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], Sarah Acolatse[2], 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

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.

39 **Introduction**

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

49 Why is identifying immune correlates of protection following BCG difficult? TB is primarily
50 a human disease that is transmitted person-to-person by aerosolized *Mycobacterium tuberculosis*
51 (Mtb). No animal model recapitulates all the features of human TB, although non-human primates
52 come close (7, 8). Although imperfect, the mouse model has facilitated mechanistic studies that
53 defined many of the critical immunological pathways required for host resistance, permitted in
54 vivo antibiotic testing, and accelerated preclinical vaccine development. As inbred mouse strains
55 are genetically identical, a strength of the mouse model lies in its "infinite" reproducibility. For
56 experimental immunology, C57BL/6 mice are the preferred inbred strain as most genetically
57 modified mice are produced using its genetic background. It is also the most frequently used
58 mouse strain for TB research. However, several features of C57BL/6 mice suggest it is an outlier
59 strain for TB research. Immunologically, it only expresses a single class II MHC molecule (i.e., I-
60 A^b). Pathologically, it develops atypical pathology (i.e., pneumonitis) following Mtb infection.
61 Finally, and perhaps most importantly, it is among the most resistant of all inbred strains to Mtb
62 infection, which makes it difficult to identify conditions that make C57BL/6 mice more resistant to
63 disease. Investigators have addressed these limitations by using F1 hybrids and susceptible
64 strains to increase MHC and genetic diversity and identify strategies to increase host resistance

65 (9-12). However, even among different inbred strains, the genetic diversity is limited (13). This is
66 a problem not only for TB research, but for many areas of biomedical research.

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

77 We previously reported that the CC strains have a large variation in primary susceptibility
78 to TB (15, 16). Here we leverage the CC model to understand the how host genetics affects
79 vaccine-induced protection against Mtb. We find that the genetic background of the host has a
80 major effect on whether BCG induces protection. There was a poor correlation between inherent
81 host susceptibility and BCG-mediated protection, indicating that vaccines can mediate protection
82 even in individuals with a genetic propensity to TB. Interestingly, the T cell response following
83 BCG vaccination and Mtb challenge was diverse, with some immune features universally
84 correlating with outcome of vaccination while others only were important in subsets of CC strains.
85 We propose that the use of CC mice as a preclinical model can facilitate the development of
86 vaccines that protect a diverse range of genotypes against TB.

87 **Results.**

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

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

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

107 We next asked whether the protective effect of BCG vaccination correlated with the
108 inherent ability (i.e., in the unvaccinated state) of each CC strain to control Mtb replication in their
109 lungs. We categorized the 24 CC strains relative to C57BL/6 mice, which we used as a reference
110 strain. The lung CFU of unvaccinated mice defined the inherent susceptibility or resistance to Mtb
111 (susceptibility, y-axis). Next, we considered whether each CC strain was protected or not based
112 on the $\Delta\log_{10}$ lung CFU (relative BCG-induced protection, x-axis). The plot was centered on

113 C57BL/6 mice, which represented a median strain in terms of lung susceptibility and protection
114 and divided the 24 CC strains into four different groups, represented by four quadrants (Fig.1C).
115 Based on lung CFU, there is a poor correlation between intrinsic susceptibility of the mouse strain
116 and protection conferred by BCG at this time point (r=0.104, p=0.62).

117 The same analysis was performed using the spleen bacterial burden from infected mice.
118 As was observed in the lung, the CC population was similarly protected as C57BL/6 (Fig.1D). The
119 range of spleen protection extended from most protected (CC026, $\Delta\log_{10}$ CFU = 2.56) to least
120 protected (CC045, $\Delta\log_{10}$ CFU = 0.13) (Fig.1E). BCG significantly protected against systemic
121 infection in 17 of the 24 CC strains (Fig.1E). There was also no correlation between susceptibility
122 and protection in the spleen (r=0.089, p=0.67). Interestingly, while C57BL/6 mice were a "median"
123 strain in the lung analysis, C57BL/6 mice had among the highest spleen CFU, suggesting they
124 disseminated more, but were among the most protected mouse strain following BCG vaccination
125 (Fig.1F). Among the CC strains, protection in the lung and the spleen were moderately correlated
126 (r=0.513, p=0.0087).

127 In follow-up studies, 13 of the 24 CC strains were retested to confirm the initial phenotype,
128 with emphasis placed on strains that represented varying levels of protection and susceptibility
129 (Table 1). Thus, using 24 CC strains, with diverse genotypes, we find that less than 50% of the
130 strains were protected by BCG, and that protection did not correlate with the intrinsic susceptibility
131 of the mouse to Mtb infection. Finally, the lack of correlation between the reduction in lung versus
132 spleen highlights the importance of tissue-specific immunity and the problems inherent in
133 protecting the lung against infection. We next assessed how BCG vaccination altered the
134 histological appearance of TB lesions in the lung.

135

136 ***BCG modifies the lung lesions caused by Mtb in a strain-specific manner.***

137 We assessed how BCG vaccination modified lung lesion development after Mtb infection
138 in 19 different CC strains plus C57BL/6 mice. A board-certified veterinarian pathologist (G.B.),

139 blinded to vaccination status of each strain, examined the lung for necrosis, neutrophilic and
140 lymphocytic infiltrates, lesion size and number. Within each pair, the vaccinated group was
141 correctly identified in 15 of the 20 strains analyzed based on reduced severity of the
142 histopathology. In some CC strains that were protected by BCG, increased lymphocytic infiltration,
143 and reduced necrosis were frequently observed after vaccination (Table 2). CC025 (Fig. 2, top
144 left) and CC037 (Fig. 2, middle left) were both susceptible to primary Mtb infection, and the
145 infected lung showed innate immune cells and tissue necrosis. BCG vaccination protects these
146 mice by promoting lymphocyte recruitment and in CC025 BCG vaccination reduced necrosis.
147 However, a reduction in lung CFU associated with BCG vaccination was not always associated
148 with an improved histological appearance of the lung. Despite a significant reduction in lung CFU
149 after BCG, no discernable histopathological differences were observed between unvaccinated
150 and BCG vaccinated CC072 mice, and both groups had non-necrotizing lymphocyte-rich
151 granulomas (Fig. 2, bottom left).

152 BCG did not significantly reduce lung CFU in eight of 20 CC strains. Two of the non-
153 protected CC strains, CC003 (Fig. 2, top right) and CC024 (Fig. 2, middle right) had no discernable
154 histological differences in granuloma structure, content, or severity between control and BCG
155 vaccination when evaluated blindly. In these 2 strains, lesions were generally small to medium-
156 sized, non-necrotizing, with abundant perivascular and peribronchiolar lymphocytes. Given the
157 intermediate susceptibility of the CC003 strain, and BCG's inability to alter lung CFU or change
158 the lesions, the data suggests that BCG does not change CC003 innate or adaptive immune
159 responses to Mtb challenge. Likewise, CC024 is unprotected by BCG vaccination, nor are lung
160 lesions altered, suggesting that neither the innate nor adaptive responses of this resistant strain
161 to primary Mtb infection is further enhanced by BCG vaccination. The natural resistance of CC024
162 strain was not further improved. In contrast, CC004 (Fig. 2, bottom right) did have discernable
163 histopathological changes when controls and BCG vaccinees were evaluated blinded. The
164 differences attributable to BCG vaccination were the greater lymphocytic infiltration with denser

165 perivascular and peribronchiolar lymphocytes. The altered lesions detected after vaccination
166 suggests that the immunity induced by BCG in CC004 mice fails to control bacilli. Together, these
167 data show that the host genetic background impacts the granuloma content, structure, and quality
168 of the immune response to BCG and Mtb challenge in a manner that is not wholly predictable.
169

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

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

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

185 We next asked if there was a correlation between the different cytokines in the lung
186 homogenate and lung CFU, in either unvaccinated or BCG-vaccinated mice. We asked if this
187 correlation differed among protected and unprotected CC strains, and C57BL/6 mice. (Fig.3B, top
188 left). There was a strong correlation between CFU and lung IFNγ levels ($R=0.71$, $p=0.0004$), in
189 unvaccinated C57BL/6 mice but not in BCG-vaccinated C57BL/6 mice. Thus, IFNγ correlated with
190 bacterial burden but not with protection, as observed in prior studies (18, 19). The CC strains

191 protected by BCG had a similar pattern. However, for unprotected CC strains, a similar correlation
192 between IFNγ and CFU was identified in unvaccinated, and BCG vaccinated mice, which we
193 suggest is evidence that these strains were unprotected by BCG vaccination. This pattern held
194 for T cell cytokines (IFNγ and IL-17), mediators produced by myeloid cells (IL-1β, IL-6, KC, and
195 MIP-1α), and TNF (Fig.3B and not shown). These data suggest that protected and unprotected
196 CC strains respond differently to BCG, and that the protected lung environment differs
197 immunologically from an unprotected environment.

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

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

212

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

215 We hypothesized that changes in the frequency of T cell subsets would correlate with
216 protection induced by BCG vaccination. An 18-parameter flow cytometry panel was used to

217 identify and enumerate different immune subsets (Figure S1). As T cells are important in
218 protection against TB, we determined the frequency and relative number of different T cell subsets
219 modulated by BCG vaccination. Not all parameters were obtained for all 24 CC strains (see
220 methods). We quantified CD4 and CD8 T cells, MAIT cells and NKT cells (Fig.4A, and Figure S2).
221 Memory and effector T cell populations were defined as central memory (T_{CM};
222 $CD44^{+}CD62L^{+}CD127^{+}$), effector memory ($T_{EM}$; $CD44^{+}CD62L^{-}CD127^{+}$), effector ($T_{Eff}$;
223 $CD44^{+}CD62L^{-}CD127^{-}$), and resident memory ($T_{RM}$; $CD44^{+}CD103^{+}CD69^{+}$) (Fig.4B, 4C). CD4 T
224 cells subsets were defined in all CC strains based on their expression of the transcription factors
225 Foxp3, Eomes, Tbet and RORγt to identify Treg, Th1, Th2, and Th17 cells, respectively (Fig.4D).
226 For these T cell subsets, we observed tremendous strain-to-strain variation. With one or two
227 exceptions, BCG vaccination did not modulate the distribution of T cells subsets nor was there
228 any correlation between the relative frequency of these T cell subsets and the outcome of BCG
229 vaccination. Thus, the frequency of different T cells subsets in the lungs of CC mice after Mtb
230 infection was largely determined by each strain's genetic background.

231 Expression of CXCR3 and CX_3CR1, which have been used to discriminate between T
232 cells in the lung parenchymal vs. the vasculature, was analyzed in a limited number of CC strains
233 (20). CXCR3 also marks Th1 cells, and in conjunction with CCR6, can differentiate between Th1
234 and Th17 cells, respectively (21). Like other markers, we observed variation in CXCR3 expression
235 between different CC strains, but little or no variation that correlated with vaccination status or
236 susceptibility. In addition to allelic variation, CC strains also carry private mutations. For example,
237 we previously reported that CC042 mice lacks cell surface CD11a expression due to a mutation
238 in the *Itgal* gene on Chromosome 7 (22). Here, we report that CC003 mice don't express CXCR3
239 (Fig.4E). This phenotype is associated with the CXCR3 allele of the PWK founder. Finally, a
240 reduction in $CX3CR1^{+}$ CD4 and CD8 T cells was observed in several of the protected CC strains
241 but not in unprotected strains (Fig.4E). This pattern mirrored the reduction in IFNγ levels observed
242 in some protected CC mice (Fig.3), suggesting that these changes reflect Th1 modulation.

243 Finally, we used principal components analysis to holistically visualize the variation in the
244 T cell phenotyping data. What is striking is the small effect of BCG vaccination in altering T cell
245 phenotypes, in comparison to large effect of genetic strain differences. These data all indicate
246 that genetics differences between the CC strains determine the differences among T cell subsets
247 recruited to the lung, and not BCG vaccination or Mtb infection. Therefore, we next wished to
248 determine whether there were functional differences among the T cell subsets in protected vs.
249 nonprotected CC strains.
250
251 ***The patterns of cytokine production differ in CC strains after BCG vaccination and Mtb***
252 ***infection.***
253 As the cytokines produced by T cells, including IFNγ, TNF, and IL-21, among others, are
254 important in control of Mtb infection, we hypothesized that changes in T cell function would be
255 associated with BCG-induced protection. To measure function, we stimulated lung mononuclear
256 cells (MNC) from each mouse with anti-CD3 or the P300 megapool, which contains 300 peptides
257 representing epitopes from 70 Mtb proteins that are frequently recognized by human CD4 T cells
258 and are also recognized by murine CD4 and CD8 T cells (23, 24). Twenty-six cytokines and
259 chemokines in the culture supernatants were measured using the Legendplex platform 24 hours
260 after stimulation. There was little difference in the ability of lung MNC from mock or BCG
261 vaccinated Mtb-infected C57BL/6 mice to secrete different cytokines or chemokines following
262 P300 stimulation. In contrast, we detected differential production of IFNγ, TNF, IL-2 or IL-17, all
263 cytokines that have been associated with immunity to Mtb infection, among mock vs. BCG
264 vaccinated CC strains (Fig.5A). Interestingly, the MNC from BCG vaccinated CC037, CC031,
265 CC72 and CC001 strains all produced more IL-17 than their respective controls. In contrast, none
266 of the unprotected strains produced significant amounts of IL-17, except for CC023, and its lung
267 MNC produced less IL-17 after BCG vaccination.

