Title: Genetic and Environmental interactions contribute to immune variation in rewilded mice

Authors: Oyebola Oyesola[1#*], Alexander E. Downie[2#], Nina Howard[1], Ramya S. Barre[2], Kasalina Kiwanuka[1], Kimberly Zaldana[1], Ying-Han Chen[3], Arthur Menezes[2], Soo Ching Lee[1], Joseph Devlin[3], Octavio Mondragón-Palomino[1], Camila Oliveira Silva Souza[1], Christin Herrmann[3], Sergei Koralov[3], Ken Cadwell[3,5], Andrea L. Graham[2,4]*, P'ng Loke[1]*

Affiliations:

[1] Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institute of Health; Bethesda, MD, USA.

[2] Department of Ecology and Evolutionary Biology, Princeton University; Princeton, NJ, USA.

[3] Kimmel Center for Biology and Medicine at the Skirball Institute, New York University, Grossman School of Medicine; New York, NY, USA.

[4] Santa Fe Institute, Santa Fe, NM, USA

[5]University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA.

# These authors contributed equally

*Corresponding Author(s). Email: oyebola.oyesola@nih.gov, algraham@princeton.edu, png.loke@nih.gov.

One-Sentence Summary: Rewilding mice reveals how genotype and environment combine to determine inter-individual immune variation.

Abstract

The relative and synergistic contributions of genetics and environment to inter-individual immune response variation remain unclear, despite its implications for understanding both evolutionary biology and medicine. Here, we quantify interactive effects of genotype and environment on immune traits by investigating three inbred mouse strains rewilded in an outdoor enclosure and infected with the parasite, *Trichuris muris*. Whereas cytokine response heterogeneity was primarily driven by genotype, cellular composition heterogeneity was shaped by interactions between genotype and environment. Notably, genetic differences under laboratory conditions can be decreased following rewilding, and variation in T cell markers are more driven by genetics, whereas B cell markers are driven more by environment. Importantly, variation in worm burden is associated with measures of immune variation, as well as genetics and environment. These results indicate that nonheritable influences interact with genetic factors to shape immune variation, with synergistic impacts on the deployment and evolution of defense mechanisms.

INTRODUCTION

An individual's immune phenotype is shaped by some combination of genetic factors and nonheritable influences such as environmental exposure (including infection history and the microbiome) (1-9). However, the relative and potentially interacting contributions of heritable and nonheritable factors to inter-individual immune variation remain controversial despite the importance of such variation for both medicine and evolutionary biology. For example, variation in immune responses can determine whether an individual will experience severe or asymptomatic infection (10, 11), and whether severity arises due to failure to control pathogens or excessive collateral tissue damage following defective immune regulation (3).

Recent studies on the human immune system have aimed at identifying the relative contributions of genetic and environmental factors to variation in immune phenotypes among healthy individuals (1, 6, 7), as well as during infection (7) or inflammatory conditions (12). Such studies draw upon analysis of immunological divergence between identical twins (7) or genetic heritability estimation for immune traits through functional genomics (1). However, the design of these studies often makes the interactive effects of genetics and environment challenging to quantify(13, 14) and generally, has not been the well-examined in most immunological studies. For example, variation not attributable to genetics is generally attributed to environment alone rather than the possibility of Genotype-by-Environment interactions (G*E), which are inferred if effects of environment are differentially amplified in different genotypes, or vice-versa (15, 16). Important context-dependency in immune function (17) is thus often missing from these calculations. For example, what if the impact of environment upon memory T cell frequencies depends upon host genotype, or if, conversely, the impact of genotype upon memory T cell frequencies depends upon environment? Evolutionary biology is explicitly interested in such context dependencies because they provide the raw materials for adaptive evolution and diversification, indeed, genotype-by-environment interactions are common and substantial in effect for a variety of traits (18, 19) and disease outcomes (15, 16, 20). Here, we use mouse experiments together with statistical frameworks that quantify interactions in addition to the independent (or main) effects of genetics and environment to elucidate causes of variation in experimental immunology.

Controlled experiments with mice could help decipher interactions between genetic and environmental effects on the immune system, but most studies in mice instead aim to reduce

69 environmental variation to discover genetic factors regulating cellular and molecular components
70 of immunity (3, 21-23). Most times, this approach ignores interactions and provides only partial
71 insight into direct genetic effects by ignoring the extent of the measured genetic effect that is
72 mediated by the particular environment. We have taken a decidedly different approach of using an
73 outdoor enclosure system to introduce laboratory mice of different genotypes into a natural
74 environment that we term "rewilding" (24). In this particular study year, we tracked behavior
75 outdoors(25), recover the mice for analysis, and then investigate genetic and environmental
76 contributions to immune phenotypes (23, 24, 26). Previously, using mice with mutant alleles in
77 inflammatory bowel disease susceptibility genes (*Nod2* and *Atg16l1*), we found that the genetic
78 mutations affected the production of cytokines in response to microbial stimulation, whereas
79 immune cell composition was more influenced by environment (23, 26). We also found that
80 rewilded C57BL/6 mice become more susceptible to infection with the intestinal nematode
81 parasite *Trichuris muris* (24). However, those experiments explored limited genetic variation and
82 did not examine whether *interactions* between genetics and environment would influence immune
83 phenotype and helminth susceptibility (4, 27).

84 Here, to quantify relative and interactive contributions of genetic and environmental
85 influences on heterogeneity in immune profiles and helminth susceptibility, we compared
86 C57BL/6, 129S1, and PWK/PhJ mice kept in a conventional vivarium versus those that were
87 rewilded and measured inter-individual heterogeneity in immune cellular composition, cytokine
88 responses and parasite burdens. Single cell RNA sequencing (scRNAseq) from individual mice
89 also provided both cellular and functional readouts, which was associated with susceptibility to
90 parasite infection. Together, our results demonstrate that interactions between genetics and
91 environment are an important source of variation for specific immune traits, but there are also
92 tissue dependent differential effects of environment versus genetics on specific cellular
93 compartments such as T cells and B cells. Hence, we argue that our rewilding approach can
94 demonstrate that the effect of an extreme environmental shift on immune phenotype is modulated
95 by genetics, and the genetic effect is modulated by an environment. Such interactions are an
96 important source of inter-individual immune variation.
97

98 **RESULTS**

99 **Experimental Design**

100 To quantify sources of heterogeneity in immune profiles and helminth susceptibility, we
101 compared 3 different strains of mice - C57BL/6, 129S1, and PWK/PhJ mice – housed in two
102 different environments - a conventional vivarium but kept in summerlike temperatures and
103 photoperiods (hereafter "Lab" controls) versus those that were outdoor (hereafter "rewilded")
104 (Fig1A). These host strains differ by up to 50 million single nucleotide polymorphisms (SNPs)
105 and short insertions/deletions (indels) *(28, 29)*; for perspective on this range of variation, the entire
106 human population is estimated to contain approximately 90 million SNPs and indels *(30)*. We
107 rewilded mice (n = 72) or kept them in laboratory housing (n = 63) for 2 weeks and then infected
108 them with approximately 200 eggs of the intestinal helminth *Trichuris muris* (n = 61) or left them
109 uninfected (n = 74), returning them to the outdoor or vivarium environment for a further 3 weeks.
110 We conducted two replicate experiments across different periods during the summer months
111 (Block 1, n = 61, ending in July; Block 2, n = 74, ending in August). We collected blood and
112 mesenteric lymph nodes (MLN) for immune phenotyping and fecal samples for microbiota
113 analyses. To assess immune cell composition, we analyzed complete blood counts with differential
114 (CBC/DIFF) of total blood, peripheral blood mononuclear cells (PBMCs) by flow cytometry with
115 a lymphocyte panel (Table S1) and MLN cells with both a lymphocyte and myeloid cell panel
116 (Table S1 and S2). To assess cytokine responses, we measured plasma cytokine concentrations
117 and stimulated MLN cells with microbial antigens and measured cytokines released in the
118 supernatant. Single cell RNA sequencing (scRNAseq) of MLN cells enabled phenotyping of both
119 immune cell composition and function. We also assessed worm burden and worm prevalence in
120 all mice exposed to *T. muris* at day 19-21 post infection prior to full worm maturation, which was
121 necessary to prevent shedding of *T. muris* eggs into the rewilded environment. Serology and PCR
122 screening panels tested for over 30 pathogens indicated that the mice had no other detectable
123 infections (Table S3 and S4).

124 **Interactions between Genotype, Environment and Infection contribute to variation in**
125 **immune composition in murine PBMCs.**

126 Since most human immunology studies are conducted by analysis of PBMCs, first, we used
127 this same approach to assess the immune phenotype of laboratory and rewilded mice. The immune
128 cell composition of PBMCs in the peripheral blood were analyzed by spectral cytometry with a

129 lymphocyte panel (Table S1). To quantify the relative contributions of Genotype (i.e. strain),
130 Environment (i.e. Lab vs Rewilded) and Infection (i.e. exposure to *Trichuris muris*) and their
131 interactions to the high-dimensional spectral cytometry data from the PBMC analysis, we used
132 Multivariate Distance Matrix Regression (MDMR) analysis, a statistical approach used to
133 determine factors contributing to variations for high-dimensional data (31-33). The MDMR model
134 we used incorporated Genotype, Environment and Infection as fixed effects and the two
135 independent experiments in July or August (Block) as a random effect to calculate the interactive
136 and independent contribution of these factors to the outcomes. The MDMR model calculates the
137 effect size of each variable on the outcome measure (here, cellular composition based on spectral
138 cytometry data) to generate a pseudo-R^2 value that quantifies the effect of the predictor variable
139 on the dissociated outcome variable. The cellular composition data for the PBMCs of each
140 individual mouse is determined by unsupervised K means clustering to group cells into clusters
141 based on similarities of cellular parameters (Fig S1A). We calculate the composition of cells for
142 each individual sample based on cluster membership established by K means clustering and this
143 unbiased cluster composition data (Fig S1B) is then used as the outcome variable for the MDMR
144 analysis.

145 The results from the MDMR analysis on PBMC cellular composition (Table S5) showed
146 that Genotype and Environment had a significant effect on variation in cellular composition, not
147 only as independent variables (Fig. 1B top) but also through interactions between Genotype and
148 the Environment (Gen*Env) (Fig. 1B bottom). These patterns can be visualized through a principal
149 component analysis (PCA) on cellular composition data of individual mice (Fig. 1C, 1D, Fig. S1B
150 and Fig. S1C). The PCA plot indicated strong effects of environment on variation along the PC1
151 axis (Fig. 1C. and Fig. 1D) and of genetics on variation along the PC2 axis (Fig. 1C. and Fig. 1D),
152 while infection displayed minimal effect (Fig. 1B, Fig. S1D and Fig. S1E). Variation along the
153 PC1 axis for rewilded mice is substantially greater than for lab mice (Fig. 1C and 1D). The PCA
154 plot also suggested that variance on the PC2 axis between mouse strains was greater in laboratory
155 mice than for rewilded mice (Fig. 1C and 1D).

