nature portfolio

Peer Review File

LysG-Driven Transcriptional Network Rewiring Underlies Lineage-Specific Phenotypes in Mycobacterium tuberculosis

Corresponding Author: Dr Amir Banaei-Esfahani

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This study provides a comprehensive multi-omics analysis of Mycobacterium tuberculosis Lineage L1 and Lineage L2 clinical strains, revealing how genetic differences translate to significant phenotypic variations. The research integrates genomic, transcriptomic, and proteomic data to elucidate lineage-specific regulatory mechanisms, particularly in stress response pathways. The findings highlight how limited genetic diversity can lead to substantial functional differences through differential regulation of key transcriptional networks.

This work is interesting and contributes to the development of new bioinformatics tools to understand how certain genetic variations are reflected in molecular differences that underlie phenotypic changes such as increased or decreased virulence of pathogenic bacterial strains. However, there are some areas where greater clarity or additional details could improve the overall understanding of the results.

Comments:

1) The L1 strain 0072 exhibits greater variability compared to the others based on Figures S2B and S2C. The authors should consider discussing this variability in more detail and evaluate whether it could affect the accuracy of the results obtained for Lineage L1.

2) In the legend of Figure 2, the strains represented by the different symbols should be specified. As discussed above, it would be interesting to see if the Lineage 1 strain represented by the light blue cross symbol, which deviates from the other two strains, corresponds to the strain named 0072. Additionally, line 898 has (D) repeated twice - the second occurrence should be corrected to (E) and the following (E) to (F).

3) At line 168, it should be specified that the modest correlation refers to Lineage L1.

4) It is unclear where the information on T cell antigens from lines 181 to 185 is derived from. Table S1? However, this table is cited for the first time in the following paragraph. It would be helpful to include a table with the gene names and respective values used for the graphs in Figure S3.

5) From lines 190 to 193, it is stated that most changes occur in non-essential genes. Adding a column to the table specifying which genes are essential and non-essential would clarify this statement.

6) At line 210, the work by Yossef Av-Gay mentioned should be cited.

7) Clarify whether the numerical values in Table S1C refer to proteins or mRNAs.

8) There is an empty space between parentheses at line 310.

9) The fold change values reported at lines 360-361 for SigB do not match the values listed for SigB in Table S3. Furthermore, Table S3 highlights RsdA and UsfX as proteins that are significantly differentially expressed between Lineages L1 and L2, in addition to SigB. However, the text does not provide any discussion about the potential roles or

significance of the differential expression of RsdA and UsfX between the two lineages. Additionally, at line 341-342 they say that ”only the protein RsfB (anti-anti sigma factor) showed significant differential expression between the two lineage among all the sigma factors …”, however RsfB shows no significant p-value in table S3. They should contextualize the differential expression of SigB, RsdA, and UsfX, and their potential connection to the lineage-specific differential regulation mediated by the DosR regulon, which is a major focus of the study. The current section on sigma factors appears somewhat disconnected and lacks integration of these related findings

10) The section from line 382 to 387 is very confusing. The source of this information is unclear, and the experimental conditions of hypoxia/normoxia for the Transcriptional Regulator Induced Phenotype (TRIP) assay are not described in the materials and methods section. The authors refer to reference 69, where the TRIP screen was conducted using isoniazid as a stress condition. Therefore, it is unclear how the authors reached the conclusion that based on a TRIP assay under hypoxic and normoxic conditions, only 5 transcription factors, including Rv1985c, showed a significant phenotype in terms of altered growth rate. Without this crucial information, it is unclear where the authors obtained the evidence that led them to select Rv1985c for further time-course experiments under hypoxia and reaeration conditions. Additionally, there is no information about the plasmid vector where Rv1985c was cloned in front of an indubible promoter. This entire paragraph needs to be rewritten in a clearer and more transparent manner.

11) It is not clear from which data the authors conclude that Rv1985c regulates approximately 80 other transcription factors upstream (lines 425-426).

12) The part from lines 581 to 587 should be moved to the results or discussion section.

13) In Figure 6, it should be specified what the numerical values on the y-axis correspond to.

14) In the legends of Figure 7 and Figure S9, the meaning of OE and EP should be explained.

15) Some bibliographic references are incomplete, missing page numbers.

Reviewer #2

(Remarks to the Author)
This study from Banaei-Esfahani et al reported correlations between transcript and protein across three L1 and three L2 M. tuberculosis complex representative clinical isolates, indicating lineage-specific post-translational regulation. Rv1985c has been identified as main master transcription regulator responsible for differential target gene expression in L2 compared to L1 strains.

Minor comments
- Methods, Mtb strains and bacterial culture: please provide rationale and/or references related to the use of modified 7H9 medium.
- Methods, BWA, samtools, Python and R: please consider including appropriate references. I would also suggest considering including any specific settings used within each tool/script/module (e.g. default vs custom).
- Methods, H37Rv refence genome: please specify the accession number and version considered.
- Methods: the Authors mentioned the use of three biological replicates. Please consider to specify the number of technical replicates for each method for improved clarity.
- Methods/Results, EGRIN: given the insufficiency of the EGRIN model, please consider either to remove or reduce the Results section on its output. Also, given the relatively poor overlap between EGRIN and GenSTrans, it does not seem to be meant as a further supporting evidence.

