuation of two Mas-like polyketide synthase mutants of Mycobacterium tuberculosis.* Microbiology-Sgm, 2003. **149**: p. 1837-1847.
56. Camacho, L.R., et al., *Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis.* Molecular Microbiology, 1999. **34**(2): p. 257-267.
57. Domenech, P., M.B. Reed, and C.E. Barry, *Contribution of the Mycobacterium tuberculosis MmpL protein family to virulence and drug resistance.* Infection and Immunity, 2005. **73**(6): p. 3492-3501.
58. Astarie-Dequeker, C., et al., *Phthiocerol Dimycocerosates of M. tuberculosis Participate in Macrophage Invasion by Inducing Changes in the Organization of Plasma Membrane Lipids.* Plos Pathogens, 2009. **5**(2).
59. Rousseau, C., et al., *Production of phthiocerol dimycocerosates protects Mycobacterium tuberculosis from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection.* Cellular Microbiology, 2004. **6**(3): p. 277-287.
60. Braunstein, M., et al., *Two nonredundant SecA homologues function in mycobacteria.* Journal of bacteriology, 2001. **183**(24): p. 6979-6990.
61. Braunstein, M., et al., *SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis.* Molecular microbiology, 2003. **48**(2): p. 453-464.
62. Sullivan, J.T., et al., *The Mycobacterium tuberculosis SecA2 system subverts phagosome maturation to promote growth in macrophages.* Infection and immunity, 2012. **80**(3): p. 996-1006.
63. Saint-Joanis, B., et al., *Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of Mycobacterium tuberculosis, increases β-lactam susceptibility and virulence.* Journal of bacteriology, 2006. **188**(18): p. 6669-6679.
64. Pym, A.S., et al., *Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti.* Molecular Microbiology, 2002. **46**(3): p. 709-717.
65. Abdullah, Z., et al., *RIG-I detects infection with live Listeria by sensing secreted bacterial nucleic acids.* Embo Journal, 2012. **31**(21): p. 4153-4164.
66. Brodin, P., et al., *ESAT-6 proteins: protective antigens and virulence factors?* Trends in Microbiology, 2004. **12**(11): p. 500-508.
67. Volkman, H.E., et al., *Tuberculous Granuloma Induction via Interaction of a Bacterial Secreted Protein with Host Epithelium.* Science, 2010. **327**(5964): p. 466-469.
68. Junqueira‐Kipnis, A.P., et al., *Mycobacteria lacking the RD1 region do not induce necrosis in the lungs of mice lacking interferon‐γ.* Immunology, 2006. **119**(2): p. 224-231.

69. Manzanillo, P.S., et al., *Mycobacterium Tuberculosis Activates the DNA-Dependent Cytosolic Surveillance Pathway within Macrophages.* Cell Host & Microbe, 2012. **11**(5): p. 469-480.
70. Stanley, S.A., et al., *The Type I IFN Response to Infection with Mycobacterium tuberculosis Requires ESX-1-Mediated Secretion and Contributes to Pathogenesis.* J Immunol, 2007. **178**(5): p. 3143-3152.
71. Prados-Rosales, R., et al., *Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice.* Journal of Clinical Investigation, 2011. **121**(4): p. 1471-1483.
72. Beveridge, T.J., *Structures of gram-negative cell walls and their derived membrane vesicles.* Journal of Bacteriology, 1999. **181**(16): p. 4725-4733.
73. Beveridge, T.J. and J.L. Kadurugamuwa, *Periplasm, periplasmic spaces, and their relation to bacterial wall structure: Novel secretion of selected periplasmic proteins from Pseudomonas aeruginosa.* Microbial Drug Resistance-Mechanisms Epidemiology and Disease, 1996. **2**(1): p. 1-8.
74. Mashburn-Warren, L.M. and M. Whiteley, *Special delivery: vesicle trafficking in prokaryotes.* Molecular Microbiology, 2006. **61**(4): p. 839-846.
75. Kuehn, M.J. and N.C. Kesty, *Bacterial outer membrane vesicles and the host-pathogen interaction.* Genes & Development, 2005. **19**(22): p. 2645-2655.
76. Lee, E.Y., et al., *Proteomics in gram-negative bacterial outer membrane vesicles.* Mass Spectrometry Reviews, 2008. **27**(6): p. 535-555.
77. Prados-Rosales, R., et al., *Role for Mycobacterium tuberculosis Membrane Vesicles in Iron Acquisition.* Journal of Bacteriology, 2014. **196**(6): p. 1250-1256.
78. Ziegenbalg, A., et al., *Immunogenicity of mycobacterial vesicles in humans: Identification of a new tuberculosis antibody biomarker.* Tuberculosis, 2013. **93**(4): p. 448-455.
79. Comstock, G.W., *Epidemiology of tuberculosis.* The American review of respiratory disease, 1982. **125**(3 Pt 2): p. 8.
80. Havlir, D.V. and P.F. Barnes, *Tuberculosis in patients with human immunodeficiency virus infection.* New England Journal of Medicine, 1999. **340**(5): p. 367-373.
81. Prevention, C., *treatment of tuberculosis among patients infected with human immunodeficiency virus: principles of therapy and revised recommendations. Centers for Disease Control and Prevention.* MMWR, 1998. **47**: p. 1-58.
82. Raviglione, M.C. and A. Pio, *Evolution of WHO policies for tuberculosis control, 1948–2001.* The Lancet, 2002. **359**(9308): p. 775-780.
83. WHO, *Global tuberculosis control 2010*. WHO/HTM/TB/2010. 2010, Geneva: World Health Organization.
84. Maher, D., A. Harries, and H. Getahun, *Tuberculosis and HIV interaction in sub‐Saharan Africa: impact on patients and programmes; implications for policies.* Tropical Medicine & International Health, 2005. **10**(8): p. 734-742.
85. Singh, H., N.K. Natt, and N. Garewal, *Bedaquiline: a new weapon against MDR and XDR-TB.* International Journal of Basic & Clinical Pharmacology, 2013. **2**(2): p. 96-102.
86. Tsara, V., E. Serasli, and P. Christaki, *Problems in diagnosis and treatment of tuberculosis infection.* Hippokratia, 2009. **13**(1): p. 20-22.

87. Bates, I., et al., *Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part II: Determinants operating at environmental and institutional level.* The Lancet infectious diseases, 2004. **4**(6): p. 368-375.
88. Espinal, M.A., *The global situation of MDR-TB.* Tuberculosis, 2003. **83**(1): p. 44-51.
89. Migliori, G.B., et al., *Totally drug-resistant and extremely drug-resistant tuberculosis: the same disease?* Clinical infectious diseases, 2012. **54**(9): p. 1379-1380.
90. Udwadia, Z.F., et al., *Totally Drug-Resistant Tuberculosis in India.* Clinical Infectious Diseases, 2012. **54**(4): p. 579-U156.
91. Velayati, A.A., et al., *Emergence of New Forms of Totally Drug-Resistant Tuberculosis BacilliSuper Extensively Drug-Resistant Tuberculosis or Totally Drug-Resistant Strains in Iran.* CHEST Journal, 2009. **136**(2): p. 420-425.
92. Andersen, P. and T.M. Doherty, *The success and failure of BCG—implications for a novel tuberculosis vaccine.* Nature Reviews Microbiology, 2005. **3**(8): p. 656-662.
93. Rodrigues, L.C., V.K. Diwan, and J.G. Wheeler, *Protective effect of BCG against Tuberculosis meningitis and military Tuberculosis- a metaanalysis.* International Journal of Epidemiology, 1993. **22**(6): p. 1154-1158.
94. Trunz, B.B., P.E.M. Fine, and C. Dye, *Effect of BCG vaccination on childhood tuberculous meningitis and miliarytuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness.* Lancet, 2006. **367**(9517): p. 1173-1180.
95. Hawkridge, T., et al., *Safety and immunogenicity of a new tuberculosis vaccine, MVA85A, in healthy adults in South Africa.* Journal of Infectious Diseases, 2008. **198**(4): p. 544-552.
96. Tameris, M.D., et al., *Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial.* The Lancet, 2013.
97. Vordermeier, H.M., et al., *Viral Booster Vaccines Improve Mycobacterium bovis BCG-Induced Protection against Bovine Tuberculosis.* Infection and Immunity, 2009. **77**(8): p. 3364-3373.
98. Verreck, F.A.W., et al., *MVA.85A Boosting of BCG and an Attenuated, phoP Deficient M. tuberculosis Vaccine Both Show Protective Efficacy Against Tuberculosis in Rhesus Macaques.* Plos One, 2009. **4**(4).
99. Goonetilleke, N.P., et al., *Enhanced immunogenicity and protective efficacy against Mycobacterium tuberculosis of bacille Calmette-Guerin vaccine using mucosal administration and boosting with a recombinant modified vaccinia virus Ankara.* Journal of Immunology, 2003. **171**(3): p. 1602-1609.
100. Williams, A., et al., *Boosting with poxviruses enhances Mycobacterium bovis BCG efficacy against tuberculosis in guinea pigs.* Infection and Immunity, 2005. **73**(6): p. 3814-3816.
101. Dye, C. and P.E.M. Fine, *A major event for new tuberculosis vaccines.* Lancet, 2013. **381**(9871): p. 972-974.
102. Kenneth, J., *Lack of efficacy of MVA85A TB vaccine candidate: potential outcomes.* Current Science, 2013. **104**(12): p. 1594-1595.
103. Menzies, D., H. Al Jahdali, and B. Al Otaibi, *Recent developments in treatment of latent tuberculosis infection.* The Indian journal of medical research, 2011. **133**(3): p. 257.
104. Dheda, K., et al., *Point-of-care diagnosis of tuberculosis: Past, present and future.* Respirology, 2013. **18**(2): p. 217-232.

105. Harries, A., *Tuberculosis and human immunodeficiency virus infection in developing countries.* The Lancet, 1990. **335**(8686): p. 387-390.
106. Jensen, P.A., et al., *Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005*. 2005: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention.
107. Shinnick, T.M. and R.C. Good, *Diagnostic mycobacteriology laboratory practices.* Clinical infectious diseases, 1995: p. 291-299.
108. Smithwìck, R.W., *Laboratory manual for acid-fast microscopy*. 1976: Public Health Service, Center for Disease Control, Bureau of Laboratories.
109. Kumari, S., R. Ichhpujani, and W.H. Organization, *Guidelines on standard operating procedures for Microbiology*. 2000: SEARO WHO.
110. Khan, M.S., et al., *Improvement of tuberculosis case detection and reduction of discrepancies between men and women by simple sputum-submission instructions: a pragmatic randomised controlled trial.* The Lancet, 2007. **369**(9577): p. 1955-1960.
111. Uplekar, M., S. Rangan, and J. Ogden, *Gender and tuberculosis control: towards a strategy for research and action*. 1999: World health organization (WHO).
112. Fairall, L.R., et al., *Effect of educational outreach to nurses on tuberculosis case detection and primary care of respiratory illness: pragmatic cluster randomised controlled trial.* Bmj, 2005. **331**(7519): p. 750-754.
113. Leung, C.C., *Reexamining the role of radiography in tuberculosis case finding.* International Journal of Tuberculosis and Lung Disease, 2011. **15**(10): p. 1279-1279.
114. Jasmer, R.M., P. Nahid, and P.C. Hopewell, *Latent tuberculosis infection.* New England Journal of Medicine, 2002. **347**(23): p. 1860-1866.
115. Pai, M., et al., *Gamma Interferon Release Assays for Detection of Mycobacterium tuberculosis Infection.* Clinical Microbiology Reviews, 2014. **27**(1): p. 3-20.
116. Nienhaus, A., et al., *IFN-γ release assay versus tuberculin skin test for monitoring TB infection in healthcare workers.* Expert review of anti-infective therapy, 2013. **11**(1): p. 37-48.
117. Mack, U., et al., *LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement.* European Respiratory Journal, 2009. **33**(5): p. 956-973.
118. Medzhitov, R. and C. Janeway, *Innate immune recognition: mechanisms and pathways.* Immunological Reviews, 2000. **173**: p. 89-97.
119. Takeda, K. and S. Akira, *Toll receptors and pathogen resistance.* Cellular Microbiology, 2003. **5**(3): p. 143-153.
120. Schlesinger, L.S., et al., *Phagocytosis of Mycobacterium tuberculosis is mediated by human monocyte complement receptors and complement component C3.* The Journal of Immunology, 1990. **144**(7): p. 2771-2780.
121. Russell, D.G., *Mycobacterium tuberculosis: here today, and here tomorrow.* Nature Reviews Molecular Cell Biology, 2001. **2**(8): p. 569-586.
122. Schlesinger, L.S., *< i> Mycobacterium tuberculosis</i> and the complement system.* Trends in microbiology, 1998. **6**(2): p. 47-49.
123. Torrelles, J., et al., *Role of C-type lectins in mycobacterial infections.* Current drug targets, 2008. **9**(2): p. 102-112.

124. Ferguson, J.S., et al., *Complement protein C3 binding to Mycobacterium tuberculosis is initiated by the classical pathway in human bronchoalveolar lavage fluid.* Infection and immunity, 2004. **72**(5): p. 2564-2573.
125. Schafer, G., et al., *Non-Opsonic Recognition of Mycobacterium tuberculosis by Phagocytes.* Journal of Innate Immunity, 2009. **1**(3): p. 231-243.
126. Killick, K.E., et al., *Receptor-mediated recognition of mycobacterial pathogens.* Cellular Microbiology, 2013. **15**(9): p. 1484-1495.
127. Court, N., et al., *Partial Redundancy of the Pattern Recognition Receptors, Scavenger Receptors, and C-Type Lectins for the Long-Term Control of Mycobacterium tuberculosis Infection.* Journal of Immunology, 2010. **184**(12): p. 7057-7070.
128. Ernst, J.D., *Macrophage receptors for Mycobacterium tuberculosis.* Infection and Immunity, 1998. **66**(4): p. 1277-1281.
129. Schlesinger, L.S., *Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors.* Journal of Immunology, 1993. **150**(7): p. 2920-2930.
130. Tailleux, L., et al., *DC-SIGN is the major Mycobacterium tuberculosis receptor on human dendritic cells.* Journal of Experimental Medicine, 2003. **197**(1): p. 121-127.
131. Philips, J.A. and J.D. Ernst, *Tuberculosis pathogenesis and immunity.* Annual Review of Pathology: Mechanisms of Disease, 2012. **7**: p. 353-384.
132. Brennan, P.J., *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis.* Tuberculosis, 2003. **83**(1-3): p. 91-97.
133. Nigou, J., et al., *Mycobacterial lipoarabinomannans: modulators of dendritic cell function and the apoptotic response.* Microbes and infection, 2002. **4**(9): p. 945-953.
134. Nigou, J., et al., *Mannosylated lipoarabinomannans inhibit IL-12 production by human dendritic cells: evidence for a negative signal delivered through the mannose receptor.* The Journal of immunology, 2001. **166**(12): p. 7477-7485.
135. Schlesinger, L.S., S.R. Hull, and T.M. Kaufman, *Binding of the terminal mannosyl units of lipoarabinomannan from a virulent-strain of Mycobacterium tuberculosis to human macrophages.* Journal of Immunology, 1994. **152**(8): p. 4070-4079.
136. Schlesinger, L.S., et al., *Differences in mannose receptor-mediated uptake of lipoarabinomannan from virulent and attenuated strains of Mycobacterium tuberculosis by human macrophages.* Journal of Immunology, 1996. **157**(10): p. 4568-4575.
137. Venisse, A., J.J. Fournie, and G. Puzo, *Mannosylated lipoarabinomannan interacts with phagocytes.* European Journal of Biochemistry, 1995. **231**(2): p. 440-447.
138. Geijtenbeek, T.B.H., et al., *Mycobacteria target DC-SIGN to suppress dendritic cell function.* Journal of Experimental Medicine, 2003. **197**(1): p. 7-17.
139. Wu, T.T., et al., *Interaction between mannosylated lipoarabinomannan and dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin influences dendritic cells maturation and T cell immunity.* Cellular Immunology, 2011. **272**(1): p. 94-101.
140. Matsumoto, M., et al., *A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages.* The journal of immunology, 1999. **163**(9): p. 5039-48.
141. Ishikawa, E., et al., *Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle.* Journal of Experimental Medicine, 2009. **206**(13): p. 2879-2888.
142. Heitmann, L., et al., *Mincle is not essential for controlling Mycobacterium tuberculosis infection.* Immunobiology, 2013. **218**(4): p. 506-516.

143. Miyake, Y., et al., *C-type Lectin MCL Is an FcR gamma-Coupled Receptor that Mediates the Adjuvanticity of Mycobacterial Cord Factor.* Immunity, 2013. **38**(5): p. 1050-1062.
144. Bowdish, D.M.E., et al., *MARCO, TLR2, and CD14 Are Required for Macrophage Cytokine Responses to Mycobacterial Trehalose Dimycolate and <italic>Mycobacterium tuberculosis</italic>.* PLoS Pathog, 2009. **5**(6): p. e1000474.
145. Lee, H.M., et al., *Mycobacterium abscessus activates the NLRP3 inflammasome via Dectin-1-Syk and p62/SQSTM1.* Immunology and Cell Biology, 2012. **90**(6): p. 601-610.
146. Yadav, M. and J.S. Schorey, *The beta-glucan receptor dectin-1 functions together with TLR2 to mediate macrophage activation by mycobacteria.* Blood, 2006. **108**(9): p. 3168-3175.
147. Rothfuchs, A.G., et al., *Dectin-1 interaction with mycobacterium tuberculosis leads to enhanced IL-12p40 production by splenic dendritic cells.* Journal of Immunology, 2007. **179**(6): p. 3463-3471.
148. Kanneganti, T.D., M. Lamkanfi, and G. Nunez, *Intracellular NOD-like receptors in host Defense and disease.* Immunity, 2007. **27**(4): p. 549-559.
149. Gandotra, S., et al., *Nucleotide-binding oligomerization domain protein 2-deficient mice control infection with Mycobactefium tuberculosis.* Infection and Immunity, 2007. **75**(11): p. 5127-5134.
150. Divangahi, M., et al., *NOD2-Deficient Mice Have Impaired Resistance to Mycobacterium tuberculosis Infection through Defective Innate and Adaptive Immunity.* Journal of Immunology, 2008. **181**(10): p. 7157-7165.
151. Brooks, M.N., et al., *NOD2 controls the nature of the inflammatory response and subsequent fate of Mycobacterium tuberculosis and M-bovis BCG in human macrophages.* Cellular Microbiology, 2011. **13**(3): p. 402-418.
152. Austin, C.M., X. Ma, and E.A. Graviss, *Common nonsynonymous polymorphisms in the NOD2 gene are associated with resistance or susceptibility to tuberculosis disease in African Americans.* Journal of Infectious Diseases, 2008. **197**(12): p. 1713-1716.
153. Zhao, M.Y., et al., *A novel single nucleotide polymorphism within the NOD2 gene is associated with pulmonary tuberculosis in the Chinese Han, Uygur and Kazak populations.* Bmc Infectious Diseases, 2012. **12**.
154. Pan, H., et al., *Polymorphisms of NOD2 and the risk of tuberculosis: a validation study in the Chinese population.* International Journal of Immunogenetics, 2012. **39**(3): p. 233-240.
155. Basu, J., D.M. Shin, and E.K. Jo, *Mycobacterial signaling through toll-like receptors.* Frontiers in Cellular and Infection Microbiology, 2012. **2**.
156. Drennan, M.B., et al., *Toll-like receptor 2-deficient mice succumb to Mycobacterium tuberculosis infection.* American Journal of Pathology, 2004. **164**(1): p. 49-57.
157. Abel, B., et al., *Toll-like receptor 4 expression is required to control chronic Mycobacterium tuberculosis infection in mice.* Journal of Immunology, 2002. **169**(6): p. 3155-3162.
158. Bafica, A., et al., *TLR9 regulates Th1 responses and cooperates with TLR2 in mediating optimal resistance to Mycobacterium tuberculosis.* The Journal of experimental medicine, 2005. **202**(12): p. 1715-1724.
159. Quesniaux, V., et al., *Toll-like receptor pathways in the immune responses to mycobacteria.* Microbes and Infection, 2004. **6**(10): p. 946-959.

160. Hertz, C.J., et al., *Microbial lipopeptides stimulate dendritic cell maturation via toll-like receptor 2.* Journal of Immunology, 2001. **166**(4): p. 2444-2450.
161. Pitarque, S., et al., *The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface.* Tuberculosis, 2008. **88**(6): p. 560-565.
162. Rohde, K., et al., *Mycobacterium tuberculosis and the environment within the phagosome.* Immunological Reviews, 2007. **219**: p. 37-54.
163. Bruns, H., et al., *Anti-TNF immunotherapy reduces CD8(+) T cell-mediated antimicrobial activity against Mycobacterium tuberculosis in humans.* Journal of Clinical Investigation, 2009. **119**(5): p. 1167-1177.
164. Gengenbacher, M. and S.H.E. Kaufmann, *Mycobacterium tuberculosis: success through dormancy.* Fems Microbiology Reviews, 2012. **36**(3): p. 514-532.
165. Flynn, J.L. and J. Chan, *Immunology of tuberculosis.* Annual Review of Immunology, 2001. **19**: p. 93-129.
166. Bold, T.D. and J.D. Ernst, *Who Benefits from Granulomas, Mycobacteria or Host?* Cell, 2009. **136**(1): p. 17-19.
167. Balcewicz-Sablinska, M.K., et al., *Pathogenic Mycobacterium tuberculosis evades apoptosis of host macrophages by release of TNF-R2, resulting in inactivation of TNF-alpha.* Journal of Immunology, 1998. **161**(5): p. 2636-2641.
168. Feng, C.G., et al., *NK cell-derived IFN-gamma differentially regulates innate resistance and neutrophil response in T cell-deficient hosts infected with Mycobacterium tuberculosis.* Journal of Immunology, 2006. **177**(10): p. 7086-7093.
169. Harding, C.V. and W.H. Boom, *Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors.* Nature Reviews Microbiology, 2010. **8**(4): p. 296-307.
170. Flynn, J.A.L., et al., *Major histocompatibility complex class-I restricted T-cells are necessary for protection against Mycobacterium tuberculosis in mice.* Infectious Agents and Disease-Reviews Issues and Commentary, 1993. **2**(4): p. 259-262.
171. Sousa, A.O., et al., *Relative contributions of distinct MHC class I-dependent cell populations in protection to tuberculosis infection in mice.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(8): p. 4204-4208.
172. D'Souza, C.D., et al., *A novel nonclassic beta 2-microglobulin-restricted mechanism influencing early lymphocyte accumulation and subsequent resistance to tuberculosis in the lung.* American Journal of Respiratory Cell and Molecular Biology, 2000. **23**(2): p. 188-193.
173. Muller, I., et al., *Impaired resistance to Mycobacterium tuberculosis infection after selective invivo depletion of L3T4+ and LYT-2+ T-cells.* Infection and Immunity, 1987. **55**(9): p. 2037-2041.
174. Caruso, A.M., et al., *Mice deficient in CD4 T cells have only transiently diminished levels of IFN-gamma, yet succumb to tuberculosis.* Journal of Immunology, 1999. **162**(9): p. 5407-5416.
175. Flynn, J.L., et al., *Major histocompatibility complex class-I-restricted T-cells are required for resistance to Mycobacterium tuberculosis infection.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(24): p. 12013-12017.

176. Orme, I.M. and F.M. Collins, *Adoptive protection of the "Mycobacterium tuberculosis-infected lung - dissociation between cells that passively transfer protective immunity and those that transfer delayed-type hypersensitivity to tuberculin.* Cellular Immunology, 1984. **84**(1): p. 113-120.
177. Wu, Y., et al., *Vaccine-elicited 10-kilodalton culture filtrate protein-specific CD8(+) T cells are sufficient to mediate protection against Mycobacterium tuberculosis infection.* Infection and Immunity, 2008. **76**(5): p. 2249-2255.
178. Chen, C.Y., et al., *A Critical Role for CD8 T Cells in a Nonhuman Primate Model of Tuberculosis.* Plos Pathogens, 2009. **5**(4).
179. Cho, S., et al., *Antimicrobial activity of MHC class I-restricted CD8+T cells in human tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(22): p. 12210-12215.
180. Shams, H., et al., *Contribution of CD8(+) T cells to gamma interferon production in human tuberculosis.* Infection and Immunity, 2001. **69**(5): p. 3497-3501.
181. Lalvani, A., et al., *Human cytolytic and interferon Â -secreting CD8+ T lymphocytes specific for Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(1): p. 270-275.
182. Lewinsohn, D.A., et al., *Mycobacterium tuberculosis-specific CD8(+) T cells preferentially recognize heavily infected cells.* American Journal of Respiratory and Critical Care Medicine, 2003. **168**(11): p. 1346-1352.
183. Woodworth, J.S., Y. Wu, and S.M. Behar, *Mycobacterium tuberculosis-Specific CD8(+) T Cells Require Perforin to Kill Target Cells and Provide Protection In Vivo.* Journal of Immunology, 2008. **181**(12): p. 8595-8603.
184. Schaible, U.E., et al., *Correction of the iron overload defect in beta-2-microglobulin knockout mice by lactoferrin abolishes their increased susceptibility to tuberculosis.* Journal of Experimental Medicine, 2002. **196**(11): p. 1507-1513.
185. Urdahl, K.B., D. Liggitt, and M.J. Bevan, *CD8(+) T cells accumulate in the lungs of Mycobacterium tuberculosis-infected Kb-/-Db-/- mice, but provide minimal protection.* Journal of Immunology, 2003. **170**(4): p. 1987-1994.
186. Leveton, C., et al., *T-cell-mediated protection of mice against virulent Mycobacterium tuberculosis.* Infection and Immunity, 1989. **57**(2): p. 390-395.
187. van Pinxteren, L.A.H., et al., *Control of latent Mycobacterium tuberculosis infection is dependent on CD8 T cells.* European Journal of Immunology, 2000. **30**(12): p. 3689-3698.
188. Cyktor, J.C., et al., *Clonal Expansions of CD8(+) T Cells with IL-10 Secreting Capacity Occur during Chronic Mycobacterium tuberculosis Infection.* Plos One, 2013. **8**(3).
189. Mogues, T., et al., *The relative importance of T cell subsets in immunity and immunopathology of Airborne Mycobacterium tuberculosis infection in mice.* Journal of Experimental Medicine, 2001. **193**(3): p. 271-280.
190. Bold, T.D. and J.D. Ernst, *CD4(+) T Cell-Dependent IFN-gamma Production by CD8(+) Effector T Cells in Mycobacterium tuberculosis Infection.* Journal of Immunology, 2012. **189**(5): p. 2530-2536.
191. Shedlock, D.J. and H. Shen, *Requirement for CD4 T cell help in generating functional CD8 T cell memory.* Science, 2003. **300**(5617): p. 337-339.

192. Serbina, N.V., V. Lazarevic, and J.L. Flynn, *CD4(+) T cells are required for the development of cytotoxic CD8(+) T cells during Mycobacterium tuberculosis infection.* Journal of Immunology, 2001. **167**(12): p. 6991-7000.
193. Trinchieri, G., *Type I interferon: friend or foe?* The Journal of experimental medicine, 2010. **207**(10): p. 2053-2063.
194. Perry, A.K., et al., *The host type I interferon response to viral and bacterial infections.* Cell Research, 2005. **15**(6): p. 407-422.
195. Schnare, M., et al., *Toll-like receptors control activation of adaptive immune responses.* Nature Immunology, 2001. **2**(10): p. 947-950.
196. Modlin, R.L. and B.R. Bloom, *TB or Not TB: That Is No Longer the Question.* Science translational medicine, 2013. **5**(213): p. 213sr6-213sr6.
197. Manca, C., et al., *Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(10): p. 5752-5757.
198. Manca, C., et al., *Hypervirulent M-tuberculosis W/Beijing strains upregulate type IIFNs and increase expression of negative regulators of the Jak-Stat pathway.* Journal of Interferon and Cytokine Research, 2005. **25**(11): p. 694-701.
199. Ordway, D., et al., *The hypervirulent Mycobacterium tuberculosis strain HN878 induces a potent TH1 response followed by rapid down-regulation.* Journal of Immunology, 2007. **179**(1): p. 522-531.
200. Stanley, S.A., et al., *The type I IFN response to infection with Mycobacterium tuberculosis requires ESX-1-mediated secretion and contributes to pathogenesis.* Journal of Immunology, 2007. **178**(5): p. 3143-3152.
201. Bouchonnet, F., et al., *Alpha/beta interferon impairs the ability of human macrophages to control growth of Mycobacterium bovis BCG.* Infection and Immunity, 2002. **70**(6): p. 3020-3025.
202. Taki, S., *Type I interferons and autoimmunity: lessons from the clinic and from IRF-2-deficient mice.* Cytokine & growth factor reviews, 2002. **13**(4): p. 379-391.
203. Meléndez-Hevia, E., *From the RNA world to the DNA-protein world: clues to the origin and early evolution of life in the ribosome.* Journal of Biosciences, 2009. **34**(6): p. 825-827.
204. Lehman, N., *RNA in evolution.* Wiley Interdisciplinary Reviews-Rna, 2010. **1**(2): p. 202-213.
205. Kitamura, T., M. Peyrard, and S.C. Lopez, *A model on the origin of RNA.* Physical Biology, 2005. **2**(3): p. 200-206.
206. Crick, F.H.C., *Origin of Genetic Code.* Journal of Molecular Biology, 1968. **38**(3): p. 367-&.
207. Joyce, G.F., *RNA evolution and the origins of life.* Nature, 1989. **338**(6212): p. 217-224.
208. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-563.
209. Prescott, L.M., J.P. Harley, and D.A.-. Klein, *Microbiology*. 6th ed. ed. 2005, Dubuque, IA: McGraw-Hill Higher Education. 1 v. (various pagings) :.
210. Jeremy M. Berg, J.L.T., *Biochemistry*. Fifth ed. 2002, New York: W.H. Freeman.
211. Lee, J.C. and R.R. Gutell, *Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs.* Journal of molecular biology, 2004. **344**(5): p. 1225-1249.

212. Morris, T. and J. Dodds, *Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue.* Phytopathology, 1979. **69**(8): p. 855.
213. Patton, J.T., *Segmented double-stranded RNA viruses: structure and molecular biology*. 2008: Horizon Scientific Press.
214. Koivunen, M.R., L.P. Sarin, and D.H. Bamford, *Structure–Function Insights into the RNA-Dependent RNA Polymerase of the dsRNA Bacteriophage φ6.* Segmented Double-stranded RNA Viruses: Structure and Molecular Biology, 2008: p. 239-258.
215. Noller, H.F., *RNA structure: reading the ribosome.* Science, 2005. **309**(5740): p. 1508-1514.
216. Zhou, Y. and J. Xie, *The roles of pathogen small RNAs.* Journal of cellular physiology, 2011. **226**(4): p. 968-973.
217. Simonetti, A., et al., *A structural view of translation initiation in bacteria.* Cellular and Molecular Life Sciences, 2009. **66**(3): p. 423-436.
218. Noller, H.F., V. Hoffarth, and L. Zimniak, *Unusual resistance of peptidyl transferase to protein extraction procedures.* Science, 1992. **256**(5062): p. 1416-1419.
219. Chapeville, F., et al., *On the role of soluble ribonucleic acid in coding for amino acids.* Proceedings of the National Academy of Sciences of the United States of America, 1962. **48**(6): p. 1086.
220. Crick, F., et al., *A speculation on the origin of protein synthesis.* Origins of life, 1976. **7**(4): p. 389-397.
221. Peano, C., et al., *An efficient rRNA removal method for RNA sequencing in GC-rich bacteria.* Microbial informatics and experimentation, 2013. **3**(1): p. 1.
222. Söll, D. and U. RajBhandary, *tRNA structure, biosynthesis, and function*. 1995, Washington, D.C.: ASM Press. 572.
223. Lengyel, P. and D. Söll, *Mechanism of protein biosynthesis.* Bacteriological reviews, 1969. **33**(2): p. 264.
224. Deutscher, M.P., *Degradation of RNA in bacteria: comparison of mRNA and stable RNA.* Nucleic acids research, 2006. **34**(2): p. 659-666.
225. Smith, J., *Genetics of transfer RNA.* Annual review of genetics, 1972. **6**(1): p. 235-256.
226. Phizicky, E.M. and A.K. Hopper, *tRNA biology charges to the front.* Genes & Development, 2010. **24**(17): p. 1832-1860.
227. Arraiano, C.M., A. Barbas, and M. Amblar, *Characterizing ribonucleases in vitro: examples of synergies between biochemical and structural analysis*, in *RNA Turnover in Bacteria, Archaea and Organelles*. 2008, Elsevier Academic Press Inc: San Diego. p. 131-+.
228. Wagner, E.G.H. and F. Darfeuille, *Small regulatory RNAs in bacteria*, in *Small RNAs*. 2006, Springer. p. 1-29.
229. Waters, L.S. and G. Storz, *Regulatory RNAs in Bacteria.* Cell, 2009. **136**(4): p. 615-628.
230. Vogel, J. and E.G.H. Wagner, *Target identification of small noncoding RNAs in bacteria.* Current Opinion in Microbiology, 2007. **10**(3): p. 262-270.
231. Arnvig, K.B. and D.B. Young, *Identification of small RNAs in Mycobacterium tuberculosis.* Molecular microbiology, 2009. **73**(3): p. 397-408.
232. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.

233. DiChiara, J.M., et al., *Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis.* Nucleic Acids Research, 2010. **38**(12): p. 4067-4078.
234. Pelly, S., W.R. Bishai, and G. Lamichhane, *A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection.* Gene, 2012. **500**(1): p. 85-92.
235. Harapan, H., et al., *The roles of microRNAs on tuberculosis infection: Meaning or myth?* Tuberculosis, 2013. **93**(6): p. 596-605.
236. Alexopoulou, L., et al., *Recognition of double-stranded RNA and activation of NF-[kappa]B by Toll-like receptor 3.* Nature, 2001. **413**(6857): p. 732-738.
237. Kumar, S., et al., *Recognition of bacterial infection by innate immune sensors.* Critical Reviews in Microbiology, 2012(00): p. 1-18.
238. Saitoh, S. and K. Miyake, *Regulatory molecules required for nucleotide-sensing Toll-like receptors.* Immunological Reviews, 2009. **227**(1): p. 32-43.
239. Kumagai, Y., O. Takeuchi, and S. Akira, *TLR9 as a key receptor for the recognition of DNA.* Advanced Drug Delivery Reviews, 2008. **60**(7): p. 795-804.
240. Lan, T., et al., *Stabilized immune modulatory RNA compounds as agonists of Toll-like receptors 7 and 8.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(34): p. 13750-13755.
241. Kawai, T. and S. Akira, *The roles of TLRs, RLRs and NLRs in pathogen recognition.* International Immunology, 2009. **21**(4): p. 317-337.
242. Cervantes, J.L., et al., *Human TLR8 is activated upon recognition of Borrelia burgdorferi RNA in the phagosome of human monocytes.* Journal of Leukocyte Biology, 2013. **94**(6): p. 1231-1241.
243. Davila, S., et al., *Genetic Association and Expression Studies Indicate a Role of Toll-Like Receptor 8 in Pulmonary Tuberculosis.* PLoS Genet, 2008. **4**(10): p. e1000218.
244. Guidotti, L.G. and F.V. Chisari, *Noncytolytic control of viral infections by the innate and adaptive immune response.* Annual Review of Immunology, 2001. **19**: p. 65-91.
245. Heil, F., et al., *Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8.* Science, 2004. **303**(5663): p. 1526-1529.
246. Sarvestani, S.T., B.R. Williams, and M.P. Gantier, *Human Toll-like receptor 8 can be cool too: implications for foreign RNA sensing.* Journal of Interferon & Cytokine Research, 2012. **32**(8): p. 350-361.
247. Hemmi, H., et al., *A Toll-like receptor recognizes bacterial DNA.* Nature, 2000. **408**(6813): p. 740-745.
248. Heeg, K., et al., *Structural requirements for uptake and recognition of CpG oligonucleotides.* International Journal of Medical Microbiology, 2008. **298**(1-2): p. 33-38.
249. Latz, E., et al., *TLR9 signals after translocating from the ER to CpG DNA in the lysosome.* Nature Immunology, 2004. **5**(2): p. 190-198.
250. Cheung, B.K.W., et al., *A role for double-stranded RNA-activated protein kinase PKR in Mycobacterium-induced cytokine expression.* Journal of Immunology, 2005. **175**(11): p. 7218-7225.
251. Clemens, M.J. and A. Elia, *The double-stranded RNA-dependent protein kinase PKR: Structure and function.* Journal of Interferon and Cytokine Research, 1997. **17**(9): p. 503-524.

252. Thompson, A.J.V. and S.A. Locarnini, *Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response.* Immunol Cell Biol, 2007. **85**(6): p. 435-445.
253. Meylan, E. and J. Tschopp, *Toll-like receptors and RNA helicases: Two parallel ways to trigger antiviral responses.* Molecular Cell, 2006. **22**(5): p. 561-569.
254. Yoneyama, M., et al., *The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses.* Nature Immunology, 2004. **5**(7): p. 730-737.
255. Satoh, T., et al., *LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(4): p. 1512-1517.
256. Kanneganti, T.-D., et al., *Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3.* Nature, 2006. **440**(7081): p. 233-236.
257. Moon, S.L., et al., *A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability.* RNA, 2012. **18**(11): p. 2029-40.
258. Farber, J.M. and P.I. Peterkin, *Listeria monocytogenes, a food-borne pathogen.* Microbiological Reviews, 1991. **55**(3): p. 476-511.
259. O'Connell, R.M., et al., *Type I interferon production enhances susceptibility to Listeria monocytogenes infection.* Journal of Experimental Medicine, 2004. **200**(4): p. 437-445.
260. Monroe, K.M., S.M. McWhirter, and R.E. Vance, *Induction of type I interferons by bacteria.* Cellular Microbiology, 2010. **12**(7): p. 881-890.
261. O'Riordan, M., et al., *Innate recognition of bacteria by a macrophage cytosolic surveillance pathway.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(21): p. 13861-13866.
262. Crimmins, G.T., et al., *Listeria monocytogenes multidrug resistance transporters activate a cytosolic surveillance pathway of innate immunity.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(29): p. 10191-10196.
263. Mandin, P., et al., *Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets.* Nucleic Acids Research, 2007. **35**(3): p. 962-974.
264. Feltcher, M.E. and M. Braunstein, *Emerging themes in SecA2-mediated protein export.* Nature Reviews Microbiology, 2012. **10**(11): p. 779-789.
265. Van Pittius, N.G., et al., *The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+ C Gram-positive bacteria.* Genome Biol, 2001. **2**(10): p. 44.1-44.18.
266. Hagmann, C.A., et al., *RIG-I Detects Triphosphorylated RNA of Listeria monocytogenes during Infection in Non-Immune Cells.* Plos One, 2013. **8**(4).
267. Auerbuch, V., et al., *Mice lacking the type I interferon receptor are resistant to Listeria monocytogenes.* Journal of Experimental Medicine, 2004. **200**(4): p. 527-533.
268. Carrero, J.A., B. Calderon, and E.R. Unanue, *Type I interferon sensitizes lymphocytes to apoptosis and reduces resistance to listeria infection.* Journal of Experimental Medicine, 2004. **200**(4): p. 535-540.
269. Rowbotham, T.J., *Preliminary-report on the pathogenicity of Legionella pneumophila for fresh-water and soil amebas.* Journal of Clinical Pathology, 1980. **33**(12): p. 1179-1183.
270. Isberg, R.R., T.J. O'Connor, and M. Heidtman, *The Legionella pneumophila replication vacuole: making a cosy niche inside host cells.* Nature Reviews Microbiology, 2009. **7**(1): p. 12-24.

271. Opitz, B., et al., *Legionella pneumophila induces IFN beta in lung epithelial cells via IPS-1 and IRF3, which also control bacterial replication.* Journal of Biological Chemistry, 2006. **281**(47): p. 36173-36179.
272. Chiu, Y.H., J.B. MacMillan, and Z.J.J. Chen, *RNA Polymerase III Detects Cytosolic DNA and Induces Type I Interferons through the RIG-I Pathway.* Cell, 2009. **138**(3): p. 576-591.
273. Mancuso, G., et al., *Bacterial recognition by TLR7 in the lysosomes of conventional dendritic cells.* Nature Immunology, 2009. **10**(6): p. 587-U48.
274. Heath, P.T. and A. Schuchat, *Perinatal group B streptococcal disease.* Best Practice & Research in Clinical Obstetrics & Gynaecology, 2007. **21**(3): p. 411-424.
275. Charrel-Dennis, M., et al., *TLR-Independent Type I Interferon Induction in Response to an Extracellular Bacterial Pathogen via Intracellular Recognition of Its DNA.* Cell Host & Microbe, 2008. **4**(6): p. 543-554.
276. Blaser, M.J. and J.C. Atherton, *Helicobacter pylori persistence: biology and disease.* Journal of Clinical Investigation, 2004. **113**(3): p. 321-333.
277. Blaser, M.J., *Helicobacters are indigenous to the human stomach: duodenal ulceration is due to changes in gastric microecology in the modern era.* Gut, 1998. **43**(5): p. 721-727.
278. Blaser, M.J. and D.E. Berg, *Helicobacter pylori genetic diversity and risk of human disease.* Journal of Clinical Investigation, 2001. **107**(7): p. 767-773.
279. Gantier, M.P., et al., *Genetic Modulation of TLR8 Response following Bacterial Phagocytosis.* Human Mutation, 2010. **31**(9): p. 1069-1079.
280. Molinari, M., et al., *Selective inhibition of Ii-dependent antigen presentation by Helicobacter pylori toxin VacA.* Journal of Experimental Medicine, 1998. **187**(1): p. 135-140.
281. Rad, R., et al., *Extracellular and Intracellular Pattern Recognition Receptors Cooperate in the Recognition of Helicobacter pylori.* Gastroenterology, 2009. **136**(7): p. 2247-2257.
282. Levi, T., et al., *Deer, predators, and the emergence of Lyme disease.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(27): p. 10942-10947.
283. Smith, R. and J. Takkinen, *Lyme borreliosis: Europe-wide coordinated surveillance and action needed?* Euro surveillance: bulletin Européen sur les maladies transmissibles= European communicable disease bulletin, 2006. **11**(6): p. E060622. 1.
284. Radolf, J.D., et al., *Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes.* Nature Reviews Microbiology, 2012. **10**(2): p. 87-99.
285. Biesiada, G., et al., *Lyme disease: review.* Arch Med Sci, 2012. **8**(6): p. 978-82.
286. Salazar, C.A., et al., *Human Lyme arthritis and the immunoglobulin G antibody response to the 37-kilodalton arthritis-related protein of Borrelia burgdorferi.* Infection and Immunity, 2005. **73**(5): p. 2951-2957.
287. Cervantes, J.L., et al., *Phagosomal signaling by Borrelia burgdorferi in human monocytes involves Toll-like receptor (TLR) 2 and TLR8 cooperativity and TLR8-mediated induction of IFN-β.* Proceedings of the National Academy of Sciences, 2011. **108**(9): p. 3683-3688.
288. Pandey, A.K., et al., *NOD2, RIP2 and IRF5 Play a Critical Role in the Type I Interferon Response to Mycobacterium tuberculosis.* Plos Pathogens, 2009. **5**(7).

289. Obregon-Henao, A., et al., *Stable extracellular RNA fragments of Mycobacterium tuberculosis induce early apoptosis in human monocytes via a caspase-8 dependent mechanism.* PLoS One, 2012. **7**(1): p. e29970.
290. Thada, S., V. Valluri, and S.L. Gaddam, *Influence of toll like receptor gene polymorphisms to tuberculosis susceptibility in humans.* Scandinavian journal of immunology, 2013.
291. Sioud, M., *Single-stranded small interfering RNA are more immunostimulatory than their double-stranded counterparts: A central role for 2 '-hydroxyl uridines in immune responses.* European Journal of Immunology, 2006. **36**(5): p. 1222-1230.
292. Vollmer, J., et al., *Immune stimulation mediated by autoantigen binding sites within small nuclear RNAs involves Toll-like receptors 7 and 8.* Journal of Experimental Medicine, 2005. **202**(11): p. 1575-1585.
293. Wiker, H.G., M. Harboe, and S. Nagai, *A localization index for distinction between extracellular and intracellular antigens of Mycobacterium tuberculosis.* Journal of General Microbiology, 1991. **137**: p. 875-884.

Chapter II

Development of Culture Filtrate RNA Isolation Protocol

2.1 Introduction

The first step in further evaluating the CF RNA population was to develop a more facile RNA purification scheme. Although the method used by Obregón-Henao et al. was effective in identifying the biologically active component of the *M. tuberculosis* CF, it was originally designed to look for protein and therefore not optimized for RNA purification [1]. A more streamlined method specifically designed for isolating RNA from large amounts of *M. tuberculosis* CF was needed.

In the previous studies, CF was initially treated with ConcanavalinA (ConA) to remove ManLAM, which had been shown to inhibit macrophage apoptosis [2-4]. The ManLAM depleted CF was then fractionated using DEAE-Sephadex and a stepwise gradient of sodium chloride, yielding seven fractions. The fractions were evaluated for apoptotic activity and the most active fraction, fraction seven, was then treated with Proteinase K, DNaseI, and RNaseVI. As mentioned in Chapter I, the activity of fraction seven was abolished with RNaseVI treatment, leading to the conclusion that RNA was the apoptotic factor. Fraction seven was further purified by treatment with DNaseI and, Proteinase K, followed by phenol:chloroform:isoamyl alcohol extraction (IAA). The CF RNA was then gel purified from a denaturing urea-polyacrylamide gel [1].

Many commercial kits are available for the purification of total RNA and sRNA. While these kits are effective for many applications, they did not fit the specific needs of this project. Kits are typically designed for the isolation of RNA from relatively small amounts of material

including cells and tissue; for this project it was important to develop a method that could be used to isolate RNA from large volumes of *M. tuberculosis* CF. RNA isolation kits generally utilize a guanidinium thiocyanate-phenol-chloroform (TRIzol) extraction method which can be coupled with a column based purification. Both the TRIzol and column based methods would be problematic for this project as they are inefficient for large volumes of sample. This chapter details the development of a purification scheme for RNA from large batches of H37Rv CF.

