

ARTICLE

Noncanonical T cell responses are associated with protection from tuberculosis in mice and humans

Megan K. Proulx^{1*}, Christine D. Wiggins^{2*}, Charlotte J. Reames¹, Claire Wu², Michael C. Kiritsy¹, Ping Xu³, Judith C. Gallant³, Patricia S. Grace^{4,6}, Brooke A. Fenderson⁶, Clare M. Smith⁵, Cecilia S. Lindestam Arlehamn^{7,8}, Galit Alter⁶, Douglas A. Lauffenburger², and Christopher M. Sassetti¹

While control of *Mycobacterium tuberculosis* (*Mtb*) infection is generally understood to require Th1 cells and IFN γ , infection produces a spectrum of immunological and pathological phenotypes in diverse human populations. By characterizing *Mtb* infection in mouse strains that model the genetic heterogeneity of an outbred population, we identified strains that control *Mtb* comparably to a standard IFN γ -dependent mouse model but with substantially lower lung IFN γ levels. We report that these mice have a significantly altered CD4 T cell profile that specifically lacks the terminal effector Th1 subset and that this phenotype is detectable before infection. These mice still require T cells to control bacterial burden but are less dependent on IFN γ signaling. Instead, noncanonical immune features such as Th17-like CD4 and $\gamma\delta$ T cells correlate with low bacterial burden. We find the same Th17 transcriptional programs are associated with resistance to *Mtb* infection in humans, implicating specific non-Th1 T cell responses as a common feature of *Mtb* control across species.

Introduction

In a natural population, exposure to *Mycobacterium tuberculosis* (*Mtb*) produces a spectrum of outcomes ranging from asymptomatic control of the infection to progressive inflammatory disease (Lin and Flynn, 2018). Lack of control of mycobacterial infections is associated with Mendelian defects in genes necessary for the differentiation of T helper 1 (Th1) cells or signaling via the IFN γ receptor, indicating a critical role for this immune axis in protective immunity (Boisson-Dupuis et al., 2018; Bustamante et al., 2014). Despite the importance of this canonical anti-mycobacterial pathway, recent observations suggest that the immune response to *Mtb* may be more diverse than previously appreciated. T cells lacking IFN γ can provide protection in animal models (Gallegos et al., 2011; Sakai et al., 2016), and highly exposed people have been identified that remain healthy despite the absence of a detectable IFN γ response to the pathogen (Kroon et al., 2020; Simmons et al., 2018). These individuals, termed resisters (RSTR), produce *Mtb*-specific clonally expanded T cells with functional programs similar to Th17 and regulatory T cells (Sun et al., 2024). Together, these observations suggest that multiple paths to protective immunity may be present in a diverse population.

In contrast to this apparent heterogeneity in humans, the small animal models most commonly used for mechanistic studies of tuberculosis (TB) consist of genetically homogenous strains of mice (Smith and Sassetti, 2018) that depend on the canonical Th1 immune response for protection (Cooper et al., 1993; Flynn et al., 1993; Mogues et al., 2001). Mouse models that mirror the immunological diversity observed in nature are necessary to dissect the functional importance of alternative forms of immunity. The creation and implementation of the collaborative cross (CC) mouse population provides a new strategy to interrogate the role of genetic diversity in infectious diseases. This population is a collection of recombinant inbred mouse strains developed from eight genetically diverse founder lines (Saul et al., 2019; Threadgill and Churchill, 2012). Together these 60+ CC strains represent the nucleotide diversity of an outbred population, which produces heritable differences in immune phenotypes at baseline and upon infection with bacteria or viruses. Infection of the CC founders (Smith et al., 2016), the recombinant CC strains (Smith et al., 2022), or the related diversity outbred population (Kurtz et al., 2024; Niazi et al., 2015) with *Mtb* demonstrates the profound effect of genetic diversity on the course of TB disease in mice. These populations display a

¹Department of Microbiology, University of Massachusetts Chan Medical School, Worcester, MA, USA; ²Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; ³Transgenic Animal Modeling Core, University of Massachusetts Chan Medical School, Worcester, MA, USA; ⁴University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; ⁵Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA; ⁶Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Cambridge, MA, USA; ⁷Center for Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA, USA; ⁸Department of Infectious Disease Immunology, Center for Vaccine Research, Statens Serum Institut, Copenhagen, Denmark.

*M.K. Proulx and C.D. Wiggins contributed equally to this paper. Correspondence to Christopher M. Sassetti: christopher.sassetti@umassmed.edu; Douglas A. Lauffenburger: lauffen@mit.edu.

© 2025 Proulx et al. This article is distributed under the terms as described at <https://rupress.org/pages/terms102024/>.

wide range of susceptibility to *Mtb* infection, and profiling the immune environment in the reproducible CC strains has revealed qualitative differences in immune state. Notably, our previous work found no correlation between IFN γ and bacterial control in the CC population and identified several strains that restrict bacterial replication with very low levels of IFN γ detected at the pulmonary site of infection (Smith et al., 2016, 2022). This observation suggests that non-Th1 immune responses could be important for bacterial control in some members of this diverse population.

In this work, we investigate the importance of non-Th1 immune responses by identifying and characterizing groups of CC mouse strains that control *Mtb* replication to a similar degree but produce dramatically different amounts of IFN γ . We show that low IFN γ -producing strains have a persistent CD4 T cell profile that is distinct from the canonical C57BL/6 response and specifically lack terminal effector Th1 cells. Using within-species and cross-species mathematical modeling strategies, we identify Th17-like cells as a strong correlate of bacterial control in these mice and as a key feature shared between these animals and human RSTRs. A small animal model reflecting this shared biology allowed us to test the importance of T cell functions, revealing a reduced impact of IFN γ on bacterial control in this context. Our findings illustrate a new approach to incorporate human-relevant genetic and phenotypic diversity into model systems and demonstrate that non-Th1 immunity is an important feature of bacterial control in some individuals.

Results

Hierarchical clustering defines two groups of CC mice with divergent CD4 T cell phenotypes

Here, we sought to define groups of mice that control *Mtb* bacterial replication via distinct mechanisms. Multiple genetically diverse strains were included in each group to allow the differentiation of random variation from features directly related to the *Mtb* immune response. From our previous work (Smith et al., 2022), we chose 10 strains with CFU burdens at 4 wk after infection that are similar to C57BL/6 mice but differ in the amount of IFN γ detected in lung homogenates. We define all of these strains as being able to control *Mtb* replication, relative to more susceptible strains. Additionally, one susceptible strain (WSB/Eij) was included to assess the impact of bacterial burden on immune state. Three female mice from each strain were infected with H37Rv by low dose aerosol and sacrificed at 4 wk after infection. Lung and spleen bacterial burden were quantified by plating for CFUs, cytokine levels in whole lung were measured by Luminex assay, and major innate immune cell populations and CD4 T cell subsets were enumerated by flow cytometry. Hierarchical clustering using these immune features separated the CC strains into two distinct groups (Fig. 1 A). Eliminating outliers (CAST/Eij, CC046, WSB/Eij) produced a set of four strains per group. Lung CFU clustered with a variety of inflammatory cytokines but varied minimally between the groups ($6.62 \log_{10}$ versus $6.27 \log_{10}$, $P = 0.0442$, Fig. 1 B). One group, termed “noncanonical” (NCan), produced relatively lower levels of IFN γ in the lung homogenate than the

“canonical” group that contained the standard C57BL/6 TB model (Fig. 1 C). Additional features, particularly the transcription factor profile of the CD4 T cells, were also divergent (Fig. 1 D). NCan mice produced a lower proportion of CD4 T cells expressing the Th1-specific transcription factor T-bet, supporting the low IFN γ levels produced by these animals. Additionally, NCan mice had a relatively higher proportion of CD4 T cells expressing the transcription factor Foxp3, marking the regulatory T cell lineage, as well as GATA3 indicative of the Th2 lineage. Thus, the immunological difference between these groups of genetically diverse strains was strongly correlated with T cell phenotype.

NCan mice control *Mtb* replication with low Th1 cytokines and reduced CD4 T cell polyfunctionality

Having validated the categorization of the CC strains, we sought to expand our immune profiling with the goal of identifying the phenotype-defining features of NCan mice. We used the eight strains from the initial study, along with CC018, which was predicted to conform to the NCan phenotype (Smith et al., 2022). At 2 wk after *Mtb* infection, lung bacterial burden was comparable across male and female mice of all strains (Fig. 2 A) and in aggregate was not significantly different between the two groups (NCan $2.61 \log_{10}$ versus canonical $2.60 \log_{10}$, $P = 0.9938$). Lung CFU increased in both groups by 5 wk after infection but continued to show no statistically significant difference (NCan $5.56 \log_{10}$ versus canonical $5.14 \log_{10}$, $P = 0.2347$; Fig. 2 A). Spleen bacterial burden followed the same trajectories in both groups (Fig. 2 B). We again compared the expression of the four main CD4 Th effector lineage-defining transcription factors and confirmed increased T-bet⁺ CD4 in canonical mice and increased Foxp3⁺ and GATA3⁺ CD4 in NCan mice (Fig. S1 A).

Given the differences in Th effector lineages in the NCan mice, we sought to understand if secretion of cytokines from lung CD4 T cells differentiated these groups. To this end, we evaluated a panel of cytokines, including IL17a, IL4, IL10, IL2, TNF α , IFN γ , and the T cell activation and memory markers CD107a and CD40L/CD154, by intracellular cytokine staining (ICS). As predicted from whole lung IFN γ levels (Fig. 1 C), NCan mice have a significantly lower proportion of CD4 T cells producing IFN γ after ex vivo restimulation with the *Mtb*-specific peptide pool MTB300 (Fig. 2 C). This observation also holds for the proportion of total live cells staining for IFN γ (Fig. 2 D), indicating the decreased IFN γ production in CD4 T cells in NCan strains is not being compensated for by an alternative cell type. Additionally, while the frequency of IFN γ -producing CD4 T cells in NCan mice rose substantially when polyclonally stimulated with anti-CD3 and anti-CD28 (Fig. 2 E), it was still significantly lower than in canonical mice.

As Th1 effectors are multifunctional, typically secreting TNF α and IL2 in addition to IFN γ , we used combinatorial polyfunctionality analysis of antigen-specific T cell subsets (COM-PASS) to quantify the polyfunctionality of the CD4 T cells (Lin et al., 2015). In response to MTB300 restimulation, NCan CD4 T cells had a significantly lower polyfunctionality score as compared with canonical CD4 T cells ($P < 0.001$, Fig. 2 F). The same was true for CD4 T cells restimulated polyclonally ($P < 0.001$, Fig. S1 B). The most abundant polyfunctional populations

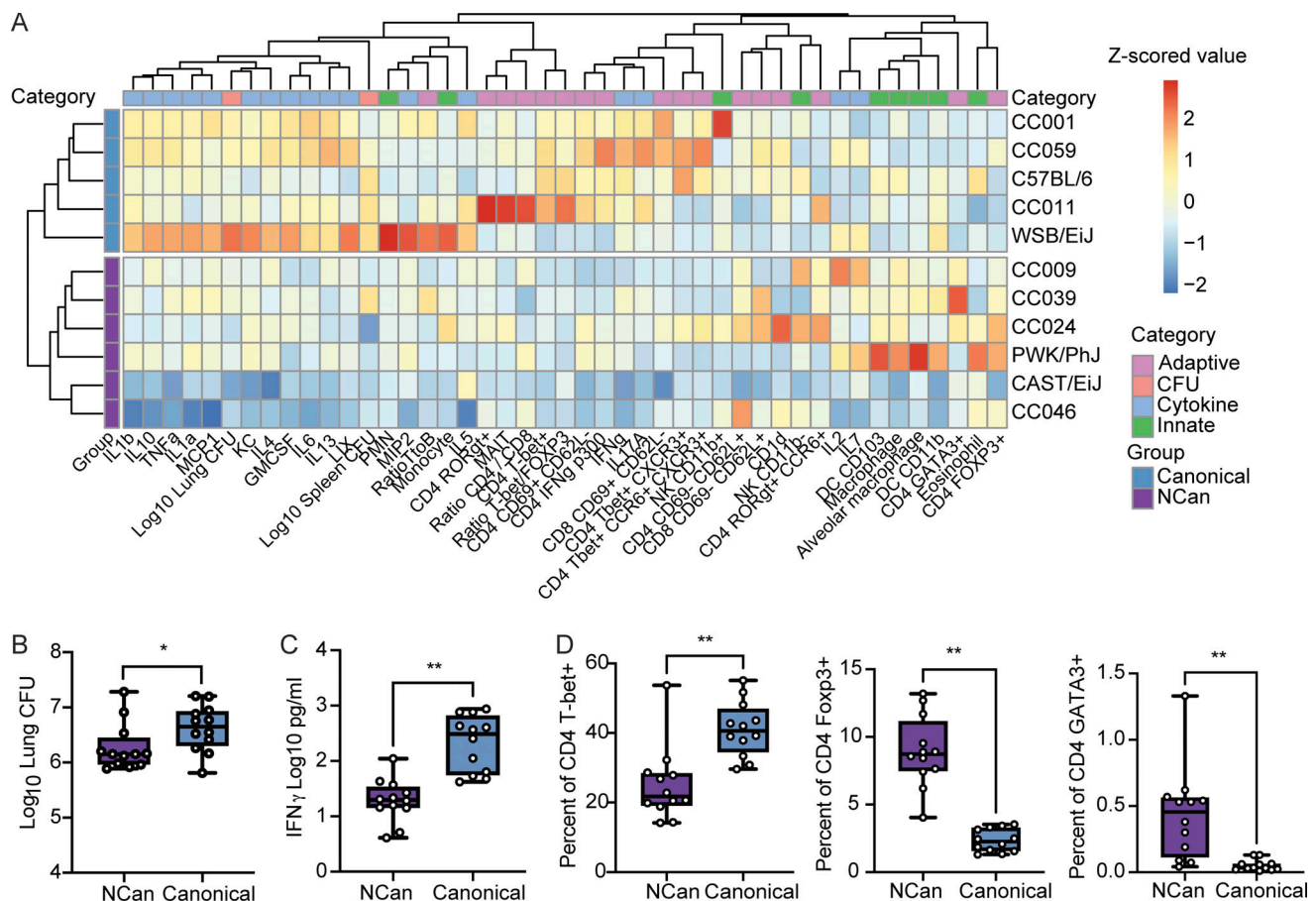


Figure 1. Hierarchical clustering defines two groups of genetically diverse mice with divergent CD4 T cell profiles. (A) Heatmap depicting immune features measured from the lungs of three female mice per strain from one experiment at 4 wk after infection. All features were scaled and centered with rows and columns hierarchically clustered via complete linkage method. **(B)** Lung bacterial burden ($\log_{10}$ transformed). **(C)** IFN γ levels from total lung homogenate as measured by multiplex Luminex assay. **(D)** Percent of lung CD4 T cells staining positive by flow cytometry for the transcription factors T-bet, Foxp3, or GATA3. **(B–D)** Box-and-whisker plots depict bounds from the 25th to 75th percentile, median line, and whiskers extending from min to max with all points shown (each point is an individual mouse). Statistical significance was assessed via Mann–Whitney U test with Benjamini–Hochberg correction (* $P < 0.05$, ** $P < 0.01$). NK, natural killer cell; DC, dendritic cell.

in both restimulation conditions were IFN γ positive and were significantly higher in proportion in canonical mice (Fig. 2 G and Fig. S1 C). However, the reduced polyfunctionality in NCan mice was not exclusively driven by cells secreting IFN γ , as the proportions of the five IFN γ -negative polyfunctional populations quantified here were also significantly reduced in NCan. These differences in cytokine production were also apparent when assessed individually, as a lower proportion of MTB300 or polyclonally stimulated NCan T cells produced TNF α , IL2, CD40L, or CD107a (Fig. S1, D and E). Thus, the decreased production of IFN γ in NCan strains extends to other cytokines, including the other Th1-associated cytokines TNF α and IL2, and suggests the NCan strains have reduced development or activation of CD4 Th1 effectors.

CD4 T cell functionality distinguishes NCan mice at multiple stages of infection as well as at baseline

The observed differences in T cell phenotype were identified at 4–5 wk after infection, during the peak of the adaptive response. To understand how these immunophenotypes developed, we

used multiparameter flow cytometry to profile the innate and adaptive immune response in the lung throughout infection and modeling methods that identified the most group-defining features at each time point.

