

# SCIENTIFIC REPORTS



OPEN

## Essential Nucleoid Associated Protein mIHF (Rv1388) Controls Virulence and Housekeeping Genes in *Mycobacterium tuberculosis*

Nina T. Odermatt, Claudia Sala, Andrej Benjak  & Stewart T. Cole

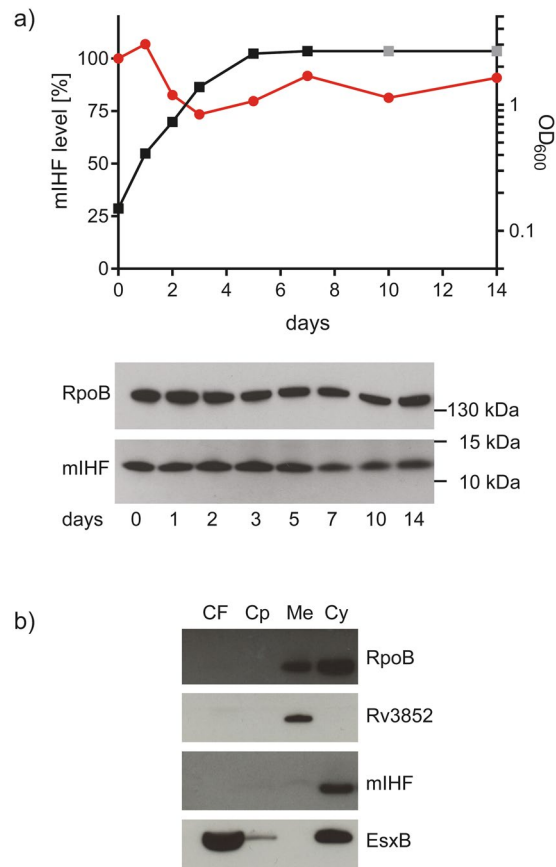
Tight control of gene expression is crucial for *Mycobacterium tuberculosis* to adapt to the changing environments encountered when infecting or exiting human cells. While three nucleoid associated proteins (NAPs) EspR, HupB and Lsr2 have been investigated, the role of a fourth, the mycobacterial integration host factor (mIHF), remains elusive. Here, we report a multidisciplinary functional analysis that exploits a conditional mIHF mutant. Gene silencing was bactericidal and resulted in elongated cells devoid of septa, with only one nucleoid. ChIP-sequencing identified 153 broad peaks distributed around the chromosome, which were often situated upstream of transcriptional start sites where EspR also bound. RNA-sequencing showed expression of 209 genes to be heavily affected upon mIHF depletion, including those for many tRNAs, DNA synthesis and virulence pathways. Consistent with NAP function, mIHF acts as a global regulator by directly and indirectly controlling genes required for pathogenesis and for housekeeping functions.

Bacterial gene expression is tightly controlled and influenced by environmental cues. In the case of the human pathogen *Mycobacterium tuberculosis*, which has to adapt to a hostile milieu upon phagocytosis by a macrophage, these stimuli include acidic pH, reactive oxygen species, nutrient limitation, fatty acid availability or the presence of lytic enzymes<sup>1</sup>. It is therefore crucial to comprehend gene expression to understand bacterial growth and survival. The regulons of traditional transcription factors comprise from one to tens of target genes. Nucleoid associated proteins (NAPs) on the other hand act at the global level impacting expression of hundreds of genes often by shaping chromatin architecture<sup>2</sup>. In *Escherichia coli*, more than twelve NAPs have been characterized so far. Their cellular abundance fluctuates during the different growth phases<sup>3</sup> and each NAP targets a specific set of genes<sup>4</sup>. In contrast, only four NAPs have been reported to date in *M. tuberculosis*. HupB is essential for growth in macrophages and for iron acquisition<sup>5</sup> and Lsr2 is an *E. coli* H-NS-like protein that preferentially binds to AT-rich sequences<sup>6</sup>. EspR regulates secretion of the main virulence factors of *M. tuberculosis*<sup>7</sup> by controlling expression of the ESX-1 related *espACD* operon among others<sup>8</sup>. These three NAPs all have multiple binding sites on the *M. tuberculosis* chromosome and regulate the vast majority of genes. The overall picture of the regulatory network is unclear, and the role of the fourth NAP, mycobacterial integration host factor mIHF, has not been defined yet.

Discovered as being essential for mycobacterial phage L5 integration into the *M. smegmatis* genome<sup>9</sup>, mIHF was therefore named after the *E. coli* counterpart, despite the two genes and their respective proteins showing no sequence similarity. mIHF is highly conserved among the *Mycobacterium* genus and even *M. leprae*, with its reduced genome, possesses a copy of *mihF*<sup>10</sup>, and orthologs occur in many other Actinobacteria. The *mihF* (*rv1388*) gene in *M. tuberculosis* was initially predicted to be 573 bp-long and to encode a ~20 kDa protein<sup>11</sup>. More recently, based on comparative genomics, Mishra and colleagues proposed that mIHF of *M. tuberculosis* contains only 105 amino acid residues<sup>12</sup> and this is supported by proteomics analysis with mIHF appearing among the top ten most abundant proteins of *M. tuberculosis*<sup>13–15</sup>. It has been reported that mIHF binds to linear and supercoiled DNA and enhances topoisomerase activity<sup>12</sup>.

The *mihF* gene was predicted to be essential for *in vitro* growth with glycerol or cholesterol as carbon sources by *Himar-1* based transposon mutagenesis<sup>16</sup>, thus suggesting an important regulatory role in *M. tuberculosis* metabolism. In this study, we investigated the biological function of mIHF thoroughly with the help of a

École Polytechnique Fédérale de Lausanne, Global Health Institute, Station 19, 1015, Lausanne, Switzerland. Correspondence and requests for materials should be addressed to S.T.C. (email: [stewart.cole@epfl.ch](mailto:stewart.cole@epfl.ch))



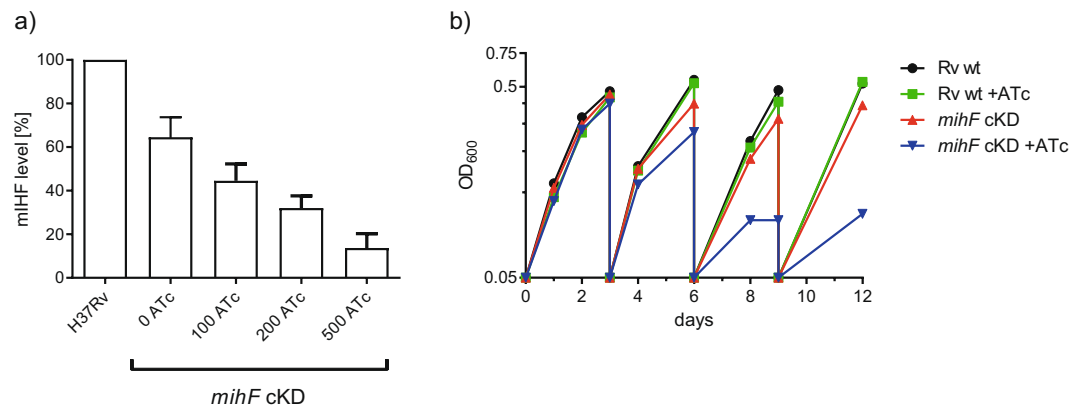
**Figure 1.** Expression and localization of mIHF. **(a)** Time-course analysis by immunoblot of mIHF levels (red bullets, left y-axis), and optical density at 600 nm ( $OD_{600}$ , black squares, right y-axis) of H37Rv from exponential to stationary phase. Protein levels were calculated by density analysis of the image below, relative to RpoB and to the first time point (day 0).  $OD_{600}$  at days 10 and 14 (grey square) were set to the same value as day 7, as the culture formed aggregates typical of *M. tuberculosis* in stationary phase, which prohibited proper measurement of  $OD_{600}$ . **(b)** Immunoblot of culture filtrate (CF), capsular (Cp), membrane (Me) and cytosolic (Cy) fractions of H37Rv. Antibodies used are indicated to the right.

conditional knockdown (cKD) mutant. We show that mIHF is indeed essential for growth and affects protein and nucleic acid synthesis. Cells depleted of mIHF displayed aberrant morphology, and nucleoid segregation, before dying. The effect of mIHF depletion was investigated using ChIP-seq and RNA-sequencing thereby demonstrating that this NAP impacts *M. tuberculosis* gene regulation pleiotropically by controlling expression of housekeeping as well as virulence genes.

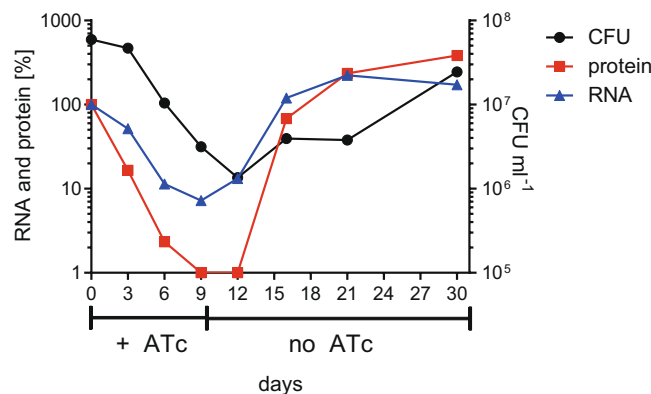
## Results

**mIHF is an abundant cytosolic protein.** While some NAPs are present throughout the entire growth cycle of a bacterium, others peak at certain stages. The mIHF protein was identified at approximately 12 kDa by immunoblotting and a time-course analysis of protein levels showed that it is constantly present from the exponential to stationary phase with little fluctuations (Fig. 1a). To probe where mIHF is localized inside the bacterial cell, *M. tuberculosis* H37Rv cell extracts were fractionated prior to subsequent immunoblotting. Proteins RpoB, Rv3852<sup>17</sup> and EsxB were used as positive controls for the cytosol, membrane and secreted fractions, respectively. The mIHF protein was detected in the cytosol only (Fig. 1b).

