Host-pathogen genetic interactions underlie tuberculosis susceptibility

Short title: Collaborative Cross and Tuberculosis

Authors:

Clare M. Smith[1,2*], Richard E. Baker[1], Megan K. Proulx[1], Bibhuti B. Mishra[1,3], Jarukit E. Long[1], Sae Woong Park[4], Ha-Na Lee[4], Michael C. Kiritsy[1], Michelle M. Bellerose[1], Andrew J. Olive[1], Kenan C. Murphy[1], Kadamba Papavinasasundaram[1], Frederick J. Boehm[1], Charlotte J. Reames[1], Rachel K. Meade[2,5], Brea K. Hampton[6,7], Colton L. Linnertz[6], Ginger D. Shaw[6,8], Pablo Hock[6], Timothy A. Bell[6], Sabine Ehrt[4], Dirk Schnappinger[4], Fernando Pardo-Manuel de Villena[6,8], Martin T. Ferris[6], Thomas R. Ioerger[9], and Christopher M. Sassetti[1,10*]

Affiliations:

[1]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA

[2]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC

[3]Department of Immunology and Microbial Disease, Albany Medical College, Albany, NY

[4]Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY

[5]University Program in Genetics and Genomics, Duke University, Durham, NC

[6]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC

[7]Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC

[8]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

[9]Department of Computer Science and Engineering, Texas A&M University, College

[10]Lead contact

*Correspondence: C.M. Sassetti (Christopher.Sassetti@umassmed.edu) and C.M. Smith (clare.m.smith@duke.edu)

Abstract:

The outcome of an encounter with *Mycobacterium tuberculosis* depends on the pathogen's ability to adapt to the variable immune pressures exerted by the host. Understanding this interplay has proven difficult, largely because experimentally tractable animal models do not recapitulate the heterogeneity of tuberculosis disease. We leveraged the genetically diverse Collaborative Cross (CC) mouse panel in conjunction with a library of *Mtb* mutants to associate bacterial genetic requirements with host genetics and immunity. We report that CC strains vary dramatically in their susceptibility to infection and produce qualitatively distinct immune states. Global analysis of *Mtb* mutant fitness across the CC panel revealed that many virulence pathways are only in specific host microenvironments, identifying the large fraction of the pathogen's genome that has been maintained to ensure fitness in a diverse

population. Both immunological and bacterial traits were associated with genetic variants distributed across the mouse genome, identifying the specific host-pathogen genetic interactions that influence pathogenesis.

Keywords: Host-pathogen interactions, tuberculosis, *Mycobacterium*, systems genetics, genomics, genetic diversity, natural variation, bacterial genetics, TnSeq, mouse models, Collaborative Cross, QTL mapping, complex traits.

Introduction

Infection with *Mycobacterium tuberculosis* (*Mtb*) produces heterogeneous outcomes that are influenced by genetic and phenotypic variation in both the host and the pathogen. Classic human genetic studies show that host variation influences immunity to tuberculosis (TB) (Abel et al., 2018; Comstock, 1978). Likewise, the co-evolution of *Mtb* with different populations across the globe has produced genetically distinct lineages that demonstrate variable virulence traits (Gagneux et al., 2006). The role of genetic variation on each side of this interaction is established, yet the intimate evolutionary history of both genomes suggests that interactions between host and pathogen variants may represent an additional determinant of outcome (McHenry et al., 2020). Evidence for genetic interactions between host and pathogen genomes have been identified in several infections (Ansari et al., 2017; Berthenet et al., 2018), including TB (Caws et al., 2008; Holt et al., 2018; Thuong et al., 2016). However, the

70 combinatorial complexity involved in identifying these relationships in natural
71 populations have left the mechanisms largely unclear.
72
73 Mouse models have proven to be a powerful tool to understand mechanisms of
74 susceptibility to TB. Host requirements for protective immunity were discovered by
75 engineering mutations in the genome of standard laboratory strains of mice, such as
76 C57BL/6 (B6), revealing a critical role of Th1 immunity. Mice lacking factors necessary
77 for the production of Th1 cells or the protective cytokine interferon gamma (IFNγ) are
78 profoundly susceptible to *Mtb* infection (Caruso et al., 1999; Cooper et al., 1993; Cooper
79 et al., 1997; Flynn et al., 1993; Saunders et al., 2002). Defects in this same immune
80 axis cause the human syndrome Mendelian Susceptibility to Mycobacterial Disease
81 (MSMD) (Altare et al., 1998; Bogunovic et al., 2012; Bustamante et al., 2014; Filipe-
82 Santos et al., 2006), demonstrating the value of knockout (KO) mice to characterize
83 genetic variants of large effect. Similarly, the standard mouse model has been used to
84 define *Mtb* genes that are specifically required for optimal bacterial fitness during
85 infection (Bellerose et al., 2020; Sassetti and Rubin, 2003; Zhang et al., 2013).
86
87 Despite the utility of standard mouse models, it has become increasingly clear that the
88 immune response to *Mtb* in genetically diverse populations is more heterogeneous than
89 any single small animal model (Smith and Sassetti, 2018). For example, while IFNγ-
90 producing T cells are critical for protective immunity in standard inbred lines of mice, a
91 significant fraction of humans exposed to *Mtb* control the infection without producing a
92 durable IFNγ response (Lu et al., 2019). Similarly, IL-17 producing T cells have been

implicated in both protective responses and inflammatory tissue damage in TB, but IL-17 has little effect on disease progression in B6 mice, except in the context of vaccination or infection with particularly virulent *Mtb* (Gopal et al., 2012; Khader et al., 2007). The immunological homogeneity of standard mouse models may also explain why only a small minority of the >4000 genes that have been retained in the genome of *Mtb* during its natural history promote fitness in the mouse (Bellerose et al., 2020). Thus, homogenous mouse models of TB fail to capture the distinct disease states, mechanisms of protective immunity, and selective pressures on the bacterium that are observed in natural populations.

The Collaborative Cross (CC) and Diversity Outbred (DO) mouse populations are new mammalian resources that more accurately represent the genetic and phenotypic heterogeneity observed in outbred populations (Churchill et al., 2004; Churchill et al., 2012). These mouse panels are both derived from the same eight diverse founder strains but have distinct population structures (Saul et al., 2019). DO mice are maintained as an outbred population and each animal represents a unique and largely heterozygous genome (Keller et al., 2018; Svenson et al., 2012). In contrast, each inbred CC strain's genome is almost entirely homozygous, producing a genetically stable and reproducible population in which the phenotypic effect of recessive mutations is maximized (Shorter et al., 2019; Srivastava et al., 2017). Together, these resources have been leveraged to identify host loci underlying the immune response to infectious diseases (Noll et al., 2019). In the context of TB, DO mice have been used as individual, unique hosts to identify correlates of disease, which resemble those observed in non-

116 human primates and humans (Ahmed et al., 2020; Gopal et al., 2013; Niazi et al.,
117 2015). Small panels of the reproducible CC strains have been leveraged to identify host
118 background as a determinant of the protective efficacy of BCG vaccination (Smith et al.,
119 2016) and a specific variant underlying protective immunity to tuberculosis (Smith et al.,
120 2019). While these studies demonstrate the tractability of the DO and CC populations to
121 model the influence of host diversity on infection, dissecting host-pathogen interactions
122 requires the integration of pathogen genetic diversity.
123
124 We combined the natural but reproducible host variation of the CC panel with a
125 comprehensive library of *Mtb* mutants to characterize the interactions between host and
126 pathogen. Using over 60 diverse mouse strains, we report that the CC panel
127 encompasses a broad spectrum of TB susceptibility and immune phenotypes, including
128 outlier lines that model non-canonical immune states. Through "Transposon
129 Sequencing" (TnSeq), we quantified the relative fitness of *Mtb* mutants across the CC
130 panel and specific immunological knockout strains. We report that approximately three
131 times more bacterial genes contribute to fitness in the diverse panel than in any single
132 mouse strain, defining a large fraction of the bacterial genome that is dedicated to
133 adapting to distinct immune states. Association of these immunological and bacterial
134 fitness traits with Quantitative Trait Loci (QTL) demonstrated the presence of discrete
135 Host-Interacting-with Pathogen QTL (*Hip*QTL) that represent inter-species genetic
136 interactions that influence the pathogenesis of this infection.
137
138 **Results:**

139 **The spectrum of TB disease traits in the CC exceeds that observed in standard**
140 **inbred mice**
141 To characterize the diversity of disease states that are possible in a genetically diverse
142 mouse population, we infected a panel of 52 CC lines and the 8 founder strains with
143 *Mtb*. To enable TnSeq studies, the animals were infected via the intravenous route with
144 a saturated library of *Mtb* transposon mutants, which in sum produce an infection that is
145 similar to the wild-type parental strain (Bellerose et al., 2020; Sassetti and Rubin, 2003).
146 Groups of 2-6 (average of n=3) male mice per genotype were infected, and the bacterial
147 burden after four weeks of infection was assessed by plating (colony forming units,
148 CFU) and quantifying the number of bacterial chromosomes in the tissue (chromosome
149 equivalents, CEQ). These two metrics were highly correlated (r=0.88) and revealed a
150 wide variation in bacterial burden across the panel (**Figure 1A** and **S1**). The
151 susceptibility of the inbred founder strains was largely consistent with previous studies
152 employing an aerosol infection (Smith et al., 2016). As expected, B6 mice were
153 relatively resistant to infection, along with 129S1/SvImJ (129) and NOD/ShiLtJ (NOD)
154 strains. In contrast to the standard inbred lines, lung bacterial burden varied by more
155 than 1000-fold across the more diverse CC panel, ranging from animals that are
156 significantly more resistant than B6, to mice that harbored more than 10^9 bacteria in
157 their lungs (**Figure 1A**). Bacterial burden in the spleen also varied several thousand-fold
158 across the panel and was moderately correlated with lung burden (r=0.43) (**Table S1**
159 and **Figure S1**). Thus, the CC panel encompasses a much greater quantitative range of
160 susceptibility than standard inbred lines.
161

162 Comparing various measures of infection progression showed many expected
163 correlations but also an unexpected decoupling of some phenotypes. As an initial
164 assessment of the disease processes in these animals, we correlated bacterial burden
165 and lung cytokine abundance with measures of systemic disease such as weight loss
166 and sufficient morbidity to require euthanasia ("earliness of death"). In general,
167 correlations between these metrics indicated that systemic disease was associated with
168 bacterial replication and inflammation (**Figure 1B** and **Figure S1**). Lung CFU was
169 correlated with weight loss, mediators that enhance neutrophil differentiation or
170 migration (CXCL2 (MIP-2; r=0.79), CCL3 (MIP-1a; r=0.77), G-CSF (r=0.78) and CXCL1
171 (KC; r=0.76)), and more general proinflammatory cytokines (IL-6 (r=0.80) and IL-1α
172 (r=0.76)) (**Figure S1)**. These findings are consistent with previous work in the DO panel,
173 that found both proinflammatory chemokines and neutrophil accumulation to be
174 predictors of disease (Ahmed et al., 2020; Gopal et al., 2013; Niazi et al., 2015).
175
176 The reproducibility of CC genotypes allowed us to quantitatively assess the heritability
177 (h^2) of these immunological and disease traits. The percent of the variation attributed to
178 genotype ranged from 56-87% (mean=73.4%; **Table S2**). The dominant role of genetic
179 background in determining the observed phenotypic range allowed a rigorous
180 assessment of outlier phenotypes that is not possible in the DO population. For
181 example, despite the correlation between lung CFU and weight loss (r=0.57), several
182 strains failed to conform to this relationship (**Figure 1C**). In particular, CC030/GeniUnc
183 (p=0.003), CC040/TauUnc (p=0.027) and A/J (p=0.03) lost significantly more weight
184 than their bacterial burdens would predict (**Figure 1C**; outlier genotypes determined by

185 studentized residuals; noted by #). Similarly, CXCL1 abundance was significantly higher
186 in CC030/GeniUnc (p=0.001) and lower in CC056/GeniUnc (p=0.040), than the level
187 predicted by their respective bacterial burden (**Figure 1C**; outlier genotypes noted by †).
188 Thus, these related disease traits can be dissociated based on the genetic composition
189 of the host.
190
191 The cluster of cytokines that was most notably unrelated to bacterial burden included
192 IFNγ and the interferon-inducible chemokines CXCL10 (IP10), CXCL9 (MIG), and CCL5
193 (RANTES) (Red cluster in **Figure 1B**; **Figure S1**) (r<0.3). Despite the clear protective
194 role for IFNγ (Cooper et al., 1993; Flynn et al., 1993), high levels have been observed in
195 susceptible mice, likely as a result of high antigen load (Barber et al., 2011; Lázár-
196 Molnár et al., 2010). While high IFNγ levels in susceptible animals was therefore
197 expected, it was more surprising to find a number of genotypes that were able to control
198 bacterial replication yet had very low levels of this critically important cytokine (**Figure**
199 **1D**). This observation is likely due the inclusion of two founder lines, CAST/EiJ (CAST)
200 and PWK/PhJ (PWK) that display this unusual phenotype (Smith et al., 2016). To further
201 investigate, a separate cohort of B6 and CAST animals was infected by the aerosol
202 route, and the number of IFNγ-producing T cells in lung and spleen was compared by
203 ELISPOT assays. At 4 weeks post-infection, B6 and CAST harbored comparable
204 burdens of *Mtb* in lung and spleen (**Figure 1E** and **1F**), and the infection elicited similar
205 numbers of *Mtb*-specific IFNγ-producing cells in the spleen (**Figure 1H**). In contrast,
206 while IFNγ producing cells were found in the lungs of B6 mice, none were detectable in
207 CAST (**Figure 1G**). Thus, while CAST animals are capable of producing IFNγ-secreting

208 cells in response to *Mtb* infection, these cells do not appear to be involved in bacterial
209 control in the lung. As CC strains that share this immune profile also produced low
210 levels of IL-17 (**Figure 1D**), the mechanism(s) conferring protection in these animals
211 remain unclear. In sum, this survey of TB-related traits in the CC demonstrated a broad
212 range of susceptibility and the presence of qualitatively distinct and genetically
213 determined disease states.
214
215 ***Tip*QTL define genetic variants that control TB immunophenotypes**
216 Tuberculosis ImmunoPhenotype Quantitative Trait Loci (*Tip*QTL), which were
217 associated with TB disease or cytokine traits, were identified and numbered in
218 accordance with previously reported *Tip*QTL (Smith et al., 2019). Of the 32 TB-disease
219 traits, we identified 9 individual metrics that were associated with reasonable statistical
220 confidence to a chromosomal locus. Of these, three were associated with high
221 confidence ($p \leq 0.053$), and six other QTL met a suggestive threshold ($p < 0.2$; **Table 1**).
222 Several individual trait QTL occupied the same chromosomal locations. For example,
223 spleen CFU and spleen CEQ, which are both measures of bacterial burden and highly
224 correlated, were associated with the same interval on distal chromosome 2 (**Table 1**,
225 *Tip5*; **Figure 2A** and **2C**). IL-10 abundance was associated with two distinct QTL (**Table**
226 **1**). While IL-10 was relatively uncorrelated with spleen CFU ($r=0.48$), one of its QTL fell
227 within the *Tip5* bacterial burden interval on chromosome 2 (**Figure 2A** and **2C**). At this
228 QTL, the NOD haplotype was associated with high values for all three traits (**Figure**
229 **2E**). Similarly, the correlated traits, CXCL1 abundance and lung CFU, were individually
230 associated to the same region on chromosome 7 (**Table 1**, *Tip8*; **Figure 2B** and **2D**). In

