

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

Received: 25 March 2022

Accepted: 5 December 2022

Published online: 14 December 2022

 Check for updates

Sergio Fabián Mosquera-Restrepo ^{1,6}, Sophie Zuberogoitia ^{2,6},
Lucie Gourette^{2,6}, Emilie Layre ², Martine Gilleron ², Alexandre Stella²,
David Rengel ², Odile Burllet-Schiltz², Ana Cecilia Caro ³, Luis F. Garcia ¹,
César Segura ⁴, Carlos Alberto Peláez Jaramillo³, Mauricio Rojas ^{1,5}  &
Jérôme Nigou ² 

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.

Tuberculosis (TB) remains one of the top ten causes of death worldwide and is the leading cause of death from a single infectious agent¹. 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 deaths¹. 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 tools¹. 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 goal².

Three diagnostic priorities have been listed by the WHO and the TB community, including development of a point-of-care biomarker-

¹Cellular Immunology and Immunogenetics Group (GICIG), Institute of Medical Research, Faculty of Medicine, University Research Headquarters (SIU), University of Antioquia (UdeA), Medellín, Colombia. ²Institute of Pharmacology and Structural Biology (IPBS), University of Toulouse, CNRS, University of Toulouse III-Paul Sabatier, Toulouse, France. ³Interdisciplinary Group for Molecular Studies (GIEM), Institute of Chemistry, Faculty of Exact and Natural Sciences, University of Antioquia (UdeA), Medellín, Colombia. ⁴Malaria Group, University Research Headquarters, University of Antioquia (UdeA), Medellín, Colombia. ⁵Flow Cytometry Core, University Research Headquarters (SIU), University of Antioquia, UdeA, Medellín, Colombia. ⁶These authors contributed equally: Sergio Fabián Mosquera-Restrepo, Sophie Zuberogoitia, Lucie Gourette.  e-mail: mauricio.rojas@udea.edu.co; Jerome.Nigou@ipbs.fr

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 condensate³. 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 disease¹.

Since 2015, the WHO has recommended urine tests based on the detection of lipoarabinomannan (LAM), a mycobacterial cell envelope lipoglycan⁶ (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 screening¹. 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 diagnosis³. Indeed, EBC collection is easy, relatively cheap, non-invasive 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 $\mu\text{g}/\text{mL}$), 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-tubesTM 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-35¹⁷). 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^- , 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^- (Fig. 1b, c). In short, EBC samples from TB patients contain μg s of LAM (corresponding to concentrations in the $\mu\text{g}/\text{mL}$ 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 LAM¹⁸ (Supplementary Fig. 1). Pentose D-arabinose is present in mycobacteria and related genera, but is absent in eukaryotes⁶. 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 Fig. 3a, 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 LAM⁶; 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,

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	
Minimal	27	19	21	
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	Isoniazid	Isoniazid	
	Rifampicin	Rifampicin	Rifampicin	
Time of EBC collection	Under antibiotic treatment for <2 week	Before antibiotic treatment	Under antibiotic treatment for <2 week	
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	
Minimal	6		6	
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		Isoniazid	
	Rifampicin		Rifampicin	
Time of EBC collection	Before antibiotic treatment		Under antibiotic treatment for <2 week, and at months 1 and 3 ^c	
Label	Ch S ⁺ /C ⁺		Ch S ⁻ C ⁻	
Control individuals				
	Community-acquired pneumonia		Healthy	
			Children	Adult
	Gram negative	Gram positive		
Number of persons	8	7	15	15
Gender M/F	4/4	4/3	10/5	12/3
Age				
Median	42	41	9	34
Minimal	18	20	7	26
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.

^aTuberculin skin test.^bMean ± SD (range).^cSix (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.

