Host genetic background is a barrier to vaccine-induced protection against infection

Rocky Lai[1], Diana Gong[2], Travis Williams[1], Abiola F Ogunsola[1], Kelly Cavallo[1], Cecilia Lindestam Arlehamn[3], Gillian L. Beamer[4], Martin Ferris[5], Christopher M. Sassetti[1], Doug Lauffenburger[2], Samuel M. Behar[1]

1. Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, Massachusetts, USA.
2. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
3. Center for infectious disease and vaccine research, La Jolla Institute for Immunology., La Jolla, CA, USA
4. Texas Biomedical Research Institute, San Antonio, TX, USA
5. Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, United States

*Correspondence:

Samuel M. Behar

E-mail address: samuel.behar@umassmed.edu

Summary (150)

The heterogeneity of immune responses observed in humans is difficult to model in standard inbred laboratory mice. We used Collaborative Cross (CC) recombinant inbred strains to capture the diversity of outbred populations and better understand how host variation affects BCG-induced immunity against *Mycobacterium tuberculosis*. Using 24 CC strains, we show host genetics has a major role in determining whether BCG elicits protective immunity to *M. tuberculosis*. Only 54% of BCG-vaccinated CC strains had significantly reduced lung bacillary burdens after infection. Furthermore, BCG efficacy is dissociable from inherent susceptibility to TB. Like people, CC mice have diverse immunological responses after BCG vaccination and *M. tuberculosis* challenge. We define immunological signatures associated with BCG-induced protection other than canonical Th1 immunity. Like interferon-γ, lung homogenate levels of IL-17 correlated with disease. However, the presence of Th1/17 CD4 T cells were an immune feature that correlated with protection. CC mice can be used to define correlates of protection and to identify vaccine strategies that protect a larger fraction of genetically diverse individuals instead of optimizing protection for a single genotype.

42 **Introduction**

43 The introduction of Bacillus Calmette Guerin (BCG) as a vaccine in the last century
44 has contributed significantly to our ability to control and contain tuberculosis (TB), which
45 has been a threat to global health for millennia. BCG is the only vaccine approved for
46 preventing tuberculosis and over 4 billion doses have been administered in over 150
47 countries (*1*). Although BCG prevents extrapulmonary TB in infants, it is less effective in
48 averting pulmonary disease in adolescents and adults (*2-5*). Complicating the use of BCG
49 is its variable efficacy in different regions of the world (*6*). While the development of more
50 effective TB vaccines is recognized by the WHO as an important yet unmet medical need,
51 the lack of identifiable correlates of protective immunity to TB is an impediment towards
52 this goal.

53 Why is identifying immune correlates of protection following BCG difficult? TB is
54 primarily a human disease that is transmitted person-to-person by aerosolized
55 *Mycobacterium tuberculosis* (Mtb). No animal model recapitulates all the features of
56 human TB, although non-human primates come close (*7, 8*). Although imperfect, the
57 mouse model has facilitated mechanistic studies that defined many of the critical
58 immunological pathways required for host resistance, permitted in vivo antibiotic testing,
59 and accelerated preclinical vaccine development. As inbred mouse strains are genetically
60 identical, a strength of the mouse model lies in its "infinite" reproducibility. For
61 experimental immunology, C57BL/6 mice are the preferred inbred strain as most
62 genetically modified mice are produced using its genetic background. It is also the most
63 frequently used mouse strain for TB research. However, several features of C57BL/6
64 mice suggest it is an outlier strain for TB research. Immunologically, it only expresses a

65 single class II MHC molecule (i.e., I-A^b). Pathologically, it develops atypical pathology
66 (i.e., pneumonitis) following Mtb infection. Finally, and perhaps most importantly, it is
67 among the most resistant of all inbred strains to Mtb infection, which makes it difficult to
68 identify conditions that make C57BL/6 mice more resistant to disease. Investigators have
69 addressed these limitations by using F1 hybrids and susceptible strains to increase MHC
70 and genetic diversity and identify strategies to increase host resistance (*9-12*). However,
71 even among different inbred strains, the genetic diversity is limited (*13*). This is a problem
72 not only for TB research, but for many areas of biomedical research.

73 The mouse genetics community responded to these limitations of inbred mice by
74 developing a mouse resource that captures the considerable genetic diversity of the *Mus*
75 species. Starting with eight founder mouse strains, including representatives from all
76 three *Mus musculus* subspecies and three wild-derived strains, an eight-way funnel
77 breeding scheme led to random assortment of the founder genomes among the progeny
78 (*14*). The progenies were used to derive recombinant inbred mouse strains, which
79 became known as Collaborative Cross (CC) strains. Each CC strain is a unique genetic
80 representation (i.e., genotype) of the eight founder strains. Importantly, as they are inbred,
81 they are a renewable resource, and an unlimited number of experiments can be
82 performed with the same genotype, which is important for identifying experimental
83 correlates and defining mechanisms.

84 We previously reported that the CC strains have a large variation in primary
85 susceptibility to TB (*15, 16*). Here we leverage the CC model to understand the how host
86 genetics affects vaccine-induced protection against Mtb. We find that the genetic
87 background of the host has a major effect on whether BCG induces protection. There was

88 a poor correlation between inherent host susceptibility and BCG-mediated protection,
89 indicating that vaccines can mediate protection even in individuals with a genetic
90 propensity to TB. Interestingly, the T cell response following BCG vaccination and Mtb
91 challenge was diverse, with some immune features universally correlating with outcome
92 of vaccination while others only were important in subsets of CC strains. We propose that
93 the use of CC mice as a preclinical model can facilitate the development of vaccines that
94 protect a diverse range of genotypes against TB.

95 **Results.**

96 ***BCG vaccination protects only a subset of CC strains against TB.***

97 To establish how the host genetic background affects vaccine mediated protection
98 against Mtb, 24 CC lines (Table 1) were vaccinated subcutaneously with BCG and rested
99 for 12 weeks. Then, the mice were infected with lose dose aerosolized H37Rv.YFP (*17*).
100 Four weeks after infection, the reduction in lung and spleen CFU were determined as a
101 measurement of protection. Mice were initially tested in four batches, with each batch
102 including six different CC strains and C57BL/6 mice as a reference strain.

103 The aggregated lung CFU from the unvaccinated and BCG-vaccinated CC strains
104 (Fig.1A, top) and C57BL/6 mice (Fig.1A, bottom) shows that as a population, CC mice
105 are protected by BCG similarly as C57BL/6 mice. However, the variance in the lung CFU
106 of CC mice is greater. When the data is segregated by individual CC strains, we can see
107 that the variance reflects different levels of protection conferred by BCG vaccination in
108 each CC strain (Fig.1B, top). The CC strains were classified as being protected by BCG
109 if there was a statistically significant difference between unvaccinated and vaccinated
110 mice. The $\Delta\log_{10}$ lung CFU was calculated for each CC strain, and they were ranked from
111 most protected (CC037, $\Delta\log_{10}$ CFU = 1.6) to least protected (CC040, $\Delta\log_{10}$ CFU = –0.6)
112 (Fig.1B, bottom). BCG vaccination exacerbated subsequent Mtb infection in the lung for
113 two CC strains (CC003 and CC040), similar to what had previously been observed in the
114 NZO founder strain (*15*). Importantly, BCG only led to significant reductions in the lung
115 Mtb burden in 13 of the 24 CC strains.

116 We next asked whether the protective effect of BCG vaccination correlated with
117 the inherent ability (i.e., in the unvaccinated state) of each CC strain to control Mtb

118 replication in their lungs. We categorized the 24 CC strains relative to C57BL/6 mice,
119 which we used as a reference strain. The lung CFU of unvaccinated mice defined the
120 inherent susceptibility or resistance to Mtb (susceptibility, y-axis). Next, we considered
121 whether each CC strain was protected or not based on the $\Delta\log_{10}$ lung CFU (relative
122 BCG-induced protection, x-axis). The plot was centered on C57BL/6 mice, which
123 represented a median strain in terms of lung susceptibility and protection and divided the
124 24 CC strains into four different groups, represented by four quadrants (Fig.1C). Based
125 on lung CFU, there is a poor correlation between intrinsic susceptibility of the mouse
126 strain and protection conferred by BCG at this time point (r=0.104, p=0.62).

127 The same analysis was performed using the spleen bacterial burden from infected
128 mice. As was observed in the lung, the CC population was similarly protected as C57BL/6
129 (Fig.1D). The range of spleen protection extended from most protected (CC026, $\Delta\log_{10}$
130 CFU = 2.56) to least protected (CC045, $\Delta\log_{10}$ CFU = 0.13) (Fig.1E). BCG significantly
131 protected against systemic infection in 17 of the 24 CC strains (Fig.1E). There was also
132 no correlation between susceptibility and protection in the spleen (r=0.089, p=0.67).
133 Interestingly, while C57BL/6 mice were a “median” strain in the lung analysis, C57BL/6
134 mice had among the highest spleen CFU, suggesting they disseminated more, but were
135 among the most protected mouse strain following BCG vaccination (Fig.1F). Among the
136 CC strains, protection in the lung and the spleen were moderately correlated (r=0.513,
137 p=0.0087).

138 In follow-up studies, 13 of the 24 CC strains were retested to confirm the initial
139 phenotype, with emphasis placed on strains that represented varying levels of protection
140 and susceptibility (Table 1). Thus, using 24 CC strains, with diverse genotypes, we find

141 that less than 50% of the strains were protected by BCG, and that protection did not
142 correlate with the intrinsic susceptibility of the mouse to Mtb infection. Finally, the lack of
143 correlation between the reduction in lung versus spleen highlights the importance of
144 tissue-specific immunity and the problems inherent in protecting the lung against
145 infection. We next assessed how BCG vaccination altered the histological appearance of
146 TB lesions in the lung.
147
148 ***BCG modifies the lung lesions caused by Mtb in a strain-specific manner.***
149 We assessed how BCG vaccination modified lung lesion development after Mtb
150 infection in 19 different CC strains plus C57BL/6 mice. A board-certified veterinarian
151 pathologist (G.B.), blinded to vaccination status of each strain, examined the lung for
152 necrosis, neutrophilic and lymphocytic infiltrates, lesion size and number. Within each
153 pair, the vaccinated group was correctly identified in 15 of the 20 strains analyzed based
154 on reduced severity of the histopathology. In some CC strains that were protected by
155 BCG, increased lymphocytic infiltration, and reduced necrosis were frequently observed
156 after vaccination (Table 2). CC025 (Fig. 2, top left) and CC037 (Fig. 2, middle left) were
157 both susceptible to primary Mtb infection, and the infected lung showed innate immune
158 cells and tissue necrosis. BCG vaccination protects these mice by promoting lymphocyte
159 recruitment and in CC025 BCG vaccination reduced necrosis. However, a reduction in
160 lung CFU associated with BCG vaccination was not always associated with an improved
161 histological appearance of the lung. Despite a significant reduction in lung CFU after
162 BCG, no discernable histopathological differences were observed between unvaccinated

163 and BCG vaccinated CC072 mice, and both groups had non-necrotizing lymphocyte-rich
164 granulomas (Fig. 2, bottom left).

165 BCG did not significantly reduce lung CFU in eight of 20 CC strains. Two of the
166 non-protected CC strains, CC003 (Fig. 2, top right) and CC024 (Fig. 2, middle right) had
167 no discernable histological differences in granuloma structure, content, or severity
168 between control and BCG vaccination when evaluated blindly. In these 2 strains, lesions
169 were generally small to medium-sized, non-necrotizing, with abundant perivascular and
170 peribronchiolar lymphocytes. Given the intermediate susceptibility of the CC003 strain,
171 and BCG's inability to alter lung CFU or change the lesions, the data suggests that BCG
172 does not change CC003 innate or adaptive immune responses to Mtb challenge.
173 Likewise, CC024 is unprotected by BCG vaccination, nor are lung lesions altered,
174 suggesting that neither the innate nor adaptive responses of this resistant strain to
175 primary Mtb infection is further enhanced by BCG vaccination. The natural resistance of
176 CC024 strain was not further improved. In contrast, CC004 (Fig. 2, bottom right) did have
177 discernable histopathological changes when controls and BCG vaccinees were evaluated
178 blinded. The differences attributable to BCG vaccination were the greater lymphocytic
179 infiltration with denser perivascular and peribronchiolar lymphocytes. The altered lesions
180 detected after vaccination suggests that the immunity induced by BCG in CC004 mice
181 fails to control bacilli. Together, these data show that the host genetic background impacts
182 the granuloma content, structure, and quality of the immune response to BCG and Mtb
183 challenge in a manner that is not wholly predictable.
184

185 **BCG-protected CC strains have an altered lung microenvironment after Mtb**
186 **infection**

187 The Mtb-infected lung is a complex environment of resident and recruited cells that
188 produce inflammatory mediators and counter-regulatory (e.g., anti-inflammatory) signals.
189 We sought to determine whether BCG vaccination primes the immune response to
190 generate a different lung microenvironment upon Mtb challenge, perhaps one that is less
191 conducive to bacillary replication and transmission. Cytokine levels in lung homogenates
192 from PBS control and BCG-vaccinated C57BL/6 or CC mice were measured 4 weeks
193 post infection using the MD31 Discovery Assay Array.

194 Given the critical role of IFNγ in Mtb control, its levels in control and BCG-
195 vaccinated CC mice were compared. In C57BL/6 mice and many of the protected CC
196 strains, including CC037, CC025, CC059, CC031, and CC011, IFNγ concentrations in
197 BCG-vaccinated mice were reduced compared to unvaccinated mice (Fig.3A). However,
198 this pattern did not hold for all protected strains. No changes in IFNγ levels were detected
199 the CC032, CC042, CC072, CC026 and CC001 strains, despite being protected by BCG.
200 Thus, when considered alone, neither the absolute level of IFNγ nor its changes after
201 vaccination could be used to predict protection.

202 We next asked if there was a correlation between the different cytokines in the lung
203 homogenate and lung CFU, in either unvaccinated or BCG-vaccinated mice. We asked if
204 this correlation differed among protected and unprotected CC strains, and C57BL/6 mice.
205 (Fig.3B, top left). There was a strong correlation between CFU and lung IFNγ levels
206 (R=0.71, p=0.0004), in unvaccinated C57BL/6 mice but not in BCG-vaccinated C57BL/6
207 mice. Thus, IFNγ correlated with bacterial burden but not with protection, as observed in

208 prior studies (*18, 19*). The CC strains protected by BCG had a similar pattern. However,
209 for unprotected CC strains, a similar correlation between IFNγ and CFU was identified in
210 unvaccinated, and BCG vaccinated mice, which we suggest is evidence that these strains
211 were unprotected by BCG vaccination. This pattern held for T cell cytokines (IFNγ and IL-
212 17), mediators produced by myeloid cells (IL-1β, IL-6, KC, and MIP-1α), and TNF (Fig.3B
213 and not shown). These data suggest that protected and unprotected CC strains respond
214 differently to BCG, and that the protected lung environment differs immunologically from
215 an unprotected environment.

216 Other cytokines were of interest. LIX (CXCL2) was negatively associated with lung
217 CFU in unvaccinated but not in BCG-vaccinated C57BL/6 mice (–0.67, p=0.0013). This
218 correlation was only observed in C57BL/6 mice, and LIX levels poorly correlated with lung
219 CFU in CC strains (Fig.3C). RANTES strongly correlated with lung CFU in C57BL/6 mice,
220 independent of vaccination status, but was only weakly correlated in CC strains (Fig.3D).
221 Like RANTES, MIP-1β was also correlated with lung CFU, independently of vaccination
222 status (Fig.3E). This was most convincing in unprotected CC mice (R=0.71-0.78), with
223 moderate correlations in vaccinated C57BL/6 and protected CC strains (R=0.46-0.49).

224 Four cytokines (G-CSF, M-CSF, MIP-2 and LIF) were strongly correlated with lung
225 CFU in unvaccinated CC strains, regardless of protection status (R=0.57-0.8, Fig.3F).
226 Again, these correlations were lost in the BCG-vaccinated protected CC strains. None of
227 these cytokines were made in significant amounts by C57BL/6 mice or were poorly
228 correlated with CFU. These data show how differences in the host genetic background,
229 result in varied lung histology and cytokines, which dramatically alter the lung
230 inflammatory environment during Mtb infection.

231

232 ***Host genetic background and not BCG vaccination determines T cell subset***
233 ***distribution in the lung after Mtb infection.***

234 We hypothesized that changes in the frequency of T cell subsets would correlate
235 with protection induced by BCG vaccination. An 18-parameter flow cytometry panel was
236 used to identify and enumerate different immune subsets (Figure S1). As T cells are
237 important in protection against TB, we determined the frequency and relative number of
238 different T cell subsets modulated by BCG vaccination. Not all parameters were obtained
239 for all 24 CC strains (see methods). We quantified CD4 and CD8 T cells, MAIT cells and
240 NKT cells (Fig.4A, and Figure S2). Memory and effector T cell populations were defined
241 as central memory (T_{CM}; $CD44^{+}CD62L^{+}CD127^{+}$), effector memory ($T_{EM}$; $CD44^{+}CD62L^{-}$
242 $CD127^{+}$), effector (T_{Eff}; $CD44^{+}CD62L^{-}CD127^{-}$), and resident memory ($T_{RM}$;
243 $CD44^{+}CD103^{+}CD69^{+}$) (Fig.4B, 4C). CD4 T cells subsets were defined in all CC strains
244 based on their expression of the transcription factors Foxp3, Eomes, Tbet and RORγt to
245 identify Treg, Th1, Th2, and Th17 cells, respectively (Fig.4D). For these T cell subsets,
246 we observed tremendous strain-to-strain variation. With one or two exceptions, BCG
247 vaccination did not modulate the distribution of T cells subsets nor was there any
248 correlation between the relative frequency of these T cell subsets and the outcome of
249 BCG vaccination. Thus, the frequency of different T cells subsets in the lungs of CC mice
250 after Mtb infection was largely determined by each strain's genetic background.