**Transcription start site identification and conditional knockdown mutant generation.** To explore the regulatory function of mIHF, we constructed an *mihF* conditional knockout mutant after first localizing the transcription site(s) since it had been suggested that *mihF* is shorter than originally annotated<sup>12</sup>. Transcripts were analysed by rapid amplification of cDNA ends (5'-RACE) and three potential transcription start sites (TSS) were detected (Fig. S1). The third of these, TSS3, situated 167 bp downstream of the currently annotated translation start site, is preceded by a TANNT -10 motif, shared by most promoters of *M. tuberculosis*<sup>18</sup>. From multiple mIHF sequence alignments and the presence of an appropriately positioned ribosome binding site (GGAGGAA)<sup>19</sup>, residue 86 of the original annotation was inferred to be the initiation codon (Fig. S1), as recently proposed<sup>12</sup>, and this will be referred to as *mihF*-86. An in-depth discussion about the *mihF* initiation codon is available in supplementary text 1.



**Figure 2.** *miHF* silencing by ATc. (a) mIHf levels detected by immunoblot normalized to RpoB and relative to H37Rv wildtype. ATc concentration is indicated in ng ml<sup>-1</sup>. (b) Growth curves of *miHF*-cKD and H37Rv wild type strains with 600 ng ml<sup>-1</sup> ATc and without ATc. Cultures were diluted to OD<sub>600</sub> = 0.05 every 3 days.



**Figure 3.** Effect of *miHF* silencing. *miHF* mRNA, mIHf protein levels and colony forming units in the presence of ATc (day 0–9, diluted every three days) and under permissive conditions (from day 9).

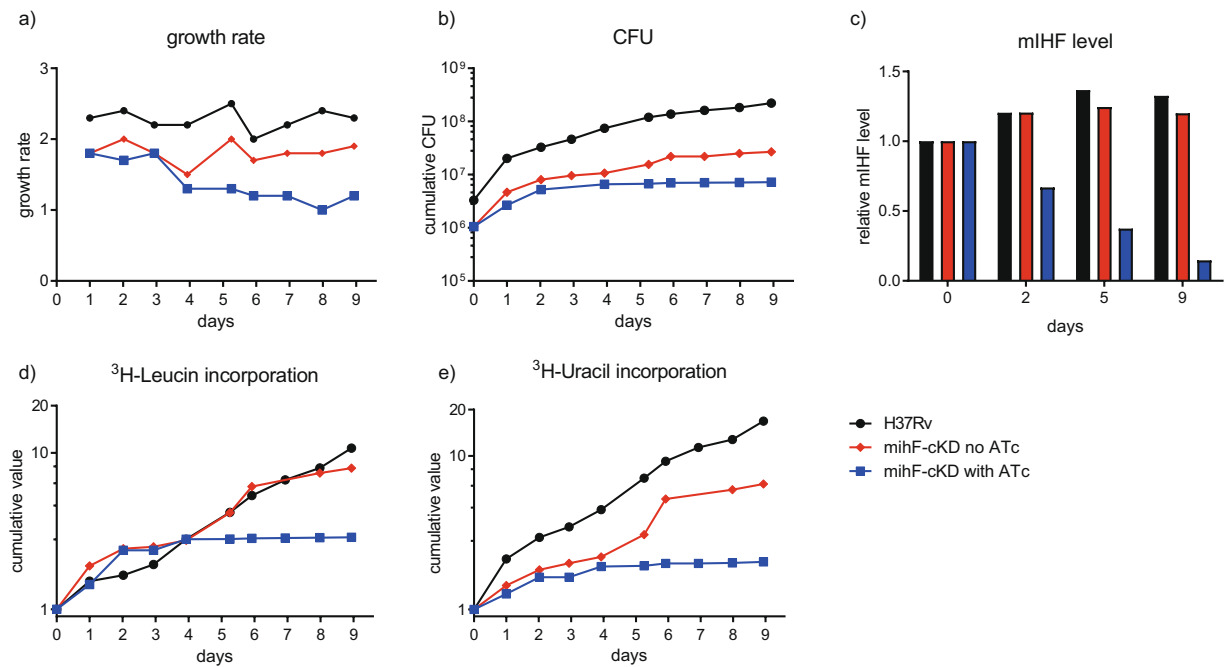
Next, we employed gene replacement to remove the annotated, full-length *miHF* gene from the chromosome. After confirming the nature of the merodiploid strain, the second crossing over event, which led to the in-frame deletion of the gene, was only successful when a copy of *miHF* was provided *in trans* under the control of the repressible TET-PIP OFF system<sup>20</sup>. Furthermore, it was necessary to include the two downstream genes, *gmk* and *rpoZ*, on the complementing construct in order to delete the chromosomal *miHF* gene and therefore obtain the conditional *miHF*-cKD mutant. The mutation was confirmed by Southern blotting (Fig. S2).

Addition of anhydrotetracycline (ATc) to the *miHF*-cKD mutant complemented with the original full-length *miHF* caused no variation in mIHf protein abundance, nor in growth dynamics (data not shown). However, controllable complementation was obtained on introduction of plasmid pNO63, carrying *miHF*-86, thus confirming that this represents the *bona fide* *miHF* gene (named *miHF* hereafter).

***miHF* is essential for growth and survival of *M. tuberculosis*.** Addition of increasing concentrations of ATc to cultures of the *miHF*-cKD mutant allowed titration of mIHf levels (Fig. 2a, S3). In the absence of ATc, the conditional mutant produced lower levels of mIHf than the wild type strain (approximately 30% less), indicating that the controllable *ptr* promoter is weaker than the natural one. The level of mIHf was reduced to less than 10% upon addition of 500 ng ml<sup>-1</sup> ATc. Of note, it was necessary to dilute the *miHF*-cKD cultures at least twice to see a reduction of the growth rate of the mutant compared to the parental strain or to the uninduced *miHF*-cKD strain (Fig. 2b). As expected, ATc had no impact on the growth of the parental H37Rv strain.

Figure 3 shows the viability of the *miHF*-cKD strain upon repression of *miHF* expression for 9 days, followed by three weeks under permissive conditions. While *miHF* RNA and mIHf protein levels rapidly decreased during the first three days, the number of colony forming units (CFU) dropped at day 6, proving that depletion of mIHf had a bactericidal effect. When ATc was removed from the bacterial culture, RNA production resumed, whereas a three-day lag was noticed before mIHf protein could be detected. Similarly, viable counts increased after a three-day lag under permissive conditions.

**Impact of mIHf depletion on macromolecular synthesis.** To examine if *de novo* protein and nucleic acid (RNA plus DNA) production was affected by mIHf-depletion, the incorporation of <sup>3</sup>H-labelled leucine



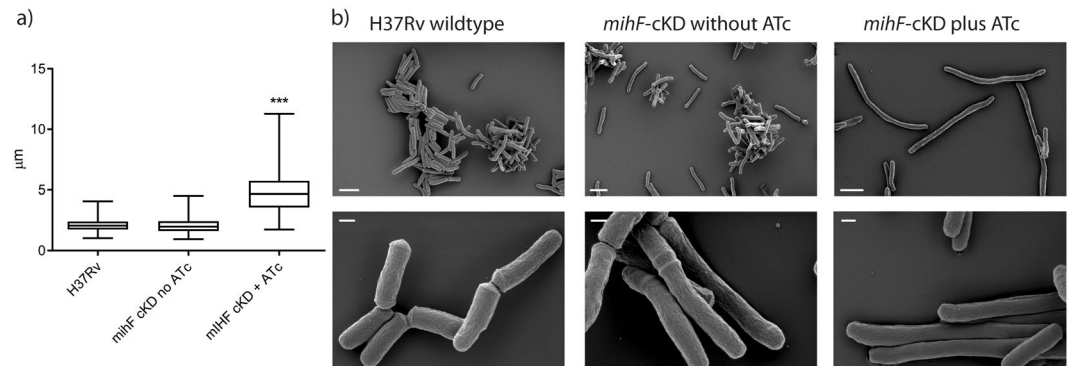
**Figure 4.** Tritium labelled-leucine and -uracil incorporation. **(a)** Growth rate of H37Rv wild type and *mihF*-cKD with and without ATc. **(b)** Cumulative CFU ml<sup>-1</sup> of the three strains. **(c)** mIHF levels at days 0, 2, 5 and 9 measured by immunoblot, relative to day 0 for each strain. **(d)** Leucine and uracil **(e)** incorporation relative to day 0.

and uracil into nucleic acids and proteins, respectively, by the *mihF*-cKD strain was monitored over nine days in non-permissive conditions. Figure 4a shows that the growth rate of the mutant strain without ATc was lower than that of H37Rv, and then further decreased in the presence of ATc. Likewise, the total number of CFU increased from  $5 \times 10^6$  to  $2 \times 10^8$  CFU ml<sup>-1</sup> for H37Rv after nine days, while *mihF*-cKD without ATc started at  $1.05 \times 10^6$  and later attained  $2.68 \times 10^7$  CFU ml<sup>-1</sup> (Fig. 4b). In contrast, growth of the *mihF*-cKD mutant in the presence of ATc reached a plateau after four days and only increased to  $7 \times 10^6$  CFU ml<sup>-1</sup> at day 9. While the mIHF protein levels rose slightly in the H37Rv and *mihF*-cKD strains without ATc, the *mihF*-cKD mutant already showed lower mIHF protein levels at day 2 after silencing and mIHF abundance dropped to ~40% at day five (Fig. 4c), when growth was arrested and nucleic acid as well as protein synthesis came to a halt (Fig. 4d and e). Depletion of mIHF had therefore a pleiotropic effect on *M. tuberculosis* physiology, as it severely compromised DNA, RNA and protein synthesis.

