Title: Housing laboratory mice in a natural environment uncovers genetic and environmental contributions to immune variation.

Authors: Jian-Da Lin[1,*], Joseph C. Devlin[1,2,*], Frank Yeung[2,3,*], Caroline McCauley[1], Jacqueline M. Leung[4], Ying-Han Chen[3], Alex Cronkite[1,3], Christina Hansen[4], Charlotte Drake-Dunn[3], Kelly V. Ruggles[5,6], Ken Cadwell[1,3,§,†], Andrea L. Graham[4,§,†] and P'ng Loke[1,§,†]

Affiliations:

[1]Department of Microbiology, New York University School of Medicine, New York, NY 10016, USA.

[2]Sackler Institute of Graduate Biomedical Sciences,

[3]Kimmel Center for Biology and Medicine at the Skirball Institute, New York University School of Medicine, New York, NY 10016, USA.

[4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA.

[5]Division of Translational Medicine, Department of Medicine,

[6]Applied Bioinformatics Laboratories, New York School of Medicine, New York, NY 10016, USA

*These authors contributed equally

†These authors contributed equally

[§]Correspondence to:

Png.Loke@nyulangone.org

Ken.Cadwell@med.nyu.edu

algraham@princeton.edu

Abstract

The relative contributions of genetic and environmental factors to variation in immune responses are still poorly understood. Here, we performed a deep phenotypic analysis of immunological parameters of laboratory mice released into an outdoor enclosure, carrying susceptibility genes (*Nod2* and *Atg16l1*) implicated in the development of inflammatory bowel diseases. Variations of immune cell populations were largely driven by environment, whereas cytokine production showed a stronger genetic component, but this genetic effect was reduced after exposure to the wild environment. Multi-omic models identified transcriptional signatures associated with differences in T cell populations. Sub-networks associated with responses against *Candida albicans* and *Bacteroides vulgatus* are also coupled with rewilding. Hence, exposing laboratory mice of different genetic backgrounds to the outdoor environment can uncover important contributors to immune variation.

One sentence summary

Natural environment exposure in laboratory mice primarily promotes variation in population frequencies of immune cells, whereas cytokine responses to microbial stimulation are affected more by genetic susceptibility to inflammatory bowel disease.

Main Text

Variation in the magnitude of the immune response is an important determinant of susceptibility to pathogen infections, as well as a predisposition to autoimmunity and cancer (*1, 2*). However, immunological studies with laboratory mice typically aim to control variation in order to focus on genetic factors that alter components of both the adaptive and innate immune system. Recent advances in technology and throughput have facilitated a systems immunology approach towards deciphering the relative genetic and environmental contributors to variation of the human immune system (*3-7*). By altering the environment in which laboratory mice are conventionally housed, combined with deep phenotypic analysis of immune responses, we describe here a resource dataset that we used to probe genetic versus environmental drivers of heterogeneity in the immune system. Controlled experiments on specific groups with the same genotype exposed to a new environment would be difficult to conduct in human studies.

Recently, there has been a growing effort to characterize wild mice and pet shop mice, which appear to share greater similarity to the immunological state of humans than laboratory mice (*8-11*). We have developed an outdoor enclosure facility that enables us to reintroduce laboratory mice into a more natural environment, a process termed "rewilding" (*12*). The enclosures include soil and vegetation as well as barnlike structures but no fully wild mammals; over the course of several weeks, mice dig burrows and otherwise adapt to their new environment (*12*). Gene-bacteria associations has been identified in both human and wild mice indicating host genetic determinants of the gut microbial composition across mammals in natural environments (*13*). To interrogate the impact of host genetic variants on microbiota and inflammatory bowel disease (IBD) susceptibility in response to environmental changes, we investigated *Atg16l1* and *Nod2* variants which are among the highest risk factors for IBD associated with a dysregulated immune response to gut microbes (*14, 15*). *Atg16l1* is a key component of the autophagy pathway, and *Nod2* is involved in bacterial sensing. We and others previously demonstrated that mice harboring mutations in *Atg16l1* or *Nod2* develop signs of inflammation in a manner dependent on exposure to infectious entities (*16-22*). Mice

with these particular mutations are sensitive to the presence of murine norovirus (MNV), *Helicobacter* species, and other commensal-like agents that are found in some animal facilities and excluded in others (*16-19, 23, 24*). To study the immunological consequences of rewilding in the context of genetic susceptibility genes for IBD, we released *Atg16l1*$^{T300A/T300A}$, *Atg16l1*$^{T300A/+}$, *Nod2*$^{-/-}$ mice in addition to C57BL/6J wild type (WT) mice to determine whether mutations in these genes alter the response to microbial exposure in a natural environment.

116 mice were released into the outdoor enclosure for 6-7 weeks, of which 104 mice (25 WT, 28 *Nod2*$^{-/-}$, 27 *Atg16l1*$^{T300A/+}$, and 24 *Atg16l1*$^{T300A/T300A}$) were recovered in time for analysis and compared to 80 matched controls (19 WT, 19 *Nod2*$^{-/-}$, 20 *Atg16l1*$^{T300A/+}$, 22 *Atg16l1*$^{T300A/T300A}$) maintained under specific pathogen free (SPF) conditions (herein referred to as lab mice) (Fig. 1A). Blood samples were collected at the time of sacrifice and analyzed by flow cytometry with a lymphocyte panel (Table S1). Additionally, cytokine levels in the plasma were also assayed (Fig. 1A). Fecal samples and cecal contents were collected for microbial profiling and for reconstitution experiments (Companion study, Yeung *et. al.*). Mesenteric lymph node (MLN) cells were isolated for flow cytometry analysis with the same lymphocyte panel as the blood and an additional myeloid cell panel (Supplementary materials). Immune activity in the MLN was measured through transcriptional profiles by RNA-seq and cytokine production assays in response to microbial stimulations (Supplementary materials). All of the lab mice were sacrificed in one week, whereas the rewilded mice were sacrificed over a 2-week period based on trapping frequency. This collection of multi-parameter datasets from nearly 200 mice provides an opportunity to examine inter-individual variation in innate and adaptive immune cell populations by flow cytometry analysis, as well as cytokine responses to microbial stimulation, integrated with high dimensional transcriptomics and microbiota data. This systems level analysis enabled us to quantify genetic and environmental influences on the heterogeneity of immunological parameters among individual mice.

In line with previous studies on feral and pet shop mice (*8*), we observed major differences in effector ($CD44^{lo}CD62L^{lo}$) and memory ($CD44^{hi}CD62L^{lo}$ or $CD44^{hi}CD62L^{hi}$) T cells (Companion study, Yeung *et. al.*). To visualize the overall landscape of flow cytometry data in an unbiased manner, phenotypic heterogeneity was analyzed on total CD45+ cells per mouse across the blood lymphocyte panel (Table S1) in lab and rewilded mice. We down sampled to 1,000 CD45+ cells from each mouse for nearly 180,000 single CD45+ cell events and single cell flow cytometry data was visualized by uniform manifold approximation and projection (UMAP) (*25*), which was superior to *t*-distributed stochastic neighborhood embedding (t-SNE) (*25*) (fig. S1A), to identify major immune cell subsets (Fig. 1, B and C, and fig. S1, B to E). Unexpected clusters of $CD44^{hi}$CD19+ cells were increased in the rewilded mice compared to lab mice (Fig. 1B), while lab mice had more $CD62L^{hi}$CD4+ cells (Fig. 1C). This unsupervised visualization strategy was useful for identifying unexpected differences (e.g. in the CD19+ compartment) in cell clusters of the peripheral blood at a single cell level, which might have been overlooked by traditional flow cytometry gating.

We also used a traditional gating strategy to systematically quantify the frequencies of immune cell subsets with a lymphocyte panel (fig. S2 and table S2). The blood lymphocyte panel identified 13 lymphocyte populations, a total myeloid cell population (CD11b/CD11c/DX5), and a total CD45+ cell population. From the proportions of these immune cell populations, we examined the inter-individual variation in lab and rewilded mice by principal coordinate analysis (PCoA) (Fig. 1D). As shown by the separation of lab and rewilded mice along principle coordinate (PC) 1 and 2, there is a strong effect of the environment on lymphocyte populations, and the differences between these populations are attributed to $CD44^{hi}CD62L^{hi}$ CD4 and $CD44^{hi}CD62L^{hi}$ CD8 T cells as inferred by examining loading factors along PC1 and PC2 (Fig. 1D). The reduced proportion of naïve T cells and increased abundance of memory T cells is consistent with previous reports (*8, 11*). We also quantified the extent of variability between lab and rewilded mice by measuring the Euclidean distances between mice in our PCoA. Notably, the within group distances in lab mice were significantly less than rewilded mice (Fig. 1E) suggesting that the natural environment

increases variability in immune cell populations. We then evaluated several metadata variables to identify covariates with the largest effect sizes on immune cell populations (Fig. 1F). As expected, the composition and activation state of immune cell populations in the peripheral blood are driven most strongly by environmental differences between the SPF facility and the outdoor enclosure. These differences far outweighed any effects of genetic deficiency in *Nod2* or *Atg16l1* in both environments (Fig. 1F). These findings are consistent with the concept that non-heritable influences explain the majority of the variation in the proportion of immune cell subsets as reported in twin studies (*3*).

Cytokine production capacity in humans has been reported to be more strongly influenced by genetic factors (*7*). To investigate the relative roles of genetics and environment on induced responses, we performed *ex vivo* stimulation experiments to examine the cytokine responses of total MLN cells from rewilded and lab mice when stimulated with six different microbes (*C. perfringens, P. aeruginosa, B. vulgatus, B. subtilis, S. aureus* and *C. albicans*), as well as αCD3/CD28 beads to activate T cells (Fig. 1A and 2A). A broad array of 13 cytokines was measured including IFN-γ, IL-1α, IL-1ß, IL-13 IL-5, IL-10, IL-17, IL-6, CCL2, CXCL1, CCL3, CCL4 and TNF-α. Relative to PBS mock-treated control cells, increases in IL-5 and IL-17 were observed after αCD3/CD28 activation, consistent with the production of these cytokines by T cells in the MLNs (Fig. 2A). In general, bacterial antigens induced CCL3, IFN-γ, TNF-α, CCL4, and IL-6 production, which differed from the increase in the regulatory cytokine IL-10 in response to the fungus *C. albicans* (Fig. 2A and fig. S3B). Bacterial antigens also induced CXCL1, which was not induced by *C. albicans* or αCD3/CD28 (Fig. 2A). As expected αCD3/CD28 was one of the most potent inducers of cytokine production, especially for IFN-γ, CCL4, TNF-α and IL-5, and rewilded mice produced higher levels of these cytokines than lab mice post stimulation (Fig. 2B). Notably, IFN-γ, IL-10, and IL-6 showed the most variation, whereas IL-1α and IL-1ß showed the least variation (fig. S3). Overall, MLN cells from rewilded mice produced more cytokines when

challenged with microbial antigens (Fig. 2B). In the companion study (Yeung *et.al.*), the effects of fungal exposure and colonization of the rewilded mice are further assessed.

We again used PCoA to exam the inter-individual variation in cytokine responses (Fig. 2C). In contrast to immune cell variations (Fig. 1, D to F), the environmental effects on cytokine production were less pronounced (Fig. 2, C and D). Instead, genotype was identified as the covariate with the largest effect size on cytokine responses (Fig. 2D). Genotype also had the greatest effect size on variation in plasma cytokine levels of individual mice (fig. S4). To investigate interactive effects of genotype and environment, we performed unsupervised clustering analysis on inter-individual correlation for the stimulated cytokine profile of MLN cells in individual mice. In the SPF facility, lab *Nod2*$^{-/-}$ mice are largely segregated from the *Atg16l1*$^{T300A/+}$ and the *Atg16l1*$^{T300A/T300A}$ mice into two separate clusters (Fig. 2E). Remarkably, when introduced into the outdoor enclosure, rewilded *Nod2*$^{-/-}$ mice, *Atg16l1*$^{T300A/+}$ and *Atg16l1*$^{T300A/T300A}$ mice are no longer segregated based on cytokine responses (Fig. 2F). These results indicate that there are genetic differences in stimulated cytokine profiles between *Nod2*$^{-/-}$, *Atg16l1*$^{T300A/+}$, and *Atg16l1*$^{T300A/T300A}$ mice under SPF conditions, but when challenged by the natural environment, these genetic differences are no longer apparent (Fig. 2, E and F), and consistent with environmental erosion of genetic effects on mucosal phenotypes of rewilded mice (*12*). Hence, the introduction of variation resulting from environmental experience may mask differences in stimulated cytokine profiles of genetically deficient mice kept under SPF conditions.

In addition to stimulated cytokine profiles, MLN cells were assessed for immune cell populations by flow cytometry analysis (fig. S5 and table S2) and gene expression by RNA-seq analysis (fig. S6 and table S2). Consistent with the peripheral blood, immune cell frequencies in MLNs showed a strong effect of environment (figs. S5 and S7). Furthermore, a group of rewilded mice all sacrificed at week 7 of rewilding, which experienced strong thunder storms and sustained high humidity right before trapping (unlike mice trapped at week 6), clustered separately (fig. S7, A and B), raising the possibility of particular events rather than cumulative environmental factors having an effect on memory T cells (fig. S7, C

and D). When the loading factors were examined, activation states of both CD4+ and CD8+ T cells were important to distinguishing lab mice from rewilded mice, as were total CD11b+CD11c+ myeloid cell frequencies (fig. S7C). RNA-seq analysis showed similarly strong environmental effects on variations (fig. S6). Differences in microbial communities in stool samples and especially fungal colonization is further discussed in the companion study (Yeung *et.al.*).