Major comments
- Methods: because bacterial gene expression can vary significantly with growth rate, growth phase, and inoculum history, the comparability of RNA-seq data depends on ensuring that cultures are physiologically matched. I therefore recommend clarifying in “Mtb strains and bacterial culture” that the inoculum source and growth phase at sampling were standardized across strains.
- Results: there are a number of sentences that would probably fit better under the discussion section. E.g. lines 169-181, but also those referred to previous studies and literature. Alternatively, cross-check whether the Journal allows for a joint results and discussion section and expand discussion points.
- Results, lines 201-211: According to the Chao’s study mentioned, under normal growth (aerobic, early log phase), the basal expression of the DosR regulon was unaffected by the pknH deletion. Under NO induction, ΔpknH showed reduced induction of DosR regulon relative to WT, suggesting PknH is required for full induction under stress. It is not fully clear whether the Authors were not able to replicate the results during basal or NO induction conditions.
- Results, “The number of SNPs is linked to the magnitude of transcript and protein expression differences between strains”: based on the data provided, it seems that only SNPs have been considered; however, I would assume that indels are also part of the genetic diversity of the strains under study. I would recommend to explain somewhere how indels have been considered and, since excluded from this analysis, this should also somehow covered under a “limitations of the study” paragraph.
- Results, “The number of SNPs is linked to the magnitude of transcript and protein expression differences between strains”: while I understand that elucidating the precise one-to-one relationship and underlying mechanism of each causative SNP is beyond the scope of this analysis, given the complexity of possible regulatory interactions, I believe that a more detailed

description of the relationship between SNPs and transcription factors (TFs) would be valuable. Since the authors focus on differentially expressed TFs, it would be informative to assess whether SNP distributions across TF genes differ significantly between lineages. Regardless of whether such differences are observed, this analysis could strengthen the link between genetic variation and transcriptional regulation. Similarly, this should somehow be applied to sigma and anti-sigma factors later used to support the role of the TFs selected in “Hierarchical organization for interactions between TFs and sigma factors”. Additionally, as other functional categories also include a high number of differentially expressed genes, the rationale for prioritizing TFs in the downstream analysis should be clarified.
- Results: looking at Table S1, it appears that strain 0153 is primarily driving the large expression differences observed for Rv1985c between L1 and L2. Given that some L1 strains from the same repository have been reported to harbor a deletion of Rv1985c (doi.org/10.1038/s41598-018-22237-5), it would be important to verify the presence or absence of this gene across all isolates included in the analysis. Notably, strain 0153 also displays higher diversity on the PCA for protein data, as well as in the intra-lineage mRNA expression profiles. Interestingly, the magnitude of its differential expression (|log2FC| approx. 10) is comparable both within its own genotype and in inter-genotype comparisons, suggesting that the expression pattern is strain-specific rather than genotype-driven. This reinforces the need to confirm that gene Rv1985c is present and intact in all strains before interpreting genotype-level effects.
- The possibility that gene Rv1985c is deleted in some L1 isolates should be addressed in the discussion. While the Authors conclude that it functions as a validated master regulator distinguishing lineages L2 and L1, the presence of a deletion in some L1 isolates challenges this interpretation and suggests that the observed differential expression may not uniformly reflect a lineage-level effect.
- Please consider reporting BCCM/ITM IDs for all strains to avoid misunderstanding.
- Discussion: please consider including a “limitations of the study” paragraph

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
-

Point-by-point response to reviewers' comments

Manuscript title: *Rv1985c-Driven Transcriptional Network Rewiring Underlies Lineage-Specific Phenotypes in Mycobacterium tuberculosis.*

We would like to sincerely thank both reviewers for their time and thoughtful evaluation of our manuscript. We deeply appreciate the reviewers' constructive and insightful comments, which have greatly contributed to improving the clarity, depth, and overall impact of our study.

We have carefully considered and addressed all comments point by point, as detailed below. In the revised manuscript, newly added text is highlighted in green and underlined, while deleted text is indicated using ~~strikethrough~~ formatting. We believe these revisions have substantially strengthened the manuscript.

Reviewer #1 (Remarks to the Author):

This study provides a comprehensive multi-omics analysis of *Mycobacterium tuberculosis* Lineage L1 and Lineage L2 clinical strains, revealing how genetic differences translate to significant phenotypic variations. The research integrates genomic, transcriptomic, and proteomic data to elucidate lineage-specific regulatory mechanisms, particularly in stress response pathways. The findings highlight how limited genetic diversity can lead to substantial functional differences through differential regulation of key transcriptional networks. This work is interesting and contributes to the development of new bioinformatics tools to understand how certain genetic variations are reflected in molecular differences that underlie phenotypic changes such as increased or decreased virulence of pathogenic bacterial strains. However, there are some areas where greater clarity or additional details could improve the overall understanding of the results.

We would like to thank the reviewer for the positive and encouraging evaluation of our work. We are pleased that the study is considered comprehensive and relevant, and we greatly appreciate the thoughtful and constructive comments provided. These insights have been invaluable in improving the clarity, completeness, and overall quality of the manuscript. We have carefully addressed all points raised in detail below and implemented the corresponding revisions in the manuscript to enhance its presentation and scientific precision.

Comments:

1) The L1 strain 0072 exhibits greater variability compared to the others based on Figures S2B and S2C. The authors should consider discussing this variability in more detail and evaluate whether it could affect the accuracy of the results obtained for Lineage L1.

We thank the reviewer for this comment. Below, we provide the exact coefficients of variation (CVs) across technical triplicates for each strain at both the protein and mRNA levels.

Protein-level CVs (median):

N0052: 0.0959; N0072: 0.1830; N0145: 0.1164; N0153: 0.1217; N0155: 0.1138; N0157: 0.1340; Overall CV: 0.1243

mRNA-level CVs (median):

N0052: 0.1486; N0072: 0.1224; N0145: 0.1183; N0153: 0.1278; N0155: 0.1247; N0157: 0.1150; Overall CV: 0.1252

As shown, at the mRNA level (Fig. S2C), the median CV for strain N0072 is slightly below the overall median CV, indicating consistent transcriptional measurements across replicates. Since the transcriptomic data form the foundation of our downstream transcriptional network analyses, this observation supports the robustness of our conclusions. At the protein level (Fig. S2B), the median CV of N0072 is somewhat higher than the overall median. To ensure that this variability did not bias the results, we processed the proteomic data using MSstats, which explicitly models variance both between strains and across technical replicates using a Linear Mixed Model (LMM). Therefore, this higher variability was already accounted for in the statistical analysis. We thank the reviewer for highlighting this point and have added a clarifying sentence to the Methods section to explicitly indicate that variance across technical replicates was incorporated during MSstats-based processing (**Lines 655-657**).

2) In the legend of Figure 2, the strains represented by the different symbols should be specified. As discussed above, it would be interesting to see if the Lineage 1 strain represented by the light blue cross symbol, which deviates from the other two strains, corresponds to the strain named 0072. Additionally, line 898 has (D) repeated twice - the second occurrence should be corrected to (E) and the following (E) to (F).

