1 **CD4+CCR6+ T cells dominate the BCG-induced transcriptional signature**

2

3 Akul Singhania[1,†], Paige Dubelko[1,†], Rebecca Kuan[1], William D. Chronister[1], Kaylin Muskat[1],

4 Jyotirmoy Das[2], Elizabeth J. Phillips[3,4], Simon A. Mallal[3,4], Grégory Seumois[1], Pandurangan

5 Vijayanand[1], Alessandro Sette[1,5], Maria Lerm[2], Bjoern Peters[1,5], Cecilia S. Lindestam

6 Arlehamn[1,*]

7

8 [1] Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La

9 Jolla, CA 92037, USA

10 [2] Division of Infection and Inflammation, Department of Biomedical and Clinical Sciences,

11 Faculty of Medicine and Health Sciences, Linköping University, Linköping, Sweden

12 [3] Institute for Immunology and Infectious Diseases, Murdoch University, Perth, WA 6150,

13 Australia

14 [4] Vanderbilt University School of Medicine, Nashville, TN 37235, USA

15 [5] Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA

16 [†] Co-first authors

17 * Corresponding author

18 Cecilia S. Lindestam Arlehamn

19 Email: cecilia@lji.org

20

21 Short title: BCG-induced transcriptional signatures

22

23

24   **ABSTRACT**

25   The century-old *Mycobacterium bovis* Bacillus Calmette-Guerin (BCG) remains the only

26   licensed vaccine against tuberculosis (TB). Despite this, there is still a lot to learn about the

27   immune response induced by BCG, both in terms of phenotype and specificity. Here, we

28   investigated immune responses in adult individuals pre and 8 months post BCG vaccination. We

29   specifically determined changes in gene expression, cell subset composition, DNA methylome,

30   and the TCR repertoire induced in PBMCs and CD4 memory T cells associated with antigen

31   stimulation by either BCG or a *Mycobacterium tuberculosis* (*Mtb*)-derived peptide pool.

32   Following BCG vaccination, we observed increased frequencies of CCR6+ CD4 T cells, which

33   includes both Th1* and Th17 subsets, and mucosal associated invariant T cells (MAITs). A

34   large number of immune response genes and pathways were upregulated post BCG

35   vaccination with similar patterns observed in both PBMCs and memory CD4 T cells, thus

36   suggesting a substantial role for CD4 T cells in the cellular response to BCG. These

37   upregulated genes and associated pathways were also reflected in the DNA methylome. We

38   described both qualitative and quantitative changes in the BCG-specific TCR repertoire post

39   vaccination, and importantly found evidence for similar TCR repertoires across different

40   subjects. The immune signatures defined herein can be used to track and further characterize

41   immune responses induced by BCG, and can serve as reference for benchmarking novel

42   vaccination strategies.

43

44 **INTRODUCTION**

45   Tuberculosis (TB), claims over 1.5 million lives every year, and is caused by infection with

46   *Mycobacterium tuberculosis* (*Mtb*). The Bacillus Calmette-Guerin (BCG) vaccine, first introduced

47   a century ago, remains the only approved vaccine against TB and most widely used vaccine in

48   the world. BCG offers variable efficacy against pulmonary TB in all age-groups, but has high

49   efficacy against severe forms of TB in young children [1-3]. The underlying cause for this variable

50   efficacy is difficult to pinpoint, and is likely due to multiple factors including geographical location

51   and exposure to environmental mycobacteria. Many approaches have been considered to

52   improve the BCG efficacy, including changes in the route of administration from the current

53   intradermal standard [4-7], booster vaccinations with either BCG [8-13], or with *Mtb*-derived antigens

54   [14-16], and recombinant BCG strains [14,17]. These results have been mixed and a clear enhanced

55   efficacy has not been achieved. Further rational vaccine improvement efforts are hindered by

56   our incomplete understanding of the mechanisms of immune protection elicited by BCG.

57   BCG was developed empirically more than a century ago [18], and yet surprisingly little is

58   known about BCG-induced immune responses to this date. The specific cell subset or subsets

59   responsible for mediating BCG's protective effects have not been clarified. While BCG-specific

60   T cell reactivity does not mediate protection alone, it can be used as an immune correlate of *Mtb*

61   infection and disease risk [19,20]. An increased level of BCG-specific cells post-vaccination is

62   frequently reported, but the BCG-specific T cell response has varied considerably between

63   studies [1,18,21,22]. Thus, studying the cellular response induced by BCG vaccination is a key

64   component of understanding how it mediates protection and what immune responses are

65   triggered.

66   Systems biology is a compelling approach that can be used to dissect the heterogeneity of

67   immune responses following various perturbations (vaccination, infection, disease,

68   autoimmunity, etc.). The resulting gene signatures and cellular profiles have proven of

69   significant diagnostic, prognostic, or mechanistic value [23]. Similar to what we have previously

70   described for individuals with latent TB infection (LTBI) as compared to TB-uninfected controls

71   [63, 64], here we used a comprehensive systems biology/multiomics approach to determine a

72   global picture of the immune responses triggered by BCG vaccination in humans. We

73   conducted a longitudinal study of immune responses associated with BCG vaccination in a

74   cohort of BCG naïve adults. This allowed the opportunity to study cell subset changes, gene

75   signatures and the TCR repertoire changes induced by BCG vaccination. The results indicate

76   that BCG induced gene signatures in adults are primarily driven by the expansion of Th1* CD4

77   memory T cells and results in both qualitative and quantitative TCR repertoire changes.

78

79   **RESULTS**

80   **The frequency of specific T cell subsets increases following BCG vaccination**

81   To determine the effects of BCG vaccination on specific T cell responses and gene

82   expression, PBMC samples were obtained from BCG naïve individuals (pre-vaccination).

83   Subsequently, these individuals were administered the BCG vaccine, and PBMC samples were

84   collected again 8 months after intradermal administration of BCG (post-vaccination), chosen as

85   a representative of a time point were immune responses are expected to have reached a steady

86   state memory phase (**Figure 1a**). PBMCs from both time points (pre- and post-vaccination)

87   were assayed by flow cytometry and in parallel stimulated for 24 hours *in vitro* with the vaccine

88   itself (BCG) or media (unstim) to identify BCG-induced immune responses based on Fluorospot

89   and RNA-sequencing (**Figure 1a**).

90   We previously described a flow cytometry panel designed to quantitate the relative

91   frequency of major PBMC subsets [24]. In addition, we have previously identified increased

92   frequencies of a specific Th subset, Th1* (CXCR3+CCR6+) [25], as well as a decrease in MR1+ T

93   cells in individuals with latent TB infection (LTBI) as compared to TB negative subjects [26]. Here,

94   we investigated the frequency changes following BCG vaccination of major PBMC subsets and

95   specific T cell subsets in the absence of antigen-specific stimulation. Upon BCG vaccination we

96   observed an increase in T cell frequency, which was specifically driven by CD4 T cells while

97   CD8 T cells remained unchanged (**Figure 1b**). In contrast, the frequency of B cells decreased

98   following vaccination (**Supplementary figure 1a**). No changes were observed for monocytes,

99   NK cells, or CD3+CD56+ T cell populations (**Supplementary figure 1a**). Memory CD4 and CD8

100 T cell populations, as defined by CCR7 and CD45RA expression, remained unchanged

101 following vaccination (**Supplementary figure 1b, c**). In terms of non-conventional T cells, we

102 found an increased frequency of MAITs (as defined by 5-OP-RU loaded MR1 tetramers)

103 following vaccination, but no changes in γδ T cells and NKT cells (**Figure 1c**). Finally, we

104 determined the frequencies of Th subsets, defined by CXCR3, CCR6 and CCR4 expression. All

105   CCR6+ populations, including Th17 and Th1* cells, increased following BCG vaccination,

106   whereas CCR6- populations, including Th1 and Th2 cells, remained unchanged (**Figure 1d**).

107   These results are consistent with our previous identification of Th1* as the Th subset that

108   contains the vast majority of *Mtb*-and non-tuberculous mycobacteria (NTM)-specific T cells [25,27],

109   and the present results point towards a role of these subsets following BCG vaccination as well.

110   **BCG-induced T cell responses and gene expression changes are enhanced upon BCG**

111   **vaccination**

112   We next investigated the magnitude and quality of BCG-induced cellular immune responses.

113   The magnitude of BCG-stimulated T cell responses, as measured by IFNγ Fluorospot assay,

114   increased post-vaccination (**Figure 2a**). In comparison, no difference was observed in the

115   magnitude of response against PHA, which was used as a positive control (**Figure 2a**).

116   To determine the effects of stimulation and vaccination on global gene expression, principal

117   component analysis (PCA) was performed on the RNA-Seq data. The PCA showed a distinct

118   separation of BCG stimulated and unstimulated samples (**Figure 2b**). In contrast, a smaller

119   separation was observed in the pre- and post- vaccination samples, in each stimulated condition

120   (**Figure 2b**).

121   In addition to the PCA analysis we performed differential gene expression analysis

122   comparing BCG stimulated and unstimulated samples, to explore BCG-induced gene

123   expression changes pre- and post- vaccination. Upon BCG vaccination, we observed an

124   increase in the number of differentially expressed genes (DEG) (**Figure 2c; Supplementary**

125   **table 1**). A total of 549 DEGs were identified when BCG-stimulated vs. unstimulated pre-

126   vaccination samples were studied; 290 downregulated and 259 upregulated. This compares

127   with a total of 927 DEGs identified post-vaccination; 527 downregulated and 400 upregulated.

128   Importantly, BCG-induced T cell responses can be detected prior to vaccination, but they were

129   enhanced 8 months post-vaccination.

130   **BCG-induced gene signatures are more pronounced after vaccination**

131   We next investigated the BCG-induced gene signatures in more detail. We focused on the

132   genes that were upregulated following BCG stimulation. A majority of the DEGs identified pre-

133   vaccination were also differentially expressed post-vaccination (**Figure 3a**; 259 genes

134   upregulated pre-vaccination and 400 genes upregulated post-vaccination, with an overlap of

135   175 genes). To determine if the unique DEGs identified both pre- and post-vaccination were

136   truly unique, or if they were below the cut-off (padj values < 0.05 and log2 fold change > 1 or < -

137   1) for reaching significance in the other group, we compared fold changes for all upregulated

138   genes identified pre- and post-vaccination. We observed that although most of these genes

139   were upregulated both pre- and post-vaccination, the majority of the genes had higher fold

140   changes post-vaccination (**Figure 3b**; **Supplementary table 1**; the 343 genes above the 45-

141   degree slope represent those that were higher post-vaccination, and the 141 genes below the

142   45-degree slope were higher pre-vaccination). This suggests that although similar genes were

143   upregulated upon BCG stimulation, the magnitude of this increase was greater post-vaccination.

144   Several immune-related genes were upregulated in response to BCG stimulation both pre-

145   and post-vaccination. These included chemokine ligands (CCL3, CCL4, CCL20, CXCL1,

146   CXCL2, CXCL3, and CXCL9), T cell activation markers (CD38, CD69, and TNFRSF4 (OX40)),

147   and cytokines (IFNG, IL1B, IL32, GZMB, and TNF), reflecting the broad immune response

148   triggered by stimulation with BCG (**Supplementary table 1**). The genes that had a higher fold

149   change post-vaccination included T cell activation genes (CD274 (PDL1) and DPP4 (CD26)),

150   cytokines (IL13, IL17F, IL22, GZMA), chemokine ligands (CXCL5, CXCL8, CXCL13), and

151   ligands for CXCR3 (CXCL10 (IP-10), and CXCL11), indicating a boosting of certain aspects of

152   the BCG-induced immune response post-vaccination (**Supplementary table 1**).

153   Functional enrichment analysis of the upregulated genes showed that similar pathways,

154   such as cytokine response, interleukin signaling and secreted factors, were identified in both the

155   genes that were higher post-vaccination (**Figure 3c; Supplementary table 2**) and in genes

156   higher pre-vaccination (**Figure 3d; Supplementary table 2**). However, as observed above, the

157 magnitude of BCG related changes was greater post-vaccination and the common pathways

158 identified in both groups had much greater significance post-vaccination. Moreover, pathways

159 such as IL-23 mediated signaling, and IL-10 and IL-17 signaling were observed only in genes

160 higher in post-vaccination (**Figure 3c**). Cell-type enrichment of upregulated genes showed an

161 enrichment of activated CD4 and CD8 activated T cells in both groups (**Figure 3c and 3d**), and

162 a slight enrichment of NK cells post-vaccination only (**Figure 3c**). These results demonstrate an

163 increase in the magnitude of BCG-specific gene expression changes post BCG vaccination.

