PNAS
www.pnas.org

Supplementary Information for

Small RNA profiling in *Mycobacterium tuberculosis* identifies MrsI as necessary for an anticipatory iron sparing response

Elias R. Gerrick[a], Thibault Barbier[a], Michael R. Chase[a], Raylin Xu[a], Josie François[a], Vincent H. Lin[a], Matthew Szucs[b], Jeremy M. Rock[a], Rushdy Ahmad[b], Brian Tjaden[c], Jonathan Livny[b], Sarah M. Fortune[a,d]

Sarah M. Fortune
Email: sfortune@hsph.harvard.edu

This PDF file includes:

Supplementary text
Figs. S1 to S5
Tables S1 to S4
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

Supplemental Materials and Methods

Plasmid construction

A list of all plasmids used in this study is located in *SI Appendix, Table S3*. Plasmid pEG79 for creating the *mrsI* deletion in *M. smegmatis* was generated by amplifying the lox-Hyg-lox cassette and the regions upstream and downstream of *M. smegmatis mrsI* using oligos EG629/630, EG627/628, and EG631/632 (Table S4), respectively. These 3 amplicons were then assembled into pEG66 linearized with PmlI using Gibson assembly master mix (NEB). Plasmid pEG101 for inducible *M. smegmatis mrsI* overexpression was generated by amplifying the *mrsI* gene from *M. smegmatis* genomic DNA using oligos EG686/687 and Gibson assembly was used to place the amplicon into pEG66 linearized with PmlI. Plasmid pEG125 for measuring *bfrA-zeoR* repression by MrsI was created by amplifying the *bfrA* promoter and 5' UTR with oligos EG785/789 and the *zeoR* gene with oligos EG788/790 and assembling into pEG76 linearized with PmeI and NdeI with Gibson assembly. Plasmid pEG164 was generated by site directed mutagenesis on plasmid pEG101 using oligos EG929/930. Plasmid pEG168 was created using site directed mutagenesis on plasmid pEG125 with oligos EG935/936. Plasmid pEG170, for measuring MrsI induction in *M. smegmatis*, was created by amplifying the promoter of *mrsI* from *M. smegmatis* using oligos EG995/996 and the luciferase gene using oligos EG997/998. The amplicons were then assembled into pEG76 linearized with PmeI and NdeI using Gibson assembly. Plasmid pTB-Sth#1_P1A2 for inducible CRISPRi knockdown of *mrsI* in *M. tuberculosis* was created by annealing oligos Sth#1_2T/ Sth#1_2B and annealing with pJR965 linearized with BsmBI (1).

Construction of the *Mycobacterium smegmatis* Δ*mrsI* and Δ*mrsI::mrsI* strains

To delete the *mrsI* gene from *M. smegmatis*, the regions flanking *mrsI* were amplified as described above to create pEG79. The deletion cassette was then amplified using oligos EG634/635 and transformed into the *M. smegmatis* recombineering strain (2). The deletion replaced the sequence of *mrsI* upstream of the rho-independent terminator with a *lox-Hyg-lox* cassette, and deletion mutants were selected for on Hygromycin B. The lox-Hyg-lox cassette was then removed using Cre-recombinase to create a clean deletion strain.

The complemented strain was created by transforming pEG101 (above) into the deletion strain. pEG101 contains the *mrsI* gene downstream of an ATc-inducible variant of the strong UV15 promoter, a strong mycobacterial expression promoter.

in vitro stress of *M. tuberculosis* for sRNA-Seq

Wild type *M. tuberculosis* was expanded in 7H9 medium (Difco) supplemented with 2% v/v glycerol, 0.05% v/v tween-80, and 10% v/v oleic acid albumin dextrose catalase (OADC, Sigma Aldrich) (*M. tuberculosis*) to mid-log phase (OD_{600} =0.5-1.0). Prior to exposure cells were pelleted and resuspended in 7H9 media supplemented with 10% vol/vol albumin dextrose NaCl (ADN), glycerol, and tyloxapol supplemented with 1mM *tert*-butyl Hydroperoxide (tBHP)(oxidative stress), 0.05% v/v SDS (membrane stress), or adjusted to pH 4.5 (acid). For carbon starvation the pelleted cells were resuspended in

PBS supplemented with 0.05% tyloxapol. Iron starvation was performed as described previously (3). Briefly, cells were grown to OD_{600}=0.5-1.0, then washed once with an equal volume of low iron medium and diluted to OD_{600}=0.1 in the same media. Cells were then grown to ~OD_{600}=1.0, diluted back to OD_{600}=0.1 in low iron media, and grown to OD_{600}=0.2-0.4 before adding 50µg/mL of the iron chelator deferoxamine (DFO). Cells were exposed to the stresses for 4 hours (oxidative stress, SDS stress, acid stress) or 24 hours (PBS starvation, DFO exposure during iron starvation) before RNA harvest. No-stress control conditions (7H9, iron-supplemented minimal media) were harvested after 24 hours.