251 Expression of CXCR3 and CX_3CR1, which have been used to discriminate
252 between T cells in the lung parenchymal vs. the vasculature, was analyzed in a limited
253 number of CC strains (*20*). CXCR3 also marks Th1 cells, and in conjunction with CCR6,

254 can differentiate between Th1 and Th17 cells, respectively (*21*). Like other markers, we
255 observed variation in CXCR3 expression between different CC strains, but little or no
256 variation that correlated with vaccination status or susceptibility. In addition to allelic
257 variation, CC strains also carry private mutations. For example, we previously reported
258 that CC042 mice lacks cell surface CD11a expression due to a mutation in the *Itgal* gene
259 on Chromosome 7 (*22*). Here, we report that CC003 mice don't express CXCR3 (Fig.4E).
260 This phenotype is associated with the CXCR3 allele of the PWK founder. Finally, a
261 reduction in CX3CR1^{+} CD4 and CD8 T cells was observed in several of the protected CC
262 strains but not in unprotected strains (Fig.4E). This pattern mirrored the reduction in IFNγ
263 levels observed in some protected CC mice (Fig.3), suggesting that these changes reflect
264 Th1 modulation.

265 Finally, we used principal components analysis to holistically visualize the variation
266 in the T cell phenotyping data. What is striking is the small effect of BCG vaccination in
267 altering T cell phenotypes, in comparison to large effect of genetic strain differences.
268 These data all indicate that genetics differences between the CC strains determine the
269 differences among T cell subsets recruited to the lung, and not BCG vaccination or Mtb
270 infection. Therefore, we next wished to determine whether there were functional
271 differences among the T cell subsets in protected vs. nonprotected CC strains.

272

273 ***The patterns of cytokine production differ in CC strains after BCG vaccination and***
274 ***Mtb infection.***

275 As the cytokines produced by T cells, including IFNγ, TNF, and IL-21, among
276 others, are important in control of Mtb infection, we hypothesized that changes in T cell

277 function would be associated with BCG-induced protection. To measure function, we
278 stimulated lung mononuclear cells (MNC) from each mouse with anti-CD3 or the P300
279 megapool, which contains 300 peptides representing epitopes from 70 Mtb proteins that
280 are frequently recognized by human CD4 T cells and are also recognized by murine CD4
281 and CD8 T cells (*23, 24*). Twenty-six cytokines and chemokines in the culture
282 supernatants were measured using the Legendplex platform 24 hours after stimulation.
283 There was little difference in the ability of lung MNC from mock or BCG vaccinated Mtb-
284 infected C57BL/6 mice to secrete different cytokines or chemokines following P300
285 stimulation. In contrast, we detected differential production of IFNγ, TNF, IL-2 or IL-17, all
286 cytokines that have been associated with immunity to Mtb infection, among mock vs. BCG
287 vaccinated CC strains (Fig.5A). Interestingly, the MNC from BCG vaccinated CC037,
288 CC031, CC72 and CC001 strains all produced more IL-17 than their respective controls.
289 In contrast, none of the unprotected strains produced significant amounts of IL-17, except
290 for CC023, and its lung MNC produced less IL-17 after BCG vaccination.

291 In parallel, a 14-parameter flow cytometry panel measured intracellular levels of
292 IFNγ, IL-2, IL-17, TNF, IL10, and IL-4, and surface expression of CD107 in both CD4 and
293 CD8 T cells (Figure S3). BCG vaccination had a minimal effect on CD8 T cell cytokine
294 responses to P300 stimulation in both C57BL/6 mice as well as most CC lines examined
295 (Fig.5B). However, BCG clearly modified the CD4 T cell cytokine responses in some CC
296 strains (Fig.5C). Patterns of cytokine production were observed in BCG vaccinated CC
297 mice that were not observed in C57BL/6 mice. For example, while BCG vaccination did
298 not alter the frequency of $IFN\gamma^+$ or TNF^+ CD4 T cells in CC037 mice, it did increase IL-2
299 and IL-17 responses. BCG vaccination of CC031 mice increased the frequency of CD4 T

300 cells producing IFNγ, IL-2 and TNF but not IL-17. Finally, CD4 T cells from BCG
301 vaccinated had CC072 had increased IFNγ and IL-2 but not TNF or IL-17. Interestingly,
302 no discernable differences in cytokine responses were observed in most of the non-
303 protected CC mice, except for a decrease in IFNγ responses in CC023. Thus, the cytokine
304 responses in both control and BCG vaccinated CC mice differ greatly from classic
305 C57BL/6 model. Increased cytokine responses after T cell stimulation were more likely to
306 be detected in protected CC strains. Moreover, BCG vaccination led to quantitative and
307 qualitative changes in the type of T cell responses that persisted after Mtb challenge.
308
309 ***Correlates of Protection.***
310 We then turned to multivariate approaches to discern how BCG-induced immune
311 changes in the lung mechanistically lead to reductions in lung CFU in CC strains. To do
312 so, the data from the lung homogenate cytokines, intracellular cytokine staining, T cell
313 phenotyping, and LEGENDplex were combined, and for each vaccinated mouse the
314 transformed feature data was calculated as: [(vaccinated CC subject–
315 mean(unvaccinated CC strain]. Then, a partial least squares regression (PLSR) model
316 with elastic net feature selection was created to identify features that changed with BCG
317 vaccination in a way that maximally covaried with BCG-induced changes in lung CFU.
318 The model performed significantly better than null models, and the first two latent
319 variables in PLSR model account for 24% and 8%, respectively, of the variability in the
320 lung features (Figure 6A, Figure S4A). The elastic net-selected features in the model
321 included features from all four datasets (Fig.6B, 6C). By plotting the scaled values of the
322 elastic net-selected features for all CC strains, ordered from most protected to least

323 protected, what is most apparent is that the protected strains all had BCG-induced
324 reductions in several inflammatory cytokines such as MIG, CCL3 and CCL4 (i.e., MIP-1α
325 and -1β), IL-6, RANTES, and IL-17. The decrease in these inflammatory mediators could
326 be a sign of resolved infection, in contrast to the unprotected strains which had more
327 inflammatory cytokines which could signify unimpeded infection. Protected strains also
328 had a greater vaccination-induced increase in activated CD4 T cells. Contrarily, protected
329 strains had decreased activated $CD4^{-}CD8^{-}$ T cells producing T-bet. Importantly, IL-2,
330 unlike several other cytokines, was increased in protected strains. Thus, BCG vaccination
331 leads to qualitative and quantitative changes in the T cell response and in the immune
332 environment of the lung after Mtb infection in CC genotypes that are associated with
333 protection.
334
335 ***Immune responses in protected and unprotected CC strains differ following Mtb***
336 ***infection***
337 While the PLSR model highlights potential correlates of lung protection across all
338 genotypes, it may miss correlates that only apply to certain subsets of CC strains. We
339 hypothesized that naturally resistant and susceptible strains may be protected in different
340 ways. To test this, we created separate partial least squares discriminant analysis
341 (PLSDA) models for resistant and susceptible CC strains and examined the features that
342 best separated the BCG-protected and non-protected strains in each of these categories.
343 (Figure 1C). The data used for the models was processed the same way as in the PLSR
344 models, although we also looked at the vaccinated and unvaccinated mice separately in
345 addition to the vaccine-induced differences. Our first analysis compared resistant CC

346 strains that were protected by BCG (n=4) or not (n=8), based on lung CFU. As described
347 earlier (Figure 4), the unvaccinated and vaccinated mice of each CC strain were very
348 similar and clustered together when the elastic net-selected features of the model were
349 hierarchically clustered, indicating that the genetic background was an important
350 determinant of outcome (Figure 7A). The protected and unprotected mice were both split
351 into multiple clusters, showing that there were multiple distinct types of immune
352 responses that were associated with protective immunity in resistant mice. The model
353 could predict protected vs non-protected CC strains accurately and thus there is a
354 multivariate signature of these categories (Figure S4B). However, there were no clear
355 individual features distinguishing the categories. In contrast, the vaccine-induced model
356 comparing protected and non-protected resistant strains did yield more univariate trends
357 in addition to the strong multivariate signature. Namely, both CD4 and CD8 T cells making
358 IL-2 are increased in protected strains, while several lung homogenate cytokines as well
359 as double negative T cells expressing CD107a and producing IL-17 decreased in
360 protected strains (Figure 7B).

361 We next analyzed CC strains that were inherently more susceptible than B6 mice,
362 and were protected by BCG (n=4) or not (n=8) (Figure 7C). This yielded two clear
363 phenotypes of susceptible, non-protected CC strains. One subcluster comprising CC004
364 and CC044 had high levels of T_{CM}, $CD4^+Foxp3^+$, $CD8^+PD\text{-}1^+$ and polyfunctional $CD4^-$
365 $CD8^-$ T cells. The other cluster had high levels of lung MIP-1a and $CD3^-Foxp3^+$ T cells,
366 and fewer polyfunctional $CD4^-CD8^-$ T cells. Interestingly, the unprotected strains with
367 high $CD4^+Foxp3^+$T cells had low $CD8^+Foxp3^+$ T cells, and the unprotected strains with
368 high $CD8^+Foxp3^+$ T cells had low $CD4^+Foxp3^+$ T cells. When looking at the protection

369 model with the susceptible strains, several features were similar to the model with all
370 mice. In this model it is evident that while IL-17 in the lung homogenate decreases with
371 BCG vaccination in protected strains, the amount of IL-17 when stimulated with the P300
372 peptide pool increases. (Figure 7D). Overall, the features that distinguish protected and
373 non-protected strains are similar for naturally resistant and susceptible mice.

374 Finally, we compared susceptible CC strains that were unprotected by BCG (i.e.,
375 poor immunity) to inherently resistant CC strains that were protected by BCG (i.e., good
376 immunity). Unsurprisingly, there is clear separation of categories in the model with all
377 mice. "Poor immunity" CC strains tended to have high levels of inflammatory monokines
378 (Figure 7E, 7F). "Good immunity" CC strains high levels of RANTES after stimulation with
379 anti-CD3. A subset of good immunity strains including CC031 and CC011 also have high
380 frequencies of IFNγ-producing CD8 T cells and various subsets of $CD4^{-}CD8^{-}$ T cells
381 including those producing IFNγ or expressing CD11a. In contrast, a subset of poor
382 immunity strains had high levels of polyfunctional $CD4^{-}CD8^{-}$ T cells and ILCregs. The
383 protection model revealed similar features with the interesting addition of activated CD4
384 T_{RM} cells increasing with BCG vaccination in poor immunity but not good immunity CC
385 strains.

386 We find that immune features can be identified using PLSDA that can categorize
387 the inherent susceptibility and response to BCG of CC strains following Mtb infection. The
388 expression of different immune features was variable between CC strains even in the
389 same protection category, indicating no single path to protection. Looking at more refined
390 subgroups of CC strains allowed us to find features that define subclasses of protected

391 and non-protected strains and highlight features that did not emerge in the PLSR model
392 such as non-T_{RM} effector and central memory T cell populations.
393
394 ***Th1/17 cells detected in a subset of CC mice that are protected by BCG vaccination***
395 An aim of this study was to identify factors that were associated with protection in
396 genetically diverse CC mice after vaccination and Mtb challenge. Vaccination enhanced
397 IL-17 responses in the CC037, CC001, CC072 and CC031 strains, which were all
398 protected by BCG (Figure 5A, 5B). As the CC strains that produced IL-17 after antigen
399 stimulation also had strong IFNγ response, we assessed these T cell responses for dual
400 production of IFNγ and IL-17. BCG vaccinated CC037 mice had the highest IL-17 levels
401 in lung homogenate (Figure 5A, 5B), and after stimulation with P300 or anti-CD3 mAb,
402 CD4 T cells were identified that produced both IFNγ and IL-17 (Figure 8A). Dual $IFN\gamma^{+}IL$-
403 17^{+} CD4 T cells were not detected in C57BL/6 mice after P300 stimulation, regardless of
404 vaccination status. BCG vaccination significantly increased the frequency of $IFN\gamma^{+}IL\text{-}17^{+}$-
405 producing CD4 T cells in the CC037, CC072 and CC001 (Figure 8A, 8B). Further analysis
406 identified BCG primed CD4 T cells from CC037 and CC001 to coexpress both Tbet and
407 RORγt after Mtb infection (Figure 8C, 8D), which are the master transcription factors for
408 Th1 and Th17 cells, respectively.
409 We next measured the polyfunctionality of the T cell response and expanded our
410 cytokine analysis to include IL-2 and TNF. In addition to IFNγ and IL-17, an increase in
411 the proportion of CD4 T cells producing IL-2 or TNF was detected (Figure 8E), indicating
412 that CD4 T cell polyfunctionality was increased in BCG vaccinated mice. A detailed
413 analysis of the different polyfunctional populations revealed different distributions of

414 Th1/17 cells within each CC strain. IFNγ$^+$IL-2$^+$TNF$^+$IL-17$^+$ cells were significantly
415 increased in CC037 following BCG vaccination (Figure S5, red box). Similarly,
416 IFNγ$^+$TNF$^+$IL-17$^+$ cells were increased following BCG vaccination in CC072 and CC001
417 (Figure S5, blue box). BCG led to a decrease in the frequency of IFNγ "single positive'
418 CD4 T cells in all four CC lines (Fig. S5, purple box). However, among these CC strains
419 only CC037 had a significant decrease in IFNγ$^+$TNF$^+$ cells. This contrasts with C57BL/6
420 mice where BCG vaccination had a minimal effect on CD4 T cell polyfunctionality.
421 Significant decreases were observed for IFNγ$^+$TNF$^+$ dual producers and IFNγ single
422 positive cells in Mtb-infected BCG-vaccinated C57BL/6 mice, consistent with the
423 observation that Th1 dominant response is associated with decrease bacterial burden in
424 this background.

425 Finally, we performed a correlation analysis between all 60 immune features and
426 the normalized lung CFU from unvaccinated control and BCG vaccinated mice. For
427 C57BL/6 mice, IFNγ-producing and polyfunctional CD4 T cells correlated with lung CFU.
428 In CC037 mice, IFNγ-producing CD8 T cells correlated with lung CFU. Importantly, many
429 features were negatively correlated with lung bacterial burden in CC037 mice (Figure 8F).
430 The frequency of CD4$^+$Roryt$^+$, and antigen-stimulated IL-17 secretion and IFNγ/IL-17
431 dual-producing CD4 T cells were significant even after correction for multiple testing,
432 suggesting that these features were associated with protection following BCG. There was
433 a strong correlation between the frequency of IFNγ/IL-17 dual-producing CD4 T cells and
434 lung CFU in CC037 mice (Spearman r = -0.8947) (Figure 8G).
435

436 **Discussion**

437 Using CC mice to model an outbred population, our study of 24 CC strains sought
438 to determine how pre-existing immunity to BCG modified the immune response to Mtb
439 infection in a genetically diverse population. An important outcome of this study is our
440 ranking of how BCG induced protection varies across the 24 CC strains were studied,
441 compared to control unvaccinated mice and relative to C57BL/6 mice, which were used
442 as a reference strain. We categorized the different CC strains as being either protected
443 or unprotected by BCG against Mtb infection, based on whether vaccination led to a
444 statistically significant reduction in lung bacillary burden measured four weeks after low
445 dose aerosol infection. Based on whether the different mice were susceptible or resistant,
446 and protected or not, relative to C57BL/6 mice, we divided the 24 CC strains into four
447 phenotypes (i.e., quadrants). We found that different immune parameters were
448 associated with protection in the four different groups. The surprising result was that BCG
449 does not protect all mouse genotypes against Mtb, and no protection was detected in
450 nearly half of the CC strains tested. Thus, these results extend our analysis of the eight
451 CC founder strains, which also varied in their ability to be protected by BCG vaccination
452 (*15*). Not only do environmental mycobacteria and genetic variations in BCG strains used
453 around the world interfere with the ability of BCG-induced protection, but we report now
454 that host genetics should be considered a third important barrier that prevents vaccine-
455 induced protection to TB.

456 An important question is why BCG doesn't induce protection in all CC strains?
457 Protection was largely defined based on reductions in lung CFU. Four weeks after low
458 dose Mtb infection was chosen as it is a standard time point used in many murine vaccine

459 studies. Although the complexity of this experiment precluded performing other time
460 points, it is possible that our rankings may have changed if other time points were used,
461 just as the rankings differ for lung vs. splenic CFU reduction. Interestingly, the CC strains
462 that were unprotected by BCG can be subdivided based on their lung pathology and
463 immune responses. None of the unprotected CC strains had immune responses that were
464 modified by BCG vaccination, but some (e.g., CC004, CC040) had changes in lung
465 histology (e.g., more lymphocytic infiltrates) that suggested that a memory recall
466 response. Thus, we suggest that BCG induced an immune response in some CC strains
467 (e.g., CC004, CC040), but it was insufficient to control lung CFU. Other strains may have
468 generated a suboptimal response to vaccination (i.e., low responders; CC003, CC024,
469 CC041).

