



HAL
open science

A Mycobacterium tuberculosis fingerprint in human breath allows tuberculosis detection

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

► **To cite this version:**

Sergio Fabián Mosquera-Restrepo, Sophie Zuberogoitia, Lucie Gouxette, Emilie Layre, Martine Gilleron, et al.. A Mycobacterium tuberculosis fingerprint in human breath allows tuberculosis detection. Nature Communications, 2022, 13 (1), pp.7751. 10.1038/s41467-022-35453-5 . hal-04260471

HAL Id: hal-04260471

<https://cnrs.hal.science/hal-04260471>

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

2

3 **Sergio Fabián Mosquera-Restrepo^{1#}, Sophie Zuberogoitia^{2#}, Lucie Gouxette^{2#}, Emilie**
4 **Layre², Martine Gilleron², Alexandre Stella², David Rengel², Odile Burlet-Schiltz², Ana**
5 **Cecilia Caro³, Luis F. Garcia¹, César Segura⁴, Carlos Alberto Peláez Jaramillo³, Mauricio**
6 **Rojas^{1,5,*} and Jérôme Nigou^{2,*}**

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

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

12 ³ Grupo Interdisciplinario de Estudios Moleculares (GIEM). Instituto de Química, Facultad de
13 Ciencias Exactas y Naturales. Universidad de Antioquia (UdeA), Medellín, Colombia.

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

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

18 # contributed equally to the work

19 * to whom correspondence should be addressed. MR, e-mail: mauricio.rojas@udea.edu.co, tel:
20 +57 (574) 219 64 61; JN, e-mail: Jerome.Nigou@ipbs.fr , tel: +33 (0)5 61 17 55 04.

21

22

23 **ABSTRACT**

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

40

41 INTRODUCTION

42 Tuberculosis (TB) remains one of the top ten causes of death worldwide and is the leading cause
43 of death from a single infectious agent¹. About a quarter of the world's population is infected
44 with *Mycobacterium tuberculosis* (Mtb), and thus at risk of developing TB. An estimated 10
45 million people developed TB in 2020 resulting in around 1.5 million TB deaths¹. 11% of these
46 TB cases occurred in children. An estimated one-third of all TB cases are, however, not
47 diagnosed or reported, in part due to significant limitations in current diagnostic tools¹. The
48 WHO's End TB Strategy aims for a 90% reduction in TB deaths and an 80% reduction in TB
49 incidence by 2030 relative to 2015 levels. Diagnostic tools are urgently needed to both monitor
50 progress and meet this goal².

51 Three diagnostic priorities have been listed by the WHO and the TB community, including
52 development of a point-of-care biomarker-based test for pulmonary TB^{3,4}. Most conventional
53 diagnostic tests rely on sputum samples, which can be difficult to obtain and have low
54 diagnostic sensitivity in children, HIV-infected individuals and patients with extrapulmonary
55 TB^{3,5}. Therefore, the ideal diagnostic would not rely on sputum samples, and can also detect
56 non-pulmonary TB. To be successfully implemented at point-of-care, a new test should use an
57 easily accessible patient sample, such as urine, blood or breath condensate³. Blood-based
58 diagnostics include interferon- γ release assays which detect the host immune response to Mtb
59 and have been relatively successful for monitoring latent TB. However blood-based tests cannot
60 accurately distinguish between Mtb infection and active TB disease¹.

61 Since 2015, the WHO has recommended urine tests based on the detection of
62 lipoarabinomannan (LAM), a mycobacterial cell envelope lipoglycan⁶ (Supplementary Fig. 1),
63 to help diagnose TB in patients who are seriously ill with HIV^{1,5,7}. However, urinary tests have
64 suboptimal sensitivity, limiting their utility in TB screening¹. One alternative is to test the liquid
65 phase of exhaled air, called the exhaled breath condensate (EBC), which can be sampled by

66 cooling⁸⁻¹⁰. Like urine and blood, the EBC is accessible and thus merits further investigation as
67 a potential fluid that can be sampled for TB diagnosis³. Indeed, EBC collection is easy,
68 relatively cheap, noninvasive and does not require specialized personnel. EBC reflects the
69 composition of the airway lining fluid and may, therefore, contain lung disease-specific markers
70 from infectious agents or infected host tissues⁸⁻¹⁰. Indeed, it was previously reported that fatty
71 acid, oxidative stress and inflammatory mediator profiles in EBC can differentiate TB-infected
72 adults and children from healthy controls^{11,12}.

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

90 **RESULTS**

91 We collected EBCs in health care centers in Medellín (Colombia) from adult and pediatric
92 patients who had been diagnosed for pulmonary TB. Diagnosis was based on clinical, and/or
93 radiographic, and/or bacteriological evidence of pulmonary disease resulting in treatment
94 initiation by the diagnosing clinician¹³⁻¹⁵. Samples were collected using R-tubesTM equipped
95 with a 0.3- μ m filter to avoid contamination with live bacteria. In 15 adult patients (Ad) out of
96 29, and in 5 pediatric patients (Ch) out of 17, the TB diagnosis was microbiologically confirmed
97 by acid-fast bacilli smear (S⁺) and/or Mtb culture (C⁺) (Table 1). The first EBC sample was
98 collected at baseline (prior to or within the first 2 weeks of initiating anti-TB treatment).
99 Additional EBC samples were obtained for 6 pediatric patients at months 1 and 3 of anti-TB
100 treatment (Table 1). Control individuals included 15 healthy adults, 15 healthy children, and 15
101 adult non-TB patients with community-acquired bacterial pneumonia (Table 1). Exclusion
102 criteria were patients positive for HIV, diabetes, cancer, autoimmune diseases,
103 immunosuppressive treatment, previous TB infections and smokers. These exclusions allowed
104 us to focus on assessing TB in the absence of factors that influence disease or might confound
105 detection. Subjects were asked to breathe at a normal frequency for 15 min, yielding at least 1
106 ml of condensate. For normalization between individuals, EBCs were lyophilized and
107 resuspended in a total volume of 250 μ l. Therefore, when available, the quantity of the bacterial
108 molecules in EBC will be provided as mass per EBC.

109

110 **EBCs from TB patients contain high levels of LAM**

111 Since LAM is an established TB biomarker in urine^{1,5,7,16}, we first attempted to detect LAM in
112 EBCs using an anti-LAM antibody (CS-35¹⁷). Dot-blot analysis showed that EBCs of all TB
113 patients at baseline exhibited a positive signal, whereas EBCs of control individuals, either
114 healthy or infected by other pulmonary bacterial pathogens, showed a negative, or extremely

115 weak, response (Supplementary Fig. 2). LAM content in EBCs was quantified relative to LAM
116 purified from *M. tuberculosis* H37Rv using a standard curve (Supplementary Fig. 2).
117 Unexpectedly, while the apparent concentration of LAM in urine from TB patients usually
118 ranges from pg to ng/ml^{5,7}, the apparent quantity of LAM in EBCs from TB patients at baseline
119 ranged from 40 ng to 947 μ g (Fig. 1a). More specifically the amount of LAM per EBC in
120 different groups ranged from 70 to 947 μ g in Ad S⁺, 4 to 19 μ g in Ad S⁻C⁻, 0.04 to 1.5 μ g in Ad
121 S⁻C⁺, 5 to 12 μ g in Ch S⁺/C⁺ and 0.48 to 6.4 μ g in Ch S⁻C⁻ (Figs. 1b and c). In short, EBC
122 samples from TB patients contain μ gs of LAM (corresponding to concentrations in the μ g/ml
123 range), manifold higher than expected based on other patient fluids.
124 These unexpectedly high apparent amounts of LAM prompted us to validate whether LAM is
125 indeed present in TB patient EBCs using cross-validation with a method other than the anti-
126 LAM antibody. We therefore tested for the presence of pentose D-arabinose, a component of
127 LAM¹⁸ (Supplementary Fig. 1). Pentose D-arabinose is present in mycobacteria and related
128 genera, but is absent in eukaryotes⁶. After acid hydrolysis, chemical derivatization, and analyses
129 by capillary electrophoresis monitored by laser-induced fluorescence (CE-LIF) and gas
130 chromatography coupled to mass spectrometry (GC-MS), we could indeed detect arabinose in
131 EBCs from TB patients but not control individuals (Supplementary Figs. 3a and b). More
132 specifically, we detected D-, but not L-arabinose (Supplementary Fig. 3c). Interestingly, the
133 monosaccharides mannose and glucose were also detected. Mannose is, in addition to
134 arabinose, a major building block of LAM⁶; glucose is also a constituent of Mtb cell envelope
135 compounds, such as the polysaccharide α -glucan or trehalose-based glycolipids. However, both
136 mannose and glucose are also normally present in the human host. Arabinose was detected only
137 after acid hydrolysis indicating that it is polymerized, whereas around 60% of glucose and 10%
138 of mannose were found as free monosaccharides (Supplementary Fig. 3d; Supplementary
139 Tables 1 and 2).

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

146 We next examined the sensitivity and specificity of immunoassay-based LAM quantification
147 in EBC for TB detection (Fig. 1a). These parameters were assessed using a receiver operating
148 characteristic (ROC) analysis (Fig. 1d). Overall ROC performance was assessed by calculating
149 the area under the curve (AUC), which was 0.997 (95% CI, 0.9928 to 1.002; $n = 46$ cases, $n =$
150 45 controls; $P < 0.0001$). At a threshold of 64 ng/EBC, this ROC analysis yielded a sensitivity
151 of 93,4% and a specificity of 100%. By these criteria, only 3 false negatives were present
152 amongst 14 Ad S⁻C⁺ patients. However, it is worth noting that Ad S⁻C⁺ patients were under
153 antibiotic treatment for 2 weeks or less (Table 1), and LAM content in EBC may have decreased
154 compared to samples collected prior to treatment initiation, as suggested by the follow up study
155 of pediatric patients (Fig. 1e). Leave-One-Out cross-validation yielded similar results,
156 providing sensitivity, specificity and accuracy prediction values of 0.93, 0.98 and 0.96,
157 respectively (Supplementary Fig. 5). Therefore, immunoassay-based LAM quantification in
158 EBC allowed us to unambiguously distinguish most TB patients at baseline from healthy
159 control subjects or non-TB patients with community-acquired bacterial pneumonia. This
160 included detecting TB patients that were smear-negative and culture-negative adults (Ad S⁻C⁻)
161 and smear- or culture-positive children (Ch S⁺/C⁺) (apparent LAM amount above 5 μg /EBC).
162 The signal-to-noise ratio was thus found to be above 75. Even smear-negative and culture-
163 negative pediatric patients (Ch S⁻C⁻) under antibiotic treatment for 15 days, with an apparent

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

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

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

177

178 **LAM in EBCs has a non-mature polysaccharide structure**

179 Quantification of LAM by immunoassay and chemical analysis showed a significant, but weak,
180 correlation ($P = 0.0002$, $r = 0.45$; Supplementary Fig. 6). In addition, as mentioned above,
181 numerical values obtained by each method were distributed in different ranges and were
182 inconsistent in many cases. This disparity may be due to difficulties in accurate quantification
183 using the CS-35 anti-LAM antibody and raises the question of the molecular structure of LAM
184 released into human lungs. Having access to the structure of Mtb cell envelope compounds *in*
185 *vivo* would give invaluable insight into the metabolic status of Mtb bacilli during human
186 infection, a key step for the development of new chemo- and immuno-therapeutic or -diagnostic
187 strategies¹⁹. We used two pooled EBC samples, one collected from 50 smear-positive adult TB
188 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 developed²⁰. 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 ¹H-¹³C 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 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-
202 β -Araf (V), 2- α -Araf \rightarrow 5 (IIIa) and 2- α -Araf \rightarrow 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 \rightarrow 5 (Xa) and t- α -Araf \rightarrow 3 (Xb) units
208 evidenced in an EmbC mutant²². These residues typify a “non-mature” LAM, containing a
209 reduced amount of branched hexa-arabinofuranosides (Ara₆)^{22,23} and an unusual Ara₅ motif
210 (Fig. 2). Ara₆ is the main epitope of the CS-35 anti-LAM antibody¹⁷. In LAM from slow-
211 growing mycobacteria, including *M. tuberculosis*, some Ara₆ chains are substituted at their non-
212 reducing ends by mannose caps⁶ (Fig. 2, Supplementary Fig. 1) that are key for LAM
213 immunomodulatory properties²⁴. 2- α -Manp are specific units of mannose caps. Corresponding

214 signals (VII) were detected in NMR spectra of LAM-enriched fractions from both Ad S⁺ pool
215 and Ch S⁺/C⁺ pool (Fig. 2; Supplementary Table 3), albeit with a weaker intensity than in
216 Mtb_broth LAM. This is consistent with the truncation of some of arabinan side chain termini
217 that bear caps (Fig. 2). The presence of mannose caps on LAM in EBCs was confirmed by CE-
218 LIF analysis after mild acid hydrolysis of LAM²⁵. The electropherogram obtained with LAM-
219 enriched fractions from both Ad S⁺ pool and Ch S⁺/C⁺ pool exhibited peaks corresponding to
220 mono- (AM), $\alpha(1\rightarrow2)$ -di- (AMM) and $\alpha(1\rightarrow2)$ -tri-mannoside units (Supplementary Fig. 7a),
221 as in Mtb_broth LAM⁶. Quantification²⁵ indicated that there are approximately 1 mannosyl, 1.5
222 dimannosyl and 0.3 trimannosyl units per LAM molecule in both the Ad S⁺ and Ch S⁺/C⁺ pools,
223 indicating a 3- to 4-fold reduced number of di- and tri-mannoside caps compared to that in
224 Mtb_broth LAM⁶. LAM purified from Mtb_broth also contains minor covalent modifications
225 on its polysaccharidic domain, such as methylthioxylose or succinyl motifs (Supplementary
226 Fig. 1; Fig. 2), whose biological significance remains largely unknown⁶. However, the signals
227 corresponding to both motifs were absent in NMR spectra recorded for LAM-enriched fractions
228 from both the Ad S⁺ and Ch S⁺/C⁺ pools (Fig. 2). Finally, the LAM polysaccharidic domain is
229 attached to a mannosyl-phosphatidyl-*myo*-inositol (MPI) lipid anchor (Supplementary Fig. 1)⁶.
230 Fatty acyl chains esterifying the MPI anchor allow LAM to migrate in SDS-PAGE²⁶. LAM-
231 enriched fractions from Ad S⁺ pool and Ch S⁺/C⁺ pool were subjected to SDS-PAGE followed
232 by western blot analysis. Bands corresponding to LAM were observed, indicating that LAM in
233 EBCs likely contains an intact MPI anchor (Supplementary Fig. 7b). LAM in EBCs showed an
234 apparent molecular weight that is slightly lower than that of Mtb_broth LAM, in agreement
235 with a partial truncation of the arabinan side chain termini.

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

239 modifications: i) a reduction in branched Ara₆ motifs with “non-mature” truncated arabinan
240 side chain termini (Ara₅), ii) no MTX motif, and iii) no succinylation detected. This analysis of
241 LAM structure in patients is helpful for the rational development or improvement of diagnostic
242 antibodies, or vaccines, targeting LAM²⁷.

