Mouse transcriptome reveals potential signatures of protection and pathogenesis in human tuberculosis

Lúcia Moreira-Teixeira[1*], Olivier Tabone[1*], Christine M. Graham[1*], Akul Singhania[1], Evangelos Stavropoulos[1], Paul S. Redford[1✦], Probir Chakravarty[2], Simon L. Priestnall[3,4], Alejandro Suarez-Bonnet[3,4], Eleanor Herbert[3,4], Katrin D. Mayer-Barber[5], Alan Sher[6], Kaori L. Fonseca[7,8,9,10], Jeremy Sousa[7,8,9], Baltazar Cá[7,8,9,10], Raman Verma[11], Pranabashis Haldar[11], Margarida Saraiva[7,8] and Anne O'Garra[1,12].

[1]Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London NW1 1AT, UK.
[2]Bioinformatics Core, The Francis Crick Institute, London NW1 1AT, UK
[3]Department of Pathobiology & Population Sciences, Royal Veterinary College, London AL9 7TA, UK
[4]Experimental Histopathology Team, The Francis Crick Institute, London NW1 1AT, UK
[5]Inflammation and Innate Immunity Unit, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda MD 20892, USA
[6]Immunobiology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
[7]i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal.
[8]IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal.
[9]ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal
[10]Programa de Pós-Graduação Ciência para o Desenvolvimento (PGCD), Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal
[11]Department of Respiratory Sciences, National Institute for Health Research Respiratory Biomedical Research Centre, University of Leicester, Leicester, United Kingdom.
[12]National Heart and Lung Institute, Imperial College London, London W2 1PG, UK.

[*]These authors contributed equally to this work.

[✦]Paul Redford's current affiliation is GSK R&D Ltd, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom.

Correspondence and requests for materials should be addressed to AOG. (email: Anne.OGarra@crick.ac.uk)

Abstract

Although mouse infection models have been extensively used to study the host response to *Mycobacterium tuberculosis*, their validity in revealing determinants of human TB resistance and disease progression has been heavily debated. Here, we show that the modular transcriptional signature in the blood of susceptible mice infected with a clinical isolate of *M. tuberculosis* resembles that of active human tuberculosis disease, with a dominance of a type I IFN response and neutrophil activation and recruitment, together with a loss in B lymphocyte, NK and T cell effector responses. In addition, resistant but not susceptible strains of mice show increased lung B, NK and T cell effector responses in the lung upon infection. Importantly, the blood signature of active disease shared by mice and humans is also evident in latent tuberculosis progressors before diagnosis suggesting that these responses both predict and contribute to the pathogenesis of progressive *M. tuberculosis* infection.

Tuberculosis (TB) results in over 1.3 million deaths annually[1], yet most individuals infected with *M. tuberculosis* remain asymptomatic. Latent TB infection (LTBI) is defined by an interferon-γ (IFN-γ)-release assay (IGRA) specific for *M. tuberculosis* antigens , although some patients may have subclinical disease and may progress to active TB[2]. Protective immune responses against *M. tuberculosis* include $CD4^+$ T lymphocytes and the cytokines IL-12, IFN-γ, TNF[3-6], and IL-1[7], but these factors do not explain why most individuals control infection, whereas a subset go on to develop active TB. A blood transcriptional signature in active TB patients has implicated type I IFN in TB pathogenesis[8-16]. Immunological heterogeneity in the blood transcriptome of a cohort of recent TB contacts has been observed, with a small proportion of contacts expressing a persistent blood TB signature and subsequently progressing to active disease (LTBI-progressors)[16], suggesting a host response evolving towards active disease[16].

How the immune response in blood[8,15] reflects that occurring at disease sites is poorly understood, and sampling the latter in humans is prohibitive. The mouse TB model, owing to the richness of genetic and immunological tools available, has been invaluable in defining immune responses in the lung influencing disease outcome

79 after infection[4,5,17]. However, a global systematic analysis to determine potential
80 common pathways of protection or pathogenesis in different TB mouse models and
81 human disease has not been reported. A role for type I IFN in TB pathogenesis[8-16] is
82 supported by mouse TB models[6,18] with elevated and sustained levels of type I IFN:
83 (i) infection of particular genetic strains of mice with clinical isolates of *M.*
84 *tuberculosis*[19-23]; (ii) infection of hosts with genetic mutations in regulators of type I
85 IFN such as Tpl2[24]; IL-1[7] or ISG15[25]; (iii) administration of adjuvants, e.g.
86 Poly(I)C[7,26]; or (iv) viral co-infection[27]. Whether it is the genetic strain of mouse or the
87 *M. tuberculosis* pathogen itself which results in an immune response that most
88 resembles human TB is unclear. Although the spectra of human[28] and mouse[29] TB
89 disease do not completely overlap, comparison of human TB with genetically diverse
90 backgrounds of mice has established points of similarity in their response to *M.*
91 *tuberculosis*. Some mouse strains recapitulate key elements of the pathogenesis of
92 human TB disease, at the level of induction of necrotic TB lesions in the lungs[29].
93 Whether the global immune response to *M. tuberculosis* in susceptible mouse strains
94 resembles that of TB in humans is as yet unclear.
95 Here, we report that the human blood TB type I IFN-inducible signature[16,8] is
96 recapitulated in susceptible C3HeB/FeJ mice infected with different strains of *M.*
97 *tuberculosis*. Increased expression of granulocyte-associated genes in blood from
98 active TB patients, TB-susceptible mice and LTBI-progressors before TB diagnosis
99 suggested their role in early disease pathogenesis. Conversely, under-abundance of
100 B, NK and effector T cell-signatures in blood from human TB patients[16], LTBI-
101 progressors and TB-susceptible mice and yet over-abundance in lungs of *M.*
102 *tuberculosis* infected C57BL/6J resistant mice reinforced their role in early disease
103 control. The translationally relevant knowledge dataset presented here on potential
104 pathways of protection and pathogenesis in human TB are easily accessible using
105 an online ShinyApp : https://ogarra.shinyapps.io/tbtranscriptome/
106
107 **RESULTS**
108 **The peak transcriptomic response in *M. tuberculosis* infected mice.** To
109 determine if a mouse blood transcriptional TB signature resembles that of human
110 disease, we tested the human blood modular transcriptional TB signature[16] on RNA-
111 Seq data from blood of different genetic inbred strains of mice, C57BL/6J (resistant)
112 and C3HeB/FeJ (susceptible), infected with low and high doses of the *M.*

113 *tuberculosis* laboratory strain H37Rv or clinical isolate HN878[21,22] (Supplementary
114 Fig.1a-c; Fig. 1, Supplementary Tables 1-3). The human blood TB signature[16] was
115 first tested on microarray data from blood of H37Rv infected BALB/c mice at different
116 time-points post-infection, to establish the peak transcriptomic response, where
117 immune signatures were barely detectable at days 14 and 21 post-infection, but
118 most significant by day 138 (Supplementary Fig. 2a; Supplementary Table 3).
119 Analysis of blood microarray data from an independent study[30], showed that the
120 blood signature in 129S2 and C57BL/6NCrl mice was again barely detectable at day
121 14 post H37Rv infection, being observed robustly by day 21, which was the end-
122 point of that study[30] (Supplementary Fig. 2a). Upon testing a lung disease modular
123 signature[31] on microarray data from lungs of H37Rv infected BALB/c mice, we
124 detected a peak response at day 56 post-infection, only starting to be detected by 28
125 days post-infection (Supplementary Fig. 2b; Supplementary Table 4). Based on
126 these data we tested the human blood transcriptional TB signature[16] and lung
127 disease modular signature[31], on the blood and lungs, respectively, from C57BL/6J
128 and C3HeB/FeJ mice infected with HN878, at days 26 to 56 post-infection
129 (Supplementary Fig. 3a and b; Supplementary Tables 3 and 4). The peak response
130 chosen was ca. 42 days post-infection, which best showed a robust signature in
131 blood and lungs from HN878 infected C57BL/6J and C3HeB/FeJ mice
132 (Supplementary Fig. 3a and b; Supplementary Tables 3 and 4; tissues from HN878
133 infected susceptible mice were harvested after 33-35 days post infection due to
134 excessive pathology).
135
136 **Blood transcriptional TB signature in mouse and humans.** Principal component
137 analysis (PCA) at the peak response depicted distinct global transcriptional
138 signatures in blood of C57BL/6J (resistant) and C3HeB/FeJ (susceptible) mice,
139 infected with low and high doses of H37Rv or HN878, with the largest distance from
140 uninfected mice observed in HN878 infected C3HeB/FeJ mice (Supplementary Fig.
141 1c). The human blood modular transcriptional TB signature[16] was recapitulated in
142 blood of HN878 infected C3HeB/FeJ mice, and high dose HN878 infected C57BL/6J
143 mice (Fig. 1; for annotation see Supplementary Table 1; for genes see
144 Supplementary Tables 2 and 3). Two over-abundant (red) IFN-inducible modules
145 (HB12 and HB23) in blood from TB patients[16], showed a graded increase from the
146 C57BL/6J to the C3HeB/FeJ mice infected with low to high dose H37Rv, further

147 increased in C57BL/6J then C3HeB/FeJ mice infected with low to high dose HN878
148 (Fig. 1). Expression levels of IFN-inducible modules (HB12 and HB23), in blood of
149 HN878 infected C3HeB/FeJ mice most closely resembled the profile in human TB
150 (Fig. 1). Likewise, other over-abundant modules of human TB, including
151 Inflammasome (HB3), Innate/hemopoietic mediators (HB5), Innate immunity/PRR/C'
152 (HB8) and Myeloid/C'/Adhesion (HB14) modules, were over-abundant in the HN878
153 infected C3HeB/FeJ mice, and to a lesser extent in high dose HN878 infected
154 C57BL/6J mice (Fig. 1). Under-abundance (blue) of the human TB modules, T cell
155 (HB4) and B cell (HB15)[16], was recapitulated in the blood of HN878 infected
156 C3HeB/FeJ mice (Fig. 1). In keeping with this, cellular deconvolution analyses[31] of
157 blood RNA-seq data from *M. tuberculosis* infected mice showed a significant
158 decrease in the percentages of B cell and $CD4^+$ T cell fractions (Supplementary Fig.
159 1d).

160 Cell-types associated with each module were identified by comparing cell-type
161 specific gene signatures using the mouse RNA-Seq dataset from ImmGen Ultra Low
162 Input (ULI) (ImmGen Consortium - GSE109125; http://www.immgen.org) as
163 described[31], analysed against the mouse gene orthologues within each human blood
164 TB module (Fig. 1, right panel). The cell-type specific enrichment data validated the
165 modular annotation for the blood T cell (HB2, HB4) modules, with enrichment for αβ-
166 and γδ-T cells; the NK and T cell (HB21) module, with enrichment for αβ- and γδ-T
167 cells and innate lymphocytes; and the B cell (HB15) module with enrichment for B
168 cells (Fig. 1). This approach also led to the discovery of previously unappreciated
169 gene signatures, most strikingly, a dominance of granulocyte-associated genes
170 within the Inflammasome (HB3) and Innate immunity/PRR/C' (HB8) modules (Fig. 1).
171 This set of granulocyte-associated genes was highly expressed in blood from HN878
172 infected C3HeB/FeJ mice and human TB cohorts (Table 1; Supplementary Table 3
173 Mouse; Supplementary Table 2 Human). Increased expression of granulocyte-
174 associated genes in blood of HN878 infected C3HeB/FeJ mice was reinforced by
175 data obtained from cellular deconvolution analyses[31](Supplementary Fig. 1d).
176

177 **Host and *M. tuberculosis* genetic differences drive lung TB signatures.** To
178 determine the transcriptional response at the site of infection, RNA-Seq data was
179 obtained from the lungs of the same C57BL/6J and C3HeB/FeJ inbred strains of

180 mice infected with H37Rv or HN878, used for the blood data from Fig. 1
181 (Supplementary Fig. 1a). PCA depicted distinct global transcriptional signatures for
182 uninfected mice and the different strains of H37Rv or HN878 infected mice, with the
183 largest distance from uninfected controls observed in HN878 infected C3HeB/FeJ
184 mice (Supplementary Fig. 4). The lung transcriptional response depicted a similar
185 but more accentuated difference between the infected and uninfected groups than in
186 blood (Supplementary Fig. 1c and Supplementary Fig. 4).

187 A lung disease modular signature[31] was tested on the lung RNA-Seq data
188 from the different groups of infected mice, to identify co-expressed groups of genes
189 across the lung (Fig. 2). The type I IFN/Ifit/Oas (L5) module was over-abundant in
190 the lungs of H37Rv and HN878 infected C57BL/6J and C3HeB/FeJ mice to similar
191 levels, as shown by Eigengene expression (Fig. 2a and b). Six modules (L10 – L15),
192 dominated by an over-abundance of granulocyte, macrophage and myeloid specific
193 genes, including modules with Myeloid/Granulocyte (L10) and IL-17
194 pathway/Granulocytes (L11) function, showed the highest Eigengene expression in
195 the HN878 infected C3HeB/FeJ mice (Fig. 2a and c; ShinyApp; Supplementary
196 Table 4). Similarly, the Inflammation/IL-1 signalling/Myeloid Cells (L12), Myeloid
197 cells/Il1b/Tnf (L13) and Myeloid cells/Other signalling (L14) modules were also over-
198 abundant in mouse lungs upon *M. tuberculosis* infection, particularly in the lungs of
199 susceptible HN878 infected C3HeB/FeJ mice (Fig. 2; Supplementary Table 4).
200 Strikingly, an Immunoglobulin h/k module (L25) was over-abundant in the lungs of
201 the C57BL/6J but only minimally in C3HeB/FeJ mice infected with low and high
202 doses of H37Rv, and in the lungs of low dose HN878 infected C57BL/6J mice (Fig.
203 2a and d; ShinyApp; Supplementary Table 4). However, this Immunoglobulin h/k
204 (L25) module was not changed in the lungs of high dose HN878 infected C57BL/6J
205 mice or C3HeB/FeJ mice (Fig. 2a and d), correlating with these mice showing
206 greater TB susceptibility (Supplementary Fig. 1b). This Immunoglobulin h/k (L25)
207 module was also highly abundant in the lungs of BALB/c mice infected with low dose
208 H37Rv, in keeping with its relatively resistant phenotype (Supplementary Fig. 2b).
209 The Ifng/Gbp/Ag presentation/C' (L7) and Cytotoxic/T cells/ILC/Tbx21/Eomes/B cells
210 (L35) modules were over-abundant in the lung across both strains of H37Rv or
211 HN878 infected mice (Fig. 2a and e; Supplementary Fig. 3b), and H37Rv infected
212 BALB/c mice (Supplementary Fig. 2b), but less abundant in lungs from HN878

