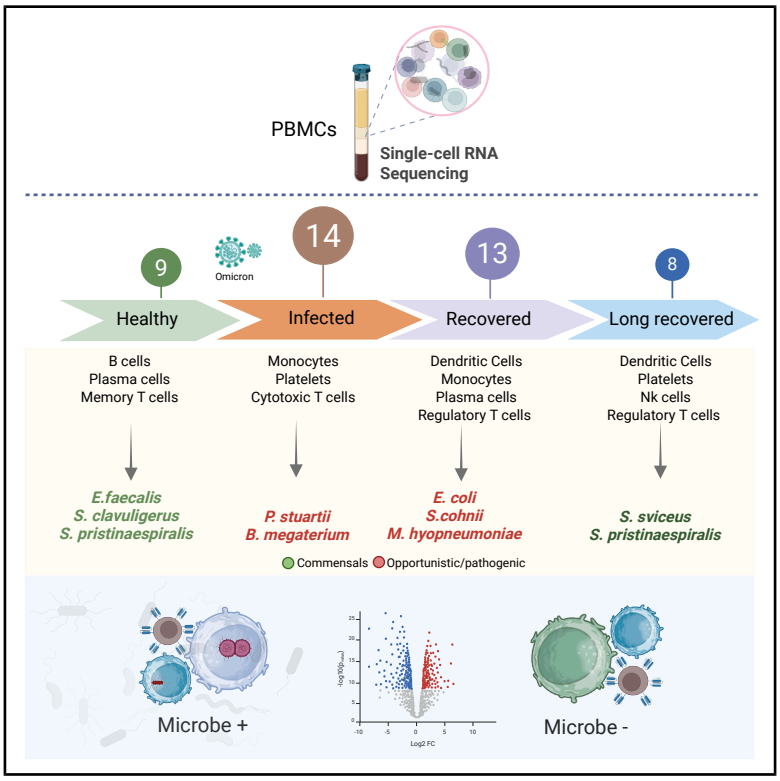
iScience

Article

Intracellular microbial shifts during COVID-19 infection and longitudinal recovery revealed by single-cell RNA sequencing

Graphical abstract



Highlights

- Opportunistic intracellular bacteria were enriched during SARS-CoV-2 infection
- Altered microbial abundance in T cells and platelets of infected individuals
- Microbial genes revealed virulence stress and AMR signatures in the immune cells
- Intracellular bacteria within immune cells modulate recovery post-infection

Authors

Jyoti Soni, Raiyan Ali, Partha Chattopadhyay, ..., Bansidhar Tarai, Sandeep Budhiraja, Rajesh Pandey

Correspondence

rajeshp@igib.in

In brief

Immunology; Microbiology; Transcriptomics

Soni et al., 2026, iScience *29*, 116344
July 17, 2026
https://doi.org/10.1016/j.isci.2026.116344

CellPress

iScience

CellPress
OPEN ACCESS

Article

Intracellular microbial shifts during COVID-19 infection and longitudinal recovery revealed by single-cell RNA sequencing

Jyoti Soni,[1,2,5] Raiyan Ali,[1,5] Partha Chattopadhyay,[1,2,4,5] Priti Devi,[1,2] Priyanka Mehta,[1,2] Aanchal Yadav,[1,2] Akshat Jaiswal,[1] Bansidhar Tarai,[3] Sandeep Budhiraja,[3] and Rajesh Pandey[1,2,6,*]

[1]Division of Immunology and Infectious Disease Biology, INtegrative GENomics of HOst-PathogEn (INGEN-HOPE) Laboratory, CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Mall Road, Delhi 110007, India

[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

[3]Max Super Speciality Hospital (A Unit of Devki Devi Foundation), Max Healthcare, Delhi 110017, India

[4]Present address: Victor Phillip Dahdaleh Heart and Lung Research Institute, Department of Medicine, University of Cambridge, Cambridge CB2 0BB, UK

[5]These authors contributed equally

[6]Lead contact

*Correspondence: rajeshp@igib.in

https://doi.org/10.1016/j.isci.2026.116344

SUMMARY

Host-microbe dynamics during SARS-CoV-2 Omicron variant infection and recovery remain poorly understood, particularly regarding intracellular microbial communities in immune cells. We performed single-cell RNA sequencing of 191,417 peripheral blood mononuclear cells (PBMCs) from 57 individuals (9 healthy, 24 Omicron-infected, 16 recently recovered, 8 long-recovered) using the BD Rhapsody platform. Microbial signatures identified with PathogenTrack revealed elevated alpha diversity in acutely infected and recently recovered groups, driven by opportunistic pathogens such as *Escherichia coli* and *Providentia stuartii.* In contrast, healthy and long-recovered individuals displayed commensal-dominated profiles, notably *Streptomyces sviceus*, indicative of restored balance. Functional analysis showed persistent microbial signatures, including *Clostridium botulinum's rpoB* in B cells of long-recovered individuals, broad expression of *E. coli* stress gene *sgrR*, and *Mycoplasma hyopneumoniae's rpsO* across 12 immune subsets. Notably, *S. sviceus dnaK* was exclusive to healthy monocytes. These results suggest enduring intracellular microbial influences, implicating them in long-COVID pathophysiology with potential therapeutic relevance.

INTRODUCTION

During viral infections, a dynamic battle for supremacy occurs between the invading virus and the host immune system. Host, viral, and immune factors in resistance against invading pathogens have been thoroughly investigated; however, the influence of the host microbiota (specifically functionally active microbes)—often overlooked—also plays a critical role.[1] Historical outbreaks, such as the influenza pandemics (1918, 1956, 1967, and 2009) and the recent COVID-19 pandemic, have demonstrated that secondary microbial infections significantly contribute to the disease severity and mortality.[2,3] SARS-CoV-2 infection has been shown to disrupt the host microbiome, with microbial disproportion persisting for at least six months post-infection.[4–7] The increased abundance of pathogenic bacterial species can promote co-infections, which are well documented to significantly increase morbidity and mortality in patients with COVID-19.[8–10] Bulk RNA-seq studies have identified *Acinetobacter baumannii* and *Pseudomonas aeruginosa* as potential biomarkers associated with disease severity, particularly in respiratory infections and bloodstream infections.[11] These disruptions correlate with the greater disease severity, elevated inflammatory markers, and an increased risk of long-term post-COVID-19 complications.[6,7,12]

Among these, intracellular microbes are particularly concerning because of their ability to persist and replicate within the host cells, making them formidable contributors to the disease progression. These hidden microbial communities may influence immune responses, inflammation, and the long-term effects of infection; however, their roles in COVID-19 pathogenesis remain largely unexplored.[13] For instance, *Bifidobacteria* and *Akkermansia muciniphila* activate STING signaling in the dendritic cells, enhancing interferon production and NK cell activity to support anti-cancer immunity.[14,15] In the context of infectious diseases, *Staphylococcus aureus* and *Klebsiella* spp. have been identified in 42.2% of ICU patients with COVID-19, which worsens outcomes by driving hyper-cytokinemia and impairing the bactericidal functions of the neutrophils and monocytes.[16] However, the role of intracellular microbes, especially those residing within the immune cells and possibly influencing their function, remains poorly understood.

Recent advancements in the single-cell RNA sequencing (scRNA-seq) have revolutionized our ability to dissect cellular heterogeneity and host-microbe interactions at a high resolution. Although scRNA-seq is primarily designed to capture host transcriptomes, a significant proportion of sequencing reads often remain unmapped to the host genome. Emerging evidence suggests that these unmapped reads frequently correspond to transcriptionally active microbes (TAMs), closely associated with the host tissues, providing an approach to study microbial activity.[17–20] By analyzing these reads, recent studies have uncovered microbial signatures that elucidate the interplay between the microbial communities and host immunity in infectious diseases, such as COVID-19. Our prior findings identified cell-type-specific enrichment of TAMs, including *E. canis, B. aphidicola, S. clavuligerus, P. tolaasii,* and *E. albertii,* in the memory B cells, naive B cells, neutrophils, naive T cells, regulatory T cells (Tregs), and platelets.[18] Leveraging these dual transcriptomic data, we established a comprehensive framework for dissecting host-microbe interactions, advancing our understanding of disease pathogenesis, and guiding the development of potential targeted therapeutic interventions.[17]

In this study, we utilized scRNA-seq to profile intracellular microbial communities across four groups: healthy individuals, COVID-19 Omicron variant-infected patients, recently recovered individuals, and individuals sampled three months' post-recovery (longitudinally recovered). We aimed to identify microbial signatures associated with immune resilience, disease severity, and prolonged immune dysregulation. Building on our previous findings[18] in SARS-CoV-2-infected and recovered individuals, we investigated whether alterations in the intracellular microbial composition drive persistent dysbiosis or gradually return to a healthy baseline. By elucidating the dynamics of host-microbe interactions, we aim to uncover the single-cell genomics-based molecular mechanisms underlying SARS-CoV-2 pathogenesis and recovery, facilitating the development of precision therapeutics for achieving immunological homeostasis.

RESULTS

Single-cell RNA-seq reveals immune landscape dynamics in SARS-CoV-2 infection and long recovered groups

To investigate the microbial communities within the circulating immune cells in the context of SARS-CoV-2 infection and recovery, we collected blood samples from 57 participants (9 healthy, 24 infected, and 24 recovered). Furthermore, the recovered group included 16 participants sampled one month after recovery (referred to as the recovered group) and a subset of 8 individuals who were sampled again at three months' post-recovery, referred to as the longitudinal-recovered group. For consistency, the "longitudinal recovered" group is hereafter referred to as "long-recovered" throughout this manuscript. Following the classification of participants into groups, we assessed the basic demographic characteristics. The median ages of the healthy, infected, recovered, and long-recovered groups were 28, 66.5, 35, and 27.5 years, respectively (Table S1). The median age of the infected individuals was significantly higher than that of the other three groups (Figure S1). While age is often considered an important variable, it cannot solely explain the observed differences, as other groups, such as healthy, recovered, and long-recovered individuals, shared similar age ranges but displayed distinct microbial compositions. To assess the potential influence of age, we performed a regression analysis between age and the differentially abundant microbes, which revealed only a weak correlation (Figure S2 and Table S2). This indicates that factors beyond age, such as infection status and immune response, are likely to drive the observed microbial shifts. The sex distribution (male-to-female ratio) was comparable across the groups and did not differ significantly. Clinical data indicated that SpO_2 levels were available for infected individuals, with a median of 95 (IQR: 96.5–92.5 = 4.25). Similarly, C-reactive protein (CRP), a marker of inflammation, was reported for 11 patients, with a median value of 82.66 mg/L and a wide interquartile range of 187.33–6.13 = 181.2 mg/L, reflecting substantial variability in the inflammatory responses (Table S1).

Immune landscape across COVID-19 infection and recovery phases

We used the BD Rhapsody Express single-cell analysis system to perform scRNA-seq, as shown in Figure 1A. Following initial processing, the dataset contained 191,417 raw cells, which were subsequently reduced to 181,309 cells after deduplication. Quality control filtering based on a minimum nGene count of >60, nUMI count of >100, <3500, and low cell counts (<500), resulted in a final dataset of 153,075 high-quality cells, including healthy (44,278), infected (39,995), recovered (27,452), and long recovered (41,350) cells from 44 samples (9 healthy, 14 infected, 13 recovered, and 8 long recovered) (Figure 1B). At the QC step, 13 samples (10 infected and 3 recovered) were excluded due to the low cell counts. We further performed clustering of the cell populations using the Louvain algorithm. With a variable resolution of 0.3–0.6, selected based on the cluster stability and biological relevance of the marker genes, we identified 13 distinct cell clusters. These cell clusters were annotated manually into major immune cell types (Figure 1C): plasma cells (16.31%), memory T cells (15.66%), B cells (9.66%), naive T cells (9.56%), activated T cells (6.91%), cytotoxic T cells (6.45%), monocytes (6.17%), NK cells (6.06%), Tregs (5.66%), dendritic cells (3.87%), NK T cells (1.99%), erythrocytes (7.36%), and platelets (4.34%), were included in our dataset. To validate these transcriptomic-based annotations, we integrated AbSeq data for surface protein expression. The cell-type-specific expression of the surface markers shown in Figure S3 illustrates the markers used for cell-type annotation across different groups and confirms the identity of the clusters identified via intracellular markers. In the infected group, we observed a notable expansion of plasma cells, Tregs, dendritic cells, and platelets, along with a reduction in NK cells, monocytes, naive and activated T cells, and B cells. This finding reflects the immune dysregulation in the infected individuals. An increase in the antibody-secreting plasma cells, coupled with the Tregs, known for calming the cytokine storm in COVID-19, indicates an immune system trending toward resolution and swift recovery.[21] However, the depletion of major T cell subsets, B cells, monocytes, and NK cells shows a dysregulated immune response during COVID-19 infection.

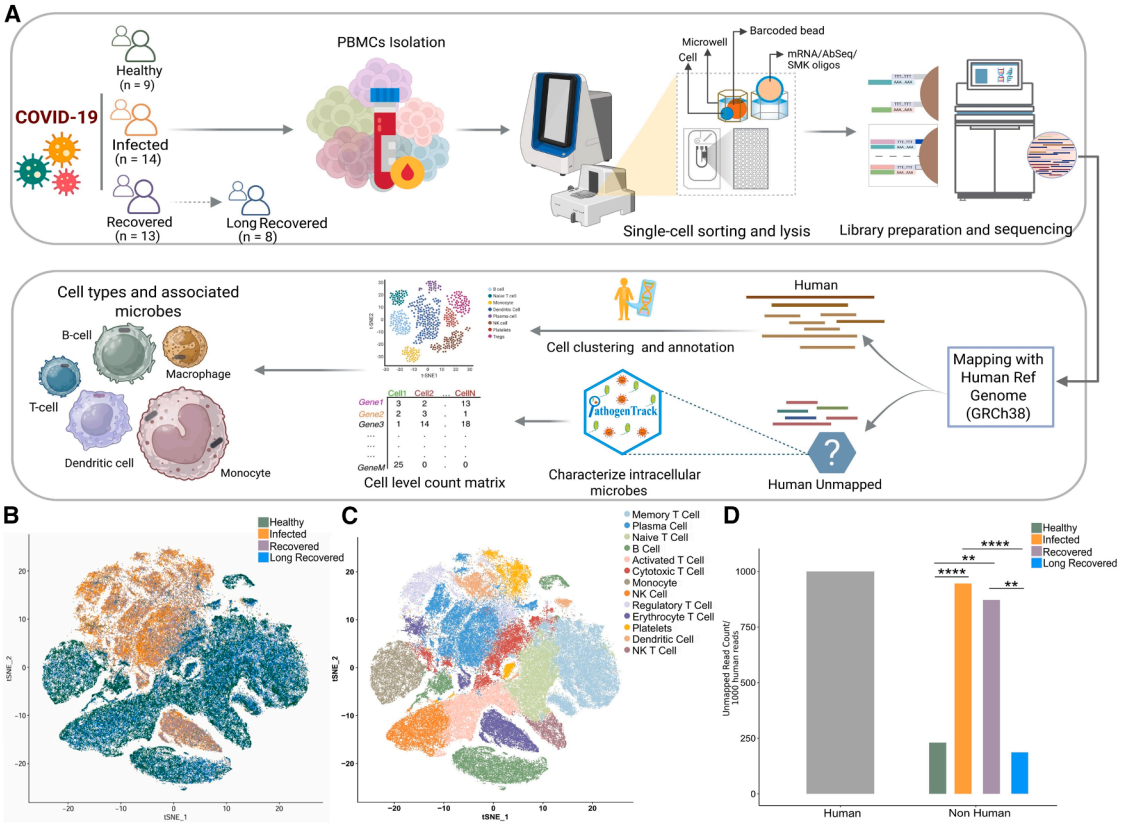


Figure 1. Summary of the study groups, experimental design, and metagenomic analyses

(A) Flowchart depicts the overall experimental and analytical design of the study. PBMCs were isolated from the blood of healthy, SARS-CoV-2-infected, recovered, and long-recovered individuals. Single-cell sorting was performed using the BD Rhapsody Cartridge, and library preparation and subsequent analysis were performed to identify the intracellular microbes.

(B) The t-SNE plot represents the distribution of 153,075 total cells after QC among the healthy (44,278), SARS-CoV-2 infected (39,995), recovered (27,452), and the long-recovered (41,350) groups.

(C) t-SNE visualization of the identified 13 major cell type clusters.

(D) The bar plot represents the unmapped read counts per thousand human reads across the disease severity groups. Bar height indicates the group means. Wilcoxon test has been used to calculate pairwise statistical significance (p adjusted value <0.05) where significance values is denoted as *, where * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$, and **** indicates $p \leq 0.0001$.