243

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

245 *M. tuberculosis* produces highly abundant and structurally diverse lipids, some of which are
246 species-specific²⁸. As shown above, EBCs contain 1.5-2 % (w/w) lipids (Supplementary Table
247 2). We tested whether *M. tuberculosis* lipids can be detected using mass spectrometry on pooled
248 EBC lipid extracts. Five families of lipids, namely phosphatidylinositol mannosides (PIMs),
249 sulfoglycolipids (SGLs), phthiocerol dimycocerosate (PDIM), mycolic acids (MA) and
250 tuberculosinyladenosine (TbAd) were detected by either MALDI-TOF and/or SFC-HRMS
251 (Supercritical Fluid Chromatography-High Resolution Mass Spectrometry) in both Ad S⁺ and
252 Ch S⁺/C⁺ pooled samples (Fig. 3). As in *Mtb*_broth, PIM molecular species were present in two
253 glyco-forms, Phosphatidylinositol di- (PIM₂) and hexa- (PIM₆) mannosides, existing in 3 main
254 acyl-forms esterified by 2 (PIM_x) to 4 (Ac₂PIM_x) fatty acyl chains (Fig. 3a)^{29,30}. However, the
255 acyl-form profile was more complex in EBCs, with a high abundance of unusual forms
256 containing stearic acid (Fig. 3a; Supplementary Table 4). The PIM biosynthetic precursor,
257 phosphatidylinositol (PI), was also detected (Fig. 3a). Negative MALDI mass spectrum also
258 revealed a large distribution of tetra-acylated forms of SGLs (Ac₄SGLs) as observed in
259 *Mtb*_broth^{31,32}, but with a shift towards higher molecular masses (Fig. 3b; Supplementary Table
260 5). This higher mass represented a mean extension of 7-carbon atoms per each of the 3
261 (hydroxy)phthioceranyl chains present on the molecule. The PDIM acylform profile was also
262 modified, showing an increased proportion of the molecular species composed of longer alkyl
263 chains (Fig. 3c; Supplementary Table 6). This likely represents a 3-carbon atom extension in

264 each of the two mycocerosic acids that compose the molecule³³. Interestingly, similar increases
265 of Ac₄SGLs and PDIM molecular masses have been previously reported for Mtb grown *in vitro*
266 with propionate or cholesterol as the limiting carbon source, or in lungs of infected mice^{33,34}.
267 Methoxy- and α -mycolic acids were detected in their free fatty acid form (Figs. 3d and e;
268 Supplementary Table 7). The distribution of the molecular species was very similar to that
269 observed for mycolic acid esters in Mtb_broth^{28,35}, the most abundant species being α -mycolic
270 acids containing 78- or 80-carbon atoms (m/z at 1136.171 and 1164.211 respectively; Figs. 3d)
271 and methoxy-mycolic acids containing 85- or 87-carbon atoms (m/z at 1252.295 and 1280.327
272 respectively; Figs. 3e). Notably, free mycolic acids are only present in trace amounts in
273 planktonically grown cells of *M. tuberculosis*, in contrast to biofilm cultures where they are
274 more abundant³⁶. Finally, TbAd was detected in both isomeric forms (1-TbAd and N⁶-TbAd)
275 at m/z 540.356³⁷ (Fig. 3f). We then analyzed individual EBCs and observed that, when detected,
276 the different *M. tuberculosis* lipids showed a profile of molecular species similar to that found
277 in pools. This finding was consistent across TB patients, whatever their clinical and
278 demographic characteristics. This lipid profile thus typifies the *in vivo* metabolism of bacilli in
279 human lungs, suggesting bacteria are growing as a biofilm with host lipids as the carbon source.
280 In order to further characterize Mtb lipids as potential biomarkers, we performed a comparative
281 analysis of the relative abundance of MA, PDIM and TbAd in the different samples based on
282 mass spectrometry signal intensities (relatively to the signal of the 1,2-ditridecanoyl-*sn*-
283 glycerol-3-phosphocholine used as internal standard) (Fig. 4). The relative proportions of
284 methoxy- vs α -mycolic acids were constant across samples. Therefore, Figs. 4a-c show the
285 cumulative relative abundances of both mycolic acid types. The overall distribution of MA,
286 PDIM and TbAd was very similar to that obtained for LAM quantification (Fig. 1). The best
287 correlation with LAM was obtained for MA ($r = 0.54$, $P < 0.0001$; Supplementary Fig. 6).
288 Measurements of MA and TbAd showed a very good correlation ($r = 0.90$, $P < 0.0001$;

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

297

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

299 We next analyzed the protein content of EBCs by proteomics. A total number of 1432 Mtb
300 proteins was detected in the 5 individual EBCs analyzed from the smear positive adults (Ad S⁺)
301 (Supplementary Table 8), with a number of proteins ranging from 147 to 1288 in different
302 patients (Fig. 5a). Only 23 Mtb proteins were detected in smear- or culture-positive children
303 (Ch S⁺/C⁺) (Supplementary Table 8), with a number ranging from 14 to 17 according to the
304 patient (Fig. 5a). In contrast, no Mtb proteins were found in the EBCs of the 5 individuals in
305 each control group (Ad healthy, Ch healthy, Ad pneumo) analyzed (Supplementary Table 8;
306 Fig. 5a; Supplementary Fig. 8). All Mtb proteins detected are listed in Supplementary Data 1.
307 Interestingly, many of them, particularly the most abundant, were previously described to be
308 released in bacterial- and/or host-derived extracellular vesicles³⁸⁻⁴¹ (Supplementary Data 2).
309 Protein abundance was determined by Label-Free Quantification (LFQ) (Supplementary Data
310 2). Changes in abundance (ratio = 1.5) of 656 proteins in adult and 22 proteins in pediatric
311 patients differentiated infected individuals from controls (p < 0.05) (Supplementary Data 2;
312 Supplementary Fig. 8). Fig. 5 shows the results for selected proteins, GroS (Figs. 5b and g),
313 GroEL2 (Figs. 5c and h), GlnA1 (Fig. 5d), HspX (Fig. 5e and i) and SodB (Fig. 5f). Their

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

330 **DISCUSSION**

331 Improving early diagnosis is key to controlling TB burden by decreasing the risk of patient
332 mortality and limiting transmission¹. Sputum is currently the main sample used for
333 microbiological diagnosis of TB. However sputum-based assays, including microscopic
334 examination, nucleic acid amplification or bacterial culture, have clear limitations, particularly
335 in children and paucibacillary patients¹. Here, we tested whether EBCs are an accessible sample
336 for the detection of TB with the potential for diagnostic development³. Whereas previous
337 studies failed to detect Mtb nucleic acids in EBC of smear- and culture-positive pulmonary TB
338 patients^{42,43}, our findings show that assay of bacterial LAM, lipids or proteins efficiently
339 distinguished baseline TB patients. Patients that were smear-negative and culture-negative
340 adults were identified as having active TB using EBC samples. In addition, culture- or smear-
341 positive, and smear-negative and culture-negative, pediatric patients were detected as TB-
342 positive by EBC analysis. The test specifically detected TB patients relative to healthy controls
343 and individuals with community-acquired bacterial pneumonia. Detection of bacterial
344 biomarkers in EBCs fulfills many criteria that are essential for implementation of point-of-care
345 tests: i) EBC collection is easy, rapid, affordable, non-invasive and does not require specialized
346 personnel; ii) although exhaled breath may contain viable Mtb⁴⁴⁻⁴⁶, EBC samples can be
347 collected in a sterile manner by adding a filter to the collecting device; thus, handling and
348 analysis do not need to be performed in a high level biosafety laboratory; iii) Mtb-derived
349 molecules in EBCs are highly abundant and very diverse; they could be assayed with previously
350 developed tools, such as LAM tests developed for urine^{5,7,47,48}. This is exemplified by the LAM
351 assay: LAM is ~1000-fold more abundant in EBC than in urine. EBCs are a promising sample
352 for clinical diagnosis and monitoring. However, widespread use still requires standardization
353 in order to reduce the inter- and intra-subject variability observed across studies for non-volatile
354 compound concentration. Thus the normal physiological ranges of the various potential human

355 biomarkers need to be defined^{8,9}. Although important, normalization appears to be less critical
356 for bacterial biomarkers that are likely to serve as positive/negative indicators of infection⁹.
357 Mtb molecules can be readily and specifically detected in EBC samples, potentially allowing
358 an immediate TB diagnosis and an initiation of treatment during the same clinical encounter.
359 In addition, this approach may address key priorities for development of TB diagnostics^{3,47,49}
360 by allowing, i) detection of TB in children and in paucibacillary specimens, ii) diagnosis of a
361 higher proportion of pediatric TB cases using a sampling method that children and parents are
362 comfortable with, iii) screening for community-based case-finding or for triaging patients in
363 the clinic.

364 To date, the only prognostic marker for end-of-treatment outcome remains culture conversion
365 status, though the sensitivity of this assay is poor⁵⁰. Our data indicate that detection of bacterial
366 biomarkers in EBCs may also provide a tool for real-time monitoring of treatment efficacy,
367 enabling the early detection of disease relapse. The EBC-based analysis would allow clinicians
368 to respond quickly to the therapeutic needs of patients, and may facilitate analysis of efficacy
369 in clinical trials for personalized, anti-TB therapies.

370 A diverse range of Mtb molecules with unexpected high levels were found in EBCs from TB
371 patients. Whereas D-arabinose derived from LAM was previously found in the range of ~10-
372 40 ng/mL in the urine of TB patients¹⁸, we detected it in the range of ~15-120 µg/mL in the
373 EBCs of TB patients at baseline. These data suggest that patients exhale a large quantity,
374 between 0.3 and 2.9 mg, of LAM per day. This represents the amount of LAM found in ~10¹¹
375 to 10¹² CFUs of Mtb grown in broth, and reflects highly metabolically active bacilli. In the
376 present study, we did not investigate EBCs from patients with extrapulmonary TB. Although
377 the amount of LAM or other Mtb molecules might be drastically reduced, the sensitivity of the
378 current tools for LAM detection in the ng/ml range might be sufficient for the diagnosis of
379 extrapulmonary TB through analysis of EBC samples, in those cases where lungs are also

380 affected. Nevertheless, where required, our data provide a basis for the rational improvement
381 of LAM detection tools^{5,7}. By performing a full NMR characterization of LAM present in the
382 lungs of TB patients, our data give insight into the molecular moieties that are best targeted for
383 efficient antibody recognition. LAM in EBCs of both adult and children exhibited an overall
384 structure similar to that found in LAM purified from *M. tuberculosis* grown in broth. However
385 the structure contains significant modifications that are important for the development of
386 diagnostic antibodies. Most particularly, LAM in EBC contains a greatly reduced amount of
387 branched Ara₆ motifs that are very immunogenic, and a major epitope for anti-LAM
388 antibodies^{17,27,51}. This may explain, at least in part, the discrepancies we observed in the LAM
389 quantity values obtained by antibody vs. chemical assays. Generating antibodies that
390 specifically recognize the “non-mature” Ara₅ motif (Fig. 2), found in both EBC and urine²³,
391 should significantly improve the sensitivity of LAM detection in these samples. LAM
392 succinylation was not detected in EBCs, although LAM was found to be hyper-succinylated in
393 the lungs of Mtb-infected C3HeB/FeJ mice²³. LAM in EBC also seems to be devoid of the
394 MTX motif. However, this motif was detected on LAM-derived oligosaccharides from a TB
395 patient’s urine²³. Accordingly, an immunoassay targeting the MTX motif and tested on urine
396 samples performed well for TB diagnosis⁵². It is worth noting that LAM structure is not specific
397 to Mtb but is shared by other mycobacteria⁶. Assessing LAM levels might be insufficient to
398 ascertain Mtb vs non-tuberculous mycobacterial infection in some cases, most particularly if
399 only minute amounts are detected. This might be improved by combining detection of LAM
400 with Mtb-specific lipids, such as SGLs or TbAd, or proteins once appropriate antibodies are
401 developed.

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

405 infection, a key step for the development of new chemo- and immunotherapeutic strategies¹⁹.
406 In addition to structural modifications in LAM mentioned above, we observed a remodeled
407 structure of PIMs, Ac₄SGLs and PDIM. Similar profiles for Ac₄SGLs and PDIM with an
408 increased proportion of high mass molecular species have been previously reported for Mtb
409 grown *in vitro* on propionate or cholesterol as the limiting carbon source, or in lungs of infected
410 mice^{33,34}. Although the structure of LAM and lipids in EBC might not be fully representative
411 of all the LAM and lipid molecules produced and circulating in the body of TB patients, our
412 data indicate that Mtb lipid anabolism is increased in the lung environment, which may support
413 bacterial virulence in this tissue³³. Interestingly, EBCs contain high amounts of free mycolic
414 acids that typify Mtb biofilm growth³⁶. Although it has long been a matter of debate, a recent
415 study shows that Mtb biofilms are relevant to human TB⁵³ and represent a drug target³⁶.
416 In conclusion, our findings indicate that detection of bacterial molecules in EBCs provides an
417 affordable and rapid approach that seems superior to the sputum assay for early diagnosis of
418 TB disease. Moreover, EBC sampling has potential for community-based case screening and
419 for monitoring treatment efficacy in real-time. In addition, our analysis of TB patient EBCs
420 indicates that these samples can be used to monitor changes in Mtb metabolism during
421 infection. Future work will develop the diagnostic method and potential biomarkers that have
422 been newly identified here in larger scale clinical studies. This will include validating the proof-
423 of-concept data obtained in the present pilot study using larger cohorts and different groups of
424 patients, including HIV-infected individuals, from different regions of the world. As a first step
425 toward a transfer to the field, we are now evaluating the efficacy of diagnosing TB using EBC
426 samples and currently commercialized LAM detection tests^{5,7}. However, we conclude that
427 sampling exhaled air is non-invasive and appears to be a powerful approach that overcomes the
428 limitations in existing assays^{45,46,54-58}. An EBC-based approach therefore has potentially broad
429 application for following other pulmonary infectious agents.

430 **METHODS**

431 **Patients and control individuals**

432 Adults and pediatric patients with pulmonary TB were recruited through the TB Control
433 Program of the Secretaría de Salud de Medellín and the Secretaría Seccional de Salud y
434 Bienestar Social de Antioquia, Colombia. The diagnostic tests for TB and HIV were performed
435 in the laboratories of health centers where the patients received primary care. Pulmonary TB
436 diagnosis was performed according to the recommendations of Colombia Ministry of Health
437 and Social Protection
438 (<https://www.minsalud.gov.co/salud/publica/PET/Paginas/Tuberculosis.aspx>), based on
439 international guidelines¹³, and was defined as clinical, and/or radiographic, and/or
440 bacteriological evidence of pulmonary disease resulting in treatment initiation by the
441 diagnosing clinician. After collection of three sputum samples and analysis for acid-fast bacilli
442 smear and mycobacterial culture, adult patients were defined to have culture-positive
443 pulmonary TB if they had Mtb growth in at least one of the three sputum cultures, and to have
444 culture-negative pulmonary TB if they had no Mtb growth in the three initial sputum samples,
445 but clinical and radiographic presentation consistent with TB^{13,14}. Smear positivity was defined
446 as having at least one positive acid-fast bacilli smear (Ziehl-Neelsen staining), regardless of
447 quantity, and smear negativity as having no acid-fast bacilli detected in the three initial sputum
448 smears. Pediatric patients were categorized as string test smear- or culture-positive if they had
449 Mtb growth or direct visualization (Ziehl-Neelsen staining) in string test samples, and as string
450 test smear- and culture- negative if they had no Mtb growth or direct visualization in string test
451 samples, but clinical and radiographic presentation consistent with TB, and improvement of
452 symptoms following anti-tuberculous treatment^{1,15}.

