Title

Network analysis identifies regulators of lineage-specific phenotypes in *Mycobacterium tuberculosis*

Authors

Amir Banaei-Esfahani [1,2], Andrej Trauner [3,4], Sonia Borrell [3,4], Sebastian M. Gygli [3,4], Tige R. Rustad [5], Julia Feldmann [3,4], Ludovic C. Gillet [1], Olga T. Schubert [1,8], David R. Sherman [5], Christian Beisel [6], Sebastien Gagneux [3,4,*], Ruedi Aebersold [1,7,*,#], Ben C. Collins [1,9,*,#]

1. Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland.
2. PhD Program in Systems Biology, Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland.
3. Swiss Tropical and Public Health Institute, Basel, Switzerland.
4. University of Basel, Basel, Switzerland.
5. Department of Global Health, University of Washington, Seattle, Washington, USA.
6. Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland.
7. Faculty of Science, University of Zurich, Zurich, Switzerland.
8. Present address: Department of Human Genetics, University of California, Los Angeles, Los Angeles, USA.
9. School of Biological Sciences, Queen's University of Belfast, UK.

* Author for correspondence: sebastien.gagneux@swisstph.ch
* Author for correspondence: aebersold@imsb.biol.ethz.ch
* Author for correspondence: collins@imsb.biol.ethz.ch
# These authors jointly supervised this work.

Keywords

Mycobacterium tuberculosis, bacterial pathogen, infectious diseases, proteomics, DIA/SWATH-MS, genomics, transcriptional network, multi-omics

Highlights

- Proteomic and transcriptomic characterization of fully sequenced diverse L1 and L2 clinical isolates of Mtb.
- Post-transcriptional control mechanisms for regulatory and virulence genes are mitigated in Mtb L2.
- By applying a genome-scale transcriptional framework, DosR, Rv1985c, Lsr2 and Rv0691c are identified as master transcription factors responsible for differential target gene expression in L2 strains compared to L1.
- L1 and L2 DosR proteins respond differently to nitric oxide stress, thus determining a relevant phenotype.

Summary

The *Mycobacterium tuberculosis* (Mtb) complex comprises seven phylogenetically distinct human-adapted lineages exhibiting different geographical distribution and degrees of pathogenicity. Among these, Lineage 1 (L1) has been associated with low virulence whereas Lineage 2 (L2) has been linked to hyper-virulence, enhanced transmission and drug resistance. Here, we conducted multi-layer comparative analyses using whole genome sequencing data combined with quantitative transcriptomic and proteomic profiling of a set of L1 and L2 clinical strains, each grown under two different conditions *in vitro*. Our data revealed different degrees of correlation between transcript and protein abundances across clinical strains and functional gene categories, indicating variable levels of post-transcriptional regulation in the tested lineages. Contrasting genomic and gene expression data showed that the magnitude of the transcriptional and translational changes was proportional to the phylogenetic distance between strains, with one out of three single nucleotide polymorphisms leading to a transcriptional and/or translational change on average. We devised a new genome-scale transcriptional regulatory model and identified several master transcription factors, strongly linked to the sigma factor network, whose targets were differentially regulated between the two lineages. These differences resulted in a higher basal expression of DosR proteins and a stronger response to nitric oxide (NO) exposure in L2 compared to L1. These patterns are most likely responsible for the shorter NO-induced growth arrest in L2 observed. Given the limited genetic variation between strains, it appears that phenotypic differences in Mtb are substantially driven by differences in the regulation of biochemical networks through master transcriptional regulators.

Introduction

Adaptive evolution is mainly driven by mutations. While the phenotypic consequences of mutations can be easily mapped in the context of strong selection such as antibiotic pressure, the phenotypes emerging from more subtle evolutionary changes are more elusive.

Mycobacterium tuberculosis (Mtb) is the etiological agent of tuberculosis (TB), the major cause of human mortality due to an infectious agent (World Health Organization, 2018). Mtb is an obligate human pathogen and comprises seven phylogenetic lineages, each with a distinct geographic distribution. Together with several animal-adapted lineages (Brites et al., 2018), these make up the "Mtb complex" (Gagneux, 2018). Even though the Mtb complex harbors little genetic diversity compared to other bacteria (Achtman, 2008), Mtb clinical strains differ significantly in virulence and transmissibility (Coscolla and Gagneux, 2010, 2014). However, linking this phenotypic variation to the limited genetic diversity in the Mtb complex has been challenging (Brites and Gagneux, 2017). Among the seven human-adapted lineages of Mtb, some have received particular attention because of their clinical significance. Of note, Lineage 2 (L2) strains are geographically widely distributed and have been associated with high virulence in macrophages and various animal models and rapid disease progression (de Jong et al., 2008), increased transmission (Holt et al., 2018) and a propensity for antibiotic resistance in humans (Borrell and Gagneux, 2009). Conversely, Lineage 1 strains (L1) are geographically localized mainly around the rim of the Indian Ocean, have been linked to reduced virulence in infection models (Bhatia et al., 1961), reduced transmission potential (Holt et al., 2018) and a reduced propensity for antibiotic resistance (Coscolla and Gagneux, 2014; Phyu et al., 2009). Importantly, in geographical areas where both lineages are present, such as in Vietnam, L2 seems to be supplanting L1 as the predominant form of Mtb (Guerra-Assunção et al., 2015; Holt et al., 2018). These features point to important physiological differences between L1 and L2. Although, many studies have described phenotypic differences in Mtb clinical strains (Coscolla and Gagneux, 2010, 2014), the molecular mechanisms underlying these differences remain to be determined. To date, the majority of insights into the molecular processes have been gained from work on a few laboratory-adapted strains of Mtb (Almeida et al., 2016; Gagneux and Small, 2007; Wang et al., 2010; Wollenberg et al., 2017; Yang et al., 2015) that do not capture the entire phylogenetic diversity of clinical strains (Chiner-Oms et al., 2018; Trauner et al., 2018).

Microbiologists are increasingly applying systems biology approaches and multi-layer omics technologies to interrogate bacterial systems and their interactions with the human host (Caron et al., 2017; Nicod et al., 2017; Penn et al., 2018). These approaches largely profile various types of biomolecules, typically genome, transcriptome, proteome and, more recently, protein complexes that collectively describe the molecular makeup of specific cells and cellular states and predict interdependencies of different levels of gene expression along the axis of the central dogma (Banaei-Esfahani et al., 2017; Comas et al., 2013; Heusel et al., 2019; Rose et al., 2013). In Mtb, genomic profiling and measurements of the transcriptional response to a particular stress

have been frequently used (Cortes et al., 2017; Manson et al., 2017; Rustad et al., 2014). Yet, computational approaches that can extract mechanistic insights from the statistical associations in a multi-layer omics dataset need further development. Moreover, a growing body of evidence shows that post-transcriptional events are of considerable importance in Mtb, and that direct functional information can be retrieved from the state of the proteome, making quantitative proteomic data an asset and a relatively recent addition to integrated multi-layered omics analyses of bacteria (Chionh et al., 2016; Cortes et al., 2017). The mass spectrometric technique SWATH-MS, combines data independent acquisition with massively parallel targeted analysis of the acquired data (Gillet et al., 2012; Röst et al., 2014) and thus offers the high degree of reproducibility, consistency, and throughput needed for large cohort studies. It has recently been optimized in model experiments with Mtb (Collins et al., 2017; Schubert et al., 2013, 2015).

Here, we generate and analyze a multi-layer omics dataset that includes transcriptome and proteome data generated from genome sequenced clinical Mtb strains. Specifically three L1 and L2 strains, respectively, were studied during two *in vitro* growth conditions. We use these data to systematically examine the strength of post-transcriptional regulation in Mtb across various functional categories. Moreover, we tailor a genome scale transcriptional framework – GenSTrans – to identify transcription factors (TFs) that regulate their target genes differently between L1 and L2 strains. We further show that ~40% of the observed significant changes can be explained by a small set of identified TFs and the number of SNPs between each pair of strains determines the extent of the significant changes at both mRNA and protein level. Our model highlights four master TFs (DosR, Rv1985c, Rv0691c, and Rv3597c) that are known to interact with several sigma factors, including SigB. We discuss how the observed differential expression of DosR proteins, most likely triggered by SigB in response to nitric oxide stress, correlates with shorter growth arrest in L2 strains compared to L1. In summary, we demonstrate the potential of network modeling to describe molecular mechanisms that differentiate Mtb L1 strains from L2. We propose that these differences may explain some of the phenotypic changes observed in clinical setting for these strains.

Results

Multi-layer omics profiling of Mtb clinical strains

To establish the data basis for the integrated, multi-layer analysis of Mtb L1 and L2 clinical strains, we selected three strains from each lineage (Borrell et al., 2019; Chiner-Oms et al., 2019) such as to represent the greatest phylogenetic diversity within each lineage based on whole genome sequence data (Fig 1a). We then grew the six strains *in vitro* to mid-log phase and profiled their transcriptomes and proteomes in biological triplicates (18 samples) using RNASeq and SWATH-MS respectively (Fig 1b, Fig S1). Illumina NGS quantified 3,933 transcripts across the samples (transcript per million, TPM > 1). SWATH-MS quantified 21,184 peptides consistently across samples from which 2,479 proteins, corresponding to 62% of Mtb's open reading frames, were inferred (Fig 2a, Fig S2a). Through the targeted data analysis pipeline *OpenSWATH* (Röst et al., 2014) and using the Mtb proteome library (Schubert et al., 2013) as prior information, three or more proteotypic peptides per protein were quantified for ~76% of the detected proteins (Fig2b). Next, we examined the quality and reproducibility of each of the omics layer individually. The median CVs (Coefficient of Variation) across the biological triplicates was ~12% at both the transcript and protein level, indicating comparable data quality between the two platforms (Fig S2b, c). Overall, we were able to quantify 94% of detected proteins across all samples by direct measurement. Where this was not possible (the remaining 6%), we inferred upper bounds of possible protein expression from local background values (Fig 2c). The correlation between biomolecular features of the replicates was high for both the transcriptomic and proteomic data (Fig S2 d, e). We assessed the technical and biological variability of the respective datasets. The results indicated an ascending molecular variation with increasing genomic distance within and between the two lineages (Fig 2d). Our analyses also resulted in a perfect unsupervised clustering of Mtb strains and lineages for both transcriptomic and proteomic data (Fig S2 f, g). Principal component analysis (PCA) suggested that the protein data are more informative with respect to separating strains by their phylogenetic relationship compared to transcriptomic data. Indeed, the first two principal components of protein data exhibited a good separation, resolving individual strains (Fig 2e). By contrast, the resolution of PCA based on transcriptomic data was more limited and reflected the phylogenetic classification (Fig 2f). In short, we have generated a dataset, which spans from the genome of fully sequenced clinical isolates of Mtb to transcriptome and proteome. The multifaceted nature and deep coverage of the high data quality provided a solid base on which to build our downstream analyses.

Post-transcriptional regulation is prevalent in Mtb clinical strains

A growing body of evidence suggests that the extent of post-transcriptional regulation in Mtb is higher than previously appreciated (Chionh et al., 2016; Cortes et al., 2017). We used our combined multi-layer omics dataset to systematically explore the extent of transcriptional and post-transcriptional regulation in Mtb clinical strains and to assess the relative impact on gene

functional categories annotated by Tuberculist (Lew et al., 2011) between the two lineages. As an indirect measure of the degree of post-transcriptional regulation, we computed the Spearman rank correlation between averaged intensities of mRNA and protein products of the same gene across biological triplicates in the six strains studied. We found that the magnitude of the correlation, and hence the assumed extent of post-transcriptional regulation, varied between functional categories. Whereas the overall correlation was 0.46, the strongest and weakest signals of post-transcriptional regulation were observed in the "*virulence, detoxification, and adaptation*" (Spearman's rho: 0.33) and "*lipid metabolism*" (Spearman's rho: 0.69) group, respectively. Remarkably, the L2 strains presented a significantly higher mRNA to protein correlation compared to L1 for both regulatory (T-Test's p-value = 0.015) and virulence genes (T-Test's p-value = 0.0036) (Fig 3a). Comparing the global correlation observed in Mtb to that in other species showed that the potential of Mtb for post-transcriptional regulation was in the same range as for humans, despite the fact that bacterial systems are usually less complex (Liu et al., 2019). An earlier study showed that mRNAs and proteins of 137 *Escherichia coli* genes were moderately correlated, although the correlation values were variable (0.54 – 0.77) depending on the different quantification methods used in the study (Taniguchi et al., 2010). Other studies in yeast and metazoan species showed mRNA-protein correlations at the gene-to-gene basis of above 0.5 (Alli Shaik et al., 2014; Beyer et al., 2004; Brockmann et al., 2007; Ghaemmaghami et al., 2003). A possible explanation for the apparent increased post-transcriptional regulation in Mtb may be the presence of the proteasome system responsible for protein degradation and homeostasis that is rarely found in other bacteria (Becker et al., 2019). Due to the substantial post-transcriptional regulation in Mtb, RNA data in isolation can lead to incomplete or misleading mechanistic understanding. This underscores the need for high quality quantitative proteome data (Cortes et al., 2017). For instance, our data revealed that T cell antigens (both MHC class one and two) are significantly regulated at the RNA level in L2 strains compared to L1 (MCHI's p-value = 4.57E-5 and MHCII's p-value = 0.0071). However, this pattern of the differential regulation could not be observed at the protein level (MCHI's p-value = 0.07 and MHCII's p-value = 0.55) (Fig S3).

