1 **DNA methylation profiling of immune cells from tuberculosis-exposed individuals**

2 **overlaps with BCG-induced epigenetic changes and correlates with the emergence of anti-**

3 **mycobacterial 'corralling cells'**

4

5 Isabelle Pehrson[1,∂], Clara Braian[1,∂], Lovisa Karlsson[1,∂], Nina Idh[1], Eva Kristin Danielsson[1],

6 Blanka Andersson[1], Jakob Paues[1,2] Jyotirmoy Das[1,•] and Maria Lerm[1,•]*

7

8

9 [1]Division of Infection and Inflammation, [2]Division of Infectious Diseases

10 Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences,

11 Linköping University, SE-58185, Linköping, Sweden.

12

13

14

15 *∂) Shared first authorship*

16 *•) Shared last authorship*

17 **) Corresponding author:*

18

19

20 Maria Lerm, Professor in Medical Microbiology, Div. of Inflammation and Infection, Lab 1,

21 floor 12, Dept. of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences

22 Linköping University, SE-58185 Linköping, Sweden, Phone: +46-732707786, E-mail:

23 maria.lerm@liu.se

24

25

26

27 **Abstract**

28 The mechanism of protection of the only approved tuberculosis (TB) vaccine, Bacillus Calmette

29 Guérin (BCG) is poorly understood. In recent years, epigenetic modifications induced by BCG

30 have been demonstrated to reflect a state of trained immunity. The concept of trained immunity

31 is now explored as a potential prevention strategy for a variety of infections. Studies on human

32 TB immunity are dominated by those using peripheral blood as surrogate markers for immunity.

33 Here, we instead studied the lung compartment by obtaining induced sputum from subjects

34 included in a TB contact tracing. CD3- and HLA-DR-positive cells were isolated from the

35 collected sputum and DNA methylome analyses performed. Unsupervised cluster analysis

36 revealed that DNA methylomes of cells from TB-exposed individuals and controls appeared as

37 separate clusters, and the numerous genes that were differentially methylated were functionally

38 connected. The enriched pathways were strongly correlated to previously reported epigenetic

39 changes and trained immunity in immune cells exposed to the BCG vaccine in human and

40 animal studies. We further demonstrated that similar pathways were epigenetically modified in

41 human macrophages trained with BCG *in vitro*. Finally, we found evidence of an *M.*

42 *tuberculosis*-triggered emergence of a non-macrophage cell population from BCG-trained

43 macrophage cultures. These cells did not phagocytose *M. tuberculosis*, but 'corralled' the

44 bacteria into focal points, resulting in limitation of bacterial growth. Altogether, our study

45 demonstrates that similar epigenetic changes are induced by *M. tuberculosis* and BCG and

46 suggests that the modifications promote transformation of macrophages (or an unknown

47 progenitor) to establish a yet undescribed cellular defense mechanism which we term

48 'corralling', based on the metaphorical resemblance to sheepdog herding.

49

Introduction

50

51   Tuberculosis (TB) is a pulmonary infection of pandemic rank and an expansion of the current

52   toolkit for diagnosis, prevention and treatment is critical for reaching the United Nations'

53   Sustainable Development Goals for 2030 of ending the TB epidemic[1]. TB is caused by

54   *Mycobacterium tuberculosis,* which transmit via aerosols and target alveolar macrophages in

55   exposed individuals[2]. In susceptible hosts, the bacteria replicate inside the macrophages and

56   use them as trojan horses in order to disseminate in the tissues [3]. Bacillus Calmette Guérin, a

57   non-virulent derivative of *Mycobacterium bovis*, has been used for almost a century as a vaccine

58   against TB, with variable efficacy. Numerous studies have failed to provide correlates of

59   protection, leaving the vaccine mechanism elusive. In recent years, the concept of trained

60   immunity has evolved as an epigenetically encoded immune memory that can be triggered by

61   a variety of stimuli and is reflected in a reprogrammed immune state characterized by a higher

62   magnitude of response to subsequent pathogen challenges. The discovery of epigenetically

63   regulated antimicrobial defense mechanisms goes beyond the classical understanding of

64   immune defense and opens up a new field of research. Along this line, we have demonstrated

65   that administration of the BCG vaccine to healthy subjects induced profound epigenetic

66   alterations in immune cells, which correlated with enhanced anti-mycobacterial activity in

67   macrophages isolated from the vaccinees[4]. The changes were reflected in the DNA methylome,

68   with the strongest response being recorded within weeks after vaccination[4]. Our observation

69   that BCG induces alterations of the DNA methylome of immune cells has later been confirmed

70   by others[5,6]. Since BCG vaccination reflects an *in vivo* interaction between immune cells and

71   viable mycobacteria, we here hypothesized that natural exposure to *M. tuberculosis* would

72   induce similar changes not only in TB patients, but also in individuals who have been exposed

73   to TB. Analyses of DNA methylomes of immune cells isolated from lungs and peripheral blood

74   allowed us to identify distinct DNA methylation (DNAm) signatures in TB-exposed

75   individuals. Pathway analyses revealed strong overlaps with previous studies on BCG-induced

76   epigenetic signatures that could be correlated with protection against *M. tuberculosis*. We

77   therefore set up an experiment to scrutinize the anti-mycobacterial mechanisms correlating with

78   BCG-induced epigenetic reprogramming in human primary macrophages. BCG-trained

79   macrophages were subjected to DNA methylome analysis and in parallel monitored in a live

80   cell imaging system during exposure to *M. tuberculosis*. Again, the DNAm change *in vitro*

81   highly overlapped with the signatures of the TB-exposed individuals and the previously

82   identified BCG-induced epigenetic changes. Guided by these findings, we designed

83   experiments to monitor the events taking place in BCG-trained cells challenged with *M.*

84   *tuberculosis*.   To our surprise, rapidly dividing, non-phagocytic cells with a size of

85   approximately 1/4th of the diameter of macrophages appeared in BCG-trained, *M. tuberculosis-*

86   infected macrophage cultures. These 'corralling cells', whose nature remains elusive,

87   apparently concentrated the bacteria into focal points by pushing them in a coordinated fashion,

88   while limiting their growth. To our knowledge, this phenomenon has not been described

89   previously.

90

91   **Results**

92   *Study design*

93   To determine epigenetic changes in the immune cells in TB-exposed individuals, we recruited

94   subjects enrolled in a routinely performed TB contact tracing at Linköping University Hospital,

95   Sweden. Age-matched individuals were included as controls (Table 1). The index case was

96   diagnosed with drug-sensitive pulmonary TB and had completed two out of six months of

97   standard treatment at the time of sample collection. All included subjects except one (a TB

98   contact) had been BCG-vaccinated previous to the study (Table 1). Interferon-Gamma Release

99   Assay (IGRA) status was determined and among the exposed individuals, two were positive

100 (including the index case) and among the controls, one individual (C2) was classified as

101 'borderline'- positive[7] (Table 1). From induced sputum, we used an established protocol for the

102 isolation of HLA-DR-positive antigen-presenting cells, dominated by macrophages[8] and T

103 cells (CD3 positive cells) (BioRXiv, 2021, pre-print, accepted for publication in *Epigenetics*),

104 whereas the PBMC fraction extracted from blood was kept as a mixed population. For the in

105 vitro experiment with BCG and *M. tuberculosis* exposure, human primary monocytes were

106 collected from blood samples from healthy blood donors and differentiated into macrophages.

107 (Fig. 1).

108

109 *DNA methylome data from TB-exposed individuals form a separate cluster*

110 DNA isolation from the studied cell populations was followed by global DNAm analysis using

111 the Illumina 450K protocol. After curation of the data[9], the datasets were subjected to

112 unsupervised hierarchical cluster analysis based on DNA CpG methylation β-values. This

113 approach accurately clustered the participants into TB-exposed and controls based on the DNA

114 methylome data derived from both HLA-DR- and CD3-positive cell populations. (Fig. 2a, b).

115 On the other hand, in the PBMC-derived dataset, the TB index case appeared outside the

116 clusters and two of the controls clustered with the other exposed individuals, one of them

117 ("Con_2") being the individual identified as borderline-positive in the IGRA test (Fig. 2c).

118

119 Next, we identified the differentially methylated CpG sites (DMCs) and differentially

120 methylated genes (DMGs) by comparing the TB-exposed and control groups for each cell

121 population. To filter out the most significantly altered DMGs in the dataset, the stringency

122 criteria of $log_2>|0.3|$ fold increased or decreased β-values and Benjamini-Hochberg (BH)-

123 corrected *p*-value <0.05 (HLA-DR), <0.1 (CD3) and <0.2 (PBMC) were applied. The results

124 are depicted as volcano plots, which show that DNA methylomes of TB-exposed most strongly

125   differ in the HLA-DR cells as compared to control subjects, followed by the CD3 population,

126   whereas PBMC datasets reveled fewer DMGs (Fig. 2d, e, f, Table 2). To highlight the locus

127   position of the DMGs, chromosome maps were constructed (Suppl. Fig. S1a-c). Using the same

128   stringency criteria as for the HLA-DR analysis, we tested whether DMGs would emerge when

129   the datasets were arranged in other possible groups as derived from the demographics (>/<

130   median age, sex, IGRA status). Neither age nor IGRA status generated any significant DMGs

131   with these settings, and gender rendered three DMGs (X and Y chromosomes were removed in

132   the initial filtering prior the analysis) in the HLA-DR positive and in the CD3 positive cells

133   (Table 2).

134

135   *Functional enrichment analysis reveals common and unique interactomes in the datasets*

136   Using the PANTHER Database, we investigated whether the identified DMGs were enriched

137   in known pathways (Fig. 3a,b,c). The analysis revealed enrichment in pathways with relevance

138   for TB infection, including hypoxia-inducible factor (HIF)1-α activation, Vitamin D

139   metabolism and p38, Wnt, Notch, interleukin, chemokine, and cytokine signaling pathways[10–

140   17]. Common pathways shared between at least two of the cell populations included B cell

141   activation, glycolysis, angiotensin II signaling, and cholecystokinin signaling. Notably, several

142   pathways named after their known functions in the nervous system were enriched in the studied

143   cell populations, including pathways involved in axon guidance and adrenaline, acetylcholine

144   and glutamate signaling. In the PBMC population but not in the lung cell populations, the

145   interferon-γ signaling pathway was identified among the enriched pathways.

146

147   *Comparisons across cell populations and species reveals the existence of a common DNA*

148   *methylome-based biosignature in mycobacteria-exposed immune cells*

149   Given the fact that the interaction between mycobacteria and eukaryotes is evolutionary ancient,

150   we predicted that highly conserved pathways exist that are common among the studied cell

151   populations. Comparing the identified DMGs from the HLA-DR, CD3 and PBMCs in a Venn

152   analysis, we discovered 185 common DMGs (Fig. 4a). We expanded the Venn analyses to

153   include data from our previous work on BCG vaccine-induced DMGs that correlated with

154   enhanced mycobacterial control[4], as natural exposure to TB and BCG vaccination both

155   represent *in vivo* encounters between mycobacteria and host immune cells. Even though the

156   routes of mycobacterial exposure differ profoundly in these settings, a set of 151 DMGs could

157   be identified as overlapping between our previous BCG study and all cell populations studied

158   here (Fig. 4b), suggesting that a highly conserved epigenetic response to mycobacterial

159   challenge exists.

