Article

CUT&Tag reveals unconventional G-quadruplex landscape in *Mycobacterium tuberculosis* in response to oxidative stress

Ilaria Maurizio[1,4], Emanuela Ruggiero[1,4], Irene Zanin[1,4], Marta Conflitti[1], Giulia Nicoletto[1], Roberta Provvedi[2] ✉ & Sara N. Richter[1,3] ✉

Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis, remains a global health threat due to increasing drug resistance and high mortality rates. To combat tuberculosis effectively, novel therapeutic targets are urgently needed. G-quadruplexes (G4s) represent promising candidates for this purpose. In this study, we successfully apply the cleavage under targets and tagmentation (CUT&Tag) technique for the first time in bacteria, mapping the G4 landscape in *Mtb* under standard and oxidative stress conditions, the latter mimicking the environment *Mtb* faces within macrophages. We validate the CUT&Tag protocol using an antibody against the RNA polymerase β-sub-unit, confirming its association with actively transcribed genes. Employing the anti-G4 antibody BG4, we discovered that *Mtb* G4s, unlike their eukaryotic counterparts, predominantly locate within gene coding sequences and consist of two-guanine tract motifs. Notably, oxidative stress increases G4 formation, correlating with reduced gene expression. Our findings provide the first evidence of G4 formation in *Mtb* cells and suggest their potential role in bacterial survival within macrophages. This study demonstrates the successful application of CUT&Tag in bacteria and unveils an unconventional G4 landscape in *Mtb*, offering new insights into bacterial stress response mechanisms and potential therapeutic targets.

Mycobacterium tuberculosis (*Mtb*) is the etiological agent of human tuberculosis (TB)[1]. It is a slow-growing bacillus characterized by a high lipid content cell wall that confers the bacterium protection in hostile environments and virulence properties[2,3].

Mtb is transmitted to susceptible individuals through droplets that are released into the air. Once inhaled, bacilli enter the respiratory tract, reach the lungs and infect alveolar machrophages[4]. Here, as defensive mechanism, the human immune system produces reactive nitrogen and oxygen species that damage bacterial structures like DNA and proteins[5]. *Mtb* has developed counteractive measures to survive and replicate within macrophages in oxidative conditions, thanks to

different antioxidant enzymes[6]. Despite global efforts, TB persists as a significant public health challenge, with recent data indicating 1.3 million fatalities and 10.6 million new cases in 2022[7]. The complexities of prolonged treatment regimens, the emergence of drug-resistant strains, and the absence of an efficacious vaccine underscore the imperative for novel therapeutic approaches.

In this scenario, G-quadruplexes (G4s) may represent alternative targets for the development of new therapeutic approaches. With roughly 4000 genes and an overall GC content around 65.6%[8], the *Mtb* genome is likely prone to G4 formation. G4s are alternative nucleic acid secondary structures that arise in single-strand, guanine-rich DNA

[1]Department of Molecular Medicine, University of Padua, Padua, Italy. [2]Department of Biology, University of Padua, Padua, Italy. [3]Microbiology and Virology Unit, Padua University Hospital, Padua, Italy. [4]These authors contributed equally: Ilaria Maurizio, Emanuela Ruggiero, Irene Zanin.
✉e-mail: roberta.provvedi@unipd.it; sara.richter@unipd.it

and RNA sequences from the planar association of four guanines (G-tetrads) through Hoogsteen hydrogen bonds; G-tetrads are linked by loops of different length and nucleotide composition, which contribute to the polymorphic nature of G4 structures[9–11]. Monovalent cations like K^+ or Na^+ promote G4 stability. G4s have been extensively studied in eukaryotic cells where they have been demonstrated to form in vivo preferentially at genes promoters and in nucleosome-depleted regions[10]. They have been implicated in various cellular functions, spanning from DNA replication, transcription, and translation to telomere maintenance, gene instability, and cancer[9–11].

Conversely, the G4 field in the prokaryotic kingdom is still poorly investigated: G4 formation has never been described in vivo, and G4 physiological role in bacteria has been only scantily elucidated so far. The *Mtb* genome has been predicted by bioinformatic analysis to potentially form more than 10,000 G4 motifs. To date in silico and in vitro studies have focused on the identification of canonical G4 motifs, composed of at least three guanine per G-tract and preferentially located at gene promoters. G4s have been predicted to form in regulatory regions of genes belonging to specific functional categories, including "cell wall and cell process" and "intermediate metabolism" or in genes that modulate the host-pathogen interaction. Additionally, known G4 ligands, like BRACO-19, c-exNDI 2 and TMPyP4, have demonstrated inhibitory effect on *Mtb* cell growth[12,13]. *Mtb* encodes for three *E. coli* ortholog helicases, namely UvrD1, UvrD2, and DinG, which have been demonstrated to unwind G4s[14–16], supporting the presence of folded G4s in *Mtb* cells.

Here we successfully performed cleavage under targets and tagmentation (CUT&Tag) genome-wide mapping in *Mtb* to define the G4 landscape. The analysis was performed in the presence and absence of oxidative stress treatment to reflect the conditions faced by the bacterium when internalized within the host macrophages. We report the CUT&Tag output of two different antibodies that target the β subunit of the RNA polymerase and G4 motifs. We show that the CUT&Tag can be efficiently applied to complex bacteria like *Mtb*, allowing the in vivo mapping of different targets. We demonstrate that, in contrast to eukaryotes, G4s are mainly located within gene coding sequences and are mostly composed of two-guanine tracts. We showed that in the oxidative stress condition, more G4s form and associate with genes with low transcript abundance, suggesting a unique role for G4s in *Mtb* and possibly in bacteria.

Results

Optimization of CUT&Tag for genome-wide profiling in *Mtb*

To map the location of G4s in *Mtb* genome-wide, we initially applied a chromatin immunoprecipitation (ChIP) protocol with BG4, the most used anti-G4 antibody[17], followed by either qPCR or sequencing. Experiments were conducted under standard laboratory growth conditions and under oxidative stress conditions to compare the G4 genomic distribution in *Mtb* cultures exposed to environments mimicking host infection.

ChIP is a multi-step antibody-based technique which allows the identification of specific features within chromatin[18]. The process involves harvesting bacterial cells, formaldehyde cross-linking to preserve biological interactions, and cell lysis before chromatin shearing. The latter is a critical step that requires fragment distribution between 100-500 bp[19]. Immunoprecipitation with the target antibody follows, and enrichment of the genomic sequences is assessed via quantitative PCR (qPCR) or next-generation sequencing (NGS). In our set-up, each ChIP experiment consisted of three samples: IP (immunoprecipitated DNA), MOCK and input. The IP was treated with the antibody of interest, while the MOCK sample underwent immunoprecipitation in the absence of the antibody, serving as control for non-specific background noise. The input sample, consisting of unprocessed sheared chromatin, enabled qPCR data normalization and target enrichment evaluation[18]. After ChIP, qPCR assessed the effectiveness of the applied

protocol, selecting three genomic regions as positive controls: *zwf1* and *mosR*, previously shown to fold into G4s in vitro[12] and *lpqS*, as an additional G4-forming region (Supplementary data 1). The G4 folding of these selected positive regions was confirmed by CD analysis (Supplementary data 1 and Fig. S1). As negative control, we identified a genomic region in the *Rv1507A* gene that, based on its sequence, could not fold into G4 in vitro. ChIP-qPCR analysis showed remarkable enrichment of IP positive samples with respect to the MOCKs, both in standard growth conditions and upon treatment with H_2O_2, indicating high BG4 specificity. Replicates were highly consistent, with positive regions at least five-fold enriched over the negative control, aligning with successful ChIP-qPCR data reported in the literature (Fig. S2A)[20]. Encouraged by these results, we proceeded with sequencing and data analysis. However, identifying a clear signal above the background proved to be challenging. Although we observed a general enrichment in putative G4 positive regions (i.e. *zwf1*) and no enrichment in the negative region *Rv1507A* compared to the input sample (Fig. S2B-C), the signal was characterized by broad enriched domains. Commonly used peak calling tools, such as MACS2[21] and Epic2[22], struggled to identify true enriched regions above the background. Interestingly, while G4-ChIP is widely and successfully applied in eukaryotes[23–27], it has never been applied in bacteria, possibly due to different chromatin organization and accessibility[28] that would explain the observed low signal-to-noise ratio. Since the noisy background limited an efficient peak calling, which is fundamental for bioinformatic analysis of the sequencing results, we moved to CUT&Tag, a technique known for providing a significantly lower background.

In fact, CUT&Tag is an advanced genomic profiling method that improves ChIP-seq in several aspects. It offers higher sensitivity, specificity, and reproducibility while requiring reduced input material and sequencing depth. Unlike ChIP-seq, CUT&Tag allows native-state chromatin mapping[29]. Its efficacy in profiling histone modifications and transcription factor binding, combined with reduced cost-effectiveness, has contributed to its widespread adoption across various organisms, including humans[26,30–32], mice[33], zebrafish[34], plants[35], and yeasts[36]. However, its application in bacteria has never been explored so far.

In the CUT&Tag protocol, cells are initially bound to magnetic beads and then mildly permeabilized to facilitate access of the target antibody and Tn5 enzyme to the cell genome[29]. We optimized *Mtb* cells binding to concanavalin A (ConA)-coated beads by measuring optical density upon varying incubation times (10 and 20 min), achieving optimal bead saturation at 20 min (Fig. S3A). To determine the optimal bacterial cell number, we performed CUT&Tag on three different sample sizes, 1, 10, and 20 million cells, quantifying bacterial DNA before and after library preparation (Fig. S3B, C). The highest DNA yield was obtained using 10 million cells.