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

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

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

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

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

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

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

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

489

490 **Exhaled Breath Condensate (EBC) collection**

491 The EBCs were collected using R-tubes (Respiratory Research, Inc). The R-tubeTM is a
492 noninvasive device that allows the collection and condensation of components of the expiration.
493 R-tubes were equipped with a 0.3- μ m filter (Respiratory Research, Inc) between the
494 mouthpiece assembly and the condensation cartridge trap to prevent bacilli contaminating
495 EBCs. Subjects breathed through a mouthpiece and a two-way non-rebreathing valve, which
496 also served as a saliva trap. They were asked to breathe at a normal frequency for 15 min. At
497 least 1 ml of sample was collected and immediately frozen at -20°C. To determine possible
498 saliva contamination in the EBC samples, α -amylase was measured by the Fishman-Doubilet
499 test using a commercial kit and following the manufacturer's instructions (Abnova, Taipei,
500 Taiwan). EBC samples that were positive for this test were discarded. Accordingly, no α -
501 amylase was detected by proteomic analysis in the EBCs used in the present study, confirming
502 the absence of saliva contamination⁵⁹.

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

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

506

507 **Dot-Blot immunoassays**

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

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

527

528 **Monosaccharide analysis and quantification by capillary electrophoresis monitored by** 529 **laser-induced fluorescence (CE-LIF)**

530 EBCs (12 to 25 μ l) were submitted to total acid hydrolysis using 2 M trifluoroacetic acid for 2 h
531 at 110°C, in the presence of 1 nmol mannoheptose (Sigma) as an internal standard. The samples
532 were dried and mixed with 0.3 μ l of 0.2 M 8-aminopyrene-1,3,6-trisulfonic acid (APTS; Sigma)
533 in 15% acetic acid and 0.3 μ l of a 1 M sodium cyanoborohydride solution dissolved in
534 tetrahydrofuran. The reaction mixture was heated at 55 °C for 90 min and subsequently
535 quenched by the addition of 20 μ l of water. Monosaccharide APTS-derivatives were analyzed
536 and quantified by capillary electrophoresis monitored by laser-induced fluorescence on a
537 P/ACE capillary electrophoresis system (Beckman Instruments, Inc.). Separations were
538 performed using an uncoated fused-silica capillary column (Sigma) of 50- μ m internal diameter
539 and 40-cm effective length (47-cm total length). Analyses were carried out at a temperature of
540 25 °C, in the reverse mode, with an applied voltage of 20 kV using 1% acetic acid (w/v)-30
541 mM triethylamine, pH 3.5, as a running electrolyte²⁵.

542

543 **Preparation of LAM-enriched fractions from EBC pools**

544 The LAM-enriched fraction from EBC pools was prepared using an approach similar to that
545 described for LAM purification from bacteria⁶¹. Lipids were removed by chloroform/methanol
546 extraction (partition chloroform/methanol/water 1.25:1.25:1). The aqueous phase was dried and
547 submitted to digestion by DNase (D4263, Sigma), RNase (R4875, Sigma) and a cocktail of
548 proteases (α -chymotrypsin (C4129, Sigma-Aldrich), *Streptomyces griseus* proteases (P8811,
549 Sigma) and trypsin (T3914, Sigma-Aldrich)) to remove nucleic acids and proteins. The digested
550 solution was dialyzed against water (MWCO 6-8 kDa) and dried, resulting in a fraction enriched
551 in LAM.

552

553 **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 trifluoroacetic acid for 2 h at 110°C) of EBC pools (5 mg), drying, and
556 derivatization with pyridine/hexamethyldisilazane/chlorotrimethylsilylamine, 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°C⁶². 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 above²⁵.

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**

578 MALDI-TOF analyses were performed on an AB Sciex TOF/TOF 5800 mass spectrometer
579 using the reflectron mode^{29,63}. Ionization was achieved by irradiation with an Nd:YAG laser
580 (349 nm) operating with a pulse rate of 400 Hz. The laser intensity was set at 3500 and
581 continuous stage motion was used with a velocity of 600 $\mu\text{m}/\text{sec}$. Data was acquired in the
582 negative ion mode. Typically, spectra from 2500 to 5000 laser shots were summed to obtain the
583 final spectrum. 2,5-dihydroxybenzoic acid (Sigma) was used as the matrix at a concentration
584 of 10 mg/mL in chloroform/methanol 9:1 or 1:1.

585

586 **NMR analysis**

587 NMR spectra were recorded at 305K without sample spinning on a Bruker 600 MHz avance III
588 HD (1H) equipped with a TCi cryoprobe (Bruker Biospin, Germany). LAM-enriched fractions
589 were exchanged in D₂O (D, 99.97% from Euriso-top, Saint-Aubin, France), with intermediate
590 lyophilisation, and then dissolved in 0.5 ml D₂O. Samples were analyzed in 200 x 5 mm UL-5
591 NMR tubes (Euriso-Top, Saint-Aubin, France). Proton and carbon chemical shifts are
592 expressed in parts per million and referenced relative to internal acetone signals at δ_{H} 2.225 and
593 δ_{C} 31.45. ¹H-¹³C correlation spectra were acquired in the echo/anti-echo acquisition mode
594 recorded in the proton-detected mode using the “hsqcetgpsisp2.2” sequence from the Topspin
595 3.5 PL7 software (Bruker Biospin, Germany). The 2D ¹H-¹³C spectra were recorded with a
596 spectral width of 22,636 Hz in ¹³C and 7,211 Hz in ¹H dimensions in order to collect a 2048x512
597 point data matrix with 16 scans per t1 value expanded to 4096x1024 by zero filling. The
598 relaxation delay was 1.5s. A sine bell window shifted by $\pi/2$ was applied in both
599 dimensions^{20,21}.

600

601 **Supercritical Fluid Chromatography (SFC)-MS lipidomic analysis**

602 EBCs (30 μ l) were mixed with 4 ng of 1,2-ditridecanoyl-*sn*-glycero-3-phosphocholine (Avanti
603 Polar Lipids), used as an internal standard allowing the quality control of extraction and
604 normalization for comparative analyses, 12 μ l of water and 105 μ l of chloroform/methanol 1:1
605 in a glass vial. After vortexing and phase separation, the organic phase was recovered for SFC-
606 MS analysis at the MetaToul Lipidomic Core Facility (I2MC, Inserm 1048, Toulouse, France)
607 instrument. The Ultra-Performance Convergence Chromatography (UPC2) was coupled on-line
608 to a Xevo G2-XS time of flight (QTOF) (Waters, Milford, MA, USA), equipped with ESI under
609 the control of MassLynx software, version 4.1. The analysis was performed in both ionization
610 modes (positive and negative) in two separate runs. 1 μ L of lipid extract was then injected into
611 the ACQUITY UPC2 Torus Diol column (100 \times 3.0 mm inner diameter (i.d.), particle size:
612 sub-1.7 μ m, Waters) at 48 $^{\circ}$ C. The lipid families were separated using a gradient of supercritical
613 CO₂/CH₃OH (1.25mg/mL ammonium formate) (1%-50%) at a 1.4 mL/min flow rate and using
614 an active back pressure regulator (ABPR) of 1.900 pounds per square inch (psi). The make-up
615 solvent was CH₃OH used at 0.25 mL/min. The source parameters were set as follows: source
616 temperature 120 $^{\circ}$ C; the capillary voltage 3.2 kV in positive mode and -3kV in negative mode;
617 the desolvation gas flow rate 690 mL/Hr in positive mode and 740 mL/Hr in negative mode;
618 the cone gas flow 40 L/Hr; the desolvation temperature 420 $^{\circ}$ C in positive mode and 390 $^{\circ}$ C in
619 negative mode. The analyses were performed in MS full scan in centroid mode from 50 to 3000
620 Da. Mtb lipids families were detected with characteristic retention times and previously
621 reported⁶⁴ acyl-form profiles and accurate *m/z*.

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

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

629

630 **Sample preparation for label-free proteomics analysis**

631 For each sample 35 µg of dried protein extracts were solubilized with 25 µl of 5% SDS. Proteins
632 were submitted to reduction and alkylation of cysteine residues by addition of TCEP and
633 chloroacetamide to a final concentration respectively of 10 mM and 40 mM. Protein samples
634 were then processed for trypsin digestion on S-trap Micro devices (Protifi) according to the
635 manufacturer's protocol, with the following modifications: precipitation was performed using
636 216 µl S-Trap buffer, 1.5 µg trypsin was added per sample for digestion in 25 µl TEAB 50 mM
637 pH 8.

638

639 **NanoLC-MS/MS analysis of proteins**

640 Tryptic peptides were resuspended in 35 µl of 2% acetonitrile and 0.05% trifluoroacetic acid
641 and analyzed by nano-liquid chromatography (LC) coupled to tandem MS, using an UltiMate
642 3000 system (NCS-3500RS Nano/Cap System; Thermo Fisher Scientific) coupled to an
643 Orbitrap Q-HFX mass spectrometer (Thermo Fisher Scientific). Around 5 µg of each sample
644 was loaded on a C18 precolumn (300 µm inner diameter × 5 mm, Thermo Fisher Scientific) in
645 a solvent made of 2% acetonitrile and 0.05% trifluoroacetic acid, at a flow rate of 20 µl/min.
646 After 5 min of desalting, the precolumn was switched online with the analytical C18 column
647 (75 µm inner diameter × 50 cm, Acclaim PepMap 2µm C18 Thermo Fisher Scientific)
648 equilibrated in 95% solvent A (5% acetonitrile, 0.2% formic acid) and 5% solvent B (80%
649 acetonitrile, 0.2% formic acid). Peptides were eluted using a 10%-45% gradient of solvent B
650 over 60 min at a flow rate of 350 nl/min. The mass spectrometer was operated in data-dependent
651 acquisition mode with the Xcalibur software (version 4.3.73.11). MS survey scans were

652 acquired with a resolution of 60,000 and an AGC target of 3×10^6 . The 12 most intense ions
653 were selected for fragmentation by high-energy collision-induced dissociation, and the resulting
654 fragments were analyzed at a resolution of 30,000, using an AGC target of 1×10^5 and a
655 maximum fill time of 54 ms. Dynamic exclusion was used within 30 s to prevent repetitive
656 selection of the same peptide.

657

658 **Bioinformatics analysis of proteomic datasets**

659 Raw MS files were processed with the Mascot software (version 2.7.0) for database search and
660 Proline⁶⁵ for label-free quantitative analysis (version 2.1.2). Data were searched against
661 MTB_H37Rv_native and Human entries of the UniProtKB protein database (release Swiss-
662 Prot+TrEMBL 2020_10, 4,080 entries and release Swiss-Prot 2020_10, 20,385 entries,
663 respectively). Carbamidomethylation of cysteines was set as a fixed modification, while
664 oxidation of methionine was set as variable modifications. Specificity of trypsin/P digestion
665 was set for cleavage after K or R, and two missed trypsin cleavage sites were allowed. The mass
666 tolerance was set to 10 ppm for the precursor and to 20 mmu in tandem MS mode. Minimum
667 peptide length was set to 7 amino acids, and identification results were further validated in
668 Proline by the target decoy approach using a reverse database at both a PSM and protein false-
669 discovery rate of 1%. For label-free relative quantification of proteins across biological
670 replicates and conditions, cross-assignment of peptide ions peaks was enabled inside a group
671 with a match time window of 1 min, after alignment of the runs with a tolerance of ± 600 s.
672 Median Ratio Fitting computes a matrix of abundance ratios calculated between any two runs
673 from ion abundances for each protein. For each pair-wise ratio, the median of the ion ratios is
674 then calculated and used to represent the protein ratio between these two runs. A least-squares
675 regression is then performed to approximate the relative abundance of the protein in each run
676 in the dataset. This abundance is finally rescaled to the sum of the ion abundances across runs.

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

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

683

684 **Statistics and Reproducibility**

685 EBC samples are unique. Values shown in graphs or tables are the mean of one to several
686 independent measurements (technical replicates) on each individual EBC sample. Statistical
687 differences between TB patient groups and controls (healthy, pneumo) were assessed using a
688 two-tailed Mann-Whitney *U*-test and a α level of 0.05. Correlations between the different Mtb
689 biomarkers were determined by Pearson correlation test. Analyses, including ROC, were done
690 using Prism 5.0 (GraphPad Software). Leave-One-Out cross-validation analysis was carried out
691 with R software (version 4.2.1)⁶⁶, using the *caret* package (version 6.0-93) for the cross-
692 validation⁶⁷, and *tidyverse* (version 1.3.2) for data wrangling and graphical output⁶⁸.