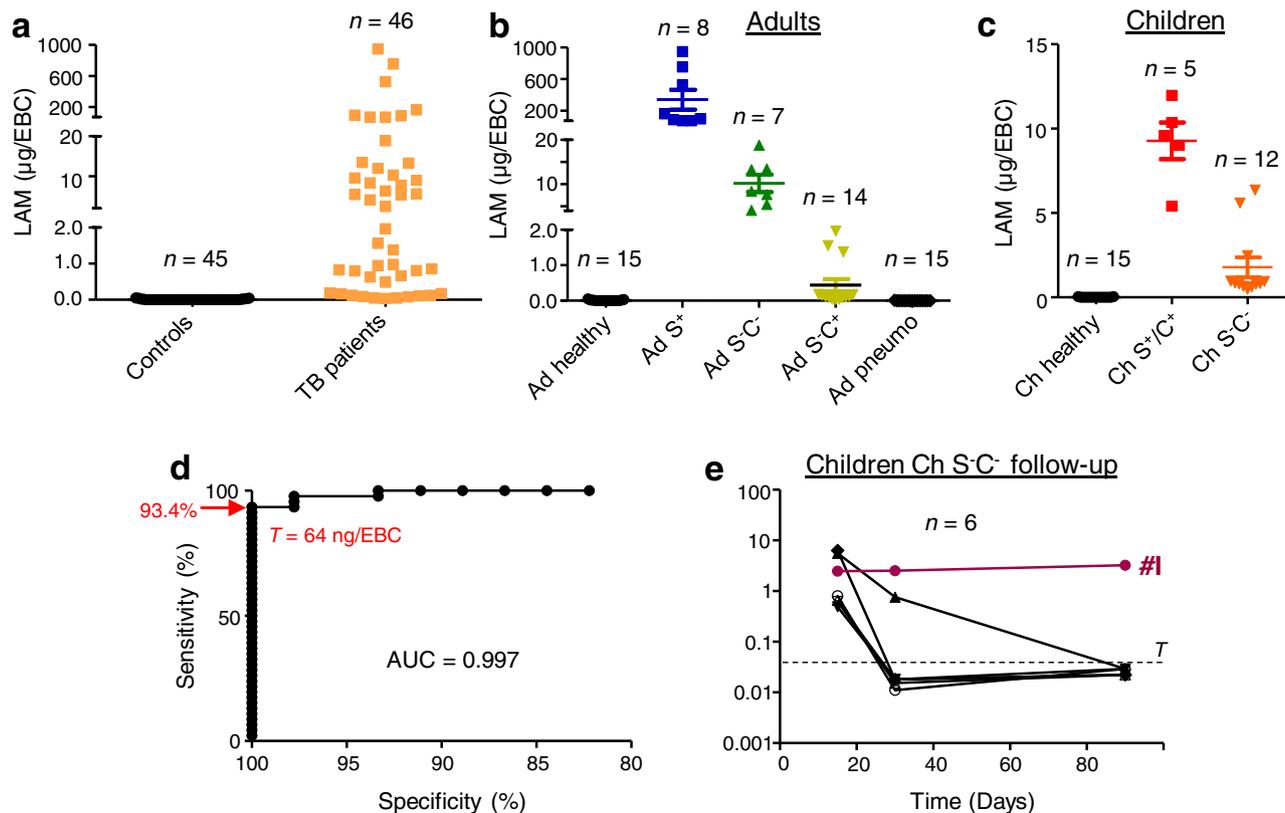


Fig. 1 | Quantification of LAM in EBC from TB patients and control individuals listed in Table 1. Quantity of LAM in EBC samples from all the subjects involved (a), adults (b), and children (c, e) was determined by an immunoassay using the CS-35 anti-LAM antibody. A receiver-operating characteristic (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.

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 (Fig. 1b, 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 #1. It is worth noting that conventional antibiotic treatment (Table 1) for this patient failed. Child #1 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 strategies¹⁹. 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 S^{+/C+} pool) (Supplementary Table 1). The pooled samples should contain levels of LAM (several tens of micrograms) that are sufficient for structural analysis using a combination of highly sensitive analytical procedures we previously developed²⁰. Ad S⁺ pool and Ch S^{+/C+} pool yielded 630 and 720 mg respectively of dried total material after freeze-drying (Supplementary Table 2). Lipids (around 1.5–2% w/w) were removed by organic solvent extraction, and the remaining material was submitted to enzymatic digestion to remove nucleic acids and proteins. The resulting LAM-enriched