160   In 2018, Hasso-Agopsowicz *et al*. described alterations in DNAm patterns in PBMCs from

161   BCG-vaccinated individuals, with concomitant enrichment in many immune-related pathways[5].

162   In order to compare that study with ours, we performed PANTHER analysis with the 185

163   common DMGs and matched the identified enriched pathways, revealing that 75% of those

164   pathways were the same as in the present study (Fig. 5a), further corroborating the relationship

165   of the altered DNAm patterns induced through TB exposure and BCG-induced changes. In a

166   recent study by Kaufmann *et al*., BCG-induced alterations of the epigenome in mice was

167   correlated with protection against *M. tuberculosis* infection[6]. In order to translate our human

168   DNA methylome signature to the signature identified in the mouse study, we searched for

169   pathway overlaps between the two we performed a Gene Ontology (GO) enrichment analysis

170   (Suppl. Fig. S2a-c). Figure 5b demonstrates that for our PBMC data, the GO terms "biological

171   processes" overlapped 100% with the mouse study (same cell population) and to 31% and 65%

172   for HLA-DR- and CD3-positive cells, respectively. In 2014, Saeed *et al* [19] demonstrated the

173   induction of trained immunity pathways by another immune-training agent, β-glucan. We

174   assessed possible pathway overlap with that study and although there were fewer overlaps as

175   compared to the BCG-induced pathways described above. Again, the strongest correlation was

176   found in the PBMC fraction, in this case in the GO terms "cellular components" (Fig. 5c and

177   Table 1c).

178   To validate how well the 284 CpG sites corresponding to the 185 overlapping DMGs performed

179   in an unsupervised cluster analysis, we included one additional TB patient and two contacts,

180   and collected HLA-DR-positive cells from induced sputum, as the DNA methylome data from

181   this cell type was clearly outperforming the others with respect to accurate separation of the

182   groups. Fig. 6 shows a *k* means-based dendrogram with a heatmap of the β values of the 284

183   CpG sites from both previous and new subjects' DNAm data, revealing a distinct separation of

184   the subjects in accordance with TB exposure.

185

186   *In vitro BCG training of macrophages induced DNAm changes corresponding to exposure to*

187   *TB*

188   To investigate whether the BCG-induced epigenetic changes can be mimicked *in vitro*, we set

189   up a BCG training experiment with macrophages isolated from donor blood that were trained

190   with BCG and subsequently infected with the virulent *M. tuberculosis* strain H37Rv (expressing

191   GFP) and monitored during the course of infection. In a subset of donors, BCG-trained cells

192   displayed an increased capacity to kill *M. tuberculosis* (Fig. 7a-b) while cell viability was

193   retained (Fig. 7c-d). DNAm analyses was performed on DNA isolated from these donors'

194   macrophages 24 hours after BCG training. We identified 7471 DMGs with the stringency

195   criteria of DNAm difference > 30% and BH corrected *p*-value < 0.01. A PANTHER pathway

196   analysis based on the identified DMGs showed significant enrichment in the Wnt signaling

197   pathway, cadherin signaling pathways and angiogenesis, overlapping with the over-represented

198   pathways in the HLA-DR+ and CD3+ cells from the TB-exposed individuals (Fig. 7e).

199

BCG training of primary human macrophages induces morphological changes during M. tuberculosis *infection*

To better understand the mechanism by which BCG enhances the anti-mycobacterial activity of macrophages, we included N-glycolylated muramyldipeptide (MPD), alongside BCG in the training experiment with macrophages from ten donors. MDP is a component of the mycobacterial cell wall and has been described as the minimal component of BCG to induce trained immunity[20,21] . The infected macrophages were monitored in the IncuCyte® S3 Live Cell Imaging system. We observed the emergence and expansion of a non-phagocytic cell type in BCG-trained macrophages from three donors (Fig. 8a-f and supplemental video S1), which seemed unaffected by high bacterial load. We termed this yet undefined cell type 'corralling cells', based on their apparent coordinated action to relocate and contain bacteria into focal points. We observed that distinct morphological changes occurred prior the emergence of the corralling cells and classified this process into 6 distinct stages shown in figure 8a-f. In staining experiments using Oil Red O, a lipid stain, in paraformaldehyde-fixed cells, we could demonstrate that the vacuole-like structures were not lipid droplets, excluding the possibility that this morphological state was reflective of foamy macrophages described previously [22] (Fig. 8g-h). The fixed cells were also stained with fluorescently conjugated antibodies (anti- α-smooth muscle actin (α-SMA), CD68, CD3, CD14, Collagen I and HLA-DR) to identify potential expression markers on the corralling cells. The HLA-DR antibody positively stained the macrophages, whereas the corralling cells were negative. (Fig 8i-j). No positive staining was observed for α-SMA, CD3, CD14 or collagen I (data not shown). Quantification of the number of corralling cells revealed that BCG-training cells but not MDP-training significantly induced the emergence of these cells (Fig. 9a). Finally, we evaluated the outcome for *M. tuberculosis* growth in wells with corralling cells. Figure 9b and supplemental video S2a

224   (corralling) and S2b (non-corralling) show that in wells with corralling cells, the GFP-

225   fluorescence of *M. tuberculosis* was strongly suppressed.

226

227   **Discussion**

228   In this study, we present data suggesting that exposure to TB generates a distinct DNAm

229   signature in pulmonary immune cells. The signature was found not only in those with active or

230   latent TB infection, but also in individuals who are exposed but IGRA-negative. The finding

231   that healthy, TB-exposed individuals also carry the signature opens the possibility that the

232   epigenetic alterations reflect a host-beneficial reprogramming of the immune mechanisms

233   rather than being induced by *M. tuberculosis* as a step to evade the immune defense. This notion

234   is supported by our observation that the DMGs identified in the present study strongly

235   overlapped with the epigenetic alterations identified in the *in vitro* BCG-trained macrophages,

236   and the previously reported DNAm changes induced during BCG vaccination, which correlated

237   with increased anti-mycobacterial capacity of macrophages[4]. In addition, we demonstrate that

238   the GO data derived from our dataset display a strong overlap with data from a study on

239   protective BCG vaccination in mice[6].

240   BCG vaccination has convincingly been shown to induce heterologous immunity protecting

241   against childhood mortality from other causes than TB[23,24]. Based on our finding that natural

242   TB exposure and BCG vaccination trigger similar epigenetic changes we propose the

243   hypothesis that a "beneficial exposure" to TB exists, which protects against other infections

244   through heterologous immunity. Along the same line, it has been shown that a substantial

245   fraction of individuals exposed to TB can be defined as 'early clearers', since they remain

246   tuberculin skin test or IGRA negative[25], suggesting effective eradication of the infection[25].

247   Identifying these early clearers and understanding the biology behind their resistance to TB

248   infection could move the field forward towards novel strategies of TB prevention.

249   In concordance with the fact that macrophages constitute the main niche for mycobacterial

250   replication, the strongest enrichment of DNAm changes was observed in the HLA-DR-positive

251   cell population, which is dominated by alveolar macrophages. The pathways identified to be

252   enriched in the HLA-DR-positive population have been described in the context of trained

253   immunity, BCG exposure and TB. For example, activation of Hypoxia-Inducible Factor 1 α

254   and glycolysis pathways (P00030 and P00024, respectively) are hallmarks of macrophages that

255   have undergone the epigenetic changes reflective of trained immunity (reviewed in [26,27]), which

256   is induced in myeloid cells upon BCG-stimulation[21]. VEGF-release (P00056) by macrophages

257   has been shown to recruit immune cells during granuloma formation[29]. Further, vitamin D has

258   been shown to strengthen the anti-mycobacterial activity of macrophages[11,30], and upregulation

259   of components of the vitamin D pathway is linked to the production of anti-microbial peptides[12],

260   providing a possible effector mechanism for mycobacterial control. Recent literature on

261   immune regulation through T cell-derived acetylcholine[31,32] attributes relevance to the

262   acetylcholine receptor pathway identified among the HLA-DR pathways.

263   Although macrophages and lymphocytes are not generally viewed as having many similarities,

264   we found 34 of the identified pathways to overlap between HLA-DR- and CD3-positive cells.

265   In data derived from the CD3 and PBMC populations, both of which represent lymphocytes,

266   overlaps were identified for glycolysis, glutamate receptor and angiotensin II pathways.

267   Interestingly, a metabolic shift towards increased glycolysis, representative of the Warburg

268   effect, has been strongly associated with trained immunity[26]. However, the literature is

269   dominated by the view that this event takes place in trained myeloid cells, while we identified

270   this circuit in CD3 cells (lymphocytes) and not in the HLA-DR cells (dominated by

271   macrophages). The glutamate receptor is widely expressed on immune cells and has been

272   described as having an important regulatory role in T cells, which can also produce and release

273   glutamate[33]. The role for angiotensin II pathway in TB remains elusive, while Angiotensin II

274   Converting Enzyme 2 is currently in the spotlight due to fact that the SARS-CoV2 virus utilizes

275   it as a receptor for entry into host cells[35]. In the PBMC population, which over all showed a

276   weaker epigenetic response, we found the interferon-γ signaling pathway, which has a well-

277   established role in anti-mycobacterial defense (reviewed in [36]), among the reprogramed

278   pathways.

279   Several studies have ascribed Wnt pathways immunomodulating functions and induction

280   during *M. tuberculosis* infection (reviewed in [14]) and M1 and M2 macrophages express distinct

281   patterns of Wnt ligands. Here, in experiments with BCG-trained macrophage cultures, we

282   observed the emergence of rapidly dividing cells with a size of approximately 1/4th of the

283   diameter of macrophages and negative for HLA-DR staining. A major limitation of the present

284   study is that the corralling cells' source of origin and mechanism for emergence could not be

285   identified, but our ongoing studies are designed to characterize these cells. Taken together, we

286   present data supportive of mycobacteria exposure induced DNAm changes that correlate with

287   previous findings from studies on BCG vaccination including TB protection and trained

288   immunity. We propose that the mycobacteria-specific DNAm changes promote the emergence

289   of corralling cells to restrict mycobacterial growth, although this hypothesis needs further

290   investigation.

291

292   **Methods**

293   *Study design and participants*

294   Patients with pulmonary TB, healthy participants with a history of TB-exposure and healthy

295   controls, with an age ranging from 18 to 53 years, were enrolled at Linköping University

296   Hospital and Linköping University, respectively. Included subjects (see Table 1 for

297   demographics) donated peripheral blood and induced sputum samples[37] (BioRxiv, pre-print,

298   2021, accepted for publication in *Epigenetics*) following oral and written informed consent

299   (ethical approval obtained from the regional ethical review board in Linköping, #2016/237-31).

300   The study protocol included questionnaires on respiratory and overall health, the evaluation of

301   IGRA-status and sputum samples for DNA extraction. The subjects' samples and

302   questionnaires were not linked to any personal information at any stage of the study. For *in*

303   *vitro* experiments, de-identified buffy coats were purchased from the blood facility at Linköping

304   University Hospital. The buffy coats were obtained from healthy blood donors, who gave

305   written consent of research use for the blood products.

306

307   *Induced sputum and pulmonary immune cell isolation*

308   Induced sputum is a well-tolerated, non-invasive method to collect cells from the surface of the

309   bronchial airways after inhalation of a hypertonic saline solution. The procedure of sputum

310   induction takes approximately 30 minutes and is both cost effective and safe with minimal

311   clinical risks[38]. Sputum specimens were collected as described by Alexis *et al* [39], with the

312   following modifications: premedication with an adrenergic β2-agonist, salbutamol (Ventoline,

313   1ml 1mg/ml) was administrated before the inhalation of hypertonic saline, using a nebulizer

314   (eFlow, PARI). The subsequent steps of sputum processing were adopted from Alexis et al.

315   (2005)[40] and Sikkeland *et al*[8]. The HLA-DR and CD3-positive cells were isolated using

316   superparamagnetic beads coupled with anti-human CD3 and Pan Mouse IgG antibodies and

317   HLA-DR/human MHC class II antibodies (Invitrogen Dynabeads, ThermoFisher, cat no 11041

318   and 14-9956-82, respectively). An initial positive selection was done with CD3 beads followed

319   by a positive HLA-DR selection. Bead-coating and cell isolation were performed according to

320   manufacturer's protocol.