231 this interval, the CAST haplotype was associated with both low bacterial burden and
232 CXCL1 (**Figure 2F**). At both *Tip5* and *Tip8*, we found no statistical evidence that the
233 positions of the associated QTL were different (*Tip5* p=0.55; *Tip8* p=0.27; 400 bootstrap
234 samples)(Boehm et al., 2019). These observations support the role of a single causal
235 variant at each locus that is responsible for a pleiotropic trait. Coincident mapping can
236 provide both additional statistical support for QTL (p values by Fisher's combined
237 probability test: Chr 7, p=0.067; Chr 2, p=0.041) and suggests potential mechanisms of
238 disease progression.
239
240 A number of factors can limit the statistical significance of QTL identified in the CC
241 population, including small effect sizes, limited genotype availability, and the genetic
242 complexity of the trait. To determine if the statistically suggestive QTL were likely to
243 reflect genuine biological effects, we took an intercross approach to validate the lung
244 CFU QTL on chromosomes 7 and 15 (*Tip8* and *Tip10*, **Table 1**). Given that the
245 associations at both QTL were driven by the CAST haplotype (**Figure 2F**), we
246 generated an F_2 population based on two CC strains, CC029/Unc and CC030/GeniUnc,
247 that contained CAST sequence at *Tip8* and *Tip10*, respectively (**Figure 3A**). 46 F_2 mice
248 that were homozygous at either locus were infected with *Mtb*, and lung CFU were
249 enumerated as per the larger CC screen (**Table S3**). Compared to F_2 mice that were not
250 CAST at either locus, mice that contained CAST at *Tip8*, *Tip10* or both loci had reduced
251 CFU burden; however, the strongest predictor of significantly lower bacterial burden
252 was CAST at *Tip8* (p=0.005, 2-factor regression; **Figure 3B**). This study provides strong

253 independent support for a CAST-driven QTL on chromosome 7 that controls lung CFU
254 and verifies the validity of the genetic mapping in this dataset.
255
256 **Table 1. Disease-related Tuberculosis ImmunoPhenotype QTL (*Tip*QTL).** Multiple
257 QTL within the same interval and clear allele effects are designated with the same
258 *Tip*QTL number. Column headings: QTL, quantitative trait loci; Chr, chromosome; LOD,
259 logarithm of the odds; CEQ, chromosomal equivalents.

QTL	Trait	Chr	LOD	P value	Interval start (Mb)	Peak (Mb)	Interval end (Mb)
Tip5	Spleen CEQ	2	9.14	2.38E-02	174.29	178.25	178.25
Tip5	Spleen CFU	2	7.04	2.19E-01	73.98	174.29	180.10
Tip6	IL-9	2	8.61	4.52E-02	33.43	41.4	41.48
Tip6	IL-9	2	7.85	1.26E-01	22.77	24.62	25.65
Tip7	IL-17	15	7.84	5.27E-02	67.98	74.14	82.11
Tip8	CXCL1	7	7.57	1.06E-01	30.43	45.22	46.72
Tip8	Lung CFU	7	7.47	1.17E-01	31.06	37.78	45.22
Tip9	IL-10	17	7.16	1.85E-01	80.98	82.47	83.55
Tip10	Lung CFU	15	7.13	1.86E-01	77.00	78.16	78.70

260
261 ***Mtb* adapts to diverse hosts by utilizing distinct gene repertoires**
262 This survey of disease-associated traits indicated that the CC panel encompasses a
263 number of qualitatively distinct immune phenotypes. To determine if different bacterial
264 functions were necessary to adapt to these conditions, we used TnSeq to estimate the

265 relative abundance of individual *Mtb* mutants after selection in each CC host genotype.
266 To serve as benchmarks of known immunological lesions, we also performed TnSeq in
267 B6 mice that were lacking the mediators of Th1 immunity, lymphocytes (*Rag2*$^{-/-}$) and
268 IFNγ (*Ifnγ*$^{-/-}$), or were lacking the immunoregulatory mediators that control disease by
269 inhibiting inflammation, nitric oxide synthase (*Nos2*$^{-/-}$) (Mishra et al., 2013) or the
270 NADPH phagocyte oxidase (*Cybb*$^{-/-}$) (Olive et al., 2018). The relative representation of
271 each *Mtb* mutant in the input pool versus the pools recovered from mouse spleens was
272 quantified (**Table S4**). Consistent with our previous work (Bellerose et al., 2020;
273 Sassetti and Rubin, 2003), we identified 234 *Mtb* genes that are required for growth or
274 survival in *Mtb* in B6 mice, based on significant underrepresentation of the
275 corresponding mutant after four weeks of *in vivo* selection. All but one of these genes
276 were found to be required in the larger panel, increasing confidence in this *Mtb* gene set
277 (**Figure 4A** and **4B**). While the number of genes found to be necessary in each
278 genotype across the panel was largely similar, the composition of these gene sets
279 varied considerably. As more CC strains, and presumably more distinct immune states,
280 were included in the analysis, the cumulative number of genes necessary for growth in
281 these animals also increased. This cumulative gene set plateaued at ~750, after the
282 inclusion of approximately 20-25 genotypes (**Figure 4A**). This number of genes far
283 outnumbered those identified in the panel of immunodeficient KO strains (**Figure 4B**
284 and **Table S4**).
285
286 To verify that our TnSeq study accurately predicted the effect of the corresponding loss-
287 of-function alleles, we assessed the phenotypes of selected bacterial deletion mutants

288 in a small set of mouse genotypes that were predicted to result in differential selection.
289 Individual *Mtb* mutants lacking genes necessary for ESX-1 type VII secretion (*eccB1*),
290 siderophore-mediated iron acquisition (*mbtA*), phosphate transport (*pstC2*), glycerol
291 catabolism (*glpK*), and RNA processing (*rnaseJ*) were generated and tagged with a
292 unique molecular barcode. These mutants were combined with a barcoded wild-type
293 parental strain, the resulting 'mini-pool' was subjected to *in vivo* selection in the same
294 manner as the TnSeq study, and the relative abundance of each mutant was
295 determined by sequencing the amplified barcodes. In each case, the difference in
296 relative abundance predicted by TnSeq was reproduced with deletion mutants (**Figure**
297 **4C**). In this simplified system, we were able to accurately quantify the expansion of the
298 bacterial population and calculate the "fitness" of each mutant relative to the wild-type
299 strain. Fitness reflects the inferred doubling time of the mutant, where a fitness of 1 is
300 defined as wild-type, and 0 represents a complete lack of growth. Even by this metric,
301 the deletion mutants displayed the differences in fitness between mouse strains that
302 was predicted by TnSeq (**Figure 4C**). The statistical significance of these differences in
303 abundance or fitness were similar for each mutant (between $p=0.009$ and $p=0.06$),
304 except for *mbtA* where the variation was higher, and confidence was modestly lower
305 ($p=0.07$ and $p=0.12$). This study also allowed us to estimate the sensitivity of the TnSeq
306 method, which could detect even the 30% fitness defect of the Δ*glpK* strain between the
307 B6 and CC018 animals (**Figure 4C**), a defect that was not observed in previous studies
308 in BALB/c mice (Bellerose et al., 2019; Pethe et al., 2010).
309

310 To also validate TnSeq predictions in a single-strain aerosol infection model, we used a
311 biotin biosynthetic mutant. *bioA* is necessary for biotin production and is essential for
312 growth in B6 mice (Woong Park et al., 2011). Our TnSeq study (**Table S4**) predicted
313 this mutant was less attenuated in the CAST background (ratio of input/selected = 12.1)
314 than in the B6 strain (ratio of input/selected = 42.2). Two weeks after aerosol infection,
315 we found that the Δ*bioA* mutant was cleared from the lungs and spleen of B6 mice but
316 displayed similar growth to wild-type in the lungs of CAST mice (**Figure 4D**). By 6
317 weeks post infection the Δ*bioA* mutant had also been largely cleared from the lungs of
318 CAST (**Figure 4D**). Thus, while TnSeq was unable elucidate the complex kinetics of this
319 phenotype, it accurately predicted the relative levels of growth attenuation in these host
320 backgrounds.
321
322 **The immunological diversity of CC mice is reflected in the pathogen's genetic**
323 **requirements**
324 The distribution of *Mtb*'s requirements across the mouse panel suggested the presence
325 of two broad categories of genes. A set of 136 "core" virulence functions were required
326 in the majority of mouse genotypes, and a second larger set of 607 "adaptive" virulence
327 genes were required in only a subset of lines (**Table S4**). The core functions included a
328 number of genes previously found to be important in B6 mice, including those
329 necessary for the synthesis of essential cofactors, such as pyridoxine (*Pdx1*) (Dick et
330 al., 2010); for the acquisition of nutrients, such as siderophore-bound iron (*irtAB*)
331 (Ryndak et al., 2010), cholesterol (*mce4*) (Pandey and Sassetti, 2008), glutamine (*glnQ*
332 and *rv2563*) (Bellerose et al., 2020); and for type VII secretion (ESX1 genes) (Stanley et

333 al., 2003). Despite the importance of these core functions, a large range in the relative
334 abundance of these mutants was observed across the panel, and in some cases
335 specific immunological requirements could be discerned. Mutants lacking the major
336 structural components of the ESX1 system were attenuated for growth in B6 mice, as
337 expected. This requirement was consistently enhanced in mice lacking *Rag2, Ifnγ*, or
338 *Nos2* (**Figure 4E**), consistent with the preferential role of ESX1 during the initial stage of
339 infection before the initiation of adaptive immunity (Stanley et al., 2003). In contrast, the
340 attenuation of mutants lacking the *glnQ* encoded glutamine uptake system was relieved
341 in all four immunodeficient mouse lines (**Figure 4E**). In both cases, the differential
342 mutant abundance observed in these KO mice was reproduced, or exceeded, in the CC
343 panel.
344
345 The adaptive virulence functions included a number of genes previously thought to be
346 dispensable in the mouse model and were only necessary in CC strains. For example,
347 the alkyl hydroperoxide reductase, AhpC has been proposed to function with the
348 adjacently encoded peroxiredoxin, AhpD and is critical for detoxifying reactive nitrogen
349 intermediates *in vitro* (Chen et al., 1998; Hillas et al., 2000). However, deletion of *ahpC*
350 has no effect on *Mtb* replication in B6 or BALB/c mice (Springer et al., 2001), and we
351 confirmed that *ahpC* and *ahpD* mutations had no effect in any of the B6-derived lines. In
352 contrast, *ahpC*, but not *ahpD* mutants were highly attenuated in a small number of CC
353 strains (**Figure 4E**). Similarly, the four phospholipase C enzymes of *Mtb* (*plcA-D*) are
354 implicated in both fatty acid uptake and modifying host cell membranes but are
355 dispensable for replication in B6 mice (Le Chevalier et al., 2015). Again, while we found

356 that none of these genes were required in B6-derived KO lines, the *plcD* mutants were
357 specifically underrepresented in a number of CC mice (**Figure 4E**). These individual
358 bacterial functions are controlled by regulatory proteins, such as the extracytoplasmic
359 sigma factors. Despite the importance of these transcription factors in the response to
360 stress, only *sigF* has been consistently been shown to contribute to bacterial replication
361 in standard inbred lines of mice (Geiman et al., 2004; Rodrigue et al., 2006). Our study
362 assesses the importance of each sigma factor in parallel across diverse host genotypes
363 and identified a clear role for several of these regulators. *sigC, sigI, sigF*, *sigL,* and *sigM*
364 mutants were each significantly underrepresented in multiple lines of mice, and several
365 of these phenotypes were only apparent in the diverse CC animals (**Figure 4E**). In
366 sum, the 607 adaptive functions that are differentially required across the host panel
367 represents nearly 20% of the non-essential gene set of Mtb, suggesting that a
368 significant fraction of the pathogen's genome is dedicated to maintaining optimal fitness
369 in diverse host environments.
370

371 **Differential genetic requirements define virulence pathways in *Mtb***

372 To more formally investigate the distinct stresses imposed on the bacterial population
373 across this host panel, we characterized the differentially required bacterial pathways.
374 Upon performing each possible pairwise comparison between the *in vivo* selected
375 mutant pools, we found 679 mutants whose representation varied significantly (FDR <
376 5%) in at least two independent comparisons (**Table S4**)**.** We then applied weighted
377 gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) to divide
378 the mutants into 20 internally-correlated modules. Further enrichment of these modules

379 for the most representative genes (intramodular connectivity > 0.6) revealed that nearly
380 all modules contained genes that are encoded adjacently in the genome and many of
381 these modules consisted of genes dedicated to a single virulence-associated function
382 (**Figure 5A**). Module 3 contains two distally encoded loci both known to be necessary
383 for ESX1-mediated protein secretion, the primary ESX1 locus (*rv3868-rv3883*) and the
384 *espACD* operon (*rv3616c-rv3614c*). Similarly, other modules consisted of genes
385 responsible for ESX5 secretion (Module 7), mycobactin synthesis (Module 4), the Mce1
386 and Mce4 lipid importers (Modules 5 and 16), phthiocerol dimycocerosate synthesis
387 (PDIM, Module 8), PDIM transport (Module 16), and phosphate uptake (Module 14).
388 The 20 genes assigned to Module 6 included two components of an important oxidative
389 stress resistance complex (*sseA* and *rv3005c*) and were highly enriched for mutants
390 predicted to be involved in this same process via genetic interaction mapping (11/20
391 genes were identified in (Nambi et al., 2015), a statistically significant overlap ($p < 2.8e-$
392 10 by hypergeometric test)). Thus, each module represented a distinct biological
393 function.
394
395 Many pathway-specific modules contained genes that represented novel functional
396 associations. For example, the gene encoding the sigma factor, *sigC*, was found in
397 Module 1 along with a non-ribosomal peptide synthetic operon. Previous genome-wide
398 ChIP-seq and overexpression screens support a role for SigC in regulating this operon
399 (Minch et al., 2015; Turkarslan et al., 2015). Similarly, *rv3220c* and *rv1626* have been
400 proposed to comprise an unusual two component system that is encoded in different
401 regions of the genome (Morth et al., 2005). Both of these genes are found in Module 2,

402 along with the PPE50 and PPE51 genes that encode at least one outer membrane
403 channel (Wang et al., 2020)(**Figure 5A**). In both cases, these associations support both
404 regulatory and obligate functional relationships between these genes. Six of the 20
405 modules were not obviously enriched for genes of a known pathway, demonstrating that
406 novel virulence pathways are important for adapting to changing host environments.
407
408 To explore the complexity of immune environments in the CC, we used the TnSeq
409 profiles of the 679 differentially fit *Mtb* mutants to cluster the mouse panel into 6 major
410 groups of host genotypes (**Figure 5B**). One mouse cluster was significantly associated
411 with high CFU (Cluster F, **Figure 5B**), which contained susceptible *Nos2*$^{-/-}$, *Cybb*$^{-/-}$, *Ifnγ*$^{-}$
412 $^{/-}$, and *Rag2*$^{-/-}$ animals. This high CFU cluster was associated with alterations in a
413 diverse set of bacterial modules and corresponded to an increased requirement for lipid
414 uptake (Modules 5 and 16) and ESX1, consistent with previous TnSeq studies in
415 susceptible *Nos2*$^{-/-}$ and C3HeB/FeJ mice (Mishra et al., 2017). In addition, we identified
416 a significant reduction in the requirement for the oxidative stress resistance (Module 6)
417 in the highest CFU cluster. Despite these associations between bacterial genetic
418 requirements and susceptibility, the clustering of mouse genotypes was largely
419 independent of overall susceptibility. Similarly, while Module 1 was significantly
420 associated with high IFNγ levels, other bacterial fitness traits were not highly correlated
421 with cytokine abundance (**Figure S3)**. Instead, each major mouse cluster was
422 associated with a distinct profile of *Mtb* genetic requirements. This observation
423 supported the presence of qualitatively distinct disease states and complex genetic
424 control of immunity.