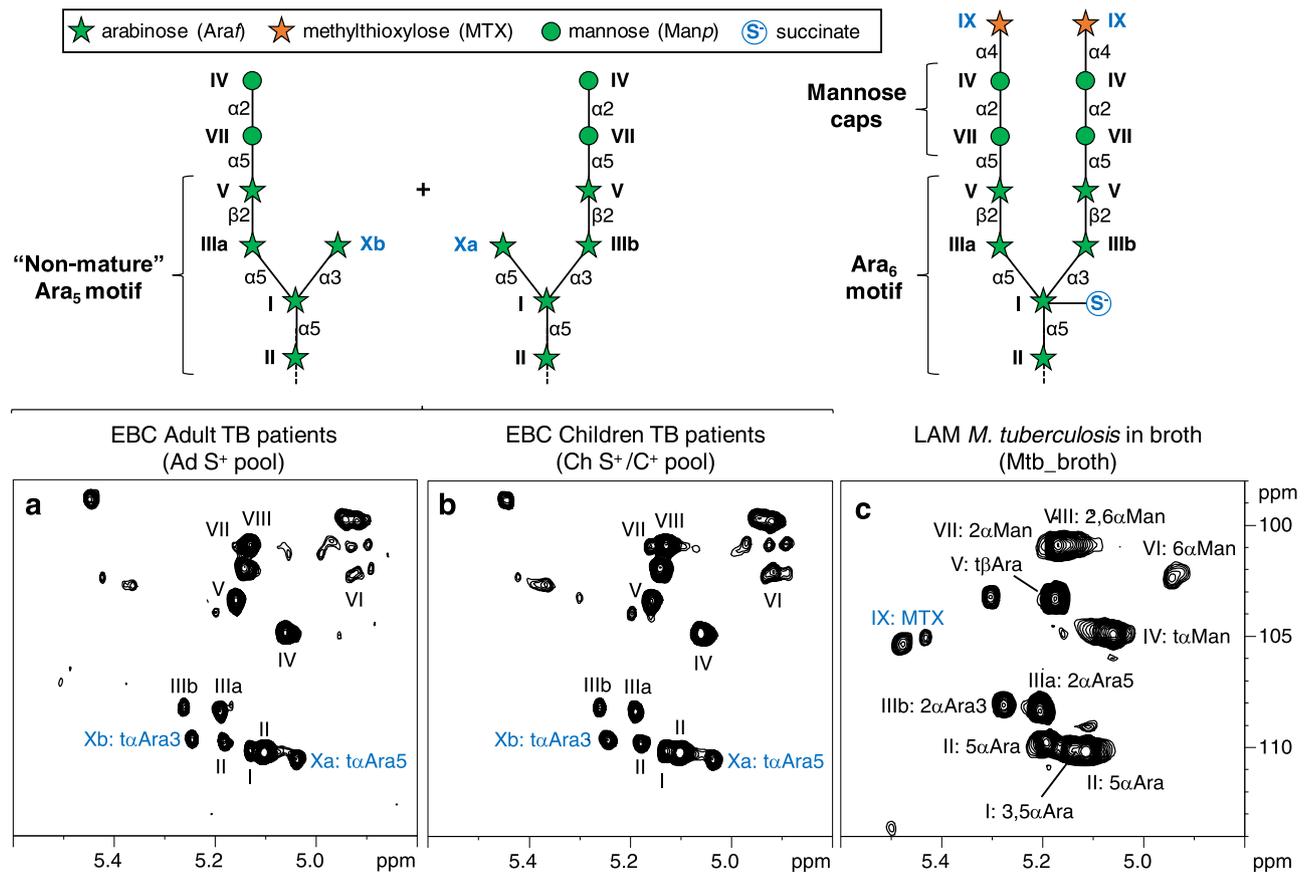


Fig. 2 | Characterization of LAM in EBC by NMR. Expanded region ($\delta^1\text{H}$: 4.80–5.50, $\delta^{13}\text{C}$ 98–114) of the 2D ^1H - ^{13}C HSQC spectrum in D_2O at 298 K of Ad S^+ pool (a) and Ch S^+/C^+ pool (b) LAM-enriched fractions, and Mtb_broth LAM (c). Cartoons show the structure of arabinan side chain termini deduced 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.

fractions contained 0.18 (Ad S^+ pool) and 0.36 (Ch S^+/C^+ pool) mg of arabinose respectively (Supplementary Table 2). The fractions were analyzed by NMR using a 600 MHz spectrometer equipped with a cryogenic probe. The anomeric region of the 2D ^1H - ^{13}C HSQC spectrum obtained with LAM-enriched fractions from both Ad S^+ pool (Fig. 2a) and Ch S^+/C^+ pool (Fig. 2b) exhibited intense cross-peaks that could be unambiguously assigned to glycosidic units building LAM (Supplementary Fig. 1), i.e., 6- α -Manp (VI), 2- α -Manp (VII), 2,6- α -Manp (VIII), t- α -Manp (IV), 5- α -Araf (II), 3,5- α -Araf (I), t- β -Araf (V), 2- α -Araf \rightarrow 5 (IIIa) and 2- α -Araf \rightarrow 3 (IIIb) residues, identical to that found in LAM purified from *M. tuberculosis* grown in broth (Mtb_broth) (Fig. 2c; Supplementary Table 3)^{20–22}. This indicates that the LAM contained in EBCs from both adult and pediatric patients has an overall intact polysaccharidic structure, in agreement with the fact that LAM in EBCs is retained in dialysis tubing with a molecular weight cut off of 8–10 kDa. However, two additional cross-peaks were observed, previously assigned to t- α -Araf \rightarrow 5 (Xa) and t- α -Araf \rightarrow 3 (Xb) units evidenced in an EmbC mutant²². These residues typify a “non-mature” LAM, containing a reduced amount of branched hexa-arabinofuranosides (Ara_6)^{22,23} and an unusual Ara_5 motif (Fig. 2). Ara_6 is the main epitope of the CS-35 anti-LAM antibody¹⁷. In LAM from slow-growing mycobacteria, including *M. tuberculosis*, some Ara_6 chains are substituted at their non-reducing ends by mannose caps⁶ (Fig. 2 and Supplementary Fig. 1) that are key for LAM immunomodulatory properties²⁴. 2- α -Manp 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 and 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 LAM²⁵. The electropherogram obtained with LAM-enriched fractions from both Ad S^+ pool and Ch S^+/C^+ pool exhibited peaks corresponding to mono- (AM), $\alpha(1\rightarrow2)$ -di- (AMM) and $\alpha(1\rightarrow2)$ -tri- (AMMM) mannoside units (Supplementary Fig. 7a), as in Mtb_broth LAM⁶. Quantification²⁵ 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 LAM⁶. LAM purified from Mtb_broth also contains minor covalent modifications on its polysaccharidic domain, such as methylthioxylose or succinyl motifs (Supplementary Fig. 1 and Fig. 2), whose biological significance remains largely unknown⁶. 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 sodium dodecyl-sulfate–polyacrylamide gel electrophoresis (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 Ara₆ motifs with “non-mature” truncated arabinan side chain termini (Ara₅), (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 LAM²⁷.