321

322   *PBMC isolation*

323 PBMCs were isolated from whole blood (TB-exposed and controls) or from leukocyte rich

324 fractions of blood obtained from healthy volunteers (Linköping University Hospital blood bank,

325 Linköping). Isolation was performed by the method of density gradient centrifugation using

326 Lymphoprep (Axis-Shield) and Sepmate-50 tubes (Stemcell Technologies) according to

327 manufacturer's protocol. IGRA status was determined on the whole-blood samples using

328 QuantiFERON-TB Gold (Cellestis) following the manufacturer's instructions.

329

330 *Cell culture and in vitro BCG training*

331 Following PBMC isolation cells were seeded in cell culture treated flasks in Dulbecco's

332 Modified Eagle's Medium (DMEM, Invitrogen) with 25 mM hepes, 100 U/ml penicillin and

333 100 μg/ml streptomycin (PEST, Gibco). Cells were incubated in 37 °C for 2 h before the non-

334 adherent lymphocytes were washed away using warm Krebs-Ringer Glucose buffer (made in

335 house). Complete DMEM supplemented with 10% pooled human serum, 2mM L-glutamine,

336 100 U/ml penicillin and 100 μg/ml streptomycin (all from Gibco) was then added to the cells

337 along with immune training agents, or medium only as negative control for 24h. For the DNA

338 methylation analysis and data presented in figure 7 a-d we used 10 μg/ml of BCG (freeze dried

339 *M. bovis bacillus Calmette-Guérin* Danish Strain). For the 10 donor experiment including MDP

340 and data presented in figures 8 and 9 we instead used 3.33 μg/ml BCG or 3.33 μg/ml N-

341 glycolylated muramyldipeptid (MDP) (Invivogen). Training agents were washed off and the

342 cells were incubated for 6-7 days in complete medium with media change every 2-3 days. The

343 cells were washed with PBS followed by trypsinization and reseeding of 5000 cells/ well in a

344 384-well plate (Corning Falcon) for infection experiments. DNAm analysis DNA was isolated

345 at day 6 after training with BCG.

346

347   *M. tuberculosis culture and infection of macrophages*

348   The laboratory *M. tuberculosis* strain, H37Rv, carrying the green fluorescent protein (GFP)-

349   encoding pFPV2 plasmid (*M. tuberculosis*-GFP) was grown in Middlebrook 7H9 broth

350   supplemented with albumin-dextrose-catalase (ADC, Becton-Dickinson), 0.05% Tween-80,

351   and the selection antibiotic Kanamycin at 20µg/ml for two to three weeks at 37°C. The bacteria

352   were then re-seeded in fresh broth for an additional 7 days to reach early log phase for use in

353   experiments. For infection, the harvested bacteria were washed, resuspended in antibiotic-free

354   complete DMEM and filtered through a Milliex-®SV 5.0µm syringe filter (Merck) to remove

355   bacterial clumps. Macrophages were infected with *M. tuberculosis* at a multiplicity of infection

356   (MOI) 0.5 and the cells imaged using an IncuCyte® S3 (IncuCyte® Live-Cell 120 Analysis

357   System, Sartorius). The relative fluorescence signal (RFU) of bacteria was measured every 4

358   hours (20x, 4 images/well) until day 7 post-infection, analyzed using IncuCyte® S3 software

359   and expressed as total integrated intensity of green objects (GCUxµm^2/image).

360

361   *Identification of morphological changes and quantification of corralling*

362   Manual identification of morphological stages was performed by importing 8-bit grayscale

363   images from the IncuCyte® S3 software to ImageJ2[41] using the Fiji package[42]. All images were

364   processed to enhance contrast by 2% for ease of viewing. The number of coralling cells was

365   manually estimated in images displayed in the IncuCyte® S3 software. Bacterial area was

366   determined using the basic analyzer tool in the IncuCyte® S3 software.

367

368   *Immunohistochemistry for characterization of corralling cells*

369   The 384-well culturing plates that were used to monitor the cells in the IncyCyte® were fixed

370   on day 10-11 from the infection. The plates were washed 2x in PBS before adding 4%

371   paraformaldehyde in PBS for 30 minutes. The plate was washed again x2 in PBS and stored in

372   PBS in 4 degrees until use. For the immunohistochemistry, the wells stained with intracellular

373   markers (α-SMA) were permeabilized with 50 ul PBS (made in house) containing 1% bovine

374   serum albumin (BSA) (Sigma-Aldrich, St. Luis, Missouri, U.S) and 0.2% Triton X-100 (Sigma

375   Aldrich)/well for 10 min. No permeabilization was used for the staining of extracellular markers

376   (CD3, CD14, CD68, HLA-DR, Collagen α2 Type I). The wells were washed 3 times with PBS

377   and then 50 ul blocking buffer containing 1% BSA in PBS was added and the plate was

378   incubated at RT for 1 h. The antibodies were diluted in PBS 1% BSA and added in a volume

379   of 10 ul/well. The antibodies used were CD68 Alexa Fluor 647 diluted 1:25 (Cat no: sc-17832,

380   Santa Cruz Biotechnology, Dallas, Texas, U.S), CD68 Alexa Fluor 647 diluted 1:20 (Cat no:

381   562111, BD biosciences, Franklin Lakes, New Jersey, U.S), α-SMA Alexa Fluor 488 diluted

382   1:50 (Cat no: sc-32251, Santa Cruz Biotechnology), CD3 FITC diluted 1:50 (Cat no: 349201,

383   BD Biosciences), CD3 FITC diluted 1:50 (Cat no: 300306, Biolegend, San Diego, California,

384   U.S), HLA-DR PE diluted 1:20 (Cat no: 327007, Biolegend), CD14 FITC diluted 1:20 (Cat no:

385   345784, BD biosciences), Collagen α2 Type I Alexa Fluor 647 diluted 1:50 (Cat no: 393573,

386   Santa Cruz Biotechnology). For isotype controls we used normal mouse isotype control IgG_1

387   Alexa Fluor 647 (Cat no: sc-24636, Santa Cruz Biotechnology) diluted 1:50, normal mouse

388   isotype control IgG_{2a} Alexa Fluor 488 (Cat no: sc-3891, Santa Cruz Biotechnology) diluted

389   1:50) and normal mouse isotype control IgG_{2b} FITC (Cat no: 0090480, BD Biosciences). The

390   plate was incubated for 1 h in RT. The pate was 3x in PBS and 20 ul Dako Mounting Medium

391   (Agilent, Santa Clara, California, U.S) was added. The plate was read in the IncuCyte® S3 Live

392   Cell Imaging system at x20.

393

394   *DNAm sequencing and data analysis*

395   DNA from PBMCs, HLA-DR-positive, CD3-positive cells, and the cultured human primary

396   macrophages was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany)

397   according to the manufacturer's instructions. Genome-wide DNAm analysis was performed

398   using the HumanMethylation450K BeadChip (Illumina, USA) array (for the HLA-DR, CD3,

399   PBMC samples) and reduced representation bisulfite sequencing (Diagenode′s RRBS) read in

400   Illumina's HiSeq 2000 (for the cultured human primary macrophages) at the Bioinformatics

401   and Expression Analysis (BEA) Core Facility at Karolinska Institute, Stockholm. The

402   methylation profiles from the HumanMethylation450K BeadChip analysis for each cell type

403   were analyzed from the raw IDAT files in R (v4.0.2) using the *minfi* (v1.36.0) with subset-

404   quantile within array (SWAN) normalization[43,44] and *ChAMP* (v2.19.3) with beta-mixture

405   quantile normalization (BMIQ) packages[45,46]. The type I and type II probes were normalized

406   using the quantile normalization method. Using the default setup of the *ChAMP* package,

407   following probes were filtered out: i) probes below the detection *p*-value (>0.01), ii) non-CpG

408   probes, iii) multi-hit probes, and iv) all probes of X and Y chromosomes. Cell type

409   heterogeneity was corrected for the PBMC cell types using the Houseman algorithm[47] and batch

410   effects were fixed using *ComBat* from the *SVA* package (v3.38.0). Differential methylation

411   analysis was performed with the linear modeling (lmFit) using the *limma* package[48] (v3.46.0)

412   in a contrast matrix of the TB-exposed and TB-non-exposed (Control) individuals. All

413   Differentially methylated CpGs (DMCs) were considered significant at the Bonferroni-

414   Hochberg (BH) corrected *p*-value < 0.05 (for HLA-DR cell types), <0.1 (for CD3 cell types)

415   and <0.2 (for PBMC cell types). The DNA from the human primary macrophages was

416   sequenced using the Diagenode′s RRBS due to a lower DNA yield. The sequenced reads were

417   quality checked using the FastQC[49] (v0.11.9). The sequences were trimmed to remove

418   artificially filled-in cytosines at the 3′ end using the TrimGalore (v.0.6.5)(

419   https://github.com/FelixKrueger/TrimGalore) with a phred score cutoff of 20 and quality

420   checked again after trimming. The trimmed sequences were aligned with the human reference

421   genome (hg38.13) using Bowtie2[50] and removed the duplicates using the Bismark (v.0.22.3)[51].

422   The methylation extractor from Bismark was used to extract the CpG methylation data from

423   the sequences. The SAMtools (v1.7) package[52] was used to sort the binary alignment files

424   (BAM) files on CpG-site chromosomal location and converted to sequence alignment map

425   (SAM) files. The methylated and unmethylated CpG counts were extracted and combined using

426   the DMRfinder[53] (v0.3) package in R (v4.0.2). The CpG-sites located in the X and Y

427   chromosome as well as CpG-sites from mitochondrial DNA were filtered out. For differential

428   methylation analysis the *methylKit*[54] (v1.18.0) package was used. CpG-sites with a read

429   coverage > 10 with both methylated and unmethylated reads were removed from the analysis.

430   The function "calculateDiffMeth" from methylKit package using logistic regression was used

431   as statistical test with "before" and "after" BCG-training as treatment used to identify

432   differentially methylated CpG-sites DMCs. The DMCs were annotated to the official gene

433   symbol using *org.Hs.eg.db* (v3.12) and AnnotationDbi (v1.52) packages using the human

434   genome version hg38. DMGs were defined as sited with a methylation difference of 30%

435   between the conditions and an BH corrected *p*-value of <0.01.

436

437   *Unsupervised cluster analysis*

438   Hierarchical clustering of the all TB-exposed and control individuals was performed with the

439   normalized β-values obtained after the data filtration in each cell type individually. The

440   distance was calculated using the Euclidean distance matrix. The *dendextend*[55] (v1.14.0) and

441   *ape*[56] (v5.4-1) packages in R were used to construct the horizontal hierarchical plots from the

442   three different cell populations using the *hclust* and *dendrogram* functions.

443

444   *Structural annotations*

445   The *EnhancedVolcano* package[57] (v1.8.0) was used to generate the individual volcano plots

446   from all cell populations. The *ChromoMap* package[58] (v0.3) was used to annotate and

447   visualize the genome-wide chromosomal distribution of the DMGs. The interactive plots were

448   generated using the *plotly* (v4.9.3) package.

449   The heatmaps were generated from the filtered DMGs with their respective CpGs for each

450   cell type using the *ComplexHeatmap* (v2.6.2) package[59]. The clustering dendrogram in

451   heatmaps were plotted using the Euclidean distance matrix.

452

453   *Pathway and functional enrichment analyses*

454   We used the PANTHER database (PantherDB v15, 16)[60] to identify the enriched pathways

455   related to our identified DMGs. In addition, to assess functional enrichment, we used the

456   *ReactomePA* (v1.34.0) package[61] with 1000 permutations and the BH-corrected *p*-values.

457   Within the package, GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used

458   and using *clusterProfiler*[62] (v3.18.1), we performed KEGG pathway[63] enrichment analysis

459   (data not shown). To enhance the visualization and better understanding of the enrichment

460   result, *GOplot*[64] (v1.0.2), another package was used. The pathway enrichment was also

461   calculated using the topology-based ontology methods using *RontoTools*[65] (v2.18.0), *SPIA*[67]

462   (v2.42.0) and *pathview*[68] (v1.30.1) was used to visualize the related pathways with the KEGG

463   pathway maps (data not shown).

