

# SCIENTIFIC REPORTS



OPEN

## *The alr-groEL1 operon in Mycobacterium tuberculosis: an interplay of multiple regulatory elements*

Aadil H. Bhat\*, Deepika Pathak\* &amp; Alka Rao

Received: 05 October 2016

Accepted: 30 January 2017

Published: 03 March 2017

Threonylcarbamoyladenosine is a universally conserved essential modification of tRNA that ensures translational fidelity in cellular milieu. TsaD, TsaB and TsaE are identified as tRNA-A<sub>37</sub>-threonylcarbamoyl (t<sup>6</sup>A)-transferase enzymes that have been reconstituted *in vitro*, in few bacteria recently. However, transcriptional organization and regulation of these genes are not known in any of these organisms. This study describes the intricate architecture of a complex multicistronic *alr-groEL1* operon, harboring essential genes, namely *tsaD*, *tsaB*, *tsaE*, *groES*, *groEL1*, and *alr* (required for cell wall synthesis), and *rimI* encoding an N- $\alpha$ -acetyltransferase in *Mycobacterium tuberculosis*. Using northern blotting, RT-PCR and *in vivo* fluorescence assays, genes *alr* to *groEL1* were found to constitute an ~6.3 kb heptacistronic operon with multiple internal promoters and an I-shaped intrinsic hairpin-like *cis*-regulatory element. A strong promoter *PtsaD* within the coding sequence of *rimI* gene is identified in *M. tuberculosis*, in addition. The study further proposes an amendment in the known bicistronic *groESL1* operon annotation by providing evidence that *groESL1* is co-transcribed as sub-operon of *alr-groEL1* operon. The architecture of *alr-groEL1* operon, conservation of the genetic context and a mosaic transcriptional profile displayed under various stress conditions convincingly suggest the involvement of this operon in stress adaptation in *M. tuberculosis*.

Survival of a cell depends on its ability to maintain translational fidelity and accuracy. Transfer-RNA (tRNA), along with the large number of modifications that adorn its structure, is a major player contributing towards accurate decoding of mRNA<sup>1</sup>. In tRNAs that decode ANN codons, adenosine at position 37 is modified with N<sup>6</sup>-threonylcarbamoyladenosine (t<sup>6</sup>A). The importance of this modification is emphasized by its presence in all domains of life. This modification stabilizes the tRNA anticodon stem loop structure, enhances codon-anticodon architecture and prevents mispairing between first base of codon and third base of anticodon. In the absence of t<sup>6</sup>A, frame shifting is found to increase especially at sequences with tandem ANN codons<sup>2,3</sup>.

While this modification was discovered four decades ago, the enzymes involved have recently been elucidated. The core set of tRNA-A<sub>37</sub>-t<sup>6</sup>A biosynthesis pathway is conserved universally and consists of two protein families, namely TsaC/Sua5 and TsaD/Kae1/Qri7. The other participating proteins of the pathway are kingdom-specific. While TsaB and TsaE are unique to bacteria (exceptions include intracellular and symbiotic bacteria), Bud32, Pcc1 and Cgi121 are specific to eukaryotes<sup>2</sup>. Homologs of TsaD and TsaB are essential for growth and survival in *Escherichia coli* and *Salmonella typhimurium*. Depletion of these proteins have led to pleiotropic effects like elongation and branching of cells, unusual distribution of DNA, appearance of nucleoid etc.<sup>4</sup>. At least in 20% of bacterial genomes, *tsaB* is physically clustered with *tsaD*<sup>5</sup>. Interestingly, genomic context of these genes is variable across eukaryotes and prokaryotes indicating differences in their transcriptional organization and control mechanisms<sup>3</sup>. However, as yet, transcriptional organization and regulation of these genes have not been studied in any of the organisms where tRNA-A<sub>37</sub>-t<sup>6</sup>A transferase pathway is experimentally characterized. Knowing that TsaB and TsaE are essential as well as unique to bacteria, these proteins are regarded as promising antibacterial and inhibitor targets<sup>3,4,6,7</sup>. Therefore, an understanding of their genetic context, transcriptional organization and control, preferably in an organism-specific manner, is highly desirable<sup>8</sup>.

CSIR-Institute of Microbial Technology, Sector 39-A, Chandigarh-160036, India. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.R. (email: raoalka@imtech.res)

*M. tuberculosis* is the causative agent of Tuberculosis (TB), a major infectious disease of human. Since late 1980s, the disease has been undergoing a resurrection driven by a variety of changes in social, medical and economic factors, thereby leading to the rise of multiple drug resistant (MDR) and extensively drug resistant (XDR) strains of *M. tuberculosis*. This fastidious pathogen can survive hostile environments. Though not much studied, modified ribonucleosides in tRNA are now implicated in virulence and as growth rate determinants in pathogenic bacteria<sup>9</sup>. Therefore, it is plausible that the pathogenic mycobacteria may also employ RNA modifications to regulate translation and survive the human innate immunity. Recently, a chromatography-coupled mass spectrometric approach to identify and quantify ribonucleosides in *M. bovis* Bacille Calmette–Guérin (BCG), a species closely related to *M. tuberculosis*, confirmed the presence of t<sup>6</sup>A modification in mycobacteria<sup>10</sup>. Accordingly, the genes implicated in tRNA-A<sub>37</sub>-t<sup>6</sup>A biosynthesis are present in all mycobacterial species including *M. tuberculosis* (Supplementary Figure S1).

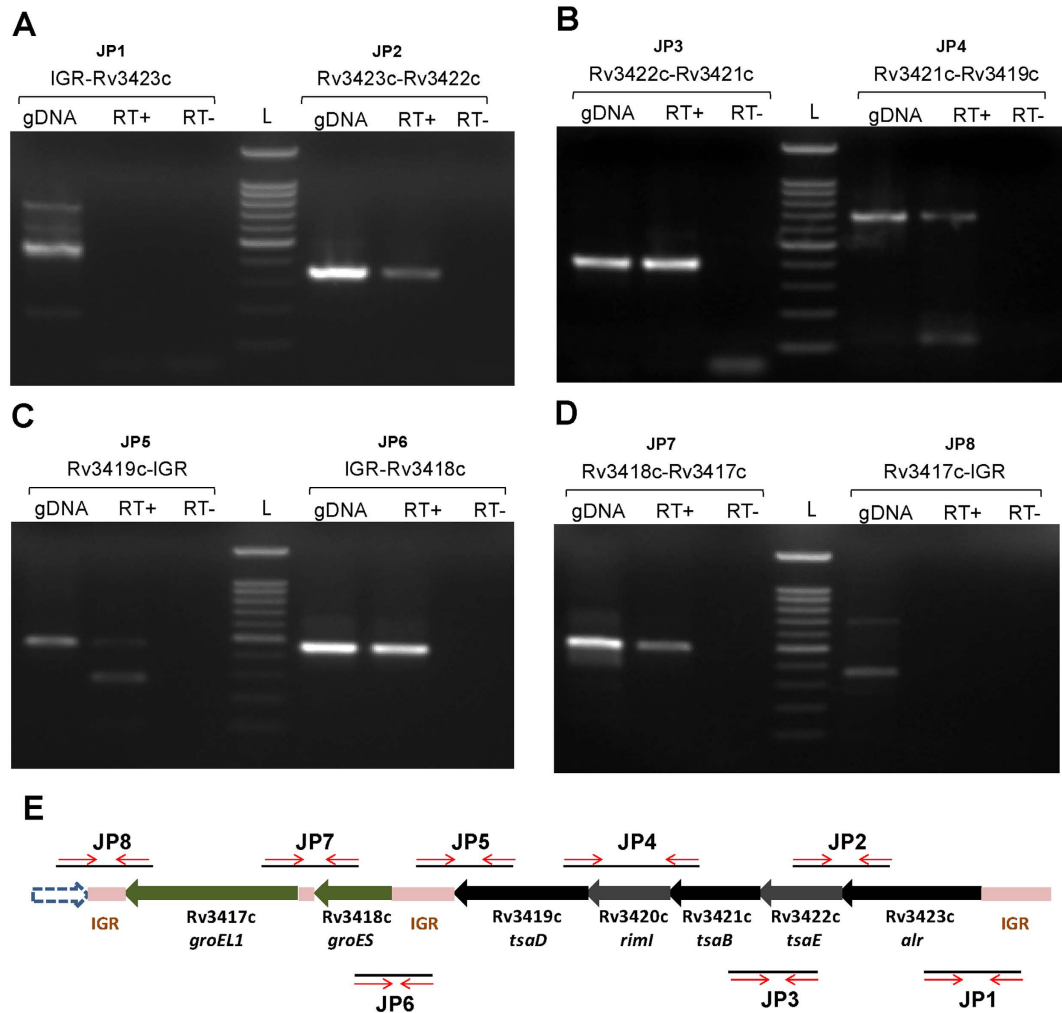
In *M. tuberculosis*, the homologs of tRNA-A<sub>37</sub>-t<sup>6</sup>A transferase machinery are present in a neighborhood where *tsaD* (Rv3419c), *tsaB* (Rv3421c) and *tsaE* (Rv3422c) are clustered together with *rimI* (Rv3420c) encoding an N- $\alpha$ -acetyltransferase and *alr* (Rv3423c) that codes for a well-characterized essential enzyme alanine racemase (Supplementary Figure S1). Presence of *rimI* in this genetic milieu is uniquely conserved in some genera like *Bacillus*, *Clostridium*, *Corynebacterium* and *Mycobacterium* (<http://string-db.org>)<sup>4,8</sup>. In a parallel study, our group has recently characterized RimI as a protein N- $\alpha$ -acetyltransferase having significantly relaxed substrate specificity, encompassing specificities of major eukaryotic NATs and mimicking NatE the closest<sup>11</sup>. Both functional and genetic context of RimI is yet little explored in bacteria, in general. Further, in *M. tuberculosis*, *tsaD* and *tsaB* are required for bacterial survival in primary murine macrophages, and essential for *in vitro* growth of bacteria on cholesterol<sup>12,13</sup>. Presence of transcripts of *tsaD* (Locus ML0379 in *M. leprae*) in the skin lesions of leprosy patients further emphasizes role of *tsaD* in pathogenicity of mycobacteria<sup>14</sup>. Similarly, *tsaE* is an essential gene according to Himar1-based transposon mutagenesis assays in H37Rv strain<sup>15</sup>. Alr catalyzes the synthesis of *d*-alanine that is needed for peptidoglycan biosynthesis<sup>4</sup>. Essentiality of *alr* for cellular viability is already known through knock-out/knock down and transposon mutagenesis studies<sup>13,15,16</sup>. The cluster of genes from *alr* to *tsaD* (termed *gcp-alr* in TBDB, <http://www.tbdb.org>) is located immediately upstream of a previously characterized bicistronic operon *groESL1* (Rv3418c and Rv3417c) encoding essential chaperones<sup>17</sup>. Despite medical importance of TB pathogen, understanding of operon architectures and regulation thereof is very poor in *M. tuberculosis*. This fact is further corroborated by our comprehensive compilation of experimentally validated and published information about operons in *M. tuberculosis* until October 2015 (Supplementary Table S5)<sup>18</sup>. Therefore, for the reasons that: (a) most genes in this cluster are essential for survival or virulence of *M. tuberculosis*, (b) they exist in an interesting context with a protein N- $\alpha$ -acetyltransferase, (c) the transcriptional organization and regulation of tRNA-A<sub>37</sub>-t<sup>6</sup>A transferase pathway is not known in bacteria yet, and d) the tRNA-A<sub>37</sub>-t<sup>6</sup>A transferase machinery is differently organized across different bacterial species, we decided to investigate this genetic context at the transcriptional level.