2.2 Materials and Methods

2.2.1 Bacterial Strains and growth conditions

M. tuberculosis H37Rv was grown from glycerol stocks on 7H11 agar plates (Difco, Detroit, MI) before passing into 100 ml glycerol alanine salts (GAS) medium at 37°C [5]. After 14 days of incubation with agitation, each culture was passed into one-liter of GAS medium. After an additional 14 days of incubation with agitation, one-liter cultures were used to upscale for large batch cultures. Each one-liter culture was split into five separate one-liter cultures, for a total of eight to ten one-liter cultures. Large batch cultures were harvested at day 14, late log phase.

For the time-course studies, one-liter cultures were split into five separate one-liter cultures, with two liters per time point. Before inoculating time course cultures, cells were centrifuged at 3,000 rpm and washed three times with GAS medium before inoculating two one-liter flasks per time point. Time course cultures were harvested at days 3, 7, 10, 14, and 21.

2.2.2 Culture filtrate

CF was obtained by removing the cell pellet from all cultures and running remaining medium through a 0.2 μm VacuCap™ bottle top filter (Pall Life Sciences, Port Washington, NY).

Large batch CF (8L) was concentrated by decreasing the volume to 150 ml using the Millipore Pellicon tangential flow filtration system and Pellicon 2 Ultrafiltration cassettes (EMD Millipore, Billerica, MA). The Pellicon system was cleaned with 0.1M sodium hydroxide for ten minutes, followed by a water flush for ten minutes. During the water flush, the pump speed and feed pressure of the system were adjusted to achieve a flux rate of 30%. The flux rate was calculated by measuring the retentate and permeate volume after 30 seconds and dividing the permeate volume by the total volume. Before concentrating CF, 10 mM ammonium bicarbonate (Sigma-Alderich, St. Louis, MO) was run through the system until the pH matched that of the ammonium bicarbonate buffer (pH 7.8). After concentration, CF was dialyzed against 10 mM ammonium bicarbonate.

Time course CF (2L) was concentrated by decreasing the volume to 50 ml using the Amicon Ultrafiltration Cell (EMD Millipore, Billerica, MA) with a Ultracel 10 kDa molecular-weight cutoff, regenerated cellulose membrane (EMD Millipore, Billerica, MA). Membranes were hydrated with 90% isopropanol for ten minutes followed by diH_2O for 30 minutes.

2.2.3 RNA Purification

Concentrated large batch samples were lyophilized, reconstituted in 15 ml RNase-free water, and dried twice using the Savant SPD2010 SpeedVac Concentrator. After the last drying step, samples were resuspended in 60 ml of 50 mM Tris HCl, 150 mM NaCl buffer.

Concentrated time course samples were reconstituted in 1 ml RNase-free water and suspended in 3 mL of the Tris NaCl buffer.

CF was then digested for two hours at room temperature with 1mM $MgCl_2$ at a final concentration of 0.4 mg/ml, with 10 µg/µl RNase-free DNaseI (Sigma-Aldrich, St. Louis, MO). DNase digestion was followed by incubation with 100 µg/ml RNase-free Proteinase K (IBI Scientific, Peosta, IA) and 10% RNase, DNase-free sodium dodecyl sulfate (SDS) (Fisher BioReagents, Fair Lawn, NJ), for a final concentration of 0.5%, for 8 hours at 55°C.

Proteinase K digestion was followed by the addition of an equal volume of phenol:chloroform:isoamyl alcohol (IAA) (25:24:1) to separate layers. Layers were separated by centrifugation at 3000 x g. The aqueous layer was transferred to a new tube and an equal volume of chloroform:IAA (24:1) was added to separate any remaining phenol. After centrifugation, the aqueous layer from this step was taken and used for further RNA isolation.

2.2.3.1 Removal of LAM

Four different methods were tested for the removal of LAM from the CF RNA: ethanol precipitation, isopropanol precipitation, passing over a ConA column, and passing over a C18 column.

For ethanol precipitation, three volumes of 100% ethanol (EMD Millipore, Billerica, MA) were added for every one volume of sample and put on ice to incubate for one hour. After incubation, samples were centrifuged at 14,000 x g for 30 minutes at 4°C to pellet RNA. Supernatant was removed, and RNA was washed twice with cold 70% ethanol and centrifuged at 14,000 x g for 10 minutes at 4°C. After the final wash the supernatant was removed, and ethanol was allowed to dry. Isopropanol precipitation was performed in a similar mannor except that

two volumes of isopropanol were used for every one volume of CF RNA; washes were also performed with 70% isopropanol.

For the ConA method, two ml of ConA Sepharose resin (GE Healthcare, Germany) was prepared and mixed with one ml of the CF RNA, after the chloroform:IAA extraction, for four hours at room temperature. After incubation, the CF RNA/ConA mixture was added to a column, and flow-through was collected. After initial flow-through was collected, the column was washed with ConA buffer, followed by ConA elution buffer.

For the C18 method of removing LAM, Sep-Pak Vac RC 500 mg and 100 mg C18 cartridges were used (Waters, Ireland). Columns were prepared for use by hydrating using one volume of LC-MS grade Methanol (Honeywell Burdick & Jackson, Muskegon, MI), and were subsequently washed four times with sterile, RNase free H_2O. For the large batch RNA purification, five ml of CF RNA were added to the 500 mg C18 column, and the flow-through was collected. For the time course RNA purification, one ml of CF RNA was added to the 100 mg C18 column, and the flow-through was collected. Each column was washed at least one time with an equal volume of 50 mM Tris HCl, 150 mM NaCl buffer to remove any remaining RNA.

2.2.3.2 CF RNA ethanol precipitation

After removal of LAM using the C18 method, CF RNA from each column was further concentrated and purified through ethanol precipitation (as described above). Once ethanol was removed, CF RNA pellets were re-suspended in 150 µl RNase-free water, combining the flow-through and wash from each column. RNA was stored at -80°C until use.

2.2.4 RNA Quality assurance/quality control

2.2.4.1 SDS-PAGE Gels: Silver Stain and Western Blot

NuPAGE 4-12% Bis-Tris (Invitrogen, Carlsbad, CA) SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) gels were used for Western blotting to address LAM contamination and silver staining to address protein contamination. CF RNA samples were added to a sample reducing agent and LDS Sample Buffer before loading onto the SDS-PAGE gel (Invitrogen, Carlsbad, CA). A pre-stained protein standard was run on each gel to indicate protein size (Bio-Rad, Hercules, CA). MES SDS Running Buffer was used to run the SDS-PAGE gels at 200 Volts (V) for approximately 45 minutes (Invitrogen, Carlsbad, CA). Silver staining was performed as previously described [6] to ensure protein had been removed during the RNA purification process.

SDS-PAGE gels were transferred onto a nitrocellulose membrane (Bio-Rad, Hercules, CA) at 50 V for 80 minutes. After transfer, membranes were blocked for one hour at room temperature with 3% BSA in 10 ml TBS. Membranes were then probed with a LAM-specific monoclonal antibody CS-35 overnight at 4°C, to test CF RNA for the presence of LAM [7]. Following primary antibody incubation, membranes were washed with 0.5% Tween-20 in TBS (TBS-T) three times for five minutes each. The membranes were then subjected to the secondary antibody, an Anti-Mouse IgG alkaline phosphatase, developed in goats, diluted 1:1000 in TBS. Membranes were incubated for 45 minutes with the diluted secondary antibody, followed by washes, as before, with TBS-T. Finally, membranes were developed using NBT/BCIP alkaline phosphatase tablets (Sigma, St. Louis, MI).

2.2.4.2 TBE-Urea Gels: Ethidium bromide staining

Novex 15% Tris Borate EDTA (TBE)-Urea gels were used for Ethidium bromide staining of RNA to visualize size and purity (Invitrogen, Carlsbad, CA). RNA was added to TBE-Urea Sample Buffer before loading onto the TBE-Urea gels, and run using TBE Running Buffer at 200 V for approximately one hour (Invitrogen, Carlsbad, CA). RNA size was estimated using the RNA Century size marker (Invitrogen, Carlsbad, CA).

TBE-Urea gels were stained in a 1 μg/ml Ethidium bromide (EtBr) bath for 15 minutes and de-stained for 10 minutes in water. EtBr stained gels were imaged on a Bio-Rad Gel Doc at 302 nm (Bio-Rad, Hercules, CA).

2.2.4.3 NanoDrop

The NanoDrop Spectrophotometer ND-1000 was used to determine the concentration of RNA. All samples were run in triplicate at three concentrations (undiluted, 1:10, and 1:100), along with two different standard curves of tRNA. Standard curves were made using tRNA from *E. coli* MRE600 (Roche, Mannheim, Germany). Standard Curve 1 consists of tRNA at: 4000 ng/μl, 3000 ng/μl, 2000 ng/μl, 1000 ng/μl (Figure 2.1 A). Standard Curve 2 consists of tRNA at: 750 ng/μl, 500 ng/μl, 250 ng/μl, 100 ng/μl, 75 ng/μl, 50 ng/μl, 25 ng/μl (Figure 2.1 B). The amount of RNA was calculated based on the appropriate tRNA standard curve.

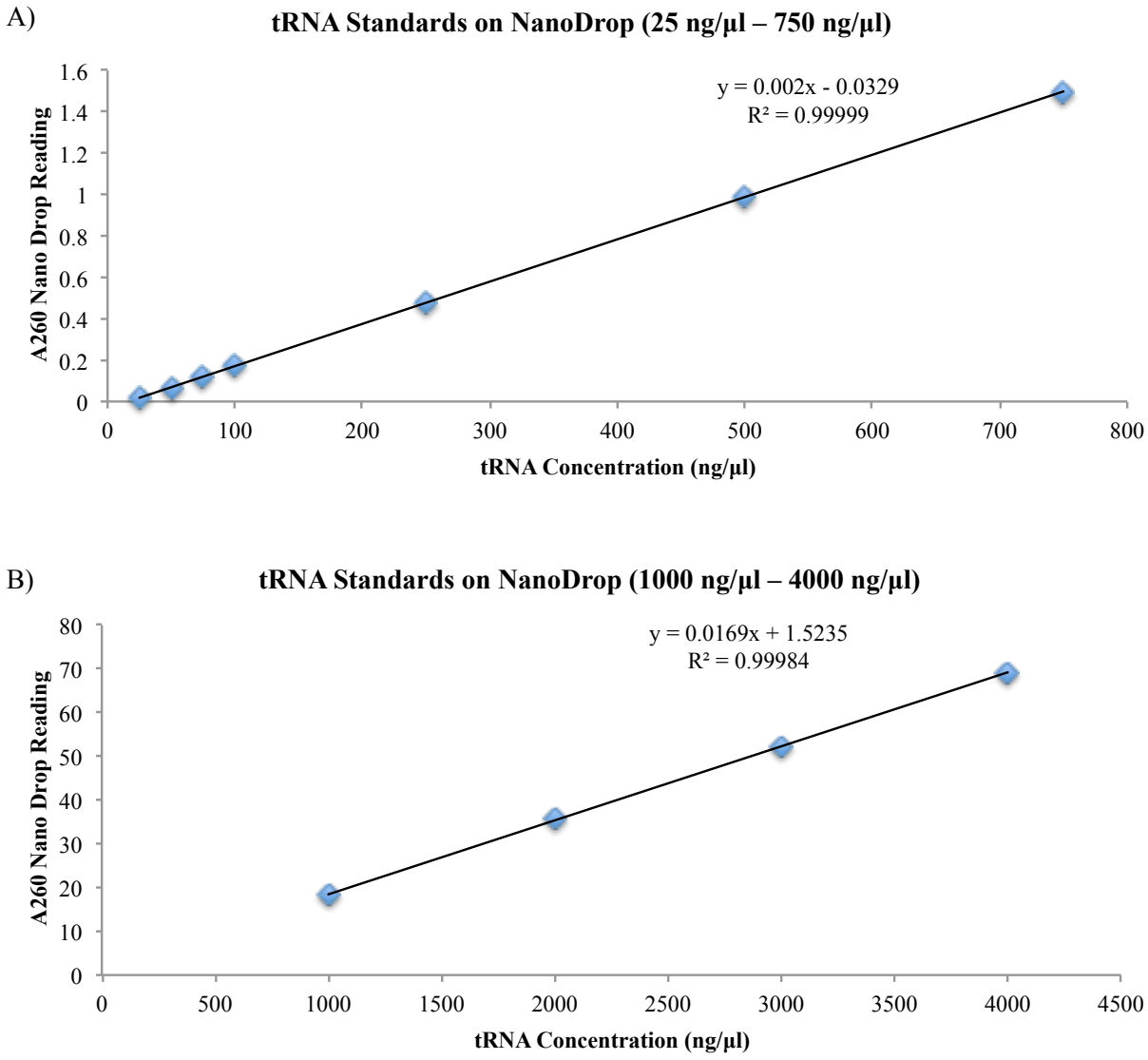


Figure 2.1 NanoDrop tRNA standard curves. NanoDrop standard curve with tRNA standards ranging between A) 25-750 ng/µl and B) 1000-4000 ng/µl.

2.2.4.4 Bioanalyzer

The Agilent 2100 bioanalyzer was used to perform on-chip gel electrophoresis, using different chips to detect different types of samples in different size ranges. The Small RNA chip (Agilent Technologies, Germany) detects RNA between 6-150 nucleotides (nt) in length, and

was used to look at sRNA. The small RNA chips have a concentration range of 50-2000 pg/µl, so the RNA samples were run at the appropriate dilution based on NanoDrop results.

CF RNA samples were also tested on an Agilent Technologies 12000 DNA chip to detect possible DNA contamination. The DNA chips are able to detect DNA between 100-12000 base pairs (bp) in length with a concentration range of 0.5-50 ng/µl.

2.2.4.5 LAL

Endotoxin levels were measured using the Limulum Amebocyte Lysate (LAL) Kinetic-QCL (Lonza, Walkersville, MD) according to the manufacturer's instructions. Each sample was tested using at least two different dilutions.

2.2.5 Kinetics of Release

In order to assess the accumulation of RNA in the CF over time compared to the rate of protein release, protein concentrations were measured using the Micro Bicinchoninic acid (BCA) Protein Assay (Thermo Scientific, Rockford, IL) following the manufacture's procedure. 96 well plates with bovine serum albumin (BSA) standards, samples, and working reagent were incubated at 37°C for 30 minutes and read using the BioTek Synergy 2 plate reader (Winooski, VT) at 562 nm. Comparing sample optical densities to the BSA standard curve using the Gen5 BioTek Software determined protein concentration.

2.2.6 Intracellular Small RNA Extraction

The FastRNA Pro Blue Kit (MP Biomedicals, Santa Ana, CA) was used to extract total *M. tuberculosis* RNA from day 14 late log phase cells, according to the manufacturer's

instructions. This kit uses a bead-beating cell membrane disruption, in combination with solutions to inactivate cellular RNases during cell lysis to prevent RNA degradation. Bead-beating is followed by purification and isolation by chloroform extraction and ethanol precipitation.

Once total RNA had been isolated, the PureLink miRNA Isolation Kit (Invitrogen, Carlsbad, CA) was used to isolate the sRNA. The kit was used according to the manufacturer's instructions, followed by a second extraction using a modified protocol. For the second extraction, one ml TRIzol reagent (Life Technologies, Carlsbad, CA) was added to the sample and incubated for five minutes at room temperature. After incubation, 0.2 ml chloroform was added and mixed by vortexing, followed by a two-minute incubation at room temperature. Samples were centrifuged at 12,000 x g for 15 minutes at 4°C, and the resulting aqueous layer was transferred to a new tube along with an equal volume of 70% ethanol. The sample was then added to the kit spin column, and the manufacturer's instructions were followed from step four.

Bioanalyzer RNA Nano chips (Agilent Technologies, Germany) were used to determine if rRNA had been removed, leaving only the sRNA. RNA samples were diluted based on the NanoDrop, as the RNA Nano chips have a concentration range of 25-500 ng/ul.

2.3 Results

2.3.1 RNA purification

The first critical step in the purification of CF RNA was to concentrate the large amount of CF harvested. The Pellicon tangential flow filtration system allows for large amounts of material to be concentrated quickly and efficiently (8-10 L in approximately two hours).

Once the material had been concentrated and reconstituted in buffer, DNase and Proteinase K digestions were done in order to digest DNA and protein in the concentrated CF. EtBr stained TBE-Urea gels were used to show the removal of DNA after DNase digestion (Figure 2.2 A), and silver-stained SDS-PAGE gels were used to confirm removal of protein (Figure 2.2 B). During method development, a faint band around the size expected of LAM was observed on the silver stained gel (Figure 2.2 B). LAM is typically not seen on silver stained gels unless the staining procedure includes a periodic acid oxidation step, used to view carbohydrates [8]. Despite the lack of the periodic acid oxidization step, a large amount of LAM can sometimes appear as a faint band. The presence of LAM in the CF samples was confirmed using the CS-35 monoclonal antibody for LAM in a Western blot (Figure 2.2 C), although the LAM detected through Western blotting (Figure 2.2 C) appears at a higher molecular weight than the faint band seen in the silver stained gel (Figure 2.2 B). The band seen through silver staining could also be lipomannan (LM), which is often associated with LAM and is smaller in size [9].

Discovery of LAM contamination in the CF RNA led to the evaluation of several methods for their ability to remove LAM (Figure 2.2 C). Ethanol precipitation and isopropanol precipitation were not successful in removing LAM. The C18 and ConA columns were both able to remove LAM from the CF RNA without greatly reducing RNA yield. Because they are commercially available and easy to use, the C18 column method was chosen for LAM removal from the CF. The final RNA purification scheme with rational for each step is depicted in Figure 2.3.

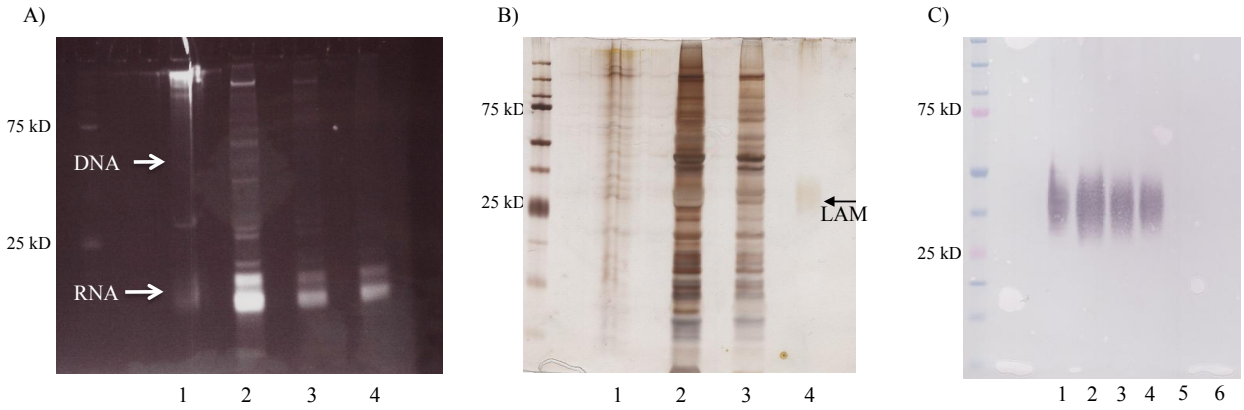


Figure 2.2 Gel based analysis of Purified CF RNA. A) Ethidium bromide stained PAGE gel shows the removal of DNA. Lane 1, CF before lyophilization; lane 2, CF after lyophilization; lane 3, CF diluted in Tris/NaCl buffer; lane 4, CF after phenol:chloroform:IAA and chloroform:IAA. B) Silver-stained PAGE gel shows the removal of protein, as well as the presence of LAM. Lane 1, CF before lyophilization; lane 2, CF after lyophilization; lane 3, CF diluted in Tris/NaCl buffer; lane 4, CF after phenol:chloroform:IAA and chloroform:IAA. C) Western blot with CS35 anti-LAM showing that C18 solid phase extraction and ConA chromatography efficiently removed LAM from the CF. Lane 1, CF after concentration; lane 2, CF after phenol:chloroform:IAA and chloroform:IAA; lane 3, after ethanol precipitation; lane 4, after isopropanol precipitation; lane 5 C18 flow-through; lane 6, after ConA chromatography.

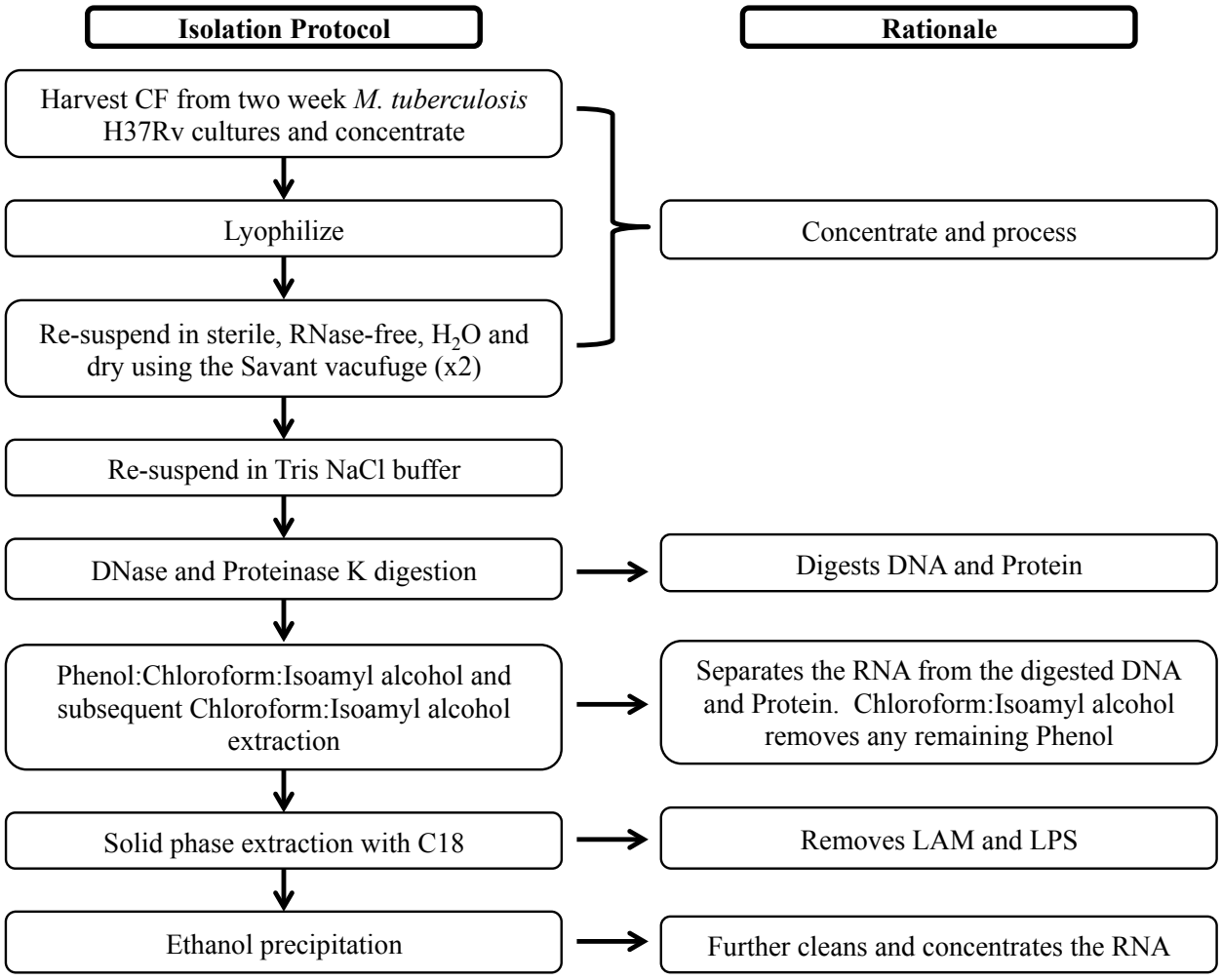


Figure 2.3 RNA purification protocol with rational. Workflow of CF RNA purification protocol developed with the rational for each step.

2.3.2 RNA quality assurance/quality control

Several methods of QA/QC were used to determine the concentration and purity of RNA in each batch including: silver staining, Western blotting, EtBr staining, NanoDrop, and Bioanalyzer analysis. The silver staining, Western blotting for LAM, and EtBr staining were all used during the method development, and were later used as visual confirmation of purity, ensuring the RNA preparations were free of protein, LAM, and DNA contaminants. The silver

stain was used to confirm the removal of protein after Proteinase K digestion, and EtBr staining was used to show removal of DNA as well as the approximate size of RNA.

For down-stream studies, it was critical to ensure consistency and exact concentration. The NanoDrop spectrophotometer was used to obtain an accurate concentration of the RNA. The secondary structure of tRNA was a concern in the gaining accurate concentrations. In order to account for the affect secondary structure may have on the accuracy of NanoDrop readings, two tRNA standard curves were used in order to calculate the concentration of the RNA, rather than the values given by the NanoDrop software. Using the NanoDrop and tRNA standards, a typical large batch of CF will yield 2-3 mg of purified RNA. It is important to note that yield can vary, as RNA can be lost throughout the isolation process.

The Bioanalyzer small RNA chip was used as an additional method of verifying the concentration given by the NanoDrop, as well as a way of obtaining a more accurate reading of the RNA size seen in the EtBr stained gel. The Bioanalyzer 12000 DNA chip was used to confirm the absence of DNA seen in the EtBr stained gel. Table 2.1 shows four different purified CF RNA samples to demonstrate the differences in concentration measurements obtained using the NanoDrop with and without the tRNA standards as well as the Bioanalyzer small RNA chip. Without the tRNA standard curves the NanoDrop calculates the concentration to be nearly two fold higher than the concentrations measured using the standards and the Bioanalyzer.

Table 2.1 Purified CF RNA sample concentrations. Shows the differences between three different methods of quantifying CF RNA sample concentration: NanoDrop without tRNA standards, NanoDrop with tRNA standards, small RNA Bioanalyzer chip.

CF RNA	Concentration given by NanoDrop without tRNA standards (µg/µl)	NanoDrop concentration calculated using tRNA standard curve (µg/µl)	Concentration given by Bioanalyzer small RNA chip (µg/µl)
CF RNA 1	2.7	1.7	1.1
CF RNA 2	1.5	0.6	0.6
CF RNA 3	2.0	1.1	1.1
CF RNA 4	3.5	2.4	1.9
Average	2.4	1.5	1.2

The CF RNA samples displayed in Table 2.1 are also shown in Figure 2.4 using a small RNA Bioanalyzer chip and an EtBr stained gel. The Bioanalyzer traces shown in Figure 2.4 A reflect both the concentrations calculated with the tRNA standards as well as the size seen in the EtBr gel (Figure 2.4 B). All of the CF RNA samples shown have similar peaks with the size predominantly between 30-70 bases.

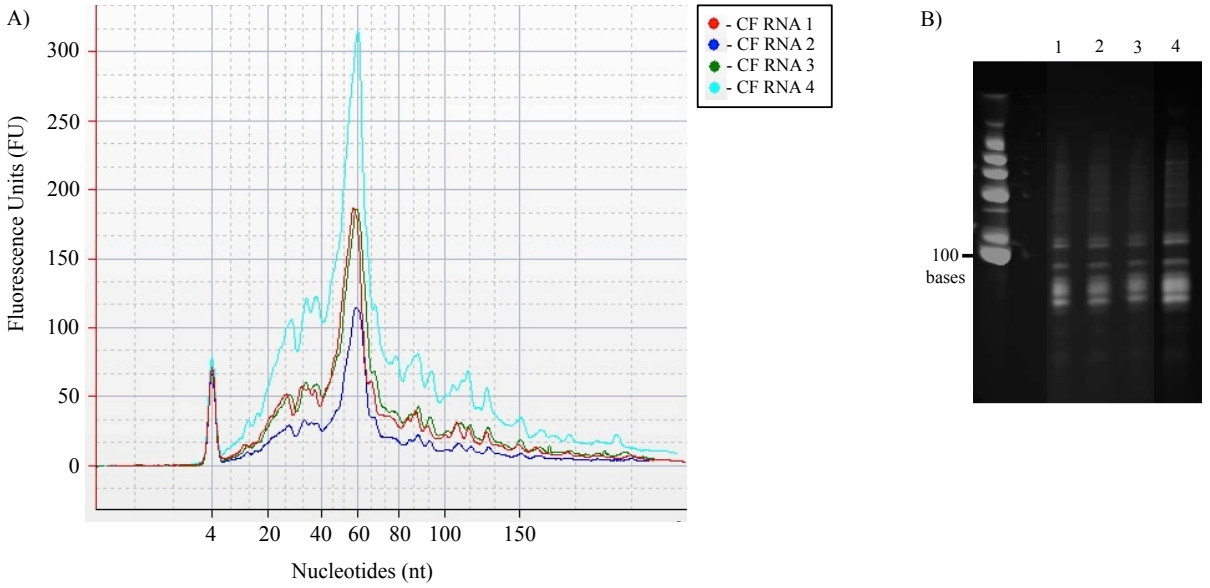


Figure 2.4 Bioanalyzer and EtBr stained gel RNA analysis. A) Bioanalyzer trace of four CF RNA samples. B) EtBr stained gel of four CF RNA samples.

The LAL endotoxin assay was used to ensure LPS had not been introduced during the purification process, which could have an effect on down stream biological studies. The amount of endotoxin in CF RNA samples was expressed in ng of endotoxin per µg of RNA. The average amount of endotoxin detected in the purified CF RNA samples was 2.6 ng/µg.

2.3.3 Time course CF RNA

Smaller cultures were evaluated for time course studies to evaluate RNA release into the CF over time. Time course studies showed low levels of CF RNA starting as early as day three, increasing over time, which is illustrated by the Bioanalyzer trace overlay in Figure 2.5. The size of the extracellular RNA is consistent throughout the time course and the large batch. The main peak is between 30 and 70 bases in length, corresponding to what was seen previously [1].

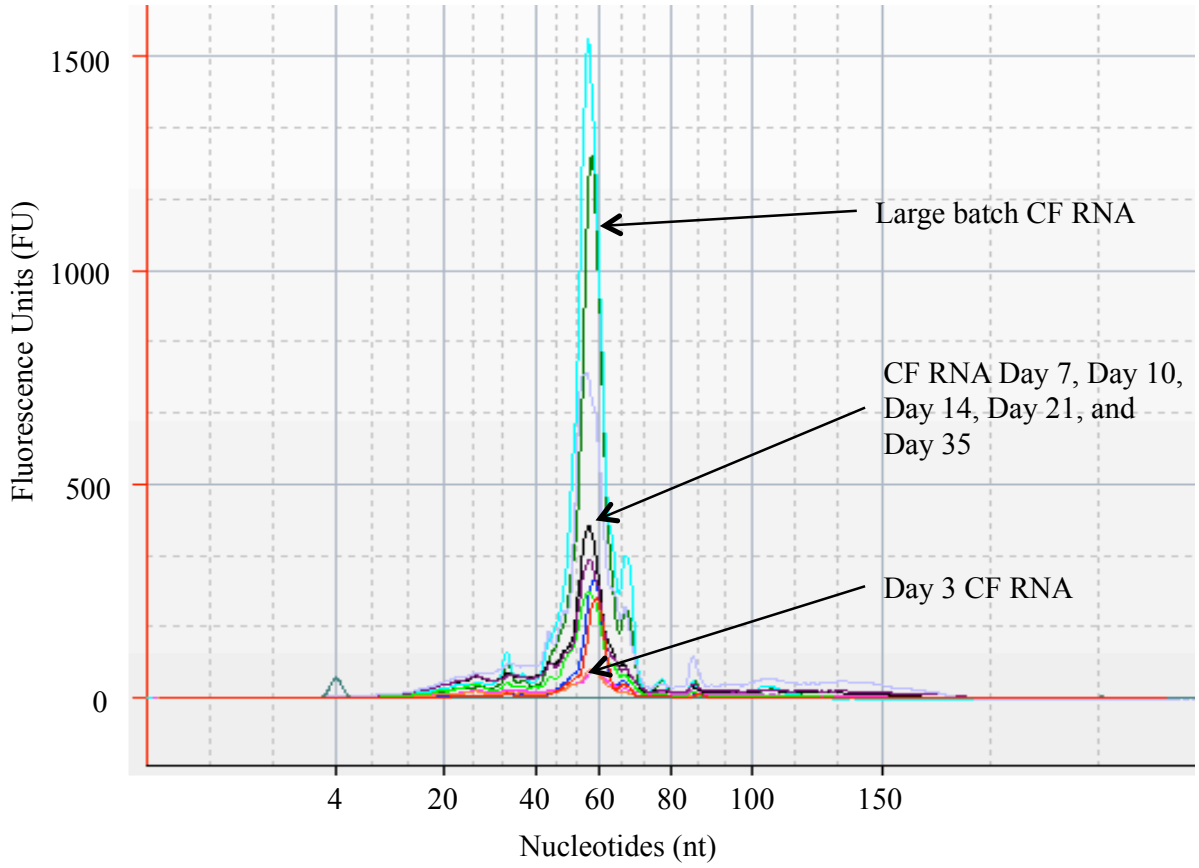


Figure 2.5 Bioanalyzer analysis of purified RNA. Bioanalyzer chip technology showed similarity between multiple batches of purified secreted RNA, including RNA purified at different times during the growth of *M. tuberculosis*.

2.3.4 Kinetics of release

The identification of extracellular RNA in the CF as early as day three lead to the investigation of the kinetics of release. To assess the kinetics of release, the rate of RNA release into the CF was compared to the rate of protein release in the CF. Protein release was measured using the BCA, and RNA release was measured using the Bioanalyzer small RNA chip. When plotted together (Figure 2.6), there is parallel trend between the rate of protein and RNA release into the CF.

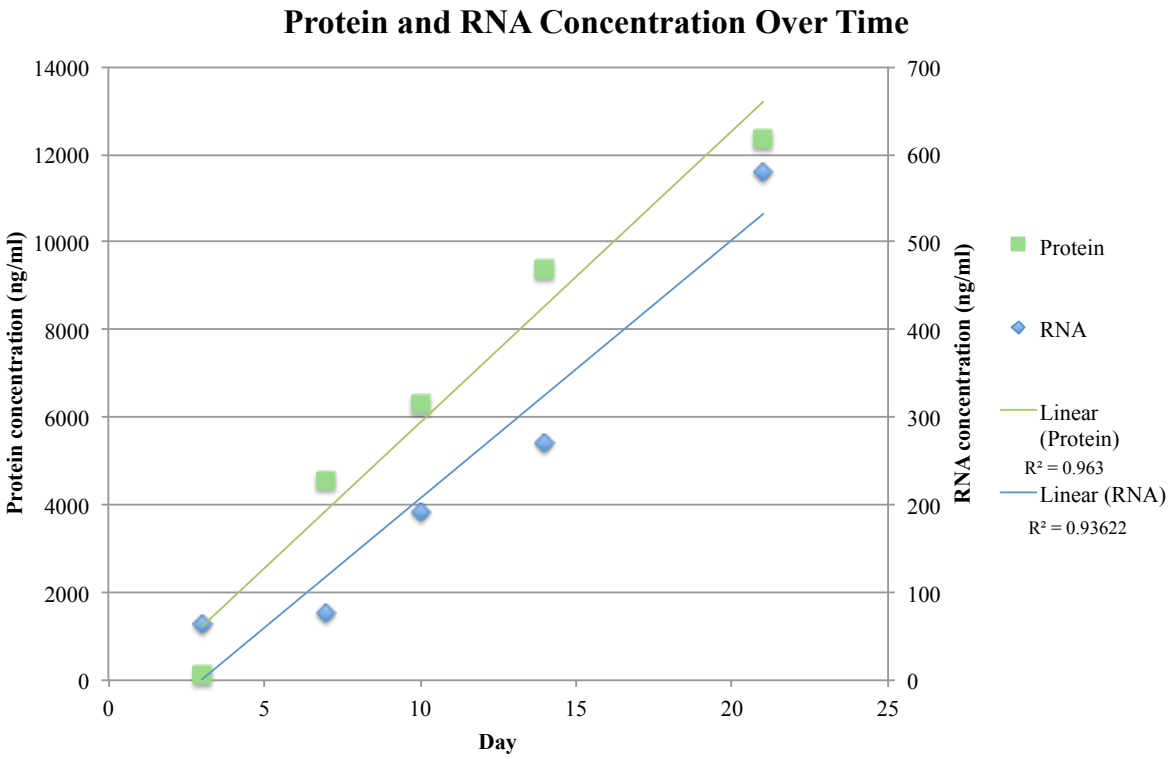


Figure 2.6 Kinetics of release. The release of RNA was measured over time using the Bioanalyzer, and compared to the release of protein in the culture filtrate, measured using the BCA. The rate of accumulation of both RNA and protein increases over a 20-day period.

While the protein has a fairly linear release rate, the RNA release is not linear. RNA from day 7-21 is more linear, so it is possible that day three could be a technical issue. The concentration of RNA in the CF of a day three culture is much lower than the other time points, which can be seen in Figure 2.5. It is possible that the concentration of day three was so low that it was outside the optimal range for the Bioanalyzer and therefore read inaccurately. By running known concentrations of tRNA on the Bioanalyzer, the instrument was found to be less accurate at lower concentrations, which could explain why the time point for day three is a trend outlier.

2.3.5 Intracellular Small RNA Extraction

The whole-cell RNA resulting from the FastRNA Pro Blue Kit was run on the Bioanalyzer showing total RNA with rRNA (Figure 2.7 A). The first isolation with the PureLink miRNA kit resulted in a decreased amount of rRNA. Because the rRNA had not been completely removed, a second isolation was preformed using a modified form of the miRNA kit protocol; after which the rRNA had been removed and only the sRNA remained (Figure 2.7 B). The workflow for the intracellular small RNA purification is depicted in Figure 2.7 C. The typical intracellular small RNA yield after two isolations with the miRNA kit is approximately 25 μg.

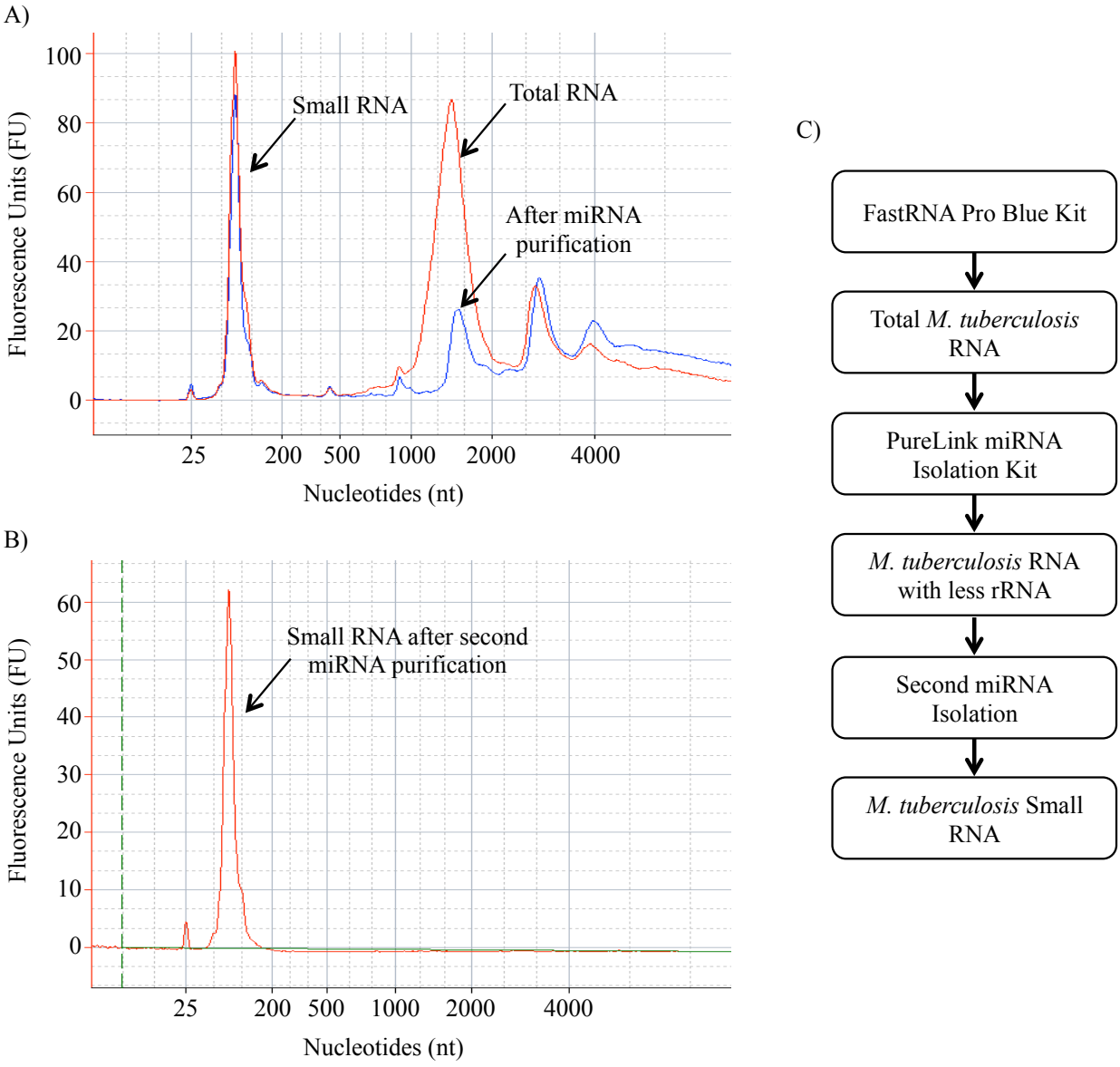


Figure 2.7 Intracellular small RNA extraction. A) Bioanalyzer RNA Nano chip results for intracellular *M. tuberculosis* RNA after the FastRNA Pro Blue kit and first PureLink miRNA Isolation kit. B) Bioanalyzer Nano chip results for intracellular *M. tuberculosis* RNA after the second miRNA kit isolation, resulting in purified, intracellular small RNA. C) Work flow of intracellular small RNA purification method.

2.4 Discussion

The development of a reliable RNA isolation was critical to moving forward with studies of extracellular RNA. The method described here allows for reliable purification of RNA from the CF of *M. tuberculosis*; free of DNA, protein, LAM, and LPS. Using the Bioanalyzer for analysis, the RNA was shown to be small in size, ranging from 30 to 70 bases in length, as was previously found [1].

The C18 solid-phase extraction binds nonpolar and moderately polar compounds, allowing the negatively charged RNA to pass through. The C18 is able to remove both LAM and LPS, two biologically active components that could potentially interfere with downstream biological assays. The ConA method was also successful in removing LAM, although the C18 method was ultimately chosen based on ease of use.

A method for removing sRNA from whole *M. tuberculosis* was developed for downstream analysis and comparison to the extracellular sRNA, which will be discussed in Chapter IV. The FastRNA Pro Blue Kit was chosen as it was used previously in whole-cell RNA isolation from *M. tuberculosis* [10, 11]. The PureLink miRNA isolation kit was used in an effort to remove the rRNA. It is important for rRNA to be removed for downstream studies, especially Next Generation Sequencing (discussed in Chapter III). There is a concern that the rRNA would overpower some of the small RNA fragments, and would need to be removed computationally. The PureLink miRNA kit is designed to isolate high-quality sRNA molecules, based on the selective binding of sRNA molecules to a silica-based membrane in the presence of 70% ethanol.

CF contains a diverse population secreted protein antigens that accumulate over time, starting as early as day two [12]. To evaluate the release of RNA in the CF, the kinetics of

release was compared to that of proteins, similar to what was done by Obregón-Henao et al. [1]. The RNA and protein have parallel release rates, which could indicate active release rather than passive release of RNA. Because the concentration of RNA in the CF at early time points is so low the kinetics of release of the RNA could be inconclusive for the early time points. It will become important to future studies to develop a method to measure specific products over time. The replication rate of *M. tuberculosis* is typically about 24 hours [13], which is important to note in looking at the kinetics of growth and the release of RNA into the CF. It has also been documented that the highest content of RNA and protein is present in two-week cultures, indicating logarithmic phase growth [14]. Because the concentration of RNA in the CF at early time points is so low the kinetics of release of the RNA could be inconclusive for the early time points. It will become important to future studies to develop a method to measure specific products over time.

H37Rv is a virulent strain of *M. tuberculosis*, although it is now considered a laboratory strain. CF RNA was also detected in *M. tuberculosis* clinical isolates HN878 and CDC1551, as well as the avirulent *M. smegmatis* (data not shown). These findings indicate that RNA release into CF is common among *Mycobacterium* spp., although it may not be a characteristic of virulence.

The development of the method described in this chapter was critical for experiments to further elucidate the composition of the small-extracellular RNA from *M. tuberculosis*, described in Chapter III. This procedure can also be used for down-stream studies to determine the biological implications of this extracellular RNA.

References

1. Obregon-Henao, A., et al., *Stable extracellular RNA fragments of Mycobacterium tuberculosis induce early apoptosis in human monocytes via a caspase-8 dependent mechanism.* PLoS One, 2012. **7**(1): p. e29970.
2. Rojas, M., et al., *Differential induction of apoptosis by virulent Mycobacterium tuberculosis in resistant and susceptible murine macrophages: role of nitric oxide and mycobacterial products.* J Immunol, 1997. **159**(3): p. 1352-1361.
3. Maiti, D., A. Bhattacharyya, and J. Basu, *Lipoarabinomannan from Mycobacterium tuberculosis promotes macrophage survival by phosphorylating Bad through a phosphatidylinositol 3-kinase/Akt pathway.* Journal of Biological Chemistry, 2001. **276**(1): p. 329-333.
4. Nigou, J., et al., *Mycobacterial lipoarabinomannans: modulators of dendritic cell function and the apoptotic response.* Microbes and infection, 2002. **4**(9): p. 945-953.
5. Takayama, K., et al., *Site of inhibitory action of isoniazid in synthesis of mycolic acids in Mycobacterium-tuberculosis.* Journal of Lipid Research, 1975. **16**(4): p. 308-317.
6. Morrissey, J.H., *Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity.* Analytical biochemistry, 1981. **117**(2): p. 307-310.
7. Kaur, D., et al., *Characterization of the epitope of anti-lipoarabinomannan antibodies as the terminal hexaarabinofuranosyl motif of mycobacterial arabinans.* Microbiology-Sgm, 2002. **148**: p. 3049-3057.
8. Dubray, G. and G. Bezard, *A highly sensitive periodic acid-silver stain for 1, 2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels.* Analytical biochemistry, 1982. **119**(2): p. 325-329.
9. Kaur, D., et al., *Biosynthesis of mycobacterial lipoarabinomannan: role of a branching mannosyltransferase.* Proceedings of the National Academy of Sciences, 2006. **103**(37): p. 13664-13669.
10. Arnvig, K.B. and D.B. Young, *Identification of small RNAs in Mycobacterium tuberculosis.* Molecular microbiology, 2009. **73**(3): p. 397-408.
11. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.
12. Andersen, P., et al., *Proteins released from Mycobacterium tuberculosis during growth.* Infection and Immunity, 1991. **59**(6): p. 1905-1910.
13. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-544.
14. Youmans, A.S. and G.P. Youmans, *Ribonucleic acid deoxyribonucleic acid and protein content of cells of different ages of Mycobacterium tuberuclosis and relationship to immunogenicity.* Journal of Bacteriology, 1968. **95**(2): p. 272-&.