213 infected C3HeB/FeJ mice (Fig. 2a; Supplementary Fig. 3b), as shown quantitatively
214 by Eigengene profiles (Fig. 2e).
215 Independent derivation and annotation yielded similar transcriptional modules
216 across all samples from uninfected and *M. tuberculosis* infected mice, resulting in 27
217 modules ((ML1 – ML27), Supplementary Fig. 5; Supplementary Tables 5 and 6). The
218 type I IFN/Stat2/Mx1 (ML2) and type I IFN signalling (ML21) modules were similarly
219 over-abundant in the lungs of H37Rv and HN878 infected C57BL/6J and C3HeB/FeJ
220 mice (Supplementary Fig. 5a and b). Over-abundance of modules ML19 and ML27,
221 enriched for Granulocyte/Macrophage specific genes, showed highest Eigengene
222 expression in HN878 infected C3HeB/FeJ mice (Supplementary Fig. 5a and c),
223 confirmed by cell-type specific enrichment analysis (Supplementary Fig. 5a). The
224 Ifng/Gbp/Ag presentation/C' (ML3) and T cell/NK/ILC/APC/B cell (ML11) modules
225 were over-abundant in lungs from both strains of H37Rv or HN878 infected mice,
226 although significantly less abundant in HN878 infected C3HeB/FeJ mice, as shown
227 quantitatively by Eigengene profiles (Supplementary Fig. 5a and d), and validated by
228 cell-type specific enrichment for T cells, DC, innate lymphocytes (ILC) and B cells
229 (Supplementary Fig. 5a). Thus, two complementary and independently derived
230 modular tools revealed similar transcriptional signatures in the lungs of *M.*
231 *tuberculosis* infected susceptible mice, indicating increased type I IFN and
232 granulocyte-associated responses and decreased IFN-γ, NK, T effector and B cell
233 responses (Fig. 2 and Supplementary Fig. 5).
234 The over-abundance of inflammatory modules associated with granulocytes
235 observed using the two independent modular approaches is in keeping with the more
236 severe inflammation observed by H&E staining in the lungs of HN878 infected
237 C3HeB/FeJ mice and high dose HN878 infected C57BL/6J mice (Fig. 3;
238 Supplementary Fig. 6). This was accompanied by greater numbers of *M.*
239 *tuberculosis* bacteria observed in the lungs of these mice by ZN staining
240 (multibacillary infections, Fig. 3; Supplementary Fig. 6).
241
242 **Degree of preservation of lung modules in human and mouse blood.** It is
243 unclear to what extent the airway transcriptional signature is reflected in the blood
244 during *M. tuberculosis* infection. Certain immune responses across a range of
245 experimental models of disease are well preserved between lung and blood, some

246 not preserved, and others only discernible in blood with prior knowledge from the
247 airway response[31]. To address this question in TB, the mouse lung modular TB
248 signature (Supplementary Fig. 5a) was tested on the RNA blood samples from the
249 different cohorts of human TB patients and from the different mouse TB models
250 (Supplementary Fig. 7a). The mouse lung modules showed significant preservation
251 in human and mouse blood as assessed by $Z_{summary}$ scores, indicating the degree of
252 preservation, with scores >10 considered strongly preserved (Supplementary Fig. 7b
253 and c). Type I IFN associated modules (ML2 and ML21) (Supplementary Fig. 5a),
254 were over-abundant in human and mouse blood, being most over-abundant in
255 HN878 infected susceptible mice (Supplementary Fig. 7a). The lung type I
256 IFN/Stat2/Mx1 module (ML2) was the most highly preserved module in human blood
257 (Supplementary Fig. 7b) and the second-most preserved module in mouse blood
258 (Supplementary Fig. 7c) and the type I IFN signalling module (ML21) stood out as
259 the third-most preserved module in both human and mouse blood (Supplementary
260 Fig. 7b and c). The lung Ifng/Gbp/Ag presentation/C' module (ML3) was weakly over-
261 abundant in the blood of human TB patients and *M. tuberculosis* infected mice
262 (Supplementary Fig. 7a), albeit to a lesser extent, but highly preserved in both
263 human and mouse blood (Supplementary Fig. 7b and c). The overall increased
264 abundance of the Ifng/Gbp/Ag presentation/C' module (ML3) was largely attributable
265 to over-expression of genes such as GBP/Gbp genes and complement genes
266 (Supplementary Fig. 7a; Supplementary Tables 2 and 3, and ShinyApp). However,
267 the *Ifng* gene itself, although upregulated in the blood of *M. tuberculosis* infected
268 resistant mice, was barely upregulated in the blood of HN878 infected susceptible
269 mice and *IFNG* was down-regulated in the blood from TB patients (Supplementary
270 Tables 2 and 3, and ShinyApp). The lung Macrophage/Granulocyte modules (ML19
271 and ML27) and Myeloid cell signalling module (ML10) were also over-abundant in
272 blood of active TB patients and most over-abundant in the blood of HN878 infected
273 susceptible mice (Supplementary Fig. 7a). While lung ML19 and ML10 modules
274 were highly preserved in both human and mouse blood, the ML27 module was only
275 highly preserved in mouse blood and to a much lesser extent in human
276 (Supplementary Fig. 7b and c). Lung modules associated with T, NK and B cells
277 (ML11 and ML13) were over-abundant in the lungs of all relatively resistant mice
278 infected with H37Rv, but to a lesser extent (ML11) or under-abundant (ML13) in
279 HN878 infected susceptible mouse lungs (Supplementary Fig. 5) and blood from all

280 TB cohorts (Supplementary Fig. 7a), ML11 being highly preserved in both human
281 and mouse blood (Supplementary Fig. 7b and c). These findings regarding the
282 preservation of over or under-abundant lung modules in the blood from human TB
283 patients and TB susceptible mouse models (Supplementary Fig. 7), are in keeping
284 with the transcriptional signatures observed on testing human blood TB modules on
285 blood from humans and mouse models of TB (Fig. 1).
286
287 **Modular gene networks in human versus mouse TB.** We further interrogated the
288 changes in gene expression of the key modules, HB3, HB15 and HB21, between the
289 blood and lungs of resistant and susceptible mice infected with the different strains of
290 *M. tuberculosis,* as compared to the human blood. To do so, we examined the
291 expression of top 50 "hub" genes with high intramodular connectivity within the
292 mouse data, on human blood from TB patients, and blood and lungs from mice
293 infected with *M. tuberculosis* (Fig. 4). In keeping with our current findings that
294 granulocyte specific genes are upregulated within the originally named
295 Inflammasome human blood TB module (HB3)[16], granulocyte-specific genes were
296 amongst the 50 "hub" genes within that module (now "Inflammasome/Granulocyte")
297 (Fig. 4). These granulocyte-specific genes include, *Cd177, Elane, Mmp8, Mpo, Ncf1,*
298 *Camp, Lcn2, S100a6, Ltf* (Fig. 4, Supplementary Fig. 8a; ShinyApp), which have
299 been associated with neutrophil recruitment and activation[32], were most highly
300 differentially expressed in blood from TB patients and *M. tuberculosis* infected
301 susceptible mice . Expression of these genes in mouse blood and lungs revealed a
302 graded increase from the C57BL/6J to the C3HeB/FeJ mice infected with low to high
303 dose H37Rv, with a further increase observed in C57BL/6J to the C3HeB/FeJ mice
304 infected with low to high dose HN878 (Fig. 4, Supplementary Fig. 8a). The 50 "hub"
305 genes within Innate immunity/PRR/C' module (HB8) also showed enrichment for
306 granulocyte-specific genes including *Mmp9, Alox5ap, Ncf2, Mxd1, S100a8* and
307 *S100a9,* also associated with neutrophil activation (Supplementary Fig. 8b), and
308 were most highly expressed in blood from human TB patients and blood and lung
309 from HN878 infected C3HeB/FeJ mice (Fig. 4, Supplementary Fig. 8b; ShinyApp).
310 Increased expression of these neutrophil-specific genes in the lungs of the TB
311 susceptible HN878 infected mice was mirrored by the increased numbers of
312 neutrophils detected in the lungs of these mice by immunohistochemistry (Fig. 5;
313 Supplementary Fig. 6), confirming the H&E data (Fig. 3; Supplementary Fig. 6).

314 Collectively these data support a major role for neutrophils in human TB
315 pathogenesis, similar to the previously reported role for neutrophils in TB susceptible
316 strains of mice[33-35].

317 The 50 top "hub" genes within the human B cell module (HB15), *Cd19, Pax5,*
318 *Spib, Cd79* and *Cd22,* were down-regulated in the blood of human TB patients and
319 *M. tuberculosis* mice (Fig. 4; Supplementary Fig. 8c; ShinyApp). Most of the B cell-
320 specific top "hub" genes were upregulated in the lungs of H37Rv infected mice, but
321 strikingly down-regulated in the lungs of high dose HN878 infected C57BL/6J and
322 C3HeB/FeJ mice (Fig. 4; Supplementary Fig. 8c; ShinyApp). This difference in
323 expression of B cell-specific genes between the lungs of relatively TB resistant and
324 susceptible mouse models, was mirrored by differences in the numbers of B cells
325 detected by B cell-specific immunofluorescent staining of lungs from these mice (Fig.
326 5; Supplementary Fig. 6). While vastly increased numbers of B cells were observed
327 in the lungs of H37Rv infected mice, with accompanying formation of B cell follicles,
328 these were practically absent in the lungs of C57BL/6J mice infected with high dose
329 HN878 and HN878 infected C3HeB/FeJ mice (Fig. 5; Supplementary Fig. 6). These
330 data support a possible role for B cells in protection against *M. tuberculosis* infection,
331 as has previously been proposed[36,37].

332 In keeping with the under-abundance of the human blood NK & T cells
333 module (HB21), the top 50 "hub" genes in this module were down-regulated in the
334 blood of patients with active TB (Fig. 4; Supplementary Fig. 8d), as previously
335 reported[16]. Although upregulated in the blood and lungs of H37Rv infected C57BL/6J
336 and C3HeB/FeJ mice and HN878 infected C57BL/6J mice, the majority of these 50
337 "hub" genes were down-regulated in the blood and either minimally or not
338 upregulated in the lungs from HN878 infected C3HeB/FeJ mice (Fig. 4;
339 Supplementary Fig. 8d). These included *Tbx21, Gzma, Eomes, Cd8a, Nfatc2, Fasl,*
340 *Nkg7, Klrd1, Klrg1, Ifng* and *Runx3*, reflecting downregulation of effector T and NK
341 cells in the blood of TB patients and HN878 infected C3HeB/FeJ mice (Fig. 4;
342 Supplementary Fig. 8d; ShinyApp). Minimally altered gene expression was mirrored
343 by a decrease in $CD3^+$ T cells in HN878 infected C3HeB/FeJ mouse lungs as shown
344 by immunofluorescence (Fig. 5; Supplementary Fig.6), reflecting an absence of
345 activated effector T cells required for protection against *M. tuberculosis* infection[4-6].

346 Heatmaps of the top 50 "hub" genes from the human blood TB modules
347 Interferon/PRR (HB12) and Interferon/C'/Myeloid (HB23) demonstrated a large

348 number of genes that were over-expressed in human blood from London and
349 Leicester TB cohorts and were similarly over-expressed in mouse blood from HN878
350 infected C3HeB/FeJ mice (Supplementary Fig. 9). In contrast, many of these type I
351 IFN-inducible genes in the HB12 module, including *Il1rn, Ifit1, Ifit2, Oas2* and *Stat2,*
352 were not upregulated, or upregulated to a lower extent, in the blood of H37Rv
353 infected C57BL/6J mice, as compared to HN878 infected C3HeB/FeJ mice
354 (Supplementary Fig. 9). The majority of the top 50 "hub" genes from the
355 Interferon/PRR (HB12) and Interferon/C'/Myeloid (HB23) human modules were
356 upregulated in the lungs of all the *M. tuberculosis* infected mice, with the highest
357 expression observed in the lungs from HN878 infected C3HeB/FeJ mice
358 (Supplementary Fig. 9).
359
360 **Blood signatures reflect the extent of lung pathology in TB.** Correlation between
361 the whole blood TB signature and the extent of lung radiographic burden of human
362 disease has been reported[8]. A quantitative measure of the transcriptional signature,
363 determined using the molecular distance to health, showed a graded increase in the
364 signature across patients categorised with no disease to minimal, moderate and
365 advanced disease[8]. Here we show that the extent of the blood modular signatures
366 associated with type I IFN-inducible genes (HB12 and HB23), shown quantitatively
367 by Eigengene expression, positively correlated with the extent of lung pathology
368 assessed by the combined relative lesion burden and percentage of tissue affected
369 scores in the TB mouse models (Fig. 6a). The type I IFN-associated blood modular
370 signature was lowest in the more resistant mouse models of TB increasing with the
371 different levels of lung pathology, peaking in the HN878 infected C3HeB/FeJ mice
372 (Fig. 6a). Similarly, the level of the type I IFN-associated blood modular signatures in
373 human TB , here shown by Eigengene expression, also positively correlated with the
374 radiographic extent of lung disease in patients with different degrees of disease (Fig.
375 6b). The neutrophil-associated (HB3 and HB8) blood modular signatures, likewise,
376 showed an increased Eigengene expression in the blood of mice in the different TB
377 models correlating with an increased lung neutrophil score (Fig. 6c), and the most
378 severe lung lesions as assessed histopathologically (Fig. 6a). The neutrophil-
379 associated modular blood signature was highest in the HN878 infected C3HeB/FeJ
380 mice correlating with the highest lung neutrophil score (Fig. 6c). Although the
381 neutrophil lung score was similarly high in the high dose HN878 infected C57BL/6J

382 mice, the blood neutrophil-associated modular signature remained low in these mice
383 (Fig. 6c). The blood neutrophil-associated signature in human TB also positively
384 correlated with the radiographic extent of lung disease in TB patients (Fig. 6d), again
385 supporting a role for neutrophils in TB pathogenesis.

386 In contrast to the increased type I IFN and neutrophil-associated blood
387 modular signatures in TB, the blood B cell (HB15) and NK & T cell (HB21) modular
388 signatures showed a decrease in the blood of *M. tuberculosis* infected mice showing
389 advanced lung disease, specifically the HN878 infected C3HeB/FeJ mice, and to a
390 lesser extent the high dose HN878 infected C57BL/6J mice (Fig. 6e). This decreased
391 blood signature in advanced disease correlated with a decrease in the lung
392 lymphocyte score, which in the more resistant mice had increased upon infection
393 (Fig. 6e). In human TB, these blood B cell (HB15) and NK & T cell (HB21) modular
394 signatures showed a similar decrease in the blood, inversely correlating with the
395 extent of lung radiographic disease (Fig. 6f).

396

397 **Modular blood signatures in LTBI-progressors.** We next set out to determine
398 whether the type I IFN (HB12, HB23), neutrophil (HB3, HB8), B cell (HB15) and NK
399 & T cell (HB21) associated modular signatures, determined in human active TB and
400 susceptible mouse models of TB, could be detected during early *M. tuberculosis*
401 infection of humans. To this end, we analysed our RNA-Seq data from the blood of
402 recent contacts of active TB patients who were subsequently shown to progress to
403 active TB (LTBI-progressors), active TB patients and healthy controls (IGRA^{-ve} and
404 IGRA^{+ve} contacts who did not progress to TB)[16] (Fig. 7).

405 The Interferon/PRR (HB12) and Interferon/C'/Myeloid (HB23) blood modular
406 signatures, shown quantitatively by Eigengene expression, were increased in the
407 blood of LTBI-progressors to the same level as in active TB patients, as compared to
408 healthy controls (Fig. 7a). As shown for the London cohort (Fig. 6b), the type I IFN-
409 associated modular signatures also correlated with the radiographic extent of lung
410 disease in this independent cohort (Fig. 7a). Type I IFN-inducible genes in these
411 modular signatures included *STAT1, STAT2, IRF9, OAS1, OASL, IFITM1, ISG15*
412 and *IL1RN* which were expressed at same level in the blood of LTBI-progressors
413 and active TB patients (Fig. 8a; ShinyApp). Again, the degree of expression of these
414 individual genes positively correlated with the extent of radiographic signs of
415 disease, being already increased in the blood of patients with minimal disease (Fig.

416 8a). We also observed increased expression of these type I IFN-inducible genes
417 (Fig. 8a) in an independent cohort of LTBI-progressors, as compared to individuals
418 with LTBI who remained healthy[38].

419 Strikingly, expression of the neutrophil-associated (HB3 and HB8) modular
420 signatures was also increased to high levels in the blood of LTBI-progressors, to the
421 same level as seen in blood of active TB patients, as compared to healthy controls
422 (Fig. 7b). The extent of these neutrophil associated blood signatures positively
423 correlated with the radiographic signs of lung disease (Fig. 7b). Confirming the
424 contribution of genes associated with neutrophil activation and recruitment, *CD177,*
425 *NCF1, NCF2, LRG1, MMP9, S100A8, S100A9* and *ALOX5AP* were upregulated in
426 the blood of LTBI-progressors from both cohorts, as compared to controls to a
427 similar level as in the blood of active TB patients, their level of expression correlating
428 with increased signs of radiographic lung disease (Fig. 8b). The increased
429 expression of genes associated with neutrophil activation and recruitment in the
430 blood of TB patients with minimal radiographic signs of disease and LTBI-
431 progressors (Fig. 8b) points to an unappreciated role for neutrophils in early disease.