Transcript and protein abundances differ substantially between Mtb L1 and L2 strains

Next, we performed differential protein and transcript expression analyses of the L1 and L2 strains. We identified 578 differentially abundant transcripts and 390 differentially abundant proteins (fold change > 1.5 and adjusted p-value < 0.01) (Table S1). The vast majority of the changes occurred in non-essential genes involved in survival, infection and persistence of the bacilli, whereas the expression of Mtb essential genes remained largely unchanged (p-values = 6.9E-7 & 2.17E-6 on mRNA and protein level). Gene set enrichment analyses of the differentially abundant transcripts and proteins highlighted regulons and functional categories that distinguish the two lineages. These included the Dormancy survival Regulator (DosR) and Iron-dependent Repressor (IdeR) regulons. DosR proteins contribute to maintenance of cell viability during transition to a non-replicating state induced by various stresses such as macrophage engulfment, hypoxia and nitric oxide (Mehra et al., 2015). The abundance of these genes was significantly

regulated in L2 strains compared to L1, both on the transcript and protein level (Fig 3b, S4a). For transcripts this observation has been reported previously (Homolka et al., 2010; Rose et al., 2013) and it is confirmed here using both transcriptome and proteome data from the same samples. Chao JD *et. al.* showed that the abundance of DosR proteins is increased upon PknH, a serine/threonine-protein kinase, knockout (Chao et al., 2010). Since we found the abundance of PknH decreased in L2 strains, we decided to examine whether there was a link between the PknH downregulation and DosR upregulation. We hypothesized that PknH could regulate DosR proteins through either DosR or an alternative transcription factor. 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. However, we could not replicate the results shown by Yossef Av-Gay and colleagues and DosR proteins remained unchanged upon PknH knockout and overexpression (Table S2).

Our analyses also pointed to the regulon IdeR, (repressed in iron-replete conditions) as being significantly differentially abundant in the two lineages (Fig 3c, S4b). Iron acquisition systems in pathogenic bacteria are critical virulence factors and Mtb requires iron to be able to grow in culture and during infection (Rodriguez and Smith, 2006; Rodriguez et al., 2002). Hence, the increased expression of IdeR proteins in L2 strains could boost the potential of the iron acquisition system and therefore strengthen L2's success during infection (Gold et al., 2008). Lipid metabolism related proteins that showed the lowest mRNA-protein correlation also appeared significantly differentially abundant between the two lineages (Fig 3d, S4c). This functional category contains 272 genes in Mtb whereas, in comparison, the *E. coli* genome encodes only ~50 such proteins. The amount of energy Mtb spends to transcribe and translate these genes indicates their critical roles in the life cycle of Mtb. Proline-glutamic acid (PE)/proline-proline-glutamic acid (PPE) genes, many of which are considered as virulence factors, also displayed differential regulation between the lineages at both RNA and protein level (Fig 3e, S4d). These proteins are challenging to quantify since they are, to a large extent, either secreted or localized on the cell surface. We next examined transcription factors (TFs) of which 138 out of 214 (%64) could be detected at the protein level. We found that at least 24 TFs showed changes in their abundance at the protein level between the lineages (Fig 3f, S4e). Taken together, our multi-omics data revealed important differences in the functional and regulatory organization of Mtb clinical isolates from lineages 1 and 2.

The number of single nucleotide polymorphisms (SNPs) is linked to the magnitude of transcript and protein expression differences between strains.

The large extent of the observed transcriptional and translational changes between the two lineages led us to examine the role of SNPs, the most common form of genetic variation in Mtb. To this end, we first calculated the genomic distance based on number of SNPs between each pair of the tested Mtb strains. The average pairwise genomic distances within L1 and L2, as well as between the two lineages, were 873, 446 and 1,761 SNPs, respectively. These results are consistent with results reported previously (Coscolla and Gagneux, 2014) that showed that the

genetic diversity within L1 was about twice that of L2. We then examined the extent to which the observed changes at the transcript and protein level correlated with the number of SNPs in the corresponding genomes. Our analysis revealed a Spearman correlation between the genomic distance, measured in number of SNPs, and the differentially regulated mRNAs and proteins of 0.73 and 0.83, respectively (Fig 4a, b). The lower proteomic coverage diminished the slope of the corresponding regression line but it almost met the slope calculated for transcripts following a coverage-based normalization. The elucidation of the one-to-one relationship and exact mechanism of each causative SNP goes beyond the scope of this analysis due to the complexity of potential mechanisms. For instance, genetic variation might affect regulatory proteins such as TFs by either changing the affinity of a given protein to DNA (both synonymous and non-synonymous SNPs), or modifying a regulatory protein (synonymous SNPs), and therefore influence gene expression. Hence, we included both synonymous and non-synonymous SNPs in our analysis. Overall, our data indicate that, on average, every three SNPs caused a significant expression change which can be detected at either transcript or protein level.

A genome-scale transcriptional model identifies divergent regulators between L1 and L2 strains.

The differential expression between the two lineages of a high number of TFs observed in our dataset led us to ask to what extent changes in transcript or protein abundance can be explained by direct regulation of TFs. To address this question, we leveraged an existing transcriptional framework, the Mtb's Environment and Gene Regulatory Influence Network (EGRIN) model, to identify TFs causally linked to the observed changes in transcript and protein abundance (Peterson et al., 2014). The EGRIN model defines clusters of co-regulated transcripts (modules) and their corresponding TFs. It requires dual enrichment analysis to connect a given regulated gene set to the modules and the modules to TFs (Fig S5a) (Peterson et al., 2016). Applying the model to our transcriptomic dataset highlighted 15 modules that were i) associated with at least one TF and ii) significantly enriched in transcripts that were differentially abundant between the two lineages (adjusted p-value < 0.05) (Fig S6a). The transcription factor DosR was identified with high significance through four different modules as a putatively causative TF for divergent gene expression in L1 and L2 strains. In addition, we also found that KstR2 (Rv3557c) could regulate its target genes differently between the two lineages. This transcription factor was identified through two different modules and controls the expression of a small set of genes involved in the utilization of cholesterol, the main carbon source used by Mtb during infection (Kendall et al., 2010) (Fig S6b). Whereas EGRIN provides a high degree of sensitivity and is well suited to pathway-level analysis provided that the model includes pathways of interest, the model could not satisfactorily address our quest for the identification of lineage-specific regulators for two reasons. First, the EGRIN model contains only 66 transcription factors and is therefore not a genome-scale model. Second, each module encompasses only 13 genes on average, which mitigates the specificity of the model. Consequently, the Mtb EGRIN based analysis can suggest several TFs regulating a given module, making it impractical to identify the causative TF with

confidence (Fig S6a). To improve on these limitations, we devised a new genome-scale transcriptional framework.

We repeated the analyses with a new genome-scale transcriptional network model, GenSTrans, that was developed based on ChIPSeq experiments (Minch et al., 2015a). The network consists of 143 transcription factors and their corresponding sets of target genes represented as specific sub-networks. Each sub-network on average consists of 64 genes. The high density of information in the network is expected to boost the specificity of the analysis compared to the EGRIN model. We overlaid the transcript data of this study on the GenSTrans network and examined the degree of enrichment of differentially abundant transcripts in specific sub-networks using hypergeometric test followed by Benjamini-Hochberg based multiple testing correction (Fig S5b). This analysis suggested 28 TFs that differentially regulated gene expression between L1 and L2 strains (Fig 5a). Of the 28 TFs identified by GenSTrans, 17 were not identified by the EGRIN model due to insufficient genome coverage (Fig 5a). Our new model could recapitulate at least one TF corresponding to 11 modules also identified by the EGRIN model (Fig S6a). Of note, GeneSTrans more precisely pointed to one or two TFs per EGRIN's module in cases where the EGRIN model has suggested several equally likely candidates. For instance, the EGRIN model identified module 446 that could potentially be regulated by either Rv0081 or Rv1460, or a combination of both (Fig 5a). Rv0081 and Rv1460 bind to 647 and 11 genes, respectively, according to ChIPSeq experiments (Minch et al., 2015a). De-convolving the signals is further complicated because the module overlaps with both sub-networks significantly. Nevertheless, the non-overlapping genes pointed to Rv1460 as the most likely causative TF explaining the observed pattern, as 9 out of 11 genes in the corresponding sub-network appeared regulated in the comparison of L1 and L2 strains.

Next, we returned to the original question about causation of differential gene expression between L1 and L2 strains. Specifically, we examined the extent to which the identified TFs could describe the observed expression differences. Our results revealed that at least 37% of the modulated transcripts could be statistically significantly linked to a relatively small set of TFs. A set of only four TFs – Rv3133c (DosR), Rv1985c, Rv3597c (Lsr2), and Rv0691c – were sufficient to explain a quarter of the transcriptomic changes between L1 and L2 strains (Fig 5b). DosR on both transcript and protein level and Rv1985c on transcript level – not identified through our proteome profiling – were differentially regulated in their abundance very significantly between the two lineages. Interestingly, and in contrast to DosR and Rv1985c, Rv3597c and Rv0691c did not show significantly different transcript/protein abundance between the lineages. We speculated that they might undergo post-translational modifications (PTMs) regulation that affect their respective level of activity. In Mtb, various types of PTMs have been identified on at least 1,200 proteins (Banaei-Esfahani et al., 2017). For instance, PknB, a serine/threonine-protein kinase, can phosphorylate Rv3597c (Turapov et al., 2018). The sub-networks of these four TFs shared several genes with each other, indicating their close functional relationship (Fig 5c, d). Of note, DosR and Rv1985c bound to 31 and 28 DosR regulon genes, respectively, according to ChIPSeq experiments (Minch et al., 2015a). These highly overlapping target genes belonging to the DosR regulon might

indicate Rv1985c as an alternative TF contributing to the DosR target regulation under certain conditions.

Hierarchical organization for sigma factors – transcription factors interactions

In addition to 214 TFs, the tubercle bacillus has one housekeeping sigma factor, SigA, and 12 accessory sigma factors leading to the reprogramming of RNA polymerase (RNAP) and consequently the initiation of transcription of particular gene sets. The sigma subunit ensures specificity of the RNAP holoenzyme for special promoter sequences (Rodrigue et al., 2006; Sachdeva et al., 2010). Gennaro and colleagues reconstructed a sigma factor regulatory network of Mtb. It consists of 13 sigma factors, seven anti-sigma factors, two anti-anti-sigma factors and 50 TFs (Chauhan et al., 2016), altogether 72 genes. Of these, we were able to quantify the abundance of all 72 at the transcript level and for 48 genes at the protein level. Hence, we asked how significantly the four master TFs could interact with the sigma factor network and generate differential response to a given stress.

We assessed whether the four master TFs identified in our analysis were present in the sigma factor network and how their interactions with sigma factors/anti-factors might affect gene expression (Fig S7). Three of the master TFs, Lsr2, DosR and Rv0691c, were among the 50 TFs represented in the sigma factor network. Therefore our four TFs of interest were significantly enriched using hypergeometric test (p-value = 0.0027). According to the model, Lsr2 and DosR both interact with three sigma factors directly, whereas Rv0691c interacts with four sigma factors indirectly through their corresponding anti-sigma factors. To examine the potential of these three TFs of regulating the whole sigma factor network, we inferred the hierarchical organization of the sigma factor network using the hierarchy score maximization algorithm (Cheng et al., 2015). This algorithm shows how a given directed network is organized in multiple levels and elucidates the position of each component at the inferred levels. The results indicated that the sigma factor network has either three or four levels, and that DosR, Lsr2 and Rv0691c fall into the top-level of the network propagating regulations towards the lower levels (Table S3). Moreover, SigB, controls the transcription factor DosR expression (Chauhan et al., 2016). Of note, only the protein RsfB (anti-anti-sigma factor) showed significant differential expression between the two lineages among all the sigma factors and their corresponding anti-sigma factors and anti-anti sigma factors quantified on transcript and protein level (Table S1). Overall, these three TFs located at the top-level of the sigma factor network interact with eight sigma factors that could, in part, facilitate stress responses of L2 strains compared to L1. As the DosR regulon showed a higher basal expression in L2 strains, likely due in some part to the transcriptional regulatory network outlined, we next considered the extent to which DosR proteins would respond differently to a relevant stress.