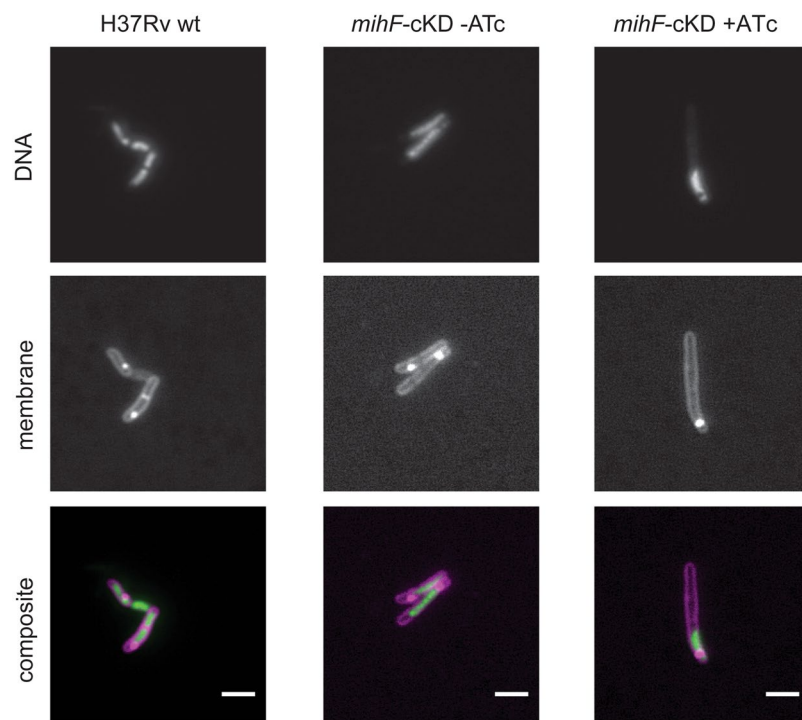
**mIHF-depleted bacteria are elongated and do not form septa.** Scanning electron microscopy allowed investigation of the ultrastructure of *mihF*-cKD mutant cells. Cells depleted of mIHF were more than twice as long as the H37Rv parental strain and the *mihF*-cKD mutant grown without ATc (Fig. 5a). The average lengths for the wild type and *mihF*-cKD strain without ATc were  $2.13 \pm 0.53 \mu\text{m}$  and  $2.09 \pm 0.67 \mu\text{m}$ , respectively, whilst *mihF*-cKD cells with ATc were  $4.98 \pm 1.96 \mu\text{m}$  long. In addition, most of the induced *mihF*-cKD cells longer than  $3 \mu\text{m}$  showed no ridge formation, which is indicative of the lack of an underlying septum<sup>21</sup>. Conversely, ridges were clearly visible in the wild type and in *mihF*-cKD cells without ATc (compare the lower panel of H37Rv and *mihF*-cKD plus ATc in Fig. 5b). Apart from the elongated, septum-less phenotype, no other abnormal shape was observed (i.e. no branching, swelling or bending) after depletion of mIHF.

Fluorescence microscopy confirmed that most of the long mIHF-depleted cells did not generate septa and additionally showed that they contained only one nucleoid (Fig. 6). On the other hand, in septum-containing cells (wild type and *mihF*-cKD grown without ATc), DNA was present in each compartment (Fig. 6). Thus, mIHF is necessary for septum formation and normal cell division as well as for DNA segregation in *M. tuberculosis*.

**mIHF has a broad impact on the *M. tuberculosis* transcriptome.** To characterize the mIHF-dependent regulon, global gene expression analysis was performed by RNA-seq and the data subjected to gene ontology (GO) analysis to identify biological processes. Effects of the inducer on the transcriptome were ruled out by comparing the transcription profile of strain H37Rv in the presence and absence of ATc. Using a false discovery rate (FDR) below 1% and a change cut-off >2-fold, no differentially expressed genes were found (Supplementary Dataset 1). Likewise, integration of the complementing plasmid did not affect gene expression significantly (Supplementary Dataset 1). The transcriptome of the *mihF*-cKD mutant strain in the presence and absence of ATc was then analysed using the same cut-off values. Depletion of mIHF had a broad impact on the global transcriptome as 679 downregulated and 464 upregulated genes were detected (Supplementary Dataset 1). By increasing the threshold to >4, the number of deregulated genes was limited to 150 downregulated and 59 upregulated (Supplementary Dataset 1).



**Figure 5.** Cell morphology of mIHf-depleted bacteria. **(a)** Cell length measured by surface electron microscopy. *mihF*-cKD cells grown with ATc for 9 days are twice as long as H37Rv wild type and *mihF*-cKD grown in the absence of ATc (\*\*\*)  $p < 0.0001$ . Whiskers represent minimum and maximum values, median is plotted at the middle of the box). **(b)** Scanning electron micrographs of H37Rv parental strain and *mihF*-cKD strains with and without ATc. Bar represents 2 μm on the upper panel and 200 nm on the lower panel.



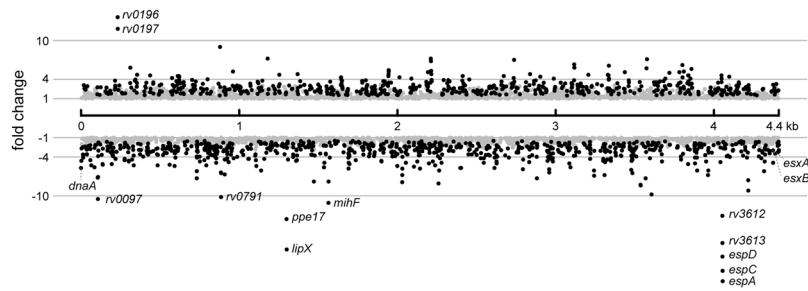
**Figure 6.** Septum and nucleoid positions. Fluorescence microscopy images of H37Rv wild type, *mihF*-cKD with and without ATc. DNA was stained with SYTO9, membranes with FM4-64. Scale bars represent 2 μm.

Figure 7 displays the distribution of the deregulated genes throughout the *M. tuberculosis* genome: no clustering was observed. The most severely affected genes are listed in Fig. 8 and discussed in greater detail below.

Strikingly, the operon showing the maximum repression was *espACD*, which is required for the ESX-1 type VII secretion system to function in *M. tuberculosis*. Additionally, various other genes encoding components of the Type VII secretion systems ESX-1, ESX-2 and ESX-5, and their substrates (EsxA, EsxB, EsxC, EsxD, EsxM, EsxN), were expressed at a lower level. mIHf has therefore a major impact on expression of virulence-related secretion systems. The second most downregulated operon was *rv1168-rv1169*, coding for PPE17 and LipX, respectively.

The majority of the stable RNAs, notably 23 out of 45 tRNAs and several small regulatory RNAs, were repressed by more than 2-fold, as well as some ribosomal protein genes (Fig. 8). Expression of the genes for the replication initiator protein DnaA and for the beta chain of the DNA polymerase III, DnaN, were >4-fold downregulated (Fig. 8). These results indicate that mIHf is involved in controlling transcription of housekeeping genes needed for protein and DNA synthesis.

Conversely, the most upregulated operon was *rv0196-rv0197*, coding for a transcriptional regulatory protein and a possible oxidoreductase, followed by two conserved hypothetical proteins (*rv0784*, *rv1057*) and the



**Figure 7.** mIHF controlled genes. (a) Differentially expressed genes identified in mIHF-depleted relative to non-depleted *M. tuberculosis* cultures are plotted at their genomic position. Grey dots represent genes with a false discovery rate (FDR) higher than 1%, black dots are significantly deregulated genes with FDR < 1%. Interesting genes discussed further in the text are labelled.

virulence-associated *mce3* operon. Moreover, several genes encoding transcription factors were found to be deregulated. For instance, expression of *lsr2*, *sigE* and *espR*, which may act downstream of mIHF, thus amplifying the regulatory signal, was diminished. The *mihF* gene itself was more than 11-fold repressed, confirming the silencing effect of ATc.

GO analysis revealed the most downregulated genes ( $p < 0.05$ ) to be associated with host-pathogen interaction, fatty acid metabolism, pathogenesis, ribosomes, transcription factors, DNA binding proteins and DNA replication initiation (Supplementary Dataset 1). The GO categories related to upregulated genes were associated with DNA damage such as DNA duplex unwinding, DNA repair and exonuclease activity.

**mIHF is a nucleoid associated protein in *M. tuberculosis*.** To evaluate if mIHF directly regulates the differentially expressed genes found in RNA-seq, mIHF binding to the *M. tuberculosis* chromosome was investigated by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). Exponentially growing H37Rv wild type cells as well as mIHF-depleted *mihF*-cKD mutant cells were assessed. Upon analysing the ChIP-seq results from exponentially growing wild type cells (Supplementary Dataset 1), we noticed that many mIHF peaks overlapped the previously published EspR binding sites<sup>7</sup>. For the sake of consistency, EspR ChIP-seq was then repeated on the same H37Rv sample used for mIHF ChIP-seq. 162 EspR binding sites were detected, of which 128 had been previously described<sup>7</sup> and the others were found close by (Fig. S4a for Pearson's correlation coefficient, PCC). mIHF bound to 153 loci in exponential phase (Supplementary Dataset 1), whereas 124 sites were contacted upon depletion of the protein (Supplementary Dataset 1). 62 mIHF binding sites were occupied in exponentially growing cells and upon mIHF depletion (PCC of 0.67, Fig. S4), and 64 of the 153 loci were shared with EspR as well. The overlap between EspR and mIHF peaks was higher in exponential phase cells (PCC of 0.6) than after mIHF depletion (PCC of 0.44, Fig. S3). We further integrated the profiles of CRP (cAMP receptor protein) and Lsr2, with 191 and 305 binding sites, respectively. While a marked overlap was observed between the regions bound by the three NAPs (Lsr2, EspR and mIHF), CRP seemed to bind at different loci (Fig. 9a).