693

694 **DATA AVAILABILITY**

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

REFERENCES

- 700
701
702 1. WHO. Global Tuberculosis Report 2021. 1-43 (2021).
703 2. WHO. Global strategy and targets for tuberculosis prevention, care and control after 2015
704 1-3 (2014).
705 3. WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a
706 consensus meeting. 1-97 (2014).
707 4. Kik, S.V., Denking, C.M., Casenghi, M., Vadnais, C. & Pai, M. Tuberculosis
708 diagnostics: which target product profiles should be prioritised? *Eur Respir J* **44**, 537-540
709 (2014).
710 5. Bulterys, M.A., *et al.* Point-Of-Care Urine LAM Tests for Tuberculosis Diagnosis: A
711 Status Update. *J Clin Med* **9**, 111 (2019).
712 6. Nigou, J., Gilleron, M. & Puzo, G. Lipoarabinomannans: from structure to biosynthesis.
713 *Biochimie* **85**, 153-166 (2003).
714 7. Correia-Neves, M., *et al.* Biomarkers for tuberculosis: the case for lipoarabinomannan.
715 *ERJ Open Res* **5**, 00115-02018 (2019).
716 8. Kuban, P. & Foret, F. Exhaled breath condensate: determination of non-volatile
717 compounds and their potential for clinical diagnosis and monitoring. A review. *Anal Chim*
718 *Acta* **805**, 1-18 (2013).
719 9. Davis, M.D. & Montpetit, A.J. Exhaled Breath Condensate: An Update. *Immunol Allergy*
720 *Clin North Am* **38**, 667-678 (2018).
721 10. Kharitonov, S.A. & Barnes, P.J. Exhaled biomarkers. *Chest* **130**, 1541-1546 (2006).
722 11. Mosquera-Restrepo, S.F., Caro, A.C., Garcia, L.F., Pelaez-Jaramillo, C.A. & Rojas, M.
723 Fatty acid derivative, chemokine, and cytokine profiles in exhaled breath condensates can
724 differentiate adult and children paucibacillary tuberculosis patients. *J Breath Res* **11**,
725 016003 (2017).
726 12. Guzman-Beltran, S., *et al.* Oxidative Stress and Inflammatory Mediators in Exhaled
727 Breath Condensate of Patients with Pulmonary Tuberculosis. A Pilot Study with a
728 Biomarker Perspective. *Antioxidants* **10**, 1572 (2021).
729 13. Lewinsohn, D.M., *et al.* Official American Thoracic Society/Infectious Diseases Society
730 of America/Centers for Disease Control and Prevention Clinical Practice Guidelines:
731 Diagnosis of Tuberculosis in Adults and Children. *Clin Infect Dis* **64**, 111-115 (2017).
732 14. Nguyen, M.H., *et al.* Factors Associated With Sputum Culture-Negative vs Culture-
733 Positive Diagnosis of Pulmonary Tuberculosis. *JAMA network open* **2**, e187617 (2019).
734 15. Starke, J.R. & Donald, P.R. (eds.). *Handbook of child and adolescent tuberculosis*,
735 (Oxford University Press, New York, 2016).
736 16. Hamasur, B., *et al.* Rapid diagnosis of tuberculosis by detection of mycobacterial
737 lipoarabinomannan in urine. *J Microbiol Methods* **45**, 41-52 (2001).
738 17. Kaur, D., Lowary, T.L., Vissa, V.D., Crick, D.C. & Brennan, P.J. Characterization of the
739 epitope of anti-lipoarabinomannan antibodies as the terminal hexaarabinofuranosyl motif
740 of mycobacterial arabinans. *Microbiology* **148**, 3049-3057 (2002).
741 18. De, P., *et al.* Estimation of D-Arabinose by Gas Chromatography/Mass Spectrometry as
742 Surrogate for Mycobacterial Lipoarabinomannan in Human Urine. *PLoS One* **10**,
743 e0144088 (2015).
744 19. Russell, D.G., *et al.* Mycobacterium tuberculosis wears what it eats. *Cell Host Microbe*
745 **8**, 68-76 (2010).
746 20. Larrouy-Maumus, G., *et al.* A glycomic approach reveals a new mycobacterial
747 polysaccharide. *Glycobiology* **25**, 1163-1171 (2015).

- 748 21. Gilleron, M., Bala, L., Brando, T., Vercellone, A. & Puzo, G. Mycobacterium
749 tuberculosis H37Rv parietal and cellular lipoarabinomannans. Characterization of the
750 acyl- and glyco-forms. *J Biol Chem* **275**, 677-684 (2000).
- 751 22. Shi, L., *et al.* The carboxy terminus of EmbC from Mycobacterium smegmatis mediates
752 chain length extension of the arabinan in lipoarabinomannan. *J Biol Chem* **281**, 19512-
753 19526 (2006).
- 754 23. De, P., *et al.* Comparative Structural Study of Terminal Ends of Lipoarabinomannan from
755 Mice Infected Lung Tissues and Urine of a Tuberculosis Positive Patient. *ACS Infect Dis*
756 **6**, 291-301 (2020).
- 757 24. Vergne, I., Gilleron, M. & Nigou, J. Manipulation of the endocytic pathway and
758 phagocyte functions by Mycobacterium tuberculosis lipoarabinomannan. *Front Cell*
759 *Infect Microbiol* **4**, 187 (2014).
- 760 25. Nigou, J., Vercellone, A. & Puzo, G. New structural insights into the molecular
761 deciphering of mycobacterial lipoglycan binding to C-type lectins: lipoarabinomannan
762 glycoform characterization and quantification by capillary electrophoresis at the
763 subnanomole level. *J Mol Biol* **299**, 1353-1362 (2000).
- 764 26. Venisse, A., Berjeaud, J.M., Chaurand, P., Gilleron, M. & Puzo, G. Structural features of
765 lipoarabinomannan from Mycobacterium bovis BCG. Determination of molecular mass
766 by laser desorption mass spectrometry. *J Biol Chem* **268**, 12401-12411 (1993).
- 767 27. Corrigan, D.T., Ishida, E., Chatterjee, D., Lowary, T.L. & Achkar, J.M. Monoclonal
768 antibodies to lipoarabinomannan/arabinomannan - characteristics and implications for
769 tuberculosis research and diagnostics. *Trends Microbiol* (2022).
- 770 28. Minnikin, D.E. & Brennan, P.J. Lipids of Clinically Significant Mycobacteria. in *Health*
771 *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids* (ed.
772 Goldfine, H.) 1-76 (Springer Nature Switzerland AG 2020, 2020).
- 773 29. Gilleron, M., Quesniaux, V.F. & Puzo, G. Acylation state of the phosphatidylinositol
774 hexamannosides from Mycobacterium bovis bacillus Calmette Guerin and
775 mycobacterium tuberculosis H37Rv and its implication in Toll-like receptor response. *J*
776 *Biol Chem* **278**, 29880-29889 (2003).
- 777 30. Gilleron, M., Nigou, J., Nicolle, D., Quesniaux, V. & Puzo, G. The acylation state of
778 mycobacterial lipomannans modulates innate immunity response through toll-like
779 receptor 2. *Chem Biol* **13**, 39-47 (2006).
- 780 31. Layre, E., *et al.* Deciphering sulfoglycolipids of Mycobacterium tuberculosis. *J Lipid*
781 *Res* **52**, 1098-1110 (2011).
- 782 32. Rhoades, E.R., Streeter, C., Turk, J. & Hsu, F.F. Characterization of sulfolipids of
783 Mycobacterium tuberculosis H37Rv by multiple-stage linear ion-trap high-resolution
784 mass spectrometry with electrospray ionization reveals that the family of sulfolipid II
785 predominates. *Biochemistry* **50**, 9135-9147 (2011).
- 786 33. Jain, M., *et al.* Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids
787 via metabolic coupling. *Proc Natl Acad Sci U S A* **104**, 5133-5138 (2007).
- 788 34. Griffin, J.E., *et al.* Cholesterol catabolism by Mycobacterium tuberculosis requires
789 transcriptional and metabolic adaptations. *Chem Biol* **19**, 218-227 (2012).
- 790 35. Marrakchi, H., Laneelle, M.A. & Daffe, M. Mycolic acids: structures, biosynthesis, and
791 beyond. *Chem Biol* **21**, 67-85 (2014).
- 792 36. Ojha, A.K., *et al.* Growth of Mycobacterium tuberculosis biofilms containing free
793 mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* **69**, 164-174 (2008).
- 794 37. Layre, E., *et al.* Molecular profiling of Mycobacterium tuberculosis identifies
795 tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc*
796 *Natl Acad Sci U S A* **111**, 2978-2983 (2014).

- 797 38. Giri, P.K., Kruh, N.A., Dobos, K.M. & Schorey, J.S. Proteomic analysis identifies highly
798 antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-
799 treated macrophages. *Proteomics* **10**, 3190-3202 (2010).
- 800 39. Lee, J., *et al.* Proteomic analysis of extracellular vesicles derived from Mycobacterium
801 tuberculosis. *Proteomics* **15**, 3331-3337 (2015).
- 802 40. Palacios, A., Gupta, S., Rodriguez, G.M. & Prados-Rosales, R. Extracellular vesicles in
803 the context of Mycobacterium tuberculosis infection. *Mol Immunol* **133**, 175-181 (2021).
- 804 41. Prados-Rosales, R., *et al.* Mycobacteria release active membrane vesicles that modulate
805 immune responses in a TLR2-dependent manner in mice. *J Clin Invest* **121**, 1471-1483
806 (2011).
- 807 42. Shenai, S., *et al.* Exploring alternative biomaterials for diagnosis of pulmonary
808 tuberculosis in HIV-negative patients by use of the GeneXpert MTB/RIF assay. *J Clin*
809 *Microbiol* **51**, 4161-4166 (2013).
- 810 43. Jain, R., Schriever, C.A., Danziger, L.H., Cho, S.H. & Rubinstein, I. The IS6110
811 repetitive DNA element of Mycobacterium tuberculosis is not detected in exhaled breath
812 condensate of patients with active pulmonary tuberculosis. *Respiration* **74**, 329-333
813 (2007).
- 814 44. Patterson, B., *et al.* Cough-independent production of viable Mycobacterium tuberculosis
815 in bioaerosol. *Tuberculosis* **126**, 102038 (2021).
- 816 45. Wood, R., *et al.* Real-Time Investigation of Tuberculosis Transmission: Developing the
817 Respiratory Aerosol Sampling Chamber (RASC). *PLoS One* **11**, e0146658 (2016).
- 818 46. Patterson, B., *et al.* Detection of Mycobacterium tuberculosis bacilli in bio-aerosols from
819 untreated TB patients. *Gates Open Res* **1**, 11 (2018).
- 820 47. Walzl, G., *et al.* Tuberculosis: advances and challenges in development of new
821 diagnostics and biomarkers. *Lancet Infect Dis* **18**, e199-e210 (2018).
- 822 48. Goletti, D., Lee, M.R., Wang, J.Y., Walter, N. & Ottenhoff, T.H.M. Update on
823 tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of
824 cure from disease. *Respirology* **23**, 455-466 (2018).
- 825 49. WHO. WHO consolidated guidelines on tuberculosis: Module 3: diagnosis - rapid
826 diagnostics for tuberculosis detection 1-82 (2020).
- 827 50. Kurbatova, E.V., *et al.* Sputum culture conversion as a prognostic marker for end-of-
828 treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis
829 of data from two observational cohort studies. *Lancet Respir Med* **3**, 201-209 (2015).
- 830 51. Choudhary, A., *et al.* Characterization of the Antigenic Heterogeneity of
831 Lipoarabinomannan, the Major Surface Glycolipid of Mycobacterium tuberculosis, and
832 Complexity of Antibody Specificities toward This Antigen. *J Immunol* **200**, 3053-3066
833 (2018).
- 834 52. Sigal, G.B., *et al.* A Novel Sensitive Immunoassay Targeting the 5-Methylthio-d-
835 Xylofuranose-Lipoarabinomannan Epitope Meets the WHO's Performance Target for
836 Tuberculosis Diagnosis. *J Clin Microbiol* **56**(2018).
- 837 53. Chakraborty, P., Bajeli, S., Kaushal, D., Radotra, B.D. & Kumar, A. Biofilm formation
838 in the lung contributes to virulence and drug tolerance of Mycobacterium tuberculosis.
839 *Nat Commun* **12**, 1606 (2021).
- 840 54. Phillips, M., *et al.* Point-of-care breath test for biomarkers of active pulmonary
841 tuberculosis. *Tuberculosis (Edinb)* **92**, 314-320 (2012).
- 842 55. Williams, C.M., *et al.* Exhaled Mycobacterium tuberculosis output and detection of
843 subclinical disease by face-mask sampling: prospective observational studies. *Lancet*
844 *Infect Dis* **20**, 607-617 (2020).

- 845 56. Chen, D., Bryden, W.A. & Wood, R. Detection of Tuberculosis by The Analysis of
846 Exhaled Breath Particles with High-resolution Mass Spectrometry. *Sci Rep* **10**, 7647
847 (2020).
- 848 57. Bruins, M., *et al.* Diagnosis of active tuberculosis by e-nose analysis of exhaled air.
849 *Tuberculosis* **93**, 232-238 (2013).
- 850 58. Maiga, M., Abaza, A. & Bishai, W.R. Current tuberculosis diagnostic tools & role of
851 urease breath test. *Indian J Med Res* **135**, 731-736 (2012).
- 852 59. Muccilli, V., *et al.* Protein profile of exhaled breath condensate determined by high
853 resolution mass spectrometry. *J Pharm Biomed Anal* **105**, 134-149 (2015).
- 854 60. Laneelle, M.A., Nigou, J. & Daffe, M. Lipid and lipoarabinomannan isolation and
855 characterization. *Methods Mol Biol* **1285**, 77-103 (2015).
- 856 61. Nigou, J., *et al.* The phosphatidyl-myo-inositol anchor of the lipoarabinomannans from
857 *Mycobacterium bovis* bacillus Calmette Guerin. Heterogeneity, structure, and role in the
858 regulation of cytokine secretion. *J Biol Chem* **272**, 23094-23103 (1997).
- 859 62. Gerwig, G.J., Kamerling, J.P. & Vliegthart, J.F.G. Determination of the D and L
860 configuration of neutral monosaccharides by high-resolution capillary G.L.C. *Carbohydr*
861 *Res* **62**, 349-357 (1978).
- 862 63. Gilleron, M., Lindner, B. & Puzo, G. MS/MS approach for characterization of the fatty
863 acid distribution on mycobacterial phosphatidyl-myo-inositol mannosides. *Anal Chem*
864 **78**, 8543-8548 (2006).
- 865 64. Layre, E., *et al.* A comparative lipidomics platform for chemotaxonomic analysis of
866 *Mycobacterium tuberculosis*. *Chem Biol* **18**, 1537-1549 (2011).
- 867 65. Bouyssie, D., *et al.* Proline: an efficient and user-friendly software suite for large-scale
868 proteomics. *Bioinformatics* **36**, 3148-3155 (2020).
- 869 66. R Core Team. R: A language and environment for statistical computing. (R Foundation
870 for Statistical Computing, Vienna, Austria, 2022).
- 871 67. Kuhn, M. caret: Classification and Regression Training. R package version 6.0-93.
872 (2022).
- 873 68. Wickham, H., *et al.* Welcome to the tidyverse. *J Open Source Softw* **4**, 1686 (2019).
874
875

876 **ACKNOWLEDGEMENTS**

877 This study was supported by the COLCIENCIAS grant (to MR): 1115-4592-1439, CODI:
878 01532, the Programa de Sostenibilidad grant from Universidad de Antioquia (to MR), the
879 Fondation pour la Recherche Médicale (Equipes FRM DEQ20180339208; Aide aux projets
880 innovants: Financement d'un ingénieur ING20160435108; to JN) and MSDAVENIR (grant
881 Fight-TB to JN). SFMR received a doctoral fellowship from COLCIENCIAS. This work was
882 also supported in part for proteomics by the French Ministry of Research with the Investment
883 for the Future program dedicated to National Infrastructures in Biology and Health (PIA,
884 Proteomics French Infrastructure, ProFI, ANR-10-INBS-08 to OBS). We thank the Integrated
885 Screening Platform of Toulouse (PICT) for providing access to NMR spectrometers which were
886 funded by CNRS, Université Paul Sabatier, Infrastructures en Biologie Santé et Agronomie
887 (IBiSA), Région Occitanie and European Structural Funds. We thank the MetaToul Lipidomic
888 Core Facility (I2MC, INSERM 1048, Toulouse, France, MetaboHUB-ANR-11-INSB-0010)
889 for assistance and access to the SFC/MS instrument. We would also like to acknowledge the
890 European Union's Horizon 2020 research and innovation program under grant agreement No
891 847762.