We appreciate the reviewer's attention to this detail. The symbols corresponding to each strain have been indicated between Fig. 2E and Fig. 2F; this information was included in the original version. We have also corrected the panel labeling in the figure legend (the second **(D)** changed to **(E)**, and the subsequent **(E)** changed to **(F)**) in the revised manuscript (**Lines 1001-1003**).

3) At line 168, it should be specified that the modest correlation refers to Lineage L1.

The modest correlation mentioned in the text refers to the genome-wide mRNA-protein correlation observed across both L1 and L2 strains (modest in comparison with *E.coli* for instance), rather than to L1 alone. For reference, the strain-specific Spearman correlation coefficients between mRNA and protein abundances are as follows:

N0052: 0.453; N0072: 0.454; N0145: 0.473; N0153: 0.448; N0155: 0.503; N0157: 0.450; Overall average: 0.463.

We have revised the corresponding sentence in the manuscript to clarify this point and avoid possible misinterpretation (**Line 172**).

4) It is unclear where the information on T cell antigens from lines 181 to 185 is derived from. Table S1? However, this table is cited for the first time in the following paragraph. It would be helpful to include a table with the gene names and respective values used for the graphs in Figure S3.

We thank the reviewer for this comment. The data underlying all volcano plots (Figs. 3, S3, and S4) are derived from **Table S1A and B**, which contains the results of the differential expression analyses at both the mRNA and protein levels. Specifically, Table S1 includes the gene names, fold changes, P values, and adjusted P values for all relevant pairwise comparisons across transcriptomic and proteomic datasets. For greater clarity, we have now cited Table S1 earlier in the text, at the point where the data were first used, ensuring that readers can directly trace the source of the information (**Line 186**).

5) From lines 190 to 193, it is stated that most changes occur in non-essential genes. Adding a column to the table specifying which genes are essential and non-essential would clarify this statement.

We thank the reviewer for this observation. Throughout the manuscript, we have integrated multiple sources of gene annotations. Including all these annotations in Table S1 could potentially create confusion for readers. The classification of essential genes was based on the work by Christopher Sassetti and colleagues (PMID: 21980284), who systematically identified essential genes in *M. tuberculosis*. We have now cited this reference in the main text to clarify the source of this information so that readers can readily access the complete list of essential genes if desired (**Line 196**).

6) At line 210, the work by Yossef Av-Gay mentioned should be cited.

We thank the reviewer for this comment. The corresponding paper by Yossef Av-Gay and colleagues (PMID: 20630871) was already cited earlier in the manuscript (**Line 207**). However, we have now cited it again in the concluding sentence of the relevant section to ensure proper clarity (**Line 214**).

7) Clarify whether the numerical values in Table S1C refer to proteins or mRNAs.

The numerical values in Table S1C refer to protein-level measurements. As clarified in the manuscript, we profiled the proteome of the PknH knockout, PknH overexpression (induced), and wild type *M. tuberculosis* H37Rv strains in technical triplicates. The relevant sentence in the text reads:

"*To test this hypothesis, we profiled the* ***proteome*** *of a PknH knockout and overexpressing strain generated from the laboratory strain H37Rv, along with the wild type strain in biological triplicates, using the same method as described above.*" (**Lines 211-213**)

8) There is an empty space between parentheses at line 310.

We thank the reviewer for this comment. The unnecessary space between the parentheses has been removed in the revised manuscript (**Line 322**).

9) The fold change values reported at lines 360-361 for SigB do not match the values listed for SigB in Table S3. Furthermore, Table S3 highlights RsdA and UsfX as proteins that are significantly differentially expressed between Lineages L1 and L2, in addition to SigB. However, the text does not provide any discussion about the potential roles or significance of the differential expression of RsdA and UsfX between the two lineages. Additionally, at line 341-342 they say that "only the protein RsfB (anti-anti sigma factor) showed significant differential expression between the two lineage among all the sigma factors …", however RsfB shows no significant p-value in table S3. They should contextualize the differential expression of SigB, RsdA, and UsfX, and their potential connection to the lineage-specific differential regulation mediated by the DosR regulon, which is a major focus of the study. The current section on sigma factors appears somewhat disconnected and lacks integration of these related findings.

We are grateful to the reviewer for several points that were raised here, and we address them one by one below.

First, the apparent discrepancy in fold change values arises from the difference in how values were reported. All fold changes in the supplementary tables (e.g. Table S3) were expressed as $\log_2$ fold changes, whereas those in the main text were presented as actual fold changes (i.e., 2 to the power of the $\log_2$ values reported in Table S3).

Second, the sigma factor analysis in the manuscript primarily focused on those sigma factors that directly interacted with the four transcription factors - Rv3133c (DosR), Rv1985c, Rv3597c (Lsr2), and Rv0691c - identified as master regulators distinguishing L1 and L2 strains (Fig. S8). RsdA and UsfX were not discussed in detail because they did not directly interact with these four transcription factors. They appear in Table S3 highlighted because they also exhibited differential responses to nitric oxide (NO) between L1 and L2 strains, which was not the main focus of that section. We have now clarified this point in the revised manuscript (**Lines 375-377**).

Third, it is important to note that Table S3B presents differential expression analyses comparing NO-perturbed strains (both L1 and L2) to their respective unperturbed controls, whereas Table S1 reports differential expression between L1 and L2 under control (non-perturbed) conditions. RsfB is significantly differentially expressed between L1 and L2 strains under control conditions (as reported in Table S1), which is why Table S1 is cited in that context (**Line 357**).

Finally, among the sigma factors, SigB is the only candidate that shows a lineage-specific differential response to NO exposure (Table S3) while remaining unchanged between L1 and L2 under control conditions (Table S1). This observation suggests that SigB most likely plays a key regulatory role in mediating the lineage-specific activation of the DosR regulon in response to NO, as summarized in Fig. 6A.