The GC content of regions where EspR and/or mIHF bound ranged between 58% and 60%, which is significantly lower than the genome average of ~66%, ( $p < 0.0001$ ). This is consistent with the preferential location of mIHF binding near the transcriptional start sites (TSS, Fig. 9b). Indeed, mIHF-bound loci were mostly situated immediately upstream of the TSS, and a minor fraction was found after the transcriptional termination site (Fig. 9b). Moreover, mIHF binding sites appeared as broad rather than as sharp peaks in exponentially growing as well as in mIHF-depleted cells.

Loci enriched for mIHF were distributed all over the genome, with some regions harbouring several peaks next to each other. Figure 10 shows the binding profile for mIHF in exponential growth phase and after depletion, with regions of interest highlighted. The genomic region around the top upregulated genes *rv0196*-*rv0197* was not bound by mIHF (Fig. 10b), while the long intergenic region upstream of the most downregulated gene *espA* showed peaks in both conditions (Fig. 10e).

Despite being named integration host factor, mIHF did not bind close to either of the prophages PhiRv1 (spanning *rv1573*-*rv1587*) or PhiRv2 (*rv2650*-*rv2659*) in the chromosome. On the other hand, the so-called genomic islands (GI) defined by Becq *et al.*<sup>22</sup>, which represent 4.5% of the whole *M. tuberculosis* genome, were contacted by mIHF in both growth conditions. Almost one quarter (23%) of mIHF binding sites in exponential phase, and 18% after depletion, were near GI.

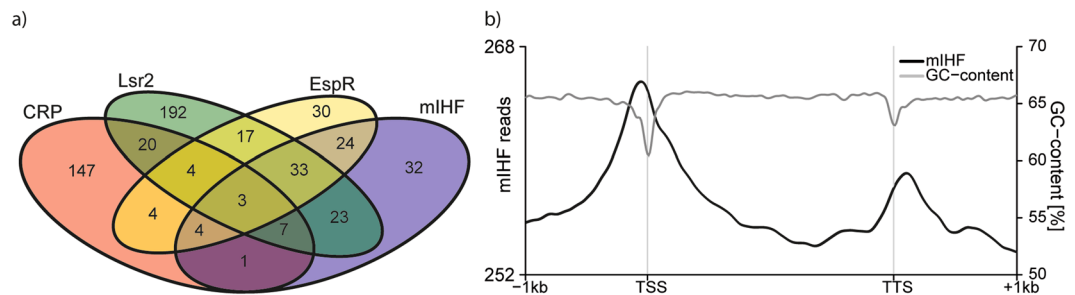
The impact of mIHF as a potential direct activator or repressor was defined for features where mIHF peaks either overlapped the genes or were within 500 bp of the coding sequence boundaries. Genes associated with an mIHF binding site were more frequently deregulated as compared to genes unlinked to an mIHF-enriched locus ( $p < 0.0001$ ). Genes associated with mIHF following depletion had a mean expression level of  $-1.91$ , representing an almost 2-fold downregulation, compared to genes unaffected by mIHF depletion with a mean expression level of 0.05, indicating no transcriptional change. Similarly, genes in close proximity to mIHF binding sites in exponential phase were expressed at  $-2.0$ , and more distant genes at 0.02. This indicates that mIHF binding directly impacts gene regulation.

	gene	description	FC	
<b>stable RNAs</b>				
MTB000049	<i>serT</i>	tRNA-Ser (GCT)	-9.2	
MTB000048	<i>argU</i>	tRNA-Arg (ACG)	-7.6	
MTB000007	<i>metT</i>	tRNA-Met (CAT)	-7.3	
<b>replication and transcription</b>				
Rv0001	<i>dnaA</i>	Chromosomal replication initiator protein DnaA	-5.7	
Rv0002	<i>dnaN</i>	DNA polymerase III (beta chain) DnaN (DNA nucleotidyltransferase)	-4.6	
Rv3711c	<i>dnaQ</i>	Probable DNA polymerase III (epsilon subunit) DnaQ	3.2	
Rv3370c	<i>dnaE2</i>	Probable DNA polymerase III (alpha chain) DnaE2	5.1	
Rv3202c	<i>Rv3202c</i>	Possible ATP-dependent DNA helicase	7.1	
<b>virulence</b>				
Rv3616c	<i>espA</i>	ESX-1 secretion-associated protein A, EspA	-23	
Rv3615c	<i>espC</i>	ESX-1 secretion-associated protein EspC	-22	
Rv3614c	<i>espD</i>	ESX-1 secretion-associated protein EspD	-19	
Rv3874	<i>esxB</i>	10 kDa culture filtrate antigen EsxB (LHP) (CFP10)	-3.4	
Rv3875	<i>esxA</i>	6 kDa early secretory antigenic target EsxA (ESAT-6)	-3.1	
Rv1963c	<i>mce3R</i>	Probable transcriptional repressor (probably TetR-family) Mce3R	1.4	
Rv3660c	<i>Rv3660c</i>	Conserved hypothetical protein	2.6	
Rv1966	<i>mce3A</i>	Mce-family protein Mce3A	5.2	
Rv1969	<i>mce3D</i>	Mce-family protein Mce3D	5.3	
Rv1968	<i>mce3C</i>	Mce-family protein Mce3C	6.8	
Rv1967	<i>mce3B</i>	Mce-family protein Mce3B	7.2	
<b>regulation</b>				
Rv1388	<i>mihF</i>	Putative integration host factor MihF	-11	
Rv3597c	<i>lsr2</i>	Iron-regulated H-NS-like protein Lsr2	-4.8	
Rv1221	<i>sigE</i>	Alternative RNA polymerase sigma factor SigE	-4.5	
Rv3849	<i>espR</i>	ESX-1 transcriptional regulatory protein EspR	-3.6	
Rv0757	<i>phoP</i>	Possible TCS response transcriptional positive regulator PhoP	-2.7	
Rv2986c	<i>hupB</i>	DNA-binding protein HU homolog HupB	-1.1	
Rv0758	<i>phoR</i>	Possible TCS response sensor kinase membrane associated PhoR	1	
Rv0196	<i>Rv0196</i>	Possible transcriptional regulatory protein	13.6	
<b>lipid metabolism</b>				
Rv1169c	<i>lipX</i>	PE family protein. Possible lipase LipX.	-18	
Rv1168c	<i>PPE17</i>	PPE family protein PPE17	-14	
Rv0469	<i>umaA</i>	Possible mycolic acid synthase UmaA	-5.2	
Rv0166	<i>fadD5</i>	Probable fatty-acid-CoA ligase FadD5	-3.4	
<b>ribosomal protein genes</b>				
Rv0634B	<i>rpmG2</i>	50S ribosomal protein L33 RpmG2	-6.2	
Rv2785c	<i>rpsO</i>	30S ribosomal protein S15 RpsO	-5.6	
Rv0055	<i>rpsR1</i>	30S ribosomal protein S18-1 RpsR1	-4.5	
Rv0704	<i>rpLB</i>	50S ribosomal protein L2 RplB	-4.1	
<b>others</b>				
Rv3613c	<i>Rv3613c</i>	Hypothetical protein	-17	
Rv3612c	<i>Rv3612c</i>	Conserved hypothetical protein	-13	
Rv0197	<i>Rv0197</i>	Possible oxidoreductase	11.8	

**Figure 8.** Selected top-scoring and differentially expressed genes in *mihF*-cKD upon ATc treatment. FDR <1%, fold change = FC, light squares represent downregulated and dark squares upregulated genes.

## Discussion

NAPs have a major effect on bacterial gene expression by acting as global transcription regulators and chromosome architects. In *M. tuberculosis*, three NAPs - EspR, Lsr2 and HupB - were characterized in depth previously, while the function of a fourth NAP candidate, Rv3852, was disproved<sup>17</sup>. In this report, the systematic investigation of mIHF is presented and its potential importance for *M. tuberculosis* pathogenesis described. Like EspR<sup>7</sup>, mIHF is exclusively cytosolic, and present throughout the bacterial growth cycle, which contrasts with the findings of Sharadamma *et al.*, who found increased mIHF levels in stationary phase<sup>23</sup>. The number of mIHF binding sites (153) was similar to that of EspR (165), but less than the 305 loci contacted by Lsr2<sup>24</sup>. No common motif



**Figure 9.** Binding sites of global transcription factors in *M. tuberculosis*. **(a)** Overlap of binding sites between CRP, Lsr2, EspR and mIHF. **(b)** mIHF binding peaks relative to the transcriptional start site (TSS) and transcriptional termination site (TTS), y-axis indicates number of reads. Black profile depicts mIHF binding in exponential phase (“mIHF”), grey profile shows GC content of the H37Rv genome (“GC-content”, right y-axis).

was found for the mIHF binding sites, confirming the non-specific DNA-binding demonstrated *in vitro*<sup>12</sup> and the broad mIHF ChIP-peaks may indicate binding to structured DNA rather than to a specific sequence.