MrsI regulon identification

For transcriptomics in *M. smegmatis* to identify the MrsI regulon, wild type::empty, *M. smegmatis* Δ*mrsI*::empty, and complemented strains were grown in 7H9 medium with ADC, tween-80, and glycerol supplemented with Hygromycin B (100µg/mL) to mid log phase (OD_{600}=0.5-1.0), then pelleted, washed once with low iron media, and resuspended in low iron media supplemented with hygromycin B and ATc to OD_{600}=0.1. Cells were grown in low iron media for 6 hours before harvesting RNA for transcriptomics. For RT-qPCR validation of *bfrA*, cells were grown as described above, with a RNA being extracted at the indicated timepoints.

For transcriptomics in Mtb to identify the MrsI regulon, the two biological replicates of the empty guide control strain and MrsI knockdown strain were grown in 7H9+OADC supplemented with glycerol, kanamycin, and tween-80 to mid-log phase (OD_{600}=0.5-1). For iron deprivation transcriptomics, cells were pelleted, washed one time with iron starvation medium, then resuspended in iron starvation medium to OD_{600}=0.1. The cultures for both strains were then split into plus and minus CRISPRi induction cultures, with 200ng/mL ATc added to the plus induction cultures. Cells were then grown to late log phase (OD_{600}=1.0) and diluted back to OD_{600}=0.1 in low iron media supplemented with 50µg/mL DFO, and ATc in the plus induction cultures. Cells were grown for 24 hours before harvesting RNA for transcriptomics. For oxidative stress and SDS stress transcriptomics, the empty guide control strain and MrsI knockdown strain were grown to early log phase (OD_{600}=0.2) and then were split into plus and minus induction cultures, with 200µg/mL ATc being added to the plus induction cultures. After 24 hours of pre-induction, cultures were pelleted and resuspended in an equal volume of 7H9 media supplemented with ADN, tyloxapol, and tBHP or SDS at the same concentrations used for sRNA discovery and expression profiling. ATc was also added to the plus induction cultures. Cultures were exposed to oxidative and SDS stress for 4 hours before harvesting RNA.

RNA-Seq data analysis

Reads were aligned using the BWA-MEM algorithm (4). Samtools was used to sort and split the reads into plus and minus strands, and the bedtools genomecov tool was used to generate per-base read depth files (5, 6). sRNAs were identified using the BS_Finder default parameters, as described previously (7). Feature counts were calculated using a

custom python script and differential expression analysis was performed using DESeq2 (8). Total RNA-Seq data analysis was performed identically, with the exception that the NC_000962 (Mtb) or NC_008596 genome feature files (gff) were used for feature calling and differential expression analysis.

For Mtb differential expression analysis during MrsI regulon identification, transcriptomics was performed on both an empty guide control strain and a MrsI knockdown strain, with and without induction of CRISPRi. DESeq2 was performed on each strain with and without induction. To determine the effects of MrsI knockdown without nonspecific effects of CRISPRi induction, normalized fold-changes were determined by subtracting the fold change of the empty guide control from the fold change of the knockdown strain. Normalized p-values were calculated using the geometric mean of the two individual adjusted p-values. To define a gene as differentially expressed in both *M. smegmatis* and *M. tuberculosis*, a 1.5-fold change cutoff was used in addition to an adjusted p-value cutoff of $p<0.05$.