432 The expression of the B cell (HB15) and NK & T cell (HB21) associated
433 modular signatures was decreased in the blood of LTBI-progressors to the same
434 extent as in active TB (Fig. 7c), as compared to controls, again correlating with
435 increased radiographic signs of disease (Fig. 7c). Expression of the NK & T cell
436 specific genes *IFNG, EOMES, TBX21, GZMA, KLRD1, NKG7,* was similarly
437 decreased in the blood of LTBI-progressors in both cohorts, as compared to the
438 healthy controls and in the blood of patients with minimal signs of disease, although
439 further decreased in those with advanced signs of radiographic disease (Fig. 8c).
440 Since T cell and NK cell genes convey protection against *M. tuberculosis* infection[4-
441 6,39,40], their loss may contribute to progression to active TB.

442 Collectively our findings predict that a dominance of gene expression
443 associated with a type I IFN response and neutrophil activation and recruitment,
444 together with a loss of NK and T cell effector responses, early after infection with *M.*
445 *tuberculosis*, may contribute to progression to active TB.

446

447 **DISCUSSION**

448 Here we show that the IFN-inducible human blood TB transcriptional signature[16] is
449 recapitulated in blood from *M. tuberculosis* HN878 infected TB susceptible

450 C3HeB/FeJ mice, whereas this signature is minimal in blood from *M. tuberculosis*
451 H37Rv infected resistant C57BL/6J mice. Combining our modular signature data with
452 cell-type specific signatures[31] we reveal an increase in neutrophil-associated genes
453 in the blood of TB susceptible mice and TB patients. Genes associated with type I
454 IFN responses and with neutrophil recruitment and activation were increased in
455 LTBI-progressors before diagnosis, suggesting an unappreciated role for neutrophils
456 in early disease. Decreased B, NK and T cell-signatures of human active TB[8,16] were
457 observed in the blood of infected TB susceptible mice and LTBI-progressors, whilst
458 upregulated upon infection in the lungs of TB resistant mice, suggesting that their
459 early loss contributes to progression to active TB.

460 Neutrophils are abundant in the lung lesions of *M. tuberculosis* infected
461 susceptible mice contributing to TB pathogenesis[33,34] whereas lesions of infected
462 resistant mice contain only scattered neutrophils, instead dominated by lymphocytes
463 and macrophages[41]. *M. tuberculosis* infected neutrophils have been detected within
464 inflammatory lung granulomas of patients with active TB[42,43]. We herein reveal low
465 levels of a neutrophil-associated signature in lungs of *M. tuberculosis* infected
466 C57BL/6J mice, which was maximally increased in HN878 infected susceptible
467 C3HeB/FeJ mice. This was validated by histological analysis, although S100A9
468 neutrophil staining was lost due to the necrotic nature of the lesions. This led to our
469 discovery of increased expression of neutrophil-associated genes within the over-
470 abundant human TB blood modules, originally annotated as "Inflammasome" and
471 "Innate immunity/PRR/C'"[16], which we now rename as "Inflammasome/Granulocyte"
472 and "Innate immunity/PRR/C'/Granulocyte". Previous studies showed no change by
473 flow cytometry in neutrophil numbers in the blood of active TB patients[8], suggesting
474 that the over-abundance of this granulocyte-associated signature of activation and
475 recruitment may be attributable to a subset of activated neutrophils which have
476 circulated to the blood from the lung. Whether these neutrophils are carriers of *M.*
477 *tuberculosis* to the blood in human TB, where the bacteria have been recently shown
478 to be detected in early disease[44], remains to be investigated. This granulocyte-
479 associated signature was also increased in blood from LTBI-progressors before
480 diagnosis, suggesting a previously unappreciated role for neutrophils in early
481 progression to disease.

482 The type I IFN-associated signature widely reported in blood of active TB
483 patients[8-16] was also present in blood from *M. tuberculosis* infected mice, with the

484 highest levels observed in the more susceptible models, correlating with more
485 severe lung pathology. This type I IFN-inducible signature resulted from the host
486 genetic background and the dose and strain of *M. tuberculosis,* possibly explaining
487 differing reports regarding the role of type I IFN in TB pathogenesis[18,20-22,24,45,46]. The
488 enhanced type I IFN-associated signature in the C3HeB/FeJ mice is in keeping with
489 a recent report that the B6.Sst1S congenic mice carrying the C3H "sensitive" allele of
490 the Sst1 locus that renders them highly susceptible to *M. tuberculosis* infections[47],
491 exhibit markedly increased type I IFN signalling which contributes to their high TB
492 susceptibility via induction of the IL-1 receptor antagonist (IL-1Ra)[48]. We show that
493 the *Il1rn* gene expression is increased in mouse blood and lung upon infection,
494 correlating with increasing susceptibility to TB in C3HeB/FeJ mice infected with
495 HN878, a *M. tuberculosis* strain reported to enhance type I IFN induction and TB
496 pathogenesis[21,22]. The *IL1RN* gene was highly expressed in blood from TB patients,
497 but also in the LTBI-progressors, along with other type I IFN-inducible genes such as
498 *OAS1, IFITM1* and *ISG15*, suggesting that type I IFN-inducible genes may contribute
499 to early TB pathogenesis. Genes of the complement cascade were also upregulated
500 in the blood from LTBI-progressors, in keeping with previous reports[15,49].

501 Upregulation of both type I and II IFN have been reported before diagnosis of
502 TB patients[15]. However, we herein report that in TB patients the *IFNG* gene itself is
503 down-regulated in the blood, alongside a number of key molecules, including *TBX21,*
504 *EOMES, GZMA, GZMB, NKG7 and KLRD1,* suggesting a loss of the protective
505 effector function of T and NK cells[5,6,39,40]. This decrease was also observed in LTBI-
506 progressors, suggesting that decreased expression of *IFNG* and other genes
507 associated with effector and cytotoxic functions early after *M. tuberculosis* infection
508 may contribute to disease progression. This supports reports that IFN-γ, cytotoxic
509 effector molecules and NK cells are important for protection against *M. tuberculosis*
510 infection in both mouse models and human disease[5,6,39,40]. In keeping with this,
511 genes associated with effector and cytotoxic NK and T cell responses (*Nkg7, Klrd1,*
512 *Gzma, Gzmb, Tbx21*) as well as *Ifng* were upregulated in the blood and lungs of *M.*
513 *tuberculosis* infected TB resistant C57BL/6J mice but drastically reduced in the blood
514 and lungs from HN878 infected susceptible C3HeB/FeJ mice, similarly to in blood
515 from LTBI-progressors and active TB patients. Decreased *IFNG* expression in the
516 blood of TB patients and *Ifng* expression in the blood and lungs of susceptible mice

517 parallels the increase in neutrophils, supporting previous reports that IFN-γ regulates
518 neutrophil function[35] thus limiting lung inflammation and TB exacerbation.
519 Our findings of a decrease in the B cell-associated modular expression in the
520 blood of *M. tuberculosis* infected susceptible mice are in keeping with reports on the
521 reduction in abundance of total B cells in human TB[8,40] largely driven by a reduction
522 in circulating naive B cells[40]. This under-abundance of the B cell-associated module
523 was also observed in blood from LTBI-progressors, although maximal in TB patients
524 and susceptible mice with advanced signs of lung disease. Reduction in peripheral B
525 cells could be due to preferential sequestration of these cells at the site of infection
526 or diminished output of B cells from the bone marrow[36,37]. Our data support a
527 combination of both, depending on the extent of the disease. The top 50 interacting
528 "hub" genes in the B cell-associated module showed increased expression in the
529 lungs from *M. tuberculosis* infected resistant mice , but were decreased in lungs from
530 HN878 infected susceptible mice, as verified by histopathology. B cells at the site of
531 infection could be contributing to control of *M. tuberculosis* infection in the resistant
532 mice as has been proposed[36,37].
533 Using a combination of mouse TB models and human TB cohorts we provide
534 data to suggest that dominance of a type I IFN response and neutrophil activation
535 and recruitment, together with a loss of B cell, NK and T cell effector responses may
536 contribute to the pathogenesis of progressive *M. tuberculosis* infection. Mouse
537 models of TB have been employed for decades as tools for elucidating mechanisms
538 of host resistance and pathogenesis. While failing to recapitulate many of the
539 features of clinical TB and in several cases protective vaccine responses, they have
540 been remarkably useful in identifying both effector and regulatory responses that
541 have emerged to be important in human infection and disease. The data reported
542 here comparing the host transcriptomic responses of *M. tuberculosis* infected mice
543 and humans offer further compelling characterization and validation of the mouse
544 model for further mechanistic studies and suggest a peripheral signature associated
545 with progression to clinical disease in TB.
546
547 **ACKNOWLEDGEMENTS.** We wish to thank S. Hadebe (The Francis Crick Institute
548 UK, now UCT, South Africa) and J. Pitt (MRC NIMR UK) for some of the early mouse
549 *M. tuberculosis* infections and sample processing leading up to the current study; K.

550 Potempa (MRC NIMR UK) for early analysis of some of the microarray data leading
551 up to the current study. We thank L. Gabryšová (The Francis Crick Institute UK, now
552 Novartis, Basel, Sw) for her intellectual contribution to discussion of the project. We
553 thank X. Wu for her help in organisation of mice for TB experiments. We thank The
554 Francis Crick Institute: Biological Services for breeding and maintenance of the mice
555 used for the early mouse *M. tuberculosis* infections leading up to the current study;
556 Advanced Sequencing Facility, Bioinformatics and Biostatistics Science Technology
557 Platforms for their contribution to our sequencing processing, and R. Goldstone
558 excellent project management of sequencing and D. Jackson for support of
559 sequencing; and Experimental Histopathology for their excellent work in preparing
560 lung sections for histological analyses.

561 This whole study was funded by The Francis Crick Institute which receives its
562 core funding from Cancer Research UK (FC001126), the UK Medical Research
563 Council (FC001126), and the Wellcome Trust (FC001126); before that by the UK
564 Medical Research Council (MRC U117565642); and by the European Research
565 Council (294682-TB-PATH). A.O'G., L.M-T., O.T., C.M.G, A.Si., E.S. were supported
566 by The Francis Crick Institute which receives its core funding from Cancer Research
567 UK (FC001126), the UK Medical Research Council (FC001126), and the Wellcome
568 Trust (FC001126); before that by the UK Medical Research Council (MRC
569 U117565642); L.M-T., and A.Si., were additionally funded by the European
570 Research Council (294682-TB-PATH). S.L.P., A.S-B., and E.H. were funded by the
571 Royal Veterinary College and The Francis Crick Institute. K.D.M-B., and A.Sh. were
572 funded by the Intramural Research Program of the National Institutes of Allergy and
573 Infectious Disease. M.S. was funded by FEDER-Fundo Europeu de
574 Desenvolvimento Regional funds through the COMPETE 2020 – Operational
575 Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and
576 by Portuguese funds through FCT in the frame-work of the project "Institute for
577 Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274), and
578 by FCT through *Estimulo Individual ao Emprego Científico*. K.L.F. and B.C. are
579 funded by FCT PhD scholarships SFRH/BD/114405/2016 and
580 SFRH/BD/114403/2016, respectively. P.H. and R.V. were supported by NIHR
581 Leicester Biomedical Research Centre and the University of Leicester.
582

583 **AUTHOR CONTRIBUTIONS.** A.O'G. conceived and designed the study with input
584 from L.M-T., M.S., C.M.G., and P.S.R; and lead and co-ordinated the study with help
585 from L.M-T. A.O'G co-wrote the manuscript together with L.M-T. O.T. performed the
586 major part of the bioinformatics analyses with significant input from A.Si. supervised
587 by A.O'G. P.C provided technical bioinformatics support. L.M-T coordinated the
588 logistics of the study. C.M.G. isolated RNA and prepared sequencing libraries from
589 the TB mouse models. E.S. assisted in TB mouse model experiments executed by
590 K.L.F., J.S., B.C., and P.S.R. and designed by A.O'G, L.M-T and M.S. E.S.
591 contributed to all early mouse model TB experiments executed with P.S.R. and L.M-
592 T. supervised by, A.O'G, L.M-T. and M.S. S.L.P., A.S-B, and E.H. performed
593 histopathological analysis and interpretation. K.D.M-B. and A.Sh. provided TB
594 samples from early mouse TB experiments leading up to the current study. M.S.,
595 K.D.M-B. and A.Sh., also provided important discussion for the project throughout
596 and critical feedback on the manuscript. R.V. and P.H. provided clinical analysis of
597 TB patients, and also provided critical feedback on the manuscript and important
598 discussion for the project. All co-authors have read, reviewed and approved the
599 manuscript.
600
601 **COMPETING INTERESTS.** The authors declare no competing interests
602
603 **REFERENCES**
604

605 1 WHO. Global Tuberculosis Report 2017. World Health Organization (2017).
606 2 Dowdy, D. W., Basu, S. & Andrews, J. R. Is passive diagnosis enough? The
607 impact of subclinical disease on diagnostic strategies for tuberculosis.
608 American journal of respiratory and critical care medicine 187, 543-551
609 (2013).
610 3 Casanova, J. L. & Abel, L. Genetic dissection of immunity to mycobacteria:
611 the human model. Annu Rev Immunol 20, 581-620 (2002).
612 4 Cooper, A. M., Mayer-Barber, K. D. & Sher, A. Role of innate cytokines in
613 mycobacterial infection. Mucosal immunology 4, 252-260 (2011).
614 5 Flynn, J. L. & Chan, J. Immunology of tuberculosis. Annu Rev Immunol 19,
615 93-129 (2001).
616 6 O'Garra, A. *et al.* The immune response in tuberculosis. Annu Rev Immunol
617 31, 475-527 (2013).
618 7 Mayer-Barber, K. D. *et al.* Host-directed therapy of tuberculosis based on
619 interleukin-1 and type I interferon crosstalk. Nature 511, 99-103 (2014).
620 8 Berry, M. P. *et al.* An interferon-inducible neutrophil-driven blood
621 transcriptional signature in human tuberculosis. Nature 466, 973-977 (2010).

