A Mycobacterium tuberculosis fingerprint in human breath allows tuberculosis detection

Sergio Fabián Mosquera-Restrepo, Sophie Zuberogoïtia, Lucie Gouxette, Emilie Layre, Martine Gilleron, Alexandre Stella, David Rengel, Odile Burlet-Schiltz, Ana Cecilia Caro, Luis Garcia, et al.

A *Mycobacterium tuberculosis* fingerprint in human breath allows tuberculosis detection

Sergio Fabián Mosquera-Restrepo[1#], Sophie Zuberogoïtia[2#], Lucie Gouxette[2#], Emilie Layre[2], Martine Gilleron[2], Alexandre Stella[2], David Rengel[2], Odile Burlet-Schiltz[2], Ana Cecilia Caro[3], Luis F. Garcia[1], César Segura[4], Carlos Alberto Peláez Jaramillo[3], Mauricio Rojas [1,5,*] and Jérôme Nigou[2,*]

[1] Grupo de Inmunología Celular e Inmunogenética (GICIG), Instituto de Investigaciones Médicas, Facultad de Medicina. Sede de Investigación Universitaria (SIU). Universidad de Antioquia (UdeA), Medellín, Colombia

[2] Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, Université Toulouse III - Paul Sabatier (UPS), Toulouse, France

[3] Grupo Insterdisciplinario de Estudios Moleculares (GIEM). Instituto de Química, Facultad de Ciencias Exactas y Naturales. Universidad de Antioquia (UdeA), Medellín, Colombia.

[4] Grupo Malaria. Sede de Investigación Universitaria, Universidad de Antioquia (UdeA), Medellín, Colombia.

[5] Unidad de Citometría de Flujo, Sede de Investigación Universitaria (SIU), Universidad de Antioquia, UdeA, Medellín, Colombia

[#] contributed equally to the work

* to whom correspondence should be addressed. MR, e-mail: mauricio.rojas@udea.edu.co, tel: +57 (574) 219 64 61; JN, e-mail: Jerome.Nigou@ipbs.fr , tel: +33 (0)5 61 17 55 04.

ABSTRACT

An estimated one third of tuberculosis (TB) cases go undiagnosed or unreported. Sputum samples, widely used for TB diagnosis, are inefficient at detecting infection in children and paucibacillary patients. Indeed, developing point-of-care biomarker-based diagnostics that are not sputum-based is a major priority for the WHO. Here, in a proof-of-concept study, we tested whether pulmonary TB can be detected by analyzing patient exhaled breath condensate (EBC) samples. We find that the presence of *Mycobacterium tuberculosis* (Mtb)-specific lipids, lipoarabinomannan lipoglycan, and proteins in EBCs can efficiently differentiate baseline TB patients from controls. We used EBCs to track the longitudinal effects of antibiotic treatment in pediatric TB patients. In addition, Mtb lipoarabinomannan and lipids were structurally distinct in EBCs compared to *ex vivo* cultured bacteria, revealing specific metabolic and biochemical states of Mtb in the human lung. This provides essential information for the rational development or improvement of diagnostic antibodies, vaccines and therapeutic drugs. Our data collectively indicate that EBC analysis can potentially facilitate clinical diagnosis of TB across patient populations and monitor treatment efficacy. This affordable, rapid and non-invasive approach seems superior to sputum assays and has the potential to be implemented at point-of-care.

INTRODUCTION

Tuberculosis (TB) remains one of the top ten causes of death worldwide and is the leading cause of death from a single infectious agent[1]. About a quarter of the world's population is infected with *Mycobacterium tuberculosis* (Mtb), and thus at risk of developing TB. An estimated 10 million people developed TB in 2020 resulting in around 1.5 million TB deaths[1]. 11% of these TB cases occurred in children. An estimated one-third of all TB cases are, however, not diagnosed or reported, in part due to significant limitations in current diagnostic tools[1]. The WHO's End TB Strategy aims for a 90% reduction in TB deaths and an 80% reduction in TB incidence by 2030 relative to 2015 levels. Diagnostic tools are urgently needed to both monitor progress and meet this goal[2].

Three diagnostic priorities have been listed by the WHO and the TB community, including development of a point-of-care biomarker-based test for pulmonary TB[3,4]. Most conventional diagnostic tests rely on sputum samples, which can be difficult to obtain and have low diagnostic sensitivity in children, HIV-infected individuals and patients with extrapulmonary TB[3,5]. Therefore, the ideal diagnostic would not rely on sputum samples, and can also detect non-pulmonary TB. To be successfully implemented at point-of-care, a new test should use an easily accessible patient sample, such as urine, blood or breath condensate[3]. Blood-based diagnostics include interferon-γ release assays which detect the host immune response to Mtb and have been relatively successful for monitoring latent TB. However blood-based tests cannot accurately distinguish between Mtb infection and active TB disease[1].

Since 2015, the WHO has recommended urine tests based on the detection of lipoarabinomannan (LAM), a mycobacterial cell envelope lipoglycan[6] (Supplementary Fig. 1), to help diagnose TB in patients who are seriously ill with HIV[1,5,7]. However, urinary tests have suboptimal sensitivity, limiting their utility in TB screening[1]. One alternative is to test the liquid phase of exhaled air, called the exhaled breath condensate (EBC), which can be sampled by

cooling[8-10]. Like urine and blood, the EBC is accessible and thus merits further investigation as a potential fluid that can be sampled for TB diagnosis[3]. Indeed, EBC collection is easy, relatively cheap, noninvasive and does not require specialized personnel. EBC reflects the composition of the airway lining fluid and may, therefore, contain lung disease-specific markers from infectious agents or infected host tissues[8-10]. Indeed, it was previously reported that fatty acid, oxidative stress and inflammatory mediator profiles in EBC can differentiate TB-infected adults and children from healthy controls[11,12].

Here, we explored whether non-volatile bacterial molecules released into the extracellular milieu during infection can be detected in the EBC of TB patients. Interestingly, we found LAM at an unexpectedly high concentration range (15 to 120 µg/mL), as well as a set of Mtb-specific lipids and proteins in EBCs of TB patients. These factors were not detected in control individuals, either healthy or with community-acquired bacterial pneumonia. These Mtb-derived molecules allowed us to efficiently distinguish TB patients at baseline, including smear-negative and culture-negative adults. These markers also allowed us to distinguish samples from Mtb culture-positive or culture-negative children. Our data suggest that Mtb molecules in EBC are potential biomarkers for the early diagnosis of TB in adults and children, even in paucibacillary patients. Moreover, the longitudinal study of children under antibiotic treatment indicates that EBC analysis may allow real-time monitoring of treatment efficacy. Finally, we observed that LAM and lipids in EBC have a distinct structure compared to *ex vivo* cultured Mtb. This chemical difference indicates that bacilli in human lungs are using a distinct metabolic process compared to cultured Mtb. These data are consistent with Mtb growing as biofilms in the lungs and using host lipids as a major carbon source. Thus, EBC analysis gives previously elusive insights into the metabolic status of Mtb during lung infection and has the potential to diagnose TB patients.

RESULTS

We collected EBCs in health care centers in Medellín (Colombia) from adult and pediatric patients who had been diagnosed for pulmonary TB. Diagnosis was based on clinical, and/or radiographic, and/or bacteriological evidence of pulmonary disease resulting in treatment initiation by the diagnosing clinician[13-15]. Samples were collected using R-tubes™ equipped with a 0.3-µm filter to avoid contamination with live bacteria. In 15 adult patients (Ad) out of 29, and in 5 pediatric patients (Ch) out of 17, the TB diagnosis was microbiologically confirmed by acid-fast bacilli smear (S^+) and/or Mtb culture (C^+) (Table 1). The first EBC sample was collected at baseline (prior to or within the first 2 weeks of initiating anti-TB treatment). Additional EBC samples were obtained for 6 pediatric patients at months 1 and 3 of anti-TB treatment (Table 1). Control individuals included 15 healthy adults, 15 healthy children, and 15 adult non-TB patients with community-acquired bacterial pneumonia (Table 1). Exclusion criteria were patients positive for HIV, diabetes, cancer, autoimmune diseases, immunosuppressive treatment, previous TB infections and smokers. These exclusions allowed us to focus on assessing TB in the absence of factors that influence disease or might confound detection. Subjects were asked to breathe at a normal frequency for 15 min, yielding at least 1 ml of condensate. For normalization between individuals, EBCs were lyophilized and resuspended in a total volume of 250 µl. Therefore, when available, the quantity of the bacterial molecules in EBC will be provided as mass per EBC.

EBCs from TB patients contain high levels of LAM

Since LAM is an established TB biomarker in urine[1,5,7,16], we first attempted to detect LAM in EBCs using an anti-LAM antibody (CS-35[17]). Dot-blot analysis showed that EBCs of all TB patients at baseline exhibited a positive signal, whereas EBCs of control individuals, either healthy or infected by other pulmonary bacterial pathogens, showed a negative, or extremely

weak, response (Supplementary Fig. 2). LAM content in EBCs was quantified relative to LAM purified from *M. tuberculosis* H37Rv using a standard curve (Supplementary Fig. 2). Unexpectedly, while the apparent concentration of LAM in urine from TB patients usually ranges from pg to ng/ml[5,7], the apparent quantity of LAM in EBCs from TB patients at baseline ranged from 40 ng to 947 µg (Fig. 1a). More specifically the amount of LAM per EBC in different groups ranged from 70 to 947 µg in Ad S^+, 4 to 19 µg in Ad S^-C^-, 0.04 to 1.5 µg in Ad S^-C^+, 5 to 12 µg in Ch S^+/C^+ and 0.48 to 6.4 µg in Ch S^-C^- (Figs. 1b and c). In short, EBC samples from TB patients contain µgs of LAM (corresponding to concentrations in the µg/ml range), manifold higher than expected based on other patient fluids.

These unexpectedly high apparent amounts of LAM prompted us to validate whether LAM is indeed present in TB patient EBCs using cross-validation with a method other than the anti-LAM antibody. We therefore tested for the presence of pentose D-arabinose, a component of LAM[18] (Supplementary Fig. 1). Pentose D-arabinose is present in mycobacteria and related genera, but is absent in eukaryotes[6]. After acid hydrolysis, chemical derivatization, and analyses by capillary electrophoresis monitored by laser-induced fluorescence (CE-LIF) and gas chromatography coupled to mass spectrometry (GC-MS), we could indeed detect arabinose in EBCs from TB patients but not control individuals (Supplementary Figs. 3a and b). More specifically, we detected D-, but not L-arabinose (Supplementary Fig. 3c). Interestingly, the monosaccharides mannose and glucose were also detected. Mannose is, in addition to arabinose, a major building block of LAM[6]; glucose is also a constituent of Mtb cell envelope compounds, such as the polysaccharide α-glucan or trehalose-based glycolipids. However, both mannose and glucose are also normally present in the human host. Arabinose was detected only after acid hydrolysis indicating that it is polymerized, whereas around 60% of glucose and 10% of mannose were found as free monosaccharides (Supplementary Fig. 3d; Supplementary Tables 1 and 2).

We then quantified arabinose using CE-LIF as a proxy for LAM in the EBC samples collected. LAM quantification by chemical analysis (Supplementary Fig. 4) yielded an overall pattern similar to that obtained for LAM quantification with the anti-LAM antibody (Fig. 1) and confirmed the µg abundance of LAM in EBCs from TB patients at baseline. However, chemical analysis provided numerical values distributed in a different range (0.7 to 30 µg *vs* 10 ng to 947 µg by immunoassay).

We next examined the sensitivity and specificity of immunoassay-based LAM quantification in EBC for TB detection (Fig. 1a). These parameters were assessed using a receiver operating characteristic (ROC) analysis (Fig. 1d). Overall ROC performance was assessed by calculating the area under the curve (AUC), which was 0.997 (95% CI, 0,9928 to 1,002; n = 46 cases, n = 45 controls; $P < 0.0001$). At a threshold of 64 ng/EBC, this ROC analysis yielded a sensitivity of 93,4% and a specificity of 100%. By these criteria, only 3 false negatives were present amongst 14 Ad S^-C^+ patients. However, it is worth noting that Ad S^-C^+ patients were under antibiotic treatment for 2 weeks or less (Table 1), and LAM content in EBC may have decreased compared to samples collected prior to treatment initiation, as suggested by the follow up study of pediatric patients (Fig. 1e). Leave-One-Out cross-validation yielded similar results, providing sensitivity, specificity and accuracy prediction values of 0.93, 0.98 and 0.96, respectively (Supplementary Fig. 5). Therefore, immunoassay-based LAM quantification in EBC allowed us to unambiguously distinguish most TB patients at baseline from healthy control subjects or non-TB patients with community-acquired bacterial pneumonia. This included detecting TB patients that were smear-negative and culture-negative adults (Ad S^-C^-) and smear- or culture-positive children (Ch S^+/C^+) (apparent LAM amount above 5 µg/EBC). The signal-to-noise ratio was thus found to be above 75. Even smear-negative and culture-negative pediatric patients (Ch S^-C^-) under antibiotic treatment for 15 days, with an apparent

LAM content > 480 ng/EBC, could be easily distinguished from control subjects (Figs. 1b and c).

Our longitudinal study followed 6 pediatric patients (Ch S^-C^-) for 3 months under antibiotic treatment. Antibody-based detection revealed a decline in apparent LAM content over the course of the treatment for 5 subjects (Fig. 1e). Four patients showed values below the threshold at month 1, and 5 at month 3. In contrast, apparent LAM content did not decrease over the time in EBCs collected from Child #I. It is worth noting that conventional antibiotic treatment (Table 1) for this patient failed. Child #I was subsequently referred to a search for primary immunodeficiency.

Altogether, these data suggest that LAM quantification by immunoassay in EBCs can be used to detect TB in adults and children, even in paucibacillary patients (smear-negative and/or culture-negative), and to monitor antibiotic treatment efficacy. These data indicate that diagnostics can potentially be based on EBC samples.

LAM in EBCs has a non-mature polysaccharide structure

Quantification of LAM by immunoassay and chemical analysis showed a significant, but weak, correlation (P = 0.0002, r = 0.45; Supplementary Fig. 6). In addition, as mentioned above, numerical values obtained by each method were distributed in different ranges and were inconsistent in many cases. This disparity may be due to difficulties in accurate quantification using the CS-35 anti-LAM antibody and raises the question of the molecular structure of LAM released into human lungs. Having access to the structure of Mtb cell envelope compounds *in vivo* would give invaluable insight into the metabolic status of Mtb bacilli during human infection, a key step for the development of new chemo- and immuno-therapeutic or -diagnostic strategies[19]. We used two pooled EBC samples, one collected from 50 smear-positive adult TB patients (Ad S^+ pool) and the other from 50 smear- or culture-positive pediatric TB patients (Ch

189 S^+/C^+ pool) (Supplementary Table 1). The pooled samples should contain levels of LAM
190 (several tens of micrograms) that are sufficient for structural analysis using a combination of
191 highly sensitive analytical procedures we previously developed[20]. Ad S^+ pool and Ch S^+/C^+
192 pool yielded 630 and 720 mg respectively of dried total material after freeze-drying
193 (Supplementary Table 2). Lipids (around 1.5-2 % w/w) were removed by organic solvent
194 extraction, and the remaining material was submitted to enzymatic digestion to remove nucleic
195 acids and proteins. The resulting LAM-enriched fractions contained 0.18 (Ad S^+ pool) and 0.36
196 (Ch S^+/C^+ pool) mg of arabinose respectively (Supplementary Table 2). The fractions were
197 analyzed by NMR using a 600 MHz spectrometer equipped with a cryogenic probe. The
198 anomeric region of the 2D ^{1}H-^{13}C HSQC spectrum obtained with LAM-enriched fractions from
199 both Ad S^+ pool (Fig. 2a) and Ch S^+/C^+ pool (Fig. 2b) exhibited intense cross peaks that could
200 be unambiguously assigned to glycosidic units building LAM (Supplementary Fig. 1), i.e. 6-α-
201 Man*p* (VI), 2-α-Man*p* (VII), 2,6-α-Man*p* (VIII), t-α-Man*p* (IV), 5-α-Ara*f* (II), 3,5-α-Ara*f* (I), t-
202 β-Ara*f* (V), 2-α-Araf→5 (IIIa) and 2-α-Araf→3 (IIIb) residues, identical to that found in LAM
203 purified from *M. tuberculosis* grown in broth (Mtb_broth) (Fig. 2c; Supplementary Table 3)[20-
204 22]. This indicates that the LAM contained in EBCs from both adult and pediatric patients has
205 an overall intact polysaccharidic structure, in agreement with the fact that LAM in EBCs is
206 retained in dialysis tubing with a molecular weight cut off of 8-10 kDa. However, two additional
207 cross-peaks were observed, previously assigned to t-α-Araf→5 (Xa) and t-α-Araf→3 (Xb) units
208 evidenced in an EmbC mutant[22]. These residues typify a "non-mature" LAM, containing a
209 reduced amount of branched hexa-arabinofuranosides (Ara_6)[22,23] and an unusual Ara_5 motif
210 (Fig. 2). Ara_6 is the main epitope of the CS-35 anti-LAM antibody[17]. In LAM from slow-
211 growing mycobacteria, including *M. tuberculosis*, some Ara_6 chains are substituted at their non-
212 reducing ends by mannose caps[6] (Fig. 2, Supplementary Fig. 1) that are key for LAM
213 immunomodulatory properties[24]. 2-α-Man*p* are specific units of mannose caps. Corresponding

signals (VII) were detected in NMR spectra of LAM-enriched fractions from both Ad S^+ pool and Ch S^+/C^+ pool (Fig. 2; Supplementary Table 3), albeit with a weaker intensity than in Mtb_broth LAM. This is consistent with the truncation of some of arabinan side chain termini that bear caps (Fig. 2). The presence of mannose caps on LAM in EBCs was confirmed by CE-LIF analysis after mild acid hydrolysis of LAM[25]. The electropherogram obtained with LAM-enriched fractions from both Ad S^+ pool and Ch S^+/C^+ pool exhibited peaks corresponding to mono- (AM), α(1→2)-di- (AMM) and α(1→2)-tri-mannoside units (Supplementary Fig. 7a), as in Mtb_broth LAM[6]. Quantification[25] indicated that there are approximately 1 mannosyl, 1.5 dimannosyl and 0.3 trimannosyl units per LAM molecule in both the Ad S^+ and Ch S^+/C^+ pools, indicating a 3- to 4-fold reduced number of di- and tri-mannoside caps compared to that in Mtb_broth LAM[6]. LAM purified from Mtb_broth also contains minor covalent modifications on its polysaccharidic domain, such as methylthioxylose or succinyl motifs (Supplementary Fig. 1; Fig. 2), whose biological significance remains largely unknown[6]. However, the signals corresponding to both motifs were absent in NMR spectra recorded for LAM-enriched fractions from both the Ad S^+ and Ch S^+/C^+ pools (Fig. 2). Finally, the LAM polysaccharidic domain is attached to a mannosyl-phosphatidyl-*myo*-inositol (MPI) lipid anchor (Supplementary Fig. 1)[6]. Fatty acyl chains esterifying the MPI anchor allow LAM to migrate in SDS-PAGE[26]. LAM-enriched fractions from Ad S^+ pool and Ch S^+/C^+ pool were subjected to SDS-PAGE followed by western blot analysis. Bands corresponding to LAM were observed, indicating that LAM in EBCs likely contains an intact MPI anchor (Supplementary Fig. 7b). LAM in EBCs showed an apparent molecular weight that is slightly lower than that of Mtb_broth LAM, in agreement with a partial truncation of the arabinan side chain termini.

Altogether, detailed structural analyses indicate that LAM in EBCs from both adult and pediatric patients has an overall intact polysaccharide structure compared to LAM purified from *M. tuberculosis* H37Rv grown in broth. However LAM from patients has significant

modifications: i) a reduction in branched Ara_6 motifs with "non-mature" truncated arabinan side chain termini (Ara_5), ii) no MTX motif, and iii) no succinylation detected. This analysis of LAM structure in patients is helpful for the rational development or improvement of diagnostic antibodies, or vaccines, targeting LAM[27].