470 We believe that the variation in host responses makes this an ideal platform for
471 testing the efficacy of preclinical vaccines in the context of a genetically diverse
472 background. Only a fraction of individuals infected with Mtb will develop disease. While
473 there exist well-defined risk factors for the development of TB, the factors that drive the
474 emergence of disease is largely unknown in most people. Similarly, the factors that
475 influence BCG efficacy remains controversial. In addition to risk factors that disrupt cell
476 mediated immunity, there are also genetic factors (i.e., MSMD, polygenic), which
477 predispose certain people to TB. Similar genetic factors could also contribute to variation
478 in BCG efficacy against pulmonary Mtb. Our analysis shows that while certain mouse
479 genotypes cannot be protected by BCG, the inherent susceptibility or resistance of the
480 mouse strain is not a predictor of vaccine efficacy. One caveat is that we have only tested
481 the efficacy of one vaccine in the CC strains. It is possible that vaccines that work through

482 alternate mechanisms may be able to protect CC mice that are refractory to BCG
483 vaccination.

484 Genetics affects host resistance to TB (*16*). Our studies show how dramatically
485 host genetics affects vaccine-induced immunity. Different inbred recombinant mouse
486 strains derived from the same eight founders can be protected or not by BCG. Importantly,
487 the ability of these mouse strains to be protected by BCG was largely independent of their
488 inherent susceptibility to Mtb infection. This suggests that a priori, vaccines are
489 appropriate interventions even in individuals that might have a genetic susceptibility to
490 TB. Importantly, we believe that diverse human populations include genotypes that
491 generate suboptimal, or even detrimental, responses to BCG. We expect that our data
492 will help to establish CC strains as a novel animal resource that can facilitate the discovery
493 of correlates of vaccine-induced protection and accelerate preclinical vaccine testing.
494 Instead of optimizing vaccine concepts in a single mouse genotype (e.g., C57BL/6 mice),
495 identifying vaccines that are effective in numerous different genotypes (i.e., genetically
496 distinct individuals), as modeled by CC or DO mice, represents an alternative paradigm
497 for preclinical vaccine development.

498 Levels of IL-17 in lung homogenate is strongly correlated with lung CFU in CC
499 strains, co-varies with IL-1β and IL-6, two cytokines that are involved in skewing of Th17
500 cells, and with TNF, KC, MIP2, and G-CSF, all which are induced by IL-17. In BCG
501 vaccinated CC strains that were protected, the correlation with IL-17 was lost. We
502 suggest that like IFNγ, IL-17 levels are a correlate of disease. In contrast, in three of the
503 12 protected strains (i.e., CC037, CC072 and CC001), we detected CD4 T cells that
504 produced IFNγ and IL-17 after Mtb infection. These cells are phenotypically similar to

505 Th1/17 cells, or Th1*, that have been identified following TB infection (*25, 26*), and have
506 recently been associated with late-forming, low burden granulomas in a NHP TB infection
507 model (*27*). None of the 13 unprotected strains had this phenotype, and these dual
508 producers were detected only in BCG vaccinated mice and not in sham controls, nor were
509 they detected in the C57BL/6 reference strain. Thus, this is an example of an immune
510 response that BCG primes but is not manifested until the CD4 T cell response is recalled
511 by Mtb infection.

512 Finally, we wished to identify the components of the immune response stimulated
513 by BCG, which were subsequently recalled after Mtb infection and associated with
514 protection. We hypothesized that CC strains protected by BCG would share common
515 immunological features. We postulated that by comparing protected to unprotected CC
516 strains would allow us to filter out components of the immune response that were elicited
517 by vaccination or infection but not associated with protection. As BCG is known to elicit T
518 cell responses that can mediate protection, we extensively characterized the T cell
519 immunity. Analysis of the different CC strains showed that there was considerable
520 diversity in the composition of the T cells recruited and maintained in the lung after
521 infection. However, most of the variability was determined by the host genetic background
522 and BCG vaccination had little impact on the T cell subsets recruited to the lung. In
523 contrast, several of the CC strains protected by BCG vaccination had evidence of
524 functional skewing of the T cell responses. Despite there being several different flavors
525 of immunity, we were able to identify an immune signature that was associated with BCG-
526 elicited protection against TB.
527

528 **Acknowledgements**

529 The following reagents were obtained through the NIH Tetramer Core Facility (contract number
530 75N93020D00005) for providing MR-1 and CD1d tetramers. We thank the Vanderbilt Medical
531 Center Digital Shared Histology Resource (DHSR) for scanning and hosting our histology data.

532

533 **Author contribution**

534

535 **Declaration of interests**

536 The authors declare no competing interests.

537

538 **Funding**

539 Supported by P01 AI123286 (C.M.S., M.F., S.M.B,), Digitization of glass slides was funded by
540 R01 HL14541 (G.L.B).

541

542

543 **Materials and Methods**

544

545 Mice

546 Female C57BL/6/J (#0664) were purchased from The Jackson Laboratory. Female mice
547 from 24 CC strains were purchased from the UNC Systems Genetics Core Facility
548 (SGCF) between July 2018 and August 2019. The 24 CC strains used in this study
549 include: CC001/Unc, CC002/Unc, CC003/Unc, CC004/TauUnc, CC010/GeniUnc,
550 CC011/Unc, CC012/GeniUnc, CC023/GeniUnc, CC024/GeniUnc, CC025/GeniUnc,
551 CC026/GeniUnc, CC031/GeniUnc, CC032/GeniUnc, CC035/Unc, CC036/GemUnc,
552 CC037/TauUnc, CC040/TauUnc, CC041/TauUnc, CC042/GeniUnc, CC044/Unc,
553 CC045/GeniUnc, CC059/TauUnc, CC061/GeniUnc,, CC072/TauUnc. More information
554 regarding the CC strains can be found at
555 http://csbio.unc.edu/CCstatus/index.py?run=AvailableLines.information.
556 All animal studies were conducted using the relevant guidelines and regulations, and
557 approved by the Institutional Animal Care and Use Committee at the University of
558 Massachusetts Medical School (UMMS) (Animal Welfare A3306-01), using the
559 recommendations from the Guide for the Care and Use of Laboratory Animals of the
560 National Institutes of Health and the Office of Laboratory Animal Welfare.

561

562 Bacterial strains

563 Bacteria expressing yellow fluorescent protein (sfYFP) were generated as described (17).
564 Mice were vaccinated with 105 CFU BCG (BCG-SSI strain, resuspended in 0.04%
565 Tween/PBS) subcutaneously (100 uL/injection).

566 Aerosolized Mtb infection of mice

567 Mice were infected by the aerosol route as previously described (*17*). Frozen bacterial
568 stocks were thawed and sonicated for 1 minute and then diluted into 5 ml of 0.01%
569 Tween-80 in PBS. The diluted bacterial suspension was aerosolized to infect mice using
570 a Glas-Col chamber (Terre Haute). The average number of bacteria delivered into the
571 lung was determined for each experiment by plating lung homogenate from 4-5 mice 24
572 hours after infection and ranged between 40-150 CFU/mouse.

573

574 Bacterial burden in lung and spleen

575 Infected mice were euthanized at pre-determined timepoints, and the left lung or whole
576 spleen were homogenized using 2 mm zirconium oxide beads (Next Advance) in a
577 FastPrep homogenizer (MP Biomedicals). Tissue homogenates were serially diluted and
578 plated on 7H11 agar plates (Hardy Diagnosis). CFU was enumerated after 19-21 days
579 of incubation at 37°C and 5% CO_2.

580

581 Lung cell preparation

582 Single cell suspensions were prepared by homogenizing lungs using a GentleMACS
583 tissue dissociator (Miltenyi), digesting with 300 U/ml collagenase (Sigma) in complete
584 RPMI (10% FBS, 2 mM L-Glutamine, 100 units/ml Penicillin/Streptomycin, 1 mM Na-
585 Pyruvate, 1X Non-essential amino acids, 0.5X Minimal essential amino acids, 25 mM of
586 HEPES, and 7.5 mM of NaOH) at 37°C for 30 minutes, and followed by a second run of
587 dissociation using the GentleMACS. Suspensions were then sequentially filtered through
588 70 μm, treated with ACK lysis buffer (Gibco), then filtered through 40 μm strainers.

589

590 Flow cytometric analysis

591 Cells were first stained with Zombie Fixable Viability dye (Biolegend) for 10 minutes at
592 room temperature (RT), after which cells were stained with antibodies in autoMACS
593 running buffer (Miltenyi) containing 5 ug/ml of anti-mouse CD16/32 (BioXcell) for 10
594 minutes at 4°C. In some experiments, MAIT cells were identified using the MR-1-PE
595 tetramer (NIH tetramer core facility) at 37°C for 1hr. Cell were then stained with surface
596 antibody cocktail for 20 minutes at 4°C. In some experiments, CD1d-APC tetramer (NIH
597 tetramer core facility) were also added at this time to identify NKT cells. To measure
598 transcription factor (TF) expression, cells were fixed and permeabilized for 30 minutes at
599 RT, followed by staining for 30 minutes with antibodies to the different TF at RT using the
600 Foxp3/TF staining buffer set (ThermoFisher). Samples were acquired on Aurora (Cytek)
601 or MACSquant (Miltenyi).

602

603 For intracellular staining, cells were first stimulated for 5 hours with either anti-CD3
604 antibodies (1 ug/mL, Biolegend) or P300 peptide megapool (*23*) which contains 300
605 peptides representing 90 antigens, and contains epitopes recognized by both murine CD4
606 and CD8 T cells. GolgiPlug (BD Biosciences) was also given at this time (0.5ug/well).
607 Viability and surface staining were performed as described, after which cells were fixed
608 instead with Fixation/Permeabilization kit for 20 minutes at 4°C. Cells were then stained
609 with an intracellular cytokine antibody cocktail for 20 minutes at 4°C.

610

611 To inactivate the bacteria, samples were fixed with 1% paraformaldehyde/PBS for 1 hour
612 at room temperature and then washed with MACS buffer. Flow data were analyzed using
613 FlowJo v10.7.1. Polyfunctionality was examined using the SPICE platform (*28*).
614
615 LEGENDplex cytokine analysis
616 Lung cells were stimulated with P300 peptide megapool (1ug/mL) or anti-CD3 (1ug/mL)
617 at a concentration of 10^6 cells/mL for 24hrs. Culture supernatants were filtered through
618 0.2um filter, after which cytokine levels were measured with LEGENDplex kits from
619 Biolegend. Specifically, we used their mouse T helper cytokine panel (IFNγ, TNF, IL-2,
620 IL-4, IL-5, IL-6, IL-9, IL-10, IL-13, IL-17a, IL-17f, IL-21, IL-22) as well as their mouse
621 proinflammatory chemokine panel (RANTES, MIP-3α, Eotaxin, TARC, KC, MCP-1, MIG,
622 IP-10, MIP-1α, MIP-1β, BLC, LIX, MDC) for a total of 26 cytokines and chemokines.
623
624 Lung homogenate
625 Cytokines and chemokiens were measured from lung homogenates using the 31-plex
626 discovery Assay (MD31) from Eve technologies (Calgary, Canada).
627
628 Histology analysis
629 The right middle lung lobe from each mouse was immersion fixed in 10% formalin
630 immediately following euthanasia. Lung tissue was processed, embedded in paraffin,
631 sectioned at 5um, and stained with hematoxylin and eosin at the Morphology Core Facility
632 at UMass Chan Medical School for histological examination by a board-certified
633 veterinary pathologist (GB). Blind classification (BCG vs PBS) within each CC strain was

634 determined based on the assumption that BCG vaccination would have three possible
635 effects on lungs: (i) reduce necrosis/neutrophilic infiltrates; (ii) increase lymphocytes
636 and/or plasma cell abundance; and (iii) reduce granuloma size. Accuracy of the blind
637 classification was assessed by an independent investigator (AO). All glass slides were
638 subsequently re-evaluated to more closely examine and describe histopathology features
639 attributable to BCG vaccination and attempt to find unique responses. Glass slides were
640 then digitized by Vanderbilt University Medical Center (Nashville, TN) and whole images
641 generated using Aperio ScanScope (Buffalo Grove, IL, USA) at 400 times normal
642 magnification with jpeg compression. Representative images were captured using Aperio
643 ImageScope magnified 40 and 400 times and imported into GraphPad Prism 9 to
644 generate the panel Figure. Digital images were not adjusted. Whole slide imaging and
645 quantification of immunostaining were performed in the Digital Histology Shared
646 Resource (DHSR) at Vanderbilt University Medical Center.
647
648 Univariate correlations
649 To evaluate how C57BL/6 mice, protected CC strains, and non-protected CC strains
650 compared in their relationships between lung cytokine production and Mtb burden, a
651 series of univariate Pearson correlations were conducted. Lung homogenate cytokines
652 and chemokines measurements in the form of concentrations were natural logged and
653 each of their Pearson correlations with the log10 lung CFU or spleen CFU were
654 evaluated, with both the cytokine and CFU measurements separated into groups of
655 C57BL/6, protected CC strains, and non-protected CC strains. A Bonferroni multiple
656 hypothesis correction was applied to account for the number of correlations calculated.

657

658 Multivariate analyses

659 Multivariate analyses involved combining all four immunological datasets in this study:
660 the lung homogenate cytokine/chemokine measurements, the ICS, the LEGENDplex
661 cytokine analysis, and the T cell phenotypic flow cytometry. The analyses also included
662 all CC mice used in the study in addition to the control C57/BL6 mice.

663

664 For the ICS, the data were analyzed as combinatorics of the seven markers measured in
665 order to best capture polyfunctionality, and further divided by CD4 T cell, CD8 T cell, $CD4^-$
666 $CD8^-$ T cell, and $CD3^-$ categories. The features were filtered by a minimum mean and
667 variance percentage of 0.000001. Only P300 stimulated samples were used from the ICS
668 dataset. The flow cytometry data was in the form of percentages of $CD3^+$ or $CD3^-$ cells,
669 while the immunoassay data was in the form of concentrations. The combined dataset
670 was natural log transformed and subsequently z-scored by measurement across all
671 samples. For models that involved BCG-induced protection, these z-scored values were
672 further transformed by subtracting the features from each vaccinated mouse and the
673 mean of the features of the unvaccinated mice from the same mouse strain.

674

675 For the PLSR and PLSDA models, elastic net feature selection was performed in 100
676 trials of 5-fold cross validation, where features that were present in at least 80% of trials
677 were included in the final model. Each model was compared to two null models: one with
678 permuted lung CFUs and one with randomly chosen features. The mean squared error

679 of the prediction of lung CFU (for the PLSR) or the classification accuracy (for the
680 PLSDAs) was compared to that of the null models in 50 trials of 5-fold cross validation.
681
682 Quantification and Statistical Analysis.
683 Statistical analysis was performed using Graphpad Prism 9. P-values were calculated
684 using unpaired t test, one-way ANOVA, or two-way ANOVA as indicated in the figure
685 legends. Multivariate analyses were performed in R (v3.6.0 and 4.0.4) using the ropls
686 Bioconductor package.
687
688
689
690

691

CC strain[a]	Batch[b]	MHC Haplotype[c]	ΔLung CFU_1 [d] reduction (logCFU)	ΔLung CFU_2 [d] reduction (logCFU)	ΔSpleen CFU_1 [e] reduction (logCFU)	ΔSpleen CFU_2 [e] reduction (logCFU)
CC001/Unc	4	b	0.94	1.45	2.79	1.45
CC002/Unc	3	b	0.83	n/a	0.46	n/a
CC003/Unc	3	g	0.07	-0.61	0.77	0.15
CC004/TauUnc	4	b	0.05	0.08	-0.16	-0.12
CC010/GeniUnc	1	b	0.53	n/a	0.92	n/a
CC011/Unc	3	g7	1.04	n/a	2.08	n/a
CC012/GeniUnc	4	wsb	0.79	n/a	0.89	n/a
CC023/GeniUnc	1	b	1.17	-0.09	1.47	0.32
CC024/GeniUnc	2	cast	0.15	0.24	0.38	0.62
CC025/GeniUnc	3	pwk	1.40	n/a	0.43	n/a
CC026/GeniUnc	2	cast	0.93	1.05	3.75	1.87
CC031/GeniUnc	3	b	1.22	1.35	1.26	1.60
CC032/GeniUnc	1	b	1.16	n/a	1.06	n/a
CC035/Unc	4	wsb	0.91	n/a	1.13	n/a
CC036/Unc	3	g7	0.62	n/a	1.84	n/a
CC037/TauUnc	1	b	1.37	1.84	1.07	1.05
CC040/TauUnc	2	b	-0.64	-0.28	-0.35	1.14
CC041/TauUnc	2	b	0.14	0.94	0.71	0.94
CC042/GeniUnc	1	b	1.65	0.96	1.00	1.32
CC044/Unc	4	b/g7	0.08	0.13	0.12	0.39
CC045/GeniUnc	4	z	0.24	n/a	0.13	n/a
CC059/TauUnc	2	b/a	1.39	n/a	1.17	n/a
CC061/GeniUnc	1	b	0.93	n/a	1.58	n/a
CC072/TauUnc	2	b	0.97	1.37	1.06	1.41

692

693 **Table 1. A description of the CC analysis batches.**

694 [a] A group of C57BL/6 mice were included in all batches as a reference strain. Data
695 was normalized to internal C57BL/6 control as indicated in the methods.
696 [b] The first cohort of CC mice were analyzed in 4 separate batches as indicated.
697 [c] MHC haplotype of each CC mouse as inherited from CC founders.
698 [d, e] Thirteen of the 24 CC strains were analyzed a second time to validate the
699 original results and perform more in-depth immune analysis. ΔLungCFU[d] and
700 ΔSpleenCFU[e] for the first experiment and the repeat experiment are listed. Mice
701 that were exacerbated by BCG vaccination are highlighted in red.
702