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

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

904 The following reagents were obtained through BEI Resources, NIAID, NIH: Monoclonal Anti-
905 *Mycobacterium tuberculosis* LAM, Clone CS-35 (produced in vitro), NR-13811; Monoclonal
906 Anti-*Mycobacterium tuberculosis* GroEL2 (Gene Rv0440), Clone CS-44 (produced in vitro),
907 NR-13813; Monoclonal Anti-*Mycobacterium tuberculosis* HspX (Gene Rv2031c), Clone IT-
908 20 (TB68) (produced in vitro), NR-13607; Monoclonal Anti-*Mycobacterium tuberculosis* KatG
909 (Gene Rv1908c), Clone IT-57 (CDA4) (culture supernatant), NR-13793; Monoclonal Anti-
910 *Mycobacterium tuberculosis* LpqH (Gene Rv3763), IT-54 (produced in vitro), NR-13792;
911 Monoclonal Anti-*Mycobacterium tuberculosis* HBHA (Gene Rv0475), Clone α -HBHA
912 (produced in vitro), NR-13804.

913

914 **AUTHOR CONTRIBUTIONS STATEMENT**

915 MR and JN conceived the study. SFRM participated to the collection of EBC samples.
916 SFRM, SZ, LG, EL, MG, AS, DR, OBS, ACC, LFG, CS, CAPIJ, MR and JN, designed and
917 performed research, and analyzed data. JN wrote the paper.

918

919 **COMPETING INTEREST STATEMENT**

920 The authors declare no competing interests.

921 Table 1. Clinical and demographic characteristics of the adult and pediatric TB patients, and
 922 control individuals, either healthy or with bacterial pneumonia.
 923

	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
Median	34	44	38
Age 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 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 ⁺

924

	Pediatric TB patients	
	String test smear- or culture-positive	String test smear- and culture-negative
	Number of patients	5
Gender M/F	3/2	6/6
Median	9	9
Age 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 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 ⁻

925

	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
Median	42	41	9	34
Age 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

926

927 N/A, not applicable

928 ^a tuberculin skin test

929 ^b mean ± SD (range)

930 ^c Six (3M/3F; BCG scar positive; 6-12 years old) of the 12 patients were followed during

931 antibiotic treatment and EBC collected after 1 and 3 months of antibiotic therapy

932 **FIGURE LEGENDS**

933
934 Fig. 1. Quantification of LAM in EBC from TB patients and control individuals listed in Table
935 1.

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

942

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

944 Expanded region (δ ^1H : 4.80-5.50, δ ^{13}C 98-114) of the 2D ^1H - ^{13}C HSQC spectrum in D_2O at
945 298 K of Ad S^+ pool (a) and Ch S^+/C^+ pool (b) LAM-enriched fractions, and Mtb_broth LAM
946 (c).

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

951

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

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

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

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

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

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

964 Data are representative of at least 2 independent experiments on each EBC pooled sample. The
965 precise stereochemistry of PIMs, SGLs, PDIM and Mycolic acids can be found in Minnikin &
966 Brennan, 2020²⁸. A detailed peak assignment is shown in Supplementary Tables 4-7.

967

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

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

978

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

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

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

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

984 In b-i, values are given as the Label-Free Quantification (LFQ) intensity (int.). The noise
985 background intensity was ~3.3 log. In j-l, values are given in arbitrary units corresponding to
986 intensity (int.) on the Dot Blot (DB) and normalized to levels of GroEL2 in the Mtb cell lysate.

987 In j and k, unless otherwise stated (ns, not significant), the difference between TB patient groups
988 and controls (healthy, pneumo) was statistically significant (Mann-Whitney *U*-test, two-tailed).