425

426 **Identification of host-pathogen genetic interactions**

427 To investigate the host genetic determinants of the bacterial microenvironment, we
428 associated *Mtb* mutant fitness profiles with variants in the mouse genome. When the
429 relative abundance of each *Mtb* mutant was considered individually, the corresponding
430 "Host Interacting with Pathogen QTL" (*Hip*QTL) were distributed across the mouse
431 genome (**Figure 6A**). 41 of these traits reached an unadjusted p-value threshold of 0.05
432 and can be considered as robust for single hypothesis testing (*Hip1-41*, **Table S2** and
433 **S5**). These included *Hip*QTL associated with both *ahpC* and *eccD1*, that explain at
434 least a portion of the observed variable abundance of these mutants (**Figure 4E**). In
435 order to reduce complexity and increase the power of this analysis, we performed QTL
436 mapping based on the first principal component of each of the previously defined
437 modules of *Mtb* virulence pathways (**Figure 5A**). Three of these "eigentraits" were
438 associated with QTL at a similar position on chromosome 10 (**Figure 6B**),
439 corresponding to Module 3 (TypeVII secretion, ESX1), Module 4 (mycobactin synthesis,
440 *mbt*), or Module 16 (cholesterol uptake, *mce4*). In all three cases, a single mutant from
441 the module was independently associated with a QTL at the same position as the
442 module eigentrait (**Table S5**, *Hip21, Hip22, Hip24*), and all genes in the corresponding
443 network cluster (**Figure 5A**) mapped to the same location (**Figure 6C-E**). While not all
444 individual traits mapped with high confidence, the coincidence of these multiple QTL
445 was statistically significant (**Figure 6B**).

446

447 Both the relative positions of the module-associated QTL and the associated founder
448 haplotypes indicated that a single genetic variant controlled the abundance of ESX1 and
449 *mbt* mutants (*Hip42*). Specifically, we found no statistical support for differentiating
450 these QTL based on position (p=0.93)(Boehm et al., 2019) and the same founder
451 haplotypes were associated with extreme trait values at both loci, though they had
452 opposite effects on the abundance of ESX1 mutants and *mbt* mutants (**Figure 6F**). We
453 conclude that a single haplotype has a pleiotropic effect on *Mtb*'s environment and has
454 opposing effects of the requirement for mycobactin synthesis and ESX1 secretion. The
455 relationship between this variant and the *mce4*-associated QTL (*Hip43*) was less clear,
456 as the statistical support for independent QTL was weak (ESX1 and *mce4* QTL p=0.17;
457 *mbt* and *mce4* QTL p=0.08) and the effects of founder haplotypes were similar but not
458 identical (**Figure 6F**). Some of this ambiguity may be related to the relatively small
459 range in trait values for *mce4*, compared to either ESX1 or *mbt* (**Figure 6G**). Based on
460 this data, we report two distinct *Hip*QTL in this region (*Hip42* and *43*, **Table S5**).
461
462 Two *Tip*QTL overlapped with *Hip*QTL (**Figure S2**), suggesting specific interactions
463 between bacterial fitness and immunity. However, most *Tip*- and *HipQTL* were distinct,
464 indicating that the fitness of sensitized bacterial mutants can be used to detect genetic
465 variants that subtly influence the bacterial environment but not overtly alter disease. We
466 chose to further investigate whether *Hip*QTL might alter overall bacterial disease using
467 the most significant *Hip*QTL on chromosome 10 (*Hip42*). We found that the founder
468 haplotypes associated with extreme trait values at this QTL could differentiate CC
469 strains with significantly altered total bacterial burden, and the NOD and WSB

470 haplotypes were associated with higher bacterial numbers (p=0.0085 for spleen CEQ;
471 p=0.027 for spleen CFU; **Figure 6H**). Thus, not only could the *Hip*QTL strategy identify
472 specific interactions between host and bacterial genetic variants, but it also appears to
473 be a sensitive approach for identifying host loci that influence the trajectory of disease.
474

475 **Identifying candidate genes underlying QTL**

476 A pipeline was designed to prioritize genetic variants based on genomic and
477 tuberculosis disease criteria. We concentrated on three QTL: two that were highly
478 significant and with clear allele effects (*Tip5, Hip42*), and the *Tip8* locus which we
479 validated by intercross. For each QTL region, we identified genes that belonged to a
480 differentially expressed transcriptional module in mouse lungs following *Mtb* infection
481 (Moreira-Teixeira et al., 2020). Next, we identified genetic variants segregating between
482 the causal CC haplotypes in the gene bodies corresponding to these transcripts, and
483 prioritized missense or nonsense variants.
484

485 For the *Tip5* QTL underlying CEQ, CFU, and IL-10 levels, we identified nine candidate
486 genes with regulatory or splicing variants and two genes with missense variants specific
487 to the NOD haplotype. Of these candidates, cathepsin Z (*Ctsz)* encodes a lysosomal
488 cysteine proteinase and has previously been associated with TB disease risk in humans
489 (Adams et al., 2011; Cooke et al., 2008). The QTL underlying lung CFU and CXCL1
490 abundance (*Tip8*), which was driven solely by the genetically divergent CAST founder
491 haplotype, contained over 50 genes (**Table S6**) and will need further refinement. The
492 QTL associated with the abundance of ESX1 and *mbt* mutants (*Hip42*) had a complex

493 causal haplotype pattern (AJ/B6/NZO vs. 129/CAST/PWK vs. NOD/WSB) suggesting
494 multiple variants might be impacting common genes. Within this interval, we identified
495 13 genes expressed in response to *Mtb* infection, three of which had SNPs fully or
496 partially consistent with at least one of the identified causal haplotype groups (**Table**
497 **S6**). *Ank3* contains several SNPs in the 3' UTR and other non-coding exons that
498 differentiated NOD/WSB from the other haplotypes. Similarly, *Fam13c* had two
499 missense mutations following the same haplotype pattern. For the AJ/B6 haplotype
500 state, we identified a missense mutation and several variants in the 3' UTR of *Rhobtb1*,
501 which belongs to the Rho family of the Ras superfamily of small GTPases (Goitre et al.,
502 2014). Overall, the evidence supports a role for *Rhobtb1* in a monogenic effect at the
503 chromosome 10 locus. This evidence includes both protein coding differences dividing
504 AJ/B6 from the other haplotypes, a potential expression/transcript regulatory difference
505 that segregates the NOD/WSB state from the remaining parental haplotypes, and a
506 plausible role for this gene in controlling intracellular trafficking (Long et al., 2020) and
507 the opposing requirements for ESX1 and mycobactin.
508

509 **Discussion**

510 Our immunological analysis of the CC panel identified correlates of TB disease
511 progression that were consistent with previous studies in both mice and human patients
512 (Ahmed et al., 2020; Niazi et al., 2015; Zak et al., 2016). More importantly, we identified
513 outlier strains that produce distinct immunological states, highlighting how our previous
514 reliance on genetically homogenous lab strains of mice has oversimplified our
515 understanding of TB pathogenesis. For example, despite the generally strong

516 correlation between lung bacterial burden and disease, CC030/GeniUnc and
517 CC040/TauUnc mice suffered from more inflammation and wasting than would be
518 predicted from the number of bacteria in their lungs or spleens. This phenotype reflects
519 a failure of disease "tolerance", which is proposed to be a critical determinant of
520 protective immunity (Ayres and Schneider, 2012; Olive et al., 2018). Similarly, we
521 identified a number of CC genotypes that produce very low, or undetectable, levels of
522 the protective cytokine IFNγ, but still control lung bacterial replication. While a growing
523 body of literature suggests that immune responses distinct from the canonical Th1
524 response can control infection (Lu et al., 2019; Sakai et al., 2016), these CC strains are
525 the first example of an animal model in which IFNγ appears to be dispensable. The
526 reproducibility of the CC strains facilitated the identification of these phenotypes and
527 provides tractable models for their characterization.
528
529 The ability to separate aspects of the immune response from disease progression
530 implied that these features are under distinct multigenic control. This conclusion is
531 supported by genetic mapping, which identified a number of variants that control distinct
532 aspects of the immune response to *Mtb*. The QTL identified in this study are generally
533 distinct from CC loci that control immunity to viruses (Ferris et al., 2013; Gralinski et al.,
534 2017; Noll et al., 2020) or another intracellular bacterial pathogen, *Salmonella* (Zhang et
535 al., 2019). However, *Tip8* and *Tip10* overlap with QTL previously defined via *Mtb*
536 infection of CC001xCC042 F_2 intercross population (Smith et al., 2019) suggesting that
537 common variants may have been identified in both studies. While the specific genetic
538 variants responsible for these QTL remain unknown, both coincident trait mapping and

539 bioinformatic analysis suggest mechanistic explanations for some QTL-phenotype
540 associations. For example, a single interval on chromosome 2 controls CFU levels and
541 IL-10, and contains a variant in the *Ctsz* gene encoding Cathepsin Z. *Ctsz* is a strong
542 candidate considering its known roles in autophagy (Amaral et al., 2018), dendritic cell
543 differentiation and function (Obermajer et al., 2008), its upregulation in non-human
544 primates (Ahmed et al., 2020) and human patients with *Mtb* (Zak et al., 2016), and the
545 association of *CTSZ* variants with disease risk in human TB studies (Adams et al.,
546 2011; Cooke et al., 2008).
547
548 Using TnSeq as a multidimensional phenotyping method across this population
549 provided insight into how the diversity of host-derived microenvironments have shaped
550 the pathogen's genome. While *Mtb* is an obligate pathogen with no significant
551 environmental niche, only a minority of the genes in its genome have been found to
552 contribute to bacterial fitness in either laboratory media or individual inbred mouse
553 models, leaving the pressures that maintain the remaining genomic content unclear. We
554 find that approximately three times more genes contribute to bacterial growth or survival
555 in the CC population than in the standard B6 model. While some bacterial genetic
556 requirements could be associated with known immune pathways, most of the differential
557 pressures on bacterial mutants could not be attributed to these simple deficiencies in
558 known mechanisms of immune control. Instead, it appears that the CC population
559 produces a spectrum of novel environments, and that a relatively large fraction of the
560 pathogen's genome is needed to adapt to changing immune pressures. Differential
561 pressures on these adaptive virulence functions are similarly apparent in genomic

562 analyses of *Mtb* clinical isolates. Signatures of selection have been detected in ESX1-
563 related genes (Holt et al., 2018; Sousa et al., 2020), *phoPR* (Gonzalo-Asensio et al.,
564 2014), and the oxidative stress resistance gene *sseA* (de Keijzer et al., 2014),
565 suggesting that *Mtb* is exposed to similarly variable host pressures in genetically diverse
566 human and mouse populations. While the combinatorial complexity of associating host
567 and pathogen genetic variants in natural populations is daunting, the identification of
568 *Hip*QTL in the CC panel indicates that these inter-species genetic interactions can be
569 important determinants of pathogenesis and can be dissected using tractable models of
570 diversity.
571
572 **Acknowledgements:** We thank all members of Sassetti Lab, past and present for
573 technical help and discussions; Dr. Nathan Hicks and Dr. Sarah Fortune for kindly
574 providing the *rnaseJ* mutant; Dr. David Tobin for insightful manuscript comments; Dr.
575 Dennis Ko for QTL acronym creativity; Erin Curtis for mouse schema expertise and the
576 Systems Genetic Core at UNC for their help in procuring CC mice in timely fashion. This
577 work was supported by NIH grants AI132130 to C. M. Sassetti and FPMV;
578 U19AI100625 to FPMV and MTF; a fellowship from the Charles H. King Foundation to
579 C. M. Smith; and a HHMI Gilliam Fellowship A20-0146 to BKH. The genetic
580 characterization of the CC strains was supported in part by NIH grant U24HG010100 to
581 FPMV.
582
583 **Author contributions:** Conceptualization**,** C.M. Smith and C.M Sassetti; Methodology,
584 R.E.B, K.C.M, K.P, T.R.L; Investigation, C.M. Smith, M.K.P, B.B.M, J.E.L, M.C.K,

585 M.M.B, A.J.O, C.J.R, C.M. Sassetti; Validation, S.W.P, H.L, S.E, D.S; Formal Analysis,
586 R.E.B, F.J.B, R.K.M, B.K.H, C.L, M.T.F, T.R.I; Resources, G.D.S, P.H, T.A.B;
587 Visualization, C.M. Smith, R.E.B, M.K.P, R.K.M, C.M. Sassetti; Writing – original draft,
588 C.M Sassetti and C.M Smith; Writing – review and editing, all authors; Supervision,
589 C.M. Sassetti; Funding acquisition, C.M. Sassetti, M.T.F, F.P.M.V.
590
591 **Declarations of interests:** None.
592
593 **Materials and Methods:**
594 *Ethics statement:*
595 Mouse studies were performed in strict accordance using the recommendations
596 from the Guide for the Care and Use of Laboratory Animals of the National
597 Institute of Health and the Office of Laboratory Animal Welfare. Mouse studies at
598 the University of Massachusetts Medical School (UMASS) were performed using
599 protocols approved by the UMASS Institutional Animal Care and Use Committee
600 (IACUC) (Animal Welfare Assurance Number A3306-01) in a manner designed to
601 minimize pain and suffering in *Mtb*-infected animals. Any animal that exhibited
602 severe disease signs was immediately euthanized in accordance with IACUC
603 approved endpoints. All mouse studies at UNC (Animal Welfare Assurance
604 #A3410-01) were performed using protocols approved by the UNC Institutional
605 Animal Care and Use Committee (IACUC).
606
607 *Mice:*