10) The section from line 382 to 387 is very confusing. The source of this information is unclear, and the experimental conditions of hypoxia/normoxia for the Transcriptional

Regulator Induced Phenotype (TRIP) assay are not described in the materials and methods section. The authors refer to reference 69, where the TRIP screen was conducted using isoniazid as a stress condition. Therefore, it is unclear how the authors reached the conclusion that based on a TRIP assay under hypoxic and normoxic conditions, only 5 transcription factors, including Rv1985c, showed a significant phenotype in terms of altered growth rate. Without this crucial information, it is unclear where the authors obtained the evidence that led them to select Rv1985c for further time-course experiments under hypoxia and reaeration conditions. Additionally, there is no information about the plasmid vector where Rv1985c was cloned in front of an indubible promoter. This entire paragraph needs to be rewritten in a clearer and more transparent manner.

We thank the reviewer for this detailed and constructive comment. Rv1985c was selected as a candidate master regulator for three independent reasons, as outlined in the manuscript:

"*However, we consider it a master regulator differentiating the two lineages for three reasons: first, it was identified by GenSTrans; second, its affected subnetwork, comprising 58 genes, was the largest (Fig 5a); and third, it is heavily involved in regulating DosR regulon genes (Fig 5e).*" (**Lines 395-398**)

We additionally referred to results from a Transcriptional Regulator Induced Phenotype (TRIP) screen performed to evaluate the impact of transcription factor induction on growth under hypoxic and normoxic conditions. This assay was used solely to demonstrate that Rv1985c, among a small subset of transcription factors (~5 TFs), significantly affected growth under these contrasting conditions, underscoring its biological relevance. The complete dataset and extended analyses from the TRIP screen fall beyond the scope of the present manuscript. Nevertheless, all data related to Rv1985c (Fig. 7A) are included here, encompassing both its effect on growth rate (Fig. 7B) and the time-course RNA-seq experiment (Fig. 7C-E), which validates the transcriptional network analysis results for this regulator.

We thank the reviewer for noting that the description of the time-course experiment under hypoxia and reaeration was missing from the Methods section. We have now added a dedicated subsection titled "Time-course experiment under hypoxia and reaeration" (**Lines 687-728**), which details the experimental setup, conditions, and the inducible expression construct used for Rv1985c. Of note, the selection of Rv1985c was based on the three reasons particularly the result of our transcriptional network analysis (first reason) as discussed clearly in the manuscript.

These revisions clarify both the rationale for selecting Rv1985c and the experimental procedures used to validate its role.

11) It is not clear from which data the authors conclude that Rv1985c regulates approximately 80 other transcription factors upstream (lines 425-426).

We thank the reviewer for highlighting this point. This conclusion was derived from the enrichment analysis performed on the results of the Rv1985c time-course experiment. The outcome of this analysis is visualized in Fig. 7D, which highlights the set of transcription factors whose expression was either directly or indirectly upregulated following the Rv1985c induction. Specifically, under reaeration conditions, approximately 80 transcription factors were found to be significantly upregulated between Day 9 (D9) and Day 12 (D12).

To enhance clarity, we have now explicitly cited Fig. 7D immediately following the sentence referring to the regulation of ~80 transcription factors in the revised manuscript. (**Line 442**)

12) The part from lines 581 to 587 should be moved to the results or discussion section.

We have already discussed the interaction between sigma factors and transcription factor networks in detail in the Results section (**Lines 341-347**). However, we agree that the lines mentioned should not appear in the Methods section. Accordingly, we have revised that part, ensuring that the methodological description remains concise (**Lines 611-617**).

13) In Figure 6, it should be specified what the numerical values on the y-axis correspond to.

We thank the reviewer for this helpful comment. We have now revised Figure 6B to clarify that the numerical values on the y-axis represent the number of hours each strain experienced growth arrest following nitric oxide (NO) exposure. The corresponding figure legend has also been updated for clarity (**Line 1044-1045**).

14) In the legends of Figure 7 and Figure S9, the meaning of OE and EP should be explained.

We thank the reviewer for drawing our attention to this issue. OE and EP refer to the OverExpressor and Empty Plasmid strains, respectively, corresponding to the inducible and uninduced (control) Rv1985c strains. We have revised the legends of Figure 7 and Figure S9 to explicitly define these abbreviations.

15) Some bibliographic references are incomplete, missing page numbers.
We thank the reviewer for this comment. We have carefully reviewed all bibliographic references and ensured that all entries now include complete information, including page numbers and other missing details.

Reviewer #2 (Remarks to the Author):

This study from Banaei-Esfahani et al reported correlations between transcript and protein across three L1 and three L2 M. tuberculosis complex representative clinical isolates, indicating lineage-specific post-translational regulation. Rv1985c has been identified as main master transcription regulator responsible for differential target gene expression in L2 compared to L1 strains.

We sincerely thank the reviewer for their thoughtful and constructive evaluation of our work. We greatly appreciate the reviewer's concise summary of our main findings. The specific comments raised are highly relevant and insightful, and we believe that addressing them has improved the clarity, precision, and overall quality of the manuscript. We have carefully considered each point and provided detailed responses below, along with the corresponding revisions in the manuscript.

Minor comments

1) Methods, Mtb strains and bacterial culture: please provide rationale and/or references related to the use of modified 7H9 medium.

The modified 7H9 medium was supplemented with 0.5% (w/v) pyruvate, 0.05% (v/v) tyloxapol, 0.2% (w/v) glucose, 0.5% bovine serum albumin (Fraction V, Roche), and 14.5 mM NaCl. We employed this modified medium to ensure optimal and consistent growth of genetically diverse *Mycobacterium tuberculosis complex* (MTBC) strains, including *M. tuberculosis*, *M. bovis*, and *M. africanum*.

This formulation also provides cross-platform consistency across our multi-omics studies. For instance, tyloxapol was used instead of Tween 80, as the latter is less compatible with proteomics workflows. Using the same culture conditions across studies facilitates direct comparison and integration of multi-omics datasets. Pyruvate was included as an alternative carbon source to support the growth of strains inhibited by glycerol. Tyloxapol prevents cell clumping while avoiding metabolically active or variable surfactants that could interfere with downstream proteomic and metabolomic analyses. Glucose supplied an additional readily metabolizable carbon source, while bovine serum albumin and NaCl maintained nutrient stability and osmotic balance.