[a] Strain	[b] BCG-induced protection	[c] Change in pathology	[d] How did the vaccinated mice differ?
CC003	N	N	No differences. Both groups had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions
CC041	N	N	No differences. Both groups had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions
CC024	N	N	No differences. Both groups had small hard-to-find lymphocytic lesions with perivascular and peribroncial lymphocytic cuffs
CC002	N	Y	Ctrl had mild lesions with perivascular lymphocytic cuffs. BCG mice had larger moderate sized macrophage-rich lesions. Lymphocytes were similarly abundant.
CC004	N	Y	Ctrl mice had mild to moderate lesions with lymphohistiocytic infiltrates. BCG mice have more lymphocytes with regions of dense perivascular/peribronchiolar lymphocytic cuffs.
CC036	N	Y	Ctrl mice have mild disease with lymphohistiocytic infiltrates. BCG mice also have lymphohistiocytic infiltrates with few very dense lymphocytic areas
CC040	N	Y	Ctrl mice have moderate necrotizing lesions with few lymphocytes. BCG mice have mild disease - small & hard to find lesions that are lymphocyte rich without necrosis.
CC044	N	Y	Unusual. Ctrl mice have lymphohistiocytic infiltrates with dense lymphocytic cuffs. BCG mice have lymphohistiocytic infiltrates with dense lymphocytic cuffs and lymphoid aggregates/follicules in parenchyma.
CC061	P	N	No differences. Unusual. Mild lesions. Both groups have dense lymphoid aggregates/follicles in parenchyma. Also, perivascular lymphocytic cuffs.
CC072	P	N	No differences. Mild disease. perivascular lymphocytic cuffs with a few macrophage-rich lesions. One mouse has a foreign-body granulomatous bronchitis/pneumonia with intralesional food/bedding/plant.
C57BL/6	P	Y	Ctrl mice had mild to moderate lymphohistiocytic lesions. BCG mice had with more, dense lymphocytic infiltrates
CC001	P	Y	Ctrl mice had moderate lesions with lymphohistiocytic infiltrates and equivocal necrosis. BCG mice had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions without necrosis
CC023	P	Y	Ctrl mice had moderate to severe lesions with necrosis and minimal lymphocytes. BCG mice have moderate lesions with more lymphocytes and less necrosis.
CC025	P	Y	Ctrl mice have moderate disease with mild lymphocytic infiltrates and necrosis. BCG mice have abundant dense perivascular/peribronchiolar lymphocytic ifiltrates and lymphohistiocytic granulomas and less necrosis.
CC026	P	Y	Ctrl mice had moderate lesions with lymphohistiocytic infiltrates. BCG mice had mild lesions with lymphohistiocytic infiltrates and a few very dense lymphocytic areas
CC031	P	Y	Ctrl mice have moderate disease with lymphohistiocytic infiltrates. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas
CC032	Protected	Y	Ctrl mice have mild disease with lymphohistiocytic infiltrates. BCG mice have more lymphocyte rich lesions that are small & hard to find
CC037	Protected	Y	Ctrl mice have moderate disease with lymphohistiocytic infiltrates. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas
CC042	Protected	Y	Ctrl mice have mild to moderate disease with few lymphocytes and necrosis. BCG mice have mild disease with more lymphocytes, less necrosis
CC059	Protected	Y	Ctrl mice had mild disease with lymphohistiocytic infiltrates with small regions of necrosis. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas without necrosis

Table 2. Summary of histological changes of CC mice.

[a] Nineteen of the 24 CC strains were available for analysis.

[b] The effect of BCG on the lung CFU is from Figure 1.

[c] Assessment by a pathologist of whether differences existed between sham and BCG vaccinated mice after Mtb infection (n=5/strain/condition).

[d] Summary of observed differences between the groups.

711 **Figure legends**

712 Figure 1. BCG-induced protection in 24 CC strains and C57BL/6 mice.

713 (A) Lung and (D) spleen CFU of all 24 CC mice (n=120/group; top) or all C57BL/6 mice
714 (n=20/group; bottom) at 4 weeks post infection. Mice were vaccinated with BCG or
715 administered PBS and rested for 12 weeks prior to infection. Each point represents an
716 individual mouse. Line, median. Student's t-test; ****, $p < 0.0001$.

717 (B) Protection in the lung and (E) was assessed by comparing the CFU in BCG vs. sham
718 vaccinated mice for each individual mouse strain (n=5/group) using the Benjamini-
719 Hochberg procedure to determine the false discovery rate (FDR) (top). The ΔCFU of BCG
720 vs. sham vaccinated was calculated (bottom). The open bar is the reference strain
721 C57BL/6 strain (n=20/group). *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

722 (C) Correlation in the lung and (F) spleen between susceptibility and protection. The
723 relative susceptibility (y-axis, defined by CFU in unvaccinated group, relative to the
724 C57BL/6 reference strain) and the protection conferred by BCG (x-axis, defined by ΔCFU
725 of vaccinated mice compared to unvaccinated controls, relative to the C57BL/6 reference
726 strain). The graph is centered on C57BL/6 mice, leading to the different CC strains being
727 placed in four quadrants based on their relative susceptibility and protection.

728

729 Figure 2. Lung histology of CC mice after BCG vaccination and Mtb challenge

730 Representative lung histology of three CC strains (CC025, CC037, CC072) that were
731 protected by BCG, after sham or BCG vaccination and Mtb challenge (Mtb and BCG+Mtb,
732 respectively) (left). Representative lung histology of three CC strains (CC003, CC004,

733 CC024) that were unprotected by BCG (right). Original magnification, 40X (inset, 400X).
734 Scale bar, 500 μm (inset, 50 μm).
735
736 Figure 3. BCG-protected CC mice have an altered lung microenvironment of CC mice
737 after Mtb infection.
738 (A) IFNγ levels were measured in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks
739 post infection. A two-way ANOVA was performed using Sidak's correction for multiple
740 comparisons. ***, $p < 0.001$; ****, $p < 0.0001$".
741 (B) – (F) Pearson correlation between lung CFU and (B) IFNγ, IL-17, TNF, IL-1β, KC and IL-6 (C)
742 LIX (D) RANTES (E) MIP-1β and (F) M-CSF, MIP-2, G-CSF, LIF, comparing PBS and BCG
743 vaccinated mice at 4 weeks post infection in C57BL/6 mice (B6), non-protected (NP) or protected
744 (P) CC mice.
745
746 Figure 4. Frequency of various T cell populations in protected and unprotected CC mice
747 following Mtb infection.
748 (A) Proportion of $CD3^+$ cell, $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells in the lungs of PBS and
749 BCG vaccinated CC mice at 4 weeks post infection.
750 (B) Proportion of memory CD4 and CD8 T cells in the lungs of either PBS or BCG
751 vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ and various
752 combinations of CD62L and CD127 expression; $CD62L^+$ $CD127^+$ (central memory),
753 $CD62L^-$ $CD127^-$ (effector), $CD62L^-$ $CD127^+$ (effector memory).
754 (C) Proportion of resident memory CD4 and CD8 T cells in the lungs of either PBS or
755 BCG vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ $CD103^+$ $CD69^+$.

756 (D) Proportion of Th1, Treg and Th17 cells in the lungs of either PBS or BCG vaccinated
757 CC mice at 4 weeks post infection, as defined by $CD44^+$ and $Tbet^+$ (Th1), $Foxp3^+$ (Treg)
758 or RoRγt (Th17) expression.
759 (E) Proportion of CD4 and CD8 T cells expressing either CXCR3 or CX_3CR1 in the lungs
760 of either PBS or BCG vaccinated CC mice at 4 weeks post infection.
761 (F) PCA of the T cell phenotyping data. Each point represents the average of usually 5
762 mice within a given mouse strain that are either BCG vaccinated or unvaccinated. Open
763 symbols, unvaccinated; closed symbols, BCG vaccinated. Each color represents a
764 different mouse strain.
765 (A–E) The data represent the mean ± SEM from one experiment. Two-way analysis of
766 variance with the Benjamini and Hochberg multiple comparsion method. The FDR was
767 set to 0.05 and the numbers in the figures are the q value.
768
769 Figure 5. Cytokine response in protected and unprotected CC strains following Mtb
770 infection.
771 (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated
772 for 24 hours with the P300 peptide pool. IFNγ, IL-2, TNF and IL-17a were measured in
773 the supernatants.
774 (B) and (C) Cells from the lungs of CC mice at 4 weeks post infection were stimulated for
775 5 hours with the P300 peptide pool. Frequencies of CD4 or CD8 T cells that expressed
776 IFNγ, IL-2, TNF or IL-17 was determined using intracellular cytokine staining.

(A–C) The data represent the mean ± SEM from one experiment. Two-way analysis of variance with Sidak's multiple comparison test was used to determine significance. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Figure 6. Multivariate correlates of protection across all 24 CC strains

(A) A partial least squares regression (PLSR) model was performed to associate BCG-induced changes in lung immune features with BCG-induced changes in lung CFU. Each data point represents a vaccinated mouse that has the mean of the unvaccinated mice from the same strain subtracted from it. The scores of the top two latent variables of the model are shown, with the color of each data point representing the changed in lung CFU induced by BCG vaccination. The percents of the variance in the lung features captured by each latent variable are shown as percentages in the axes.

(B) The bar graph depicts the loadings of the first latent variable of the PLSR model in (A). Positive loadings indicate immune features that increase from BCG vaccination in samples that are on the positive x-axis in the scores plot, which generally indicates they have worse protection. Negative loadings conversely indicate immune features that generally increase from BCG vaccination in samples that are protected by vaccination.

(C) The heatmap illustrates all features selected to be included in the model, indicating they were important for associating strains with the lung protection continuum. Colors are representations of z-scored features, with a different color representing each 10% quantile of the data. Columns are averages of the mice in the CC strains (n=5 for most cases). Feature names are colored based on the dataset from which they come. In (B)

799 and (C), features from the ICS dataset indicate production of the cytokines listed and no
800 production of any measured cytokines not listed.
801
802 Figure 7. Identifying signatures of protection in subclasses of CC strains
803 (A), (C), (E) Scores plot (top) and heatmaps (bottom) of selected features for PLSDA
804 models. Each dot represents a single mouse. The two colors in the scores plots signify
805 the two classes being compared. An ellipse is drawn around the 95% confidence interval
806 for each class. Percent of variance of the immune feature data is shown on the axes.
807 Heatmaps show features that were most important for distinguishing the two classes.
808 Colors are representations of z-scored features. Columns are averages of the mice in the
809 CC strains, with vaccinated and unvaccinated mice being separated (n=5 for most cases).
810 Feature names are colored based on the dataset from which they come (blue=T cell
811 phenotyping, brown=lung homogenate cytokines, green=ICS, magenta=LEGENDplex).
812 NP = non-protected, P = protected, R = resistant, S = susceptible.
813 (B), (D), (F) PLSDA models performed as in (A), (C), and (E), except that features are the
814 changes induced by BCG vaccination as in Figure 6.
815
816 Figure 8. Th1/17 cells correlate with BCG-mediated protection in a subset of protected
817 CC mice.
818 (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated
819 for 5 hours with either aCD3, P300 peptide pool or media control. Representative flow
820 plots of IFNγ and IL-17 producing CD4 T cells from CC037 and C57BL/6 are shown. Data
821 is representative of two independent experiments.

822 (B) Frequencies of CD4 T cells that expressed both IFNγ and IL-17 was determined using
823 intracellular cytokine staining. Data is representative of two independent experiments.
824 One-way analysis of variance with Sidak's multiple comparison test was used to
825 determine significance. **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.
826 (C) Representative flow plots of CD4 T cells expressing either T-bet or RORγt in CC037
827 or C57BL/6 mice. Data is representative of two independent experiments.
828 (D) Frequencies of CD4 T cells that expressed both T-bet or RORγt in CC037 or C57BL/6
829 mice. Data is representative of two independent experiments. One-way analysis of
830 variance with Sidak's multiple comparison test was used to determine significance. **, p
831 < 0.01; ***, $p < 0.001$.
832 (E) Proportion of CD4 T cells that express various combinations of IFNγ, TNF, IL-2 or IL-
833 17 in either PBS or BCG vaccinated mice at 4 weeks post Mtb infection. Plots are
834 generated using the SPICE platform (*28*).
835 (F) Volcano plots. Correlation between 60 immune features and the lung CFU determined
836 for CC037 and C57BL/6 mice (from Batch 1). Dotted line, $p = 0.05$. *, $q < 0.05$, corrected
837 with the Benjamini and Hochberg multiple comparison method (FDR = 0.05)
838 (G) Correlation of $IFN\gamma^{+}IL\text{-}17^{+}$ cells and bacterial burden in the lungs of either PBS or
839 BCG vaccinated CC037 mice at 4 weeks post Mtb infection. Colors represent BCG
840 vaccinated mice (red) or unvaccinated mice (blue). Spearman *r* and p value are shown.
841
842

843 **SUPPLEMENTAL INFORMATION**

844

845 **Supplementary tables**

846 Table S1. –Pathological descriptions of changes lung histology after BCG vaccination in
847 CC mice.

848 Table S2. A description of select CC strains that were repeated for in-depth immune
849 analysis.

850

851 **Supplementary figure legends**

852 Figure S1. Gating strategy for the analysis of the cell surface T cell phenotype.

853 (A) Lymphocytes were identified using FSC and SSC properties, after which viability dyes
854 were used to identify live cells. Singlet gates were used to exclude doublets, after which
855 T cells and non-T cells were identified using the lineage marker CD3.

856 (B) NKT and MAIT cells were identified from the $CD3^+$ gate using their antigen-loaded
857 CD1d and MR1 tetramers, respectively. $CD4^+$, $CD8^+$ and $CD4^-CD8^-$ T cells were also
858 identified from this population for further analysis.

859 (C) Expression of various markers were also examined within the $CD3^-$ population. As an
860 example, flow plot represents expression of T-bet within the $CD3^-$ population.

861 (D) Identification of memory T cells. CD44 was used to identify antigen-experienced T
862 cells. Resident memory T cells (T_{RM}) were identified based on CD103 expression. From
863 the $CD103^-$ fraction, CD127 and CD62L expression identified central memory (T_{CM}),
864 effector (T_E) and effector memory (T_{EM}). Subset identification of various T cell populations

865 (ex: Tbet^{+}CD4^{+} T cells) as well as activation markers (ex: PD1, CD11a) were also
866 examined in CD4^{+}, CD8^{+} and CD4^{-}CD8^{-} T cell subsets.
867 (E) Naïve T cells identified within the CD44^{-} compartment using CD62L expression.
868
869 Figure S2. Frequencies of non-conventional T cells in CC mice. Frequencies of MR-1-
870 restricted MAIT cells and CD1d-restricted NKT cells were measured in the lungs of PBS
871 or BCG vaccinated CC mice at 4 weeks post Mtb infection. The data in the graphs
872 represent the mean ± SD of one experiment. Two-way analysis of variance with the
873 Benjamini and Hochberg multiple comparsion method. The FDR was set to 0.05 and the
874 numbers in the figures are the q value.
875
876 Figure S3. Gating strategy for the analysis of T cell function (intracellular cytokine
877 staining).
878 (A) Lymphocytes were identified as using FSC and SSC properties, after which viability
879 dyes were used to identify live cells. Singlet gates were used to exclude doublets, after
880 which T cells were identified using the lineage marker CD3.
881 (B) CD4, CD8 and CD4^{-}CD8^{-} T cells were also identified from the CD3^{+} population for
882 further analysis. Cytokine expression (ex: IFNγ) were examined in all 3 T cell sub-
883 populations.
884 (C) Cytokine expression was also examined within the CD3^{-} population.
885
886 Figure S4. Multivariate model accuracies

887 (A) The accuracy of our protection model (Figure 6) was compared to that of two different
888 types of null model, where our measure of accuracy was mean squared error of prediction
889 of Δlog (lung CFU). Our PLSR model performed significantly better than both a null model
890 consisting of random features instead of our elastic-net chosen features and a null model
891 where the outcomes of Δlog (lung CFU) were scrambled. Our PLSR model and each of
892 the null models were run 50 times, and each dot represents one of these trials. Labeled
893 p-values are from one-way ANOVAs.
894 (B) Our models highlighting the immune features that distinguish pairs of
895 protection/susceptibility categories (Figure 7) were compared to two null models as in (A).
896 Each of our models, both the ones involving all mice and the protection models were
897 significantly better than both null models. Here, the accuracy metric is classification
898 accuracy. Labeled p-values are from one-way ANOVAs.
899
900 Figure S5. Examining the Polyfunctionality of CD4 T cells in select CC strains. Boolean
901 gating was used to identify various sub-populations amongst the CD4 T cells following
902 P300 stimulation in certain CC mice. An increase in various IFNγ^{+}IL-17^{+} CD4 T cells (that
903 also produced TNF or IL-2) were detected in CC037, CC072, CC031 and CC001 but not
904 in C57BL/6. In contrast, a noticeable decrease in the frequency of IFNγ "single positive"
905 CD4 T cells was observed in all strains including C57BL/6. The data in the graphs
906 represent the mean ± SD of two experiment. One-way analysis of variance with Sidak's
907 multiple comparison test was used to determine significance. *, $p < 0.05$; **, $p < 0.01$; ***,
908 $p < 0.001$; ****, $p < 0.0001$.
909

References

1. WHO, Global Tuberculosis Report 2021. (2021).

2. G. A. Colditz *et al.*, Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* **271**, 698-702 (1994).

3. A. Roy *et al.*, Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ* **349**, g4643 (2014).

4. P. Mangtani *et al.*, Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis* **58**, 470-480 (2014).

5. I. Abubakar *et al.*, Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guerin vaccination against tuberculosis. *Health Technol Assess* **17**, 1-372, v-vi (2013).