Shifts in human reads across infection and recovery

Approximately 3 billion reads were generated from 44 samples, of which ~2.12 billion aligned with the human genome (GRCh38), whereas the remaining ~0.97 billion did not map and were referred to as "unmapped" reads. The distribution of human and unmapped reads differed notably across the sample groups. Healthy and long-recovered samples showed comparable read distributions, with approximately 85% human reads (84.8% and 85.5%, respectively) and 15.2% and 14.5% unmapped reads, respectively. In contrast, the recovered samples displayed a noticeable reduction in human reads to 58.9%, with unmapped reads comprising 41.1%. This trend was more pronounced in the infected samples, where human reads decreased to 55%, and unmapped reads increased to 45%. The striking increase in unmapped reads in the infected and recently recovered groups prompted us to investigate their composition to uncover potential insights into the infection dynamics (Figure S4). We further quantified unmapped reads per 1,000 human reads from the unmapped fraction across the groups. This analysis revealed statistically significant (Wilcoxon test; $p < 0.05$) differences between the multiple group pairs: healthy vs. infected, infected vs. recovered, recovered vs. healthy, and recovered vs. long recovered. Notably, no significant difference was observed between healthy and long-recovered individuals (Figure 1D). To determine whether non-host transcriptional signals are unique to COVID-19 or represent a broader feature of systemic infection, we conducted a comparative analysis using publicly available PBMC scRNA-seq datasets from dengue and COVID-19 cohorts (Figure S5A).[22] Pseudobulk analysis showed a modest increase in unmapped reads in dengue-infected samples relative to healthy controls (~4.3% vs. ~3.7%). Conversely, COVID-19 samples demonstrated significantly elevated proportions of unmapped reads (~10–22%), signifying increased variability in the non-host transcriptional signals. Furthermore, phylum-level profiling in the dengue dataset revealed infection-associated shifts in microbial composition compared to controls (Figure S5B). Collectively, these results suggest that non-host (putative microbial) transcripts

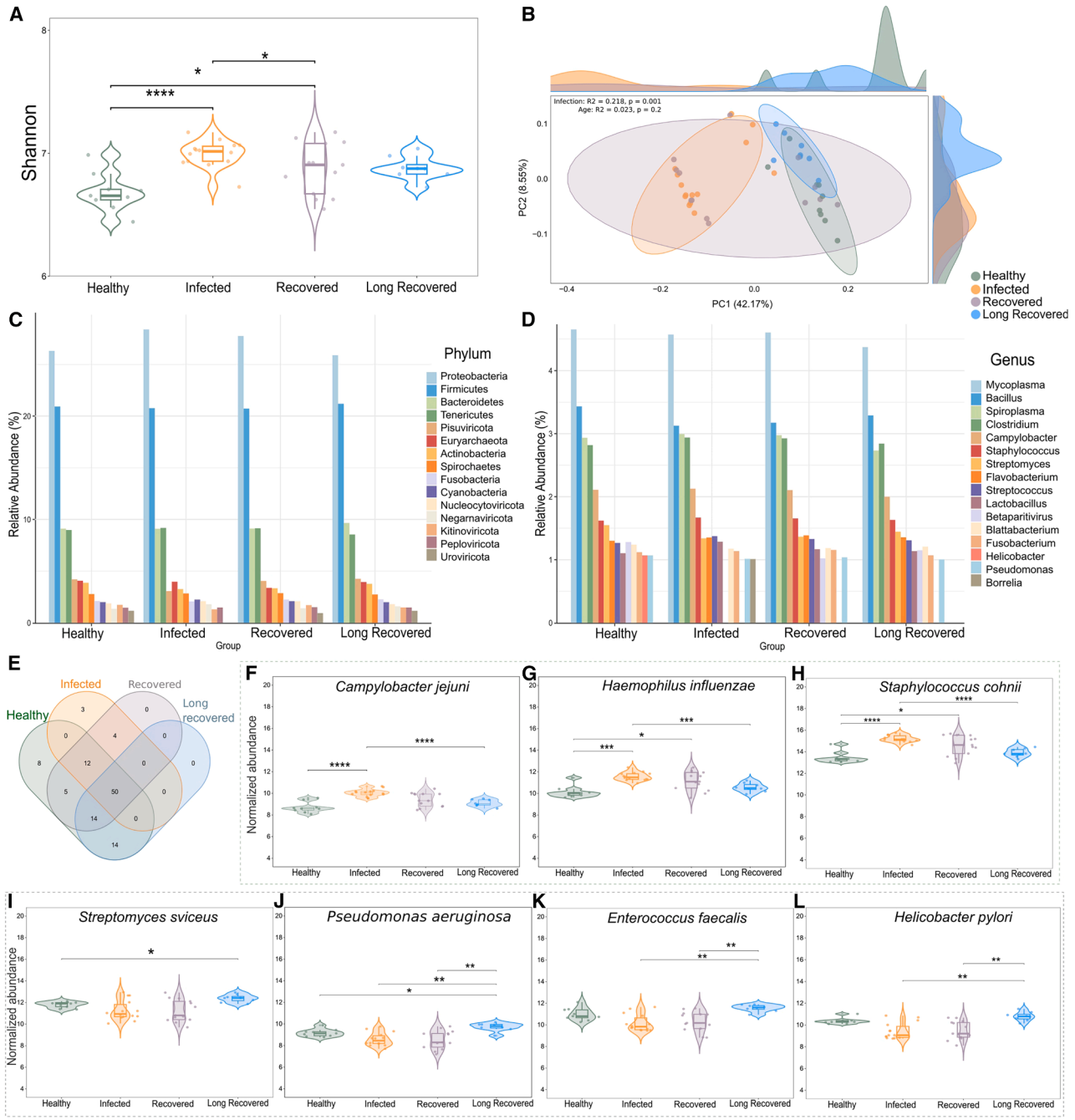


Figure 2. Microbial abundance, diversity, and taxonomic composition across healthy, infected, recovered, and long-recovered individuals

(A) The Shannon index, used to assess alpha diversity, is presented using combined violin-box plots to illustrate the within-sample microbial diversity across the study groups. Each point represents a sample; the central line shows the median, the box indicates the interquartile range (IQR), and the upper and lower whiskers correspond to the 75th (Q3) and 25th (Q1) percentiles, respectively. The violin outlines the distribution. Pairwise Wilcoxon tests were used to assess statistically significant differences in alpha diversity between groups.

(B) Principal coordinates analysis (PCoA) based on beta diversity metrics reveals distinct microbial community structures among healthy, infected, recovered, and long-recovered individuals. Each point represents a sample; ellipses denote 95% confidence intervals around group centroids, and the density overlay shows each group's distribution. Group differences were tested using multivariate PERMANOVA (adonis2), incorporating infection status and age, which identified infection status as the primary driver of variation ($R^2 = 0.218$, $p = 0.001$), while age had a smaller, non-significant effect ($R^2 = 0.023$, $p = 0.21$). Bar plots depict the relative abundance of microbial taxa at the phylum (C) and genus (D) levels, including only those with a mean relative abundance >1%.

(E) Venn diagram demonstrates the number of microbial species (relative abundance >0.2%) among the comparison groups, highlighting both overlapping and unique taxa.

(legend continued on next page)

are a common feature of systemic infection, with disease-specific differences in their abundance and taxonomic composition. This insight led us to further understand the dynamics of the unmapped fraction, aiming to determine whether intracellular microbial signatures in healthy, infected, recovered, and long-recovered individuals reflect distinct ecological states or transitional points along the infection-recovery continuum.

Dual microbial signatures characterize the transition from infection to long recovery and a healthy state

Compelled by the distinct differential distribution of the unmapped reads across our groups, we conducted non-canonical *meta*-transcriptomic analysis of the unmapped reads using the Kraken2 algorithm within the PathogenTrack pipeline, enabling precise identification of the intracellular microbes at single-cell resolution.[20] This analysis yielded 18,822,578 (~18.8M) high-quality, "human-unmapped" reads, which were used as candidate microbial reads for the downstream analysis. On average, 427,786 microbial reads were detected per sample (Table S3). We observed significantly higher microbial reads in infected and recovered groups, compared to the healthy and long-term recovered. The distribution pattern of microbial reads closely mirrored that of the total unmapped reads across the groups (Figures S6 and S7). This concordance suggests that the observed microbial dynamics are largely representative of the overall unmapped read distribution, supporting the reliability of the microbial read-based comparisons. To eliminate the possibility of any reagent contamination, we also included a negative control and proceeded with the random priming and extension steps in the library preparation; however, we did not yield any significant concentration of the library to perform further steps (Figure S8). To assess microbial abundance and trends, we initially analyzed the combined scRNA-seq data from each group. We observed a substantial number of microbial reads mapping to bacteria, viruses, and archaea across all four healthy, infected, recovered, and long recovered.

We next assessed both within-group (alpha) and between-group (beta) microbial diversity to explore community structure across the conditions. Alpha diversity, measured using the Shannon index, was significantly higher in the infected group compared to both the healthy ($p < 0.001$) and recovered individuals ($p < 0.05$, Wilcoxon test). The recovered group also showed significantly higher diversity than the healthy group. In contrast, the long-recovered group did not show significant differences from any other group and closely resembled the healthy profile (Figure 2A), aligning with the findings from the beta diversity analysis, which revealed significant differences in the microbial composition between the groups ($p < 0.01$, PERMANOVA). Principal coordinate analysis (PCoA) demonstrated distinct clustering among the four groups, with PC1 and PC2 accounting for 42.17% and 8.55% of the total variance, respectively (Figure 2B). Notably, the healthy and long recovered groups showed overlapping ellipses, indicating a return journey toward the baseline microbial composition over time. In contrast, infected individuals maintained a distinct microbial signature with a partial overlap with the recovery group. These patterns suggest that intracellular microbial diversity progressively increases during infection and gradually normalizes during recovery, ultimately resembling the healthy profile in long recovered individuals. COVID-19 is characterized by substantial heterogeneity. To assess whether microbial profiles vary with disease severity, the cohort was stratified based on oxygen saturation ($SpO_2 \geq 94\%$: mild; $< 94\%$: severe). Comparative analysis of read distribution, alpha diversity, and beta diversity revealed no statistically significant differences between mild and severe groups (Figure S9). These findings suggest that systemic microbial signatures in PBMCs do not exhibit clear severity-associated stratification.

Taxonomic composition across comparison groups

Building on the observed variations in the microbial community composition from our diversity analysis, we next quantified the relative abundance of taxa to discern the balance between dominant and low-abundance microbes across the groups. We filtered out phyla and genera with relative abundances below 1% to reduce noise and minimize the impact of low-abundance taxa, focusing instead on the potentially biologically relevant microbes. Across all the groups, the microbial community was predominantly composed of Proteobacteria (26.31%, 28.39%, 27.75%, and 25.89%) and Firmicutes (20.93%, 20.77%, 20.72%, and 21.19%) in healthy, infected, recovered, and long recovered groups, respectively. Additional enriched phyla, including Bacteroidetes, Tenericutes, Pisuviricota, and Euryarchaeota, were detected in the samples from the healthy, infected, recovered, and long-recovered individuals. These phyla have been previously reported as abundant constituents of the blood microbiome in both healthy individuals and those with metabolic or inflammatory disorders (Table S4). At the genus level, relative abundance profiles revealed a broadly conserved microbial architecture across all the clinical groups, with dominant genera such as *Mycoplasma* (4.65%, 4.57%, 4.60%, 4.37%), *Bacillus* (3.43%, 3.12%, 3.17%, 3.28%), *Spiroplasma* (2.93%, 2.99%, 2.97%, 2.84%), *Clostridium* (2.81%, 2.93%, 2.92%, 2.84%), and *Campylobacter* (2.11%, 2.13%, 2.10%, 2.0%) showing relatively stable distributions across healthy, infected, recovered, and long recovered individuals, respectively (Figures 2C and 2D). Despite this overall similarity, distinct group-associated patterns were observed. For example, *Helicobacter*, a genus commonly linked to gastrointestinal colonization and immune modulation, was found to be selectively abundant in the healthy group, suggesting a possible role in maintaining microbial homeostasis. In contrast, *Borrelia*, typically associated with vector-borne infections, has >1% relative abundance only in the infected individuals, potentially reflecting opportunistic

(F–H) Violin with boxplots represents taxa following pattern 1: infected > recovered > long-recovered > healthy, indicating a progressive reduction in abundance across clinical states.

(I–L) Violin plots show pattern 2: long-recovered > healthy > recovered/infected, suggesting the selective enrichment of specific taxa in the long-recovered group. Pairwise Wilcoxon test has been used to assess statistical significance. Significance value is denoted as *, where * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$, and **** indicates $p \leq 0.0001$.

expansion during immune perturbation. *Streptomyces*, a genus known for its antimicrobial metabolite production and commensal behavior, exhibited higher abundance in healthy (1.55%) and long recovered (1.44%) individuals compared to the infected (1.37%) and recovered (1.39%) groups, potentially contributing to microbial resilience. To further dissect these group-specific trends, we advanced to species-level profiling to identify taxa more precisely associated with the infection and recovery states.

Shared and differentially abundant microbial species

These genus-level findings prompted us to explore deeper taxonomic resolution. As a result, in total, we identified 3784 transcriptionally active microbial species, including 3072 bacterial species, 566 viruses, and 146 archaea (Table S5). To increase the reliability at the species-level, a stringent filtering criterion was applied, excluding taxa with relative abundances of less than 0.2% with the prevalence in at least 90% of the samples in the respective group. This filtering step resulted in a total of 110 microbial species. Among them, eight species— *Staphylococcus simulans, Fastidiosipila sanguinis, Thermoanaerobacter kivui, Treponema pedis, Pandoravirus dulcis, Arcobacter* sp. *L, Parachlamydia acanthamoebae,* and *Mycoplasma cynos*—were exclusively detected in the healthy group. In contrast, three species—*Serratia plymuthica, Citrobacter* sp., and *Spiroplasma phoeniceum*, were unique to the infected patients. Notably, no species were uniquely associated with either the recovered or long-recovered groups. Interestingly, 50 microbial species were consistently present across all four groups (Figure 2E). We sought to reduce sample-specific noise and emphasize conserved microbial patterns that change in abundance, rather than presence or absence, as most often infection and recovery progress by focusing on these overlapping species.

Although these species were commonly present, differential abundance analysis revealed that 47 species exhibited statistically significant variations in at least one group-wise comparison. Notably, two distinct patterns of these 47 species based on the relative abundance were observed: (1) a subset of species displayed high abundance in the infected group, followed by the recovered, long-recovered, and healthy individuals; and (2) another subset showed the highest abundance in the long-recovered group, followed by the healthy, recovered, and infected groups. Species such as *Escherichia albertii*, *Priestia megaterium, Campylobacter jejuni*, *Mycoplasma ruminantium,* and *Haemophilus influenzae* were significantly more abundant in the infected patients, followed by the recovered group—pattern 1 (Figures 2F–2H and Data S1). Intriguingly, these microbes are known to be opportunistic and pathogenic. For example, *E. albertii* is known to cause bacteremia, and *B. megaterium* is less virulent but has been reported to cause infections in immunocompromised human hosts. Similarly, *C. jejuni* causes gastroenteritis and *H. influenzae* causes sinusitis and conjunctivitis. The increase in opportunistic and pathogenic species in the infected and recovered groups highlights the suboptimal immune response or the ability of these microbes to evade the immune response in the presence of a primary infection. In contrast, species typically considered part of the normal microflora—*Streptococcus sviceus, Enterococcus faecalis, Clostridium perfringens,* and *Pseudomonas aeruginosa*—were more prevalent in the long-recovered group, followed by healthy—pattern 2, suggesting a microbial shift toward the restoration and re-establishment of a healthy microbiome during longer periods of recovery (Figures 2I–2L).

Increased opportunistic species in the immune cells of infected and recovered groups

Building on the observed distinct species patterns across the groups, we next assessed their distribution across immune cell types to understand whether they play a role in modulating the host's immune function and creating an immunosuppressive environment. We analyzed unmapped reads using PathogenTrack, which generated a cell-level count matrix. Across all the QC-passed cells, 20.93% contained microbial reads. The distribution of microbe-containing cells, calculated as the proportion of cells with microbial reads out of the total number of cells in each group, was as follows: healthy—7,605 cells (17.1%), infected—8,612 cells (21.5%), recovered—6,292 cells (22.9%), and long-recovered—9,538 cells (23.01%) (Figure 3A). While these values reflect a higher proportion of microbe-harboring cells in the long-recovered, infected, and recovered groups than in the healthy group, it is worth noting that the microbial read content within those cells revealed a different pattern. The average proportion of microbial reads per cell per condition was highest in the recovered group (32.45%), followed by the infected (28.3%), long-recovered (8.94%), and healthy (6.9%) groups (Figure S10). This observation aligns with our earlier findings of abundant alpha diversity in the infected and recovered groups compared to the healthy and long-recovered groups.