608 Male and female Collaborative Cross parental strains (A/J #0646; C57BL/6J #0664;
609 129S1/SvImJ #02448; NOD/ShiLtJ #01976; NZO/HiLtJ #02105; CAST/EiJ #0928,
610 PWK/PhJ #3715 and WSB/EiJ #01145) and single gene immunological knockout mice
611 were purchased from The Jackson Laboratory (*Nos2*$^{-/-}$ #2609, *Cybb*$^{-/-}$ #2365, *Ifnγ*$^{-/-}$
612 #2287) or Taconic (*Rag*N12) and bred at UMASS. Male mice from 52 CC strains were
613 purchased from the UNC Systems Genetics Core Facility (SGCF) between July 2013
614 and August 2014. The 52 CC strains used in this study include: CC001/Unc,
615 CC002/Unc, CC003/Unc, CC004/TauUnc, CC005/TauUnc, CC006/TauUnc,
616 CC007/Unc, CC008/GeniUnc, CC009/Unc, CC010/GeniUnc, CC011/Unc,
617 CC012/GeniUnc, CC013/GeniUnc, CC015/Unc, CC016/GeniUnc, CC017/Unc,
618 CC018/Unc, CC019/TauUnc, CC021/Unc, CC022/GeniUnc, CC023/GeniUnc,
619 CC024/GeniUnc, CC025/GeniUnc, CC027/GeniUnc, CC028/GeniUnc, CC029/Unc,
620 CC030/GeniUnc, CC031/GeniUnc, CC032/GeniUnc, CC033/GeniUnc, CC034/Unc,
621 CC035/Unc, CC037/TauUnc, CC038/GeniUnc, CC039/Unc, CC040/TauUnc,
622 CC041/TauUnc, CC042/GeniUnc, CC043/GeniUnc, CC044/Unc, CC045/GeniUnc,
623 CC046/Unc, CC047/Unc, CC051/TauUnc, CC055/TauUnc, CC056/GeniUnc,
624 CC057/Unc, CC059/TauUnc, CC060/Unc, CC061/GeniUnc, CC065/Unc,
625 CC068/TauUnc. More information regarding the CC strains can be found at
626 http://csbio.unc.edu/CCstatus/index.py?run=AvailableLines.information.
627
628 CC030 x CC029 F_2 mice were generated at UNC by crossing CC030 and CC029
629 mice (purchased from the SGCF in 2016) to generate F_1s (both CC030 dam by
630 CC029 sires as well as CC029 dam by CC030 sires). The resulting F_1s were

631 subsequently intercrossed to generate F_2 mice with all possible grandparental
632 combinations. Male and female F_2 mice were shipped to UMASS for *Mtb*
633 infections.
634
635 All mice were housed in a specific pathogen-free facility under standard
636 conditions (12hr light/dark, food and water ad libitum). Mice were infected with
637 *Mtb* between 8-12 weeks of age. Male and female mice were used, unless
638 otherwise noted.
639
640 *M.tuberculosis strains:*
641 All *M. tuberculosis* strains were grown in Middlebrook 7H9 medium containing
642 oleic acid-albumin-dextrose-catalase (OADC), 0.2% glycerol, and 0.05% Tween
643 80 to log-phase with shaking (200 rpm) at 37°C. Hygromycin (50 ug/ml) or
644 kanamycin (20 ug/ml) was added when necessary. The Δ*bioA* strain was made
645 by homologous recombination as previously described (Woong Park et al.,
646 2011). For pooled mutant infections, deletion strains were constructed using
647 ORBIT (Murphy et al., 2018), which included gene replacement by the vector
648 pKM464 carrying unique q-Tag sequences to identify each mutant for deep
649 sequencing. The *rnaseJ* mutant was also made by ORBIT and was kindly
650 provided by Dr. Nathan Hicks and Dr Sarah Fortune. Prior to all *in vivo* infections,
651 cultures were washed, resuspended in phosphate-buffered saline (PBS)
652 containing 0.05% Tween 80, and sonicated before diluting to desired
653 concentration (see below).

654

655 *Mouse Infections:*

656 For TnSeq experiments, $1x10^6$ CFU of saturated *Himar1* transposon mutants
657 (Sassetti et al., 2003) was delivered via intravenous tail vein injection. Mice were
658 infected over 3 infection batches, as denoted in **Table S1**). At indicated time
659 points mice were euthanized, and organs were harvested then homogenized in a
660 FastPrep-24 (MP Biomedicals). CFU was determined by dilution plating on 7H10
661 agar with 20 ug/mL kanamycin. For library recovery, approximately $1x10^6$ CFU
662 per mouse was plated on 7H10 agar with 20ug/mL kanamycin. After three weeks
663 of growth, colonies were harvested by scraping and genomic DNA was extracted.
664 The relative abundance of each transposon mutant was estimated as described
665 (Long et al., 2015).

666

667 Single strain aerosol infections were performed in a Glas-Col machine to deliver
668 50-150 CFU/mouse. At indicated time points, mice were euthanized, and organs
669 were harvested then homogenized in a FastPrep-24 (MP Biomedicals). CFU was
670 determined by dilution plating on 7H10 agar with 20 ug/mL kanamycin or
671 50ug/mL hygromycin as required.

672

673 Chromosomal equivalent (CEQ) was enumerated according to previously
674 published protocol (Lin et al., 2014; Muñoz-Elías et al., 2005). Cytokines and
675 chemokines were assayed from organ homogenates using the pro-inflammatory
676 focused 32-plex (Eve Technologies, Calgary, CA) or IFNγ DuoSet ELISA (R&D

677 Systems) according to manufacturer protocol. IFNγ release was assayed by
678 ELISPOT (BD Biosciences #551083) according to manufacturer protocol. Single
679 cell suspensions of lung or spleen tissue were stimulated for 18-24 hours with
680 4ug/ml purified protein derivative (PPD) and after development, spots were
681 quantified using a CTL Immunospot S5 Analyzer.

682 For pooled mutant infections, mice were infected with a pool of deletion mutants at
683 equal ratios via the intravenous route ($1x10^6$ CFU/mouse). At indicated time points,
684 approximately 10,000 CFU from the spleen homogenate of each mouse was plated on
685 7H10 agar. Genomic DNA was extracted for sequencing as described previously (Long
686 et al., 2015). Sequencing libraries spanning the variable region of each q-Tag were
687 generated using PCR primers binding to regions common among all q-Tags, similar to
688 previously described protocols (Bellerose et al., 2020; Martin et al., 2017). In each PCR,
689 a unique molecular counter was incorporated into the sequence to allow for the
690 accurate counting of input templates and account for PCR jackpotting. The libraries
691 were sequenced to 1,000-fold coverage on an Illumina NextSeq platform using a 150-
692 cycle Mid-Output kit with single-end reads. Total abundance of each mutant in the
693 library was determined by counting the number of reads for each q-Tag with a unique
694 molecular counter. Relative abundance of each mutant in the pool was then calculated
695 by dividing the total abundance of a mutant by the total abundance of reads for wild-
696 type H37Rv. The relative abundance was normalized to relative abundance at initial
697 infection (Day 0) and log_2 transformed. Fitness was calculated as previously described
698 (Palace et al., 2014).

699

700 QUANTIFICATION AND STATISTICAL ANALYSIS

701 *TnSeq analysis:*

702 TnSeq libraries were prepared and counts of each transposon mutant were estimated
703 as described (Long et al., 2015). NCBI Reference Sequence NC_018143.1 was used
704 for H37Rv genome and annotations. In the majority of cases, two replicate mouse
705 libraries were used per genotype. Only a single TnSeq library was obtained for CC010,
706 CC031, CC037, CC059, CC016, and PWK/PhJ. Insertion mutant counts across all
707 libraries were normalized by beta-geometric correction (DeJesus et al., 2015), binned
708 by gene, and replicate values for each mouse genotype averaged. Mean values for
709 each gene were divided by the grand mean then $\log_2$ transformed and quantile
710 normalized. The resulting phenotype values were used for both WGCNA and QTL
711 mapping.

712

713 To eliminate genes having no meaningful variation across the mouse panel, statistical
714 tests of $\log_2$ fold change (LFC) in counts between all possible pairs of mouse genotypes
715 were performed by resampling (DeJesus et al., 2015). 679 “significantly varying genes”
716 were identified whose representation varied significantly (FDR < 5%) in at least two
717 independent comparisons. For relative mutant abundance estimates, LFC in counts
718 between *in vitro*-grown H37Rv (6 replicate libraries) vs libraries from each mouse
719 genotype were determined by resampling as above.

720

721 *WGCNA analysis:*

722 Weighted gene correlation network analysis (WGCNA) was applied to categorize the
723 679 significantly varying genes into 20 internally-correlated modules (Langfelder and
724 Horvath, 2008). Modules were filtered (intramodular connectivity > 0.6) to obtain the
725 most representative genes. First principal component scores of module eigengenes
726 were used as phenotype values for QTL mapping after first winsorizing (q=0.05) using
727 the R package *broman* (https://cran.r-project.org/web/packages/broman/index.html).
728
729 In order to perform association analysis between modules of genes and clusters of mice
730 (**Figure 5B**), the mice were clustered based on the matrix of TnSeq LFCs for
731 significantly varying genes using *hclust* in R (with the "Ward.D2" distance metric). Then,
732 for each module of genes, the LFCs in each cluster of mice were pooled and compared
733 to all the other mice using a T-test, identifying modules with a mean LFC in a specific
734 mouse cluster that is significantly higher or lower than the average across all the other
735 mice. The resulting p-values over all combinations of gene modules and mouse
736 clusters were adjusted using Benjamini-Hochberg for an overall FDR < 0.05.
737
738 *Disease-related trait analysis and Heritability estimation:*
739 For the trait heatmap, trait values were clustered (*hclust* in R package *heatmaply;* traits
740 scaled as per default function) and dendrogram nodes colored by 3 k-means.
741 Correlation between disease-related TB traits was determined by Pearson's correlation
742 and visualized using *corrplot* (ordered by *hclust*). Heritability (h^2) of the immunological
743 and TB disease-related traits was calculated by estimating the percent of variation
744 attributed to a genotype as previously described (Noll et al., 2020).

745

746 *Genotyping and QTL mapping:*

747 A subset of the inbred CC mice used in the analysis were genotyped on the GigaMUGA
748 array (Morgan et al., 2015) available from Neogen Inc. The inbred parents, F_1s and F_2
749 mice from the CC030xCC029 cross were genotyped on the MiniMUGA array (Sigmon et
750 al., 2020) at Neogen Inc.

751

752 For CC030 x CC029 F_2 analysis, markers were filtered, restricting only to those
753 diagnostic between these strains (See **Table S3**). To assess the relative impact of the
754 Chr7 and 15 loci, a linear regression on CFU was run with CAST haplotypes on Chr7,
755 Chr15, and their interaction as factors.

756

757 For QTL mapping in the CC panel, the Most Recent Common Ancestor (Srivastava et
758 al., 2017) 36-state haplotypes were downloaded from the UNC Systems Genetics Core
759 Facility and simplified to 8-state haplotype probabilities (for the 8 CC founder strains),
760 which is appropriate for additive genetic mapping. We generated 36-state haplotype
761 probabilities from the individual CC mice genotyped on GigaMUGA and combined these
762 data with the MRCA data to obtain a common genome cache.

763

764 For CC QTL analysis, genotype and phenotype data were imported into R (version
765 3.6.1) and reformatted for R/qtl2 (version 0.20) (Broman et al., 2019). Individual TnSeq
766 and clinical trait phenotype values were winsorized (q = 0.006) as above. GigaMUGA
767 annotations were downloaded from the Jackson Laboratory, and markers were thinned

768 to a spacing of 0.1 cM using the reduce_markers function of R/qtl2. The final genetic
769 map contained 10,067 markers. QTL mapping was carried out using a linear mixed
770 model with LOCO (leave one chromosome out) kinship. For clinical trait scans, batch
771 (denoted by “block” in Table S1) was included as an additive covariate. Significance
772 thresholds for QTL were estimated using 10,000 permutations (scan1_perm function).
773 For each trait, the maximum LOD scores from the permutation scans were used to fit
774 generalized extreme value distributions, from which genome-wide permutation p-values
775 were calculated. LOD profiles and effect plots were generated using the plotting
776 functions of the R/qtl2 package. Multiple QTL at similar genetic locations were assessed
777 for independence using qtl2pleio with 400 bootstrap samples (Boehm et al., 2019). The
778 quantile-based permutation thresholding method of Neto et al. (Neto et al., 2012) was
779 used to assess the statistical significance of co-mapping traits. The NL-method, which
780 determines the LOD thresholds controlling genome wide error rate for a given p-value
781 and “hotspot” size, was employed.
782
783 *Candidate gene prioritization:*
784 To identify potential candidate genes, we focused on three QTL that were either
785 statistically significant (*Tip5*, *Hip42*) or were validated by intercross (*Tip8*). For each
786 QTL interval (determined by Bayesian interval in qtl2), we identified mouse genes that
787 were in differentially expressed modules between infected lungs of resistant and
788 susceptible mouse strains (Moreira-Teixeira et al., 2020). Of these genes, we next used
789 the Sanger sequence data (Keane et al., 2011) to filter on genetic variants segregating

790 between CC haplotypes. Where there were clear causal haplotypes, we further filtered
791 to genes with missense or nonsense variants.
792
793 **Data availability:**
794 All relevant data to support the findings of this study are located within the paper and
795 supplemental files or are in the process of being made publicly available at the time of
796 this submission; all mouse phenotype data are being deposited in Mouse Phenome
797 Database (https://phenome.jax.org), and sequence data is being deposited in the NCBI
798 Gene Expression Omnibus (GEO).
799
800 **Supplemental information titles and legends:**
801
802 **Table S1 - CC TB phenotypes**. TB disease-related phenotypes measured in the CC
803 and parental strains. Recorded values are the average of measurements from 2-6 mice
804 per genotype (average n=3). Mice were infected over 3 batches (denoted by "block").
805 "Blaze" denotes genotypes with white head-spotting coat color trait (WSB haplotype for
806 *Kitl*; used as a positive control/proof-of-concept for QTL mapping as per (Aylor et al.,
807 2011; Smith et al., 2019).
808
809 **Table S2 - Heritability (h^2) estimates for each measured TB-disease associated**
810 **phenotype (Tuberculosis ImmunoPhenotypes)**. h^2 was calculated from the
811 percentage of variation attributed to strain differences in each trait across the CC
812 strains, as previously described (Noll et al., 2020). Weight change is the percentage of

813 weight (grams), CFU/CEQ is log_{10} transformed, cytokines are measured in pg/mL lung
814 homogenate and log_{10} transformed**.**
815
816 **Table S3 - F_2 Intercross phenotype data**. Lung CFU measured in 46 F_2 mice derived
817 from CC030xCC029 intercross strategy. F_2 mice were CAST at chromosome 7, CAST
818 at Chromosome 15, CAST at both or CAST at neither locus. The infected F_2 cohort
819 included both male and female mice, as indicated.
820
821 **Table S4 - TnSeq Summary Table**. LFC values represent the log_2 fold-change (LFC)
822 between input and mouse-selected pools. "NA" indicates genes with fewer than 3
823 occupied TA transposon insertion sites for the indicated comparison. Qvals represent
824 adjusted p-values comparing mutant abundance in input and selected pools. "NA"
825 indicates genes with fewer than 3 occupied TA transposon insertion sites for the
826 indicated comparison. Required *in vivo*: "TRUE" indicates the mutant is significantly
827 underrepresented (Qval<0.05) after in mouse-selection in at least two mouse strains.
828 Required in B6: "TRUE" indicates the mutant is significantly underrepresented (Qval
829 <0.05) after in selection in B6 mice. Required in KO mice: "TRUE" indicates the mutant
830 is significantly underrepresented (Qval<0.05) after in selection in *Rag*$^{-/}$-, *Nos2*$^{-/-}$, *Cybb*$^{-/-}$,
831 or *Ifng*$^{-/-}$ mice. Core gene set: TRUE indicates the mutant is significantly
832 underrepresented (Qval<0.05) in 30 mouse strains. "Module" corresponds to WGCNA
833 module number as illustrated in **Figure 5A**. Mouse strains are listed in the same order
834 as **Figure 5B**, with the corresponding cluster designation.
835