268 In parallel, a 14-parameter flow cytometry panel measured intracellular levels of IFNγ, IL-
269 2, IL-17, TNF, IL10, and IL-4, and surface expression of CD107 in both CD4 and CD8 T cells
270 (Figure S3). BCG vaccination had a minimal effect on CD8 T cell cytokine responses to P300
271 stimulation in both C57BL/6 mice as well as most CC lines examined (Fig.5B). However, BCG
272 clearly modified the CD4 T cell cytokine responses in some CC strains (Fig.5C). Patterns of
273 cytokine production were observed in BCG vaccinated CC mice that were not observed in
274 C57BL/6 mice. For example, while BCG vaccination did not alter the frequency of IFNγ$^+$ or TNF$^+$
275 CD4 T cells in CC037 mice, it did increase IL-2 and IL-17 responses. BCG vaccination of CC031
276 mice increased the frequency of CD4 T cells producing IFNγ, IL-2 and TNF but not IL-17. Finally,
277 CD4 T cells from BCG vaccinated had CC072 had increased IFNγ and IL-2 but not TNF or IL-17.
278 Interestingly, no discernable differences in cytokine responses were observed in most of the non-
279 protected CC mice, except for a decrease in IFNγ responses in CC023. Thus, the cytokine
280 responses in both control and BCG vaccinated CC mice differ greatly from classic C57BL/6
281 model. Increased cytokine responses after T cell stimulation were more likely to be detected in
282 protected CC strains. Moreover, BCG vaccination led to quantitative and qualitative changes in
283 the type of T cell responses that persisted after Mtb challenge.
284
285 ***Correlates of Protection.***
286 We then turned to multivariate approaches to discern how BCG-induced immune changes
287 in the lung mechanistically lead to reductions in lung CFU in CC strains. To do so, the data from
288 the lung homogenate cytokines, intracellular cytokine staining, T cell phenotyping, and
289 LEGENDplex were combined, and for each vaccinated mouse the transformed feature data was
290 calculated as: [(vaccinated CC subject– mean(unvaccinated CC strain]. Then, a partial least
291 squares regression (PLSR) model with elastic net feature selection was created to identify
292 features that changed with BCG vaccination in a way that maximally covaried with BCG-induced
293 changes in lung CFU. The model performed significantly better than null models, and the first two

294 latent variables in PLSR model account for 24% and 8%, respectively, of the variability in the lung
295 features (Figure 6A, Figure S4A). The elastic net-selected features in the model included features
296 from all four datasets (Fig.6B, 6C). By plotting the scaled values of the elastic net-selected
297 features for all CC strains, ordered from most protected to least protected, what is most apparent
298 is that the protected strains all had BCG-induced reductions in several inflammatory cytokines
299 such as MIG, CCL3 and CCL4 (i.e., MIP-1α and -1β), IL-6, RANTES, and IL-17. The decrease in
300 these inflammatory mediators could be a sign of resolved infection, in contrast to the unprotected
301 strains which had more inflammatory cytokines which could signify unimpeded infection.
302 Protected strains also had a greater vaccination-induced increase in activated CD4 T cells.
303 Contrarily, protected strains had decreased activated $CD4^{-}CD8^{-}$ T cells producing T-bet.
304 Importantly, IL-2, unlike several other cytokines, was increased in protected strains. Thus, BCG
305 vaccination leads to qualitative and quantitative changes in the T cell response and in the immune
306 environment of the lung after Mtb infection in CC genotypes that are associated with protection.
307

308 ***Immune responses in protected and unprotected CC strains differ following Mtb infection***

309 While the PLSR model highlights potential correlates of lung protection across all
310 genotypes, it may miss correlates that only apply to certain subsets of CC strains. We
311 hypothesized that naturally resistant and susceptible strains may be protected in different ways.
312 To test this, we created separate partial least squares discriminant analysis (PLSDA) models for
313 resistant and susceptible CC strains and examined the features that best separated the BCG-
314 protected and non-protected strains in each of these categories. (Figure 1C). The data used for
315 the models was processed the same way as in the PLSR models, although we also looked at the
316 vaccinated and unvaccinated mice separately in addition to the vaccine-induced differences. Our
317 first analysis compared resistant CC strains that were protected by BCG (n=4) or not (n=8), based
318 on lung CFU. As described earlier (Figure 4), the unvaccinated and vaccinated mice of each CC
319 strain were very similar and clustered together when the elastic net-selected features of the model

320 were hierarchically clustered, indicating that the genetic background was an important
321 determinant of outcome (Figure 7A). The protected and unprotected mice were both split into
322 multiple clusters, showing that there were multiple distinct types of immune responses that were
323 associated with protective immunity in resistant mice. The model could predict protected vs non-
324 protected CC strains accurately and thus there is a multivariate signature of these categories
325 (Figure S4B). However, there were no clear individual features distinguishing the categories. In
326 contrast, the vaccine-induced model comparing protected and non-protected resistant strains did
327 yield more univariate trends in addition to the strong multivariate signature. Namely, both CD4
328 and CD8 T cells making IL-2 are increased in protected strains, while several lung homogenate
329 cytokines as well as double negative T cells expressing CD107a and producing IL-17 decreased
330 in protected strains (Figure 7B).

331 We next analyzed CC strains that were inherently more susceptible than B6 mice, and
332 were protected by BCG (n=4) or not (n=8) (Figure 7C). This yielded two clear phenotypes of
333 susceptible, non-protected CC strains. One subcluster comprising CC004 and CC044 had high
334 levels of T_{CM}, $CD4^{+}Foxp3^{+}$, $CD8^{+}PD\text{-}1^{+}$ and polyfunctional $CD4^{-}CD8^{-}$ T cells. The other cluster
335 had high levels of lung MIP-1a and $CD3^{-}Foxp3^{+}$ T cells, and fewer polyfunctional $CD4^{-}CD8^{-}$ T
336 cells. Interestingly, the unprotected strains with high $CD4^{+}Foxp3^{+}$T cells had low $CD8^{+}Foxp3^{+}$ T
337 cells, and the unprotected strains with high $CD8^{+}Foxp3^{+}$ T cells had low $CD4^{+}Foxp3^{+}$ T cells.
338 When looking at the protection model with the susceptible strains, several features were similar
339 to the model with all mice. In this model it is evident that while IL-17 in the lung homogenate
340 decreases with BCG vaccination in protected strains, the amount of IL-17 when stimulated with
341 the P300 peptide pool increases. (Figure 7D). Overall, the features that distinguish protected and
342 non-protected strains are similar for naturally resistant and susceptible mice.

343 Finally, we compared susceptible CC strains that were unprotected by BCG (i.e., poor
344 immunity) to inherently resistant CC strains that were protected by BCG (i.e., good immunity).
345 Unsurprisingly, there is clear separation of categories in the model with all mice. "Poor immunity"

346 CC strains tended to have high levels of inflammatory monokines (Figure 7E, 7F). "Good
347 immunity" CC strains high levels of RANTES after stimulation with anti-CD3. A subset of good
348 immunity strains including CC031 and CC011 also have high frequencies of IFNγ-producing CD8
349 T cells and various subsets of $CD4^{-}CD8^{-}$ T cells including those producing IFNγ or expressing
350 CD11a. In contrast, a subset of poor immunity strains had high levels of polyfunctional $CD4^{-}CD8^{-}$
351 T cells and ILCregs. The protection model revealed similar features with the interesting addition
352 of activated CD4 T_{RM} cells increasing with BCG vaccination in poor immunity but not good
353 immunity CC strains.

354 We find that immune features can be identified using PLSDA that can categorize the
355 inherent susceptibility and response to BCG of CC strains following Mtb infection. The expression
356 of different immune features was variable between CC strains even in the same protection
357 category, indicating no single path to protection. Looking at more refined subgroups of CC strains
358 allowed us to find features that define subclasses of protected and non-protected strains and
359 highlight features that did not emerge in the PLSR model such as non-T_{RM} effector and central
360 memory T cell populations.

361

362 ***Th1/17 cells detected in a subset of CC mice that are protected by BCG vaccination***

363 An aim of this study was to identify factors that were associated with protection in
364 genetically diverse CC mice after vaccination and Mtb challenge. Vaccination enhanced IL-17
365 responses in the CC037, CC001, CC072 and CC031 strains, which were all protected by BCG
366 (Figure 5A, 5B). As the CC strains that produced IL-17 after antigen stimulation also had strong
367 IFNγ response, we assessed these T cell responses for dual production of IFNγ and IL-17. BCG
368 vaccinated CC037 mice had the highest IL-17 levels in lung homogenate (Figure 5A, 5B), and
369 after stimulation with P300 or anti-CD3 mAb, CD4 T cells were identified that produced both IFNγ
370 and IL-17 (Figure 8A). Dual $IFN\gamma^{+}IL\text{-}17^{+}$ CD4 T cells were not detected in C57BL/6 mice after
371 P300 stimulation, regardless of vaccination status. BCG vaccination significantly increased the

372 frequency of IFNγ$^+$IL-17$^+$-producing CD4 T cells in the CC037, CC072 and CC001 (Figure 8A,
373 8B). Further analysis identified BCG primed CD4 T cells from CC037 and CC001 to coexpress
374 both Tbet and RORγt after Mtb infection (Figure 8C, 8D), which are the master transcription
375 factors for Th1 and Th17 cells, respectively.

376 We next measured the polyfunctionality of the T cell response and expanded our cytokine
377 analysis to include IL-2 and TNF. In addition to IFNγ and IL-17, an increase in the proportion of
378 CD4 T cells producing IL-2 or TNF was detected (Figure 8E), indicating that CD4 T cell
379 polyfunctionality was increased in BCG vaccinated mice. A detailed analysis of the different
380 polyfunctional populations revealed different distributions of Th1/17 cells within each CC strain.
381 IFNγ$^+$IL-2$^+$TNF$^+$IL-17$^+$ cells were significantly increased in CC037 following BCG vaccination
382 (Figure S5, red box). Similarly, IFNγ$^+$TNF$^+$IL-17$^+$ cells were increased following BCG vaccination
383 in CC072 and CC001 (Figure S5, blue box). BCG led to a decrease in the frequency of IFNγ
384 "single positive' CD4 T cells in all four CC lines (Fig. S5, purple box). However, among these CC
385 strains only CC037 had a significant decrease in IFNγ$^+$TNF$^+$ cells. This contrasts with C57BL/6
386 mice where BCG vaccination had a minimal effect on CD4 T cell polyfunctionality. Significant
387 decreases were observed for IFNγ$^+$TNF$^+$ dual producers and IFNγ single positive cells in Mtb-
388 infected BCG-vaccinated C57BL/6 mice, consistent with the observation that Th1 dominant
389 response is associated with decrease bacterial burden in this background.

390 Finally, we performed a correlation analysis between all 60 immune features and the
391 normalized lung CFU from unvaccinated control and BCG vaccinated mice. For C57BL/6 mice,
392 IFNγ-producing and polyfunctional CD4 T cells correlated with lung CFU. In CC037 mice, IFNγ-
393 producing CD8 T cells correlated with lung CFU. Importantly, many features were negatively
394 correlated with lung bacterial burden in CC037 mice (Figure 8F). The frequency of CD4$^+$Roryt$^+$,
395 and antigen-stimulated IL-17 secretion and IFNγ/IL-17 dual-producing CD4 T cells were
396 significant even after correction for multiple testing, suggesting that these features were
397 associated with protection following BCG. There was a strong correlation between the frequency

398 of IFNγ/IL-17 dual-producing CD4 T cells and lung CFU in CC037 mice (Spearman r = -0.8947)

399 (Figure 8G).

400

401 **Discussion**

402 We sought to determine how pre-existing immunity to BCG modifies the immune response
403 to Mtb infection in a genetically diverse population using CC mice to model an outbred population.
404 An important outcome is our ranking of how BCG induced protection varies across 24 CC strains,
405 compared to control unvaccinated mice and relative to C57BL/6 mice, which were used as a
406 reference strain. We kept constant the BCG strain, growth conditions, dose, and vaccination
407 route. All mice were rested for 12 weeks after vaccination, challenged with the same Mtb strain,
408 and analyzed four weeks after infection. The different CC strains were categorized as being either
409 protected or unprotected by BCG against Mtb infection, based on whether vaccination led to a
410 statistically significant reduction in lung bacillary burden measured four weeks after low dose
411 aerosol infection. After controlling for these variables, the surprising result was that BCG protected
412 only half of the CC strains tested. As these strains differed primarily in the genes and alleles they
413 inherited from the CC founder strains, we conclude that the host genetic background has a major
414 influence on whether BCG confers protection against Mtb infection. These results extend our
415 analysis of the eight CC founder strains, which also vary in their ability to be protected by BCG
416 vaccination (15). Not only do environmental mycobacteria and genetic variations in BCG strains
417 used around the world interfere with the ability of BCG-induced protection, but we suggest that
418 host genetics should be considered a third important barrier to vaccine-mediated protection to TB.

419 Why BCG doesn't induce protection in all CC strains? Protection was largely defined
420 based on reductions in lung CFU four weeks after aerosol Mtb infection, which is a standard time
421 point in murine vaccine studies. Interestingly, the CC strains that were not protected by BCG can
422 be categorized based on their lung pathology. In half of the unprotected CC strains (e.g., CC004),
423 the BCG-vaccinated mice had changes in lung histology (e.g., more lymphocytic infiltrates) that
424 differed from the unvaccinated controls. We conclude that immunization generated a memory
425 recall response but it was insufficient to control lung CFU. Histological changes couldn't be

426 identified in other strains (e.g., CC024), and we suggest that BCG generated a suboptimal
427 immune response.

428 Importantly, we wished to identify the components of the immune response stimulated by
429 BCG, which were subsequently recalled after Mtb infection and associated with protection. We
430 hypothesized that CC strains protected by BCG would share common immunological features.
431 We postulated that by comparing protected to unprotected CC strains would allow us to filter out
432 components of the immune response that were elicited by vaccination or infection but not
433 associated with protection. The T cell immune response following BCG vaccination and Mtb
434 challenge was extensively characterized as T cells are essential for protection against Mtb.
435 Although considerable diversity was observed, BCG vaccination had little impact on the
436 composition of T cells recruited and maintained in the lung after infection. Instead, the variability
437 was largely shaped by the genetic background. In contrast, functional skewing of the T cell
438 responses by BCG was evident in several of the protected CC strains. While many of the
439 protected CC strains had unique flavors of immunity, we identified immune signatures associated
440 with BCG-elicited protection against TB. We found that different immune parameters were
441 associated with protection in the four different groups

442 Like lung homogenate IFNγ levels, lung IL-17 levels strongly correlated with lung CFU.
443 However, the correlation between IL-17 and lung CFU was lost in BCG protected CC strains after
444 vaccination and challenge. These data suggest that like IFNγ, IL-17 levels are a correlate of
445 disease, and not of protection. Interestingly, IL-17 concentrations co-varies with IL-1β and IL-6,
446 two cytokines that are involved in skewing of Th17 cells, and with TNF, KC, MIP2, and G-CSF,
447 all which are induced by IL-17. IL-17 could be more generally important in driving lung
448 inflammation than predicted by the C57BL/6 model. In contrast, the CD4 T cells from four of the
449 13 protected strains by BCG co-produced IFNγ and IL-17 after Mtb infection. None of the 11
450 unprotected strains had this phenotype, and these dual producers were detected only in BCG
451 vaccinated mice and not in sham controls, nor were they detected in the C57BL/6 reference strain.

