# Host-strain compatibility influences transcriptional responses in *Mycobacterium tuberculosis* infections

Paula Ruiz-Rodriguez[1☯], Marta Caballer-Gual[1☯], Guillem Santamaria[2], Hellen Hiza[1], Mireia Coscolla[1*]

1. Institute for Integrative Systems Biology (I[2]SysBio), University of Valencia-CSIC, FISABIO Joint Research Unit Infection and Public Health, Valencia, Spain.
2. Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-Belval Esch-sur-Alzette, Luxembourg.

*Corresponding author: mireia.coscolla@csic.es

☯These authors contributed equally

**Keywords**

*Mycobacterium tuberculosis*, host-pathogen interactions, RNA-seq, transcriptomics, infection ratio, Chimpanzee bacillus, *Mycobacterium africanum*, *Mycobacterium bovis.*

# Abstract

Tuberculosis, caused by *Mycobacterium tuberculosis*, is a leading cause of infectious mortality. It affects humans and other mammals, including cattle, with significant health, ecological, and economic impacts. Understanding the drivers of its host-specificity shapes the landscape of infectivity, which could potentially inform strategies for tuberculosis management. We hypothesize that host-strain compatibility influences host responses to *M. tuberculosis* and we aim to reveal these transcriptional responses during *in vitro* infections.

We infected human and bovine cell lines with two human-associated *M. tuberculosis* strains from lineages 5 and 6, as well as two animal-associated strains (*Mycobacterium bovis* and the Chimpanzee bacillus), and quantified infection ratios, cell death, and transcriptional responses. By integrating those data and comparing responses from different compatibility combinations, we identified global transcription profiles determined by strain-host compatibility.

Our results show that the most compatible host-strain combinations had higher infection rates and a greater proportion of necrotic cells, indicating a more pathogenic phenotype *in vitro.* We also identified distinct transcriptional patterns for high and low compatibility infections. Both hosts had similar transcriptional responses to their corresponding strains, aiming to increase cell proliferation. This proliferation could favor bacterial growth, explaining the success of the strain in its corresponding host. Conversely, hosts responded to non-associated strains with a more effective immune response, among other strategies, explaining their lower success in the non-associated host.

In conclusion, our hypothesis was confirmed, showing that bacteria-host compatibility determines common transcriptional responses, but also specific responses that depend on the infecting strain. This study enhances the understanding of host-specific adaptation mechanisms in different species.


# Introduction

Pathogen specificity, the phenomenon where pathogens exhibit a preference for infecting certain hosts over others, is a cornerstone concept in understanding the dynamics of infectious diseases. There are ecological, epidemiological and biological drivers of such specificity, where hosts provide more permissive conditions to support the pathogen's life cycle. Biologically, this specificity can also result from molecular incompatibilities between interacting molecules resulting in reduced efficiency entering the cell, immune system evasion, or replication within the host[1–3]. The implications of host specificity extend beyond basic science, affecting disease control strategies, vaccine development, and the management of emerging infectious diseases, especially within the context of One-Health.

Pathogens displaying host specificity, such as influenza A, exhibit discrete molecular changes that are responsible for the host specificity[4–6]. Another example could be the specificity of coronavirus for different hosts, where research into cross-species transmission events improves understanding of potential zoonotic spillover events[7–9] . However, in other pathogens like Brucellaceae, where different species infect different hosts, pinpointing such changes is challenging presumably because of the ancestral nature of host adaptations[10]. *Mycobacterium tuberculosis* also displays host specificity across different mammals[11], and the study of the molecular determinants of such specificity extends beyond basic science, and can potentially inform the management of tuberculosis.

Tuberculosis ranks among the leading causes of death from a single infectious agent, and it is caused by members of the *M. tuberculosis*[12]. This complex includes human-associated *M. tuberculosis*, which consists of ten phylogenetic lineages (L1 to L10), and four animal-associated lineages: A1 to A4[13–16]. Lineage A1 infects wild terrestrial animals in Africa, and is the least studied among the animal-associated lineages despite being the closest relative to the human-associated Lineage 6[17]. In contrast, Lineage A4, which includes *Mycobacterium bovis* and *Mycobacterium caprae,* is the most studied animal-associated clade and is mostly found in cattle and goats, respectively. Despite the preference of certain lineages for specific hosts, there is no absolute exclusivity in host tropism[18]. For example, wildlife reservoirs such as wild boars, badgers, or ferrets among others, can also harbor *M. bovis*, leading to spillback into cattle, which can, in turn, infect wildlife[19]. *M. bovis* does occasionally infect humans, but human-to-human transmission of *M. bovis* is uncommon[20,21]. On the other hand, human associated lineages are only occasionally isolated from cattle or other animals that come in contact with humans[13,17,22,23]. Additionally, human-associated lineages are attenuated in animals, and transmission from an animal back to humans is extremely rare[24,25].

Host jumps are concentrated in the clade encompassing all four animal-associated lineages as well as the human-associated Lineages (L) 5, 6, 9, and 10, which are among the least studied members of the species. Among these, L5, L6, and A4 have the most data available. Human-associated L5 and L6 are often referred to as *Mycobacterium africanum*, though they differ significantly. While L5 is associated with Ewe Ethnicity in West Africa[26,27], Lineage 6 is more diverse, shares a more recent common ancestor with the animal-associated clades, and it has been associated with HIV infection[28,29]. These factors suggest a potential difference in host specificity between L5 and L6, with L6 possibly having a more generalistic nature compared to L5, which has been associated with the Ewe ethnicity[27]. Conversely, Lineage A4 is more commonly associated with cattle, but the host range of other animal-associated lineages is not as clear due to the scarce data available. The lack of data is particularly significant for A1, including the Chimpanzee bacillus[30] which has been isolated only once despite being very closely related to L6.

Our goal is to advance the understanding of the molecular determinants of *M. tuberculosis* specificity using an *in vitro* infection system. We hypothesize that the Lineage 5 strain (L5) and the *M. bovis* (A4) strain will display a common pattern of high-virulence-related

phenotypes in human and bovine cells, respectively, due to their higher compatibility. Conversely, Lineage 6 (L6) and the Chimpanzee bacillus (A1) will not exhibit such a clear pattern of virulence-related phenotypes in either host cell. To test this, we will examine whether the epidemiological compatibility between four different strains and their most common hosts can be partially mimicked by more successful infections in *in vitro* experiments. If so, we aim to identify the molecular mechanisms involved through expression analysis.

# Results

## *In vitro* infection ratio and cell death matches with preferences by host

To assess the compatibility between *M. tuberculosis* lineages with their host cells, we compared infection ratios across the four bacterial-host combinations at four hours post infection (hpi), one day and three days post-infection (dpi). We could detect that L5 showed significantly higher median infection ratio than *M. bovis* in human THP-1 cells (49.92% vs 18.92%; Mann-Whitney U test p-value = 0.00017; Figure 1A), while *M. bovis* showed significantly higher median infection ratios in bovine BoMac cells than L5 (41.52% vs 25.82%, Mann-Whitney U test p-value = 0.043; Figure 1A) at 3 dpi. Interestingly, L6 and the Chimpanzee bacillus did not show differences in infection ratio between cell lines, although the observed tendency suggests a higher infection ratio in human THP-1 cells. (Figure 1A).

When we group the high compatibility combinations (*M. bovis* in BoMac and L5 in THP-1) and compare it with the lower compatibility (*M. bovis* in THP-1 and L5 in BoMac), or with the infections of the Chimpanzee bacillus and L6 for which the compatibility is not as clear (undetermined from now on). We detected that high compatibility infections showed significantly higher infection ratios than the rest (Figure 1B) at all time points. Specifically, at 3 dpi, the undetermined compatibility infections showed an intermediate infection ratio, being significantly smaller than the high compatibility combinations (28.10% vs 47.40%, Dunn test padj-value = 0.045; Figure 1B) and significantly higher than the low compatibility infections (28.10% vs 18.03%, Dunn test padj-value = 0.045; Figure 1B).