To find immune features at 5 wk after infection that best discriminate between NCan and canonical mice, a partial least squares discriminant analysis (PLS-DA) model was built utilizing a least absolute shrinkage and selection operator (LASSO) feature selection algorithm. NCan and canonical mice are well separated in a reduced dimensional space (Fig. 3 A). While models built with selected features were better than models built with size-matched randomly selected features ($P < 0.01$), the predictive capability of the random model was high (Pearson $R = 0.85$, Fig. S2 A), indicating that the number of immune features that distinguish the two groups of mice was too large for meaningful selection. Partially, this is due to a high level of interrelatedness between measured features. To remedy this, immune features were organized into groups, which were then used to build a unique principal components analysis (PCA) model for each feature group. Gaussian mixture model

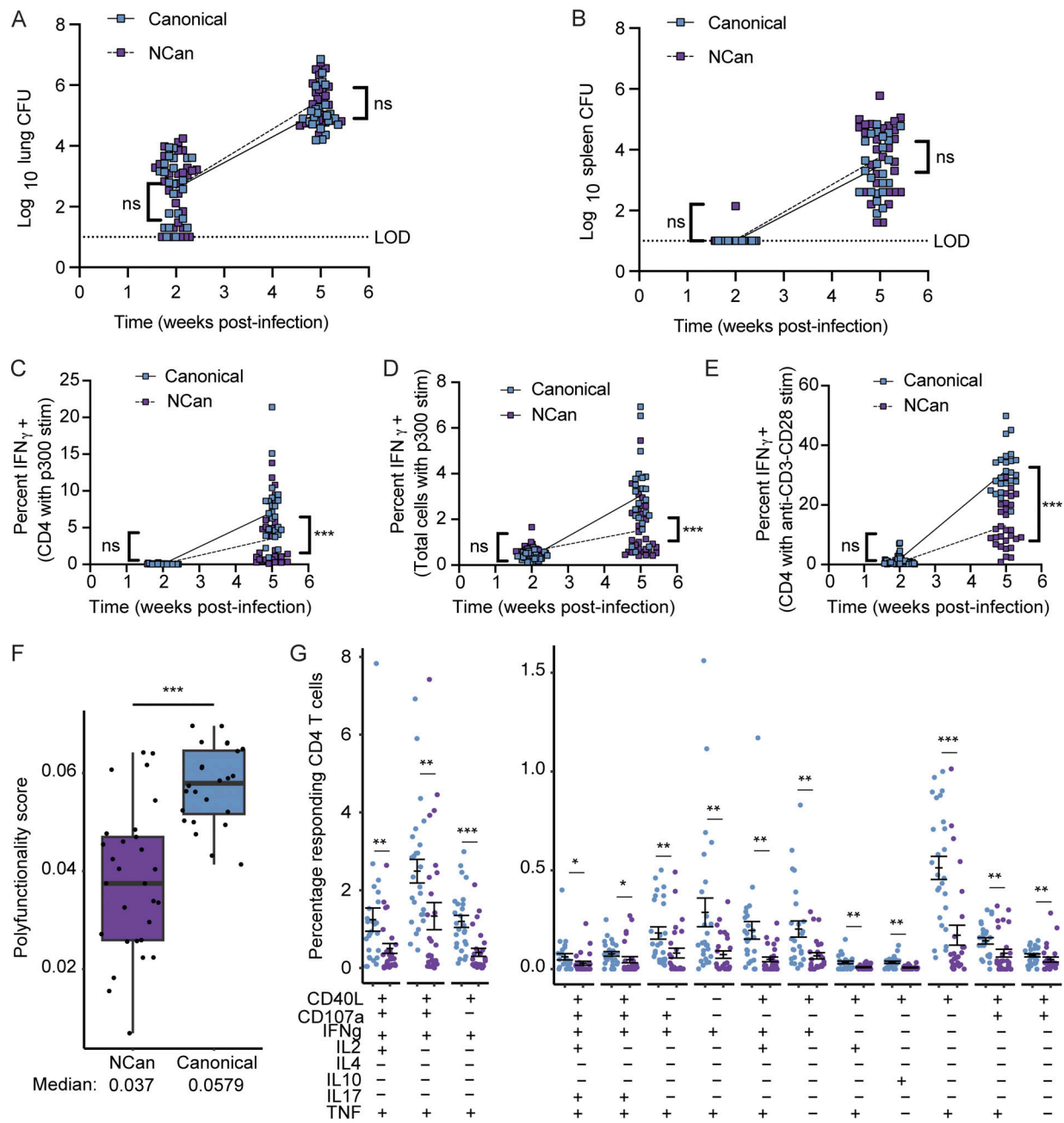


Figure 2. NCan mice control *Mtb* replication with low Th1 cytokines and reduced CD4 T cell polyfunctionality. (A–C and E) (A) Lung and (B) spleen bacterial burden (log₁₀ transformed) at 2 and 5 wk after infection. LOD, limit of detection. Percent of lung CD4 T cells staining for IFN γ by flow cytometry after stimulation with (C) MTB300 peptide pool or (E) polyclonally stimulated with anti-CD3 and anti-CD28 antibodies. (D) Percent of total lung cells staining for IFN γ by flow cytometry after stimulation with MTB300 peptide pool. (A–E) Each point depicts an individual mouse ($n = 2$ –4 mice per sex per strain per time point from one experiment). Canonical mice are represented by blue symbols with solid connecting line. NCan mice are indicated by purple symbols with dotted connecting line. Significance was determined by two-way ANOVA with Sidak multiple comparisons test. (F) Polyfunctionality score calculated by COMPASS for lung CD4 T cells after restimulation with MTB300 peptide pool. Box-and-whisker plots depict bounds from the 25th to 75th percentile, median line, and whiskers extending from min to max. Statistical significance assessed via Mann–Whitney U test. (G) Percentage of responding lung CD4 T cells after restimulation with MTB300 peptide pool. Data are split over two graphs to accommodate visualization of the lowly expressed populations. P values are derived from Mann–Whitney U test with Benjamini–Hochberg correction (ns = not significant [$P > 0.05$], * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Mean $\pm$ SE are visualized as crossbars.

clustering was performed in the reduced dimensionality PCA space to generate two clusters, and these clusters were compared with NCan and canonical groupings via the Adjusted Rand Index (ARI). This approach allows for the collapsing of

each cellular subset to its directions of maximum variation, thus reducing the impact of both interconnectedness and variable depth of feature measurement. This approach identified the CD4 T cell feature group (i.e., all features identified in CD4⁺

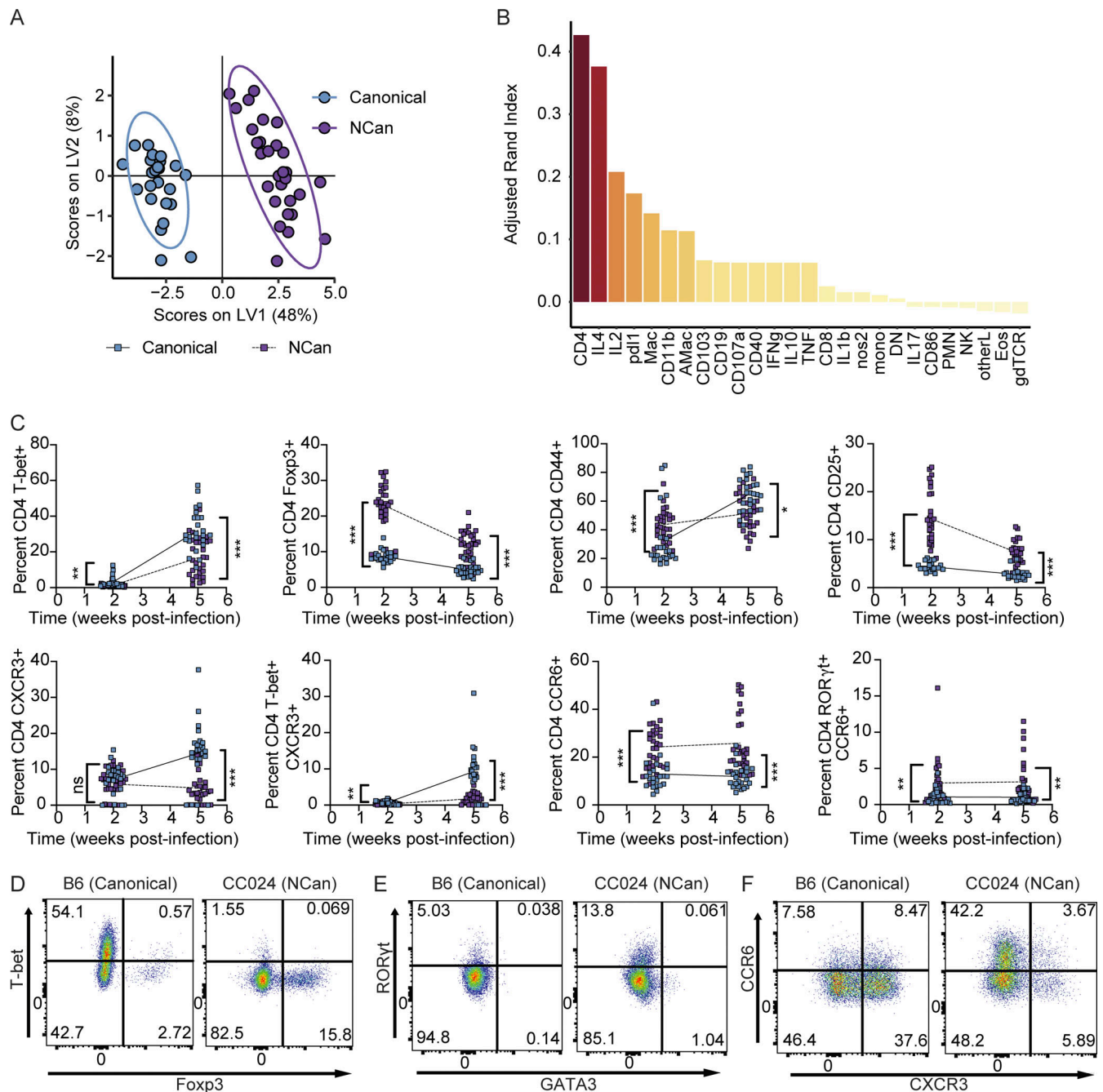


Figure 3. CD4 T cell functionality distinguishes NCan mice at multiple stages during infection. (A) Scores plot of selected lung features for PLS-DA model comparing canonical mice (blue) and NCan mice (purple) at 5 wk after infection (same mice from Fig. 2, $n = 2-4$ mice per sex per strain per time point from one experiment). Each dot represents an individual mouse, and the ellipse marks the 95% group confidence interval. (B) Bar plot depicting the ARI for indicated lung feature groups. Bars are colored by ARI value. (C) Scatter plots depicting indicated lung features measured by flow cytometry at 2 or 5 wk after infection. Canonical mice are represented by blue symbols with solid connecting line. NCan mice are indicated by purple symbols with dotted connecting line. Each point is an individual mouse, and statistical significance was assessed via Mann-Whitney U test with Benjamini-Hochberg correction (ns = not significant [$P > 0.05$], * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (D-F) Representative flow plots (log₁₀ fluorescence) of lung CD4 T cells staining for (D) T-bet versus Foxp3, (E) RORγt versus GATA3, and (F) CCR6 versus CXCR3 at 5 wk after infection. In each panel, the left plot is a representative canonical strain (B6) and the right plot is a representative NCan strain (CC024). Numbers in each quadrant indicate the percent of CD4 T cells staining positive for the respective marker. NK, natural killer cell; DN, double negative ($CD4^- CD8\alpha^-$); LV, latent variable.

cells) as the most distinguishing (Fig. 3 B). Within this CD4 T cell feature group, significant univariate differences included higher CD4 T cells expressing the Th1 transcription factor T-bet or the chemokine receptor CXCR3 in canonical strains (Fig. 3,

C-F). Expression of the activation marker CD44 was also higher in canonical mice. On the other hand, NCan strains had higher CD4 T cells expressing the regulatory T cell transcription factor Foxp3, the IL2 receptor alpha chain CD25, or the Th17-related

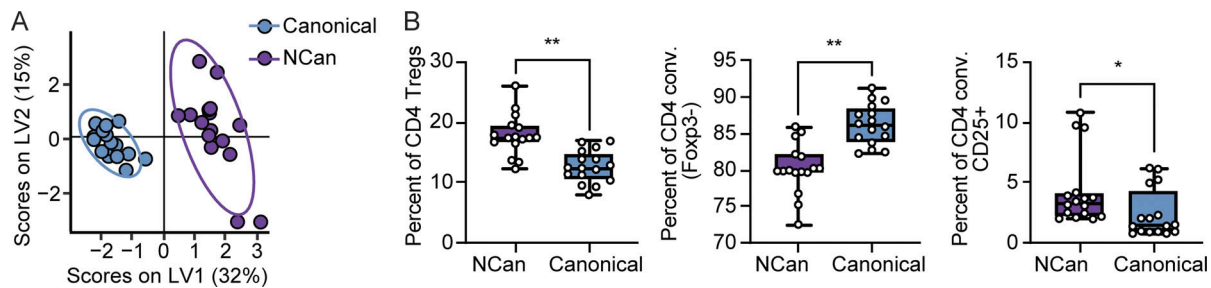


Figure 4. CD4 T cell functionality distinguishes NCan mice at baseline. (A) Scores plot of selected spleen features for PLS-DA model comparing canonical mice (blue) and NCan mice (purple) at baseline. Each dot represents an individual mouse, and the ellipse marks the 95% group confidence interval. Data are from [Graham et al. \(2024\)](#) with $n = 2$ mice per sex per strain. (B) Box-and-whisker plots depicting indicated features measured from the spleens of uninfected mice. Bounds extend from the 25th to 75th percentile, median line is marked, and whiskers extend from min to max with all points shown. Each point is an individual mouse, and statistical significance was assessed via Mann-Whitney U test with Benjamini-Hochberg correction (* $P < 0.05$, ** $P < 0.01$).

transcription factor ROR γ t and chemokine receptor CCR6. This contrast in CD4 T cell effector subsets, with canonical mice enriched for Th1-related features and NCan mice enriched for regulatory T cell and Th17-related features, complements the conclusions drawn from cytokine profiles.

2 wk after infection is an early stage in the priming of the adaptive immune response. At this time after infection, PLS-DA again illustrates that NCan and canonical mice can clearly be separated ([Fig. S2 B](#)). LASSO selection identified the minimal set of features necessary to separate NCan and canonical groups ([Fig. S2 C](#)) as well as those that are highly correlated with selected features ($|\text{Spearman } R| > 0.7$, [Fig. S2 D](#)). Even at this early stage in infection, many of the features identified in the 5 wk after infection analysis are also distinguishing groups at 2 wk. NCan mice have more regulatory T cells and more CD4 T cells expressing CD25 or CCR6 ([Fig. 3, C–F](#)). On the other hand, canonical mice have more CD4 Th1 effector cells, marked by expression of T-bet and CXCR3.

This early skewing in T cell profiles prompted us to ask if NCan and canonical mice could be distinguished in the uninfected state. A recent study ([Graham et al., 2024](#)) profiled the innate and adaptive immune populations in the spleens of 63 uninfected CC strains and illustrated extensive baseline variation across strains. Repeating our LASSO PLS-DA approach on the eight relevant strains indicated that NCan and canonical mice can also be separated at baseline ([Fig. 4 A](#)). Among the 20 features driving this separation are more conventional CD4 (CD4 Foxp3 $^{-}$) T cells in the canonical strains and more regulatory T cells in the NCan strains ([Fig. 4 B](#); and [Fig. S2, E and F](#)). NCan mice also have a higher proportion of CD25 $^{+}$ CD4 T cells, consistent with the infected mouse groups. Thus, the NCan CD4 T cell phenotype is detectable prior to infection and indicates a persistent, genetically determined alteration in immune tone that is maintained through the infection.

Single-cell RNA sequencing (scRNAseq) defines the early differentiation state and lack of Th1 terminal effector CD4 T cells in NCan mice

To more precisely define the differentiation state of NCan T cells and cell-cell communication pathways activated during infection, we generated a series of scRNAseq datasets. At 4 wk after

infection, lungs from two NCan strains (PWK/PhJ and CC024) and two canonical strains (C57BL/6 and CC011) were processed to capture gene expression (GEX) by scRNAseq. After filtering empty or dead cells, we analyzed a set of 58,689 high-quality cells. Unsupervised clustering identified major immune cell types ([Fig. S3, A and B](#)), which exhibited no significant differences in proportions between NCan and canonical strains ([Fig. S3 C](#)).

Given our flow cytometry findings identifying a distinct CD4 T cell profile in NCan mice, the CD4 T cell clusters (10,290 cells) were extracted and re-clustered. This identified 10 subsets of CD4 T cells, which were annotated using previously published criteria ([Fig. 5 A](#); and [Fig. S4, A and B](#)) ([Akter et al., 2022](#)). As expected, the proportion of regulatory T cells (expressing *Foxp3*, *Ctla4*, and *Irf2*) was higher in the NCan mice ([Fig. 5 B](#)). There were five clusters of CD4 T cells expressing IFN γ : CD4_act1a, CD4_act1b, CD4_act1c, CD4_act2, and CD4_IFN ([Fig. S4 C](#)). Again, consistent with our immune profiling, NCan mice had significantly lower proportions of three out of five of these IFN γ -producing subsets ([Fig. 5 B](#)). Two of these, CD4_act2 and CD4_act1a, were almost completely absent in NCan mice. CD4_act1a is a small cluster distinguished by expression of two heat shock proteins, *Hsp90a* and *Hsp90b*. The CD4_act2 cluster is defined by the expression of *Cx3cr1* and *Klrg1* and, in the canonical strains, is the highest producer of IFN γ ([Fig. S4 C](#)). This expression profile is consistent with terminal effector Th1 cells, which localize to the lung vasculature of C57BL/6 mice ([Sakai et al., 2014](#)). Given this result, we performed independent validation using intravascular CD45 staining combined with a panel of T cell markers in the same four strains. At 4 wk after infection, the proportion of this i.v. $^{+}$ KLRG1 $^{+}$ CX3CR1 $^{+}$ CD4 population was reduced by over 10-fold in the lungs of both NCan strains ([Fig. 5, C and D](#)).