To integrate multi-omics data from flow cytometry analysis (Table S2), transcriptomics (Table S2), cytokine profiles (Table S2), and microbial profiles (Table S2), we concatenated these four data types (Table S3) and used supervised classification models to identify features that are predictive of the environment or genotype (Fig. 3). To understand how the different predictive features interact with each other, we computed pairwise correlation values between these predictors to generate a network for the environment and genotype models (Fig. 3, B and D). For the environment model, this revealed a hub of transcription factor genes (e.g. *Nfkbia*, *Egr1*, *Jun*, *Klf6*, *Atf4* and *Fosb*) being tightly correlated with T cell populations in the blood and MLN that distinguish the environmental differences (Fig. 3, A and B and fig. S8A). For the genotype model, our classification was not as accurate as the environment, as measured by area under the receiver operator curve (AUC) of 0.81 versus 1 on the test set (Supplementary materials). In addition to immune cell populations and transcriptional signatures, certain cytokine profiles (e.g. IFN-γ/*P. aeruginosa*, IL-5/PBS) were also predictive of the genotype (Fig 3, C and D, and fig. S8B). Hence, while the features associated with environmental classification are predominantly frequency of T cells and transcription factors, the features required to classify genotype differences on *Nod2*$^{-/-}$, *Atg16l1*$^{T300A/+}$ and *Atg16l1*$^{T300A/T300A}$ mice include the ability to produce cytokines, at baseline (PBS alone) or when presented with specific microbial stimuli (*P. aeruginosa*).

To evaluate the interactions between data types, we next utilized an unsupervised sparse partial least squares (sPLS) regression model to integrate the different data types into a multi-component network (Fig. 4A and fig. S9). Data matrices (Table S4) from flow cytometry data (blood and MLNs), transcriptional profiles (MLNs), microbial profiles (16S)

and stimulated cytokine profiles were used as inputs for sPLS-regression to generate a covariance network that included 252 nodes and 844 edges (Fig. 4A and fig. S9). In the resulting model, all microbial taxa fell below our 0.6 covariance threshold and were therefore removed (Table S4). This indicated that bacterial composition was not an important component of the network. IL-5 production in response to *B. vulgatus* stimulation had the most connections within the network and was found to be higher in rewilded mice (Fig. 4A; figs. S9 and S10). The $CD44^{hi}CD62L^{hi}$ CD4 T cell population in the flow cytometry data set was highly connected within the network and was also higher in rewilded mice (Fig. 4A; figs. S9 and S10). This approach groups transcriptional expression data into modules based on co-expression. A gene module (11.1) consisting of 206 co-regulated genes was the most connected transcriptional signature within the network and was lower in rewilded mice (Fig. 4A; figs. S9 and S10). This module was enriched for genes involved in transferrin transport, cellular response to interferon-beta, leukocyte cell-cell adhesion and myeloid cell activation (Fig. 4, B and C).

We then probed our multi-component network using a gene set analysis (GSA) method to identify sub-network differences driven by environment (Fig. 4D) and genotype (fig. S11). A twelve node sub-network, which included production of CCL2 in response to *C. albicans* and IL-1α in response to *B. vulgatus*, had the highest enrichment score distinguishing between lab and rewilded mice (Fig. 4D). Components of this subnetwork separated mice based on environmental conditions (Fig. 4E). The gene module 11.1.2, while connected to CCL2/*C. albicans* and IL-1α/*B. vulgatus*, was part of a subnetwork enriched for the effects of genotype (fig. S11) and was generally higher in lab mice. The gene module 39 (consisting of 21 genes) included several involved in mitochondrial respiration (e.g. *Atp5k*, *Atp5j2*) that were enriched in rewilded mice, while the gene module 22 (consisting of 71 genes) was higher in lab mice and appears to be related to B cells (fig. S12). Notably, genotype differences were less pronounced, potentially because they may be masked by the effects of the environment (Fig. 2 and fig. S11). This multi-omics network analysis therefore substantiated the omics-by-omics inferences (Figs. 1 and 2) but also provided new insight by

identifying an important role for *C. albicans* responses during rewilding, which is addressed in the companion study (Yeung *et.al.*).

In summary, we found that changing the environment profoundly contributes to the inter-individual variations on immune cell frequencies, whereas cytokine responses to pathogen stimulation were more affected by genetic deficiencies in the IBD susceptibility genes *Nod2* and *Atg16L1* and primarily under SPF conditions. Perhaps environment is a primary driver of the composition of the immune system, but genetics is a stronger driver of per-cell responsiveness. Our finding that genotype-specific differences were reduced upon a change in the environment highlights the need to investigate environmental variables when assessing gene function and suggests that certain types of exposure mask the contribution of susceptibility genes to inflammatory diseases. A systems level analysis identified interconnected networks of transcriptional signatures, immune cell populations and cytokine profiles to microbial stimulation, highlighting a potentially important role for fungal stimulation from rewilding (Companion study, Yeung *et. al.*). Hence, rewilding laboratory mice of different genetic backgrounds and careful monitoring of other non-heritable influences (e.g. differential behavior via which divergent immunological experience of individuals may accrue, by analogy with causes of divergence in neurological phenotype during rewilding (*26, 27*)) could be a powerful new approach towards dissecting drivers of immune response heterogeneity under homeostatic as well as during disease settings or infectious challenges.

References:

1. **M. Schirmer, V. Kumar, M. G. Netea, R. J. Xavier, *Curr Opin Immunol* 54, 50 (Oct, 2018).**
2. **P. Brodin, M. M. Davis, *Nat Rev Immunol* 17, 21 (Jan, 2017).**
3. **P. Brodin *et al.*, *Cell* 160, 37 (Jan 15, 2015).**
4. **O. B. Bakker *et al.*, *Nat Immunol* 19, 776 (Jul, 2018).**
5. **M. Schirmer *et al.*, *Cell* 167, 1897 (Dec 15, 2016).**
6. **R. Ter Horst *et al.*, *Cell* 167, 1111 (Nov 3, 2016).**
7. **Y. Li *et al.*, *Cell* 167, 1099 (Nov 3, 2016).**
8. **L. K. Beura *et al.*, *Nature* 532, 512 (Apr 28, 2016).**
9. **S. P. Rosshart *et al.*, *Cell* 171, 1015 (Nov 16, 2017).**
10. **S. P. Rosshart *et al.*, *Science* 365, (Aug 2, 2019).**
11. **S. Abolins *et al.*, *Nature communications* 8, 14811 (May 3, 2017).**
12. **J. M. Leung *et al.*, *PLoS Biol* 16, e2004108 (Mar, 2018).**
13. **T. A. Suzuki *et al.*, *Mol Ecol* 28, 3197 (Jul, 2019).**
14. **M. Wlodarska, A. D. Kostic, R. J. Xavier, *Cell Host Microbe* 17, 577 (May 13, 2015).**
15. **S. Y. Wong, K. Cadwell, *PLoS Pathog* 14, e1007215 (Sep, 2018).**
16. **K. Cadwell *et al.*, *Cell* 141, 1135 (Jun 25, 2010).**
17. **D. Ramanan, M. S. Tang, R. Bowcutt, P. Loke, K. Cadwell, *Immunity* 41, 311 (Aug 21, 2014).**
18. **A. Biswas *et al.*, *Proc Natl Acad Sci U S A* 107, 14739 (Aug 17, 2010).**
19. **J. Pott, A. M. Kabat, K. J. Maloy, *Cell Host Microbe* 23, 191 (Feb 14, 2018).**
20. **Y. Matsuzawa-Ishimoto *et al.*, *J Exp Med*, (Oct 31, 2017).**
21. **D. Ramanan *et al.*, *Science* 352, 608 (Apr 29, 2016).**
22. **S. Lavoie *et al.*, *eLife* 8, (Jan 22, 2019).**
23. **E. Kernbauer, Y. Ding, K. Cadwell, *Nature* 516, 94 (Dec 4, 2014).**
24. **R. Caruso *et al.*, *Sci Immunol* 4, (Apr 19, 2019).**
25. **E. Becht *et al.*, *Nat Biotechnol*, (Dec 3, 2019).**
26. **J. Freund *et al.*, *Science* 340, 756 (May 10, 2013).**
27. **E. C. Cope *et al.*, *Hippocampus* 29, 366 (Apr, 2019).**

Acknowledgements

We wish to thank William Craigens, Daniel Navarrete Prado, Allison Lee, and Veena Chittamuri for assistance with trapping and husbandry in the field, the PU Lab Animal Resources staff for logistical support. We wish to thank the NYU School of Medicine Flow Cytometry and Cell Sorting, Microscopy, Genome Technology, and Histology Cores for use of their instruments and technical assistance (supported in part by National Institute of Health (NIH) grant P31CA016087, S10OD01058, and and S10OD018338). **Funding:** this research was supported by US National Institute of Health (NIH) grants DK103788 (K.C. and P.L.), AI121244 (K.C.), HL123340 (K.C.), DK093668 (K.C.), AI130945 (P.L.), R01 HL125816 (K.C.), HL084312, AI133977 (P.L.), research station and research rebate awards from PU EEB (A.L.G.), pilot award from the NYU CTSA grant UL1TR001445 from the National Center for Advancing Translational Sciences (NCATS) (K.C., P.L.), pilot award from the NYU Cancer Center grant P30CA016087 (K.C.), AI100853 (Y.C.), and DK122698 (F.Y.). This work was also supported by the Department of Defense grant W81XWH-16-1-0256 (P.L.), Faculty Scholar grant from the Howard Hughes Medical Institute (K.C.), Crohn's & Colitis Foundation (K.C.), Merieux Institute (K.C.), Kenneth Rainin Foundation (K.C.), Stony-Wold Herbert Fund (K.C.), and Bernard Levine Postdoctoral Research Fellowship in Immunology (Y.C.). K.C. is a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Diseases. **Author contributions:** Design of experiments, data analysis, data discussion, and interpretation: J.D.L., J.C.D., F.Y., K.C., A.L.G, and P.L.; primary responsibility for execution of experiments: J.D.L., F.Y., C.M., J.M.L., Y.H.C., A.C., C.H., and C.D.D.; MLN cell RNA and 16S analysis: J.C.D., K.V.R., J.D.L., and F.Y. Supervised and Unsupervised machine learning model analysis: J.C.D., K.V.R., and J.D.L. All authors discussed data and commented on the manuscript. **Competing interests:** K.C. and P.L. receive research funding from Pfizer. K.C. has consulted for or received an honorarium from Puretech Health, Genentech, and Abbvie. P.L. consults for and has equity in Toilabs. **Data and materials availability:** Raw sequence data from 16*S* , ITS, and RNA sequencing experiments are deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA559026 and gene expression omnibus (GEO) accession number GSE135472.

Figure Legends:

Fig. 1. Environmental change drives inter-individual variation in immune cell populations. (**A**) 25 C57BL/6$^{+/+}$, 28 *Nod2*$^{-/-}$, 27 *Atg16l1*$^{T300A/+}$, 24 *Atg16l1*$^{T300A/T300A}$ mice (total=104) were housed in the wild enclosure (Rewilded) for 6-7 weeks and successfully trapped for flow cytometry analysis of lymphoid and myeloid cell populations in the blood and mesenteric lymph nodes (MLNs). Age matched 19 C57BL/6$^{+/+}$, 19 *Nod2*$^{-/-}$, 20 *Atg16l1*$^{T300A/+}$, 22 *Atg16l1*$^{T300A/T300A}$ mice (total=80) housed under SPF conditions (Lab) were analyzed as controls. Feces were collected for 16S rRNA sequencing. MLN cells were cultured with indicated stimulates for 48 hours and supernatants were collected for cytokine profiling. Samples that fail quality control are not included in downstream analyses. (**B** and **C**) UMAP projections of ~180,000 CD45+, ~110,000 CD19+, and ~36,000 CD4+ cells on flow cytometry data of the blood from lab and rewilded mice. Events are color-coded according to CD19, CD44 (B), CD4, and CD62L (C) fluorescence level. Box plots show the abundance of CD44^{hi} CD19+ cells (B) and CD62L^{hi} CD4+ cells (C) in the blood of individual lab and rewilded mice. (**D**) Principal coordinates analysis (PCoA) of gated immune cell populations in the blood and the density of each population along the principal coordinates (PC). Right panel indicate biplots of the gated immune cell populations that are projected onto PC1 and PC2. (**E**) Box plot of pairwise Euclidean distance measures based on blood immune cell population in the blood of lab versus lab, rewilded versus rewilded and lab versus rewilded mice. (**F**) Bar plot showing the pseudo R^2 measure of effect size on the entire distance matrix used to calculate the PCoA of immune cell populations in the blood (D). ***$P < 0.001$ by Kruskal-Wallis test between groups (B, C) and ***$P < 0.001$ by Mann-Whitney-Wilcoxon test between groups (E).