Bacterial phenotyping experiments

For growth curves in *M. smegmatis*, cells were grown to mid log phase in minimal medium supplemented with 50µM $FeCl_3$ and pelleted for 10min at 4000 rpm. Cell pellets were washed with an equal volume of low iron medium, pelleted and resuspended in the same medium to OD_{600}=0.01 in low iron medium with or without 50µM $FeCl_3$, with hygromycin and ATc. Growth was monitored using an automated plate reader (Growth Curves USA). In *M. tuberculosis*, growth curves in low iron were performed by growing cells to mid log phase in 7H9 medium, pelleting cells and washing one time with low iron medium. Cells were resuspended in low iron medium supplemented with 5µg/mL DFO to an OD of 0.1, with and without 200ng/mL ATc. For all *M. tuberculosis* phenotyping, growth of three biological replicates was monitored by OD_{600} measurements.

Global Proteome Analysis

Three replicates of Δ*mrsI (ΔmrsI::empty)*, three replicates of complemented (Δ*mrsI::mrsI*) and three replicates of wild type (*wt::empty*) were detected by quantitative LC-MS/MS methods. Cells were grown to mid log phase (OD_{600} =0.5-1.0) in iron-supplemented minimal media with hygromycin B (50µg/mL). Cells were then pelleted, washed once with low iron minimal media supplemented with hygromycin B (50µg/mL) and ATc (100ng/mL), and resuspended in the same media. Cells were then grown in low iron media for 10 hours before pelleting and harvesting protein. All samples were lysed in 6M urea with protease inhibitors and clarified. The denatured protein was reduced, and alkylated, and double digested with both Lys-C and Trypsin overnight. Equivalent amount of tryptic peptides from each sample were labeled with TMT-10 reagent (Thermo Fisher Scientific) and the individual label incorporation was checked via LC-MS/MS. All samples, as expected, had greater than 95% label incorporation. The labeled digests were combined and basic reverse phase (bRP) fractionated into 24 fractions to decrease sample complexity and increase the dynamic range of detection. This global proteome detection

and quantification method was developed at the Broad Institute(9). The proteome data was acquired on a Q-Exactive+ mass spectrometer (Thermo Fisher Scientific). Peptide spectrum matching and protein identification was performed using Spectrum Mill (Agilent). Peptide identification false discovery rates (FDRs) were calculated at 3 different levels: spectrum, distinct peptide, and distinct protein. Peptide FDRs were calculated in SM using essentially the same pseudo-reversal strategy evaluated by Elias and Gygi (10) and shown to perform the same as library concatenation. A false distinct protein ID occurs when all the distinct peptides that group together to constitute a distinct protein have a deltaForwardReverseScore ≤ 0. We adjust settings to provide peptide FDR of 1-2% and protein FDR of 0-1%. SM also carries out sophisticated protein grouping using the methods previously described (11). Only proteins with >2 peptides and at least 2 TMT ratios in each replicate are counted as being identified and quantified. From the initial protein report generated by Spectrum Mill, the report was first filtered by species ensuring that each protein identified was from the organism Mycobacterium smegmatis str. MC2 155. The entries were then filtered by unique peptides ensuring that each of the proteins had >= 2 unique peptides positively identified. A total of **4704** proteins were confidently identified with >=2 unique peptides. Next, the median reporter ion intensity ratios were median-normalized to ensure that the distributions were centered on zero. The three normalized median reporter ion intensity ratios that corresponded to each of the biological replicates were processed using a 1-sample moderated T-test to generate the differential list of proteins.

TargetRNA2 prediction of sRNA targets

For agnostic prediction of MrsI targets, we ran *M. smegmatis* MrsI in TargetRNA2 using default parameters against the NC_008596 genome (12). For forced interactions between MrsI and experimentally determined targets, we changed the seed region length parameter to 6nt and used the 'Single Target' option.
For CopraRNA, the MrsI alleles from *M. tuberculosis*, *M. smegmatis*, and *M. bovis* were used as input.

Oxidative stress pre-exposure experiments and Nanostring gene expression analysis

25ng of RNA was used as input for Nanostring nCounter assays using a SPRINT profiler (Nanostring Technologies) using custom designed probes. Data was analyzed using nSolver version 4 by normalizing raw counts to internal positive controls and three housekeeping genes (Rv1568, Rv1538c, Rv1915). Furthermore, the effects of ATc were accounted for by using a non-targeting sgRNA control strain (TB965, Table S2). The counts of the *mrsI*-knockdown strain with ATc were normalized using the effects of ATc on the counts for each gene in the non-targeting control strain.