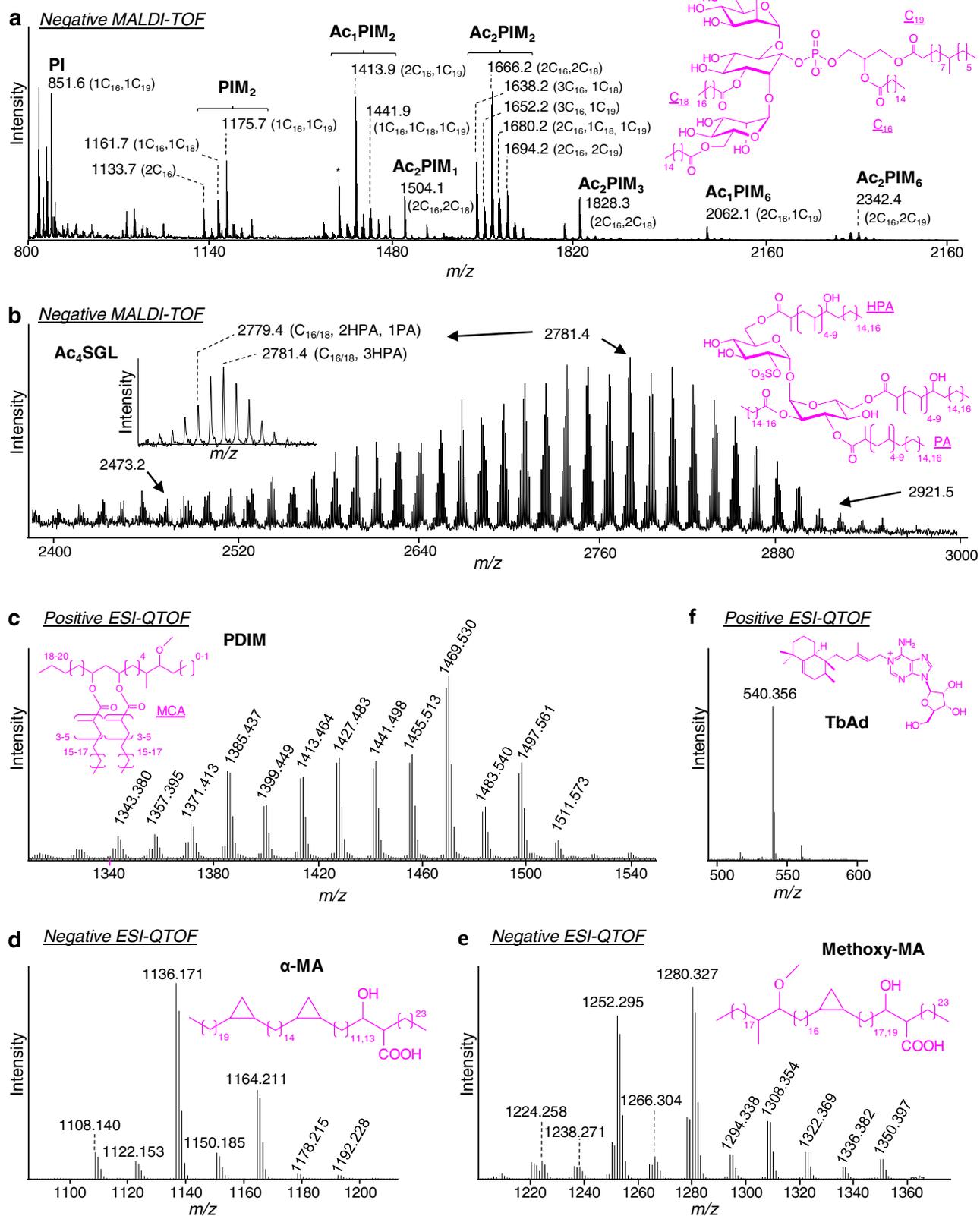
Structure of *Mtb* lipids in TB patient EBCs typifies bacterial metabolism in human lungs