To understand how microbial associations varied across the immune cell types, we first calculated the relative proportion of each immune cell type within the individual groups. As erythrocytes do not contribute to immune function, they were excluded from further analysis. Following this exclusion, the remaining 29,356 cells were further analyzed. We observed that activated T cells, memory T cells, naive T cells, and NK/NKT cells were present at lower proportions in the infected patients. In contrast, the infected patients showed an increased proportion of antibody-secreting plasma cells, dendritic cells, platelets, and Tregs compared to the healthy, recovered, and long-recovered groups (Figure 3B). To further explore the potential microbial associations in these cell types, we quantified the proportion of microbe-containing cells within each immune cell type by calculating the number of microbe-positive cells relative to the total number of cells of that type. Counterintuitively, despite their reduced overall abundance in the infected group, these T cell subsets, along with NK and NKT cells, exhibited a higher proportion of microbial reads than other cell types, indicating a preferential enrichment of intracellular microbes within these immune populations during infection (Table S6). Similarly, B cells and cytotoxic T cells also demonstrated elevated microbial associations in infected, recovered, and long-recovered individuals compared to the healthy controls, with slightly higher levels in the long-recovered group compared to the recovered (e.g., B cells: 23.8% vs. 22.7%). Strikingly, dendritic cells and monocytes, both key antigen-presenting cells, contained fewer

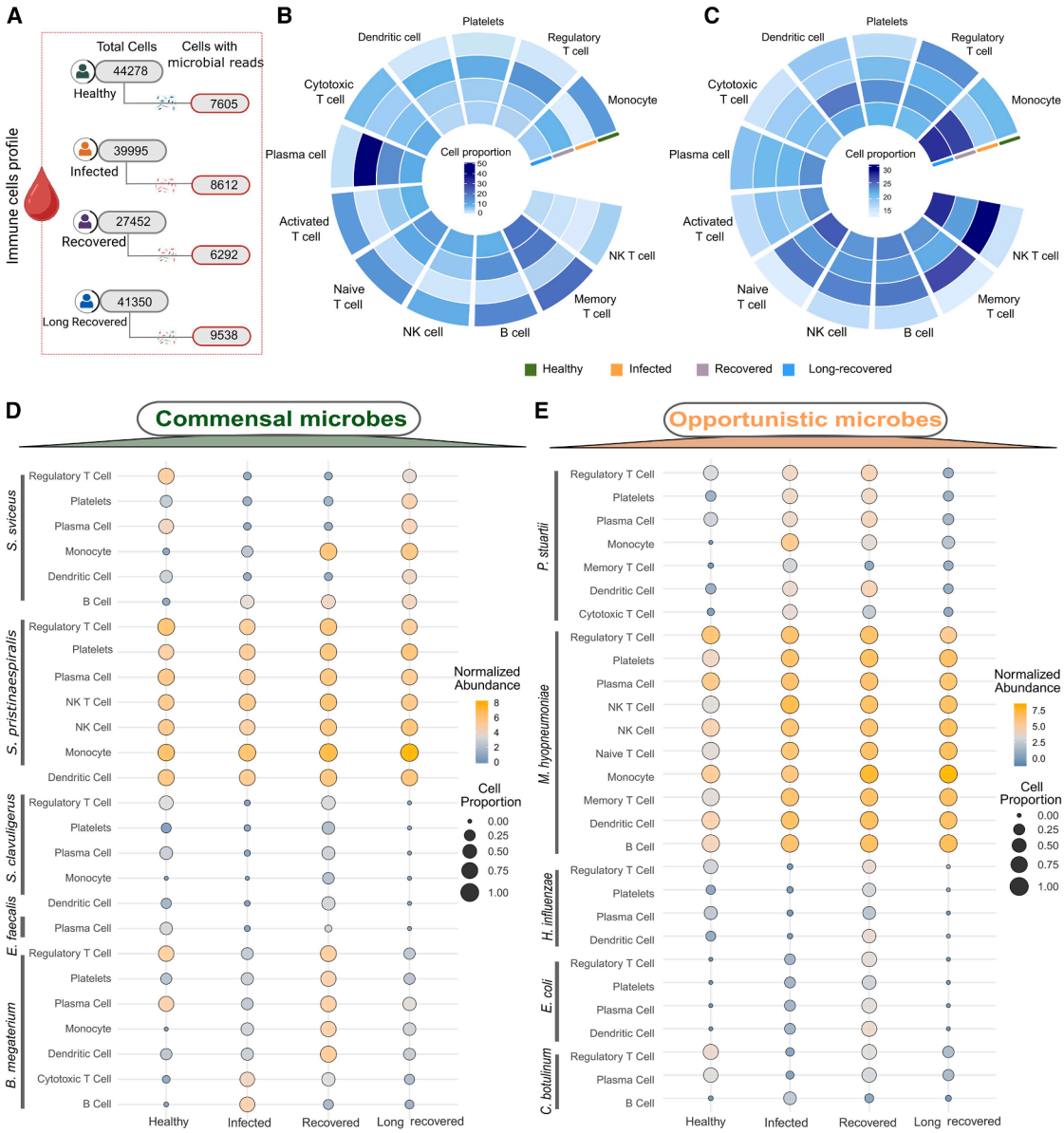


Figure 3. Immune cell distribution and cell-type-specific microbial associations across clinical groups

(A) Overview of total cells and cells containing microbial reads across the healthy, infected, recovered, and long-recovered groups.

(B) Circular heatmap shows the relative proportion of immune cell types within each group.

(C) Proportion of microbe-positive (M^+) cells across immune cell populations and clinical conditions.

(D and E) Dot plot shows the cell-type-specific distribution and normalized abundance of intracellular microbial transcripts across the study groups. Dot size represents the proportion of cells harboring microbial reads, and color intensity indicates normalized abundance.

(D) Commensal microbial signatures (e.g., *Streptomyces spp., Enterococcus faecalis,* and *Bacillus megaterium*) are enriched in the healthy and long-recovered individuals across multiple immune cell types.

(E) Opportunistic microbial signatures (e.g., *Providencia stuartii, Mycoplasma hyopneumoniae, Haemophilus influenzae, Escherichia coli,* and *Clostridium botulinum*) are predominantly enriched in infected and recovered groups, indicating disease-associated microbial shifts.

microbial reads in the infected individuals compared to the healthy, recovered, and long-recovered groups. Notably, platelets with microbial reads were specifically found to be significantly increased in the long-recovered individuals compared to both the infected and recovered groups (e.g., plasma cells: 23.2% vs. 21.1%). In contrast, no significant differences were observed in the proportion of microbial-containing Tregs across the groups (Figure 3C). Tregs are essential for maintaining immunologic tolerance and regulating excessive immune responses. The stable microbial load within the Tregs across disease states may suggest that these cells preserve their homeostatic role and are less susceptible to microbial infiltration, even under

conditions of immune perturbation. These findings indicate that immune surveillance is disrupted during infection and persists beyond the acute phase. This decrease in the number of innate and adaptive immune cells during infection points toward immunosuppression and potentially creates a window of opportunity for various microbes to invade and persist within the immune cells, highlighting a critical vulnerability in the host defense.

Microbial dysbiosis in infected and recovered individuals

To explore the types of microbes present in each cell type, we applied a stringent cutoff on the cell-level count matrix; a microbe was considered valid if it had at least 10 microbial reads and was expressed in at least 10% of the cells of each sample. Applying this filter retained 24 unique species, including 21 bacterial species and three viruses. Furthermore, we stratified these 21 bacteria into commensal, opportunistic, and others based on their function across the cell types and clinical groups (Figures 3D and 3E). Notably, commensal microbes, such as *Streptococcus sviceus,* were enriched in monocytes, plasma cells, dendritic cells, and platelets, with persistent presence in the monocytes and Tregs during long recovery, suggesting a role in chronic immune modulation. *S. pristinaespiralis* exhibited stable colonization across all the groups, indicating a potentially regulatory or tolerogenic interaction. While *S. clavuligerus* showed a phase-dependent profile, with loss during infection and long recovery, there was recolonization of dendritic cells and platelets during the recovery phase. In contrast, *Bacillus megaterium* was broadly elevated during infection and persisted in the memory T cells, Tregs, and plasma cells, implying an immune imprinting that persisted for a long term. Conversely, opportunistic pathogens displayed an inverse trend, predominating in the infected and recovered groups, highlighting a potential microbial shift associated with the infection. *M. hyopneumoniae* is abundant in most of the cells in infected, recovered, and long-recovered groups as compared to the healthy, suggesting increased access to the intracellular environment, post-COVID-19. In the ''others'' category, we observed a higher abundance of *Nostocales cyanobacterium, Mycoplasma bovigenitalium, Mycoplasma anatis, Caldicellulosiruptor saccharolyticus,* and *Buchnera aphidicola* (Figure S11). Additionally, three viral species, *Bacillus phage Stitch*, *Hop trefoil cryptic virus 2*, and *Piscine myocarditis-like virus*, were also identified. The extent of possible virus transmission between humans and other kingdoms remains largely unknown. Further investigation is required to elucidate their potential role.

Viral signatures and host-virus associations across the clinical groups

We also analyzed the viral component of the *meta*-transcriptomic dataset. Viral read abundance was normalized using cumulative sum scaling (CSS), and prevalence differences across the clinical groups were assessed using pairwise Wilcoxon tests. *Piscine myocarditis-like virus* showed a significant increase in abundance in infected individuals compared to the healthy, recovered, and long-recovered groups ($p < 0.0001$). In contrast, *Bacillus phage Stitch* and *Hop trefoil cryptic virus 2* were detected exclusively in the infected and recovered individuals but were absent in healthy and long-recovered groups, indicating a condition-associated emergence pattern (Figure S12). To explore host-virus relationships, we analyzed the co-presence of *Bacillus phage Stitch* with its putative host, *Bacillus megaterium,* at the single-cell resolution. We found 423 cells that had both phage and host transcripts, which supports the idea that they are in the same cellular microenvironment (Figure S13). Despite this co-occurrence, correlation analysis revealed only a weak, non-significant association between the phage and host abundance (Spearman $\rho = 0.088$, $p = 0.071$), suggesting non-linear or transient interactions consistent with the phage replication dynamics. Together, these findings highlight distinct viral signatures associated with the infection and recovery, as well as evidence of cell-level host-virus co-localization.

Transcriptionally active microbes (TAMs) persist via virulent and resistance pathways in immune microenvironments

To investigate the potential role of these bacterial species in the immune cells, we analyzed microbial genes originating from the species that were significantly differentially abundant in our dataset. This analysis yielded 201 genes derived from 18 bacterial species (Figure 4A). After excluding a substantial proportion of hypothetical protein-coding genes with undefined functions, we retained 30 unique genes from 8 microbial species. These genes were categorized into eight functional groups based on their annotated roles: cellular processes, metabolism, stress response, antimicrobial resistance (AMR), transcription, virulence, replication, and translation (Figure 4B and Table 1).

To further determine the cellular origin of the detected microbial genes, we traced their expression in the specific immune cell types using associated cell barcodes. Nine genes from four bacterial species were confidently mapped to distinct immune cell populations (Figure 4C). For *C. botulinum*, the transcription-related gene *rpoB* was expressed in the B cells of long-recovered individuals, whereas the AMR-associated gene *abc-f* was detected in the memory T cells of both the recovered and long-recovered groups. *E. coli* showed a broad distribution of its stress response gene *sgrR* across the memory T cells, plasma cells, B cells, platelets, and naive T cells in all the clinical groups (healthy, infected, recovered, and long-recovered). Additionally, its transporter gene *ycjO* was found in B cells, naive T cells, NK cells, Tregs, cytotoxic T cells, and monocytes in healthy, infected, and long-recovered individuals. *M. hyopneumoniae* exhibited the most extensive expression profile, with its translation-associated gene *rpsO* present in all 12 immune cell types across all clinical groups. The adhesin gene *p216* was detected in monocytes, plasma cells, platelets, B cells, memory T cells, and NK cells across all the groups, whereas the DNA repair gene *rpoE* was found in the dendritic cells, plasma cells, and Tregs in infected and recovered individuals. Finally, the stress-response gene *dnaK* from *S. sviceus* was specifically mapped to monocytes in the healthy individuals.

To investigate the intracellular relationship between the microbial activity and host response, we performed an *in silico* correlation analysis at the single-cell level, specifically focusing on the *rpsO* gene (encoding the 30S ribosomal protein S15) as a proxy for bacterial metabolic activity and protein synthesis. We first

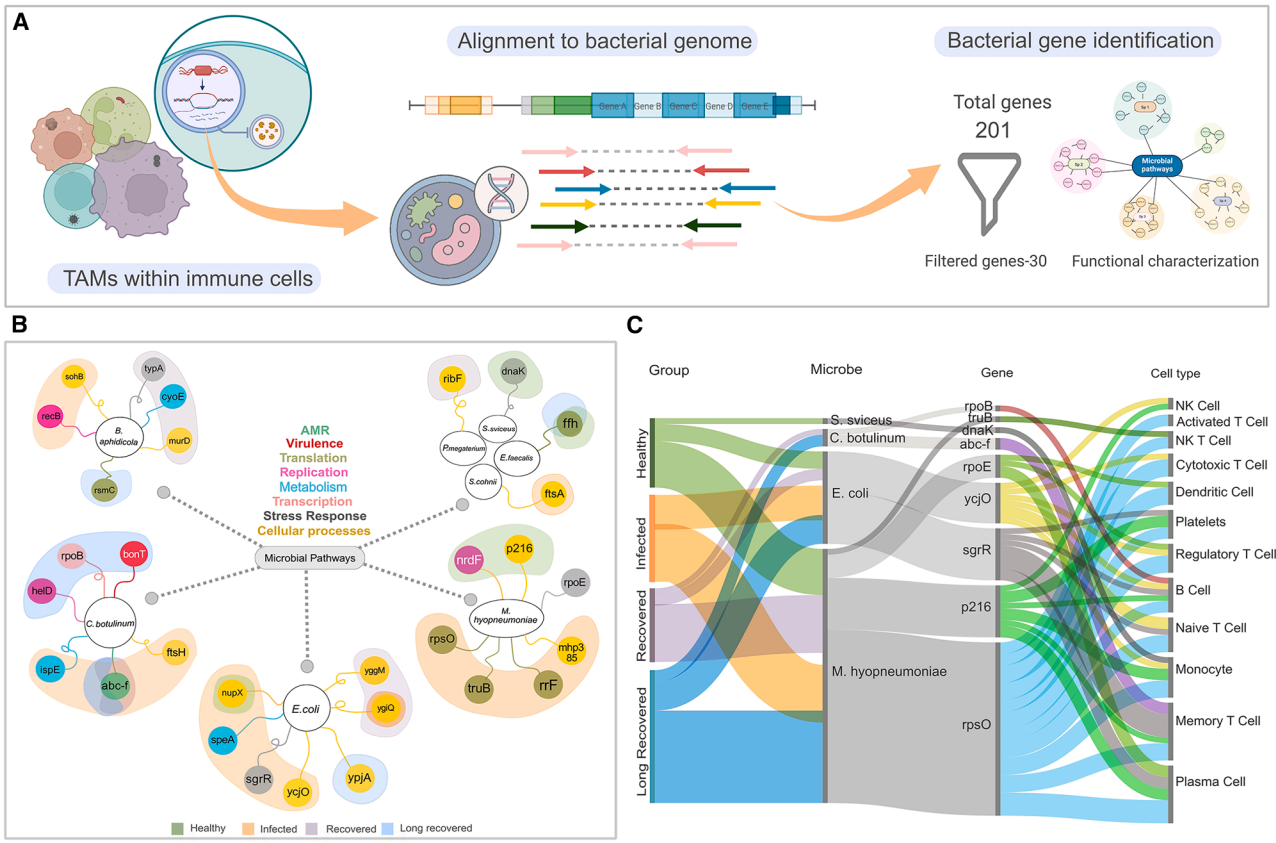


Figure 4. Functional characterization and cell-type distribution of microbial gene expression

(A) Schematic overview of the workflow for identifying transcriptionally active microbial genes (TAMs) within the immune cells, including the alignment of unmapped reads to bacterial reference genomes and downstream functional annotation. A total of 201 microbial genes were identified and filtered for functional analysis.

(B) Network-based functional categorization of microbial genes across the bacterial species. Nodes represent genes grouped by species, and colors indicate functional classes, including antimicrobial resistance (AMR), virulence, translation, replication, metabolism, transcription, stress response, and cellular processes. Background shading denotes the clinical groups (healthy, infected, recovered, and long-recovered).

(C) Sankey diagram illustrates the flow of microbial transcripts from clinical groups to bacterial species, gene identities, and host immune cell types. This highlights the distribution of microbial gene expression across specific immune cell types and conditions.

identified the specific subset of host cells containing detectable *rpsO* transcripts and employed a pseudobulk approach for these cells, filtering for host genes with greater than or equal to 10 reads across samples to ensure a robust signal for the downstream analysis. This gene had a high prevalence across the cells compared to others. Our analysis revealed 144 significant host genes (55 positively correlated and 89 negatively correlated) that directly correlate with bacterial *rpsO* gene expression (Table S7). The correlations point to two distinct host transcriptional shifts: *rpsO* positively correlated host genes are associated with erythrocyte function, heme metabolism, and stress-response pathways, as well as key immune signaling processes. The negative correlation of *rpsO* with host genes revealed the enrichment of pathways related to mRNA splicing, RNA processing, and protein metabolism, as well as immune signaling (e.g., IL-12/JAK-STAT) and cellular stress responses. Furthermore, we performed functional enrichment analysis of the 201 detected microbial genes using InterPro and Gene Ontology (GO) databases, which revealed a significant overrepresentation of domains associated with DNA-binding and restriction-modification (R-M) systems (Figure S14). Notably, Type III N4/N6-methyltransferases and Type I specificity domains were among the most enriched features, indicating that TAMs within the immune microenvironment predominantly engage in genome defense and epigenetic regulation. The enrichment of these domains suggests robust mechanisms for maintaining genomic integrity and modulating gene expression. This is further supported by the presence of stress-response regulators such as *sgrR* and *dnaK,* which likely contribute to microbial persistence under intracellular stress conditions within the host immune cells. These findings suggest that microbes are actively proliferating and engaging in various cellular processes to support their survival and persistence within the immune cells.