Northern Blots

Northern blot analysis was performed using DIG labeled probes. Probe templates for MrsI and 5S RNA for *M. tuberculosis* were created by PCR amplifying *M. tuberculosis* H37Rv gDNA with PCR primers EG1057/8 and EG1097/8, respectively. DIG-labeled

probes were synthesized using the DIG RNA Labeling Kit (Sigma-Aldrich). RNA samples were run on Novex TBE-Urea 6% gels (Thermo Fisher Scientific) and transferred to nitrocellulose membranes using the Invitrogen iBlot DNA Transfer Stacks (Thermo Fisher Scientific). Membranes were prehybridized with ULTRAhyb buffer (Thermo Fisher Scientific) for 2 hours at 65˚C before addition of probe and overnight incubation at 65˚C. Membranes were then washed with 2x SSC, 0.1% SDS solution at room temperature followed by 0.5x SSC, 0.1% SDS solution at 65˚C. Membranes were then washed and blocked with the DIG Wash and Block Buffer Set (Roche) and bound with anti-DIG-AP antibody (Sigma-Aldrich), according to manufacturer's instructions. Blots were developed using CDP-Star (Roche).

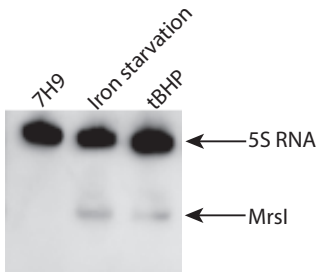

Fig. S1. Northern blot analysis of MrsI from *M. tuberculosis* after growth in 7H9, iron starvation, and oxidative stress. Arrows indicate the 5S RNA and MrsI bands.

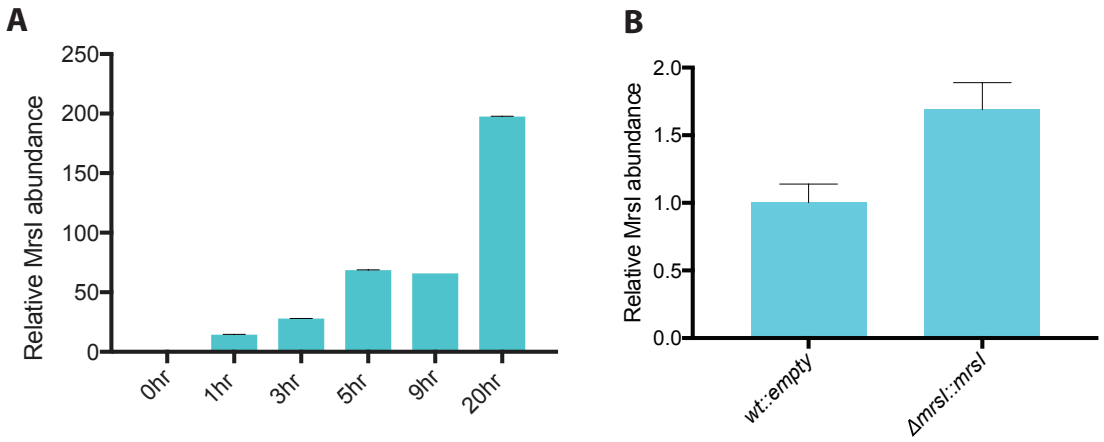

Fig. S2. Expression levels of MrsI in *M. smegmatis* during iron starvation. A) Wild type *M. smegmatis* was starved of iron for 20 hours and the levels of MrsI were measured at multiple time points during growth by RT-qPCR. Error bars represent the SD of 3 replicates. B) Expression levels of MrsI were measured by RT-qPCR in the *wt::empty* and *ΔmrsI::mrsI* strains after 20 hours of iron starvation.