164 Moreover, pathway enrichment suggested that a Th1*/Th17-like signature is associated with

165 BCG vaccination, which was also observed by flow cytometry analysis.

166 **BCG-induced DNA methylation changes reflect the gene signatures**

167 A higher magnitude of response upon microbial challenge is expected in epigenetically

168 reprogrammed cells [28], and BCG has been shown to induce DNA methylation (DNAm) changes

169 [29]. Therefore, we next assessed the methylation status of >850,000 CpG sites in DNA derived

170 from PBMC pre- and post-vaccination. First, using the Houseman algorithm to deconvolute cell

171 types from PBMC using DNAm data [30], we found, similarly to the flow cytometry analysis an

172 increased frequency of CD4 T cells post vaccination (**Supplementary figure 2a**). Unlike the

173 PCA analysis for RNAseq of unstimulated PBMCs, the PCA analysis of the DNAm data

174 revealed a clear separation between pre- and post-vaccination samples (**Figure 4a**). We next

175 determined the differentially methylated CpG sites and identified 15,679 hypomethylated and

176 15,309 hypermethylated CpG sites (**Supplementary Figure 2b**), which were both primarily

177 found in the gene bodies and intragenic regions (**Supplementary Figure 2c**). The CpG sites

178 were mapped to genes to identify differentially methylated genes (DMGs), and a subsequent

179 KEGG pathway analysis revealed the pathways affected by the DNAm changes (**Figure 4b**).

180 We identified an IFNγ- and IL-17-related pathway (inflammatory bowel disease), a TNF- and

181 CXCL10-related pathway (TNF signaling pathway), and a NOD2-like receptor signaling

182 pathway. A comparison between the DMGs and the DEGs identified post-vaccination revealed

183   that the majority of the hypo-methylated/upregulated genes were found in the identified

184   pathways (**Figure 4c, d**). These results suggest a correlation between genes identified as

185   upregulated in the RNAseq analysis and hypo-methylated genes.

186   **MTB300-specific reactivity is not boosted by BCG vaccination**

187   Next, we wanted to determine whether reactivity against a peptide pool defined in healthy

188   *Mtb*-infected individuals, and with peptides homologous to peptides found in BCG, was boosted

189   following BCG vaccination. PBMC samples from both time points (pre- and post- vaccination)

190   were stimulated for 24h *in vitro* with the MTB300 peptide pool, as described above, and

191   underwent Fluorospot analysis and RNA-Sequencing. The PCA revealed little separation

192   between the MTB300 stimulated and unstimulated samples, suggesting stimulation with the

193   MTB300 peptide pool does not have a large impact on overall global gene expression (**Figure**

194   **5a**). The lack of separation also suggested lower MTB300-specific reactivity as compared to

195   that seen upon BCG stimulation. As previously noted, there was a smaller, but consistent,

196   separation between pre- and post- vaccination samples within each stimulation condition

197   **(Figure 5a)**. Moreover, while the magnitude of MTB300-specific IFNγ response increased post-

198   vaccination (**Figure 5b**), it was lower than what was observed for BCG stimulation (**Figure 2a**).

199   Differential gene expression analysis comparing MTB300 stimulated to unstimulated

200   samples further revealed a lower number of DEGs overall (**Figure 5c; Supplementary table 1**).

201   Additionally, no increase in the number of DEGs was observed post-vaccination (**Figure 5c;**

202   **Supplementary table 1**). Moreover, there was decreased overlap between pre- and post-

203   vaccination, and comparison of upregulated genes did not show an increased magnitude of log2

204   fold changes post-vaccination, with distinct sets and similar number of genes pre- and post-

205   vaccination (**Figure 5d; Supplementary table 1**; 113 genes above the 45-degree slope, and

206   106 genes below the 45-degree slope). Functional enrichment analysis for genes with higher

207   fold changes post-vaccination included pathways such as cytokine and interleukin signaling,

208   similar to that observed in BCG stimulated samples post-vaccination, albeit with lower

209 significance (**Figure 5e; Supplementary table 2)**. There was a high enrichment for interferon

210 signaling upon MTB300 stimulation post-vaccination, with type I, type II and Jak-STAT signaling

211 pathways being significant. In contrast, genes with higher fold changes pre-vaccination were not

212 specifically enriched for any biological functions, with three pathways barely reaching the

213 significance cutoff (**Figure 5f; Supplementary table 2**). These results suggest that, although

214 there is some similarity in the pathways being affected, the MTB300-specific responses are

215 weakly induced following BCG vaccination, especially when compared to the responses upon

216 stimulation with BCG.

Gene signatures reflect specific cell subset increases post-vaccination

218 We have previously defined a Th1*-specific gene signature [25]. Given the likely importance of

219 these Th1* cells in Mtb infection, we further assessed the increase in Th1* cells observed upon

220 BCG vaccination *in silico* in the RNA-sequencing data. Moreover, we have defined a MAIT-

221 specific gene signature [26] which was also used *in silico* in the RNA sequencing data. Only

222 upregulated genes from the Th1* and MAIT signatures were used to identify cell-specific gene

223 expression (**Supplementary table 3**). We detected an overall increase in the MAIT and Th1*

224 cell signatures in the BCG stimulated samples compared to the MTB300 stimulated and

225 unstimulated samples (**Supplementary figure 3a and b**). This increase was further enhanced

226 in the BCG post-vaccination samples, similar to the observation in the flow cytometry data.

227 These results suggest an increase in the CD4 T cell frequency following BCG vaccine,

228 specifically in CCR6+ (Th1* and Th17) cell subsets.

BCG-induced gene expression in PBMCs is primarily driven by CD4 memory T cells

230 As CD4 T cells are increased upon BCG vaccination, and as memory T cell responses are

231 enhanced upon stimulation with BCG, we also performed RNA-sequencing on CD4 memory T

232 cells isolated from the same individuals as above, pre- and post-vaccination. As with PBMC

233 samples, CD4 memory T cells were stimulated with DMSO (unstimulated), MTB300 peptide

234 pool, and BCG. Differential gene expression analysis comparing BCG stimulated samples to

235   unstimulated samples showed an increase in the number of DEGs post-vaccination (**Figure 6a;**

236   **Supplementary table 1;** 286 genes upregulated pre-vaccination and 530 genes upregulated

237   post-vaccination, with an overlap of 205 genes), similar to that observed in PBMCs (**Figure 3a**).

238   In contrast, and similar to the observations in PBMCs (**Figure 5c**), MTB300 stimulated samples

239   had a much small number of genes differentially expressed compared to unstimulated samples,

240   and this did not increase post vaccination (**Figure 6a**). Focusing on the effect of BCG

241   stimulation on gene expression, PCA showed a separation between the BCG stimulated and the

242   unstimulated samples across PC1 (**Figure 6b**), however, to a lesser extent than what was

243   observed in PBMCs (**Figure 3a**). Moreover, a reduced separation was observed in pre- and

244   post- vaccination samples within each stimulation condition (**Figure 6b**). Correlation between

245   the unique and shared upregulated DEGs between pre- and post-vaccination BCG stimulated

246   samples showed that the majority of genes had a higher fold change post-vaccination (**Figure**

247   **6c; Supplementary table 1;** 467 genes above the 45-degree slope, and 144 genes below the

248   45-degree slope), also consistent with PBMC analysis (**Figure 3b**). Functional enrichment for

249   genes with higher fold changes post-vaccination identified similar pathways as those observed

250   in PBMC post-vaccination, such as cytokine signaling and interaction and interleukin signaling

251   (**Figure 6d; Supplementary table 2**). Moreover, a Th1*/Th17 type signature was also observed

252   here as evidenced by the IL-23 mediated signaling pathway (**Figure 6d,** showing the 10 most

253   significant pathways), and IL-17 signaling pathway (**Supplementary table 2**). In contrast, the

254   smaller number of genes with higher fold changes pre-vaccination did not result in any

255   significant pathways. The involvement of Th1*/Th17 cells were further strengthened by the

256   observation that there was an increase in BCG-induced IFNγ and IL-17 production in CD4+ T

257   cells following BCG vaccination (**Figure 6e**). No increase in IFNγ production was observed in

258   CD8+ or CD4-CD8- T cells (**Figure 6e**). *In silico* analysis of the Th1* signature (**Supplementary**

259   **table 3**), as before, showed an increase in the Th1*-specific gene expression in BCG stimulated

260   samples, with a further increase post-vaccination (**Figure 6f**).

261   With the analysis in CD4 memory T cells showing similar results as those observed in

262   PBMCs, we directly compared the DEGs obtained from BCG stimulated compared to

263   unstimulated samples, post-vaccination. Majority of the genes showed similar perturbations in

264   CD4 memory T cells and PBMCs, albeit to differing fold change and significance levels (**Figure**

265   **6g; Supplementary table 1**). Only 23 genes were identified to be significant in both PBMC and

266   CD4 memory T cells that had $\log_2$ fold changes in opposite direction (upregulated in CD4

267   memory T cells, but downregulated in PBMCs, and vice versa). These results indicate an

268   increase in the magnitude of BCG related gene expression changes post BCG vaccination in

269   CD4 memory T cells. Moreover, the results observed in CD4 memory T cells were similar to

270   those in PBMCs, suggesting that majority of the BCG related gene expression changes

271   observed in PBMCs are driven primarily by CD4 memory T cells, upon BCG vaccination.

272   **Stimulation with BCG leads to the expansion of specific clonotypes post BCG**

273   **vaccination**

274   To determine whether BCG vaccination leads to the specific expansion of BCG-reactive

275   clonotypes we performed TCR Sequencing on BCG naïve individuals pre, and 8 months post

276   BCG vaccination. To expand antigen-specific T cells, PBMCs were stimulated with MTB300, a

277   tetanus pool, and BCG for 14 days *in vitro*. Cultures were harvested and DNA was then purified

278   for TCR sequencing using the ImmunoSEQ service from Adaptive Biotechnologies

279   (**Supplementary Table 4**). Every sample had a culture replicate and the ex vivo repertoire of

280   CD4 T cells was utilized as a comparison. We assessed the productive repertoire of each

281   sample, i.e. the unique in-frame rearrangements that do not contain a stop codon, as well as the

282   frequency of these productive clonotypes. We identified a similar number of clonotypes covering

283   eighty percent of the productive repertoire across both replicates across all donor samples in

284   pre- and post- vaccination samples, in all three stimulation conditions (**Figure 7a**). In each

285   stimulated sample there was a selection and expansion of stimuli-specific clonotypes, resulting

286   in fewer unique rearrangements compared to the ex vivo CD4 sample (**Figure 7a**).

287   Next, we compared the productive repertoire of samples stimulated with the peptide

288   pools to the corresponding ex vivo CD4 samples, and identified clonotypes that were

289   significantly perturbed as a result of BCG (**Supplementary Figure 4**), and MTB300 and

290   Tetanus stimulation (**Supplementary Figure 5**). A similar number of clonotypes were

291   significantly expanded upon BCG stimulation across all donors, pre-vaccination

292   (**Supplementary Table 5**), and this number was increased post-vaccination (**Supplementary**

293   **table 5**). To identify reproducible clonotypes, we only retained those clonotypes that expanded

294   significantly across both replicates (-log$_2$ OR>1 and FDR p-val<0.05 in both replicates) for an

295   individual donor within each stimulation condition. These reproducible clonotypes within a

296   stimulation condition were overlapped across all three stimulation conditions, to obtain

297   clonotypes that expanded upon BCG stimulation only, MTB300 stimulation only, and clonotypes

298   expanded in both BCG and MTB300 stimulation. Any clonotype expanded upon tetanus

299   stimulation was excluded from further analysis, to remove any non-specific bystander effects

300   associated with the *in vitro* culture (**Figure 7b**). This analysis was performed independently in

301   pre- and post-vaccination samples.

302   To determine if the clonotypes that expanded post-vaccination corresponded to an

303   increased immune response or simply a result of stimulation, we compared the expanded

304   clonotypes, pre- and post- vaccination, based on the exact clonotype rearrangement sequence.

305   This indicated that although there was an overlap between clonotypes expanded pre- and post-

306   vaccination, this was a very small proportion, and majority of the clonotypes expanded post-

307   vaccination were unique (**Figure 7c**). This suggests that the unique clonotypes that expanded

308   post-vaccination are a result of vaccination, and not stimulation with either the BCG or MTB300

309   peptide pools. As this overlap was based on the exact clonotype rearrangement sequence, we

310   also performed GLIPH analysis, which clusters TCRs based on similar MHC-restricted peptide

311 antigen binding. The GLIPH analysis revealed a small overlap between pre- and post-

312 vaccination clonotypes (**Supplementary Table 6**), further indicating that the expanded

313 clonotypes post-vaccination are a direct result of the vaccination.