Differential response of L1 and L2 strains to nitric oxide stress

To determine whether the molecular differences observed between L1 and L2 strains, such as in the DosR regulon, are functionally relevant, we exposed the strains to nitric oxide (NO) stress for 24 hours and analyzed the samples with respect to their growth status and protein expression levels. NO stress simulates the physiological environment that the bacilli experience following engulfment by macrophages (Nathan, 2006). Specifically, we (i) quantified DosR target proteins and (ii) verified the growth arrest duration after NO stress (see Materials and Methods). The expression of DosR target proteins increased more strongly in response to NO in L2 strains compared to L1 (Wilcoxon signed rank test's p-value = 0.00022) (Fig 6a). Given the tight interaction with DosR and sigma factors we also examined differential expression of sigma factors in L1 and L2 on treatment with NO. Our differential protein expression analysis showed that the strength of the SigB induction to NO differed between L1 (fold change = 1.24, adjusted p-value = 0.006) and L2 (fold change = 1.94, adjusted p-value = 3.37E-6) strains while the other quantified sigma factors remained unchanged in response to NO within the both lineages (Table S4). Altogether, DosR proteins showed not only a higher basal expression but also a stronger induction following NO challenge in L2 strains when compared to L1.

Based on these molecular data, we hypothesize that following NO exposure, L2 strains, would be better capable to restart their growth compared to L1 strains. We tested this hypothesis by measuring the duration of growth arrest in two strains from each lineage in biological triplicates. The results demonstrated a significantly shorter growth arrest in L2 strains compared to L1 (Wilcoxon signed rank test's p-value = 0.0181), supporting our hypothesis (Fig 6b). In summary, our findings show that network-based analyses generate a mechanistic hypothesis that was not directly apparent from the multi-omics data. We then validated the network-driven hypothesis experimentally and found that differential regulation of bio-molecular networks might drive distinct phenotypes in Mtb clinical isolates.

Discussion

We used a multi-omics approach combined with network analyses and validation experiments to compare clinical strains of Mtb belonging to L1 and L2 to better understand what causes the differences in clinically relevant phenotypes such as virulence, transmission and drug resistance that have been observed between these lineages (Borrell et al., 2019). For this, we went beyond the conventional multi-layer omics data analysis approaches and introduced a new integrative, network-based analysis. Our results revealed that (i) the Mtb genetic distance computed based on SNPs determines largely the number of significant changes at both mRNA and protein level, (ii) the extent of post-transcriptional events in Mtb differs across functional categories, (iii) L2 strains compared to L1 strains exhibit reduced post-transcriptional regulatory potential for gene sets annotated with "regulatory" and "virulence" function, (iv) four TFs explain 25% of the expression changes observed between the two lineages, and (v) the observed molecular changes can translate to relevant phenotypes, as exemplified by nitric oxide stress-induced growth arrest, which was reduced in L2 strains.

First, we sought to elucidate to what extent SNPs cause significant changes at the mRNA and protein level. Hence, we correlated genomic distance (i.e. the number of distinct SNPs between a given pair of strains) and the number of significant changes on both transcript and protein levels. We showed that every third SNP on average is responsible for one significant change in gene expression. This observation could trigger a large community effort to map genotype – proteotype association in Mtb via protein quantitative trait locus (pQTL) analysis. This type of study has been implemented in various organisms ranging from yeast to mouse to human but not yet in bacterial systems (Großbach et al., 2019; Picotti et al., 2013; Williams et al., 2016).

Second, we investigated the significance of post-transcriptional regulation, both systems-wide and for different functional categories. The analyses revealed that the degree of Mtb post-transcriptional regulation as determined by the correlation between transcript and protein measurements resembles the patterns observed in higher organisms (Liu et al., 2019). While for some functional categories, such as lipid metabolism, the transcript – protein correlation was relatively high, for others such as virulence genes, it was very low. Therefore, transcriptomic measurements in isolation might be misleading for Mtb studies. For instance, Mtb regulates thousands of its transcripts in response to nitric oxide but only a few hundred of these changes are detectable on the protein level and the rest of them are buffered out (Cortes et al., 2017). Our analyses revealed that for virulence and regulatory genes, post-transcriptional events are less pronounced in L2 strains when compared to L1. This might suggest that L2 strains might have evolved to react more rapidly to various stresses and therefore post-translational mechanisms are most likely at play for the mentioned gene groups.

Third, we examined the extent to which significant changes between L1 and L2 strains can be explained by transcription factors (TFs). We started with the published EGRIN model, but this

pipeline was insufficient to identify all causative TFs for two reasons, namely, the lack of specificity and its sparse coverage of TFs. Hence, we devised a genome-scale transcriptional network analysis, GenSTrans, which recapitulated the vast majority of TFs identified by the EGRIN model. It also shed some new light on 17 further TFs that were not represented in the EGRIN model. The analysis resulted in four transcription factors – DosR, Rv1985c, Rv0691c and Lsr2 – being linked to a quarter of the significant changes in gene expression. These transcription factors were significantly enriched in the sigma factor network and can regulate seven sigma factors, reprogramming RNA polymerase to modulate its affinity to promoter sequences under various stresses and conditions. Hierarchical analysis of the sigma factor network showed that the master TFs identified in this study fall into the top-level organization of the sigma factor network. This creates numerous possibilities for L2 strains to respond differently to a given stress compared to L1. Of note, SigB was the only sigma factor known to control the expression of one of our proposed master TFs. Since our analyses highlighted the differential basal expression of DosR genes between L1 and L2 on both transcript and protein level, we hypothesized that DosR target proteins might reveal a similarly differential trend in response to a relevant stress. Hence, we tested and validated this hypothesis in nitric oxide (NO) stress replicating a physiological stress that the tubercle bacilli experience following uptake by macrophages. The involvement of the sigma factors as enriched in our transcriptional regulatory analysis, and, in particular, the previously shown SigB regulation of DosR, is supported by the finding that SigB displays a stronger induction with respect to NO stress in L2 vs L1 strains. This is also consistent with previous findings showing that SigB knockout strains are deficient in a variety of stress responses (Datta et al., 2011; Manganelli et al., 2004). Overall, the data showed that DosR proteins hold not only a higher basal expression but also a stronger response to NO stress in L2 strains compared to L1. We further elucidated that this molecular behavior empowers L2 strains to restart their growth following exposure to NO more rapidly. Altogether, our network analysis identified four major TFs orchestrating a large fraction of differential gene expression between L1 and L2 strains that could drive distinct phenotypes.

In summary, this study further challenged the dogma suggesting that *Mycobacterium tuberculosis* is a phenotypically homogenous clone. We showed that in clinical isolates of Mtb L1 and L2, despite being genetically rather similar, hundreds of mRNAs and proteins are differentially expressed and that these molecular differences are being translated to phenotypes likely to be of relevance in the clinic. This study also provided important insights to guide future efforts to link specific genetic loci with mRNA and/or protein-level changes. Such a map of genetic features and their effect on molecular and macroscopic phenotypes in Mtb provides an alternative regulatory archetype to those published; one derived from a genetically fairly conserved bacterium which may serve as a valuable counter point for future studies.

Materials and Methods

Mtb Strains and bacterial culture

We used six clinical isolates of Mtb, three from L1 and three from L2, that are part of a recently described set of reference strains (Borrell et al., 2019). These strains were chosen in an attempt to capture the global genetic diversity of Mtb, both within and between lineages. We grew bacteria in a modified 7H9 medium 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. With respect to conventional 7H9 culture medium, we omitted glycerol, tween 80, oleic acid and catalase. For global expression profiling experiments we cultured the bacteria in 1l bottles containing large glass beads to avoid clumping and 100 ml of media. We incubated the cultures at 37°C and rotated them continuously on a roller. For nitric oxide stress experiments we grew the bacteria in 50ml conical screwcap tubes containing either 17ml (proteomic profiling) or 10ml of culture medium. The cultures were incubated at 37°C on an orbital shaker.

Transcriptomic profiling

We transferred a 40 ml aliquot of bacterial culture in mid-log phase (OD600 = 0.5 ± 0.1) into a 50ml Falcon conical tube containing 10 ml ice. We harvested the cells by centrifugation (3,000×g, 7 min, 4°C), re-suspended the pellet in 1 ml of RNApro solution (MP Biomedicals) and transferred the suspension to a Lysing matrix B tube (MP Biomedicals). We disrupted the bacterial cells using a FastPrep24 homogeniser (40s, intensity setting 6.0, MP Biomedicals). We clarified the lysate by centrifugation (12,000×g, 5 min, 4°C), transferred the supernatant to a clean tube and added chloroform. We separated the phases by centrifugation (12,000×g, 5 min, 4°C) and precipitated the nucleic acids from the aqueous phase by adding ethanol and incubating at -20C overnight. We performed a second acid phenol extraction to enrich for RNA. We treated our samples with DNAse I Turbo (Ambion), and removed stable RNAs by using the RiboZero Gram Positive ribosomal RNA depletion kit (Epicentre). We prepared the sequencing libraries using the TruSeq stranded Total RNA kit (Illumina) and sequenced them on a HiSeq2500 run in high output mode (50 cycles, single end).

The resulting reads were mapped to the Mtb H37Rv reference genome using BWA (ver. 0.7.13); the resulting mapping files were processed with samtools (ver. 1.3.1). Per-feature read counts were performed using the Python module htseq-count (ver. 0.6.1p1) and Python (ver. 2.7.11). We performed differential expression analysis using the R package edgeR and R (ver. 3.4.0) to identify Lineage-specific transcriptional changes.

Environment and gene regulatory influence network (EGRIN) based transcriptional model

We used the EGRIN model to analyze our transcriptional data in the exact same way that described before (Peterson et al., 2016). The model describes a set of modules that each one contains a dozen of co-regulated genes. The ChIPSeq and TFOE data paved the way to correspond each module to its potential regulators namely transcription factors (TFs) (Peterson et al., 2014). Here, we projected our significant differentially regulated mRNAs (fold change > 1.5 and adjusted p-value < 0.01) between L1 and L2 strains onto the modules to figure out how significantly each module was enriched. The enrichment analysis was performed using hypergeometric test followed by Benjamini-Hochberg (BH) multiple testing correction (adjusted p-value < 0.05). To visualize a given identified TFs, we included all its corresponding modules even if they were not significantly enriched.

Genome-scale transcriptional model and networks integration

To reconstruct the Mtb transcriptional network, we retrieved a ChIPSeq data publicly available at the Mtb portal V2 (Minch et al., 2015a). We included both operon and direct interactions in the final network, which offered a great framework for genome-scale analysis. In contrast to the EGRIN model, the TFOE data was totally excluded as they largely feature indirect effects of a transcription factor overexpression to the Mtb transcriptome. The reconstructed network comprised 143 transcription factors and 2,943 genes and can easily be extended upon the availability of ChIPSeq data for the remaining transcription factors. Next, a new set of regulons/modules, each contained the target genes of a given TF, were defined. Such a definition provided a one-to-one relationship between the modules and the TFs. Moreover, the average size of each module became ~65 genes, five times that of the EGRIN model, offering a larger specificity. The significantly regulated transcripts in L2 strains compared to L1 were tested against each module using a hypergeometric test followed by Benjamini-Hochberg (BH) multiple testing correction (adjusted p-value < 0.05). The identified TFs were compared to the EGRIN model results (see the color codes in the corresponding figures). Four TFs with the largest regulated gene size explaining 25% of the significant changes were introduced as the master regulators. This approach is easily portable to other bacterial systems upon the availability of respective ChIPSeq data. The TFs of the identified sub-network was integrated into the Mtb sigma factors network constructed by Gennaro and colleagues (Chauhan et al., 2016). This networks integration revealed the interactions between our four identified TFs (three of these were presented in the sigma factor network) and the Mtb sigma factors. Sigma factors can reprogram RNA polymerases and change their affinity to a given promoter. Hence, such a network integration showed how significantly the identified TFs are capable to present different phenotypic features in response to a perturbation. Notably, SigB as the only sigma factor which could modulate the transcription factor DosR provided a great opportunity to validate the capability such analyses.

Proteomic profiling using SWATH-MS

We aliquoted 20 OD equivalents from mid-log phase (OD_{600} = 0.5 ± 0.1) bacterial cultures (e.g. 40ml of OD_{600} 0.5) into 50ml conical tubes. In the case of nitric oxide stress we harvested 8ml of bacterial culture. We harvested the bacteria by centrifugation (3,000×g, 7 min, 4°C) and washed the pellet twice with cold PBS to remove tyloxapol from the samples. The washed pellets were re-suspended in lysis buffer, which contained 0.1M ammonium bicarbonate, 8M urea and 0.1% RapiGest (#186001861, Waters). We then transferred the suspension to a Lysing matrix B tube (MP Biomedicals) and disrupted the bacterial cells using a FastPrep24 homogeniser (40s, intensity setting 6.0, MP Biomedicals). We clarified the lysate by centrifugation (12,000×g, 5 min, 4°C) and sterilized it by filtering it twice through a 0.22 µm syringe filter (Milipore) prior to further processing.

