Lineage-Specific Proteomic Signatures in the *Mycobacterium tuberculosis* Complex Reveal Differential Abundance of Proteins Involved in Virulence, DNA Repair, CRISPR-Cas, Bioenergetics and Lipid Metabolism

Solomon Abebe Yimer[1,2]*, Shewit Kalayou[3,4], Håvard Homberset[1], Alemayehu Godana Birhanu[1,3,5], Tahira Riaz[1], Ephrem Debebe Zegeye[6], Timo Lutter[1], Markos Abebe[7], Carol Holm-Hansen[8], Abraham Aseffa[7] and Tone Tønjum[1,3]

[1] Unit for Genome Dynamics, Department of Microbiology, University of Oslo, Oslo, Norway, [2] Coalition for Epidemic Preparedness Innovations, Oslo, Norway, [3] Division of Laboratory Medicine, Department of Microbiology, Oslo University Hospital, Oslo, Norway, [4] International Centre of Insect Physiology and Ecology, Nairobi, Kenya, [5] Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia, [6] NORCE Norwegian Research Centre AS, Centre for Applied Biotechnology, Bergen, Norway, [7] Armauer Hansen Research Institute, Addis Ababa, Ethiopia, [8] Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway

Despite the discovery of the tubercle bacillus more than 130 years ago, its physiology and the mechanisms of virulence are still not fully understood. A comprehensive analysis of the proteomes of members of the human-adapted *Mycobacterium tuberculosis* complex (MTBC) lineages 3, 4, 5, and 7 was conducted to better understand the evolution of virulence and other physiological characteristics. Unique and shared proteomic signatures in these modern, pre-modern and ancient MTBC lineages, as deduced from quantitative bioinformatics analyses of high-resolution mass spectrometry data, were delineated. The main proteomic findings were verified by using immunoblotting. In addition, analysis of multiple genome alignment of members of the same lineages was performed. Label-free peptide quantification of whole cells from MTBC lineages 3, 4, 5, and 7 yielded a total of 38,346 unique peptides derived from 3092 proteins, representing 77% coverage of the predicted proteome. MTBC lineage-specific differential expression was observed for 539 proteins. Lineage 7 exhibited a markedly reduced abundance of proteins involved in DNA repair, type VII ESX-3 and ESX-1 secretion systems, lipid metabolism and inorganic phosphate uptake, and an increased abundance of proteins involved in alternative pathways of the TCA cycle and the CRISPR-Cas system as compared to the other lineages. Lineages 3 and 4 exhibited a higher abundance of proteins involved in virulence, DNA repair, drug resistance and other metabolic pathways. The high throughput analysis of the MTBC proteome by

super-resolution mass spectrometry provided an insight into the differential expression of proteins between MTBC lineages 3, 4, 5, and 7 that may explain the slow growth and reduced virulence, metabolic flexibility, and the ability to survive under adverse growth conditions of lineage 7.

INTRODUCTION

Despite the discovery of *Mycobacterium tuberculosis* (Mtb) more than 130 years ago, Mtb physiology and the mechanisms of virulence are still not fully understood. More than 10.4 million individuals develop active tuberculosis (TB) globally every year, causing 1.3 million deaths annually (World Health Organization [WHO], 2019). The high morbidity and mortality associated with disease caused by Mtb reflects challenges in detecting and diagnosing TB at early stages, as well as poor therapeutic options. The urgent need for improved diagnostics, chemotherapeutics and vaccines for TB is well recognized. A better understanding of the biology of Mtb may expedite the development of the new and innovative tools in demand.

Mtb is a facultative intracellular organism that can adapt to the human host and exerts virulence and immune surveillance (Barth et al., 2013). A better understanding of the mechanisms by Mtb evades the host defense system may contribute to the development of novel therapeutics for TB.

Even though the major cause of TB in humans is Mtb, there are also other members of the human-adapted *Mycobacterium tuberculosis complex* (MTBC) that are responsible for TB. These include the five human-adapted lineages representing *M. tuberculosis* sensu stricto (lineages 1–4 and lineage 7), and lineage 5 and 6 which are commonly named as *M. africanum*. It also included some of the nine animal-adapted lineages already described earlier (O'Neill et al., 2019; Bottai et al., 2020; Coscolla et al., 2020). In addition, a recent study revealed the discovery of a human-adapted Mtb lineage 8 (Ngabonziza et al., 2020). Therefore, high quality genomic and proteomic data from human-adapted MTBC lineages-infected cells are warranted.

The MTBC lineages are classified as ancestral and "modern." The genomes of the "modern" lineages 2 (Beijing), 3 (CAS/Delhi) and 4 (Haarlem) have undergone a large deletion known as TbD1. Lineages 2, 3, and 4 are associated with major TB epidemics (Firdessa et al., 2013; Yimer et al., 2013; Bottai et al., 2020) and have higher transmission rates than lineages 1 (Indo Oceanic), 5 (*M. africanum* 1) and 6 (*M. africanum* 2) (Chacon-Salinas et al., 2005), and lineage 7 (Yimer et al., 2015). Lineage 3 has a higher anti-inflammatory phenotype than lineage 4 (Portevin et al., 2011). "Modern" Mtb lineages induce less of an early inflammatory response than lineages 1 and 6 (Portevin et al., 2011). Strains of *M. africanum* (lineage 6) develop drug resistance less frequently (Tientcheu et al., 2016), while lineage 2 (Beijing) acquires drug resistance more frequently (Nguyen et al., 2016). In order to elucidate unique and shared phenotypes and levels of virulence of MTBC lineages, high resolution proteomic data providing direct and highly detailed functional information

about cellular metabolism is imperative (Kelkar et al., 2011; Jena et al., 2013; Bespyatykh et al., 2016; Jhingan et al., 2016; Banaei-Esfahani et al., 2017; Schubert et al., 2018).

As part of our ongoing TB research with strains from the Horn of Africa, several genomics and proteomics studies have been conducted to better understand the clinical relevance of Mtb lineage 7. Mtb lineage 7 strains appear to grow more slowly *in vitro* than other Mtb lineages (Yimer et al., 2015). Patients infected with Mtb lineage 7 were associated with long delays in seeking health care (Yimer et al., 2015). Mtb lineage 7 cells frequently carry a high number of mutations in genes involved in DNA repair, replication and recombination, carbohydrate transport and metabolism, and energy production and conversion (Yimer et al., 2016). Compared to H37Rv, Mtb lineage 7 exhibits reduced abundance and hypoacetylation of proteins involved in growth and virulence (Birhanu et al., 2017; Yimer et al., 2017). However, available genomic and proteomic data do not explain the observation that lineage 7 is associated with a high rate of TB transmission, despite an apparent slower rate of growth (Yimer et al., 2015). To address the unique features of different MTBC lineages, we undertook a comprehensive proteomic study of MTBC lineages 3, 4, 5, and 7, focusing on proteins relevant to lipid metabolism, virulence, DNA repair, oxidative phosphorylation, CRISPR-Cas systems and drug resistance.

The differential expression of 3092 MTBC proteins, covering 77% of the predicted MTBC proteome, were included in this study. Principal components analysis, pathway clustering and other bioinformatics methods were used to identify unique protein signatures and potentially lineage-specific protein-protein interactions. Proteins involved in the TCA cycle and CRISPR-Cas system were more abundant in lineage 7 than in other lineages. The significance and implications of the results for MTBC lineages metabolism and pathogenesis are discussed.

MATERIALS AND METHODS

Cell Culture, Lysis and In-Gel Digestion

Strains from MTBC lineages 3 (ID: Sol 101), 4 (ID: FSP 471) and 5 and two strains from lineage 7 (ID: L735 and ID: L728) were cultivated and harvested as previously described (Yimer et al., 2017). The MTBC isolates were cultured on Middlebrook 7H10 agar plates in triplicate for 32 days, harvested, rinsed in PBS, and heat-inactivated (80°C for 90 min). Cell pellets were resuspended in lysis buffer; 10 mM Tris-HCl (pH 7.5), 2% SDS, 1 tablet per 50 mL EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich), 1 tablet per 10 mL PhosSTOP Phosphatase Inhibitor

Cocktail (Roche). The samples were transferred to Lysing Matrix B tubes (Roche) and mechanically disrupted using a bead beater (MagNa Lyser, Roche Diagnostics). The lysates were clarified by centrifugation (15,000 × g for 15 min) at 21°C, and the supernatant was transferred to 2 mL screw-cap micro-tubes (Sarstedt, Nümbrecht, Germany).

Protein samples (100 µg) were precipitated overnight by dilution with four volumes of ice-cold acetone. Precipitated proteins were centrifuged at 15,000 × g for 15 min at 4°C. The supernatant was aspirated and the protein pellet air-dried. Protein pellets were re-solubilized in NuPAGE LDS sample buffer (4x) (Life Technologies, United States) and reduced using 10 mM DL-dithiothreitol. Solubilized samples were incubated for 10 min at 70°C and proteins size-separated by SDS-PAGE (1.0 mm, 4–12% NuPAGE Novex Bis-Tris SDS-PAGE gel [Life Technologies]) at 80 V for 5 min followed by 20 min at 200 V. Gels were stained with colloidal Coomassie blue according to the manufacturer's instructions. Stained protein lanes were excised from the gel and cut into six pieces using a scalpel. The gel pieces were transferred to a micro-centrifuge tube and destained sequentially with 500 µl 50% propanol, 500 µl 100% propanol and then centrifuged at 600 rpm at room temperature. In-gel reduction, alkylation, and tryptic digestion were performed as previously described (Shevchenko et al., 2006), and customized for Mtb (Yimer et al., 2017).

Prior to alkylation, 100 µL of 10 mM DL-dithiothreitol in 100 mM ammonium bicarbonate (ABC) was added to the samples and incubated at 56°C for 1 h. Samples were alkylated in 55 mM iodoacetamide (Sigma-Aldrich, Cleveland, OH, United States) in 100 mM ABC for 1 h at room temperature in the dark. Gel pieces were washed sequentially with propanol and thereafter dehydrated under vacuum (Eppendorf, Speedvac concentrator 5301). Proteins were digested with sequencing-grade trypsin (1 µg trypsin/50 µg protein, Promega, WI, United States) for 16 h at 37°C in 50 mM ABC. Peptides were extracted from the gel by two-fold sequential treatment with 50% and 100% acetonitrile (ACN). Trypsin was inactivated by addition of 10% trifluoroacetic acid (TFA). Samples were dehydrated under vacuum as above and re-suspended in 0.05% TFA. Peptides were desalted using C18 stage tips (Rappsilber et al., 2003) with minor modifications (Yimer et al., 2017). Samples were resuspended in 20 µL 0.1% formic acid (FA), transferred to nano-liquid chromatography vials and stored at −20°C until further analysis.

Liquid Chromatography and Mass Spectrometry

Mass spectrometry (MS) analysis was performed using a Q Exactive plus hybrid quadrupole-orbitrap instrument coupled with an EASY-nLC 1200 UHPLC system (Thermo-Fisher Scientific, Biberach, Germany). Samples (6 µL) were injected onto C_{18} Acclaim PepMap columns (3 µm particles, 25 cm × 75 µm ID), followed by PepMap RSLC C_{18} (2 µm particles), both from Thermo Fisher Scientific (Biberach, Germany). Separation was performed in a 65 min solvent gradient (0.1% FA in 80/20 acetonitrile/water to 0.1% FA) at

0.3 µL/min at 40°C. MS was performed in the data-dependent acquisition mode with automatic switching between MS and MS/MS scans of peptide precursors from 380 to 1500 m/z with a 3 × 10^6 ion count target and a maximum injection time of 50 ms. The full MS scans were acquired at 70,000 resolution. The automatic gain control (AGC) target was 3 × 10^6 ions with maximum injection time of 50 ms. The AGC target for MS/MS was 10^1 and maximum injection time 100 ms with dynamic exclusion of 30 s.

Protein Identification and Quantification

Raw tandem MS/MS data were analyzed using MaxQuant (MQ) version 1.4.0.5 (Cox and Mann, 2008) and the peptide database search engine Andromeda (Cox et al., 2011) to search the Mtb H37Rv reference proteome UP000001584 (3993 entries, accessed in July 2017, UniProt-proteome), reverse decoy databases (Elias and Gygi, 2007), and select laboratory contaminants. Enzyme specificity as trypsin without proline with a maximum of two missed cleavages and carbamidomethylation of cysteine as a fixed modification, was used for peptide identification. Acetylation of the N-terminus, conversion of N-terminal glutamine and glutamic acid to pyroglutamic acid, and oxidation of methionine were included as variable modifications. The first search for precursor ions was performed with a mass tolerance of 20 ppm for calibration, while 6 ppm was applied for the main search. For protein identification, at least 1 unique peptide was required per protein group. A minimum peptide length of 7 amino acids was required for identification. To increase the number of peptides that can be used for quantification, the "requantify" and "match between runs" features of MQ were utilized with a retention time alignment window of 3 min. The default global Label-free Quantification (LFQ) algorithm was enabled (Cox et al., 2014) with the minimum LFQ ratio count for the quantification of a protein set to 1. In addition, unique peptides were used for peptide quantification. The maximum false discovery rate (FDR) cutoff of 0.01 (1%) was set at both the peptide spectra matches and the protein group levels.