452 Thus, this is an example of an immune response that BCG primes but is not manifested until the
453 CD4 T cell response is recalled by Mtb infection. These cells are similar to Th1/17 cells, or Th1*,
454 that are identified following TB infection, and are associated with late-forming, low burden
455 granulomas in a NHP TB infection model (25-27). However, an important difference is that IL-17
456 is not produced by Th1* cells. Importantly, these dual IFNγ/IL-17 producing CD4 T cells were
457 among the strongest correlates of protection. These data present a paradox. On the one hand,
458 IL-17 is a correlate of disease; conversely, IL-17-producing or RoRγt-expressing CD4 T cells are
459 a correlate of protection. We hypothesize that a cellular feature other than IL-17 production, such
460 the capacity for self-renewal, migration, or plasticity, is associated with Mtb control. The CC strains
461 will be invaluable in further unraveling the role of IL-17 and IL-17-producing T cells in immunity to
462 TB.

463 While extrinsic factors, many that disrupt cell mediated immunity, increase the risk of
464 developing TB, what drives the emergence of disease in most people is largely unknown. Host
465 genetics affects susceptibility to TB both in mice and people, and genetic loci have been identified
466 that increase TB risk (i.e., MSMD, polygenic) (16)[ref]. The factors that influence BCG efficacy
467 remains controversial. We hypothesize that genetic factors contribute to variation in the
468 effectiveness of BCG against pulmonary Mtb. Our analysis shows that while certain mouse
469 genotypes cannot be protected by BCG, the inherent susceptibility or resistance of the mouse
470 strain is not a predictor of vaccine efficacy. This suggests that a priori, vaccines are appropriate
471 interventions even in individuals that might have a genetic susceptibility to TB.

472 One caveat is that we have tested the efficacy of single vaccine in CC mice. The variation
473 in host responses of CC mice could be leveraged as an ideal platform for testing the efficacy of
474 preclinical vaccines in the context of a genetically diverse background. It is possible that vaccines
475 that work through alternate mechanisms may be able to protect CC mice that are refractory to
476 BCG vaccination. Instead of optimizing vaccine concepts in a single mouse genotype (e.g.,
477 C57BL/6 mice), identifying vaccines that are effective in numerous different genotypes (i.e.,

478 genetically distinct individuals), as modeled by CC or DO mice, represents an alternative
479 paradigm for preclinical vaccine development. Our data establish CC strains as a novel animal
480 resource that can facilitate the discovery of correlates of vaccine-induced protection and
481 accelerate preclinical vaccine testing.
482

483 **Acknowledgements**

484 We thank Clare Smith and Megan Proulx for helpful discussions, as well as other members of the
485 Behar lab. The MR-1 and CD1d tetramers were obtained through the NIH Tetramer Core Facility
486 (contract 75N93020D00005). The Vanderbilt Medical Center Digital Shared Histology Resource
487 (DHSR) digitized and hosted our histology data.

488

489 **Author contribution**

490

491 **Declaration of interests**

492 The authors declare no competing interests.

493

494 **Funding**

495 Supported by P01 AI123286 (C.M.S., M.F., S.M.B.), NSF Graduate Research Fellowship Grant
496 No.1745302 (D.G.), and R01 HL14541 (G.L.B). This work is also supported by NIAID IMPAcTB
497 and Army Institute for Collaborative Biotechnologies (D.L., D.G.).

498

499 **Materials and Methods**

500

501 Mice

502 Female C57BL/6/J (#0664) were purchased from The Jackson Laboratory. Female mice from 24
503 CC strains were purchased from the UNC Systems Genetics Core Facility (SGCF) between July
504 2018 and August 2019. The 24 CC strains used in this study include: CC001/Unc, CC002/Unc,
505 CC003/Unc, CC004/TauUnc, CC010/GeniUnc, CC011/Unc, CC012/GeniUnc, CC023/GeniUnc,
506 CC024/GeniUnc, CC025/GeniUnc, CC026/GeniUnc, CC031/GeniUnc, CC032/GeniUnc,
507 CC035/Unc, CC036/GemUnc, CC037/TauUnc, CC040/TauUnc, CC041/TauUnc,
508 CC042/GeniUnc, CC044/Unc, CC045/GeniUnc, CC059/TauUnc, CC061/GeniUnc,,
509 CC072/TauUnc. More information regarding the CC strains can be found at
510 http://csbio.unc.edu/CCstatus/index.py?run=AvailableLines.information.
511 All animal studies were conducted using the relevant guidelines and regulations, and approved
512 by the Institutional Animal Care and Use Committee at the University of Massachusetts Medical
513 School (UMMS) (Animal Welfare A3306-01), using the recommendations from the Guide for the
514 Care and Use of Laboratory Animals of the National Institutes of Health and the Office of
515 Laboratory Animal Welfare.

516

517 Bacterial strains

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

521

522 Aerosolized Mtb infection of mice

523 Mice were infected by the aerosol route as previously described (17). Frozen bacterial stocks
524 were thawed and sonicated for 1 minute and then diluted into 5 ml of 0.01% Tween-80 in PBS.

525 The diluted bacterial suspension was aerosolized to infect mice using a Glas-Col chamber (Terre
526 Haute). The average number of bacteria delivered into the lung was determined for each
527 experiment by plating lung homogenate from 4-5 mice 24 hours after infection and ranged
528 between 40-150 CFU/mouse.
529
530 Bacterial burden in lung and spleen
531 Infected mice were euthanized at pre-determined timepoints, and the left lung or whole spleen
532 were homogenized using 2 mm zirconium oxide beads (Next Advance) in a FastPrep
533 homogenizer (MP Biomedicals). Tissue homogenates were serially diluted and plated on 7H11
534 agar plates (Hardy Diagnosis). CFU was enumerated after 19-21 days of incubation at 37°C and
535 5% CO_2.
536
537 Lung cell preparation
538 Single cell suspensions were prepared by homogenizing lungs using a GentleMACS tissue
539 dissociator (Miltenyi), digesting with 300 U/ml collagenase (Sigma) in complete RPMI (10% FBS,
540 2 mM L-Glutamine, 100 units/ml Penicillin/Streptomycin, 1 mM Na-Pyruvate, 1X Non-essential
541 amino acids, 0.5X Minimal essential amino acids, 25 mM of HEPES, and 7.5 mM of NaOH) at
542 37°C for 30 minutes, and followed by a second run of dissociation using the GentleMACS.
543 Suspensions were then sequentially filtered through 70 µm, treated with ACK lysis buffer (Gibco),
544 then filtered through 40 µm strainers.
545
546 Flow cytometric analysis
547 Cells were first stained with Zombie Fixable Viability dye (Biolegend) for 10 minutes at room
548 temperature (RT), after which cells were stained with antibodies in autoMACS running buffer
549 (Miltenyi) containing 5 ug/ml of anti-mouse CD16/32 (BioXcell) for 10 minutes at 4°C. In some

550 experiments, MAIT cells were identified using the MR-1-PE tetramer (NIH tetramer core facility)
551 at 37°C for 1hr. Cell were then stained with surface antibody cocktail for 20 minutes at 4°C. In
552 some experiments, CD1d-APC tetramer (NIH tetramer core facility) were also added at this time
553 to identify NKT cells. To measure transcription factor (TF) expression, cells were fixed and
554 permeabilized for 30 minutes at RT, followed by staining for 30 minutes with antibodies to the
555 different TF at RT using the Foxp3/TF staining buffer set (ThermoFisher). Samples were acquired
556 on Aurora (Cytek) or MACSquant (Miltenyi).
557
558 For intracellular staining, cells were first stimulated for 5 hours with either anti-CD3 antibodies (1
559 ug/mL, Biolegend) or P300 peptide megapool (23) which contains 300 peptides representing 90
560 antigens, and contains epitopes recognized by both murine CD4 and CD8 T cells. GolgiPlug (BD
561 Biosciences) was also given at this time (0.5ug/well). Viability and surface staining were
562 performed as described, after which cells were fixed instead with Fixation/Permeabilization kit for
563 20 minutes at 4°C. Cells were then stained with an intracellular cytokine antibody cocktail for 20
564 minutes at 4°C.
565
566 To inactivate the bacteria, samples were fixed with 1% paraformaldehyde/PBS for 1 hour at room
567 temperature and then washed with MACS buffer. Flow data were analyzed using FlowJo v10.7.1.
568 Polyfunctionality was examined using the SPICE platform (28).
569
570 LEGENDplex cytokine analysis
571 Lung cells were stimulated with P300 peptide megapool (1ug/mL) or anti-CD3 (1ug/mL) at a
572 concentration of 10^6 cells/mL for 24hrs. Culture supernatants were filtered through 0.2um filter,
573 after which cytokine levels were measured with LEGENDplex kits from Biolegend. Specifically,
574 we used their mouse T helper cytokine panel (IFNγ, TNF, IL-2, IL-4, IL-5, IL-6, IL-9, IL-10, IL-13,

575 IL-17a, IL-17f, IL-21, IL-22) as well as their mouse proinflammatory chemokine panel (RANTES,
576 MIP-3α, Eotaxin, TARC, KC, MCP-1, MIG, IP-10, MIP-1α, MIP-1β, BLC, LIX, MDC) for a total of
577 26 cytokines and chemokines.
578
579 Lung homogenate
580 Cytokines and chemokiens were measured from lung homogenates using the 31-plex discovery
581 Assay (MD31) from Eve technologies (Calgary, Canada).
582
583 Histology analysis
584 The right middle lung lobe from each mouse was immersion fixed in 10% formalin immediately
585 following euthanasia. Lung tissue was processed, embedded in paraffin, sectioned at 5um, and
586 stained with hematoxylin and eosin at the Morphology Core Facility at UMass Chan Medical
587 School for histological examination by a board-certified veterinary pathologist (GB). Blind
588 classification (BCG vs PBS) within each CC strain was determined based on the assumption that
589 BCG vaccination would have three possible effects on lungs: (i) reduce necrosis/neutrophilic
590 infiltrates; (ii) increase lymphocytes and/or plasma cell abundance; and (iii) reduce granuloma
591 size. Accuracy of the blind classification was assessed by an independent investigator (AO). All
592 glass slides were subsequently re-evaluated to more closely examine and describe
593 histopathology features attributable to BCG vaccination and attempt to find unique responses.
594 Glass slides were then digitized by Vanderbilt University Medical Center (Nashville, TN) and
595 whole images generated using Aperio ScanScope (Buffalo Grove, IL, USA) at 400 times normal
596 magnification with jpeg compression. Representative images were captured using Aperio
597 ImageScope magnified 40 and 400 times and imported into GraphPad Prism 9 to generate the
598 panel Figure. Digital images were not adjusted. Whole slide imaging and quantification of
599 immunostaining were performed in the Digital Histology Shared Resource (DHSR) at Vanderbilt
600 University Medical Center.

601

602 Univariate correlations

603 To evaluate how C57BL/6 mice, protected CC strains, and non-protected CC strains compared
604 in their relationships between lung cytokine production and Mtb burden, a series of univariate
605 Pearson correlations were conducted. Lung homogenate cytokines and chemokines
606 measurements in the form of concentrations were natural logged and each of their Pearson
607 correlations with the log10 lung CFU or spleen CFU were evaluated, with both the cytokine and
608 CFU measurements separated into groups of C57BL/6, protected CC strains, and non-protected
609 CC strains. A Bonferroni multiple hypothesis correction was applied to account for the number of
610 correlations calculated.

611

612 Multivariate analyses

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

617

618 For the ICS, the data were analyzed as combinatorics of the seven markers measured in order to
619 best capture polyfunctionality, and further divided by CD4 T cell, CD8 T cell, $CD4^{-}CD8^{-}$ T cell,
620 and $CD3^{-}$ categories. The features were filtered by a minimum mean and variance percentage of
621 0.000001. Only P300 stimulated samples were used from the ICS dataset. The flow cytometry
622 data was in the form of percentages of $CD3^{+}$ or $CD3^{-}$ cells, while the immunoassay data was in
623 the form of concentrations. The combined dataset was natural log transformed and subsequently
624 z-scored by measurement across all samples. For models that involved BCG-induced protection,
625 these z-scored values were further transformed by subtracting the features from each vaccinated
626 mouse and the mean of the features of the unvaccinated mice from the same mouse strain.

627

628 For the PLSR and PLSDA models, elastic net feature selection was performed in 100 trials of 5-
629 fold cross validation, where features that were present in at least 80% of trials were included in
630 the final model. Each model was compared to two null models: one with permuted lung CFUs and
631 one with randomly chosen features. The mean squared error of the prediction of lung CFU (for
632 the PLSR) or the classification accuracy (for the PLSDAs) was compared to that of the null models
633 in 50 trials of 5-fold cross validation.

634

635 Quantification and Statistical Analysis

636 Statistical analysis was performed using Graphpad Prism 9. P-values were calculated using
637 unpaired t test, one-way ANOVA, or two-way ANOVA as indicated in the figure legends.
638 Multivariate analyses were performed in R (v3.6.0 and 4.0.4) using the ropls Bioconductor
639 package.