Structure of *Mtb* lipids in TB patient EBCs typifies bacterial metabolism in human lungs

M. tuberculosis produces highly abundant and structurally diverse lipids, some of which are species-specific[28]. As shown above, EBCs contain 1.5-2 % (w/w) lipids (Supplementary Table 2). We tested whether *M. tuberculosis* lipids can be detected using mass spectrometry on pooled EBC lipid extracts. Five families of lipids, namely phosphatidylinositol mannosides (PIMs), sulfoglycolipids (SGLs), phthiocerol dimycocerosate (PDIM), mycolic acids (MA) and tuberculosinyladenosine (TbAd) were detected by either MALDI-TOF and/or SFC-HRMS (Supercritical Fluid Chromatography-High Resolution Mass Spectrometry) in both Ad S^+ and Ch S^+/C^+ pooled samples (Fig. 3). As in Mtb_broth, PIM molecular species were present in two glyco-forms, Phosphatidylinositol di- (PIM_2) and hexa- (PIM_6) mannosides, existing in 3 main acyl-forms esterified by 2 (PIM_x) to 4 (Ac_2PIM_x) fatty acyl chains (Fig. 3a)[29,30]. However, the acyl-form profile was more complex in EBCs, with a high abundance of unusual forms containing stearic acid (Fig. 3a; Supplementary Table 4). The PIM biosynthetic precursor, phosphatidylinositol (PI), was also detected (Fig. 3a). Negative MALDI mass spectrum also revealed a large distribution of tetra-acylated forms of SGLs (Ac_4SGLs) as observed in Mtb_broth[31,32], but with a shift towards higher molecular masses (Fig. 3b; Supplementary Table 5). This higher mass represented a mean extension of 7-carbon atoms per each of the 3 (hydroxy)phthioceranyl chains present on the molecule. The PDIM acylform profile was also modified, showing an increased proportion of the molecular species composed of longer alkyl chains (Fig. 3c; Supplementary Table 6). This likely represents a 3-carbon atom extension in

each of the two mycocerosic acids that compose the molecule[33]. Interestingly, similar increases of Ac_4SGLs and PDIM molecular masses have been previously reported for Mtb grown *in vitro* with propionate or cholesterol as the limiting carbon source, or in lungs of infected mice[33,34]. Methoxy- and α-mycolic acids were detected in their free fatty acid form (Figs. 3d and e; Supplementary Table 7). The distribution of the molecular species was very similar to that observed for mycolic acid esters in Mtb_broth[28,35], the most abundant species being α-mycolic acids containing 78- or 80-carbon atoms (*m/z* at 1136.171 and 1164.211 respectively; Figs. 3d) and methoxy-mycolic acids containing 85- or 87-carbon atoms (*m/z* at 1252.295 and 1280.327 respectively; Figs. 3e). Notably, free mycolic acids are only present in trace amounts in planktonically grown cells of *M. tuberculosis*, in contrast to biofilm cultures where they are more abundant[36]. Finally, TbAd was detected in both isomeric forms (1-TbAd and N^6-TbAd) at *m/z* 540.356[37] (Fig. 3f). We then analyzed individual EBCs and observed that, when detected, the different *M. tuberculosis* lipids showed a profile of molecular species similar to that found in pools. This finding was consistent across TB patients, whatever their clinical and demographic characteristics. This lipid profile thus typifies the *in vivo* metabolism of bacilli in human lungs, suggesting bacteria are growing as a biofilm with host lipids as the carbon source. In order to further characterize Mtb lipids as potential biomarkers, we performed a comparative analysis of the relative abundance of MA, PDIM and TbAd in the different samples based on mass spectrometry signal intensities (relatively to the signal of the 1,2-ditridecanoyl-*sn*-glycero-3-phosphocholine used as internal standard) (Fig. 4). The relative proportions of methoxy- *vs* α-mycolic acids were constant across samples. Therefore, Figs. 4a-c show the cumulative relative abundances of both mycolic acid types. The overall distribution of MA, PDIM and TbAd was very similar to that obtained for LAM quantification (Fig. 1). The best correlation with LAM was obtained for MA ($r = 0.54$, $P < 0.0001$; Supplementary Fig. 6). Measurements of MA and TbAd showed a very good correlation ($r = 0.90$, $P < 0.0001$;

Supplementary Fig. 6). The sensitivity of detection was higher for MA and TbAd than for PDIM. Indeed, as shown in Figs 4h and i, we did not detect PDIM for smear-negative and culture-negative children (Ch S^-C^-), in contrast to LAM (Figs 1c and e), MA (Figs 4b and c) and TbAd (Figs 4e and f). However, we only used 6% of the EBC for lipidomic analyses; there is therefore at least a 15-fold possible improvement in sensitivity. As shown for LAM, measurement of Mtb lipids in EBC may be used to diagnose TB in adult and pediatric patients, and to monitor antibiotic treatment efficacy. In contrast to LAM, which is shared by other mycobacterial species, some of these lipids are Mtb-specific.

EBCs from TB patients contain *M. tuberculosis* proteins found in extracellular vesicles

We next analyzed the protein content of EBCs by proteomics. A total number of 1432 Mtb proteins was detected in the 5 individual EBCs analyzed from the smear positive adults (Ad S^+) (Supplementary Table 8), with a number of proteins ranging from 147 to 1288 in different patients (Fig. 5a). Only 23 Mtb proteins were detected in smear- or culture-positive children (Ch S^+/C^+) (Supplementary Table 8), with a number ranging from 14 to 17 according to the patient (Fig. 5a). In contrast, no Mtb proteins were found in the EBCs of the 5 individuals in each control group (Ad healthy, Ch healthy, Ad pneumo) analyzed (Supplementary Table 8; Fig. 5a; Supplementary Fig. 8). All Mtb proteins detected are listed in Supplementary Data 1. Interestingly, many of them, particularly the most abundant, were previously described to be released in bacterial- and/or host-derived extracellular vesicles[38-41] (Supplementary Data 2). Protein abundance was determined by Label-Free Quantification (LFQ) (Supplementary Data 2). Changes in abundance (ratio = 1.5) of 656 proteins in adult and 22 proteins in pediatric patients differentiated infected individuals from controls ($p < 0.05$) (Supplementary Data 2; Supplementary Fig. 8). Fig. 5 shows the results for selected proteins, GroS (Figs. 5b and g), GroEL2 (Figs. 5c and h), GlnA1 (Fig. 5d), HspX (Fig. 5e and i) and SodB (Fig. 5f). Their

abundance was higher in adults, except for SodB, which was slightly higher in children. Among the pediatric patients who were followed during the course of antibiotic treatment (Ch S^-C^-, Table 1), the EBCs of 2 (# I and D) were analyzed by proteomics. As observed for LAM and lipids (Figs. 1 and 4), the abundance of all Mtb proteins, other than HspX, remained high, or even increased, during treatment in EBCs of patient #I (Figs. 5g-i; Supplementary Data 2). Finally, we determined the relative abundance of selected Mtb proteins by immunoassay. We tested several monoclonal antibodies against Mtb proteins provided by BEI (www.beiresources.org), including antibodies against GroEL2 (CS-44), HspX (NR-13607), KatG (NR-13793), LpqH (NR-13792) and HBHA (NR-13804). Among these, only the anti-GroEL2 antibody showed enough sensitivity in our dot-blot assay. Abundance of GroEL2 in EBCs (Figs. 5j-l) gave a distribution pattern very similar to that obtained for LAM (Fig. 1) and lipids (Fig. 4), and showed a very good correlation with quantity of LAM ($r = 0.90$, $P < 0.0001$; Supplementary Fig. 6). Although this approach lacked the sensitivity to detect smear-negative patients (Ad S^-C^+ and Ch S^-C^-), that were under antibiotic treatment for 2 weeks or less (giving positive signals for LAM, MA and TbAd), it allowed us to distinguish the other baseline TB patients from control individuals.

DISCUSSION

Improving early diagnosis is key to controlling TB burden by decreasing the risk of patient mortality and limiting transmission[1]. Sputum is currently the main sample used for microbiological diagnosis of TB. However sputum-based assays, including microscopic examination, nucleic acid amplification or bacterial culture, have clear limitations, particularly in children and paucibacillary patients[1]. Here, we tested whether EBCs are an accessible sample for the detection of TB with the potential for diagnostic development[3]. Whereas previous studies failed to detect Mtb nucleic acids in EBC of smear- and culture-positive pulmonary TB patients[42,43], our findings show that assay of bacterial LAM, lipids or proteins efficiently distinguished baseline TB patients. Patients that were smear-negative and culture-negative adults were identified as having active TB using EBC samples. In addition, culture- or smear-positive, and smear-negative and culture-negative, pediatric patients were detected as TB-positive by EBC analysis. The test specifically detected TB patients relative to healthy controls and individuals with community-acquired bacterial pneumonia. Detection of bacterial biomarkers in EBCs fulfills many criteria that are essential for implementation of point-of-care tests: i) EBC collection is easy, rapid, affordable, non-invasive and does not require specialized personnel; ii) although exhaled breath may contain viable Mtb[44-46], EBC samples can be collected in a sterile manner by adding a filter to the collecting device; thus, handling and analysis do not need to be performed in a high level biosafety laboratory; iii) Mtb-derived molecules in EBCs are highly abundant and very diverse; they could be assayed with previously developed tools, such as LAM tests developed for urine[5,7,47,48]. This is exemplified by the LAM assay: LAM is ~1000-fold more abundant in EBC than in urine. EBCs are a promising sample for clinical diagnosis and monitoring. However, widespread use still requires standardization in order to reduce the inter- and intra-subject variability observed across studies for non-volatile compound concentration. Thus the normal physiological ranges of the various potential human

biomarkers need to be defined[8,9]. Although important, normalization appears to be less critical for bacterial biomarkers that are likely to serve as positive/negative indicators of infection[9]. Mtb molecules can be readily and specifically detected in EBC samples, potentially allowing an immediate TB diagnosis and an initiation of treatment during the same clinical encounter. In addition, this approach may address key priorities for development of TB diagnostics[3,47,49] by allowing, i) detection of TB in children and in paucibacillary specimens, ii) diagnosis of a higher proportion of pediatric TB cases using a sampling method that children and parents are comfortable with, iii) screening for community-based case-finding or for triaging patients in the clinic.

To date, the only prognostic marker for end-of-treatment outcome remains culture conversion status, though the sensitivity of this assay is poor[50]. Our data indicate that detection of bacterial biomarkers in EBCs may also provide a tool for real-time monitoring of treatment efficacy, enabling the early detection of disease relapse. The EBC-based analysis would allow clinicians to respond quickly to the therapeutic needs of patients, and may facilitate analysis of efficacy in clinical trials for personalized, anti-TB therapies.

A diverse range of Mtb molecules with unexpected high levels were found in EBCs from TB patients. Whereas D-arabinose derived from LAM was previously found in the range of ~10-40 ng/mL in the urine of TB patients[18], we detected it in the range of ~15-120 µg/mL in the EBCs of TB patients at baseline. These data suggest that patients exhale a large quantity, between 0.3 and 2.9 mg, of LAM per day. This represents the amount of LAM found in ~10^{11} to 10^{12} CFUs of Mtb grown in broth, and reflects highly metabolically active bacilli. In the present study, we did not investigate EBCs from patients with extrapulmonary TB. Although the amount of LAM or other Mtb molecules might be drastically reduced, the sensitivity of the current tools for LAM detection in the ng/ml range might be sufficient for the diagnosis of extrapulmonary TB through analysis of EBC samples, in those cases where lungs are also

affected. Nevertheless, where required, our data provide a basis for the rational improvement of LAM detection tools[5,7]. By performing a full NMR characterization of LAM present in the lungs of TB patients, our data give insight into the molecular moieties that are best targeted for efficient antibody recognition. LAM in EBCs of both adult and children exhibited an overall structure similar to that found in LAM purified from *M. tuberculosis* grown in broth. However the structure contains significant modifications that are important for the development of diagnostic antibodies. Most particularly, LAM in EBC contains a greatly reduced amount of branched Ara_6 motifs that are very immunogenic, and a major epitope for anti-LAM antibodies[17,27,51]. This may explain, at least in part, the discrepancies we observed in the LAM quantity values obtained by antibody vs. chemical assays. Generating antibodies that specifically recognize the "non-mature" Ara_5 motif (Fig. 2), found in both EBC and urine[23], should significantly improve the sensitivity of LAM detection in these samples. LAM succinylation was not detected in EBCs, although LAM was found to be hyper-succinylated in the lungs of Mtb-infected C3HeB/FeJ mice[23]. LAM in EBC also seems to be devoid of the MTX motif. However, this motif was detected on LAM-derived oligosaccharides from a TB patient's urine[23]. Accordingly, an immunoassay targeting the MTX motif and tested on urine samples performed well for TB diagnosis[52]. It is worth noting that LAM structure is not specific to Mtb but is shared by other mycobacteria[6]. Assessing LAM levels might be insufficient to ascertain Mtb *vs* non-tuberculous mycobacterial infection in some cases, most particularly if only minute amounts are detected. This might be improved by combining detection of LAM with Mtb-specific lipids, such as SGLs or TbAd, or proteins once appropriate antibodies are developed.

From a fundamental biology perspective, EBCs provide an unprecedented window into Mtb lifestyle in human lungs. It represents an invaluable opportunity to understand how the bacterium realigns its metabolism in response to the environments it encounters during

infection, a key step for the development of new chemo- and immunotherapeutic strategies[19]. In addition to structural modifications in LAM mentioned above, we observed a remodeled structure of PIMs, Ac_4SGLs and PDIM. Similar profiles for Ac_4SGLs and PDIM with an increased proportion of high mass molecular species have been previously reported for Mtb grown *in vitro* on propionate or cholesterol as the limiting carbon source, or in lungs of infected mice[33,34]. Although the structure of LAM and lipids in EBC might not be fully representative of all the LAM and lipid molecules produced and circulating in the body of TB patients, our data indicate that Mtb lipid anabolism is increased in the lung environment, which may support bacterial virulence in this tissue[33]. Interestingly, EBCs contain high amounts of free mycolic acids that typify Mtb biofilm growth[36]. Although it has long been a matter of debate, a recent study shows that Mtb biofilms are relevant to human TB[53] and represent a drug target[36].

In conclusion, our findings indicate that detection of bacterial molecules in EBCs provides an affordable and rapid approach that seems superior to the sputum assay for early diagnosis of TB disease. Moreover, EBC sampling has potential for community-based case screening and for monitoring treatment efficacy in real-time. In addition, our analysis of TB patient EBCs indicates that these samples can be used to monitor changes in Mtb metabolism during infection. Future work will develop the diagnostic method and potential biomarkers that have been newly identified here in larger scale clinical studies. This will include validating the proof-of-concept data obtained in the present pilot study using larger cohorts and different groups of patients, including HIV-infected individuals, from different regions of the world. As a first step toward a transfer to the field, we are now evaluating the efficacy of diagnosing TB using EBC samples and currently commercialized LAM detection tests[5,7]. However, we conclude that sampling exhaled air is non-invasive and appears to be a powerful approach that overcomes the limitations in existing assays[45,46,54-58]. An EBC-based approach therefore has potentially broad application for following other pulmonary infectious agents.

METHODS

Patients and control individuals

Adults and pediatric patients with pulmonary TB were recruited through the TB Control Program of the Secretaría de Salud de Medellín and the Secretaría Seccional de Salud y Bienestar Social de Antioquia, Colombia. The diagnostic tests for TB and HIV were performed in the laboratories of health centers where the patients received primary care. Pulmonary TB diagnosis was performed according to the recommendations of Colombia Ministry of Health and Social Protection (https://www.minsalud.gov.co/salud/publica/PET/Paginas/Tuberculosis.aspx), based on international guidelines[13], and was defined as clinical, and/or radiographic, and/or bacteriological evidence of pulmonary disease resulting in treatment initiation by the diagnosing clinician. After collection of three sputum samples and analysis for acid-fast bacilli smear and mycobacterial culture, adult patients were defined to have culture-positive pulmonary TB if they had Mtb growth in at least one of the three sputum cultures, and to have culture-negative pulmonary TB if they had no Mtb growth in the three initial sputum samples, but clinical and radiographic presentation consistent with TB[13,14]. Smear positivity was defined as having at least one positive acid-fast bacilli smear (Ziehl-Neelsen staining), regardless of quantity, and smear negativity as having no acid-fast bacilli detected in the three initial sputum smears. Pediatric patients were categorized as string test smear- or culture-positive if they had Mtb growth or direct visualization (Ziehl-Neelsen staining) in string test samples, and as string test smear- and culture- negative if they had no Mtb growth or direct visualization in string test samples, but clinical and radiographic presentation consistent with TB, and improvement of symptoms following anti-tuberculous treatment[1,15].

The first EBC sample was collected at baseline (before or within the first 2 weeks of beginning anti-TB treatment). Additional EBC samples were obtained for children at different time-points

during the anti-TB treatment. Subjects participating in the study, or their legal guardians, read and provided written informed consent, as previously approved by the Ethics Committee of the Facultad de Medicina, Universidad de Antioquia. They did not receive any compensation. Exclusion criteria were: patients positive for HIV, diabetes, cancer, autoimmune diseases, immunosuppressive treatment, smoking, or previous TB. These exclusions allowed us to focus on assessing TB in the absence of factors that influence disease or might confound detection. The groups of individuals who participated in this study include:

Smear-positive adult patients. Eight patients between 27 and 62 years old, diagnosed with pulmonary TB by direct Ziehl-Neelsen staining and culture of sputum (Table 1). These patients were under anti-TB treatment for fifteen days or less.

Smear-negative adult patients. Twenty-one patients with at least three sputum samples negative for Ziehl-Neelsen staining (Table 1). Seven patients (aged 19-70) were culture-negative. Fourteen patients (aged 21-68) were either positive for sputum culture or, for four of them, positive for bronchoalveolar lavage (BAL) direct staining or BAL Mtb culture. These fourteen patients were under anti-TB treatment for fifteen days or less.

Pediatric TB patients. The diagnostic criteria included a history of contact with a tuberculous adult, cough lasting longer than 2 weeks, reactive tuberculin skin test, and radiographic findings compatible with TB[1,15]. The children (aged 6-12) were separated into two groups based on the detection or not of mycobacteria as indicated above and in Table 1. The latter were under an anti-TB treatment regimen for fifteen days or less. EBC samples were collected for 6 of these patients 1 and 3 months after treatment.

Control patients with pneumonia. Fifteen adults (aged 18-61) who presented with Community-Acquired Pneumonia were included (Table 1). Inclusion criteria were age ≥ 18, fever (>38 °C), cough, increased respiratory rate or respiratory distress, infiltrates on chest radiographs and bacterial pneumonia confirmed in sputum. Exclusion criteria were immunodeficiency, chronic

lung or heart diseases, neoplasia, hospital-acquired pneumonia or viral pneumonia. Twelve patients had pneumonia with consolidation and three without. All of them presented complicated pneumonia, including 8 bacteremias, 4 empyemas, and 3 pleural effusions.

Healthy control individuals. Fifteen healthy adult individuals (aged 26-49; Table 1) who did not have respiratory symptoms nor a history of tuberculosis were recruited by the medical and laboratory personnel from the Facultad de Medicina, Universidad de Antioquia. Fifteen healthy children (aged 7-12; Table 1) who had no previous contact with TB patients nor respiratory symptoms were also recruited. The parents consented to participation of children in this study and informed the researchers that the children were BCG vaccinated.

Exhaled Breath Condensate (EBC) collection

The EBCs were collected using R-tubes (Respiratory Research, Inc). The R-tube™ is a noninvasive device that allows the collection and condensation of components of the expiration. R-tubes were equipped with a 0.3-µm filter (Respiratory Research, Inc) between the mouthpiece assembly and the condensation cartridge trap to prevent bacilli contaminating EBCs. Subjects breathed through a mouthpiece and a two-way non-rebreathing valve, which also served as a saliva trap. They were asked to breathe at a normal frequency for 15 min. At least 1 ml of sample was collected and immediately frozen at -20ºC. To determine possible saliva contamination in the EBC samples, α-amylase was measured by the Fishman-Doubilet test using a commercial kit and following the manufacturer's instructions (Abnova, Taipei, Taiwan). EBC samples that were positive for this test were discarded. Accordingly, no α-amylase was detected by proteomic analysis in the EBCs used in the present study, confirming the absence of saliva contamination[59].