Experimental Design

Protein samples (100 µg) were analyzed by SDS-PAGE. To reduce sample complexity and improve comprehensiveness of protein profile (Jafari et al., 2012), each biological experiment was fractionated into six gel bands, resulting in a total of 72 analytical runs (four lineages * three biological experiments * six fractions). Two categorical groupings were formed to represent (1) lineage identifier and (2) biological experiment ID. Analysis of quantitative changes in protein abundance was determined by the label-free algorithm.

Bioinformatics and Statistical Analysis

Data filtration, transformation, normalization, visualization, exploratory analysis and statistical analysis were performed in the Perseus platform (Version 1.5.5.3) (Tyanova et al., 2016). LFQ intensities from «Protein Group» output of

---

[1]https://www.geneious.com

MQ were loaded into the Perseus software. Filters were applied to remove protein contaminants, reverse database identifications and proteins identified only by their modified site. LFQ intensity values were subjected to log2 transformation. Quantitative accuracy of LFQ intensities in the biological experiment was assessed by Pearson correlation coefficient and the distribution visualized by histogram (**Supplementary File S1**). Principal component analysis was performed on protein abundances to determine the inter-lineage variance. Venn diagrams were plotted in the Venn diagram tool[2] (Heberle et al., 2015).

For statistical analysis, at least two valid LFQ intensities out of the three biological experiments were required. Missing values were imputed with random numbers from a normal distribution (parameters Width = 0.3, Shift = 1.8) (Hubner et al., 2010). Multiple-sample analysis of grouped biological replicates (permutation-based FDR cutoff of 0.01, S_0 = 2.) was performed to identify significantly changed proteins across the different lineages. Welch's *t*-test with an FDR value of 0.01 and S_0 = 2 was applied to identify proteins when the abundance was significantly changed between any of two MTBC lineages.

Analysis of GO annotation enrichment terms of the significant proteins were performed as previously described (Birhanu et al., 2017). Briefly, a two-tailed Fisher's exact test was used to assess the significance of enrichment terms. Proteins assigned to enriched term categories (*p*-value <0.05) were grouped according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification and Tuberculist's protein functional category (Kapopoulou et al., 2011). To identify molecular complexes in the large protein interaction network, MCODE clustering algorithm based on Bader and Hogue (2003) was used. The complete experimental workflow applied including in-gel digestion image is depicted in **Figure 1A**.

The Search Tool for the Retrieval of Interacting Genes version 10.0 (Franceschini et al., 2013; STRING[3]) was used to interpret the biological relevance of DA proteins in terms of predicted protein-protein interaction networks. Predicted interaction networks were finally visualized using the Cytoscape software[4] (Shannon et al., 2003).

Multiple-Genome Alignment Analysis

GenBank files of MTBC L1, L3, L4 (including H37Rv), L5 and L6 were downloaded from NCBI. For lineage 2 (ID no. DA 3 and DA 10) and lineage 7 (ID no. L735 and L728), Illumina MiSeq raw data was quality assessed with FastQC, quality- and adapter-trimmed with BBDuk, *de novo* assembled with SPAdes, and contigs mapped against H37Rv with Bowtie2. Analysis of multi-genome alignment of all lineages focusing on CRISPR-Cas locus was performed using the Geneious Prime software (Geneious Prime 2020.1[4]).

---

[2]http://www.interactivenn.net/

[3]http://string-db.org/

[4]https://cytoscape.org/index.html

Data Handling

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD020383.

Immunoblotting

A sample of 15 μg from each lysate was separated on a SDS-PAGE gel and transferred by blotting to a PVDF membrane. The membrane was blocked with 5% dry skim milk in TBS-T (10 mM Tris pH = 7.5, 150 mM NaCl, 0.05% TWEEN 20), incubated with ESAT-6 monoclonal antibody (11G4) from Thermo Fisher Scientific or GroEL2 mouse monoclonal antibody (BDI578) from OriGene diluted 1:1000 in TBS-T with 5% skim milk, and further incubated with HRP conjugated secondary antibodies (Santa Cruz SC-516102) diluted 1:5000 in TBS-T with 5% milk. Washing with TBS-T was done in between each step. The blot was photographed with luminescence detection using ECL[TM] Prime (GE Healthcare) and ChemiDoc Touch (Bio-Rad Laboratories). A stain-free image (Bio-Rad Laboratories) of the blot was taken to ensure equal loading and transfer to membrane.

Ethical Approval

The study obtained ethical approval from the Regional Committee for Medical Research Ethics in Eastern Norway (REK Øst) and the Ethiopian Science and Technology Ministry in Addis Ababa, Ethiopia. Written informed consent was obtained from the participants who provided the Mtb samples before the study was conducted.

RESULTS

Protein Identification

A total of 38,346 unique MTBC peptides were identified, corresponding to 3,092 proteins that represent 77% coverage of the predicted MTBC lineage proteome. More than 81% (2522/3092) of the proteins were detected in all four lineages (**Supplementary File S2**). The spectral quality analysis from each biological replicate is presented in a histogram (**Figure 1A**). The total number of proteins identified in MTBC lineages 5, 7, 3, and 4 were 2807, 2806, 2796, and 2722, respectively (**Figure 1B**).

The proteomic data were analyzed using hierarchical clustering analysis (HCA) (**Figure 1C**) and principal component analysis (PCA) (**Figure 1D**). For HCA, LFQ intensities were used to calculate Pearson correlations (*R*-values) for each biological replicate (**Figure 1C**) and the resulting *R*-values were sufficiently high for reliable comparisons of protein abundances. Using proteins with 100% PCA valid values, the four lineages clustered together thus confirming data reproducibility. Considerable variation was detected among all the lineages (**Figure 1D**). Proteins expressed uniquely or in common between biological replicates and MTBC lineages are summarized in the Venn diagram (**Figure 2**).

FIGURE 1 | (A–D) Proteomics analysis of *M. tuberculosis* complex (MTBC) lineages. (A) Experimental workflow used in this study includes lysis, protein precipitation, tryptic digestion, peptide desalting, and ultra-high-performance liquid chromatography followed by mass spectrometry on a Q exactive plus mass spectrometer (Thermo-Fischer, Inc.). Label free quantification and the downstream statistical analysis were performed using MaxQuant and Perseus software, respectively. (B) Overlap of differentially abundant proteins in MTBC Lineages 3, 4, 5, and 7. (C) iBAQ-based ranking of protein abundances in various lineages. (D) Principal component analysis for protein abundance data from MTBC lineages. Samples were separated by lineage (component 1, 43.9% variance) and by biological groups (component 2, 26.6% variance). Only proteins identified in all samples were considered for the analysis. Data from three biological replicates are represented.
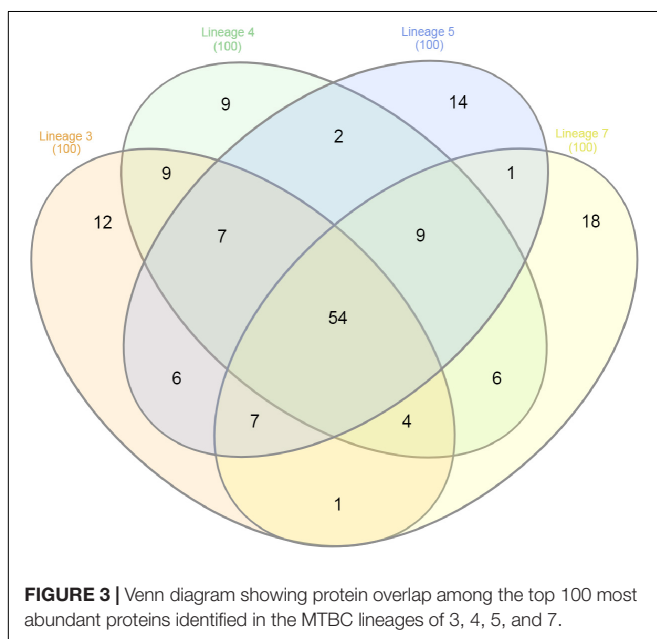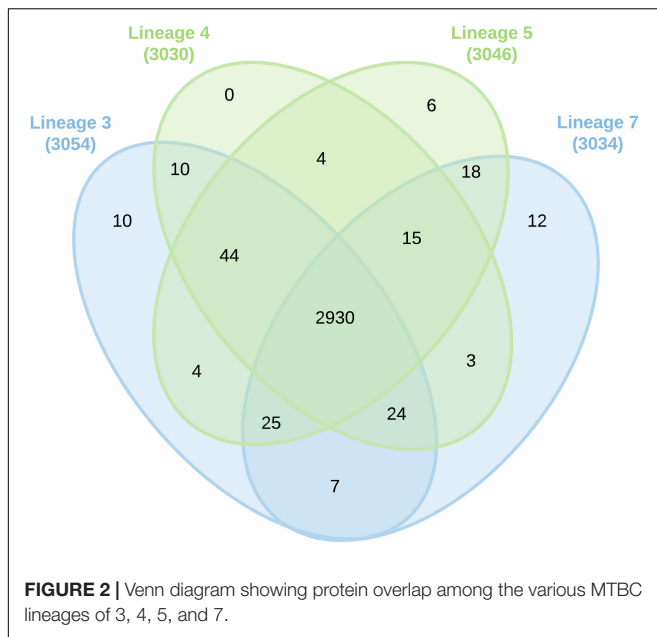
Dynamic Range of Protein Abundance

The iBAQ intensity data were used to rank the most abundant and least abundant proteins in each MTBC lineage. The 10 most abundant protein groups in lineage 7 included chaperones, kinase, acyl carrier and elongation factor, hemagglutinin and DNA ligases, and the least abundant proteins included oxidoreductases, ATP-binding protein, aminotransferase, nucleotide transferases and other uncharacterized proteins. The most abundant protein groups in lineage 4 comprised chaperones, acyl carrier, ESAT-6 like protein (ExsB), hydrolase, methyltransferase, and antioxidant proteins, and the least abundant protein groups included acyltransferase, nucleotide transferase, DNA-binding proteins and oxidoreductases. The iBAQ intensities in all proteomes covered four orders of magnitude (comparing the most and least abundant proteins) (**Supplementary File S3**).

The 100 most abundant proteins in each lineage were also analyzed; 9, 12, 14, and 18 proteins were unique to lineages 3, 4, 5, and 7, respectively (**Figure 3**). EsxA and EsxB proteins were more abundant in lineage 4 than other lineages, while the EsxB protein showed lowest abundance in lineage 7 (**Figure 4**). Tuberculist

was used to assign functional categories for the top 100 most abundant proteins for each lineage (**Supplementary File S4**). Lineages 5 and 7 exhibited higher abundance of proteins involved in intermediary metabolism and respiration than lineages 3 and 4. Proteins involved in lipid metabolism showed higher abundance in lineage 7 than other lineages. Cell wall and cell process-related proteins were more abundant in lineage 4 than in other lineages.

Metabolic Proteins

Welch's *t*-test revealed the proteins that were differentially abundant between the MTBC lineages (**Supplementary Files S5–S10**). The most distinct difference was observed between lineages 3 and 7 in which 221 proteins were differentially expressed; 138 were more abundant and 83 were less abundant (**Supplementary File S5**). The second highest difference was observed between lineages 4 and 7 that showed 177 differentially expressed proteins of which 97 were abundant upregulated and 80 less abundant (**Supplementary File S6**). Less difference was observed between lineages 3 and 4 where among the 87 differentially abundant proteins, 49 were more abundant and 38 less abundant (**Supplementary File S7**).

FIGURE 2 | Venn diagram showing protein overlap among the various MTBC lineages of 3, 4, 5, and 7.

FIGURE 3 | Venn diagram showing protein overlap among the top 100 most abundant proteins identified in the MTBC lineages of 3, 4, 5, and 7.

ANOVA was used to identify the 539 differentially abundant proteins (**Supplementary Files S11, S12**) that were further analyzed by unsupervised HCA to yield eight clusters/categories of differentially abundant proteins. The largest four clusters comprised 170, 124, 76, and 55 proteins. The major functional categories associated with these clusters include DNA repair and recombination, lipid metabolism, energy metabolism, ESX-1 and ESX-3 secretion system (**Supplementary File S11**). Gene ontology analysis among the lineages was performed to predict biological effects. Pathway sub-clusters that were significantly enriched are listed in **Supplementary File S11**. A subset of differentially expressed proteins were assigned to

functional categories, including intermediary metabolism and respiration (23.01%), conserved hypothetical proteins (20.9%), lipid metabolism (16.3%), and cell wall and cell processes-related (19.2%) (**Supplementary File S11B**).

Analysis using the online STRING protein query database suggested potential protein-protein interactions involving 536 proteins resulting in 3021 putative interactions (e.g., protein network). Potentially important protein hubs involving RecA, FaS, FabD, PpsD, PpsC, EccB3, and EccC3 were identified (**Figures 5A,B**). Six network sub-clusters, playing potential roles in type VII ESX3-secretion, DNA repair, inorganic phosphate uptake, CRISPR-Cas system, lipid metabolism, bioenergetics, and energy metabolism were identified.