836 **Table S5 - *Hip*QTL for single *Mtb* mutant QTL and eigentrait/module QTL**. *Hip1-41*
837 each represent host loci associated with the relative abundance of a single mutant
838 ($p<0.05$). *Hip42-46* correspond to *Mtb* eigentraits identified in network analysis in **Figure**
839 **5** (including significant $p<0.05$ and suggestive $p<0.25$). Figure column headings: QTL,
840 quantitative trait loci; *Mtb*, *Mycobacterium tuberculosis*; Module #, number determined
841 from WGCNA modules; ORF, open reading frame; ID, identification number; LOD,
842 logarithmic of the odds; Chr, chromosome.
843
844 **Table S6 - Candidate genes within QTL regions**. Prioritized candidates shown for
845 selected QTL. Candidates were prioritized by filtering on 1) differential expression
846 during *Mtb* infection, and 2) variants within TB-expressed genes that segregated
847 between informative CC haplotypes. Genes listed below contain non-synonymous
848 variants (ie. amino acid changes, regulatory mutations or splicing mutations) consistent
849 with the identified singly causal haplotype (NOD for *Tip5;* CAST for *Tip8*). *Hip42*
850 displayed a more complex haplotype pattern (WSB/NOD vs AJ/B6/NZO), and candidate
851 selection is discussed in the main text. Genes with missense or nonsense variants
852 (denoted by *).
853
854 **Figure S1 - Phenotypic relationships between TB disease-related traits**. Correlation
855 between 32-measured TB traits was determined by Pearson's correlation and visualized
856 using corrplot version 0.84 (ordered by hclust method "complete") in R version 4.0.3.
857 Violet indicates a positive correlation, and yellow indicates negative correlations. The

858 correlation coefficient for each trait comparison (r value) is noted on each square. EoD
859 (earliness of Death); CEQ (Chromosomal equivalents).
860
861 **Figure S2 - Visual representation of QTL mapped in the CC TnSeq infection**
862 **screen**. Tuberculosis ImmunoPhenotypes (*Tip*) QTL (QTL mapped by disease-
863 associated traits in CC mice), are shown in green. *Tip*QTL mapped by separate traits
864 that share similar founder effects were considered to be the same QTL and were named
865 accordingly. Host Interacting with Pathogen (*Hip*) QTL, (QTL mapped by individual
866 TnSeq mutant relative abundance profiles), are shown in purple. After WGCNA mutant
867 clustering and mapping with representative eigengenes from each module, QTL
868 mapped by module eigengenes are shown in magenta.
869
870 **Figure S3 - Module-trait associations**. Rows correspond to modules, columns to
871 clinical traits. Numbers in each cell give the Pearson correlation between the module
872 eigengene and the trait values across the 60-mouse panel (p-values in parentheses).
873 Cells are colored by correlation as shown in the color legend (right).
874
875 **References**

876 Abel, L., Fellay, J., Haas, D.W., Schurr, E., Srikrishna, G., Urbanowski, M., Chaturvedi,
877 N., Srinivasan, S., Johnson, D.H., and Bishai, W.R. (2018). Genetics of human
878 susceptibility to active and latent tuberculosis: present knowledge and future
879 perspectives. Lancet Infect Dis *18*, e64-e75.

880 Adams, L.A., Möller, M., Nebel, A., Schreiber, S., van der Merwe, L., van Helden, P.D.,
881 and Hoal, E.G. (2011). Polymorphisms in MC3R promoter and CTSZ 3'UTR are
882 associated with tuberculosis susceptibility. Eur J Hum Genet *19*, 676-681.

883 Ahmed, M., Thirunavukkarasu, S., Rosa, B.A., Thomas, K.A., Das, S., Rangel-Moreno,
884 J., Lu, L., Mehra, S., Mbandi, S.K., Thackray, L.B.*, et al.* (2020). Immune correlates of
885 tuberculosis disease and risk translate across species. Sci Transl Med *12*.

886 Altare, F., Durandy, A., Lammas, D., Emile, J.F., Lamhamedi, S., Le Deist, F., Drysdale,
887 P., Jouanguy, E., Döffinger, R., Bernaudin, F.*, et al.* (1998). Impairment of
888 mycobacterial immunity in human interleukin-12 receptor deficiency. Science *280*, 1432-
889 1435.

890 Amaral, E.P., Riteau, N., Moayeri, M., Maier, N., Mayer-Barber, K.D., Pereira, R.M.,
891 Lage, S.L., Kubler, A., Bishai, W.R., D'Império-Lima, M.R.*, et al.* (2018). Lysosomal
892 Cathepsin Release Is Required for NLRP3-Inflammasome Activation by Mycobacterium
893 tuberculosis in Infected Macrophages. Front Immunol *9*, 1427.

894 Ansari, M.A., Pedergnana, V., C, L.C.I., Magri, A., Von Delft, A., Bonsall, D., Chaturvedi,
895 N., Bartha, I., Smith, D., Nicholson, G.*, et al.* (2017). Genome-to-genome analysis
896 highlights the effect of the human innate and adaptive immune systems on the hepatitis
897 C virus. Nat Genet *49*, 666-673.

898 Aylor, D.L., Valdar, W., Foulds-Mathes, W., Buus, R.J., Verdugo, R.A., Baric, R.S.,
899 Ferris, M.T., Frelinger, J.A., Heise, M., Frieman, M.B.*, et al.* (2011). Genetic analysis of
900 complex traits in the emerging Collaborative Cross. Genome Res *21*, 1213-1222.

901 Ayres, J.S., and Schneider, D.S. (2012). Tolerance of infections. Annu Rev Immunol *30*,
902 271-294.

903 Barber, D.L., Mayer-Barber, K.D., Feng, C.G., Sharpe, A.H., and Sher, A. (2011). CD4
904 T cells promote rather than control tuberculosis in the absence of PD-1-mediated
905 inhibition. J Immunol *186*, 1598-1607.

906 Bellerose, M.M., Baek, S.-H., Huang, C.-C., Moss, C.E., Koh, E.-I., Proulx, M.K., Smith,
907 C.M., Baker, R.E., Lee, J.S., Eum, S.*, et al.* (2019). Common Variants in the Glycerol
908 Kinase Gene Reduce Tuberculosis Drug Efficacy. mBio *10*, e00663-00619.

909 Bellerose, M.M., Proulx, M.K., Smith, C.M., Baker, R.E., Ioerger, T.R., and Sassetti,
910 C.M. (2020). Distinct Bacterial Pathways Influence the Efficacy of Antibiotics against
911 Mycobacterium tuberculosis. mSystems *5*.

912 Berthenet, E., Yahara, K., Thorell, K., Pascoe, B., Meric, G., Mikhail, J.M., Engstrand,
913 L., Enroth, H., Burette, A., Megraud, F.*, et al.* (2018). A GWAS on Helicobacter pylori
914 strains points to genetic variants associated with gastric cancer risk. BMC Biol *16*, 84.

915 Boehm, F.J., Chesler, E.J., Yandell, B.S., and Broman, K.W. (2019). Testing Pleiotropy
916 vs. Separate QTL in Multiparental Populations. G3 (Bethesda) *9*, 2317-2324.

917 Bogunovic, D., Byun, M., Durfee, L.A., Abhyankar, A., Sanal, O., Mansouri, D., Salem,
918 S., Radovanovic, I., Grant, A.V., Adimi, P.*, et al.* (2012). Mycobacterial disease and
919 impaired IFN-γ immunity in humans with inherited ISG15 deficiency. Science *337*, 1684-
920 1688.

921 Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S.,
922 and Churchill, G.A. (2019). R/qtl2: Software for Mapping Quantitative Trait Loci with
923 High-Dimensional Data and Multiparent Populations. Genetics *211*, 495-502.

924 Bustamante, J., Boisson-Dupuis, S., Abel, L., and Casanova, J.L. (2014). Mendelian
925 susceptibility to mycobacterial disease: genetic, immunological, and clinical features of
926 inborn errors of IFN-γ immunity. Semin Immunol *26*, 454-470.

927 Caruso, A.M., Serbina, N., Klein, E., Triebold, K., Bloom, B.R., and Flynn, J.L. (1999).
928 Mice deficient in CD4 T cells have only transiently diminished levels of IFN-gamma, yet
929 succumb to tuberculosis. J Immunol *162*, 5407-5416.

930 Caws, M., Thwaites, G., Dunstan, S., Hawn, T.R., Lan, N.T., Thuong, N.T.,
931 Stepniewska, K., Huyen, M.N., Bang, N.D., Loc, T.H.*, et al.* (2008). The influence of
932 host and bacterial genotype on the development of disseminated disease with
933 Mycobacterium tuberculosis. PLoS Pathog *4*, e1000034.

934 Chen, L., Xie, Q.W., and Nathan, C. (1998). Alkyl hydroperoxide reductase subunit C
935 (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. Mol
936 Cell *1*, 795-805.

937 Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis,
938 W.D., Belknap, J.K., Bennett, B., Berrettini, W.*, et al.* (2004). The Collaborative Cross, a
939 community resource for the genetic analysis of complex traits. Nat Genet *36*, 1133-
940 1137.

941 Churchill, G.A., Gatti, D.M., Munger, S.C., and Svenson, K.L. (2012). The Diversity
942 Outbred mouse population. Mamm Genome *23*, 713-718.

943 Comstock, G.W. (1978). Tuberculosis in twins: a re-analysis of the Prophit survey. Am
944 Rev Respir Dis *117*, 621-624.

945 Cooke, G.S., Campbell, S.J., Bennett, S., Lienhardt, C., McAdam, K.P., Sirugo, G.,
946 Sow, O., Gustafson, P., Mwangulu, F., van Helden, P.*, et al.* (2008). Mapping of a novel
947 susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis. Am J
948 Respir Crit Care Med *178*, 203-207.

949 Cooper, A.M., Dalton, D.K., Stewart, T.A., Griffin, J.P., Russell, D.G., and Orme, I.M.
950 (1993). Disseminated tuberculosis in interferon gamma gene-disrupted mice. J Exp Med
951 *178*, 2243-2247.

952 Cooper, A.M., Magram, J., Ferrante, J., and Orme, I.M. (1997). Interleukin 12 (IL-12) is
953 crucial to the development of protective immunity in mice intravenously infected with
954 mycobacterium tuberculosis. J Exp Med *186*, 39-45.

955 de Keijzer, J., de Haas, P.E., de Ru, A.H., van Veelen, P.A., and van Soolingen, D.
956 (2014). Disclosure of selective advantages in the "modern" sublineage of the
957 Mycobacterium tuberculosis Beijing genotype family by quantitative proteomics. Mol
958 Cell Proteomics *13*, 2632-2645.

959 DeJesus, M.A., Ambadipudi, C., Baker, R., Sassetti, C., and Ioerger, T.R. (2015).
960 TRANSIT--A Software Tool for Himar1 TnSeq Analysis. PLoS Comput Biol *11*,
961 e1004401.

962 Dick, T., Manjunatha, U., Kappes, B., and Gengenbacher, M. (2010). Vitamin B6
963 biosynthesis is essential for survival and virulence of Mycobacterium tuberculosis. Mol
964 Microbiol *78*, 980-988.

965 Ferris, M.T., Aylor, D.L., Bottomly, D., Whitmore, A.C., Aicher, L.D., Bell, T.A., Bradel-
966 Tretheway, B., Bryan, J.T., Buus, R.J., Gralinski, L.E.*, et al.* (2013). Modeling host
967 genetic regulation of influenza pathogenesis in the collaborative cross. PLoS Pathog *9*,
968 e1003196.

969 Filipe-Santos, O., Bustamante, J., Haverkamp, M.H., Vinolo, E., Ku, C.L., Puel, A.,
970 Frucht, D.M., Christel, K., von Bernuth, H., Jouanguy, E.*, et al.* (2006). X-linked
971 susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-
972 dependent IL-12 production. J Exp Med *203*, 1745-1759.

973 Flynn, J.L., Chan, J., Triebold, K.J., Dalton, D.K., Stewart, T.A., and Bloom, B.R. (1993).
974 An essential role for interferon gamma in resistance to Mycobacterium tuberculosis
975 infection. J Exp Med *178*, 2249-2254.

976 Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B.C., Narayanan, S.,
977 Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C.*, et al.* (2006). Variable host-
978 pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *103*,
979 2869-2873.

980 Geiman, D.E., Kaushal, D., Ko, C., Tyagi, S., Manabe, Y.C., Schroeder, B.G.,
981 Fleischmann, R.D., Morrison, N.E., Converse, P.J., Chen, P.*, et al.* (2004). Attenuation
982 of late-stage disease in mice infected by the Mycobacterium tuberculosis mutant lacking
983 the SigF alternate sigma factor and identification of SigF-dependent genes by
984 microarray analysis. Infect Immun *72*, 1733-1745.

985 Goitre, L., Trapani, E., Trabalzini, L., and Retta, S.F. (2014). The Ras superfamily of
986 small GTPases: the unlocked secrets. Methods Mol Biol *1120*, 1-18.

987 Gonzalo-Asensio, J., Malaga, W., Pawlik, A., Astarie-Dequeker, C., Passemar, C.,
988 Moreau, F., Laval, F., Daffé, M., Martin, C., Brosch, R.*, et al.* (2014). Evolutionary
989 history of tuberculosis shaped by conserved mutations in the PhoPR virulence
990 regulator. Proc Natl Acad Sci U S A *111*, 11491-11496.

991 Gopal, R., Lin, Y., Obermajer, N., Slight, S., Nuthalapati, N., Ahmed, M., Kalinski, P.,
992 and Khader, S.A. (2012). IL-23-dependent IL-17 drives Th1-cell responses following
993 Mycobacterium bovis BCG vaccination. European journal of immunology *42*, 364 - 373.

994 Gopal, R., Monin, L., Torres, D., Slight, S., Mehra, S., McKenna, K.C., Fallert Junecko,
995 B.A., Reinhart, T.A., Kolls, J., Báez-Saldaña, R.*, et al.* (2013). S100A8/A9 proteins
996 mediate neutrophilic inflammation and lung pathology during tuberculosis. Am J Respir
997 Crit Care Med *188*, 1137-1146.