464

465   *Venn and overlap analyses*

466   Venn analyses were performed to detect the DMGs overlapping between cell populations and

467   between studies. We constructed the Venn diagrams by using *matplotlib-venn* package

468   (https://github.com/konstantint/matplotlib-venn) using in-house python script. The overlap

469   analyses were calculated and plotted using the *go.Sunburst* function from plotly using an in-

470   house python script.

471

472   *Statistical analyses*

473   All differences with a p-value < 0.05 were considered significant if not otherwise stated. We

474   calculated family-wise error rate (FWER) using the BH correction method. All analyses were

475   performed in R (v4.0.2) with the mentioned packages.

476

21

Acknowledgements

Author contributions

M.L., N.I., J.P. and C.B. designed the study, I.P., N.I., C.B., L.K., B.A. and E.K.D performed the laboratory work and the related analyses, J.D. and M.L. designed and performed the bioinformatic analyses of the data, J.D. and L.K wrote the scripts for analysis and created figures. N.I., I.P., J.D., L.K., C.B., E.K.D. and M.L wrote the manuscript. J.D and M.L are co-authors of a patent application "Biomarker for detection of mycobacterial infection and exposure" filed on February 2nd with the Swedish Patent Registry (#100692).

References

1. Fitchett, J. R., MacPherson, P. & Corbett, E. L. Implementing the end TB strategy and the intersection with the sustainable development goals, 2016-2030. *Transactions of the Royal Society of Tropical Medicine and Hygiene* (2015) doi:10.1093/trstmh/trw010.

2. Donald, P. R. *et al.* Droplets, dust and Guinea pigs: An historical review of tuberculosis transmission research, 1878-1940. *Int. J. Tuberc. Lung Dis.* (2018) doi:10.5588/ijtld.18.0173.

3. Davis, J. M. & Ramakrishnan, L. The Role of the Granuloma in Expansion and Dissemination of Early Tuberculous Infection. *Cell* (2009) doi:10.1016/j.cell.2008.11.014.

4. Verma, D. *et al.* Anti-mycobacterial activity correlates with altered DNA methylation pattern in immune cells from BCG-vaccinated subjects. *Sci. Rep.* **7**, (2017).

5. Hasso-Agopsowicz, M., Scriba, T. J., Hanekom, W. A., Dockrell, H. M. & Smith, S. G. Differential DNA methylation of potassium channel KCa3.1 and immune signalling pathways is associated with infant immune responses following BCG vaccination. *Sci. Rep.* (2018) doi:10.1038/s41598-018-31537-9.

6. Kaufmann, E. *et al.* BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* (2018) doi:10.1016/j.cell.2017.12.031.

7. Jonsson, J. *et al.* A borderline range for Quantiferon Gold In-Tube results. *PLoS One* (2017) doi:10.1371/journal.pone.0187313.

8. Sikkeland, L. I. B., Kongerud, J., Stangeland, A. M., Haug, T. & Alexis, N. E. Macrophage enrichment from induced sputum [3]. *Thorax* (2007) doi:10.1136/thx.2006.073544.

9. Das, J., Verma, D., Gustafsson, M. & Lerm, M. Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naïve subjects. *Epigenetics* (2019) doi:10.1080/15592294.2019.1603963.

10. Ogryzko, N. V. *et al.* Hif-1α–Induced Expression of Il-1β Protects against Mycobacterial Infection in Zebrafish. *J. Immunol.* **202**, (2019).

11. Eklund, D. *et al.* Vitamin D enhances IL-1β secretion and restricts growth of Mycobacterium tuberculosis in macrophages from TB patients. *Int. J. Mycobacteriology* (2013) doi:10.1016/j.ijmyco.2012.11.001.

12. Liu, P. T. *et al.* Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science (80-.).* **311**, 1770–1773 (2006).

13. Hölscher, C. *et al.* Chemical p38 MAP kinase inhibition constrains tissue inflammation and improves antibiotic activity in Mycobacterium tuberculosis-infected mice. *Sci.*

Rep. (2020) doi:10.1038/s41598-020-70184-x.

14. Brandenburg, J. & Reiling, N. The Wnt blows: On the functional role of Wnt signaling in mycobacterium tuberculosis infection and beyond. *Frontiers in Immunology* (2016) doi:10.3389/fimmu.2016.00635.

15. Castro, R. C. *et al.* NOTCH1 and DLL4 are involved in the human tuberculosis progression and immune response activation. *Tuberculosis* (2020) doi:10.1016/j.tube.2020.101980.

16. Eklund, D. *et al.* Human gene variants linked to enhanced NLRP3 activity limit intramacrophage growth of mycobacterium tuberculosis. in *Journal of Infectious Diseases* (2014). doi:10.1093/infdis/jit572.

17. Raffetseder, J. *et al.* Replication rates of mycobacterium tuberculosis in human macrophages do not correlate with mycobacterial antibiotic susceptibility. *PLoS One* (2014) doi:10.1371/journal.pone.0112426.

18. Kaufmann, E. *et al.* BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* **172**, 176-190.e19 (2018).

19. Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science (80-.).* (2014) doi:10.1126/science.1251086.

20. Mourits, V. P. *et al.* BCG-Induced Trained Immunity in Healthy Individuals: The Effect of Plasma Muramyl Dipeptide Concentrations. *J. Immunol. Res.* (2020) doi:10.1155/2020/5812743.

21. Kleinnijenhuis, J. *et al.* Bacille Calmette-Guérin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl. Acad. Sci. U. S. A.* (2012) doi:10.1073/pnas.1202870109.

22. Russell, D. G., Cardona, P. J., Kim, M. J., Allain, S. & Altare, F. Foamy macrophages and the progression of the human tuberculosis granuloma. *Nature Immunology* (2009) doi:10.1038/ni.1781.

23. Benn, C. S., Fisker, A. B., Whittle, H. C. & Aaby, P. Revaccination with Live Attenuated Vaccines Confer Additional Beneficial Nonspecific Effects on Overall Survival: A Review. *EBioMedicine* (2016) doi:10.1016/j.ebiom.2016.07.016.

24. Thysen, S. M. *et al.* Neonatal BCG vaccination and child survival in TB-exposed and TB-unexposed children: A prospective cohort study. *BMJ Open* (2020) doi:10.1136/bmjopen-2019-035595.

25. Verrall, A. J. *et al.* Early Clearance of Mycobacterium tuberculosis: The INFECT Case Contact Cohort Study in Indonesia. *J. Infect. Dis.* (2019) doi:10.1093/infdis/jiz168.

26. Arts, R. J. W., Joosten, L. A. B. & Netea, M. G. Immunometabolic circuits in trained immunity. *Seminars in Immunology* (2016) doi:10.1016/j.smim.2016.09.002.

27.   Kumar, R. *et al.* Immunometabolism of Phagocytes During Mycobacterium tuberculosis Infection. *Frontiers in Molecular Biosciences* (2019) doi:10.3389/fmolb.2019.00105.

28.   Kleinnijenhuis, J. *et al.* Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl. Acad. Sci.* **109**, 17537–17542 (2012).

29.   Harding, J. S. *et al.* VEGF-A from Granuloma Macrophages Regulates Granulomatous Inflammation by a Non-angiogenic Pathway during Mycobacterial Infection. *Cell Rep.* (2019) doi:10.1016/j.celrep.2019.04.072.

30.   Rao Muvva, J., Parasa, V. R., Lerm, M., Svensson, M. & Brighenti, S. Polarization of Human Monocyte-Derived Cells With Vitamin D Promotes Control of Mycobacterium tuberculosis Infection. *Front. Immunol.* (2020) doi:10.3389/fimmu.2019.03157.

31.   Cox, M. A. *et al.* Beyond neurotransmission: acetylcholine in immunity and inflammation. *Journal of Internal Medicine* (2020) doi:10.1111/joim.13006.

32.   Cox, M. A. *et al.* Choline acetyltransferase–expressing T cells are required to control chronic viral infection. *Science (80-.).* (2019) doi:10.1126/science.aau9072.

33.   Ganor, Y. & Levite, M. The neurotransmitter glutamate and human T cells: Glutamate receptors and glutamate-induced direct and potent effects on normal human T cells, cancerous human leukemia and lymphoma T cells, and autoimmune human T cells. *J. Neural Transm.* (2014) doi:10.1007/s00702-014-1167-5.

34.   Ganor, Y. & Levite, M. The neurotransmitter glutamate and human T cells: Glutamate receptors and glutamate-induced direct and potent effects on normal human T cells, cancerous human leukemia and lymphoma T cells, and autoimmune human T cells. *J. Neural Transm.* **121**, 983–1006 (2014).

35.   Kuhn, J. H., Li, W., Radoshitzky, S. R., Choe, H. & Farzan, M. Severe acute respiratory syndrome coronavirus entry as a target of antiviral therapies. *Antiviral Therapy* (2007).

36.   Lake, M. A., Ambrose, L. R., Lipman, M. C. I. & Lowe, D. M. ' 'Why me, why now?' Using clinical immunology and epidemiology to explain who gets nontuberculous mycobacterial infection. *BMC Med.* (2016) doi:10.1186/s12916-016-0606-6.

37.   Das, J., Idh, N., Sikkeland, L. I. B., Paues, J. & Lerm, M. DNA methylome-based validation of induced sputum as an effective protocol to study lung immunity: construction of a classifier of pulmonary cell types. *bioRxiv* (2021) doi:10.1101/2021.03.12.435086.

38.   Ugarte-Gil, C., Elkington, P. T., Gotuzzo, E., Friedland, J. S. & Moore, D. A. J. Induced sputum is safe and well-tolerated for TB diagnosis in a resource-poor primary healthcare setting. *American Journal of Tropical Medicine and Hygiene* (2015)

doi:10.4269/ajtmh.14-0583.

39. Alexis, N., Soukup, J., Ghio, A. & Becker, S. Sputum phagocytes from healthy individuals are functional and activated: A flow cytometric comparison with cells in bronchoalveolar lavage and peripheral blood. *Clin. Immunol.* (2000) doi:10.1006/clim.2000.4911.

40. Alexis, N. E. *et al.* In vivo particle uptake by airway macrophages in healthy volunteers. *Am. J. Respir. Cell Mol. Biol.* (2006) doi:10.1165/rcmb.2005-0373OC.

41. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* (2017) doi:10.1186/s12859-017-1934-z.

42. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* (2012) doi:10.1038/nmeth.2019.

43. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* (2012) doi:10.1186/gb-2012-13-6-r44.

44. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu049.

45. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts680.

46. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).

47. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu029.

48. Ritchie, M. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies | Nucleic Acids Research | Oxford Academic. *limma powers Differ. Expr. Anal. RNA-sequencing microarray Stud.* (2015).

49. Andrews S. FastQC A Quality control tool for high throughput sequence data. *Babraham Bioinfo* (2018).

50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* (2012) doi:10.1038/nmeth.1923.

51. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* (2011) doi:10.1093/bioinformatics/btr167.

52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp352.

53.   Gaspar, J. M. & Hart, R. P. DMRfinder: Efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics* (2017) doi:10.1186/s12859-017-1909-0.

54.   Akalin, A. *et al.* MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* (2012) doi:10.1186/gb-2012-13-10-R87.

55.   Galili, T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv428.

56.   Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* (2019) doi:10.1093/bioinformatics/bty633.

57.   Blighe, K, S Rana, and M. L. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. (2018).

58.   Anand, L. & Rodriguez Lopez, C. M. ChromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. *bioRxiv* (2019) doi:10.1101/605600.

59.   Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016) doi:10.1093/bioinformatics/btw313.

60.   Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).

61.   Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* (2016) doi:10.1039/c5mb00663e.

62.   Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* (2012) doi:10.1089/omi.2011.0118.

63.   Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* (2000) doi:10.1093/nar/28.1.27.

64.   Walter, W., Sánchez-Cabo, F. & Ricote, M. GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv300.

65.   Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* (2009) doi:10.1186/1471-2105-10-161.