## Results

### Homologs of *tsaD*, *tsaB* and *tsaE* are co-transcribed with *alr*, *rimI* and *groESL1* sub-operon in *M. tuberculosis*.

A consecutive 4 bp overlap of stop and start codons of genes *alr* to *tsaD* conjectures a five-gene long operon organization that is predicted as *gcp-alr* operon in TBDB. To test this, we first verified the presence of the individual transcripts corresponding to *alr*, *tsaE*, *tsaB*, *rimI* and *tsaD* in the total RNA pool of *M. tuberculosis* H37Rv representing constitutive condition using gene specific primers. Extensive RT-PCR experiments using appropriate primers (Supplementary Table S1) corresponding to all possible junctions between genes of the *gcp-alr* cluster with suitable controls, helped us to ascertain the beginning of the ~6–6.5 kb transcript at *alr* and the end just after *groEL1* (Fig. 1; see also Supplementary Figure S2). These experiments conclusively confirmed the co-transcription of five genes, namely *alr*, *tsaE*, *tsaB*, *rimI*, *tsaD* along with *groEL1* and *groES*, in constitutive condition (Fig. 1). In order to determine the size of the longest cotranscript, northern blotting was performed on total RNA using a 300 bp dsDNA probe (radiolabeled JP2) encompassing the junction of *alr* and *tsaE* genes. A band was spotted on the blot at around ~6–6.5 kb, thereby providing the evidence of a transcription unit longer than those predicted by *in silico* methods (Fig. 2A). These results led to the conclusion that the transcript harbours all the seven genes with *alr* as the first while *groEL1* as the last gene. The *groES* and *groEL1* genes have already been shown to form an independent bicistronic *groESL1* operon with two TSSs annotated within the intergenic region between *tsaD* and *groES*<sup>17</sup>. Our results go beyond *in silico* predictions and elucidate *alr-gcp* transcript and previously known, bicistronic *groESL1* transcript as products of larger heptacistronic *alr-groEL1* operon.

**Mapping the transcription start sites in the *alr-groEL1* operon.** To identify transcription start sites (TSSs) in the *alr-groEL1* operon, 5' RLM-RACE was performed on total RNA isolated from *M. tuberculosis* H37Rv culture under physiological conditions. As multiple transcripts originating from the operon were observed in northern experiments, and primer extension experiments were not conclusive (Supplementary Figure S4), we employed various gene-specific and nested gene-specific primers along with 5' generacer primers to locate the possible TSSs in the operon. RACE reactions were compared with mock reactions lacking TAP treatment to facilitate the identification of unambiguous TSSs. Two transcription start sites were mapped in the operon- one at nucleotide G located at –263 position (3841684 genome position) relative to the *alr* initiation codon GTG while the other was at the nucleotide T located at –275 position (3838864 genome position) with respect to the *tsaD* initiation codon ATG (Supplementary Figure S3 and Table S2). Corresponding to these start sites, –10 hexamer sequences, namely TAGGTT and TACGCT are also identified upstream of *alr* and *tsaD* genes, respectively (Supplementary Table S3). The sequences of –10 hexamers here conform to the mycobacterial promoter

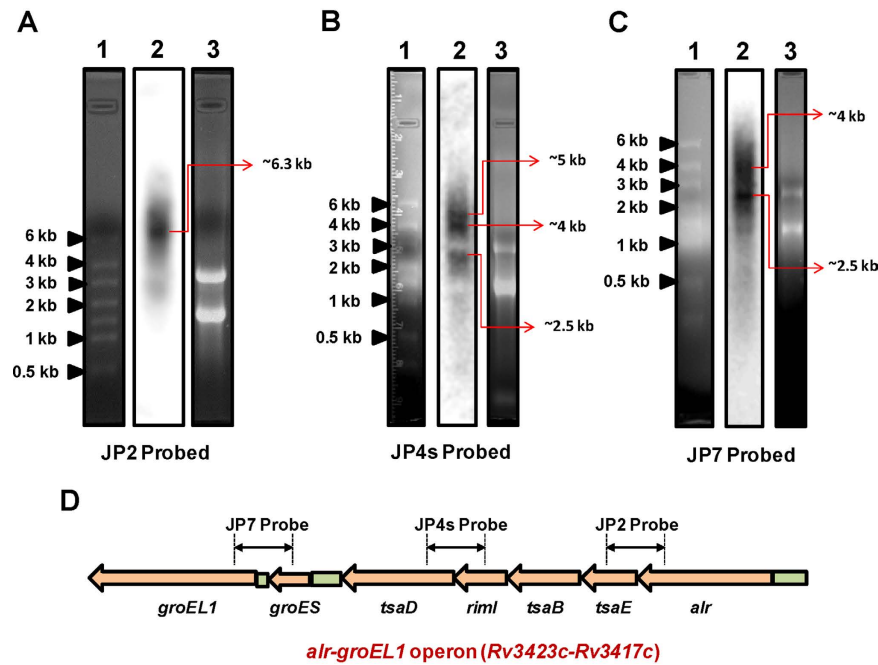


**Figure 1. Cotranscription of *alr-gcp* gene cluster as determined by RT-PCR.** (A–D) Total RNA from *M. tuberculosis* H37Rv in exponential phase was reverse transcribed and the cDNA thus made was used as template for PCR. Primers were designed in order to amplify the junctions between adjacent genes, shown above each gel. Each panel is constituted by three lanes – gDNA lane, which contains amplified product from genomic DNA and serves as positive control; RT+ lane, contains PCR product from cDNA; RT– lane, the negative control, represents the PCR product from possible gDNA present as contaminant in RNA preparation used for cDNA synthesis. (E) A schematic representation of the position of primers designed for cotranscription analysis of the gene cluster.

consensus (TANNNT at –10 box) as defined by Cortes *et al.*<sup>19</sup>. In Supplementary Table S2, we have tabulated all the TSSs located within *alr-groEL1* operon that are identified by us and or others, using different methods in independent studies<sup>19,20</sup>.

**Multiple transcripts suggest intricate operon architecture.** With *groESL1* operon already annotated<sup>17</sup>, it was anticipated that at least two differently sized transcripts should originate from this newly identified heptacistronic *alr-groEL1* operon. To authenticate the same, northern blotting was performed using two new radiolabeled probes, namely JP4s (300 bp) and JP7 (500 bp) as explained in Fig. 2B and C, in addition to previously employed probe JP2 (Fig. 2A). Multiple bands were obtained on the blot representing possible transcripts originating from the *alr-groEL1* operon. The probes JP2, JP4s and JP7 together, broadly cover the entire operon and provide an overall information about transcripts arising from the *alr-groEL1* operon. In this experiment, we could successfully detect the smaller bicistronic transcript (~2–2.5 kb; Fig. 2C) originating from the *groESL1* sub-operon. Presence of transcripts of different lengths strongly hinted at the presence of more than one regulatory element (promoters, terminators) in the heptacistronic operon.