MTBC Type VII Secretion Systems ESX-1 and ESX-3 Are Differentially Expressed

The ESX-3 secretion component proteins including EccA3, EccB3, EccC3, EccD3, and EccE3 proteins were less abundant in lineage 7 than in other lineages (**Table 1**). The ESX-3 component proteins identified showed protein-protein interactions as shown in **Figure 5B**. The differentially expressed ESX-1 component proteins included EsxA, EsxB, EsxG, EccB, EspA, EspC, and EspD (**Table 2**). The EsxA protein was more abundant in lineage 4 than in lineages 3 and 5, as corroborated by immunoblotting, independent of biological or technical replicates (**Figure 6** and **Supplementary File S13**). The EsxB proteins were less abundant in lineages 7 (**Figure 6**). The EspD and EspC proteins required for ESAT-6 secretion were more abundant in lineage 7 than in other lineages (**Table 2**).

Other virulence-related proteins including the mammalian cell entry family proteins (Mce1F, Mce1A, Mce1B and Mce1D) were more abundant in lineage 4. The anti-apoptotic component protein, Mpt64, and the type II toxin-antitoxin (TA) system (VapB10 and VapC10) were less abundant in lineage 7 (**Table 2**). Data on clusters 511, 515 and 528 suggest differential expression of type VII ESX-3 and ESX-1 secretion system/virulence-related proteins and other proteins (**Supplementary File S11**).

DNA Repair and Recombination

Proteins that play essential roles in DNA recombination and repair (RecA, RuvA, RuvB, RecR, and UvrC) (Dos Vultos et al., 2009) were less abundant in lineage 7 than other lineages (**Table 3**). Low abundance in lineage 7 included DnaE1 (DNA replication proofreading), FtsK (DNA conjugation, DNA unwinding), NrdF2 (deoxyribonucleotide biosynthesis), DnaJ2 (heat shock, hyperosmotic shock), Rv0043c, Rv0795, NucS and CspA (DNA binding), and NadA, NadB, and NadC (DNA synthesis precursors). The majority of DNA repair proteins form protein–protein interactions and form a hub with RecA (**Figure 7A**). Cluster 530 also includes many DNA repair proteins (**Supplementary File S11**).
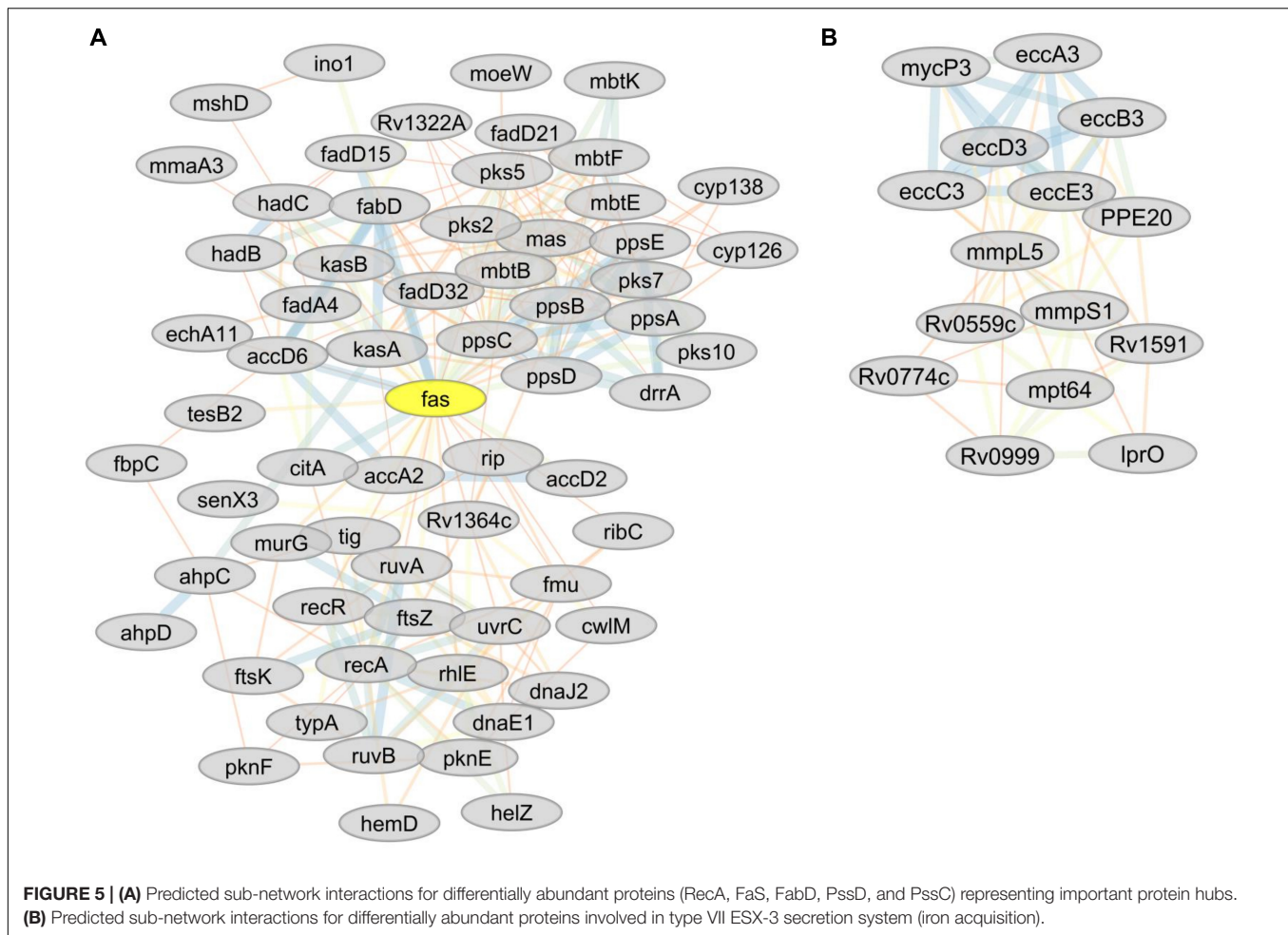
Lipid Metabolism

Proteins involved in mycolic acid and phthiocerol dimycocerosate (PDIM) biosynthesis including FaS, FadD15,

FIGURE 4 | Boxplot showing difference in EsxA and EsxB proteins abundance in the various MTBC lineages.
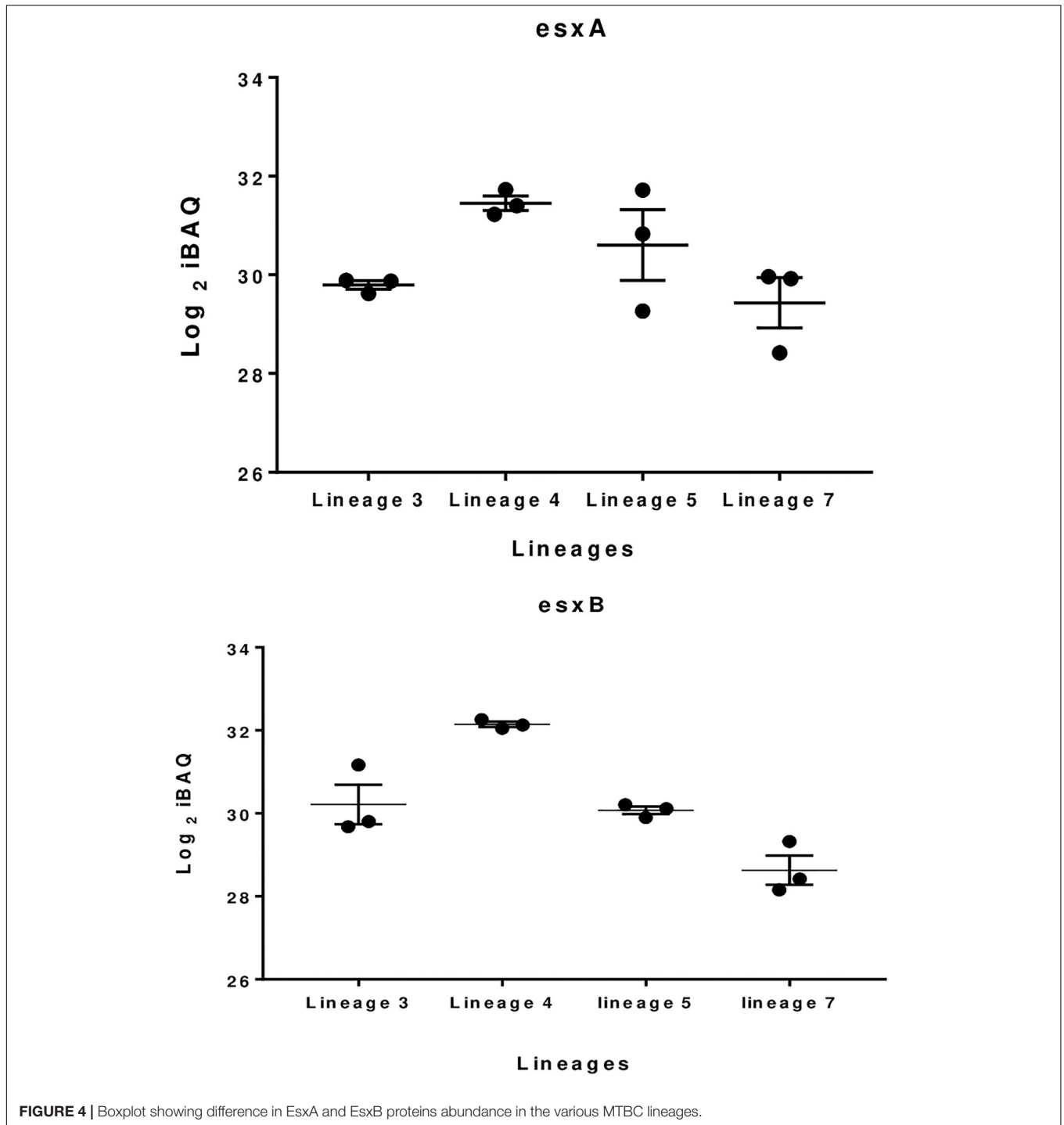
EchA11, HadB, MmaA3, and Pks13 were less abundant in lineage 7 (**Table 4**). Fas-II component proteins (KasA, KasB, and HadB), carboxylases (AccA2, AccD2, and AccD6) (Duan et al., 2014), and RV2006 involved in trehalose biosynthesis (Kapopoulou et al., 2011) were less abundant in lineages 4 and 7. Fas forms an important interaction hub (**Figure 5A**).

Proteins required for PDIM/phenolglycolipid (PGL) biosynthesis including PpsA, PpsB, PpsD, PpsE, MaS, DrrA,

Pks7, FadA5, and FadD26 (Domenech and Reed, 2009) were less abundant in lineages 4 and 7. MaS and Rv2952 proteins involved in PGL synthesis (Domenech and Reed, 2009) were less abundant in lineage 7. FadE28 and Cyp125 that catalyze the activation of long-chain fatty acids, LppX that is involved in PDIM translocation, and AftB responsible for arabinogalactan biosynthesis (Kapopoulou et al., 2011) were more abundant in lineage 5. Rv2959c that catalyzes the O-methylation linked to the

FIGURE 5 | (A) Predicted sub-network interactions for differentially abundant proteins (RecA, FaS, FabD, PssD, and PssC) representing important protein hubs. **(B)** Predicted sub-network interactions for differentially abundant proteins involved in type VII ESX-3 secretion system (iron acquisition).

phenolic group of glycosylated PGL (Simeone et al., 2013) was more abundant in lineages 4 and 5.

Among proteins involved in cell wall peptidoglycan biosynthesis, MurE, MurF, FadD5, and RipA were more abundant in lineage 4 than other lineages. In addition, IniC and IniA proteins that are targets for peptidoglycan biosynthesis inhibitors were more abundant in lineages 3 and 4 than other lineages (**Table 4**). The heatmaps corresponding to clusters 530 and 515 depict the differential abundance of lipid biosynthesis-related proteins and other proteins (**Supplementary File S11**).

Energy Metabolism

Several proteins responsible for energy metabolism (glycolysis, TCA cycle, and oxidative phosphorylation) were differentially expressed (**Table 5**). PgI is involved in glycolysis (Kapopoulou et al., 2011), PfkA required for gluconeogenesis, the AdhA which catalyzes the reversible oxidation of ethanol for the reduction of NAD, and LldD1 involved in the conversion of lactate into pyruvate (Kapopoulou et al., 2011) were more abundant in lineage 7 than other lineages.

Among the proteins involved in the oxidative direction of the TCA cycle, CitA in lineage 7 and Icl in lineages 4 and 7 were

TABLE 1 | Differentially abundant proteins in the type VII ESX-3 secretion systems in MTBC lineages.