640

641

642

643

644

645

[a] CC strain	[b] Batch	[c] MHC Haplotype	[d] log(ΔCFU) Exp1 (lung)	[d] log(ΔCFU) Exp2 (lung)	[e] log(ΔCFU) Exp1 (spleen)	[e] log(ΔCFU) Exp2 (spleen)
CC001	4	b	0.94	1.45	2.79	1.45
CC002	3	b	0.83	n/a	0.46	n/a
CC003	3	g	0.07	-0.61	0.77	0.15
CC004	4	b	0.05	0.08	-0.16	-0.12
CC010	1	b	0.53	n/a	0.92	n/a
CC011	3	g7	1.04	n/a	2.08	n/a
CC012	4	wsb	0.79	n/a	0.89	n/a
CC023	1	b	1.17	-0.09	1.47	0.32
CC024	2	cast	0.15	0.24	0.38	0.62
CC025	3	pwk	1.40	n/a	0.43	n/a
CC026	2	cast	0.93	1.05	3.75	1.87
CC031	3	b	1.22	1.35	1.26	1.60
CC032	1	b	1.16	n/a	1.06	n/a
CC035	4	wsb	0.91	n/a	1.13	n/a
CC036	3	g7	0.62	n/a	1.84	n/a
CC037	1	b	1.37	1.84	1.07	1.05
CC040	2	b	-0.64	-0.28	-0.35	1.14
CC041	2	b	0.14	0.94	0.71	0.94
CC042	1	b	1.65	0.96	1.00	1.32
CC044	4	b/g7	0.08	0.13	0.12	0.39
CC045	4	z	0.24	n/a	0.13	n/a
CC059	2	b/a	1.39	n/a	1.17	n/a
CC061	1	b	0.93	n/a	1.58	n/a
CC072	2	b	0.97	1.37	1.06	1.41

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

647 [a] A group of C57BL/6 mice were included in all batches as a reference strain. Data was normalized
648 to internal C57BL/6 control as indicated in the methods.
649 [b] The first cohort of CC mice were analyzed in 4 separate batches as indicated.
650 [c] MHC haplotype of each CC mouse as inherited from CC founders.
651 [d, e] Thirteen of the 24 CC strains were analyzed a second time to validate the original results and
652 perform more in-depth immune analysis. ΔLungCFU[d] and ΔSpleenCFU[e] for the first experiment
653 and the repeat experiment are listed. Mice that were exacerbated by BCG vaccination are
654 highlighted in red.
655

[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.

663 **Figure legends**

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

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

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

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

680

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

682 Representative lung histology of three CC strains (CC025, CC037, CC072) that were protected
683 by BCG, after sham or BCG vaccination and Mtb challenge (Mtb and BCG+Mtb, respectively)
684 (left). Representative lung histology of three CC strains (CC003, CC004, CC024) that were
685 unprotected by BCG (right). Original magnification, 40X (inset, 400X). Scale bar, 500 μm (inset,
686 50 μm).

687

688 Figure 3. BCG-protected CC mice have an altered lung microenvironment of CC mice after Mtb
689 infection.
690 (A) IFNγ levels were measured in the lungs of either PBS or BCG vaccinated CC mice at 4 weeks
691 post infection. A two-way ANOVA was performed using Sidak's correction for multiple
692 comparisons. ***, $p < 0.001$; ****, $p < 0.0001$".
693 (B) – (F) Pearson correlation between lung CFU and (B) IFNγ, IL-17, TNF, IL-1β, KC and IL-6 (C)
694 LIX (D) RANTES (E) MIP-1β and (F) M-CSF, MIP-2, G-CSF, LIF, comparing PBS and BCG
695 vaccinated mice at 4 weeks post infection in C57BL/6 mice (B6), non-protected (NP) or protected
696 (P) CC mice.
697
698 Figure 4. Frequency of various T cell populations in protected and unprotected CC mice following
699 Mtb infection.
700 (A) Proportion of $CD3^+$ cell, $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells in the lungs of PBS and BCG
701 vaccinated CC mice at 4 weeks post infection.
702 (B) Proportion of memory CD4 and CD8 T cells in the lungs of either PBS or BCG vaccinated CC
703 mice at 4 weeks post infection, as defined by $CD44^+$ and various combinations of CD62L and
704 CD127 expression; $CD62L^+$ $CD127^+$ (central memory), $CD62L^-$ $CD127^-$ (effector), $CD62L^-$
705 $CD127^+$ (effector memory).
706 (C) Proportion of resident memory CD4 and CD8 T cells in the lungs of either PBS or BCG
707 vaccinated CC mice at 4 weeks post infection, as defined by $CD44^+$ $CD103^+$ $CD69^+$.
708 (D) Proportion of Th1, Treg and Th17 cells in the lungs of either PBS or BCG vaccinated CC mice
709 at 4 weeks post infection, as defined by $CD44^+$ and $Tbet^+$ (Th1), $Foxp3^+$ (Treg) or RoRγt (Th17)
710 expression.
711 (E) Proportion of CD4 and CD8 T cells expressing either CXCR3 or CX_3CR1 in the lungs of either
712 PBS or BCG vaccinated CC mice at 4 weeks post infection.

713 (F) PCA of the T cell phenotyping data. Each point represents the average of usually 5 mice within
714 a given mouse strain that are either BCG vaccinated or unvaccinated. Open symbols,
715 unvaccinated; closed symbols, BCG vaccinated. Each color represents a different mouse strain.
716 (A–E) The data represent the mean ± SEM from one experiment. Two-way analysis of variance
717 with the Benjamini and Hochberg multiple comparsion method. The FDR was set to 0.05 and the
718 numbers in the figures are the q value.
719
720 Figure 5. Cytokine response in protected and unprotected CC strains following Mtb infection.
721 (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated for 24
722 hours with the P300 peptide pool. IFNγ, IL-2, TNF and IL-17a were measured in the supernatants.
723 (B) and (C) Cells from the lungs of CC mice at 4 weeks post infection were stimulated for 5 hours
724 with the P300 peptide pool. Frequencies of CD4 or CD8 T cells that expressed IFNγ, IL-2, TNF
725 or IL-17 was determined using intracellular cytokine staining.
726 (A–C) The data represent the mean ± SEM from one experiment. Two-way analysis of variance
727 with Sidak's multiple comparison test was used to determine significance. *, $p < 0.05$; **, $p < 0.01$;
728 ***, $p < 0.001$; ****, $p < 0.0001$.
729
730 Figure 6. Multivariate correlates of protection across all 24 CC strains
731 (A) A partial least squares regression (PLSR) model was performed to associate BCG-induced
732 changes in lung immune features with BCG-induced changes in lung CFU. Each data point
733 represents a vaccinated mouse that has the mean of the unvaccinated mice from the same strain
734 subtracted from it. The scores of the top two latent variables of the model are shown, with the
735 color of each data point representing the changed in lung CFU induced by BCG vaccination. The
736 percents of the variance in the lung features captured by each latent variable are shown as
737 percentages in the axes.

738 (B) The bar graph depicts the loadings of the first latent variable of the PLSR model in (A). Positive
739 loadings indicate immune features that increase from BCG vaccination in samples that are on the
740 positive x-axis in the scores plot, which generally indicates they have worse protection. Negative
741 loadings conversely indicate immune features that generally increase from BCG vaccination in
742 samples that are protected by vaccination.
743 (C) The heatmap illustrates all features selected to be included in the model, indicating they were
744 important for associating strains with the lung protection continuum. Colors are representations
745 of z-scored features, with a different color representing each 10% quantile of the data. Columns
746 are averages of the mice in the CC strains (n=5 for most cases). Feature names are colored
747 based on the dataset from which they come. In (B) and (C), features from the ICS dataset indicate
748 production of the cytokines listed and no production of any measured cytokines not listed.
749
750 Figure 7. Identifying signatures of protection in subclasses of CC strains
751 (A), (C), (E) Scores plot (top) and heatmaps (bottom) of selected features for PLSDA models.
752 Each dot represents a single mouse. The two colors in the scores plots signify the two classes
753 being compared. An ellipse is drawn around the 95% confidence interval for each class. Percent
754 of variance of the immune feature data is shown on the axes. Heatmaps show features that were
755 most important for distinguishing the two classes. Colors are representations of z-scored features.
756 Columns are averages of the mice in the CC strains, with vaccinated and unvaccinated mice
757 being separated (n=5 for most cases). Feature names are colored based on the dataset from
758 which they come (blue=T cell phenotyping, brown=lung homogenate cytokines, green=ICS,
759 magenta=LEGENDplex). NP = non-protected, P = protected, R = resistant, S = susceptible.
760 (B), (D), (F) PLSDA models performed as in (A), (C), and (E), except that features are the changes
761 induced by BCG vaccination as in Figure 6.
762
763 Figure 8. Th1/17 cells correlate with BCG-mediated protection in a subset of protected CC mice.

764 (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated for 5
765 hours with either aCD3, P300 peptide pool or media control. Representative flow plots of IFNγ
766 and IL-17 producing CD4 T cells from CC037 and C57BL/6 are shown. Data is representative of
767 two independent experiments.
768 (B) Frequencies of CD4 T cells that expressed both IFNγ and IL-17 was determined using
769 intracellular cytokine staining. Data is representative of two independent experiments. One-way
770 analysis of variance with Sidak's multiple comparison test was used to determine significance. **,
771 $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.
772 (C) Representative flow plots of CD4 T cells expressing either T-bet or RORγt in CC037 or
773 C57BL/6 mice. Data is representative of two independent experiments.
774 (D) Frequencies of CD4 T cells that expressed both T-bet or RORγt in CC037 or C57BL/6 mice.
775 Data is representative of two independent experiments. One-way analysis of variance with Sidak's
776 multiple comparison test was used to determine significance. **, $p < 0.01$; ***, $p < 0.001$.
777 (E) Proportion of CD4 T cells that express various combinations of IFNγ, TNF, IL-2 or IL-17 in
778 either PBS or BCG vaccinated mice at 4 weeks post Mtb infection. Plots are generated using the
779 SPICE platform (28).
780 (F) Volcano plots. Correlation between 60 immune features and the lung CFU determined for
781 CC037 and C57BL/6 mice (from Batch 1). Dotted line, $p = 0.05$. *, $q < 0.05$, corrected with the
782 Benjamini and Hochberg multiple comparison method (FDR = 0.05)
783 (G) Correlation of $IFN\gamma^{+}IL\text{-}17^{+}$ cells and bacterial burden in the lungs of either PBS or BCG
784 vaccinated CC037 mice at 4 weeks post Mtb infection. Colors represent BCG vaccinated mice
785 (red) or unvaccinated mice (blue). Spearman *r* and p value are shown.
786
787

788 **SUPPLEMENTAL INFORMATION**

789

790 **Supplementary tables**

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

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

793

794 **Supplementary figure legends**

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

796 (A) Lymphocytes were identified using FSC and SSC properties, after which viability dyes were

797 used to identify live cells. Singlet gates were used to exclude doublets, after which T cells and

798 non-T cells were identified using the lineage marker CD3.

799 (B) NKT and MAIT cells were identified from the $CD3^+$ gate using their antigen-loaded CD1d and

800 MR1 tetramers, respectively. $CD4^+$, $CD8^+$ and $CD4^-CD8^-$ T cells were also identified from this

801 population for further analysis.

802 (C) Expression of various markers were also examined within the $CD3^-$ population. As an

803 example, flow plot represents expression of T-bet within the $CD3^-$ population.

804 (D) Identification of memory T cells. CD44 was used to identify antigen-experienced T cells.

805 Resident memory T cells (T_{RM}) were identified based on CD103 expression. From the $CD103^-$

806 fraction, CD127 and CD62L expression identified central memory (T_{CM}), effector (T_E) and effector

807 memory (T_{EM}). Subset identification of various T cell populations (ex: $Tbet^+CD4^+$ T cells) as well

808 as activation markers (ex: PD1, CD11a) were also examined in $CD4^+$, $CD8^+$ and $CD4^-CD8^-$ T cell

809 subsets.

810 (E) Naïve T cells identified within the $CD44^-$ compartment using CD62L expression.

811

812 Figure S2. Frequencies of non-conventional T cells in CC mice. Frequencies of MR-1-restricted

813 MAIT cells and CD1d-restricted NKT cells were measured in the lungs of PBS or BCG vaccinated

814 CC mice at 4 weeks post Mtb infection. The data in the graphs represent the mean ± SD of one
815 experiment. Two-way analysis of variance with the Benjamini and Hochberg multiple comparsion
816 method. The FDR was set to 0.05 and the numbers in the figures are the q value.
817
818 Figure S3. Gating strategy for the analysis of T cell function (intracellular cytokine staining).
819 (A) Lymphocytes were identified as using FSC and SSC properties, after which viability dyes were
820 used to identify live cells. Singlet gates were used to exclude doublets, after which T cells were
821 identified using the lineage marker CD3.
822 (B) CD4, CD8 and $CD4^{-}CD8^{-}$ T cells were also identified from the $CD3^{+}$ population for further
823 analysis. Cytokine expression (ex: IFNγ) were examined in all 3 T cell sub-populations.
824 (C) Cytokine expression was also examined within the $CD3^{-}$ population.
825
826 Figure S4. Multivariate model accuracies
827 (A) The accuracy of our protection model (Figure 6) was compared to that of two different types
828 of null model, where our measure of accuracy was mean squared error of prediction of Δlog (lung
829 CFU). Our PLSR model performed significantly better than both a null model consisting of random
830 features instead of our elastic-net chosen features and a null model where the outcomes of Δlog
831 (lung CFU) were scrambled. Our PLSR model and each of the null models were run 50 times,
832 and each dot represents one of these trials. Labeled p-values are from one-way ANOVAs.
833 (B) Our models highlighting the immune features that distinguish pairs of protection/susceptibility
834 categories (Figure 7) were compared to two null models as in (A). Each of our models, both the
835 ones involving all mice and the protection models were significantly better than both null models.
836 Here, the accuracy metric is classification accuracy. Labeled p-values are from one-way
837 ANOVAs.
838

839 Figure S5. Examining the Polyfunctionality of CD4 T cells in select CC strains. Boolean gating
840 was used to identify various sub-populations amongst the CD4 T cells following P300 stimulation
841 in certain CC mice. An increase in various IFNγ^{+}IL-17^{+} CD4 T cells (that also produced TNF or
842 IL-2) were detected in CC037, CC072, CC031 and CC001 but not in C57BL/6. In contrast, a
843 noticeable decrease in the frequency of IFNγ "single positive" CD4 T cells was observed in all
844 strains including C57BL/6. The data in the graphs represent the mean ± SD of two experiment.
845 One-way analysis of variance with Sidak's multiple comparison test was used to determine
846 significance. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.
847
848

References

1. WHO. Global Tuberculosis Report 2021. 2021.

2. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA.* 1994;271(9):698-702.