Chapter III

Analysis of Extracellular RNA by Next Generation Sequencing

3.1 Introduction

In Chapter I, the ability of mycobacterial extracellular RNA to induce an apoptotic response in human monocytes was discussed. The work described here is a continuation of previous studies, in an effort to further define the extracellular RNA population of *M. tuberculosis*. Information gained from this research will allow for the further definition of the biological importance of the extracellular RNA, as well as its contribution in the pathogenesis of *M. tuberculosis*.

In order to better understand the biologically active RNA found in the CF, it is important to fully define the RNA population. While there are several methods that could be utilized for the further definition of the extracellular RNA, Next Generation Sequencing (NGS) was the most appropriate for the needs of this study. Microarrays can be used to detect large numbers of RNA species, and are often used in transcriptional studies. However, microarrays were not ideal in this situation, as the probes on the chip must be defined, and this study seeks to find any type of RNA that may be present. Similarly, for a PCR based method of detection, targeted DNA sequences are needed as probes. Microarray or PCR could be used to investigate previously described *M. tuberculosis* extracellular RNA, although that does not meet the needs of this study. For these reasons, NGS technology was used to further elucidate the extracellular RNA population.

NGS technologies provide large volumes of data with more accurate sequence information than conventional Sanger sequencing approaches [1]. The nature of the NGS

technology minimizes false positives, giving confidence in the sequences identified. It is possible for bias to be introduced, with library preparation and PCR amplification increasing specific RNA species to unrepresentative levels, which could influence the experiment in a detrimental fashion. The possibility of bias was investigated using a quantitation bias study of DNA oligonucleotides at specific concentrations. What is generally accepted with confidence is that nothing will be sequenced that was not initially present, meaning the results are somewhat qualitative although not necessarily quantitative.

Previous studies and Bioanalyzer analyses, discussed in Chapter II, indicated the RNA of interest was small in size, between 30-70 bases in length, and made up primarily of tRNA and rRNA fragments [2]. For most NGS involving sRNA, total RNA is used and subsequently size selected. Because the extracellular RNA from *M. tuberculosis* was already small in size, and the entire population was to be interrogated, no size selection was necessary. To determine the optimal concentration of CF RNA for NGS, a pilot study was done with a range of CF RNA concentrations. The goal of these experiments was to determine the composition of the CF RNA samples and the relative abundance of the various types of RNA present.

NGS of the CF RNA confirmed the presence of tRNA and rRNA fragments found previously [2], also revealing a third group of small RNAs (sRNA). As discussed in Chapter I, there have been recent research efforts to investigate the sRNAs of *M. tuberculosis* [3-8]. The first experimental evidence of sRNAs in *M. tuberculosis* was published in 2009 [3], and has since been the focus of several research groups. The first evidence was quickly followed by a NGS study of sRNA from *M. tuberculosis* total RNA, reveling a large number of previously unknown sRNAs [4]. Since these initial studies, other groups have identified additional sRNA candidates either through experimental evidence or bioinformatic prediction [5, 7, 8]. Despite

the identification of many previously unknown sRNAs, little is known about their function, and research is ongoing. The wealth of recent information on *M. tuberculosis* sRNAs was beneficial to the study discussed here, as the products identified by NGS could be compared to the literature.

3.2 Materials and Methods

3.2.1 Library preparation for SOLiD sequencing: CF RNA

Purified CF RNA (Chapter II) was used for NGS on the Applied Biosciences Support Oligonucleotide Ligation Detection (SOLiD) Sequencer (Life Technologies, Grand Island, NY). A range of concentrations was used from large batch and time course CF RNA samples. For the large batch CF RNA (harvested at day 14), 100 ng, 50 ng, 10 ng, 5 ng, 1 ng, 0.5 ng, and 0.1 ng were used. For samples from the time course experiment, days 3, 7, 10, and 14, were tested with 5 ng and 1 ng concentrations for each time point. The only exception was day 14, which was only tested at 1 ng because the data for the 5 ng concentration were encapsulated within the large batch samples. Sequencing libraries were prepared using the Applied Biosciences SOLiD Total RNA-Seq Kit and protocol (Life Technologies, Grand Island, NY). Briefly, CF RNA was hybridized and ligated to SOLiD™ adaptors, followed by reverse transcription (RT) to synthesize complementary DNA (cDNA) utilizing adaptor-based priming. Resulting cDNA was purified using gel extraction and amplified using PCR. Amplified cDNA was tested for yield and size distribution using the NanoDrop spectrophotometer and the Agilent 2100 Bioanalyzer with the DNA 1000 kit (Agilent Technologies, Germany). Once the cDNA had been amplified and passed gel and bioanalyzer QC assessment, fragments were ligated to template beads and deposited onto a glass slide for sequencing using the SOLiD EZ Bead kit (Life Technologies,

Grand Island, NY) (Figure 3.1). The Next Generation Sequencing Core at CSU performed library preparation and sequencing.

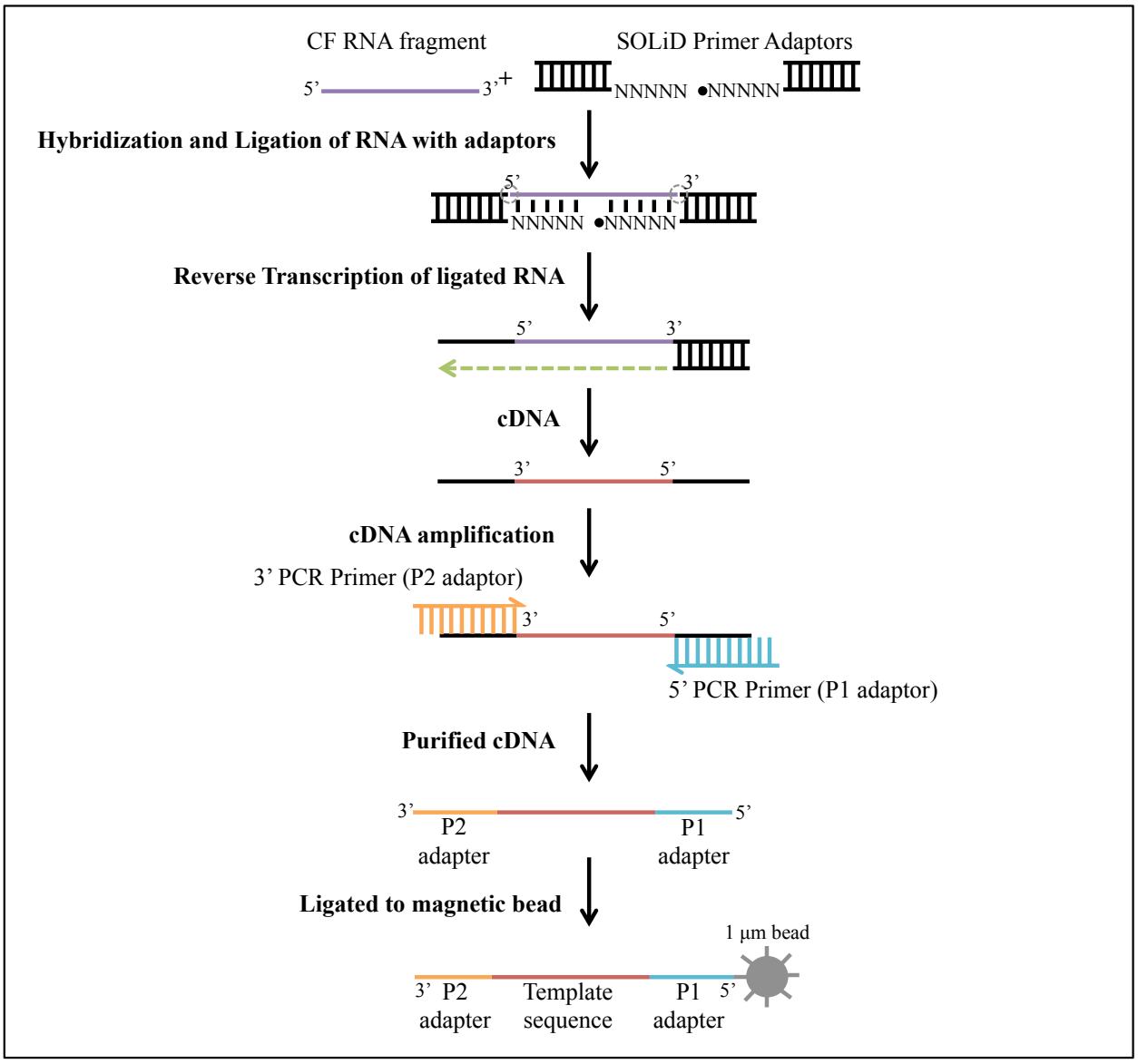


Figure 3.1 SOLiD sRNA library preparation. RNA is hybridized and ligated to adapters followed by RT to produce cDNA. cDNA is amplified using SOLiD specific primers to the adapters added in the beginning of library preparation. Amplified cDNA is purified and ligated to magnetic beads before being deposited onto glass slides for sequencing (Figure adapted from the Applied Biosystems SOLiD total RNA-Seq kit protocol).

3.2.2 Library preparation for SOLiD sequencing: Quantitation bias study

The 45 *M. tuberculosis* tRNAs were synthesized as DNA oligonucleotides (stDNA oligos). A mixture of the stDNA oligos was made to a final concentration of 5 µg, with groups of stDNA oligos at different concentrations. At random, 33 of the stDNA oligos were chosen to be at a concentration of 0.15 µg. The remaining stDNA oligos were split into groups of three for decreasing dilutions: 0.075 µg, 0.03 µg, 0.015 µg, and 0.003 µg (Table 3.1). Sequencing libraries were prepared using the Applied Biosciences SOLiD Fragment Library DNA Kit and Oligo Primer Set, according to manufacturers instructions (Life Technologies, Grand Island, NY) (Figure 3.2). The Next Generation Sequencing Core at CSU performed library preparation and sequencing.

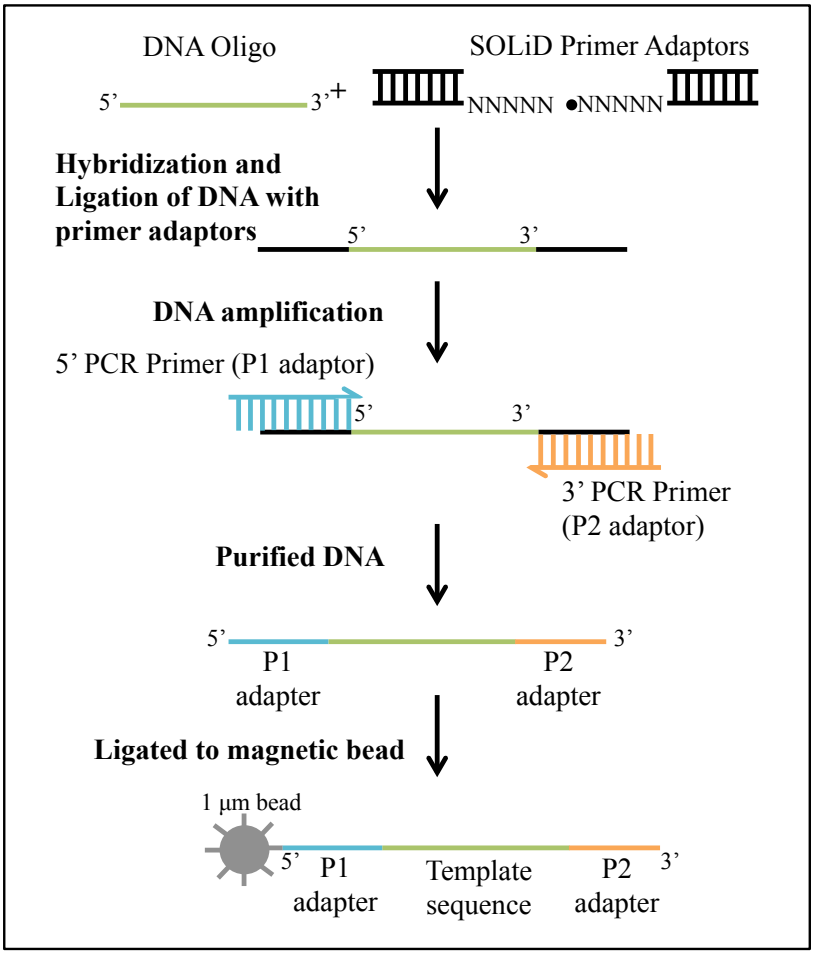


Figure 3.2 SOLiD DNA oligo library preparation. DNA oligo is hybridized and ligated to adapters followed by amplification using SOLiD specific primers to the adapters added in the beginning of library preparation. Amplified DNA is purified and ligated to magnetic beads before being deposited onto glass slides for sequencing (Figure adapted from the Applied Biosystems SOLiD Fragment Library DNA kit protocol).

Table 3.1 stDNA oligonucleotides and concentration. stDNA oligos grouped by concentration added to equal a 5 µg mixture and sequenced.

stDNA Oligo	Concentration (µg)
alaT	0.150
alaV	
argV	
argW	
asnT	
aspT	
cysU	
glnU	
gluU	
glyT	
glyU	
hisT	
ileT	
leuT	
leuU	
leuW	
lysU	
metT	
metV	
proT	
proU	
proY	
serT	
serU	
serV	
serX	
thrT	
thrU	
thrV	
trpT	
tyrT	
valT	
valV	
alaU	0.075
glnT	
pheT	
argT	0.030
gluT	
leuV	
argU	0.015
glyV	
leuX	
lysT	0.003
metU	
valU	

3.2.3 Data Analysis

Colorspace data files (.csfasta) from SOLiD sequencing were uploaded to the NextGENe data analysis server (Soft Genetics, State College, PA) for analysis. Once loaded into the NextGENe server, the 35 base sequence reads were aligned to the *M. tuberculosis* H37Rv GenBank reference genome (.gbk file) (NCBI reference sequence: NC_0009623) using the following parameters: the matching requirement parameters were base number >= 12 and base percentage >= 90%, with rigorous alignment. The base number indicates the minimum number of bases in each read that must match the reference sequence, and 12 is a standard setting. The base percentage for these studies was set high so that a minimum of 90% of each read much match the reference sequence. Rigorous analysis parameters were necessary to have confidence in the location of each read along the reference sequence.

Following alignment, a Peak Identification Report was generated for any peaks with five or more 35 base reads aligned to a region. The report provided information on the chromosome region; length of the peak consensus sequence; depth of coverage, or reads; and the consensus sequence for each peak, or fragment, identified. The whole data set was then split into three groups: tRNA, rRNA, and sRNA, using additional data analysis methods.

First, the genomic tRNA database (GtRNAdb) was used was used to identify tRNAs within the data set by using the BLAST function for each consensus sequence obtained from the Peak Identification Report [9]. For the quantitation bias study, the GtRNAdb was used to identify the individual stDNA oligos [9].

Remaining sequences were visualized using the Artemis genome browser and annotation tool [10]. The H37Rv GenBank file was loaded onto Artemis as the reference genome, and each fragment identified in the peak identification report was entered as a new feature. Each new

feature was then visualized on the genome browser, identifying the location on the genome in relation to specific genes. The information gathered from this mapping allowed for the further division of the data set into rRNA and sRNA sequences. An overview of the data analysis pipeline is shown in Figure 3.3.

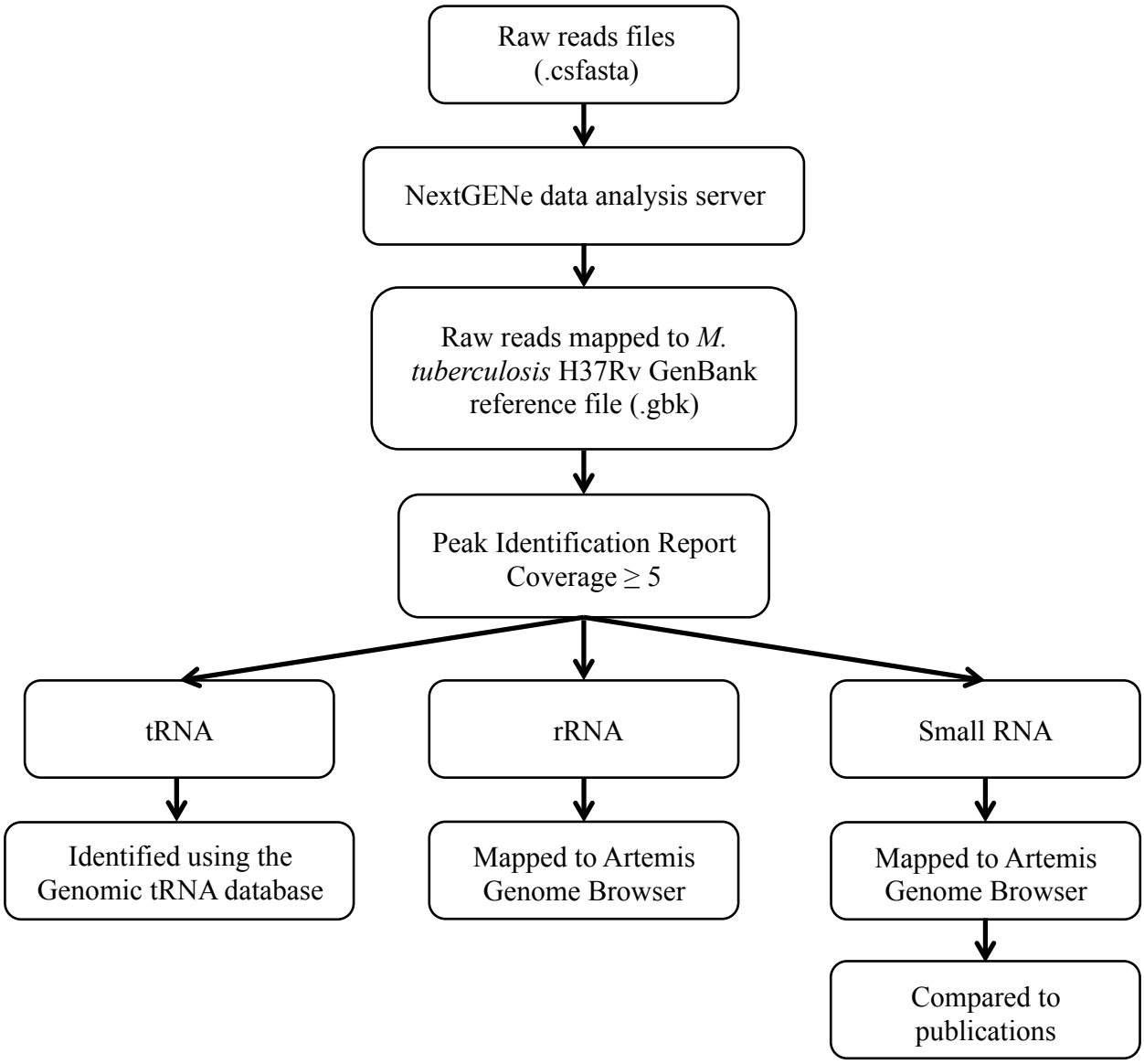


Figure 3.3 SOLiD Sequencing Data Analysis. Workflow for NGS data analysis.

3.3 Results

3.3.1 CF RNA sequencing data

Once the raw sequencing data were aligned to the H37Rv reference file, the Peak Identification Report with a cut off of five or greater reads, was generated and analyzed for each sample. The large batch samples were analyzed first, and of the *M. tuberculosis* CF RNA concentrations tested, the 100 ng and 50 ng samples were found to be the optimal concentrations. Lower concentrations tested were lacking in depth of coverage, identifying few, if any sequences with five or greater reads (Table 3.2). The 100 ng sample was preferentially used for analysis because of the depth of coverage obtained.

Table 3.2 Number of unique fragments and average reads per fragment for large batch samples. Displays the number of unique fragments as well as the average reads per fragment for each large batch sample sequenced. The 100 ng and 50 ng samples yielded the highest number of unique fragments of all concentrations tested. The 5 ng, 1 ng, and 0.5 ng samples had no unique fragments identified using the Peak Identification report.

Large Batch Sample Concentration (ng)	Number of Unique Fragments	Average Reads per Fragment
100	297	69
50	158	34
10	1	35
5	0	0
1	0	0
0.5	0	0
0.1	2	8

The time course samples were submitted for sequencing at the same time as the large batch samples when it was hypothesized that 1-5 ng of sRNA would be sufficient for sequencing. Similar to what was observed in the large batch samples, the 5 ng and 1 ng concentrations had few unique fragments identified in each sample (Table 3.3). Many of the time course samples

had more unique fragments than what was detected with the large batch samples at the same concentrations. It is likely that at such a low concentration the results would be variable and difficult to compare samples. As such, very little data were obtained and they could not be compared with any confidence.

Table 3.3 Number of unique fragments and average reads per fragment for time course samples. Each time course sample was sequenced at two different concentrations (5 ng and 1 ng) with the exception of Day 14, as it was sequenced at multiple concentrations in the large batch experiment. Time course Day 7: 1 ng and Day 10: 1ng, had no unique fragments detected using the Peak Identification report.

Time Course Day	Sample Concentration (ng)	Number of Unique Fragments	Average Reads per Fragment
3	5	19	8
3	1	1	8
7	5	6	12
7	1	0	0
10	5	15	13
10	1	0	0
14	1	7	8

As shown in Figure 3.3, after Peak Identification Reports were generated, the whole data set for each concentration was split into three groups: tRNA, rRNA, and sRNA. Table 3.4 shows the total reads, matched reads, number of reads per fragment, and average reads per fragment for the 100 ng (Table 3.4 A) and 50 ng (Table 3.4 B) samples.

Table 3.4 Number of unique fragments and reads detected for each group. A) Total reads, matched reads, number of unique fragments, and average reads per fragment from 100 ng sample. B) Total reads, matched reads number of unique fragments, and average reads per fragment from 50 ng sample.

A)

100 ng	Whole Data Set	tRNA	rRNA	sRNA
Matched Reads	20615	7998	7178	5439
Number of Unique Fragments	297	25	33	239
Average Reads per Fragment	69	320	205	23

B)

50 ng	Whole Data Set	tRNA	rRNA	sRNA
Matched Reads	21625	9223	9941	2461
Number of Unique Fragments	158	34	27	97
Average Reads per Fragment	659	342	292	25

In sequencing analysis, matched reads, or total reads, provides an idea of relative abundance of each group. The number of unique fragments gives an indication of the complexity or diversity within each group of RNA. The average reads per fragment is another indicator of abundance, illustrating the relative abundance of each unique fragment in a particular group. Although there are differences between the 50 ng and 100 ng samples, the numbers show the same trend. tRNA fragments have the highest average reads per fragment, indicating that the tRNAs are relatively more abundant. The sRNAs have the highest number of unique fragments and are therefore more diverse, although the average reads per fragment is much lower than those of the rRNA and tRNAs. Each group is discussed in more detail in the following sections.

3.3.1.1 tRNAs detected in CF RNA

The tRNAs detected using this method were the most abundant group of CF RNAs identified. Full lists of tRNA fragments from the 50 ng and 100 ng sample, with information on length, reads, genomic location, and consensus sequence, can be found in Appendix A and B respectively.

Table 3.5 shows all tRNA fragments identified in the 50 ng and 100 ng samples. Of the 45 tRNAs encoded in the *M. tuberculosis* H37Rv genome NGS identified 25 tRNAs present in the 100 ng CF RNA preparation. The 50 ng sample displays the same general trend as the 100 ng sample, although there were some outliers. The previous study by Obregón-Henao et al. identified four tRNAs, and only $tRNA^{Lys}$ was absent from those found in this study [2].

Table 3.5 tRNAs identified in 100 ng and 50 ng sample. Each tRNA detected in the 100 ng and 50 ng sample with respective number of reads. Highlighted tRNAs were also detected previously; $tRNA^{Lys}$ was the only tRNA identified in the previous study that was not found through Next Generation Sequencing [2].

tRNA	100 ng Reads	50 ng Reads
SerT	2633	6309
GlyU	2160	624
AsnT	886	518
ValU	731	114
LeuU	533	918
MetU	350	143
GlnU	107	6
ArgV	92	18
HisT	80	154
TrpT	55	0
ThrT	52	23
ThrV	47	32
LeuT	44	32
ThrU	39	30
ValV	39	7
ArgW	36	23
ProT	32	9
SerV	26	5
CysU	11	8
GlyV	11	41
AlaT	10	5
AspT	8	95
IleT	6	52
ProY	5	11
ValT	5	0

The NextGENe software was used to visualize the coverage across each tRNA gene identified during the initial data analysis. The length of the fragment is a consensus sequence, made up of 35 base reads aligned to the same region. Based on the size seen with the Bioanalyzer (Chapter II) and previous studies [2], it was expected that the sequence coverage would span the majority of the gene region. Figure 3.4 shows the top four individual tRNA gene sequence alignments for the 100 ng sample. The tRNA fragments identified in the 100 ng sample generally do not span the entire gene and the majority of the fragments are located at the

3’ end of the gene (NextGENe outputs of each tRNA from the 100 ng sample can be found in Appendix C). During NextGENe data analysis, a consensus sequence is constructed from all of the individual 35 base reads along a continuous area, and the area with the greatest depth is listed as the coverage in the Peak Identification Report. For this reason, the consensus length listed is longer than the area of the greatest coverage, which can be seen in the sequence alignment for SerT in Figure 3.4.

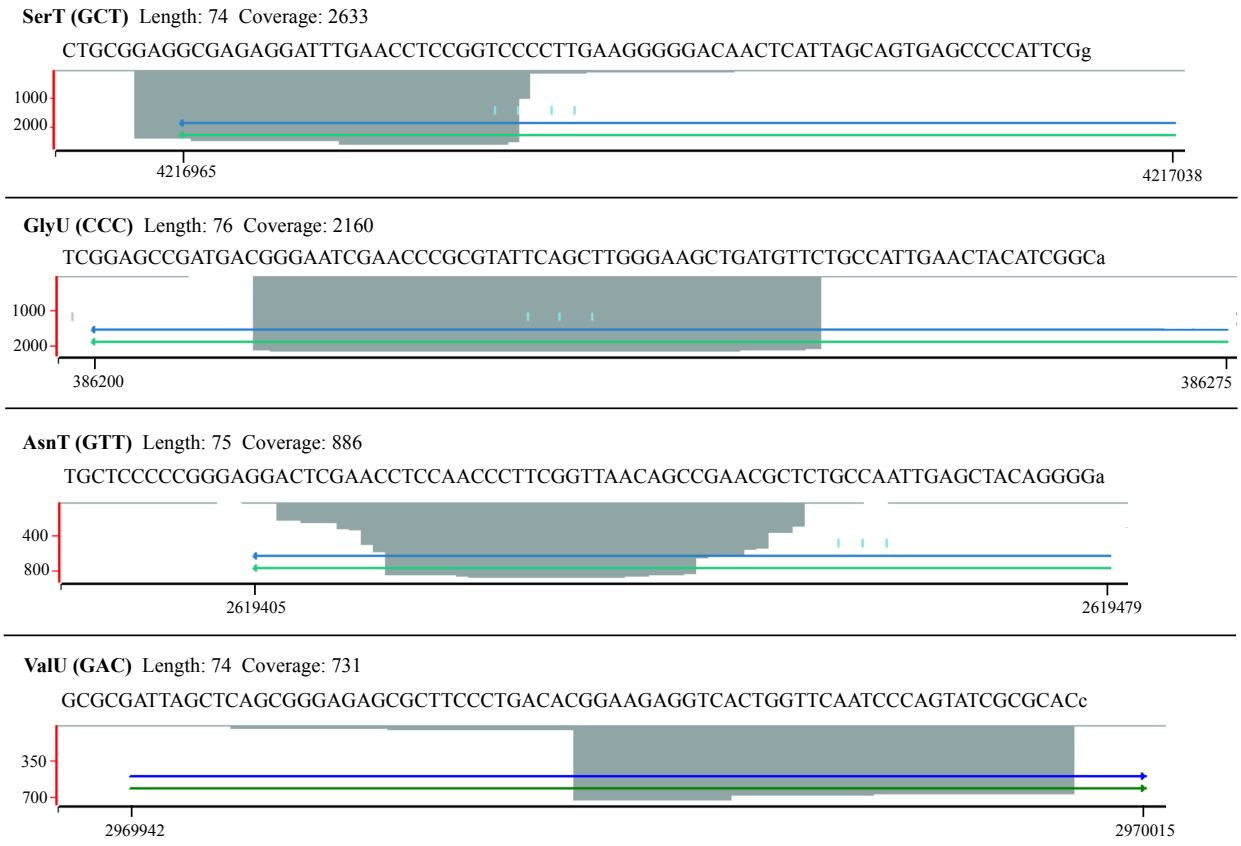


Figure 3.4 Top four individual tRNA gene sequence alignments for 100 ng sample. NGS data was aligned to the *M. tuberculosis* H37Rv GenBank reference genome. Each tRNA gene was isolated in the NextGENe viewer to show coverage. Grey regions show depth of coverage, the blue arrows are used to indicate gene locations within the reference file, and the green arrows are used to indicate mRNA. Each tRNA is labeled with name, length of the fragment, and depth of coverage. All tRNA gene sequence alignments for the 100 ng sample are displayed in Appendix C.

3.3.1.2 rRNA fragments detected in CF RNA

The rRNA group had the second-highest average reads per fragment. Full lists of rRNA fragments with their respective sequences can be found in Appendix D (50 ng) and Appendix E (100 ng).

The Artemis genome browser was used to map each of the rRNA consensus fragments to the rRNA genes (Figure 3.5). The rRNA consensus fragments for both the 100 and 50 ng samples were distributed across both the 16S and 23S rRNA genes without preference for gene location. Additionally, the 5S rRNA gene was 99% covered in both samples, although no fragments had been found for this gene in the previous study [2].

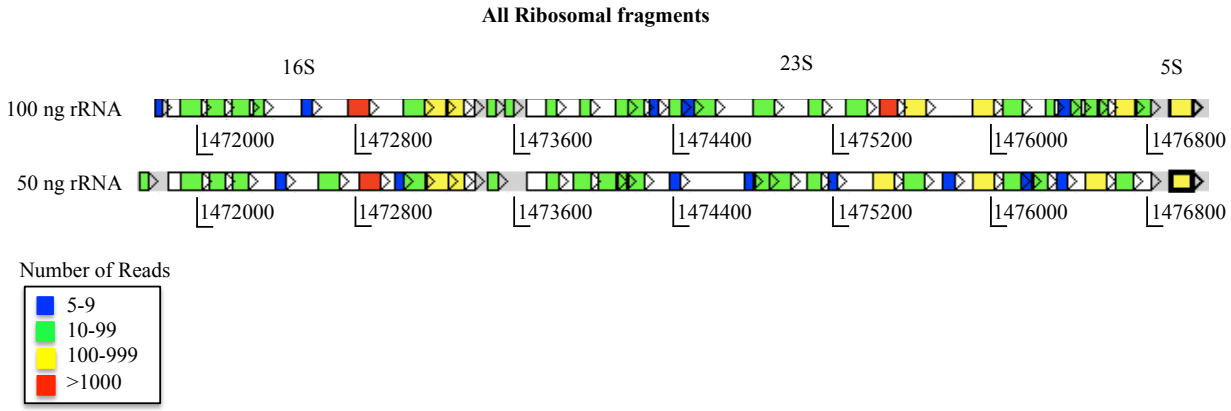
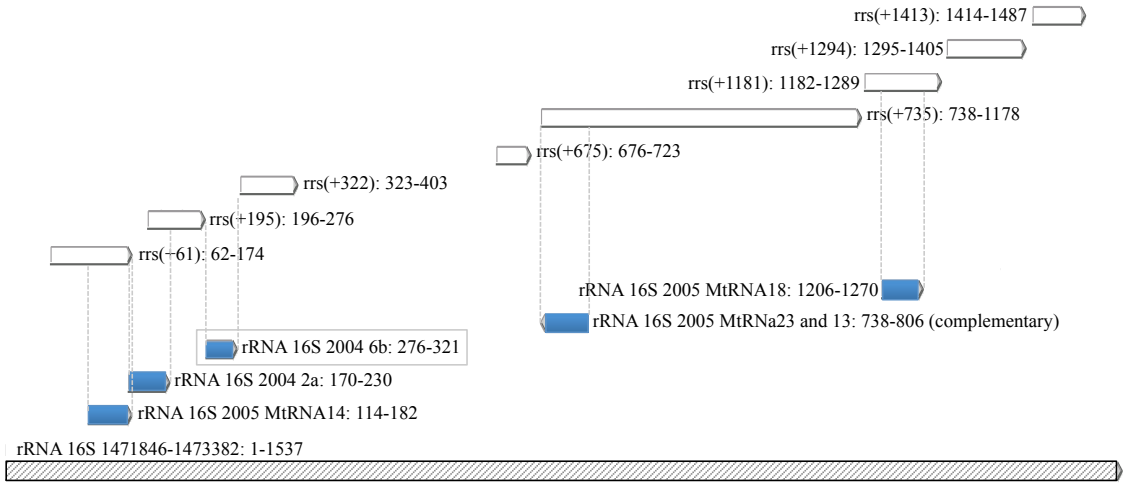


Figure 3.5 Ribosomal RNA fragments. All ribosomal consensus fragments from 100 ng and 50 ng samples, mapped using the Artemis genome browser [10]. Fragments are color-coded to show relative number of reads for each fragment.

Obregón-Henao et al. also identified rRNA fragments spanning the 16S and 23S rRNA genes and the majority of them align with the consensus sequence fragments found in this study (Figure 3.6).

A)



B)

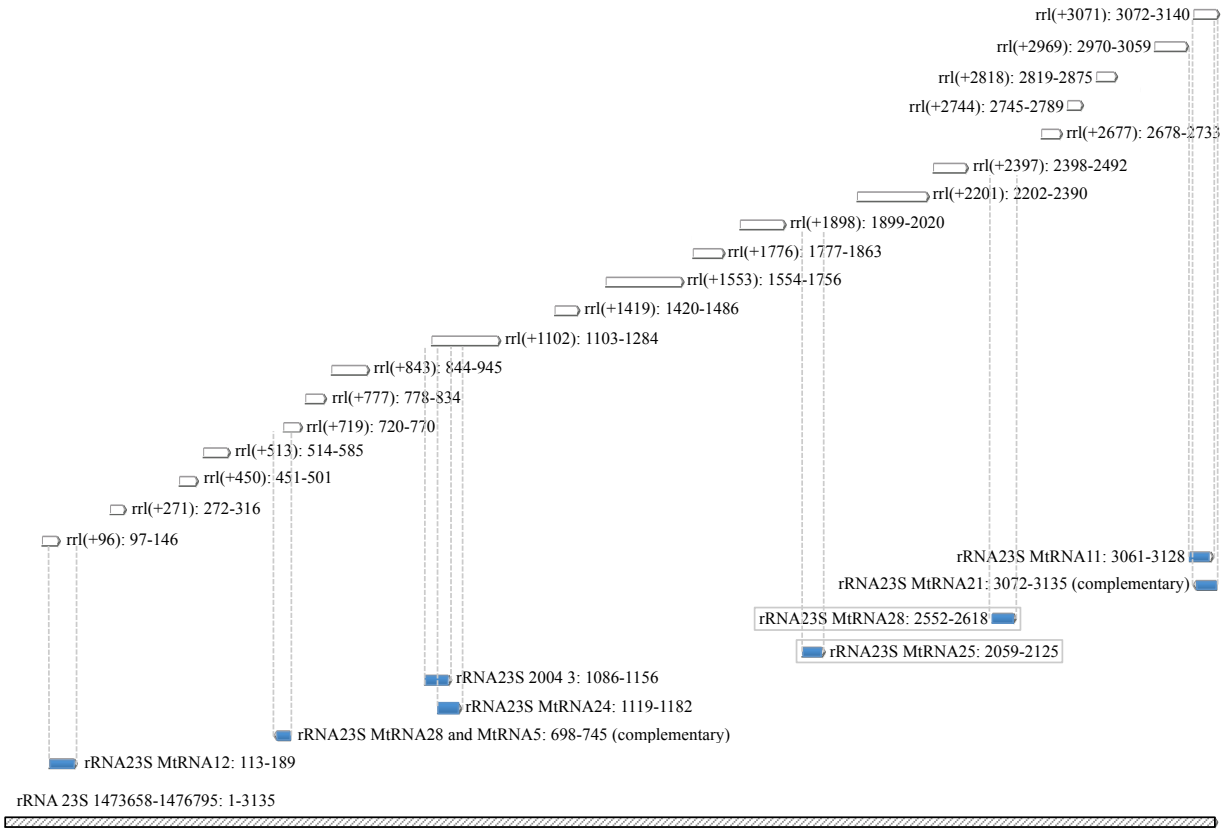


Figure 3.6 Alignments of rRNA fragments found in this study and in Obregón-Henao et al. Fragment contigs were generated using the software program Vector NTI in order to align sequences from rRNA fragments identified in this study (white arrows), with cloned RNA fragments identified in Obregón-Henao et al., 2012 (blue arrows);

connected by dotted lines [2]. Boxes indicate fragments from the pervious study that were not identified here. Reference sequences were obtained from Tuberculist [11]. A) 16S rRNA fragments B) 23S rRNA fragments.

The rRNA data analysis led to the question of why specific sections of the rRNA genes were present in the CF. One way of evaluating these consensus fragments is to use secondary structure predictions, which gives an idea of the stability of the fragments with the minimum free energy (MFE). The RNAfold web server [12] was used to predict the secondary structure of the rRNA fragments found in the 100 ng sample (Table 3.6).

Table 3.6 Secondary structure predictions of rRNA fragments found in the 100 ng sample. RNAfold web server was used to predict the secondary structure of the rRNA fragments found in the 100 ng sample. Table shows the length of the consensus sequence, coverage, transcript site, and gene distance assigned by NextGENe. For the secondary structure prediction, the table also displays the minimum free energy (MFE), free energy of the thermodynamic ensemble, as well as the ensemble diversity in each consensus fragment.

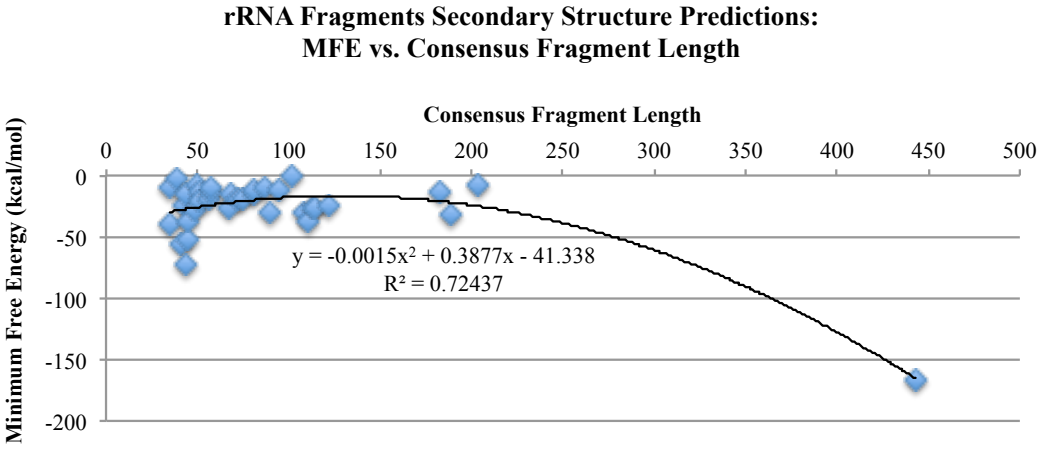
Length	Coverage	Transcript Site	Gene Distance	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
35	5	1471780..1471814	rrf(0)	-39.8	-39.13	10.06
48	6	1472521..1472568	rrl(+2397)	-28.3	-29.47	9.54
39	7	1474276..1474314	rrl(-75)	-2.2	-1.65	8.67
56	7	1476335..1476390	rrl(+2744)	-18.8	-18.89	6.56
57	7	1474435..1474491	rrl(+2818)	-13.6	-13.65	10.93
108	11	1473030..1473130	rrs(+1181)	-29.9	-26.81	16.58
203	13	1475262..1475361	rrs(+675)	-8.2	-6.4	4.35
102	18	1474501..1474601	rrs(-32)	-1.3	-1.9	5.1
51	22	1474108..1474158	rrl(+2677)	-13.2	-13.7	5.51
67	22	1475077..1475143	rrl(+2969)	-26.7	-25.36	19.07
44	23	1472278..1472321	rrl(+1553)	-72.4	-72.34	53.91
44	26	1473454..1473497	rrl(+1776)	-15.3	-15.53	22.67
81	26	1472041..1472121	rrl(+618)	-15.3	-15.71	3.28
35	29	1476276..1476310	rrl(-161)	-9.1	-9.19	10.44
57	30	1476476..1476532	rrl(+2880)	-10.6	-10.45	7.37
42	32	1476538..1476579	rrl(+1419)	-23.9	-23.02	10.57
50	36	1473754..1473803	rrl(+2618)	-7.3	-7.13	9.9
72	39	1474171..1474242	rrl(+450)	-18.5	-19.53	4.02
45	41	1476402..1476446	rrl(+2201)	-52.4	-54.08	25.5
69	41	1476729..1476797	rrl(+3071)	-15	-14	18.86
81	42	1472168..1472248	rrl(+719)	-12.2	-11.3	5.9
45	60	1473929..1473973	rrl(+1898)	-37.9	-36.99	15.88
95	74	1476055..1476149	rrl(+96)	-11.3	-9.59	13.3
51	75	1474377..1474427	rrl(+271)	-20.4	-18.81	4.94
41	77	1473543..1473583	rrl(+1102)	-55.9	-56.33	30.48
182	85	1474800..1474900	rrs(+432)	-12.6	-11.13	1.26
113	94	1471913..1472012	rrs(+1413)	-26.3	-24.69	18.17
90	134	1476627..1476716	rrl(+843)	-29.8	-29.04	28.26
111	158	1473145..1473244	rrs(+1294)	-37	-35.81	34.17
189	189	1475903..1476002	rrs(+61)	-31.7	-29.32	20.22
122	205	1475566..1475666	rrs(+322)	-23.8	-22.59	21.28
114	270	1476905..1477005	rrs(+195)	-27.1	-27.07	13.69
74	409	1473259..1473332	rrl(+513)	-18.5	-16.68	20.55
87	1303	1475434..1475520	rrl(+777)	-10	-10.81	15.38
443	3562	1472752..1472851	rrs(+735)	-166.6	-164.55	44.68

Table 3.6 displays the length of each rRNA consensus sequence identified in the 100 ng sample, along with the maximum coverage for each sequence, transcript site, and gene distance assigned by the NextGENe program. Results from the secondary structure predictions are also displayed on Table 3.6 including the minimum free energy (MFE), free energy of the thermodynamic ensemble, frequency of MFE structure in the ensemble, and the ensemble diversity for each consensus sequence entered in the program. The RNAfold program used for this analysis uses a nearest-neighbor thermodynamic model to predict the MFE. This type of modeling uses MFE to model the structure of regions that are energetically stable based on the second law of thermodynamics, which states that a system will always evolve toward thermodynamic equilibrium and that a closed system will come to thermal equilibrium with a minimum internal energy [13]. This model assumes that the energetically most stable structure is similar to the natural structure [14]. The free energy of thermodynamic ensemble is the overall free energy and sum of independent terms for different loops and base pairing interactions, or sum of all base pair free energies [14]. The MFE and free energy of the thermodynamic ensemble numbers are close to each other as they are two different methods of calculating the same thing, the stability of the RNA secondary structure. Ensemble diversity is the average distance between all structures in the thermodynamic ensemble [15]. There are several problems with prediction programs including the neglect of the RNA pseudoknots structure, and that modifications are not accounted for; both of which can impact the stability of RNA [14].

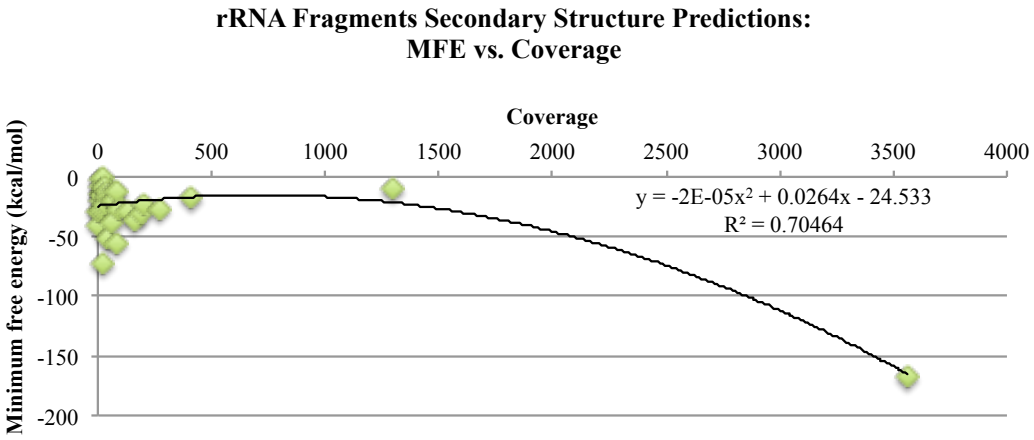
Despite the uncertainty in RNA secondary structuring model, it can be valuable to gain some idea of how stable the RNA present in the CF might be. All fragments examined were predicted to have varying degrees of secondary structure, although no correlations were apparent.

In an effort to try and better understand the results from the secondary structure predictions, the results were plotted in various ways to determine if any trends existed (Figure 3.7). The MFE for each consensus sequence was plotted against the consensus fragment length (Figure 3.7 A) and although it has been shown that the number of possible structures increases with length (L) 1.8^{L} [16], there was no trend observed with the exception of the experimental outlier that was 443 bases in length. There was also no correlation observed for MFE vs. the amount of coverage for each consensus fragment (Figure 3.7 B). To further investigate, data analysis was preformed to normalize MFE by the amount of coverage and plot against coverage (Figure 3.7 C). The consensus length was normalized by the 35 base fragment obtained from sequencing to see if this would have any effect on the correlation with the MFE, and it did not (Figure 3.7 D). Finally, the consensus length of each fragment was normalized using the amount of coverage and plotted against the MFE, which also showed no correlation (Figure 3.7 E). Based on these analyses there is no correlation between the stability of the RNA and the amount of coverage found through sequencing.