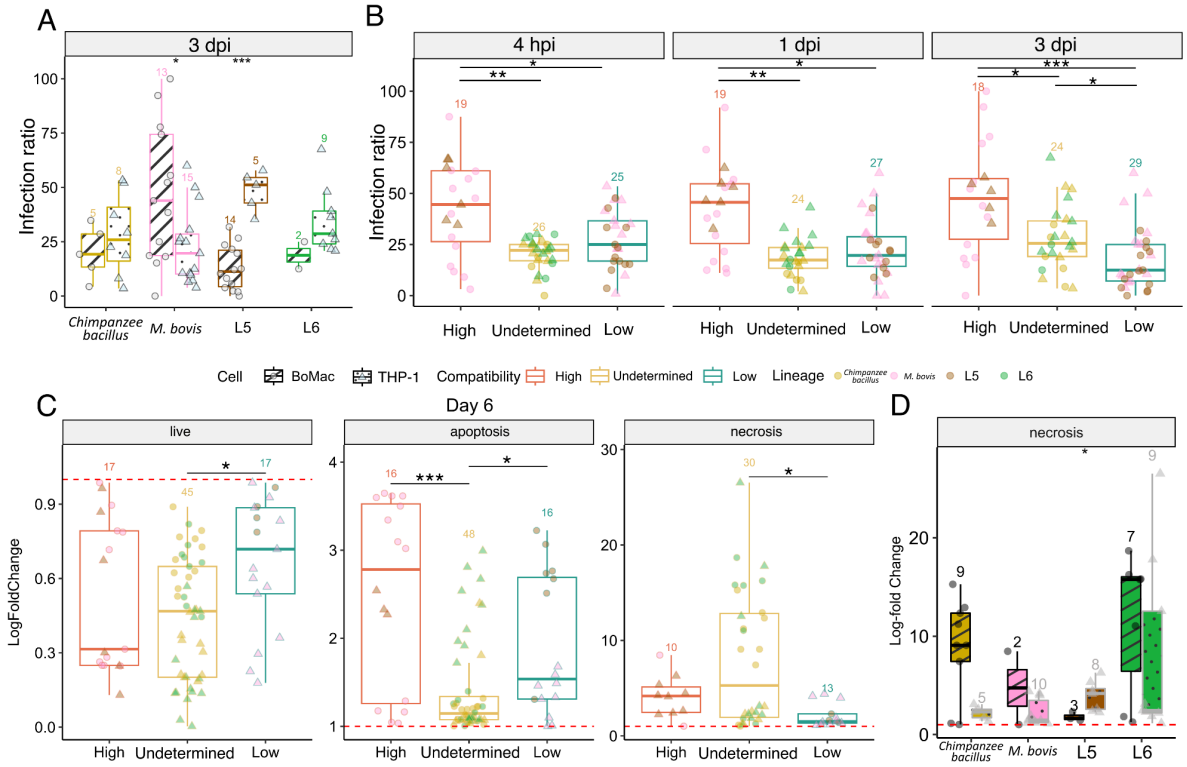



**Figure 1. Infectivity variability assessed by microscopy infection ratio in cross-infections of THP-1 and BoMac cells. A)** Differences in infection ratios between cells infected with the same lineage at 3 dpi, evaluated by the Mann-Whitney U test. **B)** Infection ratios for each compatibility group over days post-infection (dpi). The number of replicates per condition is indicated above each group. **C)** Log-fold change profile means for the three cell types given for each compatibility of interest for day 6 post-infection. **D)** Log-fold change between samples and experimental controls for day 6 post-infection, for each compatibility group in the different cell deaths. Differences between groups were evaluated using the Dunn test with the Benjamini-Hochberg correction. Only significant results are shown for clarity (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$).

We explored the effects of infection on cell death by performing cytometry analysis to categorize outcomes into necrosis, apoptosis, and live cells. Variability within each experimental group was very high, and we could not detect clear cell death differences between high and low compatibility infections (Figure 1C). However, we showed that low compatibility infections display a smaller proportion of live and apoptotic cells, but a higher proportion of necrotic cells than the undetermined group (Figure 1C). Comparing the three profiles of cell deaths, we observed the most determining value of cell death was in intermediate and high compatibility for necrosis death (8.04 for intermediate, 4.12 for high groups: see Figure S1). Therefore we focussed on the proportion of necrotic cells. L5 showed a significantly higher proportion of necrosis in human THP-1 cells than in bovine BoMac cells (3.96 vs 1.79 log-fold change; Mann-Whitney U test p-value = 0.024; Figure 1D). That result contrasts with the reverse tendency of *M. bovis*, where proportion of

necrotic cells tends to be higher in bovine BoMac cells than in THP-1 cells (4.73 vs 2.22, log-fold change; not significant difference Figure 1D).

## Transcriptional signatures of low and high compatibility infections

Global transcription profiles differ between cell types and across strains (Figure 2). *M. bovis* strains differ from other strains in low compatibility infections in THP-1 (Figure 2 panel A). Conversely, we could not observe the same transcription profile in BoMac cell infections (Figure 2 panel B). To explore expression differences due to host compatibility, we performed a differential expression analysis of orthologous genes between high and low compatibilities (Figure 2 panel C). When comparing orthologous genes between both cell types, high compatibility infections were characterized by a common enrichment of transcription in genes related to cell proliferation, including cell cycle regulation and DNA replication activities within the cell nucleolus in both cells (135 differential expressed orthologs genes (DEOGs) upregulated; Figure 2 panel D, Figure 3 panels A and B). The molecular functions identified by gene ontologies (GOs) included damaged DNA binding, catalytic activity acting on DNA, and helicase activity (Table S1). Enriched cellular components included the condensed chromosome, midbody, nuclear envelope, centriole, and oxidoreductase complex (Table S1). KEGG database enrichment analysis revealed an enrichment of pathways related to the cell cycle and Fanconi anemia pathways (Table S1).

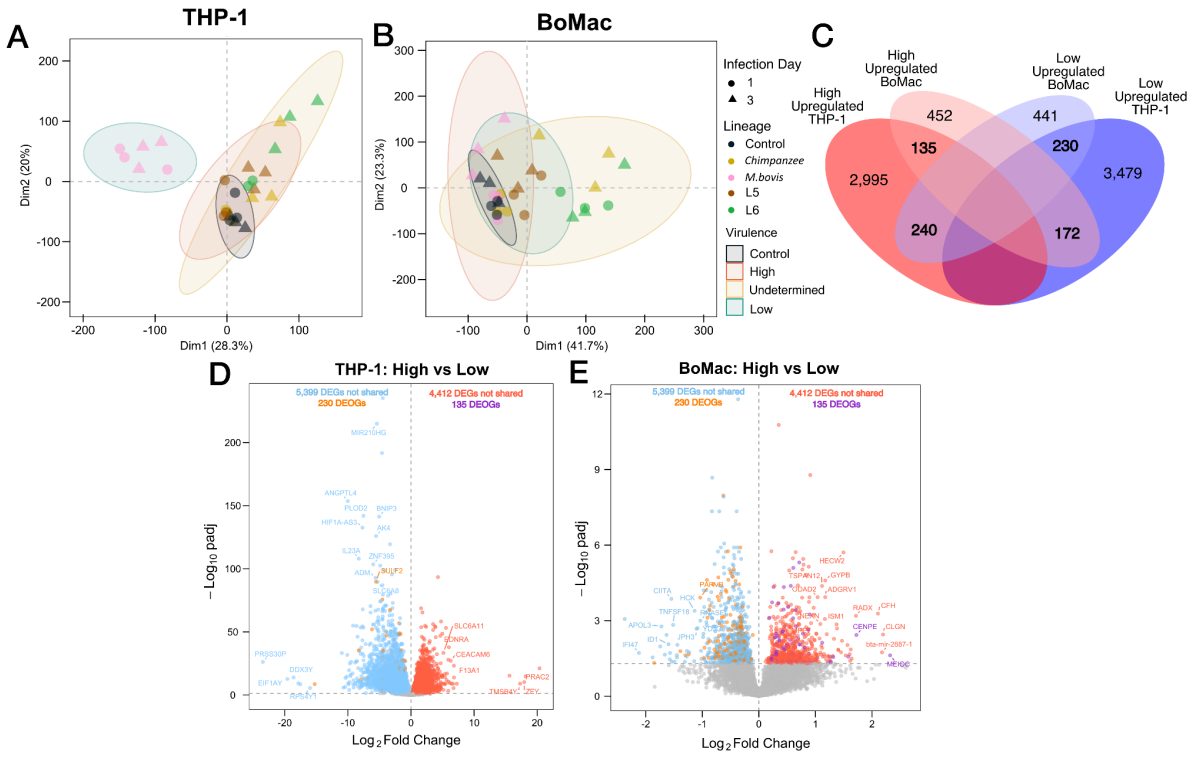



**Figure 2. Effect of gene expression of the host in the contrast of high vs low compatibilities. A)** Gene expression dimensionality for infections of all strains infected in THP-1 cells, each type in compatibility is clustered by an area. **B)** Gene expression dimensionality for infections of all strains infected in BoMac cells, each type in compatibility

is clustered by an area. **C)** Common and unique differentially expressed orthologs genes for THP-1 and BoMac cells for the contrast of high vs low compatibility**. D)** Volcano plot of differential expressed genes in the contrast in high vs low compatibilities infecting THP-1 cells. **E)** Volcano plot of differential expressed genes in the contrast in high vs low compatibilities infecting BoMac cells.