Next, we measured the protein concentration of each sample using BCA assay (A53225, Thermo Fisher Scientific). Considering the measured concentrations, we started with 60 ug protein and appropriate volume of lysis buffer was added up to 100 ul to equalize the final concentration of each sample. Then, 5 mM tris(2-carboxyethyl)phosphine (TCEP) was added to reduce protein disulfide bonds while the sample were incubated at 37°C for 30 min. Afterwards, the free cysteine residues were alkylated by adding 40 mM iodoacetamide and incubating for 30 min in the dark. The samples were diluted 6 times with 0.05 M ammonium bicarbonate to reach a final Urea concentration below 2 M. To digest proteins, 1.2 ug sequencing grade modified trypsin (V5113, Promega) was added (w/w 1:50). The samples were incubated overnight (~16 hours) at 37°C with gentle shaking of 300 rpm. To stop protein digestion, we acidified the samples (PH < 2) using formic acid followed by and incubation for 30 min with shaking of 500 rpm. We desalted the clear peptide solution using C18 MicroSpin columns (The Nest Group, 30-300 ug loading capacity). Next, we spiked in iRT peptides (Ki-3002, Biognosys) which allowed us to normalize retention time during the data processing step.

We measured the samples on a TripleTOF 5600 mass spectrometer (AB Sciex) coupled to a nanoLC system (Eksigent) as described before (Collins et al., 2017). The data acquisition was performed in SWATH-MS mode using the 64 variable windows scheme.

The OpenSWATH workflow, containing three software tools, paved the way to process the data. We used the pan Mtb spectral library generated on the same type of mass spectrometry, TripleTOF 5600 (Schubert et al., 2013). OpenSWATH extracted chromatograms and assigned peak groups according to the prior knowledge, the pan Mtb spectral library (Röst et al., 2014). Next, we performed PyProphet-cli to score the assigned peak groups using a semi-supervised algorithm. The data was filtered to global 1% FDR on both peptide and protein level. Eventually TRIC was used to integrate various information of each run and align the extracted and scored peak groups between runs offering a high degree of consistency between measurements (Röst et al., 2016). To perform absolute (based on top 3 peptides and top 5 transitions) and relative quantification,

we used the R package SWATH2stats (Blattmann et al., 2016) and eventually aLFQ (Rosenberger et al., 2014) and MSstats (Choi et al., 2014).

Assesment of nitric oxide stress

To measure the proteomic changes following nitric oxide (NO) exposure, we grew 17ml bacterial cultures to mid-log phase (OD_{600} approximately 0.5) and harvested 8ml of the culture for proteomic characterization. We then added Diethylenetriamine/NO adduct at a final concentration of 1mM to the remaining culture and incubated it for a further 24h at 37^{o}C. After this time, we harvested 8ml of the culture for proteomic characterization.

To assess the duration of NO induced growth arrest, we grew duplicate bacterial cultures to mid-log phase (OD_{600} approximately 0.5) and added Diethylenetriamine/NO adduct at a final concentration of 1mM to the treatment tube with nothing added to the control. We followed the subsequent growth dynamics by measuring OD_{600}.

Genetic, transcriptomic and proteomic distance analysis between clinical isolates

To compute strain distances on various levels, we assigned a particular cutoff to digitize them. On genome layer, we ignored various genetic insertions and deletions and only counted the numbers of distinct SNPs between a pair of Mtb strains. For mRNA and protein level, we considered a given cutoff (fold change > 1.5 and adjusted p-value < 0.01) and then the number of the significant changes were calculated. For each layer, 15 comparisons consisting of six intra lineage (three pairs within each lineage) and nine inter lineages comparisons were assumed. Afterwards Spearman correlation was computed between the genomic and either transcriptomic or proteomic distance. The slope of the respective regression line showed how many significant changes on protein and transcript level are caused by a given distinct SNP.

Sigma factor network hierarchical properties

To infer the hierarchical organization of the sigma factor network recently described (Chauhan et al., 2016), we used the hierarchy score maximization algorithm (Cheng et al., 2015). We performed the algorithm for different number of levels k (2-6). In each analysis, it computes probability scores to assign a given node to layers of the hierarchical organization. Two criteria were assessed to determine the optimal choice of k. First, the enrichment of the downward direction compared to expectation that is summarized in the corrected hierarchy score. Second, an ambiguity score which quantifies the uncertainty of a node assigning to a given layer of the organization. These two measurements revealed that k=4 is the optimal.

Acknowledgements

We would like to thank John Aitchison (Seattle Children's Hospital), Xueli Guan (Nanyang Technological University) and Uwe Sauer (ETH Zurich) for their intellectual contributions. We further thank Christoph Grundner (Seattle Children's Hospital) for providing the PknH knockout and overexpressing strains. We would like to thank the Genomics Facility Basel for profiling the transcriptome of the clinical strains. We would also thank the Scientific IT Support (ID SIS) of ETH Zurich and the scientific computing center at University of Basel for support and maintenance of the laboratory-internal computing infrastructure. This work was supported by the SystemsX.ch project TbX, the National Institutes of Health project Omics4TB Disease Progression (U19 AI106761), the Swiss National Science Foundation (grants 310030_166687, IZRJZ3_164171, IZLSZ3_170834 and CRSII5_177163) and the European Research Council (309540-EVODRTB). B.C.C. was supported by a Swiss National Science Foundation Ambizione grant (PZ00P3_161435).

Author Contributions

Conceptualization: SG, RA, BCC, ABE and AT; Clinical strains cultivation: SMG and JF; PknH strains cultivation: TRR and DRS; transcriptomic samples preparation: AT; Transcriptomic data acquisition: CB; Transcriptomic data processing and differential analysis: AT and ABE; Proteomic samples preparation, data acquisition and data processing: ABE, BCC, LCG and OTS; Genome-scale transcriptional model: ABE; Growth arrest experiment: SB and JF; Other data analysis: ABE; Manuscript preparation: ABE (with critical inputs from SG, BCC and RA); Supervision: BCC, RA and SG.

Declaration of Interests

The authors declare no competing interests.

Reference

Achtman, M. (2008). Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu. Rev. Microbiol. *62*, 53–70.

Alli Shaik, A., Wee, S., Li, R.H.X., Li, Z., Carney, T.J., Mathavan, S., and Gunaratne, J. (2014). Functional Mapping of the Zebrafish Early Embryo Proteome and Transcriptome. J. Proteome Res. *13*, 5536–5550.

Almeida, D., Ioerger, T., Tyagi, S., Li, S.-Y., Mdluli, K., Andries, K., Grosset, J., Sacchettini, J., and Nuermberger, E. (2016). Mutations in *pepQ* Confer Low-Level Resistance to Bedaquiline and

Clofazimine in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *60*, 4590–4599.

Banaei-Esfahani, A., Nicod, C., Aebersold, R., and Collins, B.C. (2017). Systems proteomics approaches to study bacterial pathogens: application to Mycobacterium tuberculosis. Curr. Opin. Microbiol. *39*, 64–72.

Becker, S.H., Jastrab, J.B., Dhabaria, A., Chaton, C.T., Rush, J.S., Korotkov, K. V, Ueberheide, B., and Darwin, K.H. (2019). The Mycobacterium tuberculosis Pup-proteasome system regulates nitrate metabolism through an essential protein quality control pathway. Proc. Natl. Acad. Sci. U. S. A. *116*, 3202–3210.

Beyer, A., Hollunder, J., Nasheuer, H.-P., and Wilhelm, T. (2004). Post-transcriptional Expression Regulation in the Yeast *Saccharomyces cerevisiae* on a Genomic Scale. Mol. Cell. Proteomics *3*, 1083–1092.

Bhatia, A.L., Csillag, A., Mitchison, D.A., Selkon, J.B., Somasundaram, P.R., and Subbaiah, T. V (1961). The virulence in the guinea-pig of tubercle bacilli isolated before treatment from South Indian patients with pulmonary tuberculosis. 2. Comparison with virulence of tubercle bacilli from British patients. Bull. World Health Organ. *25*, 313–322.

Blattmann, P., Heusel, M., and Aebersold, R. (2016). SWATH2stats: An R/Bioconductor Package to Process and Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools. PLoS One *11*, e0153160.

Borrell, S., and Gagneux, S. (2009). Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis. Int. J. Tuberc. Lung Dis. *13*, 1456–1466.

Borrell, S., Trauner, A., Brites, D., Rigouts, L., Loiseau, C., Coscolla, M., Niemann, S., De Jong, B., Yeboah-Manu, D., Kato-Maeda, M., et al. (2019). Reference set of Mycobacterium tuberculosis clinical strains: A tool for research and product development. PLoS One *14*, e0214088.

Brites, D., and Gagneux, S. (2017). The Nature and Evolution of Genomic Diversity in the Mycobacterium tuberculosis Complex. In Advances in Experimental Medicine and Biology, pp. 1–26.

Brites, D., Loiseau, C., Menardo, F., Borrell, S., Boniotti, M.B., Warren, R., Dippenaar, A., Parsons, S.D.C., Beisel, C., Behr, M.A., et al. (2018). A New Phylogenetic Framework for the Animal-Adapted Mycobacterium tuberculosis Complex. Front. Microbiol. *9*, 2820.

Brockmann, R., Beyer, A., Heinisch, J.J., and Wilhelm, T. (2007). Posttranscriptional Expression Regulation: What Determines Translation Rates? PLoS Comput. Biol. *3*, e57.

Caron, E., Aebersold, R., Banaei-Esfahani, A., Chong, C., and Bassani-Sternberg, M. (2017). A Case for a Human Immuno-Peptidome Project Consortium. Immunity *47*, 203–208.

Chao, J.D., Papavinasasundaram, K.G., Zheng, X., Chávez-Steenbock, A., Wang, X., Lee, G.Q., and Av-Gay, Y. (2010). Convergence of Ser/Thr and Two-component Signaling to Coordinate Expression of the Dormancy Regulon in *Mycobacterium tuberculosis*. J. Biol. Chem. *285*, 29239–29246.

Chauhan, R., Ravi, J., Datta, P., Chen, T., Schnappinger, D., Bassler, K.E., Balázsi, G., and Gennaro, M.L. (2016). Reconstruction and topological characterization of the sigma factor regulatory network of Mycobacterium tuberculosis. Nat. Commun. *7*, 11062.

Cheng, C., Andrews, E., Yan, K.-K., Ung, M., Wang, D., and Gerstein, M. (2015). An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. Genome Biol. *16*, 63.

Chiner-Oms, Á., González-Candelas, F., and Comas, I. (2018). Gene expression models based on a reference laboratory strain are poor predictors of Mycobacterium tuberculosis complex transcriptional diversity. Sci. Rep. *8*, 3813.

Chiner-Oms, Á., Berney, M., Boinett, C., González-Candelas, F., Young, D.B., Gagneux, S., Jacobs, W.R., Parkhill, J., Cortes, T., and Comas, I. (2019). Genome-wide mutational biases fuel transcriptional diversity in the Mycobacterium tuberculosis complex. Nat. Commun. *10*.

Chionh, Y.H., McBee, M., Babu, I.R., Hia, F., Lin, W., Zhao, W., Cao, J., Dziergowska, A., Malkiewicz, A., Begley, T.J., et al. (2016). tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. Nat. Commun. *7*, 13302.

Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics *30*, 2524–2526.

Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., Chan, D.W., Gibson, B.W., Gingras, A.C., Held, J.M., et al. (2017). Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nat. Commun. *8*, 291.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat. Genet. *45*, 1176–1182.

Cortes, T., Schubert, O.T., Banaei-Esfahani, A., Collins, B.C., Aebersold, R., and Young, D.B. (2017). Delayed effects of transcriptional responses in Mycobacterium tuberculosis exposed to nitric oxide suggest other mechanisms involved in survival. Sci. Rep. *7*, 8208.

Coscolla, M., and Gagneux, S. (2010). Does M. tuberculosis genomic diversity explain disease diversity? Drug Discov. Today. Dis. Mech. *7*, e43–e59.

Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in Mycobacterium

tuberculosis. Semin. Immunol. *26*, 431–444.

Datta, P., Shi, L., Bibi, N., Balázsi, G., and Gennaro, M.L. (2011). Regulation of central metabolism genes of Mycobacterium tuberculosis by parallel feed-forward loops controlled by sigma factor E (σE). J. Bacteriol. *193*, 1154–1160.

Gagneux, S. (2018). Ecology and evolution of Mycobacterium tuberculosis. Nat. Rev. Microbiol. *16*, 202–213.

Gagneux, S., and Small, P.M. (2007). Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet. Infect. Dis. *7*, 328–337.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics *11*, O111.016717.

Gold, B., Rodriguez, G.M., Marras, S.A.E., Pentecost, M., and Smith, I. (2008). The Mycobacterium tuberculosis IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. Mol. Microbiol. *42*, 851–865.

Großbach, J., Gillet, L., Clément-Ziza, M., Schmalohr, C.L., Schubert, O.T., Barnes, C.A., Bludau, I., Aebersold, R., and Beyer, A. (2019). Integration of transcriptome, proteome and phosphoproteome data elucidates the genetic control of molecular networks. BioRxiv 703140.

Guerra-Assunção, J., Crampin, A., Houben, R., Mzembe, T., Mallard, K., Coll, F., Khan, P., Banda, L., Chiwaya, A., Pereira, R., et al. (2015). Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife *4*.

Heusel, M., Bludau, I., Rosenberger, G., Hafen, R., Frank, M., Banaei-Esfahani, A., van Drogen, A., Collins, B.C., Gstaiger, M., and Aebersold, R. (2019). Complex-centric proteome profiling by SEC-SWATH-MS. Mol. Syst. Biol. *15*, e8438.