156 The loading factors in the PCA analysis showed that variation in expression of CD44 on
157 CD4+ T cells is important for driving the differences that defines genetic variation on the PC2 axis
158 (SFig. 1C). Although there are substantial difference in expression of CD44 on CD4+ T cells
159 between the in-bred strains for lab-housed mice, these differences were no longer present between

160 the C57BL/6 mice and the PWK/PhJ mice following rewilding (Fig1E). Hence, the genetic
161 differences seen in the clean laboratory environment can be reduced following rewilding. In
162 contrast, rewilded C57BL/6 mice had more $CD4^{+}Tbet^{+}$ cells after infection compared to the
163 PWK/PhJ mice, however, in laboratory conditions, there is no difference between these two strains
164 of mice (Fig. 1F). These results indicate that a stronger T_H1 response to *T. muris* in the C57BL/6
165 mice is only observed in the rewilding condition. Hence, genetic differences in response to
166 infection can sometimes emerge only in rewilding conditions. These results illustrate how
167 Gen*Env and Gen*Env*Inf interactions affect specific immune traits.

168 Complete blood count with differential (CBC/DIFF) is a standard clinical test used to
169 assess inflammatory responses in patients. CBC/DIFF of total blood cells collected at 2 weeks
170 and 5 weeks post rewilding provided data (Table. S6) on longitudinal changes in immune cell
171 composition during rewilding, allowing us to compare the acute effects of environmental change
172 (at 2 weeks post rewilding) to when the immune system has acclimatized to the new environment
173 (at 5 weeks post rewilding). At 2 weeks post rewilding, there is a significant effect of rewilding on
174 circulating neutrophils, lymphocytes and eosinophils across all genotypes (Fig. 1G). However, by
175 5 weeks post rewilding, we a trend towards more neutrophils in the rewilded PWK/PhJ mice than
176 the other strains (Fig. 1H), indicating that this genotype might have a larger effect on neutrophil
177 abundance especially in the rewilding environment. In contrast, eosinophils are more readily
178 induced by *T. muris* infection in the laboratory mice at 5 weeks (Fig. 1H), than in the rewilded
179 mice on the 129S1 and C57BL/6 background at 5 weeks. Together, these results indicated that
180 acute environmental change (at 2 weeks) had a bigger effect on total blood cell composition than
181 at 5 weeks. Additionally, infection induced responses in the laboratory setting can be altered during
182 rewilding in specific genotypes.

183 Overall, these results suggest that there are important context dependent effects of
184 genotype, environment, infection depending on the immune traits that are being measured in the
185 peripheral blood and the timing of the environmental change.

186 **Interactions between Genotype and Environment contribute to variation in immune**
187 **composition in murine secondary Mesenteric Lymphoid Organ.**

188 In contrast to what is obtainable in most human studies, murine studies give us an
189 opportunity to easily assess immune responses in secondary lymphoid organs such as the lymph
190 nodes. Therefore, we analyzed the mesenteric lymph nodes (MLNs) because they drain the

191 intestinal tissues, which are most affected by *T. muris* infection (in the cecum), as well as by
192 alterations to the gut microbiota. In contrast with the blood, MDMR analysis of immune
193 composition of the draining MLNs based on unsupervised clustering of cells, as explained above
194 with the lymphoid panel (Table S7), showed a significant effect of Genotype, Environment and
195 Exposure to *T. muris* (Infection) in determining immune variation (Fig. 2A, left). Critically, an
196 interactive effect of Genotype, Environment and Infection (Gen*Env*Inf) also contributed to the
197 variation in immune composition in the MLN (Fig. 2A, right). We could visualize the contribution
198 of Genotype, Environment and Infection with *T. muris* to MLN immune composition through the
199 PCA analysis of immune cellular compositional data from individual mice (Fig. 2B and SFig. 2A).
200 The PCA plot showed prominent effects of Genotype on variation along the PC1 axes (Fig. 2B
201 and Fig. 2C) with effects of Environment along the PC2 axes (Fig. 2B and Fig. 2C) and *T. muris*
202 Infection along the PC3 axes (Fig. S2B and S2C).

203 In addition to the independent effect of these factors, we also observed a G*E*I interaction
204 in that for example, *T. muris* infection had a significant effect on cellular composition of the
205 draining MLNs with increased proportion and sometimes abundance of B cells, especially in the
206 129S1 and the C57BL/6 strains, and especially following rewilding (Fig. S2D and S2E).
207 Additionally, we observed that the morphology of the MLNs was quite different among mouse
208 strains after rewilding, and this is reflected in the total cellular counts from the MLNs (Fig. 2D).
209 Here, we observed that the PWK/PhJ mice had smaller lymph nodes that were not expanded in
210 size compared to C57BL/6 and 129S1 mice after rewilding and *T. muris* infection, illustrating a
211 Gen*Env interaction that could be statistically quantified by MDMR (Fig. 2E).

212 Further, we also observed based on the loading factors from the PCA analysis (Fig. 2B
213 right) that the CD4 and B cell population in the MLN showed differential effect of the environment
214 versus genetics in driving immune variation in the MLN. For example, as noted in the blood (Fig.
215 1E), we also found that in the MLN, expression of CD44 on CD4 T cells was majorly influenced
216 by Genotype (Fig. 2F and Fig. 2G) with highest expression of CD44 on the PWK/PhJ mice across
217 all environment while expression of CD44 on B cells, which usually depicts antigen experienced
218 B cells (34) was predominantly influenced by Environment and exposure to *T. muris* parasites
219 (Fig. 2H and I) with rewilded *T.muris* exposed mice of all genotypes having a higher percentage
220 of CD44 expressing B cells than their counterparts in the vivarium (Fig. 2H, 2I and Fig. S2F).

221 These results indicate that Genotype and Infection (as well as Gen*Inf interactions) have a
222 more significant effect on immune phenotypes in the draining lymph nodes than in the peripheral
223 blood. Differences in lymph node size that are reflected in variation in lymph node cell numbers
224 between different genotypes, as well as the residence of *T. muris* in the cecum, illustrates how the
225 mesenteric lymph nodes provide a more complex picture of Gen*Env*Inf interactions than the
226 peripheral blood.

227 **Genotype has the biggest effect on plasma and MLN stimulated cytokine levels.**

228 Cytokine response profiling is a common approach for immune phenotyping patients to
229 characterize immune responses. Based on our previous study (23), we hypothesized that Genotype
230 would have a larger effect on cytokine responses than on immune cell composition. Here, we
231 measured both systemic circulating plasma cytokine levels in the plasma and cytokine production
232 into the supernatants following microbial stimulation of the MLN cells. Indeed, as hypothesized,
233 MDMR analysis of multiplex plasma cytokine data assessing systemic and circulating levels of
234 IL-5, IL-17a, IL-22, IL-6, TNF-a and IFN-g (Table S8) showed that there are no statistically
235 significant interactions among Genotype, Environment, and Infection (Fig. 3A, right); and the
236 main effect of Genotype explained more variance than Environment (Fig. 3A, left). However, there
237 is a strong effect of the different experimental blocks (Fig 3A), indicating that some unaccounted
238 environmental or technical factors could also contribute to the variation. This Genotype effect on
239 plasma cytokines (circulating levels of IL-5, IL-17a, IL-22, IL-6, TNF-a and IFN-g, Table S8) can
240 be visualized on the PCA plot, where the C57BL/6 strain either in the lab or rewilded setting
241 explained most of the difference on the PC1 axis (Fig 3B). Assessment of the loading factors
242 revealed that the IFN-gamma levels were important in driving this variance (Fig 3C). Analysis of
243 the cytokine data separately shows that the circulating IFN-γ levels were especially high in infected
244 C57BL/6 mice in both lab and rewilded (Fig 3D).

245 When we characterized cytokine responses in the supernatant after *in vitro* stimulation of
246 MLN cells, either with CD3/CD28 beads or with other microbial stimulants (LPS, *C. albicans*, *C.*
247 *perfrigens*, *B. vulgatus* and *T. muris* antigens) (SFig 3, Table S9), MDMR analysis revealed
248 Genotype having the biggest effect size on variation (Fig 3E, Left), which is consistent with the
249 analysis of plasma cytokines. However, MDMR analysis of MLN cytokine responses also showed
250 that the effect of Genotype on cytokine responses following stimulation of MLN cells with
251 microbial antigens can be modulated by Environment and Infection (Gen*Env and Gen* Inf

interactions) (Fig 3E, Right). When we focused our analysis on only cytokine responses to *Trichuris muris*, we confirmed that Genotype has the biggest effect on cytokine recall responses to *T. muris* antigen (Fig 3F, Left). However, this response also shows a significant effect of interactions between Gen*Env, Gen*Inf and Gen*Env*Inf (Fig 3F, Right). Analysis of the MLN supernatant cytokine data shows that consistent with the plasma cytokine data (Fig. 3D), production of IFN-γ from MLN cells also tends to be higher in C57BL/6 mice compared to the 129S1 strain of mice following *T. muris* infection in the lab or the rewilded environment (Fig. 3G and 3H).

Together, these results support our previous observations that genetics influence cytokine responses more strongly than the environment (23). However, here, we add evidence that the environment neither amplified nor eroded genetic effects on plasma cytokine levels, but that both environment and infection can modulate cytokine production in the MLN. Again, these results highlight the differences between analysis of the peripheral blood (the most accessible samples available in human studies), with analysis of draining lymph nodes that gather additional information on infections and the environment.

Single Cell RNA Sequencing Analysis of mesenteric lymph node (MLN) cells emphasizes contribution of interactions between environment and genetics in immune variation.

Single cell RNA sequencing (scRNA-seq) is an unbiased approach to profile immune phenotypes without pre-selection for analytes and markers of interest, and this approach can estimate both immune cell composition and provide insight into cytokine responses. Here, we used scRNA-seq to examine Gen*Env*Inf interactions on immune composition and cytokine responses in the MLN cells. Mesenteric lymph node cells (n=49,727) from individual mice (n=122) identified 23 major immune cell subsets visualized by UMAP (Fig. 4A). The cellular composition for each individual mice based on cluster membership with these 23 major immune cell subsets (Fig. S4A) is then used as the outcome variable for the MDMR analysis. In accordance with the cellular composition analysis with the flow cytometric data, MDMR analysis of the scRNAseq compositional dataset (Table S10) showed a significant effect of Genotype, Environment and Infection with *T. muris* in determining immune variation as fixed predictor variables in addition to a substantial block effect (Fig. 4B, above). Genotype and Environment (Gen*Env) interactions

282 also explained significant variation in immune composition as assessed by scRNAseq (Fig. 4B,
283 below). PCA analysis of the cellular composition from the single cell sequencing analysis (Fig
284 S4A) of the different individual mice reveals contributions of Genotype and Environment to the
285 variation among individual mice along the PC1 and PC2 axes (Fig. 4C). An example of how
286 genotype effects can be modulated by environment (Gen*Env interaction), can be observed in the
287 increase of Follicular B cells following rewilding that was especially heightened in C57BL/6 mice
288 (Fig. 4D). In contrast, decrease in CD4 T cell abundance from rewilded naïve mice occurred for
289 both 129S1 mice and C57BL/6 mice, but not PWK/PhJ mice (Fig. 4D and 4E). Overall,
290 examination of cellular composition in the MLNs by scRNAseq resulted in a similar conclusion
291 to spectral cytometry, in that Gen*Env interactions are particularly important. However, since we
292 did not perform scRNAseq on the peripheral blood, we could not directly compare if Gen*Env
293 interactions are more important in explaining variance in the cellular composition of MLNs than
294 peripheral blood with this approach.