M. tuberculosis produces highly abundant and structurally diverse lipids, some of which are species-specific²⁸. 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 matrix-assisted laser desorption ionization-time of flight mass spectrometry (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- (PIM₂) and hexa- (PIM₆) mannosides, existing in 3 main acyl-forms esterified by 2 (PIM_x) to 4 (Ac₂PIM_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 and 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 (Ac₄SGLs) as observed in Mtb_broth^{31,32}, but with a shift towards higher molecular masses (Fig. 3b and 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 acyl-form profile was also modified, showing an increased proportion of the molecular species composed of longer alkyl chains (Fig. 3c and Supplementary Table 6). This likely represents a 3-carbon atom extension in each of the two mycocerosic acids that compose the molecule³³. Interestingly, similar increases of Ac₄SGLs 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 (Fig. 3d, e and 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; Fig. 3d) and methoxy-mycolic acids containing 85- or 87-carbon atoms (*m/z* at 1252.295 and 1280.327 respectively; Fig. 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 abundant³⁶. Finally, TbAd was detected in both isomeric forms (1-TbAd and N⁶-TbAd) at *m/z* 540.356³⁷ (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, Fig. 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 Fig. 4h, i, we did not detect PDIM for smear-negative and culture-negative children (Ch S^C), in contrast to LAM (Fig. 1c, e), MA (Fig. 4b, c) and TbAd (Fig. 4e, 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). Figure 5 shows the results for selected proteins, GroS (Fig. 5b, g), GroEL2 (Fig. 5c, h), GlnA1 (Fig. 5d), HspX (Fig. 5e, 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 (Fig. 5g–i and 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 (Fig. 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 transmission¹. 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 patients¹. Here, we tested whether EBCs are an accessible sample for the detection of TB with the potential for diagnostic development³. 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

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 PIM₂ (Ac₂PIM₂) that contains 2 palmitic (C₁₆), 1 stearic (C₁₈) and 1 tuberculostearic (C₁₉) acids is drawn. * indicate intense ions that do not correspond to PIM molecular species. **b** Negative MALDI-TOF mass spectrum of Ac₄SGL. 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, 2020²⁸. A detailed peak assignment is shown in Supplementary Tables 4–7.