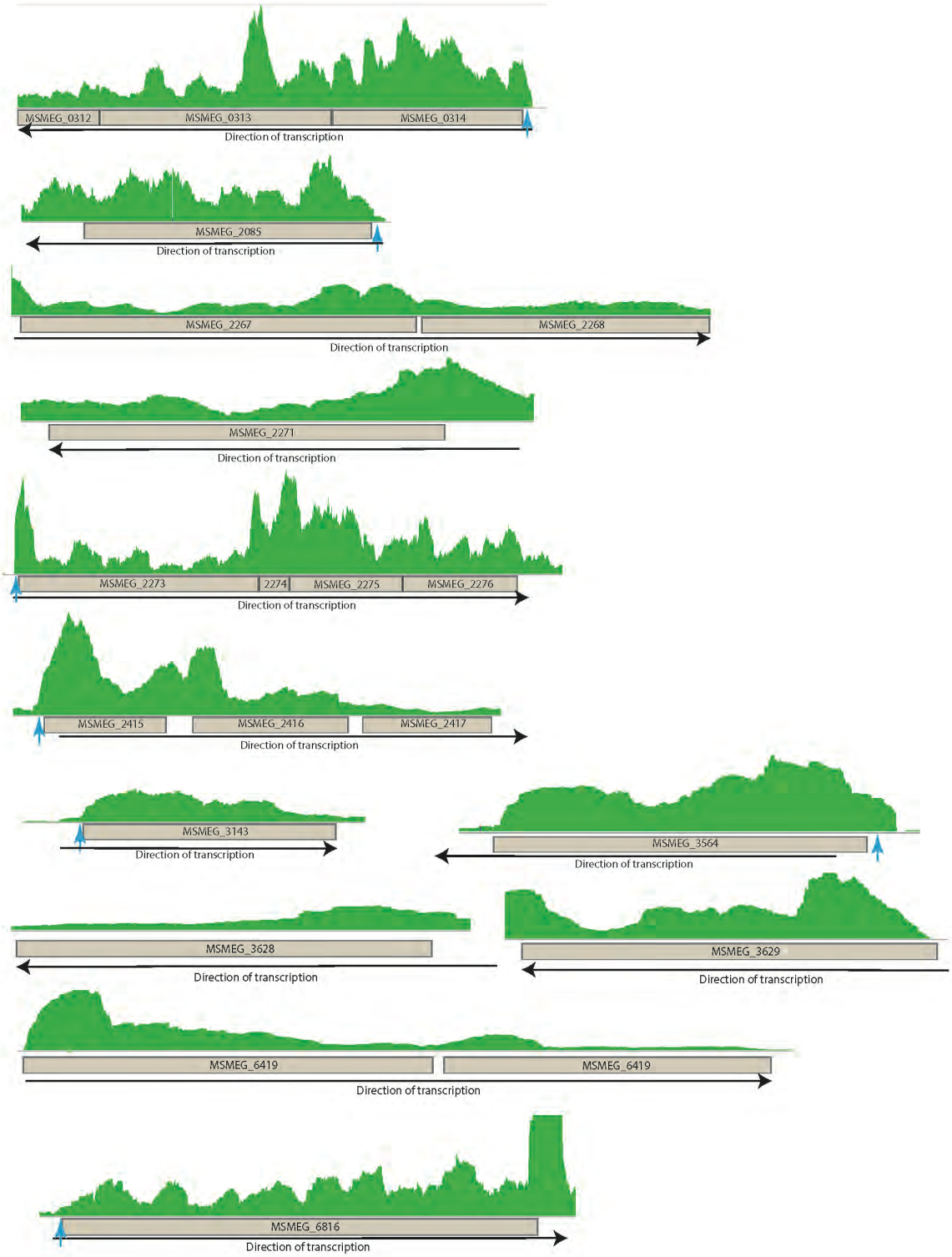

Fig. S3. Transcription units putatively regulated by MrsI in *M. smegmatis*. Visualization of each of the transcription units found by RNA-Seq to be regulated by MrsI. Direction of transcription is indicated (black arrow), as is the putative MrsI binding site in the 5' UTR of each transcript for which MrsI is predicted to bind (blue arrow).