3. Roy A, Eisenhut M, Harris RJ, Rodrigues LC, Sridhar S, Habermann S, et al. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ.* 2014;349:g4643.

4. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PE, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis.* 2014;58(4):470-80.

5. Abubakar I, Pimpin L, Ariti C, Beynon R, Mangtani P, Sterne JA, 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.* 2013;17(37):1-372, v-vi.

6. Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet.* 1995;346(8986):1339-45.

7. Flynn JL, Gideon HP, Mattila JT, and Lin PL. Immunology studies in non-human primate models of tuberculosis. *Immunol Rev.* 2015;264(1):60-73.

8. Ahmed M, Thirunavukkarasu S, Rosa BA, Thomas KA, Das S, Rangel-Moreno J, et al. Immune correlates of tuberculosis disease and risk translate across species. *Sci Transl Med.* 2020;12(528).

9. Aagaard C, Knudsen NPH, Sohn I, Izzo AA, Kim H, Kristiansen EH, 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.* 2020;205(8):2146-55.

10. Woodworth JS, Clemmensen HS, Battey H, Dijkman K, Lindenstrom T, Laureano RS, et al. A Mycobacterium tuberculosis-specific subunit vaccine that provides synergistic immunity upon co-administration with Bacillus Calmette-Guerin. *Nat Commun.* 2021;12(1):6658.

11. Kramnik I, and Beamer G. Mouse models of human TB pathology: roles in the analysis of necrosis and the development of host-directed therapies. *Semin Immunopathol.* 2016;38(2):221-37.

12. Nemeth J, Olson GS, Rothchild AC, Jahn AN, Mai D, Duffy FJ, et al. Contained Mycobacterium tuberculosis infection induces concomitant and heterologous protection. *PLoS Pathog.* 2020;16(7):e1008655.

13. Noll KE, Ferris MT, and Heise MT. The Collaborative Cross: A Systems Genetics Resource for Studying Host-Pathogen Interactions. *Cell Host Microbe.* 2019;25(4):484-98.

888 14. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative
889 Cross, a community resource for the genetic analysis of complex traits. *Nat Genet.*
890 2004;36(11):1133-7.

891 15. Smith CM, Proulx MK, Olive AJ, Laddy D, Mishra BB, Moss C, et al. Tuberculosis
892 Susceptibility and Vaccine Protection Are Independently Controlled by Host Genotype.
893 *MBio.* 2016;7(5).

894 16. Smith CM, Baker RE, Proulx MK, Mishra BB, Long JE, Park SW, et al. Host-pathogen
895 genetic interactions underlie tuberculosis susceptibility in genetically diverse mice. *Elife.*
896 2022;11.

897 17. Lee J, Boyce S, Powers J, Baer C, Sassetti CM, and Behar SM. CD11cHi monocyte-
898 derived macrophages are a major cellular compartment infected by Mycobacterium
899 tuberculosis. *PLoS Pathog.* 2020;16(6):e1008621.

900 18. Nandi B, and Behar SM. Regulation of neutrophils by interferon-gamma limits lung
901 inflammation during tuberculosis infection. *J Exp Med.* 2011;208(11):2251-62.

902 19. Sahiratmadja E, Alisjahbana B, Buccheri S, Di Liberto D, de Boer T, Adnan I, et al. Plasma
903 granulysin levels and cellular interferon-gamma production correlate with curative host
904 responses in tuberculosis, while plasma interferon-gamma levels correlate with
905 tuberculosis disease activity in adults. *Tuberculosis (Edinb).* 2007;87(4):312-21.

906 20. Sakai S, Kauffman KD, Schenkel JM, McBerry CC, Mayer-Barber KD, Masopust D, et al.
907 Cutting edge: control of Mycobacterium tuberculosis infection by a subset of lung
908 parenchyma-homing CD4 T cells. *J Immunol.* 2014;192(7):2965-9.

909 21. Becattini S, Latorre D, Mele F, Foglierini M, De Gregorio C, Cassotta A, et al. Functional
910 heterogeneity of human memory CD4(+) T cell clones primed by pathogens or vaccines.
911 *Science.* 2015;347(6220):400-6.

912 22. Smith CM, Proulx MK, Lai R, Kiritsy MC, Bell TA, Hock P, et al. Functionally Overlapping
913 Variants Control Tuberculosis Susceptibility in Collaborative Cross Mice. *mBio.*
914 2019;10(6).

915 23. Lindestam Arlehamn CS, McKinney DM, Carpenter C, Paul S, Rozot V, Makgotlho E, et
916 al. A Quantitative Analysis of Complexity of Human Pathogen-Specific CD4 T Cell
917 Responses in Healthy M. tuberculosis Infected South Africans. *PLoS Pathog.*
918 2016;12(7):e1005760.

919 24. Patankar YR, Sutiwisesak R, Boyce S, Lai R, Lindestam Arlehamn CS, Sette A, et al.
920 Limited recognition of Mycobacterium tuberculosis-infected macrophages by polyclonal
921 CD4 and CD8 T cells from the lungs of infected mice. *Mucosal Immunol.* 2020;13(1):140-
922 8.

923 25. Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA,
924 et al. Memory T cells in latent Mycobacterium tuberculosis infection are directed against
925 three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. *PLoS*
926 *Pathog.* 2013;9(1):e1003130.

927 26. Acosta-Rodriguez EV, Rivino L, Geginat J, Jarrossay D, Gattorno M, Lanzavecchia A, et
928 al. Surface phenotype and antigenic specificity of human interleukin 17-producing T helper
929 memory cells. *Nat Immunol.* 2007;8(6):639-46.

930 27. Gideon HP, Hughes TK, Tzouanas CN, Wadsworth MH, 2nd, Tu AA, Gierahn TM, et al.
931 Multimodal profiling of lung granulomas in macaques reveals cellular correlates of
932 tuberculosis control. *Immunity.* 2022;55(5):827-46 e10.

933 28. Roederer M, Nozzi JL, and Nason MC. SPICE: exploration and analysis of post-cytometric
934 complex multivariate datasets. *Cytometry A.* 2011;79(2):167-74.

935

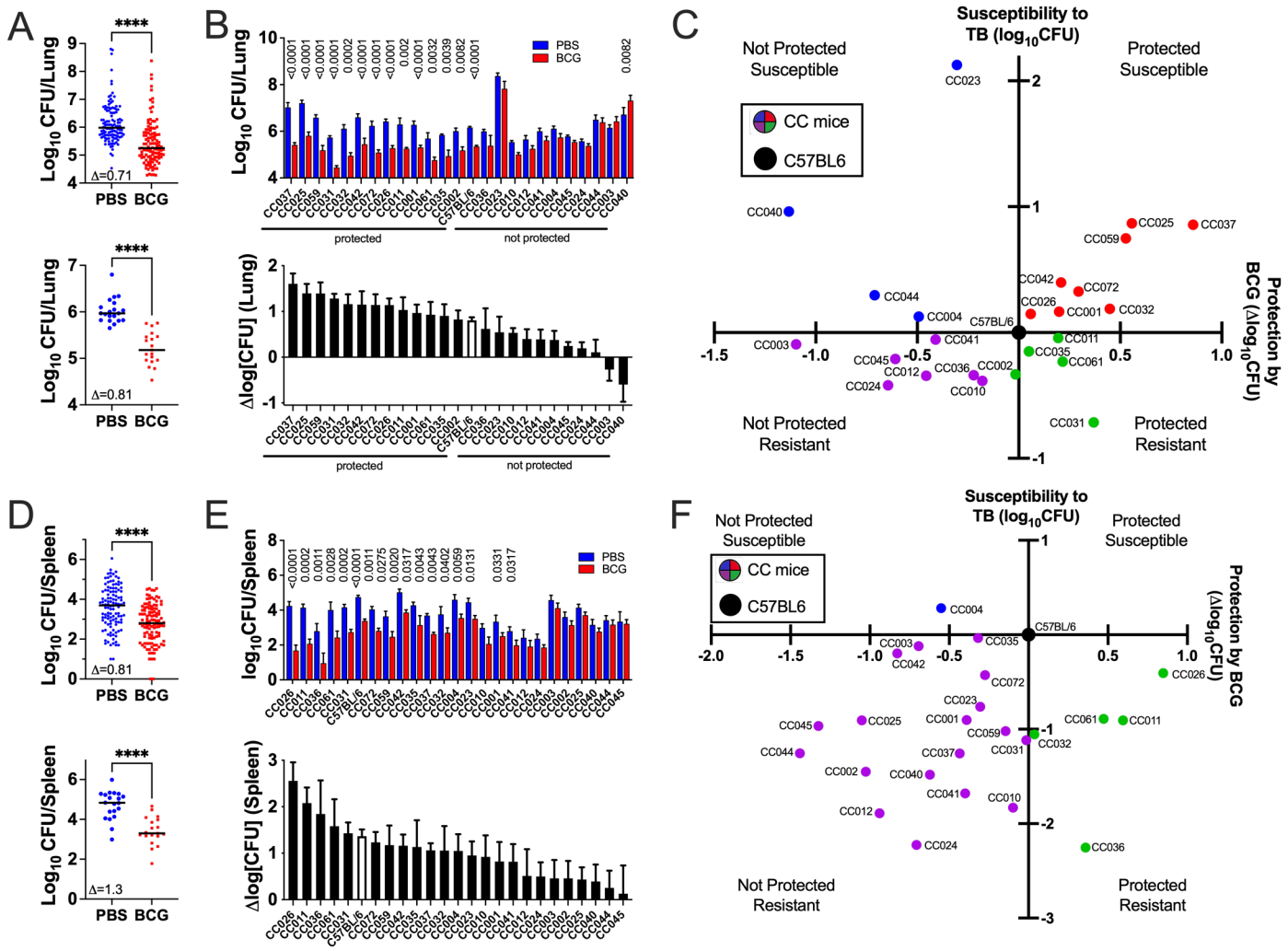

Figure 1. BCG-induced protection in 24 CC strains and C57BL/6 mice.
(A) Lung and (D) spleen CFU of all 24 CC mice (n=120/group; top) or all C57BL/6 mice (n=20/group; bottom) at 4 weeks post infection. Mice were vaccinated with BCG or administered PBS and rested for 12 weeks prior to infection. Each point 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

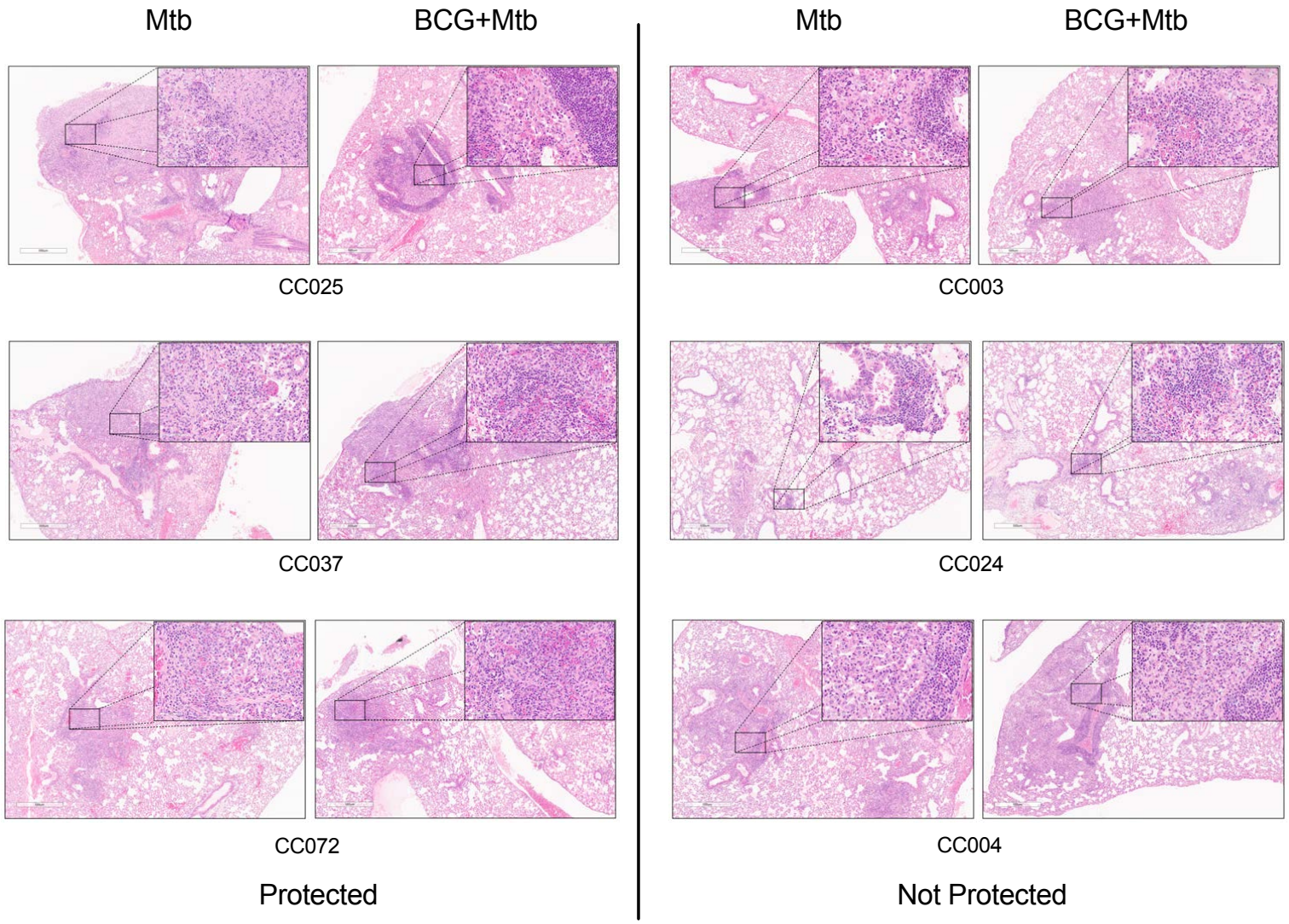

Figure 2. Lung histology of CC mice after BCG vaccination and Mtb challenge
Representative lung histology of three CC strains (CC025, CC037, CC072) that were protected by BCG, after sham or BCG vaccination and Mtb challenge (Mtb and BCG+Mtb, respectively) (left). Representative lung histology of three 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