989 Error bars represent SEM. Source data are provided as a Source Data file.

990

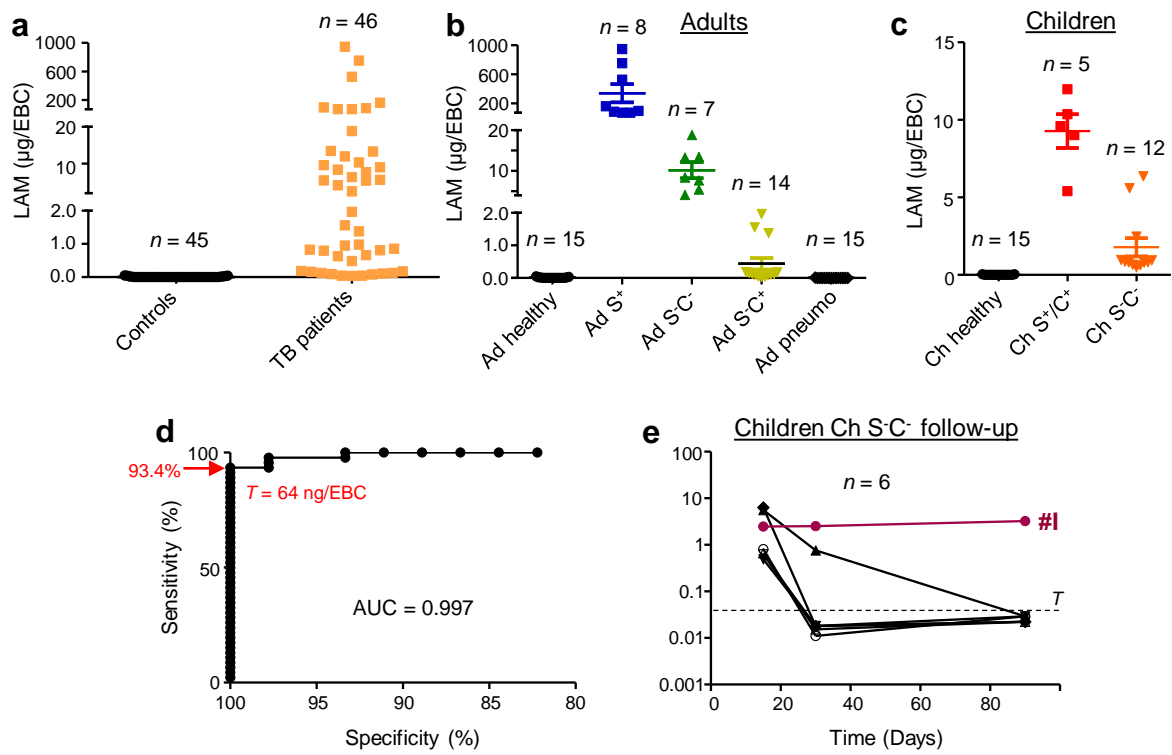


Fig. 1

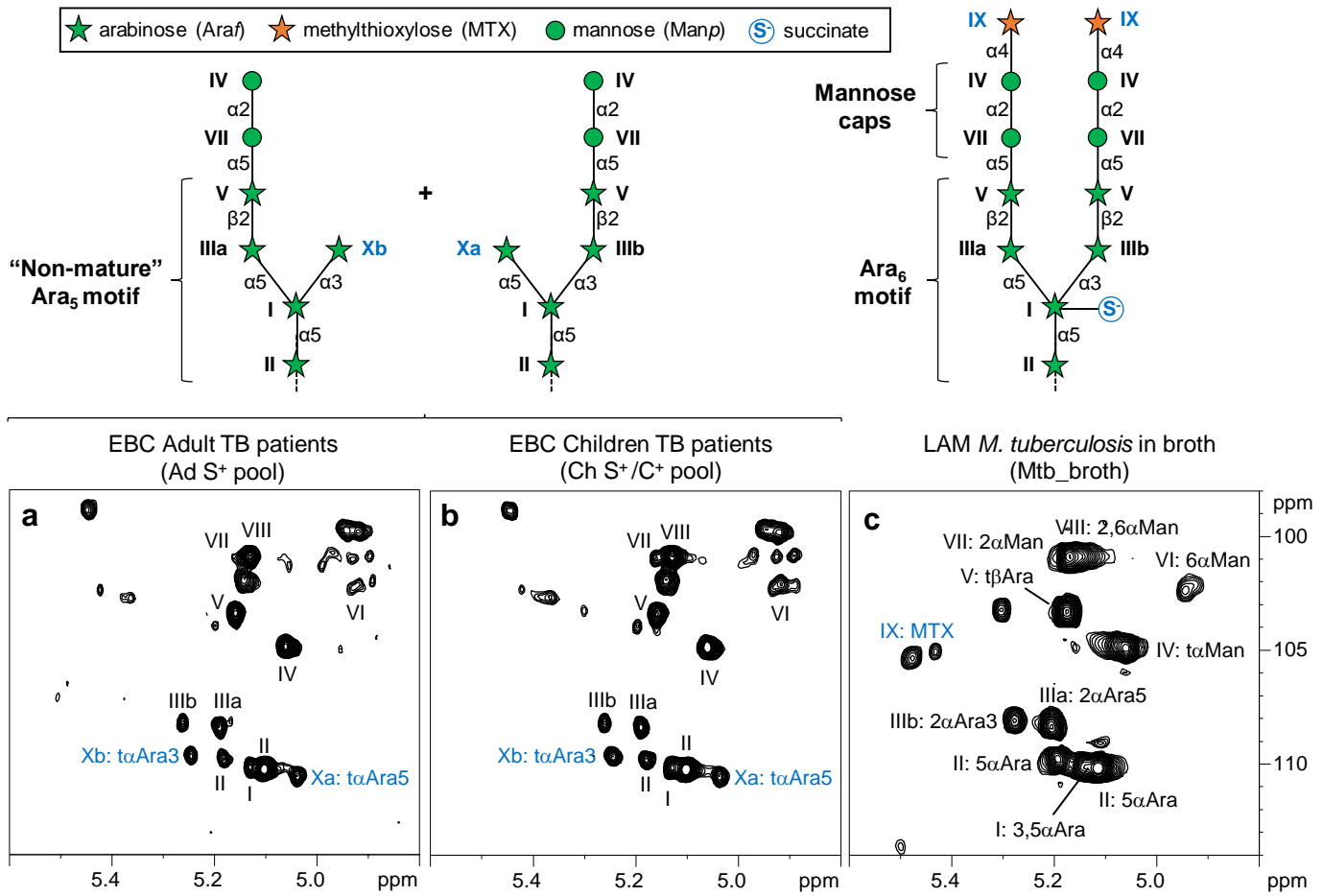


Fig. 2

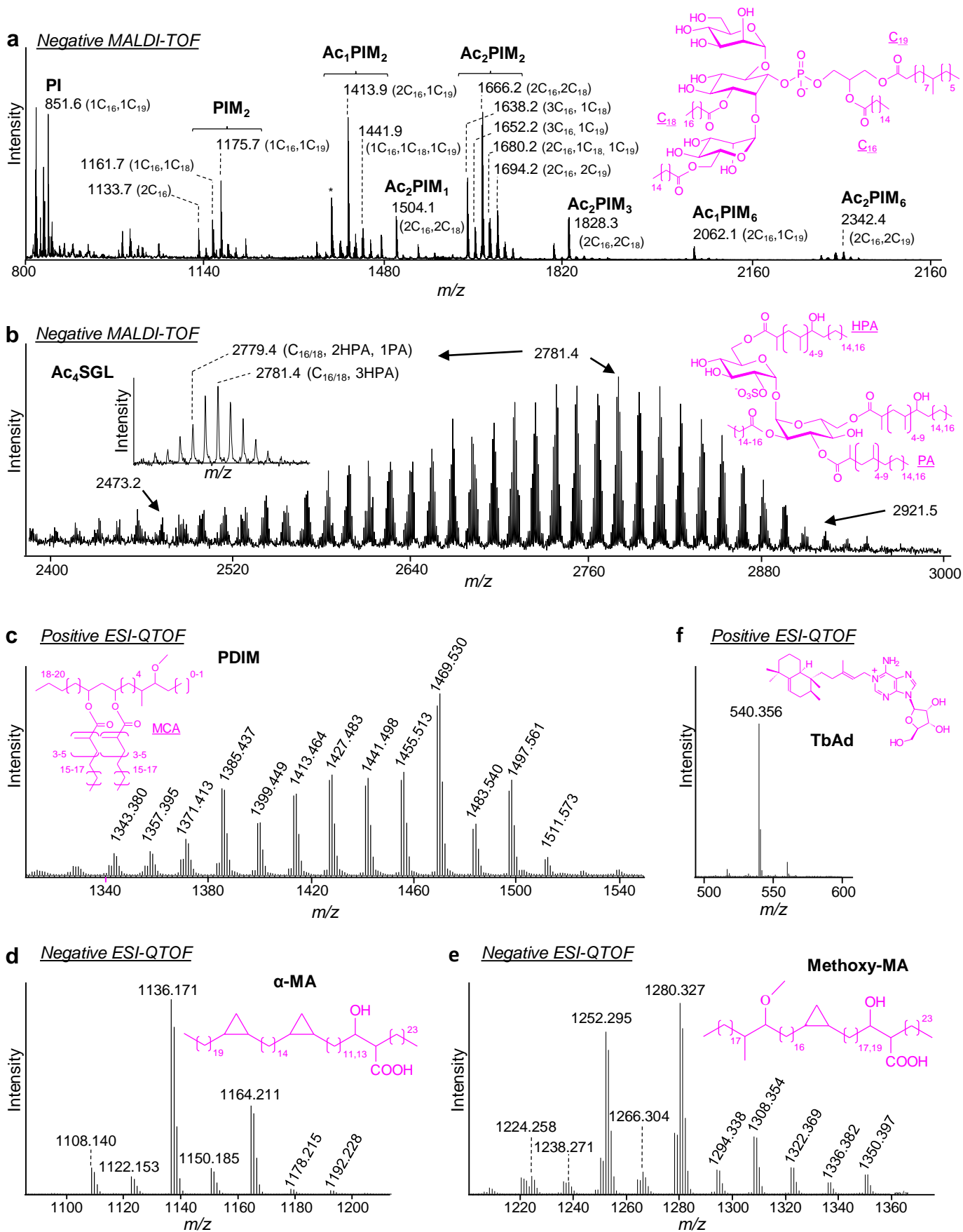


Fig. 3

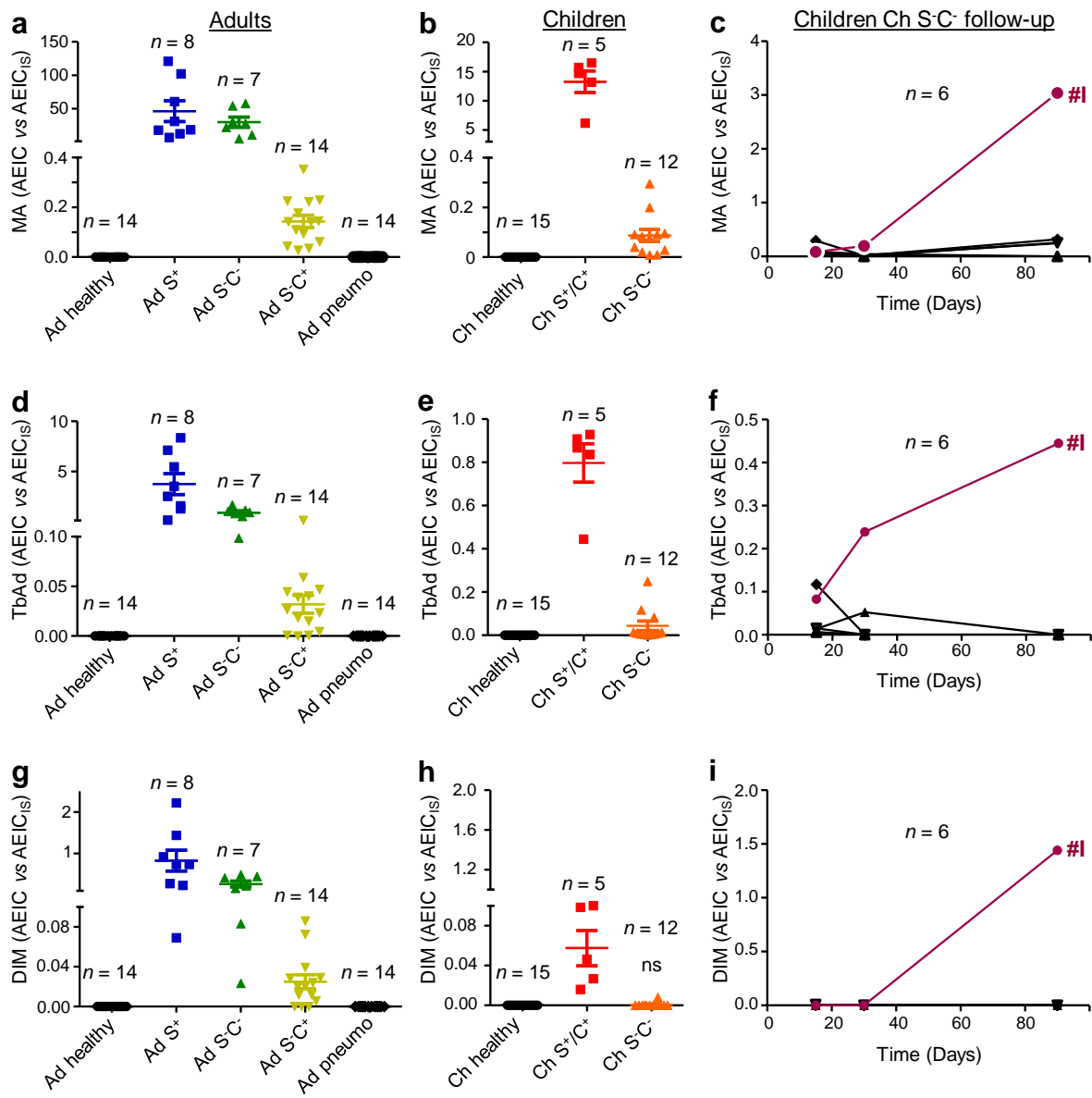


Fig. 4

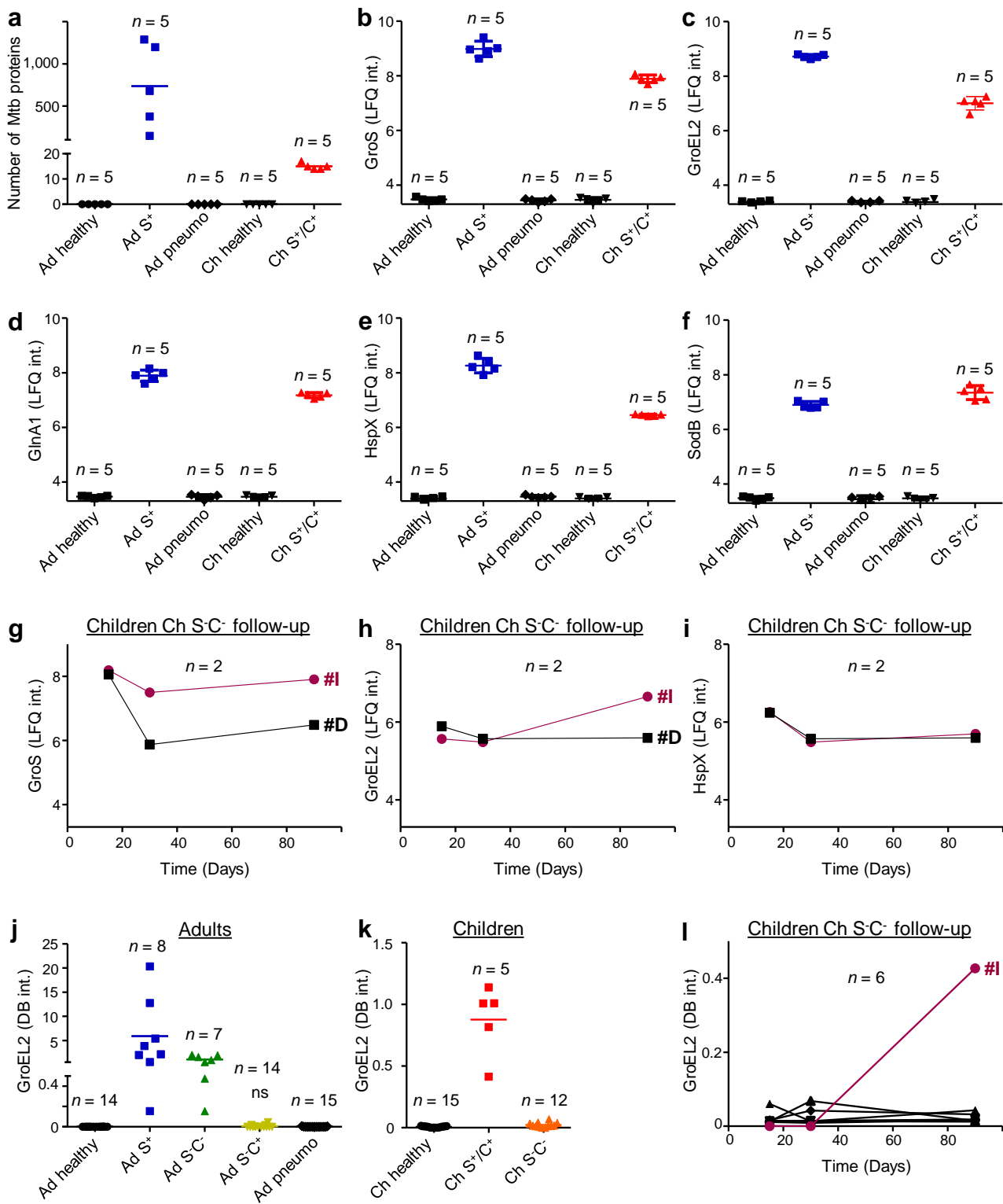


Fig. 5

Supplementary information for

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

Sergio Fabián Mosquera-Restrepo, Sophie Zuberogoitia, 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

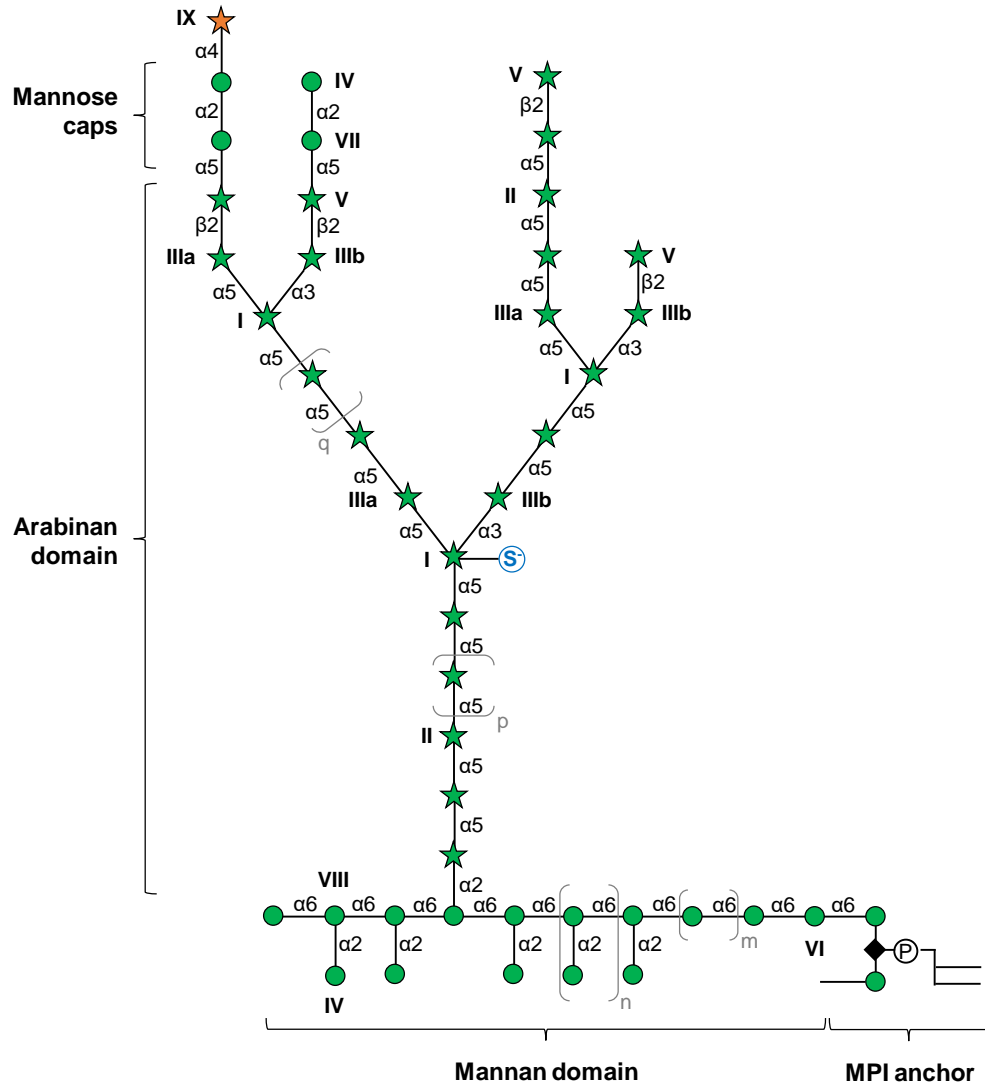
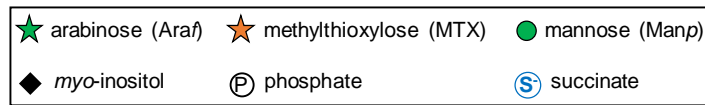
Supplementary Tables 1 to 8