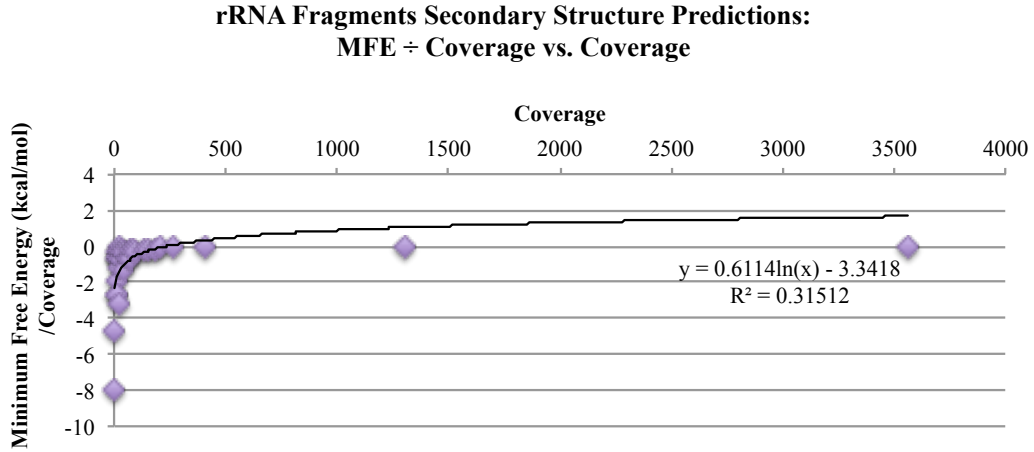
A)



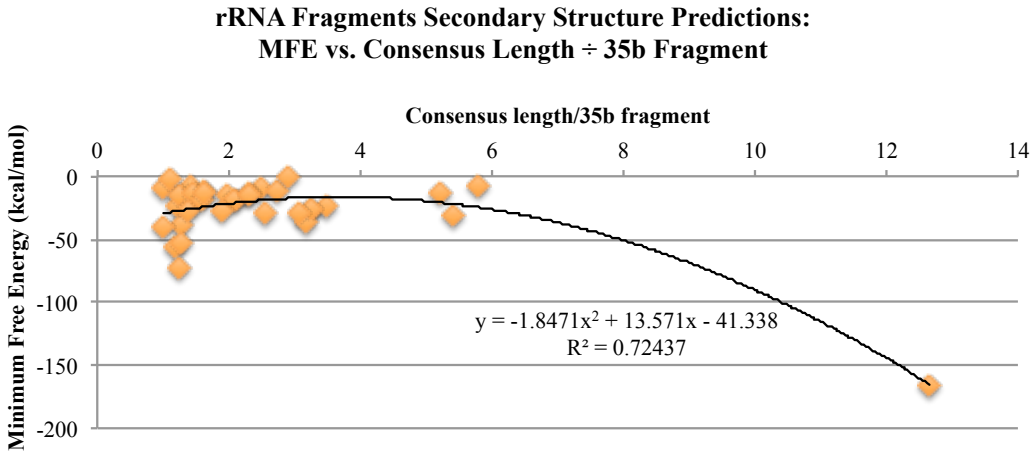
B)


rRNA Fragments Secondary Structure Predictions:
MFE vs. Coverage
Coverage
0
500
1000
1500
2000
2500
3000
3500
4000
Minimum free energy (kcal/mol)
0
-50
-100
-150
-200
y = -2E-05x² + 0.0264x - 24.533
R² = 0.70464

C)


rRNA Fragments Secondary Structure Predictions:
MFE ÷ Coverage vs. Coverage
Coverage
0
500
1000
1500
2000
2500
3000
3500
4000
Minimum Free Energy (kcal/mol)
/Coverage
4
2
0
-2
-4
-6
-8
-10
y = 0.6114ln(x) - 3.3418
R² = 0.31512

D)


rRNA Fragments Secondary Structure Predictions:
MFE vs. Consensus Length ÷ 35b Fragment
Consensus length/35b fragment
0
2
4
6
8
10
12
14
Minimum Free Energy (kcal/mol)
0
-50
-100
-150
-200
y = -1.8471x² + 13.571x - 41.338
R² = 0.72437

E)

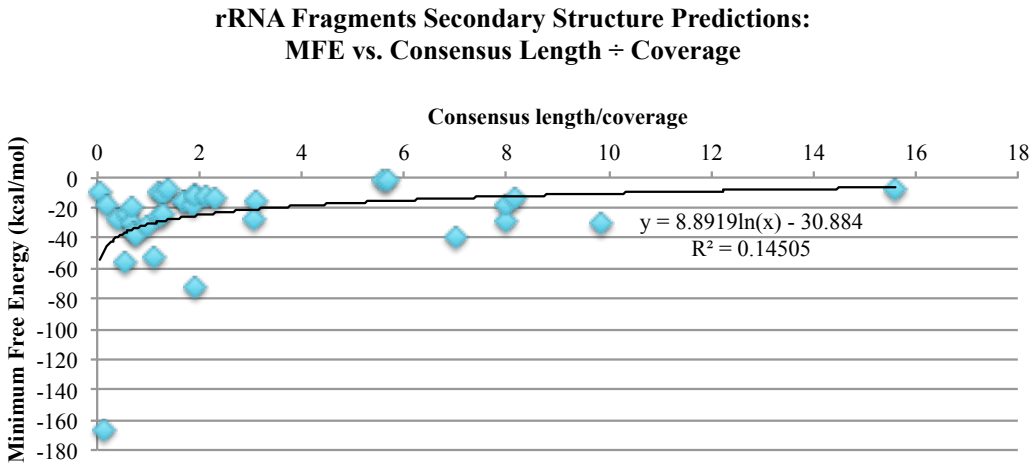


Figure 3.7 rRNA fragments secondary structure prediction analysis. A) MFE vs. the consensus fragment length. B) MFE vs. coverage for each consensus fragment. C) MFE normalized by dividing by coverage for each consensus fragment, plotted vs. coverage. D) MFE vs. consensus fragment length divided by the 35 base fragment size. E) MFE vs. consensus fragment length divided by the coverage for each consensus fragment.

3.3.1.3 sRNAs detected in CF RNA

The sRNA group had the largest number of unique fragments. Although there were many fragments, the average reads per fragment were much lower than those of the rRNA and tRNA groups, indicating a relatively low abundance overall. Table 3.7 shows the top 20 sRNA fragments identified in the 100 ng sample. Full lists of sRNA fragments for the 50 ng and 100 ng samples can be found in Appendix F and G, respectively.

sRNA fragments were compared with those reported in Arnvig et al., 2011 [4] and Miotto et al., 2012 [7] (Table 3.7). Full lists of sRNAs from 50 ng and 100 ng samples compared to the literature can be found in Appendix H and I, respectively. When sRNA fragments from the 100 ng sample were compared to those found in Arnvig et al. [4], 193 of the 2943 total sRNAs identified were also found in this study. When compared to the sRNAs from Miotto et al. [7], 43

of the 1948 identified were also found in this study. Overall, 42 unique fragments were found in this study.

Table 3.7 Top 20 sRNA fragments. Top 20 sRNA fragments found in the 100 ng sample with genome location of each fragment. Stars indicate the high abundance small RNA fragments that were also identified by Arnvig et al., 2011 [4] and/or Miotto et al., 2012 [7].

Loci	Length	Reads	Location
IS1557-2 ★	47	555	5' end of IS1557-2
IS1081-2	72	504	5' end overlapping onto Rv1199c
fadE35/Rv3798 ★	43	424	Intergenic between fadE35 and Rv3798
glyA	35	165	Towards the 5' end of glyA
Rv3660c/Rv3661 ★ ★	64	152	Intergenic between Rv3660c and Rv3661
Rv1733c/Rv1734c ★ ★	82	129	Intergenic between Rv1733c and Rv1734c
fas	34	128	In the middle of fas
Rv2144c/wag31	56	123	5' end of Rv2144c and intergenic
ideR ★	40	113	Towards the middle of ideR
phoT	67	92	3' end of phoT
Rv3894c	68	77	Towards the 3' end of Rv3894c
ssr ★ ★	40	71	In the middle of ssr
Rv1147/Rv1148c ★ ★	62	70	Intergenic between Rv1147 and Rv1148c
rpsR	37	64	Towards the 3' end of rpsR
esxA/Rv3876 ★ ★	36	62	Intergenic between esxA and Rv3876
cydA ★	45	55	Towards the 3' end of cydA
pykA	34	46	Towards the 5' end of pykA
Rv3407 ★	35	37	Towards the 3' end of Rv3407
Rv3640c ★ ★	57	35	Intergenic between Rv3640c and fic
Rv3661/Rv3662c ★ ★	167	32	Intergenic between Rv3661 and Rv3662c

★ Also found in Arnvig et al., 2011
★ Also found in Miotto et al., 2012

3.3.2 Quantitation bias study

Because there is a range in the number of reads found for different tRNAs and because this type of sequencing is considered quantitative, a study was designed to determine if bias was introduced into the sequencing of tRNAs. The study was designed by synthesizing all 45 *M.*

tuberculosis H37Rv tRNAs as stDNA oligos and sequencing a mixture of the oligos at different concentrations.

stDNA oligos were identified from the SOLiD data and grouped based on the concentration added to the mixture (Figure 3.5). It was expected that each stDNA oligo within a specific concentration group would have approximately the same number of reads. As shown in Figure 3.5, some sort of bias seems to have occurred in this study. Despite the bias seen, each of the 45 stDNA oligos added to the mixture were identified.

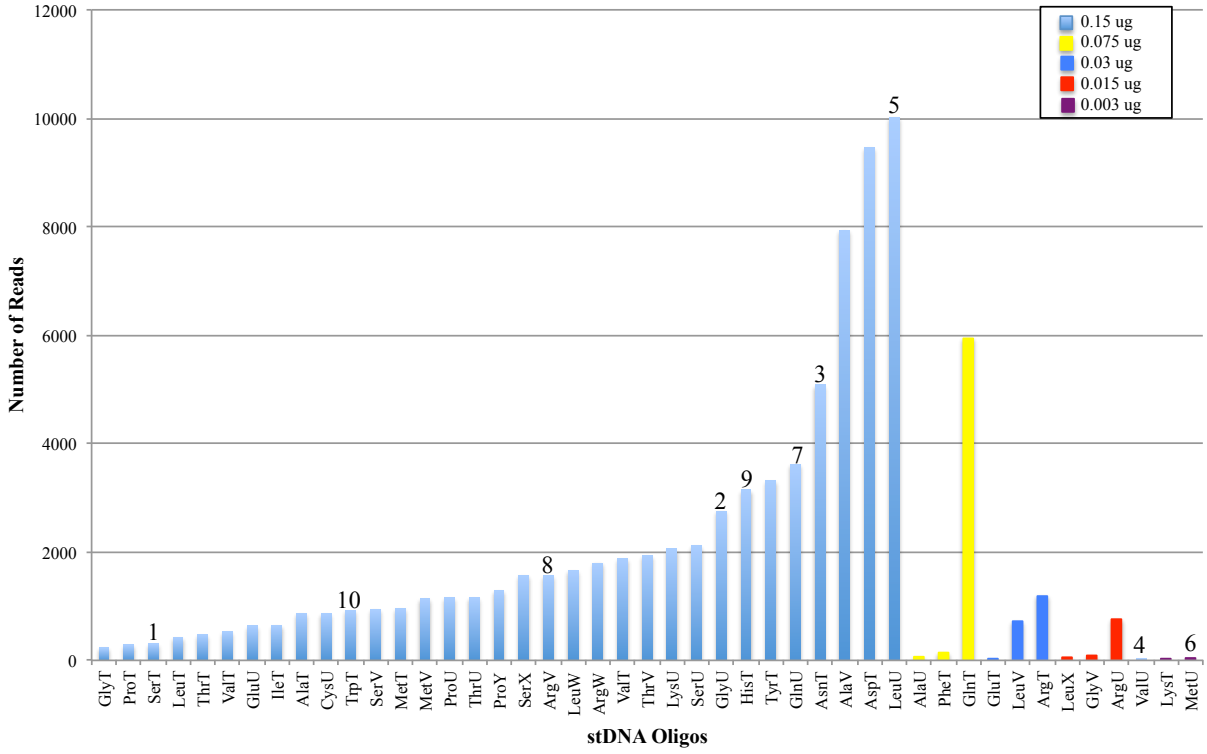


Figure 3.8 stDNA Oligonucelotides from 5 µg sample. Displays the number of reads for each of the stDNA oligos. The numbers above some of the tRNAs indicate the top ten tRNAs found in the *M. tuberculosis* CF RNA samples.

Several different types of bioinformatic analyses were performed in an attempt to identify the cause of the apparent bias. First, the bias seen with the stDNA oligos was compared to the

extracellular tRNAs identified in the previous section (Figure 3.5). The most prominent tRNA identified from the CF RNA, SerT, was one of the stDNA oligos with the lowest reads, so the bias does not follow that of secreted tRNA.

Predicted melting temperatures of each stDNA oligo were examined to see if higher melting temperatures could be a factor in the observed bias. No trend was seen, as the melting temperatures of all stDNA oligos were similar. The bias also does not correlate with free energy predictions or tRNA length. Because of the way the NextGENe software operates there was concern that the bias followed position on the genome. The tRNAs have a high degree of homology to one another, and it was possible that the program would align a fragment to the first tRNA it encountered along the genome. When the genome position of each tRNA gene was examined, the bias seen with the stDNA oligos did not correlate with genome position.

When the 5 μg stDNA oligo mixture was made it was based on concentrations given by the company who synthesized them. To help rule out a concentration issue, stDNA oligos were checked using the NanoDrop to be sure the concentrations given by IDT were correct, and that the concentration did not account for the bias (Table 3.8). Although the expected concentration and the concentration acquired using the NanoDrop are not the same, all of the stDNA oligos were off by relatively the same amount and does not account for the bias observed.

Table 3.8 Expected concentrations of stDNA oligos vs. NanoDrop readings. Each stDNA oligo was tested using the NanoDrop spectrophotometer and compared to the expected concentration given by the manufacturer, IDT.

stDNA oligo	Expected Concentration (µg/µl)	Nano Drop Concentration (µg/µl)
alaT	2.26	5.86
alaV	2.26	5.11
argV	2.26	5.66
argW	2.29	5.26
asnT	2.26	5.26
aspT	2.29	5.54
cysU	2.20	6.00
glnU	2.23	4.99
gluU	2.25	5.70
glyT	2.26	5.37
glyU	2.18	5.19
hisT	2.26	5.28
ileT	2.29	5.25
leuT	2.57	5.78
leuU	2.67	5.99
leuW	2.61	6.49
lysU	2.26	5.32
metT	2.28	5.29
metV	2.28	5.29
proT	2.30	5.95
proU	2.29	5.52
proY	2.29	5.16
serT	2.75	6.40
serU	2.68	6.19
serV	2.65	5.99
serX	2.68	6.15
thrT	2.26	5.67
thrU	2.25	5.60
thrV	2.22	6.38
trpT	2.25	5.60
tyrT	2.50	5.14
valT	2.22	5.72
valV	2.22	4.68
alaU	2.26	5.86
glnT	2.23	5.69
pheT	2.29	5.30
argT	2.22	5.70
gluT	2.28	5.50
leuV	2.28	5.48
argU	2.26	5.80
glyV	2.18	5.19
leuX	2.29	5.18
lysT	2.26	5.60
metU	2.28	5.20
valU	2.22	5.65

Discussion

The NGS pilot study provided us with a better understanding of the components of the secreted or released RNA, as well as the optimal concentrations for sequencing. The 100 ng sample was found to be the optimal concentration based on the depth of coverage obtained. Results for the 50 ng sample were slightly different from those of the 100 ng sample, with some fragments appearing in the 50 ng sample, but not the 100 ng sample. Despite some discrepancies, the two concentrations show the same trends and confirm relative abundance.

The correlation between data from Obregón-Henao et al. [2], and the tRNAs identified in the NGS run is encouraging, displaying parallels between two different methods. Only one of the tRNAs identified previously was not identified in this study, with 21 additional tRNAs identified here. This discrepancy in the number of tRNAs identified in the CF could be explained by the nature of the different methods used for detection. The NGS approach used here is more sensitive than the cloning strategy used previously, allowing for the detection of additional tRNAs [2]. Additionally, tRNAs have a high level of homology to one another, and it is possible that the $tRNA^{Lys}$ from the previous study was incorrectly identified.

The fragments that aligned to the tRNA genes did not cover the entire gene, with many tRNAs displaying greater coverage towards the 3' ends of the genes (Figure 3.4 and Appendix C). It is unknown whether the tRNAs are fragmented in this fashion in the CF, or if the fragmentation is an artifact of the sequencing. The sequencing done by Obregón-Henao et al. found the tRNAs to be truncated at the 5' and 3' ends with a length between 30-70 bases, which was confirmed by the Bioanalyzer analysis discussed in Chapter II [2]. The possible 3' bias observed with many of the tRNAs conflicts with the previous study, suggesting the possibility of a NGS sequencing issue. The structure of tRNA could be problematic for library preparation,

and it is conceivable that an issue with library preparation would explain the apparent preference for the 3' ends. Also, the RNA preparation was not mechanically fragmented during the library preparation process. Thus, if the majority of rRNAs were intact, a 3' bias would be expected based on the process of library generation (ligation and cDNA generation) and the 35 nucleotide read of the SOLiD system.

tRNAs are not typically sequenced using NGS, as many researches wish to remove tRNA and rRNA from their samples, so the protocol is not optimized for the secondary structure and stability of tRNAs. Figure 3.9 outlines points in the library preparation process where the process could be optimized or problems could occur. There have been several studies focused on library preparation for NGS, investigating where problems can arise and how this may affect bias [17-20]. Secondary structure, especially at the 3' or 5' end of the RNA can affect adapter ligation efficiency [17]. Toedling et al. found that SOLiD sequencing has a positive bias for reads with 3' end secondary structures [19]. The initial denaturation temperature of the SOLiD library preparation protocol is 65°C, which cause problems in the initial hybridization and ligation of the RNA with SOLiD adaptors, as the average predicted melting temperature for the tRNAs is 76°C. If either the 3' or 5' adaptor did not ligate properly, bias could occur at one end of the sequence during NGS, or the RNA would fail to be sequenced. A prehybidization step, or an increase in the initial denaturation temperature could eliminate secondary structure, improving adaptor ligation [17]. The number of polymerase chain reaction cycles also has the potential to influence library preparation and NGS results [17]. As shown in Figure 3.9, there are multiple stages at which the denaturation temperature could be problematic for sequencing tRNAs. These stages in the library preparation process could be optimized to ensure library preparation of the tRNAs works properly.

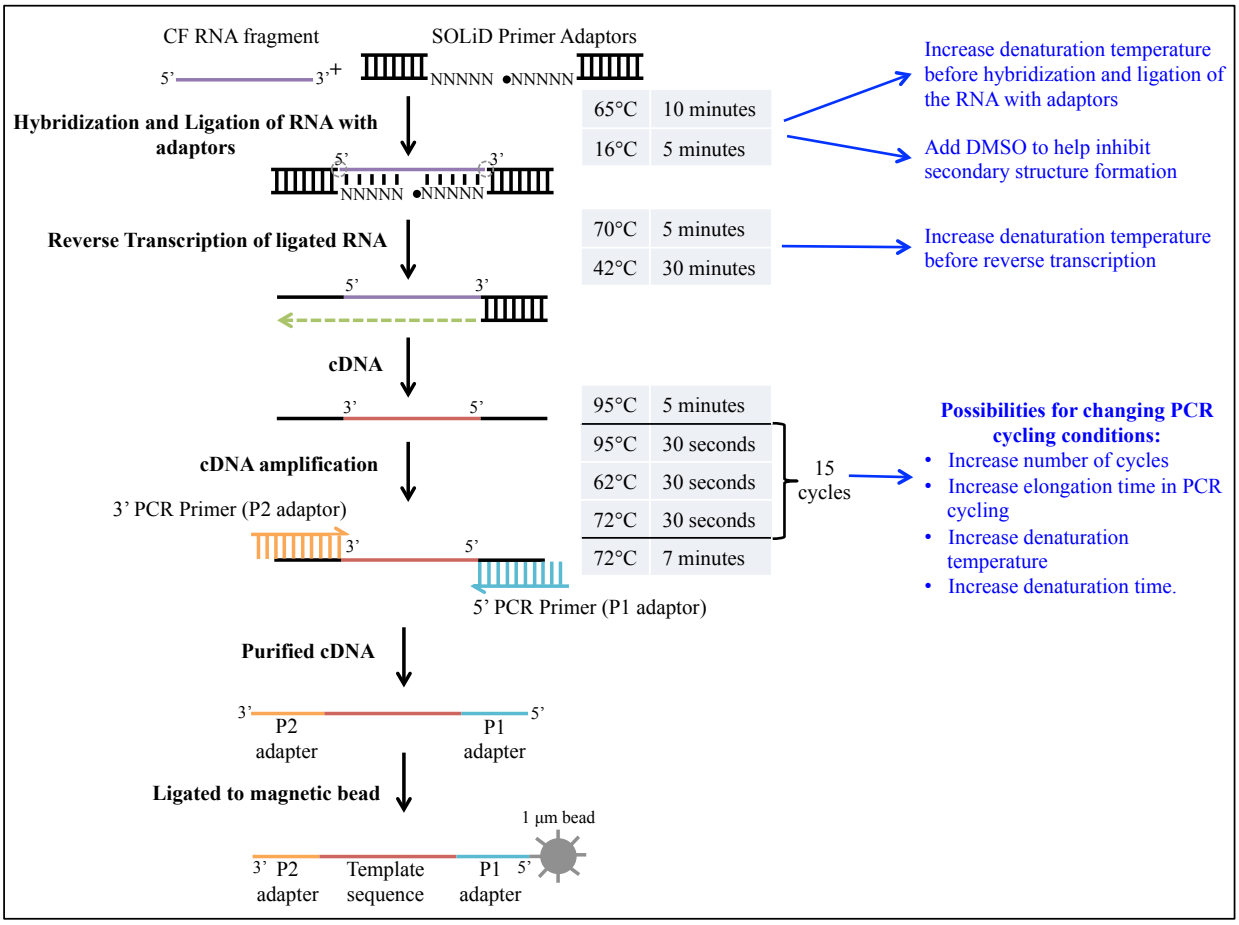


Figure 3.9 SOLiD library preparation with points where optimization and/or problems may occur. RNA library preparation with the temperatures and conditions noted. Points where optimization and/or problems may occur are indicated with suggestions for modification. Figure adapted from the Applied Biosystems SOLiD total RNA-Seq kit protocol.

The identification of rRNA fragments in the extracellular RNA also connects with the clones identified by Obregón-Henao et al. [2], with fragments distributed across the 16S and 23S rRNA genes. This study also found nearly complete coverage of the 5S rRNA gene.

It is unknown why specific fragments appear in the extracellular RNA. One hypothesis for the presence of these fragments is that they are more stable sections of the rRNA gene, capable of surviving in the extracellular environment. The RNA fold web server was used as a

method of testing this, and the fragments tested had varying levels of secondary structure, appearing to be relatively stable. Secondary structure predictions were also made for regions of rRNA genes not represented by an aligned read. The predictions for these uncovered areas had similar, if not higher levels of secondary structure and stability (Appendix J). Secondary structure analysis indicates that stability alone is not likely to be the explanation for the presence of specific rRNA fragments in the CF. In trying to find any correlation between the secondary structures of rRNA consensus sequences and the NGS sequencing results, no correlation was identified. Plotting the MFE of each predicted structure vs. the depth of coverage for that consensus sequence, yielded the inverse of what would be expected with the hypothesis that secondary structure would hinder the production of libraries. If the secondary structure of a sequence was an issue in library preparation it would be expected that more stable structures (with a more negative MFE) would have less coverage, although this was not observed.

The sRNA group contained the largest number of unique fragments, although the number of reads for each was relatively low, especially when compared to the tRNAs and rRNAs. sRNAs are of interest because they were not detected previously and have been studied in *M. tuberculosis* recently. Many of the small RNAs in the extracellular material overlapped with those previously reported in the literature [4, 7]; however, novel small RNA sequences were also identified. sRNAs were also compared with what was found by Pelly et al. from *M. tuberculosis* total RNA, discussed in Chapter I (Section 1.3.1.1) [8]. All of the clone types mentioned in Pelly et al. were detected, including the intergenic regions, although ncrMT1302 was not identified [8].

Secondary structure predictions of sRNA from the 100 ng sample using the RNAfold web server were also evaluated [12] (Appendix K). Similar to what was seen with the rRNA

consensus sequences analyzed, the sRNAs are relatively stable although there was no apparent correlation with the sequencing results. There is some correlation with the size of the sRNAs and their stability, especially as many of the sRNAs are small in size (many are the size of a single 35 base fragment) and there is not much room for secondary structure to form. The secondary structure predictions do support observations made by Obregón-Henao et al. that the CF RNA was susceptible to RNase V1, which would indicate that the CF RNA contained dsRNA hairpin structures [2].

Because of the homology of the tRNAs, an experiment was also designed to test the capabilities of our system, using stDNA oligos. This experiment indicates that some bias occurs in the NGS, although the cause of that bias is unknown. Ideally, the experiment would have been performed using synthesized RNA oligos to more accurately test the system, although this was unfortunately cost prohibitive. Another flaw in the study design was that the stDNA oligos were isolated using only standard desalting, no HPLC purification. It is possible that without the HPLC purification of the stDNA oligos, some may have degraded more than others causing differences in the library preparation and NGS. To follow up in the future, the stDNA oligos could be run on an HPLC to see if they are degraded, possibly explaining the observed bias.

The quantitative bias experiment was successful in identifying all of the stDNA oligos incorporated in the sample, even though some were at very low concentrations. The identification of all stDNA oligos gives confidence that although the NGS data may not be quantitative, they could be considered qualitative. However, based on the possibility of problems with library preparation and bias, it may be too generous to state that the technology is truly qualitative. Based on these studies it is possible to say with confidence that what is identified is present, but not that things are not missed.

The bias displayed in the quantitative bias experiment does not appear to correlate with what was found in the CF RNA, as the tRNA with the highest number of reads from the CF (SerT) was not the same as the stDNA oligo with the highest number of reads (LeuU). It is possible that the stDNA oligo bias does translate to the identification of tRNAs in the CF, for example, the SerT could be even more abundant than it appears.

Despite questions of bias created by library preparation or sequencing, the data from this study indicates some selection occurs with the extracellular RNA, especially with the tRNAs. For the stDNA oligo study, technical variability could create the observed bias. Future directions, possible methods for accounting for this variability, and the biological implications of these findings are further discussed in Chapter IV.

References

1. Metzker, M.L., *Sequencing technologies—the next generation.* Nature Reviews Genetics, 2010. **11**(1): p. 31-46.
2. Obregon-Henao, A., et al., *Stable extracellular RNA fragments of Mycobacterium tuberculosis induce early apoptosis in human monocytes via a caspase-8 dependent mechanism.* PLoS One, 2012. **7**(1): p. e29970.
3. Arnvig, K.B. and D.B. Young, *Identification of small RNAs in Mycobacterium tuberculosis.* Molecular microbiology, 2009. **73**(3): p. 397-408.
4. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.
5. Houghton, J., et al., *A Small RNA Encoded in the Rv2660c Locus of Mycobacterium tuberculosis Is Induced during Starvation and Infection.* Plos One, 2013. **8**(12).
6. Lamichhane, G., K.B. Arnvig, and K.A. McDonough, *Definition and annotation of (myco)bacterial non-coding RNA.* Tuberculosis, 2013. **93**(1): p. 26-29.
7. Miotto, P., et al., *Genome-Wide Discovery of Small RNAs in Mycobacterium tuberculosis.* Plos One, 2012. **7**(12).
8. Pelly, S., W.R. Bishai, and G. Lamichhane, *A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection.* Gene, 2012. **500**(1): p. 85-92.
9. Chan, P.P. and T.M. Lowe, *GtRNAdb: a database of transfer RNA genes detected in genomic sequence.* Nucleic acids research, 2009. **37**(suppl 1): p. D93-D97.
10. Rutherford, K., et al., *Artemis: sequence visualization and annotation.* Bioinformatics, 2000. **16**(10): p. 944-945.
11. Lew, J.M., et al., *TubercuList–10 years after.* Tuberculosis, 2011. **91**(1): p. 1-7.
12. Hofacker, I.L., *Vienna RNA secondary structure server.* Nucleic acids research, 2003. **31**(13): p. 3429-3431.
13. Dafermos, C.M., *The second law of thermodynamics and stability.* Archive for Rational Mechanics and Analysis, 1979. **70**(2): p. 167-179.
14. Eddy, S.R., *How do RNA folding algorithms work?* Nature Biotechnology, 2004. **22**(11): p. 1457-1458.
15. Freyhult, E., P.P. Gardner, and V. Moulton, *A comparison of RNA folding measures.* BMC bioinformatics, 2005. **6**(1): p. 241.
16. Mathews, D.H. and D.H. Turner, *Prediction of RNA secondary structure by free energy minimization.* Current opinion in structural biology, 2006. **16**(3): p. 270-278.
17. Tian, G., et al., *Sequencing bias: comparison of different protocols of microRNA library construction.* BMC biotechnology, 2010. **10**(1): p. 64.
18. Tariq, M.A., et al., *Whole-transcriptome RNAseq analysis from minute amount of total RNA.* Nucleic acids research, 2011. **39**(18): p. e120-e120.
19. Toedling, J., et al., *Deep-sequencing protocols influence the results obtained in small-RNA sequencing.* PLoS One, 2012. **7**(2): p. e32724.
20. Linsen, S.E., et al., *Limitations and possibilities of small RNA digital gene expression profiling.* Nature methods, 2009. **6**(7): p. 474-476.

Chapter IV

Final Discussion and Future Directions

4.1 Final Discussion

The idea of *M. tuberculosis* RNA interacting with the host immune system is not new; described first in 1964 by Anne and Guy Youmans [1]. Studies were conducted throughout the 1960s and 1970s to define the immunogenic component found in ruptured mycobacterial cells [2-8]. Through differential centrifugation, the immunologically active material was found in the ribosomal fraction [1]. When the active fraction was combined with Freund's incomplete adjuvant and given to mice, the mice had a high degree of immunity to challenge with virulent *M. tuberculosis* [4]. In an effort to further define this material, it was found to be resistant to trace amounts of ribonuclease, indicating the RNA was relatively stable; likely double-stranded or having high degrees of secondary structure [8]. The *M. tuberculosis* RNA fraction was even tested as a means for tumor reduction in mice, further showing a strong immunogenic effect on the host [9, 10]. This work resonates with the sequencing work discussed in Chapter III, as well as the work done by Obregón-Henao et al. [11]. The studies documented in this thesis have provided a better understanding of the extracellular RNA population of *M. tuberculosis*. Many of the tRNA and rRNA fragments identified by Obregón-Henao et al. were confirmed through NGS, with the discovery of additional RNAs. Despite the advances made, there are many questions that remain unanswered, including the biological relevance and how the RNAs are being released. This chapter will examine several possibilities for the mechanism of release, biological relevance, and future directions to answer these questions.

4.1.1 Active release of CF RNA

One important area to investigate is how the RNA fragments are released from *M. tuberculosis*. At this point the possibility that the extracellular RNA is released from cells through bacterial lysis cannot be ruled out, although there is some indications that this may not be the case. The presence of the RNA in the CF as early as day three indicates active release, based on work done with secreted mycobacterial proteins. In one particular study, the concentration of isocitrate dehydrogenase (ICD) in the medium was used as an indicator of autolysis [12]. ICD levels were very low early in growth, and found to rapidly increase during the late log phase of growth, peaking at day 14 [12]. The low level of ICD early in growth indicated that proteins released into the CF at that time were not a result of autolysis. Studies also showed the population profile of proteins in the CF vary during the course of *M. tuberculosis* growth. This could suggest that the extracellular RNA population also changes as a result of growth phase.

The RNA fragments found in the CF are predicted to have high degrees of secondary structure. Stability of the RNA fragments is a possible explanation as to why these specific fragments survive in the CF if their release is a result of cell lysis. Complex secondary structure could protect against host, as well as bacterial, RNase activity in the extracellular environment [13]. While this is a possibility, it is complicated by the finding that not all of the *M. tuberculosis* tRNAs were present, even though all tRNAs possess similar secondary structure. The tRNAs identified were also compared to the codon/amino acid bias in the *M. tuberculosis* genome to see if there was any correlation. *M. tuberculosis* has a statistically significant preference for amino acids that are encoded by G + C-rich codons including Ala, Gly, Pro, Arg, and Trp [14]. In contrast, *M. tuberculosis* has a reduction in A + T-rich codons like Asn, Ile,

Lys, Phe, and Tyr [14]. Although all of the preferred tRNAs were identified in the CF, the tRNA with the most reads (SerT) is not on the preferred list. Two of the tRNAs identified, including the third most relatively abundant tRNA identified in the 100 ng CF RNA sample, are on the reduced list. There were also, 15 of the 25 tRNAs identified that are neither preferred or reduced. The specific tRNAs identified and differences in relative abundance lead to the hypothesis that there is selectivity in release of small extracellular RNA fragments.

The predicted secondary structures of regions of rRNA genes not represented by an aligned read were also relatively stable, further indicating that stability may not be a factor in the presence of the extracellular RNA fragments. If the RNA were being released through autolysis, it would be expected that rRNA fragments would make up the largest percentage of the extracellular RNA population, as rRNA makes up more than 95% of the RNA population within the bacterial cell [15].

If the extracellular *M. tuberculosis* RNA is not released through autolysis, an active secretion system could be involved. The Esx-1 *M. tuberculosis* secretion system has been well studied, and is responsible for the secretion of ESAT-6 and CFP-10, encoded on the RD1 gene. There is some evidence that the RD1 locus is involved in the transfer of DNA, as well as conjugation [16]. Obergón-Henao et al. evaluated the CF of the H37RvΔRD1 mutant for RNA, and found no change in the release of extracellular RNA [11]. The analysis of this particular mutant indicates the release is not linked to the Esx-1 secretion mechanism. There are other Esx secretion mechanisms in the *M. tuberculosis* genome that could be involved in RNA release.

The Sec secretion system is another possible route of RNA release from *M. tuberculosis*. *L. monocytogenes* is able to actively secrete small RNAs through its Sec secretion system, relying on SecA2 [17]. The mycobacterial Sec system also contains SecA2, which is often

associated with virulence [18]. The presence of SecA2 in both *L. monocytogenes* and *M. tuberculosis* leads to the idea that *M. tuberculosis* is actively releasing RNA through this mechanism.

Another mechanism of RNA transport to consider is through membrane vesicles (MV). As discussed in Chapter I, MVs released by *M. tuberculosis* have been shown to contain TLR2 lipoprotein agonists, which trigger a strong inflammatory response in the host [19]. Although the exact mechanism is unknown, the MVs are proposed to act as a delivery mechanism of mycobacterial molecules through compartments of infected host cells or released from the host cell [19]. It is possible that the small-stable RNA observed in the CF of *M. tuberculosis* could be packaged in these active MVs, and moved to the cytosol for recognition of PRRs, or released from the host cell.

4.1.2 Biological Activity of CF RNA

The biological significance of the extracellular RNA and its mode of action is not well understood. Previous work found that the apoptotic activity induced by extracellular *M. tuberculosis* RNA was the result of a caspase-8, TNF-α and caspase-1 independent mechanism [11]. This knowledge gives some clues as to how the extracellular RNA interacts with the host cell, although many questions remain.

If *M. tuberculosis* RNA release occurs in the early endosomal compartment of the host phagocytic cell, the extracellular RNA has the potential to interact with TLRs 3, 7, and 8. TLR3 recognizes dsRNA, leading to the production of a type I IFN response [20]. TLR7 and 8 also recognize dsRNAs, and research has been shown that they are able to detect bacterial RNA [21, 22]. Although the *M. tuberculosis* extracellular RNA identified through NGS is single stranded

in nature, the secondary structure predictions described in Chapter III show that the ssRNA contain regions of dsRNA.

TLR8 is an attractive candidate as a means of host recognition of the *M. tuberculosis* RNA, as activation of TLR8 can lead to apoptosis and caspase-8 activation [23, 24]. The mechanism through which TLR8 recognizes *M. tuberculosis* is unknown, and is of interest as human TLR8 polymorphisms have been associated with increased susceptibility to TB [25, 26].

As discussed in Chapter I, *M. tuberculosis* tRNAs have sequence homology to sequences able to activate TLR7 and 8 [27]. The ability of TLR7 and 8 to recognize specific sequences could be a possible explanation for the presence of more tRNA fragments than other RNA species in the CF. All of the tRNAs in *M. tuberculosis* have at least one of the sequences able to activate TLR7 and 8 [27]. Ten of the tRNA fragments identified through sequencing had at least one of the sequences specific for activation of TLR8 [27]. The rRNA and sRNA fragments were also interrogated for the presence of any of the specific activating sequences. Sequences that activate both TLR7 and 8 occurred in both the sRNA and rRNA fragment populations. Likewise, sequences that specifically activate TLR8 were found in both populations. Many of the sRNA and rRNA fragments contained both TLR7/8 and the TLR8 reactive sequences. While sequence homology could help explain how the extracellular RNA is recognized by the host cell, there is some evidence that the secondary structure itself may have more of an impact on immune stimulation than the specific sequence [21].

If the *M. tuberculosis* RNA is able to translocate into the cytosol of the host cell, another group of PRRs could be responsible for recognition including PKRs, RIG-I, MDA5, Lgp2, and NOD2. Although RNA is capable of activating an inflammasome response, these responses are dependent on caspase-1, so this pathway is unlikely [23].

RNA modification is an important factor to consider when evaluating the biological significance of the *M. tuberculosis* extracellular RNA. RNA nucleosides are often modified post-transcriptionally during maturation in both prokaryotes and eukaryotes [28]. Bacterial rRNA differs from eukaryotic rRNA in degree of methylation as well as the pseudouridine content [29]. Bacterial mRNA also differs from eukaryotic mRNA in the 3'-poly(A) tail, as eukaryotic mRNA has a long tail where bacterial mRNA has a short tail or none at all [30]. tRNA is the most heavily modified subgroup of RNA, although it is modified less in prokaryotes than eukaryotes [28]. The differences in RNA modifications between eukaryotes and prokaryotes help to decrease the recognition of self-RNA, and the possibility of resulting autoimmunity. tRNA and rRNA undergo multiple species-specific modifications that could alter their ability to activate different PRRs and signaling pathways. Dephosphorylation and other modifications have been shown to change the host response to bacterial nucleic acid with fewer modifications generally leading to more immunogenic RNA [28]. Specific modified nucleosides including pseudouridine are able to abrogate the ability of RNA to signal through TLR3, TLR7, and TLR8 [28]. Modifications are also able to suppress the ability of RNA to activate DCs and induce production of TNFα and IL-12 [28]. Research performed with the mouse-specific TLR13 helps illustrate the importance of RNA modifications and their interaction with the host immune system. TLR13 was found to be a bacterial sensor that recognizes a ssRNA segment within the peptidyl transferase loop of the conserved 23S rRNA. This loop binds antibiotics, and specific modifications of this segment leads to antibiotic resistance and the bacteria cannot be detected by TLR13 [31].

RNA modifications in mycobacteria have not been widely studied, and little is known about which modifications they possess and how they may interact with the host immune system.

Chan et al. recently began investigating mycobacterial RNA modifications using a chromatography-coupled mass spectrometric approach to identifying ribonucleosides in BCG tRNA [32]. In this study, multiple candidates were identified, including a modification that had not been previously described in tRNA, N^6,N^6-Dimethyladenosine [32]. Identifying the types of modified ribonucleosides in mycobacteria is the first step in understanding their biological role.

4.1 Future Directions

Additional NGS would help to verify the information found in the study discussed here. As discussed in Chapter III, there are points of technical variability to overcome, requiring a well planned and directed research effort. Specifically, the library preparation protocol would need to be optimized to ensure adaptors are properly ligating to the RNAs, especially the tRNAs. It would be helpful to evaluate the population of small-stable extracellular RNAs with a complementary sequencing platform, which may provide more insight.

There are several different sequencing platforms available, each with a unique library preparation protocol, which may have advantages or disadvantages for different types of material including small RNA [33-36]. There is a lot of debate as to which platform is best, which is complicated by the fact that there have been few studies independently testing platforms to determine details like error rate and yield [37-41]. For the continuation of these studies it will be important to consider the read length of each platform to get longer reads that will cover the entire fragment rather than the 35 base length provided by the SOLiD. The newest SOLiD instrument (5500xi) has an improved read length of 85 bases [41]. The Illumina HiSeq and the Life Technologies Ion Proton systems are both possibilities for additional sequencing. The Illumina has a read length of ≥100 bases, which would cover all of the CF RNA based on the

size observed using the Bioanalyzer [37, 41]. The Ion Proton sequencer utilizes semiconductor technology and does not require fluorescence/camera scanning leading to a higher speed and lower cost [40-42]. The Ion Proton would also be a good candidate for the CF RNA with a read length of up to 200 bases [41].

Some studies have been done to look at NGS and the effect of both library preparation and platform on sRNAs specifically [33-36]. Linsen et al. argue that bias is generally independent of the platform used, and more dependent on the method used for sRNA library preparation [36]. Differences in library preparation protocols have been demonstrated to be responsible for differences in miRNA expression even when the same platform is used [35]. Bias has been observed for sequences with end secondary structure using both the Illumina and SOLiD library preparation protocols, although SOLiD had a slightly larger bias [33, 35]. Another study found different miRNA populations when compairing the SOLiD and Illumina platforms [34]. Deep sequencing technology continues to advance, and the amount of information that can be gained will continue to improve with the technology. For example, sequencing platforms (so called "third generation sequencing") are in development that do not require sequence amplification, which should significantly decrease the incidence of bias in the sequencing data [37]. Until then it is imperative that researchers not rely too heavily on the abilities of NGS, and carefully design experiments while being aware of the limitations of this technology.

For a new NGS study, focus should be placed on biological replicates to confirm and gain confidence in the extracellular RNA population observed in the pilot study. As discussed earlier, NGS could be used to evaluate time course samples to assess differences in RNA released during growth. Differences in the extracellular RNA over time might indicate active versus passive

release. Chapter II describes a method developed for isolating sRNA from *M. tuberculosis* cells, which could be used in NGS to compare CF RNA to low molecular weight cellular RNA.

Because of the questions of accuracy in the NGS, it is important to verify the information obtained from NGS using a different method. Quantitative Real-Time PCR (qRT-PCR) and Northern blotting could both be used to confirm select sRNA fragments found using NGS [43]. To evaluate the rRNA fragments, it would be valuable to use qRT-PCR or another method to examine areas of coverage observed for the rRNA genes, as well as for the uncovered areas of the genes.

The development of a qRT-PCR assay would be beneficial as a way to evaluate selective release of RNA from *M. tuberculosis*. This type of assay could be used to track specific RNA products over time and evaluate extracellular accumulation. In addition, a qRT-PCR assay could be utilized to evaluate mutants for the absence of RNA secretion or release. Mutants of *M. tuberculosis* secretion systems, including SecA2, could be evaluated in this manner in an attempt to identify a release mechanism. In addition, MVs can be isolated and evaluated for RNA content.

A method of fluorescent tagging could be used to visualize extracellular RNA location in the host cell, and even provide information on interaction with specific PRRs. When investigating the role of TLR8 in *B. burgdorferi* infection, Cervantes et al. used confocal microscopy and a fluorescent version of the transfection agent polyethylenimine (PEI) to visualize the cellular location of the *B. burgdorferi* RNA. TLR8 was stained red, and the PEI with *B. burgdorferi* RNA was stained green, so co-localization within the phagosomal vacuole could be observed [44]. Labeling was also used to identify *L. monocytogenes* RNA in the cytosol of infected host cells [45]. The new labeling technique described by Hagmann et al., is

non-radioactive and non-toxic, using 5'-ethynyluridine (EU) to label RNA [45]. The *L. monocytogenes* grown in the presence of EU were used to infect host cells, and the location of RNA could be tracked over time [45]. This technique could be beneficial in elucidating the release of *M. tuberculosis* extracellular RNA, and possibly the biological significance.

It would also be important to evaluate different strains of *M. tuberculosis* for CF RNA composition. While the presence of small extracellular RNA was found in both HN878 and CDC1551, the RNA population in each has not been defined. Extracellular RNA was also detected, by visualization on TBE-Urea gel, in *M. smegmatis*. The avirulent H37Ra should also be evaluated for extracellular RNA. Because of the presence of extracellular RNA in *M. smegmatis*, H37Ra would be expected to also release RNA into the CF.

Immunological studies are important to many aspects of elucidating the role of the extracellular RNA fragments and answering the broader question of biological importance. Immunological studies are critical in the illumination of the mechanism of host cell activation and could be combined with the possibilities suggested here. NGS would give some indication of differences over time and could be paired with immunological studies to look at changes in biological activity over time. Induction of type I IFNs could indicate a possible receptor for the extracellular RNA based on studies indicating TLR signaling may not be involved in type I IFN in *M. tuberculosis*, implicating cytosolic receptors [46, 47]. A mycobacterial secretion system may be responsible for the translocation of the RNA into the cytosol, although secretion relying on the RD1 locus has been ruled out in previous work [11, 48].

Identification of a specific RNA population or species that contained the active material could be advantageous for immunological studies. Chionh et al. recently developed a technique for targeted purification of a range of RNA species from total RNA, rRNA, tRNA, ncRNA,

using a multidimensional high-performance liquid chromatography (HPLC) platform [49]. Two ranges of size-exclusion chromatography (SEC) and ion-pair reverse-phase chromatography (IP RPC), allowing for the full separation of RNA species between 20 and 10,000 nucleotides in length. An additional advantage of this method is that posttranscriptional modifications of nucleosides remain intact, especially important if the modifications do play a significant role in the biological activity as is suspected.

Additional NGS, qRT-PCR, Northern blotting, and fragmentation could help identify specific fragments of interest, which could be synthesized for testing in immunological assays. Biological studies could also be used to investigate potential differences in biological activity between CDC1551, HN878, H37Ra, and *M. smegmatis*. Together with the NGS data, the biological data could implicate specific small RNA species in different biological consequences.