Holt, K.E., McAdam, P., Thai, P.V.K., Thuong, N.T.T., Ha, D.T.M., Lan, N.N., Lan, N.H., Nhu, N.T.Q., Hai, H.T., Ha, V.T.N., et al. (2018). Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat. Genet. *50*, 849–856.

Homolka, S., Niemann, S., Russell, D.G., and Rohde, K.H. (2010). Functional Genetic Diversity among Mycobacterium tuberculosis Complex Clinical Isolates: Delineation of Conserved Core and Lineage-Specific Transcriptomes during Intracellular Survival. PLoS Pathog. *6*, e1000988.

de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M., Jackson-Sillah, D.J., Fox, A., Deriemer, K., Gagneux, S., et al. (2008). Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J. Infect. Dis. *198*, 1037–1043.

Kendall, S.L., Burgess, P., Balhana, R., Withers, M., Ten Bokum, A., Lott, J.S., Gao, C., Uhia-Castro, I., and Stoker, N.G. (2010). Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. Microbiology *156*, 1362–1371.

Lew, J.M., Kapopoulou, A., Jones, L.M., and Cole, S.T. (2011). TubercuList – 10 years after. Tuberculosis *91*, 1–7.

Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E.G., Van Drogen, A., Borel, C., Frank, M., Germain, P.-L., Bludau, I., et al. (2019). Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nat. Biotechnol. *37*, 314–322.

Manganelli, R., Proveddi, R., Rodrigue, S., Beaucher, J., Gaudreau, L., and Smith, I. (2004). σ Factors and Global Gene Regulation in Mycobacterium tuberculosis. J. Bacteriol. *186*, 895–902.

Manson, A.L., Cohen, K.A., Abeel, T., Desjardins, C.A., Armstrong, D.T., Barry, C.E., Brand, J., Brand, J., Jureen, P., Malinga, L., et al. (2017). Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat. Genet. *49*, 395–402.

Mehra, S., Foreman, T.W., Didier, P.J., Ahsan, M.H., Hudock, T.A., Kissee, R., Golden, N.A., Gautam, U.S., Johnson, A.-M., Alvarez, X., et al. (2015). The DosR Regulon Modulates Adaptive Immunity and Is Essential for *Mycobacterium tuberculosis* Persistence. Am. J. Respir. Crit. Care Med. *191*, 1185–1196.

Minch, K.J., Rustad, T.R., Peterson, E.J.R., Winkler, J., Reiss, D.J., Ma, S., Hickey, M., Brabant, W., Morrison, B., Turkarslan, S., et al. (2015a). The DNA-binding network of Mycobacterium tuberculosis. Nat. Commun. *6*, 5829.

Minch, K.J., Rustad, T.R., Peterson, E.J.R., Winkler, J., Reiss, D.J., Ma, S., Hickey, M., Brabant, W., Morrison, B., Turkarslan, S., et al. (2015b). The DNA-binding network of Mycobacterium tuberculosis. Nat. Commun. *6*, 5829.

Nathan, C. (2006). Role of iNOS in human host defense. Science *312*, 1874–1875; author reply 1874-5.

Nicod, C., Banaei-Esfahani, A., and Collins, B.C. (2017). Elucidation of host-pathogen protein-protein interactions to uncover mechanisms of host cell rewiring. Curr. Opin. Microbiol. *39*, 7–15.

Penn, B.H., Netter, Z., Johnson, J.R., Von Dollen, J., Jang, G.M., Johnson, T., Ohol, Y.M., Maher,

C., Bell, S.L., Geiger, K., et al. (2018). An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses. Mol. Cell *71*, 637-648.e5.

Peterson, E.J.R., Ma, S., Sherman, D.R., and Baliga, N.S. (2016). Network analysis identifies Rv0324 and Rv0880 as regulators of bedaquiline tolerance in Mycobacterium tuberculosis. Nat. Microbiol. *1*, 16078.

Peterson, E.J.R.R., Reiss, D.J., Turkarslan, S., Minch, K.J., Rustad, T., Plaisier, C.L., Longabaugh, W.J.R.R., Sherman, D.R., and Baliga, N.S. (2014). A high-resolution network model for global gene regulation in Mycobacterium tuberculosis. Nucleic Acids Res. *42*, 11291–11303.

Phyu, S., Stavrum, R., Lwin, T., Svendsen, O.S., Ti, T., and Grewal, H.M.S. (2009). Predominance of Mycobacterium tuberculosis EAI and Beijing Lineages in Yangon, Myanmar. J. Clin. Microbiol. *47*, 335–344.

Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D.S., Schmidt, A., Deutsch, E.W., Röst, H., Sun, Z., Rinner, O., Reiter, L., et al. (2013). A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature *494*, 266–270.

Rodrigue, S., Provvedi, R., Jacques, P.-É., Gaudreau, L., and Manganelli, R. (2006). The σ factors of *Mycobacterium tuberculosis*. FEMS Microbiol. Rev. *30*, 926–941.

Rodriguez, G.M., and Smith, I. (2006). Identification of an ABC Transporter Required for Iron Acquisition and Virulence in Mycobacterium tuberculosis. J. Bacteriol. *188*, 424–430.

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., and Smith, I. (2002). ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. *70*, 3371–3381.

Rose, G., Cortes, T., Comas, I., Coscolla, M., Gagneux, S., and Young, D.B. (2013). Mapping of genotype-phenotype diversity among clinical isolates of mycobacterium tuberculosis by sequence-based transcriptional profiling. Genome Biol. Evol. *5*, 1849–1862.

Rosenberger, G., Ludwig, C., Röst, H.L., Aebersold, R., and Malmström, L. (2014). aLFQ: an R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. Bioinformatics *30*, 2511–2513.

Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., et al. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat. Biotechnol. *32*, 219–223.

Röst, H.L., Liu, Y., D'Agostino, G., Zanella, M., Navarro, P., Rosenberger, G., Collins, B.C., Gillet, L., Testa, G., Malmström, L., et al. (2016). TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat. Methods *13*, 777–783.

Rustad, T.R., Minch, K.J., Ma, S., Winkler, J.K., Hobbs, S., Hickey, M., Brabant, W., Turkarslan, S., Price, N.D., Baliga, N.S., et al. (2014). Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. Genome Biol. *15*, 502.

Sachdeva, P., Misra, R., Tyagi, A.K., and Singh, Y. (2010). The sigma factors of Mycobacterium tuberculosis: regulation of the regulators. FEBS J. *277*, 605–626.

Schubert, O.T., Mouritsen, J., Ludwig, C., Röst, H.L., Rosenberger, G., Arthur, P.K., Claassen, M., Campbell, D.S., Sun, Z., Farrah, T., et al. (2013). The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. Cell Host Microbe *13*, 602–612.

Schubert, O.T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., Gillet, L.C., Collins, B.C., Röst, H.L., Kaufmann, S.H.E., et al. (2015). Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of Mycobacterium tuberculosis. Cell Host Microbe *18*, 96–108.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Trauner, A., Banaei-Esfahani, A., Gygli, S.M., Warmer, P., Feldmann, J., Shafieechashmi, S., Eschbach, K., Zampieri, M., Borrell, S., Collins, B.C., et al. (2018). Resource misallocation as a mediator of fitness costs in antibiotic resistance. BioRxiv 456434.

Turapov, O., Forti, F., Kadhim, B., Ghisotti, D., Sassine, J., Straatman-Iwanowska, A., Bottrill, A.R., Moynihan, P.J., Wallis, R., Barthe, P., et al. (2018). Two Faces of CwlM, an Essential PknB Substrate, in Mycobacterium tuberculosis. Cell Rep. *25*, 57-67.e5.

Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., Zhang, L., Gao, C., He, Y., Li, Y., et al. (2010). Global Protein−Protein Interaction Network in the Human Pathogen *Mycobacterium tuberculosis* H37Rv. J. Proteome Res. *9*, 6665–6677.

Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science *352*, aad0189.

Wollenberg, K.R., Desjardins, C.A., Zalutskaya, A., Slodovnikova, V., Oler, A.J., Quiñones, M., Abeel, T., Chapman, S.B., Tartakovsky, M., Gabrielian, A., et al. (2017). Whole-Genome Sequencing of Mycobacterium tuberculosis Provides Insight into the Evolution and Genetic Composition of Drug-Resistant Tuberculosis in Belarus. J. Clin. Microbiol. *55*, 457–469.

World Health Organization (2018). Global tuberculosis report 2018. WHO.

Yang, M., Wang, Y., Chen, Y., Cheng, Z., Gu, J., Deng, J., Bi, L., Chen, C., Mo, R., Wang, X., et al. (2015). Succinylome Analysis Reveals the Involvement of Lysine Succinylation in Metabolism in Pathogenic *Mycobacterium tuberculosis*. Mol. Cell. Proteomics *14*, 796–811.

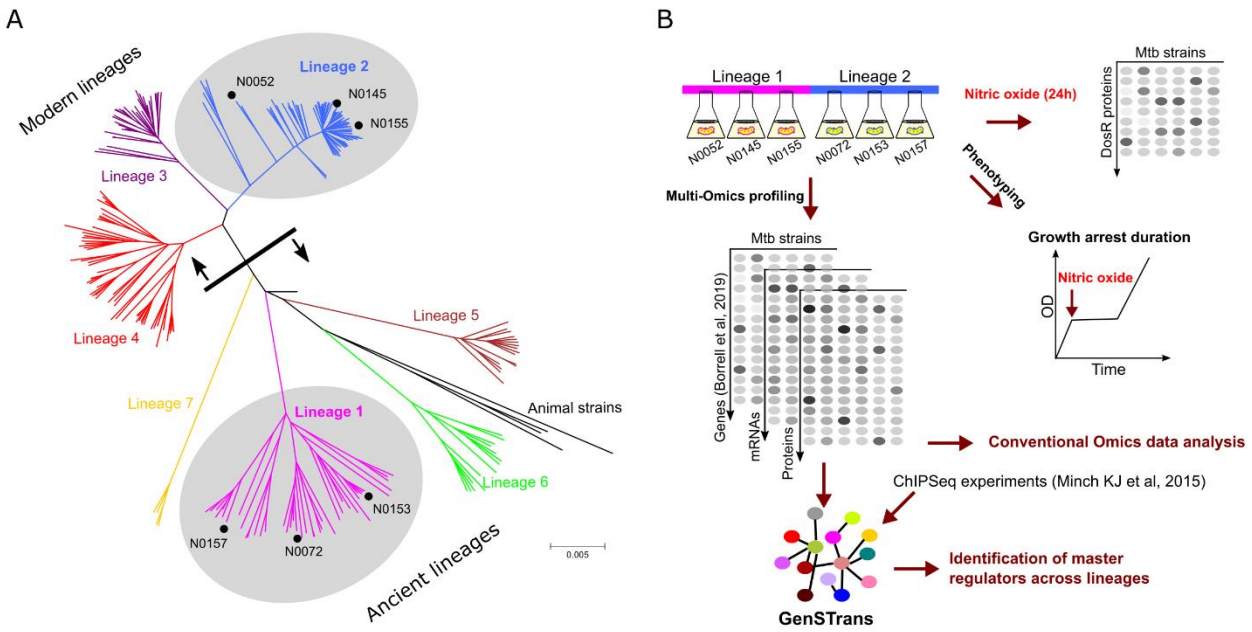

Figure 1. Schematic Illustration of Workflow. (A) Dendrogram shows the evolutionary relationships between the seven human-adapted lineages of the *Mycobacterium tuberculosis* complex and the L1 and L2 strains that were selected for this study. **(B)** Conceptual workflow. Transcriptome and proteome of six fully sequenced clinical strains belonging to L1 and L2 were measured. A genome-scale transcriptional model was developed to identify transcription factors regulating their targets differently between the two lineages. An exemplary phenotypic consequence of the differentially regulated DosR proteins between the two lineages and in response to a stress condition, the growth arrest duration following nitric oxide exposure.