314 To further examine the homogeneity of TCR repertoires before and after BCG

315 vaccination, we used TCRMatch [31] to calculate similarity scores among the CDR3β sequences

316 obtained from the subjects. CDR3β sequences obtained from each subject were separated into

317 four groups depending on vaccination status at the time of sample collection and expansion

318 status in response to BCG stimulation *in vitro*. Scores were computed for pairs of CDR3β

319 sequences within each group, and the fraction of scores exceeding thresholds of 0.84, 0.90, and

320 0.97 were computed (**Supplementary Figure 6a-c**). The fraction of scores exceeding the

321 threshold was found to be significantly higher in expanded TCRs compared to unexpanded

322 TCRs post-BCG ($p=0.016$ for 0.84, 0.0029 for 0.90, and $p<0.0001$ for 0.97). We also found a

323 significant difference between expanded vs. unexpanded TCRs pre-BCG for the lowest

324 threshold ($p=0.013$ for 0.84) and in the fraction of scores over 0.97 ($p=0.0102$), which was

325 driven by high similarity scores in one subject. To determine whether repertoire homogeneity

326 was detectable between different subjects, we used TCRMatch to compare CDR3β sequences

327 expanded pre- and post-BCG across subjects (**Supplementary Figure 6d**). The fraction of

328 scores exceeding 0.97 for each cross-subject comparison, separated by BCG vaccination

329 status, revealed that subjects post-BCG had a higher fraction of those scores (14 matches

330 above 0.97 resulting from 9 cross-subject comparisons of post-BCG expanded TCRs, compared

331 to zero matches of pre-BCG expanded TCRs). Of the 14 strong matches between subjects, 7

332 were found to be exact matches.

333 We examined the subjects with at least 1 strong match further by assessing similarity

334 between their HLA alleles. All subjects were HLA typed (**Supplementary Table 7**), and

335 therefore the number of shared HLA alleles vs. the fraction of TCRMatch scores above 0.97

336   could be compared. We found a positive correlation between the number of shared HLA class II

337   alleles and the fraction of TCRMatch scores exceeding 0.97 (**Supplementary Fig. 6e**). The 14

338   strong matches were also analyzed for evidence of concordance in gene usage. For the 7

339   identical CDR3β sequences there was an exact match between subjects in all available VDJ

340   and allele calls. For all the strong matches, 13/14 (93%) were determined to have the same V

341   family, that ranged between 2.3-12.4% in prevalence across the subjects.

342   Finally, the magnitude of expansion, represented by the sum of the number of templates

343   identified for each clonotype (total templates), of the clonotypes pre- and post-vaccination

344   indicated a much greater expansion post-vaccination (**Figure 7c,d).** This significant increase in

345   the degree of expansion was only evident upon stimulation with BCG (in BCG only, and in

346   shared BCG and MTB300), and not upon stimulation with MTB300 only (**Figure 7d**). Taken

347   together, these results indicate both a quantitative and qualitative change in clonotypes post-

348   vaccination. There were new clonotypes observed and an overall expansion of clonotypes post-

349   vaccination upon stimulation with BCG.

350

351 **DISCUSSION**

352   Here, we employed a systems biology approach to characterize the T cell subsets,

353 cytokine secretory profiles, and other cell subset markers that change following BCG

354 vaccination to yield insight into potential new correlates of protection. Specifically, total PBMC

355 and purified CD4+ T cell population transcriptomics were used in combination with DNA

356 methylome analysis and other immunological techniques, such as flow cytometry and

357 fluorospot, to characterize BCG-induced responses in a longitudinal adult cohort who provided

358 blood samples 1-2 weeks prior to and 8 months following BCG vaccination. Gaps in knowledge

359 surrounding BCG-induced immunity, both in terms of antigen-specificity and functional

360 responses, particularly the part of the response associated with protection against *Mtb* that

361 should be boosted by a vaccine [32-34], impede efforts to improve variable BCG efficacy. Few

362 studies have used multi-omics approaches in the context of BCG vaccination and were

363 conducted mostly in animals. Cortes et al. conducted a transcriptomic analysis in mice

364 comparing BCG vaccinated and naïve mice before and after *M. bovis* challenge and found

365 Th17-associated cytokines correlated with protection [35]. Darrah et al. used a combination of

366 immunohistochemistry, flow cytometry and single cell sequencing to characterize the immune

367 response in macaques after intravenous administration of BCG, a route that was largely

368 successful as 9/10 vaccinated animals were protected even after challenge with virulent *Mtb* 6

369 months post vaccination [4]. Similarly, Hoft et al. used systems immunology to show that oral and

370 mucosal BCG delivery induced distinct molecular signatures, which could potentially permit the

371 identification of genes that should be differentially targeted by vaccines geared toward inducing

372 optimal systemic or mucosal TB immunity [36]. A transcriptional signature for BCG will serve as an

373 important comparator for novel vaccination strategies, facilitating their design and evaluation.

374   Our results reaffirm the critical role of T cells, particularly CD4 T cells, in mediating

375 antimycobacterial immunity [37], both in the context of *Mtb* infection and vaccination [20,38-40]. The

376 response at the mRNA level observed from post-BCG CD4 memory T cells was very similar to

377   those in post-BCG PBMCs, suggesting that the majority of the BCG-induced gene expression

378   changes observed in PBMCs are driven primarily by CD4 memory T cells upon BCG

379   vaccination. Other studies have found that BCG induces higher CD4 T cell responses than CD8

380   responses [22,41], and with increased levels of BCG-reactive cells post-vaccination [21,22].

381   We further characterized the particular cell subset within the CD4 compartment that

382   mediates BCG-induced responses through cell surface phenotypes and gene signatures and

383   found that it is associated with Th1* (CXCR3+CCR6+CCR4-) cells. Our group previously

384   showed that this CD4+ Th subset contained the majority of *Mtb*- [42] and NTM-specific T cells [27].

385   Moreover, we later showed that Th1* was involved in the immune response following natural

386   *Mtb* infection, since Th1* was increased in individuals with LTBI compared to TB negative

387   controls [25]. Here, we found an increase of CCR6+ CD4 T cells (including Th1* and Th17

388   subsets) and CD3+MR1+ MAITs following BCG vaccination. These frequency changes could

389   also be recapitulated *in silico* in the RNA-seq data as detected by an overall increase in

390   previously characterized MAIT and Th1* cell subset signatures [25,26] post vaccination. The

391   increase in the frequency of MAIT cells observed here is in contrast to what was recently found

392   in infants following primary BCG vaccination and following BCG revaccination in tuberculin skin

393   test positive adults (consistent with prior *Mtb* infection) [43], which are both different from our

394   cohort of primary vaccinated adults. Providing further evidence for the importance of the CCR6+

395   CD4 T cells, we also found increases in IL-23 mediated and IL-17 signaling pathways following

396   BCG stimulation post vaccination both in PBMC and CD4 memory T cells. In addition, we found

397   an increase in the magnitude of BCG-induced CCR6+ associated IFNγ and IL-17 production in

398   CD4+ T cells following BCG vaccination.

399   Several studies suggest an important role for these antigen-specific Th1* cells in the

400   immune response against mycobacteria and they may be a promising candidate for a correlate

401   of protection against *Mtb*. Th1* cells also share characteristics with a CCR6+ CD4 cell subset

402   recently described as preferentially enriched in a cohort of TB non-progressors compared to

403   those who progressed to active TB [44]. In non-human primates (NHP), antigen-specific CD4 T

404   cells in the PBMCs and bronchoalveolar lavage fluid of rhesus macaques that received BCG

405   intravenously had a similar Th1/Th17 phenotype and importantly the majority of these animals

406   (9/10) were protected from *Mtb* challenge 6 months post-BCG [4]. This Th1*-like cell subset has

407   also been shown to be associated with protection against *Mtb* in other NHP studies [45,46]. Given

408   the high diversity in results and outcomes across studies and models, a correlate of vaccine-

409   induced protection will most likely not be a single marker, which makes it important to study

410   additional cell markers such as activation, migration and memory markers, as well as functional

411   secretory profiles in order to separate immunopathology from protective antigen-specific T cell

412   responses that can serve as correlates of protection [47-49].

413   In the present study, we also found ligands for CXCR3, CXCL10 (IP-10) and CXCL11

414   that exhibited a higher fold-change post-vaccination. Antigen-specific Th1* CD4 T cells express

415   the tissue homing chemokine receptor CXCR3 and could thus respond to these ligands. It has

416   been proposed that post-vaccination measurement of multifunctional responses in *Mtb*-specific,

417   relatively undifferentiated, memory T cell subsets retaining the capacity to traffic to the lung may

418   be more indicative of protective immunity against TB [50]. Murine models have implicated

419   migration markers, including CXCR3 [51,52], that confer the ability of CD4 T cells to exit the

420   circulation and enter the lung to interact with *Mtb*-infected APCs, as promising correlates of

421   protection candidates.

422   Reactivity against our *Mtb*-derived peptide pool defined in healthy *Mtb* infected interferon

423   gamma release assay (IGRA) positive individuals, MTB300, was also boosted post-BCG

424   vaccination, however to a lesser degree than BCG-induced immune responses. MTB300

425   contains 255 peptides out of a total of 300 that are homologous between *Mtb* and BCG. This

426   result provides further evidence that the antigen-specific human T cell responses triggered by

427   BCG vaccination are not fully understood. Determination of BCG-specific epitopes and antigens

428   will provide crucial immune monitoring reagents.

429   Most genes upregulated post-BCG vaccination also tend to be upregulated, albeit to a

430   lesser extent, pre-vaccination. This suggests a boosting of the response that is present pre-

431   vaccination. The BCG-specific response is probably heavily influenced by exposure to

432   mycobacteria that the immune system has been primed with in the past. Moreover, BCG

433   vaccination resulted in an overall increase in the BCG-induced response, but DEGs and BCG-

434   specific immune responses were also identified pre-vaccination following BCG stimulation. This

435   can be, at least partly, explained by previous NTM exposure resulting in cross-reactive immune

436   responses [27]. The signatures we have identified here can be compared to those found in

437   children that were vaccinated at birth and following BCG revaccination, when available.

438   Especially given the protection from severe TB observed in children and lack of efficacy in

439   adults mediated by BCG.

440   BCG vaccination resulted in expanded TCR repertoires with higher intra-repertoire

441   homogeneity than unexpanded TCRs, as well as similar TCR repertoires across different

442   subjects. Thus, the TCRs appear to converge towards similar TCR sequences more than they

443   diversify following BCG vaccination. Clustered TCRs have been described previously in the

444   responses to herpesviruses and *Mtb* [53,54]. These similar TCR repertoires may be influenced by

445   similarities in HLA alleles. Future studies can reveal how certain HLA alleles influence BCG

446   vaccine efficacy and whether this can be used as a means of predicting efficacy.

447   In conclusion, this study provides a detailed characterization of BCG-induced immune

448   responses and TCR clonotypes in adults that were vaccinated with BCG. These findings inform

449   our understanding of the immune response induced by the BCG vaccine and provides means to

450   track longitudinal changes in the specific T cells in many different settings.

451

452 **MATERIALS AND METHODS**

453 **Ethics statement**

454 All participants provided written informed consent for participation in the study. Ethical

455 approval was obtained from the Institutional review boards at Linköping University (2015/150-

456 32) and the La Jolla Institute for Immunology (VD-140).

457 **Study subjects**

458 We recruited 17 TB negative and BCG naïve individuals for participation in the study who

459 were offered the BCG vaccine through their medical school at Linköping University. Their *Mtb*

460 infection status was confirmed by a negative tuberculin skin test and IFNγ-release assay (IGRA;

461 T.Spot-TB, Oxford Immunotec).

462 Venous blood was collected in heparin-containing blood bags 1-2 weeks prior to BCG

463 vaccination and 8 months after the vaccination. Peripheral blood mononuclear cells (PBMC)

464 were purified from whole blood by density-gradient centrifugation (Ficoll-Hypaque, Amersham

465 Biosciences), according to the manufacturer's instructions. Cells were cryopreserved in liquid

466 nitrogen suspended in FBS (Gemini Bio-Products) containing 10% (vol/vol) DMSO (Sigma).

467 Cryopreserved cells were shipped from Linköping University to LJI for analysis.

468 **HLA typing**

469 Participants were HLA typed by an ASHI-accredited laboratory at Murdoch University

470 (Institute for Immunology & Infectious Diseases, Western Australia) as previously described [55].

471 HLA typing for class I (HLA A, B, C) and class II (DQA1, DQB1, DRB1, 3, 4, 5, DPB1) was

472 performed using locus-specific PCR amplification of genomic DNA. Patient-specific, barcoded

473 primers were used for amplification. Amplified products were quantitated and pooled by subject

474 and up to 48 subjects were pooled. An indexed (8 indexed MiSeq runs) library was then

475 quantitated using Kappa universal QPCR library quantification kits. Sequencing was performed

476 using an Illumina MiSeq using 2x300 paired-end chemistry. Reads were quality-filtered and

477 passed through a proprietary allele calling algorithm and analysis pipeline using the latest IMGT

478 HLA allele database as a reference. The algorithm was developed by E.J.P. and S.A.M. and

479 relies on periodically updated versions of the freely available international immunogenetics

480 information system (http://www.imgt.org) and an ASHI-accredited HLA allele caller software

481 pipeline, IIID HLA analysis suite (http://www.iiid.com.au/laboratory-testing/). The HLA type of

482 each subject is listed in **Supplementary table 7**.