295 To characterize the functional activity of the MLN cells, we used scRNAseq to identify the
296 cells expressing cytokine related genes. Based on Gene Ontology, we extracted data for 232 genes
297 defined to have molecular function in cytokine activity (GO:0005125), of which expression of 123
298 genes could be identified in the scRNAseq dataset (Table S11). Expression levels of these genes
299 (n=123) were used to subset and re-cluster the MLN cells, and they were visualized based on
300 expression of cytokine activity genes and their original cellular identity (Fig. 4F). Notably, CD4 T
301 cells and Follicular B cells, which are the largest cellular populations in the overall dataset had the
302 smallest percentage of cells that expressing cytokine genes (Fig. 4G), whereas CD8 effector cells,
303 plasmablasts, and dark zone germinal center B cells, which are less abundant in the total population
304 had higher proportions of cells expressing cytokine genes.

305 As described above, cells with cytokine activity were re-clustered based on their cytokine
306 activity profiles (Fig 4F) and cluster membership with these cytokine activity subsets (Table S12)
307 was then used as the outcome variable for the MDMR analysis. MDMR analysis showed that
308 Genotype had a significant effect on variation in cells with cytokine activity (Fig. 4H), which is
309 consistent with our previous work (23) and cytokine profiles described above. Also consistent with
310 this analysis, other variables like Environment, Infection with *T. muris* (Fig. 4H) and Gen*Env or
311 Gen*Env*Inf interactions (Fig. S4B) had no significant effect on variation in cellular cytokine
312 activity as assessed by scRNA-Seq. The Genotype effect can be observed by plotting the

313 percentage of MLN cells with cytokine activity for individual mice, with the 129S1 and PWK/PhJ
314 mice having more cells expressing genes for cytokine activity than the C57BL/6 mice (Fig. 4I).
315 PCA visualization of cellular composition based on cluster membership with cells of similar
316 cytokine activity also showed distinct Genotype differences along the PC1 axis (Fig S4C).

317 Hence, an unbiased scRNAseq approach supports the conclusion that Genotype has the
318 biggest effect on cytokine response heterogeneity, whereas cellular composition is more driven by
319 interactions between Genotype and the Environment.

320 **Variation in *Trichuris muris* worm burden is determined by genetic, environmental and**
321 **immunological factors.**

322 Ultimately, the question remains how the variance in these genetic, environmental and
323 immunological factors influences susceptibility to subsequent infection. Therefore, we
324 investigated predictors of worm burden (Fig. 5) and the contribution of genetics, environment and
325 the different immunological factors to susceptibility to worm infection. Here, we observed that
326 despite all 74 *T. muris*-exposed mice receiving approximately the same infectious dose (200 eggs),
327 worm burden was negative-binomially distributed among exposed mice (Fig 5A). Analysis of
328 worm burdens was done using Generalized Linear Models with a negative binomial error
329 distribution. We found a significant Gen*Env for worm burden (Fig 5B, *p=0.04015*), whereby
330 C57BL/6 mice harbored more worms than the other genotypes in the vivarium, but rewilding was
331 associated with higher worm burdens in all genotypes. In other words, the relative susceptibility
332 of the different host strains to *T. muris* depended upon environment (paralleling (35)). When we
333 used logistic regression to analyze worm presence/absence at the experimental endpoint (reported
334 as prevalence of infection among exposed mice in Fig. 5B), significant effects in the best model
335 included only main effects of Genotype (*p=0.0001221*) and Environment (*p=0.0044835*), plus a
336 significant effect of replicate experiment (*p=0.0329262*).

337 Interestingly, when we included PC1 and PC2 values from the MLN scRNA-seq analysis
338 (Fig. 5B) as summary measures of immune variation among individual mice, significant effects in
339 the best model of worm burden (Fig. 5C) included only main effects of Genotype *(p=0.0003322)*,
340 Environment (*p=0.0015615*), and PC2 scores *(p=0.0108213)*, which had a significant negative
341 association with worm burden. Loading factors on the PC2 axis (Fig. 5D) indicated that the dearth
342 of T cells with an interferon signature (*T.IFN*) may be a driver of the relationship between high
343 PC2 scores and decreased worm burden. Furthermore, the fact that PC2 statistically outcompeted

344 Gen*Env suggests that environment-dependent differences among genotypes in worm burdens
345 may hinge on immune factors captured on PC2 (Fig5D). These results are consistent with T_H1
346 responses being associated with increased susceptibility to helminth colonization *(36, 37)* and
347 suggest that despite complexities in how immune phenotype is influenced by genetics and
348 environment, once that immune phenotype emerges, established "rules" of infection susceptibility
349 apply (as in (24)). The results also demonstrate that genetic, environmental, and individual immune
350 variation is associated with varied infection burden.

351

352 **DISCUSSION**

353 Our results support the hypothesis that the effect of even an extreme environmental shift
354 on immune traits is modulated by genetics, and such modulations of phenotype through
355 interactions between environment and genotype are an important source of variation in immune
356 phenotypes. While we previously proposed that the immune cell composition for an individual is
357 shaped by the environment *(23)*, we find here that cellular composition in the peripheral blood is
358 shaped by interactions between genetics and the environment and those in the mesenteric lymph
359 node can shaped by interactions between genetics, environment and ongoing intestinal helminth
360 infection. The role of genotype in shaping the structural lymph node non-immune cell environment
361 as well as the lymphoid micro-environment as mediated by cytokines, chemokines and other
362 stromal factors could explain the bigger effect of genotype in the lymph nodes compared to the
363 peripheral blood. It is unsurprising that draining lymph nodes are more likely to detect and report
364 immune phenotype changes that result from local tissue infections than the systemic circulatory
365 immune cell populations. The complexity of Gen*Env*Inf interactions has important
366 ramifications for the course of natural selection on the immune system, diversity in immune
367 genotypes and efficacy of vaccine to pathogens. For example, because any given genotype may
368 produce different immune responses in different environments, environment can alter the ability
369 of individuals to resist and tolerate infections; furthermore natural selection operating on such
370 variation is likely to generate divergent alleles and allele frequencies in different environments
371 *(17)*.

372 Quantification of such interactions is rare in immunological studies and is a valuable step
373 forward in understanding the evolution and function of the immune system. The rewilding
374 approach present a unique strategy for combining controlled mouse experiments with the

375 advantages of tissue accessibility and homozygous mouse genetics, with increasing sophistication
376 of multi-dimensional immune phenotyping analyses that have already been applied in human
377 immunology. Based on these combinations, we do also identify traits for which main (i.e., non-
378 interaction) effects are dominant. For instance, heterogeneity in cytokine responses shows a
379 stronger influence of genetics, consistent with human studies (38). Furthermore, the Human
380 Functional Genomics Project produced similar results to ours, that variation in T cell phenotypes
381 is relatively more influenced by genetics, while B cell phenotypes are relatively more influenced
382 by non-heritable environmental factors (1). It is thus likely that genetics and environment likewise
383 contribute to human immune phenotype in synergistic ways. However, it is also possible that
384 differences in interpreting genetic vs environmental contribution to human immune variation arise
385 because of a focus on different immunological readouts (2, 3). We also found here that the
386 contributions of Genotype, Environment and Infection could vary depending on tissue site
387 analyzed, and since most human studies utilize peripheral blood, the effects of infection status
388 depending on the infectious agent may appear less pronounced than if tissue samples were
389 analyzed.

390 In comparing our results with human studies, it is also worth noting that human populations
391 harbor greater heterozygosity and rarely undergo such dramatic environmental shifts as the
392 laboratory mice being released into the rewilding environment. Longitudinal studies on travelers,
393 refugees or immigrants may perhaps reveal similar alterations in immune phenotypes driven by
394 environmental changes. Furthermore, the effect size of genotype versus environment on immune
395 profile might also be influenced by other factors such as the age of the individual when the
396 environmental change occurs. For example, a newborn might be more influenced by
397 environmental factors immediately after birth where they are exposed to various environmental
398 and microbial antigens as opposed to an adult (39-42). All the mice used in this study at the point
399 of rewilding were all between 5-10 weeks of age, reflecting a young adult population in
400 humans(43). Thus, perhaps releasing much younger mice or allowing sexual reproduction to occur
401 in the rewilded environment might enhance the effects of environment; this remains a question for
402 further investigation. Others have shown that interindividual variation tends to accumulate with
403 age(44), thus, it will be interesting to evaluate the how other factors such as age can interact with
404 environment and genetics to influence immune variation.

405 Conversely, using inbred strains of mice with homozygous alleles may represent an
406 extreme test of genetic influences on immune variation. Compared to the inbred strains of mice
407 that we have used in this study; most human genomes exist in a predominantly heterozygous state.
408 Hence, despite these important differences between our approach and studies on human
409 populations, there is surprising consistency between our findings here and previous conclusions
410 from studies on human immune variation describing the differentially role of environment and
411 genetics on B and T cell immune cell trait as well as the role of genetics in determine variation in
412 cytokine responses (1, 38). However, our analysis and experimental design also present us an
413 opportunity to assess the contribution of interaction to various immune traits and in different tissue
414 sites. Together, this suggest that the rewilding experiments might be a bridge to study human
415 immune response in murine models(21).

416 These experiments were performed in two experimental blocks over one summer. The
417 effects of these two experimental blocks could either be technical variation in the analysis of
418 samples, or biological variation resulting from changes to the environment over the two periods.
419 Indeed, the block effects that we discuss in results such as the CBC/DIFF, cytokine levels and
420 worm burden data could be driven by any number of technical or biotic factors that influence the
421 temperature, diet and other commensals, which could subsequently influence immune responses
422 in the mice, as well as differences in worm burden. With some assays (e.g. FACS and cytokine
423 responses after MLN stimulations), the effects of experimental block is too great, so we analyzed
424 the data generated as separate experiments instead of pooling data. Nonetheless, it is important to
425 note that by including and removing the block effects statistically, we can quantify independent
426 effects of Gen, Env and Inf, as well as the interactions between these variables while excluding
427 experimental variation from the block effects. These findings further emphasize the role of
428 environmental changes in influencing immune outcomes. A block effect could also explain some
429 of the difference between the effect of rewilding on *T. muris* worm burden previously reported
430 (24) and our findings within this study. In general, while some previous observations such as the
431 significant expansion of the neutrophil pool across all genotype previously reported (26) and
432 qualitative effect of rewilding on proportion of mice with *T. muris* (24) remain consistent across
433 experimental time lapse, other outcomes such as number of worms per mice especially in the
434 rewilded environment seems to be more dependent on the block effect and further emphasizing
435 effect of the environment on immune outcomes.

436 Variation in adult worm burden of soil transmitted helminths between individuals typically
437 follow a negative binomial distribution in humans (45) and wild animals (46). It is interesting to
438 note that our quantification of larval stages in the process of being expelled also follow a negative
439 binomial distribution. This raises the possibility that the negative binomial distribution in worm
440 burdens in natural population may result from immune consequences of Gen*Env interactions as
441 much as from differences in egg exposure. Hence, in addition to studies demonstrating that host
442 genetics (47-49) and environmental factors (24) could influence susceptibility to helminth
443 infection, we show here that interactions between genetics and environmental factors could also
444 influence helminth infection outcomes. Variation in outcome to helminth infection is complex and
445 maybe influenced by various other factors such as heterogeneity in parasite genetics factors (48-
446 50), parasite dose and frequency of parasite exposure (51-54), host microbiome factors (55-57) as
447 well as an individual's infection history(4, 54) in nature. Importantly, natural helminth infection
448 typically occurs from trickle infection of multiple small doses of egg exposures; therefore, the
449 experimental system here of a high dose *T. muris* infection may not be representative a real-world
450 exposure.