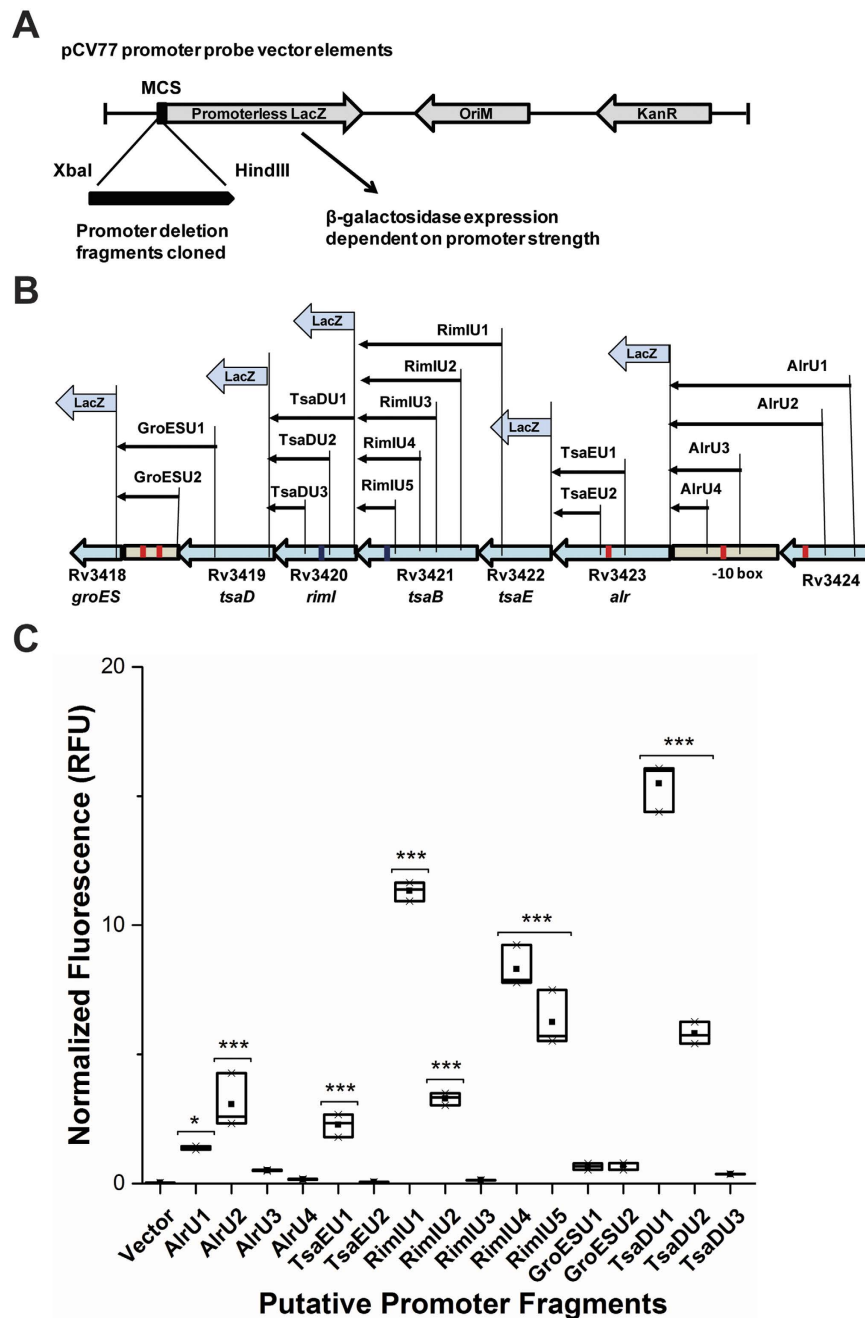
**Six different promoter elements of varying strength control the operon.** The promoter elements and their architecture in *M. tuberculosis* are variable and not always obvious upon sequence inspection using *in silico* tools (ex. BPRM) that are tailored for organisms like *E. coli*. Manual inspection of the genomic sequence of the operon in accordance with available literature<sup>18–21</sup> helped us to identify possible promoter elements in the *alr-groEL1* operon. After making a careful judgment from all predictions and northern blot results



**Figure 2. Multiple transcripts are produced from *alr-groEL1* operon.** Northern blotting of *M. tuberculosis* H37Rv total RNA was performed. About 5 to 8  $\mu$ g total RNA was resolved on formaldehyde agarose gel (1.2% in panel A and 1.5% in panels B and C) and transferred to nylon membrane. Blotted RNA was then separately probed with radiolabeled JP2 (panel A), JP4s (panel B) and JP7 (panel C) dsDNA probes. In each panel, lane 1 represents the RNA ladder, lane 2 is the blot and lane 3 shows the RNA sample. (D) Schematic representation of the *alr-groEL1* operon showing the regions used for probe generation.

put together, series of genomic fragments across the length of the operon were cloned in an *in vivo* promoter assay system (Fig. 3A and Table 1). The resulting clones were transformed into *M. smegmatis* cells and assayed for  $\beta$ -galactosidase activity using fluorescent substrate. Among the promoter fragments upstream of *alr*, the minimal promoter was identified as  $P_{alrU2}$  that contains mycobacterial consensus sequence TAGGTT at  $-10$  position. With fragment larger than  $P_{alrU2}$  i.e.  $P_{alrU1}$ , a decrease in activity was observed indicating a possibility of repressive motifs within the 466–584 bp region upstream of *alr* gene (Fig. 3B). It has earlier been studied that repression elements in the DNA region upstream of a promoter can lead to irregularities in the activity of the deletion fragments of that promoter<sup>22</sup>. Similarly, three more promoters were detected in the coding regions of *alr*, *tsaB* and *rimI* genes that are upstream of *tsaE*, *rimI* and *tsaD*, respectively. The minimal promoters identified are designated as  $P_{tsaEU1}$ ,  $P_{rimIU5}$  and  $P_{tsaDU2}$ , respectively. While a conserved  $-10$  hexamer sequence, namely TAYGAT (Y is pyrimidine) is found in mycobacterial promoters, little homology exists among  $-35$  sequences indicating the absence of a defined  $-35$  sequence and its dispensability for promoter function<sup>23</sup>. However, a conserved  $-10$  motif TANNNT, was recently found to be associated with most of the primary TSSs in mycobacteria<sup>19</sup>. Accordingly, among the new promoters identified in the *alr-groEL1* operon and summarized in Supplementary Table S3, three, namely  $P_{alrU2}$ ,  $P_{rimIU5}$  and  $P_{tsaDU2}$  harbour mycobacterial canonical promoter elements. This consensus is slightly relaxed compared to the TATAAT consensus found in *E. coli* and several other bacteria. About half of the TANNNT promoters in *M. tuberculosis* have been shown to be associated with a 3 bp extended  $-10$  motif responsible for the upsurge in the activity of the concerned promoter<sup>19</sup>. We detected this 3 bp consensus motif, [G/C][A/G]N (SRN), in the promoters located upstream of *tsaD* and *alr* genes.  $P_{tsaDU1}$  also exhibited the highest  $\beta$ -galactosidase activity (at least 23-fold higher than that of the heat shock promoter  $P_{groESU2}$ ). The activity displayed by the minimal promoter  $P_{tsaDU2}$  (with extended consensus motif) was comparable to that seen in case of minimal  $P_{rimIU5}$  promoter (without extended motif). In the COG category of RNA processing and modification genes, the proportion of leaderless genes (having very short or missing 5'-UTRs) is much higher than in other categories, in bacteria, in general<sup>24</sup>. Though the genes in *alr-groEL1* operon belong to RNA processing/modification category of genes, yet leaderless architecture is completely absent here.

**An I-shaped hairpin-like *cis*-regulatory element is integral to the operon.** Promoters alone did not explain northern blot results completely (Fig. 2). Since *groESL1* operon has already been described<sup>17</sup>, its upstream sequence was analyzed for the presence of internal terminators/ pause signals, if any, using GeSTer algorithm (Genome Scanner for Terminators)<sup>25</sup>. The tool predicted a non-canonical I-shaped, factor-independent terminator  $T_{tsaD}$ , downstream of *tsaD* gene. A series of fragments harbouring predicted and known reference terminators were cloned between the coding sequence of GFP and its promoter in the pVVGFPHis vector (Table 1). Reduction in GFP fluorescence was measured to decipher termination efficiencies (Fig. 4A and B). Significant reduction in GFP fluorescence as compared to *CTalr* control confirmed reliable *in vivo* transcriptional



**Figure 3. Multiple promoters control the transcription of *alr-groEL1* operon.** (A) pCV77 vector design and cloning of promoter deletion fragments. (B) A schematic representation of promoter deletion fragments design. (C) Relative strength/activity of putative promoter fragments from *alr-groEL1* operon as determined by normalized fluorescence intensity (in RFU) which is a measure of  $\beta$ -galactosidase expression. *M. smegmatis* mc<sup>2</sup>155 was transformed with promoter plasmids (fragments cloned in pCV77 promoter probe vector) and grown to approx. OD<sub>600</sub> 1.0. The cells were then sonicated for 3 min 15 s and  $\beta$ -galactosidase activities measured using FluoReporter *lacZ*/Galactosidase Quantitation Kit. Fluorescence intensity values were normalized against protein concentration and a given reference standard. Box chart has been plotted to show median (represented by horizontal line), mean (represented by black filled box) interquartile range (represented by box), and maximum and minimum values (represented by whiskers). The results shown were generated from three independent experiments, each with three technical replicates. The asterisks shown above each box plot represent the statistically significant difference from the vector control and are derived from *p* values with \* for *p* < 0.05, \*\* for *p* < 0.01 and \*\*\* for *p* < 0.001. Statistics was performed using one-way ANOVA with Dunnett's multiple comparison test.