To validate the novel CUT&Tag protocol in *Mtb*, we initially targeted the RNA polymerase (RNA Pol) subunit β, for which genome-wide ChIP-seq data were previously reported[37]. Experiments were conducted under standard growth and oxidative stress conditions. Briefly, bacterial cultures underwent mild formaldehyde fixation, and the permeabilized cells were incubated with the antibody, tagmented with pA-Tn5 transposase, and the PCR-amplified fragments were subjected to NGS. Each CUT&Tag experiment consisted of three biological replicates, exhibiting high Pearson's correlation coefficients, indicative of robust reproducibility (Fig. 1A). Principal component analysis (PCA) revealed proper clustering according to the growth conditions (Fig. 1B), indicating high data quality and reproducibility among replicates performed in the two conditions. All samples achieved appropriate sequencing depth, with FRiP values largely exceeding 0.01, a threshold associated with successful experiments (Fig. 1C)[38]. Peak size distribution showed that the median peak size across all samples was ~200 bp, as expected[20] (Fig. 1D). Analysis of CUT&Tag signal distribution at the center of the called peaks showed a
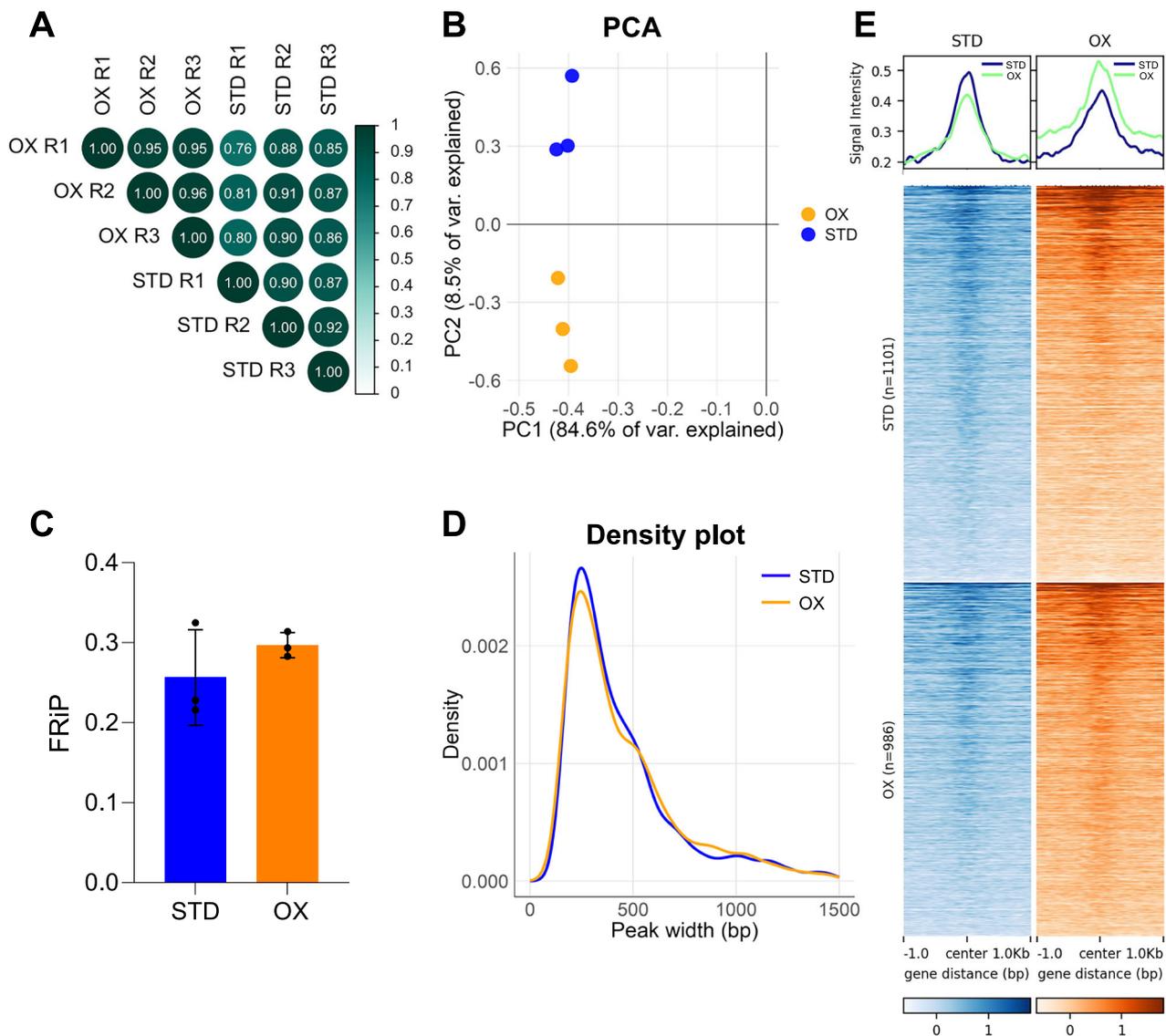
Fig. 1 | Validation of RNA Pol-CUT&Tag in the *Mtb* genome. A Correlation plot among CUT&Tag biological replicates ($n = 3$, R1, R2, R3) performed in standard growth (STD) and oxidative stress (OX) conditions. **B** Principal component analysis (PCA) for the three replicates of each condition. **C** FRiP analysis. Results are shown as mean ± s.d. of independent biological r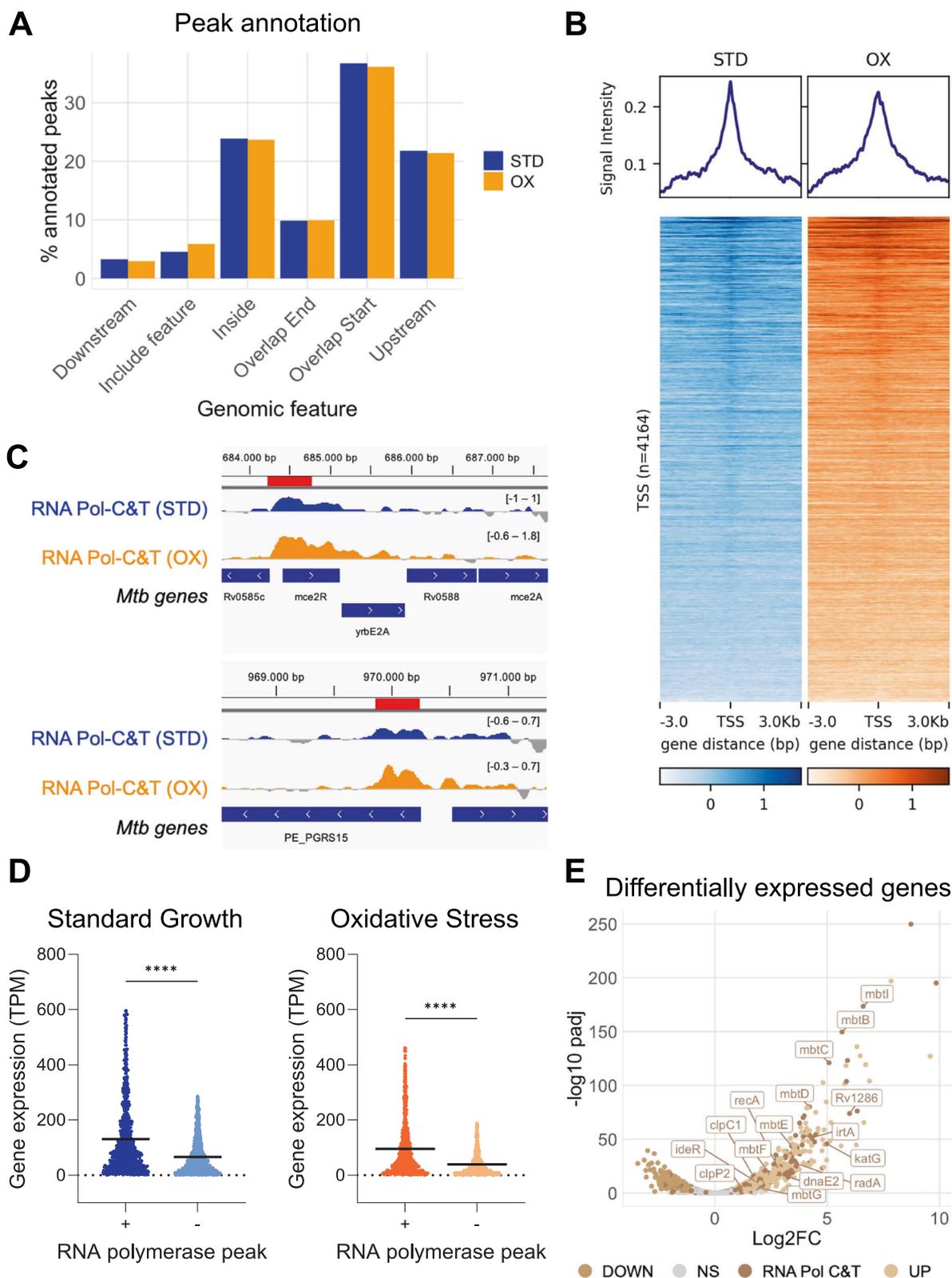eplicates ($n = 3$). **D** Density plot of peaks width distribution. **E** CUT&Tag mean signal distribution at high confidence peaks. Average plots (top panel) and density heatmaps (bottom panel) of CUT&Tag reads referring to peak center, within ± 1 kb distance. Data referring to STD and OX conditions are shown in blue and orange, respectively. Source data are provided as a Source Data file and supporting data 3.

consistent enrichment in density heatmaps, both in high confidence peaks (Fig. 1E), defined as those present in at least two of three biological replicates, and in each replicate (Fig. S4) in the two growth conditions, supporting the efficiency of immunoprecipitation.

We then conducted a comparative analysis between our CUT&Tag results and the previously published ChIP-seq data, where positive regions were identified by calculating an enrichment ratio over input at defined locations[37]. Applying an analogous analytical approach to our CUT&Tag data, using the IgG profile as background, we found that 94% (874/924) of ChIP-seq-identified regions were retrieved by CUT&Tag (Fig. S5). Furthermore, consistent with observations in eukaryotic systems[29], CUT&Tag showed enhanced sensitivity, enabling the identification of additional enriched regions. As an example, we report the visual comparison of RNA Pol coverage signals from ChIP-seq and CUT&Tag at the *rrl* locus that showed enrichment in both techniques and in both CUT&Tag conditions (STD and OX) (Fig. S5B-C). Overall, these findings strongly support the efficacy and applicability of the CUT&Tag protocol in *Mtb* for genome-wide profiling studies.