Fig. 2. The genetic contribution towards variation in cytokine production in response to microbial stimulation is masked when mice are rewilded. (**A** to **F**) MLN cells from lab and rewilded mice were *ex vivo* cultured with *B. subtilis*, *S. aureus*, *C. perfringens*, *B. vulgatus*, *P. aeruginosa*, *C. albicans*, and αCD3/CD28 beads for 48 hours and the supernatants were assayed for 13 cytokines. (**A**) Heat map of average fold changes compared to cytokine levels of MLN cells cultured with PBS controls from all mice. (**B**) Bar plot showing the top 10 $\log_2$ fold changes of cytokine responses to stimulation for MLN cells from lab and rewilded mice. (**C**) PCoA plot on cytokine stimulation profiles with the histogram of the total variance explained by each PC indicated on the axis. Right panel indicates the loading factor of each stimulated cytokine profile that contributes to each PC. (**D**) Bar plot showing the pseudo R^2 measure of effect size on the entire distance matrix used to calculate the PCoA of cytokine profiles (C). (**E** and **F**) Heatmap of pairwise correlation values from

stimulated cytokine profiles of lab (E) and rewilded (F) mice. * P <0.05, ** P <0.01, *** *P* < 0.001 by Mann-Whitney-Wilcoxon test between groups (B).

Fig. 3. Multi-omic classification models identify predictive features of environment and genotype. (**A**) Bar plot of the top 40 features predictive of the environment and (**B**) interaction network analysis of correlations ($R^2 > 0.6$ or $R^2 < -0.6$) between features (genes, immune cell populations, OTUs or cytokines) that are predictive of environmental changes. (**C**) Bar plot of the top 40 features predictive of the genotype and (**D**) network analysis of the associations between features that are associated with genotype differences. Data types are colored according to cytokine profiles (green), genes (red), immune populations (blue), and OTUs (purple). The size of the nodes in B and D are scaled proportionally to the number of connections with other nodes in each network.

Fig. 4. An unsupervised multi-component model identifies a *C. albicans* and *B. vulgatus* response network associated with environmental changes. (**A** to **C**) Data matrices from flow cytometry data, transcriptional profiles, microbial profiles and stimulated cytokine profiles were used as inputs for a sPLS-regression model to generate a covariance network (A), which identifies a 206 gene module (11.1) that is highly connected within the network. (**B**) Bar plot of top 10 biological processes that are enriched in module 11.1 according to Panther and (**C**) heatmap of the module 11.1 genes associated with these biological processes. Color bars indicate environment and genotype. (**D** and **E**) The network (A) was queried by GSA to identify the most predictive sub-network (D) for differences driven by the environment. (**E**) Unsupervised hierarchical clustering analysis and heatmap of modules in this network separated lab and rewilded mice. Color bars indicate environment and genotype. In (A, D), data types are colored according to cytokine profiles (green), genes (red), immune populations (blue), and OTUs (purple). The size of the nodes in A and D are scaled proportionally to the number of connections with other nodes in each network.

Figure S1: Identification of immune cell populations that differ between lab and rewilded mice with UMAP. (**A**) Projection of ~180,000 CD45+ cells visualized by tSNE vs UMAP. (**B**) Identification of the major lymphocyte populations on UMAP based on CD19, CD3, CD4 and CD8 expression. (**C** to **E**) Expression of cell surface molecules on CD19+ cells (C), CD4+ cells (D) and CD8+ cells (E). In (C), (D) and (E), expression of 1B11, PD1, CD25, CD44, CD127, CXCR3, KLRG1, CD27, CD69, and CD62L are shown using cut-off values based on overall distribution of fluorescence expressions.

Figure S2: Flow cytometry gating strategy for the quantification of lymphoid populations in the blood of lab and rewilded mice. (**A** to **H**) Representative flow cytometry plots illustrating gating strategy (A) of lymphoid populations in the blood including the total CD45+ cells (B), total myeloid cells (C), total B and T cells (D), total CD4 and CD8 T cells (E), $CD44^{lo}CD62L^{hi}$, $CD44^{hi}CD62L^{hi}$, $CD44^{hi}CD62L^{lo}$, $CD44^{lo}CD62L^{lo}$ CD8 T cells (F), $CD44^{lo}CD62L^{hi}$, $CD44^{hi}CD62L^{hi}$, $CD44^{hi}CD62L^{lo}$, $CD44^{lo}CD62L^{lo}$ CD4 T cells (G), and CD25+CD4+ T cells (H).

Figure S3: Hierarchy of variance for cytokine production after stimulation with microbial antigens *in vitro*. (**A**) Bar plots showing average variance for cytokine production after stimulations. (**B**) Unsupervised hierarchical clustering and heatmap of variance in cytokine production in response to *B. subtilis*, *S. aureus*, *C. perfringens*, *B. vulgatus*, *P. aeruginosa*, *C. albicans*, and αCD3/CD28 stimulation.

Figure S4: A genetic contribution towards variation in plasma cytokine levels of lab and rewilded mice. (**A** and **B**) Levels of RANTES, MIP-3a, KC, TNF-α, MCP-1, IP-10, MIP-1a, MIP-1b, IL-1α, IL-6 cytokines were measured from the plasma of lab and rewilded mice at the time of sacrifice with bead-based immunoassays (LEGENDplex). (**A**) PCoA of plasma cytokine levels in the blood of individual mice and the density of each population along the principal coordinates (PC). Right panel indicate biplots of the plasma cytokines are projected onto PC1 and PC2. (**B**) Bar plot showing the pseudo R^2 measure of the effect size of different variables on the variance of plasma cytokine levels.

Figure S5: Memory CD4+ and CD8+ T cells are increased in the mesenteric lymph nodes (MLNs) of rewilded mice. (**A** to **F**) Representative flow cytometry plots for immune cell populations after gating for CD45+ cells in the MLNs of lab or rewilded C57BL/6 mice; for the total abundance of myeloid cells (A), B and T cells (B), CD4 and CD8 T cells (C), $CD44^{lo}CD62L^{hi}$, $CD44^{hi}CD62L^{hi}$, $CD44^{hi}CD62L^{lo}$, $CD44^{lo}CD62L^{lo}$ CD4 T cells (D),

$CD44^{lo}CD62L^{hi}$, $CD44^{hi}CD62L^{hi}$, $CD44^{hi}CD62L^{lo}$, $CD44^{lo}CD62L^{lo}$ CD8 T cells (E), and CD25+CD4+ T cells (F).

Figure S6: Environment has the greatest effect on variation of transcriptional profiles. (**A** to **B**) Transcriptional profiles of MLNs were determined by Celseq. (**A**) PCoA based on gene profiles in the MLN and the density of each population along the principal coordinates (PC). Right panel indicates biplots of the genes projected onto PC1 and PC2. (**B**) Bar plot showing the pseudo R^2 measure of effect size on the entire distance matrix used to calculate the PCoA of gene expression levels.

Figure S7: Unmeasured environmental effects drive additional variation of MLN immune cell populations for the rewilded mice on the week 6 of rewilding. (**A** to **D**) A thunderstorm over the weekend between week 6 and week 7 of rewilding may have introduced additional variation into immune cell populations. (**A** and **B**) PCoA based on lymphoid populations in the MLN and the density of each population along the principal coordinates (PC). In plot (A), dot color coded by rewilded (red) and lab (purple) and in plot (B), dot color coded by lab (green), rewilded week 6 (yellow), rewilded week 7 (blue) (**C**) Biplots of the MLN lymphocyte populations are projected onto PC1 and PC2. (**D**) Bar plot showing the pseudo R^2 measure of effect size on the entire distance matrix used to calculate the PCoA of immune populations.

Figure S8: Multi-omic classification models identify transcriptional factors and certain cytokines are predictive of environmental and genotype differences. (**A** and **B**) Top predictive features of environment (A) and genotype (B). * P <0.05, ** P <0.01, *** $P <$ 0.001 by Mann-Whitney-Wilcoxon test between groups.

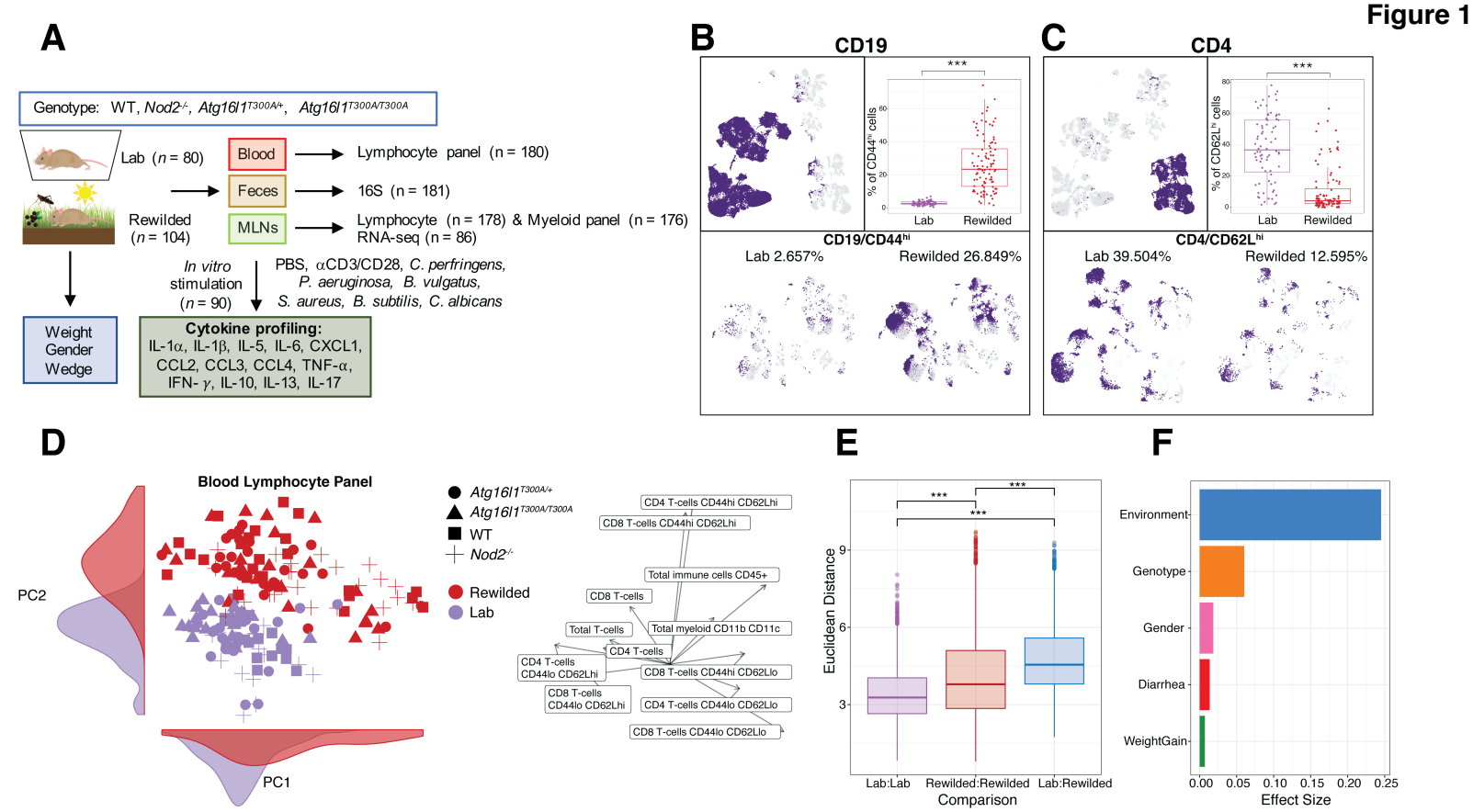
Figure S9: An unsupervised PLS-regression multi-component network model for multi-omic integration. A fully annotated multi-component network as shown in Fig. 4A.

Figure S10: The most connected features from each data type within multi-component network. (**A** to **C**) Bar plots indicate IL-5 production of MLN cells in response to *B. vulgatus* (A), gene module 11.1 (B) and $CD44^{hi}CD62L^{hi}$ CD4 T cells in the blood (C) are the most connected features within the multi-component network. ***$P < 0.001$ by Mann-Whitney-Wilcoxon test between lab and rewilded mice.

Figure S11: Heatmap of gene modules associated with rewilding. (**A** and **B**) Heatmap of gene expression levels for transcripts in gene modules 39 (A) and 22 (B). Color bars indicate environment and genotype.

Figure S12: Subnetwork associated with genotype differences. (**A**) The subnetwork most associated with genotype differences is shown. Data types are colored according to cytokine profiles (green), genes (red), immune populations (blue), and OTUs (purple). The sizes of the nodes are scaled proportionally to the number of connections with other nodes in the network. (B) Unsupervised hierarchical clustering analysis and heatmap of modules in this network separate by genotype and environment. Color bars indicate environment and genotype.