Bacterial strain or plasmid	Description	Source or reference
<i>Escherichia coli</i> TOP10	F <sup>-</sup> <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) φ80 <i>lacZ</i> Δ <i>M15</i> Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara</i> <i>leu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> ( <i>StrR</i> ) <i>endA1</i> <i>nupG</i>	Lab Collection
<i>Escherichia coli</i> XL-1 Blue MRA	Δ( <i>mcrA</i> ) 183 Δ( <i>mcrCB-hsdSMR-mrr</i> ) 173 <i>endA1</i> <i>supE44</i> <i>thi-1</i> <i>recA1</i> <i>gyrA96</i> <i>relA1</i> <i>lac</i>	Stratagene
<i>Mycobacterium smegmatis</i> mc <sup>2</sup> 155	High-transformation mutant of <i>M. smegmatis</i> ATCC 607	ATCC 700084
<i>Mycobacterium tuberculosis</i> H37Rv	Laboratory strain	ATCC 27294
pCV77 plasmid	Replicating <i>E. coli</i> - <i>Mycobacteria</i> shuttle vector with promoterless LacZ gene, Kan <sup>R</sup> due to <i>aph</i> gene	MedImmune
pCVAlrU1	584 bp (3841420–3842004 genomic region) upstream <i>alr</i> cloned in pCV77 with AlrU1F and R primers	This study
pCVAlrU2	466 bp (3841420–3841886 genomic region) upstream <i>alr</i> cloned in pCV77 with AlrU2F and R primers	This study
pCVAlrU3	188 bp (3841420–3841608 genomic region) upstream <i>alr</i> cloned in pCV77 with AlrU3F and R primers	This study
pCVAlrU4	79 bp (3841420–3841499 genomic region) upstream <i>alr</i> cloned in pCV77 with AlrU4F and R primers	This study
pCVTsaEU1	424 bp (3840197–3840621 genomic region) upstream Rv3422c cloned in pCV77 with TsaEU1F and R primers	This study
pCVTsaEU2	338 bp (3840197–3840535 genomic region) upstream Rv3422c cloned in pCV77 with TsaEU2F and R primers	This study
pCVRimIU1	725 bp (3839062–3839787 genomic region) upstream Rv3420c cloned in pCV77 with RimIU1F and R primers	This study
pCVRimIU2	483 bp (3839062–3839545 genomic region) upstream Rv3420c cloned in pCV77 with RimIU2F and R primers	This study
pCVRimIU3	335 bp (3839062–3839397 genomic region) upstream Rv3420c cloned in pCV77 with RimIU3F and R primers	This study
pCVRimIU4	262 bp (3839062–3839324 genomic region) upstream Rv3420c cloned in pCV77 with RimIU4F and R primers	This study
pCVRimIU5	120 bp (3839062–3839182 genomic region) upstream Rv3420c cloned in pCV77 with RimIU5F and R primers	This study
pCVGroESU1	926 bp (3837288–3838214 genomic region) upstream Rv3418c cloned in pCV77 with GroESU1F and R primers	This study
pCVGroESU2	277 bp (3837288–3837565 genomic region) upstream Rv3418c cloned in pCV77 with GroESU2F and R primers	This study
pCVTsaDU1	500 bp (3838589–3839089 genomic region) Rv3420c gene cloned in pCV77 with TsaDU1F and R primers	This study
pCVTsaDU2	310 bp (3838589–3838899 genomic region) upstream Rv3419c cloned in pCV77 with TsaDU2F and R primers	This study
pCVTsaDU3	140 bp (3838589–3838729 genomic region) upstream Rv3419c cloned in pCV77 with TsaDU3F and R primers	This study
pVVGFPHis plasmid	Mycobacterial constitutive expression shuttle vector containing <i>gfp</i> and Kan <sup>R</sup> genes. Developed from pVV16 plasmid.	BEI Resources, NIAID, NIH
pVVTsaDF	123 bp (3837555–3837432 genomic region) downstream of Rv3419c cloned in pVVGFPHis vector	This study
pVVTsaDR	275 bp (3837555–3837280 genomic region) downstream of Rv3419c cloned in pVVGFPHis vector in reverse orientation.	This study
pVVCtaI <sub>r</sub>	156 bp (3840806–3840962 genomic region) coding region of Rv3423c cloned in pVVGFPHis vector.	This study
pVVTsynA	40 bp synthetic hairpin sequence with a U-tract cloned in pVVGFPHis vector	This study
pVVTtrpA	27 bp synthetic hairpin sequence with a U-tract cloned in pVVGFPHis vector (based on <i>E. coli</i> tryptophan operon terminator)	This study

**Table 1. Bacterial strains and plasmids used in this study.**

regulation by  $T_{tsaD}$  (Fig. 4C). A typical *E. coli* intrinsic terminator has a U-trail post palindromic stem loop/hairpin. Non-canonical I-shaped intrinsic hairpins lack a U-trail (Supplementary Table S3). Such hairpins are fairly common among GC rich mycobacterial genomes. As the sequence of  $T_{tsaD}$  lacks any U-tract and reversing the orientation of the hairpin sequence ( $T_{tsaD}R$ ) didn't lead to any significant reduction in the fluorescence of the gene downstream of the fragment (Fig. 4C), the element is akin to an I-shaped terminator in accordance with the description provided by Unniraman *et al.*<sup>26</sup>

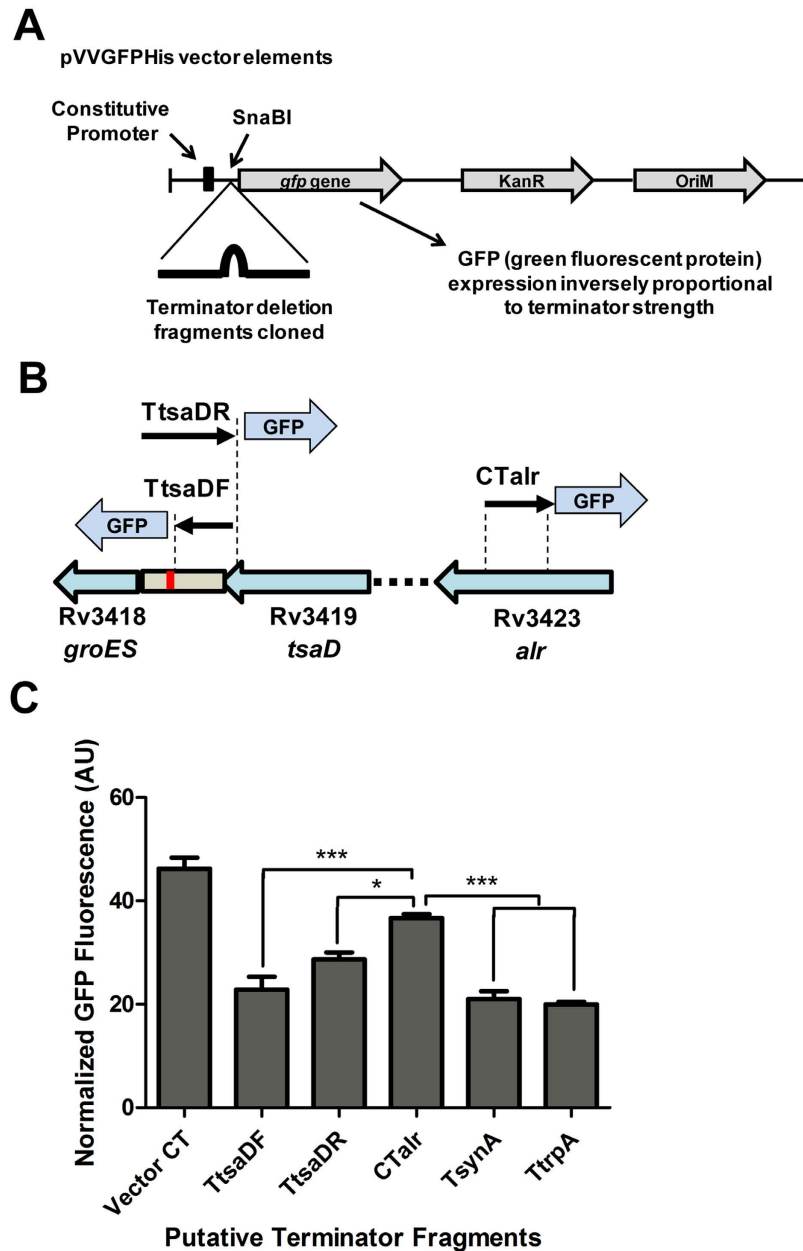
Low or partial *in vivo* termination efficiency in such I-shaped hairpins is also known<sup>25</sup>. Recently, it was demonstrated that *M. bovis* RNA polymerase requires termination-stimulating activity of mycobacterial NusG and a U-tract, which may be imperfect for termination at intrinsic terminators *in vitro*. It was also demonstrated that intrinsic terminators lacking imperfect or perfect U-tract are not terminated by *M. bovis* RNA polymerase<sup>27</sup>. However, in our experiments, the *in vivo* termination efficiencies of  $T_{tsaD}$ ,  $T_{synA}$  and  $T_{trpA}$  (previously *in vitro* validated terminators are explained in Table 1 and ref. 27) were found to be comparable (Fig. 4C). Therefore, in view of the above discussed contradictory evidences about the functionality of I-shaped terminators *in vivo*<sup>26–28</sup>,  $T_{tsaD}$  is defined as an I-shaped hairpin-like *cis*-regulatory element that might bring about reduction in transcription