Genome-wide RNA polymerase profiling by CUT&Tag in *Mtb*

The sequencing profile of the RNA Pol-CUT&Tag revealed peaks across the entire *Mtb* genome: 1101 high confidence peaks were identified in standard growth, while oxidative stress yielded 986 peaks (Supplementary data 3). Regardless of the growth condition, RNA Pol-peaks showed a similar peak distribution in genomic feature annotation with a predominant location around gene start sites (36.7% and 36.1% in standard and oxidative stress condition, respectively), within genes (23.9% and 23.7%, respectively), and upstream of genes (21.8% and 21.4%, respectively) (Fig. 2A). As the β subunit constitutes an essential component of the holoenzymatic complex, the predominant location of peaks near gene start sites indicates that the enzyme is preparing to initiate transcription. The significant abundance of peaks residing within gene bodies or overlapping gene ends indicates that the RNA polymerase complex is either initiating or finishing the transcription process. This peak distribution was confirmed by density heatmaps of coverage at gene transcription start site (TSS) (Fig. 2B) and aligns with the pivotal role of the β subunit throughout the transcription process,

A Peak annotation

B

C

D Standard Growth / Oxidative Stress

E Differentially expressed genes

in both conditions, from initiation to termination[39]. In addition, RNA Pol-CUT&Tag identified previously reported positive regions for RNA Pol binding in *Mtb* genome, *rv0586/mceR2* and *rv0872c/PE_PGRS15*[40] (Fig. 2C), further validating the protocol.

Since RNA Pol is the primary enzyme involved in transcription, we integrated CUT&Tag results with *Mtb* gene expression levels from RNA-seq data. Raw read counts were converted into transcripts per

million (TPM) for individual genes and associated with their corresponding genetic locus. Gene expression levels were then correlated with the presence or absence of RNA Pol-CUT&Tag peaks. A statistically significant increase in gene expression was observed in regions exhibiting RNA Pol-peaks versus those lacking them, both under standard and oxidative stress conditions (Fig. 2D). Genes were categorized into three groups based on gene expression interquartile (IQR)

Fig. 2 | Genome-wide RNA Polymerase profiling by CUT&Tag in *Mtb*. A RNA Pol-CUT&Tag high confidence peak annotation in standard growth condition (STD, blue bars) and upon oxidative stress (OX, orange bars). "Feature" stands for "gene"; "Include feature" indicates the peaks that extend over the whole gene coding sequence. **B** RNA Pol-CUT&Tag mean signal distribution at high confidence peaks. Average plots (top panel) and density heatmaps (bottom panel) of CUT&Tag reads in standard growth (STD, blue) and oxidative stress (OX, orange) conditions referring to peaks center, within ± 3 kb distance. **C** Visualization tracks of RNA Pol-CUT&Tag profiles, referred to known RNA Pol binding sites. Gene annotation is reported at the bottom of each track. **D** Expression levels of *Mtb* genes presenting (+) or not presenting (−) RNA Pol-CUT&Tag peaks in standard growth (left panel) and in oxidative stress conditions (right panel). Statistical differences were analyzed using unpaired t-test with two tails (STD: n (+) = 866, n (−) = 2649, p-value < 0.0001 (****); OX: n (+) = 740, n (−) = 2637, p-value < 0.0001 (****)); Mean values are reported as horizontal black lines. **E** Volcano plot illustrating differentially expressed genes upon oxidative treatment. Significant up/downregulated genes were identified using Deseq2 with p-value = 0.05 and Log2FC = 1. Dark brown dots indicate RNA polymerase-retrieved genes identified in oxidative stress conditions. Genes reported to be implicated in the response to oxidative stress are highlighted. Source data are provided as a Source Data file and supporting data 3.

values (low, medium, and high). A substantial increase in the expression levels was observed for all RNA Pol-peaks across all categories (Fig. S6), consistently associating the presence of RNA Pol-peaks with enhanced transcriptional activity.

Differential gene expression analysis was subsequently performed between standard and oxidative stress conditions. It is established that *Mtb* treatment with 5 mM H_2O_2 substantially alters the bacterial expression profile[41]. Genes were considered differentially expressed when meeting criteria of log2-fold change exceeding 1 or falling below −1, with adjusted p-value < 0.05. Under these criteria, 732 and 650 genes were found to be upregulated and downregulated, respectively. Among these, 211 upregulated and 82 downregulated genes were associated with an RNA Pol-peak, constituting 28.8% and 12.6%, respectively, of the differentially expressed genes in response to oxidative stress (Fig. 2E). In these conditions, the induction of scavenging enzymes becomes also crucial for *Mtb* survival[42,43]. Indeed, *katG*, which efficiently reduces reactive oxygen species (ROS) concentrations, was strongly upregulated, along with *cysN*, responsible for sulfate activation in cysteine biosynthesis, and enzymes involved in the DNA damage response pathway. Other upregulated genes under oxidative stress conditions included iron-related genes such as the *Mtb* gene cluster encoding mycobactin (iron-chelating siderophore[41]), *irtA* and *ideR* (involved in iron import and regulation, respectively), DNA repair enzymes (*recA*, *radA*, and *dnaE2*[41,42]), and protein repair or degradation systems (*clpC1* and *clpP2*).

These results collectively validate the efficiency of the CUT&Tag protocol in profiling the *Mtb* genome, both under standard growth and oxidative stress conditions.

## G4-CUT&Tag reveals a G4 landscape enriched in two-tetrad G4s

We then applied the CUT&Tag protocol using the BG4 antibody to define the G4 landscape in the *Mtb* genome under standard and oxidative stress growth conditions. The G4-CUT&Tag replicates showed high correlation values among them (Fig. S7A), expected fragment distribution[44] (Fig. S7B), and exceeded the accepted threshold for library complexity and sequencing depth (Fig. S7C). Quantification of CUT&Tag signal at G4-peaks showed that the BG4 signal was significantly higher than the negative control (IgG-CUT&Tag): linear regression analysis of antibody coverage correlation yielded non-linear signal distribution with significant adjusted R^2 values of 0.63 and 0.67 for standard and oxidative conditions, respectively, indicating that the BG4 antibody is specifically and reliably detecting the G4 structures in both conditions, with a slightly stronger correlation in oxidative condition (Fig. S7D). BG4 enrichment was further supported by positive values obtained from the difference of the log2 signal at the majority of G4 sites (Fig. S7D). Coverage distribution analysis revealed high signal density at the center of both high confidence peaks (Fig. S7E) and peaks separately identified in the biological replicates (Fig. S9), validating the G4-CUT&Tag approach in both tested conditions (Fig. S7E). Visual inspection of G4-CUT&Tag sequencing profiles showed remarkable improvement in signal-to-noise ratio (Fig. S7F) with respect to G4-ChIP data (Fig. S2C), enabling the identification of well-defined peaks distributed across the entire *Mtb* genome.

Notably, G4 peak annotation showed an unprecedented distribution of G4s, with almost 60% of peaks in both conditions located within gene bodies. The remaining peaks mostly overlapped with gene start and end sites (Fig. 3A). This distribution strongly differs from the typical eukaryotic pattern, where G4s predominantly accumulate at promoter regions[23,25], thus suggesting different regulatory mechanisms in *Mtb*. Motif analysis of high confidence peaks confirmed G-rich patterns capable of G4 formation in both conditions (Fig. 3B). Consequently, we applied the G4 prediction algorithm Quadparser[45] on high confidence peaks. We screened for motifs with the following characteristics: i) G-tracts of at least 2 or 3 guanines each; ii) loop length up to 7 (short) or 12 (long) nucleotides. Surprisingly ~99% of putative G4-forming sequences consisted of two-guanine tracts and short loops (Fig. 3C), generally associated with weaker G4s. Almost all G4 peaks contained this type of G4-forming sequences (Fig. S8). To assess the significance of our G4 prediction, we conducted a randomization analysis. We extracted random fragments from the entire *Mtb* genome, with abundance and average length comparable to that of CUT&Tag peaks (2000 fragments of approximately 260 bp), and applied the Quadparser algorithm to these random fragments, in accordance with the analysis on G4 peaks. We then calculated the fold enrichment of CUT&Tag samples over random fragments and observed that 2-tetrad G4s, except for those with short loops, exhibited a fold enrichment higher than 1, indicating that their identification is statistically significant and did not occur by chance. On the other hand, 3-tetrad G4s showed a fold enrichment much lower than 1, confirming the statistical significance of their absence from the pool of the identified peaks (Fig. 3D).

CD spectroscopy validated G4 folding for six representative immunoprecipitated sequences from both conditions (Supplementary data 1, Fig. 3E). Most sequences showed the typical CD signature of a G4 with antiparallel topology, characterized by two positive peaks at λ ~ 240 and 290 nm and a negative one at λ ~ 260 (STD2, STD3, STD4, STD6, OX1, OX6). The remaining sequences showed a mixed G4 topology, with positive peaks at λ ~ 260 and 290 nm, with different intensities. Overall, all tested sequences were confirmed to fold into G4s, further validating the reliability of our G4-CUT&Tag in *Mtb*.

G4-bearing genes under oxidative stress are associated with lower transcript levels than in standard growth conditions