Protein name	Gene symbol	Rv number	Less abundance lineage/s	Increased abundance lineage/s
ESX-3 secretion system protein EccA3	*eccA3*	Rv0282	7	3, 4, 5
ESX-3 secretion system protein EccC3	*eccC3*	Rv0284	7	3, 4, 5
ESX-3 secretion system protein EccD3	*eccD3*	Rv0290	7	3, 4, 5
ESX-3 secretion system protein EccB3	*eccB3*	Rv0283	7	3, 4, 5
ESX-3 secretion system protein EccE3	*eccE3*	Rv0292	7	3, 4, 5
Membrane-anchored mycosin MycP	*mycp3*	Rv0291	7	3, 4, 5
Siderophore exporter MmpL5	*mmpl5*	Rv0675	7	3, 4, 5

TABLE 2 | Differentially abundant proteins in the ESX-1 secretion system and other virulence related proteins in MTBC lineages.

Protein name	Gene symbol	Rv number	Less abundance (lineage/s)	Increased abundance lineage/s
6 kDa early secretory antigenic target	*esxA*	Rv3873	3, 5	4, 7
ESAT-6-like protein EsxB	*esxB*	Rv3874	5, 7	3, 4
ESAT-6-like protein EsxG	*esxG*	Rv0287	5, 7	3, 4
ESX-1 secretion-associated protein EspA	*espA*	Rv3626c	3, 4	5, 7
ESX-1 secretion-associated protein EspC	*espC*	Rv3881c	3, 4, 5	7
7ESX-1 secretion-associated protein EspD	*espD*	Rv3614c	3, 4, 5	7
Putative antitoxin VapB10	*vapB10*	Rv1398c	5, 7	3, 4
Ribonuclease VapC15	*vapC10*	Rv1397c	5, 7	3, 4
PPE family protein PPE18	PPE18	Rv1196	7	3, 4, 5
PE family immunomodulator PE15	PE15	Rv1386	7	3, 4, 5
Uncharacterized PPE family protein PPE38	PPE38	Rv2352c	7	3, 4, 5
Cholesterol oxidase	*choD*	Rv3409c	7	3, 4, 5
Mce-family protein Mce1A	*mce1A*	Rv0169	3, 5, 7	4
Mce-family protein Mce1B	*mce1B*	Rv0170	3, 5, 7	4
Mce-family protein Mce1D	*mce1D*	Rv0172	3, 5, 7	4
Mce-family protein Mce1F	*Mce1F*	Rv0174	3, 5, 7	4
Immunogenic protein MPT64	*mpt64*	Rv1980C	7	3, 4, 5

less abundant than other lineages. Proteins involved in alternative pathways of the TCA cycle were differentially abundant, among which the PrpD and PrpC proteins required for methyl citrate cycle pathway, and the GnD1 and GnD2 (Kapopoulou et al., 2011) proteins involved in the pentose phosphate pathway were more abundant in lineages 5 and 7. Zwf2 was more abundant in lineage 7 than in other lineages.

The fumarate reductases (FrdA and FrdB) that play vital roles in the maintenance of an energized membrane during hypoxia (Watanabe et al., 2011) and GabD2, an essential protein for generating succinate from succinic semialdehyde (SSA), were more abundant in lineages 4 and 7. The interaction of these proteins is depicted in **Figure 7B**. In contrast, lineage 5 showed low expression of GltA2 and the succinate dehydrogenase enzymes (SdhA, SdhB, SdhD) compared to other lineages.

Among the proteins involved in oxidative phosphorylation (electron transport chain), lineages 4 and 7 showed increased abundance of Ndh. FadH was less abundant in both lineages 5 and 7, and Rv1812c was less abundant in lineage 7 than other lineages. A number of highly expressed proteins involved in energy metabolism showed direct protein-protein interaction (**Figure 7B**). The heatmap in cluster 511, 513, 523, and 528 shows the differential abundance of various proteins among which proteins involved in energy metabolism are also listed (**Supplementary File S11**).

Stress Responses Proteins

Several proteins that have critical roles in protection against metabolic stress were differentially regulated (**Table 6**). MshD, Tpx, AhpC and AhpD were less abundant in lineage 7 than in other lineages. The serine/threonine-protein kinases (PknE, PknF, Rv0043, SenX3) involved in regulatory functions were less abundant in lineage 7 than other lineages. All proteins involved in metabolic stress response are shown in the heatmaps

corresponding to clusters 530 together with other proteins (**Supplementary File S11**).

ABC Transporters

Ammonium bicarbonate transporter proteins PstS1, PstS2, and PstB that are essential for the high-affinity capture of periplasmic inorganic phosphate (Pi) in Mtb (Peirs et al., 2005) and synthesis of DNA molecules were less abundant in lineage 7 (**Table 3**). Proteins involved in copper homeostasis including CtpV, CtpG multi-copper oxidase (MmcO) and CsoR were more abundant in lineage 5 than other lineages. The protein-protein interaction network of the Pst system proteins is depicted in **Figure 7C**.
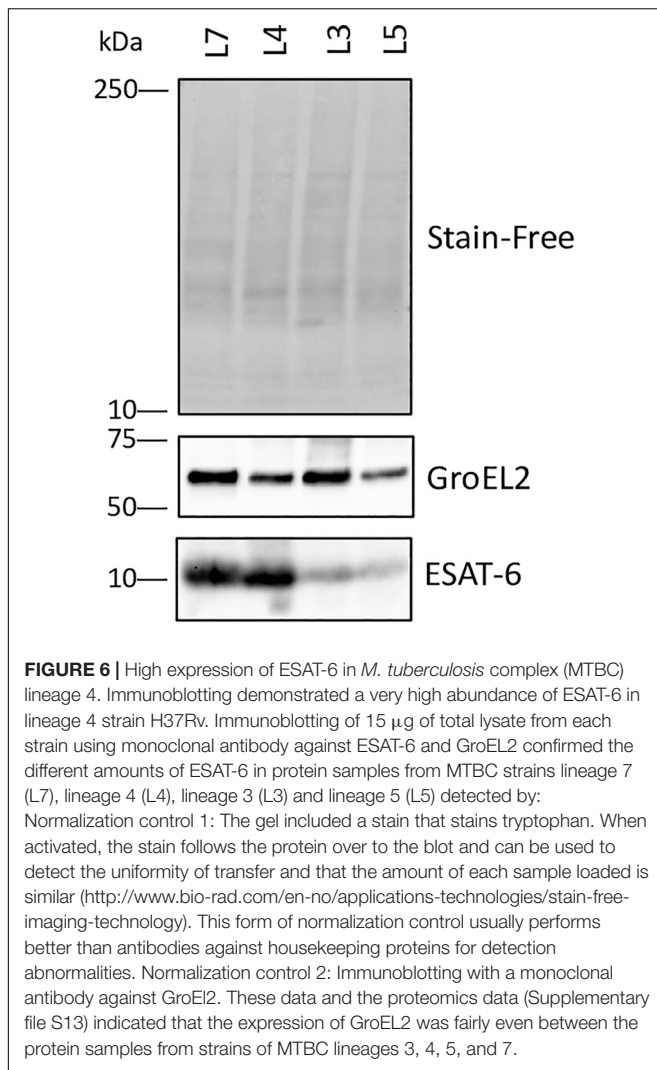
CRISPR-Cas Systems

Mtb lineage 7 harbors an abundance of CRISPR-Cas systems, and some of these components were expressed in an MTBC lineage-specific manner. Csm3 and Csm4 were more abundant in lineage 7, while the expression of the Cas6 was upregulated in lineages 4 and 7. Cas10, Csm3, Csm4, Cas6, and Cas10 proteins were also more abundant in lineage 7. Csm3, Csm4, Csm5, Cas6 exhibit pathways interaction with each other (**Figure 7D**), and are found in clusters 511, 528, and 529 (**Supplementary File S10**).

Multiple genome alignment analysis of the CRISPR-Cas locus revealed the presence of single nucleotide polymorphisms (SNPs) between the lineages. Notably, the number of SNPs observed in Csm3, Csm4, Csm5, and Cas6 in the two lineage 7 strains was reduced as compared to other lineages (**Figure 8**).

DISCUSSION

This study documented MTBC lineage-specific protein abundances in clinical strains of lineages 3, 4, 5, and 7. The coverage of the predicted MTBC proteome (77%) is similar to

FIGURE 6 | High expression of ESAT-6 in *M. tuberculosis* complex (MTBC) lineage 4. Immunoblotting demonstrated a very high abundance of ESAT-6 in lineage 4 strain H37Rv. Immunoblotting of 15 μg of total lysate from each strain using monoclonal antibody against ESAT-6 and GroEL2 confirmed the different amounts of ESAT-6 in protein samples from MTBC strains lineage 7 (L7), lineage 4 (L4), lineage 3 (L3) and lineage 5 (L5) detected by: Normalization control 1: The gel included a stain that stains tryptophan. When activated, the stain follows the protein over to the blot and can be used to detect the uniformity of transfer and that the amount of each sample loaded is similar (http://www.bio-rad.com/en-no/applications-technologies/stain-free-imaging-technology). This form of normalization control usually performs better than antibodies against housekeeping proteins for detection abnormalities. Normalization control 2: Immunoblotting with a monoclonal antibody against GroEl2. These data and the proteomics data (Supplementary file S13) indicated that the expression of GroEL2 was fairly even between the protein samples from strains of MTBC lineages 3, 4, 5, and 7.

that reported elsewhere (Bespyatykh et al., 2016; Jhingan et al., 2016; Yimer et al., 2016; Schubert et al., 2018).

The Esx-3 secretion system is essential for siderophore-mediated iron uptake and is critical for MTBC isolates growth *in vitro* (Houben et al., 2014). Mycobactin and carboxymycobactin are the main siderophore molecules in MTBC that are required for iron uptake (De Voss et al., 1999). Biosynthesis of mycobactin and carboxymycobactin requires MbtB, MbtE and MbtF proteins, and we observed that MbtE and MbtF were less abundant in lineage 7. In addition, IdeR, MmpL5 and MycP3 are involved in iron homeostasis, siderophore biosynthesis and secretion, and *in vitro* growth, respectively (De Voss et al., 1999; Pandey and Rodriguez, 2014; Siegrist et al., 2014). The decreased abundance of proteins involved in ESX-3 secretion suggests reduced uptake of iron, which may affect growth and virulence in lineage 7.

Several proteins involved in the ESX-1 secretion system were differentially abundant in a lineage-specific manner. MTBC requires ESAT6 and EsxB for full virulence (Forrellad et al., 2013; Houben et al., 2014). ESAT6 and EsxB facilitate the translocation

of Mtb from the phagosome into the host cell cytoplasm (van der Wel et al., 2007; Houben et al., 2012; Simeone et al., 2012; Jamwal et al., 2016). ESAT6 inhibits antigen-presentation and TLR signaling, IL-12 production and macrophage-dependent apoptosis (Pathak et al., 2007; Mtb with defects in secretion of EsxA and/or EsxB are not translocated into the cytosol and show decreased host cell lysis and cell-to-cell invasion (Guinn et al., 2004; Houben et al., 2012). EsxG inhibits phagosome maturation (Mehra et al., 2013). The increased abundance of EsxA, EsxB and EsxG in lineages 4 and 3 contribute to enhanced growth and increased virulence. These lineages are widely distributed globally (Coscolla and Gagneux, 2014).
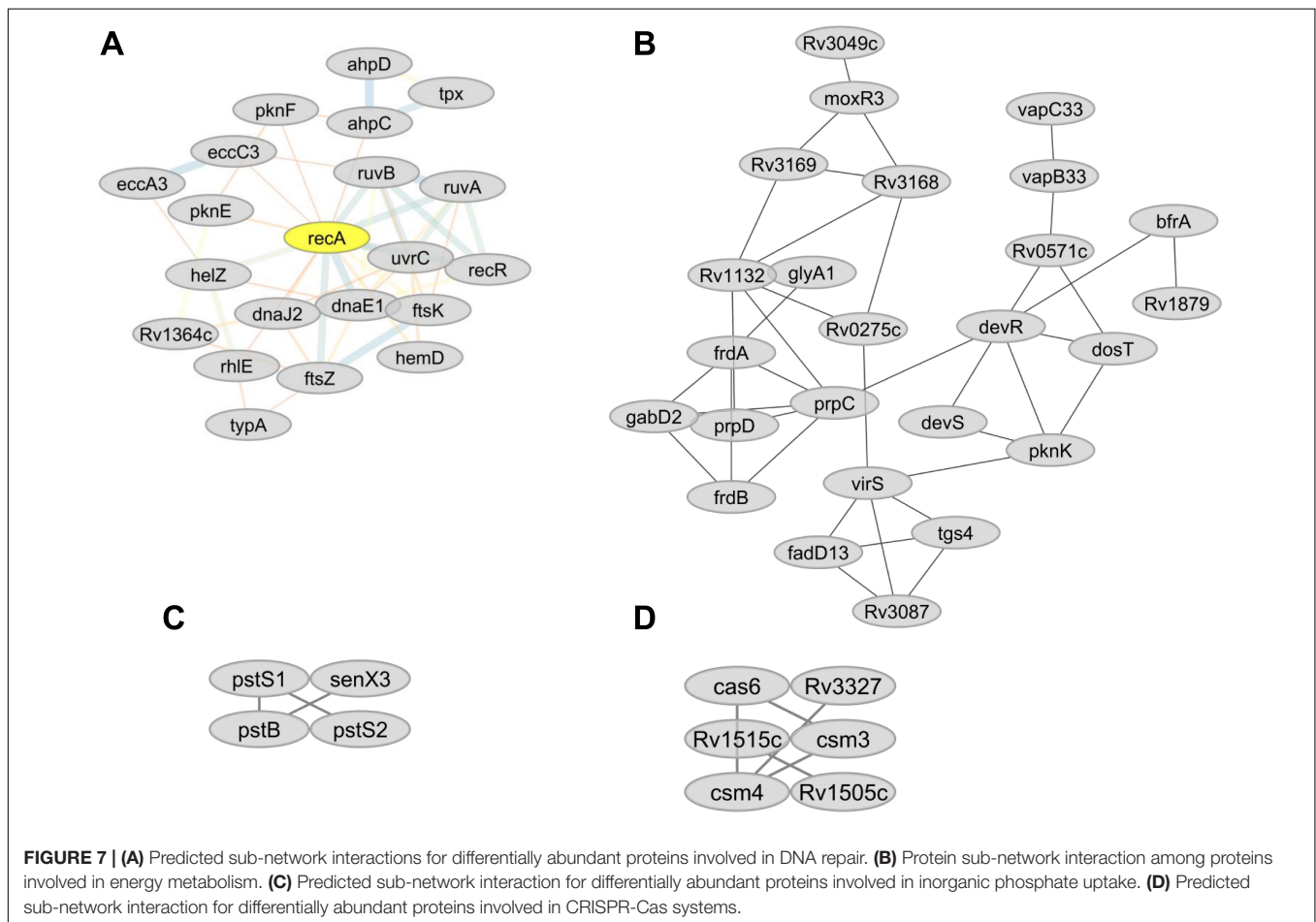
Efficient ESAT-6 secretion in Mtb is dependent on abundant expression of the EspA, EspB, EspC, and EspD proteins (Fortune et al., 2005; MacGurn et al., 2005). EspD stabilizes the cellular levels of EspA and EspC. The EspC is a potent antigen in both active and latent TB infection. The immunodominance of EspC is similar to ESAT6 as has been proposed to be a promising TB vaccine candidate (Millington et al., 2011). The high abundance of EspA, EspC, and EspD in lineages 5 and 7 may suggest a compensatory mechanism for the relatively low abundance of EsxA and EsxB observed in these two lineages.