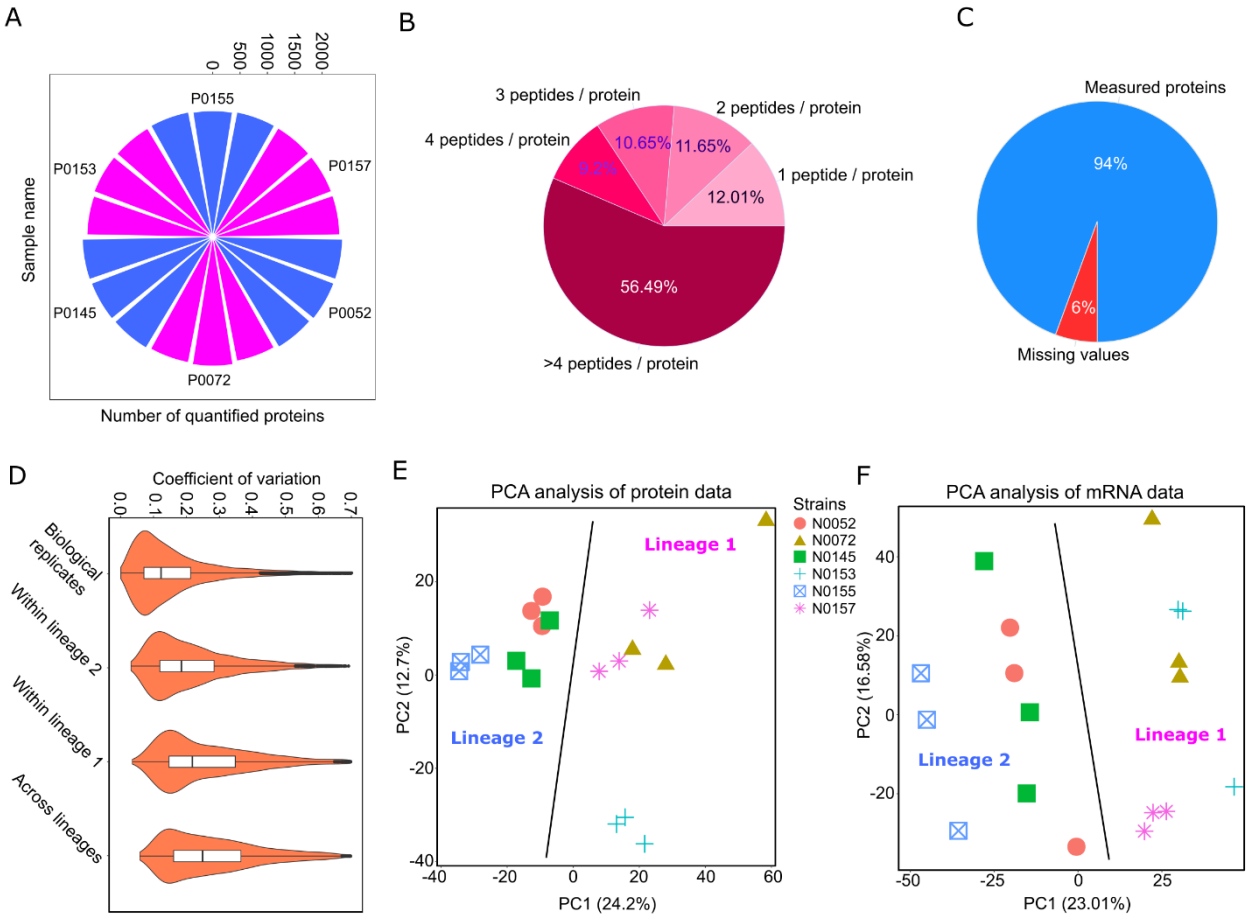

Figure 2. Cellular Protein Landscape in Mtb Clinical Strains. (A) Number of proteins quantified in each biological replicate of a clinical strain at 1% protein FDR. **(B)** Pie chart exhibits the deep coverage of peptides identified by SWATH-MS at 1% protein FDR. **(C)** Completeness of protein measurements. 94% of proteins across the samples were quantified by direct peptide measurements while the remaining (6%) were inferred using background signals. **(D)** Variation in SWATH-MS measurements due to different factors: technical (median CVs = 12%), within L2 (median CVs = 19%), within L1 (median CVs = 23%) and across all measurements (median CVs = 26%). **(D)** PCA analysis of protein data. PCA plot shows a decent separation resolution at the strain level. **(E)** PCA analysis of mRNA data. PCA analysis of mRNA data could separate clinical strains based on only their phylogenetic lineages.

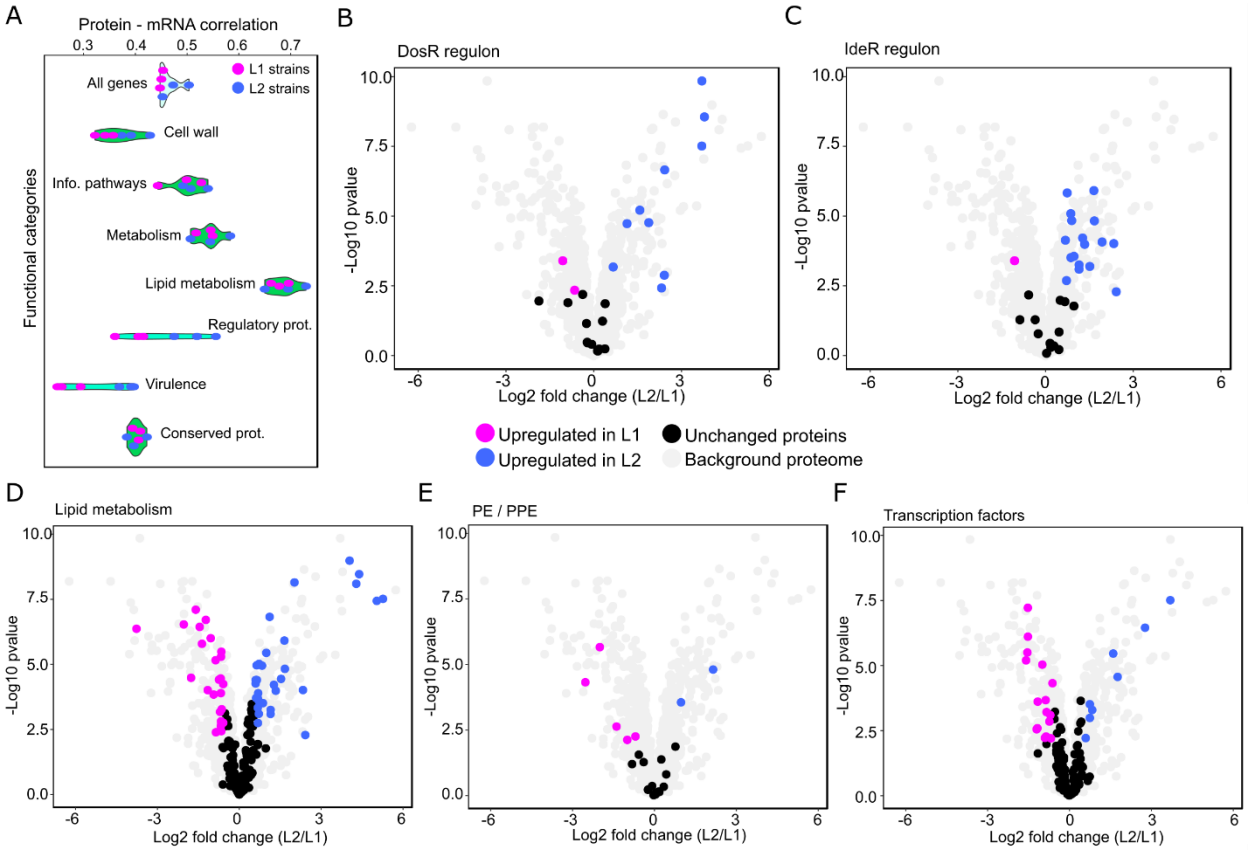

Figure 3. Enrichment and Correlation Analyses of Mtb Functional Categories. (A) Spearman correlation analyses of protein vs. mRNA abundances at the gene-to-gene basis for different functional categories: all genes (average R = 0.46), cell wall and cell processes (average R = 0.37), information pathways (average R = 0.50), Intermediary metabolism and respiration (average R = 0.54), lipid metabolism (average R = 0.69), regulatory genes (average R = 0.46), virulence, detoxification, and adaptation genes (average R = 0.33) and conserved genes (average R = 0.40). Volcano plots showing proteins of different functional groups that are differentially expressed between L1 and L2: **(B)** Dormancy survival Regulator (DosR) regulon (p-value = 2.5E-5), **(C)** Iron-dependent Repressor (IdeR) regulon (p-value = 8.3E-9), **(D)** lipid metabolism (p-value = 3.6E-5) and **(E)** PE/PPE genes (p-value = 0.014). **(F)** 24 transcription factors (out of 138 quantified) showed significant changes in their abundance in L2 strains compared to L1. For (B) to (F), purple and blue dots denote proteins significantly (p-value < 0.01 and fold change > 1.5) upregulated in L1 and L2 strains respectively, and black dots denoting proteins of the respective groups that do not significantly change. Figure S4 contains the corresponding mRNA data.

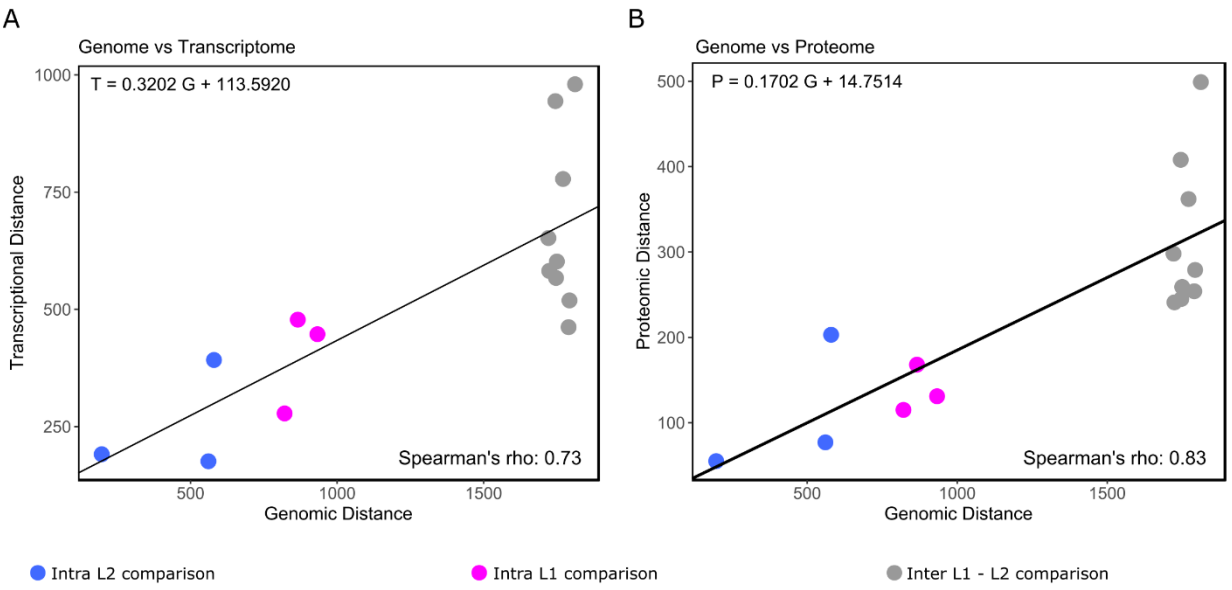

Figure 4. Genetic Distance Determines Number of Significant Changes on mRNA and Protein Level. (A) transcriptomic distance and **(B)** proteomic distance, measured by number of significant changes on mRNA and protein level respectively, as a function of genetic distance, calculated based on number of distinct SNPs for a given pair of clinical strains. Their slopes (following coverage-based normalization for proteomic slope) and high correlations showed that every three SNPs, on average, causes a significant change at the transcript and protein level.