KEGG pathway enrichment analysis was also performed; however, no pathways reached statistical significance following Benjamini-Hochberg correction (FDR $p < 0.05$). This lack of enrichment is likely due to the compact and specialized nature of the gene set, which does not encompass sufficient metabolic

Table 1. Microbe-associated genes and their known functions

Microbe name	Gene	Category	Function	Reference
E. coli	*nupX*	cellular processes	nucleoporin facilitates nucleocytoplasmic transport.	Craig et al.[23]
	ygiQ	cellular processes	involved in protein translation.	Ghatak et al.[24]
	yggM	cellular processes	yggM mediates betaine export.	Pushpker et al.[25]
	ypjA	cellular processes	function unclear, possibly membrane transport or adhesion.	Sidorczuk et al.[26]
	speA	metabolism	arginine decarboxylase produces agmatine for polyamine biosynthesis.	He et al.[27]
	ycjO	cellular processes	Ycjo encodes a subunit of an ABC transporter, specifically a sugar transporter.	Mukherjee et al.[28]
	sgrR	stress response	a transcriptional regulator controls sugar transport/metabolism under stress.	Sun and Vanderpool[29]
C. botulinum	*abc-f*	AMR	ABC-F family protein, likely involved in antibiotic resistance via translation control.	Fostier et al.[30]
	ispE	metabolism	enzymes in the MEP pathway synthesize isoprenoid precursors.	Kuwahara et al.[31]
	rpoB	transcription	β-subunit of RNA polymerase, essential for transcription.	Jacobson et al.[32]
	bonT	virulence	botulinum neurotoxin causes botulism.	Rawson et al.[33]
	ftsH	cellular processes	membrane-bound protease regulates cell division and is involved in stress response.	Liang et al.[34]
	helD	replication	DNA helicase unwinds DNA during replication/repair.	Carrasco et al.[35]
B. aphidicola	*typA*	stress response	GTPase regulates ribosome function and stress response.	Chakraborty et al.[36]
	rsmC	translation	16S rRNA methyltransferase modifies rRNA for ribosome assembly.	Chatterjee et al.[37]
	recB	replication	subunit of RecBCD, involved in DNA repair and homologous recombination.	Tamas and Dillingham[38,39]
	murD	cellular processes	enzyme for peptidoglycan synthesis, essential for cell wall formation.	Burger and Liang[40,41]
	cyoE	metabolism	protoheme IX farnesyltransferase, involved in heme biosynthesis for respiration.	Charles et al.[42]
	sohB	cellular processes	periplasmic protease, involved in protein turnover.	Cassone et al.[43]
M. hyopneumoniae	*mhp385*	cellular processes	function unclear, possibly Mycoplasma-specific metabolism or pathogenesis.	Deutscher et al.[44]
	p216	cellular processes	function unclear, potentially linked to virulence or Mycoplasma-specific processes. A cilium and heparin binding protein of Mycoplasma hyopneumoniae.	Siqueira and Leal[45,46]
	truB	translation	tRNA pseudouridine synthase, modifies tRNA for stability/function.	Yokobori et al.[47]
	rpoE	stress response	sigma factor regulates stress response gene expression.	Madeira and Gabriel[48]
	nrdF	replication	ribonucleotide reductase subunit provides dNTPs for DNA synthesis.	Chen et al.[49]
	rrf	translation	ribosome recycling factor releases ribosomes after translation.	Seely and Gagnon[50]
	rpsO	translation	ribosomal protein S15, part of the small ribosomal subunit.	Mathy et al.[51]
S. sviceus	*dnaK*	stress response	a molecular chaperone aids protein folding under stress (e.g., heat shock).	Bucca et al.[52]

(*Continued on next page*)

Table 1. *Continued*

Microbe name	Gene	Category	Function	Reference
S. cohnii	*ftsA*	cellular processes	actin-like protein, essential for septum formation during cell division.	Lara et al.[53]
P. megaterium	*ribF*	metabolism	a bifunctional enzyme synthesizes FMN and FAD for flavoprotein function.	Matern et al.[54]
E. faecalis	*ffh*	translation	signal recognition particle protein targets proteins to membranes during translation.	Michaux et al.[55]

Italicized text indicates scientific names of microbes and their gene names, following standard nomenclature conventions.

diversity for broad KEGG pathway mapping. In contrast, InterPro-based domain enrichment effectively captured the functional landscape of the microbial community, highlighting overrepresented strategies consistent with active protein synthesis, stress adaptation, and defense mechanisms.

Dysregulated host cell response associated with the presence or absence of intracellular microbes in healthy, SARS-CoV-2-infected, recovered, and long-recovered groups

Lastly, we examined host transcriptional responses associated with intracellular microbial presence by comparing M^+ and M^- cells across all the conditions. To stringently define microbial status, cells with $\geq$10 microbial UMIs were classified as M^+, whereas those with 0 microbial UMIs were designated as M^- (Figure 5A). Differential expression analysis was first performed at the pseudobulk level to understand the overall effect of microbial species at group levels and further delineated at the cellular level.

Impact of intracellular microbes on host-gene expression profiles among clinical groups

This analysis revealed a markedly higher number of upregulated genes in the healthy (377 DEGs) and long recovered (409 DEGs) groups, compared to the infected (25 DEGs) and recovered (13 DEGs) states. In contrast, downregulated genes were most prominent in the recovered group (2,182 DEGs), followed by healthy (162 DEGs), long recovered (31 DEGs), and infected (9 DEGs) conditions (Figures 5B and 5C; adj. p value <0.05, $\log_2$FC $\geq$ 1.5).

To further characterize the functional relevance of these DEGs, gene set enrichment analysis was performed. The number of enriched pathways varied markedly across conditions, with healthy (3 up, 148 down), infected (17 up, 1 down), recovered (43 up, 478 down), and long recovered (75 up, 2 down) groups, and only pathways with more than five genes are shown. In the healthy group, M+ cells exhibited upregulation of oxygen and carbon dioxide transport pathways, alongside widespread downregulation of key signaling cascades, including MAPK/RAF signaling and TP53-mediated transcriptional regulation. Importantly, apoptosis-associated genes such as *TP53, ATM,* and *GADD45A* were suppressed, suggesting a microbiome-associated metabolic enhancement and downregulation of inflammatory and stress response pathways, potentially promoting cellular homeostasis and microbial presence (Figure 5D).

In contrast to the healthy state, M+ cells in the SARS-CoV-2 infected group were enriched for pathways associated with acute innate immune activation, including neutrophil degranulation and cellular responses to chemical stress, driven by upregulation of oxidative stress regulators (*CAT*, *PRDX2, PRDX5*, and *PRDX6*) (Figure 5E and Table S8). Concomitant downregulation of RUNX3-mediated pathways suggests impaired coordination of leukocyte migration and adaptive immune signaling, indicative of a shift toward a predominantly innate immune response.

Upon recovery, M^+ cells displayed a transcriptional program consistent with immune resolution, marked by the upregulation of IL-10, IL-4, and IL-13 signaling, along with enhanced antigen processing and Toll-like receptor regulation and suppression of mRNA processing and nucleocytoplasmic transport pathways. In the long-recovered state, M^+ cells retained features of sustained innate immune priming, including persistent activation of the NLRP3 inflammasome, neutrophil degranulation, and chemokine-mediated recruitment, alongside downregulation of TCF/LEF:CTNNB1 signaling, collectively suggesting maintenance of a heightened immune surveillance state beyond recovery (Figures 5F and S15).

Transcriptional heterogeneity between microbial-positive and microbial-negative immune cell types

To define these effects at single-cell resolution, we compared transcriptional profiles between the M^+ and M^- cells across immune cell subsets in all conditions, excluding cell-type populations with fewer than 25 cells in each group. After filtering, 2 cell types in the healthy group (Tregs and plasma cells), 6 in the infected group, 7 in the recovered group, and 3 in the long-recovered group were retained (Figure 5G and Table S9). Across conditions, the transcriptional response was predominantly characterized by gene upregulation in M^+ cells, with relatively few genes showing reduced expression.

In the healthy group, M^+ plasma cells showed increased expression of *ARSA, SMPD4, SLC19A1, FKBP8*, and *AREL1*, linked to lipid and cofactor metabolism, post-translational modification, and antigen processing, consistent with an active intracellular presence of microbial species. In contrast, *FYB1*, involved in TCR signaling and immune activation, was reduced, indicating decreased adaptive signaling. In Tregs, *GREM1* and *PCDH17* were selectively upregulated, pointing to context-specific transcriptional modulation in M^+ populations.

In the infected state, 595 genes were significantly upregulated across the plasma cells, cytotoxic T cells, memory T cells, Tregs, platelets, and dendritic cells, with no genes meeting the threshold for downregulation ($\log_2$FC $\geq$ 1.5; adj. p value <0.05 (Figure 5G). At the cell level, M^+ cytotoxic T cells showed

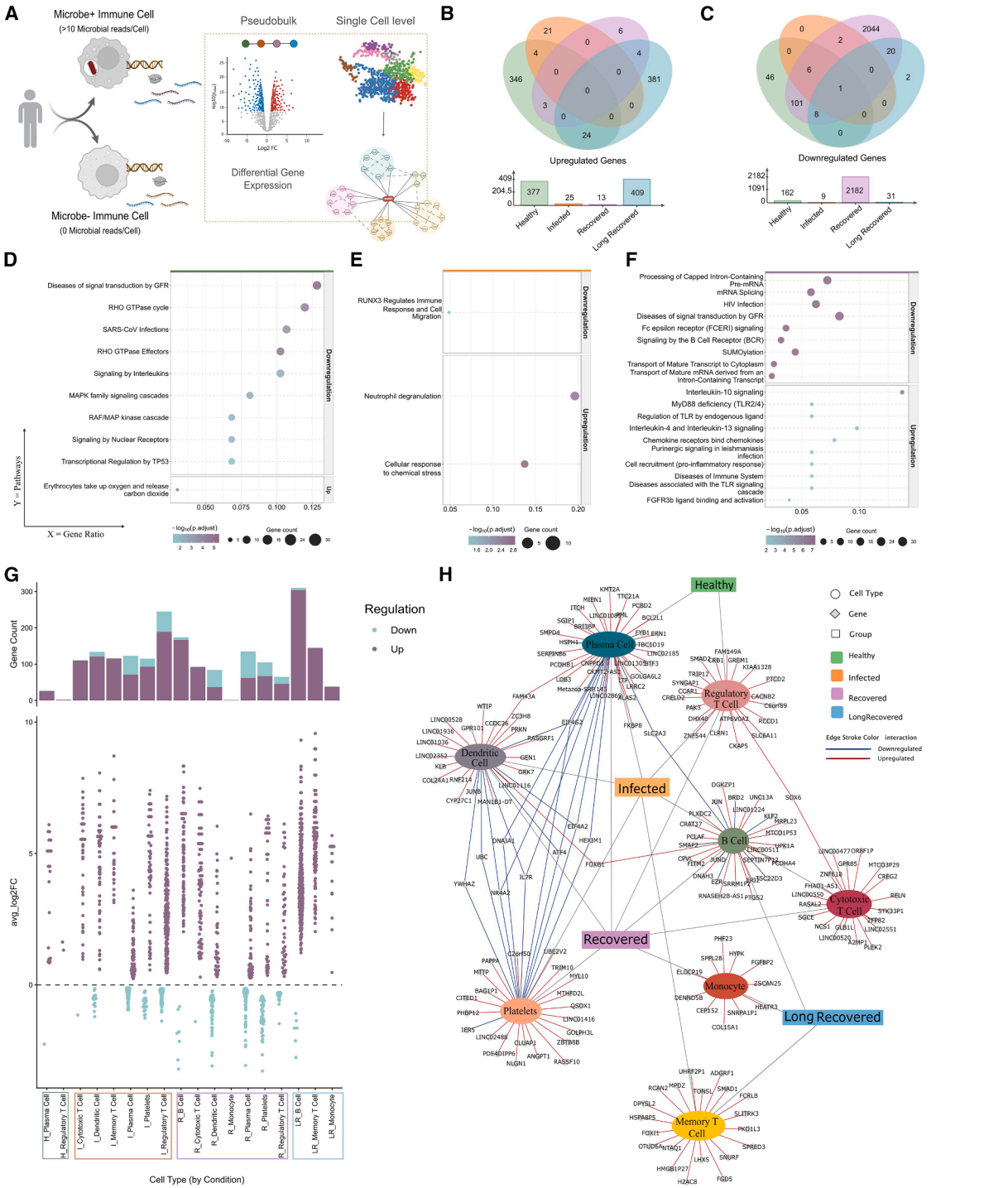


Figure 5. Transcriptional alterations associated with intracellular microbial presence across comparison groups and immune cells

(A) Schematic overview of the analysis workflow. Immune cells were classified as microbe-positive (M^+; ≥10 microbial reads per cell) or microbe-negative (M^-; 0 microbial reads), followed by differential gene expression analysis at the pseudobulk and single-cell levels.

(B and C) Venn diagrams show the overlap of upregulated (B) and downregulated (C) genes in M^+ vs. M^- cells across the healthy, infected, recovered, and long-recovered conditions. Bar plots indicate the total number of differentially expressed genes in each condition.

(D–F) Pathway enrichment analysis of differentially expressed genes in M^+ cells across conditions. Dot plots display significantly enriched pathways for healthy (D), infected (E), and recovered (F) samples. Dot size represents gene count, and color indicates statistical significance (−log10 adjusted *p* value).

(legend continued on next page)

increased expression of chemotactic and inflammatory mediators (*CXCR2, CCL2,* and *FPR1*) together with cell cycle regulators (*E2F1* and *MAP2K6*). M^+ dendritic cells were enriched for pathways associated with metabolic processes (peroxisomal transport and glycosylation), extracellular matrix remodeling, and intracellular trafficking, alongside increased expression of genes involved in interferon signaling, antigen presentation (*HLA-DRB1*), and innate immune pathways, including inflammasome and NOD signaling, with antimicrobial responses mediated by LTF. These changes were accompanied by the modulation of transcriptional regulators (*RUNX1, KMT2A,* and *PML)* and cell death pathways (*BCL2L1* and *ITCH*). In platelets, M^+ cells exhibited increased expression of genes involved in lipid metabolism and lipoprotein remodeling (*MTTP* and *ANGPTL4*), as well as protein quality control and ubiquitination (*SELENOS* and *DERL1*). In Tregs, M^+ populations showed increased expression of nuclear transport components (*NUP210*, *NUP153,* and *KPNA1*), together with pathways related to viral interactions, RNA processing, autophagy, and cellular stress responses. In memory T cells, M^+ cells displayed increased expression of genes associated with FGFR2 and PI3K-AKT signaling (*FGF7, FGF18*, and *ESR2*), consistent with increased proliferative and survival signaling. Collectively, these findings indicate distinct, cell-type-specific transcriptional changes associated with intracellular microbial presence during infection.

In the recovered state, 361 genes were upregulated and 24 downregulated across plasma cells, B cells, cytotoxic T cells, monocytes, Tregs, platelets, and dendritic cells. M^+ B cells exhibited enrichment of purinergic signaling (*P2RY6, and LPAR4*) and antigen cross-presentation (*MRC1,* and *PSMA2*), while M^+ cytotoxic T cells showed increased cGMP-NO signaling (*PDE11A,* and *PRKG1*) and IL-23–associated pathways (IL12B). In contrast, M^+ dendritic cells displayed broad suppression of cytokine, interleukin, and interferon signaling, and monocytes showed minimal transcriptional changes, with ZSCAN25 as the sole DEG. M^+ plasma cells were characterized by the enrichment of BMP4-driven developmental and extracellular matrix programs, together with immune receptor signaling (*TYROBP*) and autophagy (*PCNT*), alongside downregulation of interferon, cytokine, and stress-response pathways (*EIF4A2, EIF4G2, IL7R, ATF4,* and *HSPH1*). M^+ platelets showed increased small GTPase signaling (RHOF, RHOB) and metabolic activity, with concomitant suppression of immune and antiviral pathways, indicating a metabolically active yet immunologically restrained state. M^+ Tregs exhibited the enrichment of SMAD2-dependent TGF-β/NODAL signaling and post-transcriptional regulation (*EXOSC8* and *NLRC5*). In the longitudinal recovered group, M^+ B cells retained the activation of cytokine and chemokine signaling, antigen presentation, and inflammasome pathways (NLRP3, CASP1), alongside PI3K-AKT signaling, whereas *EZR, JUN, SMAP2*, and *KLF2* were downregulated, consistent with reduced cytoskeletal dynamics and migratory capacity (Figure 5H).

In the long-recovered state, M^+ monocytes exhibited the enrichment of FLT3 signaling (*FLT3LG*), including the activation of SRC family kinases and STAT5, together with metabolic (*BDH1*) and growth factor signaling (FGFBP2–FGFR2), consistent with sustained immune regulation and metabolic adaptation. In parallel, M^+ memory T cells showed the enrichment of erythrocyte-associated pathways (*HBB* and *HBA1*), alongside increased DNA repair and damage tolerance (*POLE2, GTF2H4,* and *DDB2*) and nucleotide metabolism, indicating enhanced genomic maintenance. Concomitant activation of lipid metabolic (*PLA2G6*) and apoptotic pathways (*FASLG*) further suggests ongoing metabolic remodeling and controlled cellular turnover during longitudinal recovery.