2) Methods, BWA, samtools, Python and R: please consider including appropriate references. I would also suggest considering including any specific settings used within each tool/script/module (e.g. default vs custom).

We thank the reviewer for this helpful comment. We have now included the appropriate references for all tools mentioned in the Methods section.

3) Methods, H37Rv refence genome: please specify the accession number and version considered.

We thank the reviewer for this comment. The reference genome used for M. tuberculosis H37Rv was NC_000962.3, and we have now included the corresponding accession number and version in the Methods section (**Lines 577 and 723**).

4) Methods: the Authors mentioned the use of three biological replicates. Please consider to specify the number of technical replicates for each method for improved clarity.

We appreciate the reviewer's careful assessment of this aspect. All measurements were performed in technical triplicates, including all proteomics and transcriptomics experiments, as well as the Rv1985c time-course experiment. The only exceptions are: (i) the Empty Plasmid (control) strain at Days 10 and 11, which was conducted in two technical replicates, and (ii) the NO proteomics measurements (DosR regulon and relevant sigma factors), which were performed in technical duplicates for each strain and condition; the resulting comparison involved 6 vs. 6 samples (3 strains × 2 replicates) for NO-treated versus control conditions. We have revised all descriptions of technical and biological replicates mentioned in the manuscript. Technical replicates refer to independent cultures and measurements of the same strain under a given condition, while biological replicates correspond to different MTBC strains (e.g. N0052, N0145). We have now added these details in the corresponding Methods sections to improve clarity (Several instances such as **Lines 559-562**).

5) Methods/Results, EGRIN: given the insufficiency of the EGRIN model, please consider either to remove or reduce the Results section on its output. Also, given the relatively poor overlap between EGRIN and GenSTrans, it does not seem to be meant as a further supporting evidence.

The EGRIN model represents the former transcriptional regulatory network of MTBC (PMID: 27573104), and our intention was to provide a comparative overview between EGRIN and our GenSTrans model. As shown in Fig. 5A, several transcription factors identified by both models, including DosR (Rv3133c), are highlighted (green), demonstrating decent concordance between the two approaches. Of note, **TFs highlighted in blue were not identifiable through the EGRIN model (Lines 292-295)**, which underscores one of the major ways our modeling strategy provides additional insight and value to the TB research community. We believe that including this comparison provides readers with a broader perspective on transcriptional network analyses available for MTBC.

Major comments

6) Methods: because bacterial gene expression can vary significantly with growth rate, growth phase, and inoculum history, the comparability of RNA-seq data depends on ensuring that cultures are physiologically matched. I therefore recommend clarifying in "Mtb strains and bacterial culture" that the inoculum source and growth phase at sampling were standardized across strains.

We appreciate the reviewer's attention to this methodological detail. Our experimental design explicitly accounted for potential variability due to growth rate, growth phase, and inoculum history. As mentioned in the Methods section under "Transcriptomic profiling", all strains were cultivated under identical conditions and harvested at the same physiological state (mid-log phase; OD_{600} = 0.5 ± 0.1). The inoculum preparation and culture handling were standardized across all strains to ensure comparability of RNA-seq and proteomics measurements. This is described in the main text and Methods section:

Main text:

"*We then grew the six strains in vitro to mid-log phase and profiled their transcriptomes and proteomes in biological triplicates using RNA sequencing and DIA-MS, respectively (Fig. 1b, Fig. S1).*"

Methods part:

"*We transferred a 40 ml aliquot of bacterial culture in mid-log phase (OD_{600} = 0.5 ± 0.1) into a 50 ml Falcon conical tube containing 10 ml ice.*"

Furthermore, the OD_{600} measurements recorded at harvest are summarized in **Fig. S1**, confirming that all samples were collected at comparable physiological states. Nevertheless, we have now reiterated this point under 'Mtb strains and bacterial culture' for full clarity (**Lines 556-559**).

7) Results: there are a number of sentences that would probably fit better under the discussion section. E.g. lines 169-181, but also those referred to previous studies and literature. Alternatively, cross-check whether the Journal allows for a joint results and discussion section and expand discussion points.

We thank the reviewer for this suggestion. We agree that several of the sentences mentioned contain discussion-like elements. However, this particular section presents a key result, namely, the overall mRNA-protein correlation (r = 0.46) across MTBC strains, which is best interpreted when placed in context with similar analyses in other bacteria (e.g., *E. coli*, r ≈ 0.8). Including this contextual information directly after the result helps convey that the relatively modest correlation observed in MTBC likely reflects pronounced post-translational regulation rather than post-transcriptional effects, consistent with the high correlation (r = 0.93) observed between mRNA and ribosome-protected fragment (RPF) levels in MTBC.

Moving this information entirely to the Discussion section might mislead readers, as they would only realize at the very end of the manuscript that the modest correlation is not driven by post-transcriptional events. Therefore, we prefer to retain this discussion-like explanation immediately following the corresponding result to ensure clarity and avoid potential confusion, of course, provided this is acceptable within the journal's formatting guidelines.

8) Results, lines 201-211: According to the Chao's study mentioned, under normal growth (aerobic, early log phase), the basal expression of the DosR regulon was unaffected by the pknH deletion. Under NO induction, ΔpknH showed reduced induction of DosR regulon relative to WT, suggesting PknH is required for full induction under stress. It is not fully clear whether the Authors were not able to replicate the results during basal or NO induction conditions.

We thank the reviewer for this comment. The description in our manuscript is derived from the study by Chao et al. (PMID: 20630871). As shown in Table 1 (both untreated and NO-treated ratios are summarized) of that paper, the DosR untreated ratio (ΔpknH / WT) was 1.69.