We compared G4 enrichment between standard growth and oxidative stress conditions. Remarkably, oxidative stress yielded more G4 high confidence peaks (1087) than standard growth (748) (Supplementary data 4). We identified 475 peaks shared between the two conditions, representing consistently folded G4s. Notably, over 50% of G4 peaks were unique to the oxidative condition, indicating that G4 folding was induced in response to the oxidative treatment. Conversely, less than 40% of peaks were unique to standard growth condition, suggesting that G4s may be more relevant in oxidative stress response (Fig. 4A). Genomic distribution analysis of unique and common peaks confirmed previous observations, with approximately 60% enrichment in gene bodies (Fig. 4B) and low abundance of BG4 signal at gene TSS (Fig. 4C). Visual inspection of unique G4-peaks in oxidative conditions
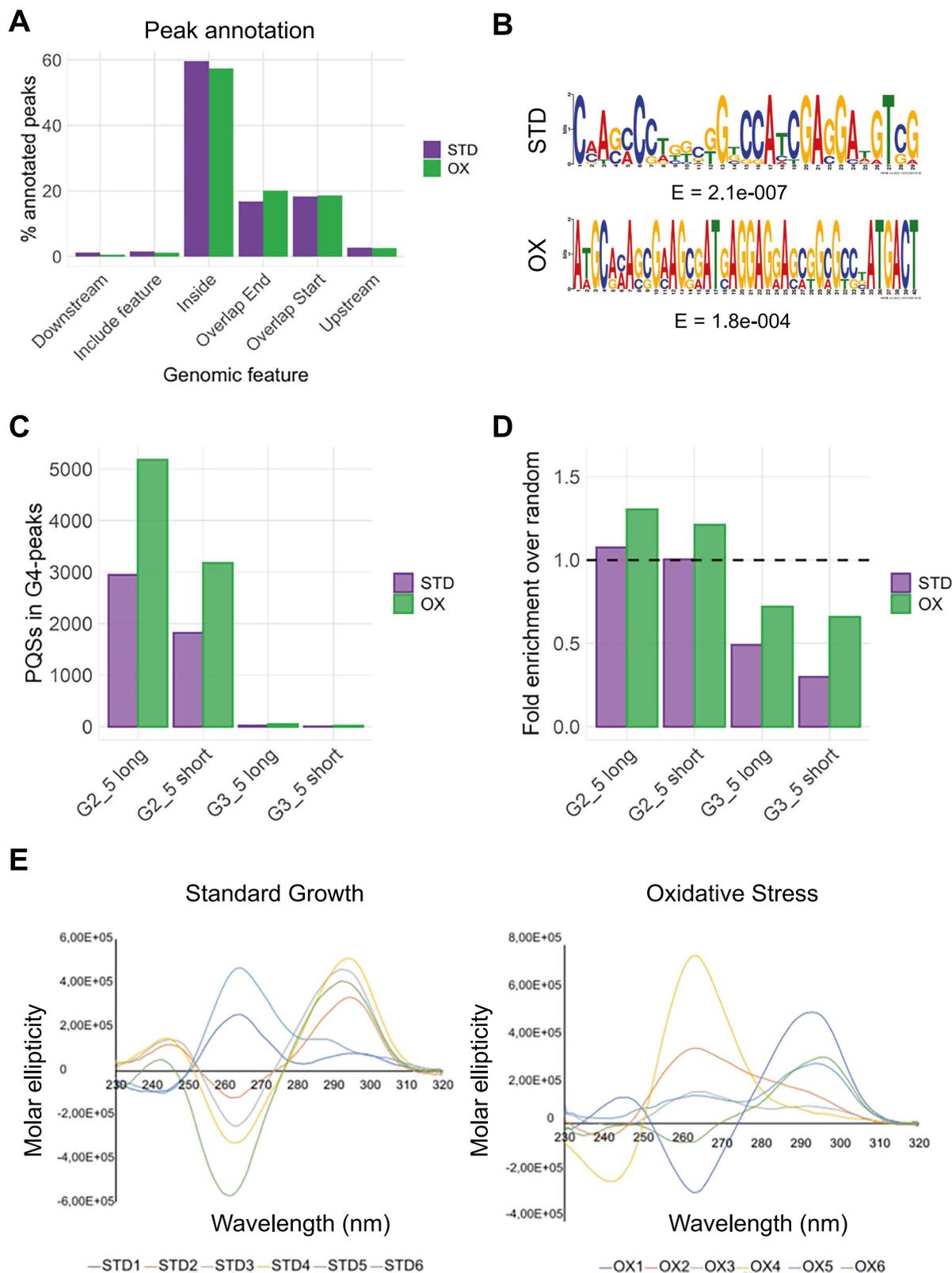
Fig. 3 | G4-CUT&Tag in the *Mtb* genome. A G4-peak annotation in standard (STD, purple bars) and oxidative conditions (OX, green bars). "Feature" stands for "gene"; "Include feature" indicates the peaks that extend over the whole gene coding sequence. **B** MEME motifs obtained from G4-peaks in standard (STD) and oxidative conditions (OX). Motif logo and E-values are reported. **C** G4 motifs prediction of retrieved peak sequences, performed with the Quadparser mapper tool, in both tested conditions. G tracts composed of at least 2 or 3 Gs and up to 5 Gs were considered, with short (0-7 nucleotides) and long (0-12 nucleotides) loop lengths. **D** Fold enrichment of Quadparser-identified sequences over randomly generated fragments in the *Mtb* genome. **E** CD spectra of representative sequences derived from standard growth (left panel) and oxidative stress (right panel) conditions. Source data are provided as a Source Data file and supporting data 4.
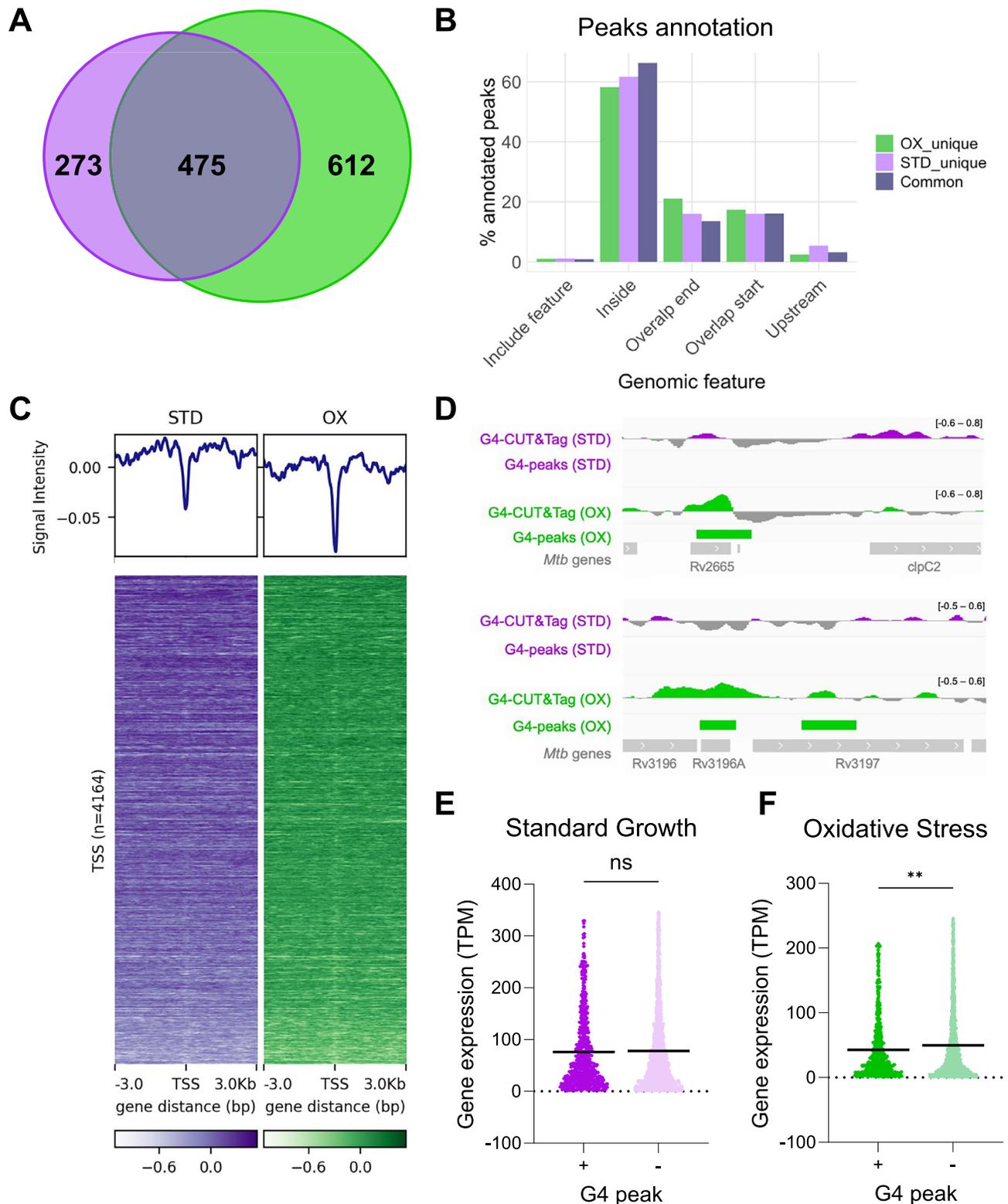
Fig. 4 | G4 enrichment in oxidative stress versus standard growth conditions.
A Venn diagram representing the G4-CUT&Tag peaks shared between standard (STD, purple) and oxidative (OX, green) conditions. **B** Genomic annotation of shared and unique peaks identified in standard and oxidative conditions. **C** Average plot (top) and density heatmap (bottom) of G4-CUT&Tag peaks in standard (purple) and oxidative (green) conditions. Peak signal distribution refers to gene TSS within ± 3 kb distance. **D** Visualization of G4-CUT&Tag profiles for oxidative unique peaks. Genomic tracks are reported as the difference between G4-CUT&Tag and IgG-CUT&Tag (grey) signals, both derived from the mean of three biological replicates. Expression levels of *Mtb* genes presenting (+) or not presenting (−) G4-CUT&Tag peaks in standard (**E**) and oxidative conditions (**F**). Statistical differences were analyzed using the unpaired t-test with two tails (STD: n (+) = 592, n (−) = 2906, p-value = 0.5028 (ns); OX: n (+) = 779, n (−) = 2590, p-value = 0.0014 (**)). Mean values are reported as horizontal black lines. Source data are provided as a Source Data file and supporting data 4.
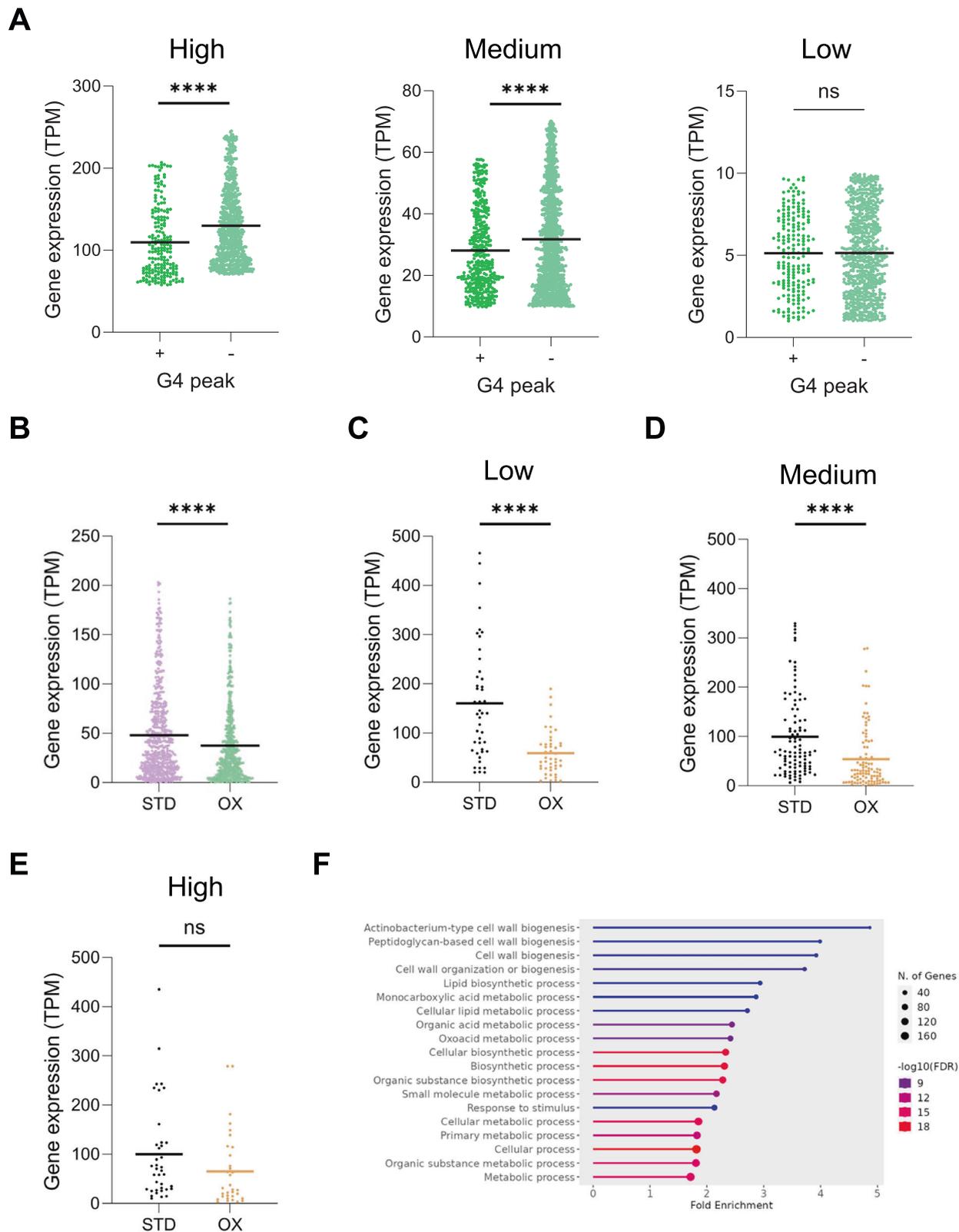
further confirmed the high signal-to-noise ratio and efficiency of the CUT&Tag analysis (Fig. 4D).