We identified three clusters of naïve CD4 T cells marked by their expression of *Sell*, *Lef1*, *Tcf7*, and *Il7r*. In all three clusters, NCan mice had a higher proportion of cells as compared with canonical mice ([Fig. 5 B](#)). This, combined with the lack of terminal effector Th1 and lower proportions of other activated CD4 T cells, indicates that the NCan CD4 T cell profile is skewed to a less differentiated state. To understand the transcriptional programs associated with this phenotype, we analyzed the

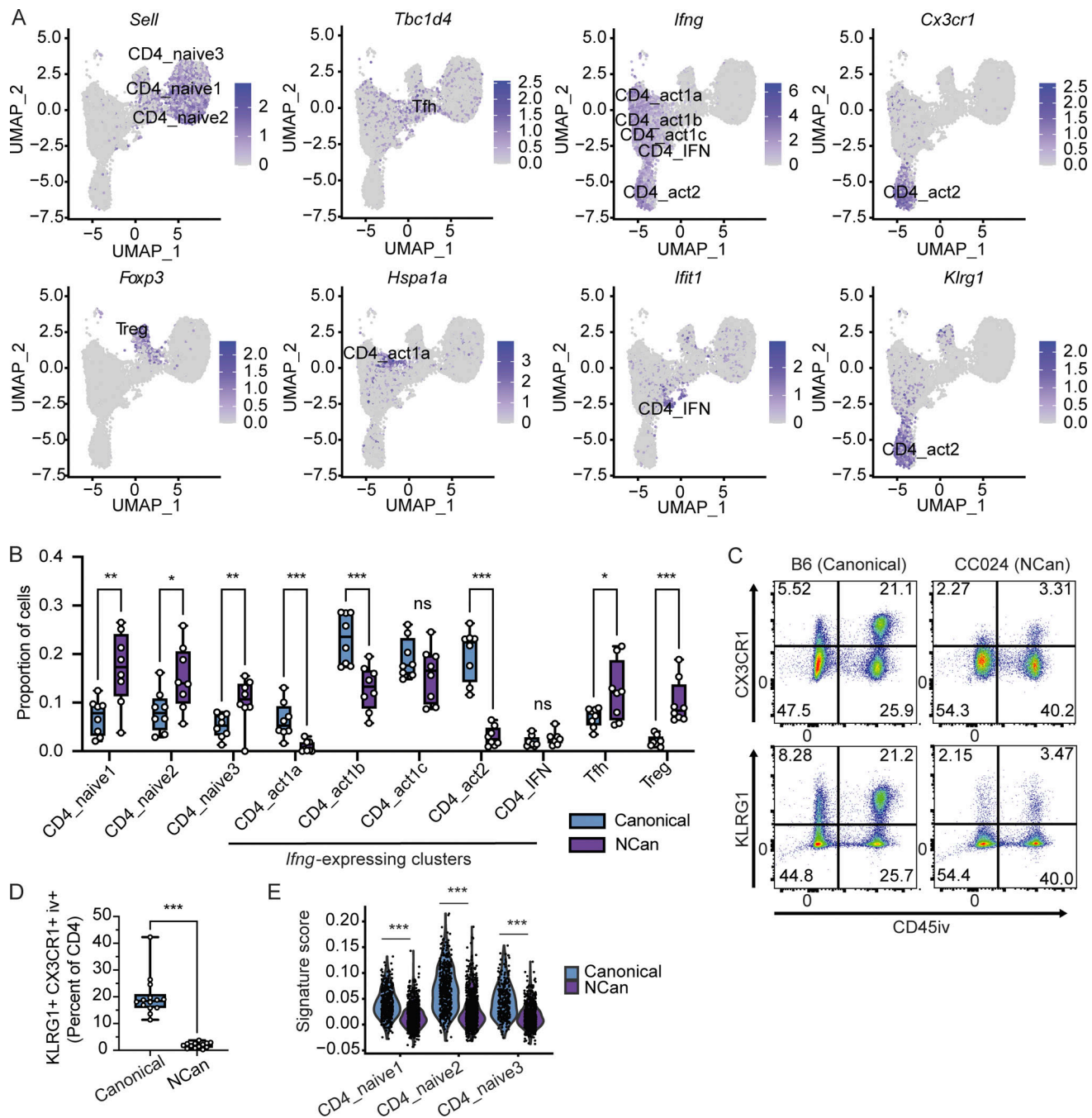


Figure 5. **scRNAseq defines the early differentiation state and lack of Th1 terminal effector CD4 T cells in NCan mice.** (A) UMAPs of the lung CD4 T cell subset from $n = 4$ male mice from two canonical strains (C57BL/6 and CC011) and two NCan strains (PWK/Phj and CC024) with GEX for cluster-defining markers from one experiment. (B) Proportion of lung CD4 T cells per cluster. Significance assessed by Mann-Whitney U test with Benjamini-Hochberg correction. (C and D) (C) Representative flow plots of KLRG1 (bottom) and CX3CR1 (top) staining in lung vasculature (CD45i.v.+) or lung parenchymal (CD45i.v.-) compartments with (D) quantification of percent of CD4+KLRG1+CX3CR1+CD45i.v.+ T cells. Data are pooled from two independent experiments ($n = 3-4$ mice per strain per experiment), with PWK/Phj and CC024 representing NCan mice and C57BL/6 and CC011 representing canonical mice. Significance assessed by Mann-Whitney U test. (E) Violin plots of the type II IFN signature score from Moreira-Teixeira et al. (2020) for the naïve lung CD4 T cell clusters. Significance assessed by Mann-Whitney U test with Benjamini-Hochberg correction (ns = not significant [$P > 0.05$], * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Treg, regulatory T cell; Tfh, T follicular helper cell.

differentially expressed genes (DEGs) between NCan and canonical mice in the three naïve CD4 T cell clusters (Fig. S4 D). Notably, many genes preferentially expressed in canonical mice were associated with type I or type II IFN signaling,

including *Ifngr1*, *Igtp*, *Irgm1*, *Iigp1*, *Irf1*, *Gpb2*, and *Stat1*. Supporting this, these clusters scored significantly higher for a type II IFN gene signature (Moreira-Teixeira et al., 2020) as compared with those in NCan mice (Fig. 5 E). The strong expression

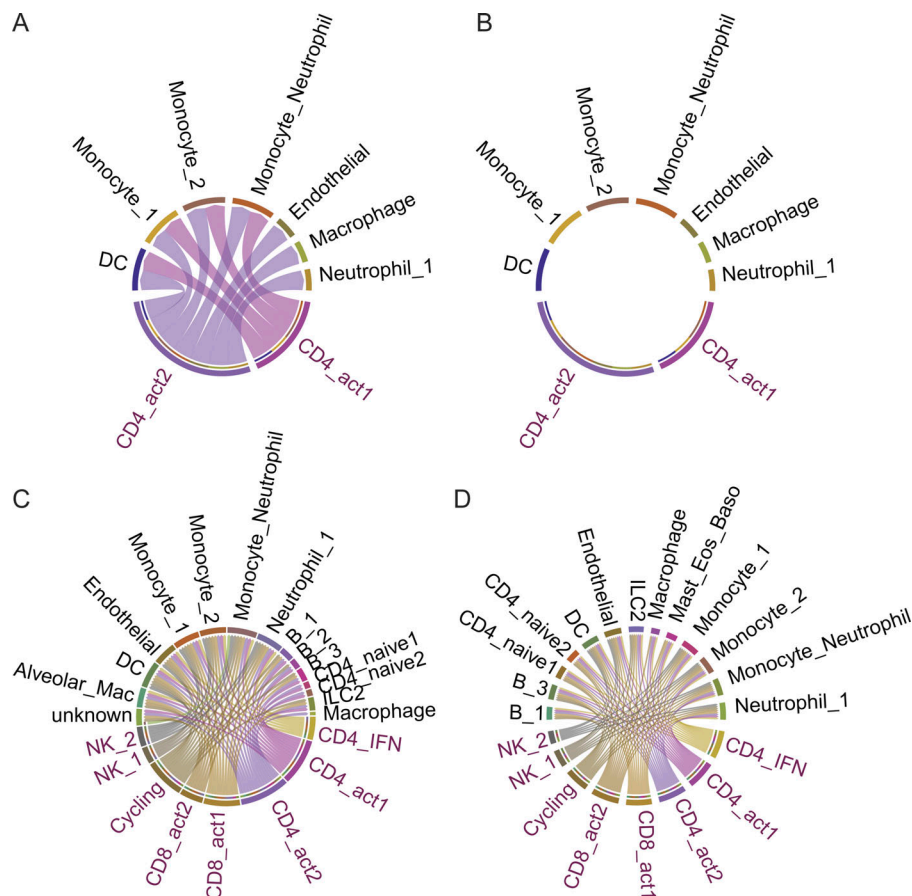


Figure 6. Cell-cell communication analysis reveals differential IFN γ cell signaling patterns between NCan and canonical mice at 5 wk after infection. (A and B) Chord diagrams represent LIANA consensus ligand–receptor signaling events for *Ifng* and *Ifngr1* or *Ifngr2* between lung cell types after filtering for consensus specificity rank (<0.05) and magnitude rank (<0.25) for (A) canonical mice or (B) NCan mice. Pink labels denote sender cell types. **(C and D)** Signaling relationships without magnitude rank filtering are represented as chord diagrams for (C) canonical mice or (D) NCan mice. All chord sizes are normalized to the sum of values, so link width represents relative value. NK, natural killer cell; DC, dendritic cell.

of the type II IFN signature was not limited to the naïve CD4 T cells, as every cluster in the lung uniform manifold approximation and projection (UMAP) scored significantly higher in the canonical mice (Fig. S4 E). The more robust presence of the type II IFN signature in the canonical mice is consistent with IFN γ levels and the Th1-skewed immune response we have described thus far.

Understanding the communication networks utilized by cells during *Mtb* infection can also highlight key pathways regulating the immune response. To this end, we used LIANA (Dimitrov et al., 2022) to explore the ligand–receptor communication pathways in the lungs of NCan and canonical mice, specifically focusing on the IFN γ signaling pathway. To look only at the most prominent IFN γ signaling relationships, interactions were filtered by specificity and magnitude. Canonical mice exhibit expected cross talk, with CD4_act1 and CD4_act2 cells acting as the strongest sender cells (Fig. 6 A). Both cell types are inferred to communicate with dendritic cell and monocyte subsets, while CD4_act2 is additionally inferred to signal to macrophages, neutrophils, and endothelial cells. However, no IFN γ signaling relationships are predicted to be amongst the strongest interactions in the NCan mice (Fig. 6 B). We next looked at all inferred IFN γ signaling relationships, filtering only for specificity and not magnitude (Fig. 6, C and D). Known IFN γ secretors such as natural killer cells and activated CD4 and CD8 T cells appear as senders for both canonical and NCan mice. This finding implies that, while ligands and receptors are present in enough

abundance in both mouse groups that signaling relationships are theoretically feasible, the canonical mice immunologically prioritize this signaling. Further, looking at IFN γ signaling to macrophages, NCan mice have inferred signaling coming from CD8_act2, CD4_act2, CD4_act1, and CD4_IFN, while canonical mice only show a relationship with CD4_act2. Thus, NCan mice have a relatively low magnitude of IFN γ signaling to macrophages, and it appears to derive from more cellular sources than canonical mice.

T cells are required for bacterial control in NCan mice, but IFN γ has a reduced impact

The availability of an experimentally tractable animal model of NCan immunity to *Mtb* allowed us to test the functional impact of T cell features. We first depleted CD4 and/or CD8 α T cells to confirm the importance of these cells in NCan mice. Animals from two NCan strains (PWK/PhJ and CC024) and one canonical strain (C57BL/6) were infected with *Mtb*, and cells were depleted starting at 2 wk after infection. At 5 wk after infection, we found that depletion was efficient (Fig. S4, F and G) and that simultaneous loss of CD4 and CD8 α cells increased bacterial burden to a similar degree in all three strains. Individual depletion of CD4 cells had a varying effect, as CD4 depletion alone increased CFU in C57BL/6 (Fig. 7 A) and CC024 (Fig. 7 B), but not PWK/PhJ (Fig. 7 C). Thus, while the relative dependence on CD4 and CD8 α varied by strain, all animals depended upon these cells for control of bacterial replication.

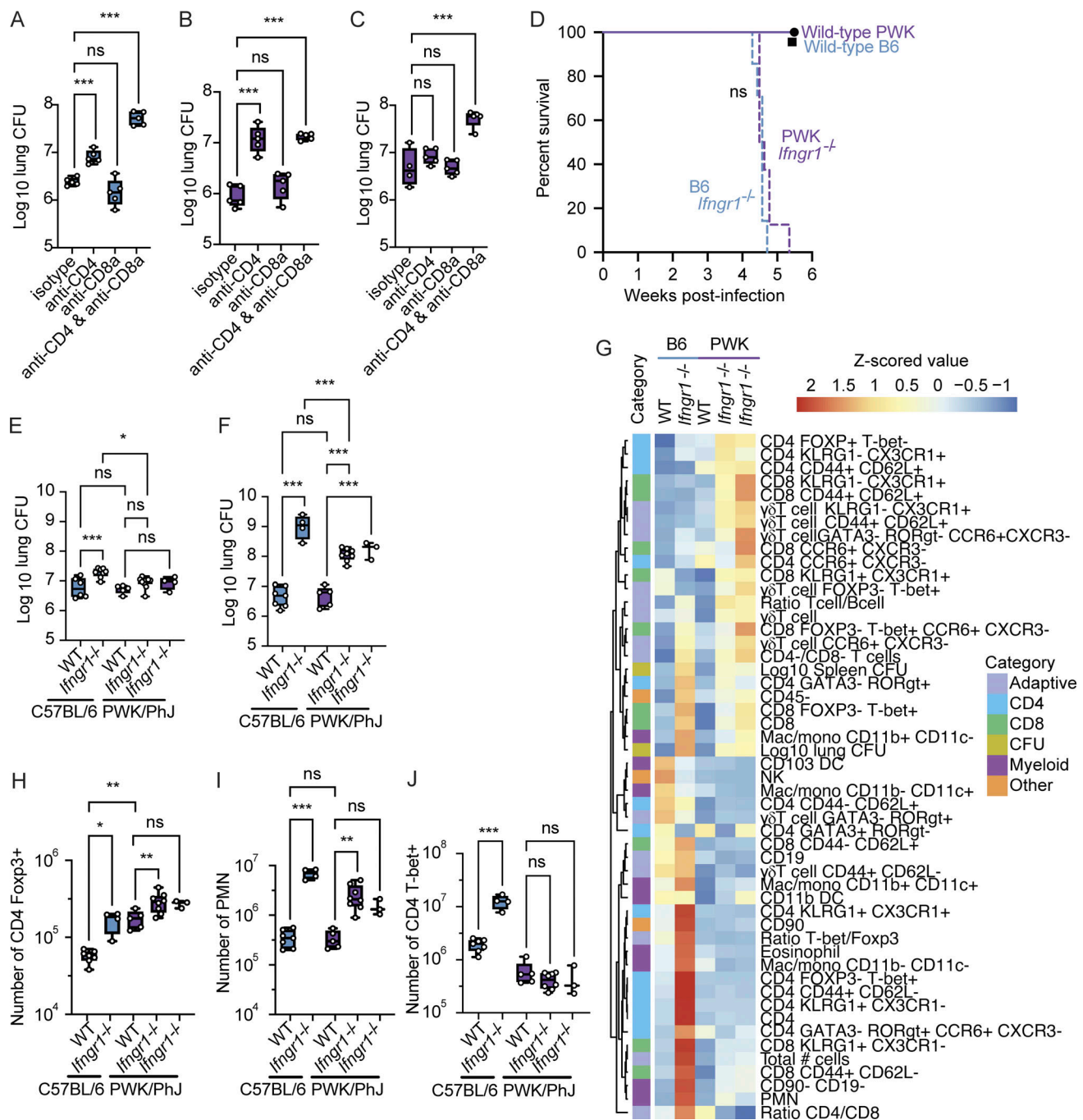


Figure 7. T cells are required for bacterial control in NCan mice, but IFN γ has a reduced impact. (A–C) Log₁₀ transformed lung CFU of (A) C57BL/6, (B) CC024, or (C) PWK/PhJ mice treated with isotype, CD4, or CD8 α -depleting antibodies ($n = 4$ –5 male mice per condition). **(D)** Kaplan–Meier survival curve with $n = 6$ –8 male mice per strain. Significance calculated with Mantel–Cox logrank test. **(E and F)** Log₁₀ transformed lung CFU at (E) 3 wk and (F) 4 wk after infection ($n = 2$ –4 mice per sex per strain per time point). **(G)** Heatmap depicting immune features measured from lung by flow cytometry with statistically significant changes at 4 wk after infection. Statistical significance assessed by Mann–Whitney U test with Benjamini–Hochberg (BH) correction (C57BL/6) or Kruskal–Wallis test with BH correction (PWK/PhJ). All features were scaled and centered with rows and columns hierarchically clustered via complete linkage method. **(H–J)** Number of (H) CD4 T cells staining for Foxp3, (I) neutrophils, and (J) CD4 T cells staining for T-bet as measured from the lung by flow cytometry at 4 wk after infection. **(E, F, and H–J)** Statistical significance assessed by one way ANOVA with Bonferroni correction (ns = not significant [$P > 0.05$], * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). As indicated in the methods, the two PWK/PhJ *Ifngr1*^{-/-} are independent strains, C57BL/6 wild-type mice were purchased from The Jackson Laboratories, and C57BL/6 *Ifngr1*^{-/-} and wild-type PWK/PhJ mice were purchased from The Jackson Laboratories and bred at UMass Chan. All box-and-whisker plots have bounds extend from the 25th to 75th percentile, median line is marked, and whiskers extend from min to max with all points shown. All data in this figure are representative of two replicate experiments. NK, natural killer cell; DC, dendritic cell; PMN, polymorphonuclear cell (neutrophil).

We next investigated the role of IFN γ signaling in the NCan immune response by generating an IFN γ receptor knockout in the PWK/PhJ background. Two independent mutant strains were generated via Cas9 editing of the *Ifngr1* gene, and each was backcrossed to wild-type PWK/PhJ mice for five generations. We confirmed that editing produced a nonfunctional receptor by treating bone marrow-derived macrophages with IFN γ and finding that induction of PD-L1 was abrogated in knockout cells (Fig. S4 H).

IFN γ protects from TB pathology both by restricting bacterial growth and by limiting inflammatory tissue damage in the lung and extrapulmonary sites (Cooper et al., 2002; Dalton et al., 2000; Desvignes and Ernst, 2009; MacMicking et al., 2003; Nandi and Behar, 2011). To compare the ultimate impact of IFN γ signaling on protection from disease, previously generated *Ifngr1*-deficient C57BL/6 and the new Cas9-edited *Ifngr1*^{-/-} PWK/PhJ mice were infected with *Mtb*, and their survival time was compared with each other and with wild-type mice (Fig. 7 D). The *Ifngr1*^{-/-} C57BL/6 mice succumbed rapidly, with a median survival time of 4.5 wk as expected (Nandi and Behar, 2011). PWK/PhJ *Ifngr1*^{-/-} also succumbed to infection with similar kinetics ($P = 0.5316$), indicating that the low level of IFN γ produced by these mice is still playing an important role.

To specifically assess the impact of IFN γ in the lungs of NCan mice, we measured bacterial burden and profiled immune features by flow cytometry. At 3 wk after infection, loss of IFN γ R1 signaling resulted in a significant increase in lung CFU of $\sim 0.5 \log_{10}$ in the C57BL/6 background but produced no discernible difference in the PWK/PhJ background (Fig. 7 E). This difference in IFN γ dependence was also apparent in the immune profile, as a PCA model of immune features separated wild-type and *Ifngr1*^{-/-} mice on the C57BL/6 background but not PWK/PhJ (Fig. S4 I). By 4 wk after infection, the difference in lung burden on the C57BL/6 background was substantial, over $2.5 \log_{10}$ (Fig. 7 F), compared with a more modest increase of $\sim 1 \log_{10}$ in PWK/PhJ. At this time point, PCA modeling of immune features distinguished wild-type and *Ifngr1*^{-/-} mice of both backgrounds, but separation was relatively reduced in PWK/PhJ (Fig. S4 J).

Immune features collected via flow cytometry at 4 wk after infection were compared to define the specific features influenced by IFN γ signaling. We found that the loss of *Ifngr1* in the C57BL/6 background resulted in a number of previously described effects (Nandi and Behar, 2011), including large increases in the number of neutrophils, total CD4 T cells, CD4 T cells expressing KLRG1, T-bet, or Foxp3 (Fig. 7 G). Of these, we only observed increases in Foxp3⁺ CD4 T cells and neutrophils in PWK/PhJ (Fig. 7, H and I), while other T cell features, such as the increase in CD4 T cells expressing T-bet, occurred exclusively in the C57BL/6 background (Fig. 7 J).