66.   Voichita, C., Ansari, S. & Draghici, S. *ROntoTools: The R Onto-Tools suite*. https://bioconductor.riken.jp/packages/3.10/bioc/vignettes/ROntoTools/inst/doc/rontoto ools.pdf (2019).

67.   Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* (2009)

doi:10.1093/bioinformatics/btn577.

68. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).

Tables

Characteristics	TB-exposed (n = 4)	Controls (n = 6)
Mean age (year)†	36±12	27±6
Mean height (cm) †	173±4	176±7
Mean weight (kg) †	69±8	77±16
Mean Body Mass Index (BMI) †	23±3	24±5
Sex (male/female)	1/3	4/2
Smoking (current/previous/never)	0/1/3	0/0/6
BCG (yes/no)	3/1	6/0
IGRA-positive/IGRA-negative	2/2	1*/5

Table 1: Demographic data of the participants.

†The standard deviation of the mean values is added to the age, height, weight and BMI.

* Borderline-positive.

Characteristics		Cell type	DMGs	Gene name	p-value
TB Exposure	exposed vs. controls	HLA-DR	9,786	(not shown)	0.05
		CD3	5,045	(not shown)	0.05
		PBMC	496	(not shown)	0.2
Age	< 30 vs. ≥ 30	HLA-DR	0		
		CD3	0		
		PBMC	0		
Gender	male vs.female	HLA-DR	3	GALNTL5; MIR3201; OPN3	0.05
		CD3	3	COL9A1; ACTG1; REAT1E	
		PBMC	0		
IGRA	positive vs.negative	HLA-DR	0		
		CD3	0		
		PBMC	0		
BMI	> 25 vs ≤ 25	HLA-DR	0		
		CD3	0		
		PBMC	0		

Table 2: Differentially methylated genes (DMGs) identified through comparison of different study participant characteristics (TB-exposure, age, gender, IGRA status and BMI) in the three different cell populations, CD3+, HLA-DR+ and PBMC. DMGs were only identified when comparing the characteristic TB-exposure and gender.

Characteristics	TB patients (n = 2)	TB-exposed (n = 2)
Mean age (year)[†]	19.5 ± 1.5[†]	35.5 ± 17.5
Mean height (cm)[†]	176 ± 7	171 ± 3
Mean weight (kg)[†]	62.5 ± 6.5	80 ± 10
Mean Body Mass Index (BMI)[†]	19.9 ± 0.6	27.3 ± 2.5
Sex (male/female)	2/0	1/1
Smoking (current/previous/never)	2/0/0	1/0/1