Integration of RNA-seq analysis with G4-CUT&Tag data revealed no significant difference in transcript levels between G4-bearing and non-G4 genes under standard growth conditions (Fig. 4E). In contrast, oxidative stress induced an overall reduction in gene expression likely induced by the treatment[41,42], with G4-bearing genes showing

statistically significant lower levels of transcripts compared to non-G4 genes (Fig. 4F).

Detailed analysis revealed that high and medium-expressing G4-bearing genes underwent significant downregulation compared to non-G4 counterparts, while low-expressing genes showed no significant difference (Fig. 5A). To further investigate this finding, we analyzed the G4 peaks unique to the oxidative condition and

Fig. 5 | Genes presenting G4s only in oxidative stress condition. A Expression levels of *Mtb* genes presenting (+) or not presenting (−) G4-CUT&Tag peaks in the oxidative stress condition grouped into three IQR categories according to their expression levels: High (left panel), Medium (central panel) and Low (right panel). Statistical differences were analyzed using the unpaired t-test with two tails (High: *n* (+) = 195, *n* (−) = 648, *p*-value < 0.0001 (****); Medium: *n* (+) = 390, *n* (−) = 1294, *p*-value < 0.0001 (****); Low: *n* (+) = 194, *n* (−) = 648, p-value = 0.9753 (ns)); Mean values are reported as horizontal black lines. **B** Expression levels in both standard (*n* = 542) and oxidative (*n* = 526) conditions of genes presenting G4s only in oxidative conditions and normalized to housekeeping *sigA* expression (*p*-value < 0.0001 (****)). Statistical differences were analyzed using the unpaired t-test with two tails and mean values are reported as horizontal black lines. **C–E** Expression levels of genes presenting G4s only in the oxidative stress conditions and grouped into three IQR categories according to RNA occupancy difference between oxidative and standard conditions. Statistical differences were analyzed using the unpaired t-test with two tails (High: *n* (+) = 38, *n* (−) = 33, *p*-value = 0.1010 (ns); Medium: *n* (+) = 100, *n* (−) = 98, *p*-value < 0.0001 (****); Low: *n* (+) = 45, *n* (−) = 46, *p*-value < 0.0001 (****)); Mean values are reported as horizontal black lines. **F** GO enrichment analysis of unique oxidative stress G4-bearing genes (FDR = 0.05). The horizontal axis represents the fold enrichment; the vertical axis represents the GO term or functional category; the size of the dot represents the number of genes in the GO term and the color of the dot represents the p-adjust value. Source data are provided as a Source Data file and supporting data 4.

compared their transcript levels in both standard and oxidative conditions: we found a significant downregulation upon G4 folding (Fig. 5B), therefore further supporting that G4 formation in the *Mtb* genome during oxidative stress is associated with gene expression downregulation.

We next analysed RNA Pol occupancy (Fig. 2) at genes presenting G4s only in the oxidative condition. We calculated the difference in RNA Pol coverage between oxidative and standard conditions at these G4 sites, and we defined three subsets according to the IQR method: Low, representing lower RNA Pol occupancy in oxidative condition; Medium, representing similar RNA Pol occupancy between the oxidative and standard conditions; High, showing higher RNA Pol occupancy in oxidative stress condition. Next, the genes included in the subsets were associated with their relative expression values in the two conditions. This analysis revealed that genes in the Low dataset (Fig. 5C) showed significant gene downregulation under oxidative stress. This behavior reflects genomic sites where G4 folding induced by oxidative stress is associated with a block in transcription. This result was obtained also with genes in the Medium dataset (Fig. 5D). Notably, genes in the High dataset (Fig. 5E) did not show an increase in gene transcription, even in the presence of high RNA Pol coverage, suggesting that the presence of G4s reduced RNA Pol activity also in this dataset.

We then investigated the possible role of genes presenting G4s only in oxidative conditions (Supplementary data 2) by performing GO enrichment analysis, which mostly showed association with metabolic and biosynthetic processes and cell wall biogenesis (Fig. 5F). Detailed analysis on expression-based subgroups (Fig. S10A) revealed that downregulated G4-bearing genes were mainly involved in cell wall organization and biogenesis, while non-differentially expressed genes were associated with biosynthetic pathways (Fig. S10B, C). The up-regulated group included too few genes for significant analysis.

In summary, our findings indicate that the G4-bearing genes recovered during *Mtb* stress response predominantly engage in various metabolic processes, as well as catalytic and binding functions. Interestingly, the identified enriched pathways are typical of *Mtb* stress-induced response, both at transcriptomic and proteomic levels[46], suggesting that G4s might be constitutive elements for *Mtb* response to stress.

Discussion

This study marks the first successful application of CUT&Tag in a prokaryotic system, specifically in *Mtb*, where we demonstrated its versatility by locating the *Mtb* transcriptional complex and folded G4s. CUT&Tag has been previously applied to human, yeast and plant cells, where it was shown to offer several advantages over traditional immunoprecipitation techniques like ChIP-seq. It streamlines the experimental process through simultaneous employment of different antibodies and contextual library preparation. The elimination of chromatin crosslinking and shearing steps provides a more physiological setting and soft treatment, avoiding epitope masking or biases caused by formaldehyde and sonication[29,47]. It requires less starting material and lower sequencing depth, resulting in an overall lower background and an improved signal-to-noise ratio, crucial for reliable peak identification. This latter aspect proved critical in *Mtb*, where G4-ChIP-seq showed excessive background signal (Fig. S2), while CUT&Tag provided a much-improved profile (Fig. 2C). Our experiments efficiently mapped genomic sites of the *Mtb* transcriptional complex with approximately 5 million reads per sample, whereas ChIP-seq required at least five times more sequencing reads. Furthermore, CUT&Tag demonstrated a higher sensitivity than ChIP-seq, retrieving additional regions besides the ChIP enriched ones (Fig. S5A). The versatility of the CUT&Tag will enable mapping of genomic sites crucial for the understanding of *Mtb* pathophysiology, such as methylation sites and binding sites of sigma factors and other transcription factors, and it will likely be applied to other bacterial species.

We analyzed axenic *Mtb* cultures under standard laboratory conditions and oxidative stress, mimicking the environment faced after macrophage phagocytosis. By mapping the RNA polymerase β subunit, we determined the precise genome-wide location of the transcriptional complex in each condition, with peaks generally associated with highly expressed genes (Fig. 2D-E). Oxidative stress induced a shift in the RNA polymerase binding profile, highlighting the presence of the transcriptional complex in genes responsible for detoxification[30,48]. RNA Pol-CUT&Tag efficiently mapped the shift of the transcriptional complex in response to the oxidative environment, proving to be a suitable method to identify regulatory elements in *Mtb*.

Our main goal was to evaluate changes in the G4 landscape under physiological conditions and during redox-dependent genome condensation, which regulates gene expression in response to stress[6,49,50]. We first ascertained that only a small fraction out of the thousands of bioinformatically predicted G4 motifs[51] were actually present in the *Mtb* genome, in line with eukaryotes[23]. We found that 60% of folded *Mtb* G4s were located within gene bodies, while less than 3% were enriched at promoter regions (Fig. 3A), and in fact, in the immuno-precipitated samples, we did not find the G4s previously characterized at promoters in vitro. This distribution is in contrast with G4s in eukaryotic cells that specifically locate at gene promoters and in nucleosome-depleted regions[23,25]. Our data align with a recent in silico analysis of *M. smegmatis*, a non-pathogenic mycobacterium, that reported higher enrichment of putative G4-forming sequences in mRNA-like strands with respect to promoter regions[52]. This unconventional distribution suggests that G4s in *Mtb*, and possibly in other mycobacteria and prokaryotes, might have different regulatory roles from those characterized so far in eukaryotic cells. It has been recently shown that bacterial histones bend DNA while compacting it[53]. It is possible that this bending also favors G4 folding even in closed chromatin regions, such as those in gene bodies of non-transcribed genes, contrary to eukaryotic histones, where in fact G4s are mostly observed in open chromatin regions.