A better understanding of the pathogenesis of *M. tuberculosis* is critical to the spread and control of TB. Continued research of sRNAs in *M. tuberculosis* could lead to new targets for drugs, vaccines, or diagnostics. sRNAs in *M. tuberculosis* are currently being studied by several groups, and there is still a lot unknown about the specific function of individual sRNAs. Continued study of *M. tuberculosis* sRNAs and their functions will also help to target specific extracellular sRNAs and their biological function.

References

1. Youmans, A.S. and G.P. Youmans, *Nature of the labile immunogenic substance in the particulate fraction isolated from Mycobacterium tuberculosis.* Journal of bacteriology, 1964. **88**(4): p. 1030-1037.
2. Youmans, A.S. and G.P. Youmans, *Immunogenic activity of a ribosomal fraction obtained from Mycobacterium tuberculosis.* Journal of Bacteriology, 1965. **89**(5): p. 1291-&.
3. Youmans, A.S. and G.P. Youmans, *Preparation of highly immunogenic ribosomal fractions of Mycobacterium tuberculosis by use of sodium dodecyl sulfate.* Journal of bacteriology, 1966. **91**(6): p. 2139-2145.
4. Youmans, A.S. and G.P. Youmans, *Effect of trypsin and ribonuclease on the immunogenic activity of ribosomes and ribonucleic acid isolated from Mycobacterium tuberculosis.* Journal of bacteriology, 1966. **91**(6): p. 2146-2154.
5. Youmans, A.S. and G.P. Youmans, *Preparation and effect of different adjuvants on the immunogenic activity of mycobacterial ribosomal fraction.* Journal of bacteriology, 1967. **94**(4): p. 836-843.
6. Youmans, A.S. and G.P. Youmans, *Ribonucleic acid deoxyribonucleic acid and protein content of cells of different ages of Mycobacterium tuberuclosis and relationship to immunogenicity.* Journal of Bacteriology, 1968. **95**(2): p. 272-&.
7. Youmans, A.S. and G.P. Youmans, *Factors affecting immunogenic activity of mycobacterial ribosomal and ribonucleic acid preparations.* Journal of bacteriology, 1969. **99**(1): p. 42-50.
8. Youmans, A.S. and G.P. Youmans, *Immunogenic mycobacterial ribosomal and ribonucleic Acid preparations: chemical and physical characteristics.* Infection and immunity, 1970. **2**(5): p. 659-668.
9. Millman, I., et al., *Effect of the H37Ra strain of M. tuberculosis and of a mycobacterial RNA fraction on tumor growth.* Experimental Biology and Medicine, 1974. **147**(3): p. 765-768.
10. Millman, I., et al., *Mycobacterial RNA. A comparison with intact mycobacteria for suppression of murine tumor growth.* Journal of medicine, 1976. **7**(3-4): p. 249.
11. Obregon-Henao, A., et al., *Stable extracellular RNA fragments of Mycobacterium tuberculosis induce early apoptosis in human monocytes via a caspase-8 dependent mechanism.* PLoS One, 2012. **7**(1): p. e29970.
12. Andersen, P., et al., *Proteins released from Mycobacterium tuberculosis during growth.* Infection and Immunity, 1991. **59**(6): p. 1905-1910.
13. Deutscher, M.P., *Degradation of RNA in bacteria: comparison of mRNA and stable RNA.* Nucleic acids research, 2006. **34**(2): p. 659-666.
14. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-544.
15. Peano, C., et al., *An efficient rRNA removal method for RNA sequencing in GC-rich bacteria.* Microbial informatics and experimentation, 2013. **3**(1): p. 1.
16. Flint, J.L., et al., *The RD1 virulence locus of Mycobacterium tuberculosis regulates DNA transfer in Mycobacterium smegmatis.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(34): p. 12598-12603.

17. Abdullah, Z., et al., *RIG-I detects infection with live Listeria by sensing secreted bacterial nucleic acids.* Embo Journal, 2012. **31**(21): p. 4153-4164.
18. Feltcher, M.E. and M. Braunstein, *Emerging themes in SecA2-mediated protein export.* Nature Reviews Microbiology, 2012. **10**(11): p. 779-789.
19. Prados-Rosales, R., et al., *Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice.* Journal of Clinical Investigation, 2011. **121**(4): p. 1471-1483.
20. Alexopoulou, L., et al., *Recognition of double-stranded RNA and activation of NF-[kappa]B by Toll-like receptor 3.* Nature, 2001. **413**(6857): p. 732-738.
21. Sarvestani, S.T., B.R. Williams, and M.P. Gantier, *Human Toll-like receptor 8 can be cool too: implications for foreign RNA sensing.* Journal of Interferon & Cytokine Research, 2012. **32**(8): p. 350-361.
22. Heil, F., et al., *Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8.* Science, 2004. **303**(5663): p. 1526-1529.
23. Kanneganti, T.-D., et al., *Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3.* Nature, 2006. **440**(7081): p. 233-236.
24. Takahashi, K., et al., *Cutting edge: Roles of caspase-8 and caspase-10 in innate immune responses to double-stranded RNA.* Journal of Immunology, 2006. **176**(8): p. 4520-4524.
25. Davila, S., et al., *Genetic Association and Expression Studies Indicate a Role of Toll-Like Receptor 8 in Pulmonary Tuberculosis.* PLoS Genet, 2008. **4**(10): p. e1000218.
26. Thada, S., V. Valluri, and S.L. Gaddam, *Influence of toll like receptor gene polymorphisms to tuberculosis susceptibility in humans.* Scandinavian journal of immunology, 2013.
27. Sioud, M., *Single-stranded small interfering RNA are more immunostimulatory than their double-stranded counterparts: A central role for 2 '-hydroxyl uridines in immune responses.* European Journal of Immunology, 2006. **36**(5): p. 1222-1230.
28. Kariko, K., et al., *Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA.* Immunity, 2005. **23**(2): p. 165-175.
29. Kiss, T., *Small nucleolar RNA‐guided post‐transcriptional modification of cellular RNAs.* The EMBO journal, 2001. **20**(14): p. 3617-3622.
30. Koski, G.K., et al., *Cutting edge: Innate immune system discriminates between RNA containing bacterial versus eukaryotic structural features that prime for high-level IL-12 secretion by dendritic cells.* Journal of Immunology, 2004. **172**(7): p. 3989-3993.
31. Oldenburg, M., et al., *TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification.* Science, 2012. **337**(6098): p. 1111-5.
32. Chan, C.T.Y., et al., *Identification of N-6,N-6-Dimethyladenosine in Transfer RNA from Mycobacterium bovis Bacille Calmette-Guerin.* Molecules, 2011. **16**(6): p. 5168-5181.
33. Tian, G., et al., *Sequencing bias: comparison of different protocols of microRNA library construction.* BMC biotechnology, 2010. **10**(1): p. 64.
34. Fehniger, T.A., et al., *Next-generation sequencing identifies the natural killer cell microRNA transcriptome.* Genome research, 2010. **20**(11): p. 1590-1604.
35. Toedling, J., et al., *Deep-sequencing protocols influence the results obtained in small-RNA sequencing.* PLoS One, 2012. **7**(2): p. e32724.
36. Linsen, S.E., et al., *Limitations and possibilities of small RNA digital gene expression profiling.* Nature methods, 2009. **6**(7): p. 474-476.

37. Glenn, T.C., *Field guide to next-generation DNA sequencers.* Molecular Ecology Resources, 2011. **11**(5): p. 759-769.
38. Mardis, E.R., *Next-generation DNA sequencing methods.* Annu. Rev. Genomics Hum. Genet., 2008. **9**: p. 387-402.
39. Metzker, M.L., *Sequencing technologies—the next generation.* Nature Reviews Genetics, 2010. **11**(1): p. 31-46.
40. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC genomics, 2012. **13**(1): p. 341.
41. Liu, L., et al., *Comparison of next-generation sequencing systems.* BioMed Research International, 2012. **2012**.
42. Rothberg, J.M., et al., *An integrated semiconductor device enabling non-optical genome sequencing.* Nature, 2011. **475**(7356): p. 348-352.
43. Pelly, S., W.R. Bishai, and G. Lamichhane, *A screen for non-coding RNA in Mycobacterium tuberculosis reveals a cAMP-responsive RNA that is expressed during infection.* Gene, 2012. **500**(1): p. 85-92.
44. Cervantes, J.L., et al., *Human TLR8 is activated upon recognition of Borrelia burgdorferi RNA in the phagosome of human monocytes.* Journal of Leukocyte Biology, 2013. **94**(6): p. 1231-1241.
45. Hagmann, C.A., et al., *RIG-I Detects Triphosphorylated RNA of Listeria monocytogenes during Infection in Non-Immune Cells.* Plos One, 2013. **8**(4).
46. Stanley, S.A., et al., *The Type I IFN Response to Infection with Mycobacterium tuberculosis Requires ESX-1-Mediated Secretion and Contributes to Pathogenesis.* J Immunol, 2007. **178**(5): p. 3143-3152.
47. Pandey, A.K., et al., *NOD2, RIP2 and IRF5 play a critical role in the type I interferon response to Mycobacterium tuberculosis.* PLoS pathogens, 2009. **5**(7): p. e1000500.
48. Monroe, K.M., S.M. McWhirter, and R.E. Vance, *Induction of type I interferons by bacteria.* Cellular Microbiology, 2010. **12**(7): p. 881-890.
49. Chionh, Y.H., et al., *A multidimensional platform for the purification of non-coding RNA species.* Nucleic Acids Research, 2013. **41**(17).

APPENDIX A

Appendix A: tRNA fragments identified in 50 ng CF RNA sample. Sequence reads from *M. tuberculosis* CF RNA 50 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, tRNA name, and the consensus sequence for each fragment.

Chromosome Region	Length	Coverage	tRNA Name	tRNA Name	Sequence
11149..11185	37	5	AlaTGC	AlaT	AGGCAGGGGTCAGGGGTTCGAGTCCCCTAGGCTCCAc
4168343..4168387	45	5	SerGGA	SerV	GCGGAGGATTCGCCTAGTGGCCTATGGCGCTCGCCTGGAACGCGg
3348659..3348693	35	6	GlnCTG	GlnU	CTGGGGTACCAGGACTCGAACCTAGAATGGCTGAa
2510597..2510670	74	7	ValTAC	ValV	GTGGGCGCGGACGGGATCGAACCGCCGACCGCTGGTGTGTAAAACCAGAGCTCTACCGCTGAGCTACGCGCCCa
2980965..2981000	36	7	ValGAC	ValU	AGTGAAACTGGTTCAATCCCAGTATCGCGCACCACg
2969894..2969928	35	8	CysGCA	CysU	CGTGCACACGGGTTCGATTCCCGTCTCCACCTCCa
1512767..1512809	43	9	LeuTAG	LeuW	GCCGGTGCTCGAAAGAGTTTGAGGGTTCGAGTCCCTCCGCCCg
1946652..1946687	36	9	ProGGG	ProT	TCAAGTGGTCGCAGGTTCAAATCCTGTCAGCCCGAc
2765331..2765366	36	9	ProTGG	ProU	TCGGGGTGACAGGATTTGAACCTGCGGCCTTCCGCt
4126582..4126616	35	11	ProCGG	ProY	AAGAGGCCGTGGGTTCAAATCCCGCCACCCCGACc
1446232..1446265	34	18	ArgCCG	ArgV	GAGGCAGACGCTCTATCCCCTGAGCTACGGGGGc
731510..731569	60	23	ThrGGT	ThrT	CGGCAGAGCGTTTCCATGGTAAGGAAAAGGTCAACGGTTCGATTCCGTTAGGGGGCTCGg
2794179..2794231	53	23	ArgTCT	ArgW	TCCGTAGCTCAGGTGGATAGAGCAAGGGCCTTCTAATCCCTAGGTCGCACGTt
924253..924286	34	28	PheGAA	PheU	CGGAAGGTCGGCGGTTCGATCCCGCCCCTGGCCa
4081365..4081430	66	30	ThrCGT	ThrU	AGCCGCCTAGGAGAATCGAACTCCTGACCTATTCATTACGAGTGAATCGCTCTACCGACTGAGCTa

Chromosome Region	Length	Coverage	tRNA Name	tRNA Name	Sequence
25684..25727	44	32	LeuCAG	LeuT	CAGTGTCCTTCGGGACGTGGGGGTTCAAGTCCCCCTTCGCCCAc
850661..850710	50	32	ThrTGT	ThrV	ACTGACGACCGCTCGCTTACAAGGCGAGTGCTCTACCACTGAGCTAAGGa
2969754..2969787	34	41	GlyGCC	GlyU	CGGATGTAGCGCAGTTGGTAGCGCATCACCTTGc
10907..10961	55	52	IleGAT	IleT	TAGAGCGCTTCGCTGATAACGAAGAGGTCGGAGGTTCGAGTCCTCCTAGGCCCAc
924131..924187	57	95	AspGTC	AspT	AGCGCGCCGCCCTGTCACGGCGGAGGTCGCGGGTTCGAGTCCCGTCAGGGTCGCCAg
2969974..2970012	39	107	ValGAC	ValU	GACACGGAAGAGGTCACTGGTTCAATCCCAGTATCGCGc
3559362..3559443	82	143	MetCAT	MetU	TCCGCTGGTAGCGGGGACAGGATTCGAACCTGCGACCTCTGGGTTATGAGCCCAGCGAGCTACCGAGCTGCTCCACCCCGCg
2827865..2827934	70	154	HisGTG	HisT	TCAGTTGGTAGAGCACCAGGTTGTGATCCTGGGTGTCGCGGGTTCGAGTCCCGTCACTCACCCCAACAGg
2619405..2619461	57	518	AsnGTT	AsnT	TGCTCCCCCGGGAGGACTCGAACCTCCAACCCTTCGGTTAACAGCCGAACGCTCTGc
386201..386275	75	624	GlyCCC	GlyU	CGGAGCCGATGACGGGAATCGAACCCGCGTATTCAGCTTGGGAAGCTGATGTTCTGCCATTGAACTACATCGGCa
2402027..2402076	50	918	LeuGAG	LeuU	TAGTGCCCTACTAATGGGCGTGGGGGTTCAAGTCCCCCCTCGGACACAAc
4216965..4217038	74	6309	SerGCT	SerT	CTGCGGAGGCGAGAGGATTTGAACCTCCGGTCCCCTTGAAGGGGGACAACTCATTAGCAGTGAGCCCCATTCGg

APPENDIX B

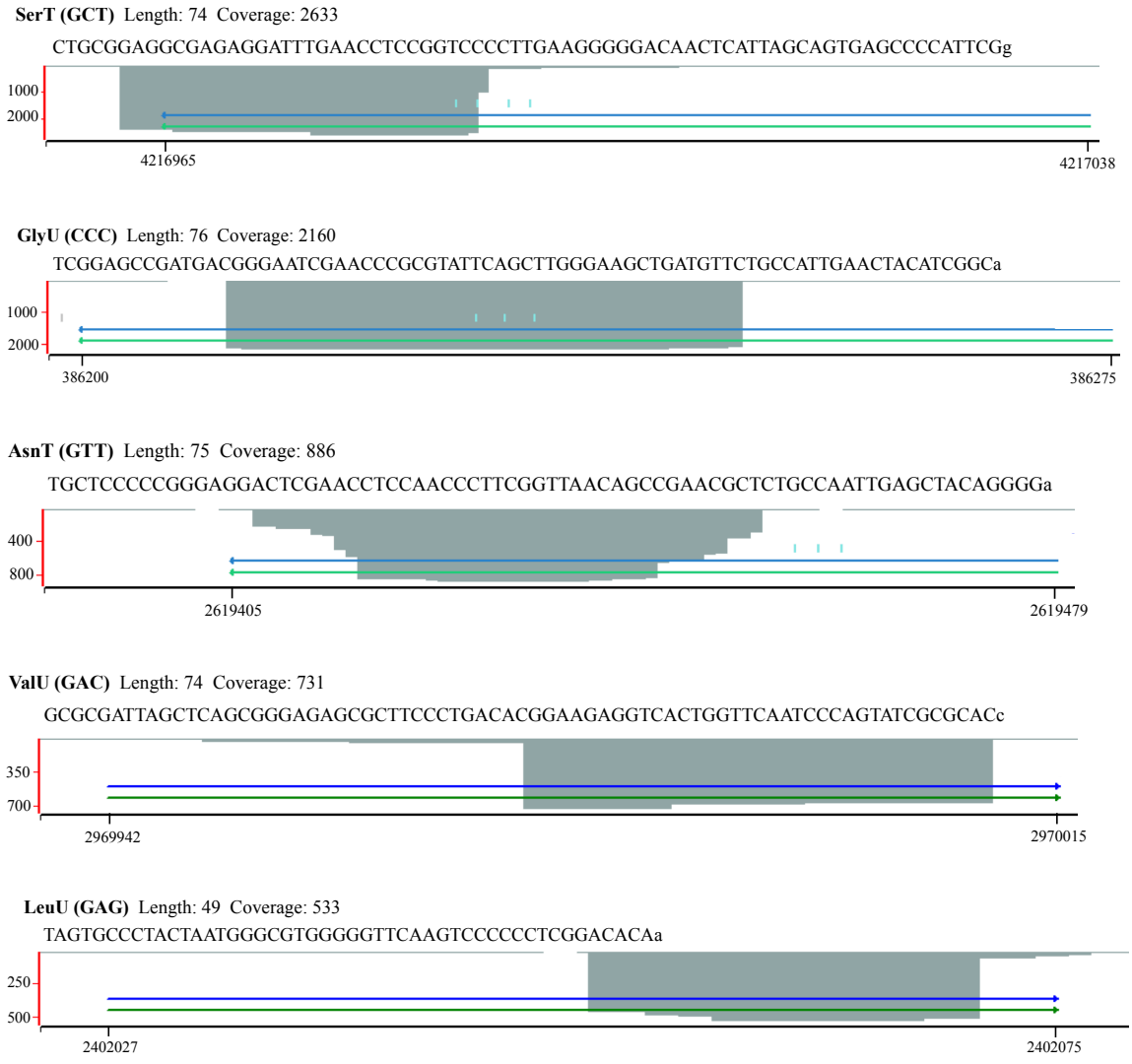
Appendix B: tRNA fragments identified in 100 ng CF RNA sample. Sequence reads from *M. tuberculosis* CF RNA 100 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, tRNA name, and the consensus sequence for each fragment.

Chromosome Region	Length	Coverage	tRNA Name	tRNA Name	Sequence
4216965..4217038	74	2633	Ser GCT	SerT	CTGCGGAGGCGAGAGGATTTGAACCTCCGG TCCCCTTGAAGGGGGACAACTCATTAGCAGT GAGCCCCATTCGg
386200..386275	76	2160	Gly CCC	GlyU	TCGGAGCCGATGACGGGAATCGAACCCGCG TATTCAGCTTGGGAAGCTGATGTTCTGCCATT GAACTACATCGGCa
2619405..2619479	75	886	Asn GTT	AsnT	TGCTCCCCCGGGAGGACTCGAACCTCCAAC CCTTCGGTTAACAGCCGAACGCTCTGCCAAT TGAGCTACAGGGGa
2969942..2970015	74	731	Val GAC	ValU	GCGCGATTAGCTCAGCGGGAGAGCGCTTCC CTGACACGGAAGAGGTCACTGGTTCAATCC CAGTATCGCGCACCc
2402027..2402075	49	533	Leu GAG	LeuU	TAGTGCCCTACTAATGGGCGTGGGGGTTCAA GTCCCCCCTCGGACACAa
3559362..3559444	83	350	Met CAT	MetU	TCCGCTGGTAGCGGGGACAGGATTCGAACC TGCGACCTCTGGGTTATGAGCCCAGCGAGCT ACCGAGCTGCTCCACCCCGCGt
3348659..3348693	35	107	Gln CTG	GlnU	CTGGGGTACCAGGACTCGAACCTAGAATGG CTGAa
1446232..1446266	35	92	Arg CCG	ArgV	GAGGCAGACGCTCTATCCCCTGAGCTACGGG GGCg
2827853..2827928	76	80	His GTG	HisT	GGTGAGTGTAGTTCAGTTGGTAGAGCACCA GGTTGTGATCCTGGGTGTCGCGGGTTCGAGT CCCGTCACTCACCCc
733525..733597	73	55	Trp CCA	TrpT	GGGGCGTAGCTCAACTGGCAGAGCAGCGGT CTCCAAAACCGCAGGTTGCAGGTTCAAGTC CTGTCGCCCCTGc
731495..731569	75	52	Thr GGT	ThrT	CCCCCTTAGCTCAGTCGGCAGAGCGTTTCCA TGGTAAGGAAAAGGTCAACGGTTCGATTCC GTTAGGGGGCTCGg

Chromosome Region	Length	Coverage	tRNA Name	tRNA Name	Sequence
850643..850714	72	47	Thr TGT	ThrV	GCCCCCTGTCAGGATTGAACTGACGACCGCT CGCTTACAAGGCGAGTGCTCTACCACTGAGC TAAGGAGGCc
25684..25722	39	44	Leu CAG	LeuT	CAGTGTCCTTCGGGACGTGGGGGTTCAAGT CCCCCTTCg
4081365..4081434	70	39	Thr CGT	ThrU	AGCCGCCTAGGAGAATCGAACTCCTGACCTA TTCATTACGAGTGAATCGCTCTACCGACTGA GCTAAGGc
2510597..2510670	74	39	Val TAC	ValV	GTGGGCGCGGACGGGATCGAACCGCCGACC GCTGGTGTGTAAAACCAGAGCTCTACCGCTG AGCTACGCGCCCa
2794187..2794254	68	36	Arg TCT	ArgW	TCAGGTGGATAGAGCAAGGGCCTTCTAATCC CTAGGTCGCACGTTCGAGTCGTGCCGGGGG CACTGTg
1946652..1946687	36	32	Pro GGG	ProT	TCAAGTGGTCGCAGGTTCAAATCCTGTCAGC CCGAc
4168343..4168418	76	26	Ser GGA	SerV	GCGGAGGATTCGCCTAGTGGCCTATGGCGCT CGCCTGGAACGCGGGTTGGGTTAACAGCCC TCGCGGGTTCAAATc
2969894..2969930	37	11	Cys GCA	CysU	CGTGCACACGGGTTCGATTCCCGTCTCCACC TCCAGg
2765541..2765595	55	11	Gly TCC	GlyV	GCGGGCGTAGCTCAATGGTAGAGCCCTAGTC TTCCAAACTAGCGACGCGGGTTCg
11145..11186	42	10	Ala TGC	AlaT	TGCAAGGCAGGGGTCAGGGGTTCGAGTCCC CTAGGCTCCACa
924131..924184	54	8	Asp GTC	AspT	AGCGCGCCGCCCTGTCACGGCGGAGGTCGC GGGTTCGAGTCCCGTCAGGGTCGc
10907..10964	58	6	Ile GAT	IleT	TAGAGCGCTTCGCTGATAACGAAGAGGTCG GAGGTTCGAGTCCTCCTAGGCCCACGAc
4126577..4126615	39	5	Pro CGG	ProY	GGACGAAGAGGCCGTGGGTTCAAATCCCGC CACCCCGAc
2969503..2969588	86	5	Val CAC	ValT	CGGCTGGGATCGAACCAGCGACCTTCCGCG TGTGAAGCGGACGCTCTCCCACTGAGCCAC GGGACCGGCGCCGAGGAGATGAACGa

APPENDIX C

Appendix C: Individual tRNA gene sequence alignments for 100 ng sample. NGS data was aligned to the *M. tuberculosis* H37Rv GenBank reference genome. Each tRNA gene was isolated in the NextGENe viewer to show coverage. Grey regions show depth of coverage, the blue arrows are used to indicate gene locations within the reference file, and the green arrows are used to indicate mRNA. Each tRNA is labeled with name, length of the fragment, and depth of coverage.

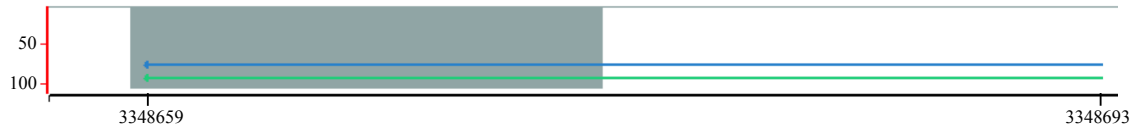


MetU (CAT) Length: 83 Coverage: 350

TCCGCTGGTAGCGGGGACAGGATTCGAACCTGCGACCTCTGGGTTATGAGCCCAGCGAGCTACCGAGCTGCTCCACCCCGCGt

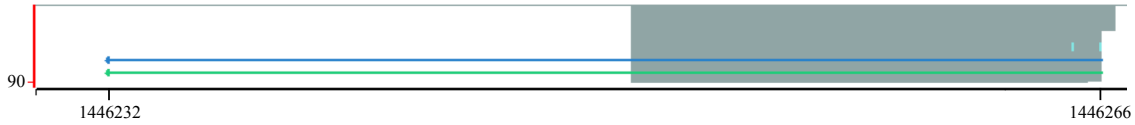
GlnU (CTG) Length: 35 Coverage: 107

CTGGGGTACCAGGACTCGAACCTAGAATGGCTGAa



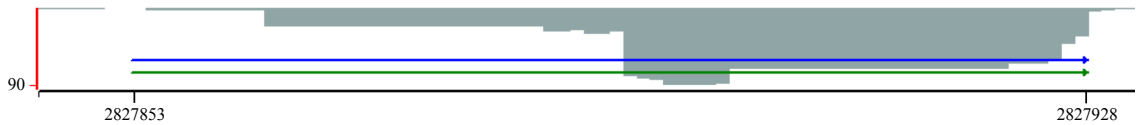
ArgV (CCG) Length: 35 Coverage: 92

GAGGCAGACGCTCTATCCCCTGAGCTACGGGGGCg



HisT (GTG) Length: 76 Coverage: 80

GGTGAGTGTAGTTCAGTTGGTAGAGCACCAGGTTGTGATCCTGGGTGTCGCGGGTTCGAGTCCCGTCACTCACCCc



TrpT (CCA) Length: 73 Coverage: 55

GGGGCGTAGCTCAACTGGCAGAGCAGCGGTCTCCAAAACCGCAGGTTGCAGGTTCAAGTCCTGTCGCCCCTGc

ThrT (GGT) Length: 75 Coverage: 52

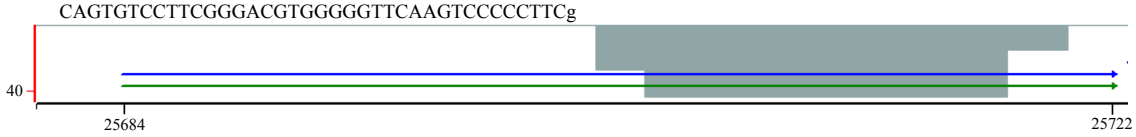
CCCCCTTAGCTCAGTCGGCAGAGCGTTTCCATGGTAAGGAAAAGGTCAACGGTTCGATTCCGTTAGGGGGCTCGg

ThrV (TGT) Length: 72 Coverage: 47

GCCCCCTGTCAGGATTGAACTGACGACCGCTCGCTTACAAGGCGAGTGCTCTACCACTGAGCTAAGGAGGCc

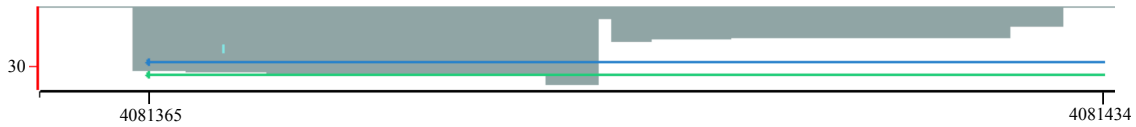
LeuT (CAG) Length: 39 Coverage: 44

CAGTGTCCTTCGGGACGTGGGGGTTCAAGTCCCCCTTCg



ThrU (CGT) Length: 70 Coverage: 39

AGCCGCCTAGGAGAATCGAACTCCTGACCTATTCATTACGAGTGAATCGCTCTACCGACTGAGCTAAGGc



ValV (TAC) Length: 74 Coverage: 39

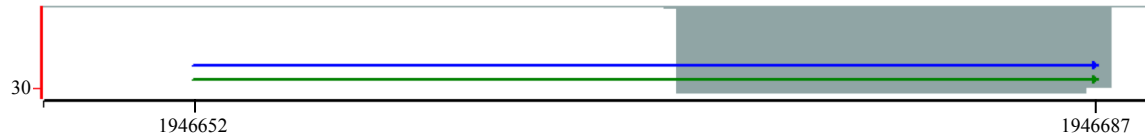
GTGGGCGCGGACGGGATCGAACCGCCGACCGCTGGTGTGTAAAACCAGAGCTCTACCGCTGAGCTACGCGCCCa

ArgW (TCT) Length: 68 Coverage: 36

TCAGGTGGATAGAGCAAGGGCCTTCTAATCCCTAGGTCGCACGTTCGAGTCGTGCCGGGGGCACTGTg

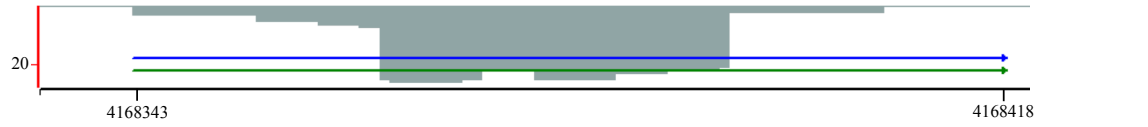
ProT (GGG) Length: 36 Coverage: 32

TCAAGTGGTCGCAGGTTCAAATCCTGTCAGCCCGAc



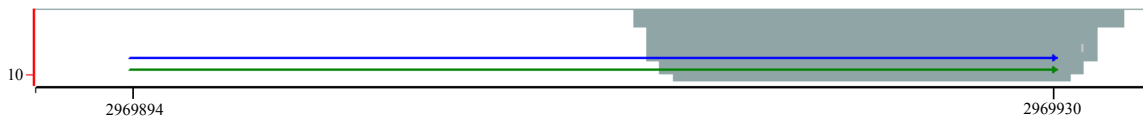
SerV (GGA) Length: 76 Coverage: 26

GCGGAGGATTCGCCTAGTGGCCTATGGCGCTCGCCTGGAACGCGGGTTGGGTTAACAGCCCTCGCGGGTTCAAATc



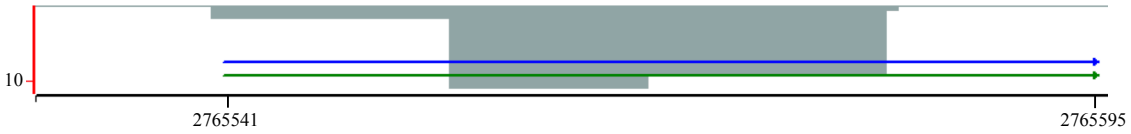
CysU (GCA) Length: 37 Coverage: 11

CGTGCACACGGGTTCGATTCCCGTCTCCACCTCCAGg



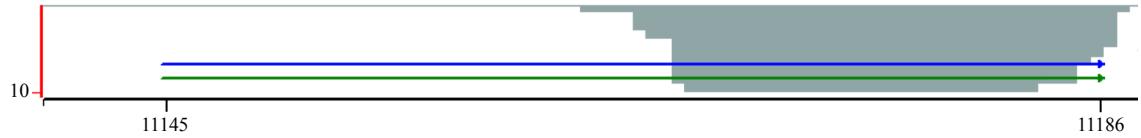
GlyV (GCA) Length: 55 Coverage: 11

GCGGGCGTAGCTCAATGGTAGAGCCCTAGTCTTCCAAACTAGCGACGCGGGTTCg



AlaT (TGC) Length: 42 Coverage: 10

TGCAAGGCAGGGGTCAGGGGTTCGAGTCCCCTAGGCTCCACa



AspT (GTC) Length: 54 Coverage: 8

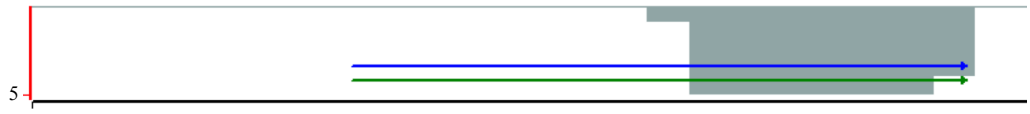
AGCGCGCCGCCCTGTCACGGCGGAGGTCGCGGGTTCGAGTCCCGTCAGGGTCGc

IleT (GAT) Length: 58 Coverage: 6

TAGAGCGCTTCGCTGATAACGAAGAGGTCGGAGGTTCGAGTCCTCCTAGGCCCACGAc

ProY (CGG) Length: 39 Coverage: 5

GGACGAAGAGGCCGTGGGTTCAAATCCCGCCACCCCGAc



ValT (CAC) Length: 86 Coverage: 5

TAGAGCGCTTCGCTGATAACGAAGAGGTCGGAGGTTCGAGTCCTCCTAGGCCCACGAc

APPENDIX D

Appendix D: rRNA fragments identified in 50 ng CF RNA sample. Sequence reads from *M. tuberculosis* CF RNA 50 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, gene name given by NextGENe, gene, location on the gene, and the consensus sequence for each fragment.

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1475756..1475806	51	5	rrl(+2098)	rrl	towards the middle	TAACCCGCAAGGGTGAAGCGGAGAATTTAAGCCCCAGTAAACGGCGGTGGt
1476332..1476370	39	5	rrl(+2674)	rrl	towards the end	GTACCCCGGGGATAACAGGCTGATCTTCCCCAAGAGTCc
1472989..1473023	35	5	rrs(+1143)	rrs	towards the end	GAGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGg
1474757..1474791	35	6	rrl(+1099)	rrl	close to the middle	CGAAAGGGAAACAGCCCAGATCGCCGGCTAAGGCc
1476145..1476190	46	6	rrl(+2487)	rrl	towards the end	GGGGCGGTTGCCTCCTAAAATGTAACGGAGGCGCCCAAAGGTTCCc
1474377..1474416	40	7	rrl(+719)	rrl	towards the middle	GGGGTAGCCGCAGCGAAAGCGAGTCTGAATAGGGCGACCc
1472388..1472432	45	7	rrs(+542)	rrs	towards the middle	TGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGGTGGTTTGTCGc
1475175..1475208	34	8	rrl(+1517)	rrl	middle	GGGTTGATATTCCCGTACCCGTGTGTGGGCGCCc
1472053..1472121	69	10	rrs(+207)	rrs	towards the beginning	AGCGCTTTAGCGGTGTGGGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCTACCAAGGc
1471704..1471738	35	11	rrs(-108)	Intergenic	Intergenic region between murA and rrs	CAATAGTGTGTTTGGTGGTTTCACATTTTTGTTGt
1474878..1474977	103	11	rrl(+1219)	rrl	middle	GCGTAATAGCTCACTGGTCAAGTGATTGTGCGCCGATAATGTAGCGGGGCTCAAGCACACCGCCGAAGCCGCGGCACATCCACCTTGTGGTGGGTGTGGGtag

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1473453..1473499	47	17	rrl(-159)	intergenic	between rrs and rrl. Closer to rrs	GAGCCGGGTGCATGACAACAAAGTTGGCCACCAACACACTGTTGGGt
1476208..1476271	64	17	rrl(+2550)	rrl	towards the end	CAGGTGGCGAGTGTAAATGCACAAGGGAGCTTGACTGCGAGACTTACAAGTCAAGCAGGGACGa
1474108..1474150	43	18	rrl(+450)	rrl	towards the beginning	CAATTCCCGAGTAGCAGCGGGCCCGTGGAATCCGCTGTGAATc
1472604..1472703	147	20	rrs(+735)	rrs	middle	agtaactgacgctgaggagcgaAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATCCGTGCCGTAGCTAacgcattaagtaccccgcctgggga
1474165..1474241	77	22	rrl(+507)	rrl	towards the beginning	GGTAAGCCTAAATACTCCTCGATGACCGATAGCGGATTAGTACCGTGAGGGAATGGTGAAAAGTACCCCGGGAGGGg
1473754..1473812	59	30	rrl(+96)	rrl	close to the beginning	GCGTGGATCCGAGGATTTCCGAATGGGGAAACCCAGCACGAGTGATGTCGTGCTACCCg
1473887..1473973	87	30	rrl(+229)	rrl	towards the beginning	TTGTGATTCCGCAAGTAGTGGCGAGCGAACGCGGAACAGGCTAAACCGCACGCATGGGTAACCGGGTAGGGGTTGTGTGTGCGGGGt
1474016..1474103	88	33	rrl(+358)	rrl	towards the beginning	CAGAAAGTGTCGTGGTTAGCGGAAGTGGCCTGGGATGGTCTGCCGTAGACGGTGAGAGCCCGGTACGCGAAAACCCGGCACCTGCCTa
1473037..1473134	98	33	rrs(+1191)	rrs	towards the end	CATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAGCGAATCCTTAAAAGCCGGTCt
1472173..1472240	68	43	rrs(+327)	rrs	towards the beginning	ACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCg
1475071..1475128	58	45	rrl(+1413)	rrl	middle	GAGAACCTTGCCCGCCGAAAGACCAAGGGTTCCTGGGCCAGGCCAGTCCGCCCAGGGt
1475553..1475652	139	70	rrl(+1876)	rrl	towards the middle	ggcgtacgagataactatGGTTAAGGAACTCGGCAAAATGCCCCCGTAACTTCGGGAGAAGGGGGACCGGAATATCGTGAACACCCTTGCGGTGGGAGCGGGATCCGGTCGCAGAAACcagtgaggagcgactgtttac
1471911..1472009	99	80	rrs(+65)	rrs	close to the beginning	CGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATAAGCCTGGGAAACTGGGt
1476055..1476135	81	89	rrl(+2397)	rrl	towards the end	CGTTGTTGAAATACCACTCTGATCGTATTGGGCATCTAACCTCGAACCCTGAATCGGGTTTAGGGACAGTGCCTGGCGGGt
1476625..1476708	84	89	rrl(+2967)	rrl	close to the end	CGGTCAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTCCAAGATCAGGTTTCTCACCCACTTGGTGGGATAAGGCCCc

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1474806..1474873	68	97	rrl(+1148)	rrl	middle	AAGTGGGAAAGGATGTGCAGTCGCAAAGACAACCAGGAGGTT GGCTTAGAAGCAGCCACCCTTGAAAg
1476476..1476571	96	107	rrl(+2818)	rrl	close to the end	GGGTTTAGAACGTCGTGAGACAGTTCGGTCTCTATCCGCCGCG CGCGTCAGAAACTTGAGGAAACCTGTCCCTAGTACGAGAGGA CCGGGACGGAc
1473259..1473327	69	179	rrs(+1413)	rrs	close to the end	CGGTAACACCCGAAGCCAGTGGCCTAACCCTCGGGAGGGAGC TGTCGAAGGTGGGATCGGCGATTGGGa
1475903..1476002	189	272	rrl(+2201)	rrl	towards the end	gcacgaatggcgtaacgacttctcaactgtctcaaccatagacTCGGCGAAATTGCACTA CGAGTAAAGATGCTCGTTACGCGCGGCAGGACGAAAAGACCC CGGGACCTTCACTACAACTTGGTATTGATGTTCGGTACGGTttgtgt aggataggtgggagactgtgaaacctcgacgccagttggg
1476906..1477005	111	435	rrf(+2)	rrf	covers nearly the entire gene	acggCGGCCACAGCGGCAGGGAAACGCCCGGTCCCATTCCGAAC CCGGAAGCTAAGCCTGCCAGCGCCGATGATACTGCCCCTCCGG GTGGAAAAGTAGGACACcgccgaa
1473150..1473250	104	448	rrs(+1303)	rrs	towards the end	GTCTGCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGC AGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGTACA CACCGCCCGTCACGTcat
1475399..1475499	140	894	rrl(+1722)	rrl	towards the middle	cgataggcaaatccgtcgCTCACTAATCCTGAGAGGTGACGCATAGCCGG TTGAGGCGAATTCGGTGATCCTCTGCTGCCAAGAAAAGCCTCT AGCGAGCACACACACGGCCCGTACCCcaaaccgacacaggtggtcag
1472808..1472907	243	6792	rrs(+891)	rrs	towards the middle	caaggctaaaactcaaaggaattgacgggggcccgcacaagcggcggagcatgtggattaattcgat gcaACGCGAAGAACCTTACCTGGGTTTGACATGCACAGGACGCG TCTAGAGATAGGCGTTCCCTTGTGGCCTGTGTGCAGGTGGTGCA TGGCTGTCGTCAGCTcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccctt gtctcatgttgccagcacgtaatggtg

APPENDIX E

Appendix E: rRNA fragments identified in 100 ng CF RNA sample. Sequence reads from *M. tuberculosis* CF RNA 100 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, gene name given by NextGENe, gene, location on the gene, and the consensus sequence for each fragment.

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1471780..1471814	35	5	rrf(0)	Intergenic	between rrs and murA	ttacgGCGGCCACAGCGGCAGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGCCTGCCAGCGCCGATGATACTGCCCCTCCGGGTGGAAAAGTAGGACACcgccgaac
1472521..1472568	48	6	rrl(+2397)	rrs	close to the middle	CGTTGTTGAAATACCACTCTGATCGTATTGGGCATCTAACCTCGAACCCTGAATCGGGTTTAGGGACAGTGCCTGGCGGGTAGTTTAACTGGGGc
1474276..1474314	39	7	rrl(-75)	rrl	close to the beginning	CTTGGTGGTGGGGTGTGGTGTTTGAGAACTGGATAGTGGTt
1476335..1476390	56	7	rrl(+2744)	rrl	towards the end	CGGCTCGTCGCATCCTGGGGCTGGAGCAGGTCCCAAGGGTTGGGc
1474435..1474491	57	7	rrl(+2818)	rrl	towards the middle	GGGTTTAGAACGTCGTGAGACAGTTCGGTCTCTATCCGCCGCGCGCGTCAGAAACTt
1473030..1473130	108	11	rrs(+1181)	rrs	towards the end	gaCGTCAAGTCATCATGCCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGTTAAGCGAATCCTTAAAAGCCggtct
1475262..1475361	203	13	rrs(+675)	rrl	towards the middle	GTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGc

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1474501..1474601	102	18	rrs(-32)	rrl	towards the middle	GGCCGTTTGTTTTGTCAGGATATTTCTAAATACCt
1474108..1474158	51	22	rrl(+2677)	rrl	towards the beginning	CCCCGGGGATAACAGGCTGATCTTCCCCAAGAGTCCATATCGACG GGATGGTTTGg
1475077..1475143	67	22	rrl(+2969)	rrl	middle	GTCAGGATAACCGCTGAAAGCATCTAAGCGGGAAACCTTCTCCA AGATCAGGTTTCTCACCCACTTGGTGGGATAAGGCCCCCCGCAG Aa
1472278..1472321	44	23	rrl(+1553)	rrs	towards the middle	gacgaatcagcggtactaaccacccaaaaccggatcgatcactcccctteGGGGGTGTGGAG TTCTGGGGCTGCGTGGGAACTTCGCTGGTAGTAGTCAAGCGAAG GGGTGACGCAGGAAGGTAGCCGTACCAGTCAGTGGTAACACTG Gggcaagccggtagggagagcgataggcaaatccgtcgctcactaatcctgaga
1473454..1473497	44	26	rrl(+1776)	intergenic	intergenic between rrs and rrl. Closer to rrs.	GGCGAATTCGGTGATCCTCTGCTGCCAAGAAAAGCCTCTAGCGA GCACACACACGGCCCGTACCCCAAACCGACACAGGTGGTCAGg
1472041..1472121	81	26	rrl(+618)	rrs	close to the beginning	CCGTCAGAGCCTCCTTTTCCTCTCCGGAGGAGGGTGGTg
1476276..1476310	35	29	rrl(-161)	rrl	close to the end	AGCCGGGTGCATGACAACAAAGTTGGCCACCAACACACTGTTGg
1476476..1476532	57	30	rrl(+2880)	rrl	towards the end	AACCTGTCCCTAGTACGAGAGGACCGGGACGGACGAACCTCt
1476538..1476579	42	32	rrl(+1419)	rrl	close to the end	CTTGCCCGCCGAAAGACCAAGGGTTCCTGGGCCAGGCCAGTCCG CCCAGGGTGAGTCGGGACCTAAg
1473754..1473803	50	36	rrl(+2618)	rrl	close to the beginning	CGGGATTAGTGATCCGGCACCCCCGAGTGGAAGGg
1474171..1474242	72	39	rrl(+450)	rrl	towards the beginning	CAATTCCCGAGTAGCAGCGGGCCCGTGGAATCCGCTGTGAATCC GCCGGGa
1476402..1476446	45	41	rrl(+2201)	rrl	close to the end	gcacgaatggcgtaacgacttctcaactgtctcaaccatagacTCGGCGAAATTGCACTAC GAGTAAAGATGCTCGTTACGCGCGGCAGGACGAAAAGACCCCG GGACCTTCACTACAACTTGGTATTGATGTTCGGTACGGTttgtgtaggat aggtgggagactgtgaaacctcgacgccagttggg
1476729..1476797	69	41	rrl(+3071)	rrl	end	AGGTCAGACCTGGAAGCTCAGTAATGGGTGTAGGGAACTGGTGC TAACCGGCCGAAAACTTACAACACc
1472168..1472248	81	42	rrl(+719)	rrs	towards the beginning	GGGGTAGCCGCAGCGAAAGCGAGTCTGAATAGGGCGACCCACA CGCGCATa
1473929..1473973	45	60	rrl(+1898)	rrl	close to the beginning	aaggaactcGGCAAAATGCCCCCGTAACTTCGGGAGAAGGGGGACCG GAATATCGTGAACACCCTTGCGGTGGGAGCGGGATCCGGTCGCA GAAACCAGTGAGGAGCGACtgtttactaaaa
1476055..1476149	95	74	rrl(+96)	rrl	towards the end	GCGTGGATCCGAGGATTTCCGAATGGGGAAACCCAGCACGAGTG ATGTCg

Transcript Site	Length	Coverage	Gene Distance	Gene	Location on Gene	Sequence
1474377..1474427	51	75	rrl(+271)	rrl	towards the middle	AAACCGCACGCATGGGTAACCGGGTAGGGGTTGTGTGTGCGGGGt
1473543..1473583	41	77	rrl(+1102)	intergenic	intergenic between rrs and rrl	aagggaaacagcccagatcgccggctaaggcccccaagcGTGTGCTAAGTGGGAAAGGATGTGCAGTCGCAAAGACAACCAGGAGGTTGGCTTAGAAGCAGCCACCCTTGAAAGAGTGCGTAATAGCTCACTGGTCAAGTgattgtgcgccgataatgtagcggggctcaagcacaccgccg
1474800..1474900	182	85	rrs(+432)	rrl	towards the middle	CTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGGATTGACGg
1471913..1472012	113	94	rrs(+1413)	rrs	close to the beginning	CGGTAACACCCGAAGCCAGTGGCCTAACCCTCGGGAGGGAGCTGTCGAAGGTGGGATCGGCGATTGGGACGAAg
1476627..1476716	90	134	rrl(+843)	rrl	close to the end	AAGACCGCGTGGAGGCCCGAACCCACTTAGGTTGAAGACTGAGGGGATGAGCTGTGGGTAGGGGTGAAAGGCCAATCAAACTCCGTGATAGCTGGTTCTCcc
1473145..1473244	111	158	rrs(+1294)	rrs	towards the end	cggaTCGGGGTCTGCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGtcacgtc
1475903..1476002	189	189	rrs(+61)	rrl	towards the end	aagtcGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTCGGGATAAGCCTGGGAAACTGGGTCTaataccgg
1475566..1475666	122	205	rrs(+322)	rrl	towards the middle	GAGATACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCg
1476905..1477005	114	270	rrs(+195)	rrf	nearly covers the whole gene	CTTGTGGTGGAAAGCGCTTTAGCGGTGTGGGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGACGGCCTACCAAGGc
1473259..1473332	74	409	rrl(+513)	rrs	very close to the end	CCTAAATACTCCTCGATGACCGATAGCGGATTAGTACCGTGAGGGAATGGTGAAAAGTACCCCGGGAGGGGa
1475434..1475520	87	1303	rrl(+777)	rrl	towards the middle	GTGAATAGTGGCGTGTTCTGGACCCGAAGCGGAGTGATCTACCCATGGCCAGGGTGa
1472752..1472851	443	3562	rrs(+735)	rrs	towards the middle	agtaactgacgctgaggagcgaaagcgtggggagcgaacaggattagataccctggtagtccacgccgtaaacggtgggtactaggtgtgggtttccttccttgggatccgtgccgtagctaacgcattaagtaccccgcctggggagtacggccgcaaggctaaaactcAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCGAAGAACCTTACCTGGGTTTGACATGCACAGGACGCGTCTagagataggcgttcccttgtggcctgtgtgcaggtggtgcatggctgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgcaaccttgtctcatgttgccagcacgtaatggtggggactcgtgagagactgccggggtcaactcggaggaaggtggg

APPENDIX F

Appendix F: sRNA fragments identified in 50 ng CF RNA sample with consensus sequence. Sequence reads from *M. tuberculosis* CF RNA 50 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, gene name given by NextGENe, gene and the consensus sequence for each fragment.