451 Critically, in this study, we also show that even systems such as this, where interactions
452 between genetics and environment are important, the basic Th1 versus Th2 immunological
453 mechanisms that govern susceptibility to *Trichuris muris* infection still predominate (24, 36).
454 Thus, in addition, this study highlights a system where basic immunological mechanism that are
455 established in the specific pathogen free facilities can be rigorously tested in a more naturalized
456 system(21).

457 An interesting observation from these studies is the reduction in genetically driven immune
458 phenotype differences in laboratory mice under rewilding conditions. This observation may be
459 relevant to the different prevalence of inflammatory conditions across geographical locations –
460 e.g., immune phenotypes may be more extreme in the absence of intensive microbial exposures
461 and therefore have a greater impact in genetically susceptible individuals. One explanation of the
462 hygiene hypothesis or the old friend's hypothesis is that improved immune-regulatory responses
463 through microbial exposure reduces the prevalence of inflammatory conditions (58-61). Our
464 results also raise the possibility that increased microbial exposure may sometimes normalize or
465 reduce the variation of immune phenotypes hence reducing the number of individuals with extreme
466 immune responses. While the mechanism of this normalization is unclear, it is certainly possible

467 that activation of immune-regulatory mechanisms through microbial exposure may normalize
468 immune variation instead of simply reducing immune responses.

469 In conclusion, our results highlight how rewilding mice with controlled genetic
470 backgrounds could be a bridge towards understanding immune variation between human
471 individuals, and that quantification of the interactions at this interface may help elucidate the
472 evolution of the immune system.

473

Figure and Figure Legends

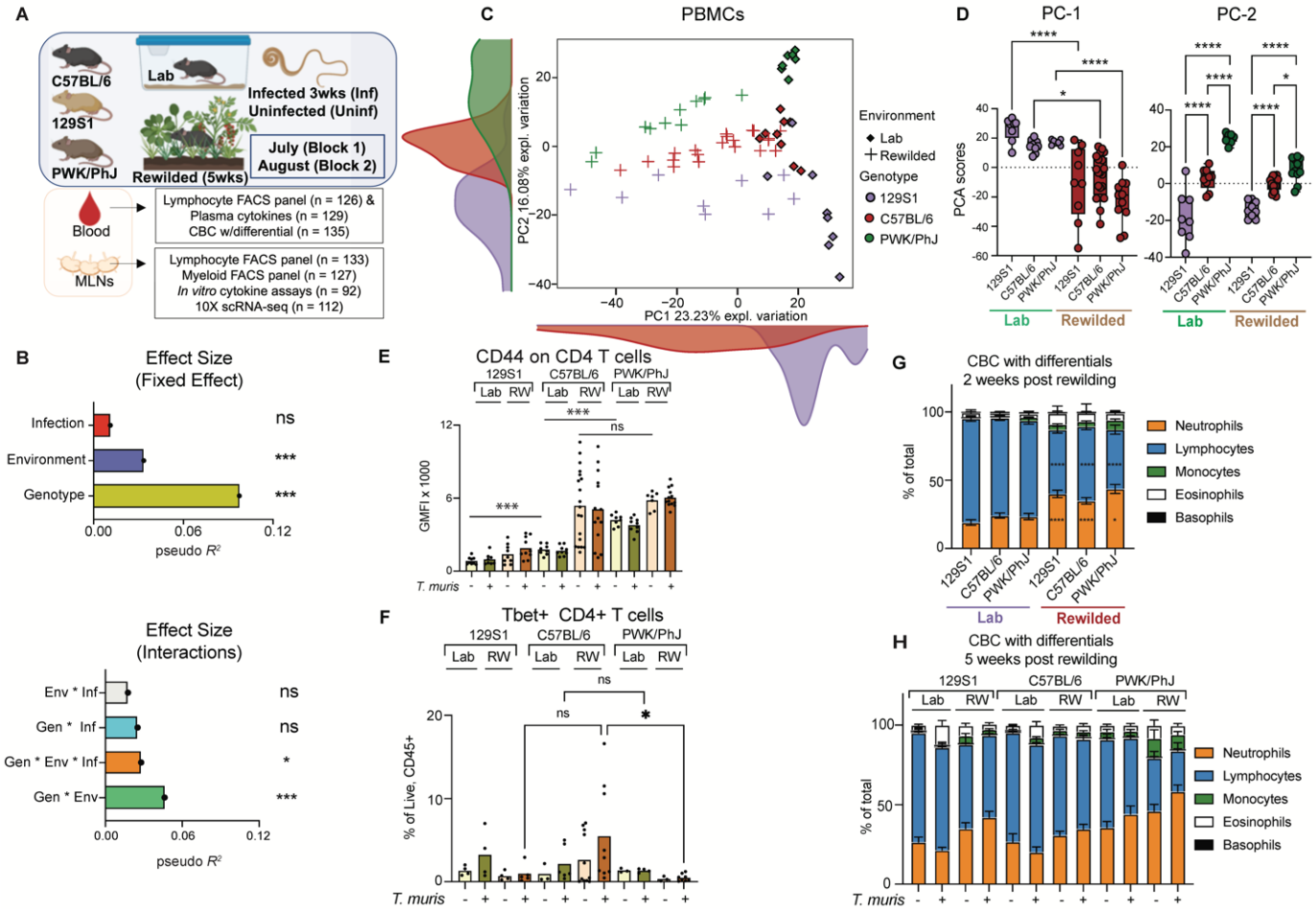


Fig. 1: Interactions between Genotype and Environment contribute to variation in immune composition in murine PBMCs. (**A**) 51 C57BL/6, 41 129S1, 43 PWK/PhJ different mice strains (total =135) were used in these experiments. Some were kept in conventional vivarium (n = 63), and some were housed in the wild enclosure (Rewilded), (n = 72) for total of 5 weeks. In addition, some were exposed

479 to 200 *T. muris* L3 eggs (n =61) while the others were left unexposed (n = 74). Experiments was repeated twice in July, Block 1 (n = 61)
480 and August, Block 2 (n = 74). Blood was collected for Flow cytometry, Plasma cytokine assessment and CBC analysis. MLN cells were
481 also collected for Single cell RNA sequencing (ScRNA-seq), Flow cytometry with two different panels - lymphocyte and myeloid panel
482 as well as cytokine profiling of supernatants from MLN stimulated cells. For Flow cytometry visualization, only Block 2 was used for
483 downstream analysis. (**B**) Bar plots showing the pseudo R^2 measure of effect size of predictor variables and interactions as calculated by
484 multivariate distance matrix regression analysis (MDMR) (**C**) PCA of immune cell clusters identified by unsupervised clustering in the
485 blood and the density of each population along the PCA and (**D**) Bar plot showing variance on PC1 and PC2 axis of PCA plots in (B). (**E**)
486 Barplot showing GMFI of CD44 on blood CD4+ T cells, (**F**) % of Tbet+ CD4 T cells of Live, CD45+ T cells. Bar plots showing percentage
487 of Neutrophils, Lymphocytes, Monocytes, Eosinophils, and Basophils out of total at (**G**) 2 weeks post re-wilding and (**H**) 5 weeks post re-
488 wilding based on assessment by CBC with differentials. Statistical significance was determined by a based on MDMR analysis with R
489 package (**B**) or based on one-way ANOVA test between different groups with Graph-Pad Software (**D**), (**E**) and (**F**). For (**E**) and (**F**) direct
490 comparison was done between groups of interest with one-way ANOVA test. Data are displayed as mean ± SEM. *ns* $p>0.05$; * $p<0.05$;
491 ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$
492
493

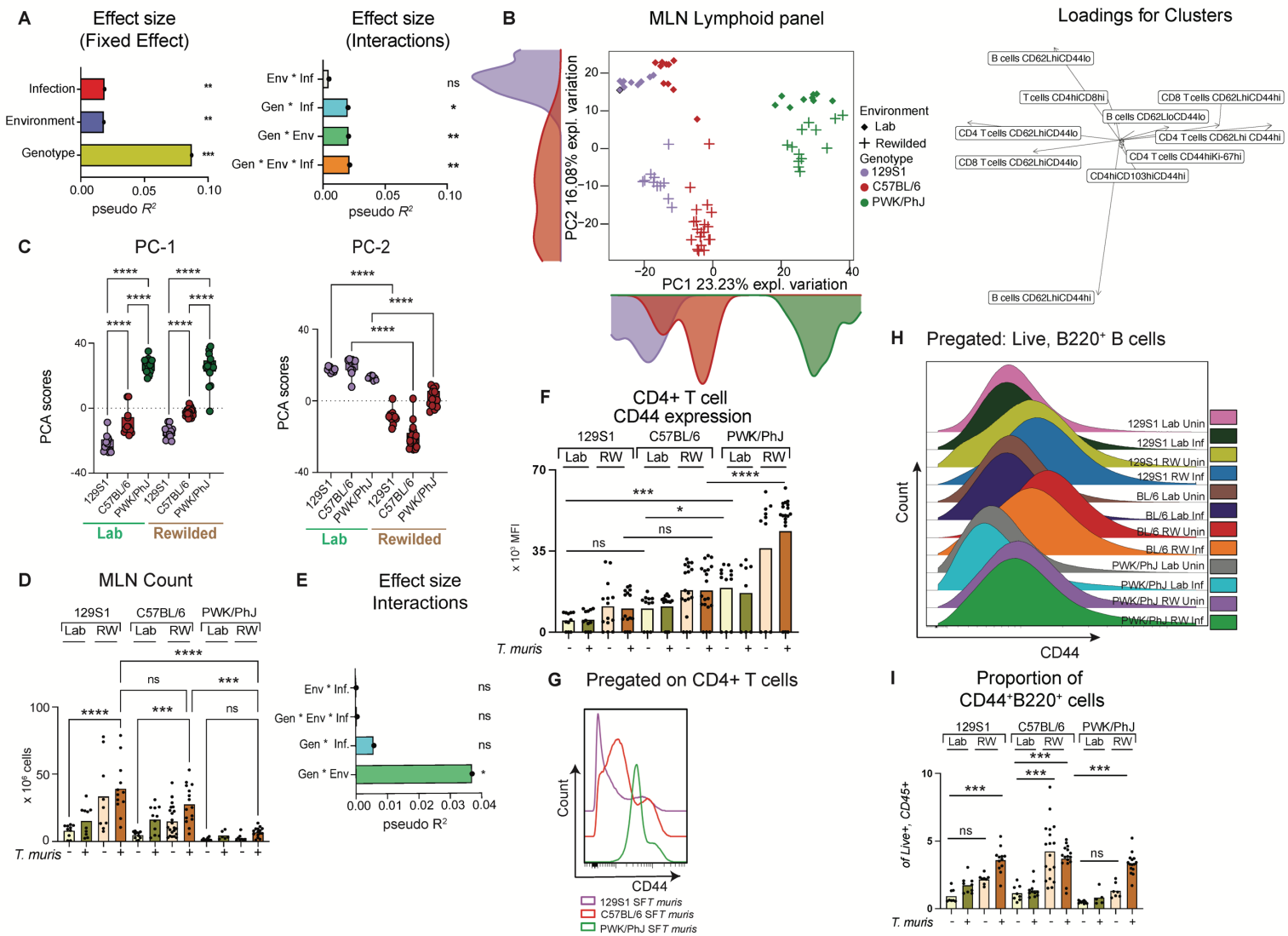


Fig. 2: Interactions between Genotype, Environment and Infection contribute to variation in immune composition in murine MLNs. (**A**) Bar plots showing the pseudo R^2 measure of effect size of predictor variables and interactions as calculated by multivariate