The ATc-dependent silencing of *mihF* expression confirmed the predicted essentiality of the gene and demonstrated that mIHF is required not only for multiplication but also for bacterial survival. Indeed, a dramatic effect on cell morphology and global physiology was noticed upon depletion of the protein. At least two passages were necessary to decrease mIHF levels and this is consistent with the remarkably high abundance of mIHF reported in exponential phase as well as during hypoxia-induced dormancy<sup>25</sup>. Intriguingly, despite the effects in cell elongation, cell shape and septum formation observed upon mIHF depletion, no gene known to be involved in these processes was heavily deregulated. Expression of *ftsZ*, coding for the cell division initiator protein, and *wag31*, whose product localizes to the septum<sup>26</sup>, were marginally downregulated, whereas *ftsK* and *ftsQ*, necessary for septum formation<sup>27</sup>, were only slightly upregulated.

Overall, genes harbouring an mIHF binding site within their coding sequence or located less than 500bp away were downregulated upon mIHF depletion although mIHF peaks were still detected. The mIHF binding profiles in exponentially growing cells as well as in non-permissive conditions were essentially the same but differed in peak width since after mIHF depletion the peaks were generally narrower. The strong, but not complete, correlation between mIHF peaks in exponential phase (153) and after depletion (124), with 62 overlapping signals, suggests that mIHF probably binds to a core set of genes with higher affinity. Examples are represented by the *rv0097* and *espA* loci, which were found to be more enriched in the depleted condition than in the exponentially growing phase. The numerous differentially regulated genes identified upon mIHF depletion indicated a pleiotropic effect with a strong bias towards virulence-related genes. Most prominent was the 20-fold repression of one of the main virulence operons, *espACD*, which is critical for ESX-1 function, EsxA/B secretion and full virulence of *M. tuberculosis*<sup>28</sup>.

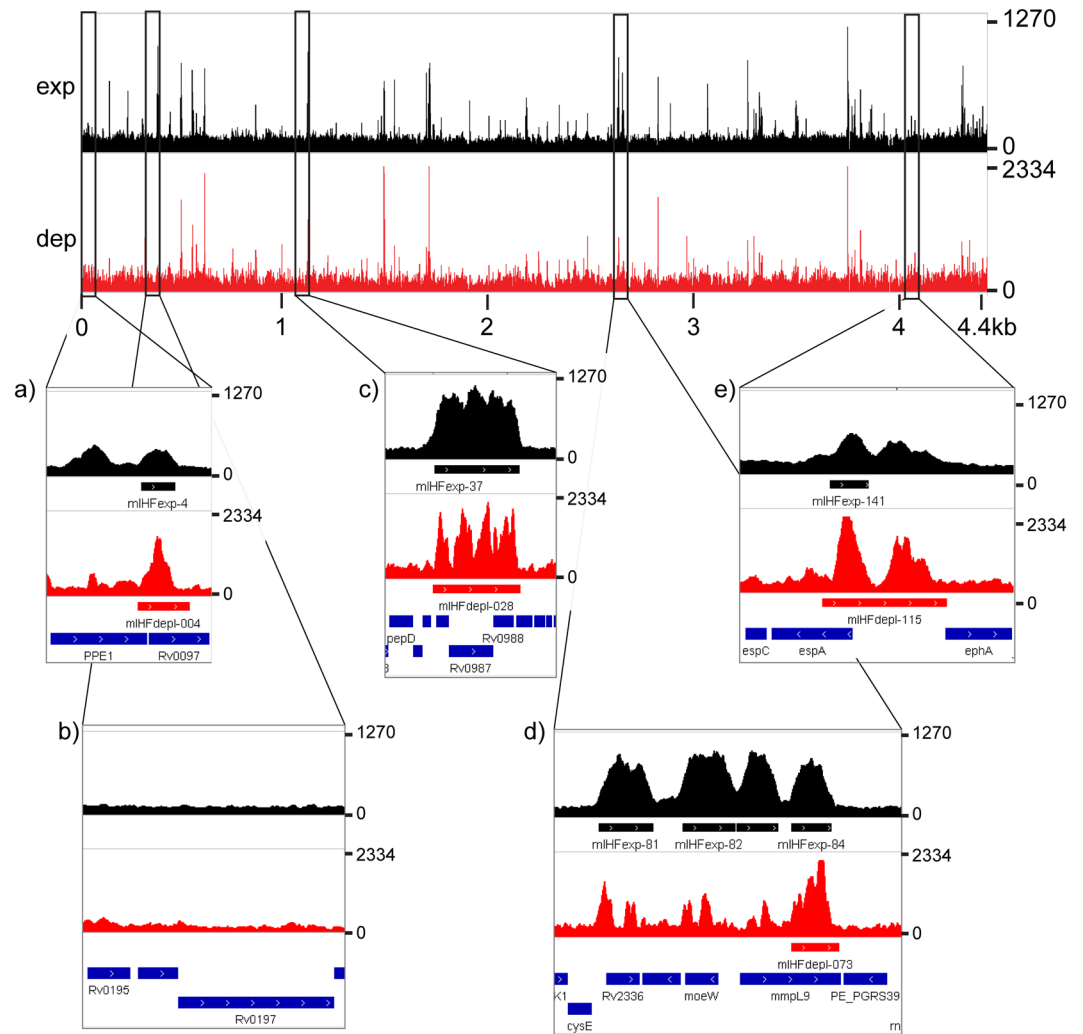
The region preceding the *espACD* operon, where mIHF binds, has been deleted from the genome of some tubercle bacilli during evolution of the *M. tuberculosis* complex and is referred to as region of difference 8, RD8<sup>29</sup>. It is especially striking that RD8 contains binding sites for numerous regulatory proteins including the NAPs EspR, Lsr2 and mIHF, as well as the activator CRP. In addition, RD8 is contacted by the response regulators MprA<sup>30</sup> and PhoP of two different two component systems, MprAB and PhoPR<sup>31</sup>. To overcome mutations in PhoR, that decrease virulence, L6 strains of *M. africanum* and all animal-adapted tubercle bacilli have deleted RD8 thereby removing a regulatory layer and restoring full virulence<sup>31</sup>.

In the same vein, expression of PE11 (*rv1169c*, LipX), which is necessary for survival of *M. tuberculosis* in *ex vivo* models<sup>32</sup>, was repressed 18-fold following mIHF depletion. LipX modulates glycolipid-synthesis<sup>33</sup>, its down-regulation reduces PDIM (phthiocerol dimycoserolate) levels in *M. tuberculosis*, leading to enhanced cellular aggregation and biofilm formation<sup>34</sup>. PDIMs are required for phagosomal escape<sup>35</sup> and to do so they act concertedly with ESX-1<sup>36</sup>. Thus, another important virulence factor is controlled by mIHF. Biofilms are associated with persistent infections of *M. tuberculosis*<sup>37</sup> and contain mycolic acids in their extracellular matrix<sup>38</sup>. In addition to *lipX*, several genes encoding mycolic acid tailoring enzymes (e.g. *umaA*, *mmaA3*) were expressed at lower levels in mIHF-depleted cells, implying that mIHF also affects biofilm formation and mycolic acid synthesis, consistent with the results of our GO survey. Furthermore, expression of *ppe17*, situated 17bp downstream of *lipX*, was repressed 13.5-fold. PPE17 was shown to be upregulated in macrophages<sup>39,40</sup>, again indicating the importance of mIHF for successful infection.

Apart from control of virulence genes, RNA, DNA and protein synthesis pathways were also found to be deregulated. For example, the gene for the epsilon subunit (*dnaQ*) of DNA polymerase III was 3.2-fold upregulated, while *dnaN* and *dnaA* were 4.6- and 5.7-fold repressed, respectively. DnaN acts as a bridge between the alpha and epsilon subunit and plays a regulatory role on the replicase<sup>41</sup>. Absence of the DnaA chromosomal replication initiator and DnaN proteins should reduce initiation of DNA replication, and this is consistent with the lower nucleic acid synthesis observed by uracil-incorporation.

Most of the sigma factors were not differentially expressed upon mIHF depletion, except for *sigE*, which was >4-fold downregulated. This alternative sigma factor is activated in various stress conditions, such as inside macrophages<sup>42</sup>, and is required for interrupting phagosome maturation<sup>43</sup>. SigE positively regulates *mprAB* and induces the stringent response in *M. smegmatis*<sup>44</sup>. As a consequence of the downregulation of *sigE* expression, neither the stringent response nor the SOS response to DNA damage were induced upon mIHF depletion, which





**Figure 10.** mIHF contacted regions on the H37Rv chromosome. The upper panel shows the global distribution of mIHF in exponentially growing cells (exp, in black) and in mIHF-depleted cells (dep, red). The number of reads is indicated on the y-axis, while gene length and names are shown on the x-axis. Zoomed images show examples of mIHF binding. (a) mIHF peak at *rv0097*, whose transcription is 10.5-fold downregulated upon mIHF depletion. (b) No peak is located close to the top upregulated genes *rv0196* and *rv0197*. (c) Third top-scoring peak cluster close to *mprA/mprB*. (d) Accumulation of peaks around *mmpL9*. (e) *EspA* extended promoter region.

implies that mIHF does not play a direct role in facing starvation. Concerning protein synthesis, the low amount of tRNAs and downregulation of ribosomal protein genes were reflected in the poor incorporation of radiolabelled leucine by mIHF-depleted cells.

Taken together, our results show that in *M. tuberculosis* mIHF is an essential, highly stable, cytosolic NAP whose gene repertoire overlaps with that of several other NAPs and regulatory proteins. mIHF likely influences gene expression by organizing chromatin structure and this will be probed using single cell analysis and chromatin conformation capture techniques to determine whether it acts alone or in association with other chromatin regulators.