For normalization between individuals, EBCs were lyophilized and adjusted to a volume of 250 µl of LPS-free water.

EBC samples were unique, but measurements were repeated between one and four times.

Dot-Blot immunoassays

EBCs (1 µl) were spotted using the Bio-Dot SF blotting apparatus (Bio-Rad) on a 0.45 µm nitrocellulose membrane (Bio-Rad) previously immersed for 5 min in Tris-buffered saline (TBS). LAM purified from *M. tuberculosis* H37Rv[21,60] or a Mtb cell lysate (obtained by sonication with probe) were used as standards and deposited at different amounts (in 1µl). The membrane was first blocked for 2 h at RT in TBS with 0.01% Tween-20 and 3% skimmed milk powder. It was subsequently incubated overnight at 4°C with the primary antibodies, mouse IgG3 anti-LAM antibody (CS-35, BEI Resources; dilution of 1:200) or mouse IgG2b anti-GroEL2 (CS-44, BEI Resources; dilution of 1:1000), in TBS supplemented with 1.5% skimmed milk powder. Horseradish peroxidase (HRP)-conjugated secondary antibodies, goat anti-mouse IgG3-HRP (ThermoFisher; dilution of 1:2000) or goat anti-mouse IgG2b-HRP (Invitrogen; dilution of 1:5000), in TBS supplemented with 1.5% skimmed milk powder were incubated with the membrane for 2 h at RT. After each antibody incubation step, the membrane was washed 3 times with TBS supplemented with 0.01% Tween-20. Membranes were processed to reveal signal using Amersham ECL Prime™ and images were captured using the ChemiDoc System (Bio-Rad). Quantification of dot density was performed using Image Lab™ Software (Bio-Rad; version 6.0.1 build 34).

The following monoclonal antibodies, provided by BEI Resources, were also tested: anti-HspX (IT-20; dilution of 1:100), anti-KatG (IT-57; dilution of 1:200), anti-LpqH (IT-54; dilution of 1:200), anti-HBHA (α-HBHA; dilution of 1:50).

Monosaccharide analysis and quantification by capillary electrophoresis monitored by laser-induced fluorescence (CE-LIF)

EBCs (12 to 25 µl) were submitted to total acid hydrolysis using 2 M trifluroacetic acid for 2 h at 110°C, in the presence of 1 nmol mannoheptose (Sigma) as an internal standard. The samples were dried and mixed with 0.3 µl of 0.2 M 8-aminopyrene-1,3,6-trisulfonic acid (APTS; Sigma) in 15% acetic acid and 0.3 µl of a 1 M sodium cyanoborohydride solution dissolved in tetrahydrofuran. The reaction mixture was heated at 55 °C for 90 min and subsequently quenched by the addition of 20 µl of water. Monosaccharide APTS-derivatives were analyzed and quantified by capillary electrophoresis monitored by laser-induced fluorescence on a P/ACE capillary electrophoresis system (Beckman Instruments, Inc.). Separations were performed using an uncoated fused-silica capillary column (Sigma) of 50-µm internal diameter and 40-cm effective length (47-cm total length). Analyses were carried out at a temperature of 25 °C, in the reverse mode, with an applied voltage of 20 kV using 1% acetic acid (w/v)-30 mM triethylamine, pH 3.5, as a running electrolyte[25].

Preparation of LAM-enriched fractions from EBC pools

The LAM-enriched fraction from EBC pools was prepared using an approach similar to that described for LAM purification from bacteria[61]. Lipids were removed by chloroform/methanol extraction (partition chloroform/methanol/water 1.25:1.25:1). The aqueous phase was dried and submitted to digestion by DNAse (D4263, Sigma), RNAse (R4875, Sigma) and a cocktail of proteases (α-chymotrypsin (C4129, Sigma-Aldrich), *Streptomyces griseus* proteases (P8811, Sigma) and trypsin (T3914, Sigma-Aldrich)) to remove nucleic acids and proteins. The digested solution was dialyzed against water (MWCO 6-8 kDa) and dried, resulting in a fraction enriched in LAM.

Monosaccharide analysis by gas chromatography coupled to mass spectrometry (GC-MS)

554 Monosaccharides were analyzed as their trimethylsilyl (TMS) derivatives after total acid
555 hydrolysis (2 M trifluroacetic acid for 2 h at 110°C) of EBC pools (5 mg), drying, and
556 derivatization with pyridine/hexamethyldisilazane/chlorotrimethylsilylane, 4:2:1, v/v/v, for 15
557 min at room temperature, and injection in GC-MS in electron ionization mode.
558 To determine the absolute configuration of arabinose, EBC pools (5 mg) were octanolysed
559 using 2 M trifluoroacetic in (R)-(−)-2-butanol (Sigma) for 16 h at 110°C[62]. Octyl-glycosides
560 were dried and converted into their TMS derivatives, as detailed above, before analysis by GC-
561 MS and comparison with D-Ara and L-Ara standards submitted to the same procedure.
562 GC-MS was performed on a Trace 1300 Thermo GC equipped with a TG-SQC column (15-m,
563 0.25 mm inner diameter, 0.25 µm film thickness) coupled to an ISQ quadrupole mass
564 spectrometer. The carrier gas was helium at a flow rate of 1 ml/min. The injector temperature
565 was 250°C and the temperature separation program was from 100 to 300 °C, at a speed of
566 15°C/min.
567
568 **Mannose cap and western blot analyses of LAM-enriched fractions**
569 For mannose cap analysis and quantification, LAM-enriched fractions (100 µg) were subjected
570 to mild acid hydrolysis using 0.1 M HCl for 20 min at 110°C, in the presence of 0.1 nmol
571 mannoheptose (Sigma) as an internal standard. Samples were dried, derivatized by APTS and
572 analyzed by capillary electrophoresis as described above[25].
573 For western blot analysis, LAM-enriched fractions (2-10 µg) were subjected to SDS-PAGE
574 (15% separating gel)[61] and transferred to a nitrocellulose membrane, which was subsequently
575 probed with CS-35 anti-LAM antibody as described above.
576
577 **MALDI-TOF analysis**

MALDI-TOF analyses were performed on an AB Sciex TOF/TOF 5800 mass spectrometer using the reflectron mode[29,63]. Ionization was achieved by irradiation with an Nd:YAG laser (349 nm) operating with a pulse rate of 400 Hz. The laser intensity was set at 3500 and continuous stage motion was used with a velocity of 600 µm/sec. Data was acquired in the negative ion mode. Typically, spectra from 2500 to 5000 laser shots were summed to obtain the final spectrum. 2,5-dihydroxybenzoic acid (Sigma) was used as the matrix at a concentration of 10 mg/mL in chloroform/methanol 9:1 or 1:1.

NMR analysis

NMR spectra were recorded at 305K without sample spinning on a Bruker 600 MHz avance III HD (1H) equipped with a TCi cryoprobe (Bruker Biospin, Germany). LAM-enriched fractions were exchanged in D_2O (D, 99.97% from Euriso-top, Saint-Aubin, France), with intermediate lyophilisation, and then dissolved in 0.5 ml D_2O. Samples were analyzed in 200 x 5 mm UL-5 NMR tubes (Euriso-Top, Saint-Aubin, France). Proton and carbon chemical shifts are expressed in parts per million and referenced relative to internal acetone signals at δ_H 2.225 and δ_C 31.45. ^{1}H-^{13}C correlation spectra were acquired in the echo/anti-echo acquisition mode recorded in the proton-detected mode using the "hsqcetgpsisp2.2" sequence from the Topspin 3.5 PL7 software (Bruker Biospin, Germany). The 2D ^{1}H-^{13}C spectra were recorded with a spectral width of 22,636 Hz in ^{13}C and 7,211 Hz in ^{1}H dimensions in order to collect a 2048x512 point data matrix with 16 scans per t1 value expanded to 4096x1024 by zero filling. The relaxation delay was 1.5s. A sine bell window shifted by $\pi/2$ was applied in both dimensions[20,21].

Supercritical Fluid Chromatography (SFC)-MS lipidomic analysis

EBCs (30 µl) were mixed with 4 ng of 1,2-ditridecanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids), used as an internal standard allowing the quality control of extraction and normalization for comparative analyses, 12 µl of water and 105 µl of chloroform/methanol 1:1 in a glass vial. After vortexing and phase separation, the organic phase was recovered for SFC-MS analysis at the MetaToul Lipidomic Core Facility (I2MC, Inserm 1048, Toulouse, France) instrument. The Ultra-Performance Convergence Chromatography (UPC2) was coupled on-line to a Xevo G2-XS time of flight (QTOF) (Waters, Milford, MA, USA), equipped with ESI under the control of MassLynx software, version 4.1. The analysis was performed in both ionization modes (positive and negative) in two separate runs. 1 µL of lipid extract was then injected into the ACQUITY UPC2 Torus Diol column (100 × 3.0 mm inner diameter (i.d.), particle size: sub-1.7 µm, Waters) at 48 °C. The lipid families were separated using a gradient of supercritical CO_2/CH_3OH (1.25mg/mL ammonium formate) (1%-50%) at a 1.4 mL/min flow rate and using an active back pressure regulator (ABPR) of 1.900 pounds per square inch (psi). The make-up solvent was CH_3OH used at 0.25 mL/min. The source parameters were set as follows: source temperature 120 °C; the capillary voltage 3.2 kV in positive mode and −3kV in negative mode; the desolvation gas flow rate 690 mL/Hr in positive mode and 740 mL/Hr in negative mode; the cone gas flow 40 L/Hr; the desolvation temperature 420 °C in positive mode and 390 °C in negative mode. The analyses were performed in MS full scan in centroid mode from 50 to 3000 Da. Mtb lipids families were detected with characteristic retention times and previously reported[64] acyl-form profiles and accurate *m/z*.

For each lipid family, a chromatogram of the main representative acyl-form was extracted at the corresponding *m/z* within 20 ppm. Area of the extracted ion chromatograms (EIC) was measured using Waters MassLynx™ Software v. 4.2. Peak area integration was done with the following parameters: window size (scans) ±3, number of smooths 2 and a mean smoothing method. The relative abundance of each lipid family was determined relatively to 1,2-

ditridecanoyl-*sn*-glycero-3-phosphocholine used as an internal standard, and calculated as the ratio of measured EIC areas of lipids *vs* internal standard.

Sample preparation for label-free proteomics analysis

For each sample 35 µg of dried protein extracts were solubilized with 25 µl of 5% SDS. Proteins were submitted to reduction and alkylation of cysteine residues by addition of TCEP and chloroacetamide to a final concentration respectively of 10 mM and 40 mM. Protein samples were then processed for trypsin digestion on S-trap Micro devices (Protifi) according to the manufacturer's protocol, with the following modifications: precipitation was performed using 216 µl S-Trap buffer, 1.5 µg trypsin was added per sample for digestion in 25 µl TEAB 50 mM pH 8.

NanoLC-MS/MS analysis of proteins

Tryptic peptides were resuspended in 35 µl of 2% acetonitrile and 0.05% trifluoroacetic acid and analyzed by nano-liquid chromatography (LC) coupled to tandem MS, using an UltiMate 3000 system (NCS-3500RS Nano/Cap System; Thermo Fisher Scientific) coupled to an Orbitrap Q-HFX mass spectrometer (Thermo Fisher Scientific). Around 5 µg of each sample was loaded on a C18 precolumn (300 µm inner diameter × 5 mm, Thermo Fisher Scientific) in a solvent made of 2% acetonitrile and 0.05% trifluoroacetic acid, at a flow rate of 20 µl/min. After 5 min of desalting, the precolumn was switched online with the analytical C18 column (75 µm inner diameter × 50 cm, Acclaim PepMap 2µm C18 Thermo Fisher Scientific) equilibrated in 95% solvent A (5% acetonitrile, 0.2% formic acid) and 5% solvent B (80% acetonitrile, 0.2% formic acid). Peptides were eluted using a 10%-45% gradient of solvent B over 60 min at a flow rate of 350 nl/min. The mass spectrometer was operated in data-dependent acquisition mode with the Xcalibur software (version 4.3.73.11). MS survey scans were

acquired with a resolution of 60,000 and an AGC target of $3x10^6$. The 12 most intense ions were selected for fragmentation by high-energy collision-induced dissociation, and the resulting fragments were analyzed at a resolution of 30,000, using an AGC target of $1x10^5$ and a maximum fill time of 54 ms. Dynamic exclusion was used within 30 s to prevent repetitive selection of the same peptide.

Bioinformatics analysis of proteomic datasets

Raw MS files were processed with the Mascot software (version 2.7.0) for database search and Proline[65] for label-free quantitative analysis (version 2.1.2). Data were searched against MTB_H37Rv_native and Human entries of the UniProtKB protein database (release Swiss-Prot+TrEMBL 2020_10, 4,080 entries and release Swiss-Prot 2020_10, 20,385 entries, respectively). Carbamidomethylation of cysteines was set as a fixed modification, while oxidation of methionine was set as variable modifications. Specificity of trypsin/P digestion was set for cleavage after K or R, and two missed trypsin cleavage sites were allowed. The mass tolerance was set to 10 ppm for the precursor and to 20 mmu in tandem MS mode. Minimum peptide length was set to 7 amino acids, and identification results were further validated in Proline by the target decoy approach using a reverse database at both a PSM and protein false-discovery rate of 1%. For label-free relative quantification of proteins across biological replicates and conditions, cross-assignment of peptide ions peaks was enabled inside a group with a match time window of 1 min, after alignment of the runs with a tolerance of +/- 600 s. Median Ratio Fitting computes a matrix of abundance ratios calculated between any two runs from ion abundances for each protein. For each pair-wise ratio, the median of the ion ratios is then calculated and used to represent the protein ratio between these two runs. A least-squares regression is then performed to approximate the relative abundance of the protein in each run in the dataset. This abundance is finally rescaled to the sum of the ion abundances across runs.

A Student t-test (two-tailed t-test, equal variances) was then performed on log2 transformed values and followed by Adjusted Benjamini–Hochberg (ABH) correction to analyze differences in protein abundance in all biologic group comparisons. Significance level was set at $p = 0.05$, and ratios were considered relevant if higher than +/- 2.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD028477.

Statistics and Reproducibility

EBC samples are unique. Values shown in graphs or tables are the mean of one to several independent measurements (technical replicates) on each individual EBC sample. Statistical differences between TB patient groups and controls (healthy, pneumo) were assessed using a two-tailed Mann-Whitney *U*-test and a α level of 0.05. Correlations between the different Mtb biomarkers were determined by Pearson correlation test. Analyses, including ROC, were done using Prism 5.0 (GraphPad Software). Leave-One-Out cross-validation analysis was carried out with R software (version 4.2.1)[66], using the *caret* package (version 6.0-93) for the cross-validation[67], and *tidyverse* (version 1.3.2) for data wrangling and graphical output[68].

DATA AVAILABILITY

Raw data are provided as a Source Data file. The mass spectrometry proteomics data were searched against MTB_H37Rv_native and Human entries of the UniProtKB protein database (release Swiss-Prot+TrEMBL 2020_10, 4,080 entries and release Swiss-Prot 2020_10, 20,385 entries, respectively), and have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD028477.

REFERENCES

1. WHO. Global Tuberculosis Report 2021. 1-43 (2021).
2. WHO. Global strategy and targets for tuberculosis prevention, care and control after 2015 1-3 (2014).
3. WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. 1-97 (2014).
4. Kik, S.V., Denkinger, C.M., Casenghi, M., Vadnais, C. & Pai, M. Tuberculosis diagnostics: which target product profiles should be prioritised? *Eur Respir J* **44**, 537-540 (2014).
5. Bulterys, M.A., *et al.* Point-Of-Care Urine LAM Tests for Tuberculosis Diagnosis: A Status Update. *J Clin Med* **9**, 111 (2019).
6. Nigou, J., Gilleron, M. & Puzo, G. Lipoarabinomannans: from structure to biosynthesis. *Biochimie* **85**, 153-166 (2003).
7. Correia-Neves, M., *et al.* Biomarkers for tuberculosis: the case for lipoarabinomannan. *ERJ Open Res* **5**, 00115-02018 (2019).
8. Kuban, P. & Foret, F. Exhaled breath condensate: determination of non-volatile compounds and their potential for clinical diagnosis and monitoring. A review. *Anal Chim Acta* **805**, 1-18 (2013).
9. Davis, M.D. & Montpetit, A.J. Exhaled Breath Condensate: An Update. *Immunol Allergy Clin North Am* **38**, 667-678 (2018).
10. Kharitonov, S.A. & Barnes, P.J. Exhaled biomarkers. *Chest* **130**, 1541-1546 (2006).
11. Mosquera-Restrepo, S.F., Caro, A.C., Garcia, L.F., Pelaez-Jaramillo, C.A. & Rojas, M. Fatty acid derivative, chemokine, and cytokine profiles in exhaled breath condensates can differentiate adult and children paucibacillary tuberculosis patients. *J Breath Res* **11**, 016003 (2017).
12. Guzman-Beltran, S., *et al.* Oxidative Stress and Inflammatory Mediators in Exhaled Breath Condensate of Patients with Pulmonary Tuberculosis. A Pilot Study with a Biomarker Perspective. *Antioxidants* **10**, 1572 (2021).
13. Lewinsohn, D.M., *et al.* Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clin Infect Dis* **64**, 111-115 (2017).
14. Nguyen, M.H., *et al.* Factors Associated With Sputum Culture-Negative vs Culture-Positive Diagnosis of Pulmonary Tuberculosis. *JAMA network open* **2**, e187617 (2019).
15. Starke, J.R. & Donald, P.R. (eds.). *Handbook of child and adolescent tuberculosis*, (Oxford University Press, New York, 2016).
16. Hamasur, B., *et al.* Rapid diagnosis of tuberculosis by detection of mycobacterial lipoarabinomannan in urine. *J Microbiol Methods* **45**, 41-52 (2001).
17. Kaur, D., Lowary, T.L., Vissa, V.D., Crick, D.C. & Brennan, P.J. Characterization of the epitope of anti-lipoarabinomannan antibodies as the terminal hexaarabinofuranosyl motif of mycobacterial arabinans. *Microbiology* **148**, 3049-3057 (2002).
18. De, P., *et al.* Estimation of D-Arabinose by Gas Chromatography/Mass Spectrometry as Surrogate for Mycobacterial Lipoarabinomannan in Human Urine. *PLoS One* **10**, e0144088 (2015).
19. Russell, D.G., *et al.* Mycobacterium tuberculosis wears what it eats. *Cell Host Microbe* **8**, 68-76 (2010).
20. Larrouy-Maumus, G., *et al.* A glycomic approach reveals a new mycobacterial polysaccharide. *Glycobiology* **25**, 1163-1171 (2015).