The authors concluded that 13 DosR-regulated proteins were identified and all were differentially expressed even under untreated conditions. Specifically, as stated in their paper:

"*Further examination of the iTRAQ data revealed that of the 48 genes commonly regulated by DosR in response to NO, hypoxia, and the Wayne model of dormancy, 13 gene products were identified by iTRAQ, all of which had higher protein levels in the ΔpknH mutant after 48-h standing conditions (Table 1). With the addition of $NaNO_2$, these 13 DosR-regulated proteins were induced to similar levels in WT and ΔpknH (Table 1).*"

In contrast, when we performed the same analysis under normal (non-NO) conditions, we did not observe differential expression of DosR regulon members between the ΔpknH and wild-type strains, and thus could not replicate the elevated basal levels reported by Chao et al.. For this reason, we extended our analyses to nitric oxide perturbed conditions, where lineage-specific differences in DosR regulon responses became evident (please also see the sigma factor part particularly SigB and the relevant network in Fig. S8).

9) Results, "The number of SNPs is linked to the magnitude of transcript and protein expression differences between strains": based on the data provided, it seems that only SNPs have been considered; however, I would assume that indels are also part of the genetic diversity of the strains under study. I would recommend to explain somewhere how indels have been considered and, since excluded from this analysis, this should also somehow covered under a "limitations of the study" paragraph.

We thank the reviewer for highlighting this point. Indeed, insertions and deletions (indels) represent an important component of genetic diversity within MTBC strains. However, indels were not included in the present analysis. The rationale for focusing exclusively on SNPs is that assessing the relative contributions of SNPs and indels to transcriptional and proteomic variation would require analysis across a substantially larger cohort of MTBC strains, likely hundreds, to achieve sufficient statistical power. Such an investigation clearly extends beyond the scope of the current study, as also mentioned in the following comment by the reviewer.

In this work, we focused on SNPs because they represent the most common and well-characterized form of genetic variation and allow a conceptual framework for understanding how genetic diversity influences transcript and protein abundance in a one-to-one manner, analogous to expression quantitative trait loci (eQTL) and protein quantitative trait loci (pQTL) analyses in other organisms. This analysis was primarily intended to provide the TB research community with a conceptual foundation for designing future large population-scale MTBC studies.

In addition to indels, differences in gene content could also contribute to functional diversity, and emerging long-read sequencing technologies are increasingly enabling such analyses. We have now added a statement in the Discussion section acknowledging that indels and gene content variations were not considered here and that their potential impact remains to be systematically explored in future large-scale MTBC population studies (**Lines 475-476**).

10) Results, “The number of SNPs is linked to the magnitude of transcript and protein expression differences between strains”: while I understand that elucidating the precise one-to-one relationship and underlying mechanism of each causative SNP is beyond the scope of this analysis, given the complexity of possible regulatory interactions, I believe that a more detailed description of the relationship between SNPs and transcription factors (TFs) would be valuable. Since the authors focus on differentially expressed TFs, it would be informative to assess whether SNP distributions across TF genes differ significantly between lineages. Regardless of whether such differences are observed, this analysis could strengthen the link between genetic variation and transcriptional regulation. Similarly, this should somehow be applied to sigma and anti-sigma factors later used to support the role of the TFs selected in “Hierarchical organization for interactions between TFs and sigma factors”. Additionally, as other functional categories also include a high number of differentially expressed genes, the rationale for prioritizing TFs in the downstream analysis should be clarified.

We thank the reviewer for this multifaceted comment.

Firstly, single-nucleotide polymorphisms (SNPs) can influence gene expression either through cis-regulatory effects (when the SNP resides within or near the gene it regulates) or trans-regulatory effects (when it impacts other genes via intermediate factors, such as transcription factors). However, a gene may harbor multiple SNPs, and while some variants may exert measurable effects, their cumulative impact on gene expression could be minimal, as the effects of individual SNPs may counteract one another. Therefore, the distribution of SNPs across functional categories, including transcription factors (TFs) and sigma factors (SFs), does not by itself provide clear biological insight without large-scale statistical power, typically achieved through very large population-level studies encompassing hundreds of strains. Such one-to-one association studies between genetic variants and molecular phenotypes clearly go beyond the scope of the present work, as the reviewer also noted.

Nevertheless, we examined our comparative genomic data between L1 and L2 strains, focusing on SNPs that are conserved within each lineage but differ between them (i.e., lineage-specific SNPs). We then analyzed the distribution of these SNPs across functional categories. Interestingly, both TFs and SFs were among the categories with relatively few nonsynonymous SNPs compared to other functional groups (**Table 1**). However, for reasons of thematic coherence, we feel that including this analysis may divert attention from the central objectives of the manuscript. We therefore prefer not to incorporate it in the main text but would be happy to provide the data upon request.

Category	Gene	Syn.	Nonsyn.	Syn. ratio	Nonsyn. ratio
Virulence, detoxification, adaptation	239	8	14	0.0335	0.0585
Sigma factor	13		1		0.0769
Transcription Factor	214	16	20	0.0748	0.0934

Conserved hypothetical proteins	1042	56	118	0.0537	0.1132
Regulatory proteins	198	19	24	0.0960	0.1212
Information pathways	242	27	31	0.1116	0.1280
Intermediary metabolism and respiration	936	79	134	0.0844	0.1431
Cell wall and cell processes	772	73	118	0.0945	0.1528
Essential Genes - Growth - In vitro	617	62	97	0.1005	0.1572
mycolic acid related protein	62	4	12	0.0645	0.1935
Lipid metabolism	272	29	57	0.1066	0.2095

Table 1. Distribution of lineage-specific SNPs across functional categories. The table summarizes the number of lineage-specific SNPs identified between MTBC lineages 1 and 2 across different functional gene categories. Ratios represent the number of synonymous (Syn.) and nonsynonymous (Nonsyn.) SNPs normalized by the number of genes within each category.

Secondly, it is important to note that these two lineages represent different levels of genetic diversity, with L1 exhibiting greater genomic variation than L2. To provide broader evolutionary context, we extracted pN/pS values from the recently published paper by Sarah M. Fortune and colleagues (PMID: 41033311), which includes ~55,000 clinical isolates spanning lineages 1–4. We specifically evaluated TFs and SFs within this dataset.