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

7. J. L. Flynn, H. P. Gideon, J. T. Mattila, P. L. Lin, Immunology studies in non-human primate models of tuberculosis. *Immunol. Rev.* **264**, 60-73 (2015).

8. M. Ahmed *et al.*, Immune correlates of tuberculosis disease and risk translate across species. *Sci Transl Med* **12**, (2020).

9. C. Aagaard *et al.*, Immunization with Mycobacterium tuberculosis-Specific Antigens Bypasses T Cell Differentiation from Prior Bacillus Calmette-Guerin Vaccination and Improves Protection in Mice. *J Immunol* **205**, 2146-2155 (2020).

10. J. S. Woodworth *et al.*, A Mycobacterium tuberculosis-specific subunit vaccine that provides synergistic immunity upon co-administration with Bacillus Calmette-Guerin. *Nat Commun* **12**, 6658 (2021).

11. I. Kramnik, G. Beamer, Mouse models of human TB pathology: roles in the analysis of necrosis and the development of host-directed therapies. *Semin Immunopathol* **38**, 221-237 (2016).

12. J. Nemeth *et al.*, Contained Mycobacterium tuberculosis infection induces concomitant and heterologous protection. *PLoS Pathog* **16**, e1008655 (2020).

13. K. E. Noll, M. T. Ferris, M. T. Heise, The Collaborative Cross: A Systems Genetics Resource for Studying Host-Pathogen Interactions. *Cell Host Microbe* **25**, 484-498 (2019).

14. G. A. Churchill *et al.*, The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* **36**, 1133-1137 (2004).

15. C. M. Smith *et al.*, Tuberculosis Susceptibility and Vaccine Protection Are Independently Controlled by Host Genotype. *MBio* **7**, (2016).

945 16. C. M. Smith *et al.*, Host-pathogen genetic interactions underlie tuberculosis susceptibility
946 in genetically diverse mice. *Elife* **11**, (2022).

947 17. J. Lee *et al.*, CD11cHi monocyte-derived macrophages are a major cellular compartment
948 infected by Mycobacterium tuberculosis. *PLoS Pathog* **16**, e1008621 (2020).

949 18. B. Nandi, S. M. Behar, Regulation of neutrophils by interferon-gamma limits lung
950 inflammation during tuberculosis infection. *J Exp Med* **208**, 2251-2262 (2011).

951 19. E. Sahiratmadja *et al.*, Plasma granulysin levels and cellular interferon-gamma production
952 correlate with curative host responses in tuberculosis, while plasma interferon-gamma
953 levels correlate with tuberculosis disease activity in adults. *Tuberculosis (Edinb)* **87**, 312-
954 321 (2007).

955 20. S. Sakai *et al.*, Cutting edge: control of Mycobacterium tuberculosis infection by a subset
956 of lung parenchyma-homing CD4 T cells. *J Immunol* **192**, 2965-2969 (2014).

957 21. S. Becattini *et al.*, Functional heterogeneity of human memory CD4(+) T cell clones primed
958 by pathogens or vaccines. *Science* **347**, 400-406 (2015).

959 22. C. M. Smith *et al.*, Functionally Overlapping Variants Control Tuberculosis Susceptibility
960 in Collaborative Cross Mice. *mBio* **10**, (2019).

961 23. C. S. Lindestam Arlehamn *et al.*, A Quantitative Analysis of Complexity of Human
962 Pathogen-Specific CD4 T Cell Responses in Healthy M. tuberculosis Infected South
963 Africans. *PLoS Pathog* **12**, e1005760 (2016).

964 24. Y. R. Patankar *et al.*, Limited recognition of Mycobacterium tuberculosis-infected
965 macrophages by polyclonal CD4 and CD8 T cells from the lungs of infected mice. *Mucosal*
966 *Immunol* **13**, 140-148 (2020).

967 25. C. S. Lindestam Arlehamn *et al.*, Memory T cells in latent Mycobacterium tuberculosis
968 infection are directed against three antigenic islands and largely contained in a
969 CXCR3+CCR6+ Th1 subset. *PLoS Pathog* **9**, e1003130 (2013).

970 26. E. V. Acosta-Rodriguez *et al.*, Surface phenotype and antigenic specificity of human
971 interleukin 17-producing T helper memory cells. *Nat Immunol* **8**, 639-646 (2007).

972 27. H. P. Gideon *et al.*, Multimodal profiling of lung granulomas in macaques reveals cellular
973 correlates of tuberculosis control. *Immunity* **55**, 827-846 e810 (2022).

974 28. M. Roederer, J. L. Nozzi, M. C. Nason, SPICE: exploration and analysis of post-cytometric
975 complex multivariate datasets. *Cytometry A* **79**, 167-174 (2011).

976

Figure 1

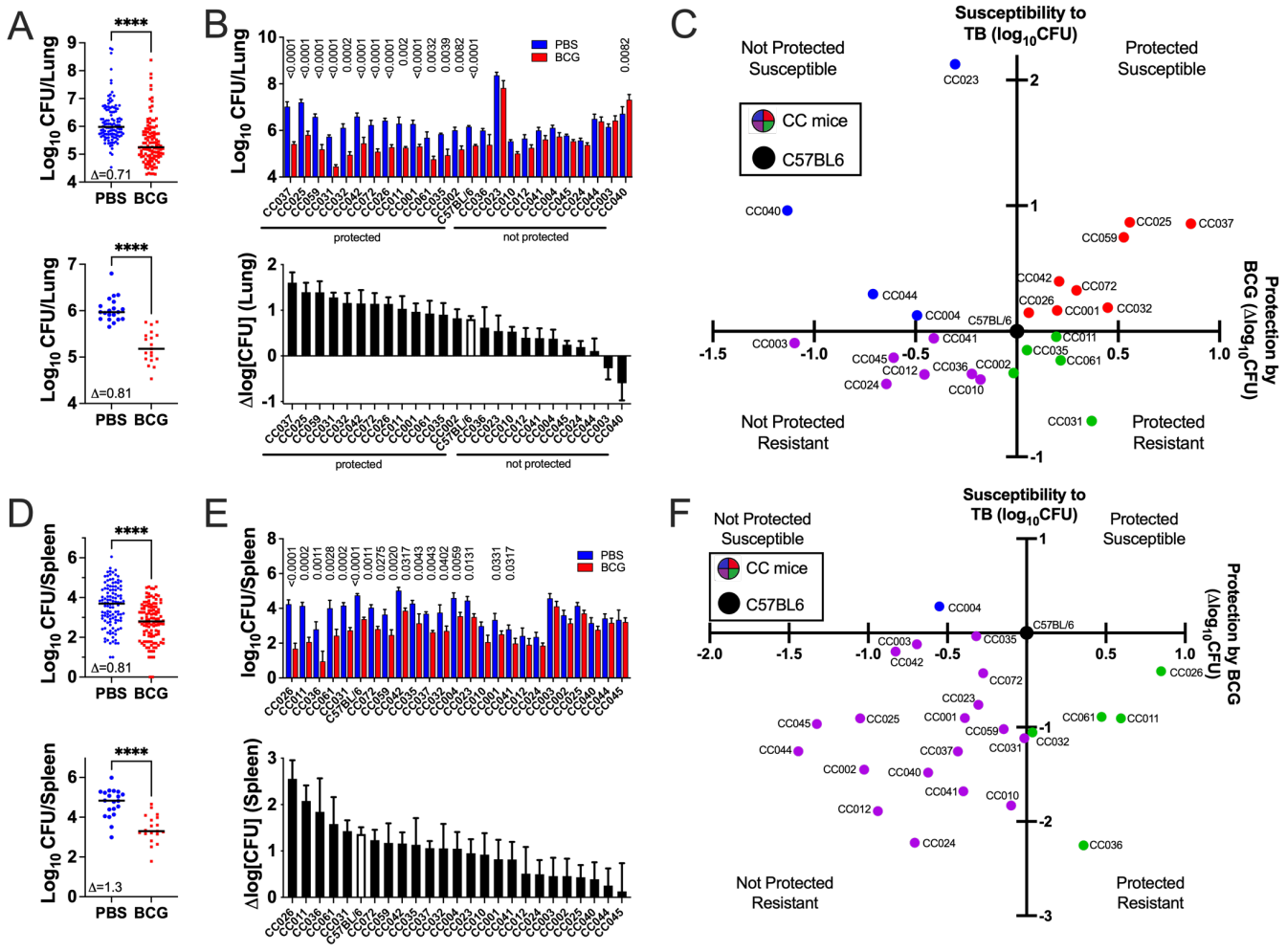

Figure 1. BCG-induced protection in 24 CC strains and C57BL/6 mice.
(A) Lung and (D) spleen CFU of all 24 CC mice (n=120/group; top) or all C57BL/6 mice (n=20/group; bottom) at 4 weeks post infection. Mice were vaccinated with BCG or administered PBS and rested for 12 weeks prior to infection. Each point represents an individual mouse. Line, median. Student's t-test; ****, $p < 0.0001$.(B) Protection in the lung and (E) was assessed by comparing the CFU in BCG vs. sham vaccinated mice for each individual mouse strain (n=5/group) using the Benjamini-Hochberg procedure to determine the false discovery rate (FDR) (top). The ΔCFU of BCG vs. sham vaccinated was calculated (bottom). The open bar is the reference strain C57BL/6 strain (n=20/group). *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. (C) Correlation in the lung and (F) spleen between susceptibility and protection. The relative susceptibility (y-axis, defined by CFU in unvaccinated group, relative to the C57BL/6 reference strain) and the protection conferred by BCG (x-axis, defined by ΔCFU of vaccinated mice compared to unvaccinated controls, relative to the C57BL/6 reference strain). The graph is centered on C57BL/6 mice, leading to the different CC strains being placed in four quadrants based on their relative susceptibility and protection.

Table 1

CC strain[a]	Batch [b]	MHC Haplotype[c]	ΔLung CFU_1 [d] reduction (logCFU)	ΔLung CFU_2 [d] reduction (logCFU)	ΔSpleen CFU_1 [e] reduction (logCFU)	ΔSpleen CFU_2 [e] reduction (logCFU)
CC001/Unc	4	b	0.94	1.45	2.79	1.45
CC002/Unc	3	b	0.83	n/a	0.46	n/a
CC003/Unc	3	g	0.07	-0.61	0.77	0.15
CC004/TauUnc	4	b	0.05	0.08	-0.16	-0.12
CC010/GeniUnc	1	b	0.53	n/a	0.92	n/a
CC011/Unc	3	g7	1.04	n/a	2.08	n/a
CC012/GeniUnc	4	wsb	0.79	n/a	0.89	n/a
CC023/GeniUnc	1	b	1.17	-0.09	1.47	0.32
CC024/GeniUnc	2	cast	0.15	0.24	0.38	0.62
CC025/GeniUnc	3	pwk	1.40	n/a	0.43	n/a
CC026/GeniUnc	2	cast	0.93	1.05	3.75	1.87
CC031/GeniUnc	3	b	1.22	1.35	1.26	1.60
CC032/GeniUnc	1	b	1.16	n/a	1.06	n/a
CC035/Unc	4	wsb	0.91	n/a	1.13	n/a
CC036/Unc	3	g7	0.62	n/a	1.84	n/a
CC037/TauUnc	1	b	1.37	1.84	1.07	1.05
CC040/TauUnc	2	b	-0.64	-0.28	-0.35	1.14
CC041/TauUnc	2	b	0.14	0.94	0.71	0.94
CC042/GeniUnc	1	b	1.65	0.96	1.00	1.32
CC044/Unc	4	b/g7	0.08	0.13	0.12	0.39
CC045/GeniUnc	4	z	0.24	n/a	0.13	n/a
CC059/TauUnc	2	b/a	1.39	n/a	1.17	n/a
CC061/GeniUnc	1	b	0.93	n/a	1.58	n/a
CC072/TauUnc	2	b	0.97	1.37	1.06	1.41

Table S1. A description of select CC strains that were repeated for in-depth immune analysis.
a A group of C57BL/6 mice were included in all batches as a reference strain. Data was normalized to internal C57BL/6 control as indicated in the methods.
b The first cohort of CC mice were analyzed in 4 separate batches as indicated.
c MHC haplotype of each CC mouse as inherited from CC founders.
d, e Thirteen of the 24 CC strains were analyzed a second time to validate the original results and perform more in-depth immune analysis. ΔLungCFU[d] and ΔSpleenCFU[e] for the first experiment and the repeat experiment are listed. Mice that were exacerbated by BCG vaccination are highlighted in red.

Table 2

Strain[a]	BCG effect[b] (lung CFU)	Were differences observed?[c]	How did the vaccinated mice differ?[d]
CC003/TauUNC	Not protected	N	No differences. Both groups had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions
CC041/TauUNC	Not protected	N	No differences. Both groups had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions
CC024/GeniUNC	Not protected	N	No differences. Both groups had small hard-to-find lymphocytic lesions with perivascular and peribroncial lymphocytic cuffs
CC002/TauUNC	Not protected	Y	Ctrl had mild lesions with perivascular lymphocytic cuffs. BCG mice had larger moderate sized macrophage-rich lesions. Lymphocytes were similarly abundant.
CC004/TauUnc	Not protected	Y	Ctrl mice had mild to moderate lesions with lymphohistiocytic infiltrates. BCG mice have more lymphocytes with regions of dense perivascular/peribronchiolar lymphocytic cuffs.
CC036/GeniUnc	Not protected	Y	Ctrl mice have mild disease with lymphohistiocytic infiltrates. BCG mice also have lymphohistiocytic infiltrates with few very dense lymphocytic areas
CC040	Not protected	Y	Ctrl mice have moderate necrotizing lesions with few lymphocytes. BCG mice have mild disease - small & hard to find lesions that are lymphocyte rich without necrosis.
CC044/GeniUnc	Not protected	Y	Unusual. Ctrl mice have lymphohistiocytic infiltrates with dense lymphocytic cuffs. BCG mice have lymphohistiocytic infiltrates with dense lymphocytic cuffs and lymphoid aggregates/follicules in parenchyma.
CC061/GeniUNC	Protected	N	No differences. Unusual. Mild lesions. Both groups have dense lymphoid aggregates/follicles in parenchyma. Also, perivascular lymphocytic cuffs.
CC072/TauUNC	Protected	N	No differences. Mild disease. perivascular lymphocytic cuffs with a few macrophage-rich lesions. One mouse has a foreign-body granulomatous bronchitis/pneumonia with intralesional food/bedding/plant.
C57BL/6	Protected	Y	Ctrl mice had mild to moderate lymphohistiocytic lesions. BCG mice had with more, dense lymphocytic infiltrates
CC001/TauUnc	Protected	Y	Ctrl mice had moderate lesions with lymphohistiocytic infiltrates and equivocal necrosis. BCG mice had mild lesions with perivascular lymphocytic cuffs and a few macrophage-rich lesions without necrosis
CC023	Protected	Y	Ctrl mice had moderate to severe lesions with necrosis and minimal lymphocytes. BCG mice have moderate lesions with more lymphocytes and less necrosis.
CC025/TauUNC	Protected	Y	Ctrl mice have moderate disease with mild lymphocytic infiltrates and necrosis. BCG mice have abundant dense perivascular/peribronchiolar lymphocytic ifiltrates and lymphohistiocytic granulomas and less necrosis.
CC026	Protected	Y	Ctrl mice had moderate lesions with lymphohistiocytic infiltrates. BCG mice had mild lesions with lymphohistiocytic infiltrates and a few very dense lymphocytic areas
CC031/GeniUnc	Protected	Y	Ctrl mice have moderate disease with lymphohistiocytic infiltrates. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas
CC032/GeniUNC	Protected	Y	Ctrl mice have mild disease with lymphohistiocytic infiltrates. BCG mice have more lymphocyte rich lesions that are small & hard to find
CC037/TauUNC	Protected	Y	Ctrl mice have moderate disease with lymphohistiocytic infiltrates. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas
CC042	Protected	Y	Ctrl mice have mild to moderate disease with few lymphocytes and necrosis. BCG mice have mild disease with more lymphocytes, less necrosis
CC059/TauUNC	Protected	Y	Ctrl mice had mild disease with lymphohistiocytic infiltrates with small regions of necrosis. BCG mice have abundant dense perivascular/peribronchiolar lymphocytes and lymphohistiocytic granulomas without necrosis

Table S2. Pathological descriptions of changes lung histology after BCG vaccination in CC mice
a Nineteen of the 24 CC strains were available for analysis.
b The effect of BCG on the lung CFU is from Figure 1.
c Assessment by a pathologist of whether differences existed between sham and BCG vaccinated mice after Mtb infection (n=5/strain/condition).
d Summary of observed differences between the groups.

.