497 distance matrix regression analysis (MDMR). (**B**) PCA of immune cell clusters identified by unsupervised clustering in the MLN with
498 the lymphoid panel and the loading factor of each population along the PCA. (**C**) Bar plot showing variance on PC1 and PC2 axis of
499 PCA plots in (A). (**D**) MLN cell count from each mouse group and (**E**) pseudo R^2 measure of effect size of predictor variables and
500 interactions as calculated by MDMR analysis based on MLN cell count (**F**) Bar plots depicting GMFI of CD44 on MLN CD4+ T cells. n
501 = 5-15 mice per group, Block = 2. (**G**) Representative Histogram from Block 1 and 2 showing concatenated files from *T. muris* infected
502 and rewilded mice of each mice strain (**H**) Representative Histogram showing concatenated files from different groups of mice in Block
503 2 with corresponding (**I**) Bar plots depicting proportion of B cells expressing CD44 on MLN cells. n = 6-18 mice per group, Block = 2.
504 Statistical significance was determined by a based on MDMR analysis with R package for (**A**) and (**E**) or based on one-way ANOVA test
505 between different groups with Graph-Pad Software for (**C**) and (**D**). For (**F**) and (**I**), one-way ANOVA test was used to test statistical
506 significance between the different groups of interest. Data are displayed as mean ± SEM. *ns* $p>0.05$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$;
507 **** $p<0.0001$.
508

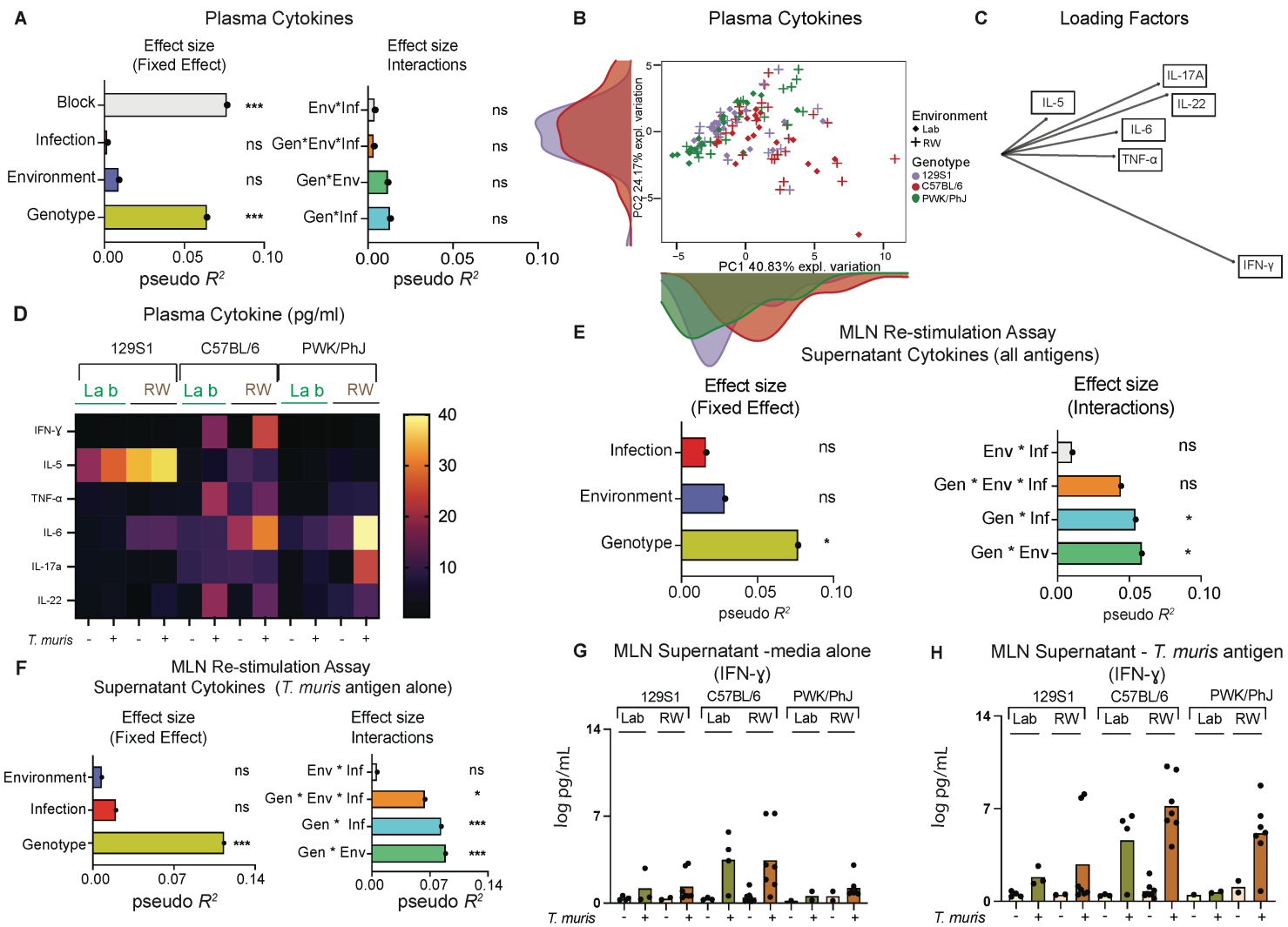


Fig. 3: Genotype has a bigger effect on cytokine responses as a fixed predictor variable. (**A**) Bar plots showing the pseudo R^2 measure of effect size of predictor variables (left) and interactions (right) as calculated by multivariate distance matrix regression analysis (MDMR) from plasma cytokine data, Block 1 and Block 2 (**B**) PCA showing circulating plasma cytokine levels from individual mice and (**C**) their

513 loading factors (**D**) Heatmap depicting levels of circulating cytokines in plasma (**E**) Bar plots showing the pseudo R^2 measure of effect
514 size of predictor variables and interactions as calculated by multivariate distance matrix regression analysis (MDMR) from cytokine
515 supernatant data of MLN cells stimulated with either CD3/CD28 beads, LPS, *Candida albicans*, *Clostridium perfringes*, *Bacteroides*
516 *vulgatus* or *T. muris* antigens. (**F**) Bar plots showing the pseudo R^2 measure of effect size of predictor variables and interactions as calculated
517 by multivariate distance matrix regression analysis (MDMR) from cytokine supernatant data of MLN cells stimulated with *T. muris* antigen.
518 Barplot showing transformed IFN-γ cytokine levels in the supernatant for (**G**) controls as well as (**H**) following stimulation with *T. muris*
519 antigen. Statistical significance was determined by a based on MDMR analysis with R package for (**A**), (**E**) and (**F**). Plasma cytokine data
520 included samples from both Block 1 and Block 2 while MLN Cytokine supernatant data included samples from Block 2 alone due to
521 technical problems with stimulation assays from Block 1. Data are displayed as mean ± SEM. ns $p>0.05$; * $p<0.05$; ** $p<0.01$; ***
522 $p<0.001$; **** $p<0.0001$
523

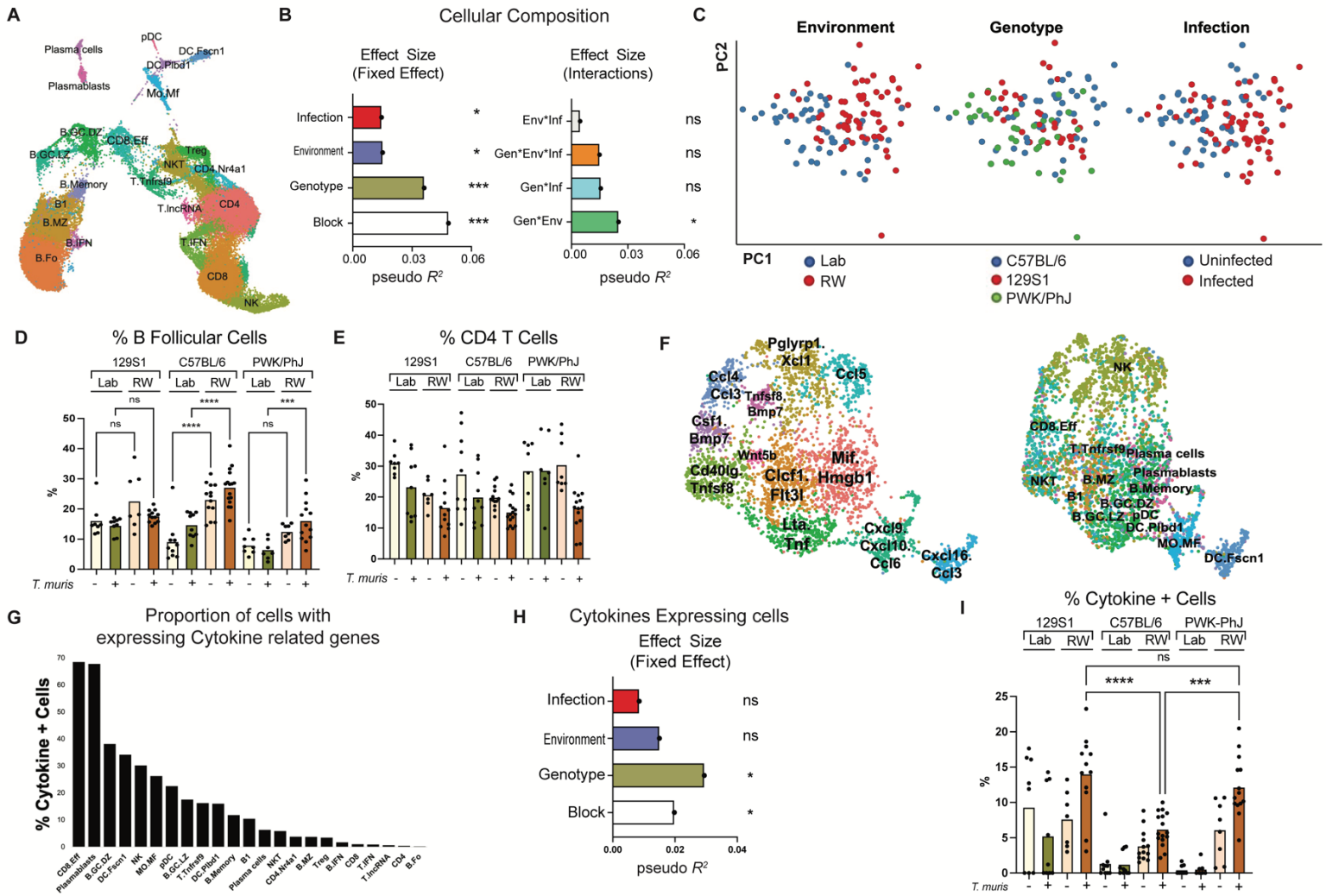
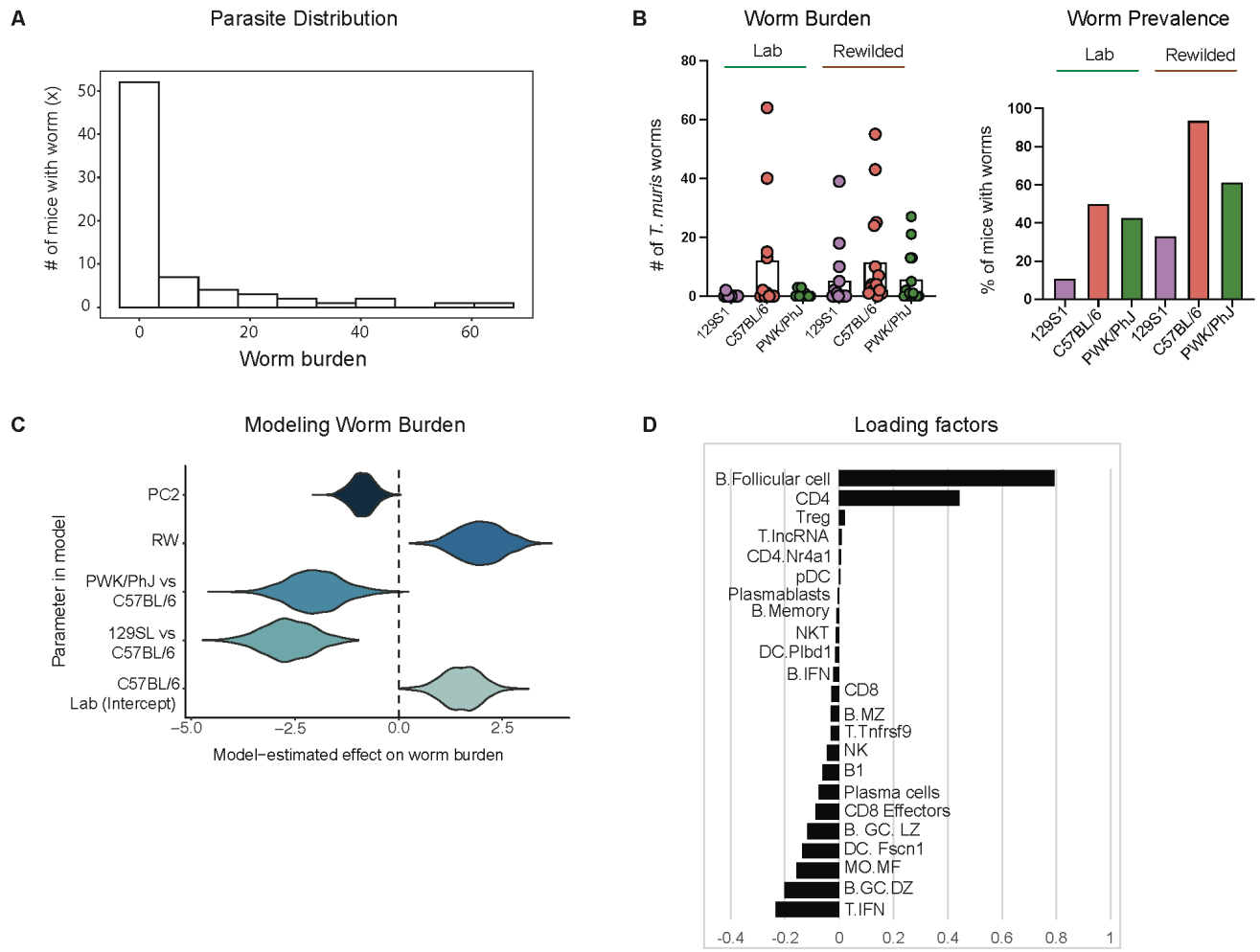