## Methods

**Strains, media and chemicals.** *M. tuberculosis* H37Rv and *mihF*-cKD strains were grown at 37 °C either in Middlebrook 7H9 broth (Difco) supplemented with 10% albumin-dextrose-catalase, 0.2% glycerol and 0.05% Tween 80 or in Sauton's liquid medium supplemented with 0.005% Tween 80. Cultures were plated on Middlebrook 7H10 (Difco) agar supplemented with 10% oleic acid-albumin-dextrose-catalase and 0.2% glycerol. Hygromycin (50 µg ml<sup>-1</sup>), kanamycin (25 µg ml<sup>-1</sup>), streptomycin (25 µg ml<sup>-1</sup>), 2.5% sucrose or Anhydrotetracycline (ATc, Clontech, 600 ng ml<sup>-1</sup>) were added when needed. For cloning procedures, One shot<sup>®</sup> TOP10 chemically competent *Escherichia coli* (Invitrogen) were grown in Luria-Bertani (LB) broth or on LB agar with hygromycin (200 µg ml<sup>-1</sup>), kanamycin (50 µg ml<sup>-1</sup>) or spectinomycin (25 µg ml<sup>-1</sup>). All chemicals were purchased from Sigma-Aldrich, unless otherwise stated. Experiments involving *M. tuberculosis* have been carried

out in a Biosafety Level 3 (BSL3) laboratory, according to the national and international guidelines (Authorization number A070027/3).

**Plasmid and conditional knockdown mutant construction.** In order to promote homologous recombination, 1 kbp up- and downstream regions of full length *mihF* were generated by PCR amplification using primers *mihF*-UF/*mihF*-UR and *mihF*-DF/*mihF*-DR (listed in Supplementary Table 1) respectively. Fragments were ligated in-frame with the AvrII site and cloned into the PacI and AscI sites of pJG1100, resulting in the suicide vector pCS35. The complementing plasmid pCS31 was constructed by cloning the full-length *mihF* gene, amplified with primers *mihF*-F/*mihF*-R, into pGA44 under control of the repressible *ptr* promoter<sup>45</sup>. The two genes located downstream of *mihF* (*gmk* and *rpoZ*), were amplified with primers *gmk*-F/*rpoZ*-R. The constitutively active promoter *PfurA102*<sup>46</sup> was amplified with primers *PfurA102*-F/*PfurA102*-R. An overlap PCR was performed to fuse *PfurA102* with *gmk-rpoZ*, and then cloned into pCS31, resulting in pNO12. The complementing vectors harbouring *mihF*-80 and *mihF*-86, as well as *PfurA102-gmk-rpoZ*, were constructed in pGA118, a derivative of pGA44, which carries a hygromycin instead of a streptomycin resistance cassette, resulting in pNO62 and pNO63, respectively. Expression of *gmk* and *rpoZ* by the complementing vector was necessary to compensate for the polar effects caused by the deletion of the *mihF* gene.

Deletion of the full-length *mihF* gene was obtained by homologous recombination using plasmid pCS35. After transformation of *M. tuberculosis* H37Rv, the first recombination event was selected on 7H10 plates, supplemented with hygromycin and kanamycin. Colonies were screened by colony PCR using CS-402/CS-403 and CS-404/CS-405 primer pairs. The merodiploid strain was generated by integration of plasmid pNO12 at the *L5 attB* site and transformants were plated on 7H10 with hygromycin, kanamycin and streptomycin. Finally, deletion of the wild type gene by allelic exchange and generation of the *mihF* cKD strain was accomplished by plating the bacteria on 7H10 supplemented with streptomycin and 2.5% sucrose. The resulting colonies were tested by PCR with primers CS-403/CS-415 for deletion of *mihF* from its native locus and confirmed by Southern blot.

**Genomic DNA extraction and Southern Blot.** Mycobacterial genomic DNA was extracted using standard protocols. To confirm successful allelic exchange of *mihF*, genomic DNA was digested with AvrII and PvuII restriction enzymes. DNA fragments were separated by 0.8% agarose gel electrophoresis before capillary blotting onto a Hybond-N+ nylon membrane (GE Healthcare) and hybridization with a probe corresponding to the same upstream and downstream regions of *mihF* cloned into pJG1100. Hybridization was carried out using the ECL Direct Nucleic Acid Labelling and Detection System (GE Healthcare) as recommended by the manufacturer.

**Growth curve measurements, colony forming unit counts, 3H-leucine and 3H-uracil incorporation.** To characterize the growth of the *mihF* cKD mutant, the strains were grown to mid-logarithmic phase and then diluted to an optical density at 600 nm ( $OD_{600}$ ) of 0.05 in 7H9 medium. ATc was added to 600 ng ml<sup>-1</sup> and the  $OD_{600}$  was recorded at different time points to obtain the growth curves. As ATc is light sensitive and depleted over time, cultures were diluted to  $OD_{600} = 0.05$  every three days.

Nucleic acid and protein synthesis were measured by incorporating tritium-labelled leucine and uracil as previously described<sup>47</sup>. As mycobacteria do not incorporate exogenous thymidine, but can use uracil for RNA as well as for DNA synthesis after methylation<sup>47</sup>, only uracil was used to assess total nucleic acid production. Strains H37Rv with ATc, *mihF*-cKD with and without ATc (600 ng ml<sup>-1</sup>) were grown to mid-exponential phase, diluted to  $OD_{600} = 0.1$  and then subsequently diluted again to the same  $OD = 0.1$  every day. Of these cultures, 1 ml was incubated daily with 1  $\mu$ Ci <sup>3</sup>H-uracil and <sup>3</sup>H-leucine, respectively. After 24 h, the sample was washed once in PBS supplemented with 0.05% Tween 80, the cells were harvested by centrifugation and stored at -80 °C until further processing. Counts per min were measured by suspending the sample in 5 ml Ecoscint XR (National diagnostics) on a Beckman Coulter LS6500 Multi-Purpose Scintillation Counter. At every sampling point, the number of colony-forming units (CFU) per millilitre of culture was evaluated and protein samples were taken at days 0, 2, 5 and 9. Cumulative protein and nucleic acid incorporation was derived as the sum of the daily incorporation multiplied by the CFU for the total incorporation relative to day 0, multiplied by the dilution factor, to account for the growth rate during the 24 h incubation. CFU were evaluated from serial dilutions of *M. tuberculosis* cultures plated on 7H10 plates and cumulative CFU calculated similarly by summing the previous CFU with the daily CFU multiplied by the dilution factor necessary to reach  $OD = 0.1$  to normalize all three samples. Growth rate was calculated as the daily  $OD_{600}$  divided by the target  $OD_{600}$  of 0.1.

**Scanning electron microscopy.** For surface scanning electron microscopy, *mihF*-cKD mutant without ATc and H37Rv wild type strains were grown in 7H9 until mid-exponential phase, pelleted, washed in PBS and resuspended to  $OD_{600} = 0.5$ . The *mihF*-cKD mutant with ATc was diluted three times in fresh ATc-containing medium and then subjected to the same protocol. Samples were then fixed on a coverslip in a solution of 1.25% glutaraldehyde, 1% tannic acid in phosphate buffer (0.1 M, pH = 7.4) for 1 h, washed in PBS prior to fixing for 30 min in 1% osmium tetroxide. The samples were then dehydrated in a graded alcohol series and dried by passing through the supercritical point of carbon dioxide (Leica Microsystems CPD300) and coated with a 2 nm layer of osmium metal using an osmium plasma coater (Filgen OPC60). Scanning electron microscopy images were taken using a field emission scanning electron microscope (Merlin, Zeiss NTS) with an acceleration voltage of 2 kV and the in-lens secondary electron detector. Cell length of bacteria was measured in ImageJ (n = 110 for H37Rv wild type, n = 96 for *mihF*-cKD without ATc and n = 98 for *mihF*-cKD + ATc). Two-tailed, one-way analysis of variance (ANOVA) with Kruskal-Wallis post-test was performed with 303 degrees of freedom and F value = 184.2.

**Fluorescence microscopy.** Strains H37Rv and *mihF*-cKD were grown to exponential phase in 7H9 Middlebrook media without ATc, while *mihF*-cKD plus ATc was diluted three times in fresh medium with

600 ng ml<sup>-1</sup> ATc. Samples were incubated with SYTO9 (4 μM), which stains the DNA, and FM4-64 (5 μg ml<sup>-1</sup>) to stain the membrane for 20 min at 37 °C. Bacteria were mounted on an agarose pad and imaged with an Olympus IX81 microscope under a 100x objective. Representative images were selected and single channel and composite images were adjusted for brightness and contrast in ImageJ.

**Total RNA extraction and 5' rapid amplification of cDNA ends.** *M. tuberculosis* H37Rv and *mihF*-cKD cultures were harvested by centrifugation, pellets were resuspended in TRIzol Reagent (ThermoFisher) and stored at -80 °C until further processing. Total RNA was extracted by bead-beating as previously described<sup>48</sup>. Integrity of RNA was checked by agarose gel electrophoresis, purity and amount of RNA were assessed using a Nanodrop instrument and Qubit Fluorometric Quantitation (ThermoFisher) respectively. SuperScript III First-Strand Synthesis System (Invitrogen) was used to generate randomly primed cDNA from 500 ng of RNA, according to the manufacturer's recommendations. Primers CS-057/CS-058 for *sigA* were used to normalize the amount of cDNA template added to each sample.