The mammalian cell entry (Mce) proteins are surface-exposed secreted proteins and crucial virulence factors (Arruda et al., 1993). The abundance of Mce1 proteins in Mtb lineage 4 may contribute to the high virulence of this lineage. Mtb has anti-apoptotic capacities to counteract host microcidal activity. The abundance of PknE and Mpt64 were low in lineage 7. Mpt64 inhibits macrophage apoptosis *in vitro* (Wang et al., 2014). PknE prevents apoptosis by interfering with host signaling pathways (Jayakumar et al., 2008).

Several proteins that have critical roles in lipid biosynthesis (mycolic acid and PDIM) were differentially expressed. Mycolic acids are vital components of the MTBC cell wall contributing to inherent drug resistance, growth and virulence (Duan et al., 2014; Marrakchi et al., 2014). The synthesis of mycolic acids requires type I and type II fatty acid synthase (FAS). Fas-II involves KasA, KasB, HadABC, Rv0635, Rv0636, Rv0637, Rv0033, and the InhA proteins (Duan et al., 2014). Among the Fas-II proteins, KasA, KasB, InhA, HadB, Rv0033, and Pks13 were less abundant in lineages 4 and 7. KasA catalyzes the early steps of fatty acid elongation reaction, and KasB extends the elongation products to complete the length of fatty acid chains (Schiebel et al., 2013). Deletion of the *kasB* gene resulted in colony morphology defect, loss of acid fastness and subclinical latent TB in mice (Bhatt et al., 2007). InhA is required in the final step of fatty acid extension (Gurvitz et al., 2008). Pks13 catalyzes the condensation reaction to produce mycolic acids precursors (Duan et al., 2014). Mycolic acid biosynthesis also requires building blocks (acylCoAs) that are synthesized from carboxylases. The MTBC genome contains *accA1, accA2 and accA3* biotin carboxylase subunits and six carboxyltransferase subunits (*accD1–accD6*). *accD6* is part of the FAS-II complex genetic locus (37). Lineage 7 exhibited a lower abundance of AccA2, AccD2 and AccD6. RV2006, which is less abundant in lineage 7, is used in trehalose

TABLE 3 | Differentially abundant proteins involved in the DNA repair and Pst systems in MTBC lineages.

Protein name	Gene ID	Rv number	Less abundance (lineage/s)	Increased abundance (lineage/s)
Protein RecA	*recA*	Rv2737c	7	3, 4, 5
Holliday junction ATP-dependent DNA helicase RuvA	*ruvA*	Rv2593c	7	3, 4, 5
Holliday junction ATP-dependent DNA helicase RuvB	*ruvB*	Rv2592c	7	3, 4, 5
Recombination protein RecR	*recR*	Rv3715c	7	3, 4, 5
UvrABC system protein C	*uvrC*	Rv1420	7	3, 4, 5
DNA polymerase III subunit alpha	*dnaE1*	Rv1547	7	3, 4, 5
DNA translocase FtsK	*ftsk*	Rv2748c	7	3, 4, 5
Putative helicase HelZ	*helZ*	Rv2101	7	3, 4, 5
Ribonucleoside-diphosphate reductase subunit beta nrdF2	*nrdF2*	Rv3048c	7	3, 4, 5
Chaperone protein DnaJ 2	*dnaJ2*	Rv2373c	7	3, 4, 5
Uncharacterized HTH-type transcriptional regulator Rv0043c	*Rv0043c*	Rv0043c	7	3, 4, 5
Insertion element IS6110 uncharacterized 12.0 kDa protein	*Rv0795*	Rv0795	7	3, 4, 5
Phosphate-binding protein PstS 2	*pstS2*	Rv0932c	7	3, 4, 5
Phosphate-binding protein PstS 1	*pstS1*	Rv0934	7	3, 4, 5
Phosphate import ATP-binding protein PstB 1	*pstsB1*	Rv0933	7	3, 4, 5

FIGURE 7 | (A) Predicted sub-network interactions for differentially abundant proteins involved in DNA repair. **(B)** Protein sub-network interaction among proteins involved in energy metabolism. **(C)** Predicted sub-network interaction for differentially abundant proteins involved in inorganic phosphate uptake. **(D)** Predicted sub-network interaction for differentially abundant proteins involved in CRISPR-Cas systems.

biosynthesis. Trehalose is a major free sugar molecule which is an integral part of cell wall glycolipids (Murphy et al., 2005). Trehalose is crucial in mycolic acid transport during cell wall biogenesis.

The MTBC cell wall is composed of complex lipids such as PDIM in addition to mycolic acids. Proteins including the Pks1, Pk5, Pks7, FadD26, FadD28, and Pks10 that showed decreased abundance in lineage 7 are involved in PDIM

TABLE 4 | Differentially abundant proteins involved in lipid metabolism in MTBC lineages.

Protein name	Gene symbol	Rv number	Less abundance (lineage/s)	Increased abundance lineage/s
Fatty acid synthase Fas	*fas*	Rv2524c	7	3, 4, 5
Long-chain-fatty-acid–CoA ligase FadD15	*fasD15*	Rv2187	7	3, 4, 5
Probable enoyl-CoA hydratase EchA11	*EchA11*	Rv1141c	7	3, 4, 5
(3R)-Hydroxyacyl-ACP dehydratase subunit HadB	*hadB*	Rv0636	7	3, 4, 5
Polyketide synthase Pks13	*pks13*	Rv3800c	7	3, 4, 5
Methoxy mycolic acid synthase MmaA3	*mmaA3*	Rv0643c	7	3, 4, 5
3-Oxoacyl-[acyl-carrier-protein] synthase 1 KasA	*kasA*	Rv2245	7	3, 4, 5
3-Oxoacyl-[acyl-carrier-protein] synthase 2	*ksaB*	Rv2246	7	3, 4, 5
Probable acetyl-/propionyl-coenzyme A carboxylase alpha chain (Alpha subunit) AccA2	*accA2*	Rv0973c	4, 7	3, 5
Probable acetyl-/propionyl-CoA carboxylase (Beta subunit) AccD2	*accD2*	Rv0974c	4, 7	3, 5
Propionyl-CoA carboxylase beta chain 6 AccD6	*accD6*	Rv2247	4, 7	3, 5
Malonyl CoA-acyl carrier protein transacylase	*fadD*	Rv0166	5, 7	3, 4
Long-chain-fatty-acid–AMP ligase FadD32	*fadD32*	fadD32	5, 7	3, 4
Galactofuranosyltransferase GlfT2	*GlfT2*	Rv3808c	5, 7	3, 4
Phthiocerol synthesis polyketide synthase type I PpsA	*ppsA*	Rv2931	4, 7	3,5
Phthiocerol synthesis polyketide synthase type I PpsB	*ppsB*	Rv2932	4, 7	3, 5
Phenolpthiocerol synthesis type-I polyketide synthase PpsC	*ppsC*	Rv2933	4, 7	3, 5
Doxorubicin resistance ATP-binding protein DrrA	*drrA*	Rv2936	4, 7	3, 5
Phthiocerol synthesis polyketide synthase type I PpsD	*ppsD*	Rv2934	4, 7	3, 5
Phthiocerol synthesis polyketide synthase type I PpsE	*ppsE*	Rv2935	4, 7	3, 5
Steroid 3-ketoacyl-CoA thiolase	*fadA5*	Rv3546	4, 7	3, 5
Polyketide synthase Pks7	*Pks7*	Rv1661	4, 7	3, 5
Long-chain-fatty-acid–AMP ligase FadD26	*fadD26*	Rv2930	4, 7	3, 5
Polyketide synthase-like Pks10	Pks10	Rv1660	5, 7	3, 4
Phthioceranic/hydroxyphthioceranic acid synthase	Pks2	Rv3825c	5, 7	3, 4
Phthiotriol/phenolphthiotriol dimycocerosates methyltransferase	*Rv2952*	Rv2952	7	3,4, 5
Putative phthiocerol dimycocerosate transporter LppX	*lppx*	Rv2945c	3,7	4, 5
Mycocerosic acid synthase-like polyketide synthase	*pks5*	Rv1527c	7	3, 4
Arabinofuranosyltransferase AftB	*aftb*	Rv3805c	3, 4, 7	5
Putative diacyglycerol *O*-acyltransferase Rv2285	*Rv2285*	Rv2285	7	3, 4, 5
Peptidoglycan endopeptidase RipA	*ripA*	Rv1477	4	3, 7
Peptidoglycan endopeptidase RipB	*ripB*	Rv1478	7	3, 4, 5
Probable penicillin-binding protein DacB1	*dacB1*	Rv3330	7	3, 4, 5
Phosphoglucosamine mutase	*glmM*	Rv3441c	7	3, 4, 5
Rhamnosyl *O*-methyltransferase	*Rv2959c*	Rv2959c	3, 7	4, 5
Uncharacterized glycosyl hydrolase Rv2006	*Rv2006*	Rv2006	4, 7	3, 5
Putative inactive phenolphthiocerol synthesis polyketide synthase type I Pks1	*pks1*	Rv2946c	3, 4, 7	5
Steroid C26-monooxygenase	*cyp125*	Rv3545c	3,4, 7	5
Long-chain-fatty-acid–AMP ligase FadD30	*fadD30*	Rv0404	3, 4, 7	5
Acyl-CoA dehydrogenase FadE28	*fadE28*	Rv3544c	3, 4, 7	5
Probable NADPH dependent 2,4-dienoyl-CoA reductase FadH	*fadh*	Rv1175c	5, 7	3, 4
UDP-*N*-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligase	*murF*	Rv2157c	3, 5, 7	4
UDP-*N*-acetylmuramoylalanine–D-glutamate ligase	*murD*	Rv2155c	5, 7	3, 4
UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase	*murE*	Rv2158c	3, 5, 7	4
Isoniazid-induced protein IniC	*iniC*	Rv0343	5, 7	3, 4
Isoniazid-induced protein IniA	*iniA*	Rv0342	5, 7	3, 4

biosynthesis (Azad et al., 1997). One study reported that the *pks1 and pks10* mutants displayed deficient PDIM production and attenuated virulence in mice (Forrellad et al., 2013). The *fadD28* mutant showed defect in PDIM biosynthesis (Forrellad et al., 2013).