We next asked whether microbial presence is associated with altered host transcriptional states within specific immune cell populations. We restricted the analysis to high-confidence microbe-positive cells (>10 reads per cell from the same microbe), retaining only *Mycoplasma hyopneumoniae* and *Staphylococcus cohni*i for downstream analysis. *M. hyopneumoniae* marked distinct, context-dependent transcriptional states, with enrichment in plasma cells during infection and monocytes in recovered and long-recovered individuals. In infected microbe-positive plasma cells, increased expression of genes associated with innate immune activation and cellular stress (e.g., ZBTB20, CAPN2, ATMIN, and PTGS1) was exhibited. By contrast, monocytes from the recovered groups showed signatures consistent with interferon responsiveness, damage sensing, and membrane repair (e.g., *CLEC12A, UBE2L6, PPP2CB,* and *CHMP1A*). Similarly, *S. cohnii* was preferentially detected in the plasma cells from infected and recovered groups, where microbe-positive cells displayed transcriptional features indicative of immune activation, metabolic adaptation, and cell survival (e.g., *VCAN, RGCC, STK17B, SLC2A3,* and *KPNB1*).

Together, these findings indicate that intracellular microbial presence is linked to distinct and persistent host immune transcriptional programs, spanning acute inflammatory activation during infection, regulatory remodeling during recovery, and sustained innate immune vigilance after recovery.

DISCUSSION

Our prior studies during the COVID-19 pandemic have highlighted the disease's profound impact on human health, revealing lasting effects such as persistent respiratory dysfunction, dysregulated immune responses, and alterations in the host microbiome.[8,10] In this study, we characterized the intracellular microbial communities residing within the immune cells across healthy, SARS-CoV-2-infected, recovered, and long-recovered individuals, with a focus on understanding the effects of SARS-CoV-2 on these communities and evaluating whether long-recovered individuals restore their baseline microbiome. Generally, microbiome disruptions are known to influence

(G) Distribution of differentially expressed genes across the immune cell types and conditions. The upper panel shows gene counts for upregulated and downregulated genes. In contrast, the lower panel displays the average $\log_2$FC across cell types.
(H) Cell gene interaction network highlights cell-type-specific transcriptional programs across the disease stages. Nodes represent genes and immune cell types, with edges indicating regulatory associations. Edge color denotes direction of regulation (upregulated in red; downregulated in blue).

disease severity, yet their long-term health implications warrant further exploration.[56] While intracellular microbes are recognized for modulating immune cell function, their role in shaping infection outcomes remains underexplored.

To investigate intracellular microbial communities, we utilized unmapped reads from the scRNA-seq data, a non-canonical yet increasingly adopted approach in recent studies to identify microbial signatures at the single-cell level.[18,19] Leveraging PathogenTrack, we identified intracellular microbes (viruses, bacteria, and fungi) and their associations with specific immune cell types. This approach aligns with the recent studies employing computational tools such as CSI-microbes to uncover microbial genera and their interactions with the host immune responses.[4,5,17,57]

Our single-cell level analysis revealed three key insights into the dynamic interplay between the intracellular microbes and immune cell function during and after SARS-CoV-2 infection. First, the overlap of cellular and unmapped component proportions in healthy and long-recovered individuals. Second, distinct microbial profiles were seen in recovered and long-recovered groups, characterized by the commensals and opportunistic taxa. Third, we found transcription, translation, and replication genes along with the virulence and stress response from the immune cell-specific microbes.

Looking into the first insight, we observed higher unmapped reads in the COVID-19-infected and recently recovered individuals compared to the healthy and long-recovered groups. Notably, alpha diversity was elevated in the infected group, contrasting with prior reports of reduced microbial diversity in the respiratory tract and gut during COVID-19.[58,59] This discordance between the microbiome findings can be due to different sample sites, as our study focused on immune cell-associated microbes rather than mucosal surfaces. Notably, our beta diversity analysis shows a clear separation of microbial communities in patients with COVID-19 compared to the healthy and long-recovered individuals. Furthermore, the 50 species shared across all four groups displayed two distinct patterns. Differential profiling revealed a notable increase in opportunistic and pathogenic taxa, including *Escherichia albertii, Campylobacter jejuni, Haemophilus influenzae, Escherichia coli, Providencia stuartii, Clostridium botulinum, Mycoplasma ruminantium,* and *Pseudomonas tolaasii,* specifically in the COVID-19-infected and recently recovered individuals. The presence of these bacteria during SARS-CoV-2 infection and recovery indicates potential coinfection, which can exacerbate disease severity, complicate the treatment regimen, prolong recovery, and increase mortality risk.[60,61] This shift suggests a dysbiotic microbial environment triggered by the SARS-CoV-2 infection persists post-recovery, potentially contributing to the long COVID symptoms. For instance, *E. albertii,* a known enteropathogen, drives gastroenteritis with symptoms like watery diarrhea and fever, thriving in the disrupted gut environment of patients with COVID-19.[62] Similarly, *C. jejuni* leverages genes like *cipA* for gut cell invasion and immune evasion, fueling both acute and persistent infections without relying on typical toxins.[63] *H. influenzae,* typically a nasopharyngeal commensal, crosses epithelial barriers to cause otitis, pneumonia, or meningitis, particularly in the immunocompromised individuals, by producing immune-disrupting effector molecules.[64,65] *E. coli* and *P. stuartii* are linked to urinary tract and bloodstream infections, reflecting their opportunistic nature in a dysregulated immune setting.[66] While several microbes, including *M. ruminantum* and *Pseudomonas tolaasii, were* also found comparatively more in the infected group, these microbes are not typically associated with human infections. Their presence suggests a potential zoonotic transmission, indicating a possible cross-species microbial transfer, which will be clarified with more studies in this direction.

In contrast, S. *sviceus, P. aeruginosa, Helicobacter pylori,* and *C. perfringens* were found to be increased in the long-recovered and healthy groups. These microbes, often part of the normal human flora, likely contribute to microbiome stability and immune homeostasis. The consistent abundance of *S. pristinaespiralis* across all the groups, potentially due to its production of antibiotics like pristinamycin, highlights its role as a stable microbiome component.[67,68] These findings emphasize that the distribution and abundance of the shared microbial species, rather than unique taxa alone, drive the observed microbiome shifts during and after SARS-CoV-2 infection.

To better understand the potential role of these microbes, we further explored the immune cell types associated with these microbes, as successful evasion of these microbes allows them to make immune cells a safe habitat.[69,70] The intracellular microbes play important roles in training the host immune response to promote symbiosis with the host cells.[71,72] However, they can modulate the immune cell functions to overcome the immune response generated in defense.[73,74] The overall immune cell profile from our dataset uncovers a fascinating immune landscape that echoes prior studies done on SARS-CoV-2. In infected individuals, we observed a striking surge in plasma B cells and Tregs, contrasted by a sharp decline in various T cell subsets compared to the healthy and recovered individuals. Plasma B cells, which generate antibodies in response to infection, likely increase in battling SARS-CoV-2.[75] Tregs are important for maintaining homeostasis within the immune system.[76] They suppress both innate and adaptive immune responses by inhibiting antigen-presenting cell maturation, reducing IL-2 availability to conventional T cells via high-affinity IL-2 receptors and releasing immunosuppressive cytokines.[21] Our data revealed a significant reduction in the natural killer (NK) cell populations in the SARS-CoV-2-infected individuals, consistent with the prior studies reporting decreased NK cell counts, impaired cytolytic activity, and a correlation with the increased disease severity.[77] Our analysis of immune cell profiles revealed significant alterations in the SARS-CoV-2-infected individuals, including a surge in the plasma B cells and Tregs and a decline in the T cell subsets and NK cells, consistent with the prior studies. Notably, microbial reads were disproportionately enriched in the NK T cells, memory T cells, B cells, NK cells, and naive T cells in the infected individuals, suggesting that opportunistic pathogens exploit the immune disarray induced by SARS-CoV-2 to colonize these cells. Monocytes and dendritic cells, critical for antigen presentation, exhibited functional impairment during acute infection, further facilitating microbial persistence.[78] The elevated microbial presence in the plasma cells, particularly in the infected and recently recovered groups, may indicate broader immune dysfunction, as these cells are primarily responsible for antibody

production. Combined, these results suggest that during immunosuppression, microbes can evade elimination by the host immune system and can persist inside these immune cells.

Our study underscores the prolonged time required for microbiome restoration following SARS-CoV-2 infection, with long-recovered individuals exhibiting microbial profiles akin to those of the healthy controls after over three months. The persistence of opportunistic pathogens in the immune cells during and after infection may contribute to long COVID symptoms by perpetuating immune dysregulation. The identification of 18 significant microbial species, categorized into commensals, opportunists, and others (not known to cause infections in humans or animals), highlights the complexity of the intracellular microbiome and its role in host immunity.

These species were categorized into commensals, opportunists, and others (not known to cause infections in humans or animals). Among the commensal microbes, *S. sviceus, S. clavuligerus, E. faecalis,* and *B. megaterium* exhibited significantly higher abundance in healthy and long-recovered individuals compared to those who were infected or recently recovered. These species are known for producing secondary metabolites, such as clavulanic acid, cephamycin C, pristinamycin, bacteriocins, and lipopeptides (e.g., surfactins and fengycins), which may contribute to microbiome balance and host health. Intriguingly, *S. pristinaespiralis* maintained consistent abundance across all the groups, suggesting its potential as a stable component of the microbiome, possibly due to its antibiotic production capabilities.[79] Furthermore, the elevated presence of certain microbes in plasma cells and Tregs indicates possible interactions with the host immune system. On the other hand, *E. coli* and *P. stuartii* show a clear surge in the platelets, plasma cells, dendritic cells, and monocytes in the infected and recovered groups. These cells interact with the intracellular microbes in distinct ways, with monocytes and dendritic cells serving as primary hosts or responders, as they perform antigen presentation, while platelets and plasma cells support immune responses indirectly. Dendritic cells are reported to become functionally impaired during acute COVID-19, particularly in their ability to sense pathogens, present antigens, and activate T cell responses.[80] Plasma cells, which initially showed increased microbial association during infection and recovery, also maintained elevated levels in the long-recovered group compared to the healthy individuals. Given that plasma cells are primarily responsible for antibody production, the presence of microbial reads within them may signal a broader immune dysfunction.

To understand the functional roles of these intracellular microbes, we further analyzed the genes captured in our dataset. For example, *E. coli* expressed virulence (*nupX, speA*), stress response (*sgrR*), and regulatory genes (*yggM*), potentially influencing host-microbe interactions.[81] Similarly, *C. botulinum*'s *rpoB* gene was linked to cytokine production in the macrophages, while *E. coli*'s *rpoE*-regulated virulence during bacteremia.[82] These findings suggest that intracellular microbes actively shape immune responses, contributing to disease progression and persistence. Likewise, the *abc-f* gene functions as a cytosolic nucleic acid sensor that regulates the expression of *CXCL10* and downstream type I interferon responses in the epithelial cells. In a similar context, *E. coli's rpoE* has been reported to control the expression of multiple virulent genes during bacteremia. Similar to *S. pristeinaespiralis* in the commensal group, we found *M. hyopneumoniae*, a causal organism of enzootic pneumonia, colonizes the porcine respiratory tract and is found in higher abundance in infected, recovered, and long-recovered relative to healthy. Together, these findings suggest that intracellular microbes, through expression of functional genes, may influence immune responses and disease progression. Overall, our results indicate that the intracellular microbiome within immune cells is dynamically altered during infection and may require over three months to return to a healthy baseline. One hypothesis is that detected microbial transcripts represent opportunistic survival and interfere with the host immune response in the weakened cellular environments. On the contrary, specific genes such as the stress regulator *sgrR* or the chaperone *dnaK* may actively modulate host pathways to influence immune activation and cell longevity, similar to those identified in *rpsO* gene correlation analysis. The increased *rpsO*-associated microbial activity may be linked to the suppression of host transcriptional and translational machinery, as well as the modulation of immune and stress-response pathways. These single-cell transcriptomics findings provide a strong foundation for future *in vitro* experimental validation using co-culture or infection models to determine if microbial expression directly drives changes in the immune cells.

To gain a deeper understanding of how these intracellular microbes modulate the immune response, we examined the transcriptional programs of cells with and without microbes at both pseudobulk and single-cell RNA-seq levels. In the healthy state, we observed the downregulation of apoptosis-related pathways along with MAPK/RAF signaling, suggesting a tolerogenic environment; notably, pathogenic bacteria are known to exploit host signaling by targeting MAPK pathways through type III effectors and modulating p53 activity to support their survival.[83–85] In the SARS-CoV-2 infected group, RUNX3 was downregulated, indicating impaired lymphocyte function and dysregulation of adaptive immunity[86] while concurrent upregulation of neutrophil degranulation and stress-response pathways reflected the activation of innate immune responses.[87] In the recovered group, anti-inflammatory interleukin signaling was upregulated, supporting T cell differentiation, M2 macrophage polarization, MHC class II expression, and B cell and plasma cell differentiation with antibody isotype switching.[88,89] In the long-recovered state, M^+ cells exhibited heightened immune response pathways, suggesting sustained immune activation likely aimed at resolving residual microbial imbalance.

The intracellular presence of microbes drives a transition from metabolic adaptation to acute inflammatory conflict, and finally, to long-term immune priming at the cellular level. In the healthy group, M^+ plasma cells shift toward lipid and cofactor metabolism (*ARSA, SMPD4*) and antigen processing (*AREL1*) while downregulating adaptive activation (*FYB1*), suggesting a homeostatic, permissive niche. During an active infection, this state changes to a strong fight between the host and the pathogen, which is marked by bacterial sensing (*FPR1*), nutritional immunity (*LTF*), and PI3K-AKT survival signaling to deal with stress caused by microbes. As patients move into recovery, the profile pivots toward immune resolution and metabolic restraint,

specifically through purinergic signaling (*P2RY6*) and the broad suppression of cytokine cascades. Notably, the long-recovered state is defined by sustained NLRP3 inflammasome activation and FLT3 signaling, indicating that intracellular microbes may maintain a state of heightened immune surveillance long after the initial infection has cleared. However, these findings warrant further validation through independent groups and functional studies to confirm the observed cell-type-specific effects of intracellular microbes on host immune responses.

In summary, our study highlights that a long recovery time is required following COVID-19 infection to restore a healthy-like immune state. During primary infection, a dysregulated immune repertoire appears to create opportunities for pathogenic and opportunistic microbial species to evade immune surveillance and proliferate within the immune cells. This intracellular persistence may further modulate immune responses and influence disease outcomes. These insights lay the groundwork for future investigations into the role of the intracellular microbiome in shaping host immunity and disease progression, with potential implications for therapeutic strategies targeting long COVID and related conditions.

Our findings indicate that recovery from SARS-CoV-2 infection is a prolonged process, with the microbiome and immune response of recently recovered individuals (after three months) remaining distinct from those of healthy controls. During infection and early recovery, we observed a dysregulated immune response, characterized by the increased abundance of opportunistic and pathogenic microbial species, potentially contributing to the elevated microbial reads detected in our scRNA-seq datasets. In contrast, healthy and long-recovered individuals exhibited higher abundances of non-pathogenic/beneficial bacteria, such as *Streptomyces sviceus, S. clavuligerus, Enterococcus faecalis,* and *Bacillus megaterium*. The expression of replication, transcription, and translation genes highlights their transcriptionally active stage, while stress response and virulence genes suggest complex interactions between microbial communities and host immune response. The gene expression differences between the cells with and without microbial reads further highlight the biological granularity governing immune response. These insights advance our understanding of SARS-CoV-2 pathogenesis by highlighting the functional role of the microbiome in immune regulation and its potential as a predictor of clinical outcomes.

Limitations of the study

While our study provides the single-cell resolution atlas of the intracellular microbiome in COVID-19 from healthy to infected to recovered and longitudinal recovered, we acknowledge potential limitations that provide important context for our findings and establish a baseline for future investigations. First, our analysis was restricted to PBMCs; extending this approach to respiratory or gastrointestinal tissues would offer a more comprehensive immunogenic view of systemic microbial translocation. Second, the retrospective design of our sampling constrained our capacity to stratify the infection group by detailed severity sub-phenotypes, and the limited sample size of the longitudinal recovered group indicates that our findings regarding microbial reversion would be strengthened with validation in larger, statistically robust cohorts. From a technical perspective, it is important to note that the BD Rhapsody platform uses 3′-end capture chemistry, which may under-detect microbial transcripts lacking stable poly-A tails compared to the host mRNA, potentially resulting in conservative abundance estimates. Additionally, an external spike-in positive control was lacking to test the pipeline's absolute sensitivity limit on this specific platform. However, these technical constraints do not compromise the clear, cell-type-specific microbial signals observed. We present this work as a foundational study; by identifying persistent virulence factors and distinct host-microbe associations, this study provides the critical molecular, transcriptomic baseline necessary for future research to rigorously test the mechanisms of inter-kingdom transfer and functional immune modulation proposed here.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Rajesh Pandey (rajeshp@igib.in).