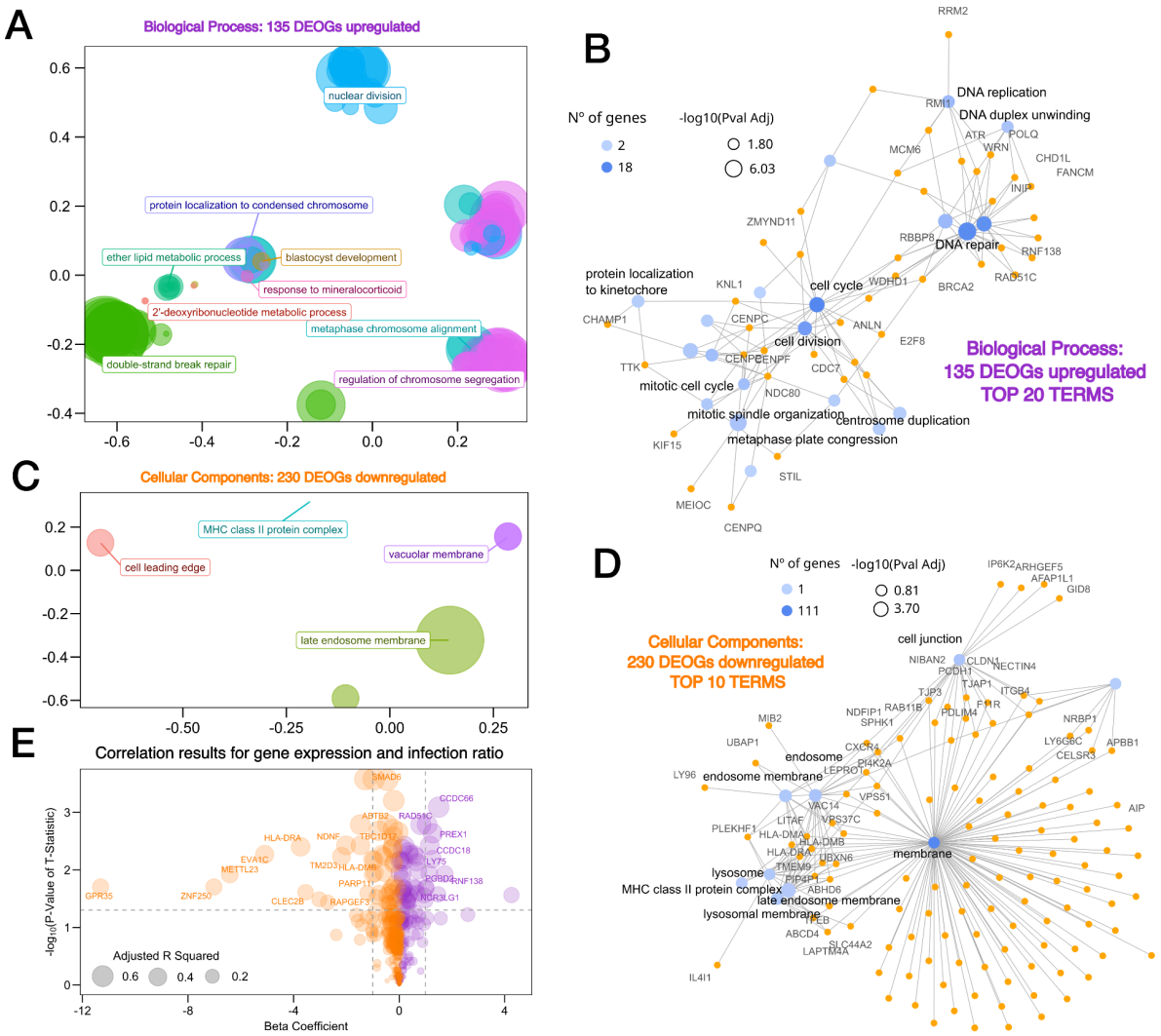



**Figure 3. Gene expression enrichment of the host in the contrast of high vs low compatibilities. A)** Gene enrichment analysis for DEOGs upregulated in high compatibility for Biological Process Ontology. **B)** Gene enrichment analysis for DEOGs downregulated in high compatibility for Cellular Components Ontology. Scatter plots in panels F and G in enriched terms show distances between points representing the similarity between terms, size of the point represents the scores of the GO term. **C)** Enriched terms networks with top 10 terms in Biological Process of DEOGs downregulated for the comparison of high vs low. **D)** Enriched terms networks with top 10 terms in Cellular Components of DEOGs downregulated for the comparison high vs low. **E)** Correlation results for gene expression with infection ratio assessed by microscopy.

But when we stratified high compatibility infections by cell type, we could detect certain genes exclusively found in one host but not the other. High compatibility THP-1 infection was characterized by increased transcription of 2,563 DEOGs that are enriched in KEGG pathways related to replication and repair: the cell cycle, DNA replication, cell growth and death, nucleotide, amino acid and carbohydrate metabolism (Table S2). Upregulated biological processes included mainly functions related with chromosome segregation, DNA replication, cell division, and responses to DNA damage. High compatibility BoMac infections also include upregulated functions related to cell cycle, DNA replication, helicase, and ATP activity (759 DEOGs), but unlike high THP-1 infections, in BoMac it also includes ferroptosis related pathways (Table S2).

Low compatibility infections were characterized by an enrichment of transcription in genes related to intracellular bacterial containment and immune evasion (230 DEOGs; Figure 2 panel E, Figure 3 panels D and E). The enrichment of cellular components included the late endosome membrane, vacuolar membrane, MHC class II protein complex, and cell leading edge, (Figure 3 panels D and E; Table S1). Again, if we stratify by cell line, low compatibility THP-1 infection is characterized by an upregulation of 3,730 DEGs. Those genes were enriched in processes of immune systems such as, signaling molecules, cellular interaction and transduction, regulation of T cell activation, cytokines and interleukins among others. Additionally, lipid metabolism and biological functions like oxygen regulation within the cell were also observed. However, high compatibility BoMac infections showed 911 genes related to antigen processing and presentation, and T helper differentiation (Table S2).

## Strain specific transcriptional signatures

We observed 240 DEOGs commonly upregulated when infecting with the L5 strain either THP-1 cells, or BoMac cells. L5 specific transcription was enriched in genes related to catalytic activity, acting on DNA and RNA, tubulin binding, helicase activity, ATP hydrolysis activity, GTPase and ribosome binding, and single-stranded DNA binding. However, *M. bovis* specific transcription (172 DEOGs), was enriched for the KEGGs pathways related to ferroptosis pathway, and molecular functions related to ubiquitin-like protein ligase activity and cadherin binding.

## Temporal transcriptional signatures depend on compatibility

We also observed transcriptional differences at different stages of infection between high and low compatibility. At early stages, high compatibility infections are characterized by increased transcription in KEGG pathways that include signal transduction, infectious disease, focal adhesion and metabolism of terpenoids and polyketides (528 DEOGs,Table S3). Enriched biological processes involved ribosome biogenesis, RNA processing, and cellular regulations such as negative regulation of apoptotic signaling pathways. Low compatibility infections are characterized by enriched KEGG pathways in pentose phosphate, thermogenesis and oxidative phosphorylation among others (365 DEOGs, Table S3). In the biological process we observe ribose phosphate biosynthetic process, energy

derivation by oxidation of organic compounds, regulation of cellular component size, response to oxygen levels and oxidative stress, regulation of leukocyte apoptotic process and vesicle-mediated transport between endosomal compartments.
At later stages of infection (3 dpi), high compatibility was characterized by enriched transcription of KEGG pathways related to lysosome, phosphatidylinositol signaling, inositol phosphate metabolism, DNA replication, and glycan degradation (641 DEOGs, Table S3). In the biological processes we obtained enriched functions related to liposaccharide metabolic processes, chromosome segregation and cellular reorganization such as nuclear division, autophagosome organization. Later stages in low compatibility infections were characterized by an enriched transcription in molecular functions enriched for protein lysine deacetylase activity (266 DEOGs, Table S3).