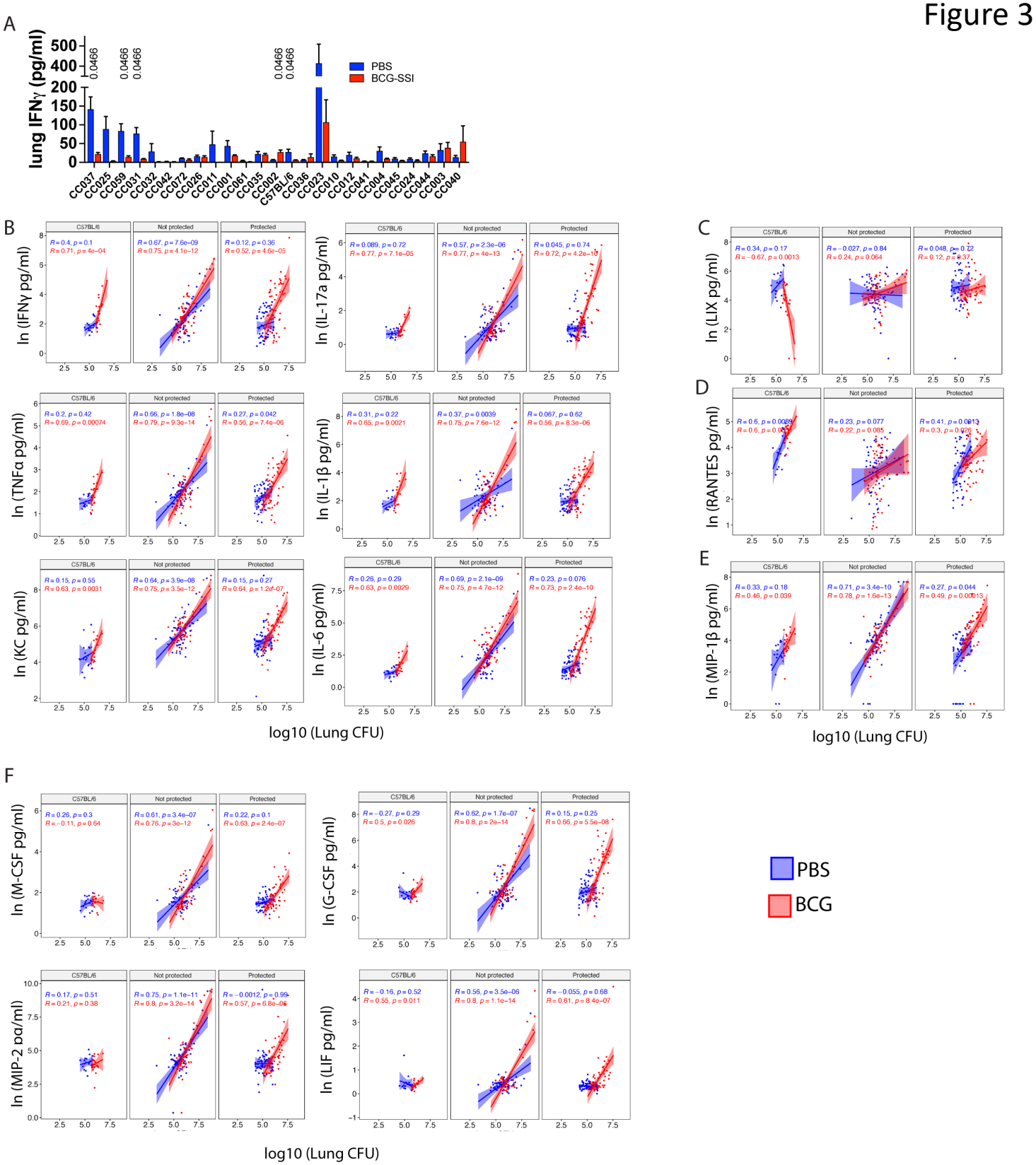

Figure 3. BCG-protected CC mice have an altered lung microenvironment of CC mice after Mtb infection.

(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

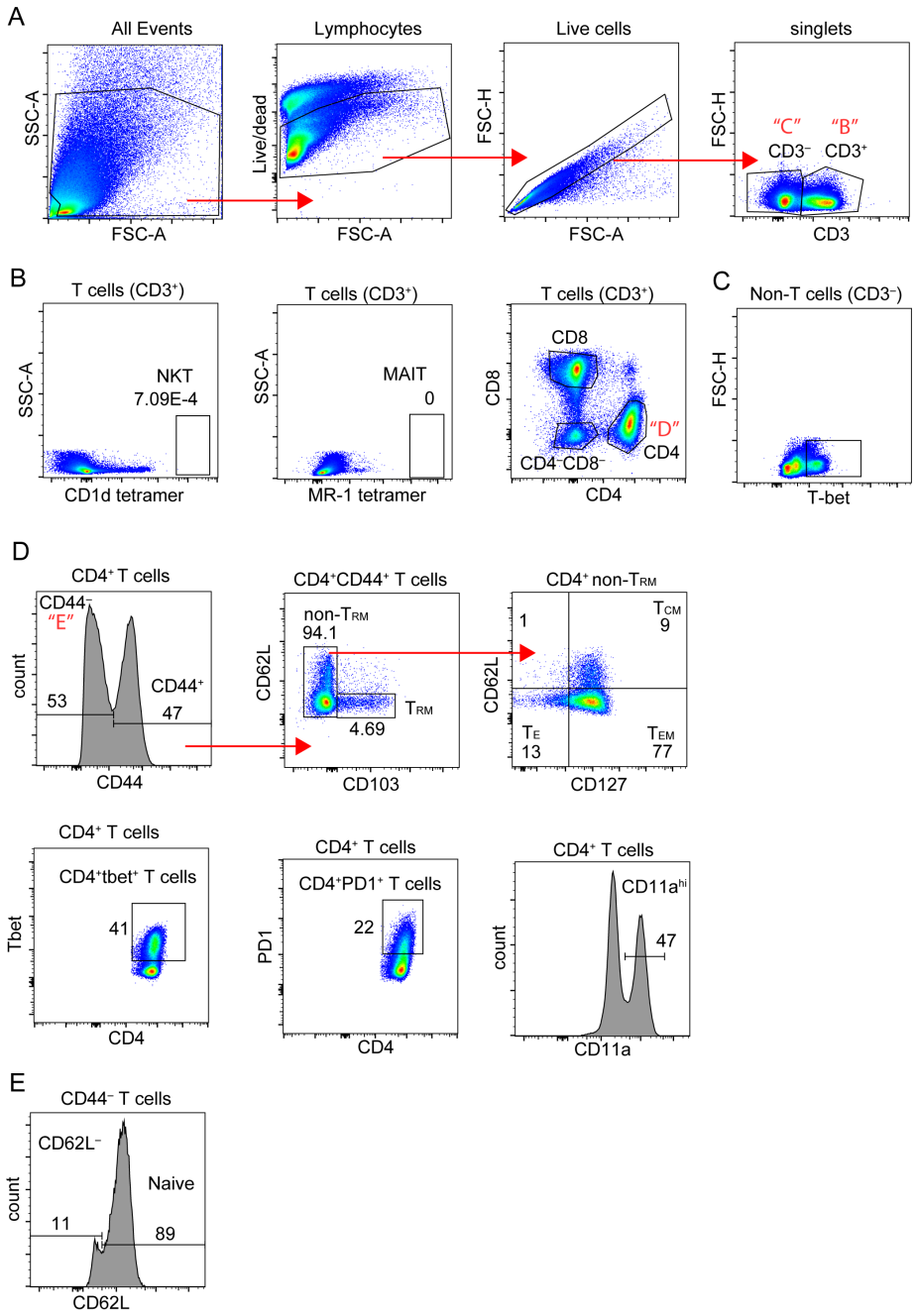

Figure S1. Gating strategy for the analysis of the cell surface T cell phenotype.
(A) Lymphocytes were identified using FSC and SSC properties, after which viability dyes were used to identify live cells. Singlet gates were used to exclude doublets, after which T cells and non-T cells were identified 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

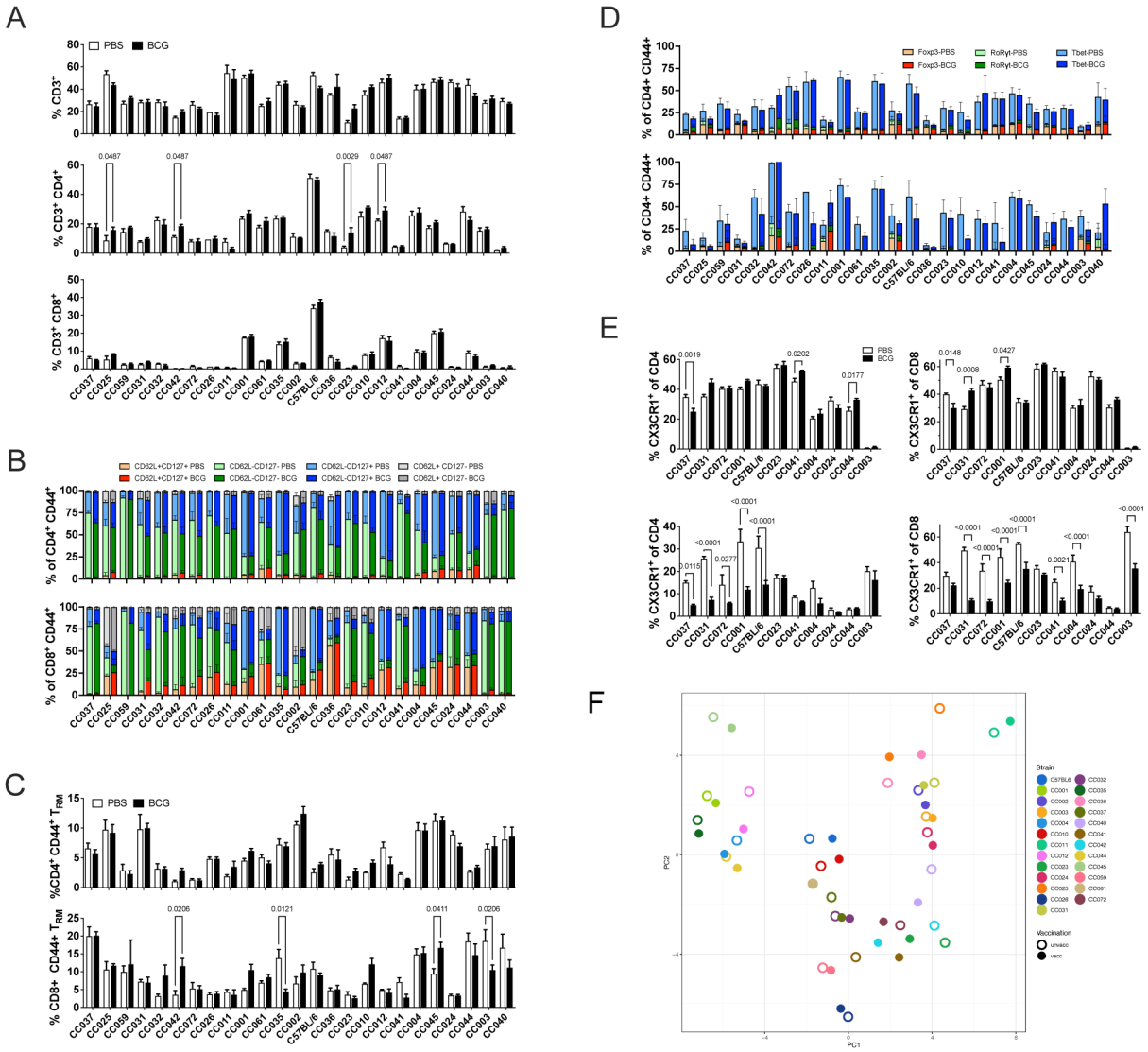

Figure 4. Frequency of various T cell populations in protected and unprotected CC mice following Mtb infection.