BCG (yes/no)	0/2	2/0

Table 3: Demographic data of the participants in the second recruitment ("test dataset")

[†] The standard deviation of the mean values of age, height, weight, and BMI.

Figures & figure legends
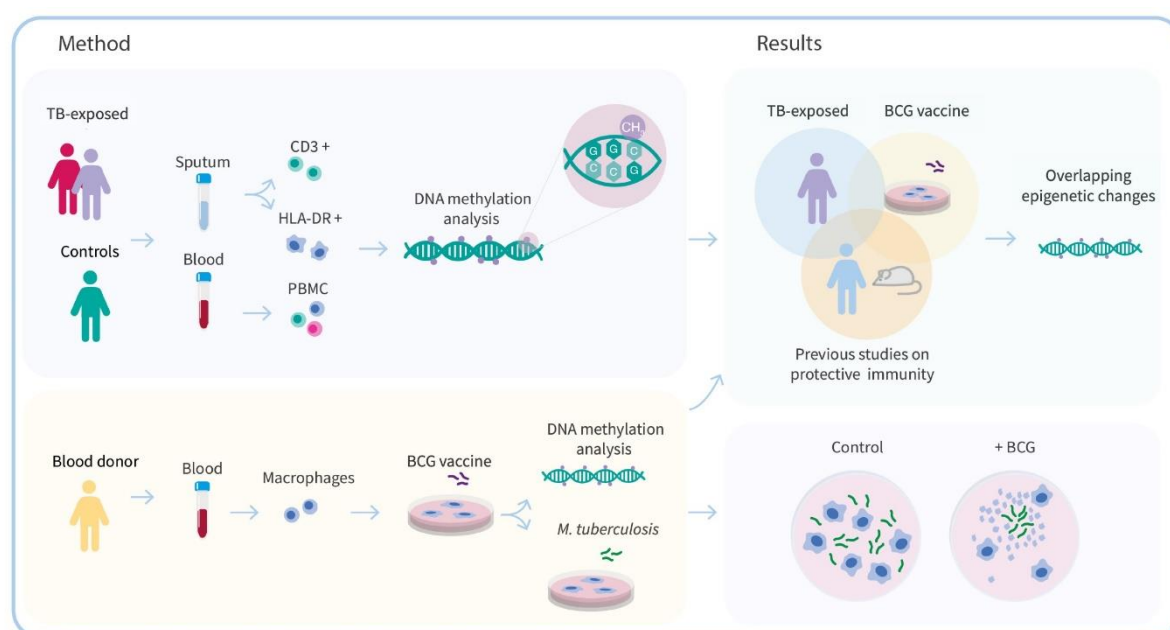
Figure 1. Schematic overview of the project workflow. Sputum and blood were isolated from TB-exposed individuals and controls and the DNA methylomes were analyzed for the different cell types. Macrophages were isolated from healthy blood donors and exposed to the Bacillus Calmette–Guérin (BCG) vaccine *in vitro.* The BCG-exposed macrophages were subjected to DNA methylome analysis and in parallel, infected with *M. tuberculosis* and monitored in a live cell imaging system.
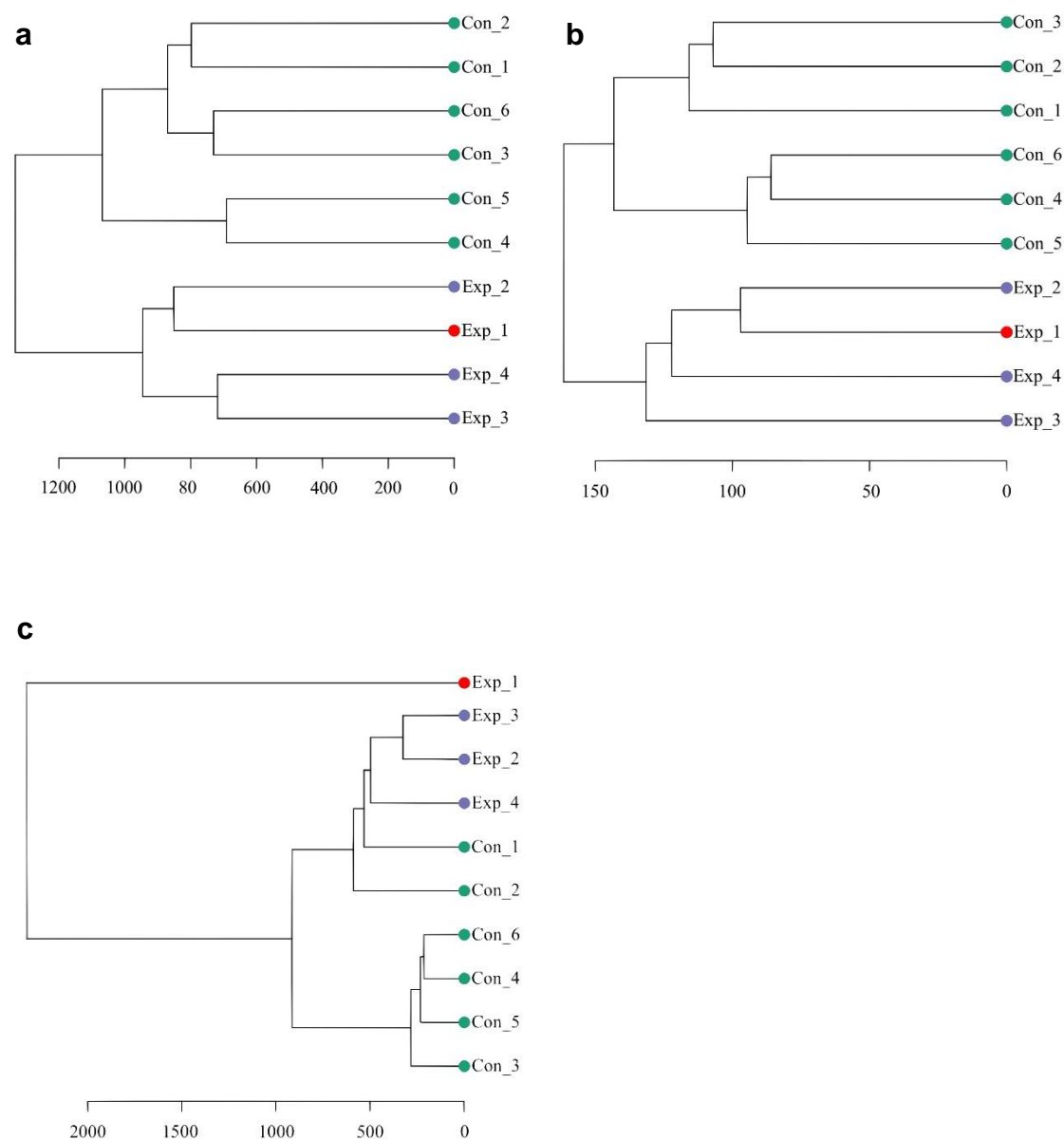
Figure 2. DNA methylome analyses of immune cells from TB-exposed individuals. Dendrograms of the unsupervised hierarchical clustering of the DNAm β-values of from **a.** HLA-DR, **b.** CD3 and **c.** PBMCs. "Con": green=controls, "Exp": purple=TB-exposed, red=TB index case. The scale defines the clustering Euclidean distance.
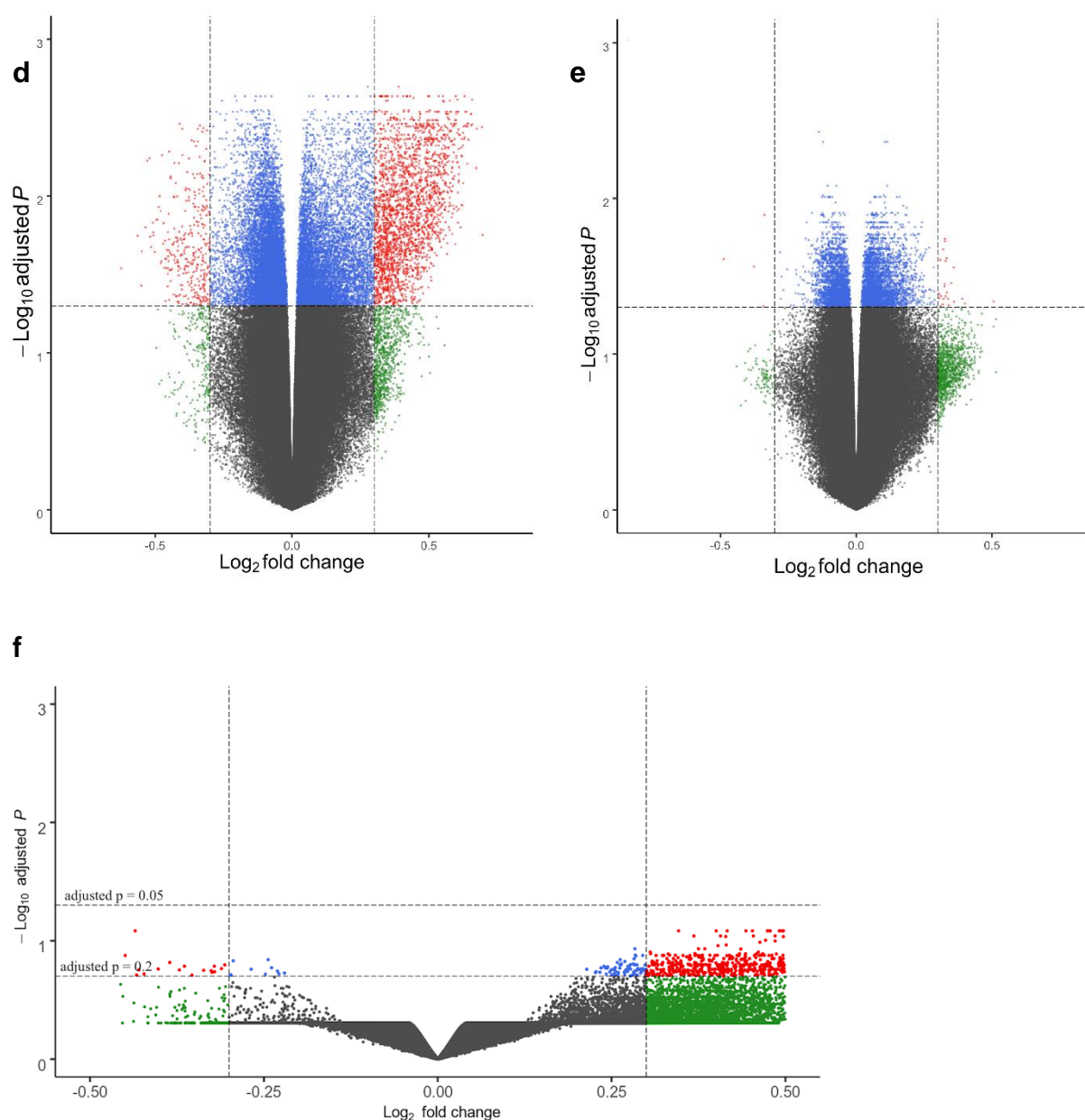
Figure 2. DNA methylome analyses of immune cells from TB-exposed individuals. Volcano plots of DMGs from **d.** HLA-DR, **e.** CD3 and **f.** PBMCs. Red dots represent DMGs above cut-offs (± 0.3 Log$_2$ fold change and BH-corrected *p*-value < 0.05, <0.1 or 0.2 as indicated).
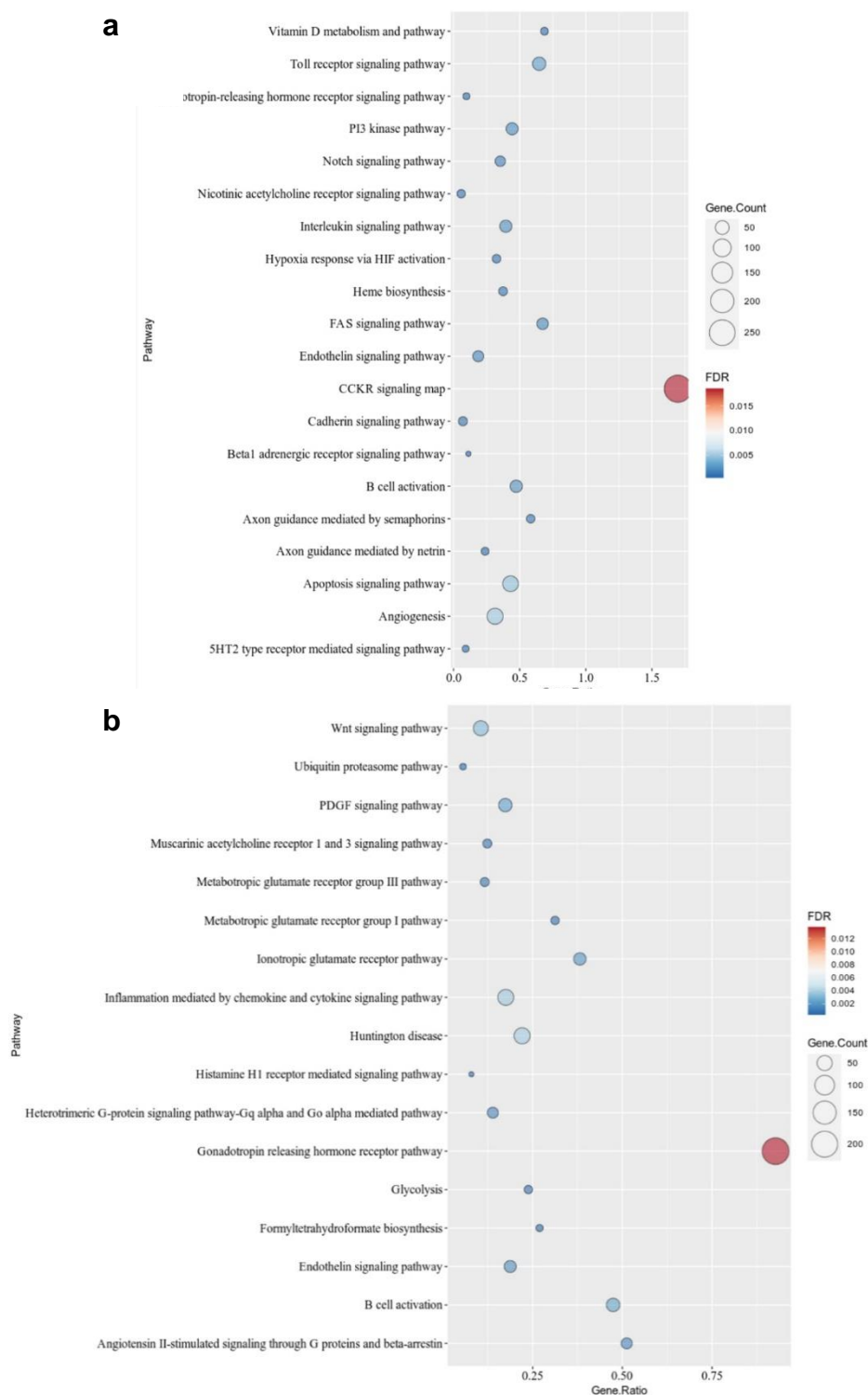
Figure 3. PANTHER pathway analysis of the identified DMGs with the cut-offs for the different cell populations given in Figure 2. Dot plots show the gene ratio, gene counts and FDR-corrected *p*-value for **a.** HLA-DR (top 20 pathways), **b.** CD3 (total 17 pathways)
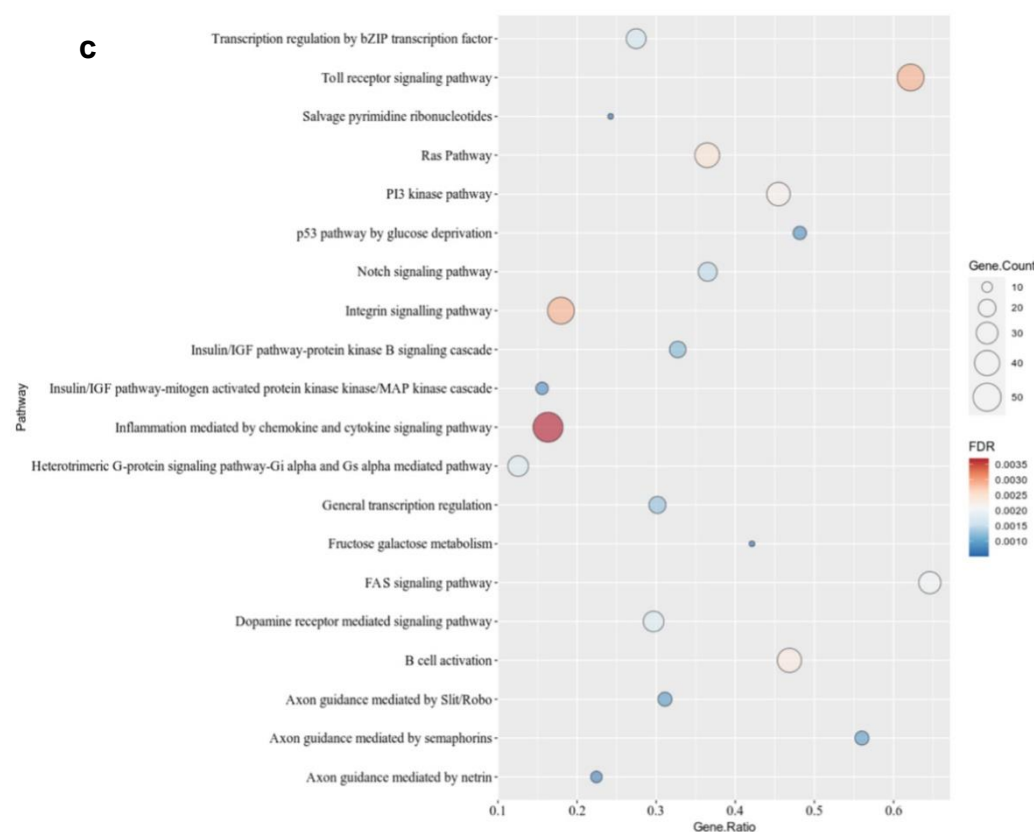
Figure 3. PANTHER pathway analysis, continued. **c.** PBMC (top 20 pathways).