by acting as a pause signal and facilitating fraying of the elongation complex or by enabling factor-dependent termination after slowing down the complex *in vivo*<sup>29</sup> or might influence transcriptional/post-transcriptional regulation of downstream genes by some other mechanism. Recently, a unique branching structure (with a trailing U-tract) was shown to function as an intrinsic terminator in *Enterococcus faecalis*<sup>30</sup>, sustaining the suggestion that greater flexibility might exist in the design of intrinsic terminator-type regulatory elements than has been contemplated so far.

**Multiple regulatory elements orchestrate a complex transcriptional profile of the operon under stress.** Multiple transcripts originating from a single operon, presence of multiple promoters or terminators and a sub-operonic architecture (with more than one independent control elements) provided enough clues about a complex yet fine mode of regulation of *alr-groEL1* operon<sup>31</sup> (Fig. 5A). Transcripts of various sizes as detected on blots were aligned with the operon in all possible combinations and corroborated with other experiments to arrive at a logical map as represented in Fig. 5B. Keeping the pathogen biology and human physiology in consideration, a quantitative transcriptional response profile of individual genes in the operon was generated under heat shock, low pH stress (pH 4.5), SDS stress (0.1% SDS), oxidative stress (5 mM H<sub>2</sub>O<sub>2</sub>, 5 mM cumene hydroperoxide) and reductive stress (1 mM DTT) to obtain a broader picture of involvement of these genes in stress adaptation in *M. tuberculosis*<sup>32,33</sup>. Results of quantitative RT-PCR of RNA isolated from bacterial cultures subjected to different stresses are summarized in Supplementary Table S4 and presented as a heat map in Fig. 5C (individual stress responses are plotted in Supplementary Figure S5). The genes of the operon were upregulated under majority of the conditions studied except for reductive (DTT) stress. The *tsaD* and *alr* genes were most notably induced by the stresses. Under detergent and oxidative stresses, significant upregulation (~3.5-fold) was observed for *tsaD* gene while under pH stress, all genes except *groEL1* were induced. Heat shock experiments carried out at 42 °C in our study led to ~7-fold upregulation (measured through RT-qPCR) of *groESL1* operon while all other genes barring *tsaD* gene of the cluster, were slightly upregulated. Similarly, Aravindhan *et al.*<sup>17</sup> had also reported 4-fold upregulation (measured through primer extension) of *groESL1* sub-operon upon heat shock<sup>17</sup>. Interestingly, *groES* was found about 9-fold upregulated in  $\Delta hspR\Delta hrcA$  strain of mycobacteria (measured through microarray) under constitutive condition previously, implying a role for HspR and HrcA in repression of *groESL1* sub-operon under constitutive condition<sup>34</sup>. This further reflects the significance of sub-operonic architecture where regulation of sub-operon is distinct from the main operon/upstream transcripts in stress especially during heat shock, pH and SDS stresses in *M. tuberculosis*. The heat map also highlights the discrete expression profile of *tsaD* gene under three of the stress conditions, namely pH, detergent and oxidative stresses. This expression pattern of *tsaD* can be explained by the presence of a strong upstream promoter, P<sub>*tsaDU1*</sub> and an I-shaped hairpin-like *cis*-regulatory element, T<sub>*tsaD*</sub> downstream of the gene. Heat map generated from the quantitative RT-PCR of operon genes as well as that of intergenic regions of the operon (see Supplementary Figure S6) corroborate subtle fine-tuning that these multiple regulatory elements could bring about to the levels of the individual transcripts or cotranscripts in the cellular milieu.

## Discussion

The tRNA-A<sub>37</sub>-t<sup>6</sup>A transferase machinery consists of universally conserved protein TsaD, and proteins specific to bacteria, namely TsaB and TsaE<sup>2,3</sup>. Physical clustering of the encoding genes is observed in the genomes of several bacteria in different contexts, yet no experimental data is available on transcriptional organization and regulation of such an important and conserved pathway except for *Neisseria gonorrhoeae* wherein *tsaD* is shown to be co-regulated with cytochrome genes<sup>35</sup>. In this study, we have characterized a new multicistronic *alr-groEL1* operon that is involved in the transcriptional regulation of *tsaD*, *tsaB* and *tsaE* homologs in *M. tuberculosis*, a persistent human pathogen. The operon is, in fact, composed of two sub-operons: *alr-tsaD* and *groESL1*, both having independent control elements like promoters and terminators. Position of a newly identified internal I-shaped hairpin-like *cis*-regulatory element, T<sub>*tsaD*</sub> has also been defined. Occurrence and importance of such sub-operon architecture in transcriptional regulation has previously been analyzed in *Bacillus subtilis* and *E. coli*<sup>31</sup>. The *alr-groEL1* operon is one of the few experimentally characterized operons where sub-operon architecture exists within the main operon, in *M. tuberculosis* (as per the information compiled in Supplementary Table S5). The sub-operonic architecture seems physiologically relevant as seen in case of heat shock, SDS and pH stress conditions where suboperon is distinctly regulated vis-à-vis main operon (Fig. 5). Complex transcriptional regulation is a characteristic feature of *M. tuberculosis* as observed in case of DNA gyrase bicistronic operon that is regulated by alternative internal promoters and competing promoters on the opposite strand<sup>36</sup>, and *recXA* operon which is controlled by multiple promoters for single gene<sup>37</sup>, and others listed in Supplementary Table S5. The case of *alr-groEL1* operon is interesting in the sense that it contains a large number of *cis*-regulatory elements, namely promoters and terminator-like *cis*-regulatory element. Whereas studies on identifying promoters and transcriptional factors are aplenty, terminators and similar *cis*-regulatory transcriptional elements are not much studied in *M. tuberculosis* (Supplementary Table S5). As seen in the case of *alr-groEL1* operon, careful analysis of such *cis*-regulatory elements is essential for unambiguous interpretation and understanding of the operon architecture in *M. tuberculosis*. Using our experimental evidence and tapping the relevant and corroborating information made available recently by others<sup>19</sup> and as cited above, we have created a comprehensive map of *alr-groEL1* operon that summarizes the sub-operonic architecture generating multiple transcripts, complex co-transcription and co-regulation of essential putative t<sup>6</sup>A transferase machinery, chaperones, cell wall biosynthesis gene *alr* and an N- $\alpha$ -acetyltransferase in *M. tuberculosis* (Fig. 5 and Supplementary Figure S1; see also Supplementary Table S3 for detailed description of individual elements). The extensive information about 40 experimentally characterized operon structures in *M. tuberculosis* (see Supplementary Table S5) along with the findings of this manuscript can serve as ready information for bioinformatics analysis/ relevant tool development.



**Figure 4.** The *alr-groEL1* operon harbors an I-shaped hairpin-like *cis*-regulatory element,  $T_{tsaD}$ . (A) Basic elements of pVVGFPHis vector used for cloning. (B) Graphical representation of tested fragments for termination efficiency. (C) Measurement of *in vivo* GFP fluorescence plotted as comparative reduction indicating termination strength of tested terminator-like fragments. *M. smegmatis* mc<sup>2</sup>155 cells were transformed with various putative terminator fragments cloned downstream of constitutive promoter of *gfp* gene in pVVGFPHis vector. The vector with no intervening foreign sequence between the promoter and GFP coding region, served as the negative control. Two synthetic terminators,  $T_{synA}$  and  $T_{trpA}$ , provided for positive controls for transcription termination. Log phase cultures were assayed for GFP fluorescence in triplicate in black 96-well microplate. The fluorescence was measured at 395 nm excitation and 509 nm emission wavelengths. The results shown are means  $\pm$  SEM of three independent experiments, each accomplished with three technical replicates. One-way ANOVA with Dunnett's multiple comparison test was employed to determine *p* values.