Figure 4: Single Cell Sequencing Analysis for assessing immune variation in cellular composition and cytokine profiles. (**A**) UMAP visualization of scRNASeq data identifying 23 major immune cell subset (**B**) Bar plot showing the pseudo R^2 measure of effect size of predictor variables and interactions as calculated by multivariate distance matrix regression analysis (MDMR) based on cellular

528 composition of cells identified in (A) in each mouse (**C**) PCA of MLN cellular compositional data as determined by scRNAseq analysis
529 Bar plots showing percentage of B Follicular cells (**D**) and CD4 T cells (**E**) based on the scRNA-Seq identified in (A). (**F**) Cytokine-
530 expressing cell clusters. (**G**) Proportion of cells expressing cytokine-related genes of those identified in (F). (**H**) Bar plots showing the
531 pseudo R^2 measure of effect size of predictor variables and interactions as calculated by multivariate distance matrix regression analysis
532 (MDMR) based on data from proportion of cytokines expressing cells identified in (F). (**I**) Bar plot showing cytokine expressing cells
533 identified in the (F). Statistical significance was determined by a based on MDMR analysis with R package for (**B**) and (**H**). Data are
534 displayed as mean ± SEM. *ns p>0.05; * p<0.05; ** p<0.01; *** p<0.001; **** p<0.0001*.
535



536

537 **Figure 5: Genetics and Environmental factors predicts outcomes during exposure with *T. muris* parasite.** Significant variation in
538 worm burden among exposed mice, 3 weeks after inoculation with 200 eggs of *Trichuris muris* per host. (**A**) Worm burden (number of
539 nematodes remaining in the caecum at that timepoint) followed a negative binomial distribution. (**B**) Worm burden depicted as Number of

540 worms per mouse and Percentage of mice (Prevalence) still infected by worms. Each was predicted by a combination of Genetic and
541 Environmental effects, including Gen*Env for worm burden (see text). (**C**) When we used PC2 from the scRNAseq data (Fig 3B) as an
542 index of immune variation among individuals in our statistical models, we found that Gen*Env was no longer significant. Instead, the best
543 model included main effects of host strain (C57BL/6 vs 129SL vs PWK/PhJ), environment (Lab vs RW), and PC2. The figure depicts 1000
544 model-estimated values for the effect of each predictor on worm burden. (**D**) Loading Factors for PC2 of the scRNAseq dataset.

Acknowledgements: We thank William Craigens, Christina Hansen and Felix Rozenberg for invaluable assistance in the field and at Stony Ford. In addition, we thank Mingming Zhao and Johnson Randall for help with setting up experiments in the lab and for help in maintaining the Joes' Flow App Software respectively. We thank Dr. Elia Tait Wojno for sharing some *Trichuris muris* parasite eggs with us. **Funding:** This research was supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, NIH and K.C is supported through the NIH grant AI130945. A.E.D. acknowledges funding support from the National Science Foundation (Award # DGE-2039656). R.S.B. and A.L.G. acknowledge funding support from NJ ACTS (New Jersey Alliance for Clinical and Translational Science), which is supported in part by the New Jersey Health Foundation, Inc., and in part by a Clinical and Translational Science Award from the National Center for Advancing Translational Science of the National Institutes of Health, under award number UL1TR003017. YC acknowledges funding from the Bernard Levine Postdoctoral Research Fellowship and the Charles H. Revson Senior Fellowship. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. **Author contributions:** Conceptualization: OO, AED, KC, ALG, PL. Methodology: OO, AED, NH, RSB, KK, YC, ALG, SCL, JD, OMP. Investigation: OO, AED, RB, NH, KK, KZ, YC, AM, SCL OMP, COSS, CH, PL. Data curation and analysis: OO, NH, AED, AG, SC. Writing – original draft: OO, AED, NH, ALG, PL. Writing – review and editing: OO, AED, KC, ALG, PL. Visualization: OO, NH, JC, AED, ALG. Supervision: SK, KC, ALG, PL. Funding acquisition: KC, ALG, PL. **Competing interests:** All other authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplemental materials. Additional raw data, code and materials used in the analysis would be made publicly available at the point of publication.

580 **Materials and Methods**

581 **Study Design**

582 **Mice and Rewilding**

583 C57BL/6J, 129S1/SV1mJ and PWK/PhJ mice were purchased from The Jackson Laboratory (Bar
584 Harbor, ME) and were housed under specific pathogen–free conditions with *ad libitum* access to
585 food and water. All mouse lines were then bred onsite in a specific pathogen free (SPF) facility at
586 National Institute of Health (NIH). The resulting littermates from the multiple breeding pairs were
587 shipped to Princeton University where they were randomly assigned to either remain in the
588 institutional vivarium (lab mice) or released into the outdoor enclosures (rewilded mice)
589 previously described and complementary manuscript (23-26). The protocols for releasing the
590 laboratory mice into the outdoor enclosure facility were approved by Princeton IACUC.

591 25- 30 mice of mixed strains and genotype, 129S1/SV1mJ, C57BL/6J and PWK/PhJ female mice
592 were used for these experiments. For rewilding, 12-15 female mice of the different strains
593 (129S1/SV1mJ, C57BL/6J and PWK/PhJ) were housed in different wedges in the enclosure for 5
594 weeks. Concomitantly, 10 mice of the different strains were left in the institutional vivarium (lab
595 mice) where the temperature and humidity were maintained. Longworth traps baited with chow
596 were used to catch the mice approximately at 2 weeks and 5 weeks after release(23). At 2 weeks
597 after release, 8-10 mice of each genotype were trapped, blood and fresh stool was collected from
598 the mice for longitudinal CBC analysis and microbiome analysis(25). At the same time 2 weeks
599 following rewilding, some of these mice, vivarium controls and the rewilded mice, were infected
600 with 200 *Trichuris muris* embryonated eggs by oral gavage. At approximately 5 weeks post-
601 rewilding and 19-21 days post *T. muris* infection, mice were recovered for analysis and worm
602 count. Blood was collected for Complete Blood Count with differentials and Flow Cytometry,
603 MLN was harvested for flow cytometry and ceca were collected and the number of adult worms
604 in each cecum were counted individually using an inverted microscope. For all analysis, samples
605 that fail quality control and/or are under limit of detection are not included in downstream
606 statistical analyses.

607 **Complete Blood Count Analysis**

608 Blood samples (approx. 30-50μL) were collected from all mice at endpoint via cheek bleeds using
609 a Medipoint Golden Rod Lancet Blade 4MM (Medipoint NC9922361) into a 1.3mL heparin coated

610 tube (Sarstedt Inc, NC9574345). Blood samples were analyzed using the element HT5, Veterinary
611 Hematology Analyzer (Heska).

612 **PBMC preparation and isolation**

613 Heparinized Whole blood collected via the cheek bleeds were mixed with blood collected via
614 cardiac puncture method. The combined blood samples were spun for 10 minutes at 1500 rpm and
615 plasma was collected and stored at -80°C for further cytokine analysis. The cellular component re-
616 suspended in PBS next underwent a density gradient separation process using the Lymphocyte
617 Separation Media (LSM™ MP Biomedicals, LLC, Irvine, CA) according to manufacturer's
618 instruction. Isolated PBMCs were washed twice in PBS and then used for downstream spectral
619 cytometric analysis.

620 **Preparation of single cell suspensions from MLN**

621 Single cell suspension from the mesenteric lymph nodes (MLNs) were prepared by mashing the
622 tissues individually through a 70µm cell strainer and washed with RPMI. Cells were then washed
623 with RPMI supplemented with 10% FCS. Live cell numbers were enumerated using the element
624 HT5, Veterinary Hematology Analyzer (Heska).

625 **Spectral cytometry and Analysis**

626 Single-cell suspensions prepared from the PBMCs, and MLN were washed twice with Flow
627 cytometry buffer (FACs Buffer) and PBS prior to incubating with Live/Dead™ Fixable Blue
628 (ThermoFisher) and Fc Block™ (clone KT1632; BD) for 10 minutes at room temperature.
629 Cocktails of fluorescently conjugated antibodies (listed in Table 1 and Table 2) diluted in FACs
630 Buffer and 10% Brilliant Stain Buffer (BD) were then added directly to cells and incubated for a
631 further 30 minutes at room temperature. For the lymphoid panel, cells were next incubated in
632 eBioscience™ Transcription Factor Fixation and Permeabilization solution (Invitrogen) for 12-18
633 hours at 4°C and stained with cocktails of fluorescently labeled antibodies against intracellular
634 antigens diluted in Permeabilization Buffer (Invitrogen) for 1 hour at 4°C.