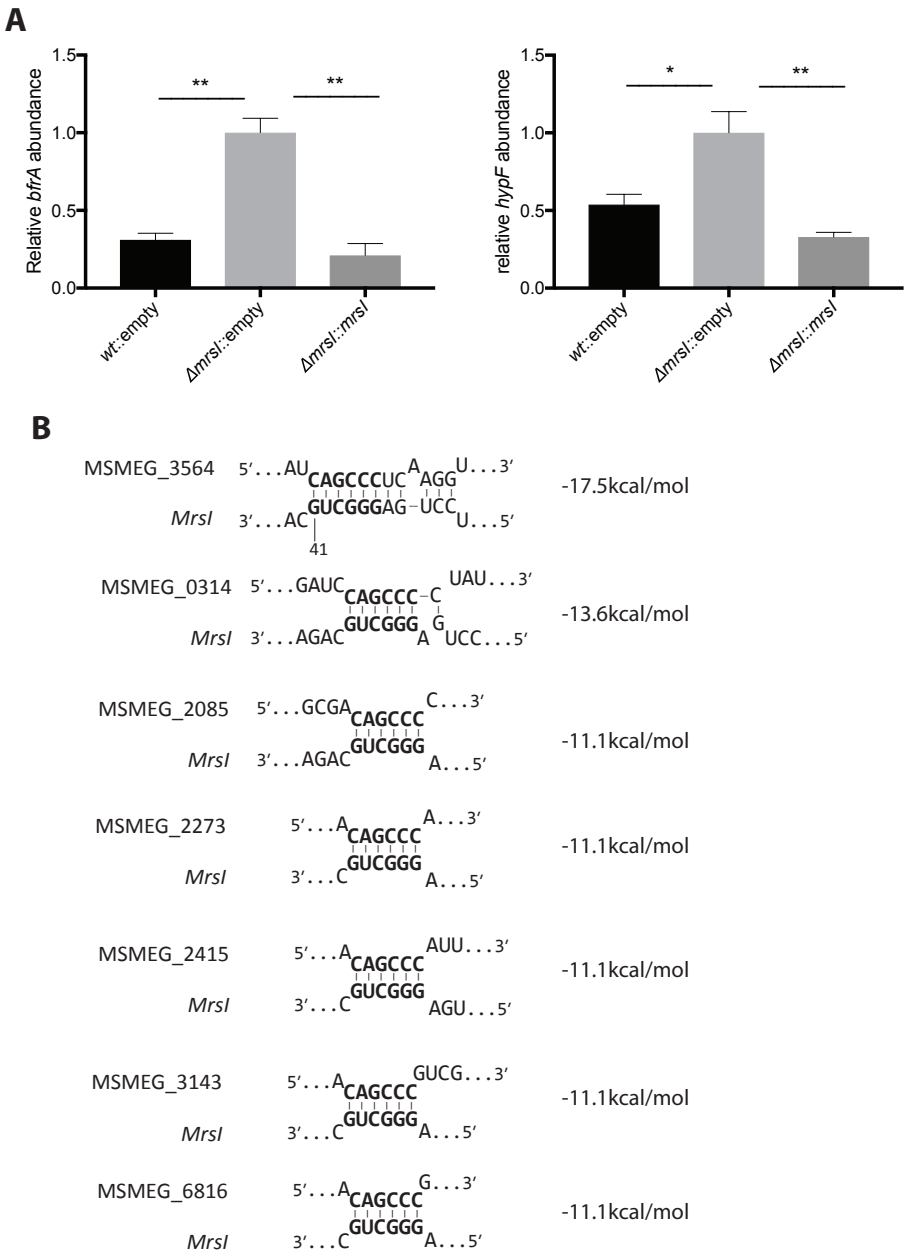

Fig. S4. Regulation of putative targets by MrsI in *M. smegmatis*. A) RT-qPCR validation of the regulation of *bfrA* and *hypF* by MrsI after 5 hours of iron starvation. *p<0.05, **p<0.005 (Unpaired T-test). Error bars represent SD of 3 replicates. B) Predicted binding interactions between MrsI and the putative targets identified by RNA-Seq. Nucleotides in bold are the seed region of the sRNA. Hybridization energies are shown to the right of each predicted interaction.

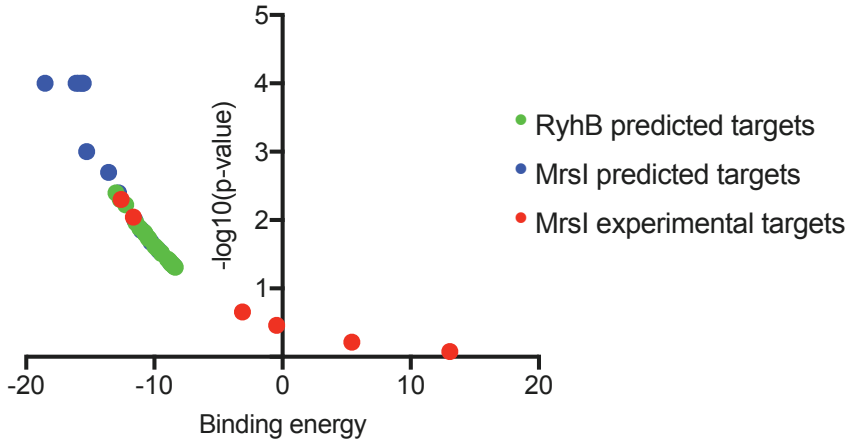

Fig. S5. Bioinformatic and manual prediction of MrsI direct targets in *M. smegmatis*. TargetRNA2 target prediction using default parameters of MrsI from *M. smegmatis* (blue) and RyhB from *E. coli* (green), and the experimentally identified MrsI targets (red). For experimentally defined targets, the seed region length was changed to 6nt and the 'Single Target' option was used.