In parallel, we could also identify  $P_{tsaDU1}$  as  $\sim$ 23-fold and  $P_{rimIU1}$  as  $\sim$ 17-fold stronger promoters than the heat shock promoter  $P_{groESU2}$  under constitutive condition in mycobacterial expression system. Characterization of the *tsaDU1* and *rimIU1* promoter fragments in this study provides for valuable tools for the expression of genes in mycobacteria.

The stress-induced mosaic transcriptional profile of individual genes depicted in heat map corroborates well with the complex architecture of the operon harboring multiple regulatory elements. The heat map also



indicates fluid dynamics of bacterial operon involved in multiple stress adaptations, wherein multiple regulatory elements can give rise to variable transcripts under different environmental conditions (Fig. 5C and Supplementary Figure S6). While it could also be a mechanism of energy and resource conservation under non-optimal survival conditions, such intricate regulatory designs may be important for accommodating the dynamic changes that happen within or around the cell. Such differences may play an important role in the evolution of a gene order across genomes<sup>31</sup>. Since the t<sup>6</sup>A modification is known in *M. bovis* and levels of t<sup>6</sup>A modification are implicated in regulating gene expression by controlling selective translation in codon-biased manner as the organism goes through hypoxia-induced latency and resuscitation later on<sup>38–41</sup>, we revisited the heat map summarizing the results of quantitative RT-PCR of *alr-groEL1* operon. Indeed, during oxidative stress, almost all genes of the operon are upregulated, including *tsaD* (an essential component in the threonylcarbamyl tRNA synthesis). A general upregulation of *tsaD*, *tsaB* and *tsaE* genes, therefore, might be a prerequisite for t<sup>6</sup>A modification during oxidative stress. Accordingly, this study opens up further interest towards understanding the transcriptional and stress regulation of t<sup>6</sup>A transferase enzymes in bacteria and in pathogenic *M. tuberculosis*, specifically.

Using mass spectrometry-based enzyme assays, we had previously identified an enzyme-substrate relationship between RimI (N- $\alpha$ -acetyltransferase) and GroES, GroEL1 and TsaD proteins as substrates wherein N-terminus of these proteins is acetylated by RimI *in vitro*<sup>11</sup>. With the help of an *in vivo* mycobacterial protein fragment complementation (MPFC) assay conducted in *M. smegmatis*, protein-protein interactions between RimI and GroES, RimI and TsaD, and TsaD and TsaE proteins are also known to us<sup>11</sup>. Although functions of the 10 kDa mycobacterial GroES/Cpn10 protein are not known in detail, GroES is reported to be N- $\alpha$ -acetylated in *M. tuberculosis* proteome<sup>42</sup>. Therefore, co-transcription of *groES1* and *rimI* under *alr-groEL1* operon supports the previously identified involvement of RimI in the functioning of the chaperone protein. Further, a weak *in vitro* interaction between a recombinant T18-RimI fusion protein and T25-TsaB fusion protein is reported in *E. coli*, previously<sup>4</sup>. Taking into account the conservation of *tsaB* and *rimI* synteny across many bacterial species (<http://string-db.org>) including pathogenic *Neisserial* and *Mycobacterial* species where a *tsaB-rimI* gene fusion is witnessed (ex. *M. leprae*), results from this study and our previously published results<sup>11</sup>, the involvement of RimI in TsaD, TsaB and TsaE functioning is anticipated. The *alr-groEL1* operon thus provides a well-defined transcriptional context to study hitherto unexplored regulation of RimI/ protein N- $\alpha$ -acetylation as well as involvement of RimI/protein acetylation in tRN A-A<sub>37</sub>-t<sup>6</sup>A modification pathway, in *M. tuberculosis*.

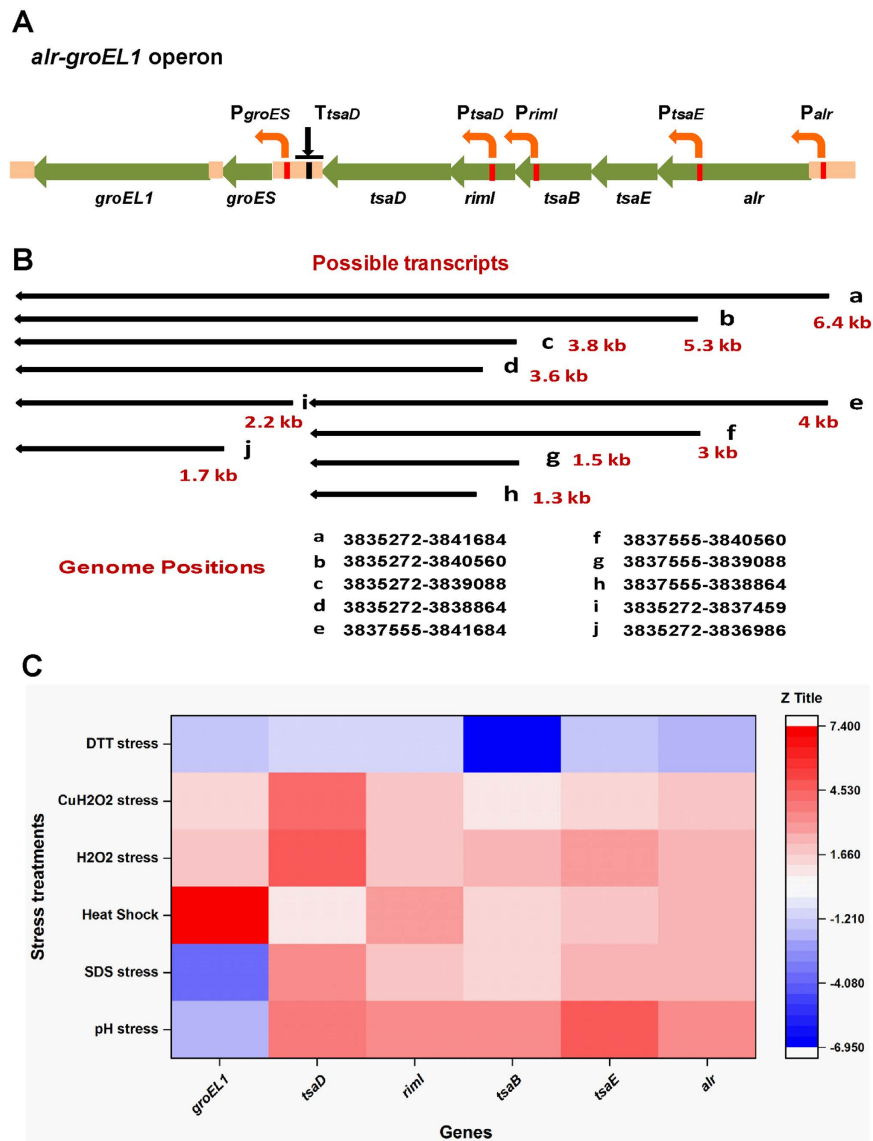
## Materials and Methods

**Bacterial strains and culturing.** The bacterial strains and plasmids used in this study are listed in Table 1. Middlebrook 7H9 broth supplemented with 10% ADC (albumin and dextrose complex), 0.2% glycerol and 0.1% Tween 80, was used for culturing *M. tuberculosis* H37Rv and *M. smegmatis* mc<sup>2</sup>155. Middlebrook 7H11 agar supplemented with 10% OADC (ADC with oleic acid) and 0.5% glycerol, was used for transformation and growth of *M. smegmatis* mc<sup>2</sup>155. The *E. coli* cells were grown in Luria-Bertani (LB) medium. *M. smegmatis* mc<sup>2</sup>155 was grown on X-Gal- and kanamycin-containing 7H11 agar plates for  $\beta$ -galactosidase reporter assays. About 800  $\mu$ g X-Gal was spread onto each 7H11 plate. *E. coli* XL1-Blue and TOP10 strains were used for transformation. Kanamycin was used at final concentrations of 50  $\mu$ g/ml for *E. coli* and 30  $\mu$ g/ml for *M. smegmatis*. All experiments were performed under constitutive condition unless mentioned otherwise.

**RNA Extraction.** *M. tuberculosis* H37Rv cultures grown at 37 °C until log phase were first treated with 35 ml GTC Buffer (5 M Guanidine thiocyanate, 0.5% Tween-80, 1%  $\beta$ -Mercaptoethanol) per 15 ml of culture followed by pelleting the cells. The cells were processed immediately for extraction of total RNA using FastRNA<sup>R</sup> Blue Kit (MP Biochemicals) as per the manufacturer's instructions, where chloroform phase separation and isopropanol precipitation were repeated twice to ensure good quality. Residual DNA was removed by treating approx. 40  $\mu$ g RNA with 4  $\mu$ l DNase I, RNase-free (Ambion) in the given reaction buffer in a total volume of 80  $\mu$ l at 37 °C for 30 min. DNase I treated samples were further purified using Qiagen Rneasy MinElute Cleanup kit and DNase I treatment was repeated again to ensure proper removal of genomic DNA. Finally, DNase I was inactivated at 75 °C for 5 min in presence of 5 mM EDTA, and RNA was used for RT-PCR and RT-qPCR.