G4 prediction on peak sequences evidenced an extraordinary abundance (~99%) of G4 motifs consisting of two tetrads (Fig. 3D). Randomization analysis confirmed the statistical significance of our findings, where 3-tetrad G4s and 2-tetrad G4s are significantly absent

and present, respectively, in the *Mtb* BG4-immunoprecipitated samples. Notably, eukaryotic organisms mostly present G4s consisting of three or more tetrads[54], further suggesting distinct regulatory roles for G4s in prokaryotes and eukaryotes. G4 motifs composed of two-guanine tracts, and therefore characterized by an overall lower stability[54], would be more dynamic, which may be important for their role in *Mtb*. However, considering the high guanine abundance characteristic of the *Mtb* genome, it was surprising to find a significant abundance of 2-tetrad G4s folded in the cells, rather than 3-tetrad ones. In this respect, the GC content has been shown to be not the main parameter impacting G4 occurrence within genomes, especially in Archaea and Bacteria kingdoms. For instance, *Leishmania major* and *Chlamydomonas reinhardtii*, which both have 60% GC content, show an opposite G4 density (low and high, respectively)[55].

We observed an increase in G4 number upon oxidative stress treatment, suggesting that G4s could be elements responsive to environmental conditions. The role of ROS in the G4 context is still debated, with studies showing both detrimental and stabilizing effects. ROS have been demonstrated to be responsible for the reduced thermal stability of G4s due to guanine oxidation and consequent disruption of the Hoogsteen hydrogen bonds[43,56]. In contrast, recent studies have shown that oxoguanines (OGs) can stabilize the G4s[57]. A study on the *bcl2* promoter indicated that the presence of OGs increased G4 stability, influencing gene transcription regulation[58]. In the human telomeric G-rich sequence, the presence of oxidative lesions was overcome by the recruitment of undamaged G-tracts that act as spare tires[59]. A similar mechanism could occur in the *Mtb* genome, given the abundance of GG-tracts, potentially explaining the increased abundance of G4s in the oxidative stress condition. It is worth noting that, upon oxidative stress conditions, the bacterial nucleoid undergoes condensation to protect DNA from ROS-induced damage[50]. Thus, DNA overwinding could contribute to double helix destabilization, promoting G4 formation. Indeed, there are specific proteins, e.g. those belonging to the WhiB family, that act as redox sensors and promote condensation of the mycobacterial genome in response to oxidative stress through the binding with G/C-rich genomic regions, thus implying a possible recognition and stimulation of G4-forming sequences[50,60]. Overall, this evidence suggests that the distinctive prevalence of G4 structures in the *Mtb* genome may serve as a protective mechanism against the effects of oxidation that *Mtb* encounters, especially in macrophages during infection.

Interestingly, G4-bearing genes showed an overall reduction in transcript levels under oxidative stress, supporting the hypothesis that G4s could act as sensors of or protectors against environmental stress. Indeed, even in the presence of RNA Pol binding, the expression of genes with G4s induced only in oxidative conditions was significantly lower with respect to the same genes without G4s, i.e. in standard conditions. *Mtb* is known to react to environmental stress by slowing down the metabolic and synthesis pathways and entering a dormancy state[46]. GO enrichment analysis showed that the oxidative stress-unique G4 peaks are indeed present in genes involved in these pathways, further supporting the hypothesis that the reduction of the expression levels of genes presenting G4s contributes to *Mtb* survival within macrophages.

Our findings show that G4 location within the genome exerts different effects on transcription: G4s embedded in gene bodies, as in *Mtb*, mediate repression of transcription, while those at promoters, as in human cells, promote transcription by enhancing transcription factor binding[25,31,61,62]. Hence, G4s may be considered epigenetic features with opposite effects to DNA methylation, which has been shown to repress and promote transcription when located at promoters and gene bodies, respectively[63].

Considering the number of new TB cases that occur annually, and the alarming emergence of TB drug resistant strains combined with the lack of innovative approaches to cure tuberculosis, there is a pressing need for alternative therapeutic targets, and G4s might serve this function. This hypothesis is further supported by the evidence that three known G4 ligands, BRACO-19, a core-extended NDI[12] and TMPyP4[13], inhibit *Mtb* growth, with a G4-related mechanism. However, achieving bacterial G4 selectivity over host G4s is crucial for therapeutic efficacy. Strategies to enhance selectivity include three-dimensional characterization of specific G4s to identify unique features for selective targeting, like a quadruplex-duplex junction[64–68], and conjugation of G4 ligands with native or modified oligonucleotides to enhance stabilization of individual G4s[69,70]. Considering that G4 ligands have proved promising antitubercular activity in vitro, further research in this direction could lead to the development of more selective antimicrobial agents. In conclusion, the present work demonstrates that G4s form in bacterial cells, where they may play a pivotal role in regulating gene expression in the *Mtb* genome and in sensing or responding to stress conditions. These findings indicate G4s as potential novel therapeutic targets to efficiently combat a deadly pathogen that still threatens humanity.

Methods

Bacterial strains and growth conditions

For CUT&Tag and RNA-seq experiments, *M. tuberculosis* H37Rv virulent strain was employed. Bacteria were grown at 37 °C in Middlebrook 7H9 (BD Difco) supplemented with 10% albumin-dextrose-sodium chloride complex, 0.2% glycerol and 0.05% Tween 80 in rolling conditions and were harvested at an early-mid exponential phase. To perform oxidative stress treatment, *Mtb* cultures in mid-exponential phase ($OD_{540} = 0.4$–0.6) were exposed to 5 mM H_2O_2 for 40 min and then harvested.

G4-chromatin immunoprecipitation (G4-ChIP)

Mtb culture was grown at early exponential phase ($OD_{540} = 0.4$–0.6) and then fixed with 1% formaldehyde (ThermoFisher Scientific, #28906) for 10 min at 37 °C rolling. Fixation was quenched by the addition of 125 mM glycine and incubated for 5 min rolling. Bacteria were centrifuged for 5 min at 1300 × g and washed twice with Tris buffered saline (20 mM Tris-HCl pH 7.5, 150 mM NaCl). 300 µL of fixed bacteria were resuspended in immunoprecipitation buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) and sonicated on Bioruptor Plus (Diagenode) at 4 °C to obtain an average fragment size distribution comprised between 100 and 500 bp; sonication cycles consisted of 30 s on followed by 30 s of recovery. Samples were centrifuged at 16,200 × g for 10 min at 4 °C and chromatin aliquots were stored at 80 °C until use.

G4-ChIP experiments were performed upon standard growth conditions and after H_2O_2 treatment. Sheared chromatin batches were derived from independent bacterial cultures. To maximize DNA recovery, each experiment consisted of three IP technical replicates pooled together. 500 ng frozen chromatin was diluted in blocking buffer (25 mM HEPES pH 7.5, 10.5 mM NaCl, 110 mM KCl, 1 mM MgCl₂, 1% BSA). Samples were incubated with 2 µL of RNase A (10 mg/mL, Thermo Fisher Scientific) at 37 °C for 20 min at 800 rpm. After RNase A treatment, INPUT sample was stored on ice until the end of the experiment. 350 ng of in-house BG4 antibody[71] was added to IP samples. IP and MOCK samples were then incubated 1 h at 16 °C, 1200 rpm. Meanwhile, anti-FLAG M2 magnetic beads (Merck, #M8823) were washed three times with blocking buffer. In the last wash, beads were resuspended in blocking buffer and incubated at 16 °C, 1200 rpm along with samples. After 1 h incubation of IP samples with BG4 antibody, 50 µL of magnetic beads were added to IP and MOCK samples, respectively, and incubated for 1 h at 16 °C, 1200 rpm. Samples were then washed three times with ice-cold wash buffer (100 mM KCl, 0.1% Tween 20, and 10 mM Tris pH 7.4). Two additional washes were performed at 37 °C, 1200 rpm for 10 min each. Elution was performed

adding 75 µL of TE buffer to the beads. Samples were treated with 2 µL of 10 mg/mL RNase A at 37 °C for 30 min, 1200 rpm. 0.5% SDS was added to the samples followed by 50 µg proteinase K and DNA decrosslinking was obtained by incubating samples first at 50 °C for 2 h and then 8 h at 65 °C. Samples were purified using MinElute PCR purification kit (QIAGEN, #28004) and qPCR was performed using PowerUP SYBR green Master Mix (Thermo Fisher Scientific, #A25741) on ABI PRISM 7000 (Applied Biosystem) using the following conditions: initial denaturation at 95 °C for 5 min followed by 40 cycles at 95 °C for 10 s and annealing at 60 °C for 30 s. Four different primer pairs (Supplementary data 1) were chosen to target both G4 positive and G4 negative regions. NEBNext Ultra II Library Prep kit for Illumina (NEB, #E7645S) was used for library preparation. Libraries' quality and size was checked with Bioanalyzer 2100 (Agilent) using Agilent DNA High Sensitivity Chips (Agilent, #5067-4626). Samples were sequenced on Illumina Nextseq 500 platform with single end chemistry, using 150 bp reads.

G4-ChIP-seq data analysis
Raw Fastq reads quality were checked on the Galaxy web platform (usegalaxy.org). Trimmomatic (v 0.39) was used to cut reads at 70 bp in length and to remove adaptors. Alignment to *Mtb H37Rv* reference genome was performed using Bowtie 2 (v2.4.2)[72]. Bamcoverage (v 3.5.4) was employed to generate bigwig from aligned bam files. RmDup[73](v 2.0.1) was used to remove duplicate sequences.

Circular dichroism (CD)
Oligonucleotides used for CD analysis were purchased from Sigma-Aldrich and are enlisted in Supplementary data 1. Oligonucleotides were resuspended at a final concentration of 3 µM in 10 mM lithium cacodylate buffer pH 7.4, supplemented with 100 mM KCl and PEG200 40% v/v (Sigma-Aldrich, #88440). Samples were denatured at 95 °C for 5 min and then cooled down overnight at room temperature (RT). CD spectra were acquired on a Chirascan Plus (Applied Photophysics) using a 5 mm quartz cell. Spectra were recorded at 20 °C in the wavelength range comprised between 230 and 320 nm. Data were adjusted for the absorption of their relative buffer and reported as molar ellipticity according to $\theta = degree \times cm^2 \times dmol^{-1}$.