## Molecular signatures of infection ratio by host compatibility

To decipher which genes could explain the lower *in vitro* infection ratio of the high compatibility infections, we correlated the expression levels of selected genes with infection ratios. We found 10 genes significantly correlated with lower infection ratio, which include 3 genes related to MHC II antigen presentation: CLEC2B, HLA-DRA and HLA-DMB (Figure 3 panel D). Conversely, we found 48 genes significantly correlated to high infection ratio, that include genes involved in double-strand break repair via homologous recombination and meiotic nuclear division (Figure 3 panel J). The most extreme values correspond to 10 genes that include NCR3LG1 that is a Natural Killer Cell Cytotoxicity Receptor 3 Ligand, and the gene RNF138 that is related to MHC I and Innate Immune System.

## Differential exon usage by host compatibility

In both, high and low compatibilities, we have identified different uses of exons leading to different isoforms in THP-1. We observed 880 genes that harbored some exons more expressed in high compatible infection and other exons in low compatible infections (Figure 4 panels A-D). These genes include enriched functions of cell growth, organization, and mobility, but also GTPase activity. In the KEGG database we observed enriched pathways the ECM-receptor interaction, focal adhesion and Cytoskeleton in muscle cells (Table S4). IL-16 which shows 28 different exons usages, 1 of them (E025) related to high compatibilities with lower log2 fold change between the two conditions, and 14 to low compatibilities with higher differences between groups. Another gene that we observed is SIGIRR related to the negative regulator of the Toll-like and IL-1R receptor signaling pathways. We observed 17 exons in the gene, 11 of them related to the high compatibility infections and 3 related to low compatibility infections. Gene PRKCD encodes a protein which can positively or negatively regulate apoptosis, and we found 23 exons where 3 exons were associated with high compatibility infections and another 3 with low compatibility infections (Figure 4 panels I-K).

Additionally, we detected 2,267 genes with all their exons more expressed in high compatible infections (Figure 4 panels E-H). They include biological process functions

related to regulation of autophagy and phospholipid metabolic process, protein localization to the nucleus, G1/S transition of mitotic cell cycle, GTPase mediated signal transduction and glucose 6-phosphate metabolic process. For the molecular functions we observed actin and DNA binding, and in the cellular components we observed functions related to the organization of the cell (Table S4).

Finally, we found 1,634 genes where all their exons were more expressed in low compatibility infection (Table sup). In this category we observed biological processes such as regulation of endocytosis, and peptidyl-serine phosphorylation and modification, and molecular functions like GTPase regulator activity, nucleoside-triphosphatase regulator activity and ATP hydrolysis activity. Cellular components were related to cleavage furrow, histone acetyltransferase complex, mitotic complex, and nucleotide-excision repair complex. In the KEGG database we observed pathways related to DNA replication, Fc gamma R-mediated phagocytosis, biosynthesis of cofactors and VEGF signaling pathway among others.

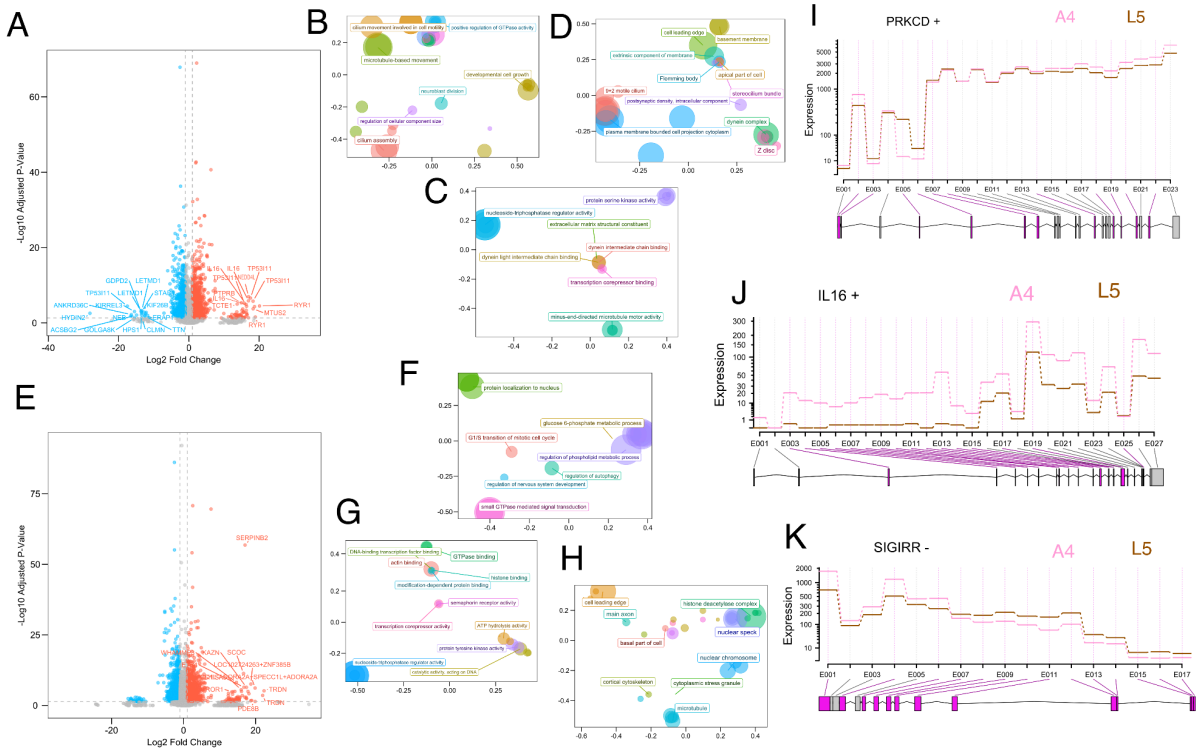



**Figure 4. Differential exon usage for the high and low compatibility infections. A)** Contrast to the differential exon usage for genes containing events related to high (blue dots) and low (dots colored in red) combinations, each exon usage is represented as a dot. **B)** Gene enriched terms in Biological Process**. C)** Gene enriched terms in Molecular Functions. **D)** Gene enriched terms in Cellular Components. Panels B, C and D involved genes with different exon usages related both compatibilities**. E)** Contrast to the differential exon usage for genes containing events only related to high (blue dots) or low (dots colored in red) combinations, each exon usage is represented as a dot. **F)** Gene enriched terms in Biological Process. **G)** Gene enriched terms in Molecular Functions. **H)** Panels F, G and H involved genes with different exon usages related to high compatibility. Scatter plots in

panels B, C, D, F, G and H displaying enriched terms show distances between points representing the similarity between terms, size of the point represents the scores of the GO term. **I)** Fitted expression values of each of the exons of the gene PRKCD. **J)** Fitted expression values of each of the exons of the gene IL16. **K)** Fitted expression values of each of the exons of the gene SIGIRR. For panels I, J and K, exons with a differential usage between high (L5) and low (A4) are highlighted with purple.

# Discussion

Pathogens have evolved to interact with specific molecules in the host's cells, like cell receptors, nutrients, toxics, or specific immune molecules that determine the molecular determinants of host-pathogens compatibility. In the case of *M. tuberculosis*, one central aspect is why specific ecotypes or lineages affect particular hosts, and if there are molecular determinants of such specificity. There is a wealth of data providing host transcriptional responses to different *M. tuberculosis* strains during *in vitro* infection[31,32]. However, our approach focuses on comparing specific patterns of different predicted compatibility between strain and host, allowing to decipher common molecular signals of different host cells to the most compatible infection and specific patterns of certain strains independently of the host.