483 **Peptides and other stimuli**

484 Peptides were synthesized as crude material on a small (1mg) scale by A&A, LLC (San

485 Diego, CA). Multi-epitope peptide pools ("megapools") were prepared as previously described

486 [56]. Individual peptides were resuspended in DMSO, and equal amounts of each peptide were

487 pooled to contruct the peptide pool. After lyophilization, the peptide pools were resuspended in

488 DMSO, aliquoted, and stored at -20°C. Two different peptide pools were used; a peptide pool

489 containing 300 Mtb-derived 15-20-mer peptides (MTB300) primarily HLA class II restricted [56],

490 and a peptide pool with 125 peptides derived from the *Clostridium tetani* toxin (TT) sequence [57].

491 In addition to peptide pool, PBMCs were also stimulated with *M. bovis* BCG-Danish or

492 Pasteur at 100 µg/ml.

493 **Fluorospot assay**

494 Antigen-specific cellular responses were measured by IFNγ Fluorospot assay with all

495 antibodies and reagents from Mabtech (Nacka Strand, Sweden). Plates were coated overnight

496 at 4°C with a mouse anti-human IFNγ (clone 1-D1K) antibody. Briefly, $2x10^5$ cells were added to

497 each well of pre-coated Immobilon-FL PVDF 96-well plates (Mabtech) in the presence of 2

498 µg/ml peptide pool, or 100µh/ml BCG, and incubated at 37°C in humidified CO_2 incubator for 20-

499 24 hrs. Cells stimulated with DMSO (corresponding to the percent DMSO in the peptide pools)

500 were used to assess non-specific/background cytokine production and PHA stimulation at 10

501 µg/ml was used as a positive control. All conditions were tested in triplicates. Fluorospot plates

502 were developed according to manufacturer's instructions (Mabtech). Briefly, cells were removed

503   and plates were washed 6 times with 200 µl PBS/0.05% Tween 20 using an automated plate

504   washer. After washing, 100 µl of antibody mixture containing anti-IFNγ (7-B6-1-FS-FITC)

505   prepared in PBS with 0.1% BSA was added to each well and plates were incubated for 2 hrs at

506   room temperature. Plates were again washed 6 times with 200 µl PBS/0.05% Tween 20 using

507   an automated plate washer and incubated with diluted fluorophores (anti-BAM-490) for 1 hr at

508   room temperature. Finally, plates were once more washed 6 times with 200 µl PBS/0.05%

509   Tween 20 using an automated plate washer and incubated with fluorescence enhancer for 15

510   mins at room temperature. The plates were blotted dry and spots were counted by computer-

511   assisted image analysis (AID iSpot, Aid Diagnostica GMBH, Strassberg, Germany). Responses

512   were considered positive if the net spot-forming cells (SFC) per 10^6 PBMC were ≥20, the

513   stimulation index ≥2, and p≤0.05 by Student's t-test or Poisson distribution test.

514   **Flow cytometry**

515   Several different flow cytometry panels were used. Cryopreserved PBMCs were thawed in

516   RPMI supplemented with 5% human serum (Gemini Bio-Products, West Sacramento, CA), 1%

517   Glutamax (Gibco,Waltham,MA), 1% penicillin/streptomycin (Omega Scientific, Tarzana, CA),

518   and 50U/ml Benzonase (Millipore Sigma, Burlington, MA). Cells were then washed and counted.

519   1 million cells were then blocked in 10% FBS for 10 mins at 4°C. After blocking, cells were

520   stained with APCef780 conjugated anti-CD4 (clone RPA-T4, eBiosciences), AF700 conjugated

521   anti-CD3 (UCHT1, BD Pharmigen), BV650 conjugated anti-CD8a (RPA-T8, Biolegend), PECy7

522   conjugated anti-CD19 (HIB19, TONBO), APC conjugated anti-CD14 (61D3, TONBO),

523   PerCPCy5.5 conjugated anti-CCR7 (G043H7, Biolegend), PE conjugated anti-CD56 (CMSSB,

524   eBiosciences), FITC conjugated anti-CD25 (M-A251, BD Pharmigen), eF450 conjugated anti-

525   CD45RA (HI100, eBiosciences) and fixable viability dye eF506 (eBiosciences) for 30 mins at

526   4°C. Cells were then washed twice and acquired on a BD FACSAria flow cytometer (BD

527   Biosciences, San Jose, CA) to measure the frequency of different cell subsets. The gating

528   strategy for this panel was performed as previously reported [24].

529   Cells were also stained with BV650 conjugated CCR6 (G034E3, BioLegend), CXCR3-APC

530   (1C6/CXCR3, BD Biosciences) for 20 min at 37°C, followed by CCR4-PE-Cy7 (1G1,

531   BDBiosciences), CCR7-PerCPCy5.5 (UCHL1, BioLegend), CD4-APCef780 (RPA-T4,

532   eBiosciences), CD3-AF700 (UCHT1, BD Pharmigen) CD45RA-eF450 (HI100, eBiosciences),

533   CD8-V500 (RPA-T8, BD Biosciences), CD14-V500 (M5E2; BD Biosciences), CD19-V500

534   (HIB19, BD Biosciences), and fixable viability dye eF506 (eBiosciences) at room temperature

535   for 30 min. Gating strategy is shown in **Supplementary Figure 7a**.

536   For non-conventional T cells PBMCs were stained with 1:100 MR1 5-OP-RU or 6-FP (as a

537   control) tetramer for 40 min at room temperature. The MR1 tetramer technology was developed

538   jointly by Dr. J. McCluskey, Dr. J. Rossjohn, and Dr. D. Fairlie [58], and the material was produced

539   by the National Institutes of Health Tetramer Core Facility, as permitted to be distributed by the

540   University of Melbourne. After 40 min, cells were also stained with fixable viability dye eF506

541   (eBiosciences) and with CD3-AF700 (UCHT1, BD Pharmigen), CD4-APCef780 (RPA-T4,

542   eBiosciences), CD8-BV650 (RPA-T8; BioLegend), CD14-V500 (M5E2; BD Biosciences), CD19-

543   V500 (HIB19, BD Biosciences), CD161-APC (HP-3G10; eBiosciences), Vα7.2-PE-Cy7 (3C10,

544   BioLegend), Vα24-PE-Dazzle594 (6B11, BioLegend), and $\gamma\delta$PAN TCR-FITC (11F2, BD

545   Biosciences) for 30 min at room temperature. The gating strategy is shown in **Supplementary**

546   **Figure 7b**. MR1+ T cells were defined as 5-OP-RU MR1 tetramer+, and CD4-, Vα24-, $\gamma\delta$PAN

547   TCR-, Vα7.2+, and CD161+ through Boolean gating.

548   For the flow cytometry measurement of IFNγ and IL-17, PBMCs were thawed and stimulated

549   with 100µg/ml BCG or left unstimulated in the presence of 1µg/ml anti-CD28 (CD28.2

550   eBioscience), and 1µg/ml anti-CD49d (9F10, BioLegend). Cells were incubated at 37°C for 5

551   hours, after which 2.5µg/ml Brefeldin A and monensin was added for another 7 hours. Cells

552   were washed and blocked in 10% FBS for 10 mins at 4°C. Cells were then stained with CD4-

553   APCef780 (RPA-T4, eBiosciences), CD3-AF700 (UCHT1, BD Pharmigen), CD8-BV650 (RPA-

554   T8; BioLegend), CD14-V500 (M5E2; BD Biosciences), CD19-V500 (HIB19, BD Biosciences),

555   fixable viability dye eF506 (eBiosciences) for 30 min at 4°C. Cells were washed twice and then

556   fixed in 4% paraformaldehyde solution for 10 min at 4°C. Saponin buffer was used to

557   permeabilize the cells by incubating them at room temperature for 10 min, followed by blocking

558   in 10% FBS for 5 min at 4°C. Cells were then stained with IFNγ-FITC (4S.B3, eBioscience) and

559   IL-17-PE-Cy7 (eBio64DEC17, eBioscience) for 30 min at room temperature. Followed by

560   washes and acquisition. The gating strategy is shown in **Supplemental figure 7c**.

561   **Fluorescence-activated cell sorting**

562   PBMCs were thawed and $2x10^6$ cells were added per well in a 96 round bottom well plate.

563   Cells were stimulated with BCG (50µg/ml), MTB300 (2µg/ml), Tetanus pool (2µg/ml), or DMSO

564   (corresponding to the percent DMSO in the peptide pools) as a control. Anti-human CD28 and

565   CD3 (1 µg/ml) was used as a positive control. The wells for the positive control was pre-coated

566   overnight at 4°C. Cells were incubated at 37°C for 24 h.

567   The following day, cells were washed and incubated in PBS with 10% FBS at 4°C for 10

568   min. They were stained with fixable viability dye eFluor 506 (eBioscience), and an antibody

569   cocktail containing anti-human CD3-Alexa Fluor 700 (UCHT1, BD Bioscience), CD4-APCeFluor

570   780 (RPA-T4, eBioscience), CD8-V500 (RPA-T8, BD Biosciences), CD45RA-eFluor 450

571   (HI100, eBioscience), and CCR7-PerCPCy5.5 (UCHL1, BioLegend) for 20 min at room

572   temperature. Cells were transferred into a 5 ml polypropylene FACS tube (BD Biosciences) and

573   PBMCs (excluding doublets) and CD4 memory T cells were sorted on a FACSAria III cell sorter

574   (Becton Dickinson) into QIAzol Lysis Reagent (QIAGEN). A total of 100,000 cells was sorted per

575   sample. For gating strategy see **Supplementary figure 7d**. Sorted cell populations were stored

576   in QIAzol Lysis Reagent at -80°C until RNA extraction.

577   CD4 T cell isolation

578   CD4+ T cells were isolated from at least $5x10^6$ PBMCs on the day of thaw by negative

579   selection using the CD4+ T cell isolation kit II (Miltenyi Biotec, Bergisch Gladbach, Germany)

580   according to manufacturer's instructions. The isolated CD4+ T cells were washed, pelleted, and

581   then stored at -80°C until DNA extraction.

582   Cell expansion for TCR sequencing

583   For *in vitro* expansion, cryopreserved PBMCs were thawed in RPMI supplemented with 5%

584   human serum (Gemini Bio-Products, West Sacramento, CA), 1% Glutamax (Gibco, Waltham,

585   MA), 1% penicillin/streptomycin (Omega Scientific, Tarzana, CA), and 50 U/ml Benzonase

586   (Millipore Sigma, Burlington, MA). The cells were then washed and viability was evaluated using

587   trypan blue dye exclusion. Briefly, at a density of $2x10^6$ cells per mL, the cells were plated in a

588   well of a 24-well plate in the presence of 50μg/ml BCG, 2μg/ml MTB300, or 2μg/ml tetanus

589   peptide pool, and were incubated in a 37°C humidified CO_2 incubator for 14 days. Every 3-4

590   days cells were supplied with 10 U/ml recombinant human IL-2. After 14 days of culture, cells

591   were harvested, counted, and pelleted. The cell pellets were stored at -80°C until DNA

592   extraction.

593   RNA sequencing

594   RNA sequencing was performed as described previously (24). Briefly, total RNA was

595   purified using an miRNeasy Micro Kit (QIAGEN) and quantified by quantitative PCR, as

596   described previously (25). Purified total RNA (1–5ng) was amplified following the Smart-Seq2

597   protocol (16 cycles of cDNA amplification) (26). cDNA was purified using AMPure XP beads

598   (Beckman Coulter). From this step, 1 ng of cDNA was used to prepare a standard Nextera XT

599   sequencing library (Nextera XT DNA sample preparation kit and index kit, Illumina). Whole-

600   transcriptome amplification and sequencing library preparations were performed in a 96-well

601   format to reduce assay-to-assay variability. Quality-control steps were included to determine

602 total RNA quality and quantity, the optimal number of PCR preamplification cycles, and

603 fragment size selection. Samples that failed quality control were eliminated from further

604 downstream steps. Barcoded Illumina sequencing libraries (Nextera; Illumina) were generated

605 using the automated platform (Biomek FXp). Libraries were sequenced on a NovaSeq 6000

606 Illumina platform to obtain 50-bp paired end reads (TruSeq Rapid kit; Illumina).

607 **TCR sequencing**

608 DNA was extracted from the cultured cells or ex vivo CD4$^+$ T cell samples using DNeasy

609 Blood and Tissue kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

610 Samples were sent to Adaptive Biotechnologies (Seattle, WA) for TCRB sequencing according

611 to their protocol. The ex vivo CD4$^+$ T cell samples were sequenced with "deep resolution" to

612 cover a maximum number of clonotypes in the repertoire. Samples that were stimulated with

613 peptide pools for 14 days and then harvested were sequenced with "survey resolution".

614 **Data analysis – RNA sequencing**

615 Paired-end reads that passed Illumina filters were filtered for reads aligning to tRNA, rRNA,

616 adapter sequences, and spike-in controls. The reads were aligned to the GRCh38 reference

617 genome and Gencode v27 annotations using TopHat v1.4.1 [59]. DUST scores were calculated

618 with PRINSEQ Lite v0.20.3 [60] and low-complexity reads (DUST > 4) were removed from BAM

619 files. The alignment results were parsed via SAMtools [61] to generate SAM files. Read counts to

620 each genomic feature were obtained with the htseq-count program v0.7.1 [62] using the "union"

621 option. Raw counts were imported into R v3.6.1 where they were subset into PBMC and CD4

622 memory, both containing unstimulated, MTB300 peptide pool, and BCG epitope pool stimulated

623 samples. The following steps were performed on PBMC and CD4 memory subsets

624 independently. R/Bioconductor package DESeq2 v.1.24.0 [63] was used to normalize raw counts.