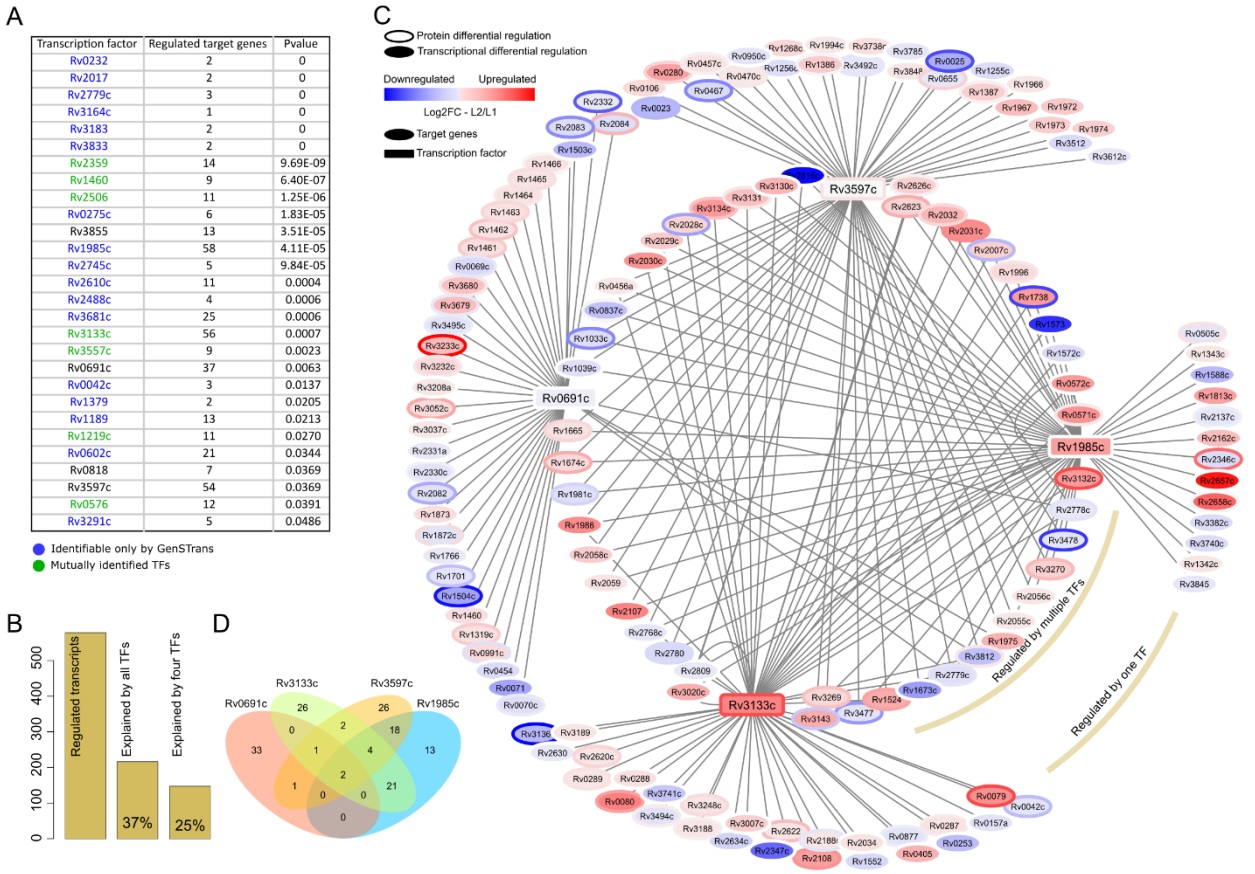

Figure 5. Transcriptional Modeling in Mtb L2 Strains Compared to L1. (A) 28 transcription factors regulating their target genes differently between L1 and L2 strains. Blue gene names denotes transcription factors that are not identifiable by the EGRIN model due to its sparse coverage. **(B)** Subnetwork of four master transcription factors explaining 25% of mRNA changes in L2 strains compared to L1. Node and border colors depict transcript and protein expression respectively. **(C)** Bar plot summarizes percentiles of regulated transcripts between the two lineages that could be explained by identified and four master transcription factors. **(D)** Venn diagram of target genes that are regulated by four master transcription factors. High degree of overlapping genes particularly between Rv3133c (DosR) and Rv1985c suggested their close functional relationships.

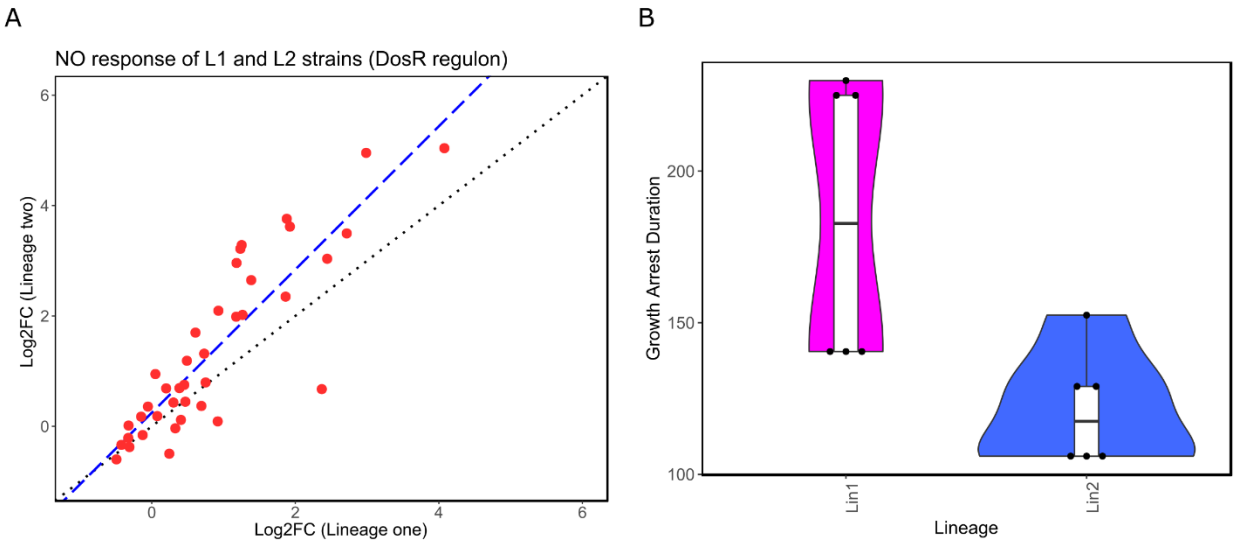

Figure 6. DosR Regulon Response to Nitric Oxide Exposure and Its Phenotypic Consequence. **(A)** Scatter plot shows bio-molecular responses of DosR proteins in Mtb clinical strains. Each dot represents a DosR protein while its respective response in L1 and L2 strains are denoted on X and Y axis respectively suggesting L2 strains responded significantly stronger to nitric oxide compared to L1 (Wilcoxon signed rank test's p-value = 0.00022). **(B)** Shorter growth arrest duration in L2 strains compared to L1 upon nitric oxide exposure (Wilcoxon signed rank test's p-value = 0.0181).

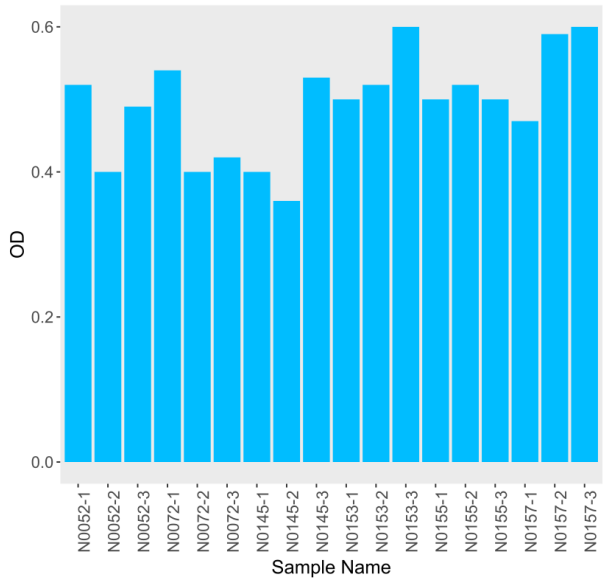

Figure S1. Bar Plot of ODs Measured at Harvest Time.

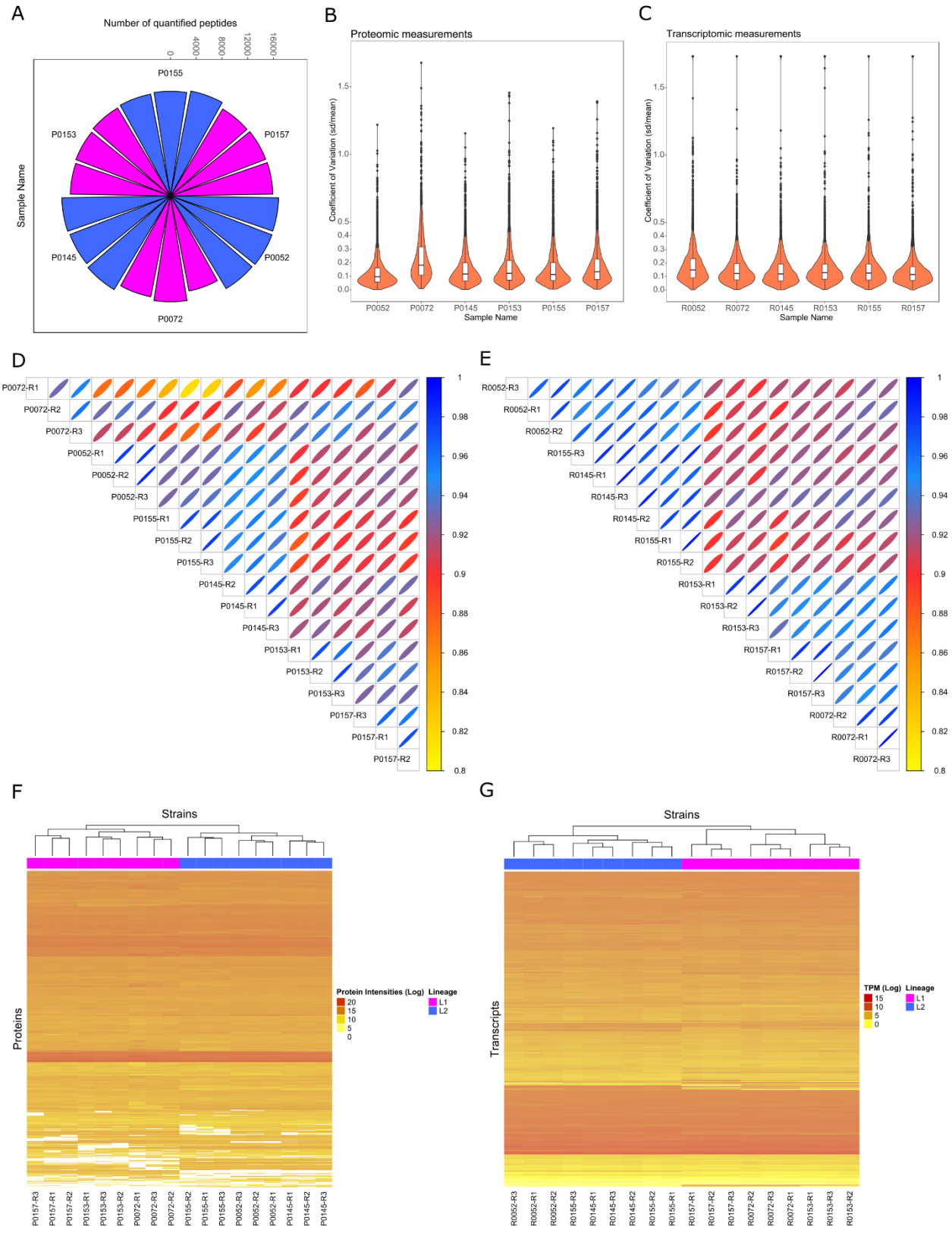

Figure S2. Transcript and Protein Data Quality Assessment. (A) Polar plot of peptide counts for each biological replicate of Mtb clinical strains. **(B-C)** Violin plots of CVs across biological replicates

of each strains. The same median CVs (12%) suggested comparable reproducibility between Illumina NGS platform and SWATH-MS used for transcriptomic and proteomic profiling of Mtb clinical strains. **(D-E)** Correlation plots of transcript (right) and protein (left) data. Spearman correlation analysis between clinical strains indicating near to perfect correlations and therefore reproducibility across biological replicates. **(F-G)** Perfect unsupervised clustering of clinical strains rooted in protein (left) and transcript (right) data.

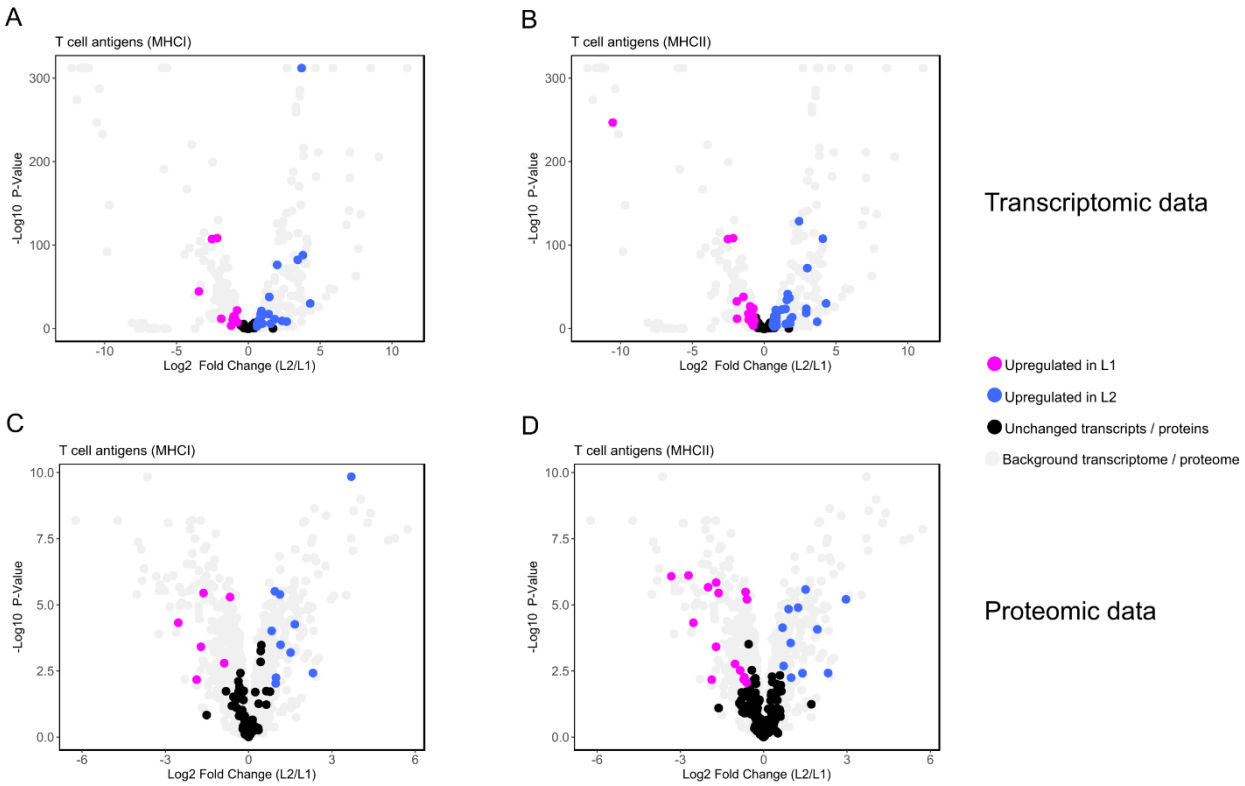

Figure S3. Enrichment Analyses of T Cell Antigens. (A-B) Volcano plot of T cell antigens presented by MHCI (left) and MHCII (right) at the mRNA level. The analysis showed that these functional categories are significantly enriched (p-value for MHCI = 4.5E-5, p-value for MHCII = 0.0071) by the differentially expressed transcripts between the two lineages. **(C-D)** Enrichment analysis of differentially expressed proteins in L2 strains compared to L1. MHCI (left) and MHCII (right) presented antigens were not significantly enriched anymore on protein level (p-value for MHCI = 0.069, p-value for MHCII = 0.55).