622 9 Blankley, S. *et al.* A 380-gene meta-signature of active tuberculosis compared
623 with healthy controls. The European respiratory journal 47, 1873-1876 (2016).
624 10 Blankley, S. *et al.* The Transcriptional Signature of Active Tuberculosis
625 Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis.
626 PloS one 11, e0162220 (2016).
627 11 Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary
628 tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PloS one
629 8, e70630 (2013).
630 12 Joosten, S. A., Fletcher, H. A. & Ottenhoff, T. H. A helicopter perspective on
631 TB biomarkers: pathway and process based analysis of gene expression data
632 provides new insight into TB pathogenesis. PloS one 8, e73230 (2013).
633 13 Kaforou, M. *et al.* Detection of tuberculosis in HIV-infected and -uninfected
634 African adults using whole blood RNA expression signatures: a case-control
635 study. PLoS medicine 10, e1001538 (2013).
636 14 Maertzdorf, J. *et al.* Human gene expression profiles of susceptibility and
637 resistance in tuberculosis. Genes Immun 12, 15-22 (2011).
638 15 Scriba, T. J. *et al.* Sequential inflammatory processes define human
639 progression from M. tuberculosis infection to tuberculosis disease. PLoS
640 pathogens 13, e1006687 (2017).
641 16 Singhania, A. *et al.* A modular transcriptional signature identifies phenotypic
642 heterogeneity of human tuberculosis infection. Nature communications 9,
643 2308 (2018).
644 17 Behar, S. M., Divangahi, M. & Remold, H. G. Evasion of innate immunity by
645 Mycobacterium tuberculosis: is death an exit strategy? Nat Rev Microbiol 8,
646 668-674 (2010).
647 18 Moreira-Teixeira, L., Mayer-Barber, K., Sher, A. & O'Garra, A. Type I
648 interferons in tuberculosis: Foe and occasionally friend. The Journal of
649 experimental medicine 215, 1273-1285 (2018).
650 19 Carmona, J. *et al.* Mycobacterium tuberculosis Strains Are Differentially
651 Recognized by TLRs with an Impact on the Immune Response. PloS one 8,
652 e67277 (2013).
653 20 Dorhoi, A. *et al.* Type I IFN signaling triggers immunopathology in
654 tuberculosis-susceptible mice by modulating lung phagocyte dynamics. Eur J
655 Immunol 44, 2380-2393 (2014).
656 21 Manca, C. *et al.* Virulence of a Mycobacterium tuberculosis clinical isolate in
657 mice is determined by failure to induce Th1 type immunity and is associated
658 with induction of IFN-alpha /beta. Proceedings of the National Academy of
659 Sciences of the United States of America 98, 5752-5757 (2001).
660 22 Manca, C. *et al.* Hypervirulent M. tuberculosis W/Beijing strains upregulate
661 type I IFNs and increase expression of negative regulators of the Jak-Stat
662 pathway. J Interferon Cytokine Res 25, 694-701 (2005).
663 23 Ordway, D. *et al.* The hypervirulent Mycobacterium tuberculosis strain HN878
664 induces a potent TH1 response followed by rapid down-regulation. Journal of
665 immunology (Baltimore, Md. : 1950) 179, 522-531 (2007).
666 24 McNab, F. W. *et al.* TPL-2-ERK1/2 signaling promotes host resistance against
667 intracellular bacterial infection by negative regulation of type I IFN production.
668 Journal of immunology (Baltimore, Md. : 1950) 191, 1732-1743 (2013).
669 25 Bogunovic, D. *et al.* Mycobacterial disease and impaired IFN-gamma
670 immunity in humans with inherited ISG15 deficiency. Science (New York,
671 N.Y.) 337, 1684-1688 (2012).

672 26 Antonelli, L. R. *et al.* Intranasal Poly-IC treatment exacerbates tuberculosis in
673 mice through the pulmonary recruitment of a pathogen-permissive
674 monocyte/macrophage population. J Clin Invest 120, 1674-1682 (2010).
675 27 Redford, P. S. *et al.* Influenza A virus impairs control of Mycobacterium
676 tuberculosis coinfection through a type I interferon receptor-dependent
677 pathway. The Journal of infectious diseases 209, 270-274 (2014).
678 28 Barry, C. E., 3rd *et al.* The spectrum of latent tuberculosis: rethinking the
679 biology and intervention strategies. Nat Rev Microbiol 7, 845-855 (2009).
680 29 Kramnik, I. & Beamer, G. Mouse models of human TB pathology: roles in the
681 analysis of necrosis and the development of host-directed therapies.
682 Seminars in immunopathology 38, 221-237 (2016).
683 30 Domaszewska, T. *et al.* Concordant and discordant gene expression patterns
684 in mouse strains identify best-fit animal model for human tuberculosis.
685 Scientific reports 7, 12094 (2017).
686 31 Singhania, A. *et al.* Transcriptional profiling unveils type I and II interferon
687 networks in blood and tissues across diseases. Nature communications 10,
688 2887 (2019).
689 32 Ley, K. *et al.* Neutrophils: New insights and open questions. Sci Immunol 3
690 (2018).
691 33 Dorhoi, A. *et al.* The adaptor molecule CARD9 is essential for tuberculosis
692 control. The Journal of experimental medicine 207, 777-792 (2010).
693 34 Eruslanov, E. B. *et al.* Neutrophil responses to Mycobacterium tuberculosis
694 infection in genetically susceptible and resistant mice. Infection and immunity
695 73, 1744-1753 (2005).
696 35 Nandi, B. & Behar, S. M. Regulation of neutrophils by interferon-gamma limits
697 lung inflammation during tuberculosis infection. The Journal of experimental
698 medicine 208, 2251-2262 (2011).
699 36 Achkar, J. M., Chan, J. & Casadevall, A. B cells and antibodies in the defense
700 against Mycobacterium tuberculosis infection. Immunological reviews 264,
701 167-181 (2015).
702 37 Lu, L. L. *et al.* A Functional Role for Antibodies in Tuberculosis. Cell 167, 433-
703 443.e414 (2016).
704 38 Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a
705 prospective cohort study. Lancet 387, 2312-2322 (2016).
706 39 Feng, C. G. *et al.* NK cell-derived IFN-gamma differentially regulates innate
707 resistance and neutrophil response in T cell-deficient hosts infected with
708 Mycobacterium tuberculosis. Journal of immunology (Baltimore, Md. : 1950)
709 177, 7086-7093 (2006).
710 40 Roy Chowdhury, R. *et al.* A multi-cohort study of the immune factors
711 associated with M. tuberculosis infection outcomes. Nature 560, 644-648
712 (2018).
713 41 Aly, S. *et al.* Oxygen status of lung granulomas in Mycobacterium
714 tuberculosis-infected mice. The Journal of pathology 210, 298-305 (2006).
715 42 Gopal, R. *et al.* S100A8/A9 proteins mediate neutrophilic inflammation and
716 lung pathology during tuberculosis. American journal of respiratory and critical
717 care medicine 188, 1137-1146 (2013).
718 43 Eum, S. Y. *et al.* Neutrophils are the predominant infected phagocytic cells in
719 the airways of patients with active pulmonary TB. Chest 137, 122-128 (2010).

720 44 Verma, R. *et al.* A novel high sensitivity bacteriophage-based assay identifies
721 low level M. tuberculosis bacteraemia in immunocompetent patients with
722 active and incipient TB. Clin Infect Dis (2019).
723 45 Cooper, A. M., Pearl, J. E., Brooks, J. V., Ehlers, S. & Orme, I. M. Expression
724 of the Nitric Oxide Synthase 2 Gene Is Not Essential for Early Control of
725 *Mycobacterium tuberculosis* in the Murine Lung. Infection and immunity 68,
726 6879-6882 (2000).
727 46 Moreira-Teixeira, L. *et al.* T Cell-Derived IL-10 Impairs Host Resistance to
728 Mycobacterium tuberculosis Infection. Journal of immunology (Baltimore, Md.
729 : 1950) 199, 613-623 (2017).
730 47 Pichugin, A. V., Yan, B. S., Sloutsky, A., Kobzik, L. & Kramnik, I. Dominant
731 role of the sst1 locus in pathogenesis of necrotizing lung granulomas during
732 chronic tuberculosis infection and reactivation in genetically resistant hosts.
733 The American journal of pathology 174, 2190-2201 (2009).
734 48 Ji, D. X. *et al.* Interleukin-1 receptor antagonist mediates type I interferon-
735 driven susceptibility to Mycobacterium tuberculosis. bioRxiv (2018).
736 49 Esmail, H. *et al.* Complement pathway gene activation and rising circulating
737 immune complexes characterize early disease in HIV-associated tuberculosis.
738 Proceedings of the National Academy of Sciences of the United States of
739 America 115, E964-E973 (2018).
740
741
742
743 **Figure legends:**
744
745 **Figure 1. Human TB blood transcriptional signature is preserved in blood of**
746 **TB susceptible mice.** Blood modules of co-expressed genes derived using WGCNA
747 from human TB datasets in Singhania *et al.* 2018[18] are shown for blood RNA-seq
748 datasets from TB patients from London (n=21 biologically independent samples),
749 South Africa (n=16 biologically independent samples) (both compared to London
750 controls; n=12 biologically independent samples) and Leicester (n=53 biologically
751 independent samples) (compared to Leicester controls; n=50 biologically
752 independent samples) (Supplementary Table 2); human blood modules were tested
753 in blood RNA-seq datasets obtained from different genetic strains of mice
754 (C57BL/6J, resistant; C3HeB/FeJ, susceptible) infected with low and high doses of
755 *M. tuberculosis* laboratory strain H37Rv or the *M. tuberculosis* clinical isolate HN878
756 (n=4 biologically independent samples per group for H37Rv infection and n=5
757 biologically independent samples per group for HN878 infection from one experiment
758 per *M. tuberculosis* infection as depicted in Supplementary Fig. 1a), compared to
759 their respective uninfected controls (Supplementary Table 3). Fold enrichment
760 scores derived using QuSAGE are depicted, with red and blue indicating modules

761 over- or under-abundant, compared to the controls. Colour intensity of the dots
762 represents the degree of perturbation, indicated by the colour scale. Size of the dots
763 represents the relative degree of perturbation, with the largest dot representing the
764 highest degree of perturbation within the plot. Only modules with fold enrichment
765 scores with FDR *p*-value < 0.05 were considered significant and depicted here (left
766 and middle panels). Module name indicates biological processes associated with the
767 genes within the module (Supplementary Table 1). C', complement. PRR, pathogen
768 recognition receptor. Cell-type associated with genes within each module were
769 identified using the mouse cell-type-specific signatures from Singhania *et al.* 2019[41]
770 (right panel). Cell-type enrichment was calculated using a hypergeometric test, with
771 only FDR p-value < 0.05 considered significant and depicted here (right panel).
772 Colour intensity represents significance of enrichment..
773
774 **Figure 2. Mouse lung disease modules tested in lungs from diverse mouse TB**
775 **models. a,** Mouse lung disease modules derived in Singhania *et al.* 2019[41] (L1-L38)
776 tested in mouse lung TB samples from different genetic strains of mice (C57BL/6J,
777 resistant; C3HeB/FeJ, susceptible) infected with low and high doses of *M.*
778 *tuberculosis* laboratory strain H37Rv or the *M. tuberculosis* clinical isolate HN878
779 (n=3 biologically independent samples per group for low dose HN878 infection of
780 C3HeB/FeJ, and n=5 biologically independent samples per group for all other groups
781 as depicted in Supplementary Fig. 1a), compared to their respective uninfected
782 controls (Supplementary Table 4). Red and blue indicate modules over- or under-
783 abundant, compared to the controls. Colour intensity of the dots represents the
784 degree of perturbation, indicated by the colour scale. Size of the dots represents the
785 relative degree of perturbation, with the largest dot representing the highest degree
786 of perturbation within the plot. Only modules with fold enrichment scores with FDR *p*-
787 value < 0.05 were considered significant and depicted here. GCC, glucocorticoid; K-
788 channel, potassium channel; TM, transmembrane; Ubiq, ubiquitination. **b-e**, Box
789 plots depicting the module eigengene expression, i.e. the first principal component
790 for all genes within the module, are shown for uninfected (Uninf) and *M. tuberculosis*
791 infected (Low dose; High dose) C57Bl/6 and C3HeB/FeJ mice, for modules (**b**) Type
792 I IFN/Ifit/Oas (L5); (**c**) IL-17 pathway/granulocytes (L11), Inflammation/IL-1 signaling/
793 Myeloid cells (L12), Myeloid cells/*Il1b*/*Tnf* (L13); (**d**) Immunoglobulin h/k enriched

794 (L25); (**e**) Cytotoxic/T cells/ILC/Tbx21/Eomes/B cells (L35) and Ifng/Gbp/Antigen
795 presentation (L7).
796
797 **Figure 3. Histological analysis of mouse lungs from *M. tuberculosis* infected**
798 **mice.** Representative photomicrographs of hematoxylin and eosin (H&E) or Ziehl–
799 Neelsen (ZN) stained lung sections from different genetic strains of mice (C57BL/6J,
800 resistant; C3HeB/FeJ, susceptible) infected with low and high doses of *M.*
801 *tuberculosis* laboratory strain H37Rv or the *M. tuberculosis* clinical isolate HN878
802 (n=2 biologically independent samples per group for H37Rv infection, HN878-
803 infected C57BL/6J mice low dose and HN878-infected C3HeB/FeJ mice high dose,
804 and n=3 biologically independent samples per group for HN878-infected C57BL/6J
805 mice high dose and HN878-infected C3HeB/FeJ mice low dose, from one
806 experiment per *M. tuberculosis* infection). From top to bottom, scale bar represents 2
807 mm, 200 µm and 100 µm for H&E staining, 20 µm for ZN staining; arrows locate
808 bacteria.
809
810 **Figure 4. Gene networks of specific TB modules in human blood from TB**
811 **patients, and blood and lung from *M. tuberculosis* infected mice.** Differential
812 expression of genes from human blood modules Inflammasome/Granulocytes (HB3),
813 B cells (HB15) and NK & T cells (HB21) depicting the top 50 "hub" network of genes
814 with high intramodular connectivity found within the mouse data (i.e., mouse genes
815 most connected with all other genes within the module), is shown for data from blood
816 from TB patients (Leicester cohort), and blood and lungs from mice infected with *M.*
817 *tuberculosis,* each against their respective controls. An enlarged representative
818 network showing human gene names is shown for human blood (top) and an
819 enlarged representative network showing mouse gene names is shown for blood
820 samples from C3HeB/FeJ mice infected with high dose of HN878 (bottom). Each
821 gene is represented as a circular node with edges representing correlation between
822 the gene expression profiles of the two respective genes. Colour of the node
823 represents log2 foldchange of the gene for human blood TB samples or mouse blood
824 and lung samples from *M. tuberculosis* infected mice compared to respective
825 controls.
826

827 **Figure 5. Histological analysis of mouse lungs from *M. tuberculosis* infected**
828 **mice for neutrophils, T and B cells**. Representative photomicrographs of lung
829 sections from different genetic strains of mice (C57Bl/6, resistant; C3HeB/FeJ,
830 susceptible) infected with low and high doses of *M. tuberculosis* laboratory strain
831 H37Rv or the *M. tuberculosis* clinical isolate HN878 (n=2 biologically independent
832 samples per group for H37Rv infection, HN878-infected C57BL/6J mice low dose
833 and HN878-infected C3HeB/FeJ mice high dose, and n=3 biologically independent
834 samples per group for HN878-infected C57BL/6J mice high dose and HN878-
835 infected C3HeB/FeJ mice low dose, from one experiment per *M. tuberculosis*
836 infection) depicting neutrophils (2B10, brown) by immunohistochemistry or T (CD3
837 positive, red) and B (B220 positive, green) cells by immunofluorescence (nuclear
838 staining depicted in blue, DAPI). Scale bar represents 100 µm (top) and 50 µm
839 (bottom) for Neutrophils, 200 µm (top) and 100 µm (bottom) for T & B cells.
840
841 **Figure 6. Quantitation of specific blood modular signatures against extent of**
842 **lung pathology in mouse models and human TB.** Box plots depicting the module
843 Eigengene expression for human blood modules Interferon/PRR (HB12) and
844 Interferon/C'/Myeloid (HB23) (**a**, **b**), Inflammasome/Granulocytes (HB3) and Innate
845 immunity/PRR/C'/ Granulocytes (HB8) (**c**, **d**), B cells (HB15) and NK & T cells
846 (HB21) (**e**, **f**), are shown for mouse blood samples from uninfected (Uninf; n = 5
847 biologically independent samples per group) and *M. tuberculosis* H37Rv or HN878
848 infected (L, low dose; H, high dose) C57Bl/6 and C3HeB/FeJ mice (n=3 biologically
849 independent samples per group for low dose HN878 infection of C3HeB/FeJ, and
850 n=5 biologically independent samples per group for all other groups as depicted in
851 Supplementary Fig. 1a) (**a**, **c**, **e**); and for human blood samples from the London TB
852 cohort divided in Healthy Control (no X-ray; n=12 biologically independent samples)
853 and TB patients grouped according to the radiographic extent of disease as No
854 disease (n=21 biologically independent samples), Minimal (n=7 biologically
855 independent samples), Moderate (n = 6 biologically independent samples) or
856 Advanced (n=8, biologically independent samples, described in Berry *et al.* 2010[9])
857 (**b**, **d**, **f**). Lung lesion global score (**a**), neutrophil (**c**) and lymphocyte (**e**) scores from
858 H&E stained lung sections are also shown for uninfected (Uninf, n=5 biologically
859 independent samples per group) and *M. tuberculosis* H37Rv or HN878 infected (L,
860 low dose; H, high dose) C57Bl/6 and C3HeB/FeJ mice (n=2 biologically independent