To understand if these differences in immune features impacted the architecture of the lung lesions, we examined H&E-stained sections from wild-type (Fig. 8, A and B) and *Ifngr1*^{-/-} (Fig. 8, C and D) mice at 4 wk after infection. Notably, the gross structure of the lung lesions differs between the two strains, with the C57BL/6 *Ifngr1*^{-/-} mice displaying lesions consisting of well-defined alveolitis that contained abundant polymorphonuclear cells and necrotic debris (Fig. 8 E). In contrast, PWK/PhJ

Ifngr1^{-/-} lungs display a more diffuse infiltration of immune cells. In regions that do appear more organized (Fig. 8 F), regions of debris are surrounded by macrophage-rich zones.

Thus, while IFN γ signaling remains important for control of TB disease in NCan mice, this cytokine's role in controlling bacterial growth and shaping the immune environment in the lung is delayed and diminished in the PWK/PhJ animals. These functional data support the computationally predicted decrease in IFN γ mediated cell-cell communication in NCan mice (Fig. 6).

Th17-like cells are associated with low bacterial burden in NCan mice

Given the reduced impact of IFN γ on immune signaling in NCan mice, we sought to identify T cell features associated with bacterial control. A partial least squares regression (PLS-R) approach was applied to the immune features previously measured at 5 wk after infection (Fig. 3), and LASSO feature selection was used to identify a reduced set of immune features capable of predicting lung CFU in NCan mice (Fig. 9 A). 10 rounds of five-fold cross-validation yielded a model with a Spearman R of 0.84 ($P = 1.038 \times 10^{-6}$) between observed lung CFU and model-predicted CFU (Fig. 9 B and Fig. S5 A). This model was significantly more predictive than models built with permuted labels or randomly selected features and was not dependent on the inclusion of any single mouse strain (Fig. S5 B). A similarly predictive model could not be built from immune features measured in canonical strains (mean Spearman R 0.60, mean $P = 0.029$) as it was dependent on the inclusion of a single strain, CC001 (Fig. S5 C). Similarly, the model built on NCan data could not be generalized to predict lung CFU in canonical strains (Spearman R 0.29) (Fig. 9 C). Collectively, this suggests cellular mechanisms of bacterial control that are specific to NCan mice. As LASSO selects a reduced feature set, we additionally generated a correlation network incorporating nonselected features that were highly correlated ($|\text{Spearman R}| > 0.8$, Benjamini-Hochberg corrected $P < 0.01$) with the selected features (Fig. 9 D). As expected, several inflammation-related features were associated with higher CFU, including neutrophils and IL1 β -producing monocytes/macrophages (nodes outlined in red). More importantly, 11 of the 14 features that were predictive of, or correlated with, low CFU in NCan strains (nodes outlined in blue) were associated with both CD4⁺ and $\gamma\delta$ T cells that express the transcription factor ROR γ t and the chemokine receptors CXCR3 and CCR6 (Fig. 9 D and Fig. S5 D). This implicates the Th17-like effector T cell programs as a key predictor of low lung bacterial burden in NCan mice.

Cross-species modeling identifies Th17 transcriptional programs in CD4 T cells as key translatable features

Several features that differentiate canonical and NCan mice are similar to those recently recognized as distinguishing the human RSTR phenotype. RSTR humans produce clonally expanded *Mtb*-specific T cells that, like NCan mice, display a unique functional and phenotypic T cell profile characterized by reduced Th1 and polyfunctional cells, and increased Th17, regulatory T cells, and CD25⁺ CD4 T cells relative to latent tuberculosis infection (LTBI) (Davies et al., 2023; Lu et al., 2019; Sun et al., 2024). Additionally, the T cells

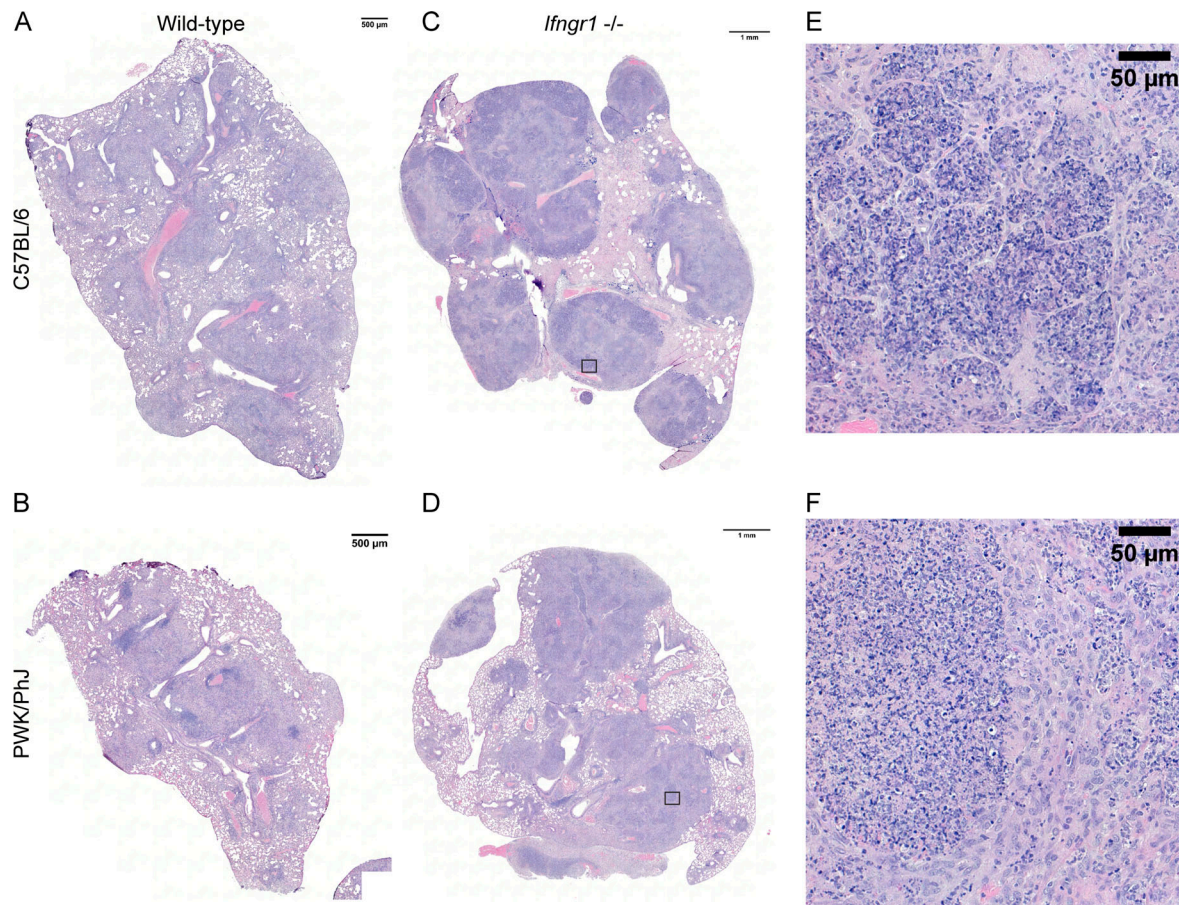


Figure 8. Loss of IFN γ R1 signaling differentially impacts lung lesion architecture at 4 wk after *Mtb* infection. (A–D) Representative lung sections stained with H&E from C57BL/6 wild-type (A) and *Ifngr1*^{−/−} (C) or PWK/PhJ wild-type (B) and *Ifngr1*^{−/−} (D) mice. **(E)** Inset of lesion architecture of C57BL/6 *Ifngr1*^{−/−} from box on panel C. **(F)** Inset of lesion architecture of PWK/PhJ *Ifngr1*^{−/−} from box on panel D. Images were taken with a TissueFAXS SL Q microscope (TissueGnostics) at 20 $\times$ magnification and region overviews, depicted in panels A–D, were exported as stitched images from TissueFACS SL viewer. Images are representative of two independent experiments of $n = 2$ –4 mice per sex per strain. Scale bars are 500 μ m (A and B), 1 mm (C and D), or 50 μ m (E and F).

of both species display an early differentiation phenotype. While these parallels between the T cells of NCan mice and human RSTRs suggested similar biology, these observations are somewhat biased by the a priori selection of immune features to compare.

To more quantitatively assess the relative importance of shared features, we compared datasets of single-cell transcriptional profiles using Translatable Components Regression (TransCompR), a mathematical framework that allows for the integrated modeling of multispecies data and the discovery of directly translatable biological pathways. While this approach has been previously used for bulk RNAseq and proteomic data (Brubaker et al., 2020; Lee et al., 2021), here TransCompR is adapted to scRNAseq for the first time.

First, a semi-supervised model was built to incorporate both human transcriptional variance and mouse phenotype. To begin, all DEGs between 522 *Mtb*-specific activated T cells from RSTR and LTBI individuals in the Sun et al. (2024). Ugandan household contacts study (Fig. 10 A) was used to build a PCA model (Fig. 10 B). The first 20 principal components (PCs) were retained from the human PC (hPC) space, capturing 25.4% of the variance present in the total human single-cell dataset and

representing human-relevant biological variation. Mouse CD4 T cells from the scRNAseq dataset generated here were then projected into the hPC space based on their expression of orthologous genes. This resulted in a model built on mouse data but with PCs defined by TB-specific human variance. Due to the nature of PCA, the direction of maximum variance in the human data is along hPC1, with decreasing variance captured by each hPC thereafter (Fig. 10 C). However, as features are expected to vary in their relative importance between species, this relationship is broken for the mouse data (Fig. 10 D), as evidenced by the lack of separation on PC1 (Fig. 10 E). Thus, any given PC could hold more importance for the mouse phenotype than the human phenotype, so we selected PCs with a Cohen's D value of 0.2 or higher, identifying eight PCs that significantly differentiated the human and/or mouse groups (Fig. 10 F).

To directly test the translational value of the model, we determined whether we could predict mouse phenotype based on the hPCs. T cells from the CD4_act1 clusters (Fig. 5 A) were used to match the activation state of the human antigen-specific cells. We used the PCs that differentiated mouse groups (PC6, PC13, PC18, and PC20) to build a linear discriminant analysis (LDA)

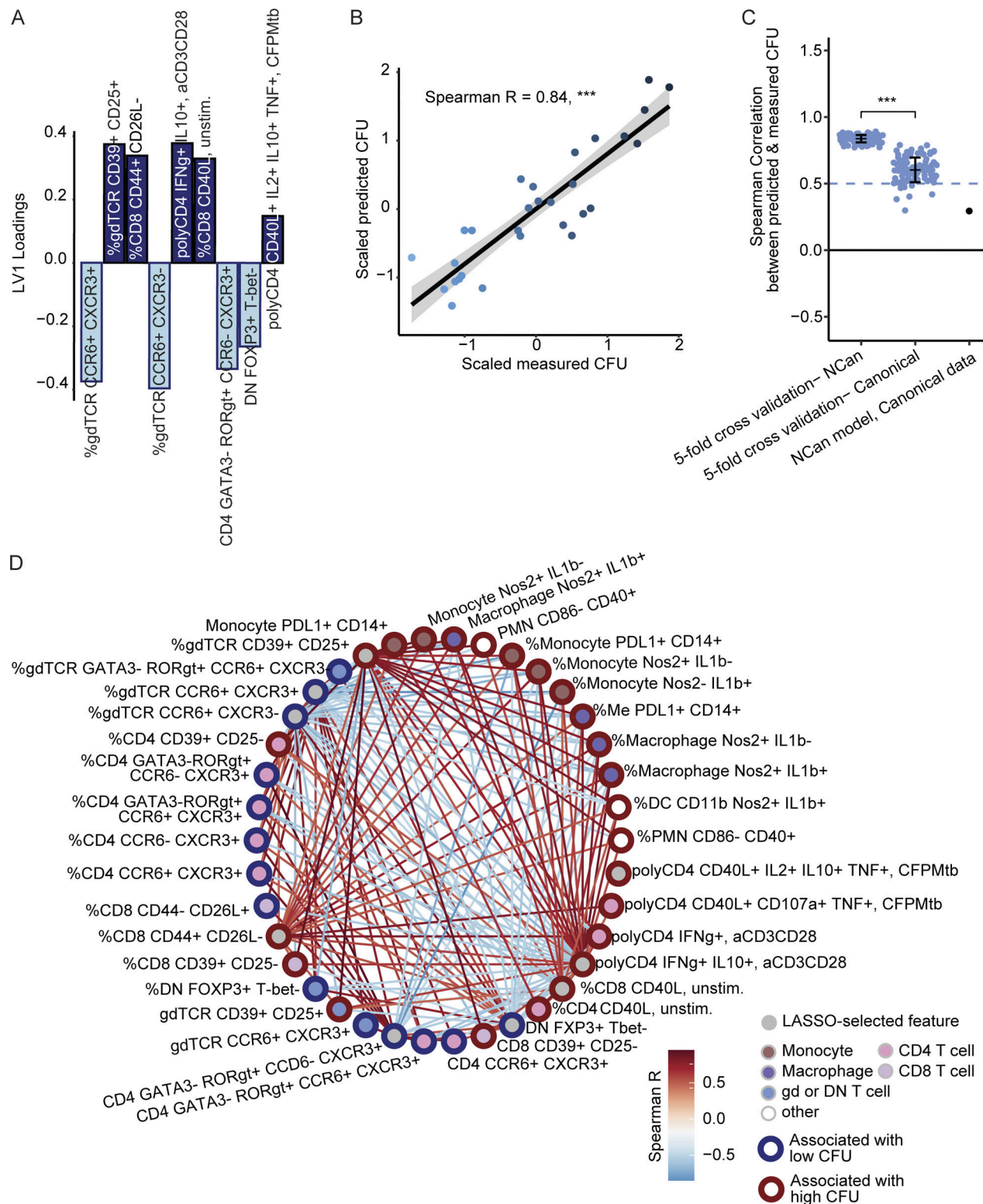


Figure 9. Th17-like subset of CD4 and $\gamma\delta$ T cells is associated with low bacterial burden in NCan mice. (A) LASSO-selected lung features plotted by their contribution to LV1, with features colored by whether they are higher in the low lung CFU samples (light blue) or the high lung CFU samples (dark blue). **(B)** Predicted lung CFU values from a PLS-R model using LASSO-selected features built on lung flow cytometry data (from Fig. 3) from NCan mice are plotted against measured lung CFU values, with dots colored by measured lung CFU value. The Spearman correlation is shown, with a 95% confidence interval in gray. **(C)** Spearman correlations from 100 rounds of fivefold cross-validation for predictive lung CFU models. Mann-Whitney U tests compare fivefold cross-validation distributions to the null distributions, with P values corrected via Benjamini-Hochberg ($***P < 0.001$). Mean ± 1 SD shown as crossbars. **(D)** Network plots show lung features highly correlated with the LASSO-selected features. Features with $|\text{correlation of coefficients}| > 0.8$ and Benjamini-Hochberg adjusted $P < 0.01$ are shown. Correlation strength is represented by line color. Node color indicates whether the feature is a LASSO-selected feature (gray) or the cell type measured by that feature, while the node outline indicates whether the feature is associated with low lung CFU (blue) or high lung CFU (red).

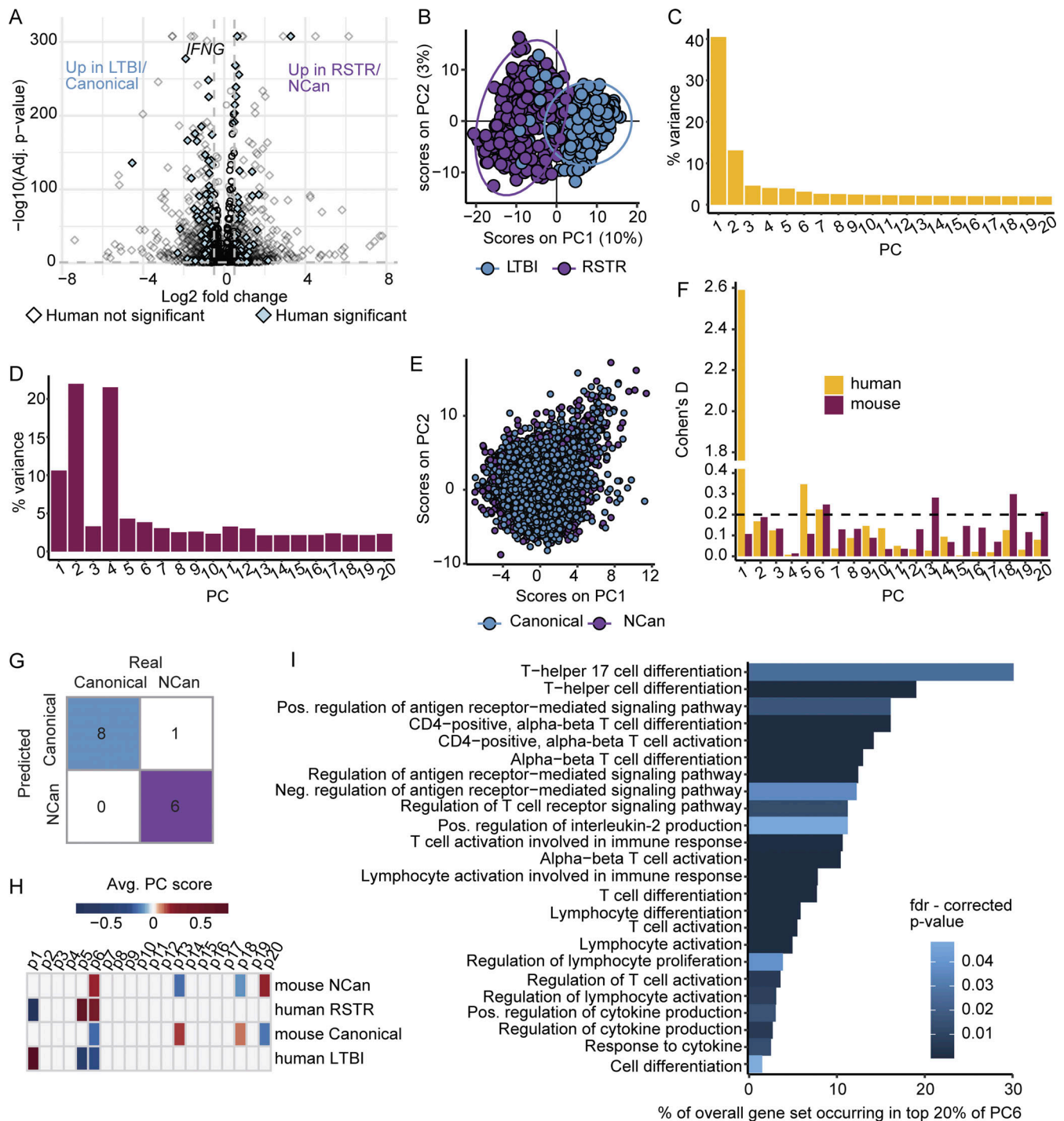


Figure 10. Cross-species modeling identifies Th17 transcriptional programs in CD4 T cells as key translatable features. (A) Volcano plot of DEGs between NCan and canonical mouse lung CD4 T cells, with genes also differentially expressed between human RSTR and LTBI CD4 T cells colored in blue. Shape denotes mouse significance (diamond = significant, circle = not significant). **(B–D)** PCA scores plot of human DEGs (RSTR = purple, LTBI = blue). % variance captured by each PC is shown for (C) human and (D) mouse PCAs. **(E)** Scores plot of mouse samples projected into human PCA space (NCan = purple, canonical = blue). **(F)** Cohen's D values for each species (human = gold, mouse = magenta). **(G)** Confusion matrix of model-predicted mouse phenotype based on LDA ensemble modeling. **(H)** Average score on each PC for each species and phenotype, colored according to inset legend. **(I)** T cell-related GO Biological Process terms are significant in NCan mice on PC6, colored by false discovery rate-corrected P value and plotted according to % of overall associated gene set per term occurring in the top 20% of PC6 loadings. fdr, false discovery rate.

model to predict the phenotype of each cell and assign each mouse to the phenotype predicted for most of its cells. This ensemble modeling approach was used because the original human PCA space was defined by a small number of antigen-specific T cells, whereas only a fraction of the large lung CD4_{act1} cluster from the CC mice are expected to be derived from antigen-specific cells that have initiated phenotypically distinct transcriptional programs. In a model-building scheme in which each mouse is left out once, while all other mice are used in a 100-round, fivefold cross-validation approach to build an LDA model, the class of the left out mouse is correctly predicted in 93% of trials, with all mice except CC24-21 correctly predicted in 100% of trials (Fig. 10 G). The average accuracy of classification of each cell for each mouse is always above 50%, except for CC24-21 (Fig. S5 E). The predictive model also compares favorably to null models built on shuffled labels ($P < 0.001$) and random features ($P < 0.001$) (Fig. S5 F). The high predictive accuracy of this model indicates that the biological information captured by this reduced set of PCs defined by human *Mtb*-specific variance has high translational value in the CC mouse model such that it is strongly phenotype defining.