**Reverse transcriptase PCR (RT-PCR).** First-strand cDNA synthesis was carried out using RevertAid M-MuLV Reverse Transcriptase kit (Fermentas) as per the instructions of the manufacturer. Nearly 1  $\mu$ g purified RNA acting as a template, was incubated with random hexamers and denatured at 65 °C for 5 min followed by snap chilling. Other reaction components were subsequently added to the tube in 20  $\mu$ l final volume. The tube was incubated in a thermal cycler at 25 °C for 5 min, 45 °C for 1 h and 70 °C for 5 min, consecutively. Along with the cDNA reaction, a negative control reaction, termed RT-, was also included from which RT enzyme was omitted to rule out the presence of any contaminating genomic DNA. PCR reactions were performed with 1  $\mu$ l template from each reaction (cDNA rxn, RT- rxn). In parallel with these reactions, another positive control PCR reaction was set up from gDNA template. The PCR products were loaded on an agarose gel followed by visualization on a UV transilluminator.

**Northern blotting.** Northern analysis was performed with *M. tuberculosis* H37Rv total RNA as per the standard protocol<sup>43</sup>. Briefly, 5 to 8  $\mu$ g of total RNA isolated from log phase cultures was resolved on 1.2% or 1.5% agarose gel containing 6% formaldehyde. RNA was transferred to Amersham Hybond-N + nylon membrane (GE Healthcare) overnight using upward capillary transfer. It was then cross-linked to membrane by exposure to UV and incubated with denatured radiolabeled probe after pre-hybridization. [ $\alpha$ -P<sup>32</sup>] dCTP-labeled DNA probes were generated near the 5' and 3' ends of the operon. Hybridization was done overnight at 42 °C, followed by



**Figure 5. Relative expression of the constituent genes of *alr-groEL1* operon under various stress conditions.** (A) A schematic representation of the positions of promoters (indicated by red bars) and possible pause signal/transcription regulatory elements (represented by black bars). (B) The *alr-groEL1* multi-transcript map representing a probable outcome of previous experiments. Genome position for each transcript is indicated to estimate its approximate size. (C) A heatmap with colour variations referring to the changes in expression levels of the individual genes. Colour scale is represented by z title. Colour approaching red denotes upregulation while downregulation is represented by increasing blue colour. RT-qPCR was done on total RNA from *M. tuberculosis* H37Rv subjected to various stresses that include heat shock at 42 °C, SDS (0.1%), pH 4.5, H<sub>2</sub>O<sub>2</sub> (5 mM), cumene hydroperoxide (5 mM), DTT (1 mM). In all cases, the bacteria were kept under stress for approx. 2 h except for heat shock which was given for 30 min. The results (fold change observed) were normalized to the expression of the 16S rRNA reference gene. The results shown are the representative of three independent experiments.

several membrane washes. Finally, the blot was exposed to phosphorimager screen which was then scanned using phosphorimager ImageQuant to visualize the hybridization signals.

**5' RLM-RACE.** RNA ligase-mediated rapid amplification of 5' cDNA ends (5' RLM-RACE) was carried out using GeneRacer kit (Invitrogen) and total RNA from *M. tuberculosis*. Manufacturer's instructions were followed with some modifications. Briefly, 5 µg total RNA was column purified after DNase-I (Ambion) digestion and then treated with calf intestinal phosphatase (CIP) and tobacco acid pyrophosphatase (TAP) sequentially. CIP was used to dephosphorylate truncated mRNAs and non-mRNAs while TAP was employed to generate a 5' monophosphate on intact mRNAs by removing the terminal pyrophosphate group. Next, the RNA was ligated to 5' RACE adaptor (RNA oligo) and reverse-transcribed to cDNA using random hexamers and Superscript III

Reverse Transcriptase at 50 °C for 1 h. To distinguish between processed and primary transcripts, a mock reaction (TAP-) was set up in parallel that was kept from TAP treatment. The cDNA was amplified with Phusion High Fidelity DNA polymerase (NEB) using GeneRacer 5' primer (complementary to a part of RNA adapter sequence) in combination with gene-specific reverse (GSR) primers, namely Rv23R (5'-CCGACATTCTCCCAGAACCG) and TsaDUR (5'-GTTCAAGCTTGACTGTCGTCATGACGGGTCC). Touch-down PCR protocol was followed for initial amplification from cDNA as: (30 sec at 98 °C) × 1 cycle, (10 sec at 98 °C, 1 min at 72 °C) × 5 cycles, (10 sec at 98 °C, 1 min at 70 °C) × 5 cycles, (10 sec at 98 °C, 30 sec at 68 °C, 1 min at 72 °C) and finally 5 min at 72 °C. Later, nested GSR primers (AlrUR, 5'-GTTCAAGCTTCTGAGCAGTATTCCGGCCTG; Qt20R, 5'-TCATTATCGGTGCGGACCTCCA) were used to generate specific products from initial PCR reactions (Supplementary Table S1). The PCR products were gel-purified and cloned in pJET1.2/blunt vector using CloneJET PCR cloning kit (Fermentas), and sequenced with pJET1.2 forward and reverse sequencing primers.

**In vivo promoter activity assay.** Putative promoter sequences were PCR amplified from *M. tuberculosis* H37Rv genomic DNA using Phusion High Fidelity DNA Polymerase (Finnzymes)/ KOD Hot Start DNA polymerase (Novagen) using appropriate primers (Supplementary Table S1). The PCR products were digested with XbaI and HindIII enzymes and cloned ahead of *lacZ* gene in promoter-less shuttle vector pCV77 in *E. coli* creating series of promoter fragment clones for *in vivo* analysis as shown in Table 1. *M. smegmatis* transformants containing individual clones were selected on X-gal-containing 7H11 agar plates. Independent transformants were grown in ~5 ml 7H9 broth until an OD<sub>600</sub> of 1. Cells were pelleted down, resuspended in 300 µl lysis buffer (1X PBS, protease inhibitor cocktail) and lysed by sonication for 3 min 15 s with 15 s pulse and 15 s rest at 30% amplitude using Sonics Vibra-Cell™ VCX500 with 3 mm stepped microtip. The supernatant from the lysate of each transformant was analysed for β-galactosidase activity using FluoReporter<sup>®</sup> *lacZ*/Galactosidase Quantitation Kit (Molecular Probes) as per given protocol. The fluorescence intensity of the samples was measured in triplicate in Nunc™ black 96-well plates using TECAN Infinite<sup>®</sup> M200 PRO microplate reader. The excitation and emission wavelengths used were 390 nm and 460 nm, respectively.

**In vivo transcriptional terminator activity assay.** Selected putative terminator sequences as listed in Table 1 were PCR amplified as above (using primers listed in Supplementary Table S1), cloned in SnaBI site after GFP promoter in pVVGFPHis vector and transformed in *M. smegmatis* mc<sup>2</sup>155. Cloned terminator fragments include: T<sub>tsaD</sub>F: 123 bp downstream region of *tsaD*- forward sequence; T<sub>tsaD</sub>R: 275 bp downstream region of *tsaD*-reverse sequence; CTalr: 156 bp coding region of *alr*; T<sub>synA</sub>: 40 bp synthetic hairpin sequence with a U-tract; and T<sub>trpA</sub>: 27 bp synthetic hairpin sequence with a U-tract (based on *E. coli* tryptophan operon terminator). Two synthetic terminators, T<sub>synA</sub> and T<sub>trpA</sub>, were included to serve as positive controls for transcription termination. Three colonies of each terminator fragment were streaked onto 7H11 agar plate with kanamycin antibiotic (30 µg/ml) and then inoculated in 2 ml 7H9 broth. The cultures were grown for 3 days in a shaking incubator set at 37 °C and 200 rpm, and then sub-cultured in fresh 7H9 media with 5% inoculum. At an OD<sub>600</sub> of 0.8, culture volumes of 120 µl were loaded into the black 96-well microplate in triplicate. GFP fluorescence intensity was measured at 395 nm excitation and 509 nm emission wavelengths in Nunc™ black 96-well plates using TECAN Infinite<sup>®</sup> M200 PRO microplate reader. The fluorescence intensity values were normalized against OD<sub>600</sub> values.

**RT-qPCR (quantitative PCR).** *M. tuberculosis* H37Rv cultures were grown until log phase. Cells were washed with 1X PBS buffer and replenished with fresh 7H9 media containing 5 mM H<sub>2</sub>O<sub>2</sub> or 5 mM cumene hydroperoxide to generate oxidative stress condition, 0.1% SDS to generate detergent stress condition, 1 mM DTT to generate reducing stress condition. Similarly, for acid stress treatment, log phase grown cells were replenished with fresh 7H9 media having pH 4.5. The freshly replenished cultures as well as appropriate controls were subjected to the respective stress condition for 2 h at 37 °C at 100 rpm. For heat shock, log phase cultures were first replenished with fresh 7H9 media and then transferred to 42 °C for 30 min.