Table S1. Genes regulated by MrsI in *M. smegmatis*

gene number	Protein function	Product binds iron	MrsI binding site
MSMEG_0312-0314	Glucose-6-phosphate dehydrogenase	Yes	Yes
MSMEG_2085	NADPH-ferredoxin reductase (*fprA*)	Yes	Yes
MSMEG_2267-2268	Unknown	Yes	No
MSMEG_2273-2276	NiFe hydrogenase maturation factors	Yes	Yes
MSMEG_2415-2417	Hemerythrin binding proteins	Yes	Yes
MSMEG_3143	Aconitase (*acnA*)	Yes	Yes
MSMEG_3564	bacterioferritin (*bfrA*)	Yes	Yes
MSMEG_3629	Unknown	Unknown	No
MSMEG_6419-6420	Unknown	Unknown	No
MSMEG_6816	molybdopterin oxidoreductase	Yes	Yes

Table S2: Bacterial strains used in this study

Organism	Strain #	Genotype
M. smegmatis	sEG1	wil-type mc^2155
M. smegmatis	sEG64	*mrsI::loxP*
M. smegmatis	sEG73	mc^2155 $pP_{UV15\text{-}Tet}$-null
M. smegmatis	sEG74	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-null
M. smegmatis	sEG80	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-*mrsI*
M. smegmatis	sEG110	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-*mrsI* L5::P_{bfrA}-*zeoR*
M. smegmatis	sEG244	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-*mrsI* L5::P_{bfrA}-*zeoR*(C27T)
M. smegmatis	sEG245	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-*mrsI*(G41A) P_{bfrA}-*zeoR*
M. smegmatis	sEG246	*mrsI::loxP* $pP_{UV15\text{-}Tet}$-*mrsI*(G41A) P_{bfrA}-*zeoR*(C27T)
M. smegmatis	sEG257	mc^2155 L5::P_{mrsI}-luciferase
M. tuberculosis	tEG1	Wild type H37Rv
M. tuberculosis	tEG16	*L5::dcas9* sgRNA-*mrsI*
M. tuberculosis	TB965	*L5::dcas9* sgRNA-null
Escherichia coli		BL21- Gold(DE3) E. coli B F- ompT hsdS(r8 - m8 -) dcm+ Tetr gal λ(DE3) endA Hte

Table S3: Plasmids used for this study

Plasmid name	Description
pEG66	p$P_{UV15\text{-}Tet}$-null (Hyg)
pEG76	null::L5(Kan)
pEG79	Δ*mrsI::lox-Hyg-lox*
pEG101	p$P_{UV15\text{-}Tet}$-*mrsI*(Hyg)
pEG125	P_{bfrA}-zeoR::L5(Kan)
pEG164	p$P_{UV15\text{-}Tet}$-*mrsI*(G41A)(Hyg)
pEG168	P_{bfrA}-zeoR(C27T)::L5(Kan)
pEG170	P_{mrsI}-luciferase::L5(Kan)
pTB-Sth#1_P1A2	$P_{UV15\text{-}Tet}$*dcas9* sgRNA-*mrsI*::L5(Kan)
pJR965	$P_{UV15\text{-}Tet}$*dcas9* sgRNA-null::L5(Kan)