21. Gilleron, M., Bala, L., Brando, T., Vercellone, A. & Puzo, G. Mycobacterium tuberculosis H37Rv parietal and cellular lipoarabinomannans. Characterization of the acyl- and glyco-forms. *J Biol Chem* **275**, 677-684 (2000).
22. Shi, L., *et al.* The carboxy terminus of EmbC from Mycobacterium smegmatis mediates chain length extension of the arabinan in lipoarabinomannan. *J Biol Chem* **281**, 19512-19526 (2006).
23. De, P., *et al.* Comparative Structural Study of Terminal Ends of Lipoarabinomannan from Mice Infected Lung Tissues and Urine of a Tuberculosis Positive Patient. *ACS Infect Dis* **6**, 291-301 (2020).
24. Vergne, I., Gilleron, M. & Nigou, J. Manipulation of the endocytic pathway and phagocyte functions by Mycobacterium tuberculosis lipoarabinomannan. *Front Cell Infect Microbiol* **4**, 187 (2014).
25. Nigou, J., Vercellone, A. & Puzo, G. New structural insights into the molecular deciphering of mycobacterial lipoglycan binding to C-type lectins: lipoarabinomannan glycoform characterization and quantification by capillary electrophoresis at the subnanomole level. *J Mol Biol* **299**, 1353-1362 (2000).
26. Venisse, A., Berjeaud, J.M., Chaurand, P., Gilleron, M. & Puzo, G. Structural features of lipoarabinomannan from Mycobacterium bovis BCG. Determination of molecular mass by laser desorption mass spectrometry. *J Biol Chem* **268**, 12401-12411 (1993).
27. Corrigan, D.T., Ishida, E., Chatterjee, D., Lowary, T.L. & Achkar, J.M. Monoclonal antibodies to lipoarabinomannan/arabinomannan - characteristics and implications for tuberculosis research and diagnostics. *Trends Microbiol* (2022).
28. Minnikin, D.E. & Brennan, P.J. Lipids of Clinically Significant Mycobacteria. in *Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids* (ed. Goldfine, H.) 1-76 (Springer Nature Switzerland AG 2020, 2020).
29. Gilleron, M., Quesniaux, V.F. & Puzo, G. Acylation state of the phosphatidylinositol hexamannosides from Mycobacterium bovis bacillus Calmette Guerin and mycobacterium tuberculosis H37Rv and its implication in Toll-like receptor response. *J Biol Chem* **278**, 29880-29889 (2003).
30. Gilleron, M., Nigou, J., Nicolle, D., Quesniaux, V. & Puzo, G. The acylation state of mycobacterial lipomannans modulates innate immunity response through toll-like receptor 2. *Chem Biol* **13**, 39-47 (2006).
31. Layre, E., *et al.* Deciphering sulfoglycolipids of Mycobacterium tuberculosis. *J Lipid Res* **52**, 1098-1110 (2011).
32. Rhoades, E.R., Streeter, C., Turk, J. & Hsu, F.F. Characterization of sulfolipids of Mycobacterium tuberculosis H37Rv by multiple-stage linear ion-trap high-resolution mass spectrometry with electrospray ionization reveals that the family of sulfolipid II predominates. *Biochemistry* **50**, 9135-9147 (2011).
33. Jain, M., *et al.* Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. *Proc Natl Acad Sci U S A* **104**, 5133-5138 (2007).
34. Griffin, J.E., *et al.* Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. *Chem Biol* **19**, 218-227 (2012).
35. Marrakchi, H., Laneelle, M.A. & Daffe, M. Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol* **21**, 67-85 (2014).
36. Ojha, A.K., *et al.* Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* **69**, 164-174 (2008).
37. Layre, E., *et al.* Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc Natl Acad Sci U S A* **111**, 2978-2983 (2014).

38. Giri, P.K., Kruh, N.A., Dobos, K.M. & Schorey, J.S. Proteomic analysis identifies highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-treated macrophages. *Proteomics* **10**, 3190-3202 (2010).
39. Lee, J., *et al.* Proteomic analysis of extracellular vesicles derived from Mycobacterium tuberculosis. *Proteomics* **15**, 3331-3337 (2015).
40. Palacios, A., Gupta, S., Rodriguez, G.M. & Prados-Rosales, R. Extracellular vesicles in the context of Mycobacterium tuberculosis infection. *Mol Immunol* **133**, 175-181 (2021).
41. Prados-Rosales, R., *et al.* Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J Clin Invest* **121**, 1471-1483 (2011).
42. Shenai, S., *et al.* Exploring alternative biomaterials for diagnosis of pulmonary tuberculosis in HIV-negative patients by use of the GeneXpert MTB/RIF assay. *J Clin Microbiol* **51**, 4161-4166 (2013).
43. Jain, R., Schriever, C.A., Danziger, L.H., Cho, S.H. & Rubinstein, I. The IS6110 repetitive DNA element of Mycobacterium tuberculosis is not detected in exhaled breath condensate of patients with active pulmonary tuberculosis. *Respiration* **74**, 329-333 (2007).
44. Patterson, B., *et al.* Cough-independent production of viable Mycobacterium tuberculosis in bioaerosol. *Tuberculosis* **126**, 102038 (2021).
45. Wood, R., *et al.* Real-Time Investigation of Tuberculosis Transmission: Developing the Respiratory Aerosol Sampling Chamber (RASC). *PLoS One* **11**, e0146658 (2016).
46. Patterson, B., *et al.* Detection of Mycobacterium tuberculosis bacilli in bio-aerosols from untreated TB patients. *Gates Open Res* **1**, 11 (2018).
47. Walzl, G., *et al.* Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect Dis* **18**, e199-e210 (2018).
48. Goletti, D., Lee, M.R., Wang, J.Y., Walter, N. & Ottenhoff, T.H.M. Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. *Respirology* **23**, 455-466 (2018).
49. WHO. WHO consolidated guidelines on tuberculosis: Module 3: diagnosis - rapid diagnostics for tuberculosis detection 1-82 (2020).
50. Kurbatova, E.V., *et al.* Sputum culture conversion as a prognostic marker for end-of-treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis of data from two observational cohort studies. *Lancet Respir Med* **3**, 201-209 (2015).
51. Choudhary, A., *et al.* Characterization of the Antigenic Heterogeneity of Lipoarabinomannan, the Major Surface Glycolipid of Mycobacterium tuberculosis, and Complexity of Antibody Specificities toward This Antigen. *J Immunol* **200**, 3053-3066 (2018).
52. Sigal, G.B., *et al.* A Novel Sensitive Immunoassay Targeting the 5-Methylthio-d-Xylofuranose-Lipoarabinomannan Epitope Meets the WHO's Performance Target for Tuberculosis Diagnosis. *J Clin Microbiol* **56**(2018).
53. Chakraborty, P., Bajeli, S., Kaushal, D., Radotra, B.D. & Kumar, A. Biofilm formation in the lung contributes to virulence and drug tolerance of Mycobacterium tuberculosis. *Nat Commun* **12**, 1606 (2021).
54. Phillips, M., *et al.* Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb)* **92**, 314-320 (2012).
55. Williams, C.M., *et al.* Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. *Lancet Infect Dis* **20**, 607-617 (2020).

56. Chen, D., Bryden, W.A. & Wood, R. Detection of Tuberculosis by The Analysis of Exhaled Breath Particles with High-resolution Mass Spectrometry. *Sci Rep* **10**, 7647 (2020).
57. Bruins, M., *et al.* Diagnosis of active tuberculosis by e-nose analysis of exhaled air. *Tuberculosis* **93**, 232-238 (2013).
58. Maiga, M., Abaza, A. & Bishai, W.R. Current tuberculosis diagnostic tools & role of urease breath test. *Indian J Med Res* **135**, 731-736 (2012).
59. Muccilli, V., *et al.* Protein profile of exhaled breath condensate determined by high resolution mass spectrometry. *J Pharm Biomed Anal* **105**, 134-149 (2015).
60. Laneelle, M.A., Nigou, J. & Daffe, M. Lipid and lipoarabinomannan isolation and characterization. *Methods Mol Biol* **1285**, 77-103 (2015).
61. Nigou, J., *et al.* The phosphatidyl-myo-inositol anchor of the lipoarabinomannans from Mycobacterium bovis bacillus Calmette Guerin. Heterogeneity, structure, and role in the regulation of cytokine secretion. *J Biol Chem* **272**, 23094-23103 (1997).
62. Gerwig, G.J., Kamerling, J.P. & Vliegenthart, J.F.G. Determination of the D and L configuration of neutral monosaccharides by high-resolution capillary G.L.C. *Carbohyd Res* **62**, 349-357 (1978).
63. Gilleron, M., Lindner, B. & Puzo, G. MS/MS approach for characterization of the fatty acid distribution on mycobacterial phosphatidyl-myo-inositol mannosides. *Anal Chem* **78**, 8543-8548 (2006).
64. Layre, E., *et al.* A comparative lipidomics platform for chemotaxonomic analysis of Mycobacterium tuberculosis. *Chem Biol* **18**, 1537-1549 (2011).
65. Bouyssie, D., *et al.* Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **36**, 3148-3155 (2020).
66. R Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2022).
67. Kuhn, M. caret: Classification and Regression Training. R package version 6.0-93. (2022).
68. Wickham, H., *et al.* Welcome to the tidyverse. *J Open Source Softw* **4**, 1686 (2019).

ACKNOWLEDGEMENTS

This study was supported by the COLCIENCIAS grant (to MR): 1115-4592-1439, CODI: 01532, the Programa de Sostenibilidad grant from Universidad de Antioquia (to MR), the Fondation pour la Recherche Médicale (Equipes FRM DEQ20180339208; Aide aux projets innovants: Financement d'un ingénieur ING20160435108; to JN) and MSDAVENIR (grant Fight-TB to JN). SFMR received a doctoral fellowship from COLCIENCIAS. This work was also supported in part for proteomics by the French Ministry of Research with the Investment for the Future program dedicated to National Infrastructures in Biology and Health (PIA, Proteomics French Infrastructure, ProFI, ANR-10-INBS-08 to OBS). We thank the Integrated Screening Platform of Toulouse (PICT) for providing access to NMR spectrometers which were funded by CNRS, Université Paul Sabatier, Infrastructures en Biologie Santé et Agronomie (IBiSA), Région Occitanie and European Structural Funds. We thank the MetaToul Lipidomic Core Facility (I2MC, INSERM 1048, Toulouse, France, MetaboHUB-ANR-11-INSB-0010) for assistance and access to the SFC/MS instrument. We would also like to acknowledge the European Union's Horizon 2020 research and innovation program under grant agreement No 847762.

The authors would like to thank the healthy individuals and patients who participated in this study, and the TB Control Program of the Secretaría de Salud de Medellín and the Secretaría Seccional de Salud y Bienestar Social de Antioquia, Colombia. Special thanks to Maria Cecilia Oquendo-Parra for her dedication in the collection of the exhaled samples and patient records from the different health institutions. We acknowledge Mrs Anne Launay (Université de Toulouse) for her assistance with GC-MS experiments, and Drs Cristina Vilaplana (Institut de Recerca Germans Trias i Pujol, Barcelona) and Jacqueline Achkar (Albert Einstein College of Medicine, New York) for highly helpful discussions and advice. We thank Life Science Editors for editorial assistance.

This article is dedicated to the memory of Prof. Stefan B. Svenson (1943–2013) and Prof. David E. Minnikin (1939-2021). Their pioneering work on the detection of LAM in urine and the biochemistry of Mtb lipids respectively remains a source of inspiration.

The following reagents were obtained through BEI Resources, NIAID, NIH: Monoclonal Anti-*Mycobacterium tuberculosis* LAM, Clone CS-35 (produced in vitro), NR-13811; Monoclonal Anti-*Mycobacterium tuberculosis* GroEL2 (Gene Rv0440), Clone CS-44 (produced in vitro), NR-13813; Monoclonal Anti-*Mycobacterium tuberculosis* HspX (Gene Rv2031c), Clone IT-20 (TB68) (produced in vitro), NR-13607; Monoclonal Anti-*Mycobacterium tuberculosis* KatG (Gene Rv1908c), Clone IT-57 (CDA4) (culture supernatant), NR-13793; Monoclonal Anti-*Mycobacterium tuberculosis* LpqH (Gene Rv3763), IT-54 (produced in vitro), NR-13792; Monoclonal Anti-*Mycobacterium tuberculosis* HBHA (Gene Rv0475), Clone α-HBHA (produced in vitro), NR-13804.

AUTHOR CONTRIBUTIONS STATEMENT

MR and JN conceived the study. SFRM participated to the collection of EBC samples.

SFRM, SZ, LG, EL, MG, AS, DR, OBS, ACC, LFG, CS, CAPJ, MR and JN, designed and performed research, and analyzed data. JN wrote the paper.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

Table 1. Clinical and demographic characteristics of the adult and pediatric TB patients, and control individuals, either healthy or with bacterial pneumonia.

	Adult TB patients		
	Smear-positive	Smear-negative	
		Culture-negative	Culture-positive
Number of patients	8	7	14
Gender M/F	6/2	5/2	8/6
Age Median	34	44	38
Age Minimal	27	19	21
Age Maximal	62	70	68
TST[a] (positive/performed)	7/8	5/7	8/14
TST Diameter (mm)[b]	13 ± 2 (9-16)	11 ± 2 (3-14)	9 ± 5 (3 – 14)
BCG scar positive/total	6/8	6/7	10/14
Primary treatment	Isoniazid Rifampicin	Isoniazid Rifampicin	Isoniazid Rifampicin
Time of EBC collection	Under antibiotic treatment for <2wk	Before antibiotic treatment	Under antibiotic treatment for <2wk
Label	Ad S^+	Ad S^-C^-	Ad S^-C^+

	Pediatric TB patients	
	String test smear- or culture-positive	String test smear- and culture-negative
Number of patients	5	12
Gender M/F	3/2	6/6
Age Median	9	9
Age Minimal	6	6
Age Maximal	12	12
TST[a] (positive/performed)	3/5	11/12
TST Diameter (mm)[b]	11 ± 5 (6-14)	9 ± 3 (4-16)
BCG scar (positive/total)	3/5	9/12
Primary treatment	Isoniazid Rifampicin	Isoniazid Rifampicin
Time of EBC collection	Before antibiotic treatment	Under antibiotic treatment for <2wk, and at months 1 and 3 [c]
Label	Ch S^+/C^+	Ch S^-C^-

	Control individuals			
	Community-acquired pneumonia		Healthy	
	Gram negative	Gram positive	Children	Adult
Number of persons	8	7	15	15
Gender M/F	4/4	4/3	10/5	12/3
Age Median	42	41	9	34
Age Minimal	18	20	7	26
Age Maximal	60	61	12	49
TST[a] (positive/performed)	unknown	unknown	10/15	14/15
TST Diameter (mm)[b]	unknown	unknown	0 ± 0.1 (0-0.3)	16 ± 5 (7-24)
BCG scar (positive/total)	unknown	unknown	10/15	14/15

Treatment	Moxifloxacin or Levofloxacin. Clindamycin	Moxifloxacin or Levofloxacin. Amikacin	N/A	N/A
Time of EBC collection	Before antibiotic treatment	Before antibiotic treatment	N/A	N/A
Label	Ad pneumo		Ch healthy	Ad healthy

N/A, not applicable

[a] tuberculin skin test

[b] mean ± SD (range)

[c] Six (3M/3F; BCG scar positive; 6-12 years old) of the 12 patients were followed during antibiotic treatment and EBC collected after 1 and 3 months of antibiotic therapy

FIGURE LEGENDS

Fig. 1. Quantification of LAM in EBC from TB patients and control individuals listed in Table 1.

Quantity of LAM in EBC samples from all the subjects involved (a), adults (b) and children (c, e) was determined by an immunoassay using the CS-35 anti-LAM antibody. A receiver operating characteristic (ROC) analysis of the LAM quantitation data is shown in (d). AUC, area under the curve; *T*, threshold. In a, b and c, the difference between TB patient groups and controls (healthy, pneumo) was statistically significant (Mann-Whitney *U*-test, two-tailed). Error bars represent SEM. Source data are provided as a Source Data file.

Fig. 2. Characterization of LAM in EBC by NMR.

Expanded region (δ ^{1}H: 4.80-5.50, δ ^{13}C 98-114) of the 2D ^{1}H-^{13}C HSQC spectrum in D_2O at 298 K of Ad S^+ pool (a) and Ch S^+/C^+ pool (b) LAM-enriched fractions, and Mtb_broth LAM (c).

Cartoons show the structure of arabinan side chain termini deducted from NMR data. The branched hexa-arabinofuranoside (Ara_6) motif is the main epitope of the CS-35 anti-LAM antibody. Structural motifs that differ between LAM in EBC and LAM purified from *M. tuberculosis* H37Rv grown in broth are highlighted in blue.

Fig. 3. *M. tuberculosis* lipids and corresponding MS signatures detected in EBC from TB patients.

a) Negative MALDI-TOF mass spectrum of PI and PIMs. A structure of tetra-acylated PIM_2 (Ac_2PIM_2) that contains 2 palmitic (C_{16}), 1 stearic (C_{18}) and 1 tuberculostearic (C_{19}) acids is drawn. * indicate intense ions that do not correspond to PIM molecular species.

b) Negative MALDI-TOF mass spectrum of Ac_4SGL. A structure that contains 2 hydroxyphthioceranyl (HPA) and 1 phthioceranyl (PA) (SL-II according to the nomenclature of Goren) is drawn.

c) Positive ESI-QTOF mass spectrum of PDIM. MCA, mycocerosic acid.

d, e) Negative ESI-QTOF mass spectrum of α- (d) and methoxy-(e) mycolic acids. The main forms are illustrated.

f) Positive ESI-QTOF mass spectrum of TbAd. 1-TbAd isomer is shown.

Data are representative of at least 2 independent experiments on each EBC pooled sample. The precise stereochemistry of PIMs, SGLs, PDIM and Mycolic acids can be found in Minnikin & Brennan, 2020[28]. A detailed peak assignment is shown in Supplementary Tables 4-7.

Fig. 4. Abundance of *M. tuberculosis* lipids in EBC from TB patients and control individuals listed in Table 1.

Abundance of MA (a-c), TbAd (d-f) and PDIM (g-i) per EBC from adults and children was determined by SFC-HRMS relatively to 1,2-ditridecanoyl-*sn*-glycero-3-phosphocholine (133 ng/ml of EBC) used as an internal standard (IS). Values are given as the ratio between areas of the extracted ion chromatograms (AEIC) of the ionized lipid molecular species and AEIC of the IS. In g, h and i, values are multiplied by 10. In a, b, d, e, g, h, unless otherwise stated (ns, not significant), the difference between TB patient groups and controls (healthy, pneumo) was statistically significant (Mann-Whitney *U*-test, two-tailed). Error bars represent SEM. Source data are provided as a Source Data file.

Fig. 5. Abundance of *M. tuberculosis* proteins in EBC from TB patients and control individuals listed in Table 1.

a) Number of Mtb proteins detected by proteomic analysis in the corresponding groups.

b-i) Abundance of selected Mtb proteins by proteomic analysis.

j-l) Abundance of GroEL2 protein by an immunoassay.

In b-i, values are given as the Label-Free Quantification (LFQ) intensity (int.). The noise background intensity was ~3.3 log. In j-l, values are given in arbitrary units corresponding to intensity (int.) on the Dot Blot (DB) and normalized to levels of GroEL2 in the Mtb cell lysate. In j and k, unless otherwise stated (ns, not significant), the difference between TB patient groups and controls (healthy, pneumo) was statistically significant (Mann-Whitney *U*-test, two-tailed). Error bars represent SEM. Source data are provided as a Source Data file.