625 Variance stabilizing transformation was applied to normalized counts to obtain $\log_2$ gene

626 expression values. Quality control was performed using boxplots and Principal component

627 analysis (PCA), using the 'prcomp' function in R, on $\log_2$ expression values. Differentially

628 expressed genes were identified using the DESeq2 Wald test, and p-values were adjusted for

629 multiple test correction using the Benjamini Hochberg algorithm [64] (**Supplementary Table 1**).

630 Genes with adjusted p values < 0.05 and log2 fold change > 1 or < -1 were considered

631 differentially expressed. Pathway enrichment analysis was performed using ToppGene Suite:

632 ToppFun [65] (**Supplementary Table 2**), and cell type enrichment was performed using DICE [66].

633 **Data analysis – TCR sequencing**

634 Pre-processing and quality control of the raw data was performed using the immunoSEQ

635 analyzer (Adaptive Biotechnologies, Inc.). Measurement metrics of processed data were

636 exported in the tsv file format and downstream data analysis was performed in Python v3.7.2

637 and in R v3.6.1. To identify clonotypes that were expanded in culture or after vaccine, each of

638 the replicates per donor was compared to the corresponding ex vivo CD4 sample, and p-values

639 and odds-ratio were calculated using a two-sided Fisher exact test, using the 'fisher_exact'

640 function in the SciPy.Stats v1.4.1 [67] and NumPy v1.17.2 [68] extensions of python. Clonotypes

641 appearing in both replicates with $-\log_2$ odds ratios (OR) >1 or <-1 and false discovery rate (FDR)

642 p-value < 0.05 corrected for multiple testing using the Benjamini–Hochberg method [64],

643 calculated using the 'fdrcorrection' function from the statsmodels module v0.9.0 [69] for Python,

644 were considered significant (**Supplementary Table 5**). For visualization purposes, all $-\log_{10}$

645 FDR p-values > 50 were set to 50. Sequence similarity by clustering was performed using

646 GLIPH v1.0 with default parameters, to identify conserved motifs and global similarity of

647 complementarity-determining region 3 (CDR3) sequences [53] (**Supplementary Table 6**).

648 We also used TCRMatch to determine sequence similarity [31]. Within each group, defined by

649 a single subject, vaccination and expansion status, and CDR3β sequences were tested against

650 each other. For groups with expanded sequences, all CDR3β were analyzed against each

651 other. For groups with unexpanded TCRs, the number of CDR3β sequences were much greater

652 than the number of expanded CDR3β sequences in the same individual, so random sampling

653   was performed. A sample of n sequences, with n=the number of expanded CDR3β sequences

654   for the corresponding individual and vaccination status, was randomly selected and

655   subsequently run through TCRMatch to assess intragroup similarity. 100 random samplings

656   were performed for each group of unexpanded sequences. Cross-subject comparisons were

657   performed by running TCRMatch on all possible pairs of sequences between all possible pairs

658   of subjects. The number of HLA alleles shared by two subjects was calculated by comparing

659   each subjects' HLA alleles for a given HLA locus against another subjects' corresponding HLA

660   alleles. All pairwise combinations were compared for each gene, with a maximum of four shared

661   allele pair combinations. For example, if subject 1 carried alleles X and Y, and subject 2 carried

662   X and Z, the number of shared alleles was recorded as 1. Meanwhile, if subject 1 were

663   homozygous for X, and subject 2 carried X and Z, the number of shared alleles was recorded as

664   2. Finally, if both subjects were homozygous for the same allele, their number of shared allele

665   pairs was counted as 4.

666   **DNA methylome data acquisition and analyses**

667   The DNA methylome data was analyzed using the HumanMethylationEPIC (850K) array

668   (Illumina, USA) as per manufacturer's instruction. The raw IDAT files of the DNA methylome

669   data was processed using the ChAMP package [70] in R (v4.0.3) after using the default filtering

670   criteria i) removing the CpGs with detection p-value >0.01, ii) filter out CpGs with <3 beads in at

671   least 5% of samples per CpGs, iii) filter out all non-CpGs contained in the dataset, iv) removing

672   all SNP-related CpGs, v) filter out all multi-hit probes, and vi) filter out all CpGs located in X and

673   Y chromosomes. The filtered data was normalized using the β mixture quantile normalization

674   (BMIQ) function using the ChAMP package. For each CpG site, the methylation β values were

675   calculated, which represents the fraction of methylated cytosines at their particular CpG site

676   (0=unmethylated, 1=fully methylated). The Houseman algorithm [30] was used to calculate the cell

677   type deconvolution as the samples were drawn from the PBMCs for "before" and "after" samples

678   separately. We performed a Shapiro-Wilk test to test the normality of the samples and used a t-

679   test to compare differences between the before and after group for each cell type. The

680   differential methylation values (mean methylation difference, mmd) were calculated using the

681   two different groups of samples, "before" and "after". A total of 139 613 CpGs were identified as

682   differentially methylated CpGs (DMCs). A volcano plot was generated using the

683   EnhancedVolcano package of hypermethylated and hypomethylated CpGs. The differentially

684   methylated CpGs were annotated using the human genome annotation (HG38.13) to find the

685   corresponding genes (Differentially Methylated Genes, DMGs). The hyper and hypomethylated

686   CpGs were annotated with the different chromosomal locations. The hyper and hypomethylated

687   genes were compared with the up and downregulated differentially expressed genes using the

688   venneuler package [71] (in house python script). A principal component analysis (PCA) was

689   performed on the normalized β values of DNA methylome dataset using the factoMineR and

690   factoExtra (https://CRAN.R-project.org/package=factoextra) packages. All DMCs were

691   considered significant with the Benjamini-Hochberg (BH) corrected p-value < 0.15, if not stated

692   otherwise. A list of DMGs intersected with DEGs were analyzed using the clusterProfiler

693   package [72] using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

694   **Data availability**

695   The RNA-seq datasets analyzed as part of this study have been deposited in the NCBI

696   Gene Expression Omnibus (GEO) database with the primary accession number GSE156422.

697

AUTHOR CONTRIBUTIONS

ASi, PD, ASe, BP, and CSLA participated in the design and direction of the study. ASi, PD, RK, KM, WC, JD, ML, and CSLA performed and analyzed experiments. GS and PV performed the RNAseq. PD performed the bioinformatics analysis, supervised by ASi. ML recruited participants and maintained participant data. EJP and SAM coordinated and performed HLA typing. ASi, PD, KM, ML, and CSLA wrote the manuscript. All authors read, edited and approved the manuscript.

ACKNOWLEDGMENTS

COMPETING INTERESTS

The authors declare no competing interests

REFERENCES

1   Colditz, G. A. *et al.* The efficacy of bacillus Calmette-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* **96**, 29-35 (1995).

2   Fine, P. E. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* **346**, 1339-1345 (1995).

3   Trunz, B. B., Fine, P. & Dye, C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* **367**, 1173-1180, doi:10.1016/S0140-6736(06)68507-3 (2006).

4   Darrah, P. A. *et al.* Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature* **577**, 95-102, doi:10.1038/s41586-019-1817-8 (2020).

5   Manjaly Thomas, Z. R. & McShane, H. Aerosol immunisation for TB: matching route of vaccination to route of infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**, 175-181, doi:10.1093/trstmh/tru206 (2015).

6   White, A. D. *et al.* Evaluation of the Immunogenicity of Mycobacterium bovis BCG Delivered by Aerosol to the Lungs of Macaques. *Clinical and vaccine immunology : CVI* **22**, 992-1003, doi:10.1128/CVI.00289-15 (2015).

7   Price, D. N., Kusewitt, D. F., Lino, C. A., McBride, A. A. & Muttil, P. Oral Tolerance to Environmental Mycobacteria Interferes with Intradermal, but Not Pulmonary, Immunization against Tuberculosis. *PLoS Pathog* **12**, e1005614, doi:10.1371/journal.ppat.1005614 (2016).

8   Barreto, M. L. *et al.* Causes of variation in BCG vaccine efficacy: examining evidence from the BCG REVAC cluster randomized trial to explore the masking and the blocking hypotheses. *Vaccine* **32**, 3759-3764, doi:10.1016/j.vaccine.2014.05.042 (2014).

9   Dourado, I. *et al.* Rates of adverse reactions to first and second doses of BCG vaccination: results of a large community trial in Brazilian schoolchildren. *Int J Tuberc Lung Dis* **7**, 399-402 (2003).

10   Rodrigues, L. C. *et al.* Effect of BCG revaccination on incidence of tuberculosis in school-aged children in Brazil: the BCG-REVAC cluster-randomised trial. *Lancet* **366**, 1290-1295, doi:10.1016/S0140-6736(05)67145-0 (2005).

11   Bekker, L. G. *et al.* A phase 1b randomized study of the safety and immunological responses to vaccination with H4:IC31, H56:IC31, and BCG revaccination in Mycobacterium tuberculosis-uninfected adolescents in Cape Town, South Africa. *EClinicalMedicine* **21**, 100313, doi:10.1016/j.eclinm.2020.100313 (2020).

12   Nemes, E. *et al.* Prevention of M. tuberculosis Infection with H4:IC31 Vaccine or BCG Revaccination. *N Engl J Med* **379**, 138-149, doi:10.1056/NEJMoa1714021 (2018).

13   Rakshit, S. *et al.* BCG revaccination boosts adaptive polyfunctional Th1/Th17 and innate effectors in IGRA+ and IGRA- Indian adults. *JCI Insight* **4**, doi:10.1172/jci.insight.130540 (2019).

14   Nieuwenhuizen, N. E. & Kaufmann, S. H. E. Next-Generation Vaccines Based on Bacille Calmette-Guerin. *Front Immunol* **9**, 121, doi:10.3389/fimmu.2018.00121 (2018).

754   15   Scriba, T. J. *et al.* Dose-finding study of the novel tuberculosis vaccine, MVA85A, in
755   healthy BCG-vaccinated infants. *The Journal of infectious diseases* **203**, 1832-1843,
756   doi:10.1093/infdis/jir195 (2011).
757   16   McShane, H. *et al.* Recombinant modified vaccinia virus Ankara expressing antigen 85A
758   boosts BCG-primed and naturally acquired antimycobacterial immunity in humans.
759   *Nature medicine* **10**, 1240-1244, doi:10.1038/nm1128 (2004).
760   17   Nieuwenhuizen, N. E. *et al.* The Recombinant Bacille Calmette-Guerin Vaccine VPM1002:
761   Ready for Clinical Efficacy Testing. *Front Immunol* **8**, 1147,
762   doi:10.3389/fimmu.2017.01147 (2017).
763   18   Abebe, F. Is interferon-gamma the right marker for bacille Calmette–Guérin-induced
764   immune protection? The missing link in our understanding of tuberculosis immunology.
765   *Clinical & Experimental Immunology* **169**, 213-219, doi:10.1111/j.1365-
766   2249.2012.04614.x (2012).
767   19   Kagina, B. M. *et al.* Specific T cell frequency and cytokine expression profile do not
768   correlate with protection against tuberculosis after bacillus Calmette-Guerin vaccination
769   of newborns. *American journal of respiratory and critical care medicine* **182**, 1073-1079,
770   doi:201003-0334OC [pii]
771   10.1164/rccm.201003-0334OC (2010).
772   20   Fletcher, H. A. *et al.* T-cell activation is an immune correlate of risk in BCG vaccinated
773   infants. *Nature communications* **7**, 11290, doi:10.1038/ncomms11290 (2016).
774   21   Soares, A. P. *et al.* Bacillus Calmette-Guerin Vaccination of Human Newborns Induces T
775   Cells with Complex Cytokine and Phenotypic Profiles. *J Immunol* **180**, 3569-3577 (2008).
776   22   Rodo, M. J. *et al.* A comparison of antigen-specific T cell responses induced by six novel
777   tuberculosis vaccine candidates. *PLoS Pathog* **15**, e1007643,
778   doi:10.1371/journal.ppat.1007643 (2019).
779   23   Burel, J. G. *et al.* Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic
780   Immune Signatures of Tuberculosis. *Front Immunol* **10**, 221,
781   doi:10.3389/fimmu.2019.00221 (2019).
782   24   Burel, J. G. *et al.* An Integrated Workflow To Assess Technical and Biological Variability
783   of Cell Population Frequencies in Human Peripheral Blood by Flow Cytometry. *J Immunol*
784   **198**, 1748-1758, doi:10.4049/jimmunol.1601750 (2017).
785   25   Arlehamn, C. L. *et al.* Transcriptional Profile of Tuberculosis Antigen-Specific T Cells
786   Reveals Novel Multifunctional Features. *J Immunol* **193**, 2931-2940,
787   doi:10.4049/jimmunol.1401151 (2014).
788   26   Pomaznoy, M. *et al.* Quantitative and Qualitative Perturbations of CD8(+) MAITs in
789   Healthy Mycobacterium tuberculosis-Infected Individuals. *Immunohorizons* **4**, 292-307,
790   doi:10.4049/immunohorizons.2000031 (2020).
791   27   Lindestam Arlehamn, C. S. *et al.* Immunological consequences of intragenus
792   conservation of Mycobacterium tuberculosis T-cell epitopes. *Proc Natl Acad Sci U S A*
793   **112**, E147-155, doi:10.1073/pnas.1416537112 (2015).
794   28   Kleinnijenhuis, J. *et al.* Bacille Calmette-Guerin induces NOD2-dependent nonspecific
795   protection from reinfection via epigenetic reprogramming of monocytes. *Proc Natl Acad
796   Sci U S A* **109**, 17537-17542, doi:10.1073/pnas.1202870109 (2012).