We confirmed our hypothesis that bacteria infecting the most compatible host cell display higher infection ratio, and also higher necrosis, both associated with more infectivity in an *in vitro* system. But, bacteria like the Chimpanzee bacillus, that was isolated only once, and their host preference is not known showed intermediate patterns of infection ratio between the other two strains, and similar infection ratio between both hosts. Similar patterns were detected in the L6 strain, despite belonging to a human associated lineage. L6 is predominantly isolated from humans, but their association with immunocompromised individuals, their phylogenetic position among animal-associated lineages and its larger genomic diversity led to the hypothesis that the L6 strain will show a different pattern compared to L5 a human-associated strain[15]. Interestingly, L6 is the closest lineage to lineage A1, which includes the Chimpanzee bacillus[30]. Our results confirm that the L6 strain and the Chimpanzee bacillus behave more similarly in *in vitro* infections than with the other strain from a human associated lineage (L5).

In order to decipher the molecular determinants of *in vitro* compatibility of different strains in human and bovine cells, we compared the expression profiles of the different compatibility profiles. Comparative genomic studies between *Bos taurus* and *Homo sapiens* face significant challenges due differences in genome reference quality and completeness. *B. taurus* genomes, despite improvements, still have inconsistencies and gaps[33], and these disparities between references complicates RNA-seq comparative analysis between species because the comparison of DEGs, and even more splicing events are challenging[34]. Despite this, we were able to focus on retrieving common transcriptional patterns of high

and low compatibility *in vitro* infections. We found that bacteria infecting the most compatible host cell can evade the immune system's weaponry because it inhibits genes related to antigen presentation (MHC II), cellular compartmentalization, and endosome creation, which are under-expressed. These pathways might play a crucial role in *M. tuberculosis* infections, as the bacteria's ability to inhibit antigen presentation and manipulate host cellular processes allows it to evade immune detection and establish chronic infections[35–38]. On the contrary, the most compatible infections result in upregulation of the host's cell replication and DNA, which we can explain as a strategy by the bacteria to make the cell replicate and thus spread the infection. This strategy was shown for *Chlamydia trachomatis* in an *in vitro* infection system that causes the proliferation of host cells to maintain a persistent infection[39]. Conversely, strains infecting their least compatible host cell display the reverse pattern. The host overexpressed compartmentalization, and specific genes such as UBXN6, ABCD4 or TFEB among others could lead to the activation and elimination of infected cells. The importance of these pathways in *M. tuberculosi*s infections has been highlighted in studies showing that effective compartmentalization in phagosomes and endosomes, and immune activation that can restrict bacterial growth and promote clearance of the infection[40]. In addition to common signatures, we could detect strain-specific patterns such as the upregulation of ferroptosis pathways in *M. bovis* infections compared to the infection with the L5 strain. Ferroptosis, characterized by iron-dependent lipid peroxidation, has been shown to be a critical mechanism of necrosis during *M. tuberculosis* infection, both *in vitro* and *in vivo*[41].

Additionally, we found important differences in the temporal response of the most and least compatible infections. Infection with the most compatible strains, results in a no specific early response more directed to cellular reorganization by increasing cellular metabolism and ribosome generation, impairing apoptosis signaling and promoting cellular mobilization in terms of cell adhesion. At later stages, these strains seemed to manipulate lysosome pathways and organize autophagosomes for survival[42], which are pathways that are important in the control of *M. tuberculosis* infection[43]. Additionally, these strains specialize in resource acquisition by degrading glycans and activating polysaccharide metabolism. This is consistent with a previous study that demonstrated *M. tuberculosis* accesses glucose *in vivo* and relies on glucose phosphorylation for survival during chronic mouse infections[44]. In contrast, infections with the least compatible strains results in more specific anti-pathogen responses early on, such as modification of oxygen levels and utilization of oxidative stress, which have been described to contribute to the antimycobacterial function of peroxisomes in the cytosol of human macrophages[45], as well as activating vesicle transport in endosomal compartments to control the bacteria[46,47].

But other transcriptional signatures, such as alternative splicing can also also play an important role in virulence specific responses, as previously found during *in vitro* infections with virulent and avirulent laboratory *M. tuberculosis* strains[48]. In our study, we observed hundreds of genes with different isoforms associated with low or high compatibility infections, and although we cannot precisely determine how these isoforms affect infection, we can identify patterns associated with compatibility that could be investigated in the

future. We detected different isoforms of IL-16 in high vs low compatibility infections, where several exon usage were related to low compatibility infections. IL-16 functions as a chemoattractant modulating T cell activation[49], and the different isoforms observed could potentially affect gene functionalities and impair the immune response. Other genes related to immune components such as SIGIRR, a negative regulator of Toll-like receptor IL-1 receptor signaling, also display different isoforms in high and low compatibility infections. Additionally, we highlight the PRKCD gene, related to apoptosis, which also presents different isoforms in high and low compatibility infections.

Our results are limited by our experimental set up. We use human-associated L5 and L6 strains and animal-associated A4 and A1 strains, which could not be representative of broader patterns across the whole *M. tuberculosis* lineage. Variability can be expected among different strains within the same lineage and among lineages[50–52]. Therefore, additional studies with more strains of the same lineage are needed to generalize our findings. Another limitation of our study is the use of THP-1 and BoMac cell lines, which may exhibit altered gene expression profiles, signaling pathways, and functional characteristics compared to primary cells[53]. Therefore, the investigation of the cellular and molecular responses of primary immune cells could provide a more realistic picture of the complexity of the immune system, including differences between host populations.

Overall, our findings have confirmed our hypothesis where more compatible strain-host combinations results in higher infection ratio and non efficient molecular response of the pathogen from early on the infection. This contributes to a deeper understanding of the molecular response of human and bovine cell lines challenged with different *M. tuberculosis* strains in the context of the complex interplay between strain-host compatibility.

# Material and Methods

## Cell lines and strains

Human THP-1 and bovine macrophages (BoMac) were used for infections. Cell cultures were maintained in respective mediums before seeding and differentiation for each infection. THP-1 monocytes were cultured at 37°C with 5% $CO_2$ to confluence in RPMI media (MEDIUM 1640 CE, Fisher Scientific 11530586), then differentiated into macrophages using 15 ng/ml of phorbol 12-myristate 13-acetate (PMA) for 24h. BoMac were thawed and maintained at 37°C with 5% $CO_2$ to confluence in IMDM media (containing GlutaMax and HEPES) (Gibco, 31980-022) supplemented with 10% FBS, 1x MEM Non-Essential Amino Acids (Gibco, 11140-035), 1x MEM Vitamins (Biochrom AG, K0373) and 50 µM β-Mercaptoethanol (Gibco, 31350-010). Macrophages were detached using trypsin-EDTA 25% (Gibco, R001100) and seeded in both 24- and 6-well cell culture-treated plates according to the intended procedure. All cells were rested for 24 hours before infection at 37°C with 5% $CO_2$ to allow macrophages to adhere on the plates.

Human-associated *M. tuberculosis* strains N1176 (included in L5) and N1202 included in L6)[28] and animal-associated *M. tuberculosis* strains G1352 (also known as *Mycobacterium bovis*)[54], and N1894 (also referred as *Chimpanzee bacillus*)[30] (see Figure S2 panel A) were used.

## *M. tuberculosis* infections

All strains were grown at 37°C to mid-log phase ($OD_{600}$ = 0.5 to 0.6) in Middlebrook 7H9 broth, enhanced with 10% (v/v) albumin/dextrose/catalase (ADC Middlebrook Enrichment – Fisher Scientific, 11756924), 0.44% (w/v) sodium pyruvate, and 0.05% (v/v) Tween 20. Single cells from each strain were prepared by resuspending and washing the cultures in phosphate-buffered saline (PBS), resuspending in either supplemented RPMI or IMDM then followed by sonication at maximum speed for 2 minutes.

Rested $7{\cdot}10^5$ THP-1 and $5{\cdot}10^5$ BoMac macrophages in 6-well plates, and $5{\cdot}10^5$ THP-1 and $1.8{\cdot}10^5$ BoMac in 24-well plates were then infected with all strain-cell combinations at different multiplicity of infection (MOI) or left uninfected. In THP-1 macrophages MOI of 10:1 while in BoMac MOI of 20:1 was used respectively. After 4 hours post-infection (hpi), extracellular bacteria were removed and the media was replaced with RPMI or IMDM supplemented with 5 µg/µL gentamicin. After 1-day post-infection (dpi), the gentamicin-supplemented media was substituted with fresh RPMI or IMDM. Infections were done in replicates whenever possible for each strain, and were maintained for 7 days.