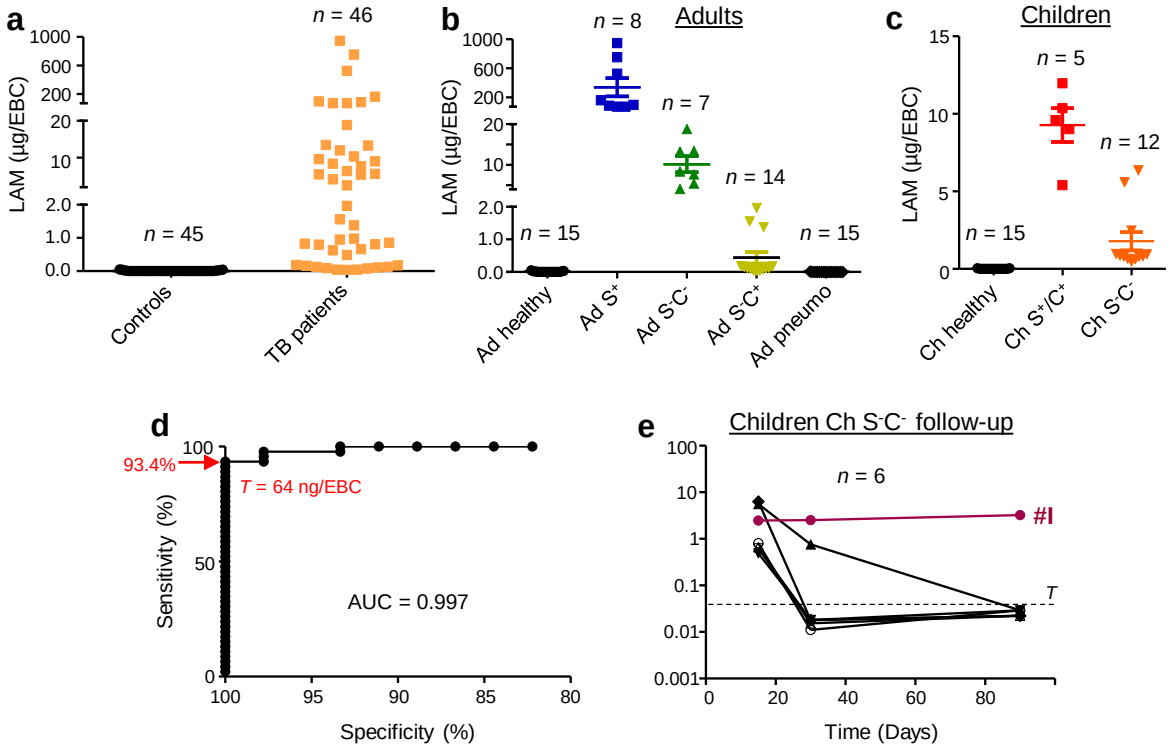

a
b
c
d
e
LAM (µg/EBC)
n = 46
n = 45
Controls
TB patients
n = 8
n = 7
n = 14
n = 15
n = 15
Adults
Ad healthy
Ad S+
Ad S-C-
Ad S-C+
Ad pneumo
n = 5
n = 12
n = 15
Children
Ch healthy
Ch S+/C+
Ch S-C-
93.4%
T = 64 ng/EBC
AUC = 0.997
Sensitivity (%)
Specificity (%)
Children Ch S-C- follow-up
n = 6
#I
T
Time (Days)

Fig. 1

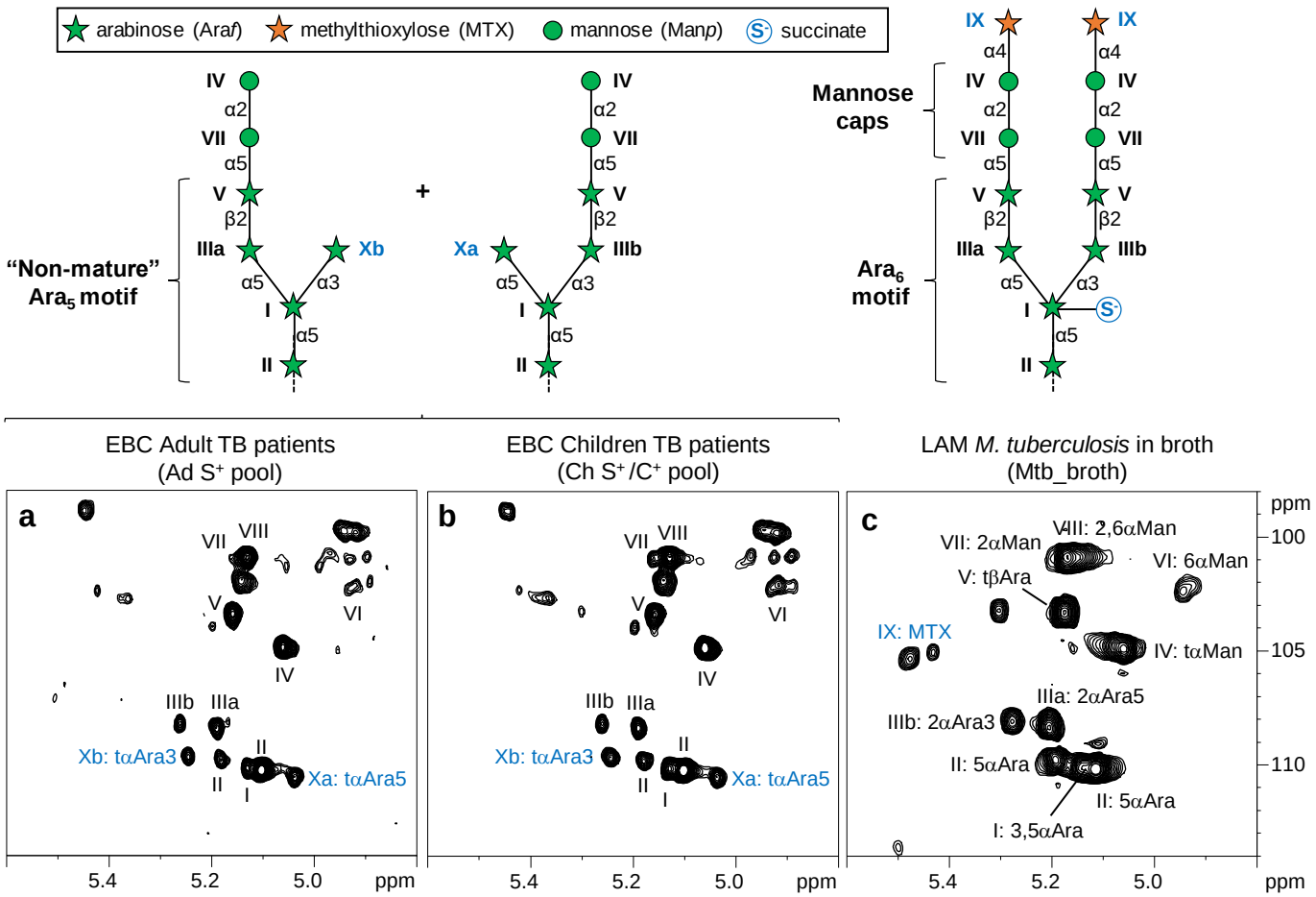

arabinose (Araf)
methylthioxylose (MTX)
mannose (Manp)
succinate
"Non-mature" Ara5 motif
Mannose caps
Ara6 motif
EBC Adult TB patients (Ad S+ pool)
EBC Children TB patients (Ch S+/C+ pool)
LAM M. tuberculosis in broth (Mtb_broth)
Xb: tαAra3
Xa: tαAra5
VII: 2αMan
VIII: 2,6αMan
VI: 6αMan
V: tβAra
IX: MTX
IV: tαMan
IIIa: 2αAra5
IIIb: 2αAra3
II: 5αAra
I: 3,5αAra
ppm

Fig. 2

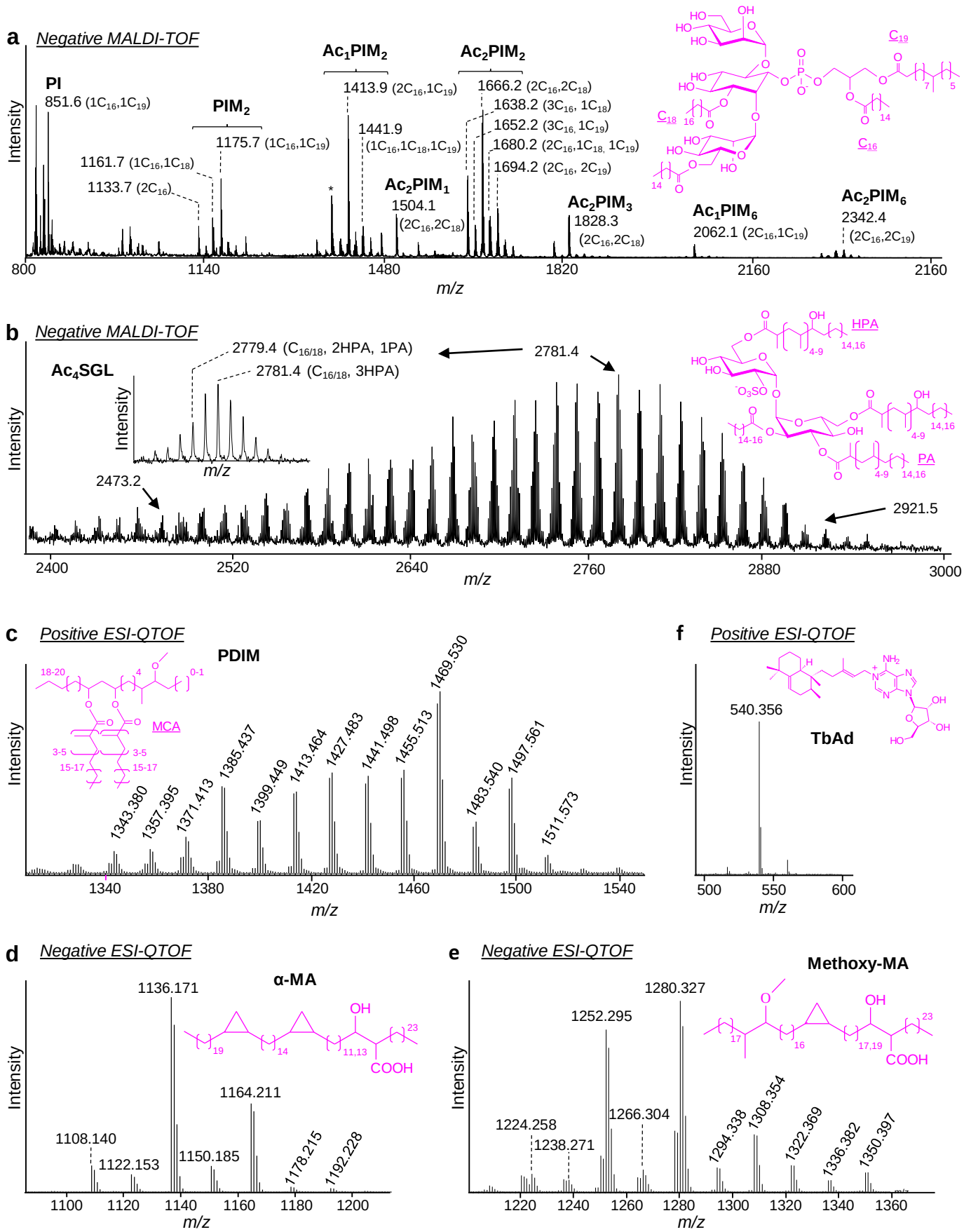

a
Negative MALDI-TOF
PI
851.6 (1C16,1C19)
PIM2
1175.7 (1C16,1C19)
1161.7 (1C16,1C18)
1133.7 (2C16)
Ac1PIM2
1413.9 (2C16,1C19)
1441.9
(1C16,1C18,1C19)
Ac2PIM2
1666.2 (2C16,2C18)
1638.2 (3C16, 1C18)
1652.2 (3C16, 1C19)
1680.2 (2C16,1C18, 1C19)
1694.2 (2C16, 2C19)
Ac2PIM1
1504.1
(2C16,2C18)
Ac2PIM3
1828.3
(2C16,2C18)
Ac1PIM6
2062.1 (2C16,1C19)
Ac2PIM6
2342.4
(2C16,2C19)
Intensity
m/z
b
Negative MALDI-TOF
Ac4SGL
2779.4 (C16/18, 2HPA, 1PA)
2781.4 (C16/18, 3HPA)
2781.4
2473.2
2921.5
HPA
PA
c
Positive ESI-QTOF
PDIM
MCA
1343.380
1357.395
1371.413
1385.437
1399.449
1413.464
1427.483
1441.498
1455.513
1469.530
1483.540
1497.561
1511.573
f
Positive ESI-QTOF
TbAd
540.356
d
Negative ESI-QTOF
α-MA
1136.171
1164.211
1108.140
1122.153
1150.185
1178.215
1192.228
e
Negative ESI-QTOF
Methoxy-MA
1280.327
1252.295
1266.304
1224.258
1238.271
1294.338
1308.354
1322.369
1336.382
1350.397
COOH

Fig. 3

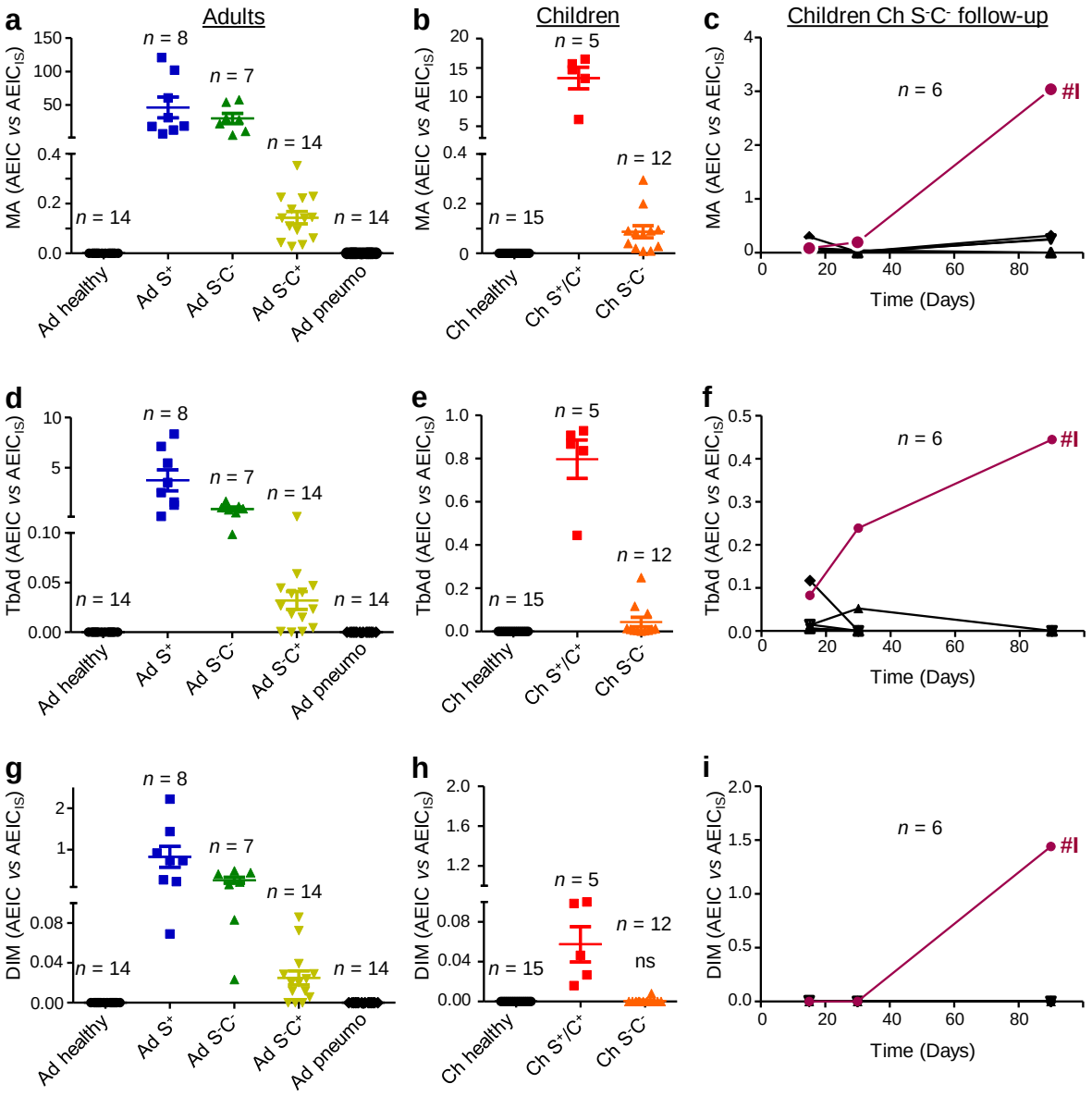

a
Adults
MA (AEIC vs AEIC$_{IS}$)
n = 8
n = 7
n = 14
n = 14
n = 14
Ad healthy
Ad S+
Ad S-C-
Ad S-C+
Ad pneumo
b
Children
n = 5
n = 12
n = 15
Ch healthy
Ch S+/C+
Ch S-C-
c
Children Ch S-C- follow-up
n = 6
#I
Time (Days)
d
TbAd (AEIC vs AEIC$_{IS}$)
e
f
g
DIM (AEIC vs AEIC$_{IS}$)
h
ns
i

Fig. 4

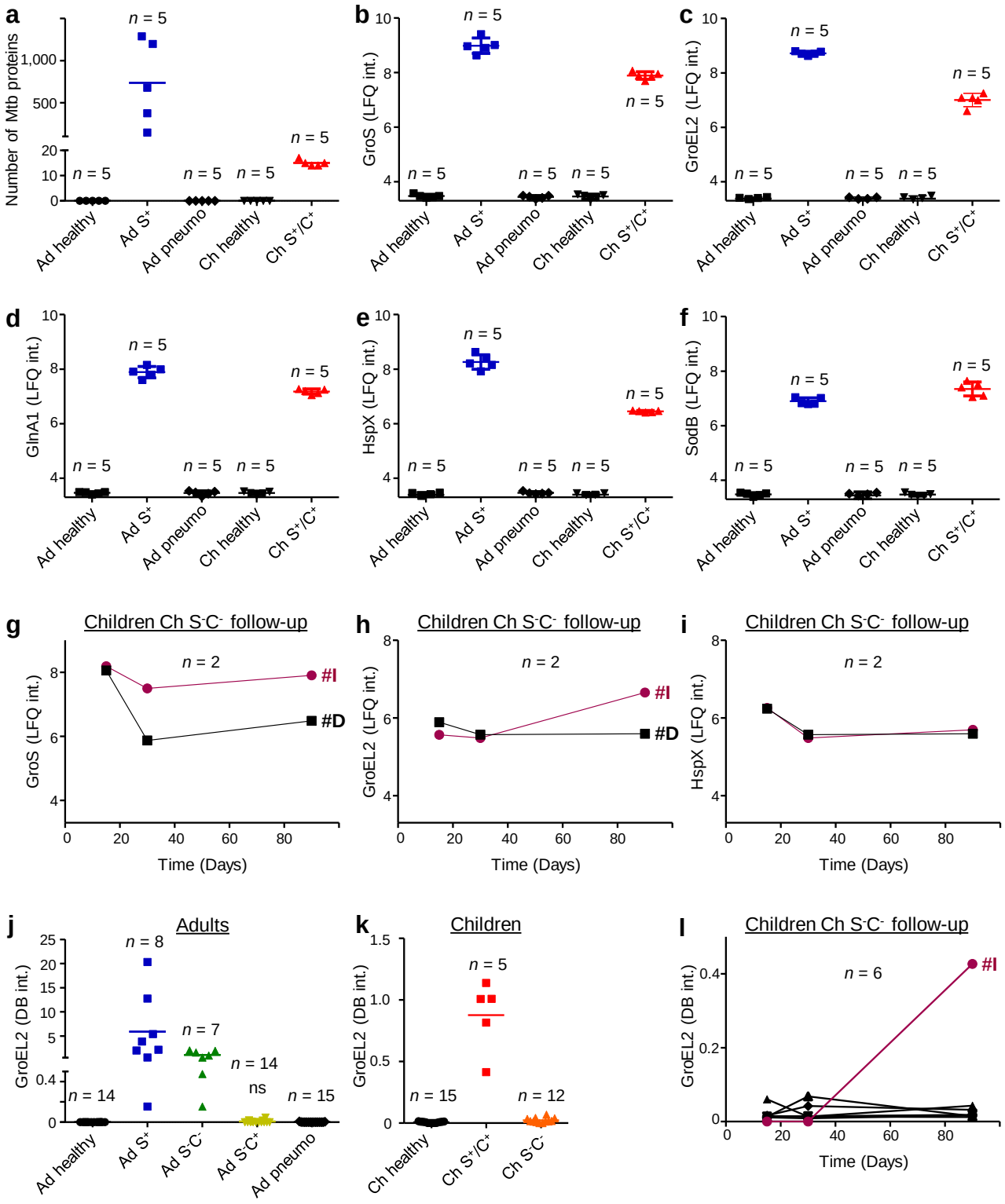

a
Number of Mtb proteins
1,000
500
20
10
0
n = 5
n = 5
n = 5
n = 5
n = 5
Ad healthy
Ad S+
Ad pneumo
Ch healthy
Ch S+/C+
b
GroS (LFQ int.)
10
8
6
4
n = 5
n = 5
n = 5
n = 5
n = 5
c
GroEL2 (LFQ int.)
n = 5
n = 5
n = 5
n = 5
n = 5
d
GlnA1 (LFQ int.)
n = 5
n = 5
n = 5
n = 5
n = 5
e
HspX (LFQ int.)
n = 5
n = 5
n = 5
n = 5
n = 5
f
SodB (LFQ int.)
n = 5
n = 5
n = 5
n = 5
n = 5
g
Children Ch S-C- follow-up
GroS (LFQ int.)
n = 2
#I
#D
0
20
40
60
80
Time (Days)
h
Children Ch S-C- follow-up
GroEL2 (LFQ int.)
n = 2
#I
#D
Time (Days)
i
Children Ch S-C- follow-up
HspX (LFQ int.)
n = 2
Time (Days)
j
Adults
GroEL2 (DB int.)
25
20
15
10
5
0.4
0
n = 8
n = 7
n = 14
ns
n = 14
n = 15
Ad healthy
Ad S+
Ad S-C-
Ad S-C+
Ad pneumo
k
Children
GroEL2 (DB int.)
1.5
1.0
0.5
0
n = 5
n = 15
n = 12
Ch healthy
Ch S+/C+
Ch S-C-
l
Children Ch S-C- follow-up
GroEL2 (DB int.)
0.4
0.2
0
n = 6
#I
0
20
40
60
80
Time (Days)