Supplementary references

Legend of Supplementary Data 1 & 2

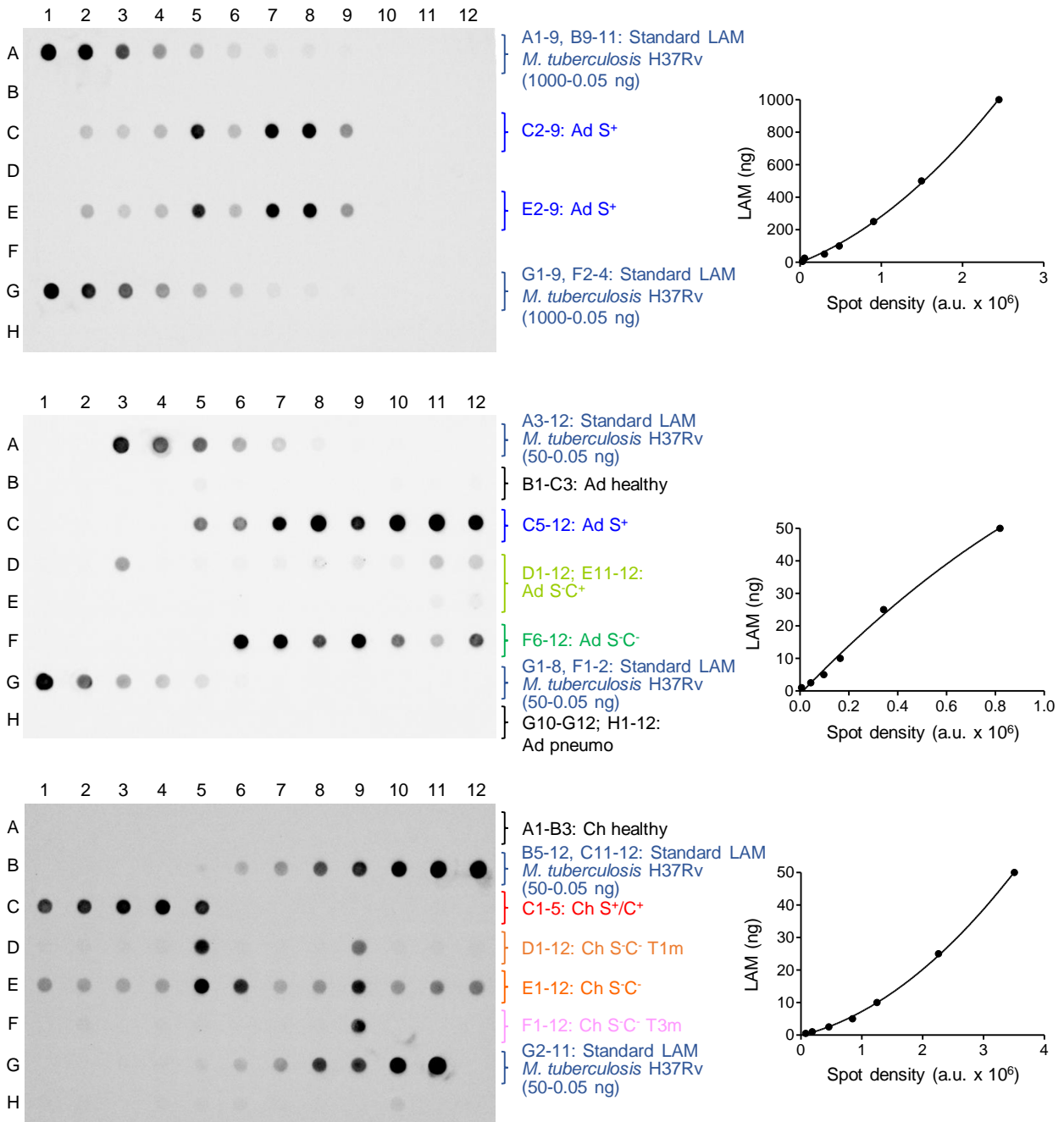
Uncropped scans of Supplementary Fig. 2

Uncropped scan of Supplementary Fig. 7



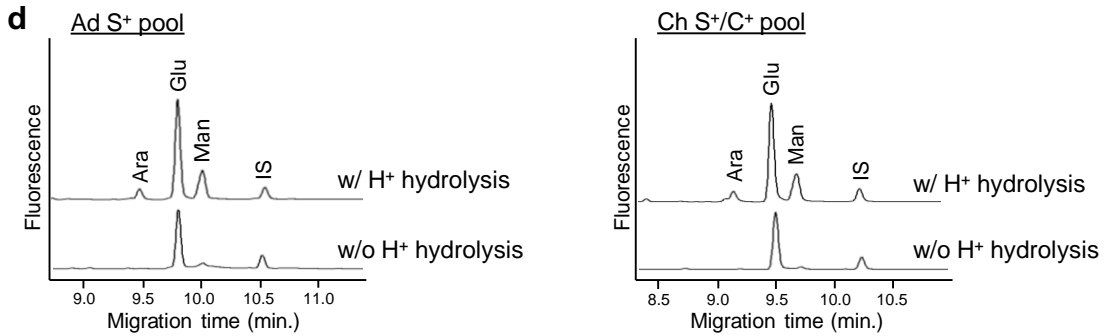
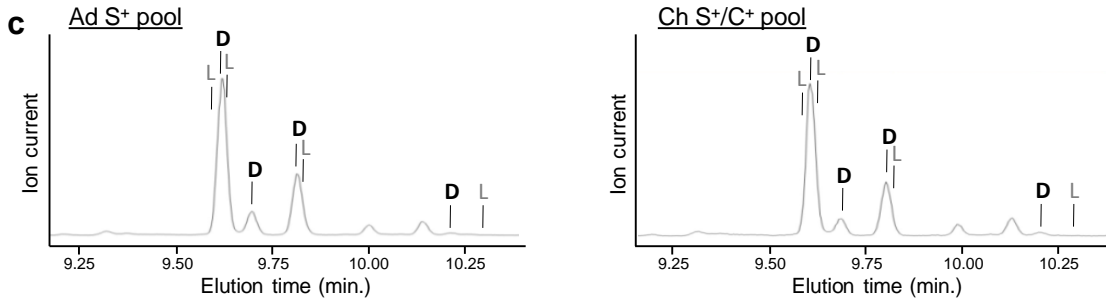
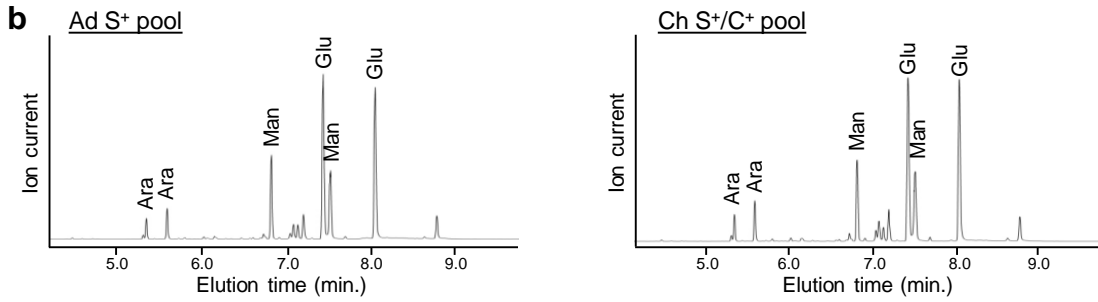
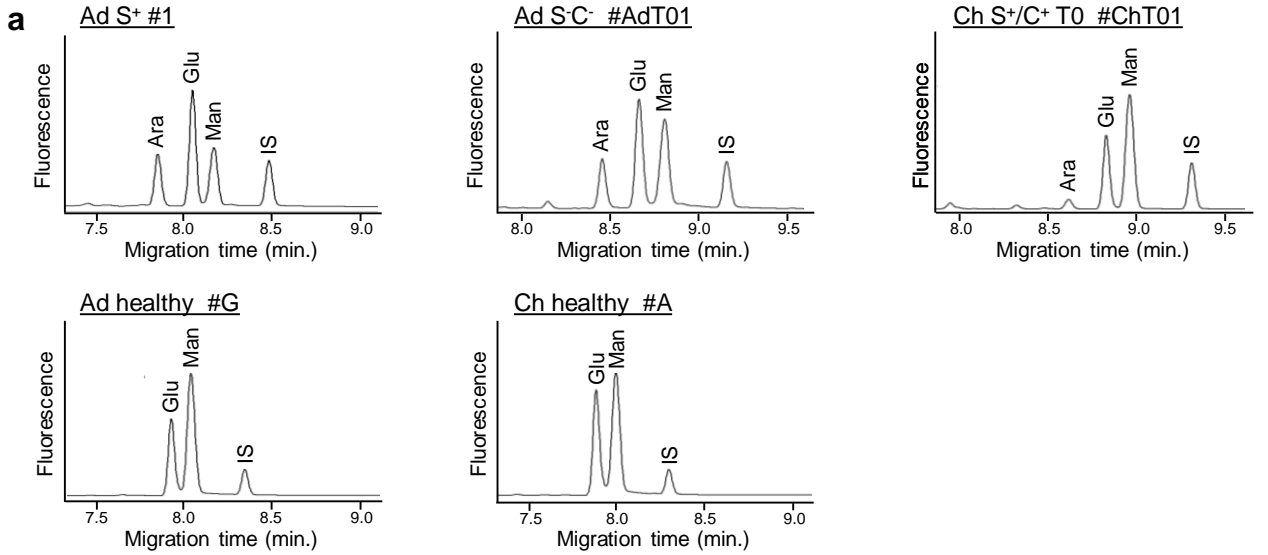
Supplementary Fig. 1. Schematic structure of LAM.

Araf, arabinofuranose; Manp, 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 determined using Image Lab™ Software (Bio-Rad). T1m, T3m, after 1 and 3 months of antibiotic treatment respectively.

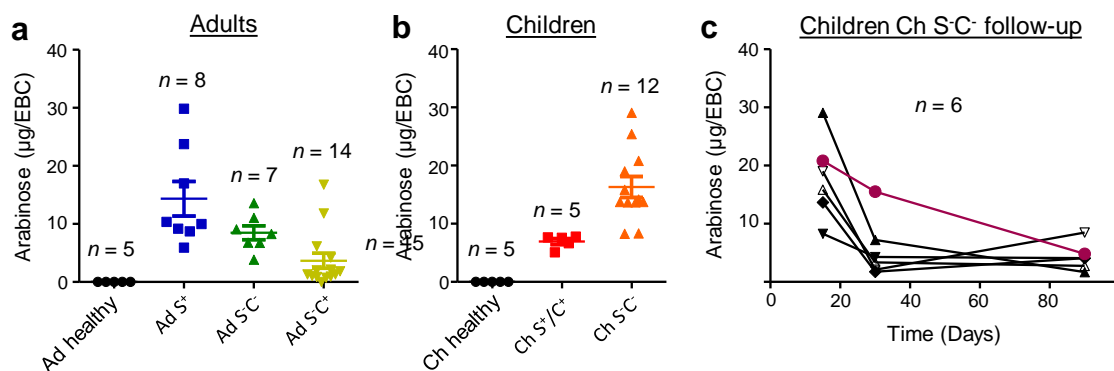


Supplementary Fig. 3. Chemical detection of arabinose and other monosaccharides in EBCs.

- 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.

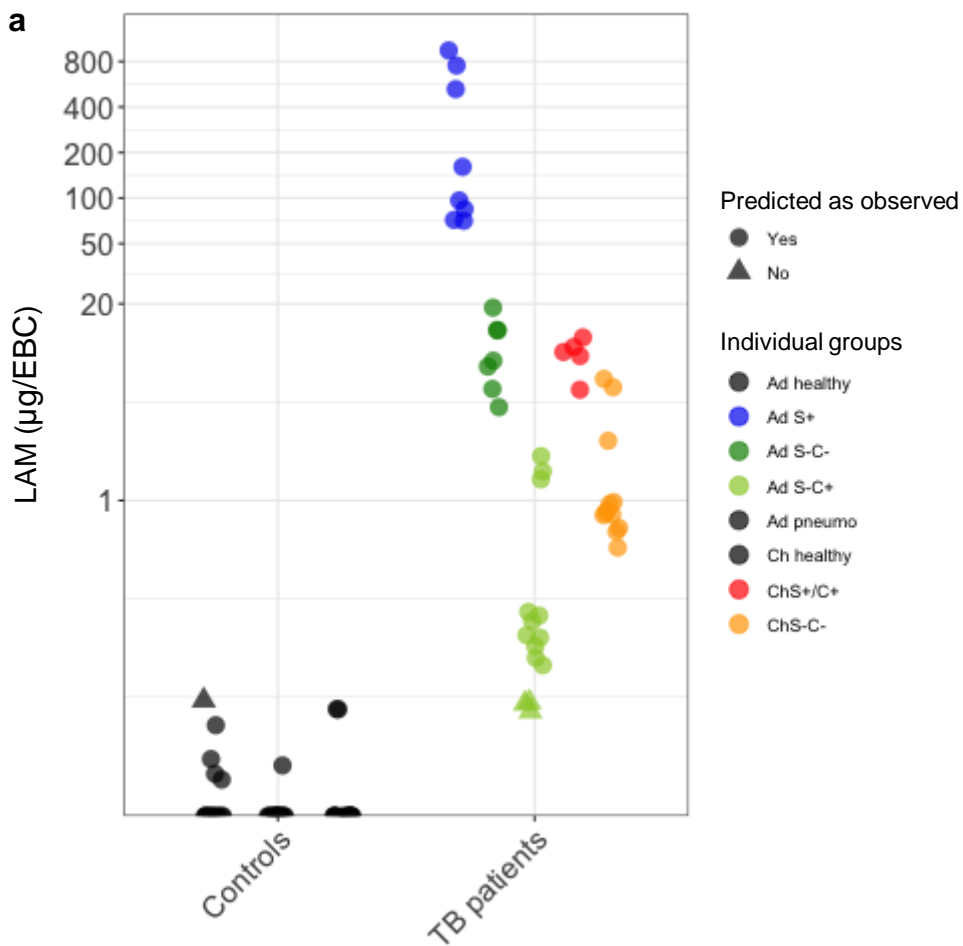
In b, c and d, analyses were performed on pooled EBCs collected from 50 adults (Ad S⁺ pool) or 50 children (Ch S⁺/C⁺ pool) TB patients (Extended Data Table 1).

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.



b

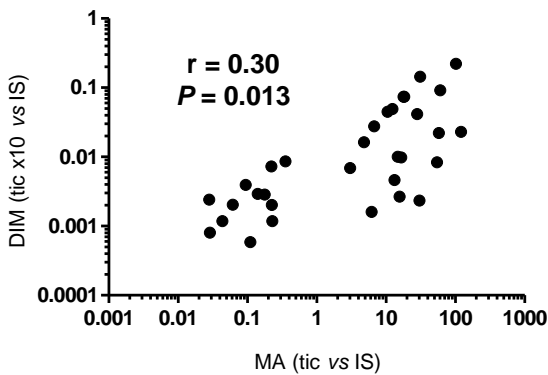
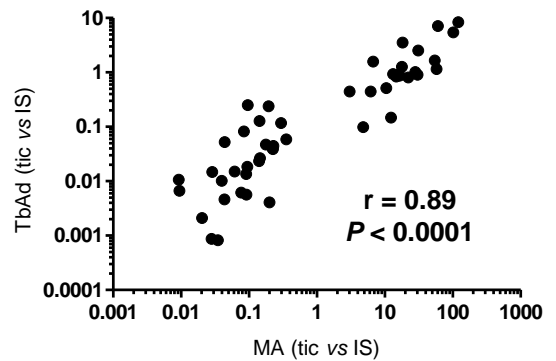
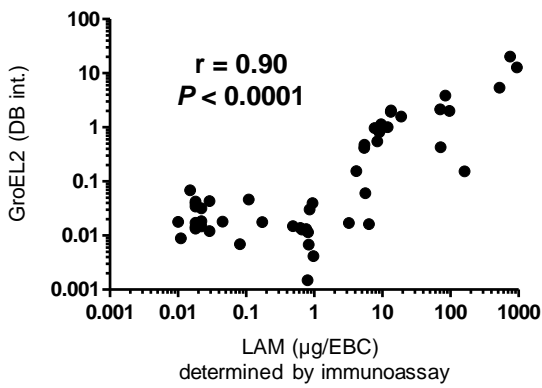
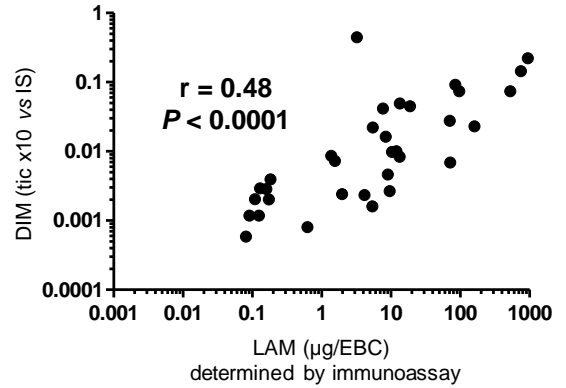
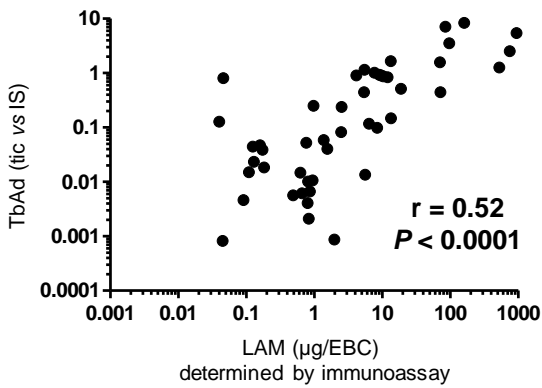
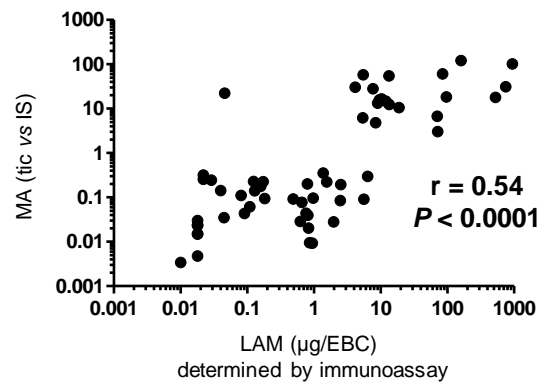
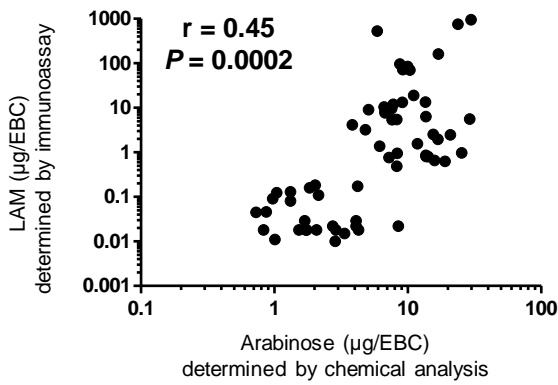
		Actual condition	
		Controls	TB patients
Prediction outcome	Controls	44	3
	TB patients	1	43

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).