## Infection ratio by microscopy

Infections for microscopy were performed from infected cells in 24 well plates, starting with $5{\cdot}10^5$ THP-1 and $1.8{\cdot}10^5$ BoMac cells per well. The infection ratio was determined by optical microscopy at 4 hpi, 1-and 3 dpi (see Figure S2 panel C). Initially, THP-1 cells or BoMac cells were fixed with 3.7% formaldehyde. Subsequently, the infected cells were stained using the Ziehl-Neelsen technique. Cells were first stained with carbol fuchsin (VWR, 1085120500), resulting in a pink stain on the infected cells. Followed by incubation in hydrochloric acid in ethanol (Merck, 100327), leading to the decolorization of macrophages but not acid-fast cells, which retained their pink coloration. Finally, the decolorized macrophages were counter-stained with methylene blue (VWR, 1012870500). We considered an infected cell as a macrophage containing at least one bacillus. To determine the infection ratio, both infected and uninfected cells were counted within a 44 $mm^2$ area. The infection ratio was calculated by dividing the number of infected cells by the total number of cells in the observed area. We obtained data across 12 experiments with at least two technical replicates for each infection combination.

## Cell death assessment by Flow cytometry

To investigate macrophage cell death induced by different strains, we performed flow cytometry analysis at 2 and 6 dpi from infected 6-well plates starting with $7{\cdot}10^5$ THP-1 and $5{\cdot}10^5$ BoMac cells per well (see Figure S2 panel C). The culture medium was removed from the wells, and the cells were detached with trypsin, and washed with PBS. The cells were then resuspended in 500 µL of PBS then stained following the Fixable Viability Stain 450 kit manufacturer instructions (BD Biosciences, 562247). Infected and uninfected macrophages were stained with Fixable Viability Stain 450, diluted 1:10 in PBS, and incubated at 37°C for 5 minutes in the dark then washed with PBS. Subsequently, the samples were stained with 80 µL CellEvent Caspase 3/7 Green detection Reagent 2 µM in PBS with 5%FBS (Thermo Fisher Scientific, C10423) and then incubated for 30 minutes at 37°C. Stained cells were fixed with 80 µL of 7.4% formaldehyde for 10 minutes at room temperature. Finally, the supernatant was removed by centrifugation, and the cells were resuspended in 500 µL of PBS and then acquired on a BD LSRFortessa the same day. FCS files obtained were exported and then analyzed using FlowJo™ Software v 10.9. After obtaining cell viability percentages for each sample, we excluded controls identified as outliers using the interquartile range (IQR) method. We then computed the mean values of the controls for each experiment. Following this, the log-fold changes for the samples were calculated by dividing the sample values by the control means. We obtained data across 12 experimental setups with at least 3 technical replicates for each infection combination. In classifying cell death, we removed values below 1 for apoptosis and necrosis, and for live cells, we excluded values above 1. Additionally, outliers in the log-fold change data were also removed. After filtering those low-quality data we analyzed data from 12 experiments with at least 2 technical replicates for each infection combination.

## RNA extraction and sequencing

RNA was extracted at 1- and 3 dpi (see Figure S2 panel C) from infected cells in 6-well plates, starting with $7{\cdot}10^5$ THP-1 and $5{\cdot}10^5$ BoMac cells per well. We used 500 µL RNAzol per $5{\cdot}10^5$ macrophages, followed by centrifugation 20 min at 10000 rpm. The supernatant was removed for eukaryotic RNA sequencing while the pellet was subjected to *M. tuberculosis* RNA extraction. Pellets from 6 wells were pulled and resuspended in 300 µL RNAzol and homogenized by vortexing for 3 minutes, followed by *M. tuberculosis* cell lysis using Lysing Matrix B containing 0.1 mm silica beads (MP biomedicals). Post-centrifugation, the supernatants containing *M. tuberculosis* RNA were transferred into a new tube. Thereafter, bacterial and eukaryotic RNAs were subjected to the RNA purification process. Isopropanol was added to induce RNA precipitation, followed by ethanol-washing steps to purify the RNA and remove impurities. Finally, the purified RNA from each sample was resuspended in 1 mM Tris-EDTA solution. We analyzed data from 8 combinations with at least 3 technical replicates for each infection combination.

After successfully extracting and purifying RNA from both the host and the bacterial samples involved in the infection, 48 samples underwent ribodepletion using the RiboZero

Gold kit (Illumina, San Diego, USA). The library preparation was elaborated using the NEBNext® SingleCell/Low Input cDNA Synthesis kit (New England Biolabs). The prepared libraries were then sequenced, utilizing an S4 150 bp configuration on the Illumina NovaSeq 6000 platform (for sequenced samples see Table S5). The obtained reads are available under the accession project ID (pending release).

## RNA-seq analysis

The reads obtained from host-infection sequencing were evaluated by quality with BBDuk function from BBMap v 39.06[55], setting a minimum length at 35 bp and a minimum quality of 35. This step ensured the preservation of the longest, high-quality reads while discarding low-quality segments. The read quality was also evaluated before and after trimming using FastQC v 0.12.1[56].

We aligned the filtered reads to their respective reference genomes using STAR v 2.7.11b[57]. Reads from infected THP-1 macrophages were mapped to *Homo sapiens* (v 104, accession number GCA_000001405.15), while those from infected BoMac macrophages were aligned to *Bos taurus* genome (v 109.12, accession number GCA_002263795.2).

The following analysis were conducted using R v 4.3.2, with the code available in PathoGenOmics-Lab/CrossInfectMTB. Gene expression quantification was performed using featureCounts function from R package Rsubread v 2.16.1[58]. This step involved mapping the high-quality reads to genomic features from the Ensembl annotation (109 releases for *B. taurus* and 104 for *H. sapiens*). We filtered the raw gene expression counts using the HTSfilter R package v 1.42.0[59] to eliminate genes with non-informative signals. Normalization of the raw counts was then carried out using the DESeq2 R package v 1.42.0[60], based on the mean of ratios method. Once we obtained the sequencing we evaluated the expression of the genes, and one sample that clustered with controls was removed.

Differentially expressed genes (DEG) between the conditions of interest (comparisons detailed in Figure S2 panel D) were obtained with DESeq2 analysis. To identify common DEGs between *H. sapiens* and *B. taurus*, we employed the biomaRt R package v 2.58.0[61], and we kept one-to-one orthologs. Each contrast's significantly differentially expressed genes were determined using thresholds of a p-value $< 0.05$ and a log-fold change $> 1$ or $< 1$. Gene Ontology Enrichment Analysis (GOEA) and KEGG Pathways Enrichment Analysis (KEA) were performed to identify the most enriched Gene Ontologies (GOs) and KEGG terms. The enrichment analysis of DEGs were analyzed with the clusterProfiler R package v 4.10.0[62] and the AnnotationDbi R package v 1.64.1, for gene enrichment we employed the org.Hs.eg.db and org.Bt.eg.db v 3.19.1 R packages. The results from GOEA were streamlined and visualized using the rrvgo R package v 1.14.1[63]. Additionally, all analyses across our study were visualized comprehensively using the ggplot2 R package v 3.5.0[64].

Correlation of differentially expressed genes and infection ratios were determined by a custom R script applying a linear model with R core basic functions, results were visualized with ggplot R package v 3.5.0[64].

### Differential exon usage analysis

Once the reads were aligned we performed a quantification of the reads using the python library HTSeq[65] v 2.0.5. Then, the differential exon usage (DEU) was assessed with the DexSeq v 1.50.0[66] R package. The steps involving alternative splicing analysis were executed with the pipeline of nf-core/rnasplice v 1.0.3. Once the differential exon usage was assessed, the contrasts between the comparisons of interest were analyzed with a custom R package and the DEUs were visualized with ggplot2 v 3.5.0[64] and DexSeq v 1.50.0[66] packages. The genes containing the DEUs were analyzed with a gene enrichment analysis for GOEA and KEA by the clusterProfiler R package v 4.10.0[62] and the AnnotationDbi R package v 1.64.1 functions.

## Acknowledgments

We extend our gratitude to: Prof. Sébastien Gagneux and Dr. Iñaki Comas for generously providing the mycobacterial strains, Dr. Raquel Conde and Dr. Esther Julián for their provision of THP-1 cells, and Dr. Matthias Schweizer for supplying the BoMac cells.
The computational analyses were performed on the HPC cluster Garnatxa at Institute for Integrative Systems Biology ($I^2$SysBio), $I^2$SysBio is a mixed research center formed by University of Valencia (UV) and Spanish National Research Council (CSIC).