For the 5'-RACE, 2 μg of *M. tuberculosis* H37Rv RNA and 1 μg of primer NO-095 were incubated at 70 °C for 5 min and then at 55 °C for 1 h in the presence of 1x cDNA synthesis buffer, 1 mM dNTPs, 40 U RNase inhibitor, 25 U Transcriptor Reverse Transcriptase (5'/3' RACE Kit, 2nd Generation, Roche). cDNA was then purified with the High Pure PCR Product Purification kit (Roche), and used in the subsequent poly(A) tailing reaction (30 min at 37 °C in the presence of 0.2 mM dATP and 80 U Terminal Transferase, Roche). Semi-nested PCR amplification on poly(A)-tailed cDNA was performed using an oligo dT-anchor primer (CS-080) and primer NO-094. Three amplification products were obtained, cloned into pTOPO (Invitrogen) and sequenced.

**RNA sequencing and analysis.** RNA was extracted from biological duplicate samples from exponential phase H37Rv and three-times diluted *mihF*-cKD as described above. The ribosomal RNA was depleted with the Ribo-Zero rRNA Removal Kit for Gram-positive Bacteria (Illumina), following the manufacturer's instructions. Libraries were prepared by the Lausanne Genomic Technologies Facility, using the Truseq Stranded mRNA Library Prep kit reagents (Illumina) according to the manufacturer's recommendations. The multiplexed libraries were sequenced on a HiSeq. 2500 instrument using TruSeq SBS Kit V4 reagents as single-end 100 nt-long reads. Sequencing data were processed using the Illumina Pipeline Software version 1.84. Reads were adapter- and quality-trimmed with Trimmomatic v0.33<sup>49</sup>. The quality settings were "SLIDINGWINDOW:5:15 MINLEN:40". Reads were aligned with Bowtie2<sup>50</sup>. Counting reads over annotated features was done with featureCounts<sup>51</sup>. Annotation was taken from TubercuList release R27 (<http://tuberculist.epfl.ch/>). Differential gene expression analysis was done using DESeq2<sup>52</sup>.

For gene ontology (GO) enrichment analysis, a cut-off of 4-fold differentially expressed genes was chosen. The GO annotation was retrieved from BioCyc<sup>53</sup> and analysis was performed with TopGO<sup>54</sup>. The conservative weighted algorithm and Fisher's exact test were used to calculate p-values for enrichment of GO terms in biological processes, molecular functions and cellular components in up- and downregulated genes upon mIHF depletion. The tree was pruned to a node size of 5 to exclude statistical artefacts of small sized GO terms.

**Protein extraction, immunoblot analysis and subcellular fractionation.** *M. tuberculosis* cells grown in 7H9 were pelleted at different time points by centrifugation, washed once in Tris-Buffered Saline (TBS, 20 mM Tris-HCl pH 7.5, 150 mM NaCl) and stored at -80 °C until further processing. Cells were sonicated in TBS supplemented with a protease inhibitor tablet (cOmplete, mini, EDTA free, Roche) for 15 min and the protein solution was then sterilized by filtration through a 0.2 μm filter to remove any residual intact cells. Protein samples were quantified using the Qubit Fluorometric Quantitation device (ThermoFisher). Equal amounts of protein preparations were loaded on SDS-PAGE 12–15% NuPAGE gels (Invitrogen) and transferred onto PVDF membranes using a semidry electrophoresis transfer apparatus (Invitrogen). Membranes were incubated in TBS-Tween blocking buffer (25 mM Tris pH 7.5, 150 mM NaCl, 0.05% Tween 20) with 5% w/v skimmed milk powder for 3 h at 4 °C prior to overnight incubation with primary antibody. Membranes were washed in TBS-Tween three times, and then incubated with secondary antibody for 2 h before washing. Signals were detected using Chemiluminescent Peroxidase Substrate 1 (Sigma-Aldrich).

Primary monoclonal anti-mIHF antibody was produced by Alere against recombinant mIHF-86 and used at a concentration of 1:2,000 in immunoblots. Horseradish peroxidase (HRP) conjugated Goat anti-mouse Kappa (SouthernBiotech) secondary antibody was used at a 1:12,000 dilution. Anti-RpoB antibodies (NeoClone) were used to detect RpoB, the internal loading control. Band intensity of immunoblots was analysed with Fiji/ImageJ and normalized to the intensity of the RpoB signal.

Cell fractions were obtained as described previously<sup>55</sup>. Briefly, H37Rv was grown in Sauton's medium with 0.005% Tween 80 to mid-exponential phase, cells were collected by centrifugation, and supernatant was filtered and concentrated 100 x to obtain the secreted fraction. The pellet was treated with 0.25% Genapol-X080 for 30 min followed by centrifugation at 14,000 g for 10 min and the proteins of the resulting supernatant precipitated with TCA, yielding the capsular fraction. The remaining pellet was subjected to sonication to break the cells, sterilized by filtration through a 0.2 μm filter followed by ultra-centrifugation at 45,000 rpm for 1 h. The supernatant contained the cytosolic fraction, while the pellet was enriched with membrane proteins.

**Chromatin Immuno Precipitation (ChIP) and library construction.** ChIP was performed as previously described<sup>56</sup>. Briefly, exponentially growing *M. tuberculosis* H37Rv liquid cultures (for input control, EspR ChIP and mIHF ChIP) or mIHF depleted *mihF*-cKD mutant were cross-linked with formalin 1% for 10 min and quenched with glycine 125 mM for 10 min, washed twice in Tris-buffered saline (TBS, pH 7.5) and sonicated on a Diagenode Bioruptor with 30 sec on/off cycles for 10 min on high settings to shear DNA to 200–500 bp fragments. Immunoprecipitation was performed with monoclonal anti-mIHF (Alere) or polyclonal rat anti-EspR

antibodies (Statens Serum Institut, Copenhagen, Denmark) in 1 ml immunoprecipitation (IP) buffer (containing 50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% Triton X-100, 0.1% (w/v) sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS) and one protease inhibitor cocktail tablet (Roche)) overnight at 4 °C. 100 µl Dynabeads sheep anti-rag IgG (DynaL Biotech) for EspR and 100 µl per sample anti-ProteinL magnetic beads (Pierce) for mIHF were pre-saturated with 1 mg ml<sup>-1</sup> bovine serum albumin and 0.1 mg ml<sup>-1</sup> salmon sperm DNA, then incubated with each corresponding sample for 4 h. The IP was washed 5 times with increasing stringency buffers as described previously<sup>7</sup>. Input control was not incubated with antibody. Libraries were prepared with the NEBNext Ultra II DNA kit (NEB) by the Lausanne Genomic Technologies Facility following the manufacturer's instructions, multiplexed and sequenced on a HiSeq 2500 instrument.

**ChIP-seq data analysis.** Alignment was performed with Bowtie2<sup>50</sup> against the H37Rv genome (NCBI NC\_000962.2) and resulted in 8.1 M uniquely aligned reads for EspR, 1.1 M for mIHF in exponentially growing cells, 7.3 M for mIHF in mIHF depleted cells and 3.8 M for the input, respectively. For mIHF ChIP in exponentially growing cells, two different concentrations of antibodies were tested. These two datasets were pooled, as the Spearman correlation of 0.96 was excellent (Fig. S5). HOMER<sup>57</sup> was used for peak calling using the dynamic peak size algorithm with the input as a control. The enrichment is calculated relative to the input sequence, which might not reflect the real peak size. Resulting peaks were manually curated, subsequently annotated and further analysed with BEDtools<sup>58</sup> and deepTools<sup>59</sup>. Profile plots of mIHF binding sites in exponentially growing cells and in mIHF-depleted cells, as well as GC content of H37Rv were generated with deepTools at binsize = 25 and visualized with IGV<sup>60</sup>. Coordinates of transcriptional start sites were taken from Cortes *et al.*<sup>18</sup>. CRP binding sites were extracted from<sup>61</sup> and Lsr2 from<sup>24</sup>. Minch *et al.* identified 305 Lsr2 binding sites by ChIP-seq, the same technique used here for mIHF and EspR. Additionally, the Minch *et al.* binding sites represent a subset of the >800 sites identified by an earlier ChIP-chip experiment<sup>62</sup>, we decided to use the Minch *et al.* dataset for our analysis. ChIPPeakAnno was used to find overlapping peaks. Connected peaks were merged, resulting in an overall lower sum of peaks for each transcription factor<sup>63</sup>.

Genes associated with mIHF binding sites (<500 bp distance to feature boundary) were analysed by ANOVA with Bonferroni's multiple comparison test to test for a correlation in gene expression. N = 331 for genes associated with mIHF binding and n = 3785 for genes not close to a mIHF peak in mIHF depleted cells. N = 330 for genes associated with mIHF binding in exponential phase, and n = 3831 for genes without any mIHF peak in close proximity. F-value was 147.2 and degrees of freedom = 8276. GC content of binding sites was extracted from the H37Rv genome (NC\_000962.3), with n = 162, 153 and 124 for EspR binding sites, mIHF binding sites in exponentially growing cells and mIHF binding sites in mIHF-depleted cells, respectively. All 4111 features were analysed for the overall GC content of *M. tuberculosis*. One-way, two-tailed ANOVA with Bonferroni's multiple comparison test was performed in GraphPad Prism with an F-value of 583.7 and 4549 degrees of freedom.

## Data Availability

Raw and processed data of RNA- and ChIP-seq data are available at the NCBI Gene Expression Omnibus repository under accession number GSE111194, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111194>.