635 Spectral Unmixing was performed for each experiment using single-strained controls using
636 UltraComp eBeads™ (Invitrogen). Dead cells and doublets were excluded from analysis. All
637 samples were collected on an Aurora™ spectral cytometer (Cytek) and analyzed using the OMIQ
638 software (https://www.omiq.ai/), data cleaning and scaling was done using algorithms like
639 FlowCut (62, 63) within the OMIQ software. Sub-sampled cells including 10,000 Live, CD45+

640 cells were re-clustered in an unsupervised version using the Joe's Flow software,
641 (Github: https://github.com/niaid/JoesFlow).

642 **MLN Cell Stimulation and Cytokine Profiling**

643 Single cell suspension of MLN cells were reconstituted in RPMI at 2×10^6 cells/mL, and 0.1 mL
644 was cultured in 96-well microtiter plates that contained 10^7 *cf*.u/mL UV-killed microbes,
645 10^5 αCD3/CD28 beads (11456F), or Lipopolysaccharide (LPS; 100ng/mL) (L2630) or PBS
646 control. The stimulated microbes and antigens included are as: *Bacteroides vulgatus* (ATCC
647 8482), *Candida albicans* (UC820), *Clostridium perfringens* (NCTC 10240), and *Trichuris*
648 *muris* antigen. Supernatants were collected after 2 days and stored at −80°C. Concentrations of
649 IL-5, IL-6, IL-22, IL-17A, IFN-γ, TNF-α, IL-2, IL-4, IL-10, IL-9 and IL-13 in supernatants were
650 measured using a commercially available murine Th cytokine LEGENDplex assay (Biolegend)
651 panel (Cat # 741044) according to the manufacturer's instructions. Plasma concentrations of IL-5,
652 IL-6, IL-22, IL-17A, IFN-γ, TNFα were measured also measured using the commercially available
653 murine Th cytokine LEGENDplex assay (Biolegend) panel (Cat #741044) according to the
654 manufacturer's instructions. Cytokines levels that were lower than limit of detection across
655 samples were excluded from further analysis.

656 **Single cell RNA sequencing**

657 Single cell suspensions were obtained from MLN as described above. 2000 cells from each
658 individual mouse (Block 1 (n = 51); Block 2 (n = 71)) were labelled with the Ab-Hash Tag
659 Oligonucleoutides (HTO). These antibodies are a mix of anti-CD45 and anti-MHCI antibodies.
660 TotalSeq™-C antibodies are used with Single Cell 5′ kit. Pooled samples from each group were
661 then loaded on a 10X Genomics Next GEM chip and single-cell GEMs were generated on a 10X
662 Chromium Controller. Subsequent steps to generate cDNA and sequencing libraries were
663 performed following 10X Genomics' protocol. Libraries were pooled and sequenced using
664 Illumina NovaSeq SP 100 cycle as per 10X sequencing recommendations.

665 The sequenced data were processed using Cell Ranger (version 6.0) to demultiplex the libraries.
666 The reads were aligned to *Mus musculus* mm10 genomes to generate count tables that were further
667 analyzed using Seurat (version 4.0). Sequencing data from the two blocks were integrated together
668 prior to further downstream analysis. Data are displayed as uniform manifold approximation and
669 projection (UMAP). The different cell subsets from each cluster were defined by the top 50

differentially expressed genes and identification using the SingleR sequencing pipeline(64). Cell types with different cytokine expression were identified based on expression of genes related to cytokine function using the Gene Ontology Browser. Seurat Analysis pipeline was used for comparisons between each of the different cell cluster of interest.

Visualization

scRNASeq analysis data were visualized using Seurat (version 4.1.2) and R Studio (version 2022.07.1). Cartoons were created using BioRender.com.

Quantification and Statistical analysis

In all cases principal component analysis was performed with in R v4.1.2. For cytokine data, log transformed data were used for generation of PCA plots and for MDMR analysis. Biplots were constructed by projecting the weighted averages of each input feature (immune cell phenotypes, cytokine level, cellular composition etc.,) along PC1, PC2 and/or PC3 derived from the biplot.pcoa function from the ape package as previously done in (23). Effect size measures were determined using the MDMR v0.5.1(65) package in R and interactions were tested using the mixed effect analysis in the MDMR package. A significant effect size is said to be present if the *p*-value was less than or equal to 0.05 (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$).

Results in graphs and barplots are displayed as mean ± SEM using Prism version 7 (GraphPad Software, Inc.) except where mentioned. Statistical analysis was performed using GraphPad Prism software (v9). Right-skewed data were log or square root transformed. For analysis of relationship between scRNAseq cell composition and worm burden in Fig. 4, worm burden was modeled as following a negative binomial distribution. Predictor variables included in the regression model were mouse strain, mouse environment (i.e., laboratory or rewilded), and loading on PC2 from analysis of scRNAseq data. 1000 model-estimated coefficient values were then plotted for each predictor variable. In some cases, data were analyzed by One-Way ANOVA with Tukey's post- when comparing three or more groups using GraphPad Prism software (v9). Experimental group was considered statistically significant if the fixed effect F test *p*-value was ≤ 0.05. Post hoc pairwise comparisons between experimental groups were made using Tukey's honestly significant difference multiple-comparison test. A difference between experimental groups was taken to be significant if the *p*-value was less than or equal to 0.05 (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$).

700 **References**

701 1. Aguirre-Gamboa R, Joosten I, Urbano PCM, van der Molen RG, van Rijssen E, van
702 Cranenbroek B, Oosting M, Smeekens S, Jaeger M, Zorro M, Withoff S, van Herwaarden
703 AE, Sweep FCGJ, Netea RT, Swertz MA, Franke L, Xavier RJ, Joosten LAB, Netea MG,
704 Wijmenga C, Kumar V, Li Y, Koenen HJPM. 2016. Differential Effects of
705 Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell reports* 17:
706 2474-87
707 2. Ter Horst R, Jaeger M, Smeekens SP, Oosting M, Swertz MA, Li Y, Kumar V,
708 Diavatopoulos DA, Jansen AFM, Lemmers H, Toenhake-Dijkstra H, van Herwaarden
709 AE, Janssen M, van der Molen RG, Joosten I, Sweep FCGJ, Smit JW, Netea-Maier RT,
710 Koenders MMJF, Xavier RJ, van der Meer JWM, Dinarello CA, Pavelka N, Wijmenga
711 C, Notebaart RA, Joosten LAB, Netea MG. 2016. Host and Environmental Factors
712 Influencing Individual Human Cytokine Responses. *Cell* 167: 1111-24.e13
713 3. Gause WC, Rothlin C, Loke Pn. 2020. Heterogeneity in the initiation, development and
714 function of type 2 immunity. *Nature Reviews Immunology* 20: 603-14
715 4. Oyesola OO, Souza COS, Loke Pn. 2022. The Influence of Genetic and Environmental
716 Factors and Their Interactions on Immune Response to Helminth Infections. *Frontiers in*
717 *Immunology* 13
718 5. Poon MML, Byington E, Meng W, Kubota M, Matsumoto R, Grifoni A, Weiskopf D,
719 Dogra P, Lam N, Szabo PA, Ural BB, Wells SB, Rosenfeld AM, Brusko MA, Brusko
720 TM, Connors TJ, Sette A, Sims PA, Luning Prak ET, Shen Y, Farber DL. 2021.
721 Heterogeneity of human anti-viral immunity shaped by virus, tissue, age, and sex. *Cell*
722 *Rep* 37: 110071
723 6. Mangino M, Roederer M, Beddall MH, Nestle FO, Spector TD. 2017. Innate and
724 adaptive immune traits are differentially affected by genetic and environmental factors.
725 *Nature Communications* 8: 13850
726 7. Brodin P, Jojic V, Gao T, Bhattacharya S, Angel Cesar JL, Furman D, Shen-Orr S,
727 Dekker Cornelia L, Swan Gary E, Butte Atul J, Maecker Holden T, Davis Mark M. 2015.
728 Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences.
729 *Cell* 160: 37-47
730 8. Abolins S, Lazarou L, Weldon L, Hughes L, King EC, Drescher P, Pocock MJO, Hafalla
731 JCR, Riley EM, Viney M. 2018. The ecology of immune state in a wild mammal, Mus
732 musculus domesticus. *PLOS Biology* 16: e2003538
733 9. Beura LK, Hamilton SE, Bi K, Schenkel JM, Odumade OA, Casey KA, Thompson EA,
734 Fraser KA, Rosato PC, Filali-Mouhim A, Sekaly RP, Jenkins MK, Vezys V, Haining
735 WN, Jameson SC, Masopust D. 2016. Normalizing the environment recapitulates adult
736 human immune traits in laboratory mice. *Nature* 532: 512-6
737 10. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, Alanio C, Kuri-
738 Cervantes L, Pampena MB, D'Andrea K, Manne S, Chen Z, Huang YJ, Reilly JP,
739 Weisman AR, Ittner CAG, Kuthuru O, Dougherty J, Nzingha K, Han N, Kim J, Pattekar
740 A, Goodwin EC, Anderson EM, Weirick ME, Gouma S, Arevalo CP, Bolton MJ, Chen F,
741 Lacey SF, Ramage H, Cherry S, Hensley SE, Apostolidis SA, Huang AC, Vella LA,
742 Betts MR, Meyer NJ, Wherry EJ. 2020. Deep immune profiling of COVID-19 patients
743 reveals distinct immunotypes with therapeutic implications. *Science* 369
744 11. Kuri-Cervantes L, Pampena MB, Meng W, Rosenfeld AM, Ittner CAG, Weisman AR,
745 Agyekum RS, Mathew D, Baxter AE, Vella LA, Kuthuru O, Apostolidis SA, Bershaw L,
746 Dougherty J, Greenplate AR, Pattekar A, Kim J, Han N, Gouma S, Weirick ME, Arevalo

CP, Bolton MJ, Goodwin EC, Anderson EM, Hensley SE, Jones TK, Mangalmurti NS, Luning Prak ET, Wherry EJ, Meyer NJ, Betts MR. 2020. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci Immunol* 5