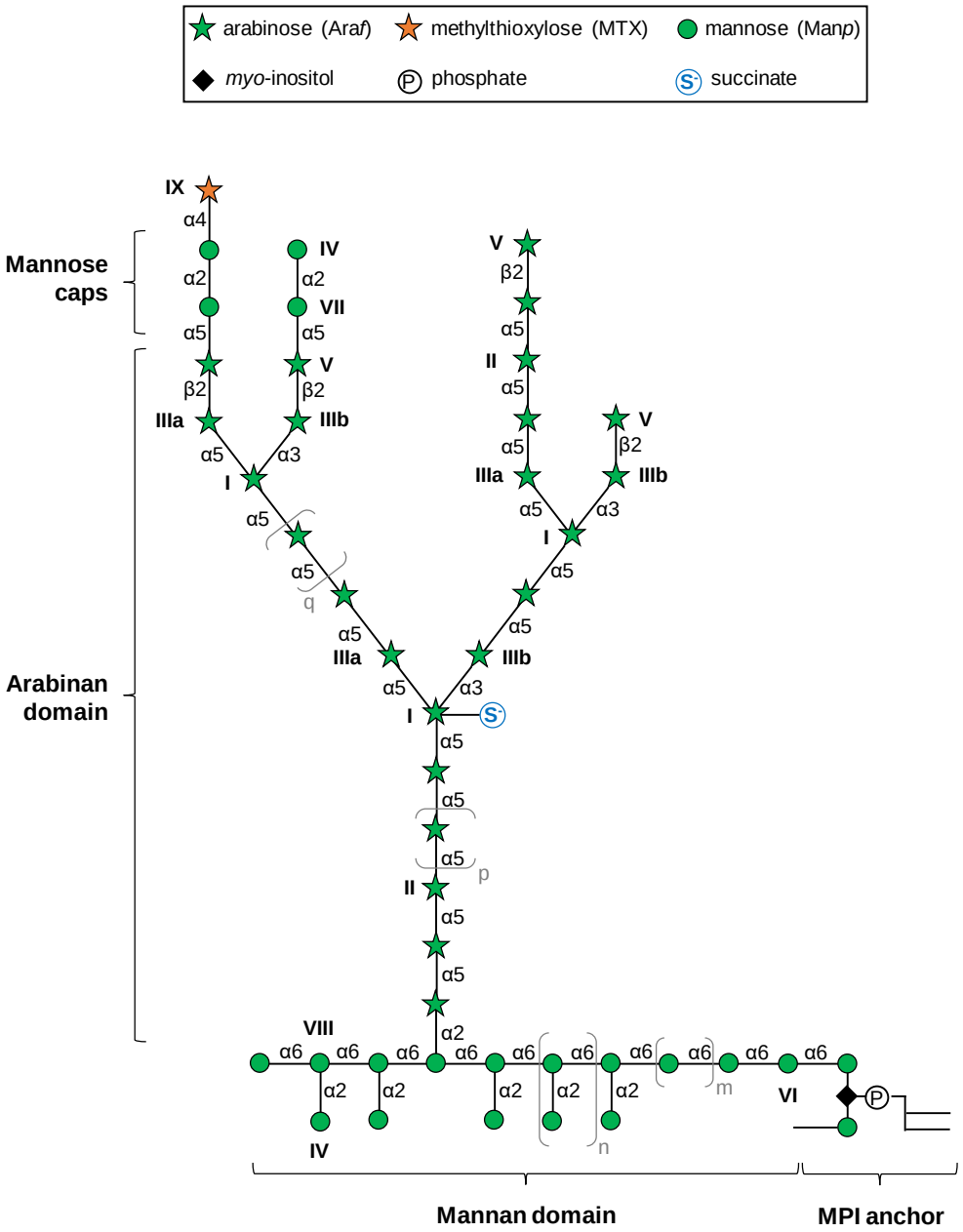
Fig. 5

Supplementary information for

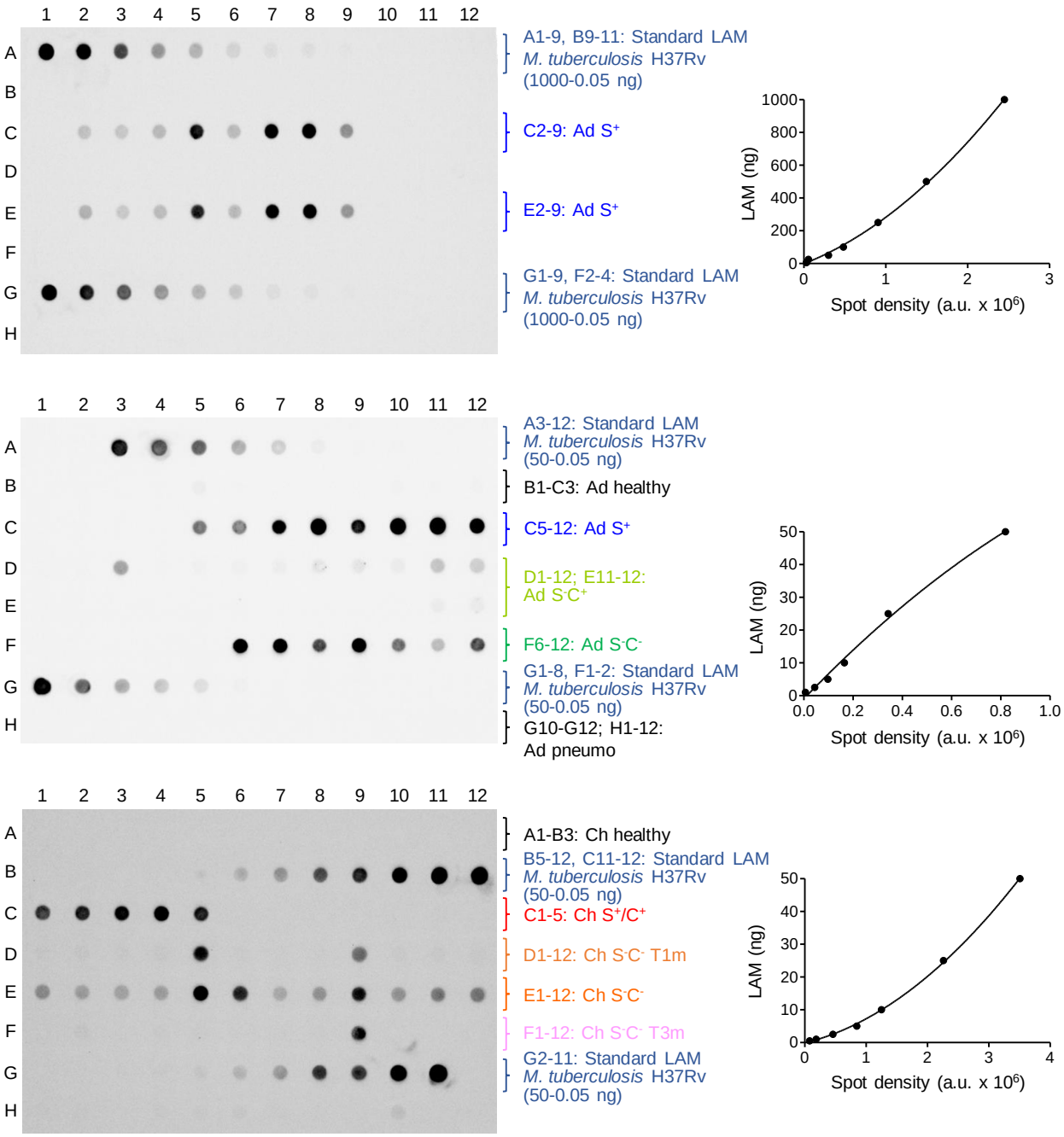
A *Mycobacterium tuberculosis* fingerprint in human breath allows tuberculosis detection

Sergio Fabián Mosquera-Restrepo, Sophie Zuberogoïtia, Lucie Gouxette, Emilie Layre, Martine Gilleron, Alexandre Stella, David Rengel, Odile Burlet-Schiltz, Ana Cecilia Caro, Luis F. Garcia, César Segura, Carlos Alberto Peláez Jaramillo, Mauricio Rojas and Jérôme Nigou

Supplementary Figures 1 to 8
Supplementary Tables 1 to 8
Supplementary references
Legend of Supplementary Data 1 & 2
Uncropped scans of Supplementary Fig. 2
Uncropped scan of Supplementary Fig. 7

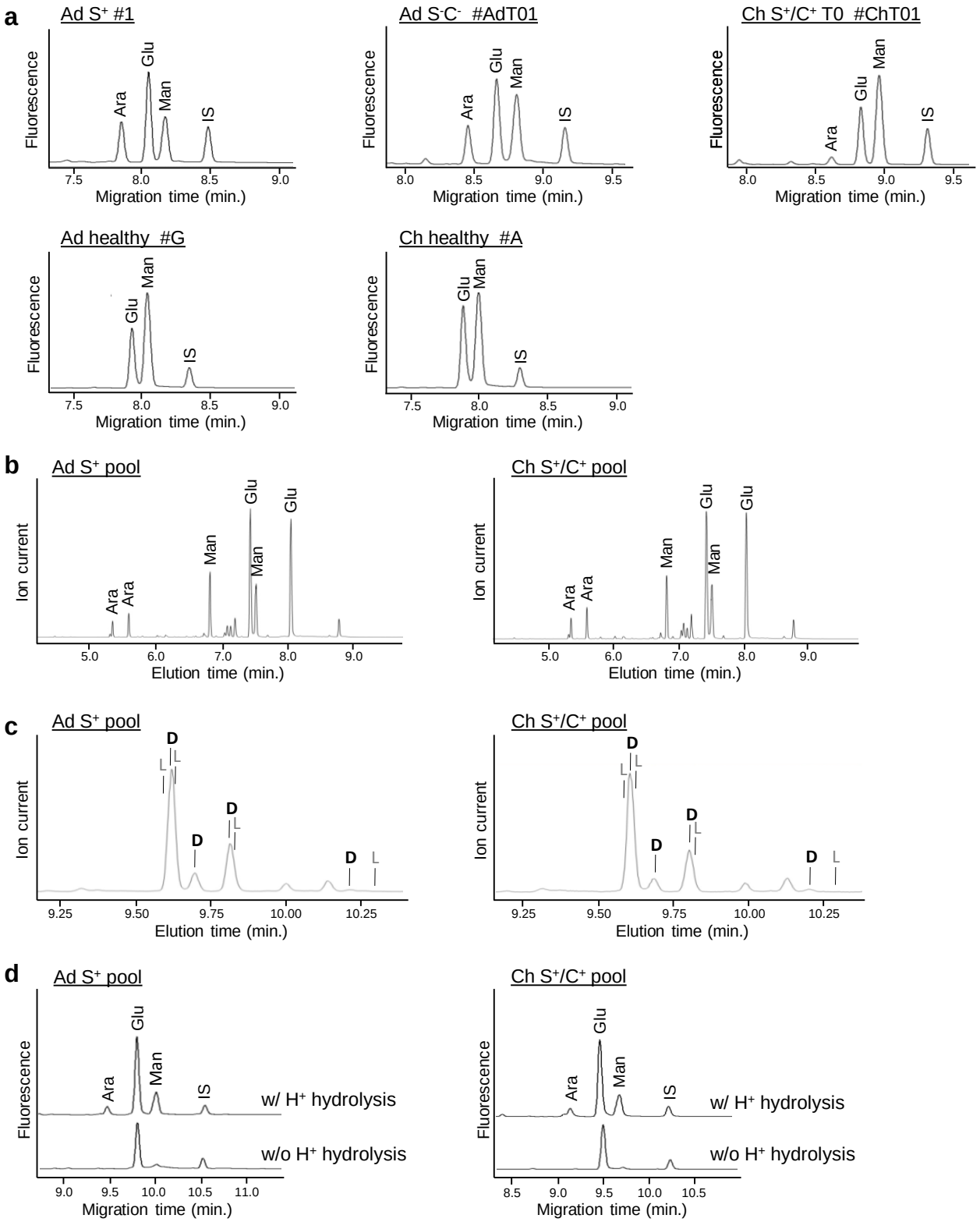


Supplementary Fig. 1. Schematic structure of LAM.
Ara*f*, arabinofuranose; Man*p*, mannopyranose; MPI, mannosyl-phosphatidyl-*myo*-inositol.



Supplementary Fig. 2. Immunoassay for LAM quantification using CS-35 anti-LAM antibody.

Examples of dot-blot images and calibration curves are shown. Spot density was dermined using Image Lab™ Software (Bio-Rad). T1m, T3m, after 1 and 3 months of antibiotic treatment respectively.

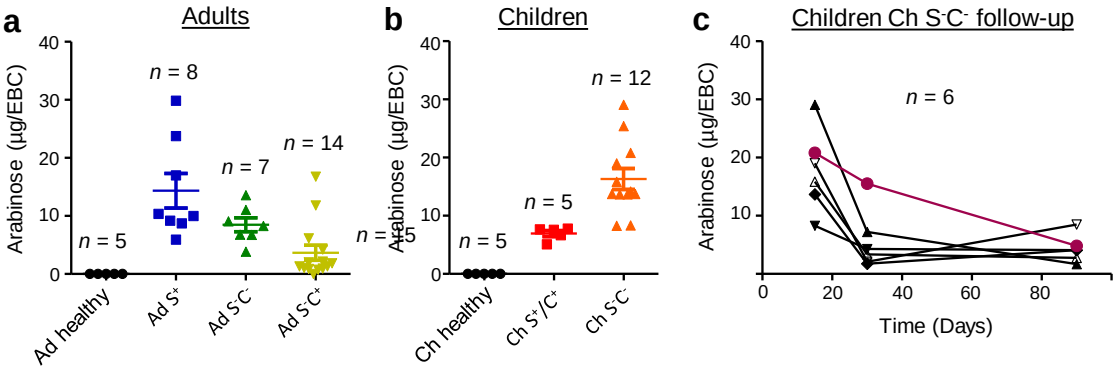

a
Ad S+ #1
Ad S-C- #AdT01
Ch S+/C+ T0 #ChT01
Ad healthy #G
Ch healthy #A
Fluorescence
Ara
Glu
Man
IS
Migration time (min.)
b
Ad S+ pool
Ch S+/C+ pool
Ion current
Elution time (min.)
c
Ad S+ pool
Ch S+/C+ pool
D
L
d
Ad S+ pool
Ch S+/C+ pool
w/ H+ hydrolysis
w/o H+ hydrolysis

Supplementary Fig. 3. Chemical detection of arabinose and other monosaccharides in EBCs.

a) Analysis of monosaccharides by CE-LIF upon total acid hydrolysis and fluorescent labeling by 8-Aminopyrene-1,3,6- trisulfonate (APTS). Representative electropherograms obtained for selected patients/individuals in different groups are shown.
b) Analysis of monosaccharides by GC/MS upon total acid hydrolysis and trimethylsilylation.
c) Analysis of arabinose configuration by GC/MS upon butanolysis and trimethylsilylation.
d) Analysis of free monosaccharides without (w/o) acid hydrolysis by CE-LIF upon fluorescent labeling by APTS.

In b, c and d, analyses were performed on pooled EBCs collected from 50 adults (Ad S^+ pool) or 50 children (Ch S^+/C^+ pool) TB patients (Extended Data Table 1).

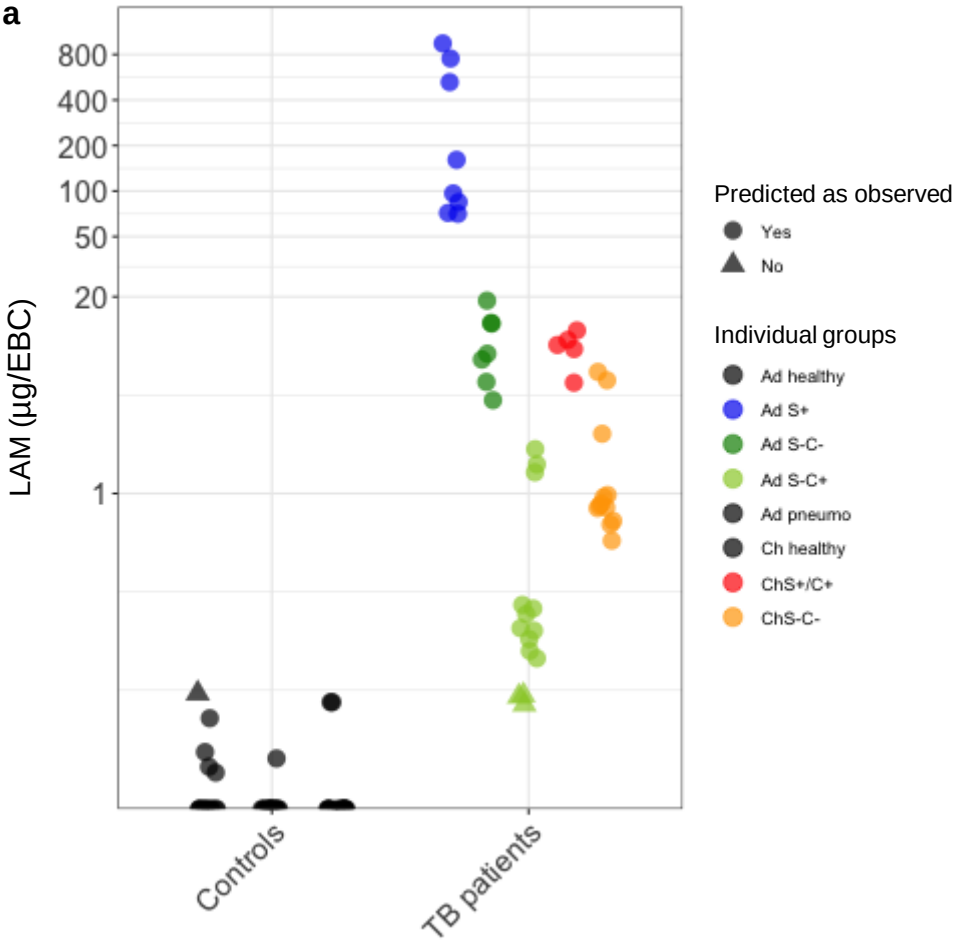
Ara, Glu, Man, and IS correspond to arabinose, glucose, mannose and internal standard (mannoheptose) derivatives respectively; D, L indicate the elution time of the trimethylsilylated R-(-)-2-butyl glycosides of D- and L-arabinose respectively.



Supplementary Fig. 4. Quantification of arabinose as a proxy of LAM in EBC from TB patients and control individuals listed in Table 1.