998 Gralinski, L.E., Menachery, V.D., Morgan, A.P., Totura, A.L., Beall, A., Kocher, J.,
999 Plante, J., Harrison-Shostak, D.C., Schäfer, A., Pardo-Manuel de Villena, F.*, et al.*
1000 (2017). Allelic Variation in the Toll-Like Receptor Adaptor Protein Ticam2 Contributes to
1001 SARS-Coronavirus Pathogenesis in Mice. G3 (Bethesda) *7*, 1653-1663.

1002 Hillas, P.J., del Alba, F.S., Oyarzabal, J., Wilks, A., and Ortiz De Montellano, P.R.
1003 (2000). The AhpC and AhpD antioxidant defense system of Mycobacterium
1004 tuberculosis. J Biol Chem *275*, 18801-18809.

1005 Holt, K.E., McAdam, P., Thai, P.V.K., Thuong, N.T.T., Ha, D.T.M., Lan, N.N., Lan, N.H.,
1006 Nhu, N.T.Q., Hai, H.T., Ha, V.T.N.*, et al.* (2018). Frequent transmission of the
1007 Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing
1008 variant in Vietnam. Nat Genet *50*, 849-856.

1009 Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A.,
1010 Agam, A., Slater, G., Goodson, M.*, et al.* (2011). Mouse genomic variation and its effect
1011 on phenotypes and gene regulation. Nature *477*, 289-294.

1012 Keller, M.P., Gatti, D.M., Schueler, K.L., Rabaglia, M.E., Stapleton, D.S., Simecek, P.,
1013 Vincent, M., Allen, S., Broman, A.T., Bacher, R.*, et al.* (2018). Genetic Drivers of
1014 Pancreatic Islet Function. Genetics *209*, 335-356.

1015 Khader, S.A., Bell, G.K., Pearl, J.E., Fountain, J.J., Rangel-Moreno, J., Cilley, G.E.,
1016 Shen, F., Eaton, S.M., Gaffen, S.L., Swain, S.L.*, et al.* (2007). IL-23 and IL-17 in the
1017 establishment of protective pulmonary CD4+ T cell responses after vaccination and
1018 during Mycobacterium tuberculosis challenge. Nat Immunol *8*, 369-377.

1019 Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation
1020 network analysis. Bmc Bioinformatics *9*, 559.

1021 Lázár-Molnár, E., Chen, B., Sweeney, K.A., Wang, E.J., Liu, W., Lin, J., Porcelli, S.A.,
1022 Almo, S.C., Nathenson, S.G., and Jacobs, W.R., Jr. (2010). Programmed death-1 (PD-
1023 1)-deficient mice are extraordinarily sensitive to tuberculosis. Proc Natl Acad Sci U S A
1024 *107*, 13402-13407.

1025 Le Chevalier, F., Cascioferro, A., Frigui, W., Pawlik, A., Boritsch, E.C., Bottai, D.,
1026 Majlessi, L., Herrmann, J.L., and Brosch, R. (2015). Revisiting the role of
1027 phospholipases C in virulence and the lifecycle of Mycobacterium tuberculosis. Sci Rep
1028 *5*, 16918.

1029 Lin, P.L., Ford, C.B., Coleman, M.T., Myers, A.J., Gawande, R., Ioerger, T., Sacchettini,
1030 J., Fortune, S.M., and Flynn, J.L. (2014). Sterilization of granulomas is common in
1031 active and latent tuberculosis despite within-host variability in bacterial killing. Nat Med
1032 *20*, 75-79.

1033 Long, J.E., DeJesus, M., Ward, D., Baker, R.E., Ioerger, T., and Sassetti, C.M. (2015).
1034 Identifying essential genes in Mycobacterium tuberculosis by global phenotypic profiling.
1035 Methods Mol Biol *1279*, 79-95.

1036 Long, M., Kranjc, T., Mysior, M.M., and Simpson, J.C. (2020). RNA Interference
1037 Screening Identifies Novel Roles for RhoBTB1 and RhoBTB3 in Membrane Trafficking
1038 Events in Mammalian Cells. Cells *9*.

1039 Lu, L.L., Smith, M.T., Yu, K.K.Q., Luedemann, C., Suscovich, T.J., Grace, P.S., Cain,
1040 A., Yu, W.H., McKitrick, T.R., Lauffenburger, D.*, et al.* (2019). IFN-γ-independent
1041 immune markers of Mycobacterium tuberculosis exposure. Nat Med *25*, 977-987.

1042 Martin, C.J., Cadena, A.M., Leung, V.W., Lin, P.L., Maiello, P., Hicks, N., Chase, M.R.,
1043 Flynn, J.L., and Fortune, S.M. (2017). Digitally Barcoding Mycobacterium tuberculosis
1044 Reveals In Vivo Infection Dynamics in the Macaque Model of Tuberculosis. mBio *8*.

1045 McHenry, M.L., Williams, S.M., and Stein, C.M. (2020). Genetics and evolution of
1046 tuberculosis pathogenesis: New perspectives and approaches. Infect Genet Evol *81*,
1047 104204.

1048 Minch, K.J., Rustad, T.R., Peterson, E.J., Winkler, J., Reiss, D.J., Ma, S., Hickey, M.,
1049 Brabant, W., Morrison, B., Turkarslan, S.*, et al.* (2015). The DNA-binding network of
1050 Mycobacterium tuberculosis. Nat Commun *6*, 5829.

1051 Mishra, B.B., Lovewell, R.R., Olive, A.J., Zhang, G.L., Wang, W.F., Eugenin, E., Smith,
1052 C.M., Phuah, J.Y., Long, J.E., Dubuke, M.L.*, et al.* (2017). Nitric oxide prevents a
1053 pathogen-permissive granulocytic inflammation during tuberculosis. Nature Microbiology
1054 *2*.

1055 Mishra, B.B., Rathinam, V.A., Martens, G.W., Martinot, A.J., Kornfeld, H., Fitzgerald,
1056 K.A., and Sassetti, C.M. (2013). Nitric oxide controls the immunopathology of
1057 tuberculosis by inhibiting NLRP3 inflammasome-dependent processing of IL-1β. Nat
1058 Immunol *14*, 52-60.

1059 Moreira-Teixeira, L., Tabone, O., Graham, C.M., Singhania, A., Stavropoulos, E.,
1060 Redford, P.S., Chakravarty, P., Priestnall, S.L., Suarez-Bonnet, A., Herbert, E.*, et al.*
1061 (2020). Mouse transcriptome reveals potential signatures of protection and
1062 pathogenesis in human tuberculosis. Nat Immunol *21*, 464-476.

1063 Morgan, A.P., Fu, C.P., Kao, C.Y., Welsh, C.E., Didion, J.P., Yadgary, L., Hyacinth, L.,
1064 Ferris, M.T., Bell, T.A., Miller, D.R.*, et al.* (2015). The Mouse Universal Genotyping
1065 Array: From Substrains to Subspecies. G3 (Bethesda) *6*, 263-279.

1066 Morth, J.P., Gosmann, S., Nowak, E., and Tucker, P.A. (2005). A novel two-component
1067 system found in Mycobacterium tuberculosis. FEBS Lett *579*, 4145-4148.

1068 Muñoz-Elías, E.J., Timm, J., Botha, T., Chan, W.T., Gomez, J.E., and McKinney, J.D.
1069 (2005). Replication dynamics of Mycobacterium tuberculosis in chronically infected
1070 mice. Infect Immun *73*, 546-551.

1071 Murphy, K.C., Nelson, S.J., Nambi, S., Papavinasasundaram, K., Baer, C.E., and
1072 Sassetti, C.M. (2018). ORBIT: a New Paradigm for Genetic Engineering of
1073 Mycobacterial Chromosomes. mBio *9*.

1074 Nambi, S., Long, J.E., Mishra, B.B., Baker, R., Murphy, K.C., Olive, A.J., Nguyen, H.P.,
1075 Shaffer, S.A., and Sassetti, C.M. (2015). The Oxidative Stress Network of
1076 Mycobacterium tuberculosis Reveals Coordination between Radical Detoxification
1077 Systems. Cell Host Microbe *17*, 829-837.

1078 Neto, E.C., Keller, M.P., Broman, A.F., Attie, A.D., Jansen, R.C., Broman, K.W., and
1079 Yandell, B.S. (2012). Quantile-based permutation thresholds for quantitative trait loci
1080 hotspots. Genetics *191*, 1355-1365.

1081 Niazi, M.K., Dhulekar, N., Schmidt, D., Major, S., Cooper, R., Abeijon, C., Gatti, D.M.,
1082 Kramnik, I., Yener, B., Gurcan, M.*, et al.* (2015). Lung necrosis and neutrophils reflect
1083 common pathways of susceptibility to Mycobacterium tuberculosis in genetically
1084 diverse, immune-competent mice. Dis Model Mech *8*, 1141-1153.

1085 Noll, K.E., Ferris, M.T., and Heise, M.T. (2019). The Collaborative Cross: A Systems
1086 Genetics Resource for Studying Host-Pathogen Interactions. Cell Host Microbe *25*, 484-
1087 498.

1088 Noll, K.E., Whitmore, A.C., West, A., McCarthy, M.K., Morrison, C.R., Plante, K.S.,
1089 Hampton, B.K., Kollmus, H., Pilzner, C., Leist, S.R.*, et al.* (2020). Complex Genetic
1090 Architecture Underlies Regulation of Influenza-A-Virus-Specific Antibody Responses in
1091 the Collaborative Cross. Cell Rep *31*, 107587.

1092 Obermajer, N., Svajger, U., Bogyo, M., Jeras, M., and Kos, J. (2008). Maturation of
1093 dendritic cells depends on proteolytic cleavage by cathepsin X. J Leukoc Biol *84*, 1306-
1094 1315.

1095 Olive, A.J., Smith, C.M., Kiritsy, M.C., and Sassetti, C.M. (2018). The Phagocyte
1096 Oxidase Controls Tolerance to Mycobacterium tuberculosis Infection. J Immunol *201*,
1097 1705-1716.

1098 Palace, S.G., Proulx, M.K., Lu, S., Baker, R.E., and Goguen, J.D. (2014). Genome-wide
1099 mutant fitness profiling identifies nutritional requirements for optimal growth of Yersinia
1100 pestis in deep tissue. mBio *5*.

1101 Pandey, A.K., and Sassetti, C.M. (2008). Mycobacterial persistence requires the
1102 utilization of host cholesterol. Proc Natl Acad Sci U S A *105*, 4376-4380.

1103 Pethe, K., Sequeira, P.C., Agarwalla, S., Rhee, K., Kuhen, K., Phong, W.Y., Patel, V.,
1104 Beer, D., Walker, J.R., Duraiswamy, J.*, et al.* (2010). A chemical genetic screen in
1105 Mycobacterium tuberculosis identifies carbon-source-dependent growth inhibitors
1106 devoid of in vivo efficacy. Nat Commun *1*, 57.

1107 Rodrigue, S., Provvedi, R., Jacques, P.E., Gaudreau, L., and Manganelli, R. (2006).
1108 The sigma factors of Mycobacterium tuberculosis. FEMS Microbiol Rev *30*, 926-941.

1109 Ryndak, M.B., Wang, S., Smith, I., and Rodriguez, G.M. (2010). The Mycobacterium
1110 tuberculosis high-affinity iron importer, IrtA, contains an FAD-binding domain. J
1111 Bacteriol *192*, 861-869.

1112 Sakai, S., Kauffman, K.D., Sallin, M.A., Sharpe, A.H., Young, H.A., Ganusov, V.V., and
1113 Barber, D.L. (2016). CD4 T Cell-Derived IFN-γ Plays a Minimal Role in Control of
1114 Pulmonary Mycobacterium tuberculosis Infection and Must Be Actively Repressed by
1115 PD-1 to Prevent Lethal Disease. PLoS Pathog *12*, e1005667.

1116 Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial
1117 growth defined by high density mutagenesis. Mol Microbiol *48*, 77-84.

1118 Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival
1119 during infection. Proc Natl Acad Sci U S A *100*, 12989-12994.

1120 Saul, M.C., Philip, V.M., Reinholdt, L.G., and Chesler, E.J. (2019). High-Diversity Mouse
1121 Populations for Complex Traits. Trends Genet *35*, 501-514.

1122 Saunders, B.M., Frank, A.A., Orme, I.M., and Cooper, A.M. (2002). CD4 is required for
1123 the development of a protective granulomatous response to pulmonary tuberculosis.
1124 Cell Immunol *216*, 65-72.

1125 Shorter, J.R., Najarian, M.L., Bell, T.A., Blanchard, M., Ferris, M.T., Hock, P., Kashfeen,
1126 A., Kirchoff, K.E., Linnertz, C.L., Sigmon, J.S.*, et al.* (2019). Whole Genome
1127 Sequencing and Progress Toward Full Inbreeding of the Mouse Collaborative Cross
1128 Population. G3 (Bethesda) *9*, 1303-1311.

1129 Sigmon, J.S., Blanchard, M.W., Baric, R.S., Bell, T.A., Brennan, J., Brockmann, G.A.,
1130 Burks, A.W., Calabrese, J.M., Caron, K.M., Cheney, R.E.*, et al.* (2020). Content and
1131 performance of the MiniMUGA genotyping array, a new tool to improve rigor and
1132 reproducibility in mouse research. bioRxiv, 2020.2003.2012.989400.

1133 Smith, C.M., Proulx, M.K., Lai, R., Kiritsy, M.C., Bell, T.A., Hock, P., Pardo-Manuel de
1134 Villena, F., Ferris, M.T., Baker, R.E., Behar, S.M.*, et al.* (2019). Functionally
1135 Overlapping Variants Control Tuberculosis Susceptibility in Collaborative Cross Mice.
1136 mBio *10*.

1137 Smith, C.M., Proulx, M.K., Olive, A.J., Laddy, D., Mishra, B.B., Moss, C., Gutierrez,
1138 N.M., Bellerose, M.M., Barreira-Silva, P., Phuah, J.Y.*, et al.* (2016). Tuberculosis
1139 Susceptibility and Vaccine Protection Are Independently Controlled by Host Genotype.
1140 mBio *7*.

1141 Smith, C.M., and Sassetti, C.M. (2018). Modeling Diversity: Do Homogeneous
1142 Laboratory Strains Limit Discovery? Trends Microbiol *26*, 892-895.

1143 Sousa, J., Cá, B., Maceiras, A.R., Simões-Costa, L., Fonseca, K.L., Fernandes, A.I.,
1144 Ramos, A., Carvalho, T., Barros, L., Magalhães, C.*, et al.* (2020). Mycobacterium

1145 tuberculosis associated with severe tuberculosis evades cytosolic surveillance systems
1146 and modulates IL-1β production. Nat Commun *11*, 1949.

1147 Springer, B., Master, S., Sander, P., Zahrt, T., McFalone, M., Song, J.,
1148 Papavinasasundaram, K.G., Colston, M.J., Boettger, E., and Deretic, V. (2001).
1149 Silencing of oxidative stress response in Mycobacterium tuberculosis: expression
1150 patterns of ahpC in virulent and avirulent strains and effect of ahpC inactivation. Infect
1151 Immun *69*, 5967-5973.