Clevage under targets and tagmentation (CUT&Tag)
CUT&Tag protocol was optimized, starting from procedures applied to eukaryotic cells[29,31,32]. *Mtb* cultures were grown until early exponential phase and harvested by centrifugation at 1200 g for 5 min. Pellets containing approximately 120 million cells were obtained, mildly fixed with 0.1% formaldehyde (ThermoScientific, #28906) for 2 min at RT and the fixing reaction was quenched by adding 75 mM Glycine. This mild fixation procedure does not result in significant chromatin cross-linking[74]. For bacteria binding, ConA beads (CliniSciences, #86057-3) were first activated by resuspension in 1 mL of binding buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM CaCl₂, 1 mM MnCl₂). Bacteria were resuspended in 1 mL wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine) and incubated for 20 min at RT on a rotating wheel. The bacterial cell envelope was permeabilized by incubating the bacteria for 1 h with 2 mg/mL of lysozyme (Serva), followed by a 5 min incubation with 0.1% Triton X-100 (Sigma-Aldrich, #T8787). Beads-bound bacteria were then resuspended in 50 µL of antibody buffer (BSA-wash buffer, 2 mM EDTA), and antibodies were added (anti-RNA Pol (Biolegend ®, #663905) 5 µg; rabbit anti-mouse IgG (Merck, #06-371) 1:100; in-house prepared BG4[71] 500 ng)). Primary antibodies were incubated for 2 h at 16-18 °C on a nutator. Three washes were performed after each antibody incubation step. BG4 samples were incubated with a 1:100 dilution of anti-FLAG M2 antibody (Merck, #F3165) for 1 h at RT. Remaining samples were resuspended in antibody buffer and incubated at RT for 1 h with 1:100 dilution of secondary antibodies. RNA polymerase and

BG4 samples were incubated with a rabbit anti-mouse IgG antibody while the negative control was incubated with an anti-rabbit guinea pig IgG (Rockland, #611-201-122). Tagmentation step required first incubation of samples for 1 h at RT in 300-BSA wash buffer (10 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1X proteinase inhibitor, 0.5% BSA) containing 1:50 dilution of pAG-Tn5 (Tebu-bio, #23615-1017) followed by beads wash and incubation in tagmentation buffer (300-BSA wash buffer with 10 mM MgCl₂) at 37 °C for 1 h. To halt tagmentation, samples were incubated at 63 °C for 1 h and then further incubated at 85 °C for 10 min to inactivate any viable bacteria.

Finally, DNA was purified through phenol-chloroform extraction and subjected to library preparation using the NEBNext High Fidelity 2X PCR Master Mix (Euroclone, #BM0541S) and using universal i5 primer and a uniquely barcoded i7 primer[75] for a total of 14 PCR cycles. Libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, #A63881) and sequenced on an Illumina Nextseq 500 instrument using a 75-bp paired-end sequencing strategy.

Bioinformatic analysis
NGS data were analyzed using both the public Galaxy Web platform and R studio (v 4.3.1). After raw data quality control with FastQC (v 0.73) reads were trimmed using Trim Galore! (v 0.6.7) and aligned to *Mtb* reference genome (NC_000962.3) with Bowtie2 (v 2.5.0)[72] with -I 10 and -X 700 and default paired-end options. Peaks were called against the negative control (IgG-CUT&Tag) using MACS2 (v 2.2.7.1) with a q-value of 0.05. All duplicates were retained. BamCoverage (v. 3.5.4) from deepTools[76] was employed to generate bigwig files which were visualized with Gviz[77] and IGV[78]. deepTools bigwigAverage and bigwigCompare functions were used to create mean signal tracks for each antibody and signal from negative control subtraction, respectively. BEDTools intersect intervals[79] was employed to obtain peak subsets: high confidence peaks among three biological replicates (Supplementary data 3 and 4), common peaks between published RNA Pol ChIP-seq enriched regions and RNA Pol-CUT&Tag ones, and common G4-peaks between standard and oxidative stress condition. Peak annotation was performed using the ChIPpeakAnno[80] package on R studio. Fraction of read in peak (FRiP) was calculated using deepTools CountReadsPerBin module and PCA and correlation plots were plotted with R studio from deepTools multiBigwigSummary function results. Peak annotation was performed using the ChIPpeakAnno[80] package on R studio. Heatmap plots were generated using deepTools, with 25 bp bin used to average the score over the length of the region. TSS regions for heatmap generation were retrieved from the literature[81].

G4 motif prediction and sequence analysis
G4 peaks in both tested conditions were organized in descending order using R studio. The DNA sequences underlying the peaks were retrieved using Extract Genomic DNA on Galaxy platform (v 3.0.3), and G4 sequence motifs were searched with MEME online tool (v 5.5.7) (https://meme-suite.org/meme/tools/meme-chip)[82]. Then, FASTA sequence files from all peaks were employed in the Quad-parser mapper tool[45] to predict G4 motifs possessing a minimum of two tetrads and featuring loop lengths ranging from 0 to 7 or 0 to 12, as indicated. Randomization analysis was performed on 2000 random fragments generated on R studio using the Biostrings package (https://bioconductor.org/packages/).

RNA extraction
For RNA sequencing, 5 mL of *Mtb* culture grown to OD₅₄₀ of 0.4-0.5 were used for RNA extraction. The samples were pelleted at 1000 g for 5 min, resuspended in 1 mL TRIzol® Reagent (Invitrogen™, #15596018), and transferred to 2 mL screw cap tubes containing 500 µL Zirconia/Silica beads (0.1 mm diameter, BioSpec Products). To achieve cellular disruption, a Mini Bead Beater (BioSpec Products) was employed, and cells underwent three rounds of 35 sec of pulses alternated with 35 sec

incubation on ice. Samples were finally centrifuged at maximum speed for 15 min at 4 °C. The supernatant was transferred into a new tube containing an equal volume of 96% Ethanol, and RNA extraction was conducted using Direct-zol RNA MicroPrep kit (Zymo Research, #R2060) following manufacturer's instructions. RNA was eluted using 20 µL nuclease-free water. RNA samples underwent a double treatment with TURBO DNase-free kit (Invitrogen, #AM1907). RNA was quantified on Qubit fluorometer, and RNA quality was evaluated through Tapestation (Agilent). Samples with RIN values above 7 were used. QIAseq FastSelect – 5S/16S/23S (Qiagen, #335925) was employed to perform rRNA depletion and stranded libraries were prepared using Stranded mRNA Prep (Illumina). Samples were sequenced with 150 bp paired-end sequencing chemistry on Illumina Novaseq platform. Three independent biological samples were produced for each experimental condition.

RNA-seq data analysis
RNA-seq data were analyzed using both the Galaxy Web platform and R studio (v 4.3.1). Quality control on samples was performed using FastQC (v 0.73) and reads were aligned to the *Mtb* reference genome (NC_000962.3) using HISAT2 (v 2.2.1). FeatureCounts[83] (v 2.0.3) was employed to count aligned reads with respect to *Mtb* genomic features and differential gene expression analysis was performed using DESeq2[84] package for R studio. Genes exhibiting an adjusted p-value below 0.05 and a Log2 Fold Change (Log2FC) above 1 or below -1 were considered as differentially expressed. Volcano plots were generated using R studio[85].

Gene ontology (GO) analysis
The functional enrichment analysis of BG4-immunoprecipitated genes identified uniquely under oxidative stress condition was achieved using the ShinyGO (v 0.8) application tool[86]. An FDR cutoff of <0.05 was used to mark the functional categories as significant.

Statistics and reproducibility
Circular dichroism data were repeatable; the average values resulting from two independent replicates were plotted. G4-ChIP was performed using sheared chromatin batches derived from independent bacterial cultures ($n = 3$). To maximize DNA recovery, each experiment consisted of three IP technical replicates pooled together. G4-ChIP enrichment was tested with qPCR which significance was calculated with Prism (version 10.0.3) using the two-tailed unpaired t-test. Bacteria binding to ConA beads was tested performing absorbance measurements derived from three independent experiments. Bioanalyzer electropherograms are representative of one CUT&Tag experiment to show libraries size distribution. CUT&Tag and RNA-seq experiments were repeatable, and data are shown for three biological replicates. The significance relative to gene expression values in the presence or absence of CUT&Tag peaks was calculated with Prism (version 10.0.3) using the unpaired t-test with two tails. CUT&Tag heatmaps and genomic profiles were generated using mean bigwig files and high confidence peaks resulting from three biological replicates. Randomization analysis on PQSs enrichment was performed using the Biostrings package to calculate the enrichment of PQSs in G4-peaks.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All genomic raw and processed data produced in the present project (G4-ChIP-seq, RNA Pol-CUT&Tag, G4-CUT&Tag and RNA-seq) have been deposited in the NCBI GEO database under accession numbers: CUT&Tag data GSE272148, ChIP-seq data GSE272149, RNA-seq data GSE272150. ChIP-seq data previously obtained with the RNA Pol antibody were downloaded from GEO at the following accession number GSE40862. Additional processed data are available in Supplementary data 3 and 4. Source data are provided with this paper.

Code availability
Custom-made R scripts are available at https://doi.org/10.5281/zenodo.15837190 and from the corresponding author upon request. Quadparser script was downloaded from https://github.com/dariober/, as indicated by Puig Lombardi et al.[45].

References
1. Alsayed, S. S. R. & Gunosewoyo, H. Tuberculosis: pathogenesis, current treatment regimens and new drug targets. *Int. J. Mol. Sci.* **24**, 5202 (2023).
2. Delogu, G., Sali, M. & Fadda, G. The biology of mycobacterium tuberculosis infection. *Mediterr. J. Hematol. Infect. Dis.* **5**, e2013070 (2013).
3. Daffé, M. & Marrakchi, H. Unraveling the structure of the mycobacterial envelope. *Microbiol. Spectr.* **7** (2019).
4. Pai, M. et al. Tuberculosis. *Nat. Rev. Dis. Prim.* **2**, 1–23 (2016).
5. Nambi, S. et al. The oxidative stress network of mycobacterium tuberculosis reveals coordination between radical detoxification systems. *Cell Host Microbe* **17**, 829–837 (2015).
6. Shastri, M. D. et al. Role of oxidative stress in the pathology and management of human tuberculosis. *Oxid. Med. Cell. Longev.* **2018**, 7695364 (2018).
7. Global Tuberculosis Report. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 (2023).
8. Ducati, R. G., Ruffino-Netto, A., Basso, L. A. & Santos, D. S. The resumption of consumption - a review on tuberculosis. *Mem. Inst. Oswaldo Cruz* **101**, 697–714 (2006).
9. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
10. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).
11. Sato, K. & Knipscheer, P. G-quadruplex resolution: from molecular mechanisms to physiological relevance. *DNA Repair* **130**, 103552 (2023).
12. Perrone, R. et al. Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions. *Sci. Rep.* **7**, 5743 (2017).
13. Mishra, S. K. et al. Characterization of G-quadruplex motifs in espB, espK, and cyp51 genes of Mycobacterium tuberculosis as potential drug targets. *Mol. Ther. Nucleic Acids* **16**, 698–706 (2019).
14. Yadav, P. et al. G-quadruplex structures in bacteria: biological relevance and potential as an antimicrobial target. *J. Bacteriol.* **203**, https://doi.org/10.1128/jb.00577-20 (2021).
15. Saha, T., Shukla, K., Thakur, R. S., Desingu, A. & Nagaraju, G. Mycobacterium tuberculosis UvrD1 and UvrD2 helicases unwind G-quadruplex DNA. *FEBS J.* **286**, 2062–2086 (2019).
16. Thakur, R. S. et al. Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach. *J. Biol. Chem.* **289**, 25112–25136 (2014).
17. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
18. Wiehle, L. & Breiling, A. Chromatin Immunoprecipitation. in *Polycomb Group Proteins: Methods and Protocols*. (eds. Lanzuolo, C. & Bodega, B.) 7–21 (Springer, New York, NY, https://doi.org/10.1007/978-1-4939-6380-5_2 (2016).