Chromosome Region	Length	Coverage	Gene Distance	Sequence
347393..347426	34	5	PE5(-2198)	CTGCCGCCCATCCCGACACTGTTCGTGGTCGCCg
352230..352281	52	5	(+81)	CCGTGGGTGCTGGCGATCACGATGCCCTATAGTGATGCCGCCCAGCGGGGTg
378209..378243	35	5	PPE6(-2498)	TCCGATGGGGGTTTACAGCCTGGACTCCGCTTTTg
409681..409723	43	5	iniB(+319)	TGGGTGCTGGGTTCCTGGGTCAGGGTGGCCTGGCTCTCGCCGc
989639..989673	35	5	citA(+188)	TCGGCAGCGGGAACGGTTCAGCCGGCGGCAGCCCg
1466925..1466959	35	5	atpD(+1084)	CCAGCTCGACCATCCTGGACCCCAGCGTTGTCGGg
1683101..1683134	34	5	mutA(+944)	CCAGCGTGATCCGTGGGTGAACATGCTGCGCTGc
2277851..2277888	38	5	(+598)	CTGCCGGAAGTCCCAAAACGACTCACCGACGGCCAGGa
2278147..2278182	36	5	(+304)	CGGCGATGTCGCGCAGTTGCTGCGGTGTGATCCGCa
2502565..2502599	35	5	rnpB(-1814)	GTCCAAACCGGCTATCAGGCCGCGGTATTCGGCGa
2626202..2626236	35	5	esxP(+283)	ATTGTGCTGCAGCGGCTGACGTTAGCTGCTGAGGa
2692351..2692385	35	5	PE_PGRS41(-414)	CCAAGCAGCTCCCTGCCGAACCAGCAGACGACGAc

Chromosome Region	Length	Coverage	Gene Distance	Sequence
2951530..2951592	63	5	(-223)	CGGGCGGTCGGCACGCTGTGCAGCGGCAGCGCAATGTCACCTACGCTGGTAG TGCTGCGCCGg
3023949..3023988	40	5	ideR(+384)	GGCAACCCGATCCCGGGCCTGGTGGAACTTGGCGTGGGCc
3495854..3495897	44	5	tgs1(+469)	CGACTGCATGGTCGTGTGGATGTTGCTCGCGAAGCTGTCGCTCa
3497036..3497070	35	5	(+521)	TGTGGCGCAGGCTGTCTTGGACCACGTGACCGACg
3606901..3606935	35	5	None	CGGCGGATGTAGCGGGCATCGCGTTCGCCGCGCGa
3826273..3826307	35	5	None	GTGGAGGCAATCGGAATCCGAGAACTAAGACAGCa
4079769..4079807	39	5	None	GCGTGCTTGGCGGTACACACCGACGATCTCGCCGACGGc
4100860..4100947	88	5	None	TACAGCTAGAAGCGTCGCAAGATCGCCGAGGCCACCCACGCAACCCCAGGAG TGCACGCTTGGTAACCGAGAACCGTGTTGGTGGGCg
353331..353365	35	6	mycP3(-1133)	CGACGGTGATCTGTTGGTGTTGCAGCCGGTGCCCg
619032..619065	34	6	(+727)	ACTATCAGGCCGACCCGGCCACTGCTGACCTGAt
918303..918341	39	6	desA1(+409)	GGCCCTGGTGGTTTTGGCCTGGGCTGAAGCCCCGGTTGa
1412992..1413025	34	6	(+1098)	GGTGACGTCAGGCTTTTTCGAGTCCGGCGAGGGg
1987052..1987085	34	6	plcD(+611)	GTTGAACGTTGACGTCATGTCGCCGGTCACACTg
2319295..2319328	34	6	cobN(+1425)	CGTGGTAGATCGCCACCGGGTTCTCCCCGAAGCc
2483389..2483425	37	6	dlaT(+1424)	CGACACCCCGATCCTGGTTCCGCCGCAGGCCGCCATg
2949955..2950022	68	6	TB31.7(+362)	TGGGCGGTGGCCGGGCCGGCTGCTCGGTTCGGTCAGTTCCGGCCTGCTCCGCC ACGCGCACTGTCCGg
2955616..2955660	45	6	(-4309)	TTTGAGTTATGAAACCCGTGGGCACATACCCTTCCGCGTCGTACt
3499671..3499705	35	6	devR(+210)	GTGAGGATCAGACAGCGCAGATCGGGCATGCGGGa
3705429..3705463	35	6	sdhA(+429)	CTGCAGACGCTGTATCAGAACTGCGTCAAGCACGa
4314343..4314391	49	6	bfrB(+165)	CGAAACCATGCAATGATGCTCGTGCAACACCTGCTCGACCGCGACCTTc
244740..244774	35	7	mmpL3(+2544)	GTTGGTGGCGGCGTGGATGCTGGCGACTGCACGGa
345471..345505	35	7	PE5(-4119)	GAGGCCCTTGGCCTGAACCCGCCCCCGGTCCCCAt
420761..420804	44	7	dnaK(+926)	CGTTGTGCTCGTGGGTGGTTCGACCCGGATGCCCGCGGTGACCg
1080848..1080906	59	7	ctpV(+2171)	CGCGATCGAACTCTCCAGGCAGACCCTGCGCACCATCTACCAGAATCTCGGCT GGGCCt
1391026..1391060	35	7	kgd(+1992)	ATGGTGCCGCCGACGCGGTAGCCCGGCAGATTCGc
2955455..2955489	35	7	(-4148)	TGCCGATCCGGTGGATGACTTTGGTCCCTATGCCt
274730..274769	40	8	None	GGTCCTCTGGCGGTGCGGCGGCCAGCGTGTCGAGAGCGGc

Chromosome Region	Length	Coverage	Gene Distance	Sequence
1220452..1220486	35	8	glyA(-52)	TGCGGTGCAGACGTGCAACGGTGGATGGCGTGTGc
1338417..1338451	35	8	(+62)	CCAGTTCCTCGATCCACTCCGGGCTGGGATTGAGg
1797585..1797619	35	8	nadC(+197)	CCTGGGCACCAACGGTTATCGGGTGCTCGACCGCg
1824736..1824770	35	8	cydA(+1117)	CGGGGCGGGCATGGTTAGCAGCGCCAGCCAGGAGa
1965776..1965820	45	8	(+119)	GGGTGTTGGCCTGGCAAGGCTCAATCCGGCCGACCGCAACGTCCc
1965823..1965861	39	8	(+166)	AGATCGGCGATGAACTCTCGGTCGCCCGAGCCTTGTCCg
2250949..2251015	67	8	otsB1(-987)	GTGCCGTCGTTGGTCGGTGAGTCCATGGCGGTAGCCAAGCCAAGTAGTCACGACTGCCGTGCCACGa
2274682..2274716	35	8	None	CCACCATGTGTACCGATCGACCCAGCCCAGCCAGa
2374240..2374279	40	8	lppK(-2292)	AGGGACAGCGCTGGGTGATGCAGACGTGGAGTCAAGTGGt
2782771..2782805	35	8	(+1237)	GGCCCGTGACGGCACTTCGGTTTGACCGACCTGGa
3046019..3046053	35	8	(-4558)	GCATAGGCAGATTCGCGGGAGCTGGGCAGCGACGa
3414771..3414804	34	8	nrdH(+154)	CCCGACCAGTGGTCGTTTCCGGCGACCACGACGg
1169366..1169429	64	9	(-3931)	GGTCATCGCGTGATCCTTCGAAACGACCAGCAAAAGTCAATCGAAGGAAATGACGCAATGACCt
1340943..1340977	35	9	esxK(+284)	CCTCAGCAGCTAACGTCAGCCGCTGCAGCACAATa
2314134..2314169	36	9	rpsR(+184)	GCAGCAGCGCCATCTCGCGGGCGTTCTTGATCGCCt
3326014..3326063	50	9	(+80)	CGACGCGTTCGATTTCCCTCGCTCGATGACGGAACCCGGCTTGGTCCGAa
3751654..3751688	35	9	PPE55(+1496)	TGGTGATGGGGCCGACGCTGATAGGTATGGCCGTa
851642..851676	35	10	phoP(+34)	CCCCAGGCGAAAACACCACACCGGAGGCTCGTGTc
1581101..1581136	36	10	(-679)	CCGTTACGTGCTGGGACGACCCGACACAGGCCGGAt
4056124..4056158	35	10	ephA(-1575)	GAGCTGACGATCGAGGTCTGCCAGTTCCTGGAAAa
15823..15857	35	11	pknB(+1613)	CTGAATCGACCGGAACTGTGGTGCCTGCGGGTGGa
88190..88244	55	11	hycD(-4084)	CGAAAGCCGAGCAGGTGGAACCGAAACGCAGTCGCCTCGTCGTATGTGCACCCGa
1166640..1166674	35	11	(-1205)	AGGCCGAGGGTGCTCAGGTATCGGGAGAAATCTGc
2219456..2219490	35	11	None	GTTGGTAAATCGGGTGCTCAGATCGGCCCTCCGGa
3826433..3826467	35	11	None	CAGGTGTCCTGATTCCGGCTCGTCGTCCACAAAAc
474296..474330	35	12	ndhA(-1657)	CCCGGTGCCGTTGCCGTGCCAAACGGTTGTTGTCg
806142..806176	35	12	rpsQ(+386)	CGAGATCCTCGAGAAGGCTAAGTAAGCCTGACGAg
1962197..1962230	34	12	narX(+1956)	GTCCTGTCCGCTCGCCGAAGACGCACTCCCGTCa

Chromosome Region	Length	Coverage	Gene Distance	Sequence
1964687..1964721	35	12	narK2(+649)	GTAGTTGCTGAACGCCACGAACCCGCCGAACACGa
2517341..2517396	56	12	fabD(+554)	TGCACTGGGTGTCGCCGGAGCGTTCCACACCGAGTTCATGGCGCCCGCACTTGACg
3496799..3496837	39	12	(+284)	AAACCGTTTCCCCGATCCCAAGGACCGCTGCCATCTGGc
3497203..3497248	46	12	(+688)	CTCAGCCGTCCGACGTCTTACCCGCTGACGACGGCGCCGCGATCCt
3666695..3666729	35	13	accA3(+338)	CGGCCCCAGCCCGCAGTCGATCCGCGACCTGGGCg
68848..68882	35	14	None	GTGCTTGCTGCCCTACTGCTGGCGTATCGCTCGCg
2949865..2949899	35	14	TB31.7(+272)	CCACAGTGAAATCGTTCCGGCGGCAGCCGTTCCCa
3468156..3468242	87	14	ssr(+177)	AGGGTCCGGGCCGAGGGCGTTCCCGGTCTGACAGACTAGCCGTCGCTTAGGCAGCGAGAGCGTAGTCGCGCTGATGTGAATCGGCGc
3415123..3415167	45	16	nrdH(-165)	CCCCGTGGATCTGGTGAGACCTGGTGAACCTGGGATCTGCCGGTa
4168234..4168282	49	16	(-106)	GCCTTCCAGCCCTGGGGGAGTTCACAGGATAGACGCCGCGCGGGGTCCa
2238981..2239015	35	17	ctpF(-1144)	ACACCCATGCAGAGGGGACGCCAATGTCAGCCCAa
164390..164424	35	18	cyp138(+1024)	CGATCATTATCAATATCGCCCAGATACATGGCGAt
1187430..1187464	35	21	0	CTCCCGTGGTTATGCCTCTTGTCACGCCAACCACc
3468051..3468090	40	21	ssr(+329)	GATCTCCCGGTCACGCGAACTAGCCGAGATGACGATCCCa
4262658..4262692	35	21	fadD32(+374)	ACGGGGCACAGTCGTCGAGCACCGCGTGTAACCGa
2767686..2767719	34	22	(-21)	GGGCACCGGCGGCGCTTCGTGGGTCCGTTCGTAc
80258..80292	35	24	glyA(-1362)	GGCGGCCACGGGAAACGGACCCGCAGCAACGCGGg
1275613..1275673	61	27	None	GGCGGTGACCGCGGCATTGTTTTCAGTTTGTGACAACTTCTCAATATGCCGCGGTCGCCGc
2040858..2040894	37	34	PPE28(+1405)	ACGGCGACCCGGCCTATGGTTATTCGACCTCGCCGCc
1960669..1960781	113	37	(-878)	cggggAAACCCGGTGATCTGCCCGAAGTGCTGGGCGATTGAGCGGGTATGTACACCCGGTTTGACCTACCGTCCCAAGACGGGGCTACCGCCTTCGGGCAGATCCtcatcctg
1993213..1993246	34	41	(+60)	GGGTTGGGGATTTCTGAACCGATCAGTCCCGGGt
3649468..3649502	35	48	ctpC(-1024)	GCCACCGGTGAGCGCATCGAGCTATCCGCGGTGAc
754758..754793	36	54	atsD(-1344)	CGTCGGAATGCGGTGCGTGGCAACGTATGAATGTTg
4099438..4099475	38	58	(-4524)	CCGGGGGGTCGGACTGATACACCCTCGGCTATGGCCGa
2404472..2404549	78	63	(-2750)	ACCAAGGCCAGGAGCCCAATCAGGGCCAAGACCAGCGCAATGATCAGCATCGGTCATCCTCCAACCGGCTAGCAGCGa
4352941..4352979	39	91	(-31)	GAGTTCGACCTTCCGTCGGTCTCGCCCTTTCTCGTGTTt

Chromosome Region	Length	Coverage	Gene Distance	Sequence
2844174..2844208	35	119	fas(+5124)	CGCTCGGCGTTGAGCACTTCCACTGTGCTGTGGGc
1342600..1342663	64	338	0	GGTCATTGCGTCATTTCCTTCGATTGACTTTTGCTGGTCGTTTCGAAGGATCACGCGATGACCg
1469500..1469574	75	357	murA(-747)	GCGCACGTCGTTGTCGGCTCCTTGTGTTTCTGATCCTTGACAAGCCAGAAACCTTAAGCCACAACGACGTGCGCc
4252929..4252971	43	431	(-22)	ACGTCGTTGTGGCTTAAGGTTTCTGGCTTGTCAAGGATCAGAa

APPENDIX G

Appendix H: sRNA fragments identified in 100 ng CF RNA sample with consensus sequence. Sequence reads from *M. tuberculosis* CF RNA 100 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, gene name given by NextGENe, gene, and the consensus sequence for each fragment.

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
303550..303583	34	5	argC	nirB(+684)	ATCGAGTCGGTGAAGCATTCGGATGGCTCGGTGt
337304..337337	34	5	Between Rv3750c and Rv3751	PE_PGRS4(+1736)	CCCGCCGGTGTCCAGGCCAAACCCGCCGCTGCCg
601708..601741	34	5	ctpC	hemA(+1267)	TCTTCGAGCTTGACCAGACCGCCGTCGATGCCGt
1564447..1564480	34	5	fadD32	gmk(+46)	AACCGGCGGCAGTGGGACGTGTGGTGGTGCTGTc
1577190..1577223	34	5	fusA1	(+260)	CGTGGTGGCGATGTGCGTTGCGTCCGCGGGGCTg
2055657..2055690	34	5	gltB	(-298)	CTTCGGCTCGCTGGTTAGCGATTGTTAGTTGCAc
2239323..2239356	34	5	gmk	ctpF(-803)	TGCACGGCCAGATAGTGCCCACGCTGGCCAACAt
2289878..2289911	34	5	hemA	lipT(+1309)	GCACGGCGGTCGGCGGCGGCGGTCAGCAATGCGc
2827612..2827645	34	5	hemH	orn(+455)	CAAGGGGCTGACGCACCGGGCGCTGGCCGACATc
2978147..2978180	34	5	Inter geneic	None	GCTGGCAGTTCCTCGGTCTCAACGATCTTGCCGc
3022467..3022500	34	5	Inter geneic	sigB(+6)	GATGCACCCACAAGGGCCACCACAAGCCGGGTTg
4198611..4198644	34	5	Inter geneic	tyrA(-1711)	CGGGGGTGGAAGTACTGTCCGTTCCTGGACTGCg
4262779..4262812	34	5	kasA	fadD32(+254)	CAGGTTCTGCGGGCACAGGATGGCGACGCGGTCa
4302868..4302901	34	5	lipT	(-79)	GCACACATCGAGGTCAGGATCTCGGTCATATCGc
4334567..4334600	34	5	mbtB	gltB(+1481)	CGACAACACCGCAACTGGAGTGTCGGTGCCCATc
4369747..4369780	34	5	mmpL5	mycP2(+390)	CAATGATCGAACTCACCACCGTGCCGTGCGCGTc

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
4398690..4398723	34	5	mycP2	sigM(-1463)	GCTGGTTCCCGGCGCCCGCATCGCTAACGGCCGc
119170..119204	35	5	nadC	(+1456)	CGCACCCGATCACCGCATTCGTCCTCTTTGTGGCc
194242..194276	35	5	ndhA	(+662)	CAGCTGCCGATGACTGTTCACCGCGTCCGCGCCCc
269454..269488	35	5	nirB	(-1690)	CGACGATGTGACCAAACACCATGTGCTGCAAAGCt
274730..274764	35	5	orn	None	GGTCCTCTGGCGGTGCGGCGGCCAGCGTGTCGAGa
348462..348496	35	5	pckA	PE5(-1128)	TCAACACCCGTAATCCGTTGCTGGCCAGGGTAACc
474296..474330	35	5	PE_PGRS4	ndhA(-1657)	CCCGGTGCCGTTGCCGTGCCAAACGGTTGTTGTCg
782891..782925	35	5	pepD	fusA1(+406)	TCGTCAACAAGATGGACAAGATCGGTGCGGACTTc
873531..873565	35	5	pks3	hemH(+188)	CGTCGATGCCCCTAACCATCTGGCCGGGCCGCCGg
884318..884352	35	5	radE14	(-268)	GCGGGCGGCGGTGGGTGCGTCGATGGTGGCGTCCa
1099314..1099348	35	5	rnpB	pepD(+248)	TCCGACCATGACGCCCCCTCCTGGGATGGTTCGCc
1511851..1511885	35	5	Rv0102	fadE14(+1005)	CTTGTCGACGAAACGACGCTTGGCAAATGGTGGCg
1736341..1736375	35	5	Rv0165c	ileS(-144)	AGGCCCAGTAGAACCGAACGGGTTGGCCCGTCACa
1866320..1866354	35	5	Rv0225	argC(+744)	GGCATCCTGGCCACCTGCACGGCACGCACCCGATc
1901178..1901212	35	5	Rv0284	dsbF(-536)	GCCTACTCGATTGGCCCGCGGTGCGACCGTGCCGa
2032991..2033025	35	5	Rv0790c	mycP5(-704)	TGGGGATTTACACCACAATCGTCATCATCGGTGCg
2056276..2056310	35	5	Rv1349	erg3(-211)	GGCTGCCCACCCACGGTGCTCGGTTCGGTCCGGGa
2477897..2477931	35	5	Rv1401	(+707)	CGACGCCGCCGTCACCGACGTGGTGCATGCCGACg
2500503..2500537	35	5	Rv1676	rnpB(+214)	CCGCAGGCTCGGGCGAGCAACCCTCAAGCGCCTGc
2502565..2502599	35	5	Rv1795	rnpB(-1814)	GTCCAAACCGGCTATCAGGCCGCGGTATTCGGCGa
2519178..2519212	35	5	Rv1813c	kasA(+1063)	CGGTGCTCACGGTGCTGACGCTGCGCGACGGCGTc
2672020..2672054	35	5	Rv1996	mbtB(+3783)	CGAGCGAACTCGAAGGCCACTATCGCACCCATGCa
2783947..2783981	35	5	Rv2144c	(+61)	GGCGCCCGGATAGAAACTCAACGTCACGTCGTCGa
2819339..2819373	35	5	Rv2212	scoA(+497)	CCCCCTGCCAGGCATGCACCAGTGCGAAGTCGGTc
3020895..3020929	35	5	Rv2228c	sigB(-1532)	CGCTGGGCTTCTTGATCTCGCTGTGGGCAGGATCg
3651551..3651585	35	5	Rv2477c	ctpC(+1025)	CTCGGCCATCGCGTTGCTGATCACCGGCGACGTGc
3900311..3900345	35	5	Rv2626c	cpsA(-2733)	ATCAACGCACCAACGTGCATCATCGTGGTCGACGa
3953877..3953911	35	5	Rv2627c	(+446)	CGAGATGCTCGCCGAACGCTACCGGGGAAGCCGCg
4221473..4221507	35	5	Rv2655c	(+384)	CTGCTCGATGCCGCCCTCGCCAAACACGCCACAGc
4231808..4231842	35	5	Rv2707	None	CGTGCTCGCCGCCGCATCGCGTCGATTGCTGCCCt
2404447..2404484	38	5	Rv3480c	(-2725)	CTGGTTGCTGGTGACCACCGCGAACACCAAGGCCAGGa
1797489..1797527	39	5	Rv3517	nadC(+101)	GGTGCCTGCCAGTGCGACGACCACCGCATCGCTGGTGAc
2500383..2500425	43	5	Rv3776	(-4984)	GGGCGGACGAGTTGGCCTGTAAGCCGGATTCTGTTCCGCGCCg

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
2952672..2952716	45	5	Rv3785	(-1365)	GCTCTGAGATGACCGGAACACGGCGGACCTGATGTTC TTCCATGa
2953710..2953754	45	5	Rv3828c	(-2403)	CGTAGATAAAGCGGCCCGGCATGGGCACCAGTGGCGT AAGTGAGa
1516874..1516919	46	5	Rv3910	fabG(-572)	CTGTTCCACGGGACGATCCGCGACAACGTGTTCGCTG CAGACCCGg
1314313..1314361	49	5	Rv4006	pks3(+588)	AGCGATCTCGCTCTTGCCGGTGGTGTGGTTGTCACGC TAGAACCGCGGa
251787..251840	54	5	scoA	pckA(+5)	CTCAGCGACCATCCCCGGTCTGGATACCGCGCCGACG AATCACCAGGGGTTGCt
776008..776071	64	5	sigB	mmpL5(+2409)	GTTGCCAGATCAGCACCGACAGCCCGAACGAGGCGC CCAGCGACAACACCACGGTGCCGACGAt
88211..88244	34	6	accA3	hycD(-4084)	CGAAACGCAGTCGCCTCGTCGTATGTGCACCCGa
215774..215807	34	6	aceAb	bglS(-462)	GTTCGTCCGGACCTTGTTCGACCGCGTCACCGCa
326352..326385	34	6	ctaD	fadE6(+2076)	GGCTGGCTAGCAGGTACTCGCCCCACTTGCCGGg
352444..352477	34	6	ctpG	(+295)	AGCTGCTACGTGGCATCGTCGCGCTGCGCACCGg
420777..420810	34	6	dlaT	dnaK(+942)	GGTTCGACCCGGATGCCCGCGGTGACCGATCTGg
631410..631443	34	6	dnaK	(-1679)	CACGCCGGTGACCACGCCACCAACGACCGTCGCc
2162128..2162161	34	6	fadE6	aceAb(+562)	GCTTGGTCGAATCTCCCTATCGCACACCGCAAAc
2276245..2276278	34	6	fas	pfkB(+146)	GACATGCACAATGCGGGCGACATTGATACCGCCg
2593248..2593281	34	6	Inter geneic	rocE(+872)	GATGGCCAGCCCGGCGGTTTCCTGCTCTGCGAAg
3497349..3497382	34	6	Inter geneic	(+834)	TGCCCGATCACCGAACCGCTGGAGATCGCCAAGa
3516906..3516939	34	6	lpqB	nuoG(+160)	TGGAGCCCGTCGGCGCCTGCCGGCAATGCCTGGt
3894312..3894345	34	6	narX	PE31(+219)	CAGTTTGTGACCACGCTGGCCACCAGCGCTAGTt
595932..595966	35	6	nuoG	serB1(+619)	TTCGCGGGTGCCGTCCCAGATCTTGTCGGCGATGa
707305..707339	35	6	otsB1	galTa(-4197)	CACGCGATCCATCCGCGGTTCGCTGTTGGTTTCCa
1004881..1004915	35	6	PE31	prrB(+926)	AGATCCGGGCCGCGTCGTGGGCGGCGCGGTCGAGc
1669140..1669174	35	6	pfkB	moxR1(-109)	ACTGAGCCAGGTGTGATTTGCCGGGCACCACCGCg
2150873..2150907	35	6	pks13	nanT(-599)	CGGCGTGATGACGCTGACCTCCCGCCCTCCAATAc
2252256..2252290	35	6	prrB	otsB1(+254)	GGCCGCCCGCGGAATCAGGCTGCCGCCGGGCTCCc
2445035..2445069	35	6	rocE	(-2718)	CCTGCGCGGAGTCGGCATTGGTGCGGTCGATGGGa
2483389..2483423	35	6	Rv0079	dlaT(+1424)	CGACACCCCGATCCTGGTTCCGCCGCAGGCCGCCa
2842511..2842545	35	6	Rv0185	fas(+6787)	CGAGCCGGGCAGCACCACGTGCAGCCGCGACGCGa
3404859..3404893	35	6	Rv0284	ctaD(+28)	GGTGCGGGCCGGGTAAGGACGAATGGCTTCGAGTt
3492964..3492998	35	6	Rv0289	(-842)	CCAGTGACAGTCGTGACATCGTGCGGGGCCTGACg
3623406..3623440	35	6	Rv0538	lpqB(+1470)	TCGTCACCGGTTCGCCAGGACAACGACACCACCGa
3666695..3666729	35	6	Rv0613c	accA3(+338)	CGGCCCCAGCCCGCAGTCGATCCGCGACCTGGGCg

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
4222194..4222228	35	6	Rv1478 and intergenic	(+1105)	TCTGCGACATCGACCACACACTGCCCTACCCACTc
4258521..4258555	35	6	Rv2182c	pks13(+2591)	TCTGGGTGGGCGCGGCGTAGACACACACCTCCAGa
346439..346476	38	6	Rv2625c	PE5(-3148)	ATTCCCGGCCGCCTCGATGCGCTGGGCCCGGCCCGCAa
4056007..4056044	38	6	Rv2777c	ephA(-1689)	CGGGATGTAGGTCAGGTCCACAGCCACCGGGCGCACGa
1963271..1963312	42	6	Rv3127	narX(+874)	TCAGCGGATCGGCACGGCGCAGCTCAGAGACCGTGCGCCCCa
2239872..2239913	42	6	Rv3127	ctpF(-246)	GCTCAGTCAGCAGAGCGGTGGTCAATTCCGGTCAGGCTCCGg
1313412..1313455	44	6	Rv3131	pks3(-270)	ATCACACGCGCGGTGCATGCTGCTGTGGCTGTCGAGCAGTGTTg
3084849..3084898	50	6	Rv3616c	ald(-1922)	AGCATGTCCAGCATCACGGGGTCTGCACCGACGGCATCGAGCTGCGGGGa
3493021..3493086	66	6	Rv3776	(-899)	CCTTGATCCGGGTAGGGATAGCCCCGCCGTTGGCAGCAGTTCCCGCCCCCACACCACGGCGGCCGc
2952398..2952466	69	6	serB1	(-1091)	ACATGGTCGCCAGACTCCAGGTGAACAACCACAGGATCACCAACACGCTCCAGTGGACGTTCACCACAa
662451..662484	34	7	acrA1	nrdZ(+1156)	AGTCATGTAATCTCGGCTCGATCAACCTCGCCCg
2073625..2073658	34	7	desA2	(+544)	CGCGGCATCTACGCGCCTTCCGCTCCGCGGCCGa
7449..7483	35	7	fumC	gyrA(+147)	CCCGTGCATCGCCGGGTGCTCTATGCAATGTTCGa
25117..25151	35	7	gyrA	TB39.8(+293)	GCTCGAATCGGACGACCACATCACCATACGTTTGc
349084..349118	35	7	Inter geneic	PE5(-506)	CCCCGACAACGAGTACACCGCCAACATCGATCGGa
867824..867858	35	7	ispF	(+493)	GGTGTCCAGCCGCTGGGTGTCCAATAGCGGAATGa
1222477..1222511	35	7	nrdZ	desA2(+518)	CGAAGAGCCCATCCTGGCCGGACTCATCGACCGCa
1227126..1227160	35	7	PPE3	fumC(+405)	GTGGCCGTCGGGAAGGTGTCGTTGGACGACTGCGa
2017524..2017558	35	7	Rv0284	malQ(-48)	CTGAGCCTTCCGTCAGCACAGCACGGTTGGCTACc
2278025..2278059	35	7	Rv0774c	(+427)	CATGCTTGCGCCGGTGGCCAAACCGTCATCGACGa
3806205..3806239	35	7	Rv0823c	acrA1(+584)	GTTCACCCCGATGCTGCTGCCGGACATTGGGCGCa
4346889..4346923	35	7	Rv0996	PE35(-3822)	CTGTTGTCGATCGTCGGCACCCAACGGCAGTGGTc
1113102..1113140	39	7	Rv1828	(+718)	TGATGGCGGTGATCCTGGTTGGCTCTGCCGCCGCGGCCt
917498..917537	40	7	Rv2030c	(-1187)	CCGACCTTCGACTGTTCCAGCTGACGACACAGCGCCCGGa
3468268..3468321	54	7	Rv3870	ssr(+98)	GTGGTCTCTTGCCTGCACCGGCACGCTTCCCTTGATTCGATGCGCGAAGTCGAa

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
4024199..4024262	64	7	ssr	ispF(+85)	TCGCACAGCGCATGAACGGCCACGTCACCGTCGGAG TGGCCCGCGCAACCGTCGGCGCTCGGGa
340528..340608	81	7	TB39.8	PPE3(+1164)	CCCTACGCCATCGCTCCGCCCGGCATCGGGTTCGGCT CGGGGATGAGCGCCAGCGCCAGCGCTCAACGCAAGG CACCACAg
1796908..1796941	34	8	devR	nadB(+1103)	CGGGTTGCACGGCGCCAACCGCCTGGCCTCCAAc
1845595..1845628	34	8	devS	uvrA(+1854)	CGTTCCGTCGACCCCCGTCGTCAACTCACCGTCg
2484655..2484688	34	8	hspX	lipB(+71)	CGCGTGGCAGCTACAGCGAGAGCTAGCCGACGCc
2955616..2955649	34	8	hycD	(-4309)	TTTGAGTTATGAAACCCGTGGGCACATACCCTTc
4220050..4220083	34	8	Inter geneic between Rv2628 and Rv2629	lipE(+365)	CTGCGAGTACCTGCCCTCCTACACCAGTCATGGc
88506..88540	35	8	kdpD	hycD(-3788)	CAGACTGGATCGACAGATCGTGCGGGCGTCGGCAc
851642..851676	35	8	lipB	phoP(+34)	CCCCAGGCGAAAACACCACACCGGAGGCTCGTGTc
1150652..1150686	35	8	lipE	kdpD(+1000)	GCCGACGACGAGCTGCGTGGCGTTCATCTCGCGGg
1235415..1235449	35	8	nadB	xseA(+11)	CCACCGCGCGAACCGGGAACGGATTCTCCGCTGAa
1796054..1796088	35	8	nadB	nadB(+249)	CCCGATGCGGTGTACTCGATCGTCGCCGACGGCTa
2979238..2979272	35	8	phoP	clpC2(-4624)	GTCTGCCATGCCGGGCCTTTCGGTCTCGTGAGAGa
3496869..3496903	35	8	pks16	(+354)	CAGGCCGATGTCGCCTTGGCGGCGGCCATACCGCg
3499671..3499705	35	8	Rv0289	devR(+210)	GTGAGGATCAGACAGCGCAGATCGGGCATGCGGGa
3606667..3606701	35	8	Rv0290	None	TAGCGCCAGTGCTTGGACGATCCGCTCATGTCCCa
4029729..4029763	35	8	Rv2657c	mutY(-730)	TAGGTAGCGGGCCACGCGGTTCCAGATCCAGCACg
1132353..1132392	40	8	Rv3131	pks16(+728)	CAAGTACCAGGGCACCATGACCGCGGCGCCCAACTTC GCc
3498580..3498619	40	8	Rv3229c	devS(+646)	CCGGACAACAACTCGGTGGCGATGTCACGGGTGGCC TCGa
353501..353542	42	8	Rv3587c / intergenic	mycP3(-956)	TGGCTACCGGTTTGACGGTGACCTATCGGGTTGCCAC CGGTg
352199..352242	44	8	Rv3879c	(+50)	CATCGCTGAAACCATTGGGGCGGGGACCTTTCCGTGG GTGCTGg
2278606..2278652	47	8	uvrA	hspX(+280)	GTCAGCACCTACCGGCAGCGACACCGTGCGAACGAA GGAACCGTACg
4359090..4359155	66	8	xseA	None	CGTTGGTGTGATCGGAGTTACCGGCGCTCCCGGGATG GGTGTGATTGGGGTTCCCGGGGTGATCGg
1197094..1197127	34	9	cfp2	(+815)	CAGTTATCTACAGAATGAGTAGCGCTCGTTGGCc
3589569..3589602	34	9	esxK	(+175)	CGCTGAAGCAGCTGTGGACCGCCGCCAGCCCAGc
356191..356225	35	9	Inter geneic	(+311)	CGCCCGGTATTTGGACCGATACGGCATCCGCGCCg
1340943..1340977	35	9	Inter geneic	esxK(+284)	CCTCAGCAGCTAACGTCAGCCGCTGCAGCACAATa

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
2103054..2103088	35	9	Inter geneic	ndh(-12)	CCTGACGGGGTCACCTCGATGAGCGAGTTCAGTTa
3246358..3246392	35	9	PPE3	ppsA(+913)	CCCCGGCGGTATTTCGCGGTTTCGACCAAGTCGGc
3548920..3548954	35	9	ppsA	mesT(-3620)	CGCCATATTTCAAGAGCAGGTTGGAGCTTCTGCCc
353322..353365	44	9	Rv0290	mycP3(-1133)	CGGTGTCGTCGACGGTGATCTGTTGGTGTTGCAGCCG GTGCCCg
1965772..1965820	49	9	Rv0292	(+115)	TGGTGGGTGTTGGCCTGGCAAGGCTCAATCCGGCCGA CCGCAACGTCCc
2256537..2256600	64	9	Rv1072	fdxA(-109)	GGGTGCAAGAAAACCGGGGCGGTTCACCCGACCGCC AGCGGGATTCACGCTCCCCCAGGCCATa
2655876..2655940	65	9	Rv1738	cfp2(+175)	GCCCTTGTTCGCAAACGACACGTTGGGATCGGCGAGG CTGTTGAGCAGGCTGGTCAACTGGGCGg
340750..340826	77	9	Rv3212	PPE3(+1386)	CCGCCCGGCGAAGACCCAGTCACATCCACGGTGGCCT CGGATCGGGGTGCCGGACATCTGGGCTTTGCCGGGAC GGc
2150918..2150951	34	10	ctaC	None	AGCACCTAACAGCACACGACGACGGGACTGCAAa
2954155..2954188	34	10	fas	(-2848)	GGCGCGGAACACCGCGAGATCCAACGGGGCCCTg
1413108..1413142	35	10	Inter geneic	None	GGGCAGACCCGGCGTGACTCTCGGGGGGCGTCTGa
2122411..2122445	35	10	Inter geneic	lldD2(+706)	GCCGACGGTGCCAGGCCAGCGATCCAGTGAAGCGa
2463741..2463775	35	10	lldD2	ctaC(+976)	CGCCCGCAGGGCCTCGGCGTTTGTCTTCCCGTCGa
2847337..2847371	35	10	PPE56	fas(+1961)	TGATCCATTGGTCGGTGCCCTGAGTGTCGACGAGc
3758308..3758342	35	10	Rv2627c	PPE56(+8760)	AGTGGGCCGGTGGCGTAGGGCGAACCGTCGGCCGa
32182..32215	34	11	atpD	bioF2(-2080)	CGCTGAGCTTTCGTCTAACACCGCCGAAACCGCc
1998195..1998228	34	11	desA1	PE_PGRS31(-2386)	CACCCCAACCTGATCCAACGACAACCGCCCCTCc
1199075..1199109	35	11	fabD	fadA3(+264)	GCCATCCGGGTGGTCTGCAGCGACGACGAACAGTa
1220452..1220486	35	11	fadA3	glyA(-52)	TGCGGTGCAGACGTGCAACGGTGGATGGCGTGTGc
1466925..1466959	35	11	fas	atpD(+1084)	CCAGCTCGACCATCCTGGACCCCAGCGTTGTCGGg
2517353..2517387	35	11	Inter geneic	fabD(+566)	CGCCGGAGCGTTCCACACCGAGTTCATGGCGCCCg
2845154..2845188	35	11	Inter geneic	fas(+4144)	GTGCACACCGCTGGCGATGATGCTGGTCGGGTTGt
4056753..4056787	35	11	Inter geneic	ephA(-946)	CGCCGGTGGCGCAATCAGGGCTGTGCTAACCCACc
378204..378242	39	11	Rv0029	PPE6(-2493)	GCCACTCCGATGGGGGTTTACAGCCTGGACTCCGCTT Tt
3093770..3093825	56	11	Rv0309	rpsO(-22)	TCGCGATCATGGATATCAGGGCACGGCCACCGCGAAC CGCAGCACGCACCGATGTc
1485598..1485663	66	11	Rv1322A	fadA4(-199)	GCGGATGCCCTGATCGTCGTTGATTTCCTCGTGGACCA GGATCATGCCAAGGTGGTCGTGATACCa

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
918303..918380	78	11	Rv1765c	desA1(+370)	GGCCCTGGTGGTTTTGGCCTGGGCTGAAGCCCCGGTTGACTACCTCGAGGCGAAGTTTCTCCAACTCGACAGGGTCGa
1389211..1389245	35	12	ctpG	(-583)	GATTGCGGACTTCGCCCAGTGGGACAACAGCCATa
2626202..2626236	35	12	Inter geneic	esxP(+283)	ATTGTGCTGCAGCGGCTGACGTTAGCTGCTGAGGa
3606901..3606935	35	12	IS1081-1	None	CGGCGGATGTAGCGGGCATCGCGTTCGCCGCGCGa
2236394..2236438	45	12	Rv1247c	ctpG(+868)	CACGAATGATGCCGTCGGTGGCCAGGCGTTCGCCCGGTTTAACGa
1197139..1197190	52	12	Rv2345	None	GTCCACCAGGCTGACCACTCGCACTTTTGCGTGGTAGACGCAGGATCAACGg
1169366..1169426	61	12	Rv3229c	(-3931)	GGTCATCGCGTGATCCTTCGAAACGACCAGCAAAAGTCAATCGAAGGAAATGACGCAATGa
2102731..2102764	34	13	Inter geneic	ndh(+278)	GAGCAATTCCGAGACGACGCACTGCCCGGCCAGg
1419706..1419740	35	13	lprA	lprA(+8)	TGGCGGTGGCGGCGGCAACAACGGAACAAGGTGGa
2646753..2646787	35	13	ndh	(+301)	CGGGTTACCTGCCCGATCGGTGACGATGATCGCAg
2756411..2756445	35	13	Rv1234	(-1665)	GGGACGGTCACCGGCGGTCAGGATGTCGTAATCGg
2959794..2959828	35	13	Rv2365c	None	GCGCTTTAATACGTCGTAGGCATTCACTTCGTTGc
4119573..4119607	35	13	Rv2455c	nth(-3679)	GGCCGGGTTGTTGACGGCCGGGGTCAAGCTTCCCg
1376990..1377028	39	13	Rv2633c and intergenic	lpqY(-496)	CCAGCCCAGACAGGTTCCCGGTTCAACACCCGCCGCCGc
4168191..4168282	92	13	Rv3679	(-63)	GAAGGGGACCCCGCGCACCCGACAGAGCCCGTTGACCCTTGCTGCCTTCCAGCCCTGGGGGAGTTCACAGGATAGACGCCGCGCGGGGTCCa
498846..498880	35	14	Inter geneic	(+739)	GGTCAGCGCAGTTCCCACCCACACCACCATGACGa
1666920..1666954	35	14	Inter geneic	acn(-874)	GATGCTTGTGGTGTCGTTGGCGTTGACCTGCGCTg
2728347..2728381	35	14	mycP3	PE25(-81)	CACGTTGAGAAAACGCAACGACTTCACTGGCATAa
2782771..2782805	35	14	Rv0412c	(+1237)	GGCCCGTGACGGCACTTCGGTTTGACCGACCTGGa
2951533..2951573	41	14	Rv2477c	(-226)	GCGGTCGGCACGCTGTGCAGCGGCAGCGCAATGTCACCTAc
3496579..3496627	49	14	Rv2625c	(+64)	CCGTGCGAACGGTTCTCACCCTGGCCGTCCGGGCCCCCTCCATCCACAa
355733..355798	66	14	Rv3131	mycP3(+1235)	CAAGCCGGTCGCCGATCCGCCGGTCCCGGCGCCCAAAGACACCACACCGCGCAACGTCGCATTCGc
3153388..3153421	34	15	ctpV	efpA(+1210)	GCCATGCCGATGCCAATCCCGCCGACGACGATCg
1166640..1166674	35	15	cysS	(-1205)	AGGCCGAGGGTGCTCAGGTATCGGGAGAAATCTGc
2624292..2624327	36	15	efpA	(+471)	TCGTCGAGTCGAGTGGTGTTGTTGGTCACGGTCGGc

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
1080094..1080138	45	15	Rv1043c	ctpV(+1417)	CGAAGAAGATCGACACCGTGGTGTTCGACAAGACCG GCACCCTCa
2391717..2391761	45	15	Rv2345	cysS(+698)	TGGATGTCGAGGCCGCTTCCGATACGACTGAGCGCGA TGGCTGCg
1993213..1993246	34	16	bfrB	(+60)	GGGTTGGGGATTTCTGAACCGATCAGTCCCGGGt
2775535..2775568	34	16	Inter geneic	(+263)	AACTGGCCAGGCCTTACCTCGCACCGGGTAACCg
3496375..3496409	35	16	Rv1760	tgs1(-9)	CCTTCCCCGGTGATCCGGATTATCTGCAACCGTCa
4314343..4314387	45	16	Rv2472	bfrB(+165)	CGAAACCATGCAATGATGCTCGTGCAACACCTGCTCG ACCGCGAc
2949865..2949938	74	16	TB31.7	TB31.7(+272)	CCACAGTGAAATCGTTCCGGCGGCAGCCGTTCCCACA TTGGTCGACATGTCCAAAGACGCAGTGCTGATGGTCg
2280314..2280348	35	17	fadD31	(+734)	CGGTCCAGTCGCGTACTTGCGAACGGATGCGCCGc
2638357..2638391	35	17	hpx	PPE40(+1144)	GAAGAAGCCGATGTTTCCATCACCCGAGTTGAAGa
2885351..2885385	35	17	narK2	(+740)	GGATCAACGGTATCCGCCGGATGCGGTCGATTCGc
3540695..3540729	35	17	PPE40	hpx(+16)	CCCGAACACCTGGGTGTGTAACGGGGTGCCGTCCg
2177096..2177147	52	17	Rv2033c	fadD31(+9)	GGCTCCCGGCAGGAACTCAGGGTTCGTAGCGGCCTAC TACAAATCGAGGACt
1964660..1964721	62	17	Rv2565	narK2(+649)	GTAGATCGTGGTGATGTAGGTGGGCAGGTAGTTGCTG AACGCCACGAACCCGCCGAACACGa
1114988..1115021	34	18	lprQ	arcA(-2164)	GCGCTGCCGGGCATGATCGTCGGCGAAATCGCGc
2274682..2274716	35	18	Rv0998	None	CCACCATGTGTACCGATCGACCCAGCCCAGCCAGa
571794..571838	45	18	Rv2028c	lprQ(+84)	CGGCGTCTGGCGTTGACGGCCCTTGGGTTTGGGGTGT TGGCACCg
65446..65480	35	19	efpA	celA1(-72)	ATTCCCGAGAACAGCCTGAAATCCTGTTCGGTTGa
3153949..3153983	35	19	Inter geneic	efpA(+648)	AGGGCGGTGCGGGCCAGGTAGATCATCACCAGCCc
3497199..3497253	55	19	Rv3131	(+684)	CTGTCTCAGCCGTCCGACGTCTTACCCGCTGACGACG GCGCCGCGATCCTGGCAc
4100706..4100778	73	20	Inter geneic	None	GCTCGATCCAGAAGAGAAGGTTCGGTCTCCCGACCCG GGCGCCCAGCATGGTTCCCGGCACCCACGCGGAGTc
3468146..3468234	89	20	ssr	ssr(+185)	AGCTGATGCCAGGGTCCGGGCCGAGGGCGTTCCCGG TCTGACAGACTAGCCGTCGCTTAGGCAGCGAGAGCGT AGTCGCGCTGATGTGa
2279214..2279247	34	21	acg	acg(+85)	CCTGGCGCTGGATAGCCGAGGACCACACGGTTGc
4404606..4404640	35	23	Rv3131	None	CCATCAAGAAGTTCGCGTCGATGATGCAGTGTTCg
3496804..3496849	46	23	Rv3916c	(+289)	GTTTCCCCGATCCCAAGGACCGCTGCCATCTGGCCAC CATCGGGGt