(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

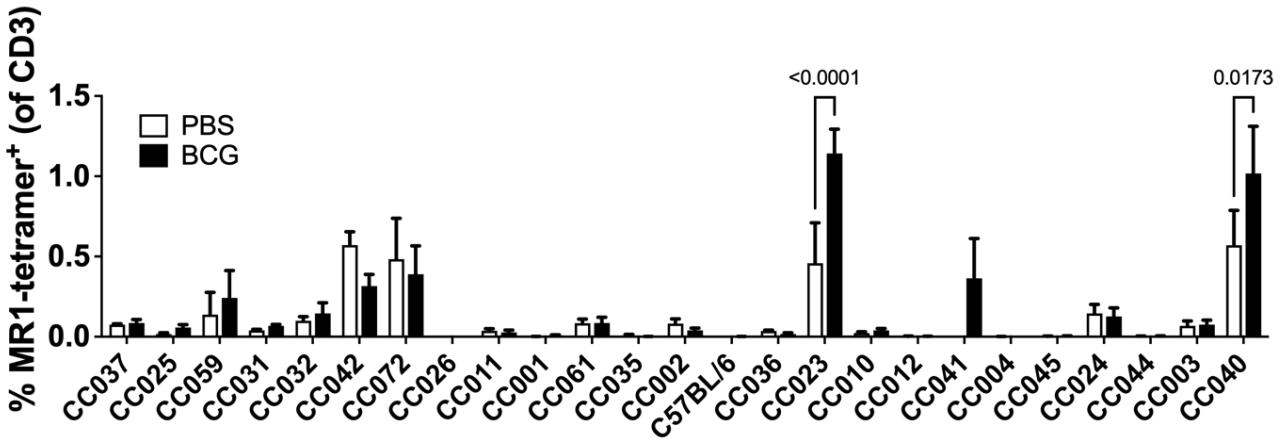

Figure S2. Frequencies of non-conventional T cells in CC mice. Frequencies of MR-1-restricted MAIT cells were measured in the lungs of PBS or BCG vaccinated CC mice at 4 weeks post Mtb infection. The data in the graphs represent the mean ± SD of one experiment. Two-way analysis of variance with the Benjamini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

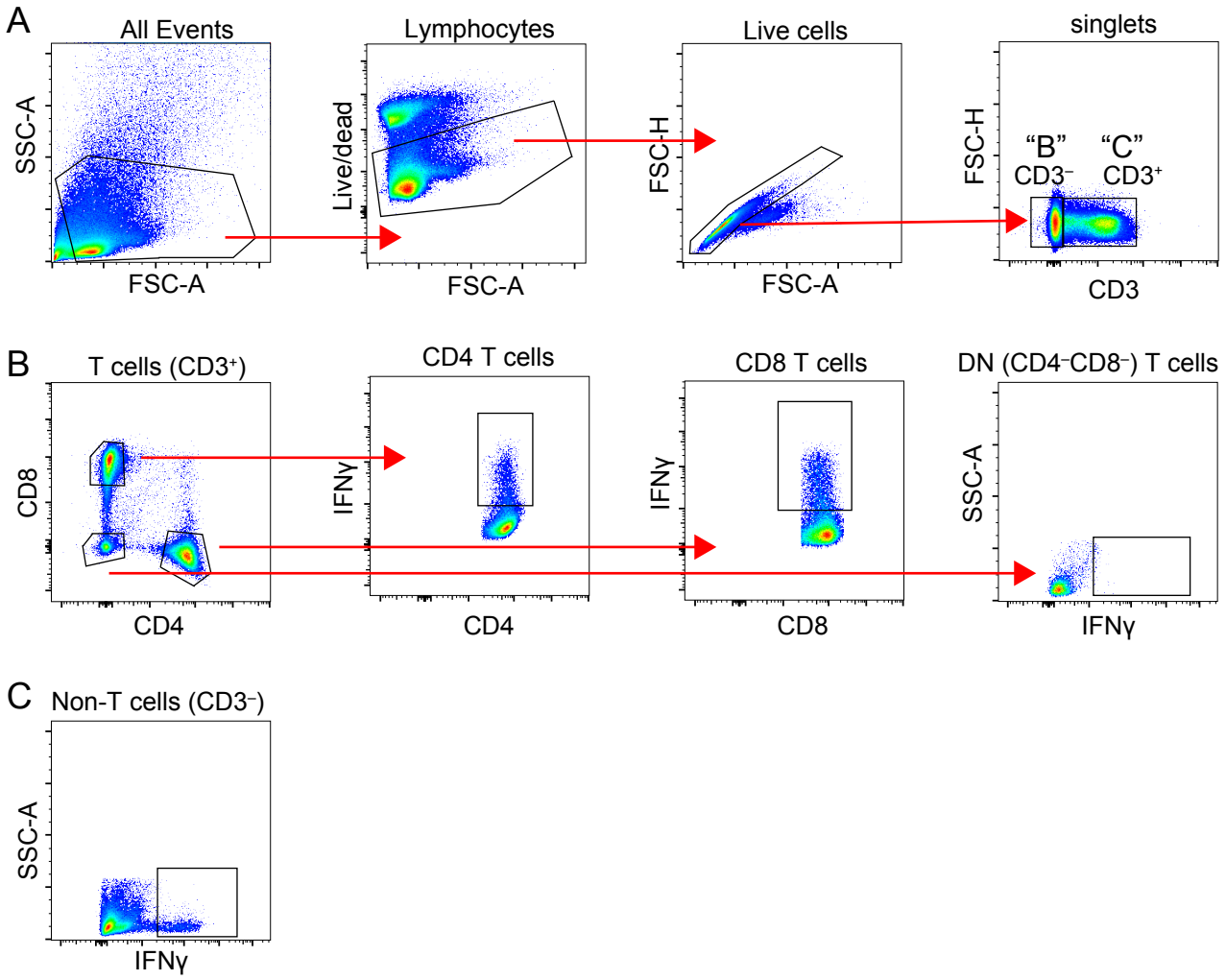

Figure S3. Gating strategy for the analysis of T cell function (intracellular cytokine staining).

(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

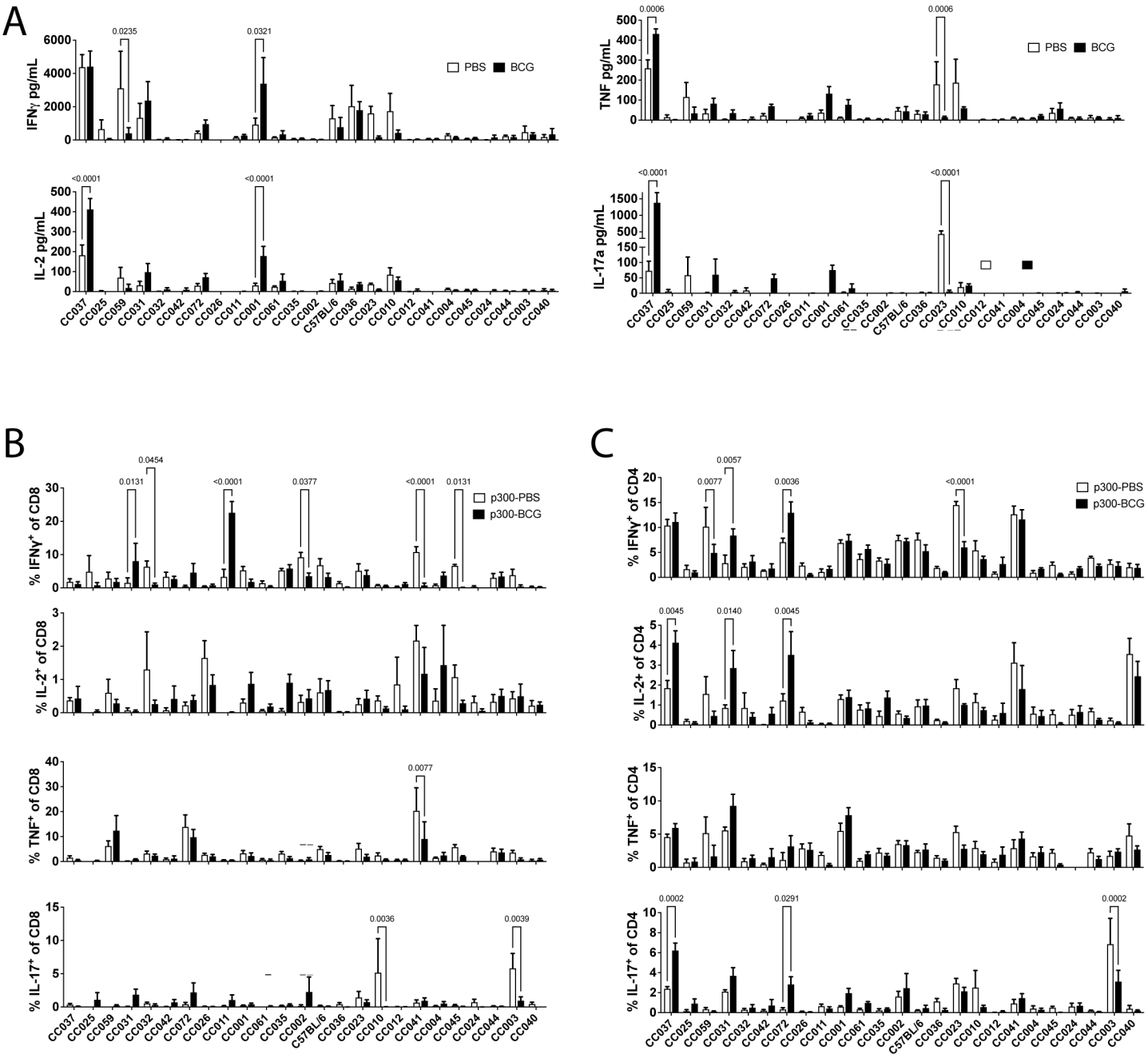

Figure 5. Cytokine response in protected and unprotected CC strains following Mtb infection. (A) Cells were isolated from the lungs of CC mice at 4 weeks post infection and stimulated for 24 hours with the P300 peptide pool. IFNγ, IL-2, TNF and IL-17a were measured in the supernatants. (B) and (C) Cells 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. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

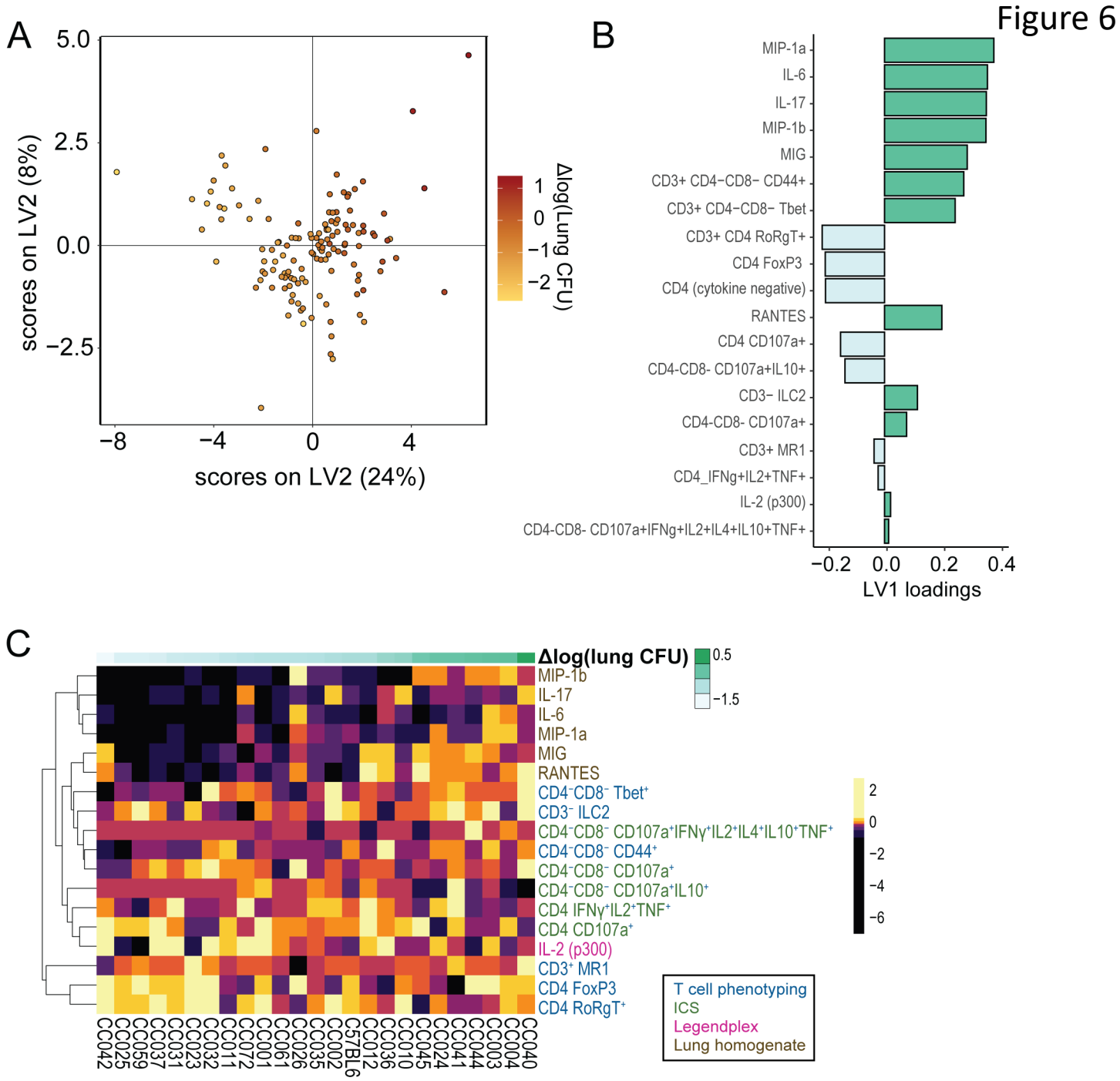

Figure 6. Multivariate correlates of protection across all 24 CC strains
(A) A partial least squares regression (PLSR) model was performed to associate BCG-induced changes in lung immune features with BCG-induced changes in lung CFU. Each data point represents a vaccinated mouse that has the mean of 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.