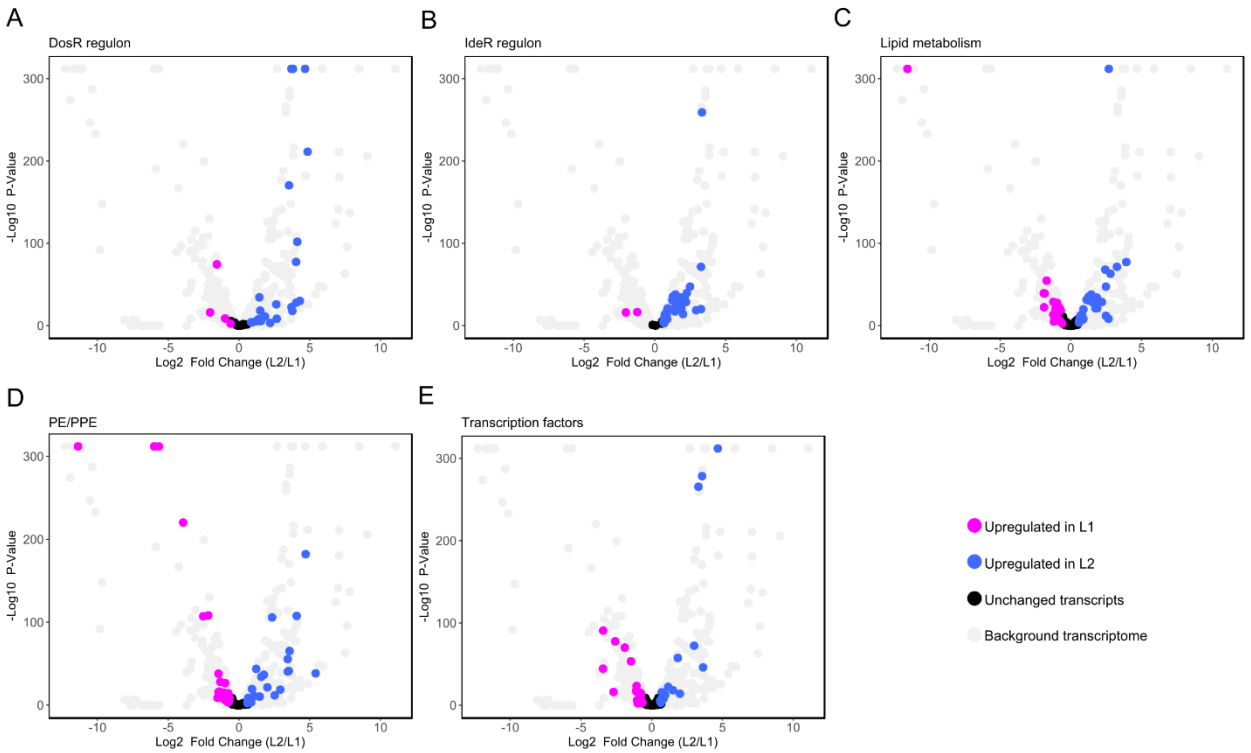

Figure S4. Enrichment Analyses of Mtb Functional Categories by mRNA data. Volcano plots display that **(A)** DosR regulon (p-value = 4.08E-14), **(B)** IdeR regulon (p-value = 3.05E-28), **(C)** lipid metabolism (p-value = 0.0025), and **(D)** PE/PPE (p-value = 9.3E-9) are significantly regulated between L1 and L2 clinical strains at the RNA level. **(E)** Regulated transcription factors between the two lineages on mRNA level. 34 (out of 211 quantified) transcription factors were significantly regulated in L2 strains in respect to L1. (see Figure 3 for protein data)

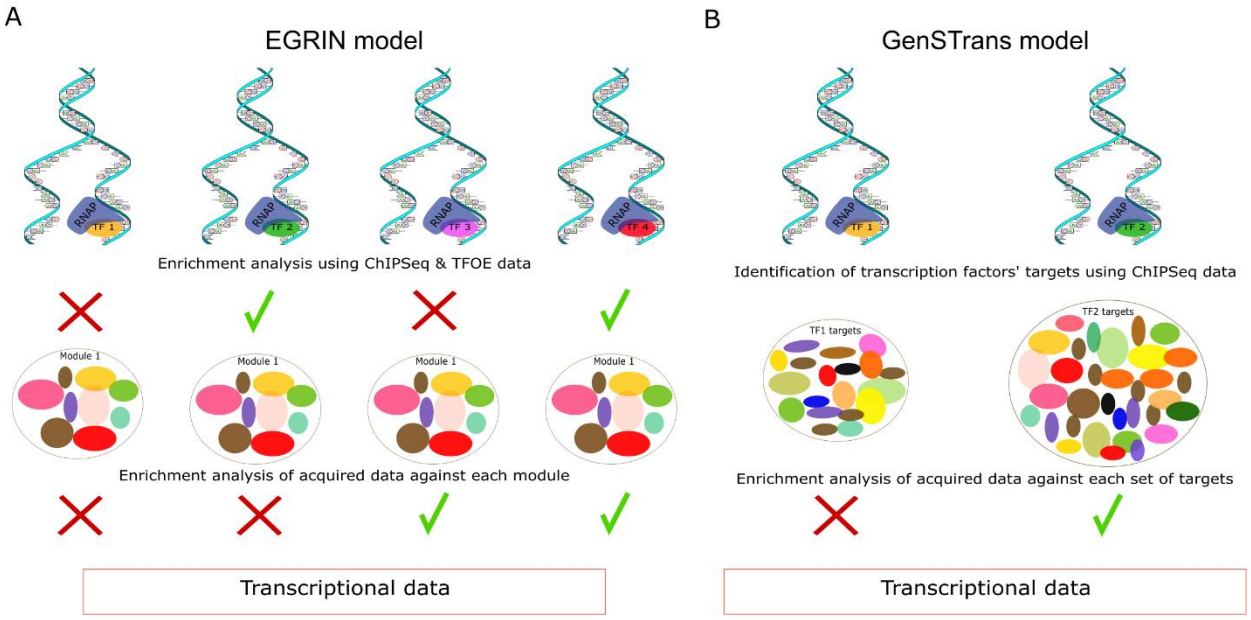

Figure S5. Conceptual Diagrams for Mtb Transcriptional Modeling. (A) Schematic illustration of the EGRIN model. The EGRIN model is a module centric approach, which demonstrates clusters of co-regulated genes (module) inferred from hundreds of RNA profiles that are publicly available. Next, each module was linked to transcription factor(s) using both Transcription Factor OverExpression (TFOE) (Rustad et al., 2014) and ChIPSeq (Minch et al., 2015b) experiments by enrichment analysis. Therefore, some of modules could not be corresponded to any transcription factor and remained as orphans. Eventually acquired RNA data for a given study should be analyzed against modules. **(B)** Concept of GenSTrans. GenSTrans was developed using ChIPSeq data (Minch et al., 2015b) and could be considered as a transcription factor centric model. This model explained sub-networks (instead of modules in the EGRIN model) which were reconstructed for each transcription factor. Therefore, each sub-network was linked to one and only one transcription factor. Finally, acquired RNA data could be analyzed through sub-networks.

A

Module	Pvalue	Transcription factor
182	4.81E-10	Rv3133c
446	6.39E-08	Rv0081, Rv1460
525	1.94E-06	Rv1773c
22	7.72E-05	Rv2359
530	0.0006	Rv0576, Rv2034, Rv2506
180	0.0006	Rv1990c
360	0.0011	Rv0022c
404	0.0035	Rv3133c
240	0.0131	Rv0757, Rv2506
406	0.0138	Rv3133c
152	0.0200	Rv3557c
380	0.0274	Rv0757, Rv1994c, Rv3765
181	0.0274	Rv3557c
78	0.0417	Rv1219c, Rv3160c
183	0.0447	Rv2250c, Rv3133c

Mutually identified TFs

B

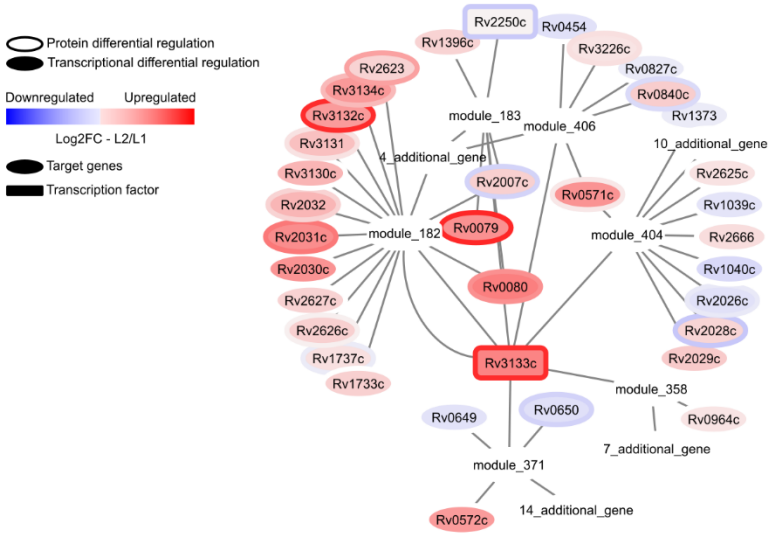

Figure S6. EGRIN Based Identification of Transcription Factors Regulating Their Target Modules Differently Between L1 and L2 Strains. (A) Identified modules and corresponding transcription factors by the EGRIN model. Green transcription factors were also identified by GenSTrans. **(B)** The transcription factor DosR that was identified through four modules significantly.

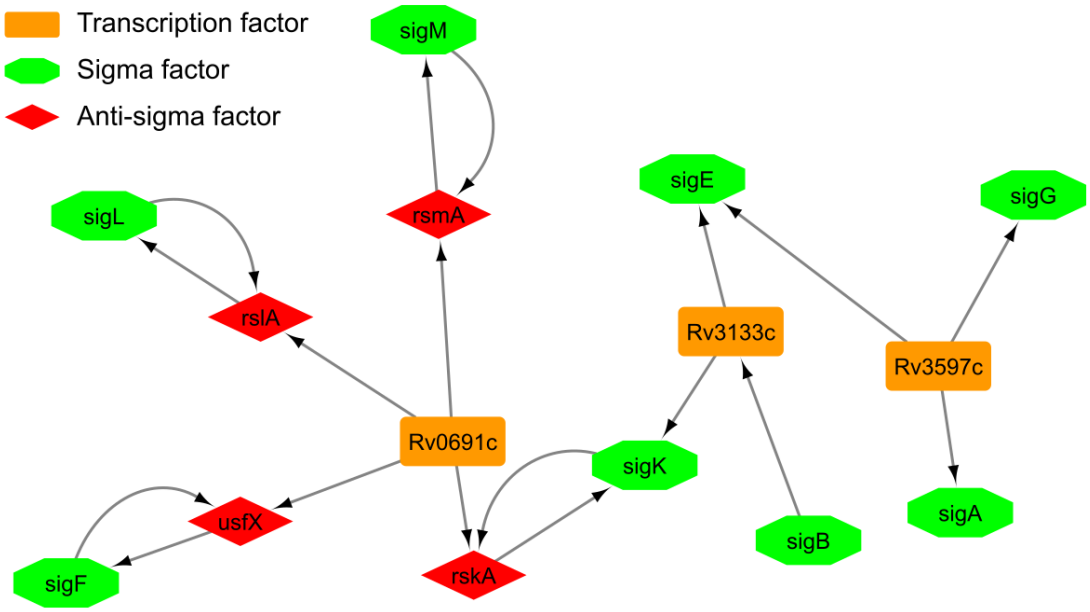

Figure S7. Interactions of Inter-L1 and L2 Master Transcription Factors with Sigma factors. Network shows that each of Rv3133c (DosR) and Rv3597c interact with three sigma factors whereas Rv0691c could regulate four anti-sigma factor and therefore modulating sigma factors indirectly.