Materials availability

This study did not generate new unique reagents and materials.

Data and code availability

- Raw and analyzed single-cell RNA-seq data generated in this study have been deposited in GEO under accession number GSE303208 and in the SRA under BioProject accession number PRJNA1293800. These datasets are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All custom scripts used for sample demultiplexing, microbial profiling integration, downstream statistical analysis, and visualization are publicly available through GitHub at: (https://github.com/INGEN-HOPE/Intracellular-Microbial-Analaysis-of-BD-Rhapsody-Wta-Abseq-data).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors duly acknowledge all the individuals who participated in the study. The authors acknowledge the help and support from Dr. Aradhita Baral toward facilitation as research manager and coordination with the funders. The support of lab manager Dr. Bharti Kumari is duly acknowledged for ensuring all reagents are available for the experiments. The authors acknowledge the support of Anil Kumar and Nisha Rawat toward sample transport and sample management. JS acknowledges UGC for the fellowship support. AY acknowledges CSIR. This research was funded by Bill and Melinda Gates Foundation, (grant no. INV-033578) to RP.

AUTHOR CONTRIBUTIONS

J.S., conceptualization, methodology, investigation, data curation, visualization, and writing – original draft. R.A., formal analysis, data curation, visualization, and writing – original draft. P.C., formal analysis and methodology. P.D., conceptualization, data curation, visualization, and writing – original draft. P.M, formal analysis. A.Y., investigation, writing – original draft. A.J., formal analysis. R.P., conceptualization, methodology, supervision, writing—review and editing, and funding acquisition. All authors contributed to the article and approved the submitted version.

DECLARATION OF INTERESTS

All the authors affirm that there is no conflict of interest while conducting the study. We also confirm that the funding body did not have any role in planning, execution, or inferences drawn from the study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human subjects and clinical protocol
- METHOD DETAILS
  - PBMC isolation
  - Single-cell capture and cDNA synthesis
  - Library preparation and sequencing
  - scRNA-seq data pre-processing, cell clustering, and annotation
  - *Meta*-transcriptomic analysis
  - Comparative analysis of dengue and COVID-19 single-cell RNA-seq datasets
  - Differential microbial gene analysis
  - Host cell transcriptomic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2026.116344.

Received: September 15, 2025
Revised: April 9, 2026
Accepted: May 26, 2026

REFERENCES

1. Stevens, E.J., Bates, K.A., and King, K.C. (2021). Host microbiota can facilitate pathogen infection. PLoS Pathog. *17*, e1009514. https://doi.org/10.1371/journal.ppat.1009514.
2. Morens, D.M., Taubenberger, J.K., and Fauci, A.S. (2008). Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. J. Infect. Dis. *198*, 962–970. https://doi.org/10.1086/591708.
3. Shafran, N., Shafran, I., Ben-Zvi, H., Sofer, S., Sheena, L., Krause, I., Shlomai, A., Goldberg, E., and Sklan, E.H. (2021). Secondary bacterial infection in COVID-19 patients is a stronger predictor for death compared to influenza patients. Sci. Rep. *11*, 12703. https://doi.org/10.1038/s41598-021-92220-0.
4. Yadav, A., Kumari, P., Devi, P., Adjele, J.J.B., Budhiraja, S., Tarai, B., and Pandey, R. (2023). Dual RNA-Seq reveals transcriptionally active microbes (TAMs) dynamics in the serum of dengue patients associated with disease severity. Front. Microbiol. *14*, 1307859. https://doi.org/10.3389/fmicb.2023.1307859.
5. Yadav, A., Devi, P., Kumari, P., Shamim, U., Tarai, B., Budhiraja, S., and Pandey, R. (2024). Metatranscriptomic insights into the dengue patient blood microbiome: Enhanced microbial diversity and metabolic activity in severe patients. PLoS Negl. Trop. Dis. *18*, e0012589. https://doi.org/10.1371/journal.pntd.0012589.
6. de Nies, L., Galata, V., Martin-Gallausiaux, C., Despotovic, M., Busi, S.B., Snoeck, C.J., Delacour, L., Budagavi, D.P., Laczny, C.C., Habier, J., et al. (2023). Altered infective competence of the human gut microbiome in COVID-19. Microbiome *11*, 46. https://doi.org/10.1186/s40168-023-01472-7.
7. Yeoh, Y.K., Zuo, T., Lui, G.C.-Y., Zhang, F., Liu, Q., Li, A.Y., Chung, A.C., Cheung, C.P., Tso, E.Y., Fung, K.S., et al. (2021). Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. Gut *70*, 698–706. https://doi.org/10.1136/gutjnl-2020-323020.
8. Devi, P., Maurya, R., Mehta, P., Shamim, U., Yadav, A., Chattopadhyay, P., Kanakan, A., Khare, K., Vasudevan, J.S., Sahni, S., et al. (2022). Increased Abundance of Achromobacter xylosoxidans and Bacillus cereus in Upper Airway Transcriptionally Active Microbiome of COVID-19 Mortality Patients Indicates Role of Co-Infections in Disease Severity and Outcome. Microbiol. Spectr. *10*, e02311-21. https://doi.org/10.1128/spectrum.02311-21.
9. Devi, P., Kumari, P., Yadav, A., Tarai, B., Budhiraja, S., Shamim, U., and Pandey, R. (2023). Longitudinal study across SARS-CoV-2 variants identifies transcriptionally active microbes (TAMs) associated with Delta severity. iScience *26*, 107779. https://doi.org/10.1016/j.isci.2023.107779.
10. Devi, P., Kumari, P., Yadav, A., Tarai, B., Budhiraja, S., Shamim, U., and Pandey, R. (2023). Transcriptionally active nasopharyngeal commensals and opportunistic microbial dynamics define mild symptoms in the COVID 19 vaccination breakthroughs. PLoS Pathog. *19*, e1011160. https://doi.org/10.1371/journal.ppat.1011160.
11. Chai, L., Wang, Q., Si, C., Gao, W., and Zhang, L. (2022). Potential Association Between Changes in Microbiota Level and Lung Diseases: A Meta-Analysis. Front. Med. *8*, 723635. https://doi.org/10.3389/fmed.2021.723635.
12. Ghamari, M., Jabalameli, F., Afhami, S., Halimi, S., Emaneini, M., and Beigverdi, R. (2025). Acinetobacter baumannii infection in critically ill patients with COVID-19 from Tehran, Iran: the prevalence, antimicrobial resistance patterns and molecular characteristics of isolates. Front. Cell. Infect. Microbiol. *14*, 1511122. https://doi.org/10.3389/fcimb.2024.1511122.
13. Thakur, A., Mikkelsen, H., and Jungersen, G. (2019). Intracellular pathogens: host immunity and microbial persistence strategies. J. Immunol. Res. *2019*, 1–24. https://doi.org/10.1155/2019/1356540.
14. Fertitta, V., Varano, B., Del Cornò, M., Fortini, P., Aureli, A., and Conti, L. (2024). Akkermansia muciniphila- and Pathogenic Bacteria-Derived Endotoxins Differently Regulate Human Dendritic Cell Generation and γδ T Lymphocyte Activation. Biomolecules *14*, 1571. https://doi.org/10.3390/biom14121571.
15. Sivan, A., Corrales, L., Hubert, N., Williams, J.B., Aquino-Michaels, K., Earley, Z.M., Benyamin, F.W., Man Lei, Y., Jabri, B., Alegre, M.-L., et al. (2015). Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science *350*, 1084–1089. https://doi.org/10.1126/science.aac4255.
16. Mairpady Shambat, S., Gómez-Mejia, A., Schweizer, T.A., Huemer, M., Chang, C.-C., Acevedo, C., Bergada-Pijuan, J., Vulin, C., Hofmaenner, D.A., Scheier, T.C., et al. (2022). Hyperinflammatory environment drives dysfunctional myeloid cell effector response to bacterial challenge in COVID-19. PLoS Pathog. *18*, e1010176. https://doi.org/10.1371/journal.ppat.1010176.
17. Soni, J., Mehta, P., Yadav, S., Chattopadhyay, P., and Pandey, R. (2025). Protocol for investigating intracellular microbial diversity using single-cell RNA-seq in immune cells of SARS-CoV-2-positive and recovered patients. STAR Protoc. *6*, 103546. https://doi.org/10.1016/j.xpro.2024.103546.
18. Yadav, S., Mehta, P., Soni, J., Chattopadhyay, P., Devi, P., Habyarimana, T., Tardalkar, K., Joshi, M., and Pandey, R. (2023). Single-cell RNA-Seq reveals intracellular microbial diversity within immune cells during SARS-CoV-2 infection and recovery. iScience *26*, 108357. https://doi.org/10.1016/j.isci.2023.108357.
19. Robinson, W., Stone, J.K., Schischlik, F., Gasmi, B., Kelly, M.C., Seibert, C., Dadkhah, K., Gertz, E.M., Lee, J.S., Zhu, K., et al. (2024). Identification of intracellular bacteria from multiple single-cell RNA-seq platforms using

CSI-Microbes. Sci. Adv. *10*, eadj7402. https://doi.org/10.1126/sciadv.adj7402.

20. Zhang, W., Xu, X., Fu, Z., Chen, J., Chen, S., and Tan, Y. (2022). PathogenTrack and Yeskit: tools for identifying intracellular pathogens from single-cell RNA-sequencing datasets as illustrated by application to COVID-19. Front. Med. *16*, 251–262. https://doi.org/10.1007/s11684-021-0915-9.

21. Saheb Sharif-Askari, F., Saheb Sharif-Askari, N., Hafezi, S., Alsayed, H.A.H., Selvakumar, B., Eladham, M.W.A., Mdkhana, B., Bayram, O.S., Temsah, M.-H., and Halwani, R. (2023). Increased blood immune regulatory cells in severe COVID-19 with autoantibodies to type I interferons. Sci. Rep. *13*, 17344. https://doi.org/10.1038/s41598-023-43675-w.

22. Arora, J.K., Opasawatchai, A., Poonpanichakul, T., Jiravejchakul, N., Sungnak, W., Sakuntabhai, A., Singhasivanon, P., Suraamornkul, S., Yingtaweesak, T., Manopwisedjaroen, K., et al. (2022). Single-cell temporal analysis of natural dengue infection reveals skin-homing lymphocyte expansion one day before defervescence. iScience *25*, 104034. https://doi.org/10.1016/j.isci.2022.104034.

23. Craig, J.E., Zhang, Y., and Gallagher, M.P. (1994). Cloning of the nupC gene of Escherichia coli encoding a nucleoside transport system, and identification of an adjacent insertion element, IS 186. Mol. Microbiol. *11*, 1159–1168. https://doi.org/10.1111/j.1365-2958.1994.tb00392.x.

24. Ghatak, S., King, Z.A., Sastry, A., and Palsson, B.O. (2019). The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. Nucleic Acids Res. *47*, 2446–2454. https://doi.org/10.1093/nar/gkz030.

25. Pushpker, R., Bay, D.C., and Turner, R.J. (2022). Small multidrug resistance protein EmrE phenotypically associates with OmpW, DcrB and YggM for osmotic stress protection by betaine in Escherichia coli. Microbiology (Reading, Engl) *168*. https://doi.org/10.1099/mic.0.001287.

26. Sidorczuk, K., Burdukiewicz, M., Cerk, K., Fritscher, J., Kingsley, R.A., Schierack, P., Hildebrand, F., and Kolenda, R. (2024). adhesiomeR: a tool for Escherichia coli adhesin classification and analysis. BMC Genom. *25*, 609. https://doi.org/10.1186/s12864-024-10525-6.

27. He, B., Choi, K.Y., and Zalkin, H. (1993). Regulation of Escherichia coli glnB, prsA, and speA by the purine repressor. J. Bacteriol. *175*, 3598–3606. https://doi.org/10.1128/jb.175.11.3598-3606.1993.

28. Mukherjee, K., Huddleston, J.P., Narindoshvili, T., Nemmara, V.V., and Raushel, F.M. (2019). Functional Characterization of the ycjQRS Gene Cluster from Escherichia coli: A Novel Pathway for the Transformation of d-Gulosides to d-Glucosides. Biochemistry *58*, 1388–1399. https://doi.org/10.1021/acs.biochem.8b01278.

29. Sun, Y., and Vanderpool, C.K. (2011). Regulation and function of Escherichia coli sugar efflux transporter A (SetA) during glucose-phosphate stress. J. Bacteriol. *193*, 143–153. https://doi.org/10.1128/JB.01008-10.

30. Fostier, C.R., Monlezun, L., Ousalem, F., Singh, S., Hunt, J.F., and Boël, G. (2021). ABC-F translation factors: from antibiotic resistance to immune response. FEBS Lett. *595*, 675–706. https://doi.org/10.1002/1873-3468.13984.

31. Kuwahara, T., Ogura, Y., Oshima, K., Kurokawa, K., Ooka, T., Hirakawa, H., Itoh, T., Nakayama-Imaohji, H., Ichimura, M., Itoh, K., et al. (2011). The lifestyle of the segmented filamentous bacterium: a non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing. DNA Res. *18*, 291–303. https://doi.org/10.1093/dnares/dsr022.

32. Jacobson, M.J., Lin, G., Whittam, T.S., and Johnson, E.A. (2008). Phylogenetic analysis of Clostridium botulinum type A by multi-locus sequence typing. Microbiology (Reading, Engl) *154*, 2408–2415. https://doi.org/10.1099/mic.0.2008/016915-0.

33. Rawson, A.M., Dempster, A.W., Humphreys, C.M., and Minton, N.P. (2023). Pathogenicity and virulence of Clostridium botulinum. Virulence *14*, 2205251. https://doi.org/10.1080/21505594.2023.2205251.

34. Liang, W.-D., Bi, Y.-T., Wang, H.-Y., Dong, S., Li, K.-S., and Li, J.-S. (2013). Gene expression profiling of Clostridium botulinum under heat shock stress. Biomed Res. Int. *2013*, 1–8. https://doi.org/10.1155/2013/760904.

35. Carrasco, B., Torres, R., Moreno-Del Álamo, M., Ramos, C., Ayora, S., and Alonso, J.C. (2024). Processing of stalled replication forks in Bacillus subtilis. FEMS Microbiol. Rev. *48*, fuad065. https://doi.org/10.1093/femsre/fuad065.

36. Chakraborty, A., Halder, S., Kishore, P., Saha, D., Saha, S., Sikder, K., and Basu, A. (2022). The structure-function analysis of Obg-like GTPase proteins along the evolutionary tree from bacteria to humans. Genes Cells *27*, 469–481. https://doi.org/10.1111/gtc.12942.

37. Chatterjee, A., Cui, Y., and Chatterjee, A.K. (2009). RsmC of Erwinia carotovora subsp. carotovora negatively controls motility, extracellular protein production, and virulence by binding FlhD and modulating transcriptional activity of the master regulator. J. Bacteriol. *191*, 4582–4593. https://doi.org/10.1128/JB.00154-09.

38. Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.-S., Wernegreen, J.J., Sandström, J.P., Moran, N.A., and Andersson, S.G.E. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. Science *296*, 2376–2379. https://doi.org/10.1126/science.1071278.

39. Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol. Mol. Biol. Rev. *72*, 642–671, Table of Contents. https://doi.org/10.1128/MMBR.00020-08.

40. Burger, N.F.V., Nicolis, V.F., and Botha, A.-M. (2024). Host-specific coevolution likely driven by diet in Buchnera aphidicola. BMC Genom. *25*, 153. https://doi.org/10.1186/s12864-024-10045-3.

41. Liang, Y., Dikow, R.B., Su, X., Wen, J., and Ren, Z. (2024). Comparative genomics of the primary endosymbiont Buchnera aphidicola in aphid hosts and their coevolutionary relationships. BMC Biol. *22*, 137. https://doi.org/10.1186/s12915-024-01934-w.

42. Charles, H., Balmand, S., Lamelas, A., Cottret, L., Pérez-Brocal, V., Burdin, B., Latorre, A., Febvay, G., Colella, S., Calevro, F., and Rahbé, Y. (2011). A genomic reappraisal of symbiotic function in the aphid/Buchnera symbiosis: reduced transporter sets and variable membrane organisations. PLoS One *6*, e29096. https://doi.org/10.1371/journal.pone.0029096.

43. Cassone, B.J., Wenger, J.A., and Michel, A.P. (2015). Whole Genome Sequence of the Soybean Aphid Endosymbiont Buchnera aphidicola and Genetic Differentiation among Biotype-Specific Strains. J. Genomics *3*, 85–94. https://doi.org/10.7150/jgen.12975.

44. Deutscher, A.T., Tacchi, J.L., Minion, F.C., Padula, M.P., Crossett, B., Bogema, D.R., Jenkins, C., Kuit, T.A., Walker, M.J., and Djordjevic, S.P. (2012). Mycoplasma hyopneumoniae Surface proteins Mhp385 and Mhp384 bind host cilia and glycosaminoglycans and are endoproteolytically processed by proteases that recognize different cleavage motifs. J. Proteome Res. *11*, 1924–1936. https://doi.org/10.1021/pr201115v.