861 samples per group for H37Rv infection, HN878-infected C57BL/6J mice low dose
862 and HN878-infected C3HeB/FeJ mice high dose, and n=3 biologically independent
863 samples per group for HN878-infected C57BL/6J mice high dose and HN878-
864 infected C3HeB/FeJ mice low dose, from one experiment per *M. tuberculosis*
865 infection).
866
867 **Figure 7. Quantitation of specific blood modular signatures in blood of healthy**
868 **controls, LTBI, LTBI-progressors and active TB patients.** Box plots depicting the
869 module Eigengene expression for human blood modules Interferon/PRR (HB12) and
870 Interferon/C'/Myeloid (HB23) (**a**), Inflammasome/Granulocytes (HB3) and Innate
871 immunity/PRR/C'/ Granulocytes (HB8) (**b**), B cells (HB15) and NK & T cells (HB21)
872 (**c**), are shown for human blood samples from the Leicester TB cohort divided in
873 Control (IGRA^{-ve} TB contacts who remained healthy; n=50 biologically independent
874 samples), LTBI (IGRA^{+ve} TB contacts who remained healthy; n=49 biologically
875 independent samples), LTBI_Progressor (TB contacts who developed TB, time point
876 just before the contact was diagnosed with active TB; n=6 biologically independent
877 samples) and Active_TB (patients with active disease; n=53 biologically independent
878 samples) (left panels) or divided in Control – LTBI (IGRA^{-ve} and IGRA^{+ve} TB contacts
879 who remained healthy) or TB patients grouped according to the radiographic extent
880 of disease as Minimal, Moderate and Advanced (right panels; Supplementary Table
881 7).
882
883 **Figure 8. Quantitation of IFN, neutrophil and lymphocyte-specific gene**
884 **expression in blood of healthy controls, LTBI, LTBI-progressors and active TB**
885 **patients.** Box plots depicting the $\log_2$ expression values of selected genes from type
886 I IFN-associated modules HB12 and HB23 (**a**), neutrophil-associated modules HB3
887 and HB8 (**b**) and NK & T cell module HB21 (**c**) are shown for human blood samples
888 from the Leicester TB cohort divided in Control (IGRA^{-ve} TB contacts who remained
889 healthy; n=50 biologically independent samples), LTBI (IGRA^{+ve} TB contacts who
890 remained healthy; n=49 biologically independent samples), LTBI_Progressor (TB
891 contacts who developed TB, time point just before the contact was diagnosed with
892 active TB; n=6 biologically independent samples) and Active_TB (patients with active
893 disease; n=53 biologically independent samples) (left panels) or divided in Control –
894 LTBI (IGRA^{-ve} and IGRA^{+ve} TB contacts who remained healthy) and TB patients

895 grouped according to the radiographic extent of disease as Minimal, Moderate and
896 Advanced (right panels; Supplementary Table 7). Box plots are also shown for
897 human blood samples of LTBI (non-progressors; n=217 biologically independent
898 samples) and LTBI_Progressor (individuals who developed TB, time point 1 to 180
899 days before diagnosis; n=17 biologically independent samples) from an independent
900 cohort (GSE79362, Zak *et al.* 2016[19]) (middle panels).
901
902
903
904
905
906
907
908
909
910
911 **Table legends**
912 **Table 1: Granulocyte associated genes** within the Inflammasome (HB3, left) and
913 the Innate immunity/PRR/C' (HB8, right) modules that are over-expressed in the
914 blood of TB patient cohorts from London, South Africa and Leicester, compared to
915 healthy controls (58 out of 87, and 53 out of 92 Granulocyte associated genes,
916 respectively).
917
918

(HB3) Inflammasome*			
ABCA13	DMXL2	KIF1B	NTNG2
AIF1	EVI2A	LCN2	OSM
APOBR	FAS	LILRB5	PLA2G4A
ASPRV1	FCGR3B	LPCAT2	PPP1R3D
ATP8B4	FGL2	LTF	PRTN3
CAMP	GAS7	LY96	RNASEL
CASP4	GPR141	MARCKS	S100A6
CCR1	GSN	MCEMP1	SELL
CD177	HIST1H2BC	MCTP1	SIGLEC9
CD300A	HP	MEFV	SOCS3
CKAP4	IL18RAP	MILR1	TFEC

CLEC4A	IL1B	MMP8	TLR5
CLEC4D	IRAK3	NAIP	VCAN
CLEC4E	KCNJ2	NCF1	
CLEC5A	KCTD12	NOD2	

919
920

(HB8) Innate immunity/PRR/C'***

ACSL1	FUT7	MXD1	REPS2
ALOX5AP	GAB2	MYBPC3	RNF19B
ANXA3	GLIPR2	NCF2	RRAGD
AQP9	HCAR2	P2RX1	S100A11
ARL11	HRH2	P2RY13	S100A8
BCL6	IFNGR2	PADI4	S100A9
BMX	IGSF6	PGLYRP1	SIPA1L2
BST1	ITGAM	PLBD1	SLC22A4
C1RL	JAML	PPP1R3B	STEAP4
C5AR1	LITAF	PTPRJ	TIMP2
CHST15	LRG1	PYGL	TLR2
CRISPLD2	LYN	RAB31	TLR6
ENTPD1	MMP9	RALB	
FOSL2	MNDA	RBM47	

921
922

923
924 **METHODS**
925
926 **Experimental animals and ethics.** C57BL/6J and C3HeB/FeJ mice were
927 purchased from Jackson Laboratories (Bar Harbour, ME) and housed under barrier
928 conditions in the Animal Biosafety Level 3 (ABSL3) facility at i3S, Porto, Portugal.
929 Experiments were performed in accordance with recommendations of the European
930 Union Directive 2010/63/EU and approved by Portuguese National Authority for
931 Animal Health – *Direção Geral de Alimentação e Veterinária*. (DGAV-
932 Ref.0421/000/000/2016). Mice were kept with food and water ad libitum and
933 humanely euthanized by CO_2 asphyxiation. Every effort was made to minimize
934 suffering. Age matched females were used in experiments.
935
936 **Mouse models of TB.** *M. tuberculosis* experiments were performed under ABSL-3
937 conditions. *M. tuberculosis* H37Rv (laboratory strain) and HN878 (clinical isolate)
938 were grown to midlog phase in Middlebrook 7H9 broth supplemented with 10% oleic
939 acid albumin dextrose complex (Difco), 0.05% Tween 80, and 0.5% glycerol before
940 being quantified on 7H11 agar plates and stored in aliquots at −80°C. Aliquots frozen
941 at −80°C were then thawed (6 aliquots) and quantified, to determine the
942 concentration of the stored inocula. Mice were infected via the aerosol route using an
943 inhalation exposure system (Glas-Col), calibrated to deliver from ~100 to 1000 CFUs
944 to the lung. The infection dose was confirmed by determining the number of viable
945 bacteria in the lungs of five mice 3 days after the aerosol infection (low dose: ~100-
946 450 CFUs/lung; high dose: ~700-900 CFUs/lung). Infected mice were monitored
947 regularly for signs of illness such as wasting, piloerection and hunching. Mice were
948 euthanized by CO_2 inhalation and blood and lung samples from each group were
949 collected from individual mice for RNA isolation, post *M. tuberculosis* infection at the
950 known peak of the blood transcriptomic response, or in the specific case of the
951 susceptible C3HeB/FeJ mice infected with HN878, when they showed signs of
952 severe illness (Supplementary Fig. 1a). Blood and lung samples from age matched
953 uninfected mice were collected at the same time and used as controls. Lung
954 samples from additional infected mice were collected for bacterial load determination
955 (Supplementary Fig. 1b) or histology (Figs. 3 and 5; Supplementary Fig. 6).
956 Determination of lung bacterial load was performed by plating serial dilution of the

957 organ homogenate on Middlebrook 7H11 agar supplemented with 10% oleic acid
958 albumin dextrose complex plus PANTA to prevent contamination with other
959 infections. CFUs were counted after 3 weeks of incubation at 37°C, and the bacterial
960 load per lung was calculated.
961
962 **Histopathological analysis of lung samples.** Lung tissues from *M. tuberculosis*
963 infected mice were perfused and fixed in 10% neutral-buffered formalin followed by
964 70% ethanol, processed and embedded in paraffin, sectioned at 4 µm and stained
965 with hematoxylin and eosin (H&E) or Ziehl-Neelsen (ZN). A single section from each
966 tissue was viewed and scored as a consensus by three board-certified veterinary
967 pathologists (S.L.P., E.H. and A.S.-B.) blinded to the groups (Supplementary Fig. 6).
968 Presence of *M. tuberculosis* bacilli detected by ZN positive staining was scored as
969 paucibacillary or multibacillary according to their abundance in the tissue. A semi-
970 quantitative scoring (0-4 points) method was devised to assess the following
971 histological features: inflammatory cells (neutrophils, lymphocytes, plasma cells,
972 macrophages), necrosis, pleuritis and fibrosis; using the following scale: 0 = not
973 present, 1 = minimal, 2 = mild, 3 = moderate, and 4 =marked changes. The relative
974 lesion burden scoring (0-5 points) was determined using the following scale: 0= no
975 lesions, 1 = focal lesion, 2 = multiple focal lesions, 3 = one or more focal severe
976 lesions, 4 = multiple focal lesions that are extensive and coalesce, and 5 = extensive
977 lesions that occupy the majority of the lung lobe. The percentage of tissue affected
978 was also scored and the lesion types graded as Type I, Type II and Type III as
979 previously described by Irwin *et al.*[50]. Representative images of each group were
980 acquired on an Olympus BX43 microscope using an Olympus SC50 camera and
981 cellSens Entry software (Ver. 1.18).
982 Lung lesion global score (Fig. 6a) was calculated by combining the relative
983 lesion burden and the percentage of tissue affected, scored from H&E stained lung
984 sections (Supplementary Fig. 6). Neutrophil and lymphocyte scores for H&E stained
985 lung sections (Supplementary Fig. 6) were plotted in Fig. 6c and 6e, respectively.
986
987 **Microscopy for neutrophils, T and B cells.** Lung sections from *M.*
988 *tuberculosis* infected mice were de-waxed and re-hydrated before staining. The
989 neutrophil staining was performed using the automated equipment Ventana
990 Discovery ULTRA. Sections were treated with Protease 1 at 37°C for 8 min for

991 antigen retrieval, incubated with primary antibody Rat anti-mouse 2B10 (in house
992 clone 2B10) at 37°C for 48 min, followed by *OmniMap anti-Rt HRP* (RUO) at room
993 temperature (RT) for 12 min. For T and B cell staining, sections were microwaved for
994 23 min (900W) with Citrate Buffer pH6 for antigen retrieval and then incubated with
995 primary antibodies Rabbit monoclonal anti-CD3G (clone ab134096, Abcam) and
996 biotin Rat anti-mouse CD45R/B220 (clone RA3-6B2, BD) for 1h at RT. Sections
997 were then incubated with donkey anti-Rabbit IgG (H+L) secondary antibody Alexa
998 Fluor™ 555 (A-31572, Invitrogen) and Streptavidin, Alexa Fluor™ 488 conjugate
999 (s32354, Invitrogen) for 45 min at RT, followed by DAPI for 15 min at RT and Sudan
1000 Black for 20 min at RT. Stained sections were mounted with Tissue-Tek® Glass™
1001 Pertex, examined and scored by two board-certified veterinary pathologists (S.L.P.
1002 and A.S.-B). Neutrophils were assessed semi-quantitatively (based on intensity of
1003 labelling) as follows: 0 = none present, 1 = low numbers, 2 = moderate numbers and
1004 3 = high numbers. Neutrophil viability was scored as viable (which label with IHC) or
1005 necrotic by assessing which subset dominated the stained tissue. Representative
1006 images of each group were acquired an Olympus BX43 microscope using an
1007 Olympus SC50 camera and cellSens Entry software (Ver. 1.18).
1008 For T and B cell quantification, slides were scanned using the objective
1009 magnification of 20x on an Olympus VS120-L100 Slide Scanner. T and B cells were
1010 assessed semi-quantitatively (based on positive labelling) as follows: 0 = none
1011 present, 1 = very low numbers, 2 = low to moderate numbers, 3 = moderate to high
1012 numbers and 4 = very high numbers. The T/B cell ratio (%/%) and presence of
1013 follicles with germinal centres were also scored for each slide.
1014
1015 **RNA isolation.** Blood was collected in Tempus reagent (Life Technologies) at 1:2
1016 ratio. Total RNA was extracted using the PerfectPure RNA Blood Kit (5 PRIME).
1017 Globin RNA was depleted from total RNA (1.5–2 µg) using the Mouse GLOBINclear
1018 kit (Thermo Fisher Scientific). Lungs were collected in TRI-Reagent (Sigma-Aldrich).
1019 Total RNA was extracted using the RiboPure™ Kit (Ambion). All RNA was stored at
1020 −80 °C until use.
1021
1022 **Quantity and quality of RNA samples.** Quantity was verified using NanoDrop™
1023 1000/8000 spectrophotometers (Thermo Fisher Scientific). Quality and integrity of
1024 the total and the globin-reduced RNA were assessed with the HT RNA Assay

1025 Reagent kit (Perkin Elmer) using a LabChip GX bioanalyser (Caliper Life
1026 Sciences/Perkin Elmer) and assigned an RNA Quality Score (RQS) or RNA 6000
1027 Pico kit (Agilent) using a BioAnalyzer 2100 (Agilent) and assigned an RNA Integrity
1028 (RIN) score. RNA with an RQS/RIN >6 was used to prepare samples for microarray
1029 or RNA-seq.
1030
1031 **Microarray.** cRNA was prepared from 200 ng globin-reduced blood RNA or 200 ng
1032 tissue total RNA using the Illumina TotalPrep RNA Amplification Kit (Ambion).
1033 Quality was checked using an RNA 6000 Nano kit (Agilent) using a BioAnalyzer
1034 2100 (Agilent). Biotinylated cRNA samples were randomized; 1.5 µg cRNA was then
1035 hybridized to Mouse WG-6 v2.0 bead chips (Illumina) according to the
1036 manufacturer's protocols.
1037
1038 **RNA-Seq.** cDNA library preparation: for blood and tissues, total/globin-reduced RNA
1039 (200 ng) was used to prepare cDNA libraries using the TruSeq Stranded mRNA HT
1040 Library Preparation Kit (Illumina). For cDNA library preparation of FACS sorted cells,
1041 total RNA (30–500 pg) was used to prepare cDNA libraries using the NEBNext®
1042 Single Cell/Low Input RNA Library Prep Kit NEBNext® Multiplex Oligos for Illumina®
1043 #E6609 (New England BioLabs). Quality and integrity of the tagged libraries were
1044 initially assessed with the HT DNA HiSens Reagent kit (Perkin Elmer) using a
1045 LabChip GX bioanalyser (Caliper Life Sciences/Perkin Elmer). Tagged libraries were
1046 then sized and quantitated in duplicate (Agilent TapeStation system) using D1000
1047 ScreenTape and reagents (Agilent). Libraries were normalized, pooled and then
1048 clustered using the HiSeq® 3000/4000 PE Cluster Kit (Illumina). The libraries were
1049 imaged and sequenced on an Illumina HiSeq 4000 sequencer using the HiSeq®
1050 3000/4000 SBS kit (Illumina) at a minimum of 25 million paired-end reads (75 bp/100
1051 bp) per sample.
1052
1053 **Microarray data analysis.** Microarray data was processed in GeneSpring GX v14.8
1054 (Agilent Technologies). Flags were used to filter out the probe sets that did not result
1055 in a "present" call in at least 10% of the samples, with the "present" lower cut-off of
1056 0.99. Signal values were then normalized using neqc function with default
1057 parameters from limma package (v 3.38.3) in R. This function performs background
1058 correction, quantile normalization and log2 transformation of intensity signals. For