Quantity of arabinose per EBC from adults (a) and children (b, c) was determined by acid hydrolysis and CE-LIF analysis. In a and b, the difference between TB patient groups and controls (healthy, pneumo) was statistically significant (Mann-Whitney *U*-test, two-tailed). Error bars represent SEM. Source data are provided as a Source Data file.

a



b

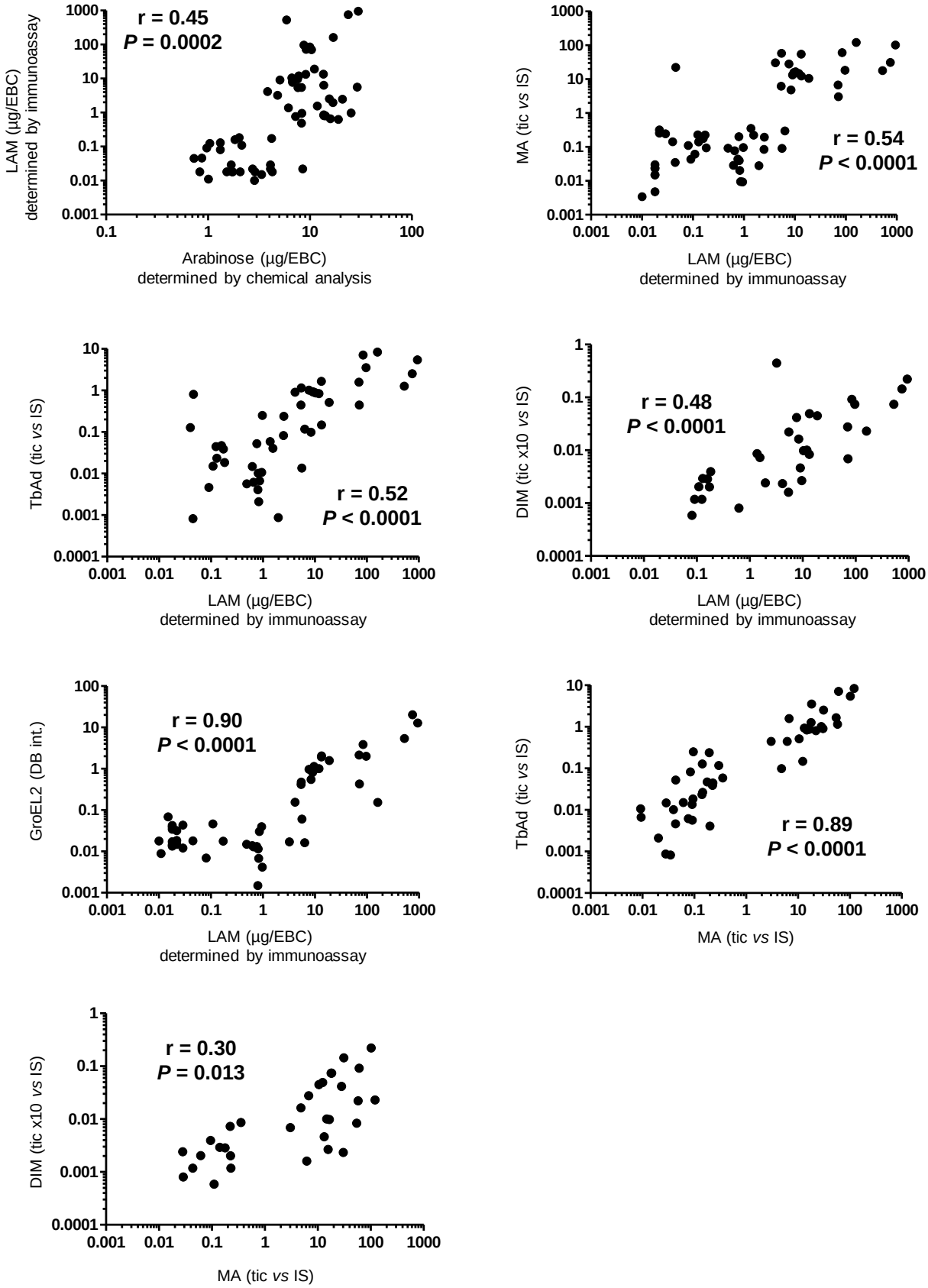
		Actual condition	
		Controls	TB patients
Prediction outcome	Controls	44	3
	TB patients	1	43

Supplementary Fig. 5. Graphical visualization of the Leave-One-Out cross-validation prediction (a) and tabular comparison between actual individual condition and Leave-One-Out cross-validation-predicted condition (b).

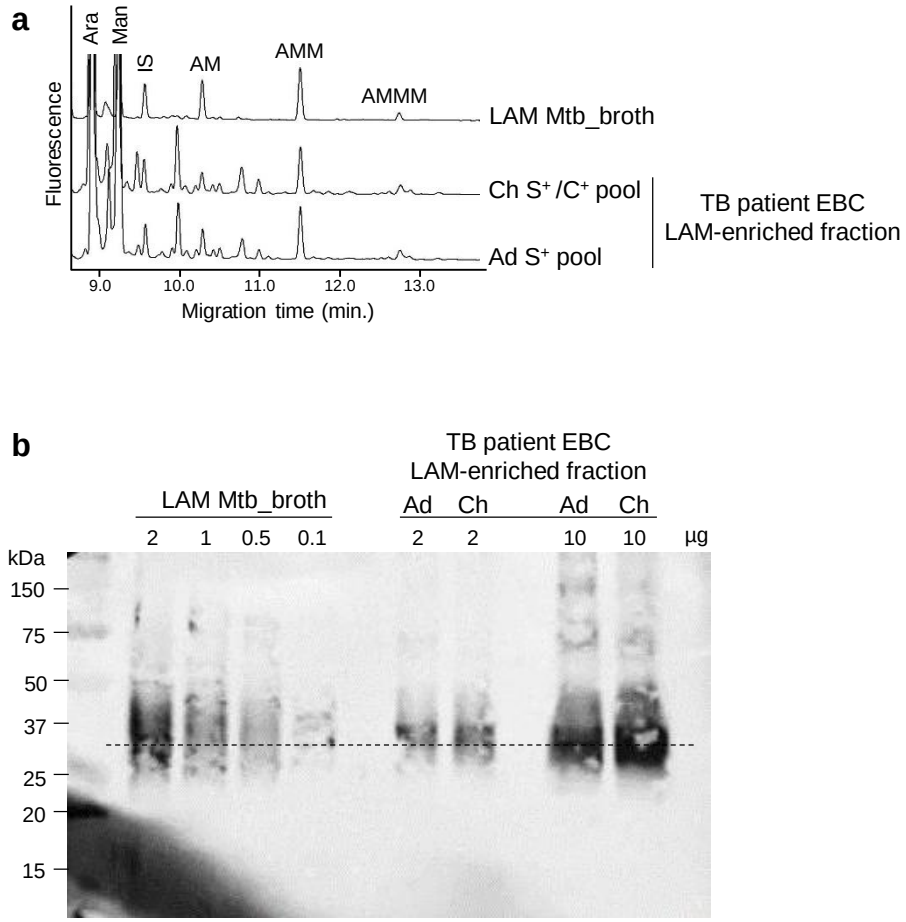
a) Quantity of LAM in EBC samples for control individuals and TB patients are shown. Symbol colors correspond to individual groups. Circles correspond to individuals whose actual condition was correctly predicted by the Leave-One-Out cross-validation (LOOCV), whereas triangles correspond to individuals whose actual condition was not correctly predicted by the LOOCV.

b) Columns refer to actual individual condition and rows to the LOOCV prediction. Numbers correspond to the frequencies of cross-grouping between actual and LOOCV-predicted conditions.

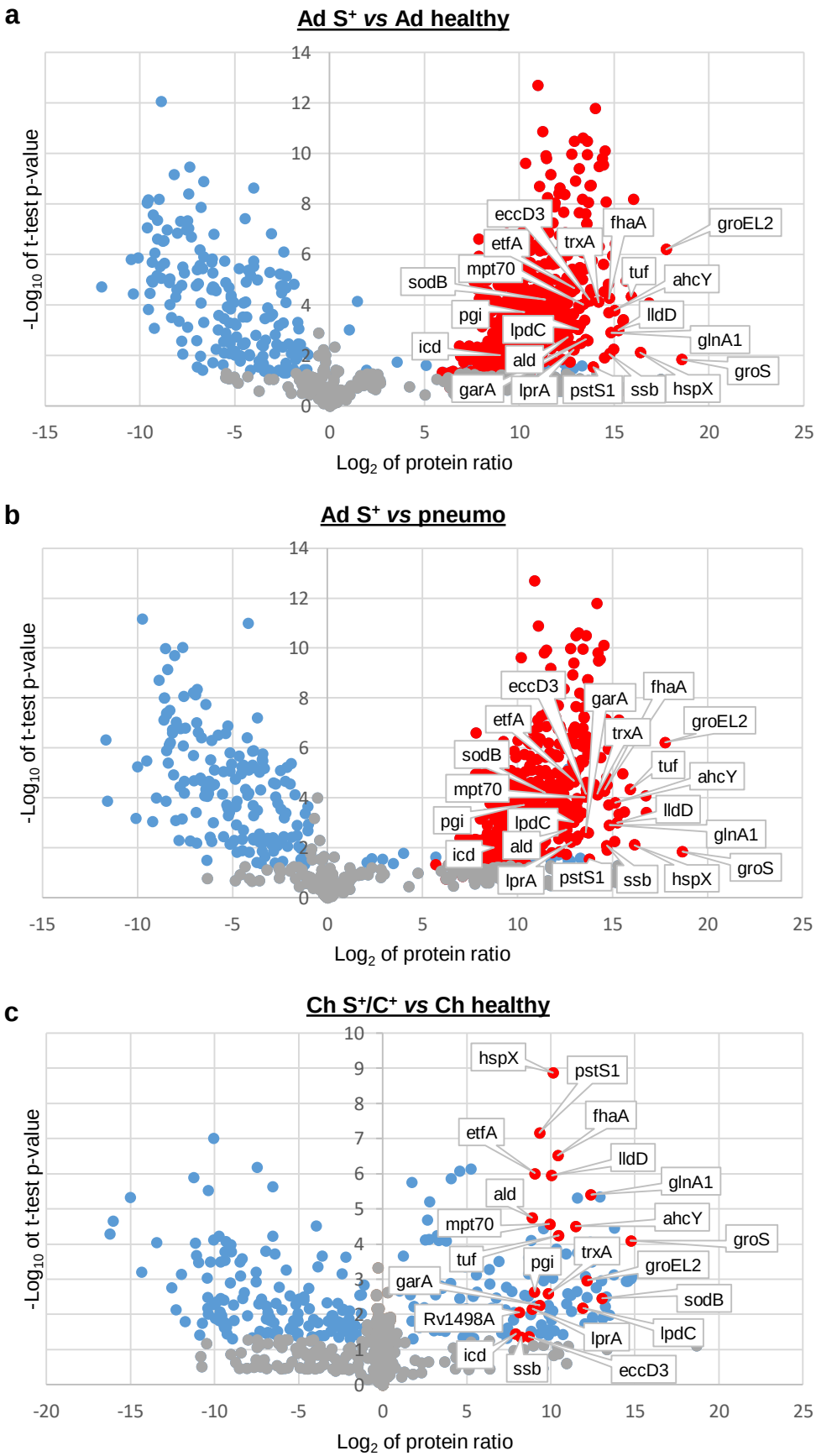
Ad, adult; Ch, child.



Supplementary Fig. 6. Correlation between the measured abundance of the different Mtb molecules in individual EBCs. Pearson's r and *P* values (two-tailed) are provided. A total of 64 EBCs was collected for TB patients (46 at baseline, 12 and 6 after 1 and 3 months of antibiotic treatment respectively for S^-C^- pediatric patients).



Supplementary Fig. 7. LAM in EBC bear mannose caps (a) and exhibit a lower apparent MW (b).
a) Analysis of mannose caps by CE-LIF upon mild acid hydrolysis and fluorescent labeling by 8-Aminopyrene-1,3,6-trisulfonate. IS, internal standard, mannoheptose-APTS; AM, Man*p*-(α1→2)-Ara-APTS; AMM, Man*p*-(α1→2)-Man*p*-(α1→2)-Ara-APTS; AMMM, Man*p*-(α1→2)-Man*p*-(α1→2)-Man*p*-(α1→2)-Ara-APTS.
b) Western blot probed with CS-35 anti-LAM antibody. The different fractions were loaded according to the quantity of arabinose determined by CE-LIF quantification (Extended Data Table 2), 0.1-2 µg for Mtb_broth LAM and 2-10 µg for LAM-enriched fraction from Ad S^+ pool and Ch S^+/C^+ pool.
Ad, adult; Ch, child.
Data are representative of 2 independent experiments.



Supplementary Fig. 8. Volcano plot of log significance (paired t-tests) versus log paired ratio of changes in protein abundance.

Grey circles represent identified proteins not showing significant differences whereas red and blue circles represent Mtb and human proteins respectively with significantly different abundances between the two indicated groups ($p < 0.05$; ratio = 1.5). A Student t-test (two-tailed t-test, equal variances) was performed on log2 transformed values and followed by Adjusted Benjamini–Hochberg (ABH) correction to analyze differences in protein abundance in all biologic group comparisons.

Supplementary Table 1. Clinical and demographic characteristics of the adult and pediatric TB patients whose EBC was pooled.

	Adults	Children
	Smear-positive	**String test smear- or culture-positive**
Number of patients	50	50
Gender M/F	24/26	30/20
Age Median	38	9
Age Minimal	15	6
Age Maximal	60	12
TST[a] (positive/performed)	44/50	6/50
TST Diameter (mm)[b]	12 ± 8 (0-25)	10 ± 2 (0-16)
BCG scar (positive/total)	40/50	6/50
Primary treatment	Isoniazid[c] Rifampicin[c]	Isoniazid[c] Rifampicin[c]
Time of EBC collection	Before antibiotic treatment	Before antibiotic treatment
Label	**Ad S^+ pool**	**Ch S^+/C^+ pool**

[a] tuberculin skin test
[b] mean ± SD (range)
[c] antibiotic treatment started after EBC collection

Supplementary Table 2. Global composition of EBC pools and LAM-enriched fractions

	Mass (mg)	Ad S^+ pool	Ch S^+/C^+ pool
EBC	Total[a]	630	720
	Proteins[b]	172 (27%)	189 (26%)
	Lipids[c]	9.4 (1.5%)	14.3 (2%)
	Arabinose[d]	0.67 (0.1%)	0.76 (0.1%)
	Glucose[d]	6.7[e]	6.6[e]
	Mannose[d]	2.9[f]	3.1[f]
LAM-enriched fraction	Total	11.3	27
	Arabinose[d]	0.18	0.36
	Glucose[d]	0.09	0.24
	Mannose[d]	0.83	1.8

[a], determined by weighing
[b], determined by BCA protein assay
[c], determined after extraction of lipids by organic solvents and weighing
[d], determined after total acid hydrolysis, or not, and quantification by CE-LIF
[e], 60% of Glu was found as a free monosaccharide (detected without total acid hydrolysis)
[f], 10% of Man was found as a free monosaccharide (detected without total acid hydrolysis)

Supplementary Table 3. Anomeric ^{1}H and ^{13}C NMR chemical shifts of LAM in EBC pools measured at 298 K in D_2O.
Differences between LAM in EBC and LAM purified from *M. tuberculosis* H37Rv grown in both (Mtb_broth) are highlighted in blue. Ara*f*, arabinofuranose; Man*p*, mannopyranose; MTX, methylthioxylose.

Residue		Ad S^+ pool	Ch S^+/C^+ pool	LAM Mtb_broth
I, 3,5-α-Ara*f*	^{13}C	110.2	110.2	110.1
	^{1}H	5.13	5.13	5.14
II, 5-α-Ara*f*	^{13}C	109.9	109.9	109.8
	^{1}H	5.18	5.18	5.19
	^{13}C	110.3	110.3	110.2
	^{1}H	5.10	5.10	5.11
IIIa, 2-α-Ara*f*→5	^{13}C	108.4	108.4	108.4
	^{1}H	5.19	5.19	5.21
IIIb, 2-α-Ara*f*→3	^{13}C	108.2	108.2	108.1
	^{1}H	5.26	5.26	5.28
IV, t-α-Man*p*	^{13}C	104.9	104.9	104.9
	^{1}H	5.06	5.06	5.06
V, t-β-Ara*f*	^{13}C	103.4	103.5	103.3/103.2
	^{1}H	5.16	5.16	5.18/5.30
VI, 6-α-Man*p*	^{13}C	102.3	102.2	102.3
	^{1}H	4.92	4.92	4.94
VII, 2-α-Man*p*	^{13}C	101.0	101.0	100.9
	^{1}H	5.15	5.16	5.19
VIII, 2,6-α-Man*p*	^{13}C	100.9	100.9	100.9
	^{1}H	5.13	5.13	5.17
IX, MTX	^{13}C	ND	ND	105.3/105.0
	^{1}H	ND	ND	5.48/5.43
Xa, t-α-Araf→5	^{13}C	110.6	110.6	ND
	^{1}H	5.04	5.04	ND
Xb, t-α-Araf→3	^{13}C	109.7	109.7	ND
	^{1}H	5.24	5.24	ND

ND, not detected

Supplementary Table 4. List of the $[M-H]^-$ ions of the PIM and corresponding major molecular species detected by MALDI-TOF mass spectrometry in Ch S^+/C^+ pool.

Peak assignment was performed according to previous studies[1-3]. The last column indicates the number and the fatty acyl chains (C_{16}, palmitoyl; C_{18}, stearoyl; C_{19}, tuberculostearoyl, i.e. 10-methyl-stearoyl) esterifying the PIM molecules (PIM_1, PIM_2, PIM_3, or PIM_6: mono-, di-, tri-, or hexa-mannosides: respectively).

In red are indicated the molecular species found in trace amount only in Mtb_broth and that show increased abundance in EBCs.

Measured *m/z* of $[M-H]^-$	Calculated *m/z* of $[M-H]^-$	Molecular formula of M	Relative Intensity (%)	Species	Fatty acyl chains
851.6	851,5655	$C_{44}H_{85}O_{13}P$	n.d.[a]	PI	$1C_{16}$, $1C_{19}$
1133.6	1133,6242	$C_{53}H_{99}O_{23}P$	3.9	PIM_2	$2C_{16}$
1161.7	1161,6555	$C_{55}H_{103}O_{23}P$	3.4	PIM_2	$1C_{16}$, $1C_{18}$
1175.7	1175,6711	$C_{56}H_{105}O_{23}P$	8.3	PIM_2	$1C_{16}$, $1C_{19}$
1189.7	1189,6868	$C_{57}H_{107}O_{23}P$	1.8	PIM_2	$2C_{18}$
1203.7	1203,7024	$C_{58}H_{109}O_{23}P$	1.8	PIM_2	$1C_{18}$, $1C_{19}$
1399.9	1399,8852	$C_{71}H_{133}O_{24}P$	2.2	Ac_1PIM_2	$2C_{16}$, $1C_{18}$
1413.9	1413,9008	$C_{72}H_{135}O_{24}P$	15.2	Ac_1PIM_2	$2C_{16}$, $1C_{19}$
1427.9	1427,9165	$C_{73}H_{137}O_{24}P$	3.0	Ac_1PIM_2	$1C_{16}$, $2C_{18}$
1441.9	1441,9321	$C_{74}H_{139}O_{24}P$	3.5	Ac_1PIM_2	$1C_{16}$, $1C_{18}$, $1C_{19}$
1456.0	1455,9478	$C_{75}H_{141}O_{24}P$	2.2	Ac_1PIM_2	$1C_{16}$, $2C_{19}$
1476.1	1476,0620	$C_{81}H_{153}O_{20}P$	3.0	Ac_2PIM_1	$3C_{16}$, $1C_{18}$
1504.1	1504,0933	$C_{83}H_{157}O_{20}P$	5.1	Ac_2PIM_1	$2C_{16}$, $2C_{18}$
1546.0	1546,1403	$C_{86}H_{163}O_{20}P$	1.6	Ac_2PIM_1	$1C_{16}$, $3C_{18}$
1638.2	1638,1148	$C_{87}H_{163}O_{25}P$	8.1	Ac_2PIM_2	$3C_{16}$, $1C_{18}$
1652.2	1652,1305	$C_{88}H_{165}O_{25}P$	3.3	Ac_2PIM_2	$3C_{16}$, $1C_{19}$
1666.2	1666,1461	$C_{89}H_{167}O_{25}P$	16.2	Ac_2PIM_2	$2C_{16}$, $2C_{18}$
1680.2	1680,1618	$C_{90}H_{169}O_{25}P$	4.3	Ac_2PIM_2	$2C_{16}$, $1C_{18}$, $1C_{19}$
1694.2	1694,1774	$C_{91}H_{171}O_{25}P$	5.1	Ac_2PIM_2	$2C_{16}$, $2C_{19}$
1800.2	1800,1677	$C_{93}H_{173}O_{30}P$	1.7	Ac_2PIM_3	$4C_{16}$
1828,3	1828,1990	$C_{95}H_{177}O_{30}P$	4.1	Ac_2PIM_3	$2C_{16}$, $2C_{18}$
2062.1	2062,1121	$C_{96}H_{175}O_{44}P$	1.2	Ac_1PIM_6	$2C_{16}$, $1C_{19}$
2342.4	2342,3887	$C_{115}H_{211}O_{45}P$	0.9	Ac_2PIM_6	$2C_{16}$, $2C_{19}$

[a] the intensity of PI was not taken into account

Supplementary Table 5. List of the [M-H]$^-$ ions of the Ac$_4$SGL and corresponding molecular species detected by MALDI-TOF mass spectrometry in Ch S$^+$/C$^+$ pool.