Total RNA was extracted from all stress-treated *M. tuberculosis* cultures as well as control cultures as described above. The cDNA made with random hexamers was used as a template and qPCR was performed with gene-specific primers using Maxima SYBR Green qPCR Master Mix (Fermentas). Threshold cycle (*C<sub>T</sub>*) of the internal 16S rRNA control was utilized for the normalization of the calculated *C<sub>T</sub>* value. The qPCR results were analyzed using REST 2009 software version 1 (see legend of Supplementary Figure S5).

## References

1. Yarian, C. *et al.* Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **277**, 16391–5 (2002).
2. Thiaville, P. C. *et al.* Cross kingdom functional conservation of the core universally conserved threonylcarbamoyladenine tRNA synthetase enzymes. *Eukaryot. Cell* **13**, 1222–31 (2014).
3. Thiaville, P. C., Iwata-Reuyl, D. & de Crécy-Lagard, V. Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t(6)A), a universal modification of tRNA. *RNA Biol.* **11**, 1529–39 (2014).
4. Handford, J. I. *et al.* Conserved network of proteins essential for bacterial viability. *J. Bacteriol.* **191**, 4732–4749 (2009).
5. El Yacoubi, B. *et al.* A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. *EMBO J.* **30**, 882–893 (2011).
6. Allali-Hassani, A., Campbell, T. L., Ho, A., Schertzer, J. W. & Brown, E. D. Probing the active site of YjeE: a vital Escherichia coli protein of unknown function. *Biochem. J.* **384**, 577–84 (2004).
7. Kopina, B. J. Investigation of essential tRNA modifications lysidine (k 2C34) and N6-threonylcarbamoyl adenosine (t 6A37) as potential antimicrobial targets, at <http://gradworks.umi.com/36/35/3635857.html> (2014).
8. Oberto, J. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**, 4 (2013).
9. Durand, J. M. B. & Björk, G. R. Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of *Shigella flexneri*: a possible role in adaptation of virulence. *Mol. Microbiol.* **47**, 519–27 (2003).
10. Chan, C. T. Y. *et al.* Identification of N6,N6-dimethyladenosine in transfer RNA from Mycobacterium bovis Bacille Calmette-Guérin. *Molecules* **16**, 5168–81 (2011).

11. Pathak, D., Bhat, A. H., Sapehia, V., Rai, J. & Rao, A. Biochemical evidence for relaxed substrate specificity of N $\alpha$ -acetyltransferase (Rv3420c/rimI) of *Mycobacterium tuberculosis*. *Sci. Rep.* **6**, 28892 (2016).
12. Rengarajan, J., Bloom, B. R. & Rubin, E. J. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA* **102**, 8327–8332 (2005).
13. Griffin, J. E. *et al.* High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7**, e1002251 (2011).
14. Ribeiro-Guimarães, M. L. *et al.* Expression analysis of proteases of *Mycobacterium leprae* in human skin lesions. *Microb. Pathog.* **43**, 249–54 (2007).
15. Sasseti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**, 77–84 (2003).
16. Awasthy, D., Bharath, S., Subbulakshmi, V. & Sharma, U. Alanine racemase mutants of *Mycobacterium tuberculosis* require D-alanine for growth and are defective for survival in macrophages and mice. *Microbiology* **158**, 319–327 (2012).
17. Aravindhan, V. *et al.* *Mycobacterium tuberculosis* groE promoter controls the expression of the bicistronic groESL1 operon and shows differential regulation under stress conditions. *FEMS Microbiol. Lett.* **292**, 42–49 (2009).
18. Roback, P. *et al.* A predicted operon map for *Mycobacterium tuberculosis*. *Nucleic Acids Res.* **35**, 5085–5095 (2007).
19. Cortes, T. *et al.* Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* **5**, 1121–1131 (2013).
20. Shell, S. S. *et al.* Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genet.* **11** (2015).
21. Agarwal, N. & Tyagi, A. K. Mycobacterial transcriptional signals: Requirements for recognition by RNA polymerase and optimal transcriptional activity. *Nucleic Acids Res.* **34**, 4245–4257 (2006).
22. Hunt, D. M. *et al.* Long-range transcriptional control of an operon necessary for virulence-critical ESX-1 secretion in *Mycobacterium tuberculosis*. *J. Bacteriol.* **194**, 2307–2320 (2012).
23. Newton-Foot, M. & Gey van Pittius, N. C. The complex architecture of mycobacterial promoters. *Tuberculosis* **93**, 60–74 (2013).
24. Zheng, X., Hu, G.-Q., She, Z.-S. & Zhu, H. Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC Genomics* **12**, 361 (2011).
25. Mitra, A., Angamuthu, K. & Nagaraja, V. Genome-wide analysis of the intrinsic terminators of transcription across the genus *Mycobacterium*. *Tuberculosis (Edinb.)* **88**, 566–75 (2008).
26. Unniraman, S., Prakash, R. & Nagaraja, V. Alternate Paradigm for Intrinsic Transcription Termination in Eubacteria. *J. Biol. Chem.* **276**, 41850–41855 (2001).
27. Czyz, A., Mooney, R. a, Iaconi, A. & Landick, R. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio* **5**, e00931 (2014).
28. Ingham, C. J., Hunter, I. S. & Smith, M. C. M. Rho-independent terminators without 3' poly-U tails from the early region of actinophage  $\phi$ C31. *Nucleic Acids Res.* **23**, 370–376 (1995).
29. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* **97**, 7090–7095 (2000).
30. Johnson, C. M. *et al.* Identification of a conserved branched RNA structure that functions as a factor-independent terminator. *Proc. Natl. Acad. Sci. USA* **111**, 3573–3578 (2014).
31. Okuda, S. *et al.* Characterization of relationships between transcriptional units and operon structures in *Bacillus subtilis* and *Escherichia coli*. *BMC Genomics* **8**, 48 (2007).
32. Gupta, S. & Chatterji, D. Stress responses in mycobacteria. *IUBMB Life* **57**, 149–159 (2005).
33. Bhat, S. A. *et al.* The mechanism of redox sensing in *Mycobacterium tuberculosis*. *Free Radic. Biol. Med.* **53**, 1625–41 (2012).
34. Stewart, G. R. *et al.* Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology* **148**, 3129–3138 (2002).
35. Hernandez, D. R. Regulation of expression of a *Neisseria gonorrhoeae* tRNA-modification enzyme (Gcp). (The University of Arizona. at <http://arizona.openrepository.com/arizona/handle/10150/242381> (2012).
36. Unniraman, S., Chatterji, M. & Nagaraja, V. DNA gyrase genes in *Mycobacterium tuberculosis*: A single operon driven by multiple promoters. *J. Bacteriol.* **184**, 5449–5456 (2002).
37. Forse, L. N., Houghton, J. & Davis, E. O. Enhanced expression of recX in *Mycobacterium tuberculosis* owing to a promoter internal to recA. *Tuberculosis* **91**, 127–135 (2011).
38. Chan, C. T. Y. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* **6**, e1001247 (2010).
39. Chan, C. T. Y. *et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* **3**, 937 (2012).
40. Chionh, Y. H., Babu, I. R., Ng, S., Alonso, S. & Dedon, P. C. Adaptive reprogramming of the mycobacterial proteome during hypoxia by selective translation of stress response genes through the modification of tRNA decoding affinities. *American Society for Microbiology* **104**, at <http://www.asmonlineeducation.com/php/asm2014abstracts/data/papers/104.htm> (2014).
41. Chionh, Y. H., Hia, F., Alonso, S. & Dedon, P. C. Multidimensional quantification of RNA modification dynamics: An integrative approach to study translational reprogramming in pathogenic mycobacteria. *Metabolomics* **132**, at <http://www.omicsonline.org/proceedings/multidimensional-quantification-of-rna-modification-dynamics-an-integrative-approach-to-study-translational-reprogramming-in-pathogenic-mycobacteria-862.html> (2013).
42. Kelkar, D. S. *et al.* Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry. *Mol. Cell. Proteomics* **10**, M111.011627–M111.011627 (2011).
43. Brown, T. Analysis of RNA by northern and slot-blot hybridization. *Curr. Protoc. Immunol.* Chapter 10, Unit 10.12 (2001).

## Acknowledgements

A.H.B. and D.P. acknowledge the Junior and Senior Research Fellowships provided by CSIR. The following reagents were obtained through BEI Resources, NIAID, NIH: Naked Plasmid pVV16 for Expression in *M. smegmatis*, NR-13402 and Genomic DNA from *M. tuberculosis*, Strain H37Rv, NR-14865. We thank Dr. Ashwani Kumar for helpful discussions and generous gift of pCV77 vector. This work was supported by DBT (RGYI Grant No. BT/PR15106/GBD/27/317/2011 to Alka Rao) and CSIR (Grant No. BSC0210C to Alka Rao). “The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication”.

## Author Contributions

A.H.B. and A.R. conceived and planned the experiments, A.H.B. and D.P. performed the experiments, A.H.B. and A.R. analyzed and interpreted the data, and A.H.B., D.P. and A.R. wrote and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Bhat, A. H. *et al.* The *alr-groEL1* operon in *Mycobacterium tuberculosis*: an interplay of multiple regulatory elements. *Sci. Rep.* 7, 43772; doi: 10.1038/srep43772 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017