45. Siqueira, F.M., Gerber, A.L., Guedes, R.L.M., Almeida, L.G., Schrank, I.S., Vasconcelos, A.T.R., and Zaha, A. (2014). Unravelling the transcriptome profile of the Swine respiratory tract mycoplasmas. PLoS One *9*, e110327. https://doi.org/10.1371/journal.pone.0110327.

46. Leal Zimmer, F.M.A., Paes, J.A., Zaha, A., and Ferreira, H.B. (2020). Pathogenicity & virulence of Mycoplasma hyopneumoniae. Virulence *11*, 1600–1622. https://doi.org/10.1080/21505594.2020.1842659.

47. Yokobori, S.i., Kitamura, A., Grosjean, H., and Bessho, Y. (2013). Life without tRNAArg-adenosine deaminase TadA: evolutionary consequences of decoding the four CGN codons as arginine in Mycoplasmas and other Mollicutes. Nucleic Acids Res. *41*, 6531–6543. https://doi.org/10.1093/nar/gkt356.

48. Madeira, H.M.F., and Gabriel, J.E. (2007). Regulation of gene expression in Mycoplasmas: contribution from Mycoplasma hyopneumoniae and

Mycoplasma synoviae genome sequences. Genet. Mol. Biol. *30*, 277–282. https://doi.org/10.1590/S1415-47572007000200016.

49. Chen, A.Y., Fry, S.R., Forbes-Faulkner, J., Daggard, G.E., and Mukkur, T.K.S. (2006). Comparative immunogenicity of M. hyopneumoniae NrdF encoded in different expression systems delivered orally via attenuated S. typhimurium aroA in mice. Vet. Microbiol. *114*, 252–259. https://doi.org/10.1016/j.vetmic.2005.12.009.
50. Seely, S.M., and Gagnon, M.G. (2022). Mechanisms of ribosome recycling in bacteria and mitochondria: a structural perspective. RNA Biol. *19*, 662–677. https://doi.org/10.1080/15476286.2022.2067712.
51. Mathy, N., Pellegrini, O., Serganov, A., Patel, D.J., Ehresmann, C., and Portier, C. (2004). Specific recognition of rpsO mRNA and 16S rRNA by Escherichia coli ribosomal protein S15 relies on both mimicry and site differentiation. Mol. Microbiol. *52*, 661–675. https://doi.org/10.1111/j.1365-2958.2004.04005.x.
52. Bucca, G., Ferina, G., Puglia, A.M., and Smith, C.P. (1995). The dnaK operon of Streptomyces coelicolor encodes a novel heat-shock protein which binds to the promoter region of the operon. Mol. Microbiol. *17*, 663–674. https://doi.org/10.1111/j.1365-2958.1995.mmi_17040663.x.
53. Lara, B., Rico, A.I., Petruzzelli, S., Santona, A., Dumas, J., Biton, J., Vicente, M., Mingorance, J., and Massidda, O. (2005). Cell division in cocci: localization and properties of the Streptococcus pneumoniae FtsA protein. Mol. Microbiol. *55*, 699–711. https://doi.org/10.1111/j.1365-2958.2004.04432.x.
54. Matern, A., Pedrolli, D., Großhennig, S., Johansson, J., and Mack, M. (2016). Uptake and Metabolism of Antibiotics Roseoflavin and 8-Demethyl-8-Aminoriboflavin in Riboflavin-Auxotrophic Listeria monocytogenes. J. Bacteriol. *198*, 3233–3243. https://doi.org/10.1128/JB.00388-16.
55. Michaux, C., Gerovac, M., Hansen, E.E., Barquist, L., and Vogel, J. (2023). Grad-seq analysis of Enterococcus faecalis and Enterococcus faecium provides a global view of RNA and protein complexes in these two opportunistic pathogens. microLife *4*, uqac027. https://doi.org/10.1093/femsml/uqac027.
56. MacCann, R., Ghosh, T.S., Garcia Leon, A.A., Li, J., Negi, R., Gaillard, C., Saini, G., Tinago, W., Feeney, E.R., Yousif, O., et al. (2024). Associations between host microbiome and inflammation suggest role for host microbiome in driving COVID-19 disease severity. Microbes Infect. *26*, 105247. https://doi.org/10.1016/j.micinf.2023.105247.
57. Yadav, A., Devi, P., Kumari, P., Maurya, R., Shamim, U., and Pandey, R. (2024). Protocol to decode the role of transcriptionally active microbes in SARS-CoV-2-positive patients using an RNA-seq-based approach. STAR Protoc. *5*, 103071. https://doi.org/10.1016/j.xpro.2024.103071.
58. Al-Momani, H., Nelson, A., Al Balawi, H., Al Balawi, D., Aolymat, I., Khasawneh, A.I., Tabl, H., Alsheikh, A., Zueter, A.M., Pearson, J., and Ward, C. (2025). Assessment of upper respiratory and gut bacterial microbiomes during COVID-19 infection in adults: potential aerodigestive transmission. Sci. Rep. *15*, 1811. https://doi.org/10.1038/s41598-025-85806-5.
59. Wang, D., Duan, Y., He, L., Jiang, J., Xian, J., Yuan, K., Zhang, R., Zhang, H., Wang, J., Li, N., et al. (2025). Altered microbiota of the lower respiratory tract and its association with COVID-19 severity analysed by metagenomics and metatranscriptomics. Commun. Biol. *8*, 804. https://doi.org/10.1038/s42003-025-08234-1.
60. Chen, X., Liao, B., Cheng, L., Peng, X., Xu, X., Li, Y., Hu, T., Li, J., Zhou, X., and Ren, B. (2020). The microbial coinfection in COVID-19. Appl. Microbiol. Biotechnol. *104*, 7777–7785. https://doi.org/10.1007/s00253-020-10814-6.
61. Mirzaei, R., Goodarzi, P., Asadi, M., Soltani, A., Aljanabi, H.A.A., Jeda, A.S., Dashtbin, S., Jalalifar, S., Mohammadzadeh, R., Teimoori, A., et al. (2020). Bacterial co-infections with SARS-CoV-2. IUBMB Life *72*, 2097–2111. https://doi.org/10.1002/iub.2356.
62. Leszczyńska, K., Święcicka, I., Daniluk, T., Lebensztejn, D., Chmielewska-Deptuła, S., Leszczyńska, D., Gawor, J., and Kliber, M. (2023). Escherichia albertii as a Potential Enteropathogen in the Light of Epidemiological and Genomic Studies. Genes *14*, 1384. https://doi.org/10.3390/genes14071384.
63. Crofts, A.A., Poly, F.M., Ewing, C.P., Kuroiwa, J.M., Rimmer, J.E., Harro, C., Sack, D., Talaat, K.R., Porter, C.K., Gutierrez, R.L., et al. (2018). Campylobacter jejuni transcriptional and genetic adaptation during human infection. Nat. Microbiol. *3*, 494–502. https://doi.org/10.1038/s41564-018-0133-7.
64. Kappler, U., Henningham, A., Nasreen, M., Yamamoto, A., Buultjens, A.H., Stinear, T.P., Sly, P., and Fantino, E. (2024). Tolerance to Haemophilus influenzae infection in human epithelial cells: Insights from a primary cell-based model. PLoS Pathog. *20*, e1012282. https://doi.org/10.1371/journal.ppat.1012282.
65. Zuluaga, N., Martínez, D., Hernández, C., Ballesteros, N., Castañeda, S., Ramírez, J.D., and Muñoz, M. (2023). Description of pathogenic bacteria in patients with respiratory symptoms associated with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Colombia. Ann. Clin. Microbiol. Antimicrob. *22*, 56. https://doi.org/10.1186/s12941-023-00595-x.
66. Choi, H.J., Jeong, S.H., Shin, K.S., Kim, Y.A., Kim, Y.R., Kim, H.S., Shin, J.H., Shin, J.H., Uh, Y., Bae, S., et al. (2022). Characteristics of Escherichia coli Urine Isolates and Risk Factors for Secondary Bloodstream Infections in Patients with Urinary Tract Infections. Microbiol. Spectr. *10*, e01660-22. https://doi.org/10.1128/spectrum.01660-22.
67. Lara, A.C., Corretto, E., Kotrbová, L., Lorenc, F., Petříčková, K., Grabic, R., and Chroňáková, A. (2021). The Genome Analysis of the Human Lung-Associated Streptomyces sp. TR1341 Revealed the Presence of Beneficial Genes for Opportunistic Colonization of Human Tissues. Microorganisms *9*, 1547. https://doi.org/10.3390/microorganisms9081547.
68. Miller, A.K., and Williams, S.M. (2021). Helicobacter pylori infection causes both protective and deleterious effects in human health and disease. Genes Immun. *22*, 218–226. https://doi.org/10.1038/s41435-021-00146-4.
69. Balboa, L., Baena, A., and Carreño, L.J. (2022). Editorial: innate immune cells in the control of intracellular bacteria. Front. Cell. Infect. Microbiol. *12*, 830203. https://doi.org/10.3389/fcimb.2022.830203.
70. Oelen, R., de Vries, D.H., Brugge, H., Gordon, M.G., Vochteloo, M., single-cell eQTLGen consortium; BIOS Consortium; Ye, C.J., Westra, H.-J., Franke, L., and van der Wijst, M.G.P. (2022). Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. Nat. Commun. *13*, 3267. https://doi.org/10.1038/s41467-022-30893-5.
71. Zheng, D., Liwinski, T., and Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. Cell Res. *30*, 492–506. https://doi.org/10.1038/s41422-020-0332-7.
72. Kaufmann, S.H.E. (2011). Intracellular pathogens: living in an extreme environment. Immunol. Rev. *240*, 5–10. https://doi.org/10.1111/j.1600-065X.2010.01001.x.
73. Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H.B., Trombetta, J.J., Satija, R., Shalek, A.K., Xavier, R.J., Regev, A., and Hung, D. (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell *162*, 1309–1321. https://doi.org/10.1016/j.cell.2015.08.027.
74. Bogdan, C. (2008). Mechanisms and consequences of persistence of intracellular pathogens: leishmaniasis as an example. Cell. Microbiol. *10*, 1221–1234. https://doi.org/10.1111/j.1462-5822.2008.01146.x.
75. Boulanger, M., Molina, E., Wang, K., Kickler, T., Xu, Y., and Garibaldi, B.T. (2021). Peripheral Plasma Cells Associated with Mortality Benefit in Severe COVID-19: A Marker of Disease Resolution. Am. J. Med. *134*, 1029–1033. https://doi.org/10.1016/j.amjmed.2021.01.040.
76. Lee, N., and Kim, W.-U. (2017). Microbiota in T-cell homeostasis and inflammatory diseases. Exp. Mol. Med. *49*, e340. https://doi.org/10.1038/emm.2017.36.

77. Amarilla-Irusta, A., Zenarruzabeitia, O., Sevilla, A., Sandá, V., Lopez-Pardo, A., Astarloa-Pando, G., Pérez-Garay, R., Pérez-Fernández, S., Meijide, S., Imaz-Ayo, N., et al. (2024). CD151 identifies an NK cell subset that is enriched in COVID-19 patients and correlates with disease severity. J. Infect. *89*, 106304. https://doi.org/10.1016/j.jinf.2024.106304.

78. Saichi, M., Ladjemi, M.Z., Korniotis, S., Rousseau, C., Ait Hamou, Z., Massenet-Regad, L., Amblard, E., Noel, F., Marie, Y., Bouteiller, D., et al. (2021). Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. Nat. Cell Biol. *23*, 538–551. https://doi.org/10.1038/s41556-021-00681-2.

79. Zhao, Y., Feng, R., Zheng, G., Tian, J., Ruan, L., Ge, M., Jiang, W., and Lu, Y. (2015). Involvement of the TetR-Type Regulator PaaR in the Regulation of Pristinamycin I Biosynthesis through an Effect on Precursor Supply in Streptomyces pristinaespiralis. J. Bacteriol. *197*, 2062–2071. https://doi.org/10.1128/JB.00045-15.

80. Winheim, E., Rinke, L., Lutz, K., Reischer, A., Leutbecher, A., Wolfram, L., Rausch, L., Kranich, J., Wratil, P.R., Huber, J.E., et al. (2021). Impaired function and delayed regeneration of dendritic cells in COVID-19. PLoS Pathog. *17*, e1009742. https://doi.org/10.1371/journal.ppat.1009742.

81. Abdelwahed, E.K., Hussein, N.A., Moustafa, A., Moneib, N.A., and Aziz, R.K. (2022). Gene Networks and Pathways Involved in Escherichia coli Response to Multiple Stressors. Microorganisms *10*, 1793. https://doi.org/10.3390/microorganisms10091793.

82. Urusova, D.V., Merriman, J.A., Gupta, A., Chen, L., Mathema, B., Caparon, M.G., and Khader, S.A. (2022). Rifampin resistance mutations in the rpoB gene of Enterococcus faecalis impact host macrophage cytokine production. Cytokine *151*, 155788. https://doi.org/10.1016/j.cyto.2021.155788.

83. Shan, L., He, P., and Sheen, J. (2007). Intercepting host MAPK signaling cascades by bacterial type III effectors. Cell Host Microbe *1*, 167–174. https://doi.org/10.1016/j.chom.2007.04.008.

84. Nandi, I., and Aroeti, B. (2023). Mitogen-Activated Protein Kinases (MAPKs) and Enteric Bacterial Pathogens: A Complex Interplay. Int. J. Mol. Sci. *24*, 11905. https://doi.org/10.3390/ijms241511905.

85. Zaika, A.I., Wei, J., Noto, J.M., and Peek, R.M. (2015). Microbial regulation of p53 tumor suppressor. PLoS Pathog. *11*, e1005099. https://doi.org/10.1371/journal.ppat.1005099.

86. Yin, S., Yu, J., Hu, B., Lu, C., Liu, X., Gao, X., Li, W., Zhou, L., Wang, J., Wang, D., et al. (2018). Runx3 mediates resistance to intracellular bacterial infection by promoting IL12 signaling in group 1 ILC and NCR+ILC3. Front. Immunol. *9*, 2101. https://doi.org/10.3389/fimmu.2018.02101.

87. Skerniskyte, J., Barroso, M.V., Chicher, J., Hammann, P., Demais, V., Wright, K., Mostowy, S., and Marteyn, B.S. (2025). Neutrophils display antibacterial defense via non-canonical LC3 decoration of extracellular bacteria. Microbes Infect. *27*, 105545. https://doi.org/10.1016/j.micinf.2025.105545.

88. Duell, B.L., Tan, C.K., Carey, A.J., Wu, F., Cripps, A.W., and Ulett, G.C. (2012). Recent insights into microbial triggers of interleukin-10 production in the host and the impact on infectious disease pathogenesis. FEMS Immunol. Med. Microbiol. *64*, 295–313. https://doi.org/10.1111/j.1574-695X.2012.00931.x.

89. Iwaszko, M., Biały, S., and Bogunia-Kubik, K. (2021). Significance of Interleukin (IL)-4 and IL-13 in Inflammatory Arthritis. Cells *10*, 3000. https://doi.org/10.3390/cells10113000.

90. Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at bioRxiv. https://doi.org/10.1101/2021.05.05.442755.

91. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

92. Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019). CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. *47*, D721–D728. https://doi.org/10.1093/nar/gky900.

93. Franzén, O., Gan, L.-M., and Björkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford) *2019*, baz046. https://doi.org/10.1093/database/baz046.

94. Lu, J., Rincon, N., Wood, D.E., Breitwieser, F.P., Pockrandt, C., Langmead, B., Salzberg, S.L., and Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. Nat. Protoc. *17*, 2815–2839. https://doi.org/10.1038/s41596-022-00738-y.

95. Breitwieser, F.P., and Salzberg, S.L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics *36*, 1303–1304. https://doi.org/10.1093/bioinformatics/btz715.

96. Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. Nat. Methods *10*, 1200–1202. https://doi.org/10.1038/nmeth.2658.

97. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

98. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

99. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

100. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

101. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

102. Radhakrishnan, A., Jain, Y., Uhler, C., and Lander, E.S. (2025). Efficiently quantifying dependence in massive scientific datasets using InterDependence Scores. Proc. Natl. Acad. Sci. USA *122*, e2509860122. https://doi.org/10.1073/pnas.2509860122.

103. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

104. Chattopadhyay, P., Khare, K., Kumar, M., Mishra, P., Anand, A., Maurya, R., Gupta, R., Sahni, S., Gupta, A., Wadhwa, S., et al. (2022). Single-cell multiomics revealed the dynamics of antigen presentation, immune response and T cell activation in the COVID-19 positive and recovered individuals. Front. Immunol. *13*, 1034159. https://doi.org/10.3389/fimmu.2022.1034159.

105. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257. https://doi.org/10.1186/s13059-019-1891-0.