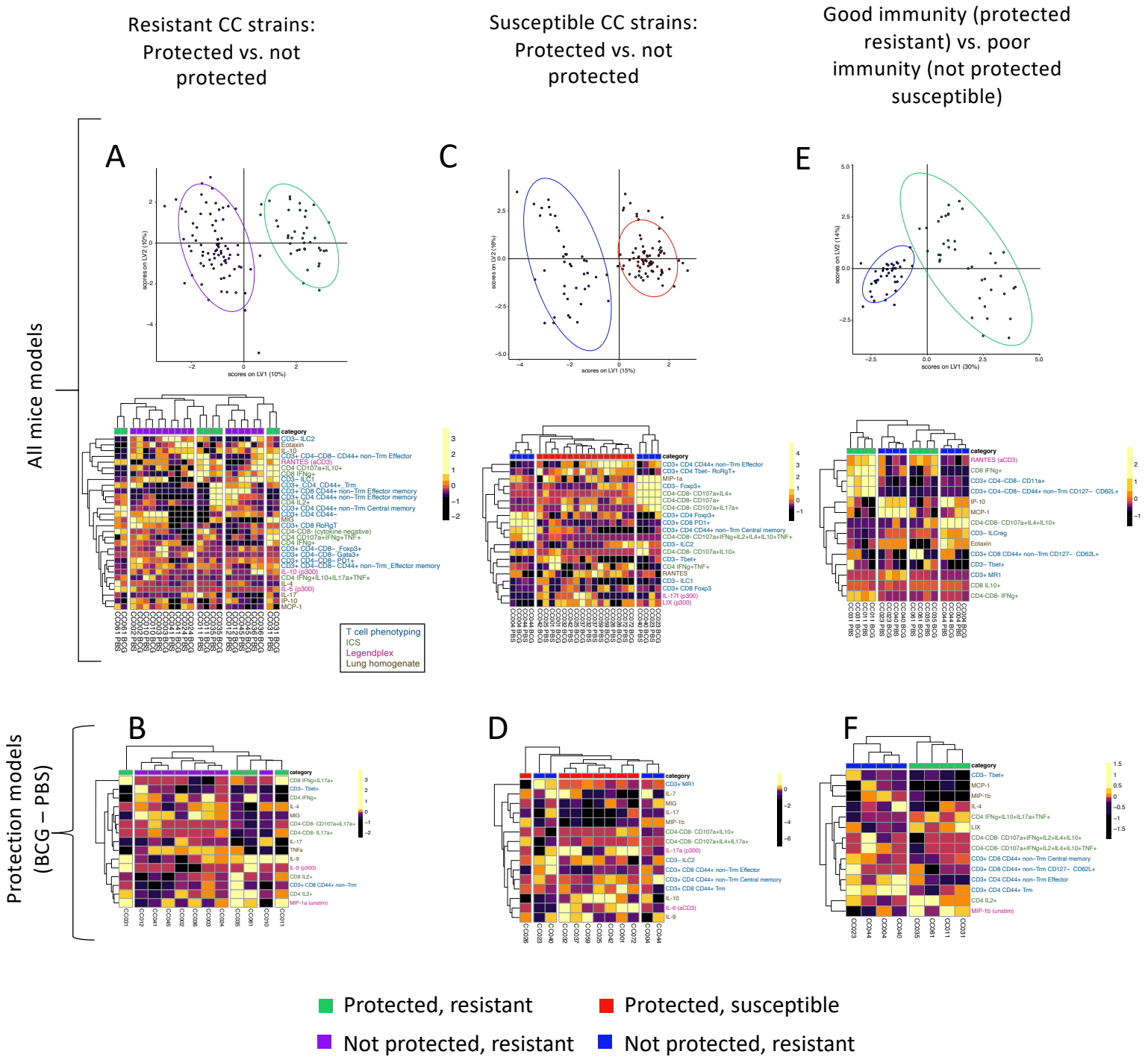

Figure 7. Identifying signatures of protection in subclasses of CC strains
(A), (C), (E) Scores plot (top) and heatmaps (bottom) of selected features for PLSDA models. Each dot represents a single mouse. The two colors in the scores plots signify the two classes being compared. An ellipse is drawn around 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

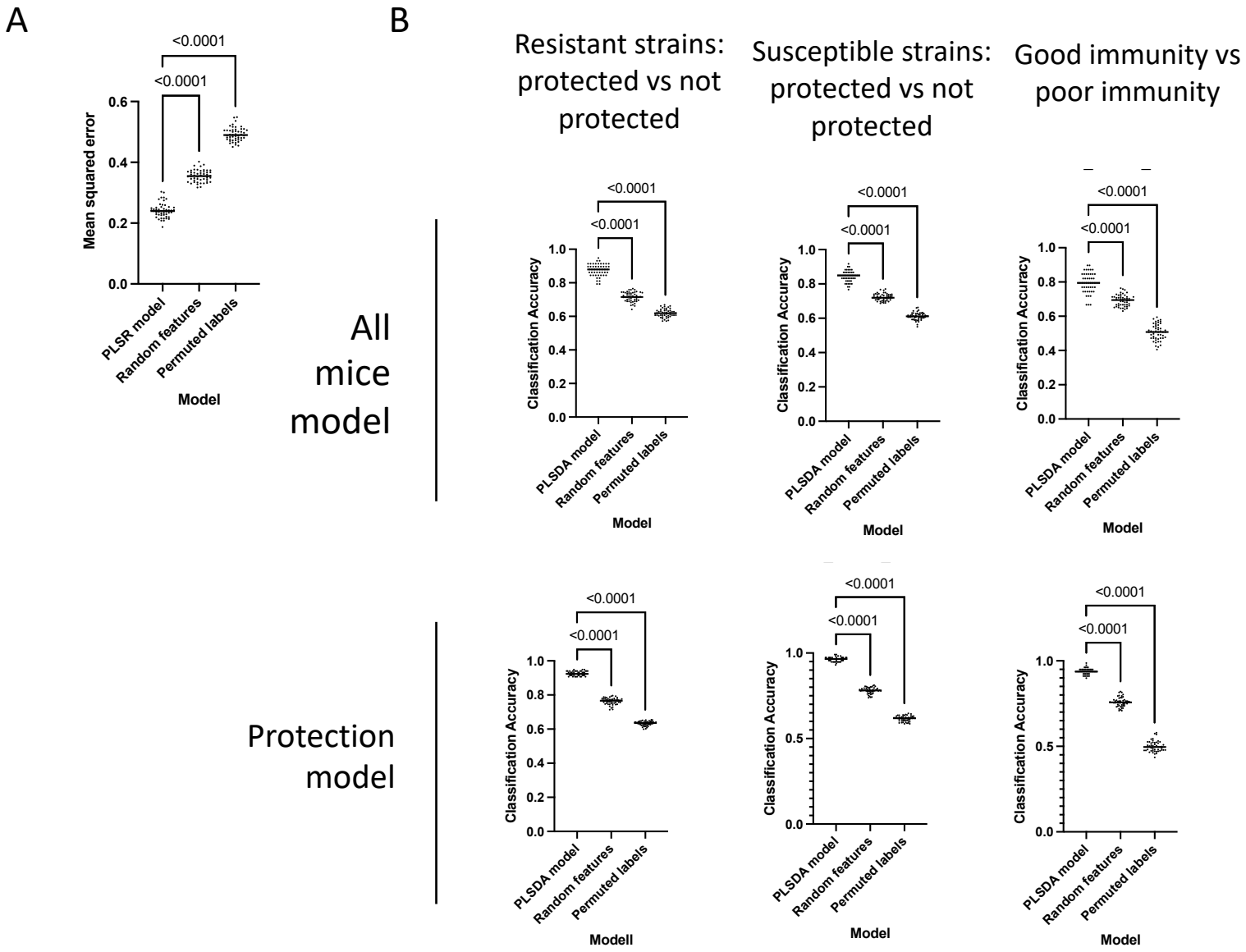

Figure S4. Multivariate model accuracies

(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.

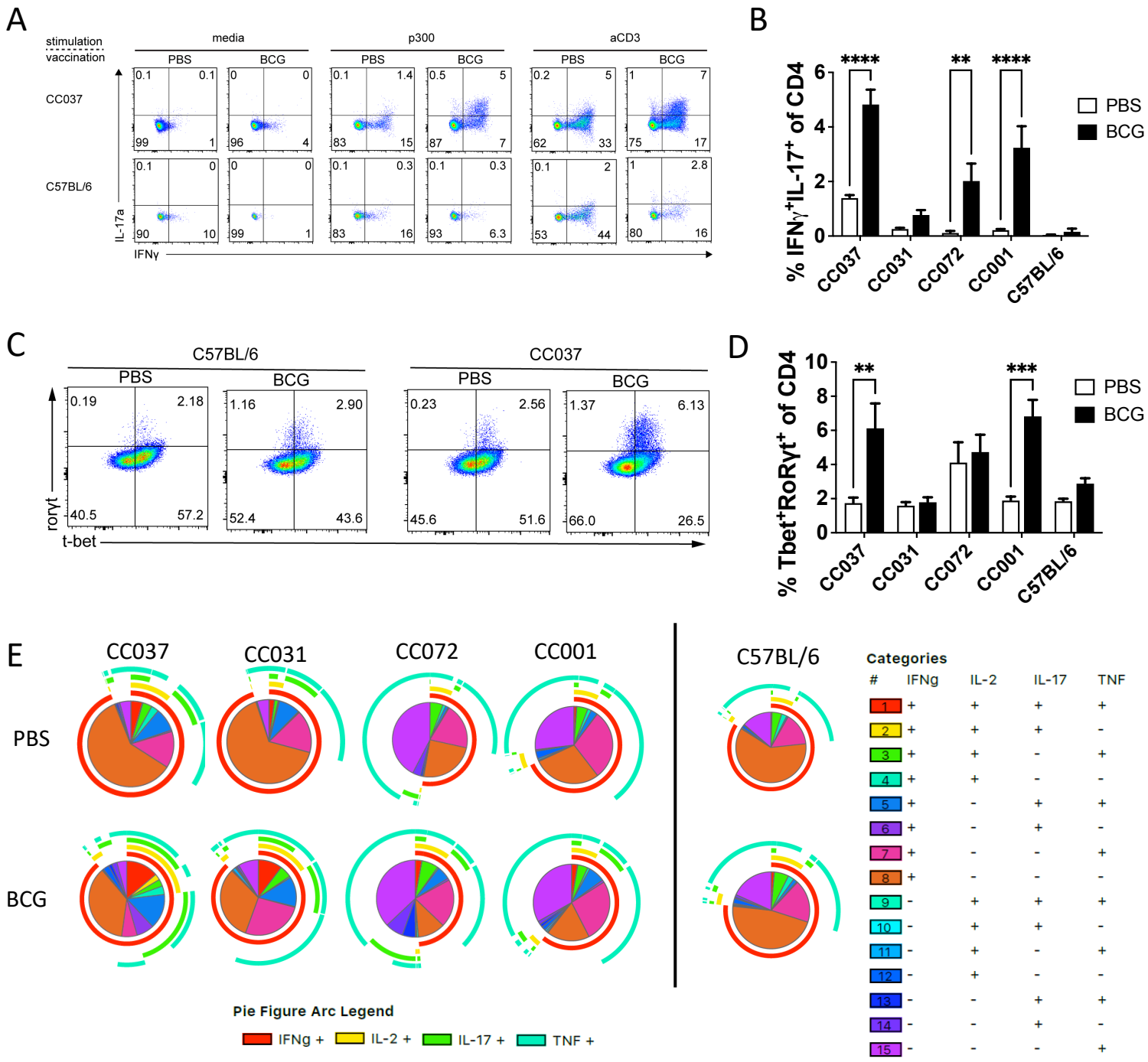
A
stimulation
vaccination
media
p300
aCD3
PBS
BCG
CC037
C57BL/6
IL-17a
IFNγ
B
% IFNγ+IL-17+ of CD4
PBS
BCG
CC037
CC031
CC072
CC001
C57BL/6
C
C57BL/6
CC037
roryt
t-bet
D
% Tbet+RoRyt+ of CD4
E
CC037
CC031
CC072
CC001
C57BL/6
PBS
BCG
Categories
IFNg IL-2 IL-17 TNF
Pie Figure Arc Legend
IFNg +
IL-2 +
IL-17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

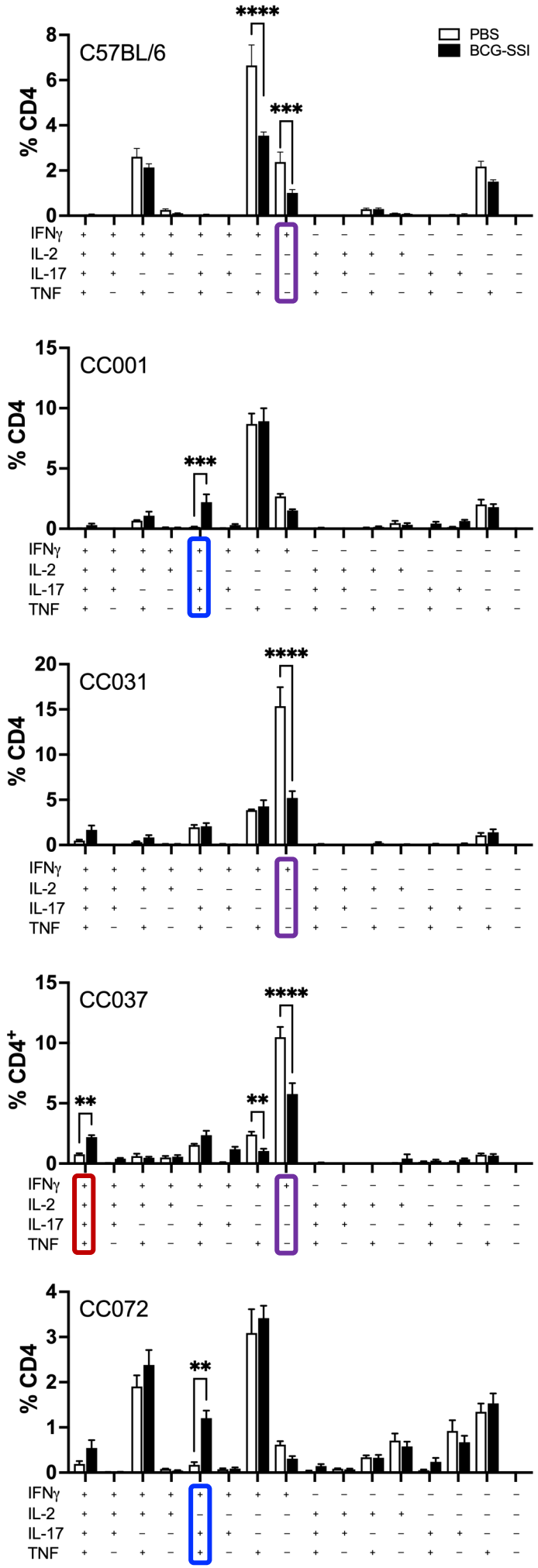

Figure S5. Examining the Polyfunctionality of CD4 T cells in select CC strains. Boolean gating was used to identify various sub-populations amongst the CD4 T cells following P300 stimulation in certain CC mice. An increase in various IFNγ$^+$IL-17$^+$ CD4 T cells (that also produced TNF or IL-2) were 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. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.