Figure 2

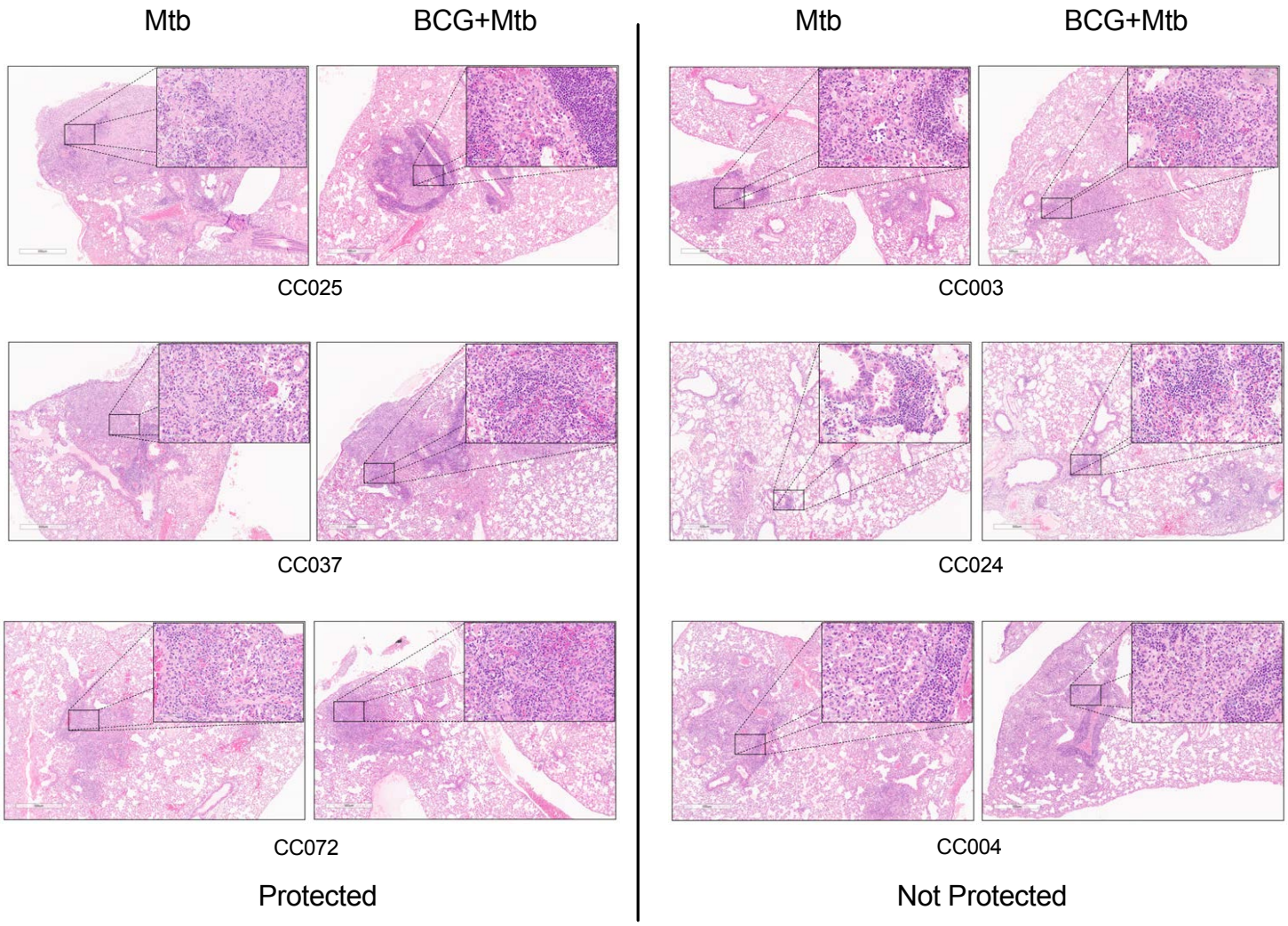

Figure 2. Lung histology of CC mice after BCG vaccination and Mtb challenge
Representative lung histology of three CC strains (CC025, CC037, CC072) that were protected by BCG, after sham or BCG vaccination and Mtb challenge (Mtb and BCG+Mtb, respectively) (left). Representative lung histology of three CC strains (CC003, CC004, CC024) that were unprotected by BCG (right). Original magnification, 40X (inset, 400X). Scale bar, 500 μm (inset, 50 μm).

Figure 3

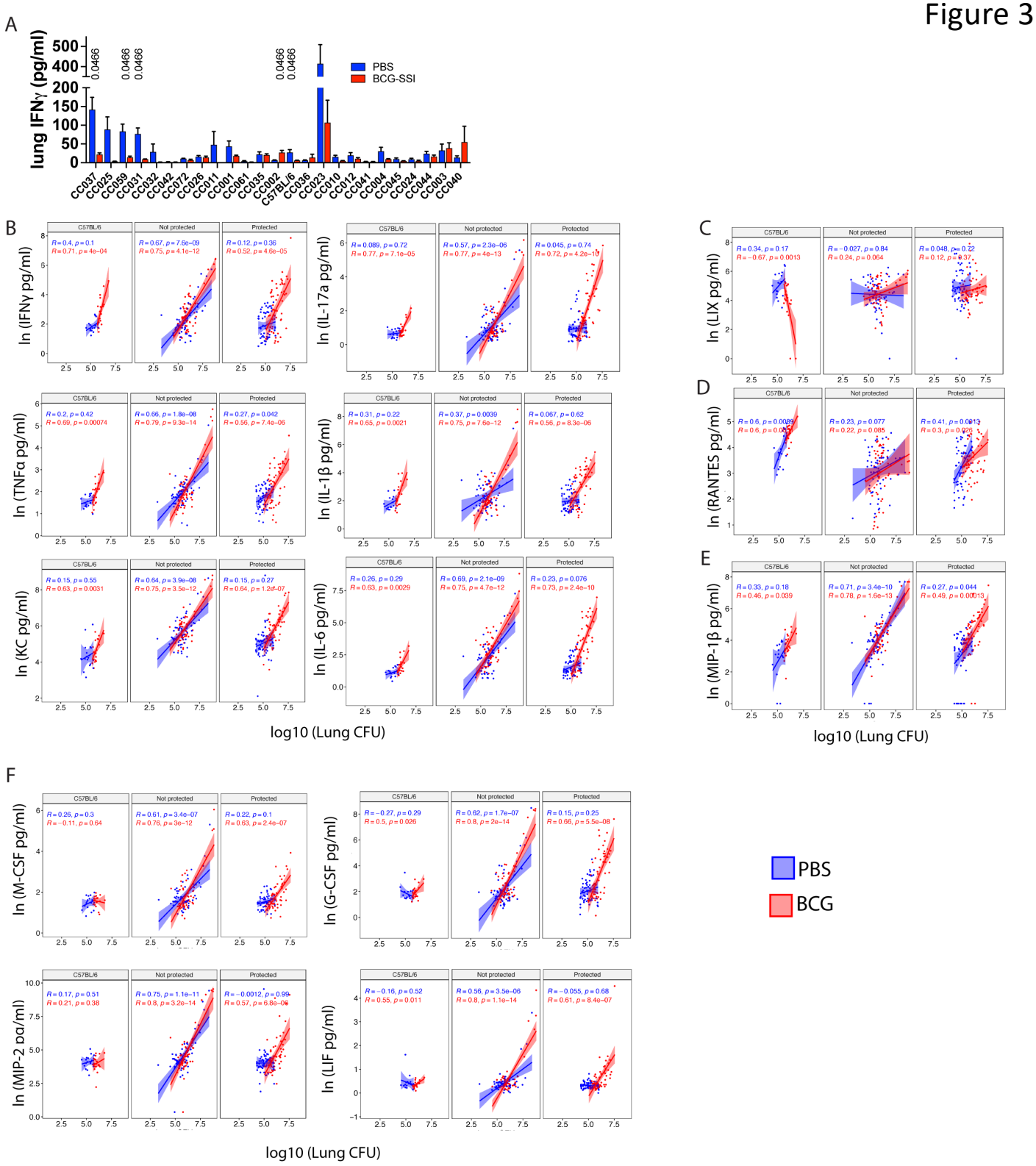

Figure 3. BCG-protected CC mice have an altered lung microenvironment of CC mice after Mtb infection.

(A) IFNγ levels were measured in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks post infection. A two-way ANOVA was performed using Sidak's correction for multiple comparisons. ***, $p < 0.001$; ****, $p < 0.0001$". (B) – (F) Pearson correlation between lung CFU and (B) IFNγ, IL-17, TNF, IL-1β, KC and IL-6 (C) LIX (D) RANTES (E) MIP-1β and (F) M-CSF, MIP-2, G-CSF, LIF, comparing PBS and BCG vaccinated mice at 4 weeks post infection in C57BL/6 mice (B6), non-protected (NP) or protected (P) CC mice.

Figure S1

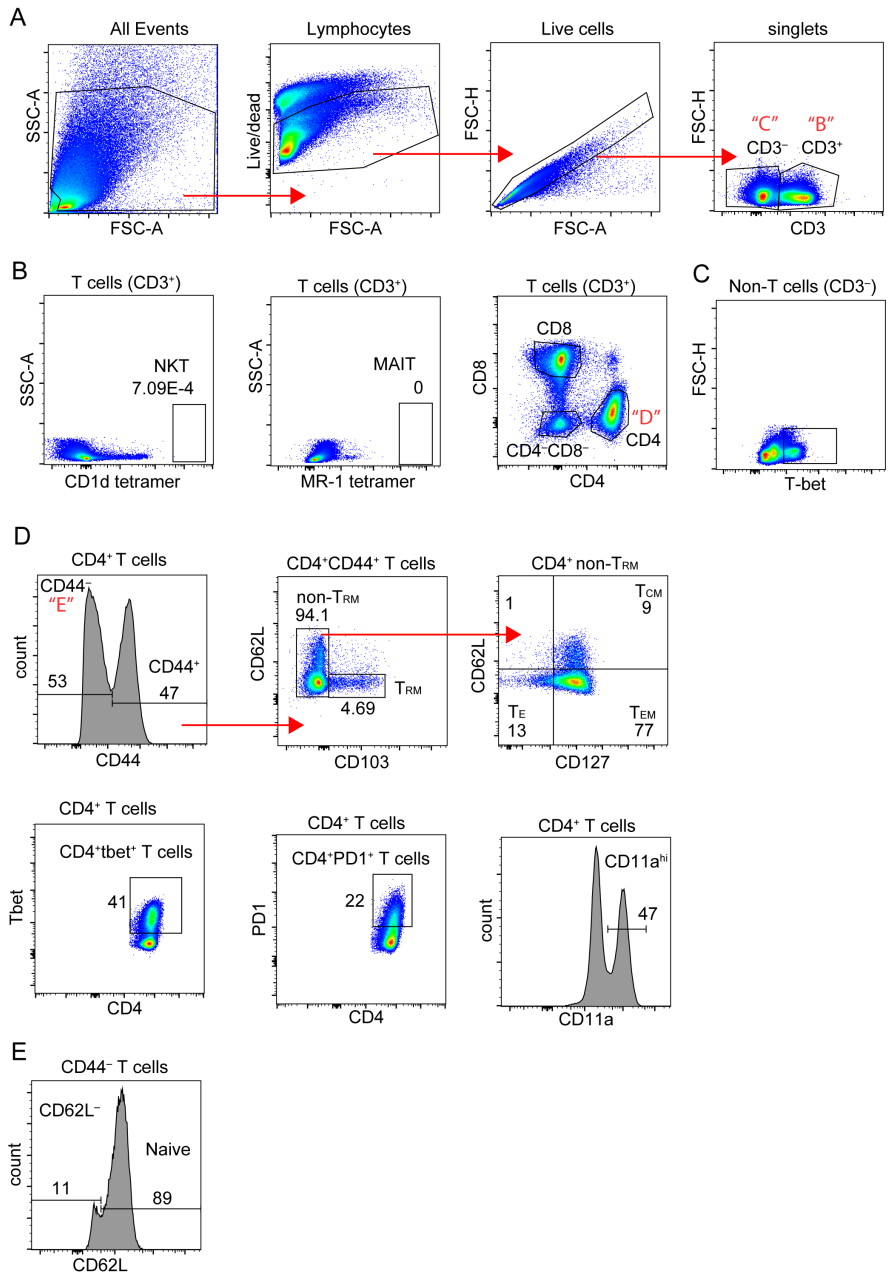

Figure S1. Gating strategy for the analysis of the cell surface T cell phenotype.
(A) Lymphocytes were identified using FSC and SSC properties, after which viability dyes were used to identify live cells. Singlet gates were used to exclude doublets, after which T cells and non-T cells were identified using the lineage marker CD3.
(B) NKT and MAIT cells were identified from the $CD3^+$ gate using their antigen-loaded CD1d and MR1 tetramers, respectively. $CD4^+$, $CD8^+$ and $CD4^-CD8^-$ T cells were also identified from this population for further analysis.
(C) Expression of various markers were also examined within the $CD3^-$ population. As an example, flow plot represents expression of T-bet within the $CD3^-$ population.
(D) Identification of memory T cells. CD44 was used to identify antigen-experienced T cells. Resident memory T cells (T_{RM}) were identified based on CD103 expression. From the $CD103^-$ fraction, CD127 and CD62L expression identified central memory (T_{CM}), effector (T_E) and effector memory (T_{EM}). Subset identification of various T cell populations (ex: $Tbet^+CD4^+$ T cells) as well as activation markers (ex: PD1, CD11a) were also examined in $CD4^+$, $CD8^+$ and $CD4^-CD8^-$ T cell subsets.
(E) Naïve T cells identified within the $CD44^-$ compartment using CD62L expression.

Figure 4

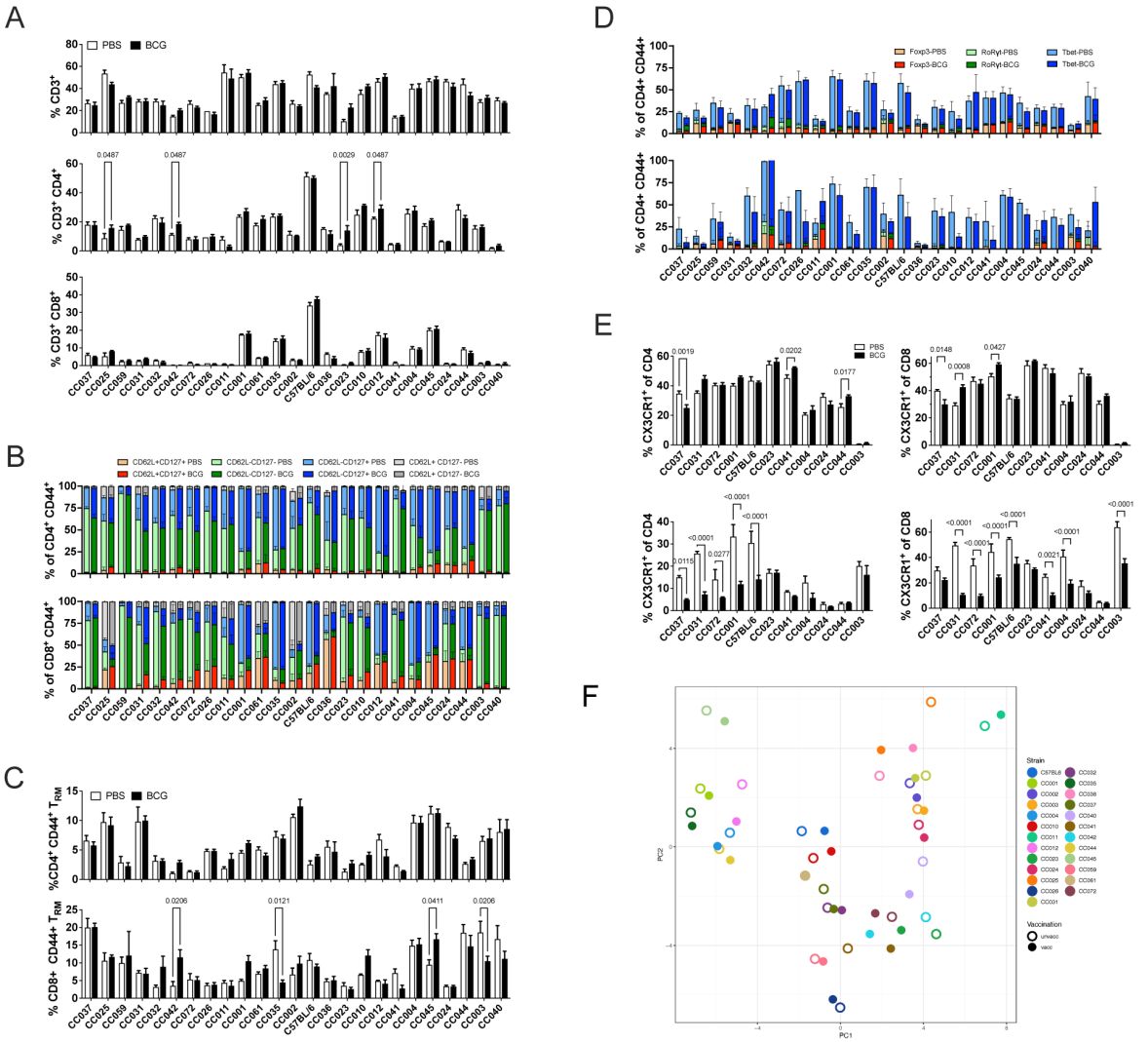

Figure 4. Frequency of various T cell populations in protected and unprotected CC mice following Mtb infection.

(A) Proportion of $CD3^+$ cell, $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells in the lungs of PBS and BCG vaccinated CC mice at 4 weeks post infection.

(B) Proportion of memory CD4 and CD8 T cells in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ and various combinations of CD62L and CD127 expression; $CD62L^+$ $CD127^+$ (central memory), $CD62L^-$ $CD127^-$ (effector), $CD62L^-$ $CD127^+$ (effector memory).

(C) Proportion of resident memory CD4 and CD8 T cells in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ $CD103^+$ $CD69^+$.

(D) Proportion of Th1, Treg and Th17 cells in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ and $Tbet^+$ (Th1), $Foxp3^+$ (Treg) or RoRγt (Th17) expression.

(E) Proportion of CD4 and CD8 T cells expressing either CXCR3 or CX_3CR1 in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks post infection.

(F) PCA of the T cell phenotyping data. Each point represents the average of usually 5 mice within a given mouse strain that are either BCG vaccinated or unvaccinated. Open symbols, unvaccinated; closed symbols, BCG vaccinated. Each color represents a different mouse strain.