Finally, the gene programs that comprise these highly predictive translational PCs can be interrogated to infer the underlying biology, as each PC represents a quantitative combination of multiple expression profiles that contribute to the prediction. PC6 was the only PC that was significantly associated with both human and mouse phenotypes (Fig. 10 H), indicating that it is comprised of cross-species translatable gene features that differentiate between the phenotypic states. Therefore, we tested the top and bottom 20% of the gene loadings of this PC for enrichment in Gene Ontology (GO) process terms using STRING (Table S1). PC6 exhibited a high number of T cell-related GO term enrichments (Fig. 10 I). As the resulting PC6 score for each sample is an aggregate of weighted gene values according to the PC loadings, the PC6 loadings can be considered a “signature” of multiple gene sets together. The gene set with the highest DEG representation at the top of PC6 was “Th-17 cell differentiation,” along with more general terms associated with T cell differentiation, regulation of antigen receptor-mediated signaling, and regulation of and response to cytokine production. At the same time, the bottom of PC6 was more defined by metabolism and cell death-related terms. NCan mice scored high on PC6 overall, indicating a relative emphasis of Th17 and T cell signaling compared with metabolic and cell death-related processes, while the low scores of canonical mice indicated the inverse. Overall, this analysis shows that, at the basic unit of control, the antigen-experienced T cell, individual cells exhibit transcriptional profiles that preview larger, organism-level trends in a predictive manner. Furthermore, these data demonstrate that features related to T cell differentiation and Th17-differentiation in particular are associated with bacterial control in NCan mice and are the quantitatively dominant trait shared with human RSTRs.

Discussion

Virtually all human diseases present as a spectrum of phenotypes. In recent years, these distinct presentations are

increasingly appreciated as “endotypes” that are linked to distinct biological mechanisms (Anderson, 2008). Understanding these mechanistic links often requires an experimentally tractable animal model that shares the relevant biological diversity, but it is not reasonable to expect a small animal model to recapitulate all aspects of a human disease. Our work describes a new approach to leverage the diverse and reproducible CC population along with mathematical modeling strategies to identify and characterize the key biological features that can be accurately translated between humans and mice sharing an endotype. To do this, we characterized a noncanonical immune state found in the diverse mouse population, used TransCompR to identify T cell differentiation programs shared between these mice and RSTR humans, and took advantage of the experimental tools available in the mouse system to begin to dissect the corresponding mechanisms of protection.

The fundamental immunological difference we characterized in the CC mouse model was a genetically determined bias in T cell differentiation, with a Th1 response predominating in the canonical mice versus regulatory T cells and Th17-like cells in NCan mice. The lack of terminally differentiated Th1 cells that we discovered in the NCan mice during *Mtb* infection is not specific to the CC population, as a similar immunophenotype was recently described in an independent cohort of wild-derived inbred mouse strains (Raveslout-Chávez et al., 2023, Preprint). Both our work and previous studies suggest possible mechanisms to explain these differences in T cell differentiation. We found that regulatory T cells are relatively abundant in NCan mice before infection and ~25% of lung CD4 T cells express Foxp3 in NCan mice after 2 wk of infection. As regulatory T cells can influence the differentiation and activity of the Th17 subset (Cardona and Cardona, 2019; Vignali et al., 2008), these cells are an attractive candidate to orchestrate the observed differences between Th1 and Th17-like cells in CC mice. A paucity of terminally differentiated Th1 cells after *Mtb* infection has also been observed in C57BL/6 mice lacking PD-1 or IL27 signaling and in mice heterozygous for a T-bet deletion (Sakai et al., 2016; Sallin et al., 2017; Torrado et al., 2015), suggesting possible molecular pathways that contribute to the observed differences in Th1 abundance and differentiation.

The resulting immune response in NCan mice controls *Mtb* replication to an equivalent degree as C57BL/6 and other canonical strains. However, this protection in the NCan mice is less dependent on IFN γ than in canonical mice. Instead, the predominant correlate of bacterial control in these mice is ROR γ t and CCR6 expression in CD4 and $\gamma\delta$ T cells, implicating Th17-related T cell programs. There is increasing evidence that IFN γ -independent responses have a role to play in the outcome of an *Mtb* infection (Cowley and Elkins, 2003; Gallegos et al., 2011; Sakai et al., 2016) and IL17-secreting CD4 T cells are a common feature of protective states in multiple species and in different contexts. IL17-producing T cells have been associated with the control of a hypervirulent strain of *Mtb* in C57BL/6 mice (Gopal et al., 2014), with cynomolgus macaque granulomas that control *Mtb* replication (Gideon et al., 2015), and with a protective vaccine response in mice (Gopal et al., 2012).

Additionally, IL17 production or ROR γ t expression in CD4 T cells correlates with vaccine-induced protection from *Mtb* infection in several CC strains (Lai et al., 2024). In this work, we report that Th17 programs are the key T cell feature that is translatable between the CC mouse model and human RSTRs. Whether these programs are indicative of bona fide Th17 cells utilizing IL17 to exert their protective effect versus a Th17-like cell acting via other mechanisms remains to be determined.

The basis of the human RSTR phenotype remains elusive. How this immune response develops and how it contributes to *Mtb* immunity are unclear, but we believe observations from our mouse studies begin to shed light on the possibilities. Firstly, we show that certain features of the CD4 T cell phenotype in NCan mice are genetically determined and preexisting before infection. This is consistent with human genome-wide association studies (GWAS) in a Ugandan cohort that identified two regions of linkage to the RSTR phenotype (Stein et al., 2008) and was replicated in a subsequent study of HIV-infected individuals (Sobota et al., 2017). A separate study evaluating purified protein derivative reactivity as a quantitative trait in *Mtb*-exposed South Africans also identified a genetic locus that overlaps with the Ugandan RSTR cohort (Cobat et al., 2009). Secondly, the NCan and canonical immune responses are equally capable of restricting bacterial replication. While the arrest of T cell differentiation due to early clearance of antigen has been proposed as a model to explain the lack of IFN γ in RSTR humans (Lu et al., 2019), our mouse data indicate that it is possible to develop this immune signature in the presence of consistent antigen burden. Finally, while the RSTR phenotype has been associated with increased innate immune responses (Simmons et al., 2022), humoral responses (Li et al., 2017), and altered T cell phenotypes (Lu et al., 2019; Sun et al., 2024), our studies highlight the importance of T cell function. We found that the Th17-like T cell phenotype was the predominant translatable feature between human RSTR and NCan mice and that T cells are critical for *Mtb* immunity in the mouse model. While we do not claim the NCan mouse phenotype to be equivalent to the human RSTRs as a whole, we believe mechanistic dissection of these specific features in the mouse will facilitate more specific human studies to understand the basis of the related human endotypes.

Our work also sheds light on the relative importance of IFN γ functions. In addition to its well-characterized role in stimulating antimicrobial mediators in infected macrophages (Cooper et al., 2002; Dalton et al., 2000; MacMicking et al., 2003), this cytokine is also known to restrict pathological neutrophil recruitment via signaling in both myeloid cells and non-hematopoietic cells in the lung (Desvignes and Ernst, 2009; Mishra et al., 2013; Nandi and Behar, 2011). These previously described functions are all represented in our cell-cell communication analysis, which illustrates IFN γ -producing CD4 T cells communicating with monocytes, macrophages, neutrophils, and endothelial cells. We speculate that the lack of these immunoregulatory functions in NCan mice are responsible for the rapid death of *Ifngr1*^{-/-} PWK/PhJ mice, despite the relatively modest bacterial burden in their lungs. Therefore, while the NCan mice may rely less on IFN γ for antimicrobial control, our data clearly illustrate that IFN γ signaling is still required for survival and

that the *Ifngr1*^{-/-} PWK/PhJ mice present a unique background to study the influence of IFN γ on bacterial control versus tissue preservation.

TB endotypes are vastly more complex than the simple terms, such as LTBI, RSTR, and active TB, classically used to define disease states. Indeed, even within the RSTR endotype, there is heterogeneity (Davies et al., 2023). While NCan mice are clearly differentiated from canonical mice by a combination of immune features, there are still quantitative differences in IFN γ production within these groupings, among other features. Similarly, while T cells are important for bacterial control, our data suggest that CD4 T cells have a more dominant role in bacterial burden control in one NCan mouse strain, whereas CD4 and CD8 α cells are redundant in another background. This heterogeneity is an essential component of our computational strategies, allowing the identification of features specifically associated with an endotype of interest. It is also likely that seemingly distinct endotypes may share similar biology. For example, recent work described CC strains that are differentially protected by Bacillus Calmette-Guerin vaccination (Lai et al., 2024; Smith et al., 2016), and this difference corresponds to the strain categorization in our study. Of the strains that overlap in this study, three canonical strains are protected by Bacillus Calmette-Guerin and three NCan strains are not, suggesting that the immunophenotypic differences we describe may be relevant to vaccine efficacy. Taken together, this work provides a new approach for understanding human endotypes by linking specific features with those present in the diverse and reproducible CC mouse population. Such an approach could be equally fruitful if expanded to the investigation of other complex disease states.

Materials and methods

Mice

Male and female C57BL/6 (#0664) mice were purchased from The Jackson Laboratory. Male and female collaborative cross mice (CC001/Unc, CC059/TauUnc, CC011/UncJ, CC009/UncJ, CC039/Unc, CC024/GeniUncJ, CC046/Unc, and CC018/UncJ) were purchased from the University of North Carolina Systems Genetics Core Facility (University of North Carolina, Chapel Hill, NC, USA). PWK/PhJ (#03715), CAST/EiJ (#0928), WSB/EiJ (#01145), and *Ifngr1*^{-/-} (#03288) mice were purchased from The Jackson Laboratory and bred at UMass Chan. All mice were housed in a specific pathogen-free facility in a 12-h light/dark cycle with access to food and water ad libitum.

An *Ifngr1*^{-/-} in PWK/PhJ was made by the UMass Chan Transgenic Animal Modeling Core using a CRISPR Cas9 single guide RNA (sgRNA) (5'-TATGTGGAGCATAACCGGAG-3', IDT #Mm.Cas9.IFNGR1.1.AL) targeting exon 5. Briefly, PWK/PhJ females were stimulated with Hyperova (0.15 ml i.p.) followed by hCG (7.5 IU i.p.) 48 h later. Eggs were collected and underwent in vitro fertilization the following morning with sperm isolated from male PWK/PhJ mice. Cas9 mRNA (TriLink Biotechnologies) plus *Ifngr1* sgRNA were delivered by pronuclear injection. Embryos were transferred to pseudopregnant Swiss Webster females. Pups were delivered naturally or by C-section, genotyped for *Ifngr1* status (forward primer 5'-GTCCTCGTATTTCACCGT

GAAG-3' and reverse primer 5'-GACCAAAACAGGCAAAGAAAAA C-3'), and analyzed for editing using TIDE (Brinkman et al., 2014). Two separate mice were selected and independently backcrossed to wild-type PWK/PhJ mice for five generations to establish two PWK/PhJ *Ifngr1*^{-/-} colonies. PWK/PhJ *Ifngr1*^{-/-} #8 has a 16-bp deletion at bp 551 that generates a premature stop after 204 aa. PWK/PhJ *Ifngr1*^{-/-} #18 has an 87-bp deletion plus insertion of an A at bp 477 that generates a premature stop at 174 aa. Wild-type *Ifngr1* is 477 aa.

Mouse infections

Mice were infected between 7 and 12 wk of age with virulent H37Rv. Briefly, H37Rv was grown in Middlebrook 7H9 medium supplemented with oleic acid-albumin-dextrose-catalase, 0.2% glycerol, and 0.05% Tween 80 to log phase with shaking at 37°C. To prepare for infection, the culture was washed, sonicated, and resuspended in PBS pH 7.4 containing 0.04% Tween 80. Approximately 50–150 CFU was delivered using an aerosol generating device (Glas-Col). To harvest organs for flow cytometry, enumeration of cytokines, and quantification of CFU, mice were humanely euthanized by isoflurane overdose followed by cervical dislocation. Whole blood was collected by cardiac puncture into EDTA-coated tubes. Lungs and spleens were aseptically harvested and individually homogenized in a FastPrep-24 bead beater (MPI Biomedical) with two pulses of 30 s at 4 M/s. The quantity of viable bacteria was enumerated by dilution plating onto Middlebrook 7H10 agar and counted after 21 days of incubation at 37°C. Multiplex cytokine analysis on lung homogenate was performed by Eve Technologies Corp. using Eve Technologies' Mouse High Sensitivity T Cell 18-Plex Discovery Assay (MDHSCT18).

Flow cytometry

Lung lobes were harvested into R10 media (RPMI 1640 plus L-glutamine [#11875093; Gibco], 10% FBS [#F4135; Sigma-Aldrich], HEPES [#15630130; Gibco], and penicillin-streptomycin [#15140122; Gibco]), treated with collagenase [#17104019; Gibco] and dissociated using a GentleMacs (Miltenyi). After incubation at 37°C for 30 min, tissues were subject to another dissociation and passed through a 70- μ m filter. RBCs were then lysed (eBioscience RBC Lysis buffer, #00-4300-54; Thermo Fisher Scientific), then cells were washed and passed through a 40- μ m filter. Cell suspensions were counted and arrayed into plates for staining. Cells were first treated with anti-CD16/CD32 (#BP0307; BioXcell) to block nonspecific antibody staining, followed by surface staining with respective antibody cocktails, including fixable Live/Dead near-infrared stain (LI0119; Life Technologies). For experiments with transcription factor staining, cells were treated with the eBioscience Foxp3/Transcription Factor Staining Buffer Set (#00-5523-00; Thermo Fisher Scientific) according to the manufacturer's instructions and stained with the respective antibody cocktail. For experiments with ICS, cells were first restimulated with the peptide pool MTB300 (Lindestam Arlehamn et al., 2016) at 1 μ g/ml for 5 h, anti-CD3 plus anti-CD28 (#100331 and #102112; Biolegend) at 1 μ g/ml for 5 h, culture filtrate protein (#NR-14825; BEI) at 1 μ g/ml plus *Mtb* lysate (#NR-14822; BEI) at 20 μ g/ml for ~18 h, or left unstimulated. All stimulation cocktails included brefeldin A

(#420601; Biolegend) and GolgiStop (#554724; BD) for no more than 5 h. After stimulation, cells were permeabilized with Cytofix/Cytoperm (#544714; BD Biosciences) before staining with antibody cocktail against respective cytokines. At the conclusion of all staining protocols, cells were fixed in 4% paraformaldehyde, washed in PBS, and stored at 4°C. All samples were run on an Aurora (Cytek) spectral flow cytometer and analyzed in FlowJo v10 (TreeStar). All antibodies used in this study are listed in Table S2. Further details regarding antibody panels and gating strategies are detailed in the data repository (<https://fairdomhub.org/studies/1320>).

T cell depletion

Depletion of T cells was achieved using neutralizing antibodies to CD4 (clone GK1.5, #BP0003-1; BioXcell), CD8 α (clone 2.43, #BP0061; BioXcell in C57BL/6 and CC024 or clone 53-6.7 #BP0004-1; BioXcell in PWK/PhJ), and a rat IgG2b isotype control (LTF-2, #BP0090; BioXcell). Antibodies were administered by subcutaneous injection of 200 μ g starting at 14 days after infection and continuing with administration of 100 μ g every 3–4 days until day 35. Depletion efficiency was verified using alternative antibody clones for CD4 (clone RM4-5, #100552; Biolegend) and CD8 β (clone YTS156.7, #126617; Biolegend).