1152 Srivastava, A., Morgan, A.P., Najarian, M.L., Sarsani, V.K., Sigmon, J.S., Shorter, J.R.,
1153 Kashfeen, A., McMullan, R.C., Williams, L.H., Giusti-Rodríguez, P.*, et al.* (2017).
1154 Genomes of the Mouse Collaborative Cross. Genetics *206*, 537-556.

1155 Stanley, S.A., Raghavan, S., Hwang, W.W., and Cox, J.S. (2003). Acute infection and
1156 macrophage subversion by Mycobacterium tuberculosis require a specialized secretion
1157 system. Proceedings of the National Academy of Sciences *100*, 13001-13006.

1158 Svenson, K.L., Gatti, D.M., Valdar, W., Welsh, C.E., Cheng, R., Chesler, E.J., Palmer,
1159 A.A., McMillan, L., and Churchill, G.A. (2012). High-resolution genetic mapping using
1160 the Mouse Diversity outbred population. Genetics *190*, 437-447.

1161 Thuong, N.T., Tram, T.T., Dinh, T.D., Thai, P.V., Heemskerk, D., Bang, N.D., Chau,
1162 T.T., Russell, D.G., Thwaites, G.E., Hawn, T.R.*, et al.* (2016). MARCO variants are
1163 associated with phagocytosis, pulmonary tuberculosis susceptibility and Beijing lineage.
1164 Genes Immun *17*, 419-425.

1165 Turkarslan, S., Peterson, E.J., Rustad, T.R., Minch, K.J., Reiss, D.J., Morrison, R., Ma,
1166 S., Price, N.D., Sherman, D.R., and Baliga, N.S. (2015). A comprehensive map of
1167 genome-wide gene regulation in Mycobacterium tuberculosis. Sci Data *2*, 150010.

1168 Wang, Q., Boshoff, H.I.M., Harrison, J.R., Ray, P.C., Green, S.R., Wyatt, P.G., and
1169 Barry, C.E., 3rd (2020). PE/PPE proteins mediate nutrient transport across the outer
1170 membrane of Mycobacterium tuberculosis. Science *367*, 1147-1151.

1171 Woong Park, S., Klotzsche, M., Wilson, D.J., Boshoff, H.I., Eoh, H., Manjunatha, U.,
1172 Blumenthal, A., Rhee, K., Barry, C.E., 3rd, Aldrich, C.C.*, et al.* (2011). Evaluating the
1173 sensitivity of Mycobacterium tuberculosis to biotin deprivation using regulated gene
1174 expression. PLoS Pathog *7*, e1002264.

1175 Zak, D.E., Penn-Nicholson, A., Scriba, T.J., Thompson, E., Suliman, S., Amon, L.M.,
1176 Mahomed, H., Erasmus, M., Whatney, W., Hussey, G.D.*, et al.* (2016). A blood RNA
1177 signature for tuberculosis disease risk: a prospective cohort study. Lancet *387*, 2312-
1178 2322.

1179 Zhang, J., Teh, M., Kim, J., Eva, M.M., Cayrol, R., Meade, R., Nijnik, A., Montagutelli,
1180 X., Malo, D., and Jaubert, J. (2019). A Loss-of-Function Mutation in the Integrin Alpha L
1181 (Itgal) Gene Contributes to Susceptibility to Salmonella enterica Serovar Typhimurium
1182 Infection in Collaborative Cross Strain CC042. Infect Immun *88*.

1183 Zhang, Y.J., Reddy, M.C., Ioerger, T.R., Rothchild, A.C., Dartois, V., Schuster, B.M.,
1184 Trauner, A., Wallis, D., Galaviz, S., Huttenhower, C.*, et al.* (2013). Tryptophan
1185 Biosynthesis Protects Mycobacteria from CD4 T-Cell-Mediated Killing. Cell *155*, 1296 -
1186 1308.
1187

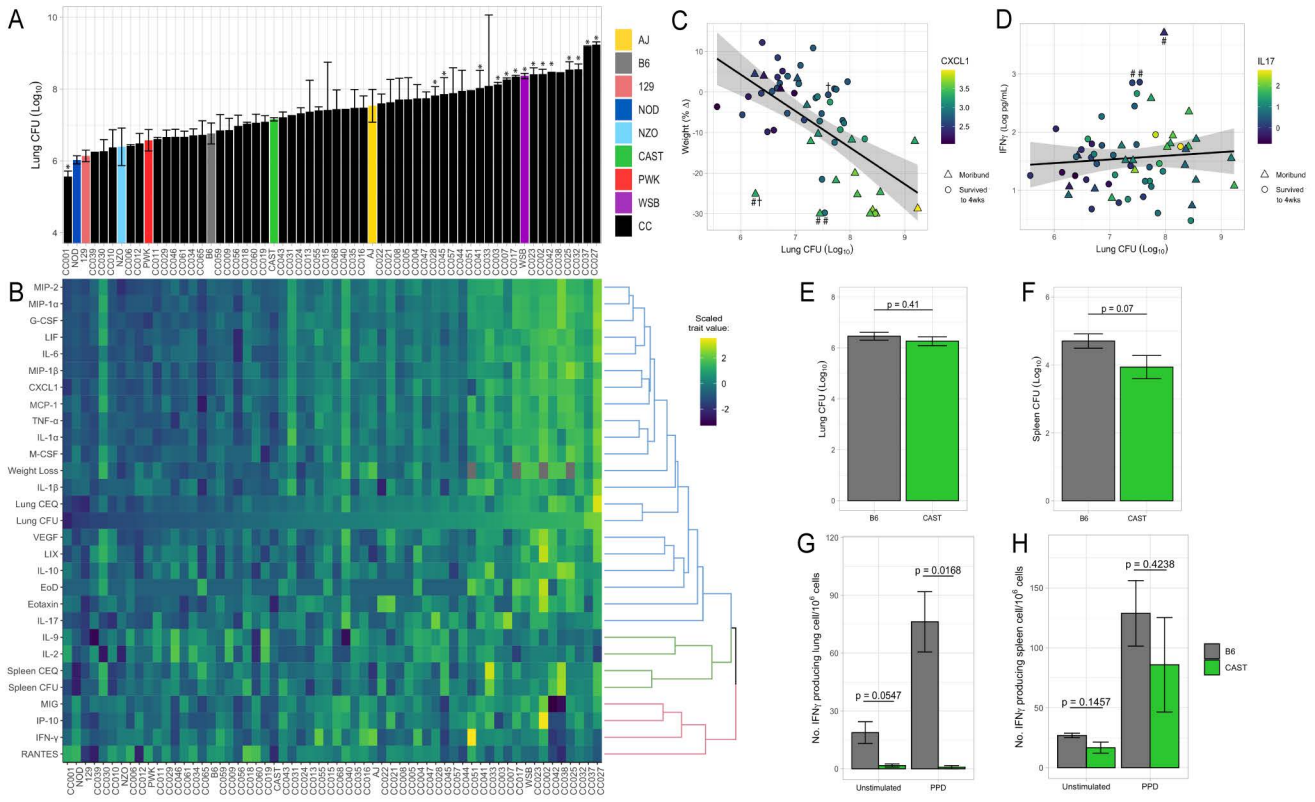

Figure 1. The spectrum of M. tuberculosis disease-related traits across the Collaborative Cross. (A) Average lung CFU (log10) across the CC panel at 4 weeks post-infection. Bars show mean +/- SD for CFU per CC or parental strain; groups of 2-6 (average of n=3) mice per genotype. Bars noted with * indicate strains that were statistically different from B6 (P<0.05; 1-factor ANOVA with Dunnett's post-test). (B) Heatmap of the 32 disease-related traits (log10 transformed) measured including: lung and spleen colony forming units (CFU); lung and spleen chromosomal equivalents (CEQ); weight loss (% change); cytokines from lung, "earliness of death" (EoD) reflecting the number of days prior to the end of experiment that moribund strains were euthanized. Mouse genotypes are ordered by lung CFU. Scaled trait values were clustered (hclust in R package heatmaply) and dendrogram nodes colored by 3 k-means. Blue node reflects correlation coefficient r>0.7; green r=0.3-0.6 and red r<0.2. (C) Correlation of lung CFU and weight (% change) shaded by CXCL1 levels. Genotypes identified as statistical outliers for weight are noted by #; CXCL1 by †. (D) Correlation of lung CFU and IFNγ levels shaded by IL-17. Strains identified as outliers for IFNγ noted by #. Each point in (C) and (D) is the average value per genotype. Outlier genotypes were identified after linear regression using studentized residuals. (E-H) B6 and CAST traits measured at 4 weeks post aerosol infection (E) lung CFU; (F) spleen CFU; (G) number of IFNγ producing lung T cells; (H) number of IFNγ producing spleen T cells. Bar plots show the mean +/- SD. Significance between groups was determined by Welch's t test. Groups consist of 4-10 mice per genotype from 2 independent experiments. All mice in the initial CC screen were male; validation studies in panels E-H included both sexes, where no statistical difference at the 4-week timepoint between sexes was observed.

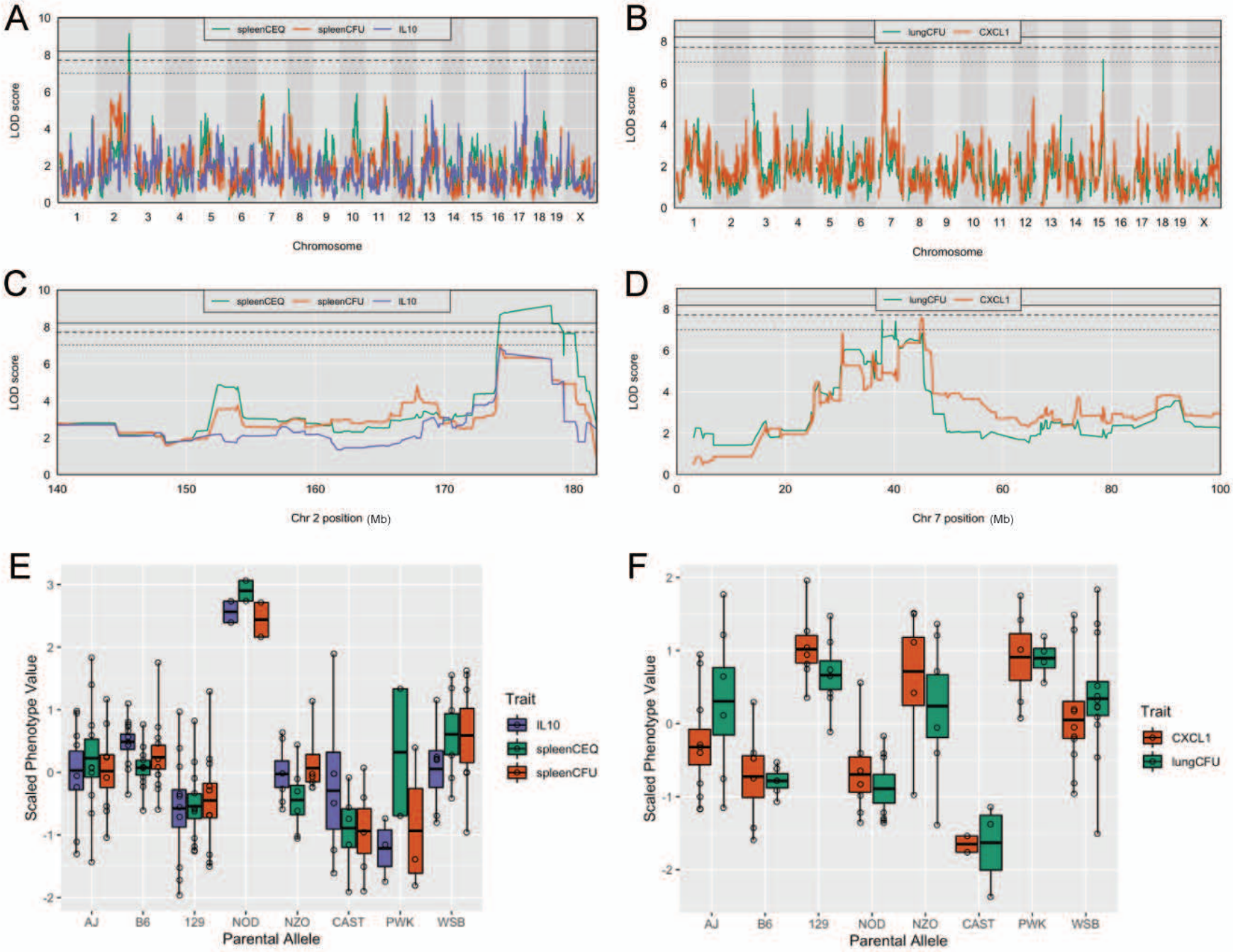

Figure 2. Host loci underlying TB disease-related traits. (A-B) Whole genome QTL scans of (A) spleen CEQ, spleen CFU and IL-10 (B) lung CFU and CXCL1. (C) Zoom of chromosome 2 loci. (D) Zoom of chromosome 7 loci. Thresholds were determined by permutation analysis; solid line, middle dashed line, and lowest dotted lines represent P = 0.05, P = 0.1, and P = 0.2. (E-F) Scaled phenotype value per haplotype at the QTL peak marker. Each dot represents the mean value for a genotype.

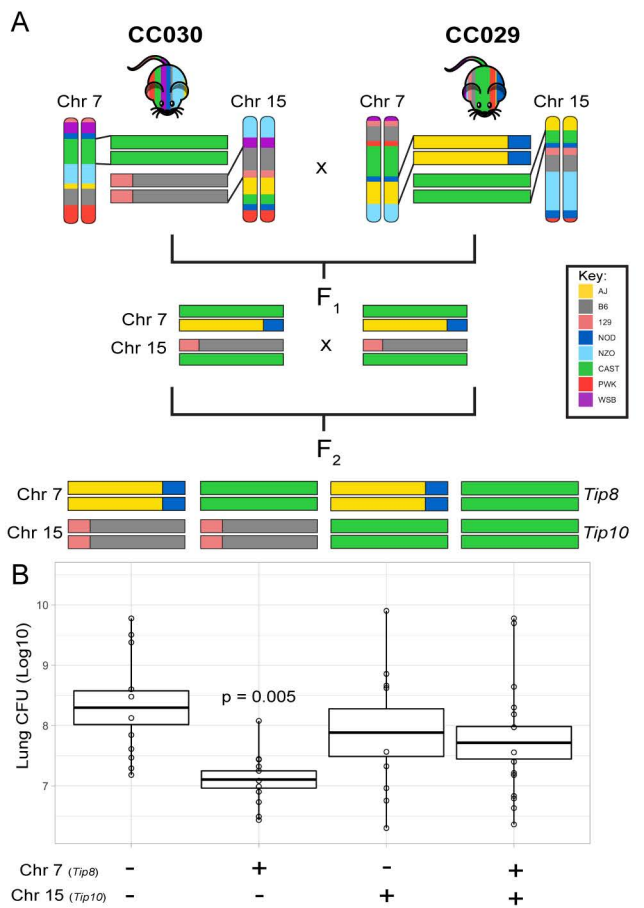

Figure 3. An F2 intercross approach to validate QTL underlying lung CFU. (A) Cross schema between CC030 and CC029 CC strains that contain CAST informative haplotypes at Chr 7 or 15. CC030 contains CAST at 30-45Mb on Chr 7 and CC029 contains CAST at 77-79Mb on Chr 15. The F2 population based on these founders were genotyped and informative segregants were infected with Mtb. (B) Lung CFU at 1-month post-infection. F2 mice are grouped by the four possible genotypes: CAST at both 7 and 15, CAST at 7, CAST at 15, or non-CAST at both loci. To determine the effect of CAST haplotype on lung CFU, data was analyzed by 2-factor ANOVA to determine statistical significance.