Peptidoglycan is an essential component of the bacterial cell wall, which is crucial in maintaining cell shape and strength (Kouidmi et al., 2014). Mur ligases (MurD, MurE, and MurF) and RipA play significant roles in peptidoglycan biosynthesis. DacB1, which is less abundant in lineage 7, is involved in crosslinking

TABLE 5 | Differentially abundant proteins involved in energy metabolism in MTBC lineages.
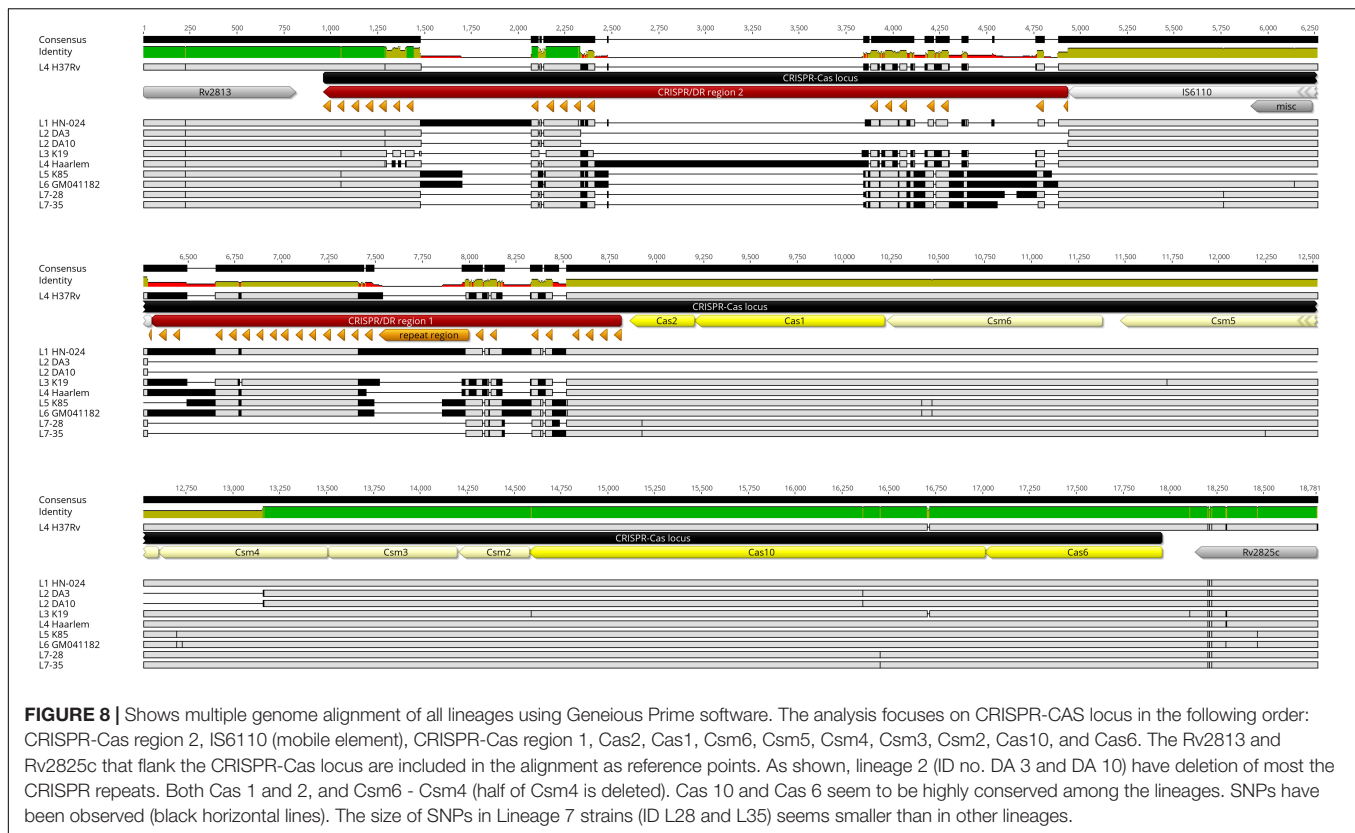
Protein name	Gene symbol	Rv number	Less abundance (lineage/s)	Increased abundance lineage/s
Glucose-6-phosphate isomerase Pgi	*pgI*	Rv 0946c	3, 4, 5	7
6-Phosphofructokinase PfkA	*pfkA*	Rv3010c	3, 4, 5	7
Probable alcohol dehydrogenase AdhA	*adhA*	Rv1862	3, 4, 5	7
Possible L-lactate dehydrogenase (cytochrome) LldD1	*LldD1*	Rv0694	3, 4, 5	7
Putative citrate synthase II CitA	*citA*	Rv0889c	7	3, 4, 5
Isocitrate lyase Icl	*Icl1*	Rv0467	4, 7	3, 5
6-Phosphogluconate dehydrogenase Gnd1	*gnD1*	Rv1844c	5, 7	5, 7
6-Phosphogluconate dehydrogenase, decarboxylating Gnd2	*gnD2*	Rv1122	5, 7	5, 7
Glucose-6-phosphate 1-dehydrogenase Zwf2 (G6PD)	*Zwf2*	Rv1447c	3, 4, 5	7
Fumarate reductase flavoprotein subunit	*frdA*	Rv1552	3, 5	4, 7
Fumarate reductase iron-sulfur subunit	*frdB*	Rv1553	3, 5	4, 7
Putative succinate-semialdehyde dehydrogenase 2	*gabD2*	Rv1731	3, 5	4, 7
Citrate synthase 1	*gltA2*	Rv0896	5	3, 4, 7
Succinate dehydrogenase 2 membrane subunit SdhC	*sdhC*	Rv3316	5	3, 4, 7
Succinate dehydrogenase flavoprotein subunit	*sdhA*	Rv3318	5	3, 4, 7
Probable succinate dehydrogenase SdhB	*sdhB*	sdhB	5	3, 4, 7
2-Methylcitrate synthase	*prpC*	Rv1131	3, 4, 5	7
2-Methylcitrate dehydratase	*prpD*	Rv1130	3, 4, 5	7
Probable NADH dehydrogenase Ndh	*ndH*	Rv1854C	4, 7	3, 5
NADH dehydrogenase-like protein Rv1812c	*nadH*	Rv1812c	7	3, 4, 5
Probable NADPH dependent 2,4-dienoyl-CoA reductase FadH	*fadH*	RV1812c	5, 7	3, 4
Quinolinate synthase A	*nadA*	Rv1594	7	3, 4, 5
L-Aspartate oxidase	*nadB*	Rv1595	7	3, 4, 5
Nicotinate-nucleotide pyrophosphorylase	*nadC*	Rv1596	7	3, 4, 5
Isocitrate lyase 2	*aceaB*	Rv1916	4	3, 5, 7
Putative isocitrate lyase subunit A	*aceaA*	Rv1915	4	3, 5, 7

TABLE 6 | Differentially abundant metabolic stress response and regulatory proteins in MTBC lineages.

Protein name	Gene symbol	Rv number	Less abundance (lineage/s)	Increased abundance lineage/s
Alkyl hydroperoxide reductase C	*ahpC*	Rv2428	4, 5, 7	3
Alkyl hydroperoxide reductase AhpD	*ahpD*	Rv2429	7	3, 4, 5
Mycothiol acetyltransferase	*mshD*	Rv0819	7	3, 4, 5
Thiol peroxidase	*tpX*	Rv1932	7	3, 4, 5
Serine/threonine-protein kinase PknE	*pknE*	Rv1743	7	3, 4, 5
Serine/threonine-protein kinase PknF	*pkmF*	Rv1746	7	3, 4, 5
Uncharacterized HTH-type transcriptional regulator Rv0043c	Rv0043c	Rv0043c	7	3,4, 5
Sensor-like histidine kinase senX3	*senX3*	Rv0490	7	3, 4, 5
DNA-binding transcriptional activator DevR/DosR	*devR/DosR*	Rv3133c	3,7	5, 7
Oxygen sensor histidine kinase response regulator DosT	*dosT*	Rv2027c	3, 4, 5	7
Oxygen sensor histidine kinase response regulator DevS/DosS	*devS*	Rv3132c	4, 5, 7	3

of cell wall peptidoglycan strands (Goodsmith et al., 2015). The cytoplasmic biogenesis of fatty acids is coordinated with a set of steps in charge of leading them to their destination out of the cell. The proper synthesis and localization of cell wall is compromised if there is any derangement in the formation and exporting steps of fatty acids. Given that several proteins involved in the essential steps of lipid biosynthesis were less abundant in lineage 7, the possibility of defects in cell wall formation and growth in this lineage is very high.

Mtb uses homologous recombination (HR) to repair DNA damage and several proteins are essential (Singh et al., 2016). Lineage 7 exhibited a low abundance of important DNA repair components including UvrC, RecA, RuvA, RuvB, RecR, DnaE1, and HelZ proteins. Our previous study showed non-synonymous SNPs in genes in lineage 7 strains involved in DNA metabolism such as *nadD* and *dnaB* (helicase) (Yimer et al., 2016). Notably, three of the five proteins involved in NAD+ cofactor synthesis (Kurnasov et al., 2003), i.e., NadA, NadB, and NadC, were

FIGURE 8 | Shows multiple genome alignment of all lineages using Geneious Prime software. The analysis focuses on CRISPR-CAS locus in the following order: CRISPR-Cas region 2, IS6110 (mobile element), CRISPR-Cas region 1, Cas2, Cas1, Csm6, Csm5, Csm4, Csm3, Csm2, Cas10, and Cas6. The Rv2813 and Rv2825c that flank the CRISPR-Cas locus are included in the alignment as reference points. As shown, lineage 2 (ID no. DA 3 and DA 10) have deletion of most the CRISPR repeats. Both Cas 1 and 2, and Csm6 - Csm4 (half of Csm4 is deleted). Cas 10 and Cas 6 seem to be highly conserved among the lineages. SNPs have been observed (black horizontal lines). The size of SNPs in Lineage 7 strains (ID L28 and L35) seems smaller than in other lineages.

less abundant in lineage 7. NAD+ cofactor starvation is detrimental in the tubercle bacilli (Boshoff et al., 2008). As DNA repair and recombination are expected to play crucial roles in the adaptative potential, evolution and pathogenicity of Mtb, it is tempting to speculate that the decreased abundance of these essential DNA repair/recombination components in lineage 7 may imply less efficient DNA repair mechanism or slow growth.

The CRISPR-Cas genome editing system protects cells against phage infection (Barrangou and Marraffini, 2014). The CRISPR-derived RNA (crRNA) is responsible for directing the invader-degrading Cas complex to the invader (Hille and Charpentier, 2016). crRNA synthesis requires Cas6, Cas10, Csm2, Csm3, Csm4, and Csm5. Mutations in Cas-Csm genes have been shown to affect immunity in *Staphylococcus aureus* (Hatoum-Aslan et al., 2014). Csm3 acts as a ruler and measures the size of incoming crRNA maturation within the Cas10-Csm complex in *S. epidermidis* (Hatoum-Aslan et al., 2013). Cas10 displays rapid conformational fluctuations on foreign RNA targets, but is locked in a static configuration on self RNA, highlighting the central role of self versus non-self discrimination and target specificity (Wang et al., 2019). The lineage 7 strains exhibited an increased abundance of proteins involved in the CRISPR-Cas system. This observation is supported by the multiple-genome alignment analysis of the CRISPR-Cas locus which revealed very low number of SNPs in the genomes of lineage 7 strains than in other lineages. Mutations in the genes encoding CRISPR-Cas systems or the lack of CRISPR-Cas systems generally contribute

to drug resistance (Wei et al., 2019). Deletion of Cas1 was found to reduce the sensitivity to anti-tuberculosis drugs by decreasing the persistence during treatment (Wei et al., 2019). Our previous study revealed that strains of lineages 3 and 4 exhibited a higher propensity for resistance to anti-TB drugs than lineage 7 strains (Yimer et al., 2015). The increased abundance of the Csm3, Csm4, Cas6 and Cas10 in lineage 7 may protect against phage infection and drug resistance. These possibilities warrant further study.

Metabolic stress adaptation is essential for MTBC survival. A number of proteins involved in metabolic stress responses were differentially abundant in a lineage-specific manner. TpX, which is less abundant in lineage 7, catalyzes the reduction of hydroperoxides and peroxynitrite (Rho et al., 2006). A *tpx*-mutant Mtb strain showed impaired replication in macrophages (Hu and Coates, 2009). MshD involved in mycothiol (MSH) biosynthesis has a similar function to glutathione (Newton et al., 2008). Mutations in *mshD* resulted in MSH production defect (Newton et al., 2005; Venkatesan et al., 2016). An *ahpC* mutant Mtb showed decreased survival in macrophages (Master et al., 2002). These proteins were less abundant in lineage 7 and may result in an inefficient metabolic response to oxidative stress, which warrants further functional study.