a

b
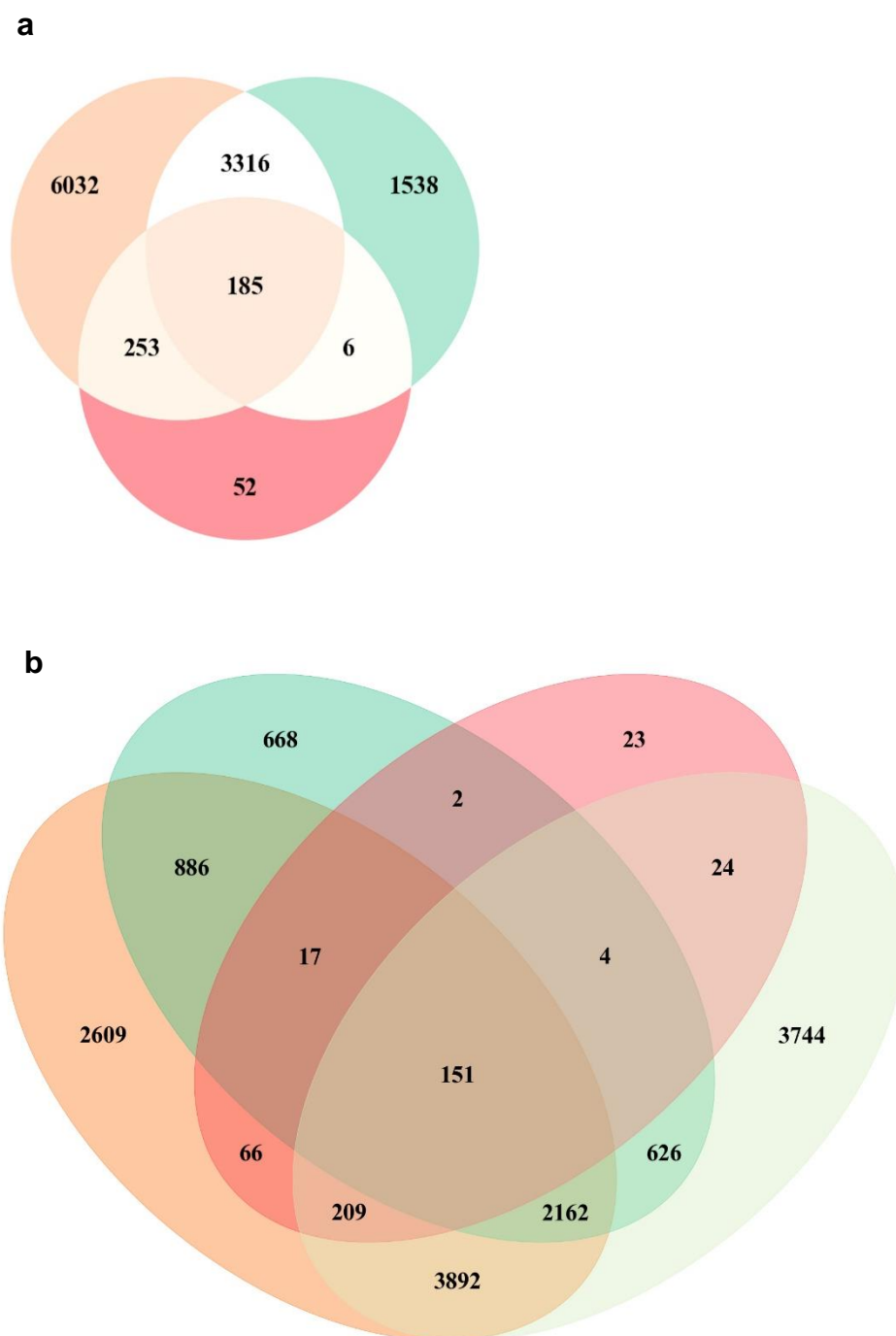
Figure 4: Venn analyses comparing DMGs, pathways and GO terms between different datasets. **a.** Overlapping DMGs derived from the HLA-DR (orange), CD3 (green) and PBMC (red) DNA methylomes. **b.** Overlapping DMGs from this study and from our previous work on BCG induced DMGs (light green).
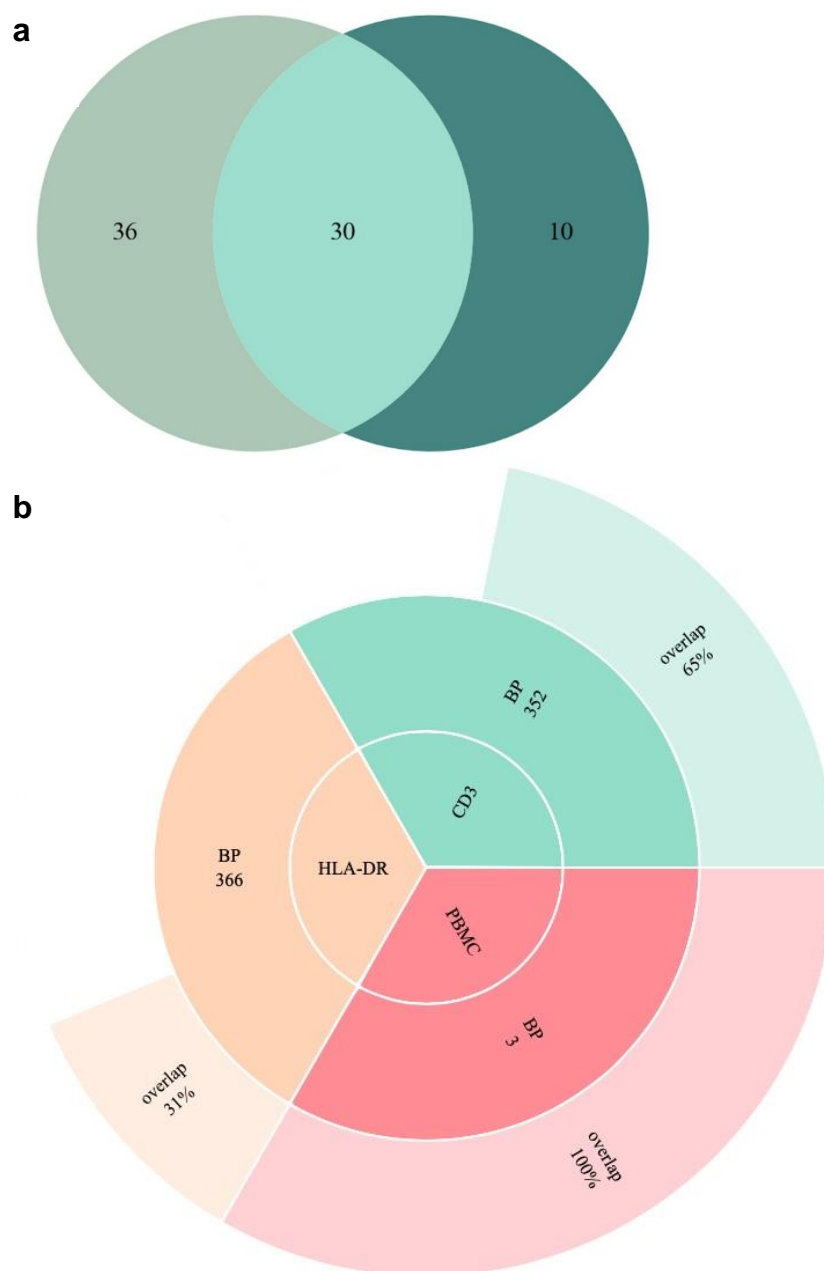
Figure 5. Pathway overlap with other studies' results. **a.** Venn diagram describing the number of Panther pathways overlapping between the ones derived from the 185 common DMGs in this study (dark green) and Hasso-Agopsowicz *et al* (human BCG vaccine study, light green) **a.** Sunburst Plot describing the overlap of enriched GO biological processes emerging from a comparison between the GO data derived from the 185 common DMGs (Figure 4a) and Kaufmann et al (BCG study performed in mouse PBMCs).
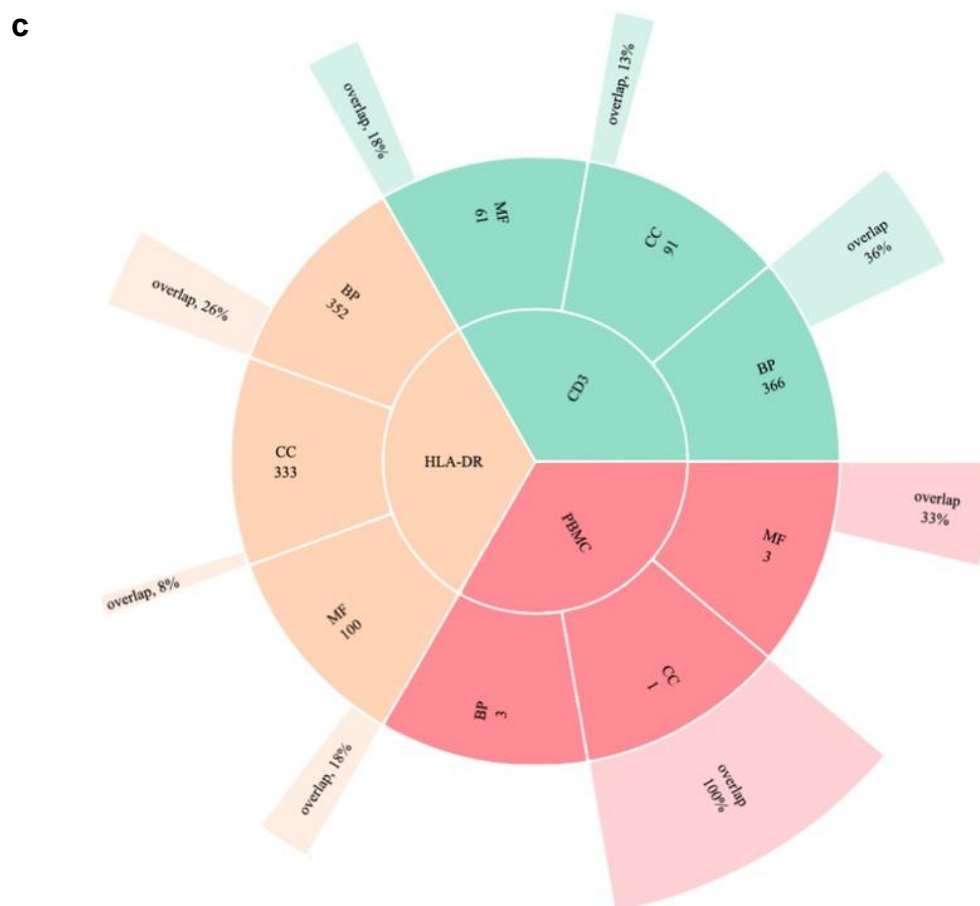
Figure 5. Pathway overlap with other studies' results, continued. **c**. Sunburst Plot describing the overlap of enriched GO biological processes emerging from a comparison between the GO data derived from the 185 common DMGs and Saeed *et al* (study on trained immunity induced by β-glucan).
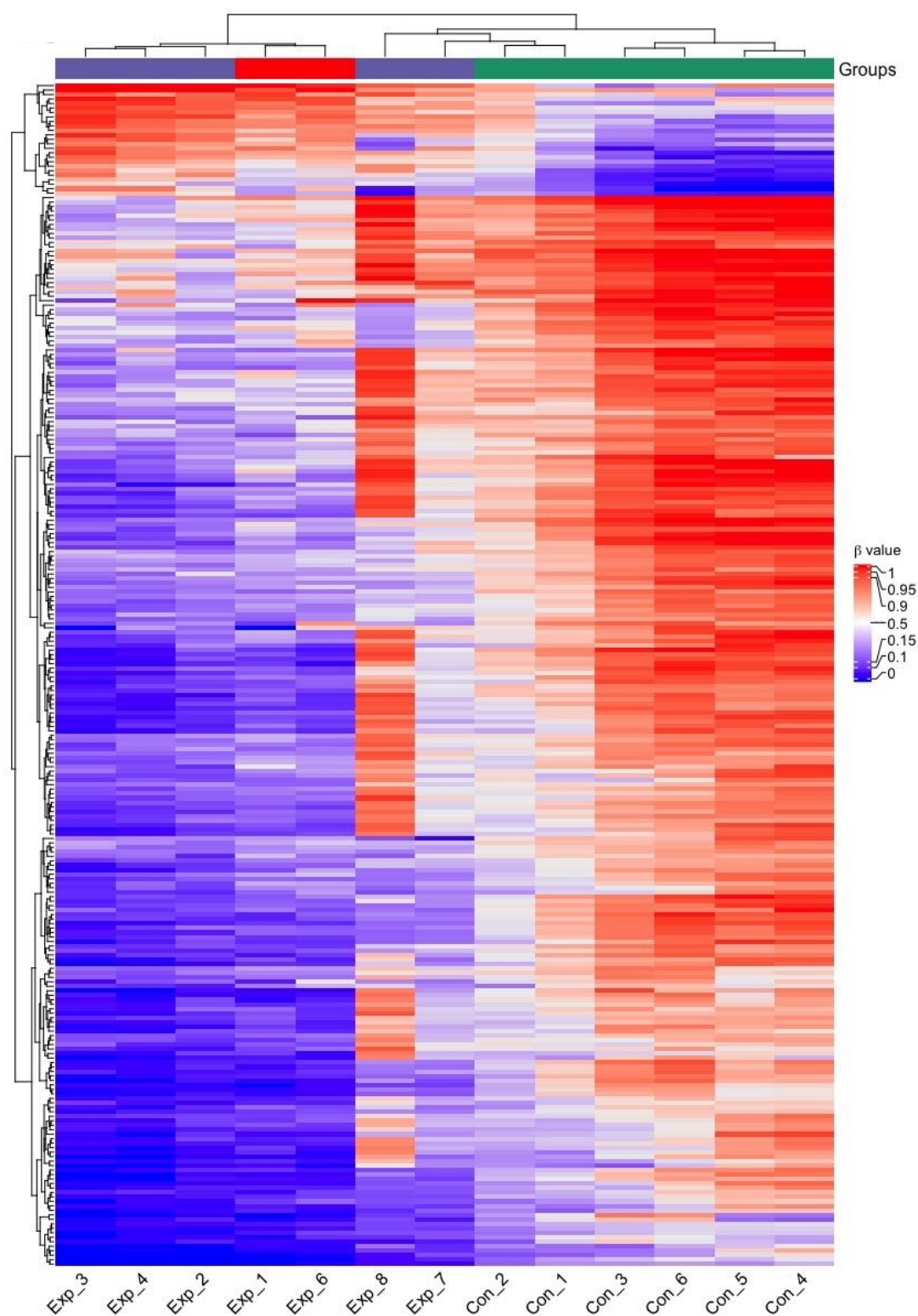
Figure 6. Heatmap of the HLA-DR-derived β values of the signature's 284 CpG sites of the 6 initial subjects (Exp_1-4 and Con_1-6) and the three additional exposed subjects (Exp_6-8). Purple=exposed, red=TB index case, green=controls.
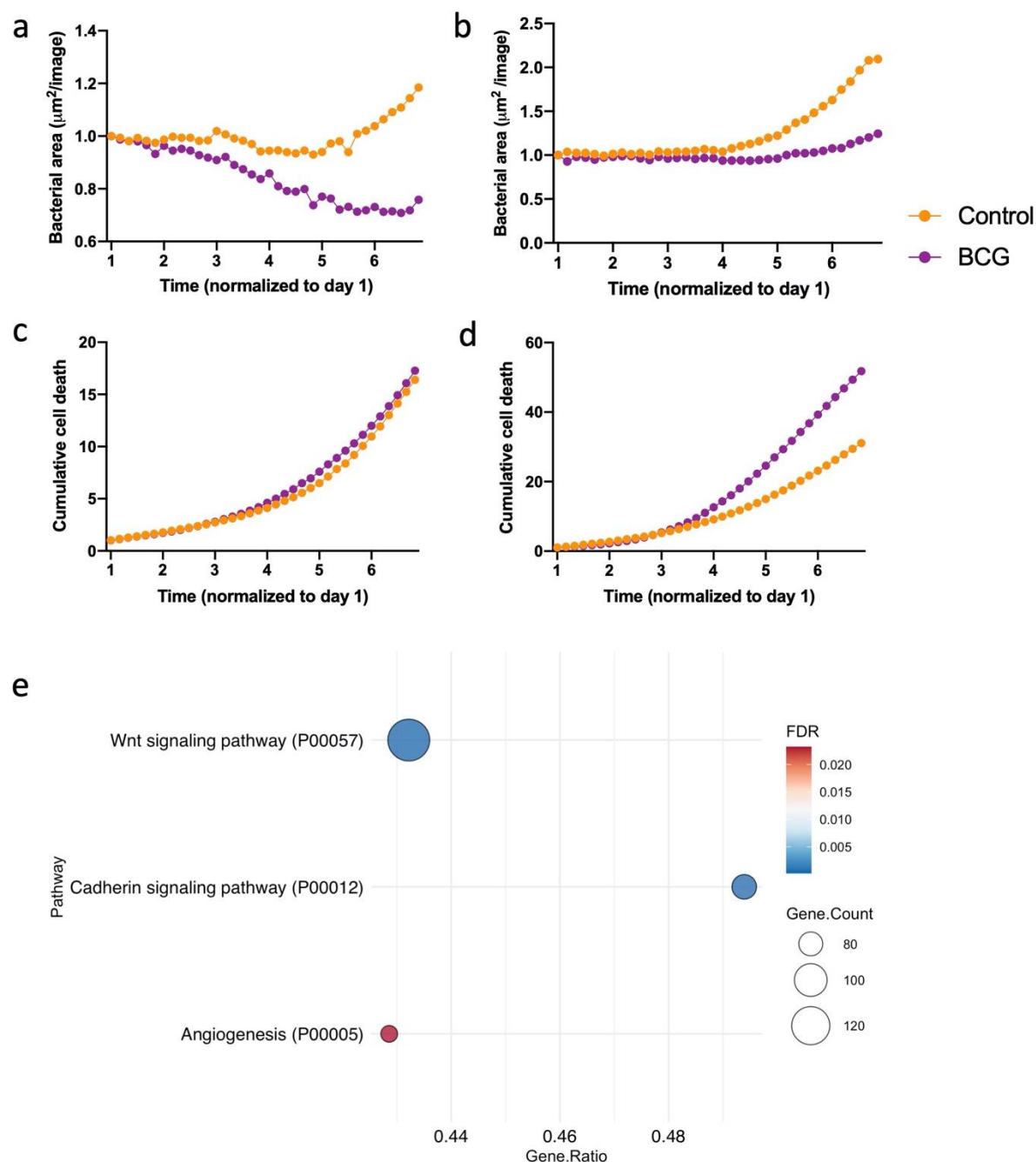
Figure 7. Macrophages were BCG-trained (purple) or left untreated (orange) before infection with *M. tuberculosis* (H37rv-GFP) or DNAm analysis. **a-b.** Total bacterial area (µm²/image) measured by green fluorescence over time. **c-d.** Cumulative cell death measured by the red fluorescent DRAQ7 DNA stain marking the nuclei of non-viable cells. **a-d** shows data from one donor separately with the median of two replicates per timepoint.**e.** PANTHER pathway analysis of the identified DMGs from DNA methylation analysis in the human primary