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
1965823..1965861	39	24	iniB	(+166)	AGATCGGCGATGAACTCTCGGTCGCCCGAGCCTTGTC Cg
409680..409728	49	24	Rv1566c	iniB(+318)	CTGGGTGCTGGGTTCCTGGGTCAGGGTGGCCTGGCTC TCGCCGCGTCAa
3496517..3496571	55	24	Rv1738	(+2)	GACCGCAGCCGTTGATGGGAAAGGCCCGGCAGCCAT GAACACCCATTTCCCGGAc
1773938..1774017	80	24	Rv3131	(-109)	CGCGCGGTGCCGAAGGGAGGCGGCTGGACGGGCGCT TGCTGGACGGGCGCTTGCTGGACCGGCGCCTGTTGGA CCGGCGc
2767686..2767728	43	26	Rv2465c	(-21)	GGGCACCGGCGGCGCTTCGTGGGTCCGTTCGTACTCG GCGAGa
50642..50676	35	27	inol	ino1(+448)	AGCACACTGGGCGTAGAACTTGTCGGCTTCCTCCg
3649468..3649502	35	27	Rv3268	ctpC(-1024)	GCCACCGGTGAGCGCATCGAGCTATCCGCGGTGAc
916241..916275	35	28	Rv0822c	(+36)	CGGTCAACACCATCGGGGAATGCTGACTCGGAGAg
754760..754793	34	30	Rv0659c	atsD(-1344)	TCGGAATGCGGTGCGTGGCAACGTATGAATGTTg
3326014..3326063	50	31	Rv2970A	(+80)	CGACGCGTTCGATTTCCCTCGCTCGATGACGGAACCC GGCTTGGTCCGAa
4100781..4100947	167	32	Inter geneic	None	agccacgataacggcagaagtgttgcgggtctGCGTAATTGCGAACAG CAGATGGCATCGACGGCCCTTTGGGTGGGGCTACAGC TAGAAGCGTCGCAAGATCGCCGAGGCCACCCACGCA ACCCCAGGAGTgcacgcttggtaaccgagaaccgtgttggtgggcg
4079755..4079811	57	35	Some on Rv3640c and some on intergenic region	None	GGATCTCCTTCGGTGCGTGCTTGGCGGTACACACCGA CGATCTCGCCGACGGCCCCt
3826433..3826467	35	37	Rv3407	None	CAGGTGTCCTGATTCCGGCTCGTCGTCCACAAAAc
1816201..1816234	34	46	pykA	pykA(+12)	GGGAAAATCGTCTGCACTCTCGGGCCGGCCACCc
1824726..1824770	45	55	cydA	cydA(+1117)	TGGCCAGGAACGGGGCGGGCATGGTTAGCAGCGCCA GCCAGGAGa
4352944..4352979	36	62	Inter geneic	(-31)	TTCGACCTTCCGTCGGTCTCGCCCTTTCTCGTGTTt
2314134..2314170	37	64	rpsR	rpsR(+183)	GCAGCAGCGCCATCTCGCGGGCGTTCTTGATCGCCTg
1275613..1275674	62	70	Inter geneic	None	GGCGGTGACCGCGGCATTGTTTTCAGTTTGTGACAAC TTCTCAATATGCCGCGGTCGCCGCg
3468051..3468090	40	71	ssr	ssr(+329)	GATCTCCCGGTCACGCGAACTAGCCGAGATGACGATC CCa
4376405..4376472	68	77	Rv3894c	(+3980)	ACAACTGGGCCACTTTCGCCGAGGTCTGGGATTTCAC CCAGGGGTCCATCGGCATCGTGGCCCAGTTc

Chromosome Region	Length	Coverage	Location on geneome	Gene Distance	Sequence
913457..913523	67	92	phoT	phoT(+731)	GAACCAGAAGGCCACCGAGGACTACATCTCCGGGCG CTTCGGCTAGGCCCGATGCCCTCGATGGCCa
3023949..3023988	40	113	ideR	ideR(+384)	GGCAACCCGATCCCGGGCCTGGTGGAACTTGGCGTG GGCc
2404494..2404549	56	123	some on Rv2144c and some on intergenic region	(-2772)	GGGCCAAGACCAGCGCAATGATCAGCATCGGTCATCC TCCAACCGGCTAGCAGCGa
2844174..2844207	34	128	fas	fas(+5125)	CGCTCGGCGTTGAGCACTTCCACTGTGCTGTGGg
1960700..1960781	82	129	Inter geneic and on Rv1734c	(-909)	GGGCGATTGAGCGGGTATGTACACCCGGTTTGACCTA CCGTCCCAAGACGGGGCTACCGCCTTCGGGCAGATCC TCATCCTg
4099412..4099475	64	152	Inter geneic	(-4498)	GAGGCGAGGGTCGTCGTGCATCAGCCCCGGGGGGTC GGACTGATACACCCTCGGCTATGGCCGa
80258..80292	35	165	glyA	glyA(-1362)	GGCGGCCACGGGAAACGGACCCGCAGCAACGCGGg
1473259..1473332	74	409	rrs	rrs(+1413)	CGGTAACACCCGAAGCCAGTGGCCTAACCCTCGGGA GGGAGCTGTCGAAGGTGGGATCGGCGATTGGGACGA Ag
4252929..4252971	43	424	Inter geneic	(-22)	ACGTCGTTGTGGCTTAAGGTTTCTGGCTTGTCAAGGA TCAGAa
1342599..1342670	72	504	IS1081-2	0	AGGTCATTGCGTCATTTCCTTCGATTGACTTTTGCTGG TCGTTTCGAAGGATCACGCGATGACCGCCCACTa
1469528..1469574	47	555	IS1557-2	murA(-747)	TCTGATCCTTGACAAGCCAGAAACCTTAAGCCACAAC GACGTGCGCc

APPENDIX H

Appendix I: sRNA fragments identified in 50 ng CF RNA sample compared to literature. Sequence reads from *M. tuberculosis* CF RNA 50 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, and gene name given by NextGENe, gene. All fragments were visualized using the Artemis genome browser to identify location on the genome [1]. Each fragment was compared to sRNAs reported in the literature: Arnvig et al., 2009 [2] and 2011 [3]; DiChiara et al., 2010 [4]; and the stable RNA list from Tuberculist [5].

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1466925..1466959	35	5	atpD(+1084)	atpD	towards the end	In Table 3 of the Arnvig et al, 2011 paper
989639..989673	35	5	citA(+188)	citA	towards the beginning	
3023949..3023988	40	5	ideR(+384)	ideR	towards the middle	
409681..409723	43	5	iniB(+319)	iniB	towards the beginning	
2692351..2692385	35	5	PE_PGRS41(-414)	intergenic	Intergenic region between Rv2395 and PE_PGRS41	From Tuberculist Stable RNA list, found in DiChiara, 2010 paper.
4100860..4100947	88	5	None	intergenic	Intergenic region between Rv3661 and Rv3662c	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv3661 and Rv3662c as flanking CDSs. They annotate it as MTS2823.
2626202..2626236	35	5	esxP(+283)	Intergenic and esxP	In the intergenic region between esxO and esxP. Some on the end of esxP	
4079769..4079807	39	5	None	IS1553 and intergenic	At the end of IS1553 and some on the intergenic region between IS1553 (and Rv3640c) and fic	IG2805 from Table S3 reads mapped to intergenic regions Arnvig 2011
1683101..1683134	34	5	mutA(+944)	mutA	middle	

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
347393..347426	34	5	PE5(-2198)	Rv0284	towards the middle	
352230..352281	52	5	(+81)	Rv0289	towards the beginning	
378209..378243	35	5	PPE6(-2498)	Rv0311	middle	
2277851..2277888	38	5	(+598)	Rv2030c	towards the beginning	In Table 4 of the Arnvig et al, 2011 paper
2278147..2278182	36	5	(+304)	Rv2030c	towards the beginning. Closer to the beginning of the gene than the (+598) fragment	In Table 4 of the Arnvig et al, 2011 paper
2502565..2502599	35	5	rnpB(-1814)	Rv2228c	towards the beginning	
2951530..2951592	63	5	(-223)	Rv2625c	towards the end	In Table 4 of the Arnvig et al, 2011 paper
3497036..3497070	35	5	(+521)	Rv3131	middle	In Table 4 of the Arnvig et al, 2011 paper
3606901..3606935	35	5	None	Rv3229c	close to the beginning	
3826273..3826307	35	5	None	Rv3407	close to the beginning	
3495854..3495897	44	5	tgs1(+469)	tgs1	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
4314343..4314391	49	6	bfrB(+165)	bfrB	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
2319295..2319328	34	6	cobN(+1425)	cobN	towards the middle	
918303..918341	39	6	desA1(+409)	desA1	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
3499671..3499705	35	6	devR(+210)	devR	towards the beginning	From Tuberculist Stable RNA list, found in the Arnvig and Young, 2009 paper. Also in Arnvig and Young, 2011 paper.
2483389..2483425	37	6	dlaT(+1424)	dlaT	towards the end	
2955616..2955660	45	6	(-4309)	intergenic	Intergenic region between Rv2628 and Rv2629. Closer to the beginning of Rv2629.	IG2022 from Table S3 reads mapped to intergenic regions Arnvig 2011
1987052..1987085	34	6	plcD(+611)	plcD	towards the end	
353331..353365	35	6	mycP3(-1133)	Rv0290	towards the beginning	
619032..619065	34	6	(+727)	Rv0528	middle	
1412992..1413025	34	6	(+1098)	Rv1264	close to the end	
3705429..3705463	35	6	sdhA(+429)	sdhA	towards the beginning	
2949955..2950022	68	6	TB31.7(+362)	TB31.7	middle	
1080848..1080906	59	7	ctpV(+2171)	ctpV	towards the end	
420761..420804	44	7	dnaK(+926)	dnaK	middle	
2955455..2955489	35	7	(-4148)	intergenic	Intergenic region between Rv2628 and Rv2629. Close to the end of Rv2628.	IG2022 from Table S3 reads mapped to intergenic regions Arnvig 2011
1391026..1391060	35	7	kgd(+1992)	kgd	middle	
244740..244774	35	7	mmpL3(+2544)	mmpL3	towards the end	

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
345471..345505	35	7	PE5(-4119)	Rv0283	close to the end	
1824736..1824770	35	8	cydA(+1117)	cydA	towards the end	
1220452..1220486	35	8	glyA(-52)	intergenic	In the intergenic region between coaA and glyA. Close to the beginning of glyA	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv3661 and Rv3662c as flanking CDSs. They annotate it as MTS0858.
1797585..1797619	35	8	nadC(+197)	nadC	towards the beginning	
3414771..3414804	34	8	nrdH(+154)	nrdH	Towards the end	
274730..274769	40	8	None	Rv0229c	towards the middle	
1338417..1338451	35	8	(+62)	Rv1194c	towards the beginning	
1965776..1965820	45	8	(+119)	Rv1738	middle	In Table 4 of the Arnvig et al, 2011 paper
1965823..1965861	39	8	(+166)	Rv1738	towards the end next to another fragment (+119)	In Table 4 of the Arnvig et al, 2011 paper
2250949..2251015	67	8	otsB1(-987)	Rv2004c, intergenic, and Rv2005c	Fragment covers the beginning of Rv2004c, the intergenic region between Rv2004c and Rv2005c, and the end of Rv2005c	
2274682..2274716	35	8	None	Rv2028c	towards the end	
2374240..2374279	40	8	lppK(-2292)	Rv2114	towards the middle	
2782771..2782805	35	8	(+1237)	Rv2477c	towards the end	
3046019..3046053	35	8	(-4558)	Rv2733c	towards the beginning	
1340943..1340977	35	9	esxK(+284)	esxK and intergenic	At the end of esxK with 22bp into the intergenic region between esxK and esxL	
1169366..1169429	64	9	(-3931)	IS1081-1	At the beginning of Rv1047 and towards the beginning of the insertion sequence IS1081 moble element	
3751654..3751688	35	9	PPE55(+1496)	PPE55	towards the beginning	
2314134..2314169	36	9	rpsR(+184)	rpsR	towards the end	
3326014..3326063	50	9	(+80)	Rv2970A	towards the middle	
1581101..1581136	36	10	(-679)	intergenic	Intergenic region between Rv1404 and Rv1405c	IG1101 from Table S3 reads mapped to intergenic regions Arnvig 2011
851642..851676	35	10	phoP(+34)	phoP	towards the beginning	
4056124..4056158	35	10	ephA(-1575)	Rv3616c	towards the beginning	In Table 4 of the Arnvig et al, 2011 paper

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
2219456..2219490	35	11	None	intergenic	Intergenic region between Rv1976c and Rv1977	IG1535 from Table S3 reads mapped to intergenic regions Arnvig 2011
15823..15857	35	11	pknB(+1613)	pknB	towards the end	
88190..88244	55	11	hycD(-4084)	Rv0079	At the beginning of the gene, overlapping onto the intergenic region between Rv0078A and Rv0079	In Table 4 of the Arnvig et al, 2011 paper
1166640..1166674	35	11	(-1205)	Rv1043c	towards the beginning	
3826433..3826467	35	11	None	Rv3407	towards the end	
2517341..2517396	56	12	fabD(+554)	fabD	towards the middle	
1962197..1962230	34	12	narX(+1956)	intergenic and narX	Intergenic region between Rv1735c and narX. 2bp on the end of narX	In Table 4 of the Arnvig et al, 2011 paper
1964687..1964721	35	12	narK2(+649)	narK2	middle	
806142..806176	35	12	rpsQ(+386)	rpsQ	at the end of the gene with 10 bp hanging over into the intergenic region between rpsQ and atsA	
474296..474330	35	12	ndhA(-1657)	Rv0394c	towards the end	
3496799..3496837	39	12	(+284)	Rv3131	towards the beginning	In Table 4 of the Arnvig et al, 2011 paper
3497203..3497248	46	12	(+688)	Rv3131	towards the end	In Table 4 of the Arnvig et al, 2011 paper
3666695..3666729	35	13	accA3(+338)	accA3	towards the beginning	
68848..68882	35	14	None	Rv0064	towards the beginning	
3468156..3468242	87	14	ssr(+177)	ssr	towards the middle. Next to another fragment (+329)	From Tuberculist Stable RNA list.
2949865..2949899	35	14	TB31.7(+272)	TB31.7	towards the middle	
3415123..3415167	45	16	nrdH(-165)	intergenic	Intergenic region between nrdH and Rv3054c	IG2356 from Table S3 reads mapped to intergenic regions Arnvig 2011
4168234..4168282	49	16	(-106)	intergenic	Intergenic region between Rv3722c and tRNA-Ser(GGA)	From Tuberculist Stable RNA list, (See Arnvig and Young, 2009; DiChiara et al., 2010
2238981..2239015	35	17	ctpF(-1144)	Intergenic and Rv1996	intergenic region between Rv1995 and Rv1996 with some at the beginning of Rv1996	IG1550 from Table S3 reads mapped to intergenic regions Arnvig 2011
164390..164424	35	18	cyp138(+1024)	cyp138	towards the end	
4262658..4262692	35	21	fadD32(+374)	fadD32	towards the beginning	

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1187430..1187464	35	21	0	Rv1065	At the beginning of Rv1065 with 5bp overlapping onto the intergenic region between lpqV and Rv1065	
3468051..3468090	40	21	ssr(+329)	ssr	middle	From Tuberculist Stable RNA list.
2767686..2767719	34	22	(-21)	Rv2465c	close to the end	
80258..80292	35	24	glyA(-1362)	repeat region	close to the beginning of a repeat region "MTV030.15, len:315bp. Probable REP'-1 pseudogene fragment, similar to many Mycobacterium proteins inside REP13E12 elements (Rv0071/Rv0072)	
1275613..1275673	61	27	None	intergenic	In the intergenic region between Rv1147 and Rv1148c. Close to the end of Rv1147.	IG903 from Table S3 reads mapped to intergenic regions Arnvig 2011
2040858..2040894	37	34	PPE28(+1405)	PPE28	towards the middle	
1960675..1960774	113	37	(-878)	intergenic	Intergenic region between Rv1733c and Rv1734c. Very close to the end of Rv1734c, with one bp on the end of Rv1734c	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv3661 and Rv3662c as flanking CDSs. They annotate it as MTS1338.
1993213..1993246	34	41	(+60)	Rv1760	towards the beginning	
3649468..3649502	35	48	ctpC(-1024)	Rv3268	towards the beginning	
754758..754793	36	54	atsD(-1344)	Rv0659c	towards the end	
4099438..4099475	38	58	(-4524)	intergenic	Intergenic region between Rv3660 and Rv3661	From Tuberculist Stable RNA list, (See Arnvig and Young, 2009; DiChiara et al., 2010.
2404472..2404549	78	63	(-2750)	Rv2144c and intergenic	At the beginning of Rv2144c and into the intergenic region between Rv2144c and wag31	
4352941..4352979	39	91	(-31)	intergenic	Intergenic region between esxA and Rv3876	IG3002 from Table S3 reads mapped to intergenic regions Arnvig 2011
2844174..2844208	35	119	fas(+5124)	fas	middle	

Transcript Site	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1342600..1342663	64	338	0	IS1081-2	5 bp at the beginning of Rv1199c, with the rest towards the beginning of IS1081-2 insertion sequence	
1469500..1469574	75	357	murA(-747)	IS1557-2	5 bp at the beginning of Rv1313c and the rest close to the beginning of Insertion sequence IS1557	
4252929..4252971	43	431	(-22)	intergenic	Intergenic region between Rv3797 (fadE35) and Rv3798	IG2937 from Table S3 reads mapped to intergenic regions Arnvig 2011

References

1. Rutherford, K., et al., *Artemis: sequence visualization and annotation.* Bioinformatics, 2000. **16**(10): p. 944-945.
2. Arnvig, K.B. and D.B. Young, *Identification of small RNAs in Mycobacterium tuberculosis.* Molecular microbiology, 2009. **73**(3): p. 397-408.
3. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.
4. DiChiara, J.M., et al., *Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis.* Nucleic Acids Research, 2010. **38**(12): p. 4067-4078.
5. Lew, J.M., et al., *TubercuList–10 years after.* Tuberculosis, 2011. **91**(1): p. 1-7.

APPENDIX I

Appendix I: sRNA fragments identified in 100 ng CF RNA sample compared to literature. Sequence reads from *M. tuberculosis* CF RNA 100 ng sample were mapped to the *M. tuberculosis* H37Rv GenBank reference file, and peak identification reports were exported in text format. Peak identification report gives chromosome region, length of the fragment, depth of coverage, and gene name given by NextGENe, gene. All fragments were visualized using the Artemis genome browser to identify location on the genome [1]. Each fragment was compared to sRNAs reported in the literature: Arnvig et al., 2009 [2] and 2011 [3]; DiChiara et al., 2010 [4]; and the stable RNA list from Tuberculist [5].

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
119170..119204	35	5	(+1456)	Rv0102	towards the end	In TableS2 of Arnvig 2011.
194242..194276	35	5	(+662)	Rv0165c	towards the end	In TableS2 of Arnvig 2011.
251787..251840	54	5	pckA(+5)	pckA	Very close to the beginning of the gene	In TableS2 of Arnvig 2011. Rv0211
269454..269488	35	5	(-1690)	Rv0225	towards the end	In TableS2 of Arnvig 2011.
274730..274764	35	5	None	Rv4006	Towards the end	
303550..303583	34	5	nirB(+684)	nirB	Close to the middle	In TableS2 of Arnvig 2011. Rv0252
337304..337337	34	5	PE_PGRS4(+1736)	PE_PGRS4	close to the middle	In TableS2 of Arnvig 2011. Rv0279c
348462..348496	35	5	PE5(-1128)	Rv0284	towards the middle	In TableS2 of Arnvig 2011.
474296..474330	35	5	ndhA(-1657)	ndhA	towards the end	In TableS2 of Arnvig 2011. Rv0392c
601708..601741	34	5	hemA(+1267)	hemA	towards the end	In TableS2 of Arnvig 2011. Rv0509

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
776008..776071	64	5	mmpL5(+2409)	mmpL5	close to the end	In TableS2 of Arnvig 2011. Rv0676c
782891..782925	35	5	fusA1(+406)	fusA1	Towards the beginning	In TableS2 of Arnvig 2011. Rv0684
873531..873565	35	5	hemH(+188)	hemH	Towards the beginning	
884318..884352	35	5	(-268)	Rv0790c	towards the middle	In TableS2 of Arnvig 2011.
1099314..1099348	35	5	pepD(+248)	pepD	Towards the beginning	In TableS2 of Arnvig 2011. Rv0983
1314313..1314361	49	5	pks3(+588)	pks3	close to the middle	In TableS2 of Arnvig 2011. Rv1180
1511851..1511885	35	5	fadE14(+1005)	radE14	towards the end	
1516874..1516919	46	5	fabG(-572)	Rv1349	towards the middle	In TableS2 of Arnvig 2011.
1564447..1564480	34	5	gmk(+46)	gmk	Towards the beginning	In Table 6 of the Arnvig et al, 2011 paper
1577190..1577223	34	5	(+260)	Rv1401	middle	In TableS2 of Arnvig 2011.
1736341..1736375	35	5	ileS(-144)	Intergenic	Intergenic region between Rv1535 and ileS	Table 8 of Arnvig 2011 paper (abundant 5' UTRs in intergenic regions) (?)
1797489..1797527	39	5	nadC(+101)	nadC	Towards the beginning	In TableS2 of Arnvig 2011. Rv1596
1866320..1866354	35	5	argC(+744)	argC	towards the end	In TableS2 of Arnvig 2011. Rv1652
1901178..1901212	35	5	dsbF(-536)	Rv1676	Towards the beginning	In TableS2 of Arnvig 2011.
2032991..2033025	35	5	mycP5(-704)	Rv1795	middle	In TableS2 of Arnvig 2011.
2055657..2055690	34	5	(-298)	Rv1813c	at the end of the gene, and a little over the region between Rv1813c and Rv1812c	In TableS2 of Arnvig 2011.
2056276..2056310	35	5	erg3(-211)	Intergenic	Intergenic region between Rv1813c and erg3	IG1406 from Table S3 reads mapped to intergenic regions Arnvig 2011
2239323..2239356	34	5	ctpF(-803)	Rv1996	towards the middle Rv1996 is before ctpF	In Table 4 of the Arnvig et al, 2011 paper
2289878..2289911	34	5	lipT(+1309)	lipT	towards the end	In TableS2 of Arnvig 2011. Rv2045c
2404447..2404484	38	5	(-2725)	Rv2144c	close to the beginning. Very close to another fragment (-2772)	

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
2477897..2477931	35	5	(+707)	Rv2212	towards the middle	
2500383..2500425	43	5	(-4984)	Intergenic	very close to the end of Rv2226 and the end of rnpB (misc_RNA)	IG1727 from Table S3 reads mapped to intergenic regions Arnvig 2011
2500503..2500537	35	5	rnpB(+214)	rnpB	towards the end	Ribonuclease P RNA. From Tuberculist Stable RNA list
2502565..2502599	35	5	rnpB(-1814)	Rv2228c	Towards the beginning	
2519178..2519212	35	5	kasA(+1063)	kasA	towards the end	In TableS2 of Arnvig 2011. Rv2245
2672020..2672054	35	5	mbtB(+3783)	mbtB	towards the end	In TableS2 of Arnvig 2011. Rv2383c
2783947..2783981	35	5	(+61)	Rv2477c	very close to the beginning of the gene	
2819339..2819373	35	5	scoA(+497)	scoA	towards the middle	In TableS2 of Arnvig 2011. Rv2504c
2827612..2827645	34	5	orn(+455)	orn	towards the end	In TableS2 of Arnvig 2011. Rv2511
2952672..2952716	45	5	(-1365)	Rv2626c	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
2953710..2953754	45	5	(-2403)	Rv2627c	towards the end	In Table 4 of the Arnvig et al, 2011 paper
2978147..2978180	34	5	None	Rv2655c	towards the middle	
3020895..3020929	35	5	sigB(-1532)	Rv2707	towards the middle.	
3022467..3022500	34	5	sigB(+6)	sigB	beginning	In TableS2 of Arnvig 2011. Rv2710
3651551..3651585	35	5	ctpC(+1025)	ctpC	middle. "Cation transport ATPase"	In TableS2 of Arnvig 2011. Rv3270
3900311..3900345	35	5	cpsA(-2733)	Rv3480c	close to the beginning	
3953877..3953911	35	5	(+446)	Rv3517	towards the middle	
4198611..4198644	34	5	tyrA(-1711)	Intergenic	Between Rv3750c and Rv3751	IG1379 from Table S3 reads mapped to intergenic regions Arnvig 2011

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
4221473..4221507	35	5	(+384)	Rv3776	towards the beginning. At the beginning of "domain of unknown function DUF 22" fragments flanking both ends of this domain	
4231808..4231842	35	5	None	Rv3785	middle	
4262779..4262812	34	5	fadD32(+254)	fadD32	Towards the beginning	In TableS2 of Arnvig 2011. Rv3801c
4302868..4302901	34	5	(-79)	Rv3828c	towards the end	
4334567..4334600	34	5	gltB(+1481)	gltB	towards the middle	In TableS2 of Arnvig 2011. Rv3859c
4369747..4369780	34	5	mycP2(+390)	mycP2	Towards the beginning	In TableS2 of Arnvig 2011. Rv3886c
4398690..4398723	34	5	sigM(-1463)	Rv3910	middle	
88211..88244	34	6	hycD(-4084)	Rv0079	Towards the beginning	In Table 4 of the Arnvig et al, 2011 paper
215774..215807	34	6	bglS(-462)	Rv0185	Towards the beginning	In TableS2 of Arnvig 2011.
326352..326385	34	6	fadE6(+2076)	fadE6	towards the end	In TableS2 of Arnvig 2011. Rv0271c
346439..346476	38	6	PE5(-3148)	Rv0284	Towards the beginning	In TableS2 of Arnvig 2011.
352444..352477	34	6	(+295)	Rv0289	close to the middle	In TableS2 of Arnvig 2011.
420777..420810	34	6	dnaK(+942)	dnaK	middle	In TableS2 of Arnvig 2011. Rv0350
595932..595966	35	6	serB1(+619)	serB1	towards the middle	In TableS2 of Arnvig 2011. Rv0505c
631410..631443	34	6	(-1679)	Rv0538	towards the end	In TableS2 of Arnvig 2011.
707305..707339	35	6	galTa(-4197)	Rv0613c	towards the end	In TableS2 of Arnvig 2011.
1004881..1004915	35	6	prrB(+926)	prrB	towards the middle	In TableS2 of Arnvig 2011. Rv0902c
1313412..1313455	44	6	pks3(-270)	Intergenic	between Rv1179 and pks3	IG930 from Table S3 reads mapped to intergenic regions Arnvig 2011

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1669140..1669174	35	6	moxR1(-109)	Rv1478 and intergenic	at the end of the the gene into the intergenic region between Rv1478 and moxR1	IG1154 Starting on1669143, the beginning of this fragment is not intergenic (some other type of regulatory element?) from Table S3 reads mapped to intergenic regions Arnvig 2011
1963271..1963312	42	6	narX(+874)	narX	middle	In Table 4 of the Arnvig et al, 2011 paper
2150873..2150907	35	6	nanT(-599)	Intergenic	Intergenic region between Rv1903 and Rv1904. Closer to Rv1904	Also IG1484 from Table S3 reads mapped to intergenic regions Arnvig 2011
2162128..2162161	34	6	aceAb(+562)	aceAb	in the middle	In TableS2 of Arnvig 2011. Rv1916
2239872..2239913	42	6	ctpF(-246)	ctpG	middle	In TableS2 of Arnvig 2011. Rv1992c
2252256..2252290	35	6	otsB1(+254)	otsB1	Towards the beginning	In TableS2 of Arnvig 2011. Rv2006
2276245..2276278	34	6	pfkB(+146)	pfkB	Towards the beginning	In TableS2 of Arnvig 2011. Rv2029c
2445035..2445069	35	6	(-2718)	Rv2182c	towards the middle	
2483389..2483423	35	6	dlaT(+1424)	dlaT	towards the end	
2593248..2593281	34	6	rocE(+872)	rocE	towards the middle	In TableS2 of Arnvig 2011. Rv2320c
2842511..2842545	35	6	fas(+6787)	fas	towards the end	In TableS2 of Arnvig 2011. Rv2524c
2952398..2952466	69	6	(-1091)	Rv2625c	close to the beginning	In Table 4 of the Arnvig et al, 2011 paper
3084849..3084898	50	6	ald(-1922)	Rv2777c	close to the middle	
3404859..3404893	35	6	ctaD(+28)	ctaD	very close to the beginning of the gene	In TableS2 of Arnvig 2011. Rv3043c
3492964..3492998	35	6	(-842)	Rv3127	towards the end of the gene. Very close to another fragment (-899)	In Table 4 of the Arnvig et al, 2011 paper
3493021..3493086	66	6	(-899)	Rv3127	towards the end	In Table 4 of the Arnvig et al, 2011 paper
3497349..3497382	34	6	(+834)	Rv3131	towards the end	In Table 4 of the Arnvig et al, 2011 paper

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
3516906..3516939	34	6	nuoG(+160)	nuoG	Towards the beginning	In TableS2 of Arnvig 2011. Rv3151
3623406..3623440	35	6	lpqB(+1470)	lpqB	Towards the end of the gene.	In TableS2 of Arnvig 2011. Rv3244c
3666695..3666729	35	6	accA3(+338)	accA3	Towards the beginning	In TableS2 of Arnvig 2011. Rv3285
3894312..3894345	34	6	PE31(+219)	PE31	towards the end	Table 8 of Arnvig 2011 paper (abundant 5' UTRs in intergenic regions)
4056007..4056044	38	6	ephA(-1689)	Rv3616c	towards the end	In Table 4 of the Arnvig et al, 2011 paper
4222194..4222228	35	6	(+1105)	Rv3776	towards the end. At the end of "domain of unknown function DUF 222"	
4258521..4258555	35	6	pks13(+2591)	pks13	middle	In TableS2 of Arnvig 2011. Rv3800c
7449..7483	35	7	gyrA(+147)	gyrA	Towards the beginning	In TableS2 of Arnvig 2011. Rv0006
25117..25151	35	7	TB39.8(+293)	TB39.8	Towards the beginning	In TableS2 of Arnvig 2011. Rv0020c
340528..340608	81	7	PPE3(+1164)	PPE3	towards the end	
349084..349118	35	7	PE5(-506)	Rv0284	towards the end	In TableS2 of Arnvig 2011.
662451..662484	34	7	nrdZ(+1156)	nrdZ	middle	In TableS2 of Arnvig 2011. Rv0570
867824..867858	35	7	(+493)	Rv0774c	middle	In TableS2 of Arnvig 2011.
917498..917537	40	7	(-1187)	Rv0823c	Towards the beginning	In TableS2 of Arnvig 2011.
1113102..1113140	39	7	(+718)	Rv0996	towards the middle	In TableS2 of Arnvig 2011.
1222477..1222511	35	7	desA2(+518)	desA2	towards the middle	In TableS2 of Arnvig 2011. Rv1094
1227126..1227160	35	7	fumC(+405)	fumC	Towards the beginning	In TableS2 of Arnvig 2011. Rv1098c
2017524..2017558	35	7	malQ(-48)	Intergenic	Intergenic region between malQ and Rv1782	IG1379 from Table S3 reads mapped to intergenic regions Arnvig 2011
2073625..2073658	34	7	(+544)	Rv1828	towards the end	In TableS2 of Arnvig 2011.

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
2278025..2278059	35	7	(+427)	Rv2030c	Towards the beginning	In Table 4 of the Arnvig et al, 2011 paper
3468268..3468321	54	7	ssr(+98)	ssr	Towards the beginning. Ssr is classified as a misc RNA	From Tuberculist Stable RNA list.
3806205..3806239	35	7	acrA1(+584)	acrA1	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
4024199..4024262	64	7	ispF(+85)	ispF	Towards the beginning	In TableS2 of Arnvig 2011. Rv3581c
4346889..4346923	35	7	PE35(-3822)	Rv3870	Towards the beginning	
88506..88540	35	8	hycD(-3788)	hycD	towards the middle	In TableS2 of Arnvig 2011. Rv0084
352199..352242	44	8	(+50)	Rv0289	close to the beginning of Rv0289	In TableS2 of Arnvig 2011.
353501..353542	42	8	mycP3(-956)	Rv0290	towards the middle	In TableS2 of Arnvig 2011.
851642..851676	35	8	phoP(+34)	phoP	Towards the beginning	In TableS2 of Arnvig 2011. Rv0757
1132353..1132392	40	8	pks16(+728)	pks16	middle	In TableS2 of Arnvig 2011. Rv1013
1150652..1150686	35	8	kdpD(+1000)	kdpD	towards the middle	In TableS2 of Arnvig 2011. Rv1028c
1235415..1235449	35	8	xseA(+11)	xseA	at the beginning	In TableS2 of Arnvig 2011. Rv1108c
1796054..1796088	35	8	nadB(+249)	nadB	Towards the beginning	In TableS2 of Arnvig 2011. Rv1595
1796908..1796941	34	8	nadB(+1103)	nadB	towards the end	In TableS2 of Arnvig 2011. Rv1595
1845595..1845628	34	8	uvrA(+1854)	uvrA	towards the middle	In TableS2 of Arnvig 2011. Rv1638
2278606..2278652	47	8	hspX(+280)	hspX	towards the end	In TableS2 of Arnvig 2011. Rv2031c
2484655..2484688	34	8	lipB(+71)	lipB	Towards the beginning	In TableS2 of Arnvig 2011. Rv2217
2955616..2955649	34	8	(-4309)	Intergenic	Intergenic between Rv2628 and Rv2629	IG2022 from Table S3 reads mapped to intergenic regions Arnvig 2011
2979238..2979272	35	8	clpC2(-4624)	Rv2657c	Towards the beginning	

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
3496869..3496903	35	8	(+354)	Rv3131	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
3498580..3498619	40	8	devS(+646)	devS	towards the middle	In TableS2 of Arnvig 2011. Rv3132c
3499671..3499705	35	8	devR(+210)	devR	towards the middle	In TableS2 of Arnvig 2011. Rv3133c
3606667..3606701	35	8	None	Rv3229c	towards the middle	
4029729..4029763	35	8	mutY(-730)	Rv3587c / intergenic	At the beginning of Rv3587c, over lapping a little into the intergenic region between Rv3587c and Rv3588c	3bp in IG2768 from Table S3 reads mapped to intergenic regions Arnvig 2011
4220050..4220083	34	8	lipE(+365)	lipE	towards the middle	In TableS2 of Arnvig 2011. Rv3775
4359090..4359155	66	8	None	Rv3879c	Towards the beginning	
340750..340826	77	9	PPE3(+1386)	PPE3	closer to the end of PPE3 than PPE3(+1164)	
353322..353365	44	9	mycP3(-1133)	Rv0290	Towards the beginning	In TableS2 of Arnvig 2011.
356191..356225	35	9	(+311)	Rv0292	Towards the beginning	In TableS2 of Arnvig 2011.
1197094..1197127	34	9	(+815)	Rv1072	at the end	In Table 3 of the Arnvig et al, 2011 paper
1340943..1340977	35	9	esxK(+284)	esxK	at the end of Rv1247c into the intergenic region between esxK and esxL	In TableS2 of Arnvig 2011. Rv1197
1965772..1965820	49	9	(+115)	Rv1738	middle	In Table 4 of the Arnvig et al, 2011 paper
2103054..2103088	35	9	ndh(-12)	Intergenic	Intergenic region between ndh and Rv1855c. Very close to the beginning of ndh	IG1440 from Table S3 reads mapped to intergenic regions Arnvig 2011
2256537..2256600	64	9	fdxA(-109)	Intergenic	very close to the end of the gene Rv2008c. between fdxA and Rv2008c	IG1562 from Table S3 reads mapped to intergenic regions Arnvig 2011
2655876..2655940	65	9	cfp2(+175)	cfp2	towards the middle	In Table 5 of Arnvig et al 2011. Antisense transcripts deried from 3' UTRs (?)
3246358..3246392	35	9	ppsA(+913)	ppsA	Towards the beginning	In TableS2 of Arnvig 2011. Rv2931

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
3548920..3548954	35	9	mesT(-3620)	Intergenic	13 bp after the end of Rv3179	IG2451 from Table S3 reads mapped to intergenic regions Arnvig 2011
3589569..3589602	34	9	(+175)	Rv3212	Towards the beginning	
1413108..1413142	35	10	None	Intergenic	Intergenic region between Rv1264 and Rv1265. Very close to the end of Rv1264	mcr11, putative small regulatory RNA (ncRNA). From Tuberculist Stable RNA list, found in the DiChiara et al., 2010 paper. Same as MTs0997 from Table 10 of Arnvig 2011 paper
2122411..2122445	35	10	lldD2(+706)	lldD2	middle	In Table 4 of the Arnvig et al, 2011 paper
2150918..2150951	34	10	None	Intergenic	before the beginning of the gene Rv1904	IG1484 from Table S3 reads mapped to intergenic regions Arnvig 2011
2463741..2463775	35	10	ctaC(+976)	ctaC	close to the end	In TableS2 of Arnvig 2011. Rv2200c
2847337..2847371	35	10	fas(+1961)	fas	Towards the beginning	In TableS2 of Arnvig 2011. Rv2524c
2954155..2954188	34	10	(-2848)	Rv2627c	middle	In Table 4 of the Arnvig et al, 2011 paper
3758308..3758342	35	10	PPE56(+8760)	PPE56	Towards the end of the gene.	In TableS2 of Arnvig 2011. Rv3350c
32182..32215	34	11	bioF2(-2080)	Rv0029	Towards the beginning	In TableS2 of Arnvig 2011. Rv0029
378204..378242	39	11	PPE6(-2493)	Rv0309	towards the middle	In TableS2 of Arnvig 2011.
918303..918380	78	11	desA1(+370)	desA1	towards the middle	From Tuberculist Stable RNA list, found in the Arnvig and Young, 2009 paper. Also in Arnvig and Young, 2011 paper.
1199075..1199109	35	11	fadA3(+264)	fadA3	Towards the beginning	In TableS2 of Arnvig 2011. Rv1074c
1220452..1220486	35	11	glyA(-52)	Intergenic	Intergenic close to the beginning of glyA. Between coaA and glyA.	Table 10 of Arnvig 2011 paper, named MTS0858. coaA=Rv1092c, glyA=Rv1093

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1466925..1466959	35	11	atpD(+1084)	atpD	towards the end	In Table 3 of the Arnvig et al, 2011 paper
1485598..1485663	66	11	fadA4(-199)	Rv1322A	towards the middle	In TableS2 of Arnvig 2011.
1998195..1998228	34	11	PE_PGRS31(-2386)	Rv1765c	towards the middle	In TableS2 of Arnvig 2011.
2517353..2517387	35	11	fabD(+566)	fabD	towards the middle	In TableS2 of Arnvig 2011. Rv2243
2845154..2845188	35	11	fas(+4144)	fas	middle	In TableS2 of Arnvig 2011. Rv2524c
3093770..3093825	56	11	rpsO(-22)	Intergenic	between rpsO and ribF	IG2148 from Table S3 reads mapped to intergenic regions Arnvig 2011
4056753..4056787	35	11	ephA(-946)	Intergenic	between Rv3616c and ephA. Closer to the beginning of Rv3616c.	IG2787 from Table S3 reads mapped to intergenic regions Arnvig 2011
1169366..1169426	61	12	(-3931)	IS1081-1	close to the beginning of insertion sequence of IS1081	
1197139..1197190	52	12	None	Intergenic	between Rv1072 and Rv1073	IG841 from Table S3 reads mapped to intergenic regions Arnvig 2011
1389211..1389245	35	12	(-583)	Rv1247c	At the beginning of the gene, with one bp in the intergenic region between Rv1247 and kgd	In TableS2 of Arnvig 2011.
2236394..2236438	45	12	ctpG(+868)	ctpG	middle	In TableS2 of Arnvig 2011. Rv1992c
2626202..2626236	35	12	esxP(+283)	Rv2345	towards the middle	
3606901..3606935	35	12	None	Rv3229c	Towards the beginning	
1376990..1377028	39	13	lpqY(-496)	Rv1234	Towards the beginning. Rv1234 is before lpqY	In TableS2 of Arnvig 2011.
1419706..1419740	35	13	lprA(+8)	lprA	very close to the beginning of the gene	In TableS2 of Arnvig 2011. Rv1270c
2102731..2102764	34	13	ndh(+278)	ndh	Towards the beginning	In TableS2 of Arnvig 2011. Rv1854c
2646753..2646787	35	13	(+301)	Rv2365c	very close to the end of the gene	
2756411..2756445	35	13	(-1665)	Rv2455c	Towards the beginning	

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
2959794..2959828	35	13	None	Rv2633c and intergenic	At the beginning of Rv2633c and overlapping onto the intergenic region between Rv2633c and PE_PGRS46	10bp in IG2029 from Table S3 reads mapped to intergenic regions Arnvig 2011
4119573..4119607	35	13	nth(-3679)	Rv3679	towards the end. Within a predicted ATPase involved in chromosome partitioning	
4168191..4168282	92	13	(-63)	Intergenic	Between Rv3722c and Rvnt41. Rvnt41 is between Rv3722c and Rv3723.	From Tuberculist Stable RNA list, (See Arnvig and Young, 2009; DiChiara et al., 2010
355733..355798	66	14	mycP3(+1235)	mycP3	close to the end	In TableS2 of Arnvig 2011. Rv0291
498846..498880	35	14	(+739)	Rv0412c	towards the middle	In TableS2 of Arnvig 2011.
1666920..1666954	35	14	acn(-874)	Intergenic	Intergenic region between Rv1476 and Rv1477. Closer to Rv1477	IG1152 from Table S3 reads mapped to intergenic regions Arnvig 2011
2728347..2728381	35	14	PE25(-81)	Intergenic	between PE25 and Rc2432c	IG1875 from Table S3 reads mapped to intergenic regions Arnvig 2011
2782771..2782805	35	14	(+1237)	Rv2477c	Towards the end	
2951533..2951573	41	14	(-226)	Rv2625c	towards the end	In Table 4 of the Arnvig et al, 2011 paper
3496579..3496627	49	14	(+64)	Rv3131	towards the beginning, very close to another segment (+2)	In Table 4 of the Arnvig et al, 2011 paper
1080094..1080138	45	15	ctpV(+1417)	ctpV	towards the middle	In TableS2 of Arnvig 2011. Rv0969
1166640..1166674	35	15	(-1205)	Rv1043c	Towards the beginning	In TableS2 of Arnvig 2011.
2391717..2391761	45	15	cysS(+698)	cysS	towards the middle	In TableS2 of Arnvig 2011. Rv2130c
2624292..2624327	36	15	(+471)	Rv2345	Towards the beginning. At the end of a "domain of unknown function (DUF477)"	
3153388..3153421	34	15	efpA(+1210)	efpA	towards the end	In TableS2 of Arnvig 2011. Rv2846c
1993213..1993246	34	16	(+60)	Rv1760	Towards the beginning	In TableS2 of Arnvig 2011.

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
2775535..2775568	34	16	(+263)	Rv2472	at the end	In Table 3 of the Arnvig et al, 2011 paper
2949865..2949938	74	16	TB31.7(+272)	TB31.7	towards the middle	In TableS2 of Arnvig 2011. Rv2623
3496375..3496409	35	16	tgs1(-9)	Intergenic	Very close to the beginning of tgs1. In the intergenic region between tgs1 and Rv3131	IG2419 from Table S3 reads mapped to intergenic regions Arnvig 2011
4314343..4314387	45	16	bfrB(+165)	bfrB	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
1964660..1964721	62	17	narK2(+649)	narK2	towards the middle	In TableS2 of Arnvig 2011. Rv1737c
2177096..2177147	52	17	fadD31(+9)	fadD31	Very close to the beginning of the gene	In TableS2 of Arnvig 2011. Rv1925
2280314..2280348	35	17	(+734)	Rv2033c	towards the end	
2638357..2638391	35	17	PPE40(+1144)	PPE40	towards the middle	In TableS2 of Arnvig 2011. Rv2356c
2885351..2885385	35	17	(+740)	Rv2565	towards the middle	
3540695..3540729	35	17	hpx(+16)	hpx	Very close to the beginning of the gene	In TableS2 of Arnvig 2011. Rv3171c
571794..571838	45	18	lprQ(+84)	lprQ	Towards the beginning	In TableS2 of Arnvig 2011. Rv0483
1114988..1115021	34	18	arcA(-2164)	Rv0998	towards the middle	In TableS2 of Arnvig 2011.
2274682..2274716	35	18	None	Rv2028c	towards the end	In TableS2 of Arnvig 2011.
65446..65480	35	19	celA1(-72)	Intergenic	between Rv0061 and celA1, closer to the end of Rv0061	IG50 from Table S3 reads mapped to intergenic regions Arnvig 2011
3153949..3153983	35	19	efpA(+648)	efpA	towards the middle	In TableS2 of Arnvig 2011. Rv2846c
3497199..3497253	55	19	(+684)	Rv3131	towards the middle	In Table 4 of the Arnvig et al, 2011 paper
3468146..3468234	89	20	ssr(+185)	ssr	towards the middle of ssr	From Tuberculist Stable RNA list.