## Data availability

The code for the bioinformatic methods and analyses described are available at https://github.com/PathoGenOmics-Lab/CrossInfectMTB (pending release).

## Funding

This work was financially supported by project PID2021-123443OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by FEDER A way to make Europe. PRR is supported by CNS2022-135116 funded by MCIN/AEI /10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. MCG is supported by a predoctoral contract from the Generalitat Valenciana (CIACIF/2021/159) and Project CIPROM2021-053 from Generalitat Valenciana.

## Author Contributions

**PRR:** data curation, formal analysis, investigation, methodology, software, validation, visualization, writing - original draft. **MCG:** investigation, validation, methodology, visualization, writing - original draft. **GS:** data curation, formal analysis, investigation,

methodology. **HH:** investigation, writing – review & editing. **MC:** conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing - original draft.

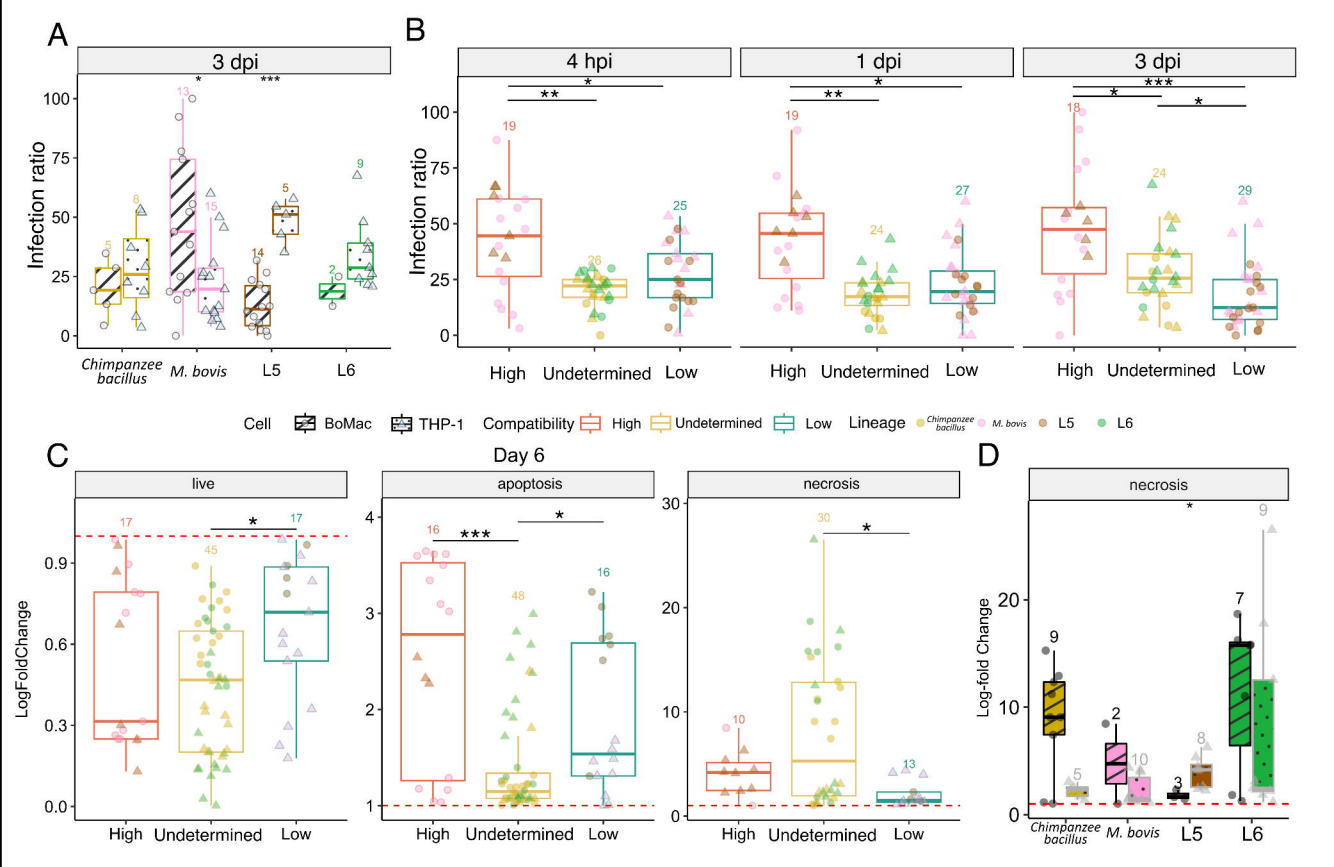
# References

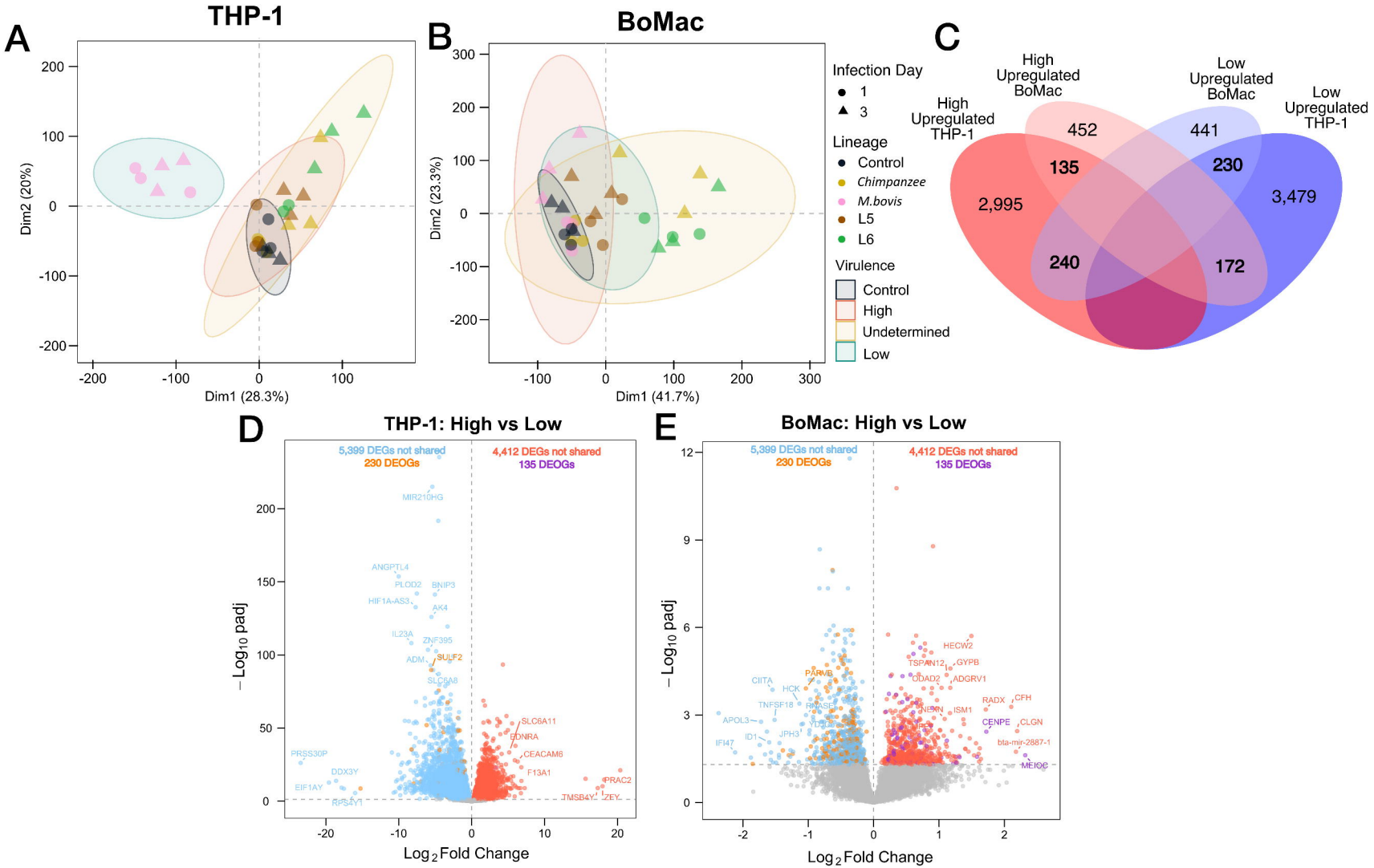

1. Speranza, E. Understanding virus-host interactions in tissues. *Nat. Microbiol.* **8**, 1397–1407 (2023).
2. Finlay, B. B. & McFadden, G. Anti-Immunology: Evasion of the Host Immune System by Bacterial and Viral Pathogens. *Cell* **124**, 767–782 (2006).
3. de Jong, H. K., Parry, C. M., van der Poll, T. & Wiersinga, W. J. Host–Pathogen Interaction in Invasive Salmonellosis. *PLoS Pathog.* **8**, e1002933 (2012).
4. Rodriguez-Frandsen, A., Alfonso, R. & Nieto, A. Influenza virus polymerase: Functions on host range, inhibition of cellular response to infection and pathogenicity. *Virus Res.* **209**, 23–38 (2015).
5. Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).
6. Gabriel, G. & Fodor, E. Molecular Determinants of Pathogenicity in the Polymerase Complex. in *Influenza Pathogenesis and Control - Volume I* (eds. Compans, R. W. & Oldstone, M. B. A.) 35–60 (Springer International Publishing, Cham, 2014). doi:10.1007/82_2014_386.
7. Muylaert, R. L. *et al.* Using drivers and transmission pathways to identify SARS-like coronavirus spillover risk hotspots. *Nat. Commun.* **14**, 6854 (2023).
8. Niu, Z. *et al.* N501Y mutation imparts cross-species transmission of SARS-CoV-2 to mice by enhancing receptor binding. *Signal Transduct. Target. Ther.* **6**, 1–3 (2021).
9. Tan, C. C. S. *et al.* Transmission of SARS-CoV-2 from humans to animals and potential host adaptation. *Nat. Commun.* **13**, 2988 (2022).
10. Wattam, A. R. *et al.* Comparative phylogenomics and evolution of the Brucellae reveal a path to virulence. *J. Bacteriol.* **196**, 920–930 (2014).
11. Gagneux, S. Ecology and evolution of Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **16**, 202–213 (2018).
12. WHO. Global Tuberculosis Report 2023. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023.
13. Brites, D. *et al.* A New Phylogenetic Framework for the Animal-Adapted Mycobacterium tuberculosis Complex. *Front. Microbiol.* **9**, (2018).
14. Guyeux, C. *et al.* Newly Identified Mycobacterium africanum Lineage 10, Central Africa. *Emerg. Infect. Dis.* **30**, 560–563 (2024).
15. Coscolla, M. *et al.* Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. *Microb. Genomics* **7**, 000477 (2021).
16. Ngabonziza, J. C. S. *et al.* A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nat. Commun.* **11**, 2917 (2020).
17. Brites, D. & Gagneux, S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunol. Rev.* **264**, 6–24 (2015).
18. Coscolla, M. & Gagneux, S. Consequences of genomic diversity in Mycobacterium tuberculosis. *Semin. Immunol.* **26**, 431–444 (2014).