797   29   Verma, D. *et al.* Anti-mycobacterial activity correlates with altered DNA methylation
798   pattern in immune cells from BCG-vaccinated subjects. *Scientific reports* **7**, 12305,
799   doi:10.1038/s41598-017-12110-2 (2017).
800   30   Houseman, E. A. *et al.* Reference-free deconvolution of DNA methylation data and
801   mediation by cell composition effects. *BMC Bioinformatics* **17**, 259, doi:10.1186/s12859-
802   016-1140-4 (2016).
803   31   Chronister, W. D. *et al.* TCRMatch: Predicting T-Cell Receptor Specificity Based on
804   Sequence Similarity to Previously Characterized Receptors. *Front Immunol* **12**, 640725,
805   doi:10.3389/fimmu.2021.640725 (2021).
806   32   Fletcher, H. A. Systems approaches to correlates of protection and progression to TB
807   disease. *Semin Immunol* **39**, 81-87, doi:10.1016/j.smim.2018.10.001 (2018).
808   33   Satti, I. & McShane, H. Current approaches toward identifying a correlate of immune
809   protection from tuberculosis. *Expert Rev Vaccines* **18**, 43-59,
810   doi:10.1080/14760584.2019.1552140 (2019).
811   34   Barker, L., Hessel, L. & Walker, B. Rational approach to selection and clinical
812   development of TB vaccine candidates. *Tuberculosis (Edinb)* **92 Suppl 1**, S25-29,
813   doi:10.1016/S1472-9792(12)70009-4 (2012).
814   35   Aranday Cortes, E., Kaveh, D., Nunez-Garcia, J., Hogarth, P. J. & Vordermeier, H. M.
815   Mycobacterium bovis-BCG vaccination induces specific pulmonary transcriptome
816   biosignatures in mice. *PLoS One* **5**, e11319, doi:10.1371/journal.pone.0011319 (2010).
817   36   Hoft, D. F. *et al.* PO and ID BCG vaccination in humans induce distinct mucosal and
818   systemic immune responses and CD4(+) T cell transcriptomal molecular signatures.
819   *Mucosal Immunol* **11**, 486-495, doi:10.1038/mi.2017.67 (2018).
820   37   Morgan, J. *et al.* Classical CD4 T cells as the cornerstone of antimycobacterial immunity.
821   *Immunol Rev* **301**, 10-29, doi:10.1111/imr.12963 (2021).
822   38   Caccamo, N. *et al.* Multifunctional CD4(+) T cells correlate with active Mycobacterium
823   tuberculosis infection. *Eur J Immunol* **40**, 2211-2220, doi:10.1002/eji.201040455 (2010).
824   39   Qiu, Z. *et al.* Multifunctional CD4 T cell responses in patients with active tuberculosis.
825   *Scientific reports* **2**, 216, doi:10.1038/srep00216 (2012).
826   40   Smith, S. G., Zelmer, A., Blitz, R., Fletcher, H. A. & Dockrell, H. M. Polyfunctional CD4 T-
827   cells correlate with in vitro mycobacterial growth inhibition following Mycobacterium
828   bovis BCG-vaccination of infants. *Vaccine* **34**, 5298-5305,
829   doi:10.1016/j.vaccine.2016.09.002 (2016).
830   41   Murray, R. A. *et al.* Bacillus Calmette Guerin vaccination of human newborns induces a
831   specific, functional CD8+ T cell response. *J Immunol* **177**, 5647-5651,
832   doi:10.4049/jimmunol.177.8.5647 (2006).
833   42   Lindestam Arlehamn, C. S. *et al.* Memory T cells in latent Mycobacterium tuberculosis
834   infection are directed against three antigenic islands and largely contained in a
835   CXCR3+CCR6+ Th1 subset. *PLoS Pathog* **9**, e1003130, doi:10.1371/journal.ppat.1003130
836   (2013).
837   43   Gela, A. *et al.* Effects of BCG vaccination on donor unrestricted T cells in humans.
838   *bioRxiv*, 2021.2004.2029.441927, doi:10.1101/2021.04.29.441927 (2021).

839 44 Nathan, A. *et al.* Multimodally profiling memory T cells from a tuberculosis cohort
840   identifies cell state associations with demographics, environment and disease. *Nat*
841   *Immunol* **22**, 781-793, doi:10.1038/s41590-021-00933-1 (2021).
842 45 Dijkman, K. *et al.* Prevention of tuberculosis infection and disease by local BCG in
843   repeatedly exposed rhesus macaques. *Nat Med* **25**, 255-262, doi:10.1038/s41591-018-
844   0319-9 (2019).
845 46 Cadena, A. M. *et al.* Concurrent infection with Mycobacterium tuberculosis confers
846   robust protection against secondary infection in macaques. *PLoS Pathog* **14**, e1007305,
847   doi:10.1371/journal.ppat.1007305 (2018).
848 47 Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M. & Lugli, E. The who's who of T-
849   cell differentiation: human memory T-cell subsets. *Eur J Immunol* **43**, 2797-2809,
850   doi:10.1002/eji.201343751 (2013).
851 48 Sutherland, J. S., Adetifa, I. M., Hill, P. C., Adegbola, R. A. & Ota, M. O. C. Pattern and
852   diversity of cytokine production differentiates between Mycobacterium tuberculosis
853   infection and disease. *Eur J Immunol* **39**, 723-729, doi:10.1002/eji.200838693 (2009).
854 49 Somoskovi, A. & Salfinger, M. Nontuberculous mycobacteria in respiratory infections:
855   advances in diagnosis and identification. *Clinics in laboratory medicine* **34**, 271-295,
856   doi:10.1016/j.cll.2014.03.001 (2014).
857 50 Srivastava, S. & Ernst, J. D. Cutting edge: Direct recognition of infected cells by CD4 T
858   cells is required for control of intracellular Mycobacterium tuberculosis in vivo. *J*
859   *Immunol* **191**, 1016-1020, doi:10.4049/jimmunol.1301236 (2013).
860 51 Anderson, K. G. *et al.* Intravascular staining for discrimination of vascular and tissue
861   leukocytes. *Nature protocols* **9**, 209-222, doi:10.1038/nprot.2014.005 (2014).
862 52 Sakai, S. *et al.* Cutting edge: control of Mycobacterium tuberculosis infection by a subset
863   of lung parenchyma-homing CD4 T cells. *J Immunol* **192**, 2965-2969,
864   doi:10.4049/jimmunol.1400019 (2014).
865 53 Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature*
866   **547**, 94-98, doi:10.1038/nature22976 (2017).
867 54 Song, I. *et al.* Broad TCR repertoire and diverse structural solutions for recognition of an
868   immunodominant CD8(+) T cell epitope. *Nat Struct Mol Biol* **24**, 395-406,
869   doi:10.1038/nsmb.3383 (2017).
870 55 Sulzer, D. *et al.* T cells from patients with Parkinson's disease recognize alpha-synuclein
871   peptides. *Nature* **546**, 656-661, doi:10.1038/nature22815 (2017).
872 56 Lindestam Arlehamn, C. S. *et al.* A Quantitative Analysis of Complexity of Human
873   Pathogen-Specific CD4 T Cell Responses in Healthy M. tuberculosis Infected South
874   Africans. *PLoS Pathog* **12**, e1005760, doi:10.1371/journal.ppat.1005760 (2016).
875 57 da Silva Antunes, R. *et al.* Definition of Human Epitopes Recognized in Tetanus Toxoid
876   and Development of an Assay Strategy to Detect Ex Vivo Tetanus CD4+ T Cell Responses.
877   *PLoS One* **12**, e0169086, doi:10.1371/journal.pone.0169086 (2017).
878 58 Corbett, A. J. *et al.* T-cell activation by transitory neo-antigens derived from distinct
879   microbial pathways. *Nature* **509**, 361-365, doi:10.1038/nature13160 (2014).
880 59 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-
881   Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

882   60   Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic
883   datasets. *Bioinformatics* **27**, 863-864, doi:10.1093/bioinformatics/btr026 (2011).
884   61   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
885   2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
886   62   Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
887   throughput sequencing data. *Bioinformatics* **31**, 166-169,
888   doi:10.1093/bioinformatics/btu638 (2015).
889   63   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
890   for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8
891   (2014).
892   64   Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
893   Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
894   *(Methodological)* **57**, 289-300 (1995).
895   65   Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list
896   enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311,
897   doi:10.1093/nar/gkp427 (2009).
898   66   Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene
899   Expression. *Cell* **175**, 1701-1715 e1716, doi:10.1016/j.cell.2018.10.022 (2018).
900   67   Jones, E., Oliphant, T. & Peterson, P. SciPy: Open source scientific tools for Python.
901   (2001).
902   68   Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for
903   efficient numerical computation. *Computing in Science & Engineering* **13**, 22 (2011).
904   69   Seabold, S. & Perktold, J. in *Proceedings of the 9th Python in Science Conference.* 61
905   (Scipy).
906   70   Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**,
907   428-430, doi:10.1093/bioinformatics/btt684 (2014).
908   71   Wilkinson, L. Exact and Approximate Area-Proportional Circular Venn and Euler
909   Diagrams. *IEEE Transactions on Visualization and Computer Graphics* **18**, 321-331,
910   doi:10.1109/TVCG.2011.56 (2012).
911   72   Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing
912   biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118
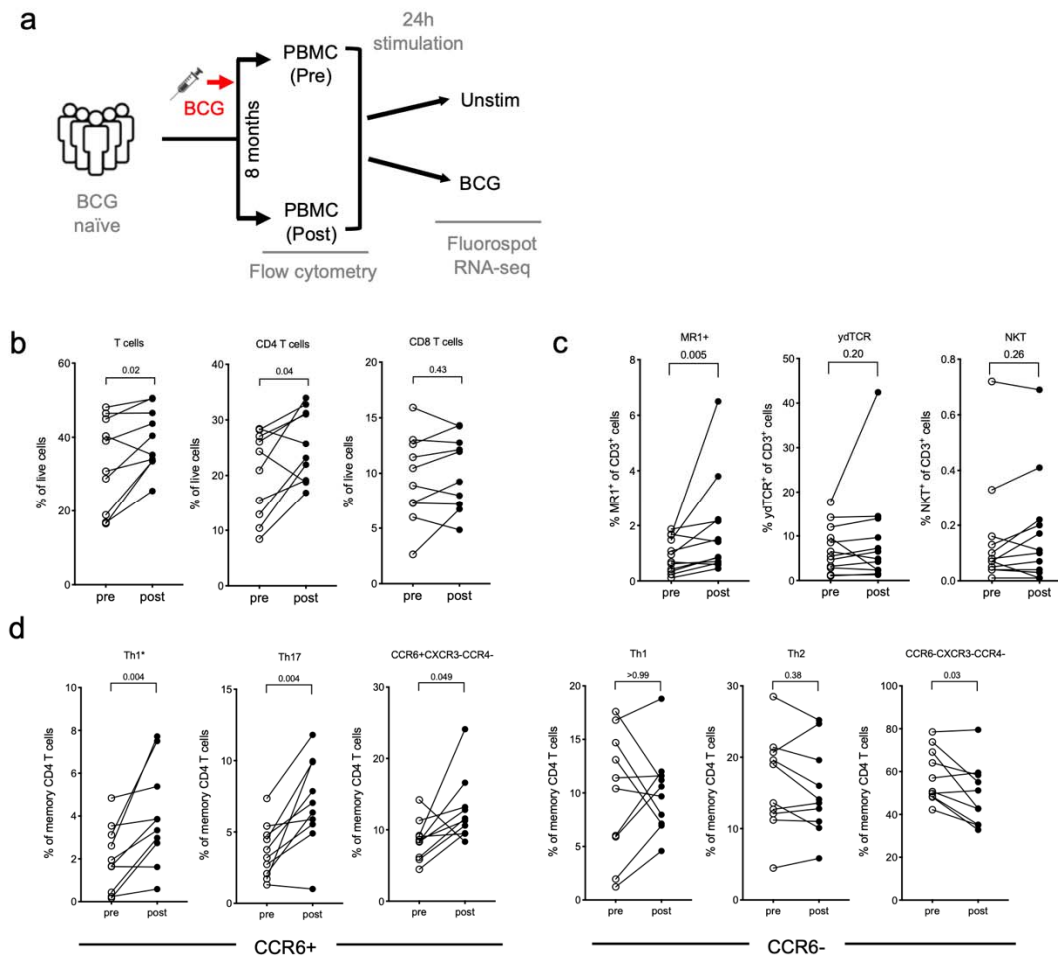913   (2012).
914