Proteins involved in cellular respiration were differentially abundant in a lineage-specific manner. The TCA pathway supplies NADH for use in oxidative phosphorylation (Cavalcanti et al., 2014). CitA and IcL proteins, which are less abundant in lineage 7, are essential to catalyze the biosynthesis of citrate and isocitrate to produce NADH. Among the enzymes involved

in oxidative phosphorylation, lineage 7 showed less abundance of key enzymes (Ndh, Rv1812C and FadH) (Cook et al., 2014). Ndh is one of the major oxidoreductases in mycobacteria (Iqbal et al., 2018) and has a catalytic function crucial for reducing the flight of electrons to oxygen during transfer from NADH to menaquinone. Decreased NADH abundance causes starvation and disruption of cellular redox homeostasis and results in cell death (Iqbal et al., 2018). The question is therefore how does lineage 7 in particular maintain fitness for survival given the decreased abundance of several key enzymes involved in the oxidative direction of TCA cycle and electron transport chain? The TCA cycle is dynamic. When Mtb encounters hostile environmental conditions such as hypoxia, the bacteria shift the oxidative TCA cycle to a reductive direction and start to secrete succinic acid to complete the respiratory cycle. This is facilitated by succinate dehydrogenase (SDH) that forms complex II of the respiratory chain. Complex II of the respiratory chain is composed of SdhA, SdhB, SdhC, and SdhD proteins (Cook et al., 2014; Iqbal et al., 2018), of which SdhA and SdhB were more abundant in lineage 7. These enzymes oxidize succinate to fumarate in the cytoplasm. Fumarate reductase (FRD), a "paralog" of SDH, catalyzes the reverse reaction. We found an increased abundance of FrdA and FrdB in lineages 4 and 7 that may indicate a shift in metabolic pathway in these two lineages. It has been shown FrdA was more abundant under hypoxic conditions in Mtb (Sharma and Tyagi, 2016) and that TCA was repressed in dormant culturable Mtb cells (Converse et al., 2009). Lineage 7 also exhibited a higher abundance of several proteins that are involved in alternative pathways of the TCA cycle. This may be considered a compensatory mechanism for generating energy to maintain basic physiological functions of Mtb lineage 7 cells and to survive harsh environmental conditions.

Lineages 3 and 4 demonstrated a higher abundance of IniC and IniA that play key roles in the development of resistance to both isoniazid and ethambutol (Colangeli et al., 2005; Almeida Da Silva and Palomino, 2011). Lineage 4 has been frequently associated with drug resistance (Fenner et al., 2012; Tessema et al., 2013; Gygli et al., 2017). In addition, lineages 3 and 4 showed higher level of drug tolerance than other Mtb lineages in Ethiopia (Yimer et al., 2015).

CONCLUSION

This study presents novel findings regarding differential protein abundance/protein expression in Mtb lineage 7 versus the other clinically relevant MTBC lineages 3, 4, and 5. The reduced abundance of proteins involved in type VII ESX-3 and ESX-1 secretion systems, TCA cycle and oxidative phosphorylation, lipid metabolism and DNA repair in lineage 7 may contribute to the slow growth and less virulent phenotype, which corroborates a previous report (Yimer et al., 2017). The increased abundance of proteins involved in the CRISPR-Cas system in lineage 7 may enable protection against phage infection and drug resistance. Given the increased abundance of several proteins in Mtb lineage 7 that are involved in alternative energy metabolism pathways, we hypothesize that increased respiratory plasticity may contribute

to the ability of this lineage to survive adverse growth conditions possibly through increased periods of dormancy.

The increased abundance of proteins involved in type VII ESX-3 and ESX-1 secretion systems, DNA repair, drug resistance and lipid metabolism may contribute to the high virulence and increased fitness for survival of lineage 4, which is consistent with the wide global distribution of this lineage. Lineage 3, prevalent in central Asia and East Africa, showed increased abundance of proteins involved in lipid metabolism, ESX-3 secretion system, the TCA cycle and other metabolic pathways associated with increased fitness and survival capacity. The results of this study may provide an insight into the phenotypic differences among clinically important MTBC lineages. However, additional studies using models for *in vivo* infection are warranted to confirm these findings.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Regional Committee for Medical Research Ethics in Eastern Norway (REK Øst) and the Ethiopian Science and Technology Ministry in Addis Ababa, Ethiopia. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SAY and TT conceived the study and study design. SAY conducted the field work and collected lineages 3, 4, and 7 isolates in Ethiopia. AGB and EDZ performed specimen cultivation. SK and EDZ performed sample inactivation and lysate preparation for MS analysis. HH performed immunoblotting. SK performed MQ analysis. TR ran MS analysis. SK and SAY performed proteome bioinformatics and statistical analysis. AGB performed cytoscape. TL ran Geneious Prime for sequence analysis. SAY, TT, and SK evaluated and interpreted the data and drafted the manuscript. All authors edited and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.550760/full#supplementary-material

FILE S1 | A multi-histogram demonstrates the distribution of protein spectra detected across all the experimental replicates. The frequency distribution of the $\log_2$-transformed LFQ intensities has a unimodal shape with the typical "bell shaped" normal distribution. This distribution remained the same after the dataset was imputed for missing values. Proteins exhibiting low abundance and/or poorly ionizing peptides are highlighted by the red color.

FILE S2 | Depicts the complete list of protein groups identified in all lineages.

FILE S3 | Shows the iBAQ intensities in all proteomes that covered a dynamic range of four orders of magnitude between the most- and least abundant proteins.

FILE S4 | The top 100 most abundant proteins assigned by their functional categories using Tuberculist (Kapopoulou et al., 2011).

FILE S5 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance difference between lineages 3 and 7.

FILE S6 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance comparison between lineages 4 and 7.

FILE S7 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance comparison between lineages 3 and 4.

FILE S8 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance comparison between lineages 5 and 7.

FILE S9 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance comparison between lineages 4 and 5.

FILE S10 | Protein abundance comparison using Welch's *t*-test between the MTBC lineages: Protein abundance comparison between lineages 3 and 5.

FILE S11 | List of the 539 differentially abundant proteins based on ANOVA analysis. The DA protein groups were further subjected to unsupervised hierarchical cluster analysis resulting in eight clustered protein groups.

FILE S12 | Gene ontology analysis among the MTBC lineages 3, 4, 5, 7. Pathway sub-clusters that were significantly enriched are listed. The list also includes significant enrichment analysis between the lineages: 5 and 7, 3 and 7, and 4 and 7.

FILE S13 | Detailed experimental procedure for immunoblotting.

REFERENCES

Almeida Da Silva, P. E., and Palomino, J. C. (2011). Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J. Antimicrob. Chemother.* 66, 1417–1430. doi: 10.1093/jac/dkr173

Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T., and Riley, L. W. (1993). Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* 261, 1454–1457. doi: 10.1126/science.8367727

Azad, A. K., Sirakova, T. D., Fernandes, N. D., and Kolattukudy, P. E. (1997). Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J. Biol. Chem.* 272, 16741–16745. doi: 10.1074/jbc.272.27.16741

Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2. doi: 10.1186/1471-2105-4-2

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. doi: 10.1016/j.mib.2017.09.013

Barrangou, R., and Marraffini, L. A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell.* 54, 234–244. doi: 10.1016/j.molcel.2014.03.011

Barth, K., Remick, D. G., and Genco, C. A. (2013). Disruption of immune regulation by microbial pathogens and resulting chronic inflammation. *J. Cell. Physiol.* 228, 1413–1422. doi: 10.1002/jcp.24299

Bespyatykh, J., Shitikov, E., Butenko, I., Altukhov, I., Alexeev, D., and Mokrousov, I. (2016). Proteome analysis of the *Mycobacterium tuberculosis* Beijing B0/W148 cluster. *Sci. Rep.* 6:28985.

Bhatt, A., Fujiwara, N., Bhatt, K., Gurcha, S. S., Kremer, L., Chen, B., et al. (2007). Deletion of kasB in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5157–5162. doi: 10.1073/pnas.0608654104

Birhanu, A. G., Yimer, S. A., Holm-Hansen, C., Norheim, G., Aseffa, A., Abebe, M., et al. (2017). Nε- and O-Acetylation in *Mycobacterium tuberculosis* lineage 7 and lineage 4 strains: proteins involved in bioenergetics, virulence, and antimicrobial resistance are acetylated. *J. Proteome Res.* 16, 4045–4059. doi: 10.1021/acs.jproteome.7b00429

Boshoff, H. I., Xu, X., Tahlan, K., Dowd, C. S., Pethe, K., Camacho, L. R., et al. (2008). Biosynthesis and recycling of nicotinamide cofactors in *Mycobacterium tuberculosis*: an essential role for nad in nonreplicating bacilli. *J. Biol. Chem.* 283, 19329–19341. doi: 10.1074/jbc.m800694200

Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., and Pawlik, A. (2020). TbD1 deletion as a driver of the evolutionary success of modern epidemic *Mycobacterium tuberculosis* lineages. *Nat. Commun.* 11:684.

Cavalcanti, J. H. F., Esteves-Ferreira, A. A., Quinhones, C. G. S., Pereira-Lima, I. A., Nunes-Nesi, A., Fernie, A. R., et al. (2014). Evolution and functional implications of the tricarboxylic acid cycle as revealed by phylogenetic analysis. *Genome Biol. Evol.* 6, 2830–2848. doi: 10.1093/gbe/evu221

Chacon-Salinas, R., Serafín-López, J., Ramos-Payán, R., Méndez-Aragón, P., Hernández-Pando, R., Van Soolingen, D., et al. (2005). Differential pattern of cytokine expression by macrophages infected in vitro with different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* 140, 443–449. doi: 10.1111/j.1365-2249.2005.02797.x

Colangeli, R., Helb, D., Sridharan, S., Sun, J., Varma-Basil, M., Hazbón, M. H., et al. (2005). The *Mycobacterium tuberculosis* iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Mol. Microbiol.* 55, 1829–1840. doi: 10.1111/j.1365-2958.2005.04510.x

Converse, P. J., Karakousis, P. C., Klinkenberg, L. G., Kesavan, A. K., Ly, L. H., and Allen, S. S. (2009). Role of the dosR-dosS two-component regulatory system in *Mycobacterium tuberculosis* virulence in three animal models. *Infect. Immun.* 77, 1230–1237. doi: 10.1128/iai.01117-08

Cook, G. M., Hards, K., Vilchèze, C., Hartman, T., and Berney, M. (2014). Energetics of respiration and oxidative phosphorylation in mycobacteria. *Microbiol. Spectr.* 2. doi: 10.1128/microbiolspec.MGM2-0015-2013

Coscolla, M., Brites, D., Menardo, F., Loiseau, C., Borrell, S., Darko Otchere, I., et al. (2020). Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *BioRxiv* [Preprint]. doi: 10.1101/2020.06.10.141788

Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26, 431–444. doi: 10.1016/j.smim.2014.09.012

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* 13, 2513–2526. doi: 10.1074/mcp.m113.031591

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805. doi: 10.1021/pr101065j

De Voss, J. J., Rutter, K., Schroeder, B. G., and Barry, C. E. (1999). Iron acquisition and metabolism by mycobacteria. *J. Bacteriol.* 181, 4443–4451. doi: 10.1128/jb.181.15.4443-4451.1999

Domenech, P., and Reed, M. B. (2009). Rapid and spontaneous loss of phthiocerol dimycocerosate (PIDM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies. *Microbiology* 155, 3532–3543. doi: 10.1099/mic.0.029199-0

Dos Vultos, T., Mestre, O., Tonjum, T., and Gicquel, B. (2009). DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS Microbiol. Rev.* 33, 471–487.

Duan, X., Xiang, X., and Xie, J. (2014). Crucial components of *Mycobacterium* type II fatty acid biosynthesis (Fas-II) and their inhibitors. *FEMS Microbiol. Lett.* 360, 87–99. doi: 10.1111/1574-6968.12597

Elias, J. E., and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214. doi: 10.1038/nmeth1019

Fenner, L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., et al. (2012). Effect of mutation and genetic background on drug resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 56, 3047–3053.

Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., et al. (2013). Mycobacterial lineages causing pulmonary and extrapulmonary *Tuberculosis*, Ethiopia. *Emerg. Infect. Dis.* 19, 460–463. doi: 10.3201/eid1903.120256

Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y García, J., Morbidoni, H. R., de la Paz Santangelo, M., et al. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3–66.

Fortune, S. M., Jaeger, A., Sarracino, D. A., Chase, M. R., Sassetti, C. M., Sherman, D. R., et al. (2005). Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10676–10681. doi: 10.1073/pnas.0504922102

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., and Roth, A. (2013). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815.