Figure 1


A
Genotype: WT, Nod2-/-, Atg16l1T300A/+, Atg16l1T300A/T300A
Lab (n = 80)
Rewilded (n = 104)
Blood
Feces
MLNs
Lymphocyte panel (n = 180)
16S (n = 181)
Lymphocyte (n = 178) & Myeloid panel (n = 176)
RNA-seq (n = 86)
In vitro stimulation (n = 90)
PBS, αCD3/CD28, C. perfringens, P. aeruginosa, B. vulgatus, S. aureus, B. subtilis, C. albicans
Weight
Gender
Wedge
Cytokine profiling:
IL-1α, IL-1β, IL-5, IL-6, CXCL1, CCL2, CCL3, CCL4, TNF-α, IFN-γ, IL-10, IL-13, IL-17
B
CD19
% of CD44hi cells
Lab
Rewilded
CD19/CD44hi
Lab 2.657%
Rewilded 26.849%
C
CD4
% of CD62Lh cells
CD4/CD62Lhi
Lab 39.504%
Rewilded 12.595%
D
Blood Lymphocyte Panel
PC2
PC1
Atg16l1T300A/+
Atg16l1T300A/T300A
WT
Nod2-/-
Rewilded
Lab
CD4 T-cells CD44hi CD62Lhi
CD8 T-cells CD44hi CD62Lhi
Total immune cells CD45+
CD8 T-cells
Total T-cells
Total myeloid CD11b CD11c
CD4 T-cells
CD4 T-cells CD44lo CD62Lhi
CD8 T-cells CD44hi CD62Llo
CD8 T-cells CD44lo CD62Lhi
CD4 T-cells CD44lo CD62Llo
CD8 T-cells CD44lo CD62Llo
E
Euclidean Distance
Lab:Lab
Rewilded:Rewilded
Lab:Rewilded
Comparison
F
Environment
Genotype
Gender
Diarrhea
WeightGain
Effect Size

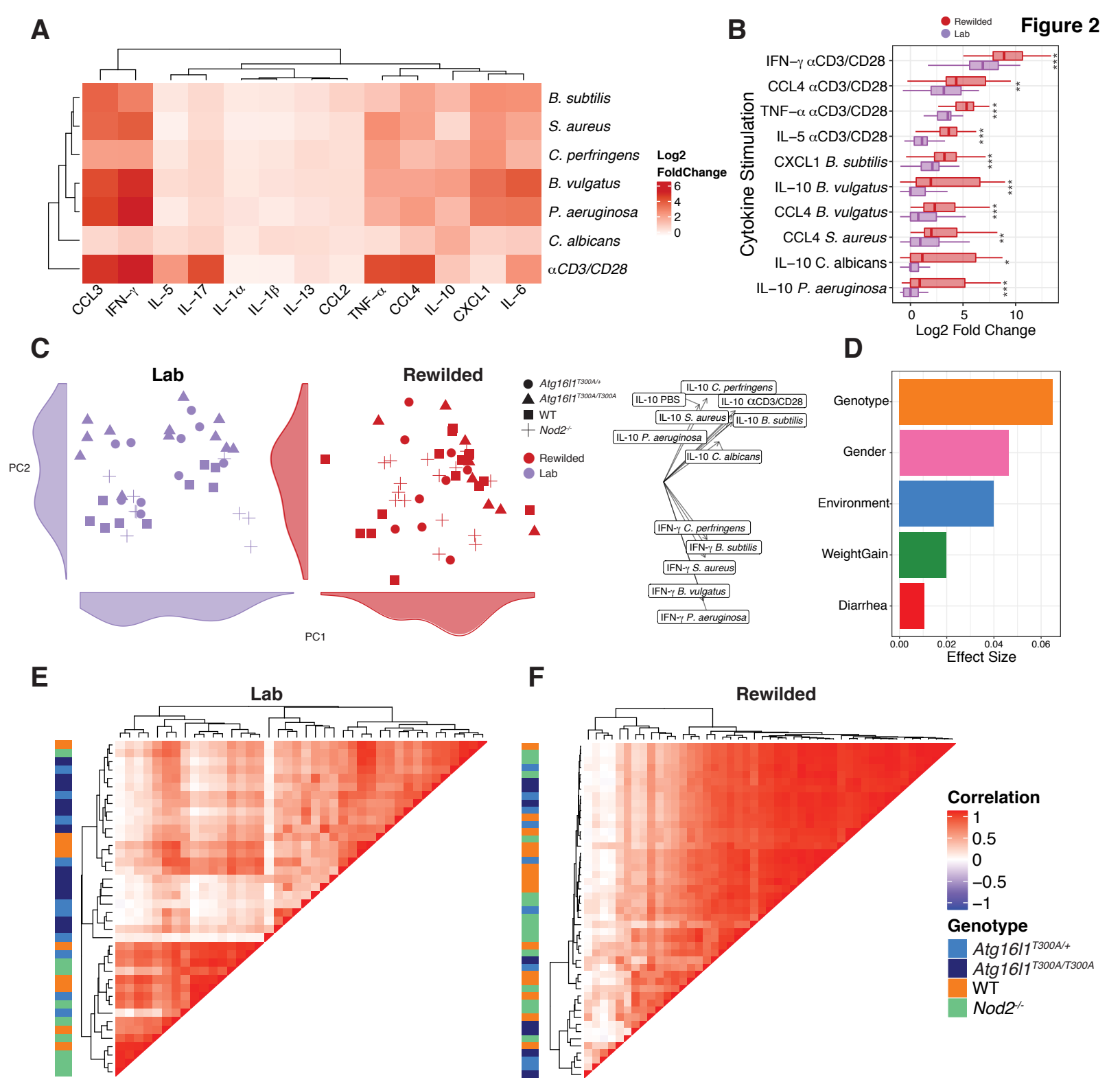


Figure 2
A
B. subtilis
S. aureus
C. perfringens
B. vulgatus
P. aeruginosa
C. albicans
αCD3/CD28
CCL3
IFN−γ
IL−5
IL−17
IL−1α
IL−1β
IL−13
CCL2
TNF−α
CCL4
IL−10
CXCL1
IL−6
Log2 FoldChange
6
4
2
0
B
Rewilded
Lab
Cytokine Stimulation
IFN−γ αCD3/CD28
CCL4 αCD3/CD28
TNF−α αCD3/CD28
IL−5 αCD3/CD28
CXCL1 B. subtilis
IL−10 B. vulgatus
CCL4 B. vulgatus
CCL4 S. aureus
IL−10 C. albicans
IL−10 P. aeruginosa
0
5
10
Log2 Fold Change
C
Lab
Rewilded
PC2
PC1
Atg16l1T300A/+
Atg16l1T300A/T300A
WT
Nod2-/-
Rewilded
Lab
IL-10 C. perfringens
IL-10 PBS
IL-10 αCD3/CD28
IL-10 S. aureus
IL-10 B. subtilis
IL-10 P. aeruginosa
IL-10 C. albicans
IFN-γ C. perfringens
IFN-γ B. subtilis
IFN-γ S. aureus
IFN-γ B. vulgatus
IFN-γ P. aeruginosa
D
Genotype
Gender
Environment
WeightGain
Diarrhea
0.00
0.02
0.04
0.06
Effect Size
E
Lab
F
Rewilded
Correlation
1
0.5
0
−0.5
−1
Genotype
Atg16l1T300A/+
Atg16l1T300A/T300A
WT
Nod2-/-

Figure 3

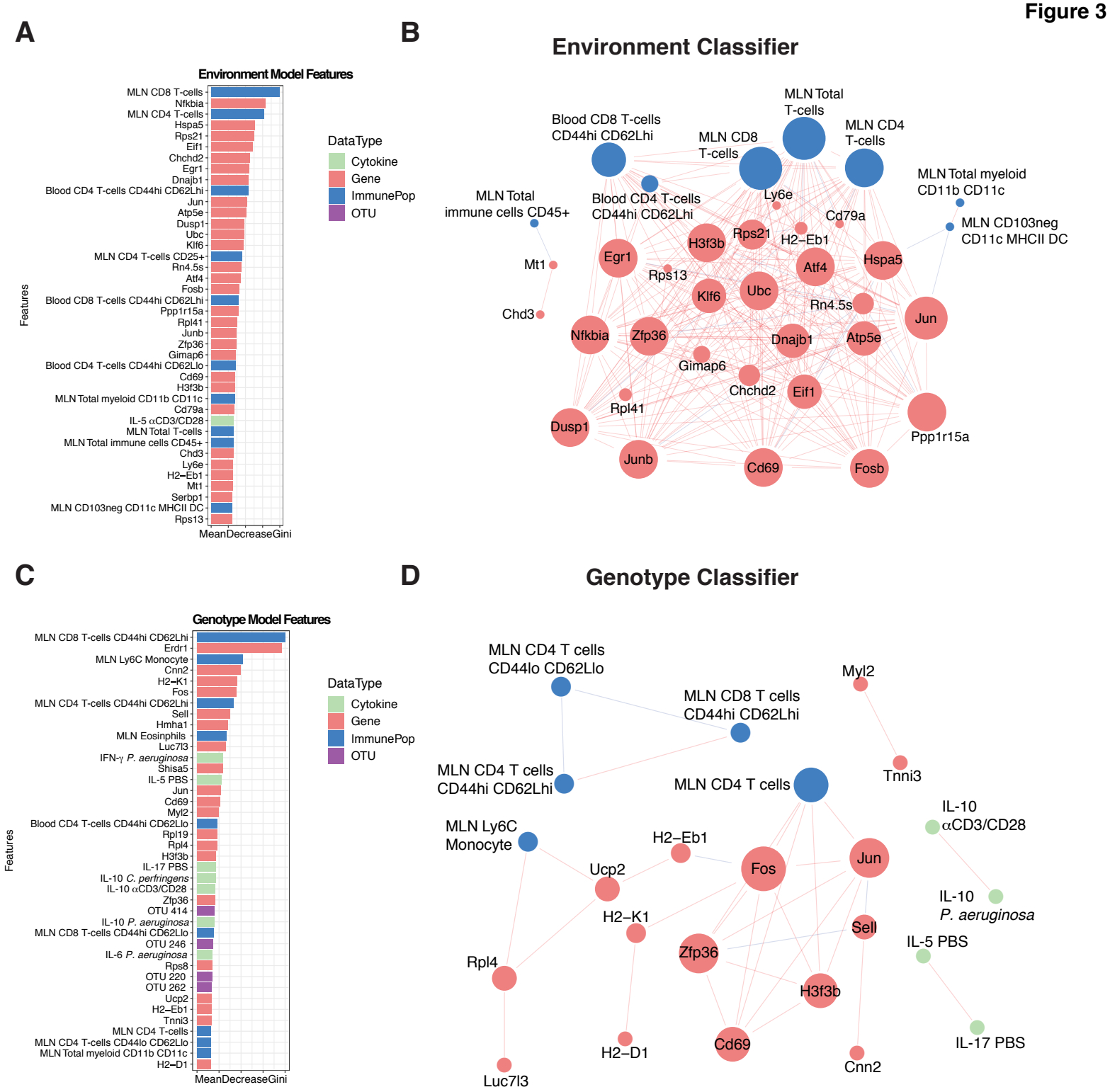
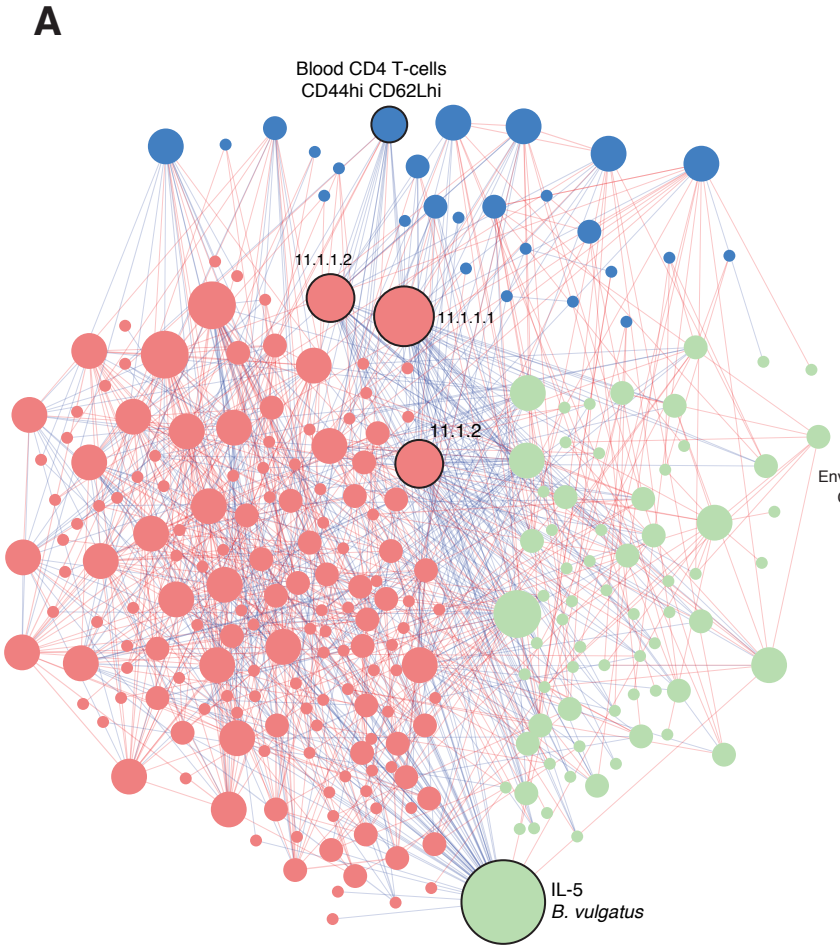
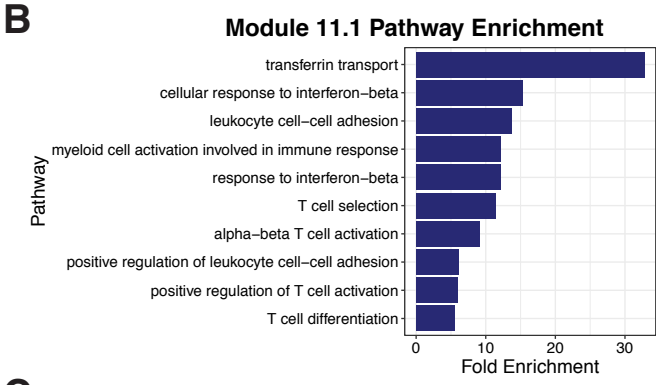
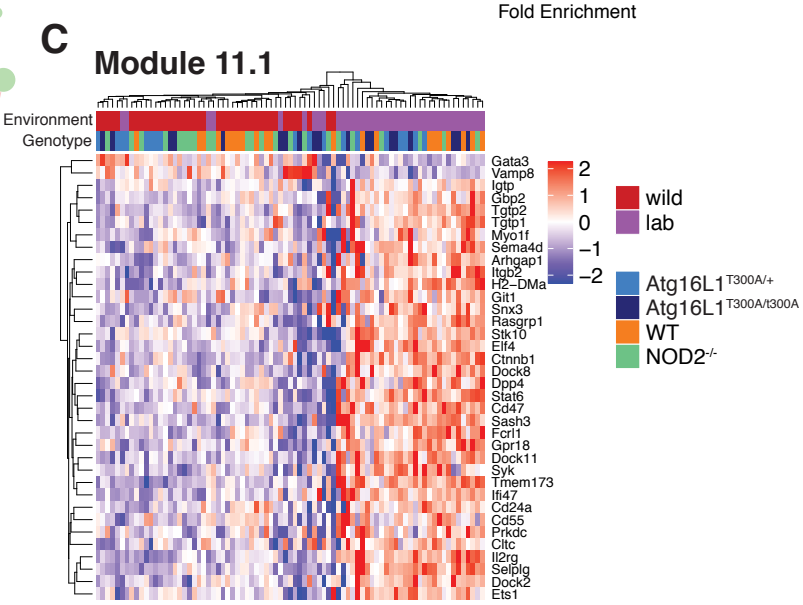
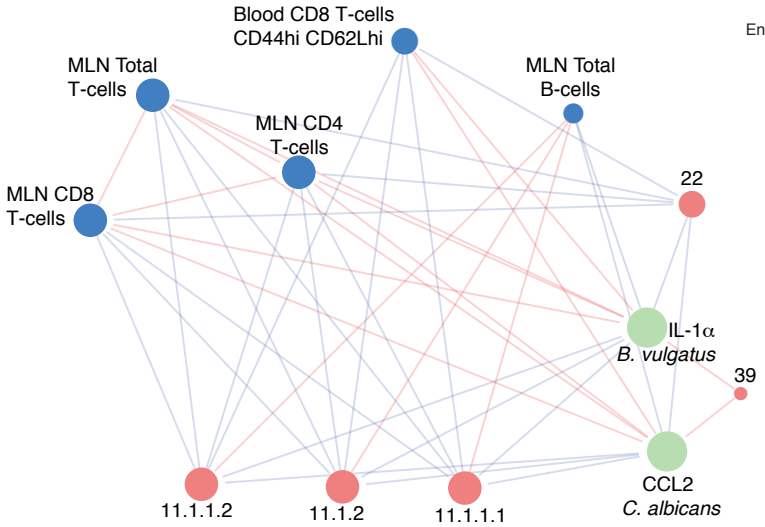