Table S4: Oligos used for this study

Oligo name	Sequence	Description
EG236	gactacaccaagggctacaag	RT-qPCR for *sigA*
EG237	ttgatcacctcgaccatgtg	RT-qPCR for *sigA*
EG493	AGCAATTGACGAGCGAACT	RT-qPCR for MSMEG_3564 *(bfrA)*
EG494	ATTCGGCACGCGTATGTT	RT-qPCR for MSMEG_3564 *(bfrA)*
EG627	ccttttgcgtttaatactgtttTTCTCACACGGCCGGTCGG	*mrsI* deletion plasmid construction
EG628	tatggcgcgcGCGGGTCCCGCGTCAAGC	*mrsI* deletion plasmid construction
EG629	cgggacccgcGCGCGCCATAACTTCGTA	*mrsI* deletion plasmid construction
EG630	gaagccttgcAGTGGATCCATAACTTCGTATAATG	*mrsI* deletion plasmid construction
EG631	tggatccactGCAAGGCTTCCCTAATTTAGC	*mrsI* deletion plasmid construction
EG632	agcgagacgaaatacgcgatCGGGTGATTCCGCTGTTG	*mrsI* deletion plasmid construction
EG634	ttctcacacggccggtcggc	*mrsI* deletion cassette amplification
EG635	cgggtgattccgctgttggt	*mrsI* deletion cassette amplification
EG686	tagataggctctgcacAAGTATCGAGCCAACGGAC	plasmid construction
EG687	agccgtgaacgacacAAAAGATTCGGGCGGGTC	plasmid construction
EG785	ccttttgcgtttaatactgtttTTCACCGGGCCTTTCCGC	plasmid construction
EG788	gctagagccgtgaacgaccaCTAGTCCTGCTCCTCGGC	plasmid construction
EG789	acttggccatGGTCACTCCTAGACACCTTGAG	plasmid construction
EG790	aggagtgaccATGGCCAAGTTGACCAGTG	plasmid construction
EG795	ACTTCGTGGAGGACGACTT	RT-qPCR for *zeoR*
EG796	CAGGCCAGGGTGTTGTC	RT-qPCR for *zeoR*
EG929	cgggggtctctgtagccctcaggac	site directed mutagenesis
EG930	gtcctgagggctacagagacccccg	site directed mutagenesis
EG935	ccttgagggctaatcaagtggtgctgtttgccg	site directed mutagenesis
EG936	cggcaaacagcaccacttgattagccctcaagg	site directed mutagenesis
EG995	ccttttgcgtttaatactgtttCACCGCGATGTGGCACCT	plasmid construction
EG996	ggctgccgtgCGATAATAGGCAAGGCTTCCCTAATTTAGC	plasmid construction
EG997	cctattatcgCACGGCAGCCCGGTGAAG	plasmid construction
EG998	gctagagccgtgaacgaccaTTACTGCTCGTTCTTCAGCACGC	plasmid construction
EG1057	TAATACGACTCACTATAGGGtgcaccaaggcacggg	Northern probe template construction
EG1058	acaacgaccgcggcc	Northern probe template construction
EG1097	TAATACGACTCACTATAGGGTACTTTTCCACC	
EG1098	ttacggcggccacagc	
Sth#1_2T	GGGAaccgggggtcactgcagccc	plasmid construction
Sth#1_2B	AAACgggctgcagtgacccccggt	plasmid construction

Additional dataset S1 (separate file)

Coordinates and expression profiles of the 189 sRNA candidates during growth in the 5 stress conditions.

Additional dataset S2 (separate file)

RNA-Seq expression data for *mrsI* strains in *M. smegmatis.*

Additional dataset S3 (separate file)

Proteomics data for on *mrsI* strains in *M. smegmatis.*

Additional dataset S4 (separate file)

RNA-Seq expression data for *mrsI* strains in *M. tuberculosis.*

References

1. Rock JM, et al. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. doi:10.1038/nmicrobiol.2016.274.
2. van Kessel JC, Hatfull GF Recombineering in Mycobacterium tuberculosis. doi:10.1038/NMETH996.
3. Kurthkoti K, et al. (2017) The Capacity of *Mycobacterium tuberculosis* To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas. *MBio* 8(4):e01092-17.
4. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 0(0):1–3.
5. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Appl NOTE* 26(6):841–84210.
6. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma Appl NOTE* 25(1610):2078–2079.
7. Dejesus MA, et al. (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio* 8(1). doi:10.1128/mBio.02133-16.
8. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15. doi:10.1186/s13059-014-0550-8.
9. Mertins P, et al. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods* 10(7):634–637.
10. Elias JE, Gygi SP (2010) Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods Mol Biol* (SUPP.60):1–16.
11. Nesvizhskii AI, Aebersold R (2005) Interpretation of Shotgun Proteomic Data. *Mol Cell Proteomics* 4(10):1419–1440.
12. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of

small regulatory RNAs in bacteria. *Nucleic Acids Res* 42(Web Server issue):W124-9.