1059 modular fold enrichment analysis, Illumina IDs were converted to Ensembl IDs using
1060 both the annotation file available from Illumina and biomaRt package (2.38.0) in R.
1061 Next, transcripts were filtered to select the 50% most variable probes across all
1062 samples.
1063
1064 **RNA-Seq data analysis.** Raw paired-end RNA-seq data was subjected to quality
1065 control using FastQC (Babraham Bioinformatics) and MultiQC[51].
1066 Trimmomatic[52] v0.36 was used to remove the adapters and filter raw reads below 36
1067 bases long, and leading and trailing bases below quality 25. The filtered reads were
1068 aligned to the *Mus musculus* genome Ensembl GRCm38 (release 86) using
1069 HISAT2[53] v2.0.4 with default settings and RF rna-strandedness, including unpaired
1070 reads, resulting from Trimmomatic, using option -U. The mapped and aligned reads
1071 were quantified to obtain the gene-level counts using HtSeq[54] v0.6.1 with default
1072 settings and reverse strandedness. Raw counts were processed using the
1073 bioconductor package DESeq2[55] v1.12.4 in R v3.3.1, and normalized using the
1074 DESeq method to remove the library-specific artefacts. Variance stabilizing
1075 transformation was applied to obtain normalized $\log_2$ gene expression values.
1076 Further quality control was performed using principal component analysis, boxplots,
1077 histograms and density plots. Differentially expressed genes were calculated using
1078 the Wald test in DESeq2[55]. Genes with $\log_2$ fold change >1 or <−1 and false
1079 discovery rate (FDR) p-value < 0.05 corrected for multiple testing using the
1080 Benjamini–Hochberg (BH) method[56] were considered significant. Log_2 fold changes
1081 in mouse blood, mouse lung and human blood datasets (Berry London, Berry South
1082 Africa and Leicester: GSE107995), using the top 50 intra-modular genes within
1083 selected human blood modules, were represented in heatmaps using the pheatmap
1084 package in R (Raivo Kolde (2019). pheatmap: Pretty Heatmaps. R package version
1085 1.0.12.) (Fig. 4; Supplementary Fig. 8 and 9). For lung module generation, and
1086 modular fold enrichment, only protein coding genes were considered
1087 (Supplementary Fig. 5). PCA plots were generated using prcomp function in R and
1088 plotted using ggplot2 package (H. Wickham. ggplot2: Elegant Graphics for Data
1089 Analysis. Springer-Verlag New York, 2016.).
1090
1091 **Cellular deconvolution.** Deconvolution analysis for quantification of relative levels
1092 of distinct cell types on a per sample basis was carried out on normalized counts

1093 using CIBERSORT[57]. CIBERSORT estimates the relative subsets of RNA transcripts
1094 using linear support vector regression. Mouse cell signatures for 25 cell types were
1095 obtained using ImmuCC[58] and grouped into 9 representative cell types based on the
1096 application of ImmuCC cellular deconvolution analysis to the sorted cell RNA-seq
1097 samples from the ImmGen ULI RNA-seq dataset (ImmGen Consortium:
1098 GSE109125; http://www.immgen.org) as previously described[31,59,60] (Supplementary
1099 Fig. 1d).
1100
1101 **Module generation.** Human blood modules were previously determined in human
1102 TB[16]. Weighted gene co-expression network analysis was performed to identify lung
1103 modules using the package WGCNA[61] in R. Modules were across all control and
1104 infected samples, using $\log_2$ RNA-seq expression values. The lung modules were
1105 constructed using the 10,000 most variable genes across all lung samples. A signed
1106 weighted correlation matrix containing pairwise Pearson correlations between all the
1107 genes across all the samples was computed using a soft threshold of $\beta = 22$ to reach
1108 a scale-free topology. Using this adjacency matrix, the topological overlap measure
1109 (TOM) was calculated, which measures the network interconnectedness[62] and is
1110 used as input to group highly correlated genes together using average linkage
1111 hierarchical clustering. The WGCNA dynamic hybrid tree-cut algorithm[63] was used to
1112 detect the network modules of co-expressed genes with a minimum module size of
1113 20, and deep split = 1. Lung modules were numbered ML1–ML27, and human blood
1114 modules previously found in human TB[16] were numbered HB1-HB23. An additional
1115 "grey" module was identified in lung modules (Supplementary Table 6, module titled
1116 NA), consisting of genes that were not co-expressed with any other genes. These
1117 grey modules were not considered in any further analysis. To create gene interaction
1118 networks, hub genes with the highest intramodular connectivity and a minimum
1119 correlation of 0.75 were calculated, with a cut-off of 50 hub genes, and exported into
1120 Cytoscape v3.4.0 for visualization.
1121 For checking either human blood modules into mouse data or mouse lung
1122 modules into human data, human Ensembl gene ID were translated into Mouse gene
1123 ID using BioMart to extract mouse ortholog genes (Supplementary Table 8).
1124
1125 **Modular annotation.** Lung modules were enriched for biological pathways and
1126 processed using IPA (QIAGEN Bioinformatics), Metacore (Thomson Reuters), and a

1127 careful manual annotation, by checking cell-type-specific enrichment and individual
1128 read counts. Significantly enriched canonical pathways, and upstream regulators
1129 were obtained from IPA (top 5). Modules were assigned names based on
1130 representative biological processes from pathways and processes from all three
1131 methods (Supplementary Table 5 and 6).
1132
1133 **Module enrichment analysis.** Fold enrichment for the WGCNA modules was
1134 calculated using the quantitative set analysis for gene expression (QuSAGE)[64] using
1135 the bioconductor package qusage v2.4.0 in R, to identify the modules of genes over-
1136 or under-abundant in a dataset, compared to the respective control group using
1137 $\log_2$ expression values. The qusage function was used with n.points parameter set to
1138 2^{15}. Only modules with enrichment scores with FDR *p*-value < 0.05 were considered
1139 significant, and plotted using the ggcorrplot function in R. Eigengene profiles, which
1140 are representative expression profiles for a given module in aparticular dataset, have
1141 been generated using the moduleEigengenes function from the WGCNA package
1142 and have plotted using ggplot2 package.
1143
1144 **Cell-type-specific enrichment.** Cell-type enrichment analysis to identify over-
1145 represented cell types in blood and lung modules was performed as previously
1146 described[31] using a hypergeometric test, using the phyper function in R. *p*-Values
1147 were corrected for multiple testing using the p.adjust function in R, using the BH
1148 method, to obtain FDR corrected *p*-values.
1149
1150 **Method for use of online WebApp**
1151 An online web application: https://ogarra.shinyapps.io/tbtranscriptome/ accompanies
1152 the manuscript to visualize the findings of the study. The app is subdivided into 4
1153 distinct pages that can be accessed through the tabs displayed on the top of the
1154 page, with a customized sidebar for user input on each page.
1155 Tab 1: "**Expression Table**" allows the user to visualize read counts, either as raw
1156 counts or log2 normalized expression values, in either the Mouse Blood TB, Mouse
1157 Lung TB, Human Blood TB (Leicester, London or South Africa) datasets. Each row
1158 represents a different gene, each column a sample in the corresponding dataset.
1159 The user can download the dataset into spreadsheet file format.
1160

1161 Tab 2: "**Average expression Table**" allows the user to visualize the average read
1162 counts by group, either as raw counts or log2 normalized expression values, in either
1163 the Mouse Blood TB, Mouse Lung TB, Human Blood TB (Leicester, London or South
1164 Africa) datasets. Each row represents a different gene, each column a group in the
1165 corresponding dataset. The user can download the dataset into spreadsheet file
1166 format.
1167
1168 Tab 3: **"Gene expression"** allows the user to visualize the expression of individual
1169 genes, either as raw or log2 normalized expression values, in either the Mouse
1170 Blood TB, Mouse Lung TB, Human Blood TB (Leicester, London or South Africa)
1171 datasets. Each dot represents the expression value for the chosen gene, in one
1172 sample.
1173
1174 Tab 4: **"Module profiles"** allows the user to visualize the expression profile
1175 (EigenGene from WGCNA R package), of a module he can select, either from
1176 Human Blood TB Modules (HB1-HB23)[16], Mouse Lung TB modules (ML1–ML27)
1177 derived *de novo* in this study, or Mouse Lung Disease modules (L1-L38)[31]. Each dot
1178 represents the EigenGene value for the chosen module, in one sample. A table
1179 below the plot displays all genes present within that module.
1180
1181 **Reporting Summary**
1182 Further information on research design is available in the Life Sciences Reporting
1183 Summary.
1184
1185 **Data availability**
1186 The materials, data and any associated protocols that support the findings of this
1187 study are available from the corresponding author upon request. The RNA-seq
1188 datasets have been deposited in the NCBI Gene Expression Omnibus (GEO)
1189 database with the primary accession number GSE140945 (TB mouse blood and
1190 lung). Publicly available datasets used in this study include GSE107995 (human TB
1191 datasets from Singhania *et al.* 2018[18]) and GSE79362 (human TB dataset from Zak
1192 *et al.* 2016[19]).
1193
1194 REFERENCES METHODS

50 Irwin, S. M. *et al.* Presence of multiple lesion types with vastly different microenvironments in C3HeB/FeJ mice following aerosol infection with Mycobacterium tuberculosis. Disease models & mechanisms 8, 591-602 (2015).
51 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047-3048 (2016).
52 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120 (2014).
53 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357-360 (2015).
54 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166-169 (2015).
55 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550 (2014).
56 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289-300 (1995).
57 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12, 453-457 (2015).
58 Chen, Z. *et al.* Inference of immune cell composition on the expression profiles of mouse tissue. Scientific reports 7, 40508 (2017).
59 Heng, T. S., Painter, M. W. & Immunological Genome Project, C. The Immunological Genome Project: networks of gene expression in immune cells. Nature immunology 9, 1091-1094 (2008).
60 Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. Cell 176, 897-912 e820 (2019).
61 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).
62 Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. BMC bioinformatics 8, 22 (2007).
63 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719-720 (2008).
64 Yaari, G., Bolen, C. R., Thakar, J. & Kleinstein, S. H. Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene-gene correlations. Nucleic Acids Res 41, e170-e170 (2013).

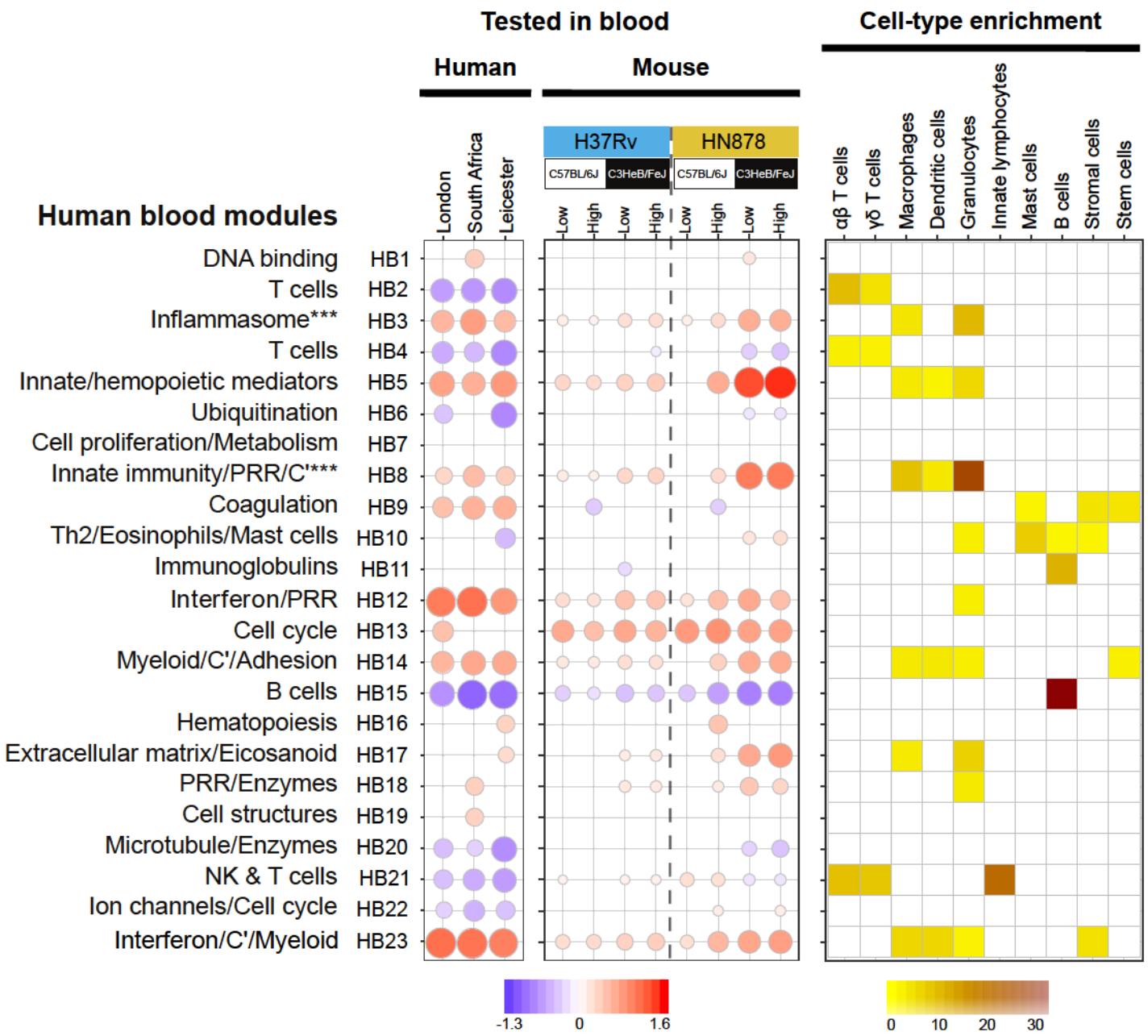
Tested in blood
Cell-type enrichment
Human
Mouse
H37Rv
HN878
C57BL/6J
C3HeB/FeJ
C57BL/6J
C3HeB/FeJ
London
South Africa
Leicester
Low
High
Low
High
Low
High
Low
High
αβ T cells
γδ T cells
Macrophages
Dendritic cells
Granulocytes
Innate lymphocytes
Mast cells
B cells
Stromal cells
Stem cells
Human blood modules
DNA binding HB1
T cells HB2
Inflammasome*** HB3
T cells HB4
Innate/hemopoietic mediators HB5
Ubiquitination HB6
Cell proliferation/Metabolism HB7
Innate immunity/PRR/C'*** HB8
Coagulation HB9
Th2/Eosinophils/Mast cells HB10
Immunoglobulins HB11
Interferon/PRR HB12
Cell cycle HB13
Myeloid/C'/Adhesion HB14
B cells HB15
Hematopoiesis HB16
Extracellular matrix/Eicosanoid HB17
PRR/Enzymes HB18
Cell structures HB19
Microtubule/Enzymes HB20
NK & T cells HB21
Ion channels/Cell cycle HB22
Interferon/C'/Myeloid HB23
-1.3
0
1.6
0
10
20
30