A
Environment Model Features
MLN CD8 T-cells
Nfkbia
MLN CD4 T-cells
Hspa5
Rps21
Eif1
Chchd2
Egr1
Dnajb1
Blood CD4 T-cells CD44hi CD62Lhi
Jun
Atp5e
Dusp1
Ubc
Klf6
MLN CD4 T-cells CD25+
Rn4.5s
Atf4
Fosb
Blood CD8 T-cells CD44hi CD62Lhi
Ppp1r15a
Rpl41
Junb
Zfp36
Gimap6
Blood CD4 T-cells CD44hi CD62Llo
Cd69
H3f3b
MLN Total myeloid CD11b CD11c
Cd79a
IL-5 αCD3/CD28
MLN Total T-cells
MLN Total immune cells CD45+
Chd3
Ly6e
H2–Eb1
Mt1
Serbp1
MLN CD103neg CD11c MHCII DC
Rps13
Features
MeanDecreaseGini
DataType
Cytokine
Gene
ImmunePop
OTU
B
Environment Classifier
MLN Total T-cells
Blood CD8 T-cells CD44hi CD62Lhi
MLN CD8 T-cells
MLN CD4 T-cells
MLN Total myeloid CD11b CD11c
MLN Total immune cells CD45+
Blood CD4 T-cells CD44hi CD62Lhi
Ly6e
Cd79a
MLN CD103neg CD11c MHCII DC
H3f3b
Rps21
H2–Eb1
Egr1
Atf4
Hspa5
Mt1
Rps13
Klf6
Ubc
Rn4.5s
Jun
Chd3
Nfkbia
Zfp36
Dnajb1
Atp5e
Gimap6
Chchd2
Eif1
Rpl41
Dusp1
Ppp1r15a
Junb
Cd69
Fosb
C
Genotype Model Features
MLN CD8 T-cells CD44hi CD62Lhi
Erdr1
MLN Ly6C Monocyte
Cnn2
H2–K1
Fos
MLN CD4 T-cells CD44hi CD62Lhi
Sell
Hmha1
MLN Eosinphils
Luc7l3
IFN-γ P. aeruginosa
Shisa5
IL-5 PBS
Jun
Cd69
Myl2
Blood CD4 T-cells CD44hi CD62Llo
Rpl19
Rpl4
H3f3b
IL-17 PBS
IL-10 C. perfringens
IL-10 αCD3/CD28
Zfp36
OTU 414
IL-10 P. aeruginosa
MLN CD8 T-cells CD44hi CD62Llo
OTU 246
IL-6 P. aeruginosa
Rps8
OTU 220
OTU 262
Ucp2
H2–Eb1
Tnni3
MLN CD4 T-cells
MLN CD4 T-cells CD44lo CD62Llo
MLN Total myeloid CD11b CD11c
H2–D1
Features
MeanDecreaseGini
DataType
Cytokine
Gene
ImmunePop
OTU
D
Genotype Classifier
MLN CD4 T cells CD44lo CD62Llo
Myl2
MLN CD8 T cells CD44hi CD62Lhi
MLN CD4 T cells CD44hi CD62Lhi
MLN CD4 T cells
Tnni3
IL-10 αCD3/CD28
MLN Ly6C Monocyte
H2–Eb1
Fos
Jun
Ucp2
IL-10 P. aeruginosa
H2–K1
Sell
IL-5 PBS
Zfp36
Rpl4
H3f3b
IL-17 PBS
H2–D1
Cd69
Cnn2
Luc7l3

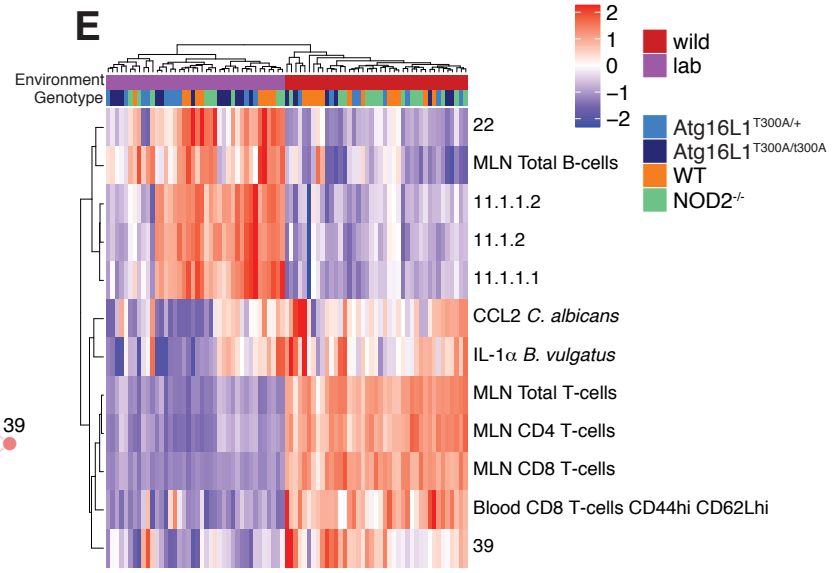
Figure 4


A
Blood CD4 T-cells
CD44hi CD62Lhi
11.1.1.2
11.1.1.1
11.1.2
IL-5
B. vulgatus

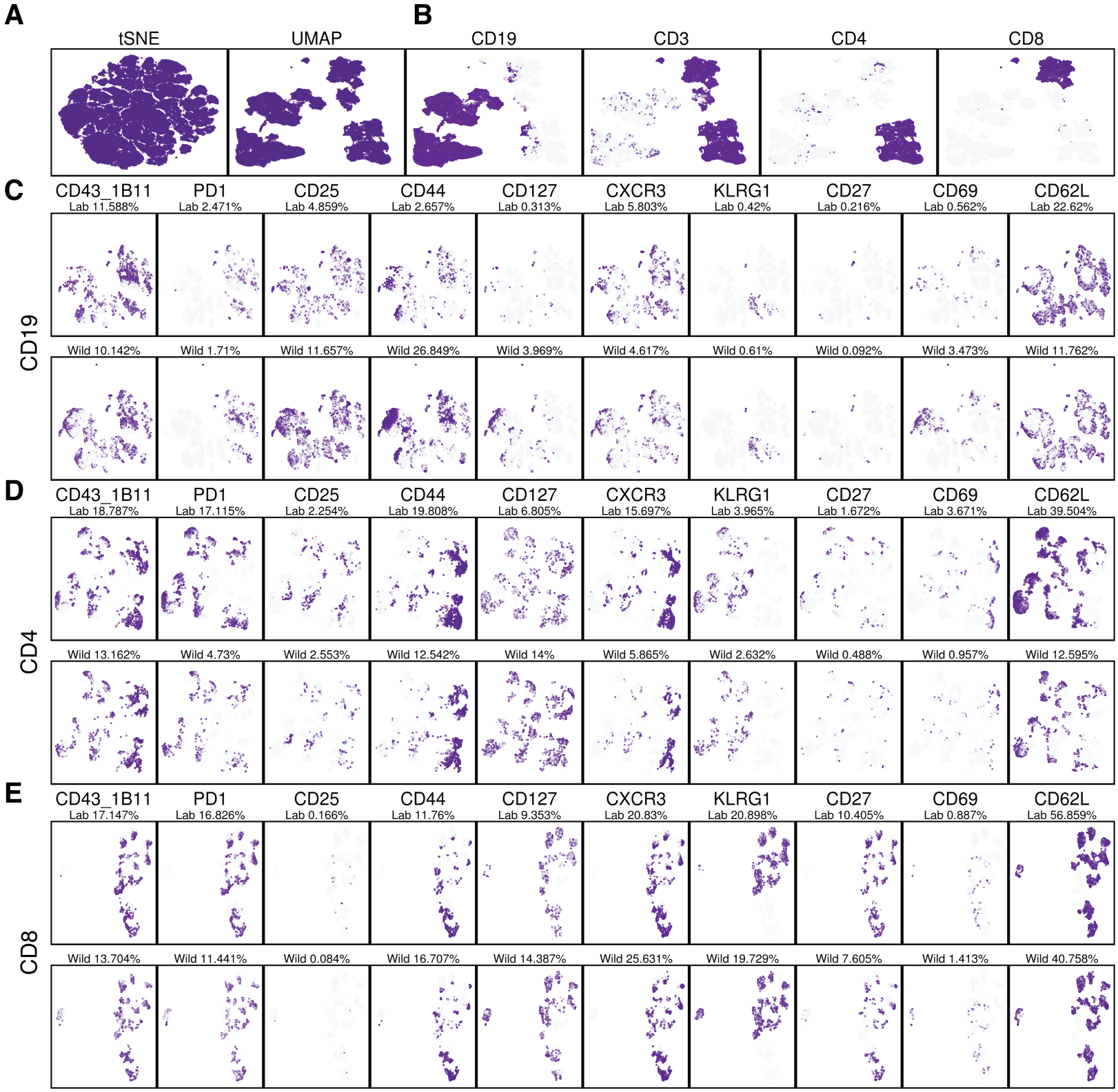

B
Module 11.1 Pathway Enrichment
Pathway
transferrin transport
cellular response to interferon−beta
leukocyte cell−cell adhesion
myeloid cell activation involved in immune response
response to interferon−beta
T cell selection
alpha−beta T cell activation
positive regulation of leukocyte cell−cell adhesion
positive regulation of T cell activation
T cell differentiation
0
10
20
30
Fold Enrichment