Goodsmith, N., Guo, X. V., Vandal, O. H., Vaubourgeix, J., Wang, R., Botella, H., et al. (2015). Disruption of an *M. tuberculosis* membrane protein causes a magnesium-dependent cell division defect and failure to persist in mice. *PLoS Pathog.* 11:e1004645. doi: 10.1371/journal.ppat.1004645

Guinn, K. M., Hickey, M. J., Mathur, S. K., Zakel, K. L., Grotzke, J. E., Lewinsohn, D. M., et al. (2004). Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* 51, 359–370. doi: 10.1046/j.1365-2958.2003.03844.x

Gurvitz, A., Hiltunen, J. K., and Kastaniotis, A. J. (2008). Function of heterologous *Mycobacterium tuberculosis* InhA, a Type 2 fatty acid synthase enzyme involved in extending C20 fatty acids to C60-to-C90 mycolic acids, during de novo lipoic acid synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 74, 5078–5085. doi: 10.1128/aem.00655-08

Gygli, S. M., Borrell, S., Trauner, A., and Gagneux, S. (2017). Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. *FEMS Microbiol. Rev.* 41, 354–373. doi: 10.1093/femsre/fux011

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L. A. (2014). Genetic characterization of antiplasmid immunity through a Type III-A CRISPR-Cas System. *J. Bacteriol.* 196, 310–317. doi: 10.1128/jb.01130-13

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L. A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J. Biol. Chem.* 288, 27888–27897. doi: 10.1074/jbc.m113.499244

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169. doi: 10.1186/s12859-015-0611-3

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150496. doi: 10.1098/rstb.2015.0496

Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeón, L., Abdallah, A. M., et al. (2012). ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell. Microbiol.* 14, 1287–1298. doi: 10.1111/j.1462-5822.2012.01799.x

Houben, E. N., Korotkov, K. V., and Bitter, W. (2014). Take five - Type VII secretion systems of mycobacteria. *Biochim. Biophys. Acta* 1843, 1707–1716. doi: 10.1016/j.bbamcr.2013.11.003

Hu, Y., and Coates, A. R. M. (2009). Acute and persistent *Mycobacterium tuberculosis* infections depend on the thiol peroxidase TPX. *PLoS One* 4:e5150. doi: 10.1371/journal.pone.0005150

Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., et al. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* 189, 739–754. doi: 10.1083/jcb.200911091

Iqbal, I. K., Bajeli, S., Akela, A. K., and Kumar, A. (2018). Bioenergetics of *Mycobacterium*: an emerging landscape for drug discovery. *Pathogens* 7:24. doi: 10.3390/pathogens7010024

Jafari, M., Primo, V., Smejkal, G. B., Moskovets, E. V., Kuo, W. P., and Ivanov, A. R. (2012). Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. *Electrophoresis* 33, 2516–2526. doi: 10.1002/elps.201200031

Jamwal, S. V., Mehrotra, P., Singh, A., Siddiqui, Z., Basu, A., and Rao, K. V. S. (2016). Mycobacterial escape from macrophage phagosomes to the cytoplasm represents an alternate adaptation mechanism. *Sci. Rep.* 6:23089.

Jayakumar, D., Jacobs, W. R. Jr., and Narayanan, S. (2008). Protein kinase E of *Mycobacterium tuberculosis* has a role in the nitric oxide stress response and apoptosis in a human macrophage model of infection. *Cell. Microbiol.* 10, 365–374.

Jena, L., Kashikar, S., Kumar, S., and Harinath, B. C. (2013). Comparative proteomic analysis of *Mycobacterium tuberculosis* strain H37Rv versus H37Ra. *Int. J. Mycobacteriol.* 2, 220–226. doi: 10.1016/j.ijmyco.2013.10.004

Jhingan, G. D., Kumari, S., Jamwal, S. V., Kalam, H., Arora, D., Jain, N., et al. (2016). Comparative proteomic analyses of avirulent, virulent, and clinical strains of *Mycobacterium tuberculosis* identify strain-specific patterns. *J. Biol. Chem.* 291, 14257–14273. doi: 10.1074/jbc.m115.666123

Kapopoulou, A., Lew, J. M., and Cole, S. T. (2011). The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis* 91, 8–13. doi: 10.1016/j.tube.2010.09.006

Kelkar, D. S., Kumar, D., Kumar, P., Balakrishnan, L., Muthusamy, B., Yadav, A. K., et al. (2011). Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry. *Mol. Cell. Proteomics* 10:M111.011627. doi: 10.1074/mcp.M111.011627

Kouidmi, I., Levesque, R. C., and Paradis-Bleau, C. (2014). The biology of Mur ligases as an antibacterial target. *Mol. Microbiol.* 94, 242–253. doi: 10.1111/mmi.12758

Kurnasov, O., Goral, V., Colabroy, K., Gerdes, S., Anantha, S., Osterman, A., et al. (2003). NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. *Chem. Biol.* 10, 1195–1204.

MacGurn, J. A., Raghavan, S., Stanley, S. A., and Cox, J. S. (2005). A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 57, 1653–1663. doi: 10.1111/j.1365-2958.2005.04800.x

Marrakchi, H., Lanéelle, M. A., and Daffé, M. (2014). Mycolic acids: structures, biosynthesis, and beyond. *Chem. Biol.* 21, 67–85. doi: 10.1016/j.chembiol.2013.11.011

Master, S. S., Springer, B., Sander, P., Boettger, E. C., Deretic, V., and Timmins, G. S. (2002). Oxidative stress response genes in *Mycobacterium tuberculosis*: role of ahpC in resistance to peroxynitrite and stage-specific survival in macrophages. *Microbiology* 148, 3139–3144. doi: 10.1099/00221287-148-10-3139

Mehra, A., Zahra, A., Thompson, V., Sirisaengtaksin, N., Wells, A., Porto, M., et al. (2013). *Mycobacterium tuberculosis* type VII secreted effector EsxH targets host ESCRT to impair trafficking. *PLoS Pathog.* 9:e1003734. doi: 10.1371/journal.ppat.1003734

Millington, K. A., Fortune, S. M., Low, J., Garces, A., Hingley-Wilson, S. M., Wickremasinghe, M., et al. (2011). Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5730–5735. doi: 10.1073/pnas.1015153108

Murphy, H. N., Stewart, G. R., Mischenko, V. V., Apt, A. S., Harris, R., McAlister, M. S., et al. (2005). The OtsAB pathway is essential for trehalose biosynthesis in *Mycobacterium tuberculosis*. *J. Biol. Chem.* 280, 14524–14529. doi: 10.1074/jbc.m414232200

Newton, G. L., Buchmeier, N., and Fahey, R. C. (2008). Biosynthesis and functions of mycothiol, the unique protective thiol of actinobacteria. *Microbiol. Mol. Biol. Rev.* 72, 471–494. doi: 10.1128/mmbr.00008-08

Newton, G. L., Ta, P., and Fahey, R. C. (2005). A mycothiol synthase mutant of *Mycobacterium smegmatis* produces novel thiols and has an altered thiol redox status. *J. Bacteriol.* 187, 7309–7316. doi: 10.1128/jb.187.21.7309-7316.2005

Ngabonziza, J. C. S., Loiseau, C., Marceau, M., Marceau, M., Jouet, A., Menardo, F., et al. (2020). A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* 11:2917.

Nguyen, V. A. T., Bañuls, A.-L., and Tran, T. H. T. (2016). *Mycobacterium tuberculosis* lineages and anti-tuberculosis drug resistance in reference hospitals across Viet Nam. *BMC Microbiol.* 16:167. doi: 10.1186/s12866-016-0784-6

O'Neill, M. B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., et al. (2019). Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol. Ecol.* 23, 3241–3256.

Pandey, R., and Rodriguez, G. M. (2014). IdeR is required for iron homeostasis and virulence in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 91, 98–109. doi: 10.1111/mmi.12441

Pathak, S. K., Basu, S., Basu, K. K., Banerjee, A., Pathak, S., Bhattacharyya, A., et al. (2007). Direct extracellular interaction between the early secreted antigen ESAT-6 of *Mycobacterium tuberculosis* and TLR2 inhibits TLR signaling in macrophages. *Nat. Immunol.* 8, 610–618. doi: 10.1038/ni1468

Peirs, P., Lefèvre, P., Boarbi, S., Wang, X. M., Denis, O., Braibant, M., et al. (2005). *Mycobacterium tuberculosis* with disruption in genes encoding the phosphate binding proteins PstS1 and PstS2 is deficient in phosphate uptake and demonstrates reduced in vivo virulence. *Infect. Immun* 73, 1898–1902. doi: 10.1128/iai.73.3.1898-1902.2005

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450.

Portevin, D., Gagneux, S., Comas, I., and Young, D. (2011). Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog.* 7:e1001307. doi: 10.1371/journal.ppat.1001307

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670. doi: 10.1021/ac026117i

Rho, B. S., Hung, L. W., Holton, J. M., Vigil, D., Kim, S. I., Park, M. S., et al. (2006). Functional and structural characterization of a thiol peroxidase from *Mycobacterium tuberculosis*. *J. Mol. Biol.* 361, 850–863. doi: 10.1016/j.jmb.2006.05.076

Schiebel, J., Kapilashrami, K., Fekete, A., Bommineni, G. R., Schaefer, C. M., Mueller, M. J., et al. (2013). Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 288, 34190–34204. doi: 10.1074/jbc.m113.511436

Schubert, O. T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., et al. (2018). Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* 18, 96–108. doi: 10.1016/j.chom.2015.06.001

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharma, S., and Tyagi, J. S. (2016). *Mycobacterium tuberculosis* DevR/DosR dormancy regulator activation mechanism: dispensability of phosphorylation, cooperativity and essentiality of 828 α10 helix. *PLoS One* 11:e0160723. doi: 10.1371/journal.pone.0160723

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860. doi: 10.1038/nprot.2006.468

Siegrist, M. S., Steigedal, M., Ahmad, R., Mehra, A., Dragset, M. S., Schuster, B. M., et al. (2014). Mycobacterial Esx-3 Requires Multiple Components for Iron Acquisition. *mBio* 5:e001073-14.

Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., et al. (2012). Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS Pathog.* 8:e1002507. doi: 10.1371/journal.ppat.1002507

Simeone, R., Huet, G., and Constant, P. (2013). Functional characterisation of three O-methyltransferases involved in the biosynthesis of phenolglycolipids in *Mycobacterium tuberculosis*. *PLoS One* 8:e58954. doi: 10.1371/journal.pone.0058954

Singh, A., Bhagavat, R., Vijayan, M., and Chandra, N. (2016). A comparative analysis of the DNA recombination repair pathway in mycobacterial genomes. *Tuberculosis* 99, 109–119. doi: 10.1016/j.tube.2016.04.011

Tessema, B., Beer, J., Merker, M., Emmrich, F., Sack, U., Rodloff, A. C., et al. (2013). Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. *BMC Infect. Dis.* 13:131. doi: 10.1186/1471-2334-13-131

Tientcheu, L. D., Haks, M. C., Agbla, S. C., Sutherland, J. S., Adetifa, I. M., Donkor, S., et al. (2016). Host immune responses differ between *M. africanum*- and *M. tuberculosis*-infected patients following standard anti-tuberculosis treatment. *PLoS Negl. Trop. Dis.* 10:e0004701. doi: 10.1371/journal.pntd.0004701

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., et al. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Method* 13, 731–740. doi: 10.1038/nmeth.3901

van der Wel, N., Hava, D., Houben, D., Fluitsma, D., and Maaike van Zon Pierson, J. (2007). tuberculosis and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 129, 1287–1298. doi: 10.1016/j.cell.2007.05.059

Venkatesan, A., Palaniyandi, K., Sharma, D., Bisht, D., and Narayanan, S. (2016). Functional characterization of PknI-Rv2159c interaction in redox homeostasis of *Mycobacterium tuberculosis*. *Front. Microbiol.* 7:1654. doi: 10.3389/fmicb.2016.01654

Wang, L., Mo, C. Y., Wasserman, M. R., Rostøl, J. T., Marraffini, L. A., and Liu, S. (2019). Dynamics of Cas10 govern discrimination between self and non-self in type III CRISPR-Cas immunity. *Mol. Cell.* 73, 278–290. doi: 10.1016/j.molcel.2018.11.008

Wang, Q., Liu, S., Tang, Y., Liu, Q., and Yao, Y. (2014). MPT64 Protein from *Mycobacterium tuberculosis* inhibits apoptosis of macrophages through NF-kB-miRNA21-Bcl-2 Pathway. Manganelli R. *PLoS One* 9:e100949. doi: 10.1371/journal.pone.0100949

Watanabe, S., Zimmermann, M., Goodwin, M. B., Sauer, U., Barry, C. E., and Boshoff, H. I. (2011). Fumarate reductase activity maintains an energized membrane in anaerobic *Mycobacterium tuberculosis*. *PLoS Pathog.* 7:e1002287. doi: 10.1371/journal.ppat.1002287

Wei, J., Lu, N., Li, Z., Wu, X., Jiang, T., Xu, L., et al. (2019). The *Mycobacterium tuberculosis* CRISPR-associated Cas1 involves persistence and tolerance to anti-tubercular drugs. *Biomed. Res. Int.* 2:7861695.

World Health Organization [WHO] (2019). Available online at: http://www.who.int/tb/publications/global_report/en/

Yimer, S. A., Birhanu, A. G., Kalayou, S., Riaz, T., Zegeye, E. D., Beyene, G. T., et al. (2017). Comparative proteomic analysis of *Mycobacterium tuberculosis* lineage 7 and lineage 4 strains reveals differentially abundant proteins linked to slow growth and virulence. *Front. Microbiol.* 8:795. doi: 10.3389/fmicb.2017.00795

Yimer, S. A., Hailu, E., Derese, Y., Bjune, G. A., and Holm-Hansen, C. (2013). Spoligotyping of *Mycobacterium tuberculosis* isolates among pulmonary tuberculosis patients in Amhara Region, Ethiopia. *APMIS* 121, 878–885. doi: 10.1111/apm.12046

Yimer, S. A., Namouchi, A., Zegeye, E. D., Holm-Hansen, C., Norheim, G., Abebe, M., et al. (2016). Deciphering the recent phylogenetic expansion of the originally deeply rooted *Mycobacterium tuberculosis* lineage 7. *BMC Evol. Biol.* 16:146. doi: 10.1186/s12862-016-0715-z

Yimer, S. A., Norheim, G., Namouchi, A., Zegeye, E. D., Kinander, W., Tønjum, T., et al. (2015). *Mycobacterium tuberculosis* Lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *J. Clin. Microbiol.* 53, 1301–1309. doi: 10.1128/jcm.03566-14