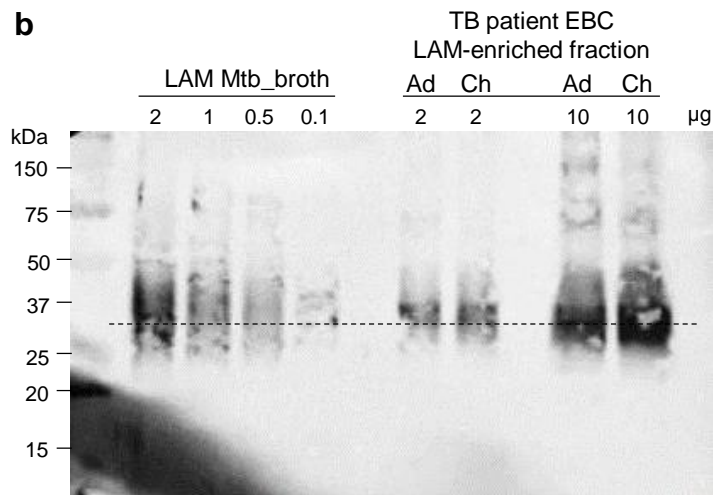
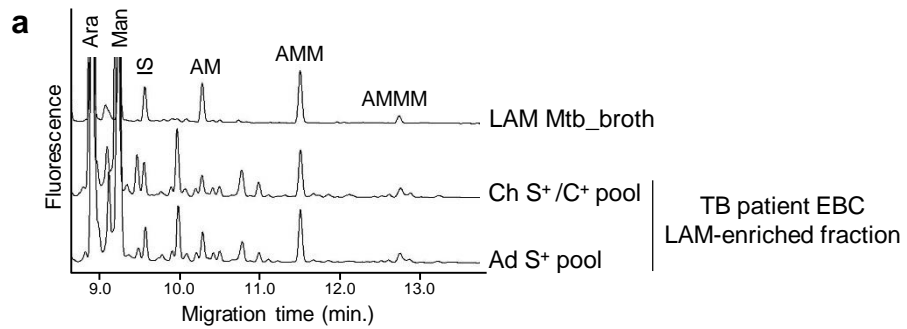
a) Quantity of LAM in EBC samples for control individuals and TB patients are shown. Symbol colors correspond to individual groups. Circles correspond to individuals whose actual condition was correctly predicted by the Leave-One-Out cross-validation (LOOCV), whereas triangles correspond to individuals 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).



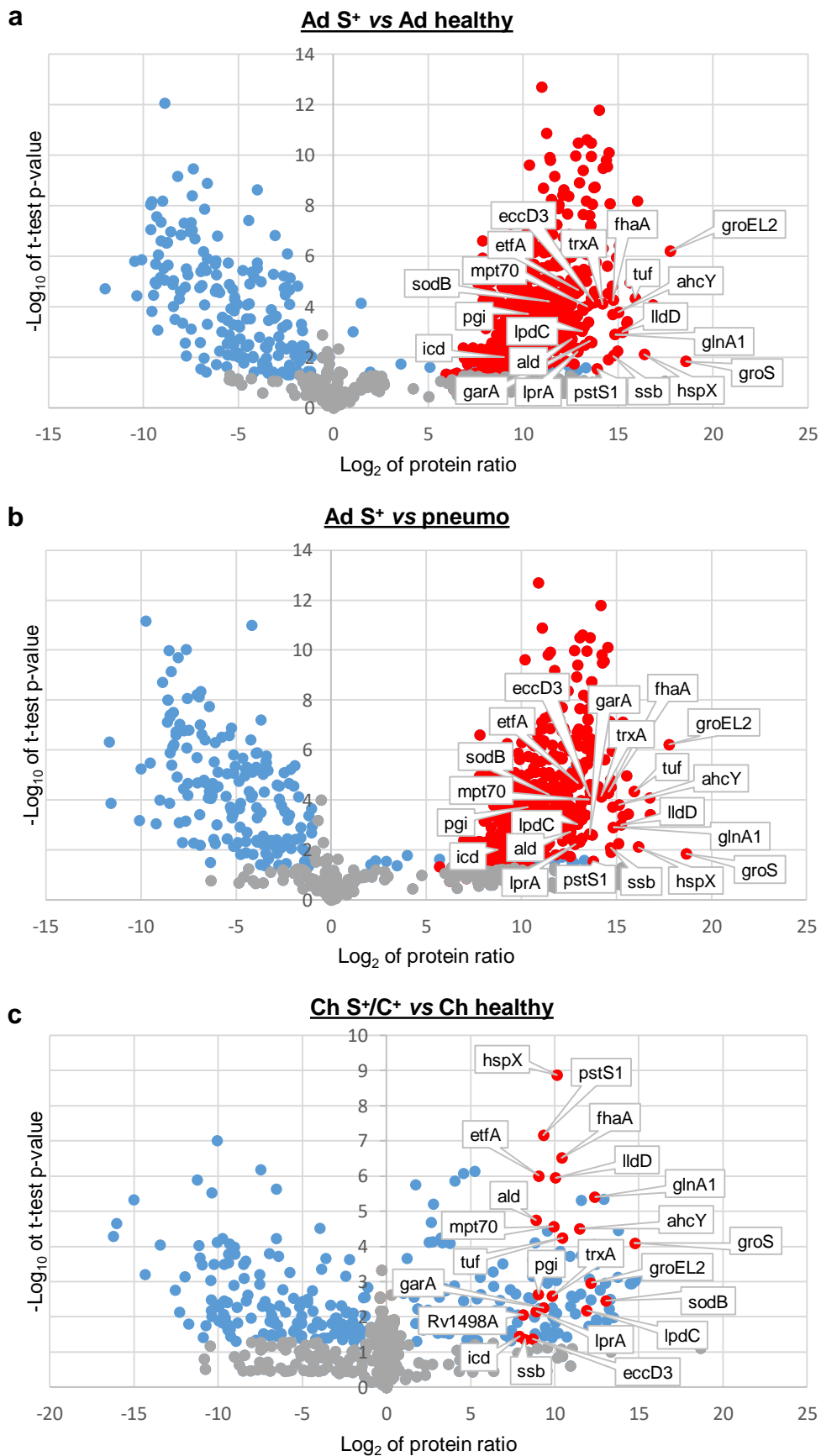
Supplementary Fig. 7. LAM in EBC bear mannose caps (a) and exhibit a lower apparent MW (b).

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 -(α 1 \rightarrow 2)-Ara-APTS; AMM, Man p -(α 1 \rightarrow 2)-Man p -(α 1 \rightarrow 2)-Ara-APTS; AMMM, Man p -(α 1 \rightarrow 2)-Man p -(α 1 \rightarrow 2)-Man p -(α 1 \rightarrow 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 μ g for Mtb_broth LAM and 2-10 μ g 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
Median	38	9
Age Minimal	15	6
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 ^1H 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. Araf, arabinofuranose; Manp, mannopyranose; MTX, methylthioxylose.

Residue		Ad S ⁺ pool	Ch S ⁺ /C ⁺ pool	LAM Mtb_broth
I, 3,5- α -Araf	^{13}C	110.2	110.2	110.1
	^1H	5.13	5.13	5.14
II, 5- α -Araf	^{13}C	109.9	109.9	109.8
	^1H	5.18	5.18	5.19
	^{13}C	110.3	110.3	110.2
	^1H	5.10	5.10	5.11
IIIa, 2- α -Araf \rightarrow 5	^{13}C	108.4	108.4	108.4
	^1H	5.19	5.19	5.21
IIIb, 2- α -Araf \rightarrow 3	^{13}C	108.2	108.2	108.1
	^1H	5.26	5.26	5.28
IV, t- α -Manp	^{13}C	104.9	104.9	104.9
	^1H	5.06	5.06	5.06
V, t- β -Araf	^{13}C	103.4	103.5	103.3/103.2
	^1H	5.16	5.16	5.18/5.30
VI, 6- α -Manp	^{13}C	102.3	102.2	102.3
	^1H	4.92	4.92	4.94
VII, 2- α -Manp	^{13}C	101.0	101.0	100.9
	^1H	5.15	5.16	5.19
VIII, 2,6- α -Manp	^{13}C	100.9	100.9	100.9
	^1H	5.13	5.13	5.17
IX, MTX	^{13}C	ND	ND	105.3/105.0
	^1H	ND	ND	5.48/5.43
Xa, t- α -Araf \rightarrow 5	^{13}C	110.6	110.6	ND
	^1H	5.04	5.04	ND
Xb, t- α -Araf \rightarrow 3	^{13}C	109.7	109.7	ND
	^1H	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 studies¹⁻³. The last column indicates the number and the fatty acyl chains (C₁₆, palmitoyl; C₁₈, stearoyl; C₁₉, tuberculostearoyl, i.e. 10-methyl-stearoyl) esterifying the PIM molecules (PIM₁, PIM₂, PIM₃, or PIM₆: 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 <i>m/z</i> of [M-H] ⁻	Calculated <i>m/z</i> 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 ₁ PIM ₂	2C ₁₆ , 1C ₁₈
1413.9	1413,9008	C ₇₂ H ₁₃₅ O ₂₄ P	15.2	Ac ₁ PIM ₂	2C ₁₆ , 1C ₁₉
1427.9	1427,9165	C ₇₃ H ₁₃₇ O ₂₄ P	3.0	Ac ₁ PIM ₂	1C ₁₆ , 2C ₁₈
1441.9	1441,9321	C ₇₄ H ₁₃₉ O ₂₄ P	3.5	Ac ₁ PIM ₂	1C ₁₆ , 1C ₁₈ , 1C ₁₉
1456.0	1455,9478	C ₇₅ H ₁₄₁ O ₂₄ P	2.2	Ac ₁ PIM ₂	1C ₁₆ , 2C ₁₉
1476.1	1476,0620	C ₈₁ H ₁₅₃ O ₂₀ P	3.0	Ac ₂ PIM ₁	3C ₁₆ , 1C ₁₈
1504.1	1504,0933	C ₈₃ H ₁₅₇ O ₂₀ P	5.1	Ac ₂ PIM ₁	2C ₁₆ , 2C ₁₈
1546.0	1546,1403	C ₈₆ H ₁₆₃ O ₂₀ P	1.6	Ac ₂ PIM ₁	1C ₁₆ , 3C ₁₈
1638.2	1638,1148	C ₈₇ H ₁₆₃ O ₂₅ P	8.1	Ac ₂ PIM ₂	3C ₁₆ , 1C ₁₈
1652.2	1652,1305	C ₈₈ H ₁₆₅ O ₂₅ P	3.3	Ac ₂ PIM ₂	3C ₁₆ , 1C ₁₉
1666.2	1666,1461	C ₈₉ H ₁₆₇ O ₂₅ P	16.2	Ac ₂ PIM ₂	2C ₁₆ , 2C ₁₈
1680.2	1680,1618	C ₉₀ H ₁₆₉ O ₂₅ P	4.3	Ac ₂ PIM ₂	2C ₁₆ , 1C ₁₈ , 1C ₁₉
1694.2	1694,1774	C ₉₁ H ₁₇₁ O ₂₅ P	5.1	Ac ₂ PIM ₂	2C ₁₆ , 2C ₁₉
1800.2	1800,1677	C ₉₃ H ₁₇₃ O ₃₀ P	1.7	Ac ₂ PIM ₃	4C ₁₆
1828.3	1828,1990	C ₉₅ H ₁₇₇ O ₃₀ P	4.1	Ac ₂ PIM ₃	2C ₁₆ , 2C ₁₈
2062.1	2062,1121	C ₉₆ H ₁₇₅ O ₄₄ P	1.2	Ac ₁ PIM ₆	2C ₁₆ , 1C ₁₉
2342.4	2342,3887	C ₁₁₅ H ₂₁₁ O ₄₅ P	0.9	Ac ₂ PIM 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 Ac₄SGL 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 hydroxyphthioceranoic acid (HPA) and 1 phthioceranoic 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 C₁₆ and the C₁₈, respectively.

Ac₄SGL from Mtb_broth show a massif of ions centered at *m/z* from 2459 to 2543^{4,5} (see below in blue), whereas Ac₄SGL 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 <i>m/z</i> of [M-H] ⁻	Calculated <i>m/z</i> 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+NH₄]⁺ 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 Mt103⁷⁻⁹, whereas PDIM in EBC show increased abundance of the forms at *m/z* 1469 and 1497 (see below in red; Fig. 3).

Measured <i>m/z</i> of [M+NH ₄] ⁺ adducts	Calculated <i>m/z</i> of [M+NH ₄] ⁺ 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 studies¹⁰.

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_broth¹⁰⁻¹².

Measured <i>m/z</i> of [M-H] ⁻	Calculated <i>m/z</i> of [M-H] ⁻	Absolute Error (ppm)	Molecular formula of M	Relative Intensity (%)	Species	Total carbon number
1108,1406	1108,1356	4,51	C ₇₆ H ₁₄₈ 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
<i>M. tuberculosis</i>					
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.

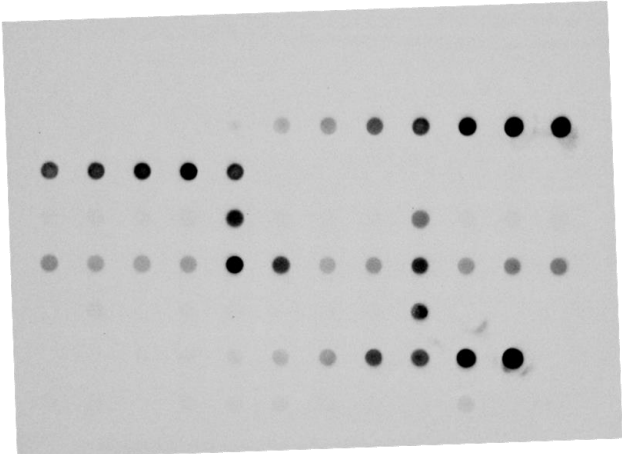
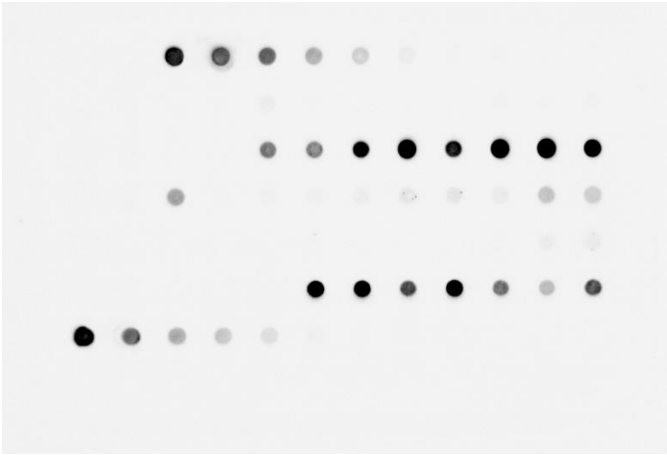
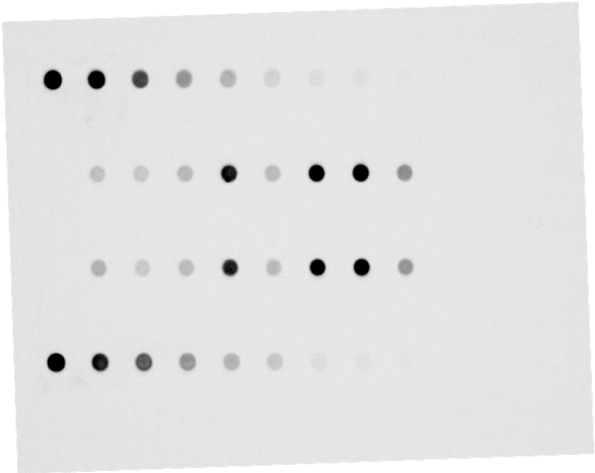
Supplementary Data 2: Abundance of proteins detected by proteomic analysis in individual EBCs.

Proteins highlighted in blue were previously described to be released in extracellular vesicles¹³⁻¹⁶. 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.