C
Module 11.1
Environment
Genotype
Gata3
Vamp8
Igtp
Gbp2
Tgtp2
Tgtp1
Myo1f
Sema4d
Arhgap1
Itgb2
H2−DMa
Git1
Snx3
Rasgrp1
Stk10
Elf4
Ctnnb1
Dock8
Dpp4
Stat6
Cd47
Sash3
Fcrl1
Gpr18
Dock11
Syk
Tmem173
Ifi47
Cd24a
Cd55
Prkdc
Cltc
Il2rg
Selplg
Dock2
Ets1
2
1
0
−1
−2
wild
lab
Atg16L1T300A/+
Atg16L1T300A/t300A
WT
NOD2-/-

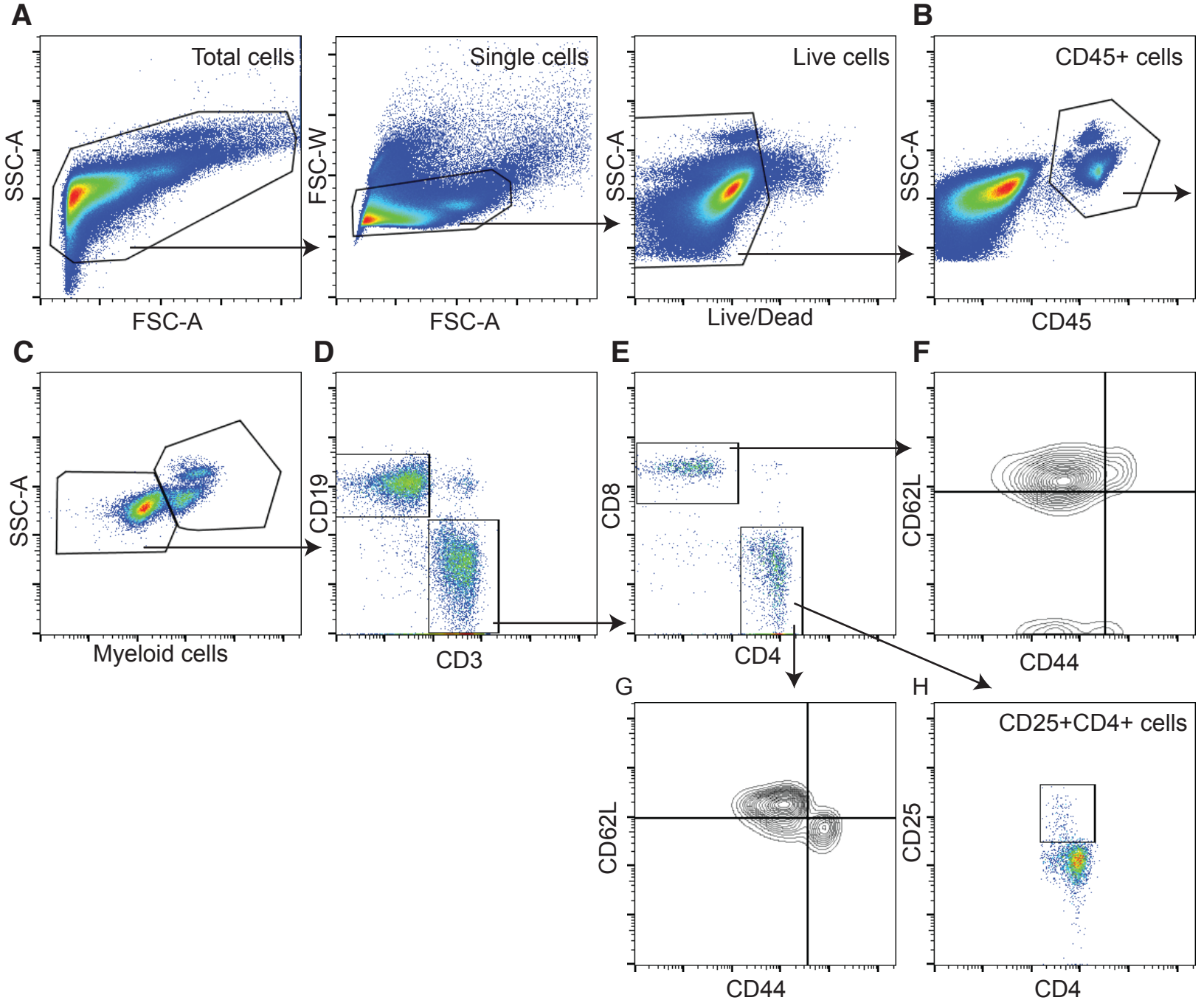

D
Blood CD8 T-cells
CD44hi CD62Lhi
MLN Total
T-cells
MLN Total
B-cells
MLN CD4
T-cells
MLN CD8
T-cells
22
IL-1α
B. vulgatus
39
CCL2
C. albicans
11.1.1.2
11.1.2
11.1.1.1


E
Environment
Genotype
22
MLN Total B-cells
11.1.1.2
11.1.2
11.1.1.1
CCL2 C. albicans
IL-1α B. vulgatus
MLN Total T-cells
MLN CD4 T-cells
MLN CD8 T-cells
Blood CD8 T-cells CD44hi CD62Lhi
39
2
1
0
−1
−2
wild
lab
Atg16L1T300A/+
Atg16L1T300A/t300A
WT
NOD2-/-