915

Figures and Legends

Figure 1

FIGURE 1. T cell subsets show an increase post BCG vaccination. a Schematic of the project workflow. Flow cytometry analysis was performed on PBMC samples obtained from BCG naïve individuals, pre and post BCG vaccination, and Fluorospot and RNA-sequencing on PBMCs stimulated with either DMSO or BCG for 24 hours. **b-d** Frequencies of cell subsets pre- (open circles) and post- (closed circles) BCG vaccination as determined by flow cytometry. Each point represents one participant, Wilcoxon matched pair signed rank test. **b** T cells, CD4 and CD8 T cells. **c** MAITs; as defined by staining with MR1 5-OP-RU tetramer, γδ T cells; defined by pan γδTCR, and NKTs; defined by Vα24 staining. **d** T-helper subsets defined by CXCR3, CCR6 and CCR4 expression.
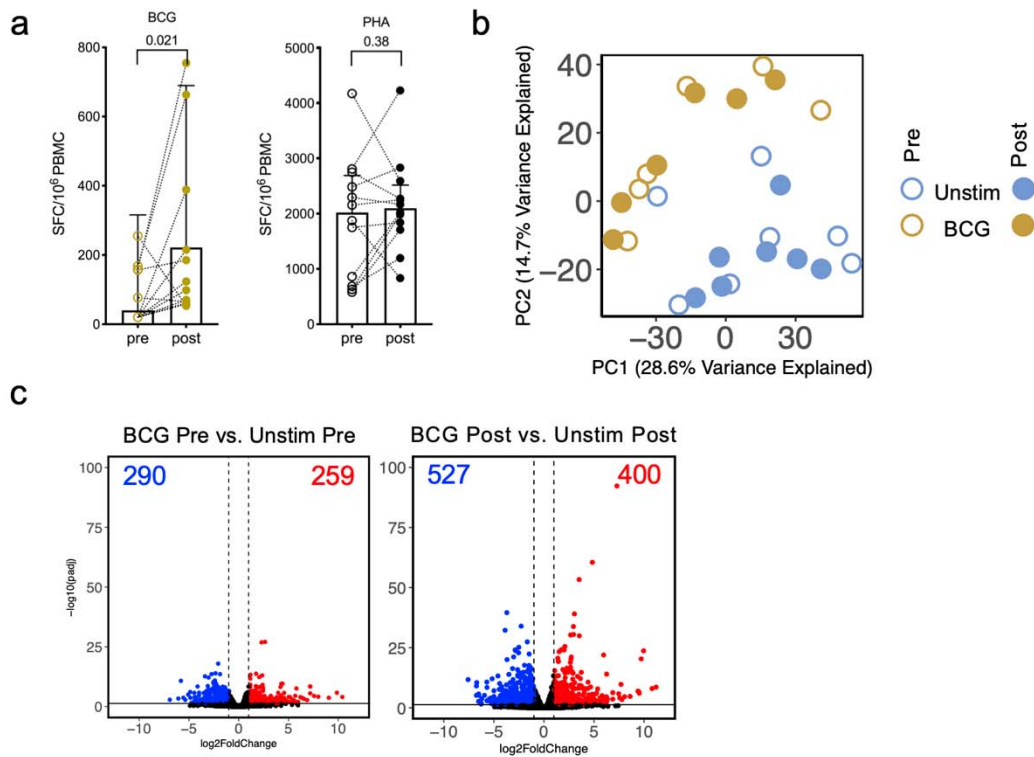
Figure 2

927

928   **FIGURE 2. BCG-induced T cell responses and gene expression changes are enhanced**

929   **upon BCG vaccination. a** Paired point graphs depicting magnitude of IFNγ responses pre- and

930   post-vaccination, against BCG and PHA as SFC per 10^6 cultured PBMC as determined by

931   Fluorospot. Each point and symbol represent one participant, median ± interquartile range is

932   shown. Wilcoxon matched pair signed rank test. **b** PCA depicting the variation in global gene

933   expression as a result of stimulation condition and vaccination. Pre-vaccination (open circles)

934   and post vaccination (closed circles), unstimulated samples (blue) and BCG stimulated samples

935   (gold). **c** Volcano plot depicting differentially expressed genes after BCG stimulation as

936   compared to unstimulated samples, pre- and post-vaccination. Red indicates upregulated genes

937   (adjusted p value < 0.05 and $\log_2$ fold change > 1) and blue indicates downregulated genes

938   (adjusted p value < 0.05 and $\log_2$ fold change < -1).

939
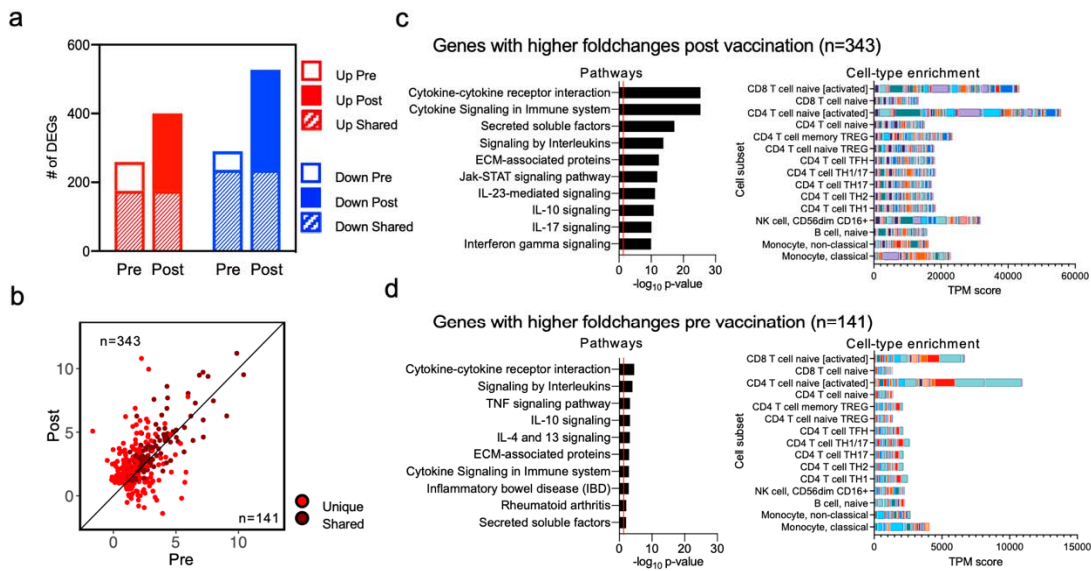
Figure 3

FIGURE 3. BCG-induced gene expression changes are enhanced upon BCG vaccination. **a** Bar plots representing the total number of differentially expressed genes pre- and post-vaccination. Empty and filled bars represent pre- and post-vaccination, respectively, color represents up or down regulation, and hatched filling represents overlap of genes between the pre and post time points. **b** Scatterplot of genes upregulated upon BCG stimulation in pre-vaccination (x-axis) and post-vaccination (y-axis). Each dot represents the $\log_2$ fold change for a gene, and color represents shared (gene significantly differentially expressed in both pre- and post-vaccination) or unique (differentially expressed in only one condition) expression. Black line at the 45° slope represents identical perturbation in pre- and post-vaccination, with genes above or below the line showing higher perturbation in the post or pre-group, respectively. **c, d** Pathway and cell-type enrichment (dice-database.org) for genes with **(c)** higher $\log_2$ fold changes post vaccination, and **(d)** higher $\log_2$ fold changes pre-vaccination. The ten most significant pathways are shown.
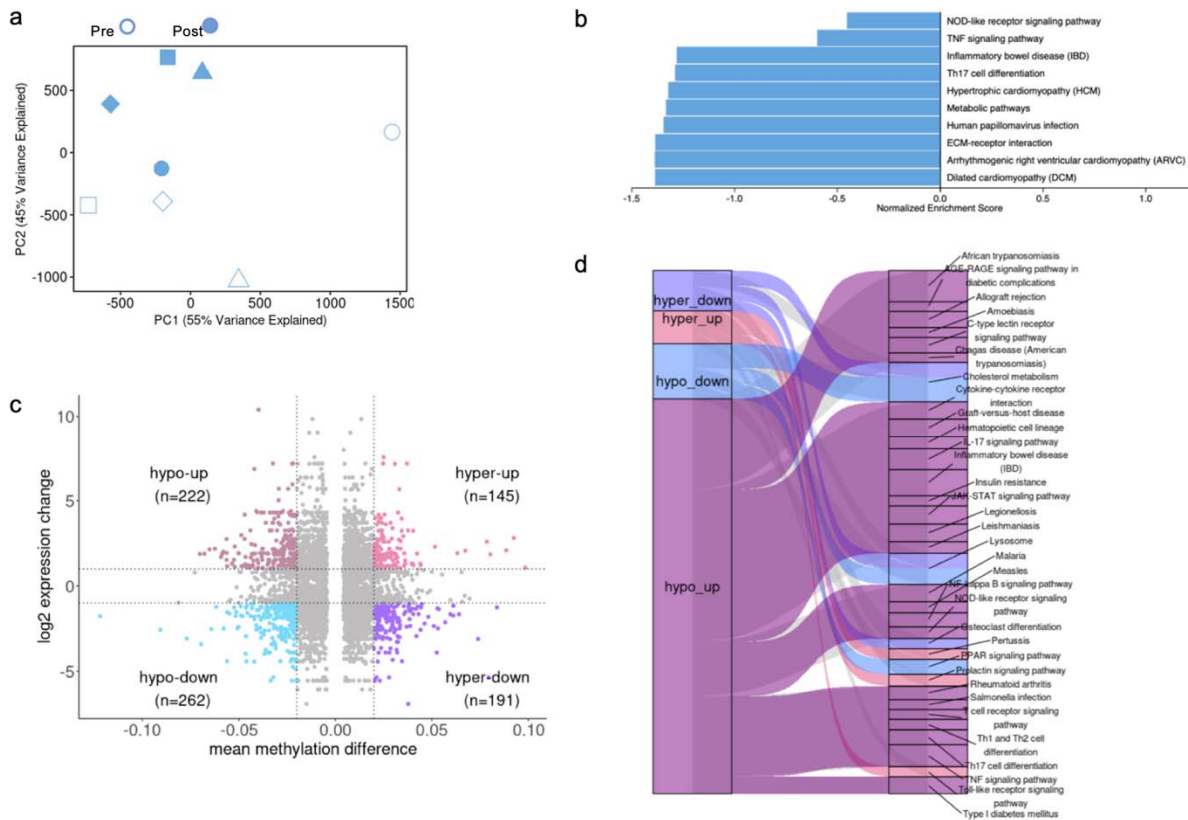
Figure 4

FIGURE 4. BCG-induced DNA methylation changes. a PCA of the overall DNA methylome data. Unstimulated pre-vaccination (open symbols) and post vaccination (closed symbols) samples**. b** Pathway enrichment for DMGs derived from hypomethylated CpGs. **c** Scatter plot representing the overlap of hypermethylated, hypomethylated, upregulated and downregulated genes post-vaccination. The vertical dotted lines represent the cut-off value of mmd ≥ |0.02| and the horizontal dotted lines represent the log2 expression value cut-off ≥ |1|. **d** Pathway enrichment analysis of overlapping significant DEGs and DMGs using the KEGG database The Gene Set Enrichment Analysis was performed using the 2000 permutations and BH-corrected p-value < 0.05 on the common gene list with mmd values. The resulting map of enriched pathways using KEGG database in different combinations of DMGs and DEGs is shown. Edge width indicates the number of significant pathways involved with the corresponding DMG-DEG list.
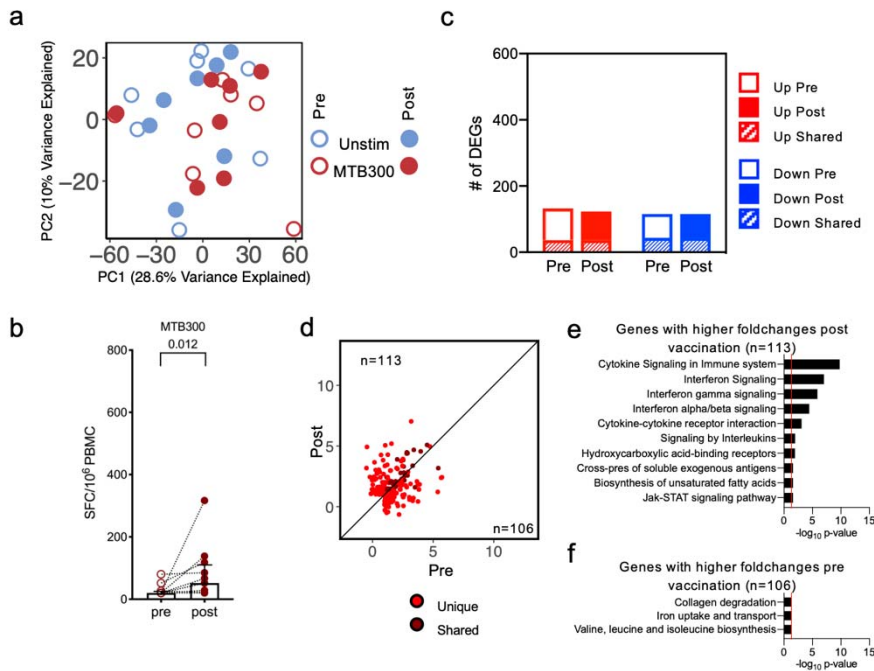
Figure 5