106. Unterman, A., Sumida, T.S., Nouri, N., Yan, X., Zhao, A.Y., Gasque, V., Schupp, J.C., Asashima, H., Liu, Y., Cosme, C., et al. (2022). Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. Nat. Commun. *13*, 440. https://doi.org/10.1038/s41467-021-27716-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TRUPCR® SARS-CoV-2 RT qPCR Kit	3B BlackBio	Cat# 3B306
BD Human single-cell multiplexing kit	Becton Dickinson	Cat# 633781
BD Rhapsody WTA amplification kit	Becton Dickinson	Cat# 633801
BD Rhapsody cDNA kit	Becton Dickinson	Cat# 633773
AMPure XP	Beckman Coulter	Cat# A63881
Qubit dsDNA HS Assay kit	Invitrogen	Cat# Q32854
Agilent 2100 Bioanalyzer	Agilent	Cat# 5067-4626
NovaSeq 6000 S4 reagent kit (300 cycles)	Illumina	Cat# 20028312
Deposited data		
Raw and analyzed data single cell data	This paper	SRA BioProject: PRJNA1293800 GEO: GSE303208
Public data analyzed for comparative analysis	https://doi.org/10.1016/j.isci.2022.104034[22]	*BioStudies*, Accession: E-MTAB-9467. Retrieved from https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-9467
Software and algorithms		
bcl2fastq 2.19	NA	GitHub - brwnj/bcl2fastq: NextSeq specific bcl2fastq2 wrapper.
STARsolo 2.7.9a	Kaminow et al.[90]	https://github.com/alexdobin/STAR/releases
STAR 2.5.2b	Dobin et al.[91]	http://biocc.hrbmu.edu.cn/CellMarker
CellMarkerDB	Zhang et al.[92]	http://biocc.hrbmu.edu.cn/CellMarker
PanglaoDB	Franzén et al.[93]	https://panglaodb.se/index.html
PathogenTrack pipeline	Zhang et al.[20]	https://github.com/ncrna/PathogenTrack
Kraken 2 2.1.2	Lu et al.[94]	https://github.com/DerrickWood/kraken2
Pavian 1.2.1	Breitwieser et al.[95]	https://github.com/fbreitwieser/pavian
Prism 9	GraphPad	https://www.graphpad.com/
Rawgraphs 2.0 beta	NA	https://app.rawgraphs.io/
metagenomeSeq 1.46.0	Paulson et al.[96]	https://github.com/HCBravoLab/metagenomeSeq
phyloSeq 1.48.0	McMurdie et al.[97]	https://github.com/joey711/phyloseq
Minikraken build version v1 dated 3/2020	NA	https://benlangmead.github.io/aws-indexes/k2
Custom microbial analysis and demultiplexing scripts	This paper	https://github.com/INGEN-HOPE/Intracellular-Microbial-Analaysis-of-BD-Rhapsody-Wta-Abseq-data
SAMtools 1.6.0	Li et al.[98]	https://github.com/samtools/samtools
Bowtie 2.4.4	Langmead and Salzberg[99]	https://github.com/BenLangmead/bowtie2
Bwa-mem2 2.2.1	Li and Durbin[100]	https://github.com/bwa-mem2/bwa-mem2
FeatureCounts 2.0.1	Liao et al.[101]	https://subread.sourceforge.net/
Seurat 5.3.0	Stuart et al., 2019	https://github.com/satijalab/seurat
R 4.5.0	NA	https://cran.r-project.org/
Vegan 2.6.1	Radhakrishnan et al.[102]	https://github.com/vegandevs/vegan
rstatix 0.7.2	NA	https://github.com/kassambara/rstatix
Stats 4.4.1	NA	https://svn.r-project.org/R/trunk/src/library/stats/
ClusterProfiler 4.18.4	Wu et al.[103]	https://github.com/YuLab-SMU/clusterProfiler
Reactome.db 1.95.0	Bioconductor	https://bioconductor.org/packages/Reactome.db
SRA-tools 3.0.0	NCBI	https://github.com/ncbi/sra-tools

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human subjects and clinical protocol

All study procedures adhered to the Declaration of Helsinki, local legislation, and institutional requirements. The study protocol was approved by the Institutional Ethics Committee (IEC) of CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB) (Ref No: CSIR-IGIB/IHEC/2020-21/01). The patients provided their written informed consent before participating in this study.

To investigate the functional relevance of intracellular microbes in COVID-19 infection and recovery, blood samples were collected from 57 participants. Participants were allocated to experimental groups based on COVID-19 qRT-PCR status and recovery time point: healthy individuals who had never tested positive for COVID-19 ($n = 9$), qRT-PCR-positive COVID-19 patients ($n = 24$), recently recovered individuals sampled after qRT-PCR negativity ($n = 16$), and long-recovered follow-up individuals sampled at a later recovery time point ($n = 8$). The patients and participants provided their written informed consent before they participated in this study. 5 mL of venous blood was drawn from the cubital vein using EDTA vials. Samples were collected from the Max Super Speciality Hospital, Saket. The clinical characteristics of these patients are provided in Table S1. Table S1 summarizes the demographic and clinical characteristics of the study cohort, including age, sex, comorbidities, oxygen saturation (SpO_2), and C-reactive protein (CRP) measurements. Both male and female participants were included across all study groups, and the sex distribution was comparable between groups. Sex-specific effects were not separately evaluated in downstream analyses due to limited subgroup sample sizes. Initially, 16 samples were collected from individuals in the recovered group at a single time point. A second set of samples was later collected from 8 of these individuals to capture a subsequent recovery phase. During quality control, 3 of the initial 16 samples were excluded due to poor quality. This included 2 individuals who also had samples collected at the second time point. Despite this, all 8 s time-point samples were retained for analysis, as comparisons were made across broader clinical groups (9 healthy, 14 infected, 13 recovered, and 8 long-recovered) rather than between time points within recovered individuals.

No cell lines were used in this study; therefore, cell line authentication and mycoplasma contamination testing are not applicable.

METHOD DETAILS

PBMC isolation

PBMCs were isolated from freshly collected blood using Histopaque-1077 (Sigma-Aldrich) density gradient centrifugation. Briefly, 10 mL of Histopaque was added to a 50-mL conical tube, followed by careful layering of 10 mL of whole blood mixed with PBS at a 1:1 ratio (5 mL blood +5 mL PBS). Samples were centrifuged at 400 × g for 20 min at room temperature without any brake. The plasma layer was removed, and the PBMCs at the interface were transferred to a sterile tube and washed twice with sterile PBS (at 400 × g, 10 min) and resuspended in freezing medium (90% FBS, 10% DMSO). Cells were initially cryopreserved at −80°C for 24 h in an isopropyl alcohol cryocooler before transfer to liquid nitrogen for long-term storage.

Single-cell capture and cDNA synthesis

We followed the protocol as described by Chattopadhyay et al. with some modification.[104] PBMCs were retrieved from the liquid nitrogen storage and transferred to an isopropyl alcohol cryocooler, where they gradually thawed until the internal temperature reached 25°C. To complete thawing, samples were briefly incubated in a 37°C water bath for approximately 2 min, ensuring the removal of any residual ice crystals. Thawed cells were immediately quenched with 2 mL of pre-warmed (37°C) 1× phosphate-buffered saline (PBS; Thermo Fisher Scientific) supplemented with 2% fetal bovine serum (FBS). The volume was then adjusted to 10 mL using PBS +2% FBS. The suspension was centrifuged at 400 × g for 5 min at room temperature, and the supernatant was discarded. The resulting cell pellet was resuspended in PBS +2% FBS, filtered through a 40 μm cell strainer (Corning), and cells were counted using trypan blue exclusion.

Briefly, 0.2 million cells per sample were labeled using the BD Single-Cell Multiplexing Kit-Human (Cat. No.: 633781) and 50 BD AbSeq Ab-Oligos, following the manufacturer's instructions (Doc ID: 214419 Rev. 2.0). After labeling, the cells were washed twice to remove unbound tags and antibodies, counted, and resuspended in the appropriate buffer for single-cell capture (4 samples at a time, ~6,000 cells per sample). The prepared cell suspension was then loaded into a cartridge on the BD Rhapsody Express Single-Cell Analysis System for single-cell capture using microwell-based partitioning. Following cell lysis, released mRNA was captured on the oligo(dT)-coated cell capture beads, and reverse transcription was carried out on the beads, generating barcoded cDNA tagged with Unique Molecular Identifiers (UMIs) (Cat. No.: 633773).[17]

In parallel, a negative control consisting of an empty cartridge loaded with nuclease-free water was processed through the same workflow up to the random priming and extension step. TapeStation profile of this control showed no detectable library peaks and insufficient DNA for further processing, indicating minimal background contamination during library preparation.

Library preparation and sequencing

Library preparation was performed according to the BD Rhapsody System mRNA Whole Transcriptome Analysis (WTA) and Sample Tag protocol, alongside the BD AbSeq protocol for antibody-derived tag (ADT) sequencing. Briefly, cDNA was amplified using the BD Rhapsody Whole Transcriptome Amplification (WTA) Kit (Doc ID: 210967 Rev. 1.0)—a random priming and extension approach, followed by an indexing step. Concurrently, extended Sample Tags and AbSeq oligonucleotide tags were denatured from the capture

beads and further amplified and indexed to generate sequencing-ready libraries. Libraries were quantified using a Qubit 4.0 fluorimeter (Invitrogen, USA). Library fragment size was checked using the Bioanalyzer (catalog no. 5067-4626). The final processed WTA, Sample Tag, and AbSeq libraries were pooled based on their indexing and sequenced on the Illumina NovaSeq 6000 platform. WTA libraries were sequenced at a depth of 30,000 reads per cell using an 85 bp Read1 and 215 bp Read2 sequencing configuration, while Sample Tag (SMK) and AbSeq libraries were sequenced at a depth of 360 reads per cell per sample tag and 500 reads per AbSeq per cell, with a 51 bp Read1 and 75 bp Read2 sequencing read length, using NovaSeq 6000.

scRNA-seq data pre-processing, cell clustering, and annotation

Illumina basecall files (*.bcl) were converted to FASTQ format using bcl2fastq2 (v2.19.0.316). Sample demultiplexing has been done using a custom script with the sample tag sequence of each sample from batch-level fastq files. The script allows for 1 base mismatch while identifying the sample tag. Reads were then checked for quality using FastQC, and low-quality reads as well as adapters were filtered out using cutadapt. R1 reads captured cell labels and UMIs, while R2 reads captured RNA sequence reads that were aligned to the human genome (GRCh38) using STARSolo (v2.7.9a). For each set, over 68.6% of the QC-passed reads aligned uniquely to the transcriptome. A total of 191,417 cells were captured from 57 samples. Cells were deduplicated, followed by quality control filtering with a filtration criterion of cells with fewer nGene of >60, nUMI >100, <3500 a mitochondrial gene content exceeding 20%, after which 156,151 cells were left. Additionally, samples containing fewer than 500 cells were excluded to ensure sufficient cellular representation and robust downstream analysis, resulting in 44 samples and 153,075 cells. Cells were clustered using a semi-supervised clustering approach, using a list of cell-type-specific immune-related genes. Cell clustering was found to be optimized at a variable resolution of 0.3–0.6, after iterative selection through comparing multiple resolutions from 0.1 to 1, which is further visualized through t-distributed Stochastic Neighbor Embedding (t-SNE). Cluster-defining marker genes were identified using the Wilcoxon rank-sum test via the FindAllMarkers function, applying a $\log_2$ fold-change cut-off of 1.5 and a p-adjusted value of less than 0.05. Annotation has been done by combining reference-based annotation with manual curation using CellMarker DB and PanglaoDB references based on cluster markers. We defined 13 distinct cell populations based on these marker profiles. AbSeq sequencing data were demultiplexed from WTA using a custom script. Paired-end FASTQ files were aligned to the antibody reference sequence using Bowtie2 (v2.4.4), and alignments were processed with SAMtools (v1.10). A cell-by-feature count matrix was generated by retaining unique combinations of cell barcode, antibody, and UMI, followed by aggregation of antibody counts per cell. Seurat objects were created from sample-wise count matrices using the ADT assay. Data were normalized using centered log-ratio (CLR) transformation and scaled for downstream analysis. Cell-level metadata were incorporated, and only annotated cells were retained for analysis and antibody marker expression visualization.

Meta-transcriptomic analysis

QC-passed reads were then analyzed in the PathogenTrack pipeline,[20] first mapped with the human reference using the STAR[91] with the human genome (GRCh38) to remove human reads and retain unmapped reads. Using the unmapped files, PathogenTrack ran Kraken2[105] with database reference Minikraken build version v1 8gb dated 3/2020 and generated a cell-level count matrix, along with the Kraken output files. Kraken reports were quantified at the species level to generate a species-level read count matrix using the Pavian tool (v 1.2.1)[94] in R software. Further, it was normalized with the CSS log normalization method using the ''cumNorm'' function of the ''metagenomeSeq'' (v1.46.0) R package. Microbial relative abundance was calculated by dividing the read count for each taxon by the total read count per sample to obtain the proportion of reads assigned to each taxon. Alpha diversity of taxa among the groups was calculated using the ''estimate richness'' function of the ''phyloseq'' (v1.48.0) R package, considering observed Shannon index measures. Beta diversity (Bray-Curtis dissimilarity) was assessed with the Bray distance calculated using the ''vegdist'' function of the ''vegan'' (v2.6.10) R package. PCoA was used to ordinate each dissimilarity matrix, and the variance among the experimental groups was assessed using the ''cmdscale'' function of the ''stats'' (v4.4.1) R package and visualized in 2D as PCoA with a density plot. From the total identified microbes, 110 abundant species with a relative abundance greater than 2% within the experimental groups. Among these, 50 overlapping species between the groups were further analyzed. Additionally, the abundant phylum and genus of the taxa across groups with greater than 1% relative abundance were selected. For severity comparison, patients were stratified based on oxygen saturation ($SpO_2 \geq 94\%$: mild; $< 94\%$: severe). Microbial profiles were compared between groups by analyzing unmapped and microbial-assigned reads, alpha diversity (richness and evenness), and beta diversity (community composition). Due to the limited sample size of the severe group, this analysis was considered exploratory.

Comparative analysis of dengue and COVID-19 single-cell RNA-seq datasets

Single-cell RNA-seq data from peripheral blood mononuclear cells (PBMCs) of dengue-infected and healthy individuals (ArrayExpress: E-MTAB-9467)[22] and COVID-19 patients (NCBI SRA: SRR16922265, SRR16922266, SRR16922269; PRJNA648991)[106] were analyzed to assess mapping efficiency. Raw sequencing reads were processed using SRA-tools (v3.0.0) and aligned to the human reference genome with STAR (v2.7.10b) in a Conda-managed environment. Alignment files were handled using SAMtools (v1.6). Mapping statistics were extracted from the STAR Log.final.out file. Data processing and visualization were performed in R (v4.5.0) using ggplot2 for comparative analysis.

Differential microbial gene analysis

The presence of microbes at the cellular level was quantified through the pathogen track and was further normalized by the CSS normalization method using the "cumNorm" function of the "metagenomeSeq" (v1.46.0) R package, which accounts for the assumption that microbes are shared across the cells. The cellular proportion of each microbe was determined by calculating the ratio of cells in which the microbe was detected (positive cells) to the total number of cells within each condition and cell type. Classified microbial reads of selected abundant microbes at the cellular level were extracted using extract_kraken_reads.py of krakentools (v1.2) and further aligned with their microbial genomes using bwa-mem2 (v2.2.1)[100] and then quantified using featureCounts (v2.0.1),[101] and their cellular identity was extracted from the aligned read IDs using samtools (v1.3.1).[98]

Host cell transcriptomic analysis

Cells were classified based on microbial read abundance, retaining only high-confidence microbe-positive ($\geq$10 UMIs) and microbe-negative (0 UMIs) cells while excluding low-confidence cells (1–9 UMIs). Differential gene expression (DGE) analysis between these groups was performed using the "FindMarkers" function of the "Seurat" package (v5.3.0) with the Wilcoxon test. Significant DEGs were defined by adjusted $p < 0.05$ and $|\text{avg\_log2FC}| \geq 1.5$. Functional enrichment analysis was conducted using clusterProfiler (v4.18.4) and ReactomePA (v1.54.0) along with reactome.db (v1.95.0) R package.

At the cellular level, DGE analysis was performed within immune cell subsets across experimental groups, excluding clusters with fewer than 25 cells per group. Enrichment analysis of these subset-specific DEGs was carried out using Enrichr (March 2026). To identify pathogen-specific host responses, we further analyzed high-confidence unique pathogen single-microbe-positive cells ($\geq$10 reads per microbe per cell) and compared them to matched microbe-negative cells within the same cell type and condition. For host microbe gene correlation analysis, rpsO was selected as the microbial feature due to its high prevalence across cells. We aggregated cells with detectable rpsO expression into a pseudobulk profile and retained host genes with $\geq$10 reads. Spearman correlation between host gene expression and rpsO levels was computed, followed by Benjamini–Hochberg correction. Significance was defined as adjusted p value <0.05 and rho $|\rho| \geq 0.5$.

QUANTIFICATION AND STATISTICAL ANALYSIS

The significance of the distribution of overlapping microbial species among the experimental group was assessed using the Wilcoxon test by the R function "wilcox_test" of the "rstatix" package (v0.7.2), and the significance of alpha diversity among groups was also calculated with the same, and beta diversity significance of multiple predictors was calculated through multivariate PERMANOVA using the "adonis2" function with 999 permutations of the "vegan" (v2.6.10) R package. Differential microbes and their significance at the cellular level were calculated using the FindMarkers function with the default Wilcoxon test in the Seurat (v5.3.0) R package.

ADDITIONAL RESOURCES

This study did not generate additional resources.