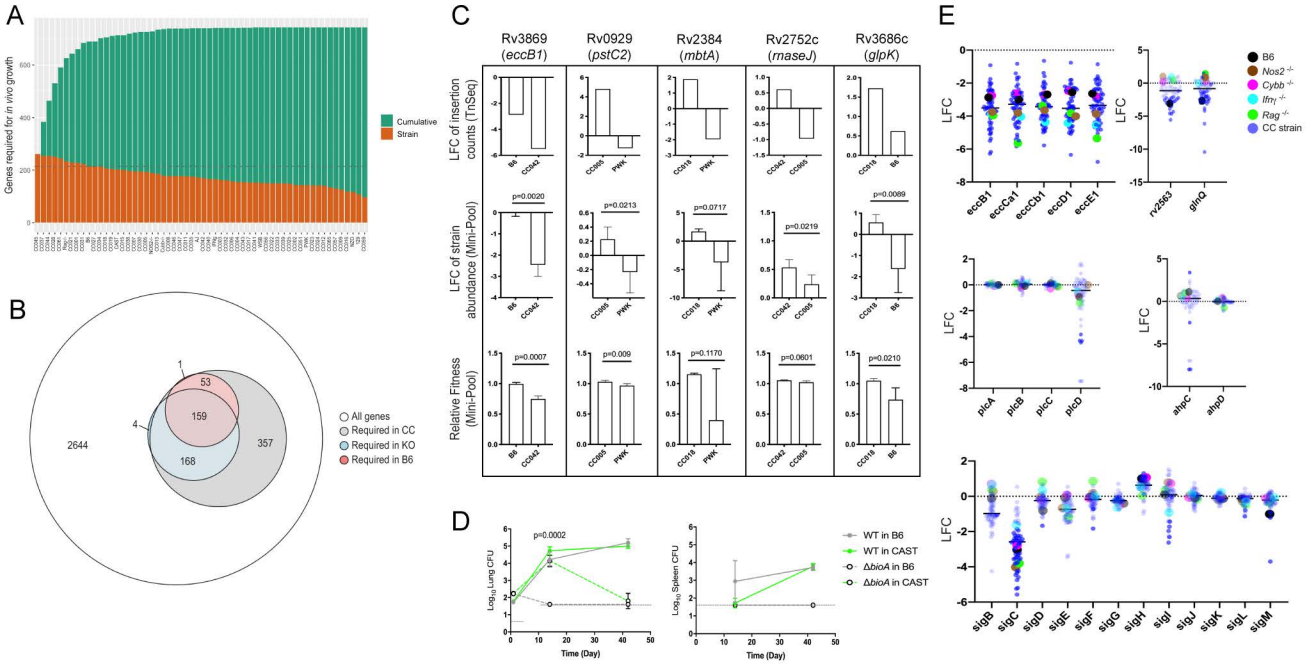

Figure 4. Mtb genetic requirements vary across diverse hosts. (A) The number of Mtb genes required for growth or survival in each diverse mouse strain across the panel (Qval ≤ 0.05). Orange indicates the mutants required for each strain; turquoise shows the cumulative requirement as each new host strain is added. (B) Venn diagram showing the composition of Mtb gene sets required in each category of host (white, largest circle), only required in the CC panel (grey), required in specific immunological KO mice (blue) and genes required in B6 mice (red). In order to be called "essential" in each mouse strain, Mtb genes had to be significantly over or unrepresented in at least 2 genotypes. (C) Each box shows Log2 fold change (LFC) of individual mutants from the TnSeq screen relative to the input pool in indicated mouse strains (top); Log2 fold change of the indicated deletion mutants relative to WT from a pooled mutant validation infection (middle panel); Relative fitness calculated from (middle panel) to account for generation differences in each host due to differential growth rate. Bars are the average of 3-6 mice per mutant/genotype +/- SD. Statistical differences between mini-pool validation groups was assessed by Welch's t test. (D) Lung CFU and spleen CFU from single strain low-dose aerosol infections of ΔbioA mutant or WT H37Rv strain in B6 and CAST mice at 2- or 5-weeks post-infection. Dashed line indicates the limit of detection. Each point indicates the average CFU +/- SD of 4-5 mice per group. (E) Log2 fold change of selected mutants from the TnSeq screen across the CC panel and immunological KO mice. Each dot represents the average LFC per mouse genotype. All mice in the complete mouse panel TnSeq screen were male; mice in the ΔbioA aerosol validation were female; mice in the mini-pool validation studies were male and female with no significant differences detected.

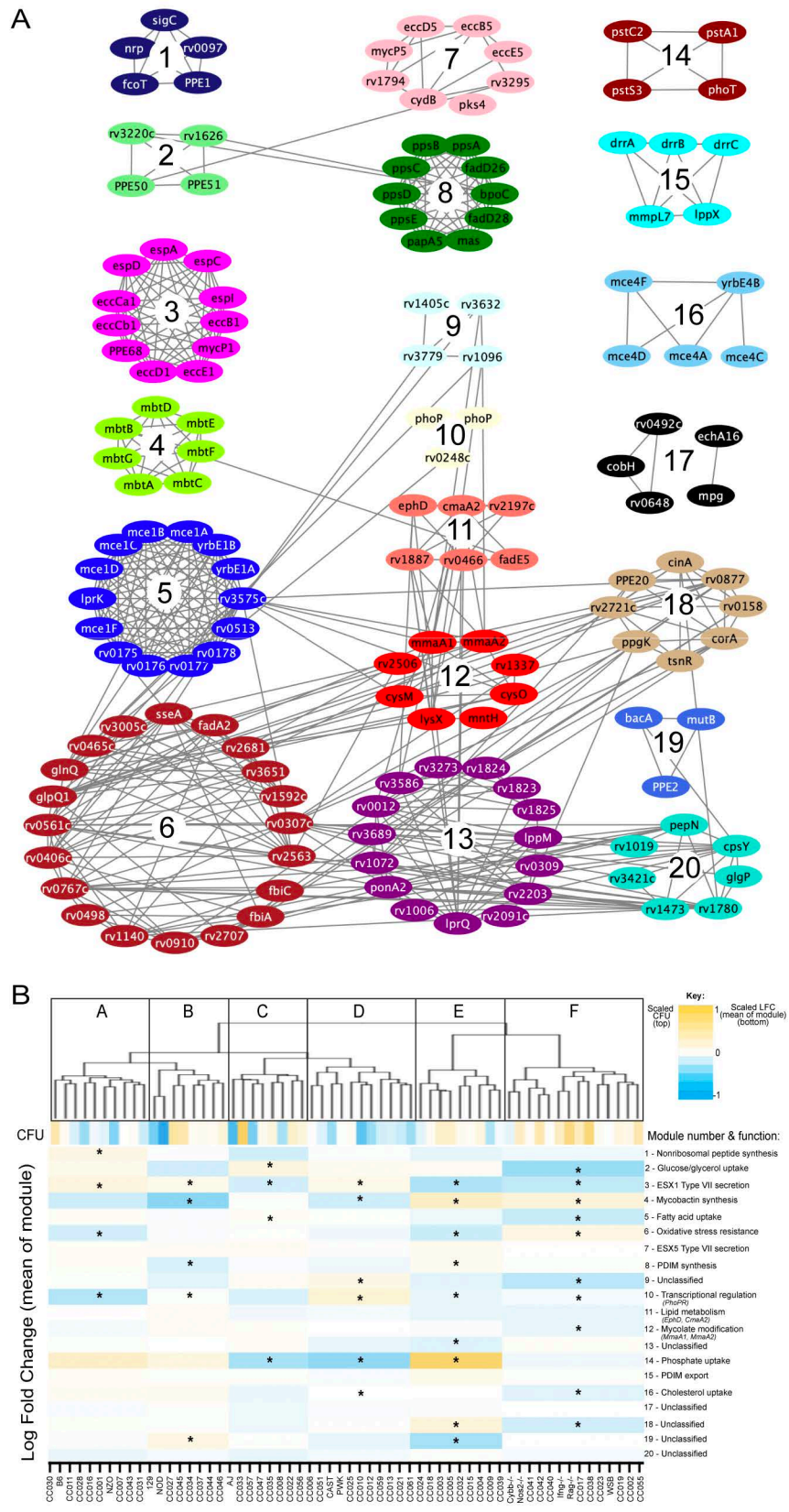

Figure 5. Mtb virulence pathways associate with distinct host immune pressures. (A) Weighted gene correlation network analysis (WGCNA) of the 679 Mtb genes that significantly vary across the diverse mouse panel. The most representative genes of each module (intramodular connectivity > 0.6) are shown. (B) Mouse genotypes were clustered based on the relative abundance (log fold change relative to *in vitro*) of the 679 variable *Mtb* mutants. The six major clusters (Cluster A-F) were associated with both CFU and the relative abundance of mutants in each bacterial module (1-20; right hand-side with known functions). Cluster F is associated with high CFU ($p<0.05$, as described in methods). * indicate modules significantly associated with specific mouse clusters ($P<0.05$).

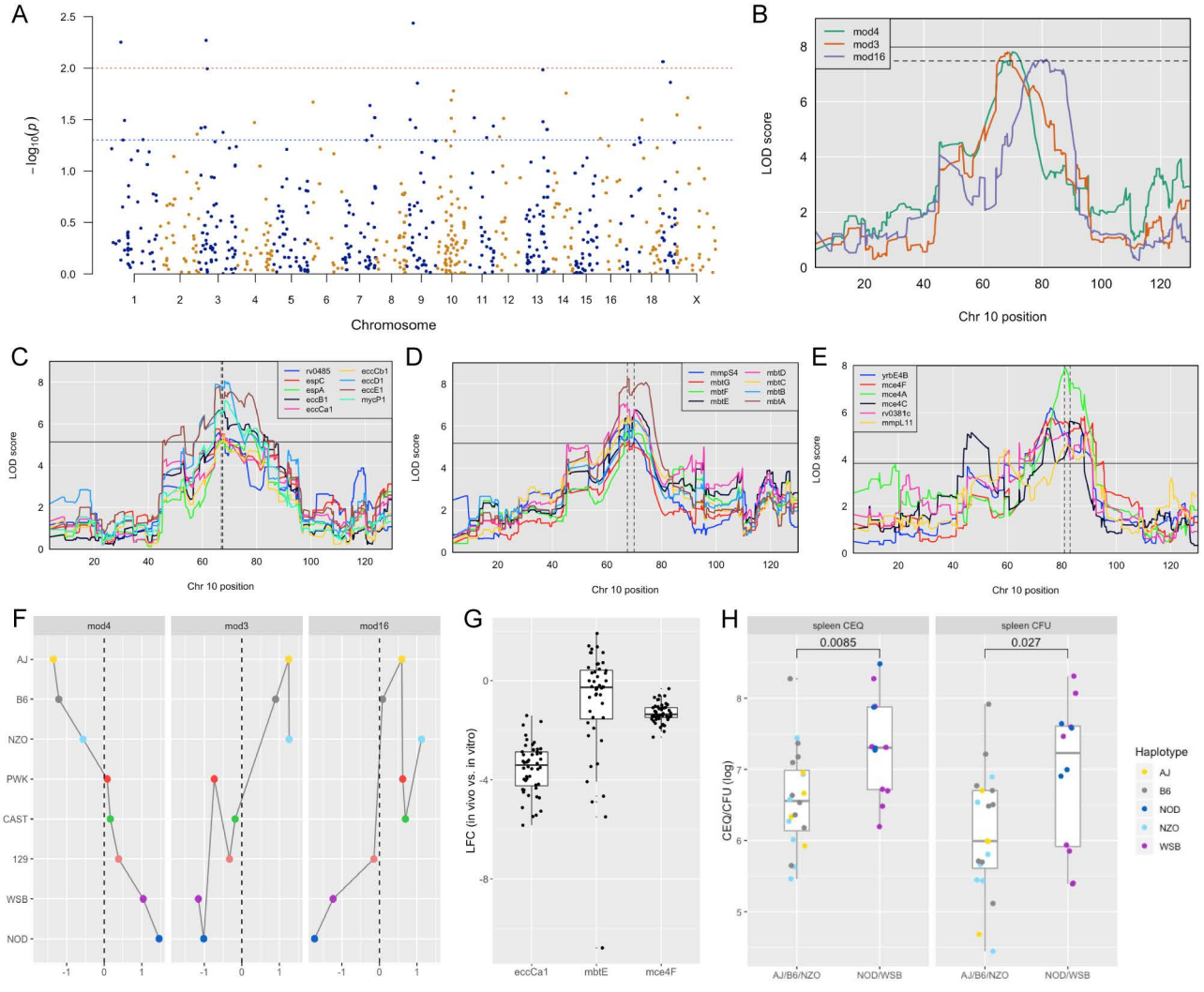

Figure 6. Identification of 'Host Interacting with Pathogen' QTL mapping (HipQTL). (A) Manhattan plot of single Mtb mutant QTL mapping across the mouse genome. Each dot represents an individual Mtb mutant plotted at the chromosomal location of its maximum LOD score. Red dashed line indicates P<0.01; Blue P<0.05. (B) Chromosome 10 QTL corresponding to Mtb eigentraits identified in network analysis in Figure 5. Module 3 (type VII secretion, ESX1 operon; orange), Module 4 (Mycobactin synthesis, mbt; green) and Module 16 (Cholesterol uptake, mce4; purple) are shown. Solid and dotted lines indicated P = 0.05 and P = 0.1, respectively. (C-E) QTL mapping of single Mtb mutants corresponding to the (C) ESX1 module, (D) mbt module and (E) mce4 modules. Coincidence of multiple QTL was assessed by the NL-method of Neto et al (Neto et al., 2012). Thresholds shown are for N=9, N=8, and N=6 for panels C, D, and E, respectively. (F) Parental founder effects underlying Module 3, 4 and 16 QTL. Allele effects were calculated at the peak LOD score marker on chromosome 10. (G) Distribution of log fold change (LFC) of representative single mutants from each module; eccCa1 (ESX1 module), mbtE (mbt module) and mce4F (mce4 module) relative to in vitro. Each dot is the LFC of the specified mutant in each CC mouse strain. Box and whiskers plots (Tukey) of each trait indicate the median and interquartile range. (H) Spleen CEQ and Spleen CFU for CC strains (box plots as in G). Mouse values are grouped by the parental haplotype allele series underlying the chromosome 10 Hip42 locus (NOD/WSB vs AJ/B6/NZO). Each dot represents the average CFU/CEQ of each CC genotype. Statistical differences in disease-associated traits and distinct haplotypes groups was assessed by t-test. LOD, logarithm of the odds; LFC, log fold change; CEQ, Chromosomal equivalents; CFU, colony forming units.