Intravascular CD45 staining

At designated time points after infection, mice were i.v. injected with 2.4 μ g of anti-CD45.1-PE (for PWK/PhJ and CC024, #110708; Biolegend) or 2.4 μ g anti-CD45.2-PE (for C57BL/6 and CC011, #109808; Biolegend) or PBS. After 2 min, mice were euthanized, and lungs were processed as detailed above.

scRNAseq

C57BL/6, CC011, PWK/PhJ, and CC024 mice were infected as described. At 4 wk after infection, 4 mice per strain were humanely euthanized, and lungs were perfused with DMEM (#11965092; Gibco) containing 10% FBS. The left lobe of the lung was processed for CFU enumeration as described above. The right lobes were harvested into DMEM 10% FBS and processed to single-cell suspension as described above. Dead cells were removed using the Dead Cell Removal Kit (#130-090-101; Miltenyi Biotec) according to the manufacturer's instructions. Cell suspensions were counted, and 1.25×10^5 cells/sample were transferred to round-bottom polystyrene tubes, treated with TruStain FcX PLUS (#156603; Biolegend) to block nonspecific antibody binding, and stained with TotalSeq-A universal mouse antibody panel (pre-catalog version #99833; Biolegend) according to the manufacturer's instructions. Cells were counted, and 10,000 cells/sample were run through the 10x Genomics Chromium Controller using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 as indicated by the manufacturer. GEX and antibody-derived (ADT) libraries were constructed following the 10x Genomics User Guide #206 and Biolegend TotalSeq-A according to the manufacturer instructions. Libraries were pooled at a ratio of 80% GEX to 20% ADT and sequenced on a Nova-Seq6000 sequencer (Illumina) targeting 50,000 paired-end reads/cell. The raw FASTQ files can be accessed on SRA with BioProject accession number PRJNA1045547.

scRNAseq analysis

FASTQ files were uploaded to the 10x Genomics Cloud Analysis application for alignment, filtering, barcode counting, and unique molecular identifier (UMI) counting using Cell Ranger count pipeline (cellranger-6.1.2). C57BL/6 was aligned to the GRCm38 reference. Custom references and annotations for PWK/PhJ, CC011, and CC024 were built using cellranger mkref from pseudogenomes generated by g2gtools (v0.2.7). References for the pseudogenomes were the GRCm38 primary assembly, GENCODE vM25 annotations, and variant calls from [Srivastava et al. \(2017\)](#). The resulting feature-barcode matrices were imported to Seurat (v5.0.3) ([Hao et al., 2024](#)) using R (v4.2.2) and Rstudio (2022.12.0+353). This yielded a dataset averaging 5,138 cells/mouse with 53,543 reads/cell and 1,573 mRNA detected/cell (Fig. S3, D–F). Cells were filtered for high mitochondrial content (<10%), empty cells (nFeature_RNA > 200), and high feature counts (nFeature_RNA < 4000). Doublets were called by DoubletFinder v2 (v2.0.3) ([McGinnis et al., 2019](#)) and excluded. GEX data were normalized using SCTransform followed by dimensionality reduction for the top 30 PCs and clustering via UMAP (resolution = 0.5). The ADT tag library was normalized using the centered log-ratio method followed by centering and scaling. Datasets were then integrated and dimensionality reduction and clustering were rerun. Clusters were annotated using SingleR ([Aran et al., 2019](#)), existing scRNAseq datasets ([Akter et al., 2022](#)), and expression of canonical markers in the ADT dataset. The annotated Seurat object can be accessed via GEO with accession number GSE277712.

Module scoring for type II IFN signature ([Moreira-Teixeira et al., 2020](#)) was performed using the AddModuleScore function with ctrl = 100. Further sub-clustering of the CD4 T cells was performed by extracting just the CD4⁺ clusters in the ADT dataset. Dimensionality reduction and clustering (resolution = 0.8) were rerun. Annotation was performed as described above. Differential expression between NCan and canonical naïve CD4 T cell clusters was calculated in Seurat by pseudobulking within each cluster using PrepSCTFindMarkers and AggregateExpression. Very lowly expressed genes (median count < 20) were filtered out before running FindMarkers using DESeq2.

Cell communication analysis was performed using LIA-NA+ (v1.2.1), using LIANA's mouse consensus databases ([Dimitrov et al., 2022](#)). Clusters were filtered to require a minimum of three cells, while cells were filtered to require expression of at least 200 genes. Counts were log1p transformed. Results were filtered by specificity_rank (specificity_rank < 0.05) and magnitude_rank (magnitude_rank < 0.25) scores. Visualizations were performed with the "chord_freq" function to show signaling pathways of interest for each group.

Macrophage infection

Bone marrow was isolated from femurs by centrifugation and differentiated in DMEM containing 10% FBS and 20% L929-conditioned supernatants for 7 days. Macrophages were harvested, counted, replated, and rested overnight. After incubation with 10 ng/ml of recombinant IFN γ (#485-MI-100; R&D Systems) overnight, macrophages were washed and harvested in PBS containing 10 nM EDTA and 2% FBS. Cells were passed

through a 40- μ m filter, blocked with anti-CD16/CD32, and stained with antibodies against CD11b (FITC, #101205; Biolegend), PD-L1 (APC, #124312; Biolegend), and fixable Live/Dead near-infrared stain. Samples were fixed, stored, and run on the flow cytometer as described above.

Lung histology

The top right lung lobe from each mouse was submerged in 10% neutral-buffered formalin (#SF100-4; Thermo Fisher Scientific) immediately after euthanasia. After fixation, lung lobes were embedded in paraffin, sectioned, and stained with H&E by the UMass Chan Medical School Morphology Core. Slides were scanned by the UMass Chan Medical School SCOPE Microscopy Core (RRID: SCR_022721) using a TissueFAXS SL Q microscope (TissueGnostics) at 20 $\times$ magnification.

Blocked PCA

Data were z-scored (mean-centered and variance-scaled). All features of a type, e.g., CD4 or IL2, were collected, and a PCA was performed on each feature type individually using "pca_ropls" from the "systemsRology" R package (v1.1). PC scores were extracted and clustered with Gaussian Mixture Models into two groups with "gmm" from the ClusterR R package (v1.3.2), using 100 iterations starting from a random seed and Euclidean distance for clustering. Group assignments were predicted for each sample, and then compared with ground truth group labels using the ARI generated by "adjustedRandIndex" in "mclust" (v6.0.1). ARI values were then compared between feature types.

COMPASS

Using the ICS data, polyfunctional gates were generated in FlowJo (v10) for all CD4 cells and exported as csv files. Data were transformed from proportions to counts by multiplying with total CD4 count and rounding to the nearest integer. COMPASS analysis was run for each stimulation using the "Simple_COMPASS" function from the "COMPASS" package (v1.36.2) with 40,000 iterations and eight replications ([Lin et al., 2015](#)). T cell subsets with an average COMPASS-determined mean response <0.01 were filtered out. The percentage of responding T cells that belong to a subset was determined by taking the higher of either zero or the proportion of responding T cells when stimulated minus the proportion of responding T cells when unstimulated. Significance for the difference in the percentage of responding T cells between NCan and canonical mice was calculated using two-tailed Mann-Whitney U tests with Bonferroni correction. The "PolyfunctionalityScore" function from COMPASS was used to determine a polyfunctionality score for each individual mouse, which summarizes both the number of antigen-specific polyfunctional subsets detected out of all possible subsets as well as the degree of polyfunctionality. Significance for the difference in polyfunctionality scores for NCan and canonical mice was calculated using the two-tailed Mann-Whitney U test, corrected via Benjamini-Hochberg. Only polyfunctional subsets are shown.

Multivariate analysis

Prior to multivariate analyses, all features were appropriately normalized. Features with >50% zero values were removed from

the dataset for multivariate analyses. All features were then mean-centered and scaled to unit variance. All heatmaps shown were generated with the “pheatmap” R package (v1.0.12).

For PCA and PLS-DA, the R package *systemssRology* (v1.1) was utilized. For classification models, LASSO variable selection was first performed via the LASSO feature selection algorithm, which was run 100 times on the entire dataset using the function “select_lasso.” Features selected in 80% or more of the 100 rounds were used in the final model to reduce the risk of overfitting the models due to the large number of available features. For regression analysis, PLS-R was performed on the data after feature reduction via LASSO regularization and variable selection. LASSO selection was performed as described above, with features selected in 90% or more of rounds used in the final model.

PLS-DA classification model and PLS-R predictive model performances were assessed using a fivefold cross-validation approach, and reported cross-validation accuracy is the mean of 10 rounds of fivefold cross-validation. Each round includes 100 repeats of feature selection per fold. To assess the importance of selected features, negative control models were built both by permuting group labels and by selecting random, size-matched features in place of true selected features. 10 rounds of cross-validation with 100 permutation and 100 random-feature trials per round (again with 100 repeats of feature selection per fold per round for the permutation trials) were performed, and distributions were compared via a Wilcoxon rank-sum test. For PLS-R models, this process was also repeated with each strain of mice held out completely from the cross-validation testing and training datasets to confirm that one mouse strain was not driving predictive ability. For PLS-DAs, accuracy is defined as the fraction of correctly predicted class in the held-out test data. For PLS-Rs, accuracy is defined as the Spearman correlation of predicted and measured values across all data. All multivariate analyses were visualized and analyzed using R (v4.3.1).

To generate a correlation network for LASSO-selected features, |Spearman correlations| >0.7 and Benjamini-Hochberg corrected $P < 0.01$ were filtered from all Spearman correlation coefficients across the dataset, including COMPASS-calculated polyfunctional CD4 T cell measurements, for only NCan animals. To more stringently reduce the number of features considered, null distributions of randomized correlations were created for each LASSO-selected feature independently by shuffling LASSO-selected feature values across samples and computing the resultant Spearman correlation matrix. Only true correlations that fell above the 95th percentile of randomized correlations were retained. To be included in the correlation plot, features had to be significantly correlated with one or more LASSO-selected features, and only correlations to those LASSO-selected features are shown. Manual positioning and color corrections were made in Adobe Illustrator 2021 for the sole purpose of better visualization.

TransCompR

Publicly available single-cell transcriptomic data generated via SELECT-seq from seven donors (three RSTR and four LTBI) (Sun et al., 2024) were retrieved as a Seurat object via Zenodo. Mouse

lung scRNAseq data generated in this work was subset to only CD4 T cell clusters. Mouse genes were converted to homologs using the Mouse Genome Database at the Mouse Genome Informatics website, The Jackson Laboratory. DEGs ($P < 0.05$) were found across the human SELECT-seq data between RSTR and LTBI individuals via a Wilcoxon rank-sum test, with P values corrected via Benjamini-Hochberg, and likewise across the mouse scRNAseq data. To build a TransCompR model, a PCA space was built on the human DEGs after filtering for homology. The PCA utilized the lesser of 20 hPCs or the number of PCs that explained at least 50% of the total SELECT-seq variance. Mouse scRNAseq data were projected into the human PCA space by multiplying the human PC loadings by the mouse data for each cell in the dataset, creating mouse principal components (mPCs). The mouse T cell data were then reduced to just the CD4_act1 subset to best align with the human data. One mouse, CC024_29, was removed from analysis due to low (<50 cells) counts of CD4 T cells detected. Cohen's D was used to assess the relationship of each mPC to the mouse phenotype (NCan or canonical). A Cohen's D of 0.2 (small effect size) was used as a cutoff, and mPCs that made the cutoff were then used to build an LDA model to predict phenotype per cell. The model was trained by holding each mouse out through 100 rounds of fivefold cross-validation using data from all other mice, down-sampled to allow for balanced class sizes. The model built was then used to predict the phenotype of every cell from the left out mouse, and the mouse is predicted to have the phenotype that most of its cells share. This same framework was used to generate null models either built on shuffled labels across the entire dataset or on random size-matched mPCs. The top 20% and bottom 20% of the loadings of each significant mPC were taken for enrichment analysis via STRING, using database version 12.0. Terms were considered enriched if the false discovery rate-corrected P value was <0.05. Resultant enriched terms were filtered to remove non-T cell subset-related terms and focus only on GO Process terms, and ratios of the number of genes detected in the top or bottom 20% of the loadings to the number of overall genes in a given gene set were computed.

Statistics

Statistical comparisons were performed using GraphPad Prism v10 or R (v4.3.2) using methods as indicated in the figure legends.

Study approval

All animal studies were approved by the Institutional Animal Care and Use Committee of the UMass Chan Medical School (#A3306-01; Animal Welfare Assurance).

Online supplemental material

Fig. S1 shows the additional data relevant to Fig. 2, including individual transcription factor staining and ICS staining as well as COMPASS polyfunctionality analysis of polyclonally stimulated lung CD4 T cells. Fig. S2 shows the additional data relevant to Fig. 3, including the scores plots and LASSO-selected features at 2 wk after infection. The scores plot and LASSO-selected features for the analysis of the baseline spleen phenotyping

study in Fig. 4 are also shown. Fig. S3 shows the additional data relevant to Fig. 5, including the scRNAseq whole lung UMAP and dot plot illustrating expression of cluster-defining genes. Fig. S4 contains the additional data relevant to Fig. 5, including the scRNAseq UMAP of re-clustered CD4 T cells, expression of *Ifng* by cluster, dot plot of key GEX by cluster, heatmap of DEGs between canonical and NCan mice in the naïve CD4 T cell cluster, and the type II IFN signature scoring of the whole lung scRNAseq object. The CD4 and CD8a T cell depletion efficiency, scores plots for PCA of flow cytometry data at 3 and 4 wk after infection, and functional confirmation of *Ifngr1* editing in the PWK/PhJ *Ifngr1*^{-/-} mouse relevant to Fig. 7 are also shown. Fig. S5 shows the multivariate model validation relevant to Figs. 9 and 10. Table S1 provides the STRING enrichment results for the significant PCs relevant to Fig. 10. Table S2 lists all the flow cytometry antibodies used in this study.

Data availability

All data from this study are available on the IMPAc-TB SEEK database at <https://fairdomhub.org/studies/1320>. Code can be found at <https://github.com/Lauffenburger-Lab>.

Acknowledgments

We would like to acknowledge Meng Sun, Chetan Seshadri, Rocky Lai, and Samuel Behar for their helpful advice and discussions, as well as the members of the Sassetti lab for their technical assistance. We are also grateful for the expertise of the UMass Chan Flow Cytometry Core. We thank the National Institutes of Health (NIH) Tetramer Core Facility (contract number 75N93020D00005) for providing MR-1 5-OP-U and CD1d PBS-57 tetramers.

This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, the NIH, and the Department of Health and Human Services, under contract No. 75N93019C00071 and grant No. AI181898.

Author contributions: M.K. Proulx: conceptualization, data curation, formal analysis, investigation, validation, visualization, and writing—original draft, review, and editing. C.D. Wiggins: data curation, formal analysis, methodology, visualization, and writing—original draft, review, and editing. C.J. Reames: investigation, validation, and writing—review and editing. C. Wu: formal analysis. M.C. Kiritsy: investigation. P. Xu: methodology and resources. J.C. Gallant: methodology and resources. P.S. Grace: investigation. B.A. Fenderson: investigation. C.M. Smith: investigation and writing—review and editing. C.S. Lindestam Arlehamn: resources and writing—review and editing. G. Alter: data curation, investigation, methodology, and supervision. D.A. Lauffenburger: conceptualization, funding acquisition, project administration, supervision, and writing—review and editing. C.M. Sassetti: conceptualization, funding acquisition, project administration, supervision, and writing—review and editing.

Disclosures: G. Alter reported personal fees from Moderna, nonfinancial support from Systems Seromyx and Leyden Labs,

and "other" from Sanofi, GSK, and Pfizer outside the submitted work. No other disclosures were reported.

Submitted: 23 September 2024

Revised: 2 January 2025

Accepted: 7 March 2025

References

- Akter, S., K.S. Chauhan, M.D. Dunlap, J.A. Choreño-Parra, L. Lu, E. Esaulova, J. Zúñiga, M.N. Artyomov, D. Kaushal, and S.A. Khader. 2022. Mycobacterium tuberculosis infection drives a type I IFN signature in lung lymphocytes. *Cell Rep.* 39:110983. <https://doi.org/10.1016/j.celrep.2022.110983>
- Anderson, G.P. 2008. Endotyping asthma: New insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet.* 372:1107–1119. [https://doi.org/10.1016/S0140-6736\(08\)61452-X](https://doi.org/10.1016/S0140-6736(08)61452-X)
- Aran, D., A.P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R.P. Naikawadi, P.J. Wolters, A.R. Abate, et al. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20:163–172. <https://doi.org/10.1038/s41590-018-0276-y>
- Boisson-Dupuis, S., N. Ramirez-Alejo, Z. Li, E. Patin, G. Rao, G. Kerner, C.K. Lim, D.N. Krementsov, N. Hernandez, C.S. Ma, et al. 2018. Tuberculosis and impaired IL-23-dependent IFN- γ immunity in humans homozygous for a common TYK2 missense variant. *Sci. Immunol.* 3:eaa8714. <https://doi.org/10.1126/sciimmunol.aau8714>
- Brinkman, E.K., T. Chen, M. Amendola, and B. van Steensel. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42:e168. <https://doi.org/10.1093/nar/gku936>
- Brubaker, D.K., M.P. Kumar, E.L. Chiswick, C. Gregg, A. Starchenko, P.N. Vega, A.N. Southard-Smith, A.J. Simmons, E.A. Scoville, L.A. Coburn, et al. 2020. An interspecies translation model implicates integrin signaling in infliximab-resistant inflammatory bowel disease. *Sci. Signal.* 13:eaay3258. <https://doi.org/10.1126/scisignal.aay3258>
- Bustamante, J., S. Boisson-Dupuis, L. Abel, and J.-L. Casanova. 2014. Mendelian susceptibility to mycobacterial disease: Genetic, immunological, and clinical features of inborn errors of IFN- γ immunity. *Semin. Immunol.* 454–470. <https://doi.org/10.1016/j.smim.2014.09.008>
- Cardona, P., and P.-J. Cardona. 2019. Regulatory T cells in Mycobacterium tuberculosis infection. *Front. Immunol.* 10:2139. <https://doi.org/10.3389/fimmu.2019.02139>
- Cobat, A., C.J. Gallant, L. Simkin, G.F. Black, K. Stanley, J. Hughes, T.M. Doherty, W.A. Hanekom, B. Eley, J.-P. Jais, et al. 2009. Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. *J. Exp. Med.* 206:2583–2591. <https://doi.org/10.1084/jem.20090892>
- Cooper, A.M., L.B. Adams, D.K. Dalton, R. Appelberg, and S. Ehlers. 2002. IFN- γ and NO in mycobacterial disease: New jobs for old hands. *Trends Microbiol.* 10:221–226. [https://doi.org/10.1016/S0966-842X\(02\)02344-2](https://doi.org/10.1016/S0966-842X(02)02344-2)
- Cooper, A.M., D.K. Dalton, T.A. Stewart, J.P. Griffin, D.G. Russell, and I.M. Orme. 1993. Disseminated tuberculosis in interferon gamma gene-disrupted mice. *J. Exp. Med.* 178:2243–2247. <https://doi.org/10.1084/jem.178.6.2243>
- Cowley, S.C., and K.L. Elkins. 2003. CD4⁺ T cells mediate IFN- γ -independent control of Mycobacterium tuberculosis infection both in vitro and in vivo. *J. Immunol.* 171:4689–4699. <https://doi.org/10.4049/jimmunol.171.9.4689>
- Dalton, D.K., L. Haynes, C.-Q. Chu, S.L. Swain, and S. Wittmer. 2000. Interferon γ eliminates responding CD4 T cells during mycobacterial infection by inducing apoptosis of activated CD4 T cells. *J. Exp. Med.* 192:117–122. <https://doi.org/10.1084/jem.192.1.117>
- Davies, L.R., M.T. Smith, D. Cizmeci, S. Fischinger, J.S.-L. Lee, L.L. Lu, E.D. Layton, A.D. Grant, K. Fielding, and C.M. Stein. 2023. IFN- γ independent markers of Mycobacterium tuberculosis exposure among male South African gold miners. *EBioMedicine.* 93:104678. <https://doi.org/10.1016/j.ebiom.2023.104678>
- Desvignes, L., and J.D. Ernst. 2009. Interferon- γ -responsive non-hematopoietic cells regulate the immune response to Mycobacterium tuberculosis. *Immunity.* 31:974–985. <https://doi.org/10.1016/j.immuni.2009.10.007>
- Dimitrov, D., D. Türei, M. Garrido-Rodriguez, P.L. Burmedi, J.S. Nagai, C. Boys, R.O. Ramirez Flores, H. Kim, B. Szalai, I.G. Costa, et al. 2022. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* 13:3224. <https://doi.org/10.1038/s41467-022-30755-0>