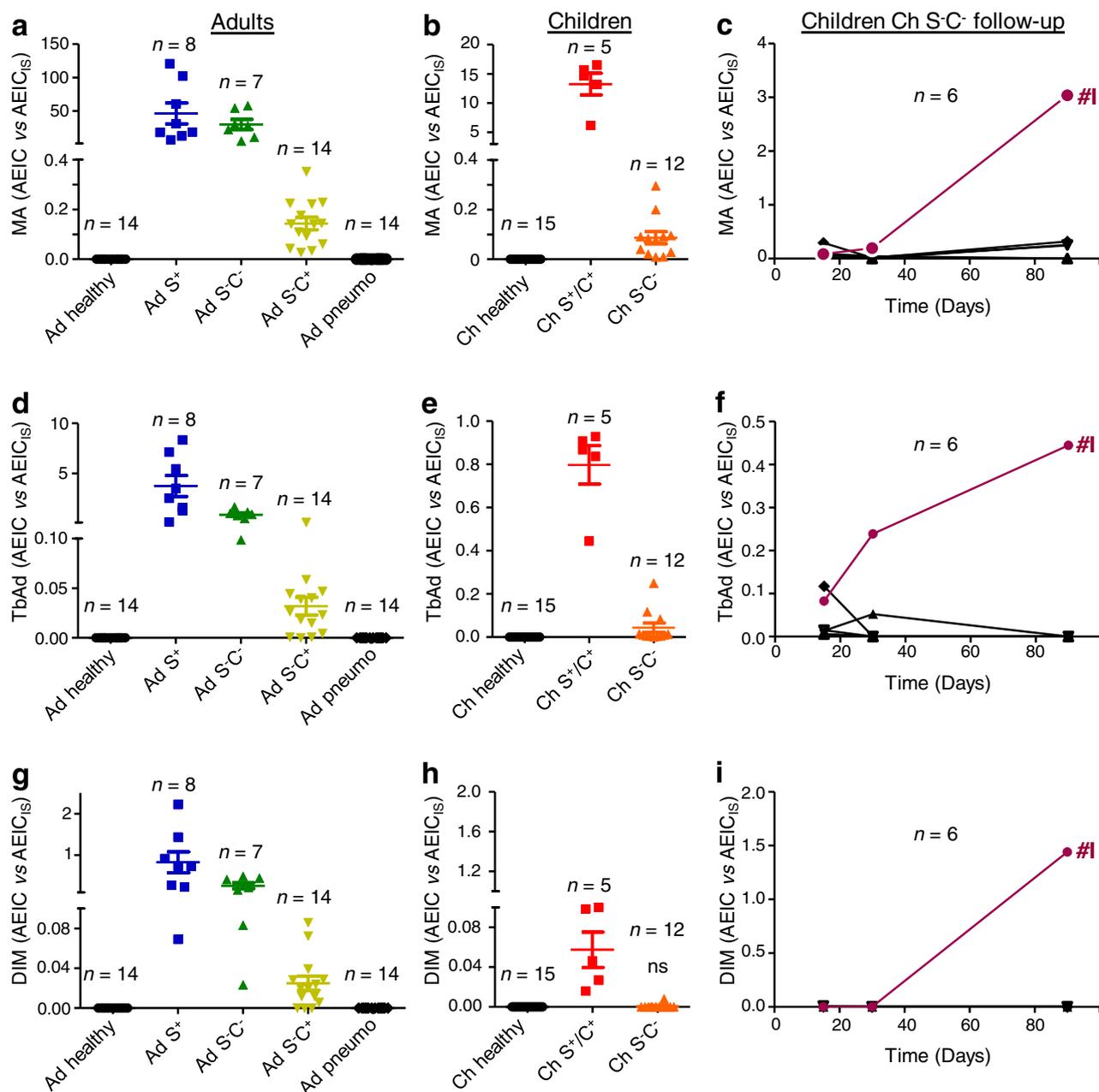


Fig. 4 | Abundance of *M. tuberculosis* lipids in EBC from TB patients and control individuals listed in Table 1. Abundance of MA (**a–c**), TbAd (**d–f**), and PDIM (**g–i**) per EBC from adults and children was determined by SFC-HRMS relatively to 1,2-ditridecanoyl-*sn*-glycero-3-phosphocholine (133 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.