Peak assignment was performed according to previous studies[4,5]. SL-I and SL-II refer to the nomenclature introduced by Goren *et al.*[6]. SL-I are acylated by 1 palmitic or 1 stearic acid ($C_{16/18}$), 2 hydroxyphthioceranoïc acid (HPA) and 1 phthioceranoïc acid (PA); SL-II are acylated by 1 palmitic or 1 stearic acid ($C_{16/18}$), and 3 HPA (See Fig. 3). The last column indicates the cumulated number of carbon atoms in the 3 HPA or PA chains; the first and second number correspond to molecules acylated by the C_{16} and the C_{18}, respectively.

Ac$_4$SGL from Mtb_broth show a massif of ions centered at *m/z* from 2459 to 2543[4,5] (see below in blue), whereas Ac$_4$SGL in EBCs show a massif of ions centered at *m/z* from 2739 to 2823 (see below in red; Fig. 3). n.d., not determined.

Measured *m/z* of [M-H]$^-$	Calculated *m/z* of [M-H]$^-$	Molecular formula of M	Relative Intensity (%)	Species SL-I : 2HPA, 1PA SL-II : 3 HPA	Total chain length of the 3 HPA/PA (carbon atom number)
2471,2	2471,1636	C_{152} H_{293} O_{20} S	n.d.	SL-I	124 / 122
2473,2	2473,1429	C_{151} H_{291} O_{21} S	1,0	SL-II	123 / 121
2485,2	2485,1793	C_{153} H_{295} O_{20} S	0,7	SL-I	125 / 123
2487,2	2487,1586	C_{152} H_{293} O_{21} S	0,8	SL-II	124 / 122
2499,2	2499,1949	C_{154} H_{297} O_{20} S	1,0	SL-I	126 / 124
2501,2	2501,1742	C_{153} H_{295} O_{21} S	1,0	SL-II	125 / 123
2513,2	2513,2106	C_{155} H_{299} O_{20} S	0,8	SL-I	127 / 125
2515,2	2515,1899	C_{154} H_{297} O_{21} S	1,0	SL-II	126 / 124
2527,2	2527,2262	C_{156} H_{301} O_{20} S	0,9	SL-I	128 / 126
2529,2	2529,2055	C_{155} H_{299} O_{21} S	0,9	SL-II	127 / 125
2541,2	2541,2419	C_{157} H_{303} O_{20} S	1,2	SL-I	129 / 127
2543,2	2543,2212	C_{156} H_{301} O_{21} S	0,8	SL-II	128 / 126
2555,2	2555,2575	C_{158} H_{305} O_{20} S	0,9	SL-I	130 / 128
2557,2	2557,2368	C_{157} H_{303} O_{21} S	1,1	SL-II	129 / 127
2569,2	2569,2732	C_{159} H_{307} O_{20} S	1,3	SL-I	131 / 129
2571,2	2571,2525	C_{158} H_{305} O_{21} S	1,7	SL-II	130 / 128
2583,2	2583,2888	C_{160} H_{309} O_{20} S	0,9	SL-I	132 / 130
2585,2	2585,2681	C_{159} H_{307} O_{21} S	1,8	SL-II	131 / 129
2597,2	2597,3045	C_{161} H_{311} O_{20} S	0,7	SL-I	133 / 131
2599,2	2599,2838	C_{160} H_{309} O_{21} S	2,1	SL-II	132 / 130
2611,3	2611,3201	C_{162} H_{313} O_{20} S	0,9	SL-I	134 / 132
2613,3	2613,2994	C_{161} H_{311} O_{21} S	1,5	SL-II	133 / 131
2625,3	2625,3358	C_{163} H_{315} O_{20} S	1,3	SL-I	135 / 133
2627,3	2627,3151	C_{162} H_{313} O_{21} S	1,5	SL-II	134 / 132
2639,3	2639,3514	C_{164} H_{317} O_{20} S	1,2	SL-I	136 / 134
2641,3	2641,3307	C_{163} H_{315} O_{21} S	1,9	SL-II	135 / 133
2653,3	2653,3671	C_{165} H_{319} O_{20} S	1,4	SL-I	137 / 135
2655,3	2655,3464	C_{164} H_{317} O_{21} S	2,0	SL-II	136 / 134
2667,3	2667,3827	C_{166} H_{321} O_{20} S	1,2	SL-I	138 / 136
2669,3	2669,3620	C_{165} H_{319} O_{21} S	2,2	SL-II	137 / 135
2681,3	2681,3984	C_{167} H_{323} O_{20} S	1,6	SL-I	139 / 137
2683,3	2683,3777	C_{166} H_{321} O_{21} S	1,8	SL-II	138 / 136
2695,3	2695,4140	C_{168} H_{325} O_{20} S	1,5	SL-I	140 / 138
2697,3	2697,3933	C_{167} H_{323} O_{21} S	2,7	SL-II	139 / 137
2709,3	2709,4297	C_{169} H_{327} O_{20} S	2,0	SL-I	141 / 139
2711,3	2711,4090	C_{168} H_{325} O_{21} S	2,9	SL-II	140 / 138
2723,3	2723,4453	C_{170} H_{329} O_{20} S	1,5	SL-I	142 / 140

2725,3	2725,4246	C_{169} H_{327} O_{21} S	2,6	SL-II	141 / 139
2737,3	2737,4610	C_{171} H_{331} O_{20} S	1,7	SL-I	143 / 141
2739,3	2739,4403	C_{170} H_{329} O_{21} S	2,9	SL-II	142 / 140
2751,3	2751,4766	C_{172} H_{333} O_{20} S	2,0	SL-I	144 / 142
2753,3	2753,4559	C_{171} H_{331} O_{21} S	3,3	SL-II	143 / 141
2765,3	2765,4923	C_{173} H_{335} O_{20} S	2,3	SL-I	145 / 143
2767,3	2767,4716	C_{172} H_{333} O_{21} S	2,8	SL-II	144 / 142
2779,4	2779,5079	C_{174} H_{337} O_{20} S	1,8	SL-I	146 / 144
2781,4	2781,4872	C_{173} H_{335} O_{21} S	3,0	SL-II	145 / 143
2793,4	2793,5236	C_{175} H_{339} O_{20} S	1,3	SL-I	147 / 145
2795,4	2795,5029	C_{174} H_{337} O_{21} S	2,9	SL-II	146 / 144
2807,4	2807,5392	C_{176} H_{341} O_{20} S	1,6	SL-I	148 / 146
2809,4	2809,5185	C_{175} H_{339} O_{21} S	3,3	SL-II	147 / 145
2821,4	2821,5549	C_{177} H_{343} O_{20} S	1,3	SL-I	149 / 147
2823,4	2823,5342	C_{176} H_{341} O_{21} S	3,0	SL-II	148 / 146
2835,4	2835,5705	C_{178} H_{345} O_{20} S	1,6	SL-I	150 / 148
2837,4	2837,5498	C_{177} H_{343} O_{21} S	2,6	SL-II	149 / 147
2849,4	2849,5862	C_{179} H_{347} O_{20} S	1,2	SL-I	151 / 149
2851,4	2851,5654	C_{178} H_{345} O_{21} S	1,7	SL-II	150 / 148
2863,4	2863,6018	C_{180} H_{349} O_{20} S	1,0	SL-I	152 / 150
2865,4	2865,5811	C_{179} H_{347} O_{21} S	1,6	SL-II	151 / 149
2877,4	2877,6175	C_{181} H_{351} O_{20} S	0,8	SL-I	153 / 151
2879,4	2879,5968	C_{180} H_{349} O_{21} S	1,0	SL-II	152 / 150
2891,4	2891,6331	C_{182} H_{353} O_{20} S	0,6	SL-I	154 / 152
2893,4	2893,6124	C_{181} H_{351} O_{21} S	1,2	SL-II	153 / 151
2905,5	2905,6488	C_{183} H_{355} O_{20} S	0,4	SL-I	155 / 153
2907,4	2907,6281	C_{182} H_{353} O_{21} S	0,6	SL-II	154 / 152
2919,4	2919,6644	C_{184} H_{357} O_{20} S	n.d.	SL-I	156 / 154
2921,5	2921,6437	C_{183} H_{355} O_{21} S	n.d.	SL-II	155 / 153

Supplementary Table 6. List of the $[M+NH_4]^+$ ions of the PDIM and corresponding major molecular species detected by ESI-QTOF mass spectrometry in Ch S^+/C^+ pool.

Peak assignment was performed according to previous studies[7,8]. In PDIMA, 2 mycocerosic acids (MCA) esterify a phthiocerol chain (Fig. 3). The last column indicates the cumulated number of carbon atoms in the 2 MCA chains.

PDIM from Mtb_broth show major forms from *m/z* 1371 to 1427 (see below in blue), depending of the *M. tuberculosis* strain, H37Rv, Erdman or Mt103[7-9], whereas PDIM in EBC show increased abundance of the forms at *m/z* 1469 and 1497 (see below in red; Fig. 3).

Measured *m/z* of $[M+NH4]^+$ adducts	Calculated *m/z* of $[M+NH4]^+$ adducts	Absolute Error (ppm)	Molecular formula of M	Relative Intensity (%)	Total chain length of the 2 MCA (carbon atom number)
1343,3804	1343,3856	3,87	$C_{89}H_{176}O_5$	2,4	53-56
1357,3953	1357,4013	4,42	$C_{90}H_{178}O_5$	2,6	54-57
1371,4132	1371,4169	2,70	$C_{91}H_{180}O_5$	4,0	55-58
1385,4377	1385,4326	3,68	$C_{92}H_{182}O_5$	9,7	56-59
1399,4492	1399,4482	0,71	$C_{93}H_{184}O_5$	5,7	57-60
1413,4642	1413,4639	0,21	$C_{94}H_{186}O_5$	8,9	58-61
1427,4838	1427,4795	3,01	$C_{95}H_{188}O_5$	10,6	59-62
1441,4980	1441,4952	1,94	$C_{96}H_{190}O_5$	10,0	60-63
1455,5133	1455,5108	1,72	$C_{97}H_{192}O_5$	10,6	61-64
1469,5308	1469,5265	2,93	$C_{98}H_{194}O_5$	18,8	62-65
1483,5408	1483,5421	0,88	$C_{99}H_{196}O_5$	5,1	63-66
1497,5616	1497,5578	2,54	$C_{100}H_{198}O_5$	9,4	64-67
1511,5732	1511,5734	0,13	$C_{101}H_{200}O_5$	1,8	65-68
1525,5882	1525,5891	0,59	$C_{102}H_{202}O_5$	0,5	66-69

Supplementary Table 7. List of the $[M-H]^-$ ions of the mycolic acids and corresponding major molecular species detected by ESI-QTOF mass spectrometry in Ch S^+/C^+ pool.
Peak assignment was performed according to previous studies[10].
The distribution of the free mycolic acids molecular species in EBC (Fig. 3) is very similar to that observed for mycolic acid esters in Mtb_broth[10-12].

Measured *m/z* of $[M-H]^-$	Calculated *m/z* of $[M-H]^-$	Absolute Error (ppm)	Molecular formula of M	Relative Intensity (%)	Species	Total carbon number
1108,1406	1108,1356	4,51	C_{76} $H1_{48}$ O_3	2.4	α-	76
1122,1537	1122,1512	2,23	C_{77} H_{150} O_3	2.1	α-	77
1136,1716	1136,1669	4,14	C_{78} H_{152} O_3	27.7	α-	78
1150,1857	1150,1825	2,78	C_{79} H_{154} O_3	3.6	α-	79
1164,2111	1164,1982	11,08	C_{80} H_{156} O_3	10.5	α-	80
1178,2152	1178,2138	1,19	C_{81} H_{158} O_3	0.5	α-	81
1192,2289	1192,2295	0,50	C_{82} H_{160} O_3	0.2	α-	82
1224,2585	1224,2557	2,29	C_{83} H_{164} O_4	2.0	Methoxy-	83
1238,2712	1238,2713	0,08	C_{84} H_{166} O_4	1.3	Methoxy-	84
1252,2955	1252,2870	6,79	C_{85} H_{168} O_4	15.9	Methoxy-	85
1266,3046	1266,3026	1,58	C_{86} H_{170} O_4	1.9	Methoxy-	86
1280,3279	1280,3183	7,50	C_{87} H_{172} O_4	18.2	Methoxy-	87
1294,3385	1294,3339	3,55	C_{88} H_{174} O_4	2.6	Methoxy-	88
1308,3542	1308,3496	3,52	C_{89} H_{176} O_4	5.6	Methoxy-	89
1322,3693	1322,3652	3,10	C_{90} H_{178} O_4	2.8	Methoxy-	90
1336,3824	1336,3809	1,12	C_{91} H_{180} O_4	0.9	Methoxy-	91
1350,3977	1350,3965	0,89	C_{92} H_{182} O_4	2.0	Methoxy-	92

Supplementary Table 8: Number of proteins detected by proteomic analysis in individual EBCs.

Number of proteins detected	Ad S^+	Ad healthy	Ad pneumo	Ch S^+/C^+	Ch healthy
M. tuberculosis					
Minimal	147	0	0	14	0
Maximal	1288	0	0	17	0
Cumulated	1432	0	0	23	0
Human					
Minimal	161	248	232	121	204
Maximal	189	282	275	203	274
Cumulated	331	348	328	227	365

Supplementary Data 1: List of Mtb proteins detected by proteomic analysis in individual EBCs.

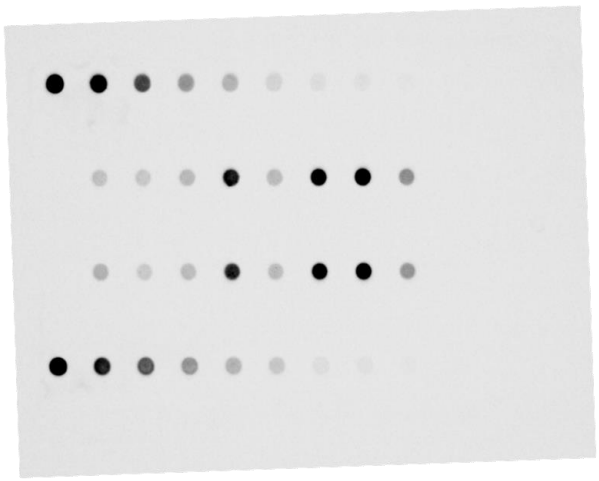
Supplementary Data 2: Abundance of proteins detected by proteomic analysis in individual EBCs.

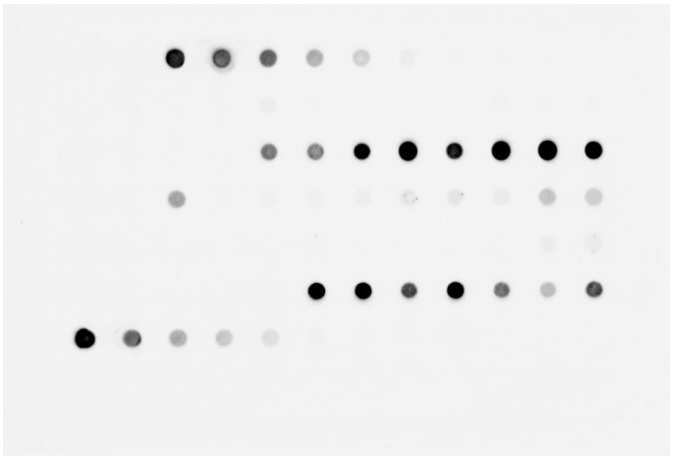
Proteins highlighted in blue were previously described to be released in extracellular vesicles[13-16]. The citing reference(s) are indicated for each protein.

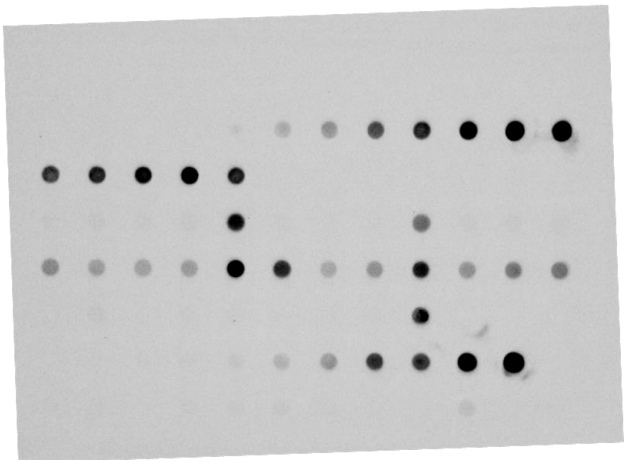
REFERENCES

1. Gilleron, M., *et al.* Acylation state of the phosphatidylinositol mannosides from Mycobacterium bovis bacillus Calmette Guerin and ability to induce granuloma and recruit natural killer T cells. *J Biol Chem* **276**, 34896-34904 (2001).
2. Gilleron, M., Quesniaux, V.F. & Puzo, G. Acylation state of the phosphatidylinositol hexamannosides from Mycobacterium bovis bacillus Calmette Guerin and mycobacterium tuberculosis H37Rv and its implication in Toll-like receptor response. *J Biol Chem* **278**, 29880-29889 (2003).
3. Gilleron, M., Lindner, B. & Puzo, G. MS/MS approach for characterization of the fatty acid distribution on mycobacterial phosphatidyl-myo-inositol mannosides. *Anal Chem* **78**, 8543-8548 (2006).
4. Layre, E., *et al.* Deciphering sulfoglycolipids of Mycobacterium tuberculosis. *J Lipid Res* **52**, 1098-1110 (2011).
5. Rhoades, E.R., Streeter, C., Turk, J. & Hsu, F.F. Characterization of sulfolipids of Mycobacterium tuberculosis H37Rv by multiple-stage linear ion-trap high-resolution mass spectrometry with electrospray ionization reveals that the family of sulfolipid II predominates. *Biochemistry* **50**, 9135-9147 (2011).
6. Goren, M.B., Brokl, O., Das, B.C. & Lederer, E. Sulfolipid I of Mycobacterium tuberculosis, strain H37RV. Nature of the acyl substituents. *Biochemistry* **10**, 72-81 (1971).
7. Camacho, L.R., *et al.* Analysis of the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**, 19845-19854 (2001).
8. Jain, M., *et al.* Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. *Proc Natl Acad Sci U S A* **104**, 5133-5138 (2007).
9. Augenstreich, J., *et al.* The conical shape of DIM lipids promotes Mycobacterium tuberculosis infection of macrophages. *Proc Natl Acad Sci U S A* **116**, 25649-25658 (2019).
10. Laval, F., Laneelle, M.A., Deon, C., Monsarrat, B. & Daffe, M. Accurate molecular mass determination of mycolic acids by MALDI-TOF mass spectrometry. *Anal Chem* **73**, 4537-4544 (2001).
11. Minnikin, D.E. & Brennan, P.J. Lipids of Clinically Significant Mycobacteria. in *Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids* (ed. Goldfine, H.) 1-76 (Springer Nature Switzerland AG 2020, 2020).
12. Marrakchi, H., Laneelle, M.A. & Daffe, M. Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol* **21**, 67-85 (2014).
13. Giri, P.K., Kruh, N.A., Dobos, K.M. & Schorey, J.S. Proteomic analysis identifies highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-treated macrophages. *Proteomics* **10**, 3190-3202 (2010).
14. Lee, J., *et al.* Proteomic analysis of extracellular vesicles derived from Mycobacterium tuberculosis. *Proteomics* **15**, 3331-3337 (2015).
15. Palacios, A., Gupta, S., Rodriguez, G.M. & Prados-Rosales, R. Extracellular vesicles in the context of Mycobacterium tuberculosis infection. *Mol Immunol* **133**, 175-181 (2021).
16. Prados-Rosales, R., *et al.* Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J Clin Invest* **121**, 1471-1483 (2011).

Uncropped scans of Supplementary Fig. 2.







Uncropped scan of Supplementary Fig. 7.