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
4100706..4100778	73	20	None	Intergenic	behind another Intergenic fragment: 4100814. between Rv3661 and Rv3662c.	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv3661 and Rv3662c as flanking CDSs. They annotate it as MTS2823.
2279214..2279247	34	21	acg(+85)	acg	Towards the beginning	In Table 4 of the Arnvig et al, 2011 paper
3496804..3496849	46	23	(+289)	Rv3131	close to the middle, several other fragments on this gene	In Table 4 of the Arnvig et al, 2011 paper
4404606..4404640	35	23	None	Rv3916c	towards the end	
409680..409728	49	24	iniB(+318)	iniB	close to the beginning	In TableS2 of Arnvig 2011. Rv0341
1773938..1774017	80	24	(-109)	Rv1566c	Very close to the beginning of the gene	In TableS2 of Arnvig 2011.
1965823..1965861	39	24	(+166)	Rv1738	towards the end. Next to another fragment (+115)	In Table 4 of the Arnvig et al, 2011 paper
3496517..3496571	55	24	(+2)	Rv3131	2bp into the beginning	In Table 4 of the Arnvig et al, 2011 paper
2767686..2767728	43	26	(-21)	Rv2465c	close to the beginning	
50642..50676	35	27	ino1(+448)	inol	towards the middle	In Table 6 of the Arnvig et al, 2011 paper
3649468..3649502	35	27	ctpC(-1024)	Rv3268	Towards the beginning	
916241..916275	35	28	(+36)	Rv0822c	beginning	In TableS2 of Arnvig 2011.
754760..754793	34	30	atsD(-1344)	Rv0659c	towards the middle	In TableS2 of Arnvig 2011.
3326014..3326063	50	31	(+80)	Rv2970A	middle/close to the end of the gene	
4100814..4100913	167	32	None	Intergenic	Between Rv3661 and Rv3662c	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv3661 and Rv3662c as flanking CDSs. They annotate it as MTS2823.
4079755..4079811	57	35	None	Some on Rv3640c and some on intergenic region	at the beginning of Rv3640c and I little on the intergenic region between Rv3640c and fic	IG2805 from Table S3 reads mapped to intergenic regions Arnvig 2011

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
3826433..3826467	35	37	None	Rv3407	towards the end of the gene. Also towards the end of "Phd YefM; Region: PhdYeFM; c109153"	In Table 3 of the Arnvig et al, 2011 paper
1816201..1816234	34	46	pykA(+12)	pykA	close to the beginning	In TableS2 of Arnvig 2011. Rv1617
1824726..1824770	45	55	cydA(+1117)	cydA	towards the end	In Table 4 of the Arnvig et al, 2011 paper
4352944..4352979	36	62	(-31)	Intergenic	between esxA and Rv3876	IG3002 from Table S3 reads mapped to intergenic regions Arnvig 2011
2314134..2314170	37	64	rpsR(+183)	rpsR	towards the end	In TableS2 of Arnvig 2011. Rv0055
1275613..1275674	62	70	None	Intergenic	close to the end of Rv1147. Between Rv1147 and Rv1148c	In Table 10 of the Arnvig et al, 2011 paper. sRNAs with Rv1147 and Rv1148c as flanking CDSs. They annotate it as MTS0903.
3468051..3468090	40	71	ssr(+329)	ssr	towards the middle	From Tuberculist Stable RNA list.
4376405..4376472	68	77	(+3980)	Rv3894c	close to the end	
913457..913523	67	92	phoT(+731)	phoT	at the end of the gene, a little over between phoT and phoY2	In TableS2 of Arnvig 2011. Rv0820
3023949..3023988	40	113	ideR(+384)	ideR	towards the middle	In TableS2 of Arnvig 2011. Rv2711
2404494..2404549	56	123	(-2772)	some on Rv2144c and some on intergenic region	at the beginning of Rv2144c and overlapping onto intergenic region between Rv2144c and wag31	
2844174..2844207	34	128	fas(+5125)	fas	towards the middle	In TableS2 of Arnvig 2011. Rv2524c.

Chromosome Region	Length	Coverage	Gene Distance	Location on geneome	Location on gene	Also Found In
1960700..1960781	82	129	(-909)	Intergenic and on Rv1734c	Between Rv1733c and Rv1734c. Some of it overlaps with the end of Rv1734c	Probably MTS1338 from Table 10 of Arnvig 2011 paper, although the paper lists it as having 1733c and Rv1735c as flanking sequences, which excludes Rv1734c.
4099412..4099475	64	152	(-4498)	Intergenic	between Rv3660c and Rv3661	From Tuberculist Stable RNA list, (See Arnvig and Young, 2009; DiChiara et al., 2010. Also in Arnvig 2011 Table 10 as MTS2822
80258..80292	35	165	glyA(-1362)	glyA	Towards the beginning	In TableS2 of Arnvig 2011. Rv0070c
1473259..1473332	74	409	rrs(+1413)	rrs	very close to the end of the gene	
4252929..4252971	43	424	.	Intergenic	between fadE35 and Rv3798. Closer to the beginning of Rv3798	IG2937 from Table S3 reads mapped to intergenic regions Arnvig 2011
1342599..1342670	72	504	0	IS1081-2	Towards the beginning, IS1081-2 overlaps Rv119c except for the beginning	
1469528..1469574	47	555	murA(-747)	IS1557-2	close to the beginning of IS1557-2	IG1037 from Table S3 reads mapped to intergenic regions Arnvig 2011

References

1. Rutherford, K., et al., *Artemis: sequence visualization and annotation.* Bioinformatics, 2000. **16**(10): p. 944-945.
2. Arnvig, K.B. and D.B. Young, *Identification of small RNAs in Mycobacterium tuberculosis.* Molecular microbiology, 2009. **73**(3): p. 397-408.
3. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS Pathog, 2011. **7**(11): p. e1002342.
4. DiChiara, J.M., et al., *Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis.* Nucleic Acids Research, 2010. **38**(12): p. 4067-4078.
5. Lew, J.M., et al., *TubercuList–10 years after.* Tuberculosis, 2011. **91**(1): p. 1-7.

APPENDIX J

Appendix J: Secondary structure predictions of regions of rRNA genes not represented by an aligned read. RNAfold web server was used to predict the secondary structure of the regions of rRNA genes not represented by an aligned read. Table shows the chromosome region evaluated, along with the minimum free energy (MFE), free energy of the thermodynamic ensemble, as well as the ensemble diversity for each region.

Chromosome Region	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
1471846:1471910	-16.3	-17.52	10.81
1472010:1472052	-14	-14.35	1.96
1472122:1472172	-15.3	-15.68	3.95
1472241:1472387	-58.4	-60	27.15
1472433:1472603	-64.7	-61.17	11.84
1472704:1472988	-107.9	-112.06	33.52
1473024:1473036	0	-0.24	1.29
1473135:1473382	-93.4	-95.99	45.48
1473658:1473753	-26.8	-27.22	26.21
1473813:1473886	-18.9	-17.59	10.38
1473974:1474015	-12.1	-10.23	7
1474151:1474164	-2.4	-1.4	0.69
1474242:1474376	-40.6	-40.95	36.93
1474417:1474756	-138.1	-131.23	39.38
1474792:1474805	-1.2	-0.43	1.52
1474978:1475070	-33.3	-31.74	9.87
1475129:1475174	-11.8	-10.24	13.63
1475209:1475552	-110.2	-110.71	110.38
1475653:1475755	-27.1	-25.55	19.83
1475807:1476054	-73.2	-73.62	23.28
1476136:1476144	0	-0.04	0.28
1476191:1476207	-0.5	-0.17	0.86

Chromosome Region	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
1476272:1476331	-17.5	-17.7	8.54
1476371:1476624	-98.7	-95.56	71.88
1476709:1476795	-20.9	-21.05	16.32

APPENDIX K

Appendix K: Secondary structure predictions of sRNA fragments found in the 100 ng sample. RNAfold web server was used to predict the secondary structure of the sRNA fragments found in the 100 ng sample. Table shows the length, coverage, chromosome region, and location on the H37Rv genome. For the secondary structure prediction, the table also displays the minimum free energy (MFE), free energy of the thermodynamic ensemble, as well as the ensemble diversity in each fragment.

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
34	5	4198611..4198644	Between Rv3750c and Rv3751	Intergenic	-10.60	-10.83	0.97
34	5	4262779..4262812	fadD32	Towards the beginning	-8.70	-9.25	2.24
34	5	4334567..4334600	gltB	towards the middle	-7.70	-8.93	8.56
34	5	1564447..1564480	gmk	Towards the beginning	-8.90	-9.74	8.73
34	5	601708..601741	hemA	towards the end	-6.80	-6.90	0.54
34	5	2289878..2289911	lipT	towards the end	-10.60	-11.11	6.58
34	5	4369747..4369780	mycP2	Towards the beginning	-5.10	-5.56	2.71
34	5	303550..303583	nirB	683 in	-11.70	-12.32	1.45
34	5	2827612..2827645	orn	towards the end	-10.20	-10.52	2.04
34	5	337304..337337	PE_PGRS4	close to the middle	-7.10	-8.00	8.47
34	5	1577190..1577223	Rv1401	middle	-12.10	-12.23	0.87
34	5	2055657..2055690	Rv1813c	at the end of the gene, and a little over the region between Rv1813c and Rv1812c	-5.90	-6.96	6.58

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
34	5	2239323..2239356	Rv1996	towards the middle Rv1996 is before ctpF	-9.90	-10.50	2.13
34	5	2978147..2978180	Rv2655c	towards the middle	-8.10	-8.83	4.58
34	5	4302868..4302901	Rv3828c	towards the end	-7.50	-7.81	1.22
34	5	4398690..4398723	Rv3910	middle	-7.90	-9.06	9.44
34	5	3022467..3022500	sigB	beginning	-7.20	-7.48	3.13
35	5	1866320..1866354	argC	towards the end	-5.70	-6.29	7.19
35	5	3651551..3651585	ctpC	middle. "Cation transport ATPase"	-15.30	-15.32	0.11
35	5	782891..782925	fusA1	Towards the beginning	-5.00	-5.37	8.24
35	5	873531..873565	hemH	Towards the beginning	-9.20	-9.66	6.07
35	5	1736341..1736375	Inter geneic	Intergenic region between Rv1535 and ileS	-9.70	-9.81	1.89
35	5	2056276..2056310	Inter geneic	Intergenic region between Rv1813c and erg3	-10.20	-10.43	4.16
35	5	2519178..2519212	kasA	towards the end	-13.50	-13.89	2.12
35	5	2672020..2672054	mbtB	towards the end	-5.20	-5.84	2.87
35	5	474296..474330	ndhA	towards the end	-6.60	-7.98	10.46
35	5	1099314..1099348	pepD	Towards the beginning	-9.00	-10.00	5.79
35	5	1511851..1511885	radE14	towards the end	-6.40	-6.88	7.32
35	5	2500503..2500537	rnpB	towards the end	-11.60	-12.13	4.86
35	5	119170..119204	Rv0102	towards the end	-5.40	-5.60	1.09
35	5	194242..194276	Rv0165c	towards the end	-6.30	-6.30	3.75
35	5	269454..269488	Rv0225	towards the end	-3.30	-4.33	5.55
35	5	348462..348496	Rv0284	towards the middle	-5.80	-6.52	4.36
35	5	884318..884352	Rv0790c	towards the middle	-10.20	-11.39	9.15
35	5	1901178..1901212	Rv1676	Towards the beginning	-10.50	-11.22	9.30
35	5	2032991..2033025	Rv1795	middle	-3.70	-5.04	9.72
35	5	2477897..2477931	Rv2212	towards the middle	-10.40	-10.97	4.04
35	5	2502565..2502599	Rv2228c	Towards the beginning	-8.50	-9.58	5.98
35	5	2783947..2783981	Rv2477c	very close to the beginning of the gene	-4.80	-5.49	4.06
35	5	3020895..3020929	Rv2707	towards the middle.	-7.40	-8.10	1.82

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
35	5	3900311..3900345	Rv3480c	close to the beginning	-7.10	-8.27	7.44
35	5	3953877..3953911	Rv3517	towards the middle	-9.00	-9.65	2.24
35	5	4221473..4221507	Rv3776	towards the beginning. At the beginning of "domain of unknown function DUF 22" fragments flanking both ends of this domain	-1.00	-2.48	8.16
35	5	4231808..4231842	Rv3785	middle	-6.70	-7.85	7.78
35	5	274730..274764	Rv4006	Towards the end	-10.30	-10.60	3.45
35	5	2819339..2819373	scoA	towards the middle	-7.10	-7.78	8.21
38	5	2404447..2404484	Rv2144c	close to the beginning. Very close to another fragment (-2772)	-13.80	-13.89	0.41
39	5	1797489..1797527	nadC	Towards the beginning	-11.20	-12.23	8.08
43	5	2500383..2500425	Inter geneic	very close to the end of Rv2226 and the end of rnpB (misc_RNA)	-18.20	-19.47	3.78
45	5	2952672..2952716	Rv2626c	towards the middle	-6.80	-8.07	8.50
45	5	2953710..2953754	Rv2627c	towards the end	-8.50	-9.56	6.05
46	5	1516874..1516919	Rv1349	towards the middle	-9.90	-11.01	7.95
49	5	1314313..1314361	pks3	close to the middle	-15.30	-17.12	6.77
54	5	251787..251840	pckA	Very close to the beginning of the gene	-14.00	-15.59	11.47
64	5	776008..776071	mmpL5	close to the end	-13.60	-14.41	9.77
34	6	2162128..2162161	aceAb	in the middle	-2.80	-3.57	5.04
34	6	420777..420810	dnaK	middle	-10.90	-11.49	3.05
34	6	326352..326385	fadE6	towards the end	-10.60	-11.10	3.05
34	6	3516906..3516939	nuoG	Towards the beginning	-11.70	-12.58	10.84
34	6	3894312..3894345	PE31	towards the end	-9.40	-9.87	2.64
34	6	2276245..2276278	pfkB	Towards the beginning	-7.10	-7.64	3.33
34	6	2593248..2593281	rocE	towards the middle	-7.00	-7.76	8.79
34	6	88211..88244	Rv0079	Towards the beginning	-5.10	-5.71	2.48
34	6	215774..215807	Rv0185	Towards the beginning	-5.80	-6.67	4.93
34	6	352444..352477	Rv0289	close to the middle	-11.10	-11.23	1.27
34	6	631410..631443	Rv0538	towards the end	-7.00	-7.33	5.18

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
34	6	3497349..3497382	Rv3131	towards the end	-8.30	-8.69	1.24
35	6	3666695..3666729	accA3	Towards the beginning	-12.30	-13.14	9.40
35	6	3404859..3404893	ctaD	very close to the beginning of the gene	-6.00	-7.22	4.37
35	6	2483389..2483423	dlaT	towards the end	-6.20	-6.91	4.98
35	6	2842511..2842545	fas	towards the end	-12.90	-13.14	2.05
35	6	2150873..2150907	Inter geneic	Intergenic region between Rv1903 and Rv1904. Closer to Rv1904	-9.10	-9.12	0.13
35	6	3623406..3623440	lpqB	Towards the end of the gene.	-5.50	-5.71	5.29
35	6	2252256..2252290	otsB1	Towards the beginning	-14.60	-15.34	2.95
35	6	4258521..4258555	pks13	middle	-7.90	-8.44	4.61
35	6	1004881..1004915	prrB	towards the middle	-18.90	-19.48	10.51
35	6	707305..707339	Rv0613c	towards the end	-8.40	-9.29	4.10
35	6	1669140..1669174	Rv1478 and intergenic	at the end of the the gene into the intergenic region between Rv1478 and moxR1	-8.10	-8.48	3.88
35	6	2445035..2445069	Rv2182c	towards the middle	-9.50	-9.86	2.69
35	6	3492964..3492998	Rv3127	towards the end of the gene. Very close to another fragment (-899)	-4.50	-6.14	6.38
35	6	4222194..4222228	Rv3776	towards the end. At the end of "domain of unknown function DUF 222"	0.00	-0.54	2.14
35	6	595932..595966	serB1	towards the middle	-6.10	-7.38	11.71
38	6	346439..346476	Rv0284	Towards the beginning	-12.30	-13.01	8.42
38	6	4056007..4056044	Rv3616c	towards the end	-7.60	-8.39	12.01
42	6	2239872..2239913	ctpG	middle	-9.80	-10.36	11.47
42	6	1963271..1963312	narX	middle	-11.20	-11.90	7.46
44	6	1313412..1313455	Inter geneic	between Rv1179 and pks3	-13.10	-13.77	12.57
50	6	3084849..3084898	Rv2777c	close to the middle	-13.50	-14.05	5.57
66	6	3493021..3493086	Rv3127	towards the end	-21.20	-22.00	7.81
69	6	2952398..2952466	Rv2625c	close to the beginning	-23.30	-23.58	1.28
34	7	662451..662484	nrdZ	middle	-3.10	-3.63	4.79

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
34	7	2073625..2073658	Rv1828	towards the end	-10.60	-10.81	5.83
35	7	3806205..3806239	acrA1	towards the middle	-11.80	-11.89	0.54
35	7	1222477..1222511	desA2	towards the middle	-7.70	-8.18	3.42
35	7	1227126..1227160	fumC	Towards the beginning	-7.90	-8.97	9.24
35	7	7449..7483	gyrA	Towards the beginning	-10.20	-10.31	0.76
35	7	2017524..2017558	Inter geneic	Intergenic region between malQ and Rv1782	-12.30	-12.64	1.14
35	7	349084..349118	Rv0284	towards the end	-5.50	-5.87	1.24
35	7	867824..867858	Rv0774c	middle	-8.20	-9.72	5.39
35	7	2278025..2278059	Rv2030c	Towards the beginning	-10.40	-10.52	0.85
35	7	4346889..4346923	Rv3870	Towards the beginning	-7.30	-8.31	8.51
35	7	25117..25151	TB39.8	Towards the beginning	-3.40	-3.50	1.87
39	7	1113102..1113140	Rv0996	towards the middle	-15.10	-15.85	1.74
40	7	917498..917537	Rv0823c	Towards the beginning	-7.00	-7.44	9.08
54	7	3468268..3468321	ssr	Towards the beginning. Ssr is classified as a misc RNA	-10.70	-11.91	10.41
64	7	4024199..4024262	ispF	Towards the beginning	-20.50	-22.40	14.95
81	7	340528..340608	PPE3	towards the end	-28.00	-28.83	19.07
34	8	2955616..2955649	Inter geneic between Rv2628 and Rv2629	inter genic	-3.00	-3.77	7.03
34	8	2484655..2484688	lipB	Towards the beginning	-14.60	-14.89	1.02
34	8	4220050..4220083	lipE	towards the middle	-3.40	-4.19	4.46
34	8	1796908..1796941	nadB	towards the end	-11.30	-11.59	3.12
34	8	1845595..1845628	uvrA	towards the middle	-3.70	-4.16	4.18
35	8	3499671..3499705	devR	towards the middle	-4.30	-4.71	3.80
35	8	88506..88540	hycD	towards the middle	-8.90	-9.26	4.75
35	8	1150652..1150686	kdpD	towards the middle	-10.80	-11.19	4.87
35	8	1796054..1796088	nadB	Towards the beginning	-10.90	-10.99	0.74
35	8	851642..851676	phoP	Towards the beginning	-3.10	-4.14	9.50
35	8	2979238..2979272	Rv2657c	Towards the beginning	-10.90	-11.16	2.53
35	8	3496869..3496903	Rv3131	towards the middle	-13.20	-14.09	5.66

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
35	8	3606667..3606701	Rv3229c	towards the middle	-5.70	-6.89	4.74
35	8	4029729..4029763	Rv3587c / intergenic	At the beginning of Rv3587c, over lapping a little into the intergenic region between Rv3587c and Rv3588c	-6.20	-6.89	7.14
35	8	1235415..1235449	xseA	at the beginning	-8.70	-9.08	2.05
40	8	3498580..3498619	devS	towards the middle	-10.00	-10.62	15.09
40	8	1132353..1132392	pks16	middle	-10.80	-11.34	2.27
42	8	353501..353542	Rv0290	towards the middle	-18.20	-18.57	1.09
44	8	352199..352242	Rv0289	close to the beginning of Rv0289	-10.30	-11.55	9.61
47	8	2278606..2278652	hspX	towards the end	-10.20	-10.41	1.90
66	8	4359090..4359155	Rv3879c	Towards the beginning	-22.30	-23.45	12.32
34	9	1197094..1197127	Rv1072	at the end	-5.10	-5.76	4.41
34	9	3589569..3589602	Rv3212	Towards the beginning	-10.20	-10.78	3.19
35	9	1340943..1340977	esxK	at the end of Rv1247c into the intergenic region between esxK and esxL	-9.30	-9.64	1.46
35	9	2103054..2103088	Inter geneic	Intergenic region between ndh and Rv1855c. Very close to the beginning of ndh	-5.90	-7.00	10.23
35	9	3548920..3548954	Inter geneic	13 bp after the end of Rv3179	-6.10	-6.30	2.72
35	9	3246358..3246392	ppsA	Towards the beginning	-8.60	-0.96	6.27
35	9	356191..356225	Rv0292	Towards the beginning	-9.40	-9.81	9.10
44	9	353322..353365	Rv0290	Towards the beginning	-13.10	-13.91	3.52
49	9	1965772..1965820	Rv1738	middle	-18.90	-19.58	7.29
64	9	2256537..2256600	Inter geneic	very close to the end of the gene Rv2008c. between fdxA and Rv2008c	-27.00	-27.33	2.74
65	9	2655876..2655940	cfp2	towards the middle	-20.70	-22.70	17.70
77	9	340750..340826	PPE3	closer to the end of PPE3 than PPE3(+1164)	-45.30	-46.46	5.59

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
34	10	2150918..2150951	Inter geneic	before the beginning of the gene Rv1904	-2.00	-2.77	6.57
34	10	2954155..2954188	Rv2627c	middle	-8.80	-9.73	3.85
35	10	2463741..2463775	ctaC	close to the end	-10.80	-10.93	0.44
35	10	2847337..2847371	fas	Towards the beginning	-7.00	-7.67	1.70
35	10	1413108..1413142	Inter geneic	Intergenic region between Rv1264 and Rv1265. Very close to the end of Rv1264	-9.90	-11.00	13.05
35	10	2122411..2122445	lldD2	middle	-4.30	-4.99	8.02
35	10	3758308..3758342	PPE56	Towards the end of the gene.	-14.90	-15.58	1.98
34	11	32182..32215	Rv0029	Towards the beginning	-2.60	-0.32	2.89
34	11	1998195..1998228	Rv1765c	towards the middle	0.00	-0.45	1.91
35	11	1466925..1466959	atpD	towards the end	-11.70	-11.90	0.83
35	11	2517353..2517387	fabD	towards the middle	-10.70	-11.11	2.13
35	11	1199075..1199109	fadA3	Towards the beginning	-6.40	-6.69	4.01
35	11	2845154..2845188	fas	middle	-12.50	-13.08	2.01
35	11	1220452..1220486	Inter geneic	inter genic close to the beginning of glyA. Between coaA and glyA.	-8.00	-8.84	10.89
35	11	4056753..4056787	Inter geneic	between Rv3616c and ephA. Closer to the beginning of Rv3616c.	-9.90	-10.52	2.96
39	11	378204..378242	Rv0309	towards the middle	-7.80	-8.58	8.15
56	11	3093770..3093825	Inter geneic	between rpsO and ribF	-11.20	-11.88	4.78
66	11	1485598..1485663	Rv1322A	towards the middle	-14.60	-16.75	16.75
78	11	918303..918380	desA1	towards the middle	-34.10	-35.16	4.61
35	12	1389211..1389245	Rv1247c	At the beginning of the gene, with one bp in the intergenic region between Rv1247 and kgd	-5.90	-6.54	4.07
35	12	2626202..2626236	Rv2345	towards the middle	-15.00	-15.71	2.61
35	12	3606901..3606935	Rv3229c	Towards the beginning	-14.40	-15.23	8.32
45	12	2236394..2236438	ctpG	middle	-13.10	-13.39	6.29
52	12	1197139..1197190	Inter geneic	between Rv1072 and Rv1073	-12.10	-13.10	13.68

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
61	12	1169366..1169426	IS1081-1	close to the beginning of insertion sequence of IS1081	-21.40	-22.42	1.59
34	13	2102731..2102764	ndh	Towards the beginning	-1.90	-3.00	7.13
35	13	1419706..1419740	lprA	very close to the beginning of the gene	-4.40	-5.12	6.19
35	13	2646753..2646787	Rv2365c	very close to the end of the gene	-11.50	-11.98	5.03
35	13	2756411..2756445	Rv2455c	Towards the beginning	-7.20	-7.78	10.47
35	13	2959794..2959828	Rv2633c and intergenic	At the beginning of Rv2633c and overlapping onto the intergenic region between Rv2633c and PE_PGRS46	-1.50	-3.05	8.96
35	13	4119573..4119607	Rv3679	towards the end. Within a predicted ATPase involved in chromosome partitioning	-14.00	-14.54	6.37
39	13	1376990..1377028	Rv1234	Towards the beginning. Rv1234 is before lpqY	-5.20	-6.18	7.85
92	13	4168191..4168282	Inter geneic	Between Rv3722c and Rvnt41. Rvnt41 is between Rv3722c and Rv3723.	-38.90	-40.30	19.32
35	14	1666920..1666954	Inter geneic	Intergenic region between Rv1476 and Rv1477. Closer to Rv1477	-9.50	-9.97	1.45
35	14	2728347..2728381	Inter geneic	between PE25 and Rc2432c	-4.90	-5.03	1.17
35	14	498846..498880	Rv0412c	towards the middle	-1.90	-2.50	3.04
35	14	2782771..2782805	Rv2477c	Towards the end	-8.50	-9.05	4.10
41	14	2951533..2951573	Rv2625c	towards the end	-14.90	-15.01	0.81
49	14	3496579..3496627	Rv3131	towards the beginning, very close to another segment (+2)	-12.70	-12.91	6.60
66	14	355733..355798	mycP3	close to the end	-17.00	-17.50	5.96
34	15	3153388..3153421	efpA	towards the end	-1.50	-2.61	8.28
35	15	1166640..1166674	Rv1043c	Towards the beginning	-6.50	-7.70	11.91

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
36	15	2624292..2624327	Rv2345	Towards the beginning. At the end of a "domain of unknown function (DUF477)"	-8.00	-9.19	3.12
45	15	1080094..1080138	ctpV	towards the middle	-8.10	-8.92	5.41
45	15	2391717..2391761	cysS	towards the middle	-7.80	-9.58	16.90
34	16	1993213..1993246	Rv1760	Towards the beginning	-8.70	-9.49	6.87
34	16	2775535..2775568	Rv2472	at the end	-7.50	-7.94	8.96
35	16	3496375..3496409	Inter geneic	Very close to the beginning of tgs1. In the intergenic region between tgs1 and Rv3131	-6.00	-6.52	6.42
45	16	4314343..4314387	bfrB	towards the middle	-7.30	-8.23	11.96
74	16	2949865..2949938	TB31.7	towards the middle	-18.10	-19.49	11.13
35	17	3540695..3540729	hpx	Very close to the beginning of the gene	-9.60	-10.16	8.93
35	17	2638357..2638391	PPE40	towards the middle	-3.10	-3.67	4.38
35	17	2280314..2280348	Rv2033c	towards the end	-11.40	-12.50	5.42
35	17	2885351..2885385	Rv2565	towards the middle	-14.90	-14.95	0.20
52	17	2177096..2177147	fadD31	Very close to the beginning of the gene	-8.90	-10.26	16.65
62	17	1964660..1964721	narK2	towards the middle	-17.60	-19.07	17.70
34	18	1114988..1115021	Rv0998	towards the middle	-11.50	-12.44	3.27
35	18	2274682..2274716	Rv2028c	towards the end	0.00	-0.79	3.85
45	18	571794..571838	lprQ	Towards the beginning	-17.10	-17.57	1.85
35	19	3153949..3153983	efpA	towards the middle	-11.10	-11.68	2.91
35	19	65446..65480	Inter geneic	between Rv0061 and celA1, closer to the end of Rv0061	-6.80	-7.37	7.90
55	19	3497199..3497253	Rv3131	towards the middle	-15.70	-16.78	13.47
73	20	4100706..4100778	Inter geneic	behind another inter geneic fragment: 4100814. between Rv3661 and Rv3662c.	-22.90	-23.54	7.23
89	20	3468146..3468234	ssr	towards the middle of ssr	-33.20	-34.80	12.54
34	21	2279214..2279247	acg	Towards the beginning	-10.80	-11.12	1.50

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
35	23	4404606..4404640	Rv3916c	towards the end	-5.80	-5.99	2.28
46	23	3496804..3496849	Rv3131	close to the middle, several other fragments on this gene	-15.10	-15.50	5.03
39	24	1965823..1965861	Rv1738	towards the end. Next to another fragment (+115)	-9.40	-11.25	5.98
49	24	409680..409728	iniB	close to the beginning	-18.10	-19.38	18.12
55	24	3496517..3496571	Rv3131	2bp into the beginning	-12.60	-12.88	6.80
80	24	1773938..1774017	Rv1566c	Very close to the beginning of the gene	-43.30	-44.21	6.54
43	26	2767686..2767728	Rv2465c	close to the beginning	-14.00	-14.97	6.48
35	27	50642..50676	inol	towards the middle	-4.20	-4.74	5.58
35	27	3649468..3649502	Rv3268	Towards the beginning	-11.10	-11.90	8.11
35	28	916241..916275	Rv0822c	beginning	-6.90	-7.55	5.45
34	30	754760..754793	Rv0659c	towards the middle	-6.70	-7.47	8.44
50	31	3326014..3326063	Rv2970A	middle/close to the end of the gene	-12.20	-12.80	4.84
167	32	4100814..4100913	Inter geneic	Between Rv3661 and Rv3662c	-59.00	-60.11	6.75
57	35	4079755..4079811	Some on Rv3640c and some on intergenic region	at the beginning of Rv3640c and I little on the intergenic region between Rv3640c and fic	-20.60	-21.08	1.84
35	37	3826433..3826467	Rv3407	towards the end of the gene. Also towards the end of "Phd YefM; Region: PhdYeFM; c109153"	-4.20	-4.85	4.22
34	46	1816201..1816234	pykA	close to the beginning	-6.50	-7.04	3.64
45	55	1824726..1824770	cydA	towards the end	-12.60	-13.20	13.41
36	62	4352944..4352979	Inter geneic	between esxA and Rv3876	-4.90	-5.34	5.66
37	64	2314134..2314170	rpsR	towards the end	-10.30	-11.02	9.69
62	70	1275613..1275674	Inter geneic	close to the end of Rv1147. Between Rv1147 and Rv1148c	-36.60	-37.37	5.94
40	71	3468051..3468090	ssr	towards the middle	-6.10	-6.92	10.36
68	77	4376405..4376472	Rv3894c	close to the end	-39.80	-40.15	1.25

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
67	92	913457..913523	phoT	at the end of the gene, a little over between phoT and phoY2	-27.90	-28.49	3.63
40	113	3023949..3023988	ideR	towards the middle	-14.80	-15.37	2.10
56	123	2404494..2404549	some on Rv2144c and some on intergenic region	at the beginning of Rv2144c and overlapping onto intergenic region between Rv2144c and wag31	-9.20	-10.63	14.82
34	128	2844174..2844207	fas	towards the middle	-10.80	-11.49	1.25
82	129	1960700..1960781	Inter geneic and on Rv1734c	Between Rv1733c and Rv1734c. Some of it overlaps with the end of Rv1734c	-25.90	-27.54	20.12
64	152	4099412..4099475	Inter geneic	between Rv3660c and Rv3661	-24.50	-25.52	18.79
35	165	80258..80292	glyA	Towards the beginning	-10.90	11.15	6.42
74	409	1473259..1473332	rrs	very close to the end of the gene	-26.30	-27.07	15.92
43	424	4252929..4252971	Inter geneic	between fadE35 and Rv3798. Closer to the beginning of Rv3798	-7.40	-8.23	10.42
72	504	1342599..1342670	IS1081-2	Towards the beginning, IS1081-2 overlaps Rv119c except for the beginning	-35.60	-35.82	0.82
47	555	1469528..1469574	IS1557-2	close to the beginning of IS1557-2	-6.40	-7.03	2.39

Length	Coverage	Chromosome region	Location on genome	Location on gene	Minimum free energy (kcal/mol)	Free Energy of thermodynamic ensemble (kcal/mol)	Ensemble diversity
35	5	3900311..3900345	Rv3480c	close to the beginning	-7.10	-8.27	7.44
35	5	3953877..3953911	Rv3517	towards the middle	-9.00	-9.65	2.24
35	5	4221473..4221507	Rv3776	towards the beginning. At the beginning of "domain of unknown function DUF 22" fragments flanking both ends of this domain	-1.00	-2.48	8.16
35	5	4231808..4231842	Rv3785	middle	-6.70	-7.85	7.78
35	5	274730..274764	Rv4006	Towards the end	-10.30	-10.60	3.45
35	5	2819339..2819373	scoA	towards the middle	-7.10	-7.78	8.21
38	5	2404447..2404484	Rv2144c	close to the beginning. Very close to another fragment (-2772)	-13.80	-13.89	0.41
39	5	1797489..1797527	nadC	Towards the beginning	-11.20	-12.23	8.08
43	5	2500383..2500425	Inter geneic	very close to the end of Rv2226 and the end of rnpB (misc_RNA)	-18.20	-19.47	3.78
45	5	2952672..2952716	Rv2626c	towards the middle	-6.80	-8.07	8.50
45	5	2953710..2953754	Rv2627c	towards the end	-8.50	-9.56	6.05
46	5	1516874..1516919	Rv1349	towards the middle	-9.90	-11.01	7.95
49	5	1314313..1314361	pks3	close to the middle	-15.30	-17.12	6.77
54	5	251787..251840	pckA	Very close to the beginning of the gene	-14.00	-15.59	11.47
64	5	776008..776071	mmpL5	close to the end	-13.60	-14.41	9.77
34	6	2162128..2162161	aceAb	in the middle	-2.80	-3.57	5.04
34	6	420777..420810	dnaK	middle	-10.90	-11.49	3.05
34	6	326352..326385	fadE6	towards the end	-10.60	-11.10	3.05
34	6	3516906..3516939	nuoG	Towards the beginning	-11.70	-12.58	10.84
34	6	3894312..3894345	PE31	towards the end	-9.40	-9.87	2.64
34	6	2276245..2276278	pfkB	Towards the beginning	-7.10	-7.64	3.33
34	6	2593248..2593281	rocE	towards the middle	-7.00	-7.76	8.79
34	6	88211..88244	Rv0079	Towards the beginning	-5.10	-5.71	2.48
34	6	215774..215807	Rv0185	Towards the beginning	-5.80	-6.67	4.93
34	6	352444..352477	Rv0289	close to the middle	-11.10	-11.23	1.27
34	6	631410..631443	Rv0538	towards the end	-7.00	-7.33	5.18

APPENDIX L

Appendix L: Permissions obtained. Copies of permissions obtained for Figures 1.1 and 1.3 in Chapter I.

Figure 1.1

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Apr 11, 2014

This is a License Agreement between Sarah W Sheldon ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3344830708466
License date	Mar 09, 2014
Licensed content publisher	Oxford University Press
Licensed content publication	Glycobiology
Licensed content title	Mycobacterial lipoarabinomannan: An extraordinary lipoheteroglycan with profound physiological effects:
Licensed content author	Delphi Chatterjee, Kay-Hooi Khoo
Licensed content date	02/01/1998
Type of Use	Thesis/Dissertation
Institution name	None
Title of your work	Identification of small extracellular RNA fragments of Mycobacterium tuberculosis
Publisher of your work	n/a
Expected publication date	Apr 2014
Permissions cost	0.00 USD
Value added tax	0.00 USD
TotalTotal	**0.00 USD**
TotalTotal	**0.00 USD**
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of

all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501245908.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:
customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Figure 1.1

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 11, 2014

This is a License Agreement between Sarah W Sheldon ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Sarah W Sheldon
Customer address	3185 Rampart Road Fort Collins, CO 80523
License number	3344830527170
License date	Mar 09, 2014
Licensed content publisher	Elsevier
Licensed content publication	Chemistry & Biology
Licensed content title	The Methyl-Branched Fortifications of*Mycobacterium tuberculosis*
Licensed content author	David E Minnikin,Laurent Kremer,Lynn G Dover,Gurdyal S Besra
Licensed content date	May 2002
Licensed content volume number	9
Licensed content issue number	5
Number of pages	9
Start Page	545
End Page	553
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of your thesis/dissertation	Identification of small extracellular RNA fragments of Mycobacterium tuberculosis
Expected completion date	Apr 2014
Estimated size (number	170

of pages)	
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. Byclicking "accept" in connection with completing this licensingtransaction, you agree that the following terms and conditions apply to thistransaction (along with the Billing and Payment terms and conditionsestablished by Copyright Clearance Center, Inc. ("CCC"), at the timethat you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementionedmaterial subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example,figures) has appeared in our publication with credit or acknowledgement toanother source, permission must also be sought from that source. If suchpermission is not obtained then that material may not be included in yourpublication/copies. Suitable acknowledgement to the source must be made, eitheras a footnote or in a reference list at the end of your publication, asfollows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title ofarticle / title of chapter, Pages No., Copyright (Year), with permission fromElsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit -"Reprinted from The Lancet, Vol. number, Author(s), Title of article, PagesNo., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media forwhich permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures andillustrations may be altered/adapted minimally to serve your work. Any otherabbreviations, additions, deletions and/or any other alterations shall be madeonly with prior written authorization of Elsevier Ltd. (Please contact Elsevierat permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived inthis instance, please be advised that your future requests for Elseviermaterials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specificallygranted in the combination of (i) the license details provided by you andaccepted in the course of this licensing transaction, (ii) these terms andconditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rightslicensed immediately upon issuance of the license at the end of the licensingprocess for the transaction, provided that you have disclosed complete andaccurate details of your proposed use, no license is finally effective unlessand until full payment is received from you (either by publisher or by CCC) asprovided in CCC's Billing and Payment terms and conditions. If fullpayment is not received on a timely basis, then any license preliminarilygranted shall be deemed automatically revoked and shall be void as if nevergranted. Further, in the event that you breach any of these terms andconditions or any of CCC's Billing and Payment terms and conditions, thelicense is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of thematerials beyond the scope of an unrevoked license, may constitute copyrightinfringement and publisher reserves the right to take any and all action toprotect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respectto the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher andCCC, and their respective officers, directors, employees and agents, from andagainst any and all claims arising out of your use of the licensed materialother than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not besublicensed, assigned, or transferred by you to any other person withoutpublisher's written permission.

12. No Amendment Except in Writing: This license may not be amended exceptin a writing signed by both parties (or, in the case of publisher, by CCC onpublisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any termscontained in any purchase order, acknowledgment, check endorsement or otherwriting prepared by you, which terms are inconsistent with these terms andconditions or CCC's Billing and Payment terms and conditions. These termsand conditions, together with CCC's Billing and Payment terms and conditions(which are incorporated herein), comprise the entire agreement between you andpublisher (and CCC) concerning this licensing transaction. In the eventof any conflict between your obligations established by these terms andconditions and those established by CCC's Billing and Payment terms andconditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny thepermissions described in this License at their sole

discretion, for any reasonor no reason, with a full refund payable to you. Notice of such denialwill be made using the contact information provided by you. Failure toreceive such notice will not alter or invalidate the denial. In no eventwill Elsevier or Copyright Clearance Center be responsible or liable for anycosts, expenses or damage incurred by you as a result of a denial of yourpermission request, other than a refund of the amount(s) paid by you toElsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted fornon-exclusive world **English** rights only unless yourlicense was granted for translation rights. If you licensed translation rightsyou may only translate this content into the languages you requested. Aprofessional translator must perform all translations and reproduce the contentword for word preserving the integrity of the article. If this license is tore-use 1 or 2 figures then permission is granted for non-exclusive world rightsin all languages.

16. **Posting licensed content on any Website**: The followingterms and conditions apply as follows: Licensing material from an Elsevierjournal: All content posted to the web site must maintain the copyright informationline on the bottom of each image; A hyper-text must be included to the Homepageof the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx orthe Elsevier homepage for books at http://www.elsevier.com; CentralStorage: This license does not include permission for a scanned version of thematerial to be stored in a central repository such as that provided byHeron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be includedto the Elsevier homepage at http://www.elsevier.com . Allcontent posted to the web site must maintain the copyright information line onthe bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clausesare applicable: The web site must be password-protected and made available onlyto bona fide students registered on a relevant course. This permission isgranted for 1 year only. You may obtain a new license for future websiteposting.

For journal authors: the following clauses are applicable inaddition to the above: Permission granted is limited to the author acceptedmanuscript version* of your paper.

***Accepted Author Manuscript (AAM) Definition:** An accepted authormanuscript (AAM) is the author's version of the manuscript of an article thathas been accepted for publication and which may include any author-incorporatedchanges suggested through the processes of submission processing, peer review,and editor-author communications. AAMs do not include other publishervalue-added contributions such as copy-editing, formatting, technicalenhancements and (if relevant) pagination.

You are not allowed to download and post the published journal article (whetherPDF or HTML, proof or final version), nor may you scan the printed edition tocreate an electronic version. A hyper-text must be included to the Homepage ofthe journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. As part of our normal production process, you will receive an e-mail noticewhen your article appears on Elsevier's online service ScienceDirect(www.sciencedirect.com). That e-mail will include the article's Digital ObjectIdentifier (DOI). This number provides the electronic link to the publishedarticle and should be included in the posting of your personal version. We askthat you wait until you receive this e-mail and have the DOI to do any posting.

Posting to a repository: Authorsmay post their AAM immediately to their employer's institutional repository forinternal use only and may make their manuscript publically available after thejournal-specific embargo period has ended.

Please also refer to Elsevier'sArticle Posting Policy for further information.

18. **For book authors** the following clauses are applicable in addition tothe above: Authors are permitted to place a briefsummary of their work online only.. You are notallowed to download and post the published electronic version of your chapter,nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors arepermitted to post a summary of their chapter only in their institution'srepository.

20. **Thesis/Dissertation**: If your license is for use in athesis/dissertation your thesis may be submitted to your institution in eitherprint or electronic form. Should your thesis be published commercially, pleasereapply for permission. These requirements include permission for the Libraryand Archives of Canada to supply single copies, on demand, of the completethesis and include permission for UMI to supply single copies, on demand, ofthe complete thesis. Should your thesis be published commercially, pleasereapply for permission.

Elsevier Open Access Terms andConditions

Elsevier publishes Open Access articles in both its Open Access journals andvia its Open Access articles option in subscription journals.

Authors publishing in an Open Access journal or who choose to make theirarticle Open Access in an Elsevier subscription journal select one of thefollowing Creative Commons user licenses, which define how a reader may reusetheir work: Creative Commons Attribution License (CC BY), Creative CommonsAttribution – Non Commercial - ShareAlike (CC BY

NCSA) and Creative Commons Attribution – Non Commercial – No Derivatives (CC BYNC ND)

Terms & Conditions applicable toall Elsevier Open Access articles:

Any reuse of the article must not represent the author as endorsing theadaptation of the article nor should the article be modified in such a way asto damage the author's honour or reputation.

The author(s) must be appropriately credited.

If any part of the material to be used (for example, figures) has appearedin our publication with credit or acknowledgement to another source it is theresponsibility of the user to ensure their reuse complies with the terms andconditions determined by the rights holder.

Additional Terms & Conditionsapplicable to each Creative Commons user license:

CC BY: You may distribute andcopy the article, create extracts, abstracts, and other revised versions,adaptations or derivative works of or from an article (such as a translation),to include in a collective work (such as an anthology), to text or data minethe article, including for commercial purposes without permission from Elsevier

CC BY NC SA: For non-commercialpurposes you may distribute and copy the article, create extracts, abstractsand other revised versions, adaptations or derivative works of or from anarticle (such as a translation), to include in a collective work (such as ananthology), to text and data mine the article and license new adaptations or creationsunder identical terms without permission from Elsevier

CC BY NC ND: For non-commercialpurposes you may distribute and copy the article and include it in a collectivework (such as an anthology), provided you do not alter or modify the article,without permission from Elsevier

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC NDlicense requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Promotional purposes (advertising or marketing)
- Commercial exploitation (e.g. a product forsale or loan)
- Systematic distribution (for a fee or free ofcharge)

Please refer to Elsevier'sOpen Access Policy for further information.

21. **Other Conditions**:

v1.7

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501245907.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:
customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Figure 1.3

Garite, Bibi

From: Mazzullo, Mala
Sent: Monday, March 10, 2014 9:15 AM
To: Garite, Bibi
Subject: FW: CSHL Press Reprint Permission Request Form - G&D

On 3/8/14 11:06 PM, "Reprint" <Reprint@cshl.edu> wrote:

>Default Intro
>Default Intro - line2
>
>
>Name: Sarah Sheldon
>CompanyInstitution: Colorado State University Library Address: 3185
>Rampart Road Library Address (line 2): Building A
>City: Fort Collins
>State (US and Canada): CO
>Country: USA
>Zip: 80523
>Title: Research Associate
>Lab/Department: Microbiology, Immunology, and Pathology. Belisle Lab
>Phone: (970) 494-8874
>Fax: (970) 491-8708
>Email: sarah.w.sheldon@gmail.com
>Title of Publication: Masters Thesis: Identification of Small
>Extracellular RNA Fragments of Mycobacterium tuberculosis
>Authors/Editors: Sarah W. Sheldon, John T. Belisle Date of
>Publication: April 4, 2014
>Publisher: Colorado State University
>Title of CSHLP Journal/Book: Genes and Development Title of
>Article/Chapter: tRNA biology charges to the front CSHL
>Authors/Editors: Eric M. Phizicky and Anita K. Hopper Page Numbers:
>1832-1860 Figure Numbers: 1 Figure Page Numbers: 1833 Copyright Date:
>2010
>Language:
>Territory:
>Format: Print
>Additional comments: I would like to use this figure in the
>Introductory chapter of my master's thesis.

Permission granted by the copyright owner, contingent upon the consent of the original author, provided complete credit is given to the original source and copyright date.

By Christina Kary, Ph.D. 3/12/14
Date

COLD SPRING HARBOR LABORATORY PRESS

1