- Flynn, J.L., J. Chan, K.J. Triebold, D.K. Dalton, T.A. Stewart, and B.R. Bloom. 1993. An essential role for interferon gamma in resistance to Mycobacterium tuberculosis infection. *J. Exp. Med.* 178:2249–2254. <https://doi.org/10.1084/jem.178.6.2249>
- Gallegos, A.M., J.W. van Heijst, M. Samstein, X. Su, E.G. Pamer, and M.S. Glickman. 2011. A gamma interferon independent mechanism of CD4 T cell mediated control of M. tuberculosis infection in vivo. *PLoS Pathog.* 7:e1002052. <https://doi.org/10.1371/journal.ppat.1002052>
- Gideon, H.P., J. Phuah, A.J. Myers, B.D. Bryson, M.A. Rodgers, M.T. Coleman, P. Maiello, T. Rutledge, S. Marino, S.M. Fortune, et al. 2015. Variability in tuberculosis granuloma T cell responses exists, but a balance of pro and anti-inflammatory cytokines is associated with sterilization. *PLoS Pathog.* 11:e1004603. <https://doi.org/10.1371/journal.ppat.1004603>
- Gopal, R., Y. Lin, N. Obermajer, S. Slight, N. Nuthalapati, M. Ahmed, P. Kalinski, and S.A. Khader. 2012. IL-23-dependent IL-17 drives Th1-cell responses following Mycobacterium bovis BCG vaccination. *Eur. J. Immunol.* 42:364–373. <https://doi.org/10.1002/eji.201141569>
- Gopal, R., L. Monin, S. Slight, U. Uche, E. Blanchard, B.A. Fallert Junecko, R. Ramos-Payan, C.L. Stallings, T.A. Reinhart, J.K. Kolls, et al. 2014. Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection. *PLoS Pathog.* 10:e1004099. <https://doi.org/10.1371/journal.ppat.1004099>
- Graham, J.B., J.L. Swarts, S.R. Leist, A. Schäfer, T.A. Bell, P. Hock, J. Farrington, G.D. Shaw, M.T. Ferris, and F.P.-M. de Villena. 2024. Unique immune profiles in collaborative cross mice linked to survival and viral clearance upon infection. *IScience.* 27:109103. <https://doi.org/10.1016/j.isci.2024.109103>
- Hao, Y., T. Stuart, M.H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, and C. Fernandez-Granda. 2024. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* 42:293–304. <https://doi.org/10.1038/s41587-023-01767-y>
- Kroon, E.E., C.J. Kinnear, M. Orlova, S. Fischinger, S. Shin, S. Boolay, G. Walzl, A. Jacobs, R.J. Wilkinson, and G. Alter. 2020. An observational study identifying highly tuberculosis-exposed, HIV-1-positive but persistently TB, tuberculin and IGRA negative persons with M. tuberculosis specific antibodies in Cape Town, South Africa. *EBioMedicine.* 61:103053. <https://doi.org/10.1016/j.ebiom.2020.103053>
- Kurtz, S.L., R.E. Baker, F.J. Boehm, C.C. Lehman, L.R. Mittereder, H. Khan, A.P. Rossi, D.M. Gatti, G. Beamer, and C.M. Sasseti. 2024. Multiple genetic loci influence vaccine-induced protection against Mycobacterium tuberculosis in genetically diverse mice. *PLoS Pathog.* 20:e1012069. <https://doi.org/10.1371/journal.ppat.1012069>
- Lai, R., D.N. Gong, T. Williams, A.F. Ogunola, K. Cavallo, C.S.L. Arlehamn, S. Acolatse, G.L. Beamer, M.T. Ferris, and C.M. Sasseti. 2024. Host genetic background is a barrier to broadly effective vaccine-mediated protection against tuberculosis. *J. Clin. Invest.* 133:e167762. <https://doi.org/10.1172/JCI167762>
- Lee, M.J., C. Wang, M.J. Carroll, D.K. Brubaker, B.T. Hyman, D.A. Lauffenburger. 2021. Computational interspecies translation between Alzheimer's disease mouse models and human subjects identifies innate immune complement, TYROBP, and TAM receptor agonist signatures, distinct from influences of aging. *Front. Neurosci.* 15:727784. <https://doi.org/10.3389/fnins.2021.727784>
- Li, H., X.X. Wang, B. Wang, L. Fu, G. Liu, Y. Lu, M. Cao, H. Huang, and B. Javid. 2017. Latently and uninfected healthcare workers exposed to TB make protective antibodies against Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA.* 114:5023–5028. <https://doi.org/10.1073/pnas.1611776114>
- Lin, L., G. Finak, K. Ushey, C. Seshadri, T.R. Hawn, N. Frahm, T.J. Scriba, H. Mahomed, W. Hanekom, P.-A. Bart, et al. 2015. COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nat. Biotechnol.* 33:610–616. <https://doi.org/10.1038/nbt.3187>
- Lin, P.L., and J.L. Flynn. 2018. The end of the binary era: Revisiting the spectrum of tuberculosis. *J. Immunol.* 201:2541–2548. <https://doi.org/10.4049/jimmunol.1800993>
- Lindestam Arlehamn, C.S., D.M. McKinney, C. Carpenter, S. Paul, V. Rozot, E. Makgutho, Y. Gregg, M. van Rooyen, J.D. Ernst, M. Hatherill, et al. 2016. A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy M. tuberculosis infected South Africans. *PLoS Pathog.* 12:e1005760. <https://doi.org/10.1371/journal.ppat.1005760>
- Lu, L.L., M.T. Smith, K.K.Q. Yu, C. Luedemann, T.J. Suscovich, P.S. Grace, A. Cain, W.H. Yu, T.R. McKittrick, D. Lauffenburger, et al. 2019. IFN- γ -independent immune markers of Mycobacterium tuberculosis exposure. *Nat. Med.* 25:977–987. <https://doi.org/10.1038/s41591-019-0441-3>
- MacMicking, J.D., G.A. Taylor, and J.D. McKinney. 2003. Immune control of tuberculosis by IFN- γ -inducible LRG-47. *Science.* 302:654–659. <https://doi.org/10.1126/science.1088063>
- McGinnis, C.S., L.M. Murrow, and Z.J. Gartner. 2019. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8:329–337.e324. <https://doi.org/10.1016/j.cels.2019.03.003>
- Mishra, B.B., V.A. Rathinam, G.W. Martens, A.J. Martinot, H. Kornfeld, K.A. Fitzgerald, and C.M. Sasseti. 2013. Nitric oxide controls the immunopathology of tuberculosis by inhibiting NLRP3 inflammasome-dependent processing of IL-1 β . *Nat. Immunol.* 14:52–60. <https://doi.org/10.1038/ni.2474>
- Mogues, T., M.E. Goodrich, L. Ryan, R. LaCourse, and R.J. North. 2001. The relative importance of T cell subsets in immunity and immunopathology of airborne Mycobacterium tuberculosis infection in mice. *J. Exp. Med.* 193:271–280. <https://doi.org/10.1084/jem.193.3.271>
- Moreira-Teixeira, L., O. Tabone, C.M. Graham, A. Singhanian, E. Stavropoulos, P.S. Redford, P. Chakravarty, S.L. Priestnall, A. Suarez-Bonnet, E. Herbert, et al. 2020. Mouse transcriptome reveals potential signatures of protection and pathogenesis in human tuberculosis. *Nat. Immunol.* 21:464–476. <https://doi.org/10.1038/s41590-020-0610-z>
- Nandi, B., and S.M. Behar. 2011. Regulation of neutrophils by interferon- γ limits lung inflammation during tuberculosis infection. *J. Exp. Med.* 208:2251–2262. <https://doi.org/10.1084/jem.20110919>
- Niazi, M.K., N. Dhulekar, D. Schmidt, S. Major, R. Cooper, C. Abeijon, D.M. Gatti, I. Kramnik, B. Yener, M. Gurcan, and G. Beamer. 2015. Lung necrosis and neutrophils reflect common pathways of susceptibility to Mycobacterium tuberculosis in genetically diverse, immune-competent mice. *Dis. Model. Mech.* 8:1141–1153. <https://doi.org/10.1242/dmm.020867>
- Ravesloot-Chávez, M.M., E. Van Dis, D. Fox, X.H. Nguyenla, S.L. Rawal, M.A. Ballinger, H. Thomas, D. Kotov, R. Vance, and M.W. Nachman. 2023. Tuberculosis susceptibility in genetically diverse mice reveals functional diversity of neutrophils. *bioRxiv.* <https://doi.org/10.1101/2023.06.29.547125> (Preprint posted June 30, 2023).
- Sakai, S., K.D. Kauffman, M.A. Sallin, A.H. Sharpe, H.A. Young, V.V. Ganusov, and D.L. Barber. 2016. CD4 T cell-derived IFN- γ plays a minimal role in control of pulmonary Mycobacterium tuberculosis infection and must be actively repressed by PD-1 to prevent lethal disease. *PLoS Pathog.* 12:e1005667. <https://doi.org/10.1371/journal.ppat.1005667>
- Sakai, S., K.D. Kauffman, J.M. Schenkel, C.C. McBerry, K.D. Mayer-Barber, D. Masopust, and D.L. Barber. 2014. Cutting edge: Control of Mycobacterium tuberculosis infection by a subset of lung parenchyma-homing CD4 T cells. *J. Immunol.* 192:2965–2969. <https://doi.org/10.4049/jimmunol.1400019>
- Sallin, M.A., S. Sakai, K.D. Kauffman, H.A. Young, J. Zhu, and D.L. Barber. 2017. Th1 differentiation drives the accumulation of intravascular, non-protective CD4 T cells during tuberculosis. *Cell Rep.* 18:3091–3104. <https://doi.org/10.1016/j.celrep.2017.03.007>
- Saul, M.C., V.M. Philip, L.G. Reinholdt, Center for Systems Neurogenetics of Addiction, and E.J. Chesler. 2019. High-diversity mouse populations for complex traits. *Trends Genet.* 35:501–514. <https://doi.org/10.1016/j.tig.2019.04.003>
- Simmons, J.D., K.A. Dill-McFarland, C.M. Stein, P.T. Van, V. Chihota, T. Ntshiq, P. Maenette, G.J. Peterson, P. Bencheq, M. Nsereko, et al. 2022. Monocyte transcriptional responses to Mycobacterium tuberculosis associate with resistance to tuberculin skin test and interferon gamma release assay conversion. *MSphere.* 7:e0015922. <https://doi.org/10.1128/msphere.00159-22>
- Simmons, J.D., C.M. Stein, C. Seshadri, M. Campo, G. Alter, S. Fortune, E. Schurr, R.S. Wallis, G. Churchyard, H. Mayanja-Kizza, et al. 2018. Immunological mechanisms of human resistance to persistent Mycobacterium tuberculosis infection. *Nat. Rev. Immunol.* 18:575–589. <https://doi.org/10.1038/s41577-018-0025-B>
- Smith, C.M., R.E. Baker, M.K. Proulx, B.B. Mishra, J.E. Long, S.W. Park, H.-N. Lee, M.C. Kiritsy, M.M. Bellerose, A.J. Olive, et al. 2022. Host-pathogen genetic interactions underlie tuberculosis susceptibility in genetically diverse mice. *eLife.* 11:e74419. <https://doi.org/10.7554/eLife.74419>
- Smith, C.M., M.K. Proulx, A.J. Olive, D. Laddy, B.B. Mishra, C. Moss, N.M. Gutierrez, M.M. Bellerose, P. Barreira-Silva, J.Y. Phuah, et al. 2016. Tuberculosis susceptibility and vaccine protection are independently controlled by host genotype. *MBio.* 7:e01516. <https://doi.org/10.1128/mBio.01516-16>
- Smith, C.M., and C.M. Sasseti. 2018. Modeling diversity: Do homogeneous laboratory strains limit discovery? *Trends Microbiol.* 26:892–895. <https://doi.org/10.1016/j.tim.2018.08.002>

- Sobota, R.S., C.M. Stein, N. Kodaman, I. Maro, W. Wieland-Alter, R.P. Igo Jr., A. Magohe, L.L. Malone, K. Chervenak, N.B. Hall, et al. 2017. A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in HIV-positive individuals from tuberculosis hyper-endemic regions in east Africa. *PLoS Genet.* 13:e1006710. <https://doi.org/10.1371/journal.pgen.1006710>
- Srivastava, A., A.P. Morgan, M.L. Najarian, V.K. Sarsani, J.S. Sigmon, J.R. Shorter, A. Kashfeen, R.C. McMullan, L.H. Williams, P. Giusti-Rodríguez, et al. 2017. Genomes of the mouse collaborative cross. *Genetics.* 206:537–556. <https://doi.org/10.1534/genetics.116.198838>
- Stein, C.M., S. Zalwango, L.L. Malone, S. Won, H. Mayanja-Kizza, R.D. Mugerwa, D.V. Leontiev, C.L. Thompson, K.C. Cartier, R.C. Elston, et al. 2008. Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS One.* 3:e4094. <https://doi.org/10.1371/journal.pone.0004094>
- Sun, M., J.M. Phan, N.S. Kieswetter, H. Huang, K.K. Yu, M.T. Smith, Y.E. Liu, C. Wang, S. Gupta, and G. Obermoser. 2024. Specific CD4⁺ T cell phenotypes associate with bacterial control in people who 'resist' infection with *Mycobacterium tuberculosis*. *Nat. Immunol.* 25:1411–1421. <https://doi.org/10.1038/s41590-024-01897-8>
- Threadgill, D.W., and G.A. Churchill. 2012. Ten years of the collaborative cross. *G3.* 2:153–156. <https://doi.org/10.1534/g3.111.001891>
- Torrado, E., J.J. Fountain, M. Liao, M. Tighe, W.W. Reiley, R.P. Lai, G. Meintjes, J.E. Pearl, X. Chen, D.E. Zak, et al. 2015. Interleukin 27R regulates CD4⁺ T cell phenotype and impacts protective immunity during *Mycobacterium tuberculosis* infection. *J. Exp. Med.* 212: 1449–1463. <https://doi.org/10.1084/jem.20141520>
- Vignali, D.A., L.W. Collison, and C.J. Workman. 2008. How regulatory T cells work. *Nat. Rev. Immunol.* 8:523–532. <https://doi.org/10.1038/nri2343>

Supplemental material

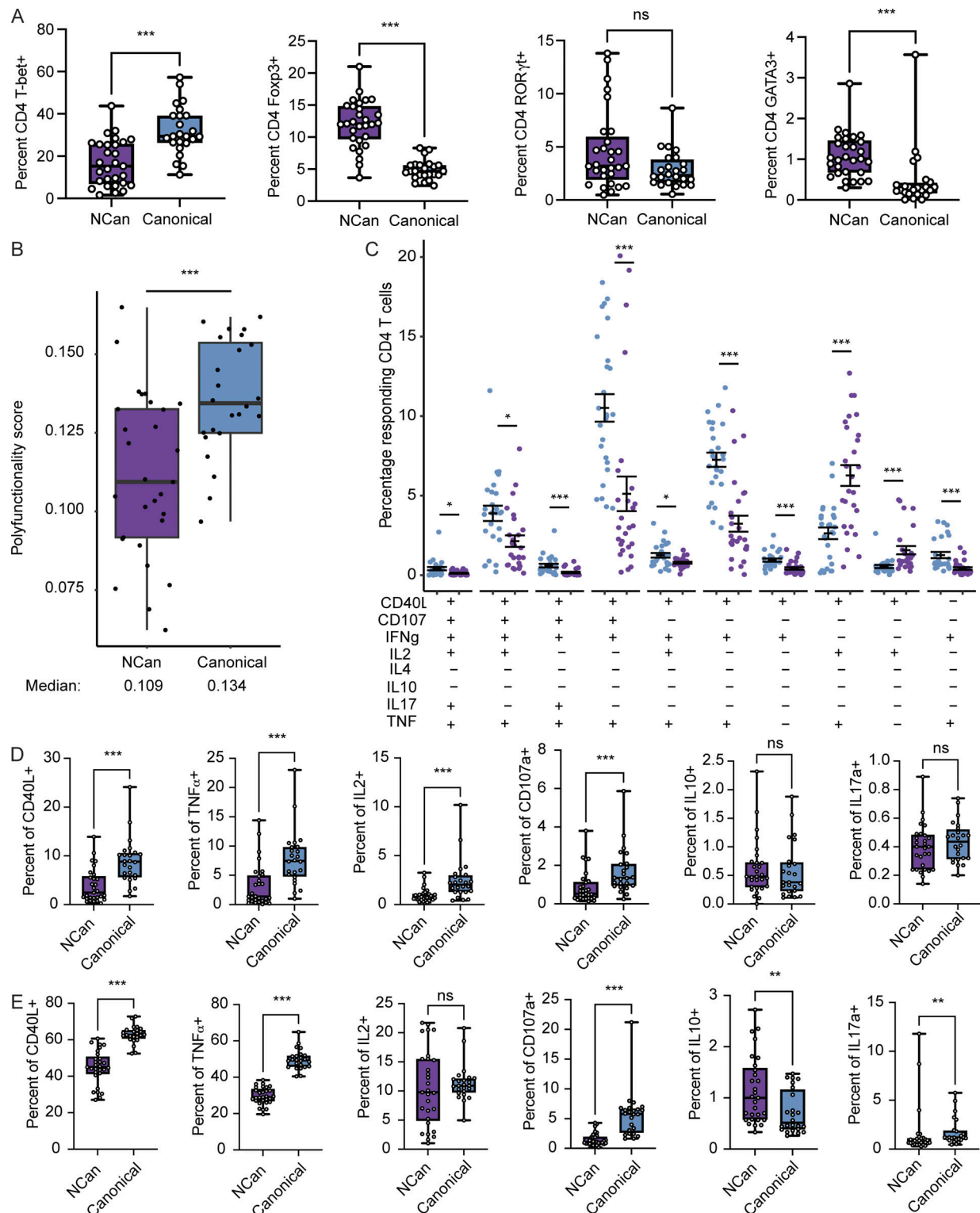


Figure S1. **NCan mice control *Mtb* replication with low Th1 cytokines and reduced CD4 T cell polyfunctionality.** (A) Percent of lung CD4 T cells staining positive by flow cytometry for the transcription factors T-bet, Foxp3, RORγt, or GATA3. (B) Polyfunctionality score calculated by COMPASS for lung CD4 T cells after restimulation with anti-CD3 and anti-CD28 antibodies. (C) Percentage of responding lung CD4 T cells after restimulation with anti-CD3 and anti-CD28 antibodies. P values derived from Mann–Whitney U test with Benjamini–Hochberg correction. Only subsets with a maximum over 2% are included. Mean $\pm$ SE are visualized as crossbars. (D and E) Percentage of lung CD4 T cells staining for indicated cytokines after ex vivo restimulation with (D) MTB 300 peptide pool or (E) anti-CD3 and anti-CD28 antibodies. (A–E) Box-and-whisker plots depict bounds from the 25th to 75th percentile, median line, and whiskers extending from min to max with all points shown. Statistical significance was assessed via Mann–Whitney U test with Benjamini–Hochberg correction (ns = not significant [$P > 0.05$], * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). All data in this figure are from 5 wk after infection ($n = 2$ –4 mice per sex per strain per time point).