19. Keller, C. A. et al. Effects of sheared chromatin length on ChIP-seq quality and sensitivity. *G3* **11**, jkab101 (2021).

20. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **13**, 551–564 (2018).

21. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).

22. Stovner, E. B. & Sætrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392–4393 (2019).

23. Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).

24. Hänsel-Hertsch, R. et al. Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat. Genet.* **52**, 878–883 (2020).

25. Lago, S. et al. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **12**, 3885 (2021).

26. Li, C. et al. Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res.* **31**, 1546–1560 (2021).

27. Nicoletto, G. et al. G-quadruplexes in an SVA retrotransposon cause aberrant TAF1 gene expression in X-linked dystonia parkinsonism. *Nucleic Acids Res.* **52**, 11571–11586 (2024).

28. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **21**, 227–242 (2020).

29. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

30. Fu, Z. et al. Cut&tag: a powerful epigenetic tool for chromatin profiling. *Epigenetics* **19**, 2293411 (2024).

31. Zanin, I. et al. Genome-wide mapping of i-motifs reveals their association with transcription regulation in live human cells. *Nucleic Acids Res.* **51**, 8309–8321 (2023).

32. Lyu, J., Shao, R., Kwong Yung, P. Y. & Elsässer, S. J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res.* **50**, e13 (2022).

33. Rhodes, C. T. et al. An epigenome atlas of neural progenitors within the embryonic mouse forebrain. *Nat. Commun.* **13**, 4196 (2022).

34. Barakat, R., Campbell, C. A. & Espin-Palazon, R. Identification of transcription factor binding sites by cleavage under target and release using nuclease in zebrafish. *Zebrafish* **19**, 104–108 (2022).

35. Wang, Q. et al. ZmICE1a regulates the defence–storage trade-off in maize endosperm. *Nat. Plants* **10**, 1999–2013 (2024).

36. Torres-Garcia, S. et al. Genome-wide profiling of histone modifications in fission yeast using CUT&Tag. *Methods Mol. Biol.* **2862**, 309–320 (2025).

37. Uplekar, S., Rougemont, J., Cole, S. T. & Sala, C. High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in Mycobacterium tuberculosis. *Nucleic Acids Res.* **41**, 961–977 (2013).

38. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).

39. Stephanie, F., Tambunan, U. S. F. & Siahaan, T. J. M. Tuberculosis transcription machinery: a review on the mycobacterial RNA polymerase and drug discovery efforts. *Life* **12**, 1774 (2022).

40. Ju, X. et al. Incomplete transcripts dominate the Mycobacterium tuberculosis transcriptome. *Nature* **627**, 424–430 (2024).

41. Voskuil, M. I., Bartek, I. L., Visconti, K. & Schoolnik, G. K. The response of mycobacterium tuberculosis to reactive oxygen and nitrogen species. *Front. Microbiol.* **2**, 105 (2011).

42. Wu, M., Shan, W., Zhao, G.-P. & Lyu, L.-D. H2O2 concentration-dependent kinetics of gene expression: linking the intensity of oxidative stress and mycobacterial physiological adaptation. *Emerg. Microbes Infect.* **11**, 573–584 (2022).

43. Wu, S. et al. Crosstalk between G-quadruplex and ROS. *Cell Death Dis.* **14**, 37 (2023).

44. Yu, F., Sankaran, V. G. & Yuan, G.-C. CUT&RUNTools 2.0: a pipeline for single-cell and bulk-level CUT&RUN and CUT&Tag data analysis. *Bioinformatics* **38**, 252–254 (2021).

45. Lombardi, E. P. & Londoño-Vallejo, A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* **48**, 1603 (2020).

46. Choudhary, E., Sharma, R., Pal, P. & Agarwal, N. Deciphering the proteomic landscape of Mycobacterium tuberculosis in response to acid and oxidative stresses. *ACS Omega* **7**, 26749–26766 (2022).

47. Ouyang, W. et al. Rapid and low-input profiling of histone marks in plants using nucleus CUT&Tag. *Front. Plant Sci.* **12**, 634679 (2021).

48. Henikoff, S. & Ahmad, K. In situ tools for chromatin structural epigenomics. *Protein Sci.* **31**, e4458 (2022).

49. Walker, A. M., Abbondanzieri, E. A. & Meyer, A. S. Live to fight another day: the bacterial nucleoid under stress. *Mol. Microbiol.* n/a,

50. Chawla, M. et al. Redox-dependent condensation of the mycobacterial nucleoid by WhiB4. *Redox Biol.* **19**, 116–133 (2018).

51. Rawal, P. et al. Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res* **16**, 644–655 (2006).

52. Shitikov, E. et al. Genome-wide transcriptional response of Mycobacterium smegmatis MC2155 to G-quadruplex ligands BRACO-19 and TMPyP4. *Front. Microbiol.* **13** (2022).

53. Hu, Y. et al. Bacterial histone HBb from Bdellovibrio bacteriovorus compacts DNA by bending. *Nucleic Acids Res.* **52**, 8193–8204 (2024).

54. Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682 (2006).

55. Vannutelli, A., Perreault, J.-P. & Ouangraoua, A. G-quadruplex occurrence and conservation: more than just a question of guanine–cytosine content. *NAR Genomics Bioinforma.* **4**, lqac010 (2022).

56. Singh, A., Kukreti, R., Saso, L. & Kukreti, S. Oxidative stress: role and response of short guanine tracts at genomic locations. *Int. J. Mol. Sci.* **20**, 4258 (2019).

57. Gros, J. et al. Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.* **35**, 3064–3075 (2007).

58. Bielskutė, S., Plavec, J. & Podbevšek, P. Oxidative lesions modulate G-quadruplex stability and structure in the human BCL2 promoter. *Nucleic Acids Res.* **49**, 2346–2356 (2021).

59. Hognon, C., Gebus, A., Barone, G. & Monari, A. Human DNA telomeres in presence of oxidative lesions: the crucial role of electrostatic interactions on the stability of guanine quadruplexes. *Antioxidants* **8**, 337 (2019).

60. Cumming, B. M. et al. The physiology and genetics of oxidative stress in Mycobacteria. *Microbiol. Spectr.* doi:10.1128/microbiolspec.mgm2-0019–2013 (2014)

61. Spiegel, J. et al. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **22**, 117 (2021).

62. Chen, Y. et al. An upstream G-quadruplex DNA structure can stimulate gene transcription. *ACS Chem. Biol.* **19**, 736–742 (2024).

63. Li, Q., Wang, Y., Hu, X., Zhao, Y. & Li, N. Genome-wide mapping reveals conservation of promoter DNA methylation following chicken domestication. *Sci. Rep.* **5**, 8748 (2015).

64. Butovskaya, E., Heddi, B., Bakalar, B., Richter, S. N. & Phan, A. T. Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc.* **140**, 13654–13662 (2018).

65. Nadai, M. et al. A catalytic and selective scissoring molecular tool for quadruplex nucleic acids. *J. Am. Chem. Soc.* **140**, 14528–14532 (2018).

66. Mazzini, S. et al. Quadruplex-duplex junction in LTR-III: a molecular insight into the complexes with BMH-21, namitecan and doxorubicin. *PLOS ONE* **19**, e0306239 (2024).

67. Díaz-Casado, L. et al. De Novo design of selective quadruplex–duplex junction ligands and structural characterisation of their binding mode: targeting the G4 Hot-Spot. *Chem. Eur. J.* **27**, 6204–6212 (2021).

68. Ryazantsev, D. Y. et al. Probing GFP chromophore analogs as anti-HIV agents targeting LTR-III G-Quadruplex. *Biomolecules* **11**, 1409 (2021).

69. Berner, A. et al. G4-ligand-conjugated oligonucleotides mediate selective binding and stabilization of individual G4 DNA structures. *J. Am. Chem. Soc.* **146**, 6926–6935 (2024).

70. Tassinari, M. et al. Selective targeting of mutually exclusive DNA G-quadruplexes: HIV-1 LTR as paradigmatic model. *Nucleic Acids Res.* **48**, 4627–4642 (2020).

71. Maurizio, I. et al. Production of the anti-G-quadruplex antibody BG4 for efficient genome-wide analyses: from plasmid quality control to antibody validation. *Methods Enzymol.* **695**, 193–219 (2024).

72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

73. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

74. Kim, T. H. & Dekker, J. Formaldehyde cross-linking. *Cold Spring Harb. Protoc.* **2018**, pdb.prot082594 (2018).

75. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

76. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

77. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol. Biol. Clifton NJ* **1418**, 335–351 (2016).

78. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

79. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12–1-34 (2014).

80. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinforma.* **11**, 237 (2010).

81. 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).

82. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

83. 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).

84. 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).

85. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html.

86. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).

Author contributions
Conceptualization, S.N.R. and R.P.; methodology, I.M. and M.C.; investigation, I.M.; data analysis, I.M., E.R., I.Z., G.N.; writing – original draft, I.M. and E.R.; writing - review & editing, E.R., I.Z. and S.N.R.; funding acquisition, S.N.R. and R.P.; resources, S.N.R. and R.P.; supervision, S.N.R. All authors contributed to data interpretation and editing the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-62485-4.

Correspondence and requests for materials should be addressed to Roberta Provvedi or Sara N. Richter.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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