FIGURE 5. MTB300 peptide pool stimulation has a small impact on BCG vaccination. a PCA depicting variation in global gene expression as a result of stimulation condition and vaccination. Empty and filled circles represent the pre- and post-vaccination time points, respectively, and color represents stimulation condition. **b** Paired point graphs depicting magnitude of responses (sum of IFNγ, IL-5, and IL-10) pre- and post-vaccination, against MTB300, as SFC per 10^6 cultured PBMC. Each point and symbol represent one participant, median ± interquartile range is shown. Wilcoxon matched pair signed rank test. **c** Bar plots representing the total number of differentially expressed genes pre- and post-vaccination. Empty and filled bars represent pre- and post-vaccination, respectively, color represents up or down regulation, and hatched filling represents overlap of genes between the pre and post time points. **d** Scatterplot of genes upregulated upon MTB300 stimulation in pre-vaccination (x-axis) and post-vaccination (y-axis). Each dot represents the $\log_2$ fold change for a gene, and color represents shared (gene significantly differentially expressed in both pre- and post-vaccination) or unique (differentially expressed in only one condition) expression. Black line at the 45° slope represents identical perturbation in pre- and post-vaccination, with genes above or below the line showing higher perturbation in the post- or pre-group, respectively. **e** Pathway enrichment for genes with higher $\log_2$ fold changes post vaccination. The ten most significant pathways are shown. **f** Pathway enrichment for genes with higher $\log_2$ fold changes pre vaccination.
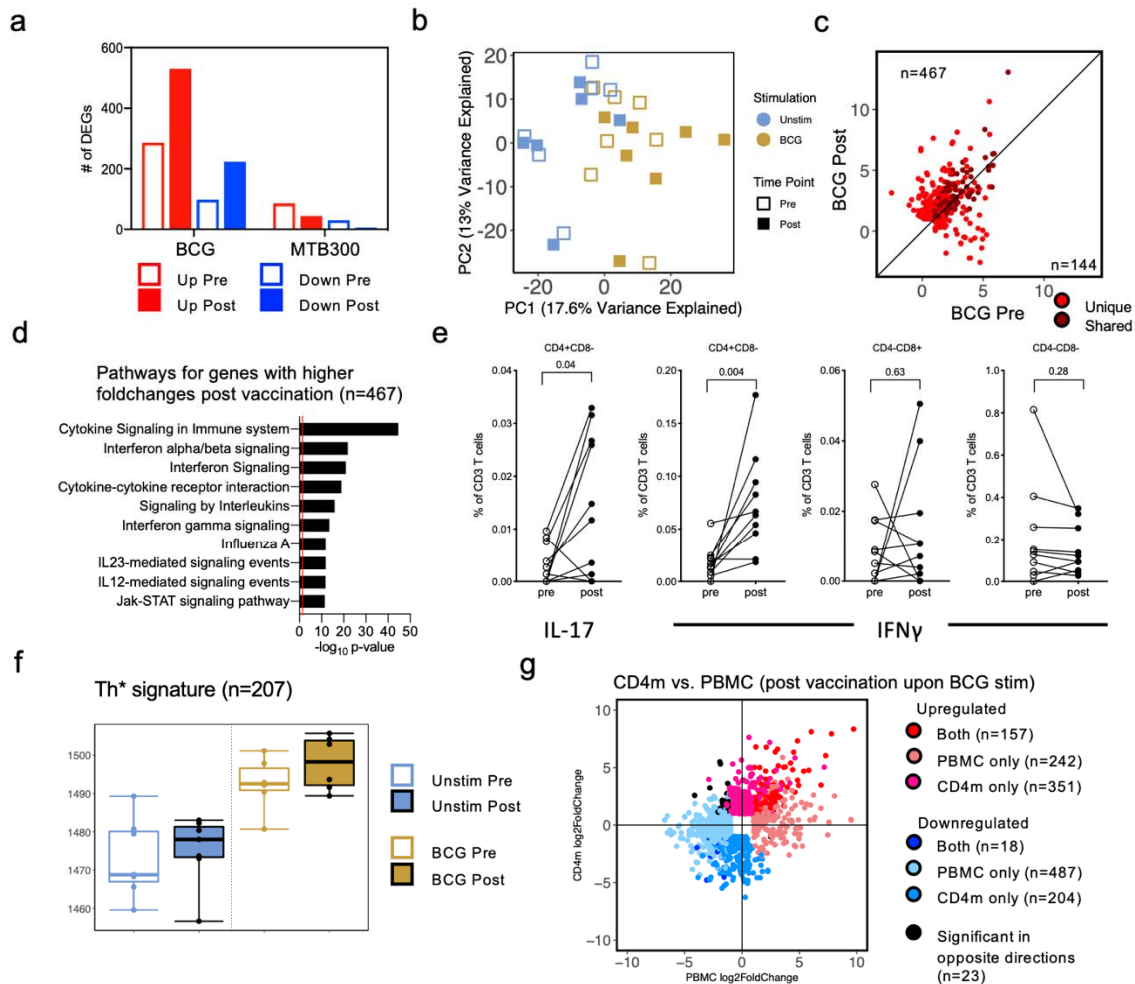
Figure 6

988

989 **FIGURE 6. CD4 memory T cells show enhanced gene expression post BCG vaccination in**

990 **BCG stimulated samples. a** Grouped bar plots representing differentially expressed genes

991 identified in BCG and MTB300 stimulation, pre- and post-vaccination. Empty bars and filled bars

992 represent pre- and post-vaccination, respectively, and color represents up or down regulation. **b**

993 Principal component analysis depicting the variation in global gene expression as a result of

994 stimulation condition and vaccination. Empty and filled squares represent the pre- and post-

995 vaccination time points, respectively, and color represents stimulation condition. **c** Scatterplot of

996 genes upregulated upon BCG stimulation in pre-vaccination (x-axis) and post-vaccination (y-

997 axis). Each dot represents the log$_2$ fold change for a gene, and color represents shared (gene

998 significantly differentially expressed in both pre- and post-vaccination) or unique (differentially

999 expressed in only one condition) expression. Black line at the 45° slope represents identical

1000   perturbation in pre- and post-vaccination, with genes above or below the line showing higher

1001   perturbation in the post- or pre-group, respectively. **d** Pathway enrichment for genes with higher

1002 $\log_2$ fold changes post vaccination. The ten most significant pathways are shown. **e** Percentage

1003   cytokine (IL-17 or IFNγ) detected from T cells in response to BCG. Each dot represents one

1004   donor, median ± interquartile range is shown. Wilcoxon matched pair signed rank test. **f** Boxplot

1005   of RNA-sequencing data depicting the sum of $\log_2$ expression values (variance stabilizing

1006   transformation, VST) for all genes in the Th1* cell signature. Empty and filled boxplot represent

1007   pre- and post-vaccination, respectively, and color represents stimulation condition. Each dot

1008   represents an individual donor. **g** Scatterplot of genes differentially expressed between post-

1009   vaccination in BCG stimulated compared to unstimulated samples, in PBMC (x-axis) and CD4

1010   memory T cells (y-axis). Each dot represents the $\log_2$ fold change for a gene, and color

1011   represents shared or uniquely differentially expressed genes in PBMC or CD4m.
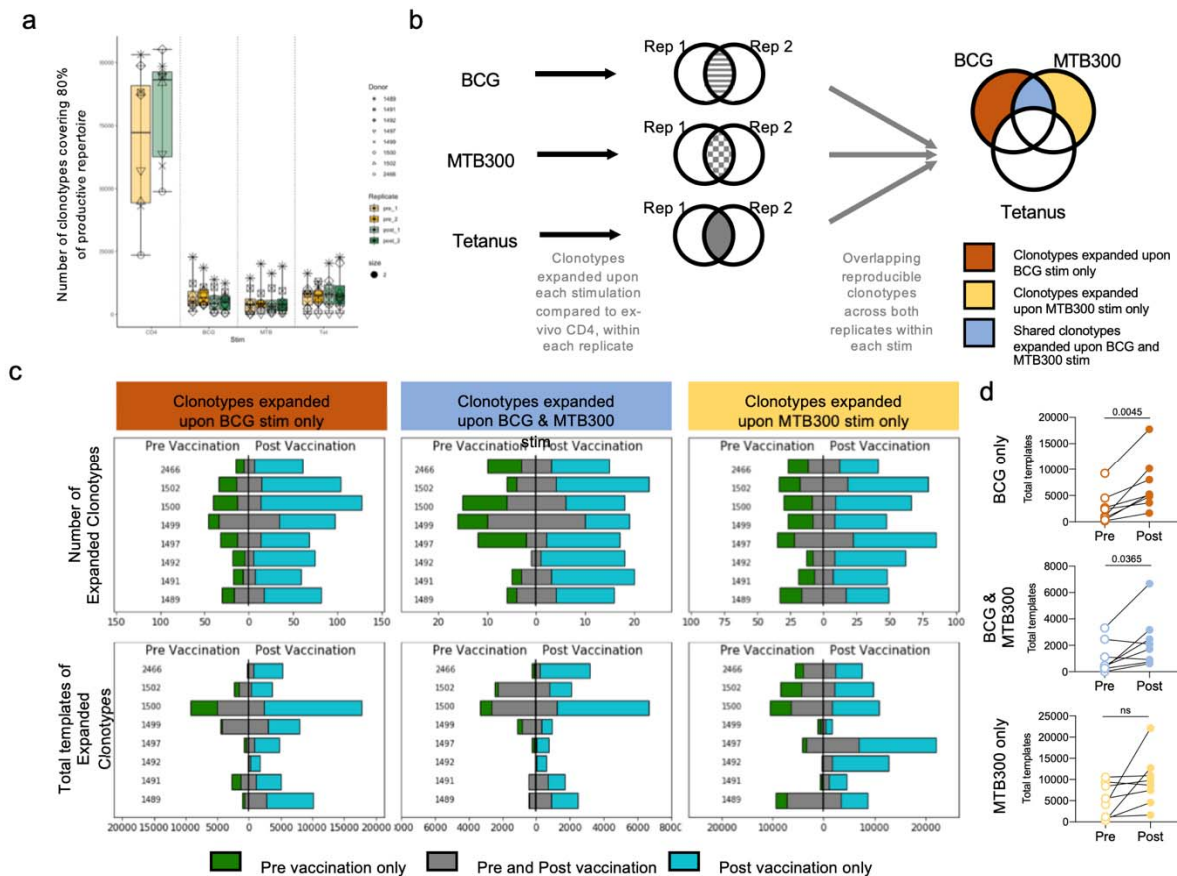
1012

Figure 7

FIGURE 7. BCG and MTB300 stimulation elicit specific T cell responses. **a** Number of TCR
clonotypes covering 80% of the productive repertoire across stimulation conditions and pre- and
post-vaccination. Color, yellow and green, represents pre- and post-vaccination, respectively.
Each symbol represents one donor. Replicate samples are indicated by the intensity of the color
(lighter and darker shade). **b** TCR-sequencing was performed on CD4$^+$ T cell samples obtained
from BCG naïve individuals (n=8), stimulated with the BCG epitope pool, MTB300 peptide pool,
or the tetanus peptide pool. Only those clonotypes that expanded upon stimulation compared to
ex-vivo CD4 samples across both replicates were retained within each stimulation condition.
The resulting clonotypes in the three stimulation conditions were overlapped to obtain
clonotypes that expanded upon BCG stimulation only, MTB300 stimulation only, and clonotypes
expanded in both BCG and MTB300 stimulation. Any clonotype expanded upon tetanus
stimulation was excluded. This process was repeated for samples obtained post-vaccination. **c**
Bar graphs showing the number of expanded clonotypes, and the total templates for these
clonotypes, across different conditions, pre- and post-vaccination. Blue indicates clonotypes

1028 unique to post-vaccination, green indicates clonotypes unique to pre-vaccination, and grey

1029 indicates clonotypes that expanded in both pre- and post-vaccination. **d** Quantification of total

1030 templates expanded in pre- and post-vaccination for individual donors in each condition. Empty

1031 and filled circles represent pre- and post-vaccination, respectively.

1032