macrophages exposed to BCG compared to controls. Dot plots show the gene ratio, gene counts

and FDR-corrected *p*-value for 3 significantly over-represented pathways.
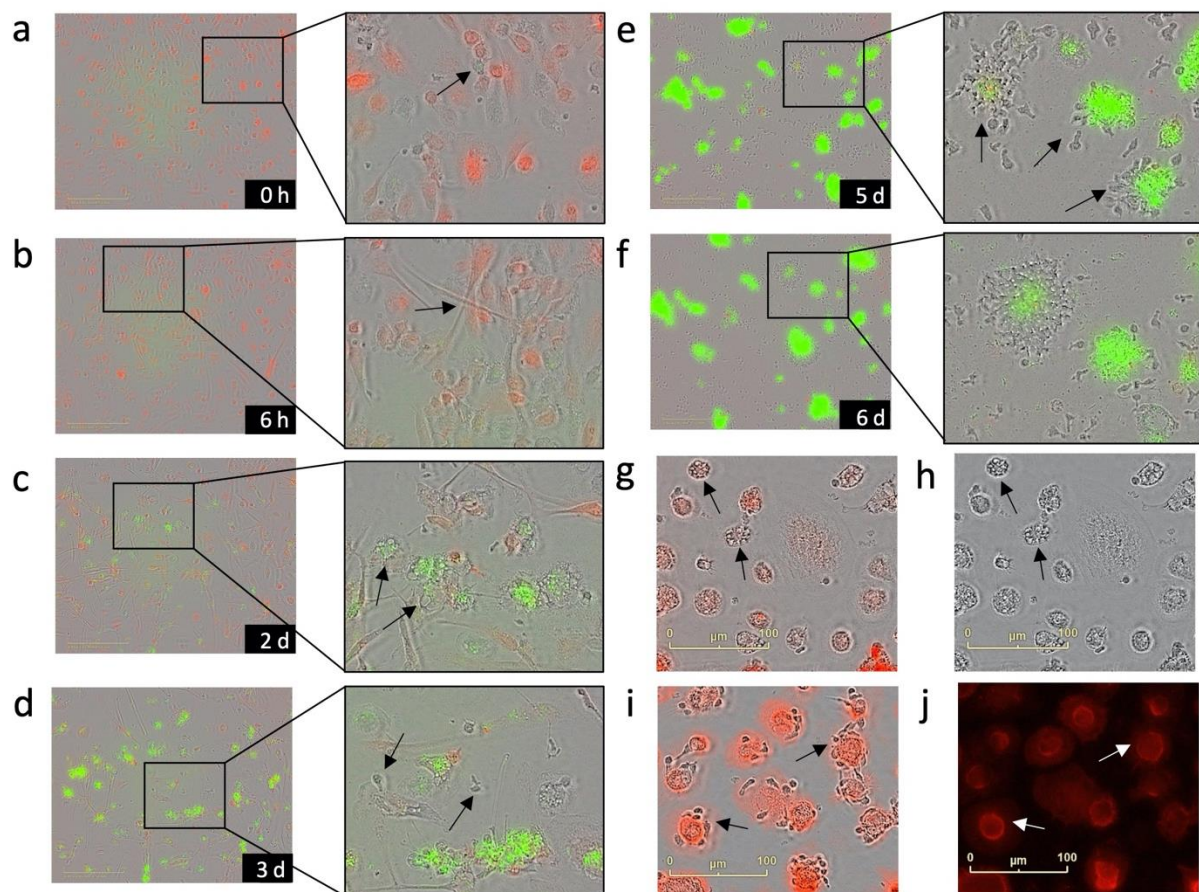
Figure 8. Morphological staging of the emergence of corralling cells (images representative of 3 donors with 4-6 technical replicates). **a.** Macrophages with red fluorescently labelled nuclei using Nuclight and *M. tuberculosis* (H37rv-GFP) fluorescent in green by GFP-expression, arrow showing intracellular *M. tuberculosis.* **b**. Stretching; Macrophages form extensions (100-300 µm long) during the first 24 hours **c**. Vacuolization; non-stretching macrophages become vacuolized (initially <1um in diameter, increasing to 15 um within 24-36 hours). **d.** Emergence; the presence of corralling cells is obvious at day 3 after which their numbers increase **e.** Corralling; corralling cells engage in pushing the bacteria into focal areas. **f.** Containment; *M. tuberculosis* bacteria are contained in the focal areas. **g-h**. Paraformaldehyde-fixed cells stained with Oil red O (red). Arrows indicate vacuolization. **i-j** Paraformaldehyde-fixed cells stained with fluorescently labelled anti-HLA-DR (red). Arrows indicate corralling cells.
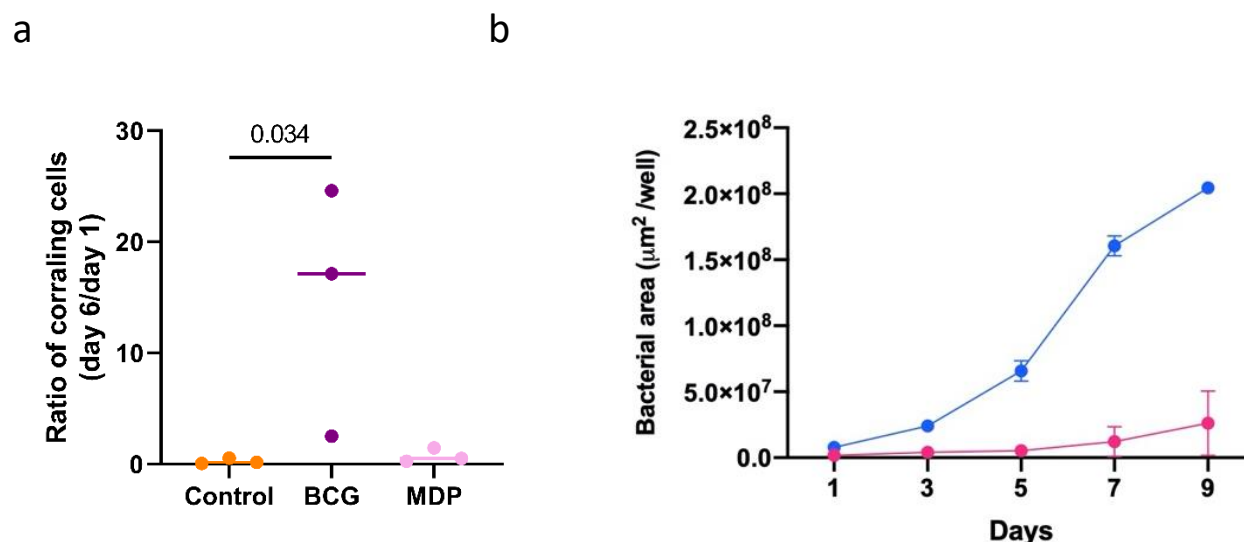
a

b

Figure 9. **a.** Quantification of corralling cells in *M. tuberculosis*-infected cultures with BCG- and MPD-trained macrophages. Cells with a size of 10-20 µm length were manually counted in 4-6 technical replicates from three donors. Statistical analysis was done with the Kruskal-Wallis test with Dunn's multiple comparisons (*p*-value as indicated). **b.** *M. tuberculosis* (strain H37rv-GFP) fluorescence area (µm²/well) measured in 12-well plates in the IncuCyte® S3 analysis tool for 2 donors with corralling cells (pink) and 2 donors without corralling cells (blue), error bars show SEM.