12. Ingelfinger F, Gerdes LA, Kavaka V, Krishnarajah S, Friebel E, Galli E, Zwicky P, Furrer R, Peukert C, Dutertre C-A, Eglseer KM, Ginhoux F, Flierl-Hecht A, Kümpfel T, De Feo D, Schreiner B, Mundt S, Kerschensteiner M, Hohlfeld R, Beltrán E, Becher B. 2022. Twin study reveals non-heritable immune perturbations in multiple sclerosis. *Nature* 603: 152-8
13. McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, Conti D, Gauderman WJ, Hsu L, Hutter CM, Jankowska MM, Kerr J, Kraft P, Montgomery SB, Mukherjee B, Papanicolaou GJ, Patel CJ, Ritchie MD, Ritz BR, Thomas DC, Wei P, Witte JS, participants obow. 2017. Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases. *American Journal of Epidemiology* 186: 753-61
14. Aschard H, Lutz S, Maus B, Duell EJ, Fingerlin TE, Chatterjee N, Kraft P, Van Steen K. 2012. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Human Genetics* 131: 1591-613
15. Virolainen SJ, VonHandorf A, Viel KCMF, Weirauch MT, Kottyan LC. 2023. Gene–environment interactions and their impact on human health. *Genes & Immunity* 24: 1-11
16. Hunter DJ. 2005. Gene–environment interactions in human diseases. *Nature Reviews Genetics* 6: 287-98
17. Lazzaro BP, Little TJ. 2009. Immunity in a variable world. *Philos Trans R Soc Lond B Biol Sci* 364: 15-26
18. Vieira C, Pasyukova EG, Zeng Z-B, Hackett JB, Lyman RF, Mackay TFC. 2000. Genotype-Environment Interaction for Quantitative Trait Loci Affecting Life Span in Drosophila melanogaster. *Genetics* 154: 213-27
19. Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565-77
20. Mitchell SE, Rogers ES, Little TJ, Read AF. 2005. HOST-PARASITE AND GENOTYPE-BY-ENVIRONMENT INTERACTIONS: TEMPERATURE MODIFIES POTENTIAL FOR SELECTION BY A STERILIZING PATHOGEN. *Evolution* 59: 70-80
21. Graham AL. 2021. Naturalizing mouse models for immunology. *Nature Immunology* 22: 111-7
22. Hamilton SE, Badovinac VP, Beura LK, Pierson M, Jameson SC, Masopust D, Griffith TS. 2020. New Insights into the Immune System Using Dirty Mice. *The Journal of Immunology* 205: 3-11
23. Lin J-D, Devlin JC, Yeung F, McCauley C, Leung JM, Chen Y-H, Cronkite A, Hansen C, Drake-Dunn C, Ruggles KV, Cadwell K, Graham AL, Loke Pn. 2020. Rewilding Nod2 and Atg16l1 Mutant Mice Uncovers Genetic and Environmental Contributions to Microbial Responses and Immune Cell Composition. *Cell Host & Microbe* 27: 830-40.e4
24. Leung JM, Budischak SA, Chung The H, Hansen C, Bowcutt R, Neill R, Shellman M, Loke Pn, Graham AL. 2018. Rapid environmental effects on gut nematode susceptibility in rewilded mice. *PLOS Biology* 16: e2004108
25. Downie AE, Oyesola O, Barre RS, Caudron Q, Chen Y-H, Dennis EJ, Garnier R, Kiwanuka K, Menezes A, Navarrete DJ, Mondragón-Palomino O, Saunders JB, Tokita CK, Zaldana K, Cadwell K, Loke P, Graham AL. 2023. Social association predicts immunological similarity in rewilded mice. *bioRxiv*: 2023.03.15.532825

26. Yeung F, Chen Y-H, Lin J-D, Leung JM, McCauley C, Devlin JC, Hansen C, Cronkite A, Stephens Z, Drake-Dunn C, Fulmer Y, Shopsin B, Ruggles KV, Round JL, Loke Pn, Graham AL, Cadwell K. 2020. Altered Immunity of Laboratory Mice in the Natural Environment Is Associated with Fungal Colonization. *Cell Host & Microbe* 27: 809-22.e6
27. Lai PS, Massoud AH, Xia M, Petty CR, Cunningham A, Chatila TA, Phipatanakul W. 2018. Gene-environment interaction between an IL4R variant and school endotoxin exposure contributes to asthma symptoms in inner-city children. *Journal of Allergy and Clinical Immunology* 141: 794-6.e3
28. Lusis AJ, Seldin MM, Allayee H, Bennett BJ, Civelek M, Davis RC, Eskin E, Farber CR, Hui S, Mehrabian M, Norheim F, Pan C, Parks B, Rau CD, Smith DJ, Vallim T, Wang Y, Wang J. 2016. The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *Journal of Lipid Research* 57: 925-42
29. Link VM, Duttke SH, Chun HB, Holtman IR, Westin E, Hoeksema MA, Abe Y, Skola D, Romanoski CE, Tao J, Fonseca GJ, Troutman TD, Spann NJ, Strid T, Sakai M, Yu M, Hu R, Fang R, Metzler D, Ren B, Glass CK. 2018. Analysis of Genetically Diverse Macrophages Reveals Local and Domain-wide Mechanisms that Control Transcription Factor Binding and Function. *Cell* 173: 1796-809.e17
30. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* 526: 68-74
31. Zapala MA, Schork NJ. 2012. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Frontiers in genetics* 3: 190-
32. McArdle BH, Anderson MJ. 2001. FITTING MULTIVARIATE MODELS TO COMMUNITY DATA: A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. *Ecology* 82: 290-7
33. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32-46
34. Camp RL, Kraus TA, Birkeland ML, Puré E. 1991. High levels of CD44 expression distinguish virgin from antigen-primed B cells. *Journal of Experimental Medicine* 173: 763-6
35. Scott ME. 1991. Heligmosomoides polygyrus (Nematoda): susceptible and resistant strains of mice are indistinguishable following natural infection. *Parasitology* 103 Pt 3: 429-38
36. Klementowicz JE, Travis MA, Grencis RK. 2012. Trichuris muris: a model of gastrointestinal parasite infection. *Seminars in Immunopathology* 34: 815-28
37. Else KJ, Grencis RK. 1991. Cellular immune responses to the murine nematode parasite Trichuris muris. I. Differential cytokine production during acute or chronic infection. *Immunology* 72: 508-13
38. Li Y, Oosting M, Smeekens SP, Jaeger M, Aguirre-Gamboa R, Le KTT, Deelen P, Ricaño-Ponce I, Schoffelen T, Jansen AFM, Swertz MA, Withoff S, van de Vosse E, van Deuren M, van de Veerdonk F, Zhernakova A, van der Meer JWM, Xavier RJ, Franke L, Joosten LAB, Wijmenga C, Kumar V, Netea MG. 2016. A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* 167: 1099-110.e14
39. Olin A, Henckel E, Chen Y, Lakshmikanth T, Pou C, Mikes J, Gustafsson A, Bernhardsson AK, Zhang C, Bohlin K, Brodin P. 2018. Stereotypic Immune System Development in Newborn Children. *Cell* 174: 1277-92.e14

843 40. Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S,
844 Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, Subbarao P, Mandhane P, Becker A,
845 McNagny KM, Sears MR, Kollmann T, Mohn WW, Turvey SE, Brett Finlay B. 2015.
846 Early infancy microbial and metabolic alterations affect risk of childhood asthma.
847 *Science Translational Medicine* 7: 307ra152-307ra152
848 41. Arrieta M-C, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. 2014. The Intestinal
849 Microbiome in Early Life: Health and Disease. *Frontiers in Immunology* 5
850 42. Djuardi Y, Wammes LJ, Supali T, Sartono E, Yazdanbakhsh M. 2011. Immunological
851 footprint: the development of a child's immune system in environments rich in
852 microorganisms and parasites. *Parasitology* 138: 1508-18
853 43. Dutta S, Sengupta P. 2016. Men and mice: Relating their ages. *Life Sciences* 152: 244-8
854 44. Kaczorowski KJ, Shekhar K, Nkulikiyimfura D, Dekker CL, Maecker H, Davis MM,
855 Chakraborty AK, Brodin P. 2017. Continuous immunotypes describe human immune
856 variation and predict diverse responses. *Proceedings of the National Academy of Sciences*
857 114: E6097-E106
858 45. Maizels RM, Bundy DAP, Selkirk ME, Smith DF, Anderson RM. 1993. Immunological
859 modulation and evasion by helminth parasites in human populations. *Nature* 365: 797-
860 805
861 46. Babayan SA, Liu W, Hamilton G, Kilbride E, Rynkiewicz EC, Clerc M, Pedersen AB.
862 2018. The Immune and Non-Immune Pathways That Drive Chronic Gastrointestinal
863 Helminth Burdens in the Wild. *Frontiers in Immunology* 9
864 47. Little MC, Hurst RJ, Else KJ. 2014. Dynamic changes in macrophage activation and
865 proliferation during the development and resolution of intestinal inflammation. *J*
866 *Immunol* 193: 4684-95
867 48. Bellaby T, Robinson K, Wakelin D. 1996. Induction of differential T-helper-cell
868 responses in mice infected with variants of the parasitic nematode Trichuris muris.
869 *Infection and Immunity* 64: 791-5
870 49. Johnston CE, Bradley JE, Behnke JM, Matthews KR, Else KJ. 2005. Isolates of Trichuris
871 muris elicit different adaptive immune responses in their murine host. *Parasite*
872 *Immunology* 27: 69-78
873 50. Koyama K, Ito Y. 1996. Comparative studies on immune responses to infection in
874 susceptible B10.BR mice infected with different strains of the murine nematode parasite
875 Trichuris muris. *Parasite Immunol* 18: 257-63
876 51. Bancroft AJ, Else KJ, Grencis RK. 1994. Low-level infection with Trichuris muris
877 significantly affects the polarization of the CD4 response. *Eur J Immunol* 24: 3113-8
878 52. Bancroft AJ, Else KJ, Humphreys NE, Grencis RK. 2001. The effect of challenge and
879 trickle Trichuris muris infections on the polarisation of the immune response. *Int J*
880 *Parasitol* 31: 1627-37
881 53. Glover M, Colombo SAP, Thornton DJ, Grencis RK. 2019. Trickle infection and
882 immunity to Trichuris muris. *PLOS Pathogens* 15: e1007926
883 54. Colombo SAP, Grencis RK. 2020. Immunity to Soil-Transmitted Helminths: Evidence
884 From the Field and Laboratory Models. *Frontiers in Immunology* 11
885 55. White EC, Houlden A, Bancroft AJ, Hayes KS, Goldrick M, Grencis RK, Roberts IS.
886 2018. Manipulation of host and parasite microbiotas: Survival strategies during chronic
887 nematode infection. *Science Advances* 4: eaap7399
888 56. Venzon M, Das R, Luciano DJ, Park HS, Kool ET, Belasco JG, Hubbard EJA, Cadwell
889 K. 2021. Microbial byproducts determine reproductive fitness of free-living and parasitic
890 nematodes. *bioRxiv*: 2021.08.02.454806

57. Hayes KS, Bancroft AJ, Goldrick M, Portsmouth C, Roberts IS, Grencis RK. 2010. Exploitation of the Intestinal Microflora by the Parasitic Nematode Trichuris muris. *Science* 328: 1391-4
58. Bach J-F. 2018. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nature Reviews Immunology* 18: 105-20
59. Maizels RM. 2020. Regulation of immunity and allergy by helminth parasites. *Allergy* 75: 524-34
60. Maizels RM, Yazdanbakhsh M. 2003. Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol* 3: 733-44
61. Santiago HC, Nutman TB. 2016. Human Helminths and Allergic Disease: The Hygiene Hypothesis and Beyond. *Am J Trop Med Hyg* 95: 746-53
62. Meskas J, Wang S, Brinkman R. 2020. flowCut — An R package for precise and accurate automated removal of outlier events and flagging of files based on time versus fluorescence analysis. *bioRxiv*: 2020.04.23.058545
63. Park LM, Lannigan J, Jaimes MC. 2020. OMIP-069: Forty-Color Full Spectrum Flow Cytometry Panel for Deep Immunophenotyping of Major Cell Subsets in Human Peripheral Blood. *Cytometry Part A* 97: 1044-51
64. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* 20: 163-72
65. McArtor DB, Lubke GH, Bergeman CS. 2017. Extending multivariate distance matrix regression with an effect size measure and the asymptotic null distribution of the test statistic. *Psychometrika* 82: 1052-77