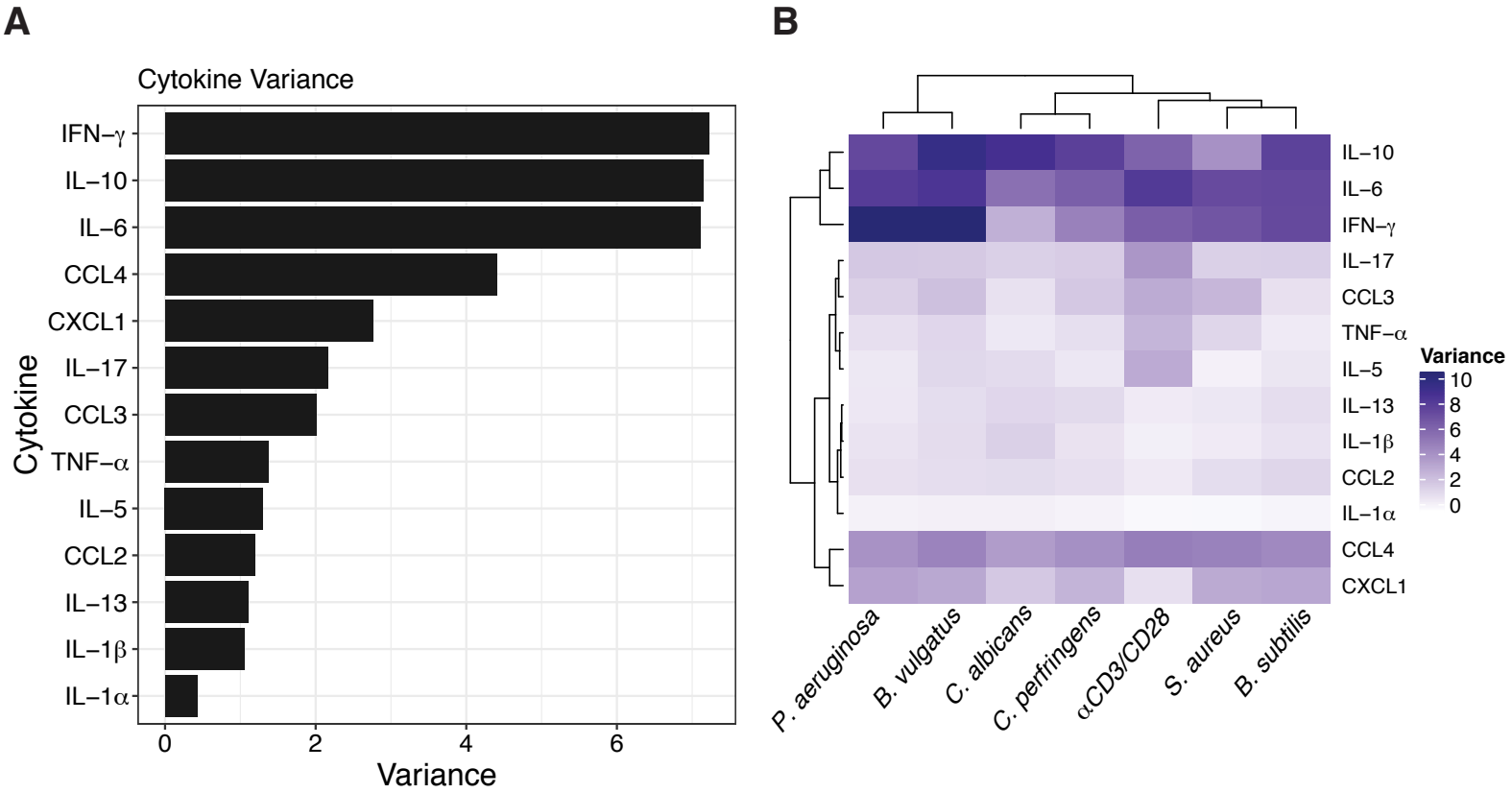
Supplementary Figure 1



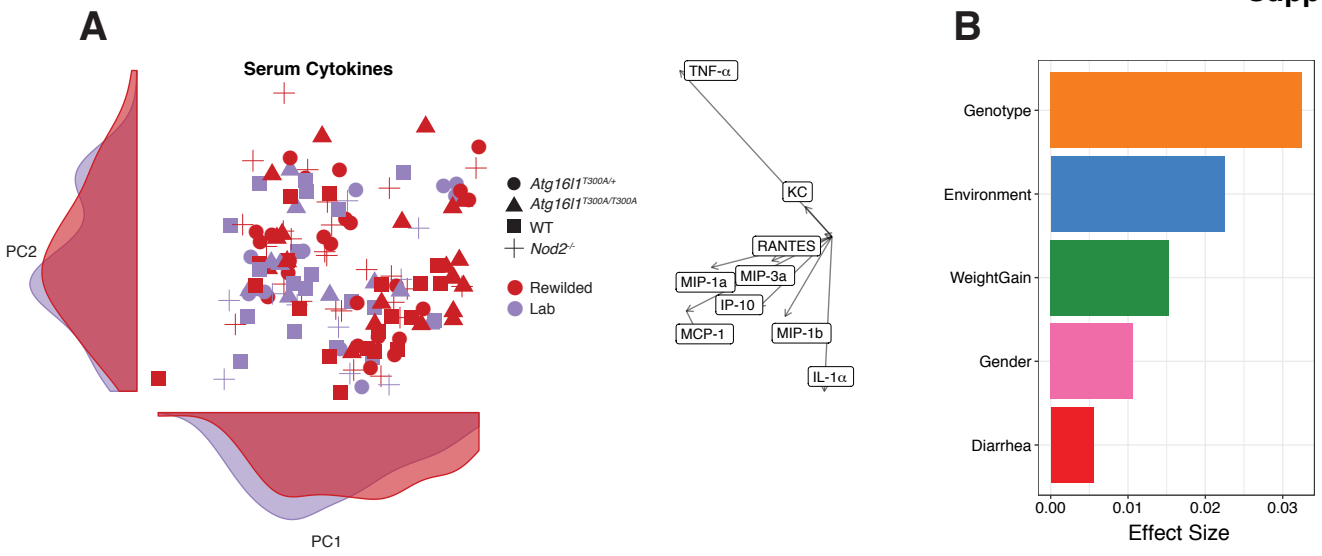
Supplementary Figure 2



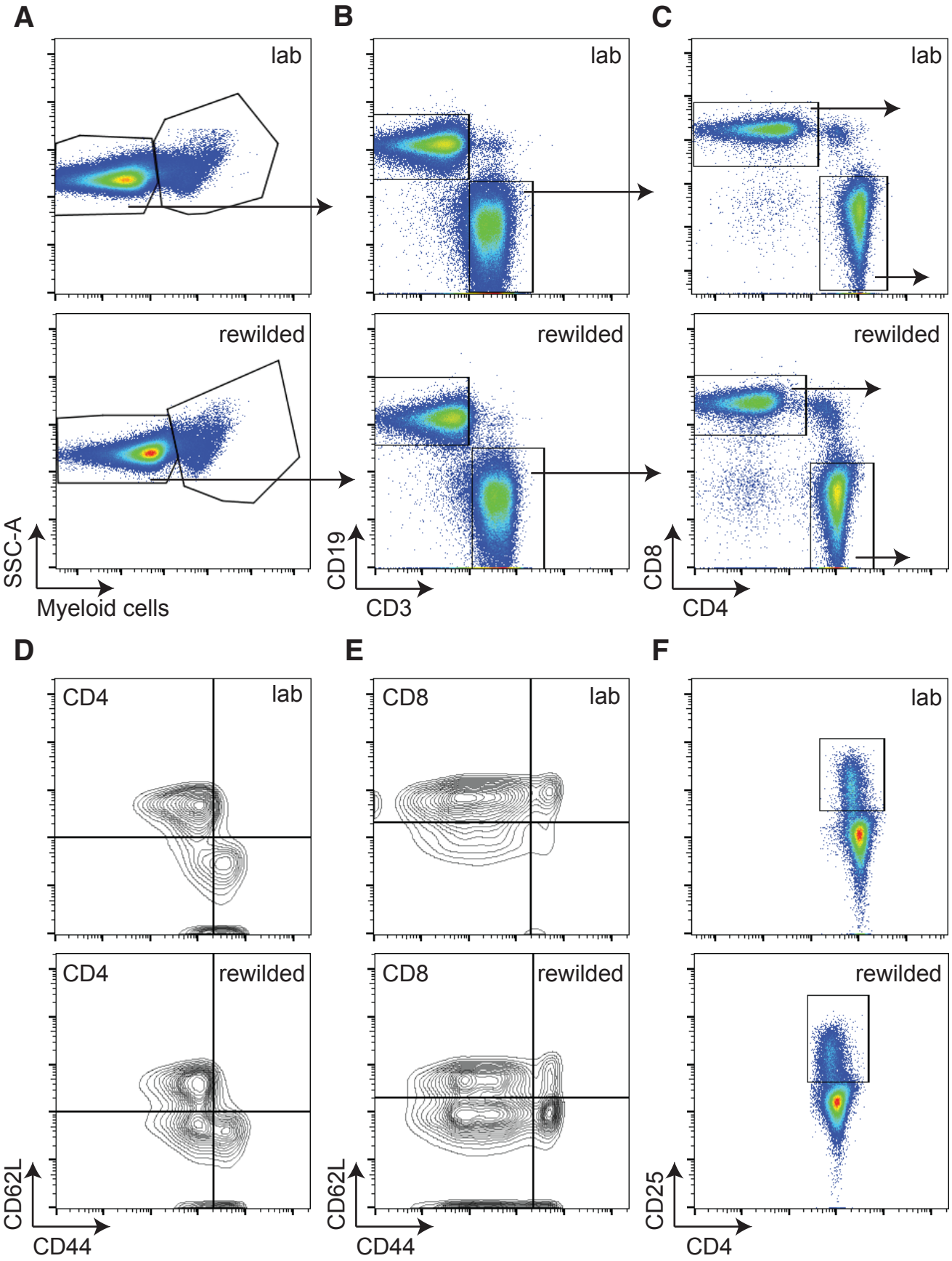
Supplementary Figure 3



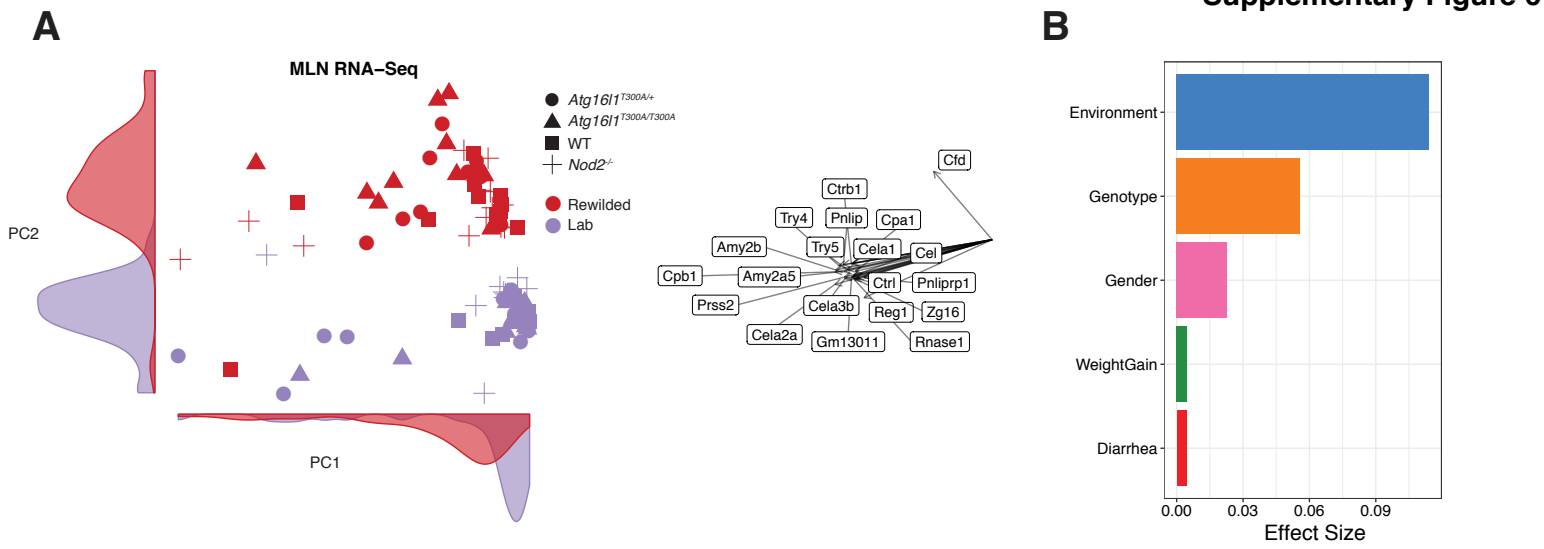
Supplementary Figure 4



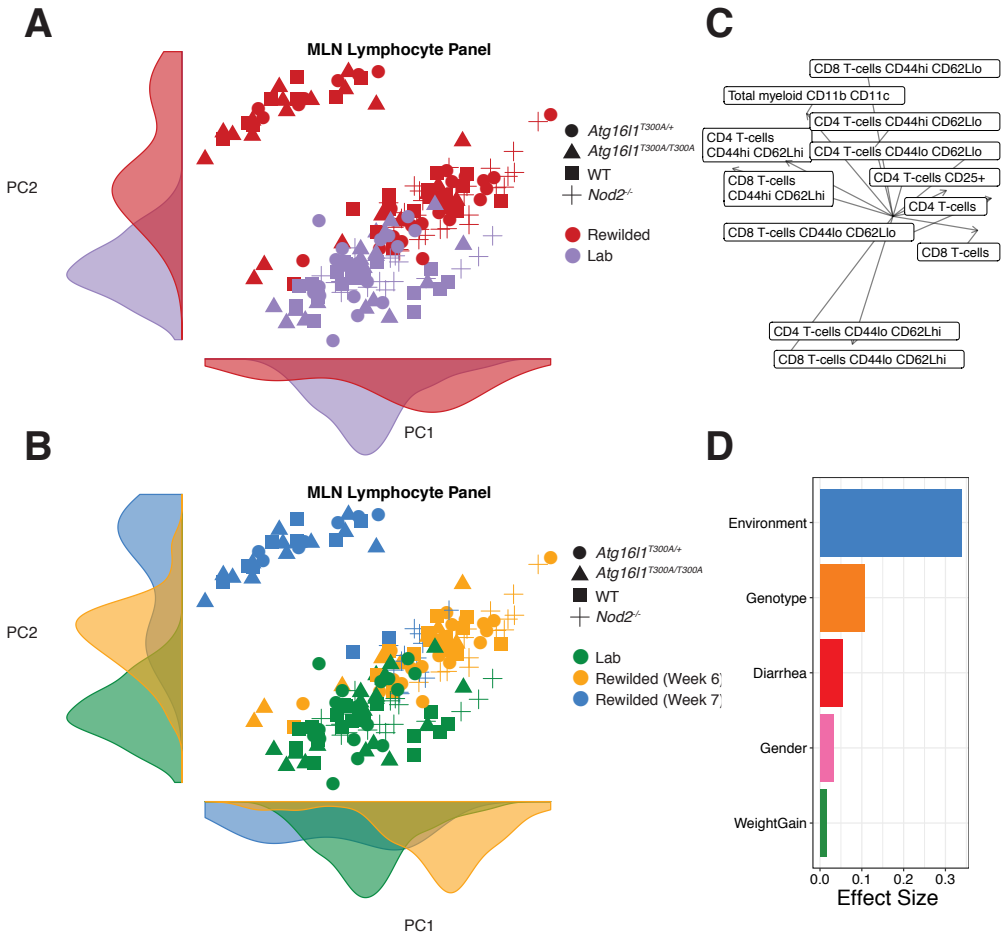
Supplementary Figure 5



Supplementary Figure 6

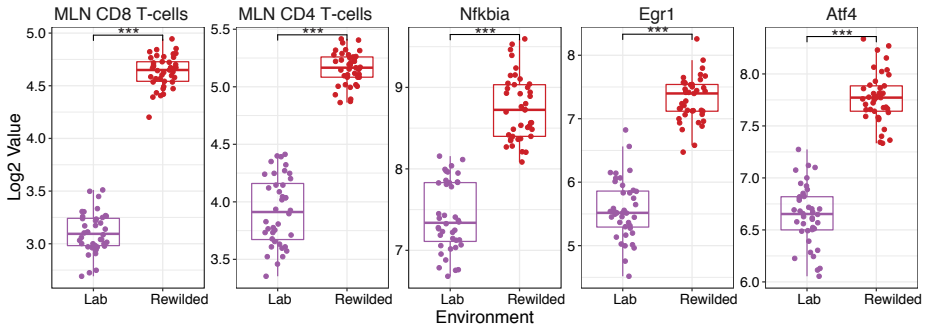


Supplementary Figure 7

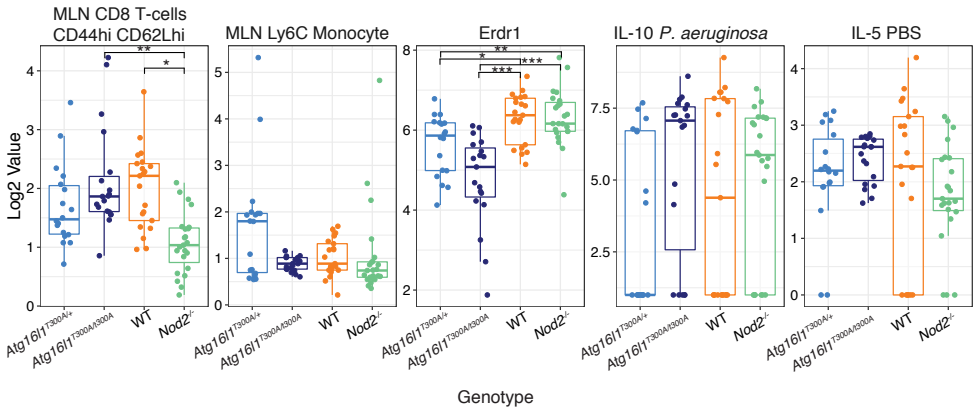


Supplementary Figure 8

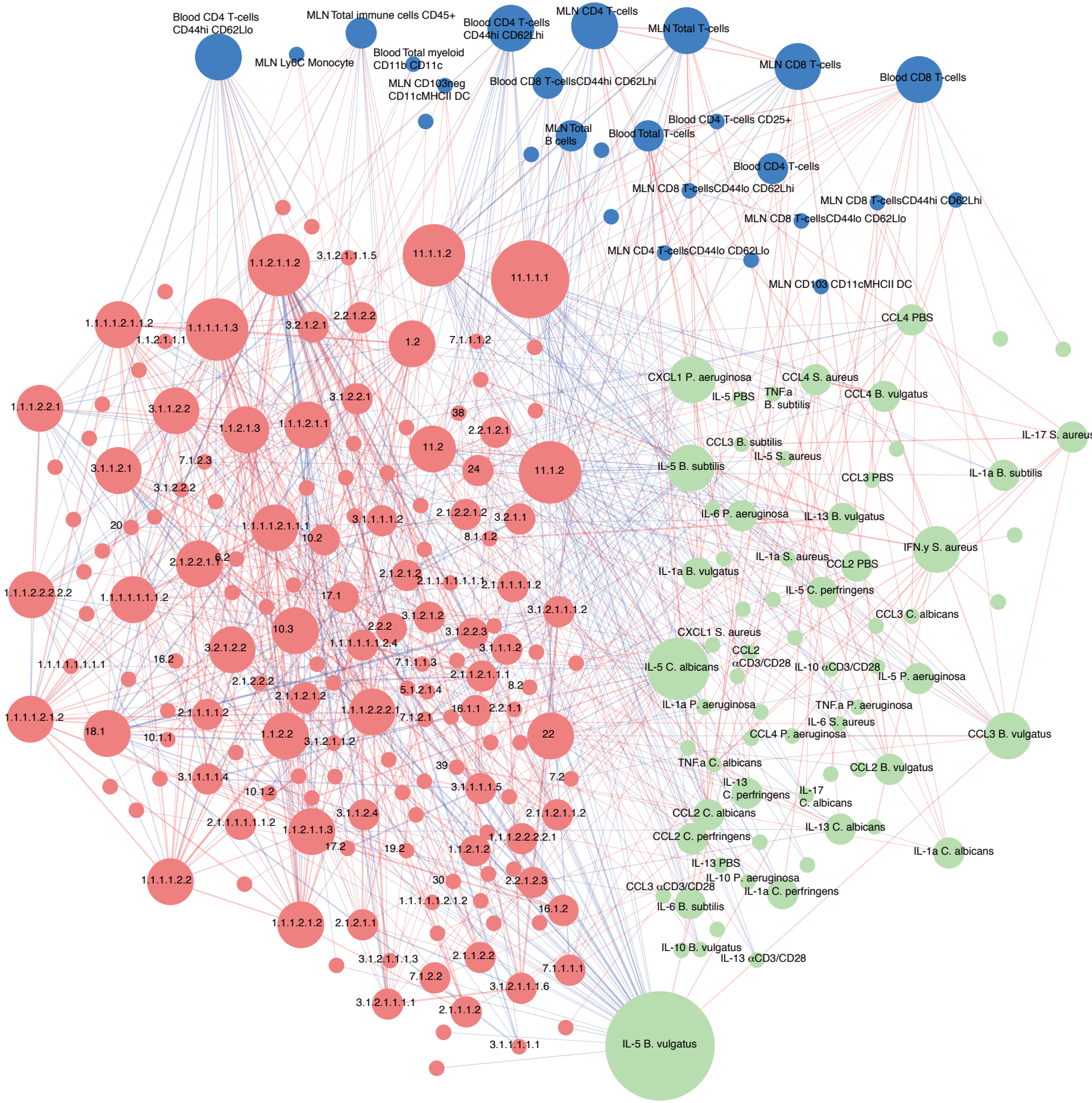
A



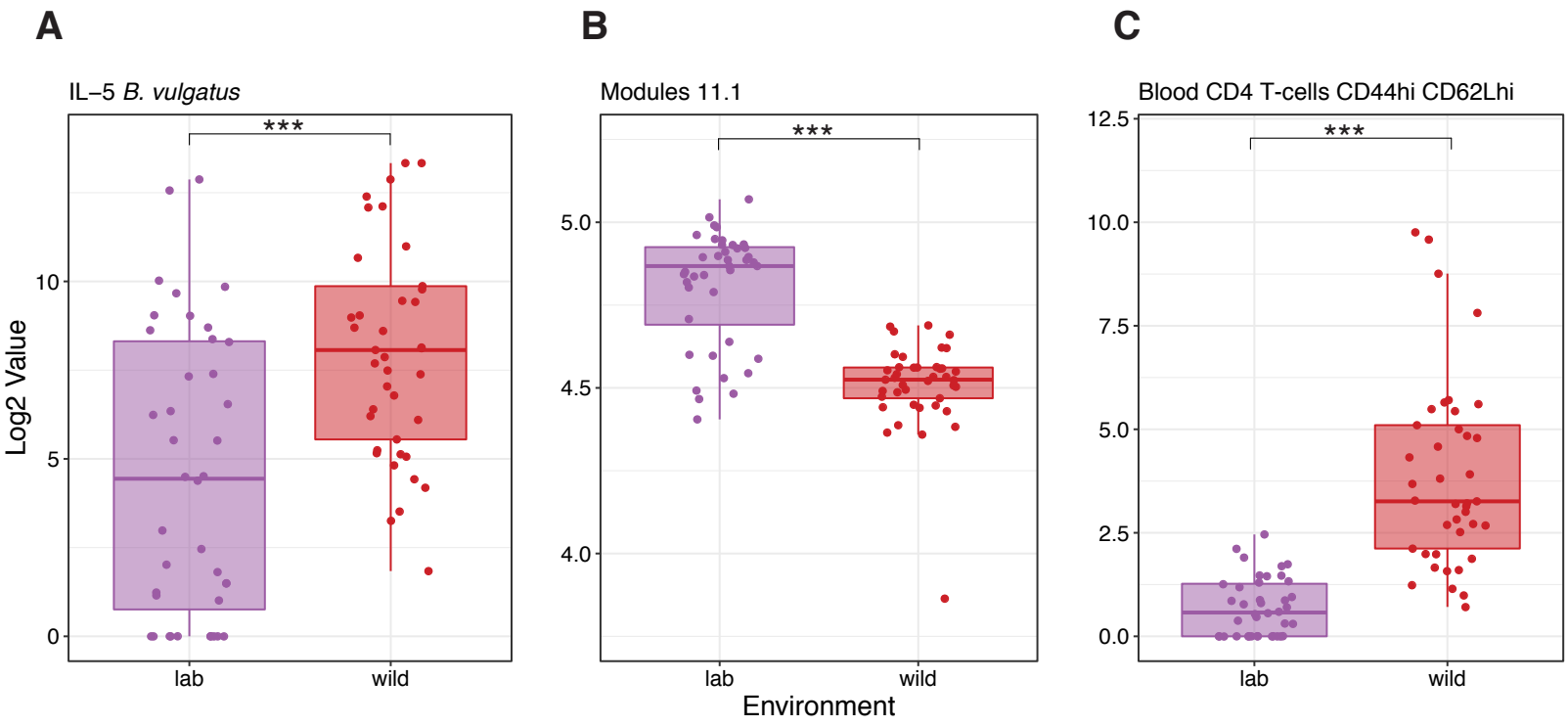
B

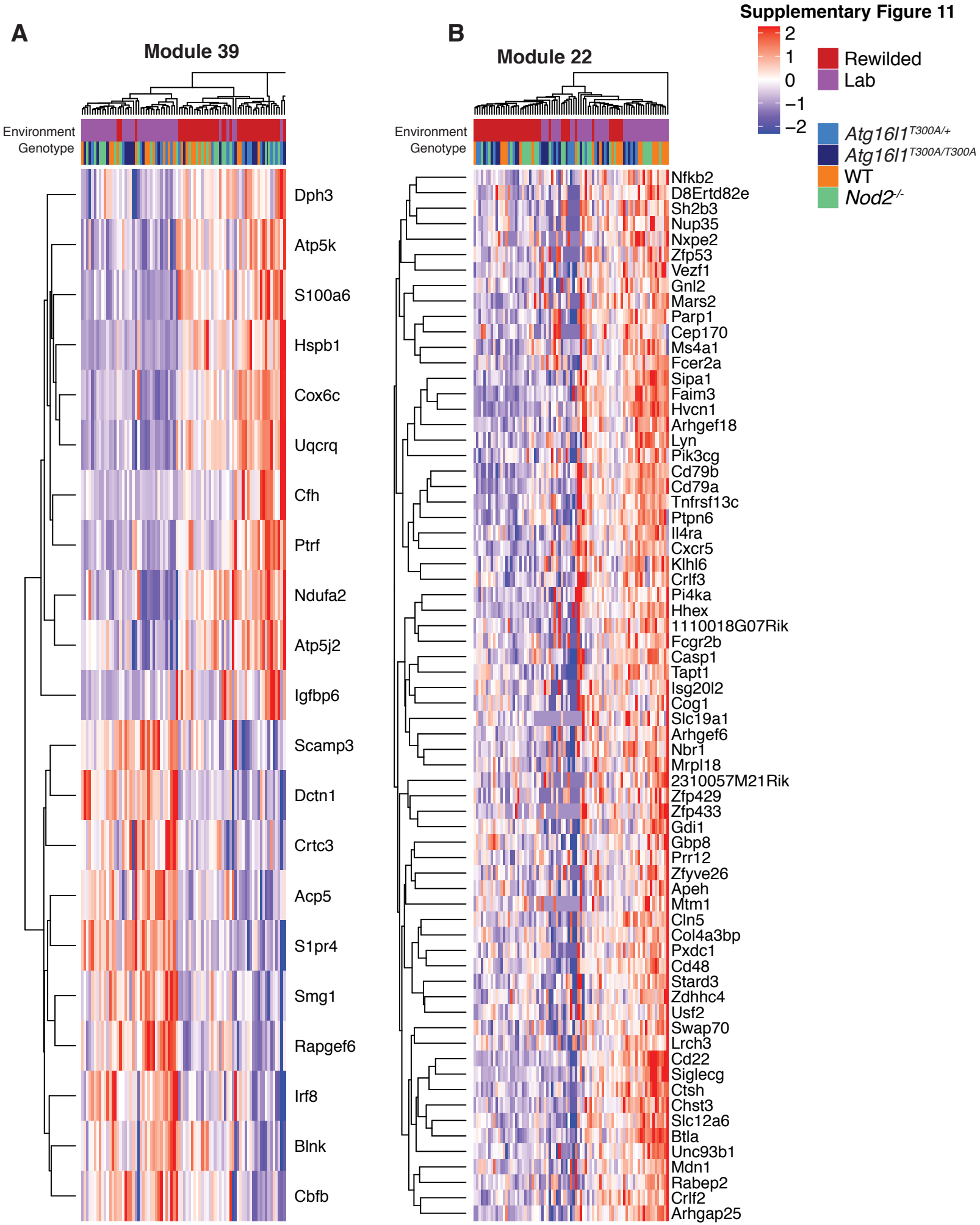
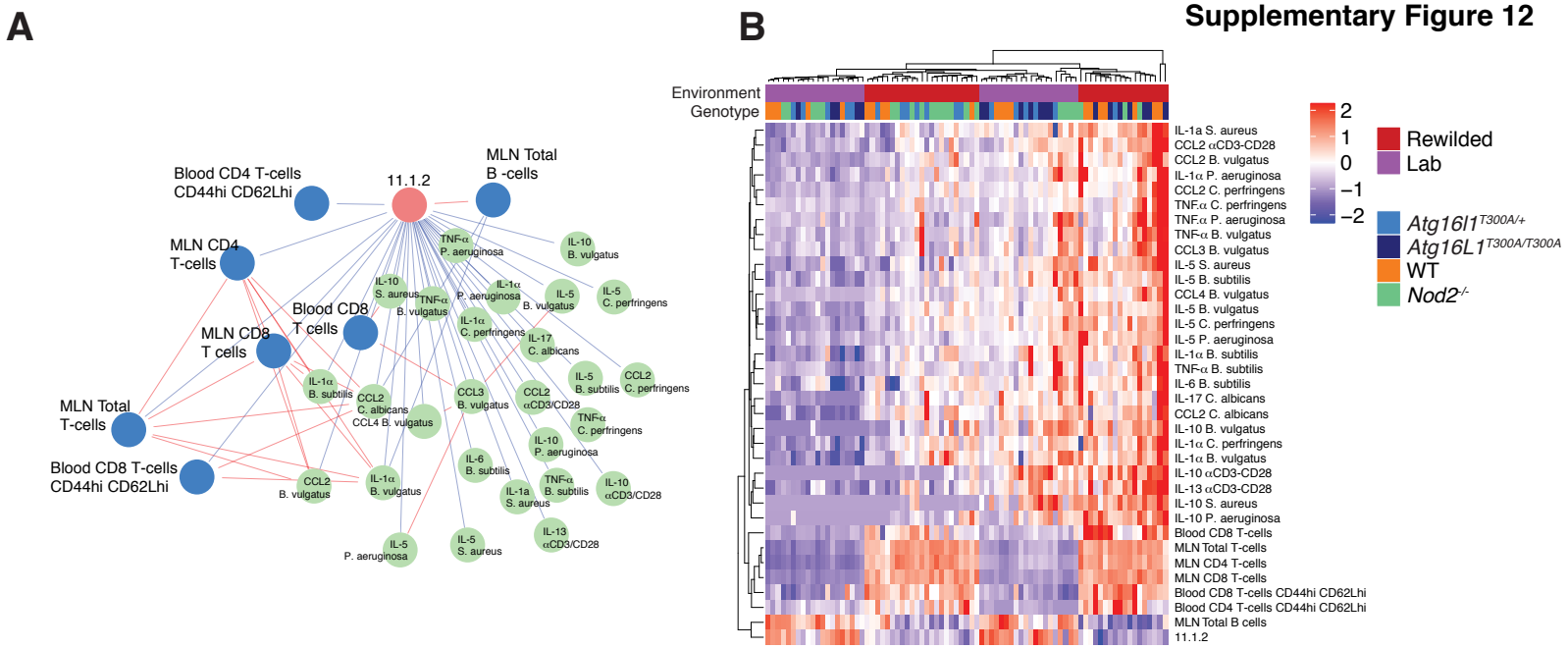


Supplementary Figure 9



Supplementary Figure 10




Supplementary Figure 11