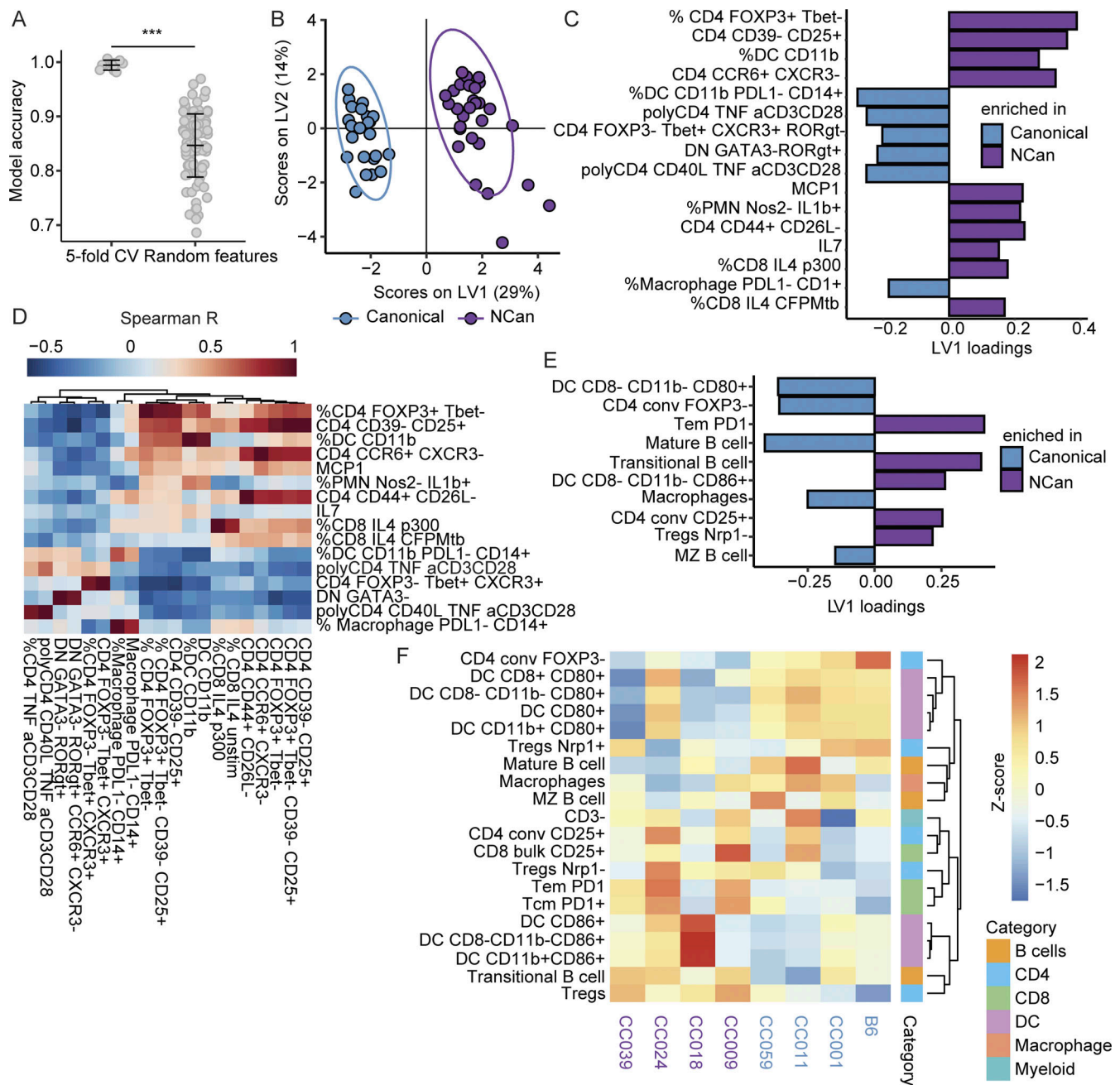


Figure S2. **NCan and canonical mice can be distinguished at 5 wk, 2 wk, and baseline.** (A) Model classification accuracy for 10 rounds of fivefold cross-validation (CV) compared with a model built on randomly selected features via Mann-Whitney U test (***) at 5 wk. Mean $\pm$ SD shown as crossbars. (B) Scores plot for LASSO PLS-DA on lung features at 2 wk after infection ($n = 2-4$ mice per sex per strain per time point). Each dot represents an individual mouse, and the ellipse marks the 95% group confidence interval. (C) LASSO-selected features for the 2 wk PLS-DA are plotted by their contribution to LV1. (D) Heatmap showing correlation values of features highly correlated ($|Spearman R| > 0.7$) with LASSO-selected features at 2 wk. (E) LASSO-selected features for baseline spleen PLS-DA model seen in Fig. 4A are plotted by their contribution to LV1. (F) Average z-scored value of each LASSO-selected spleen feature or highly correlated ($|Spearman R| > 0.8$, $P < 0.01$) feature for each mouse strain (canonical = blue, NCan = purple) at baseline. Features are colored by cell type according to inset legend, and measurements were hierarchically clustered. NK, natural killer cell; DC, dendritic cell; Treg, regulatory T cell.

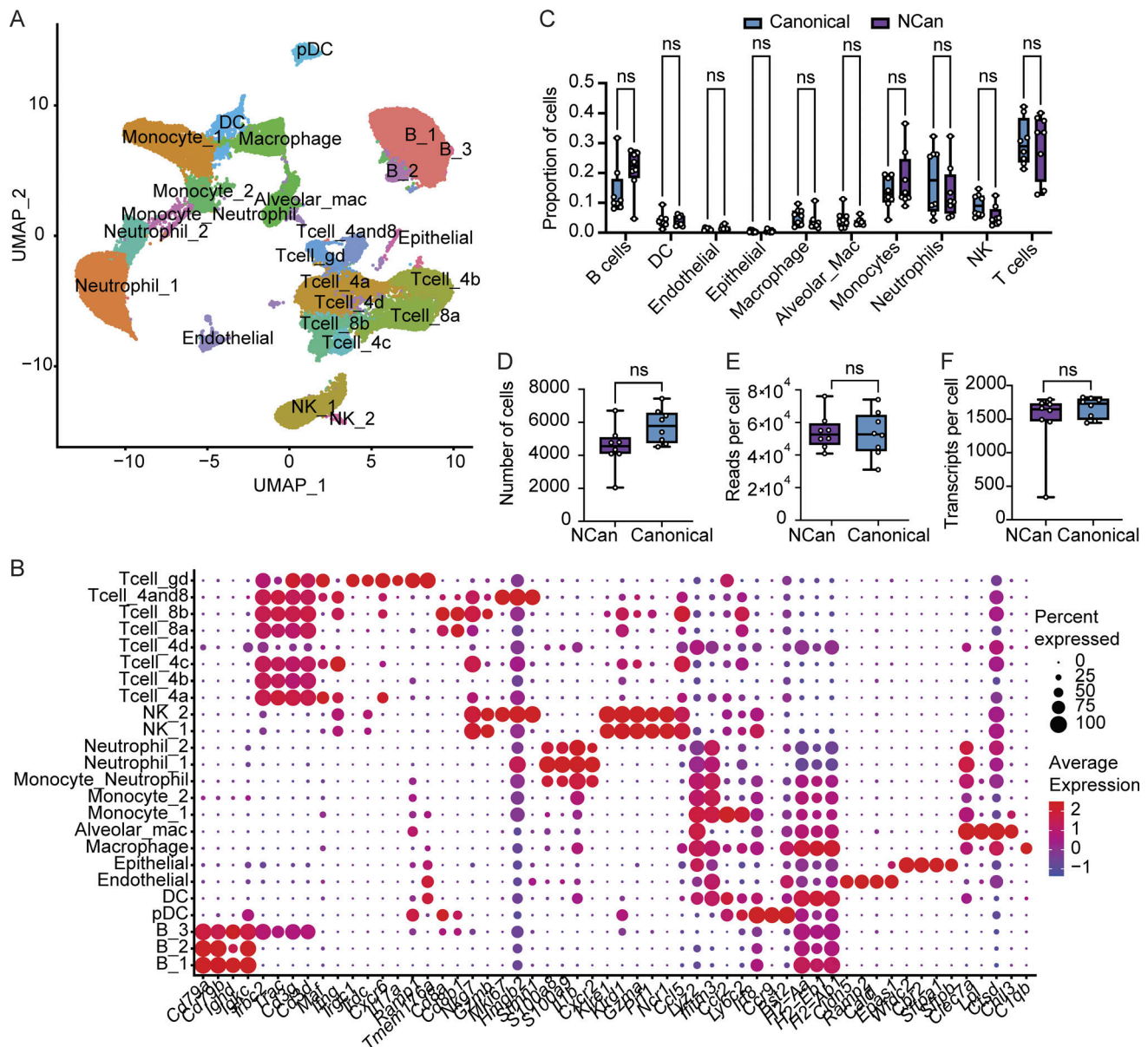
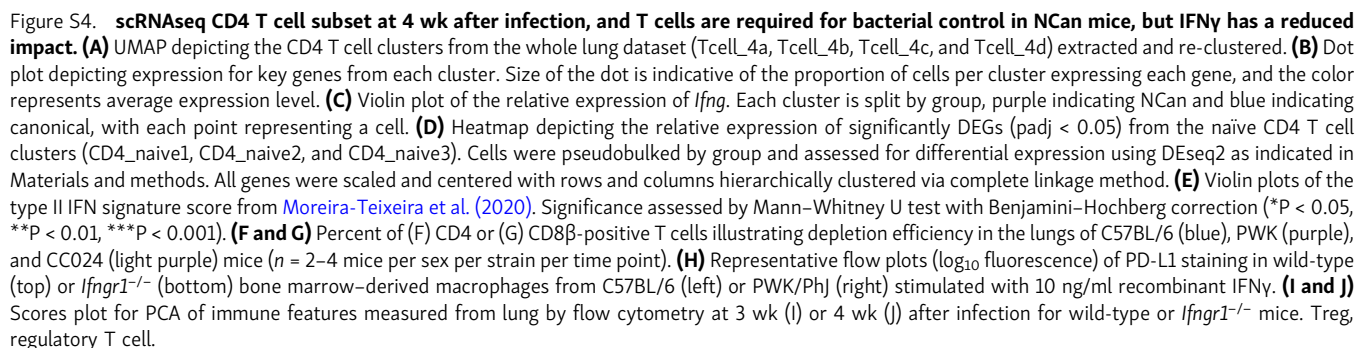


Figure S3. scRNAseq of whole lung at 4 wk after infection. Four male mice per strain from two NCan strains (PWK/PhJ and CC024) and two canonical strains (C57BL/6 and CC011) were infected with H37Rv by low-dose aerosol. At 4 wk after infection, mice were euthanized, and lungs were processed to capture GEX by scRNAseq and stained with a panel of 198 oligo-tagged antibodies to capture surface protein expression. **(A and B)** (A) UMAP of 58,689 high-quality cells retrieved from whole lung and (B) dot plot depicting expression for key genes from each cluster. Size of the dot is indicative of the proportion of cells per cluster expressing each gene, and the color represents average expression level. **(C)** Proportion of major immune cell types. Statistical significance was assessed via Mann-Whitney U test with Benjamini-Hochberg correction (ns = not significant [$P > 0.05$]). **(D-F)** Overall, the dataset averaged (D) 5,138 cells per mouse with (E) 53,543 reads/cell and (F) 1,573 mRNA detected/cell. Significance assessed using the Mann-Whitney U test. All box-and-whisker plots depict bounds from the 25th to 75th percentile, median line, and whiskers extending from min to max with all points shown. NK, natural killer cell; DC, dendritic cell.



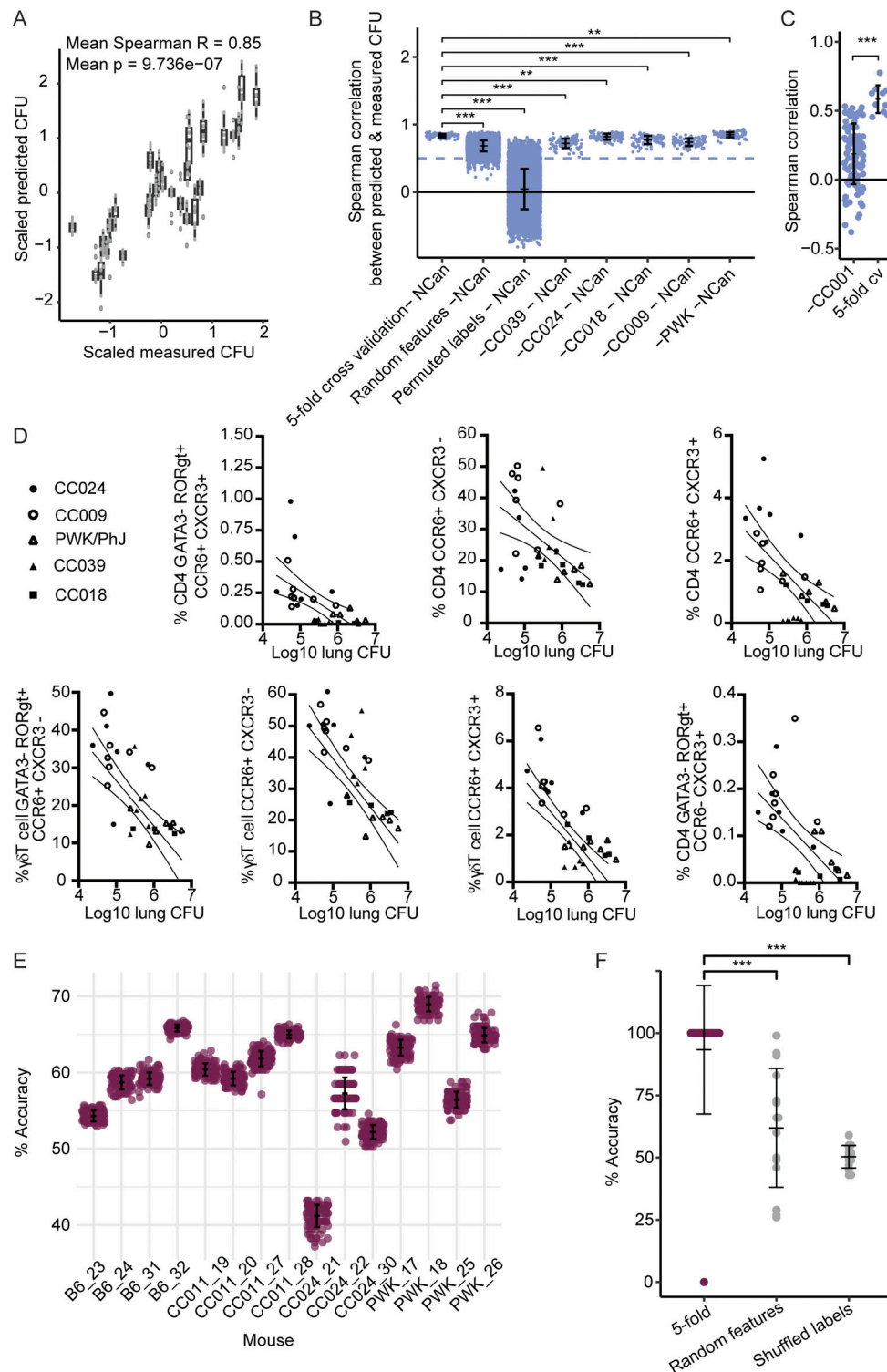


Figure S5. Multivariate model validation. (A) Z-scored values of predicted lung CFU and measured lung CFU for each of 10 rounds of cross-validation. Box-and-whisker plots depict bounds from the 25th to 75th percentile, median line, and whiskers extending to the largest value or a maximum of 150% the interquartile range from the hinge, with all points shown. (B and C) Spearman correlations from 10 rounds of fivefold cross-validation (cv) are shown for (B) predictive NCan lung CFU models compared with null models and (C) canonical lung CFU model compared with a null model. Mann-Whitney U tests compare fivefold cross-validation distributions to the null distributions or to models built without any one individual strain, with P values corrected via Benjamini-Hochberg (**P < 0.01, ***P < 0.001). Mean $\pm$ 1 SD shown as crossbars. (D) Th17-associated features plotted against lung CFU values, with linear regression lines and 95% confidence intervals shown. (E) Average classification accuracy across all cells per mouse for each of 100 rounds of model prediction using LDA models in Fig. 9. (F) Ensemble LDA model accuracy of phenotype prediction for each mouse across 100 rounds of fivefold cross-validation, compared with null models built on random features or shuffled labels. Mann-Whitney U tests compare cross-validation distributions with P values corrected via Benjamini-Hochberg (**P < 0.01, ***P < 0.001). Mean $\pm$ 1 SD shown as crossbars.

Provided online are Table S1 and Table S2. Table S1 provides the STRING enrichment results for the significant PCs relevant to Fig. 10. Table S2 lists all the flow cytometry antibodies used in this study.