Figure1

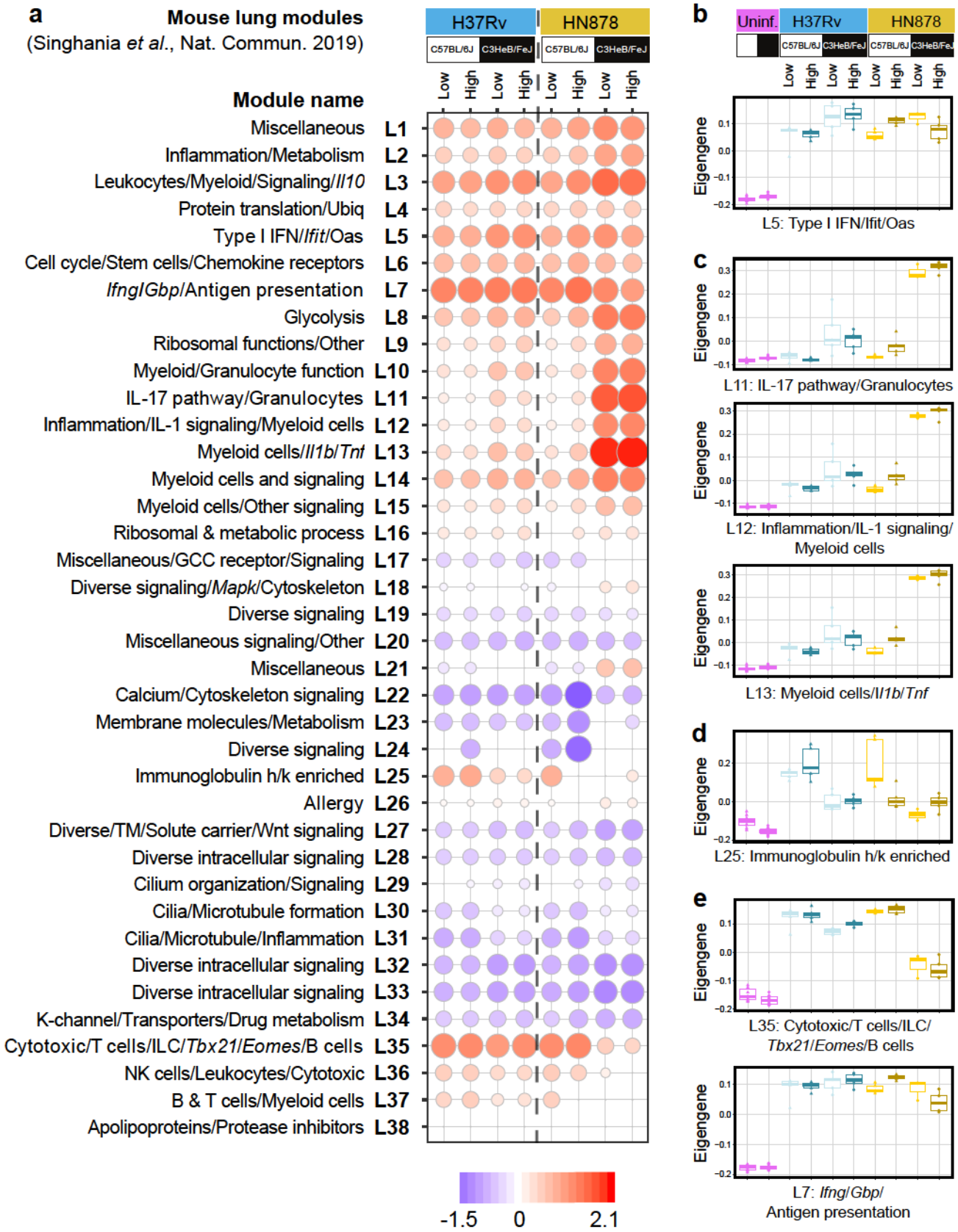
a
Mouse lung modules
(Singhania et al., Nat. Commun. 2019)
H37Rv
HN878
C57BL/6J
C3HeB/FeJ
Low
High
Module name
Miscellaneous L1
Inflammation/Metabolism L2
Leukocytes/Myeloid/Signaling/Il10 L3
Protein translation/Ubiq L4
Type I IFN/Ifit/Oas L5
Cell cycle/Stem cells/Chemokine receptors L6
Ifng/Gbp/Antigen presentation L7
Glycolysis L8
Ribosomal functions/Other L9
Myeloid/Granulocyte function L10
IL-17 pathway/Granulocytes L11
Inflammation/IL-1 signaling/Myeloid cells L12
Myeloid cells/Il1b/Tnf L13
Myeloid cells and signaling L14
Myeloid cells/Other signaling L15
Ribosomal & metabolic process L16
Miscellaneous/GCC receptor/Signaling L17
Diverse signaling/Mapk/Cytoskeleton L18
Diverse signaling L19
Miscellaneous signaling/Other L20
Miscellaneous L21
Calcium/Cytoskeleton signaling L22
Membrane molecules/Metabolism L23
Diverse signaling L24
Immunoglobulin h/k enriched L25
Allergy L26
Diverse/TM/Solute carrier/Wnt signaling L27
Diverse intracellular signaling L28
Cilium organization/Signaling L29
Cilia/Microtubule formation L30
Cilia/Microtubule/Inflammation L31
Diverse intracellular signaling L32
Diverse intracellular signaling L33
K-channel/Transporters/Drug metabolism L34
Cytotoxic/T cells/ILC/Tbx21/Eomes/B cells L35
NK cells/Leukocytes/Cytotoxic L36
B & T cells/Myeloid cells L37
Apolipoproteins/Protease inhibitors L38
-1.5
0
2.1
b
Uninf.
H37Rv
HN878
Eigengene
L5: Type I IFN/Ifit/Oas
c
L11: IL-17 pathway/Granulocytes
L12: Inflammation/IL-1 signaling/ Myeloid cells
L13: Myeloid cells/Il1b/Tnf
d
L25: Immunoglobulin h/k enriched
e
L35: Cytotoxic/T cells/ILC/ Tbx21/Eomes/B cells
L7: Ifng/Gbp/ Antigen presentation

Figure 2

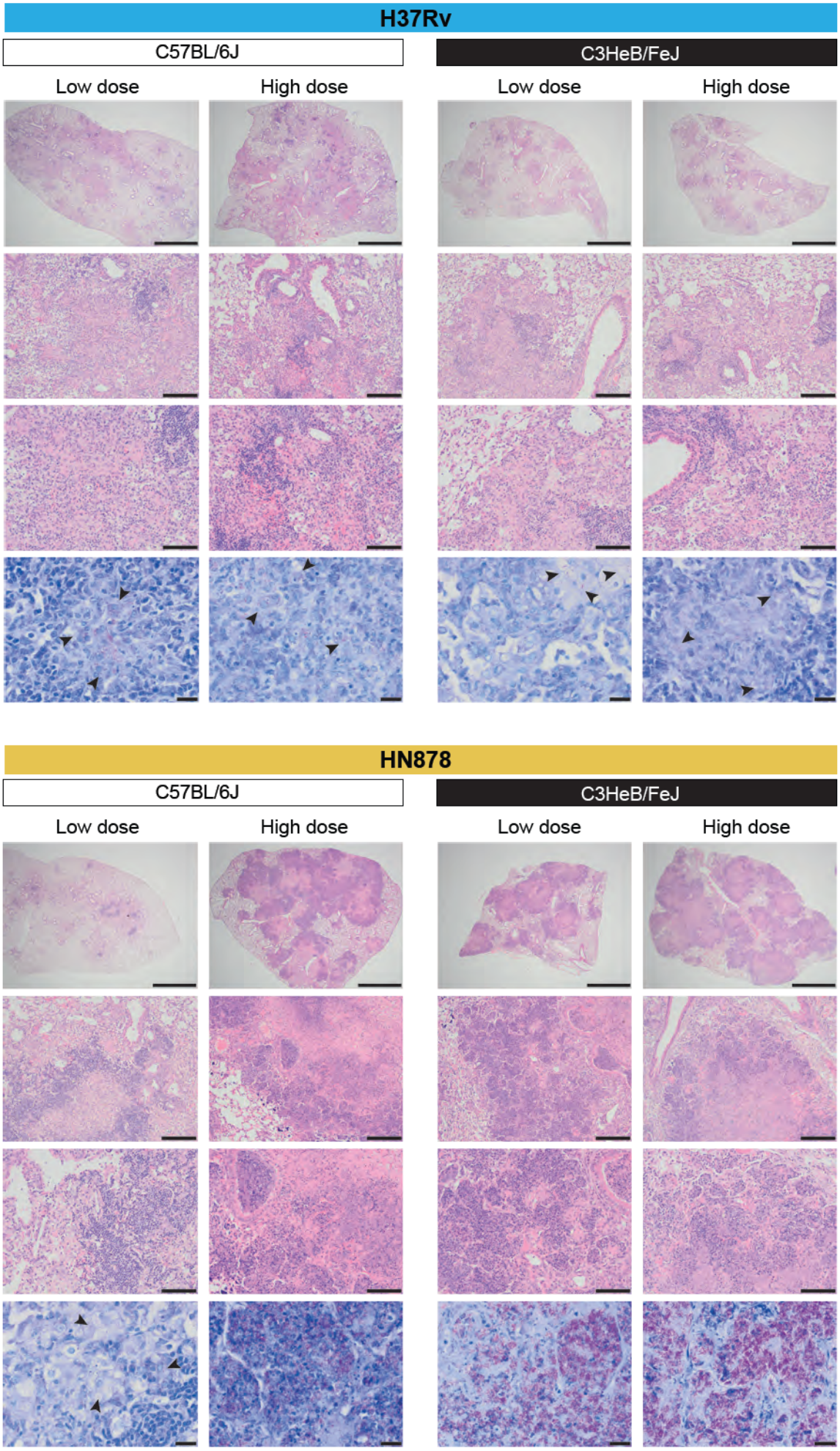
H37Rv
C57BL/6J
C3HeB/FeJ
Low dose
High dose
Low dose
High dose
HN878
C57BL/6J
C3HeB/FeJ
Low dose
High dose
Low dose
High dose

Figure 3

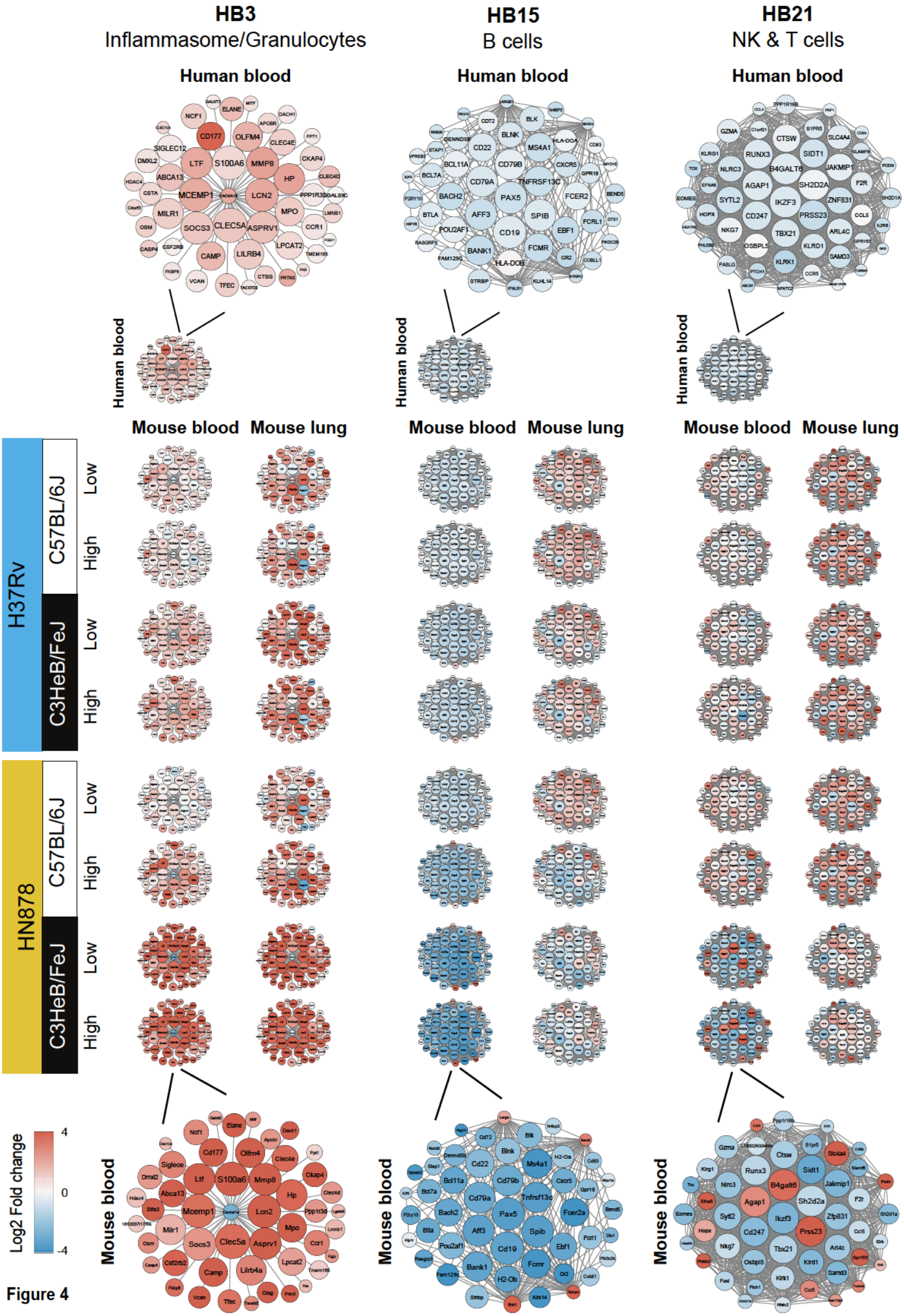
HB3
Inflammasome/Granulocytes
HB15
B cells
HB21
NK & T cells
Human blood
Human blood
Human blood
Human blood
Human blood
Human blood
Mouse blood
Mouse lung
Mouse blood
Mouse lung
Mouse blood
Mouse lung
H37Rv
C57BL/6J
C3HeB/FeJ
HN878
C57BL/6J
C3HeB/FeJ
Low
High
Low
High
Low
High
Low
High
Mouse blood
Mouse blood
Mouse blood
Log2 Fold change
4
0
-4