(A–E) The data represent the mean ± SEM from one experiment. Two-way analysis of variance with the Benjamini and Hochberg multiple comparison method. The FDR was set to 0.05 and the numbers in the figures are the q value.

Figure S2

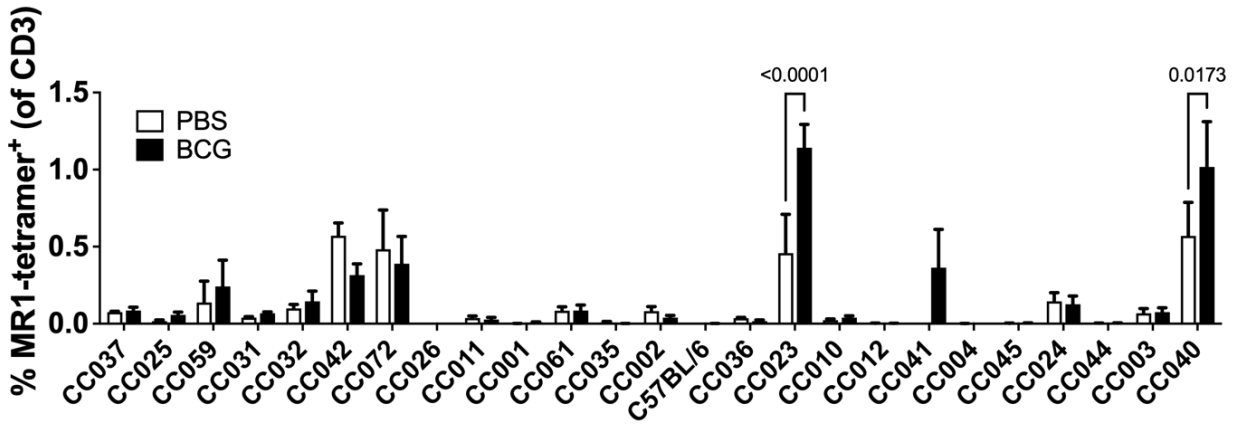

Figure S2. Frequencies of non-conventional T cells in CC mice. Frequencies of MR-1-restricted MAIT cells were measured in the lungs of PBS or BCG vaccinated CC mice at 4 weeks post Mtb infection. The data in the graphs represent the mean ± SD of one experiment. Two-way analysis of variance with the Benjamini and Hochberg multiple comparsion method. The FDR was set to 0.05 and the numbers in the figures are the q value.

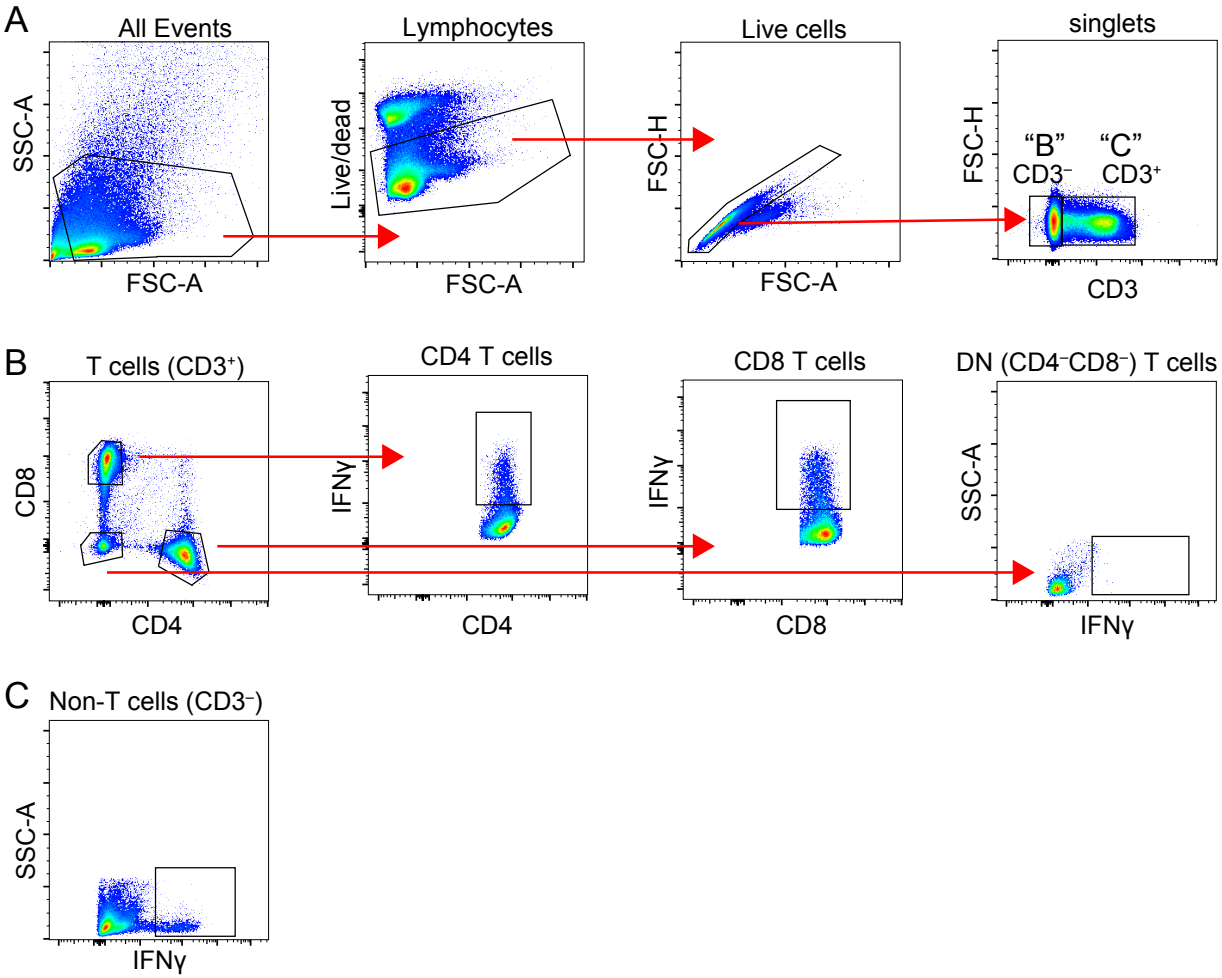

Figure S3. Gating strategy for the analysis of T cell function (intracellular cytokine staining).

(A) Lymphocytes were identified as using FSC and SSC properties, after which viability dyes were used to identify live cells. Singlet gates were used to exclude doublets, after which T cells were identified using the lineage marker CD3.

(B) CD4, CD8 and $CD4^{-}CD8^{-}$ T cells were also identified from the $CD3^{+}$ population for further analysis. Cytokine expression (ex: IFNγ) were examined in all 3 T cell sub-populations.

(C) Cytokine expression was also examined within the $CD3^{-}$ population.

Figure 5

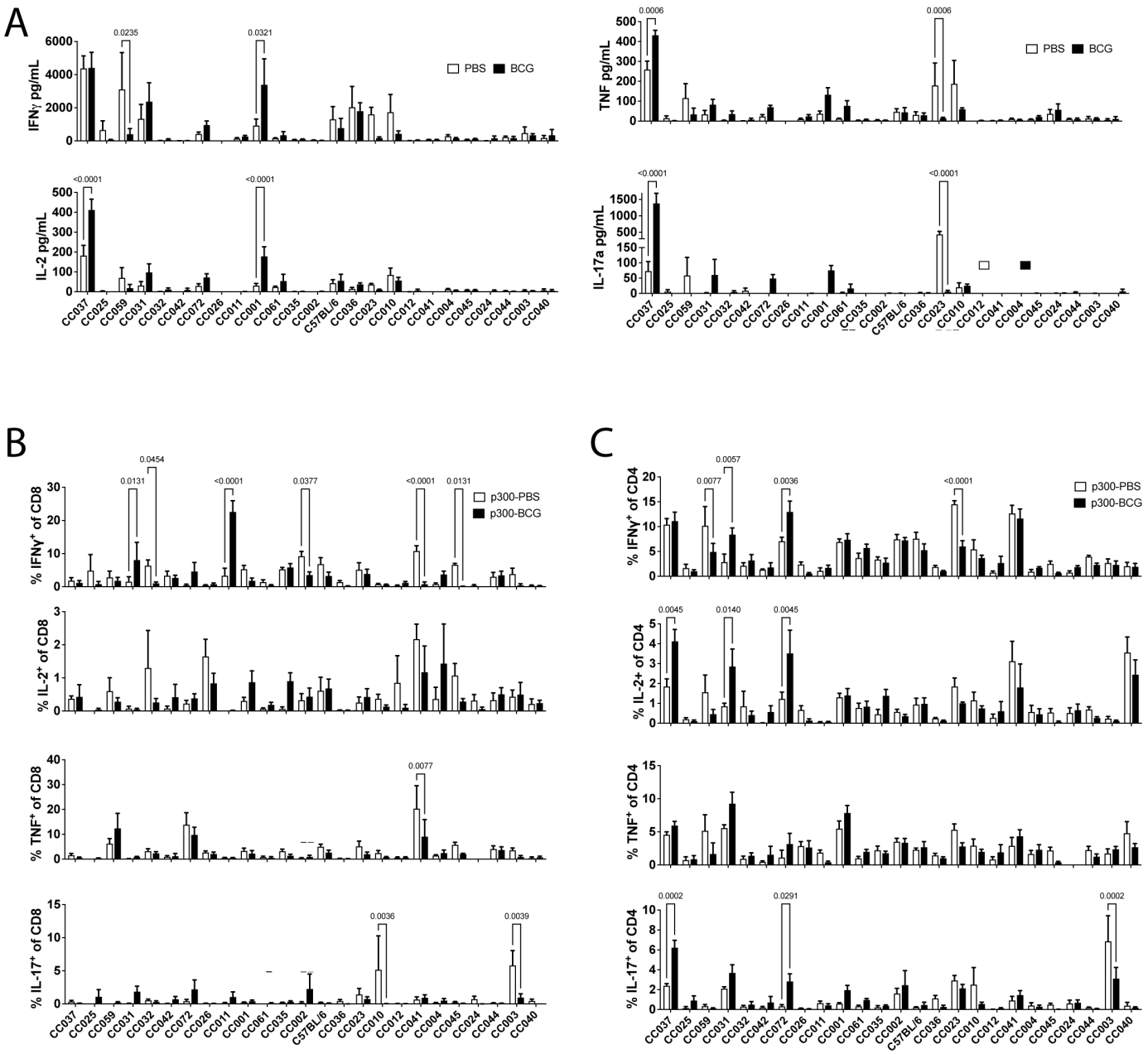

Figure 5. Cytokine response in protected and unprotected CC strains following Mtb infection. (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated for 24 hours with the P300 peptide pool. IFNγ, IL-2, TNF and IL-17a were measured in the supernatants. (B) and (C) Cells from the lungs of CC mice at 4 weeks post infection were stimulated for 5 hours with the P300 peptide pool. Frequencies of CD4 or CD8 T cells that expressed IFNγ, IL-2, TNF or IL-17 was determined using intracellular cytokine staining. (A–C) The data represent the mean ± SEM from one experiment. Two-way analysis of variance with Sidak's multiple comparison test was used to determine significance. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

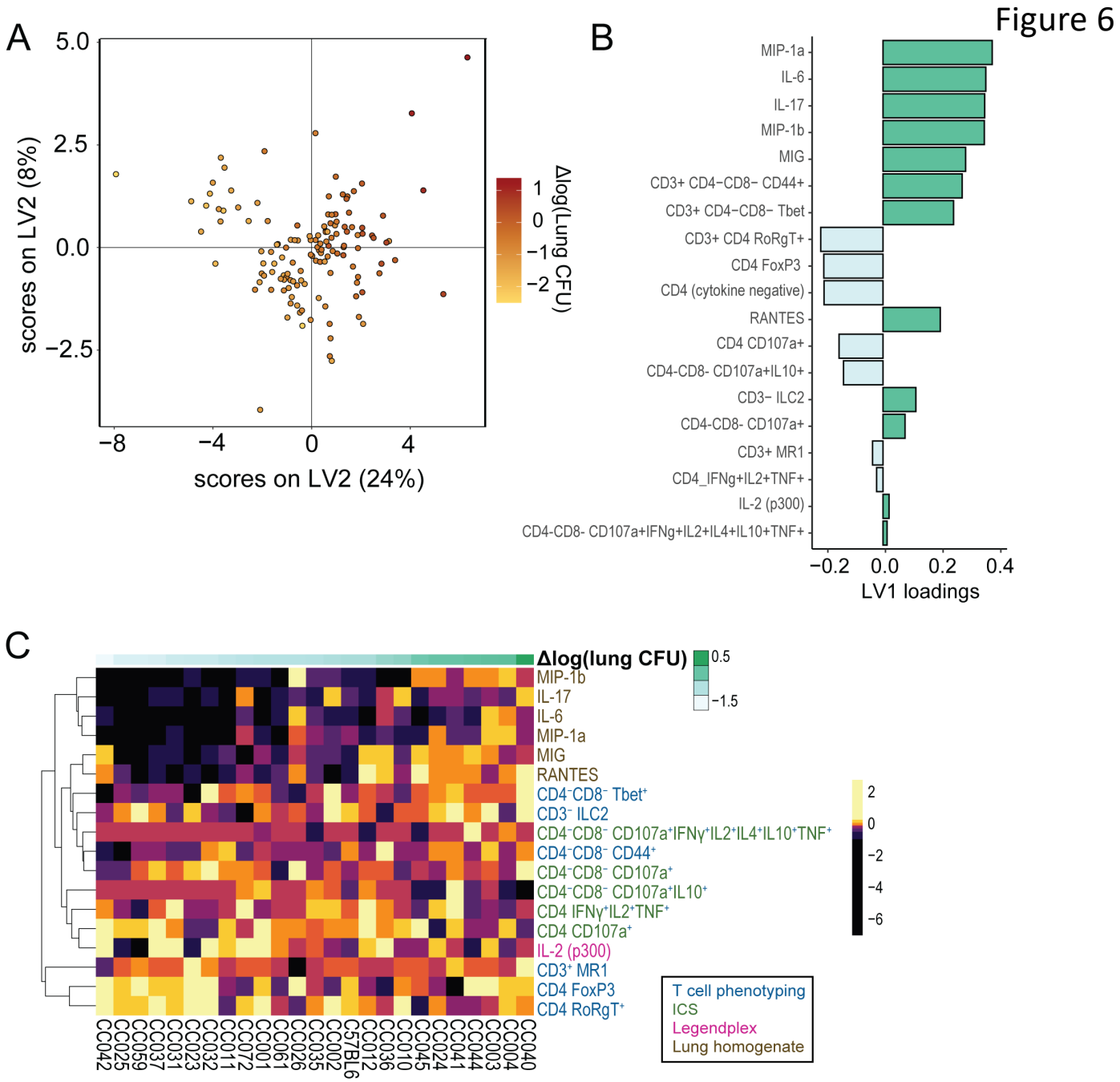

Figure 6. Multivariate correlates of protection across all 24 CC strains
(A) A partial least squares regression (PLSR) model was performed to associate BCG-induced changes in lung immune features with BCG-induced changes in lung CFU. Each data point represents a vaccinated mouse that has the mean of the unvaccinated mice from the same strain subtracted from it. The scores of the top two latent variables of the model are shown, with the color of each data point representing the changed in lung CFU induced by BCG vaccination. The percents of the variance in the lung features captured by each latent variable are shown as percentages in the axes. (B) The bar graph depicts the loadings of the first latent variable of the PLSR model in (A). Positive loadings indicate immune features that increase from BCG vaccination in samples that are on the positive x-axis in the scores plot, which generally indicates they have worse protection. Negative loadings conversely indicate immune features that generally increase from BCG vaccination in samples that are protected by vaccination. (C) The heatmap illustrates all features selected to be included in the model, indicating they were important for associating strains with the lung protection continuum. Colors are representations of z-scored features, with a different color representing each 10% quantile of the data. Columns are averages of the mice in the CC strains (n=5 for most cases). Feature names are colored based on the dataset from which they come. In (B) and (C), features from the ICS dataset indicate production of the cytokines listed and no production of any measured cytokines not listed.