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

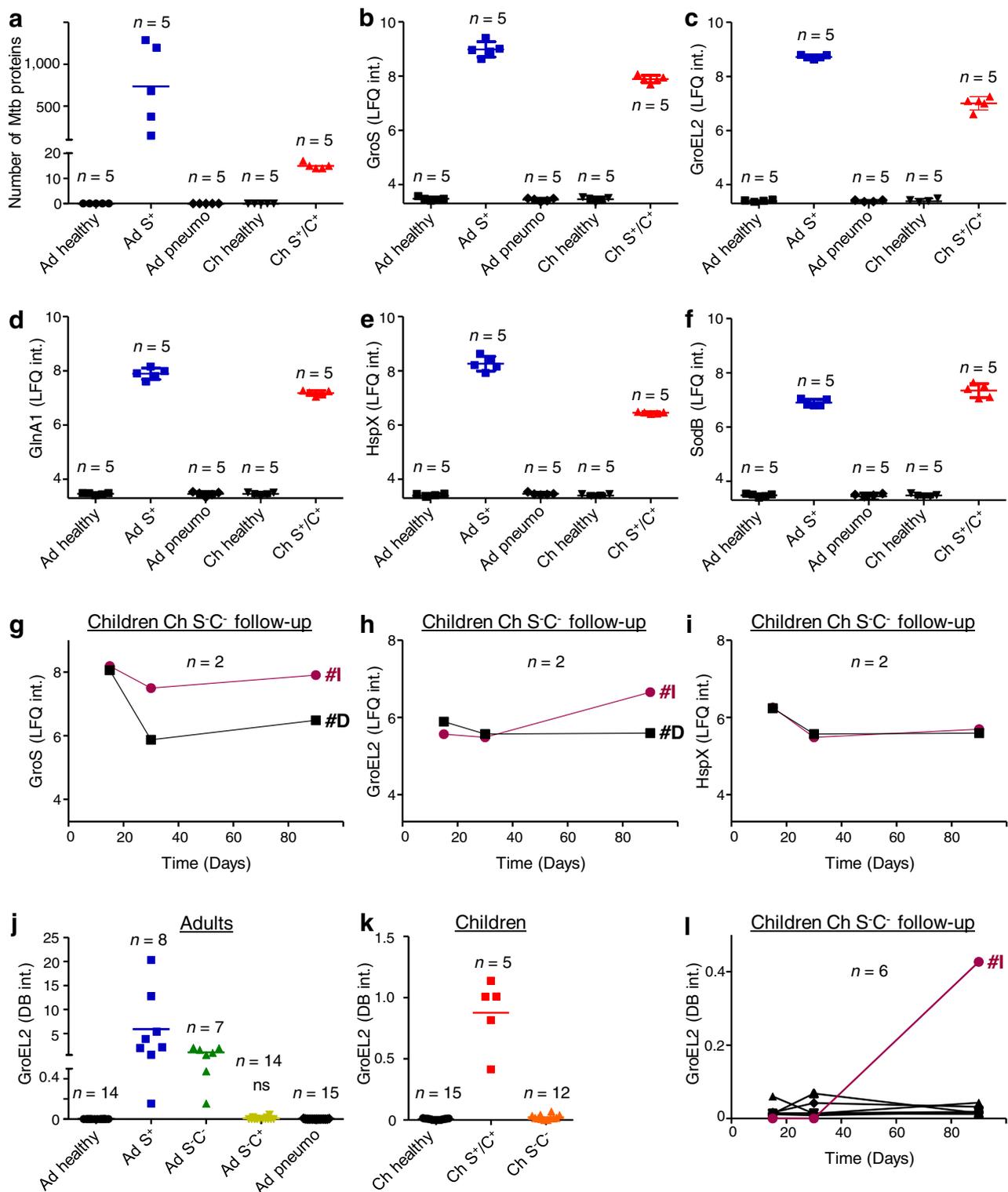


Fig. 5 | Abundance of *M. tuberculosis* proteins in EBC from TB patients and control individuals listed in Table 1. **a** Number of Mtb proteins detected by proteomic analysis in the corresponding groups. **b–i** Abundance of selected Mtb proteins by proteomic analysis. **j–l** Abundance of GroEL2 protein by an immunoblot. 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.

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 infection⁹. 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 poor⁵⁰. 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 patients¹⁸, 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 -10¹¹ to 10¹² 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 Ara₆ 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” Ara₅ motif (Fig. 2), found in both EBC and urine²³, 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 mice²³. 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 urine²³. Accordingly, an immunoassay targeting the MTX motif and tested on urine samples performed well for TB diagnosis⁵². It is worth noting that LAM structure is not specific to Mtb but is shared by other mycobacteria⁶. 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 strategies¹⁹. In addition to structural modifications in LAM mentioned above, we observed a remodeled structure of PIMs, Ac₄SGLs and PDIM. Similar profiles for Ac₄SGLs 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 tissue³³. Interestingly, EBCs contain high amounts of free mycolic acids that typify Mtb biofilm growth³⁶. Although it has long been a matter of debate, a recent study shows that Mtb biofilms are relevant to human TB⁵³ and represent a drug target³⁶.

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 guidelines¹³, 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 15 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 15 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 TB¹⁵. 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 15 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 ($>38^\circ\text{C}$), 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-tubeTM is a non-invasive 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 -20°C . 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 contamination⁵⁹.

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 1 μL). 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 PrimeTM and images were captured using the ChemiDoc System (Bio-Rad). Quantification of dot density was performed using Image LabTM 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 trifluoroacetic 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 electrolyte²⁵.

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

bacteria⁶¹. 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)

Monosaccharides were analyzed as their trimethylsilyl (TMS) derivatives after total acid hydrolysis (2 M trifluoroacetic acid for 2 h at 110 °C) of EBC pools (5 mg), drying, and derivatization with pyridine/hexamethylcyclotriazane/chlorotrimethylsilylamine, 4:2:1, v/v/v, for 15 min at room temperature, and injection in GC-MS in electron ionization mode.

To determine the absolute configuration of arabinose, EBC pools (5 mg) were octanolyzed using 2 M trifluoroacetic in (R)-(-)-2-butanol (Sigma) for 16 h at 110 °C⁶². Octyl-glycosides were dried and converted into their TMS derivatives, as detailed above, before analysis by GC-MS and comparison with D-Ara and L-Ara standards submitted to the same procedure.

GC-MS was performed on a Trace 1300 Thermo GC equipped with a TG-SQC column (15-m, 0.25 mm inner diameter, 0.25 μ m film thickness) coupled to an ISQ quadrupole mass spectrometer. The carrier gas was helium at a flow rate of 1 mL/min. The injector temperature was 250 °C and the temperature separation program was from 100 to 300 °C, at a speed of 15 °C/min.

Mannose cap and western blot analyses of LAM-enriched fractions

For mannose cap analysis and quantification, LAM-enriched fractions (100 μ g) were subjected to mild acid hydrolysis using 0.1 M HCl for 20 min at 110 °C, in the presence of 0.1 nmol mannoheptose (Sigma) as an internal standard. Samples were dried, derivatized by APTS and analyzed by capillary electrophoresis as described above²⁵.

For western blot analysis, LAM-enriched fractions (2–10 μ g) were subjected to SDS-PAGE (15% separating gel)⁶¹ and transferred to a nitrocellulose membrane, which was subsequently probed with CS-35 anti-LAM antibody as described above.