## References

- Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F. R. & Smith, I. Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **31**, 715–724 (1999).
- Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–95 (2010).
- Azam, T. A., Iwata, A., Nishimura, A., Azam, T. A. L. I. & Ueda, S. Growth Phase-Dependent Variation in Protein Composition of the *Escherichia coli* Nucleoid. *J. Bacteriol.* **181**, 6361–6370 (1999).
- Browning, D. F., Grainger, D. C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* **13**, 773–80 (2010).
- Pandey, S. D. *et al.* Iron-Regulated Protein HupB of *Mycobacterium tuberculosis* Positively Regulates Siderophore Biosynthesis and Is Essential for Growth in Macrophages. *J. Bacteriol.* **196**, 1853–65 (2014).
- Gordon, B. R. G., Imperial, R., Wang, L., Navarre, W. W. & Liu, J. Lsr2 of *Mycobacterium tuberculosis* represents a novel class of H-NS-like proteins. *J. Bacteriol.* **190**, 7052–9 (2008).
- Blasco, B. *et al.* Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. *PLoS Pathog.* **8**, e1002621 (2012).
- Garcés, A. *et al.* EspA acts as a critical mediator of ESX1-dependent virulence in *Mycobacterium tuberculosis* by affecting bacterial cell wall integrity. *PLoS Pathog.* **6**, 15–16 (2010).
- Lee, M. H. & Hatfull, G. F. Mycobacteriophage L5 Integrase-Mediated Site-Specific Integration *In Vitro*. *J. Bacteriol.* **175**, 6836–6841 (1993).
- Cole, S. T. *et al.* Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011 (2001).
- Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **396**, 651–653 (1998).
- Mishra, A. *et al.* Integration Host Factor of *Mycobacterium tuberculosis*, mIHF, Compacts DNA by a Bending Mechanism. *PLoS One* **8**, e69985 (2013).
- de Souza, G., Leversen, N., Målen, H. & Wiker, H. G. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *J. Proteomics* **75**, 502–10 (2011).
- Målen, H., Pathak, S., Sjøteland, T., de Souza, G. A. & Wiker, H. G. Definition of novel cell envelope associated proteins in Triton X-114 extracts of *Mycobacterium tuberculosis* H37Rv. *BMC Microbiol.* **10**, 132 (2010).
- Schubert, O. T. *et al.* The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe* **13**, 602–12 (2013).
- Griffin, J. E. *et al.* High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7**, e1002251 (2011).
- Odermatt, N. T. *et al.* Rv3852 (H-NS) of *Mycobacterium tuberculosis* Is Not Involved in Nucleoid Compaction and Virulence Regulation. *J. Bact.* **199**, 1–12 (2017).

18. Cortes, T. *et al.* Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* 1121–1131 <https://doi.org/10.1016/j.celrep.2013.10.031> (2013).
19. Newton-Foot, M. & Gey van Pittius, N. C. The complex architecture of mycobacterial promoters. *Tuberculosis (Edinb)*. **93**, 60–74 (2013).
20. Boldrin, F. *et al.* Development of a repressible mycobacterial promoter system based on two transcriptional repressors. *Nucleic Acids Res.* **38**, e134 (2010).
21. Dahl, J. L. Electron microscopy analysis of *Mycobacterium tuberculosis* cell division. *FEMS Microbiol. Lett.* **240**, 15–20 (2004).
22. Becq, J. *et al.* Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol. Biol. Evol.* **24**, 1861–71 (2007).
23. Sharadamma, N. *et al.* Molecular Dissection of *Mycobacterium tuberculosis* Integration Host Factor Reveals Novel Insights into the Mode of DNA Binding and Nucleoid Compaction. *J. Biol. Chem.* **289**, 34325–34340 (2014).
24. Minch, K. J. *et al.* The DNA-binding network of *Mycobacterium tuberculosis*. *Nat. Commun.* **6**, 5829 (2015).
25. Schubert, O. T. *et al.* Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* **18**, 96–108 (2015).
26. Kieser, K. J. & Rubin, E. J. How sisters grow apart: mycobacterial growth and division. *Nat. Rev. Microbiol.* **12**, 550–562 (2014).
27. Slayden, R. A., Knudson, D. L. & Belisle, J. T. Identification of cell cycle regulators in *Mycobacterium tuberculosis* by inhibition of septum formation and global transcriptional analysis. *Microbiology* **152**, 1789–1797 (2006).
28. Chen, J. M. *et al.* EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194**, 884–93 (2012).
29. Brosch, R. *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci.* **99**, 3684–3689 (2002).
30. Pang, X. *et al.* MprAB regulates the espA operon in *Mycobacterium tuberculosis* and modulates ESX-1 function and host cytokine response. *J. Bacteriol.* **195**, 66–75 (2013).
31. Orgeur, M. & Brosch, R. Evolution of virulence in the *Mycobacterium tuberculosis* complex. *Curr. Opin. Microbiol.* **41**, 68–75 (2018).
32. Narayana, Y., Joshi, B., Katoch, V. M., Mishra, K. C. & Balaji, K. N. Differential B-cell responses are induced by *Mycobacterium tuberculosis* PE antigens Rv1169c, Rv0978c, and Rv1818c. *Clin. Vaccine Immunol.* **14**, 1334–1341 (2007).
33. Singh, P. *et al.* PE11, a PE/PPE family protein of *Mycobacterium tuberculosis* is involved in cell wall remodeling and virulence. *Sci. Rep.* **6**, 21624 (2016).
34. Rastogi, S. *et al.* Down-regulation of PE11, a cell wall associated esterase, enhances the biofilm growth of *Mycobacterium tuberculosis* and reduces cell wall virulence lipid levels. 52–61 (2017).
35. Quigley, J. *et al.* The cell wall lipid PDIM contributes to phagosomal escape and host cell exit of *Mycobacterium tuberculosis*. *MBio* **8**, 1–12 (2017).
36. Augenstein, J. *et al.* ESX-1 and phthiocerol dimycocerosates of *Mycobacterium tuberculosis* act in concert to cause phagosomal rupture and host cell apoptosis. *Cell. Microbiol.* **19** (2017).
37. Richards, J. P. & Ojha, A. K. Mycobacterial Biofilms. 1–11, <https://doi.org/10.1128/microbiolspec.MGM2-0004-2013>. **Correspondence** (2014).
38. Ojha, A. K. *et al.* Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol. Microbiol.* **69**, 164–174 (2008).
39. Schnappinger, D. *et al.* Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J. Exp. Med.* **198**, 693–704 (2003).
40. Donà, V. *et al.* ThePPE Domain of PPE17 Is Responsible for Its Surface Localization and Can Be Used to Express Heterologous Proteins on the Mycobacterial Surface. *PLoS One* **8**, 1–8 (2013).
41. Gu, S. *et al.* The  $\beta$ 2 clamp in the *Mycobacterium tuberculosis* DNA polymerase III  $\alpha\beta\epsilon$  replicase promotes polymerization and reduces exonuclease activity. *Sci. Rep.* **6**, 18418 (2016).
42. Manganelli, R. Sigma Factors: Key Molecules in *Mycobacterium tuberculosis* Physiology and Virulence. *Microbiol. Spectr.* 1–23, <https://doi.org/10.1128/microbiolspec.MGM2-0007-2013>. **Correspondence** (2014).
43. Casonato, S., Proveddi, R., Dainese, E., Palù, G. & Manganelli, R. *Mycobacterium tuberculosis* Requires the ECF Sigma Factor SigE to Arrest Phagosome Maturation. *PLoS One* **9**, e108893 (2014).
44. Sureka, K. *et al.* Polyphosphate kinase is involved in stress-induced mprAB-sigE-rel signalling in mycobacteria. *Mol. Microbiol.* **65**, 261–276 (2007).
45. Kolly, G. S. *et al.* Assessing the essentiality of the decaprenyl-phospho-d-arabinofuranose pathway in *Mycobacterium tuberculosis* using conditional mutants. *Mol. Microbiol.* **92**, 194–211 (2014).
46. Sala, C. *et al.* *Mycobacterium tuberculosis* FurA Autoregulates Its Own Expression. *J. Bacteriol.* **185**, 5357–5362 (2003).
47. Wayne, L. G. Synchronized replication of *Mycobacterium tuberculosis*. *Infect. Immun.* **17**, 528–530 (1977).
48. Jungwirth, B. *et al.* High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in *Corynebacterium glutamicum*. *Microbiology* **159**, 12–22 (2012).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
51. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
53. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* **44**, D471–D480 (2016).
54. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology (2016).
55. Lou, Y., Rybnik, J., Sala, C. & Cole, S. T. EspC forms a filamentous structure in the cell envelope of *Mycobacterium tuberculosis* and impacts ESX-1 secretion. *Mol. Microbiol.* **103**, 1–39 (2016).
56. Hartkoorn, R. C. *et al.* Genome-wide definition of the SigF regulon in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194**, 2001–9 (2012).
57. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).
58. Quinlan, A. R. BEDTools: The Swiss-Army tool for genome feature analysis. *Current Protocols in Bioinformatics* **2014** (2014).
59. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, 160–165 (2016).
60. Robinson, J. T. *et al.* Integrative Genomic Viewer. *Nature* **29**, 24–26 (2011).
61. Kahramanoglou, C. *et al.* Genomic mapping of cAMP receptor protein (CRPMt) in *Mycobacterium tuberculosis*: Relation to transcriptional start sites and the role of CRPMt as a transcription factor. *Nucleic Acids Res.* **42**, 8320–8329 (2014).
62. Gordon, B. R. G. *et al.* Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **107**, 5154–9 (2010).
63. Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).

## Acknowledgements

We thank the Lausanne Genomic Technologies Facility at the University of Lausanne and the BIOEM platform at EPFL for technical assistance, Joe Buechler from Alere for production of antibodies and Gaëlle Kolly for the construction of pGA118. This work was supported by the Swiss National Science Foundation (grant 31003A-162641 to STC).

## Author Contributions

N.T.O. and C.S. generated the *mihF*-cKD strain and conceived the experiments, N.T.O. performed all experiments. A.B. mapped the RNA-seq data and N.T.O. mapped ChIP-seq data and performed all bioinformatics analysis of RNA- and ChIP-seq and all other results. N.T.O., C.S. and S.T.C. wrote the manuscript. All authors read and commented on the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-32340-2>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018