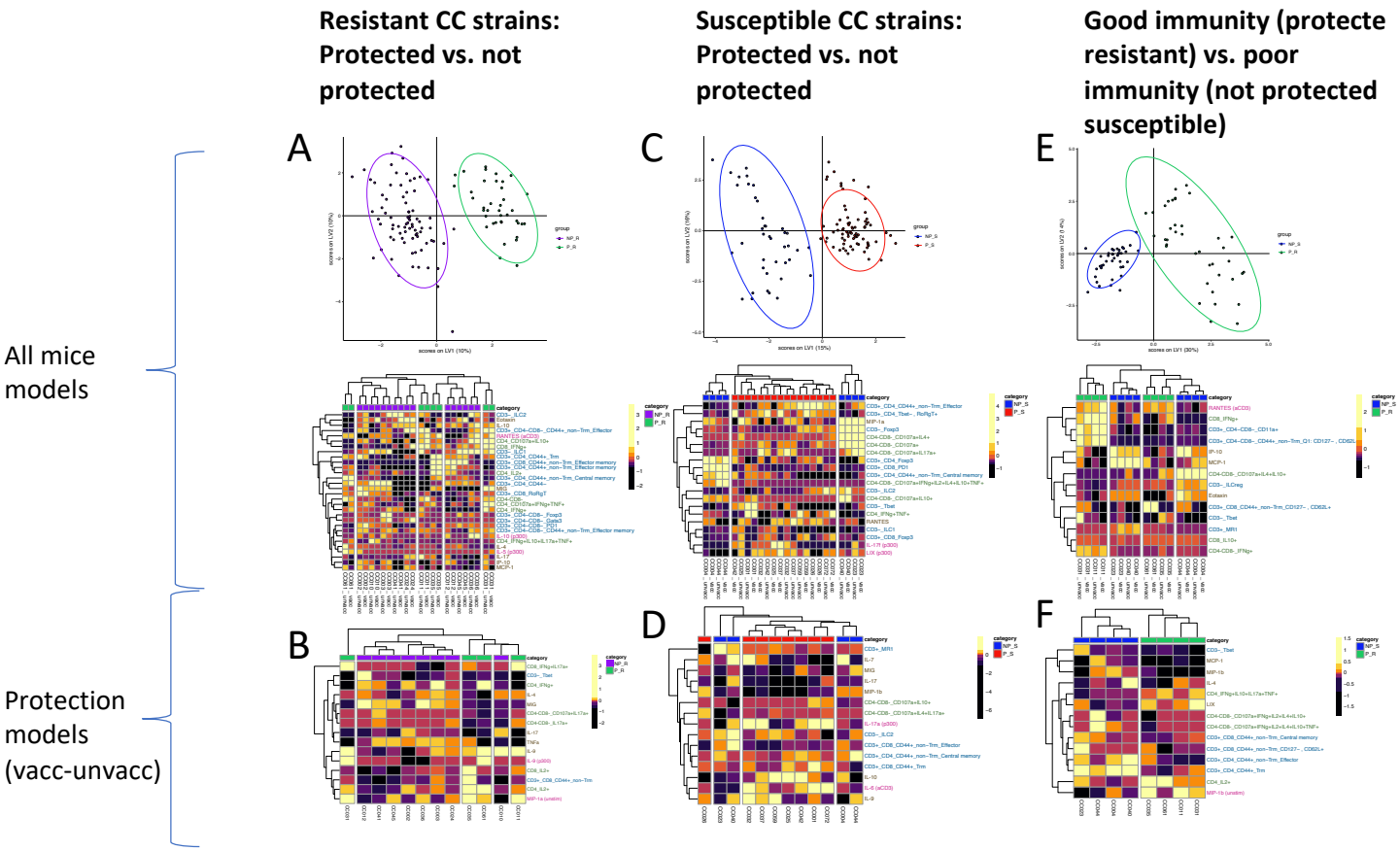

Figure 7. Identifying signatures of protection in subclasses of CC strains
(A), (C), (E) Scores plot (top) and heatmaps (bottom) of selected features for PLSDA models. Each dot represents a single mouse. The two colors in the scores plots signify the two classes being compared. An ellipse is drawn around the 95% confidence interval for each class. Percent of variance of the immune feature data is shown on the axes. Heatmaps show features that were most important for distinguishing the two classes. Colors are representations of z-scored features. Columns are averages of the mice in the CC strains, with vaccinated and unvaccinated mice being separated (n=5 for most cases). Feature names are colored based on the dataset from which they come (blue=T cell phenotyping, brown=lung homogenate cytokines, green=ICS, magenta=LEGENDplex). NP = non-protected, P = protected, R = resistant, S = susceptible. (B), (D), (F) PLSDA models performed as in (A), (C), and (E), except that features are the changes induced by BCG vaccination as in Figure 6.

Figure S4

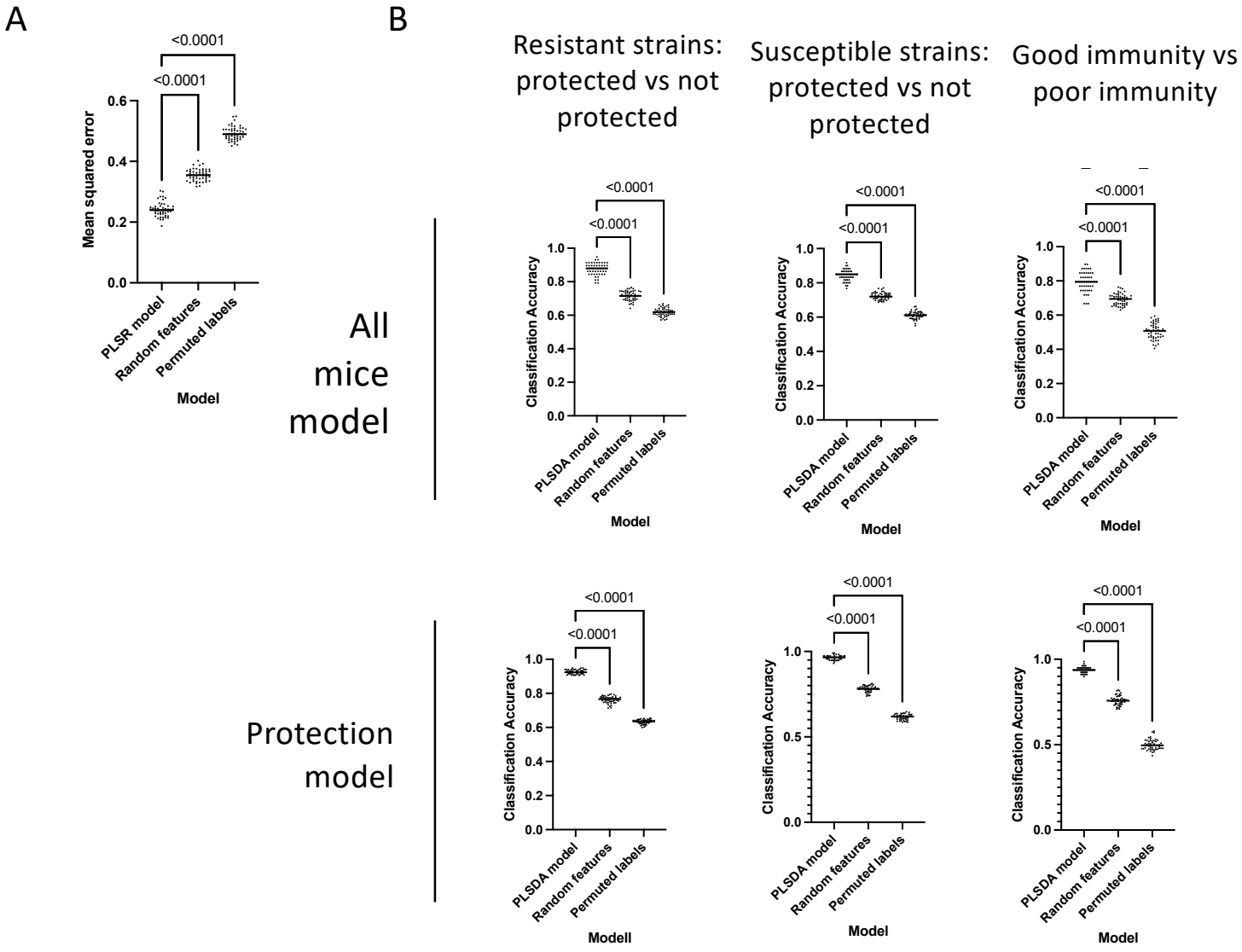

Figure S4. Multivariate model accuracies

(A) The accuracy of our protection model (Figure 6) was compared to that of two different types of null model, where our measure of accuracy was mean squared error of prediction of Δlog (lung CFU). Our PLSR model performed significantly better than both a null model consisting of random features instead of our elastic-net chosen features and a null model where the outcomes of Δlog (lung CFU) were scrambled. Our PLSR model and each of the null models were run 50 times, and each dot represents one of these trials. Labeled p-values are from one-way ANOVAs.

(B) Our models highlighting the immune features that distinguish pairs of protection/susceptibility categories (Figure 7) were compared to two null models as in (A). Each of our models, both the ones involving all mice and the protection models were significantly better than both null models. Here, the accuracy metric is classification accuracy. Labeled p-values are from one-way ANOVAs.

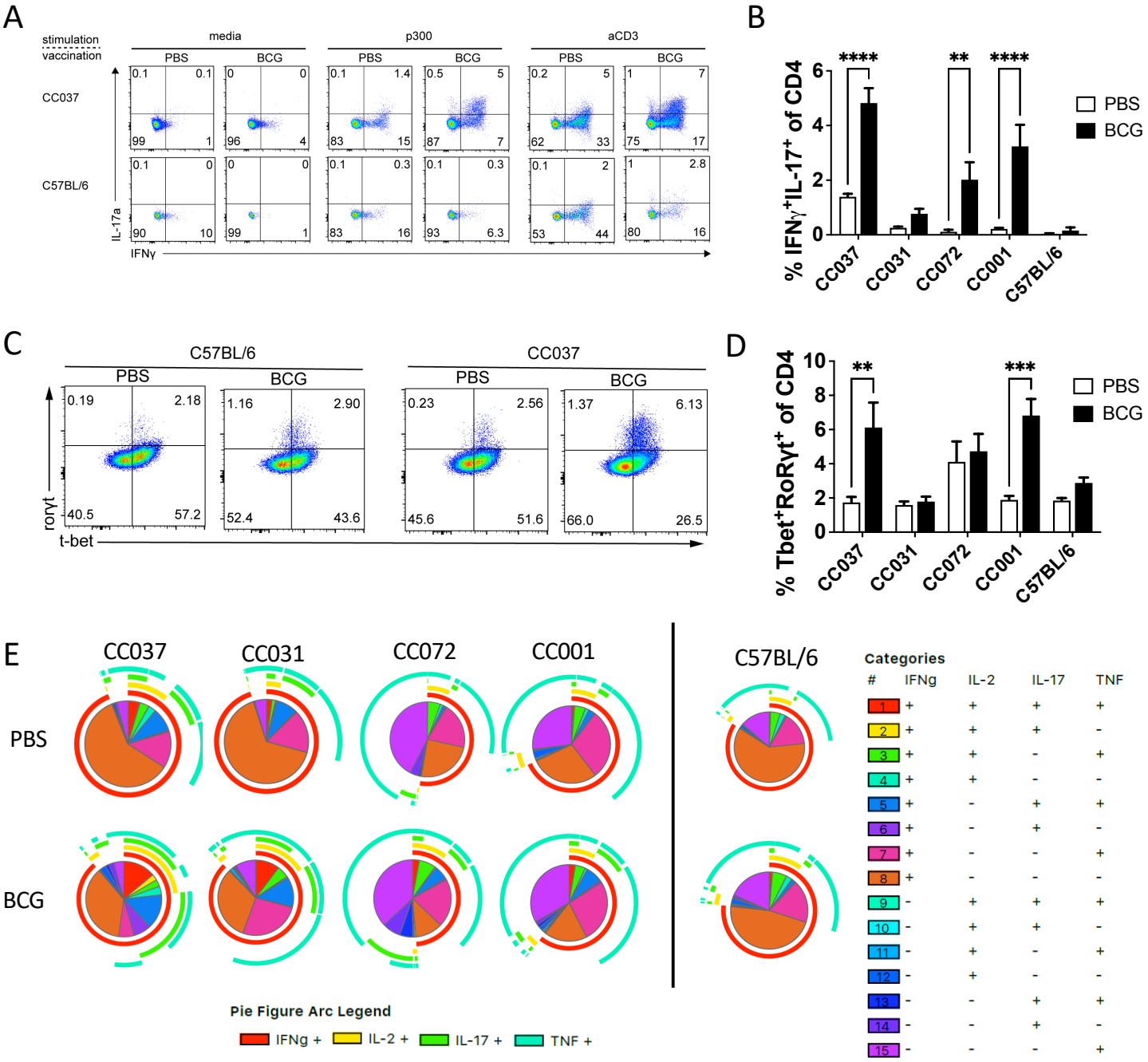
A
stimulation
vaccination
media
p300
aCD3
PBS
BCG
CC037
C57BL/6
IL-17a
IFNγ
B
% IFNγ+IL-17+ of CD4
PBS
BCG
CC037
CC031
CC072
CC001
C57BL/6
C
C57BL/6
CC037
roryt
t-bet
D
% Tbet+RoRyt+ of CD4
E
CC037
CC031
CC072
CC001
C57BL/6
PBS
BCG
Categories
Pie Figure Arc Legend
IFNg +
IL-2 +
IL-17 +
TNF +
F
C57BL/6
CC0037
P (uncorrected)
Spearman correlation with Lung CFU
G
CC037
IFNγ+ IL-17+
-0.8947
<0.0001
log10[CFU]

Figure 8

Figure 8. Th1/17 cells correlate with BCG-mediated protection in a subset of protected CC mice.

(A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated for 5 hours with either aCD3, P300 peptide pool or media control. Representative flow plots of IFNγ and IL-17 producing CD4 T cells from CC037 and C57BL/6 are shown. Data is representative of two independent experiments.

(B) Frequencies of CD4 T cells that expressed both IFNγ and IL-17 was determined using intracellular cytokine staining. Data is representative of two independent experiments. One-way analysis of variance with Sidak's multiple comparison test was used to determine significance. **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

(C) Representative flow plots of CD4 T cells expressing either T-bet or RORγt in CC037 or C57BL/6 mice. Data is representative of two independent experiments.

(D) Frequencies of CD4 T cells that expressed both T-bet or RORγt in CC037 or C57BL/6 mice. Data is representative of two independent experiments. One-way analysis of variance with Sidak's multiple comparison test was used to determine significance. **, $p < 0.01$; ***, $p < 0.001$.

(E) Proportion of CD4 T cells that express various combinations of IFNγ, TNF, IL-2 or IL-17 in either PBS or BCG vaccinated mice at 4 weeks post Mtb infection. Plots are generated using the SPICE platform (*28*).

(F) Volcano plots. Correlation between 60 immune features and the lung CFU determined for CC037 and C57BL/6 mice (from Batch 1). Dotted line, $p = 0.05$. *, $q < 0.05$, corrected with the Benjamini and Hochberg multiple comparison method (FDR = 0.05)

(G) Correlation of $IFN\gamma^{+}IL\text{-}17^{+}$ cells and bacterial burden in the lungs of either PBS or BCG vaccinated CC037 mice at 4 weeks post Mtb infection. Colors represent BCG vaccinated mice (red) or unvaccinated mice (blue). Spearman *r* and p value are shown.

Figure S5

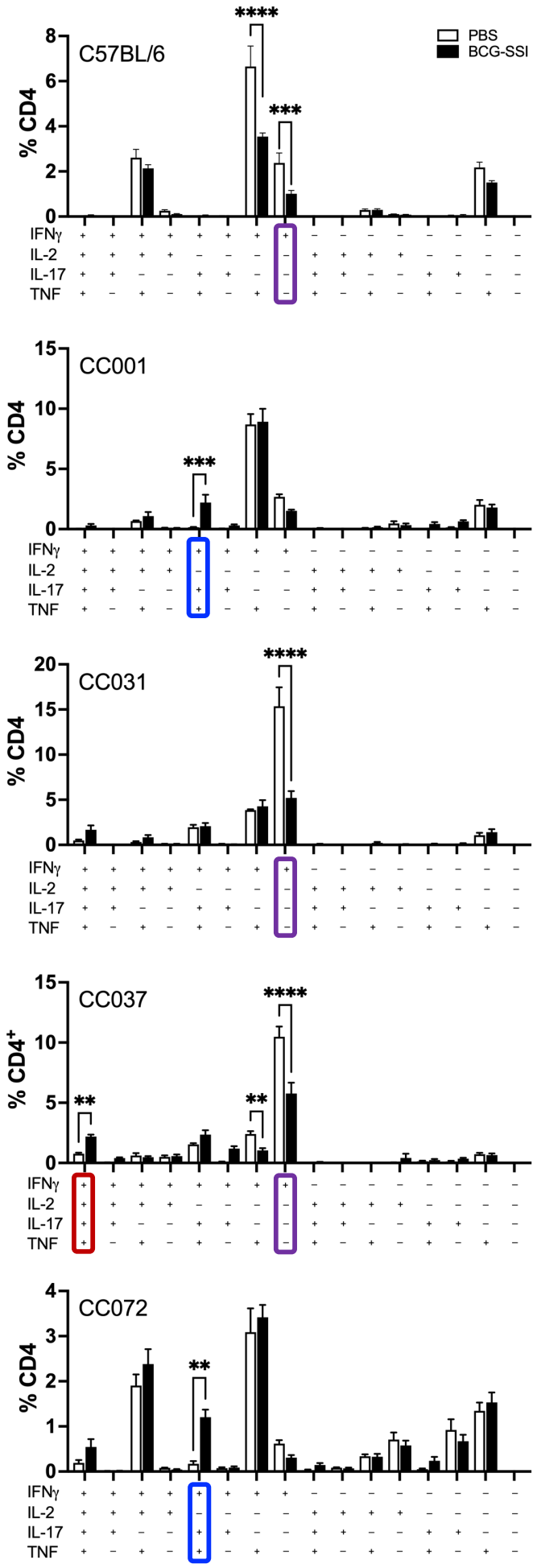

Figure S5. Examining the Polyfunctionality of CD4 T cells in select CC strains. Boolean gating was used to identify various sub-populations amongst the CD4 T cells following P300 stimulation in certain CC mice. An increase in various IFNγ+IL-17+ CD4 T cells (that also produced TNF or IL-2) were detected in CC037, CC072, CC031 and CC001 but not in C57BL/6. In contrast, a noticeable decrease in the frequency of IFNγ "single positive" CD4 T cells was observed in all strains including C57BL/6. The data in the graphs represent the mean ± SD of two experiment. One-way analysis of variance with Sidak's multiple comparison test was used to determine significance. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.