Figure 4

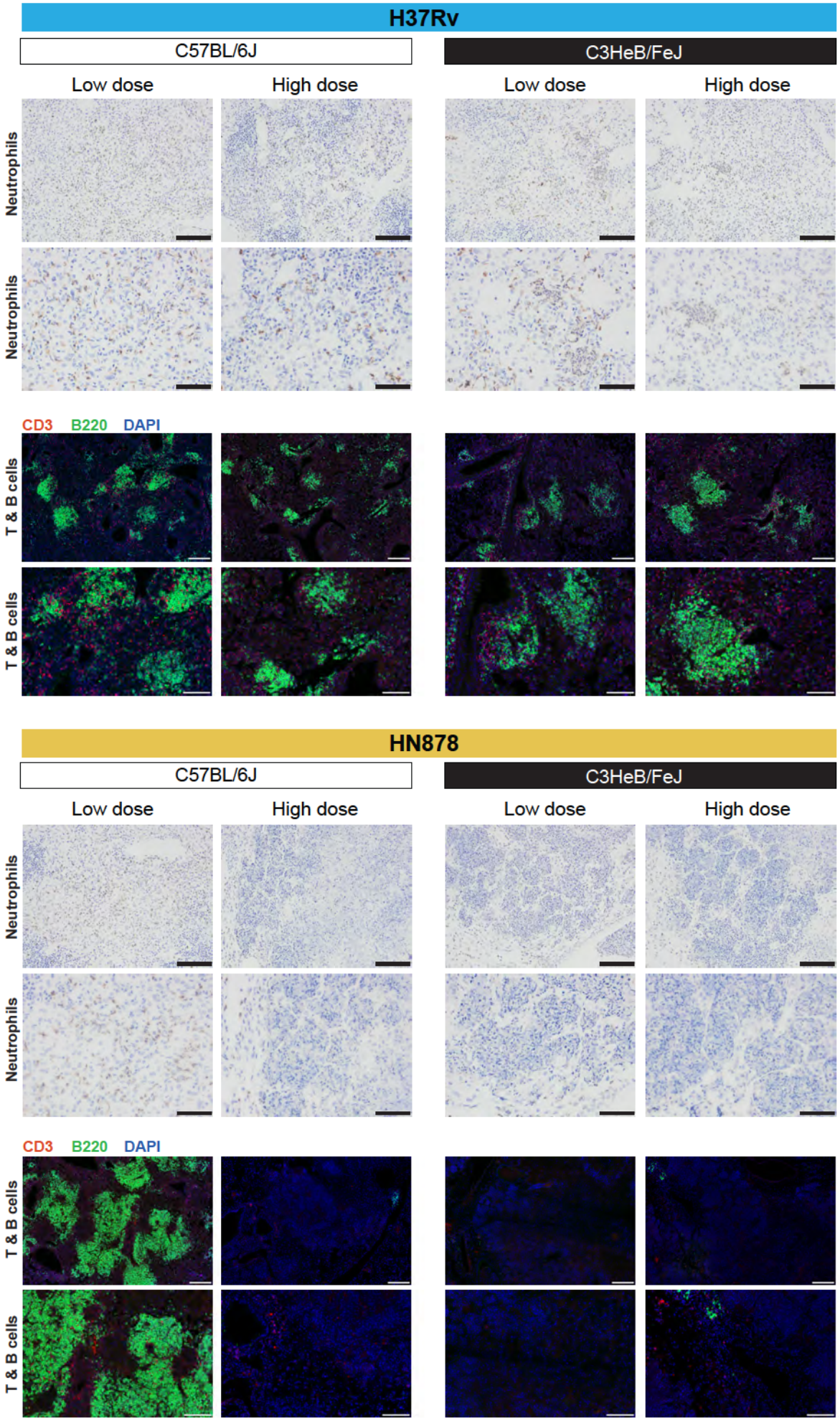
H37Rv
C57BL/6J
C3HeB/FeJ
Low dose
High dose
Low dose
High dose
Neutrophils
Neutrophils
CD3 B220 DAPI
T & B cells
T & B cells
HN878
C57BL/6J
C3HeB/FeJ
Low dose
High dose
Low dose
High dose
Neutrophils
Neutrophils
CD3 B220 DAPI
T & B cells
T & B cells

Figure 5

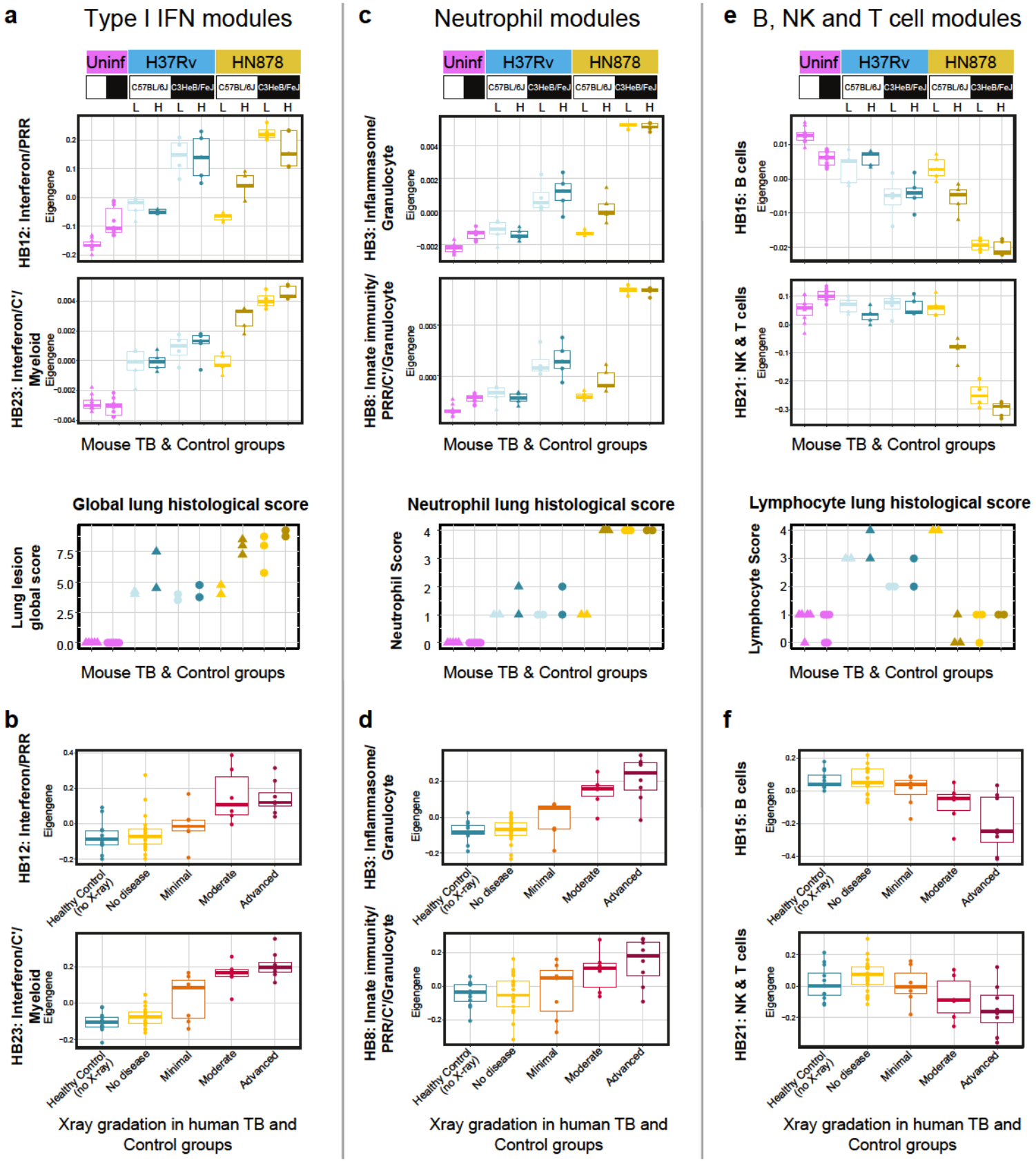
a
Type I IFN modules
c
Neutrophil modules
e
B, NK and T cell modules
Uninf
H37Rv
HN878
C57BL/6J
C3HeB/FeJ
L H L H L H L H
HB12: Interferon/PRR
HB23: Interferon/C'/Myeloid
HB3: Inflammasome/Granulocyte
HB8: Innate immunity/PRR/C'/Granulocyte
HB15: B cells
HB21: NK & T cells
Eigengene
Mouse TB & Control groups
Global lung histological score
Lung lesion global score
Neutrophil lung histological score
Neutrophil Score
Lymphocyte lung histological score
Lymphocyte Score
b
d
f
Healthy Control (no X-ray)
No disease
Minimal
Moderate
Advanced
Xray gradation in human TB and Control groups

Figure 6

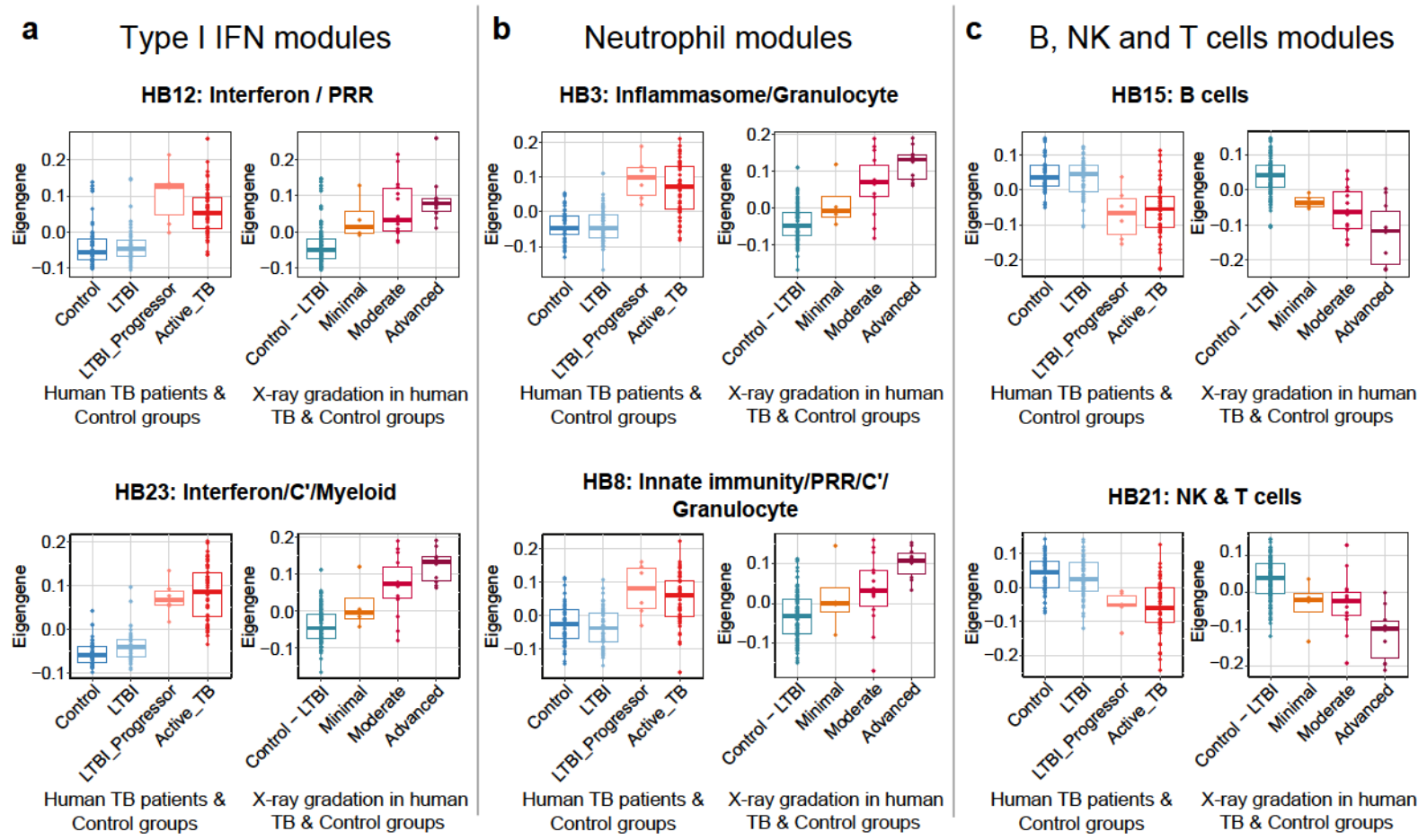
a
Type I IFN modules
HB12: Interferon / PRR
HB23: Interferon/C'/Myeloid
b
Neutrophil modules
HB3: Inflammasome/Granulocyte
HB8: Innate immunity/PRR/C'/ Granulocyte
c
B, NK and T cells modules
HB15: B cells
HB21: NK & T cells
Eigengene
Control
LTBI
LTBI_Progressor
Active_TB
Control – LTBI
Minimal
Moderate
Advanced
Human TB patients & Control groups
X-ray gradation in human TB & Control groups

Figure 7

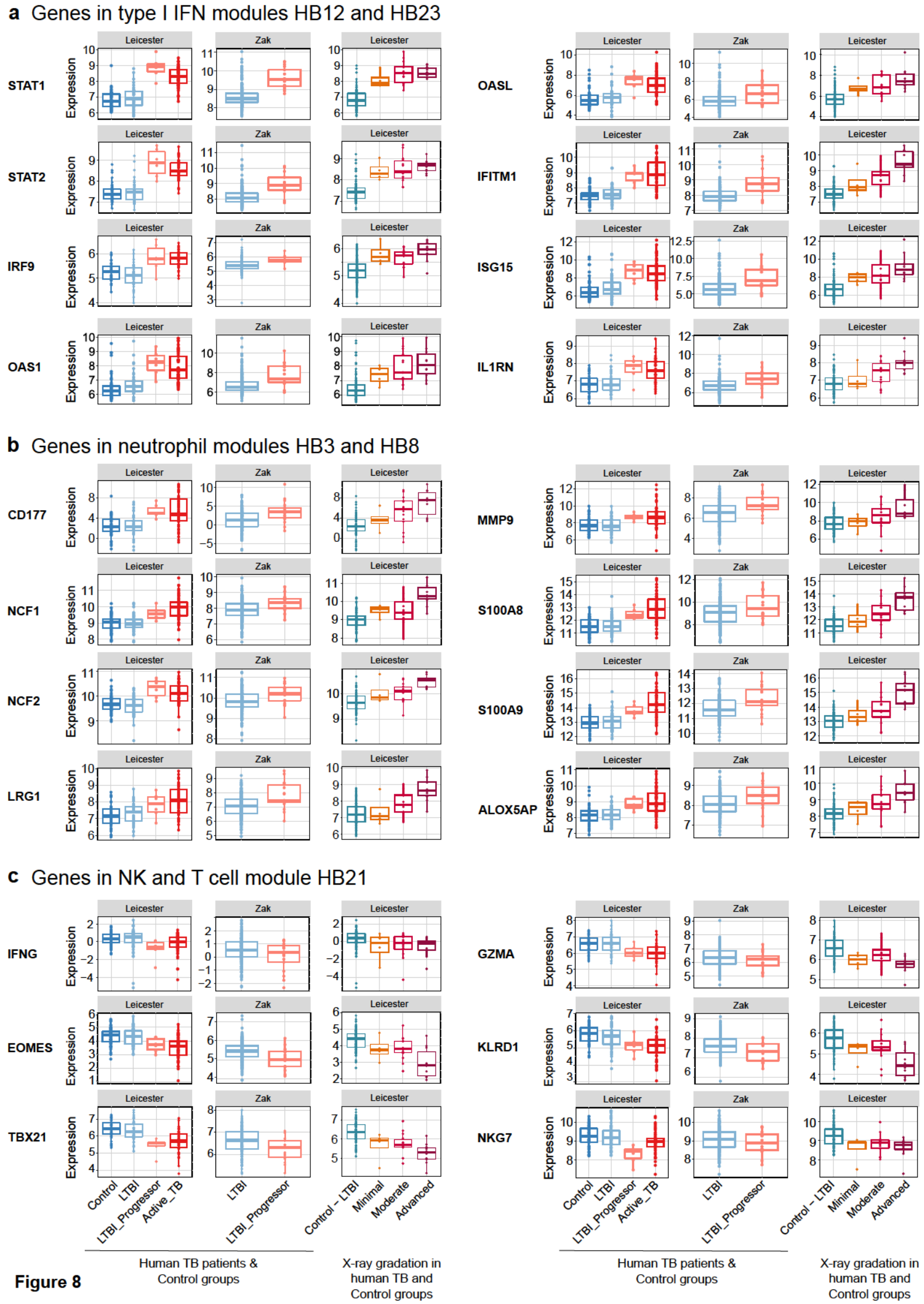
a Genes in type I IFN modules HB12 and HB23
STAT1
STAT2
IRF9
OAS1
OASL
IFITM1
ISG15
IL1RN
b Genes in neutrophil modules HB3 and HB8
CD177
NCF1
NCF2
LRG1
MMP9
S100A8
S100A9
ALOX5AP
c Genes in NK and T cell module HB21
IFNG
EOMES
TBX21
GZMA
KLRD1
NKG7
Leicester
Zak
Leicester
Expression
Control
LTBI
LTBI_Progressor
Active_TB
Control - LTBI
Minimal
Moderate
Advanced
Human TB patients & Control groups
X-ray gradation in human TB and Control groups

Figure 8