The ratio of nonsynonymous to synonymous polymorphisms (pN/pS) reflects the selective pressure acting on protein-coding genes, where pN/pS < 1 typically indicates purifying (negative) selection. As summarized in the violin plots below (**Fig. 1**), both TFs and SFs exhibit pN/pS distributions comparable to the genome-wide averages across all lineages, suggesting that these gene classes are under similar purifying selection rather than lineage-specific adaptive diversification. In L2, slightly higher median and mean pN/pS values were observed for both TFs and SFs, but this trend parallels the overall distribution across all genes, indicating that some genes, including TFs and SFs, may experience relatively weaker purifying selection. While these results are of evolutionary interest, we believe they are not central to the main focus of this study and could potentially be misinterpreted if overemphasized.

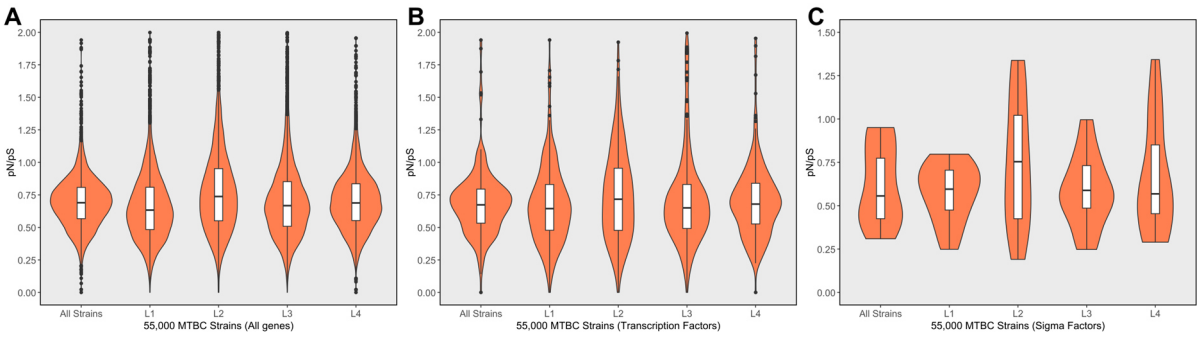

Figure 1. Distribution of pN/pS ratios across MTBC lineages and functional categories. Violin plots show the distribution of the ratio of nonsynonymous to synonymous polymorphisms (pN/pS) across

~55,000 MTBC strains. Lower pN/pS ratios indicate stronger purifying (negative) selection. Each panel represents a distinct gene subset: **(A)** all coding genes, **(B)** transcription factors, and **(C)** sigma factors. Within each panel, the first violin corresponds to all strains combined, followed by lineage-specific distributions (L1–L4).

Finally, regarding the transcriptional model, we emphasize that it incorporated **all differentially regulated genes**, not only TFs. The TFs discussed in the manuscript were identified as outputs of the model. The main objective was to assess the extent to which transcriptional variation between L1 and L2 strains could be explained by TF-mediated regulatory control. Please see "Regulated target genes" in Fig. 5A, which illustrates how many genes differentially regulated between L1 and L2 strains are regulated by the corresponding TFs identified through our transcriptional network analysis.

11) Results: looking at Table S1, it appears that strain 0153 is primarily driving the large expression differences observed for Rv1985c between L1 and L2. Given that some L1 strains from the same repository have been reported to harbor a deletion of Rv1985c (doi.org/10.1038/s41598-018-22237-5), it would be important to verify the presence or absence of this gene across all isolates included in the analysis. Notably, strain 0153 also displays higher diversity on the PCA for protein data, as well as in the intra-lineage mRNA expression profiles. Interestingly, the magnitude of its differential expression (|log2FC| approx. 10) is comparable both within its own genotype and in inter-genotype comparisons, suggesting that the expression pattern is strain-specific rather than genotype-driven. This reinforces the need to confirm that gene Rv1985c is present and intact in all strains before interpreting genotype-level effects.

12) The possibility that gene Rv1985c is deleted in some L1 isolates should be addressed in the discussion. While the Authors conclude that it functions as a validated master regulator distinguishing lineages L2 and L1, the presence of a deletion in some L1 isolates challenges this interpretation and suggests that the observed differential expression may not uniformly reflect a lineage-level effect.

We thank the reviewer for these closely related comments (11 and 12), which provided a great opportunity to further clarify and strengthen the interpretation of our findings.

We confirm that the Rv1985c gene is absent in strain N0153 (L1), consistent with previously reported deletions in certain L1 isolates. However, we maintain that interpreting Rv1985c as a lineage-associated transcriptional regulator, rather than a purely strain-specific effect, remains valid for several reasons.

Firstly, the identification of Rv1985c originated from our transcriptional network analysis, in which four transcription factors, including Rv1985c, collectively explained ~25% of the differentially expressed genes between L1 and L2 strains. This network analysis was based on experimentally validated TF-target relationships derived from ChIP seq experiments in H37Rv (PMID: 25581030). Importantly, the model does not select only TFs that are themselves

differentially expressed. Instead, it identifies TFs whose direct downstream targets exhibit significant differential regulation between lineages. Consequently, even if a TF's own expression does not change, the model can detect its regulatory footprint if its target genes collectively show significant differential expression. In this context, the size of the target set, the subset of differentially expressed targets, and the overall number of significant changes are all relevant.

For example, Rv3597c emerged as a major TF in our analysis despite not being differentially expressed between lineages (Fig. 5B). This aligns with established evidence that many bacterial TFs undergo post-translational modifications or other mechanisms that modulate activity independently of transcript abundance. Therefore, the identification of Rv1985c as a key TF is not invalidated by its absence in a single L1 strain; rather, it reflects lineage-level transcriptional differences that are consistent with its regulatory network footprint.

TFs	Degree	L2 / L1	N0052 / N0072	N0052 / N0153	N0052 / N0157	N0145 / N0072	N0145 / N0153	N0145 / N0157	N0155 / N0072	N0155 / N0153	N0155 / N0157
Rv0275c	10	1	1	1	1	1	1	1	1	1	1
Rv2017	10	1	1	1	1	1	1	1	1	1	1
Rv2506	10	1	1	1	1	1	1	1	1	1	1
Rv3183	10	1	1	1	1	1	1	1	1	1	1
Rv1985c	8	1	1	1	1	1	1	1			1
Rv2359	8	1	1	1	1	1	1	1		1	
Rv3164c	8	1	1	1	1	1	1		1	1	

Table 2. Consistency of transcription factor (TF) identification across all pairwise lineage comparisons. The table summarizes the top TFs identified by network-based transcriptional analysis across all nine possible pairwise comparisons between L1 and L2 strains. "Degree" indicates the number of times a given TF is identified among all nine pairwise comparisons, in addition to the overall L2/L1 comparison.