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 μ m/s. 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 305 K 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 D₂O (D, 99.97% from Euriso-top, Saint-Aubin, France), with intermediate lyophilisation, and then dissolved in 0.5 mL D₂O. Samples were analyzed in 200 \times 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. ¹H-¹³C 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 ¹H-¹³C spectra were

recorded with a spectral width of 22,636 Hz in ¹³C and 7211 Hz in ¹H dimensions in order to collect a 2048 \times 512 point data matrix with 16 scans per t1 value expanded to 4096 \times 1024 by zero filling. The relaxation delay was 1.5 s. A sine bell window shifted by $\pi/2$ was applied in both dimensions^{20,21}.

Supercritical fluid chromatography (SFC)-MS lipidomic analysis

EBCs (30 μ L) 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 μ L of water and 105 μ L 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 Meta-Toul 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. One microliter of lipid extract was then injected into the ACQUITY UPC2 Torus Diol column (100 \times 3.0 mm inner diameter (i.d.), particle size: sub-1.7 μ m, Waters) at 48 °C. The lipid families were separated using a gradient of supercritical CO₂/CH₃OH (1.25 mg/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 CH₃OH 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 –3 kV in negative mode; the desolvation gas flow rate 690 mL/h in positive mode and 740 mL/h in negative mode; the cone gas flow 40 L/h; 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 reported⁶⁴ 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) \pm 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 μ g of dried protein extracts were solubilized with 25 μ L 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 μ L S-Trap buffer, 1.5 μ g trypsin was added per sample for digestion in 25 μ L TEAB 50 mM pH 8.

NanoLC-MS/MS analysis of proteins

Tryptic peptides were resuspended in 35 μ L 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 μ g of each sample was loaded on a C18 precolumn (300 μ m inner diameter \times 5 mm, Thermo Fisher Scientific) in a solvent made of

2% acetonitrile and 0.05% trifluoroacetic acid, at a flow rate of 20 $\mu\text{L}/\text{min}$. After 5 min of desalting, the precolumn was switched online with the analytical C18 column (75 μm inner diameter \times 50 cm, Acclaim PepMap 2 μm 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 3×10^6 . 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 1×10^5 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 Proline⁶⁵ 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, 4080 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 \pm 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 \pm 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-validation⁶⁷, and *tidyverse* (version 1.3.2) for data wrangling and graphical output⁶⁸.

Reporting summary

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

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, 4080 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. Source data are provided with this paper.

References

1. WHO. *Global Tuberculosis Report 2021*. 1–43 (WHO, 2021).
2. WHO. *Global Strategy and Targets for Tuberculosis Prevention, Care and Control after 2015*. 1–3 (WHO, 2014).
3. WHO. *High-priority Target Product Profiles for New Tuberculosis Diagnostics: Report of A Consensus Meeting*. 1–97 (WHO, 2014).
4. Kik, S. V., Denking, C. M., Casenghi, M., Vadrnais, 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 Netw. open* **2**, e187617 (2019).
15. Starke, J. R. & Donald, P. R. (eds.). *Handbook of Child and Adolescent Tuberculosis* (Oxford University Press, 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 *Mycobacterium arabinans*. *Microbiology* **148**, 3049–3057 (2002).
18. De, P. et al. Estimation of D-arabinose by gas chromatography/mass spectrometry as surrogate for *Mycobacterium 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.* SO966-842X(22)00174-3 (2022).
28. Minnikin, D. E. & Brennan, P. J. 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. USA* **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 tuberculosis nucleoside products of the virulence-associated enzyme Rv3378c. *Proc. Natl Acad. Sci. USA* **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 (WHO, 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**, e01338-18 (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. *Carbohydr. 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 M.R.): 1115-4592-1439, CODI: 01532, the Programa de Sostenibilidad grant from Universidad de Antioquia (to M.R.), the Fondation pour la Recherche Médicale (Equipes FRM DEQ20180339208; Aide aux projets innovants: Financement d'un ingénieur ING20160435108; to J.N.) and MSDAVENIR (grant Fight-TB to J.N.). S.F.M.R. 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 O.B.S.). 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-INBS-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

M.R. and J.N. conceived the study. S.F.R.M. participated to the collection of EBC samples. S.F.R.M., S.Z., L.G., E.L., M.G., A.S., D.R., O.B.S., A.C.C., L.F.G., C.S., C.A.P.J., M.R., and J.N., designed and performed research, and analyzed data. J.N. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35453-5>.

Correspondence and requests for materials should be addressed to Mauricio Rojas or Jérôme Nigou.

Peer review information *Nature Communications* thanks Laura Via, Novel Chegou, Michal Holcapek and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022