A
Module 39
Environment
Genotype
Dph3
Atp5k
S100a6
Hspb1
Cox6c
Uqcrq
Cfh
Ptrf
Ndufa2
Atp5j2
Igfbp6
Scamp3
Dctn1
Crtc3
Acp5
S1pr4
Smg1
Rapgef6
Irf8
Blnk
Cbfb
B
Module 22
Environment
Genotype
Nfkb2
D8Ertd82e
Sh2b3
Nup35
Nxpe2
Zfp53
Vezf1
Gnl2
Mars2
Parp1
Cep170
Ms4a1
Fcer2a
Sipa1
Faim3
Hvcn1
Arhgef18
Lyn
Pik3cg
Cd79b
Cd79a
Tnfrsf13c
Ptpn6
Il4ra
Cxcr5
Klhl6
Crlf3
Pi4ka
Hhex
1110018G07Rik
Fcgr2b
Casp1
Tapt1
Isg20l2
Cog1
Slc19a1
Arhgef6
Nbr1
Mrpl18
2310057M21Rik
Zfp429
Zfp433
Gdi1
Gbp8
Prr12
Zfyve26
Apeh
Mtm1
Cln5
Col4a3bp
Pxdc1
Cd48
Stard3
Zdhhc4
Usf2
Swap70
Lrch3
Cd22
Siglecg
Ctsh
Chst3
Slc12a6
Btla
Unc93b1
Mdn1
Rabep2
Crlf2
Arhgap25
2
1
0
−1
−2
Rewilded
Lab
Atg16l1T300A/+
Atg16l1T300A/T300A
WT
Nod2-/-

Supplementary Figure 12