Secondly, to rigorously assess whether the signal for Rv1985c reflects a lineage-level or strain-specific effect, we repeated the network-based transcriptional analysis across all nine pairwise comparisons between L1 and L2 strains (3 × 3 combinations). In six of these comparisons, N0153 (the Rv1985c-absent strain) was excluded. The top TFs consistently identified across multiple comparisons are summarized in **Table 2** and **Fig. 2** (below). Notably, Rv1985c remained among the most consistently identified regulators even when N0153 was excluded, supporting that the observed signal reflects a lineage-level effect rather than being driven solely by a single strain. Interestingly, Rv1985c was not significantly identified in the N0155/N0153 comparison, despite being largely upregulated in N0155 relative to N0153 ($Log_2FC = 10.33$, adjusted P-value = 4.22×10^{-72}), which further illustrates the principles of the network-based analysis.

Thirdly, we assessed the expression of Rv1985c using TPM (transcripts per million) data, excluding N0153. Mean TPM values per strain (across technical replicates) were 37.17 (N0052), 43.41 (N0145), and 45.93 (N0155) in L2 strains, and 32.01 (N0072) and 34.32 (N0157) in L1 strains. Based on the median of these strain-level means, expression in L2 strains was ~1.4-fold higher than in L1 strains.

Finally, in our Rv1985c time-course experiment, which included normoxia (D0), hypoxia (D2-D7), and reaeration (D8–D12), we overlaid L2/L1 differentially expressed genes on those affected by Rv1985c induction (Fig. 7C and Fig. S10). On D0 and D9–D12, both L2/L1 up- and downregulated genes were significantly enriched among the corresponding up- and downregulated genes influenced by Rv1985c induction. During D2–D8, L2/L1 downregulated genes were significantly enriched among the genes downregulated in the time-course experiment upon Rv1985c induction. Of note, Rv1985c most likely acts as an upstream transcription factor, as it appears to significantly upregulate approximately 80 other TFs during D9–D12. (**Fig. 7C**). These results indicate that Rv1985c activity reflects lineage-level transcriptional effects, further supporting its role as a lineage-associated regulator.

Overall, these analyses further support the conclusion that Rv1985c represents a lineage-level transcriptional regulator, and we have discussed this point in the relevant Results and Discussion sections of the manuscript (**Lines 312-319, 498-504**).

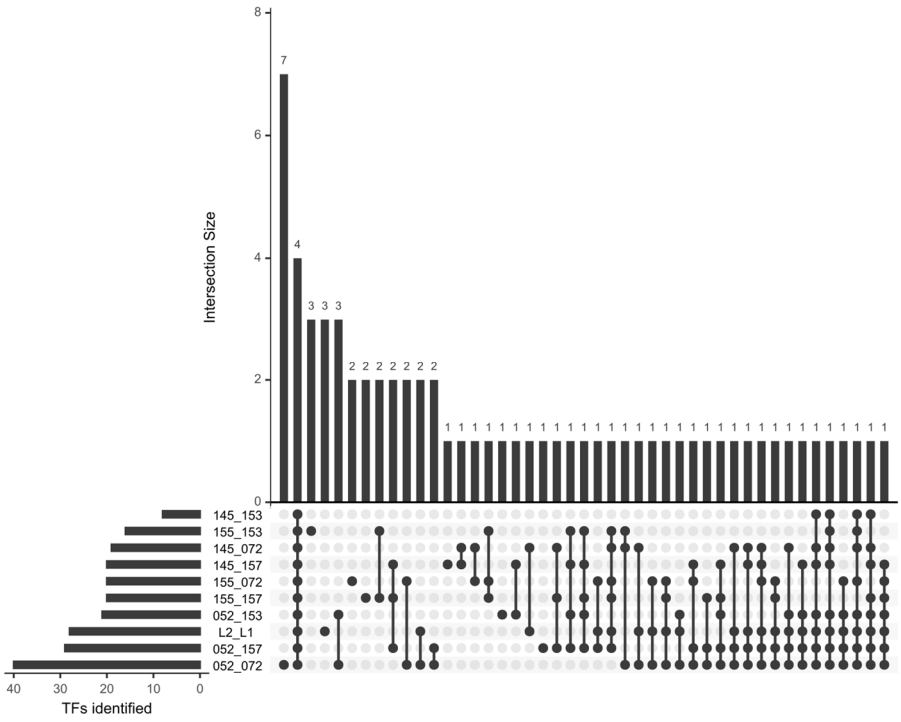

Figure 2. Overlaps of transcription factors (TFs) identified across L2 and L1 strains. UpSet plot illustrating the number of TFs shared between different strain-specific L2 vs L1 comparisons (3 L2 strains × 3 L1 strains), as well as TFs identified in the overall L2 vs L1 comparison.

13) Please consider reporting BCCM/ITM IDs for all strains to avoid misunderstanding.

We thank the reviewer for drawing our attention to this issue. We have now added the BCCM/ITM strain identifiers to the corresponding Methods section to avoid any potential misunderstanding. Of note, the full description of the used clinical strains has been discussed in following paper as already cited (PMID: 30908506) (**Lines 545-546**)

L1 strains:

- N0072: ITM-500942; CT2018-00083
- N0153: ITM-500943; CT2018-00084
- N0157: ITM-500941; CT2018-00082

L2 strains:

- N0052: ITM-601826; CT2018-03241
- N0145: ITM-500944; CT2018-00085
- N0155: ITM-500946; CT2018-00088

14) Discussion: please consider including a “limitations of the study” paragraph

We thank the reviewer for this suggestion. We have now added several sentences to highlight the limitations of the study in the Discussion section.