19. Palmer, M. V., Thacker, T. C., Waters, W. R., Gortázar, C. & Corner, L. A. L. Mycobacterium bovis: A Model Pathogen at the Interface of Livestock, Wildlife, and Humans. *Vet. Med. Int.* **2012**, 236205 (2012).

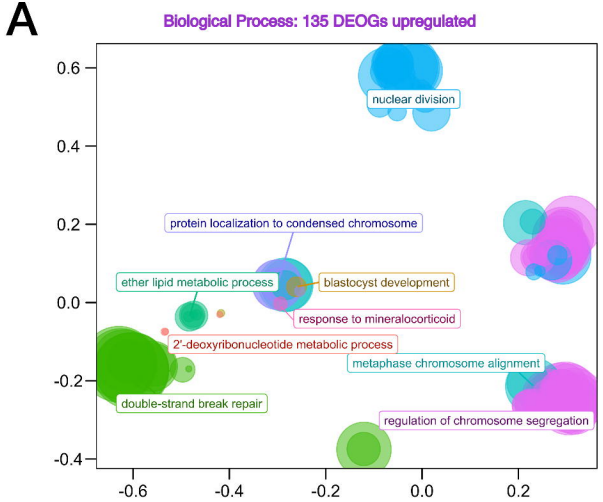
20. Berg, S. & Smith, N. H. Why doesn't bovine tuberculosis transmit between humans? *Trends Microbiol.* **22**, 552–553 (2014).
21. Evans, J. T. *et al.* Cluster of human tuberculosis caused by Mycobacterium bovis: evidence for person-to-person transmission in the UK. *The Lancet* **369**, 1270–1276 (2007).
22. Ghodbane, R. & Drancourt, M. Non-human sources of Mycobacterium tuberculosis. *Tuberc. Edinb. Scotl.* **93**, 589–595 (2013).
23. Ghielmetti, G. *et al.* Tuberculosis in Swiss captive Asian elephants: microevolution of Mycobacterium tuberculosis characterized by multilocus variable-number tandem-repeat analysis and whole-genome sequencing. *Sci. Rep.* **7**, 14647 (2017).
24. Whelan, A. O. *et al.* Revisiting Host Preference in the Mycobacterium tuberculosis Complex: Experimental Infection Shows M. tuberculosis H37Rv to Be Avirulent in Cattle. *PLOS ONE* **5**, e8527 (2010).
25. Villarreal-Ramos, B. *et al.* Experimental infection of cattle with Mycobacterium tuberculosis isolates shows the attenuation of the human tubercle bacillus for cattle. *Sci. Rep.* **8**, 894 (2018).
26. Osei-Wusu, S. *et al.* Macrophage susceptibility to infection by Ghanaian Mycobacterium tuberculosis complex lineages 4 and 5 varies with self-reported ethnicity. *Front. Cell. Infect. Microbiol.* **13**, (2023).
27. Yeboah-Manu, D. *et al.* Spatio-Temporal Distribution of Mycobacterium tuberculosis Complex Strains in Ghana. *PloS One* **11**, e0161892 (2016).
28. Diarra, B. *et al.* Mycobacterium africanum (Lineage 6) shows slower sputum smear conversion on tuberculosis treatment than Mycobacterium tuberculosis (Lineage 4) in Bamako, Mali. *PloS One* **13**, e0208603 (2018).
29. Otchere, I. D. *et al.* Comparative genomics of Mycobacterium africanum Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. *Sci. Rep.* **8**, 11269 (2018).
30. Coscolla, M. *et al.* Novel Mycobacterium tuberculosis Complex Isolate from a Wild Chimpanzee. *Emerg. Infect. Dis.* **19**, 969–976 (2013).
31. Koo, M.-S., Subbian, S. & Kaplan, G. Strain specific transcriptional response in Mycobacterium tuberculosis infected macrophages. *Cell Commun. Signal. CCS* **10**, 2 (2012).
32. Leisching, G. *et al.* RNAseq reveals hypervirulence-specific host responses to M. tuberculosis infection. *Virulence* **8**, 848–858 (2017).
33. Zhou, S. *et al.* A clone-free, single molecule map of the domestic cow (Bos taurus) genome. *BMC Genomics* **16**, 644 (2015).
34. Florea, L., Souvorov, A., Kalbfleisch, T. S. & Salzberg, S. L. Genome assembly has a major impact on gene content: a comparison of annotation in two Bos taurus assemblies. *PloS One* **6**, e21400 (2011).
35. Ramachandra, L. *et al.* Phagosomal Processing of Mycobacterium tuberculosis Antigen 85B Is Modulated Independently of Mycobacterial Viability and Phagosome Maturation. *Infect. Immun.* **73**, 1097–1105 (2005).
36. Harding, C. V. & Boom, W. H. Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors. *Nat. Rev. Microbiol.* **8**, 296–307 (2010).
37. Roy, S. *et al.* Transcriptional landscape of Mycobacterium tuberculosis infection in macrophages. *Sci. Rep.* **8**, 6758 (2018).
38. Pu, W. *et al.* Comparative transcriptomic analysis of THP-1-derived macrophages infected with Mycobacterium tuberculosis H37Rv, H37Ra and BCG. *J. Cell. Mol. Med.* **25**, 10504–10520 (2021).
39. Kessler, M. *et al.* Chlamydia trachomatis disturbs epithelial tissue homeostasis in fallopian tubes via paracrine Wnt signaling. *Am. J. Pathol.* **180**, 186–198 (2012).

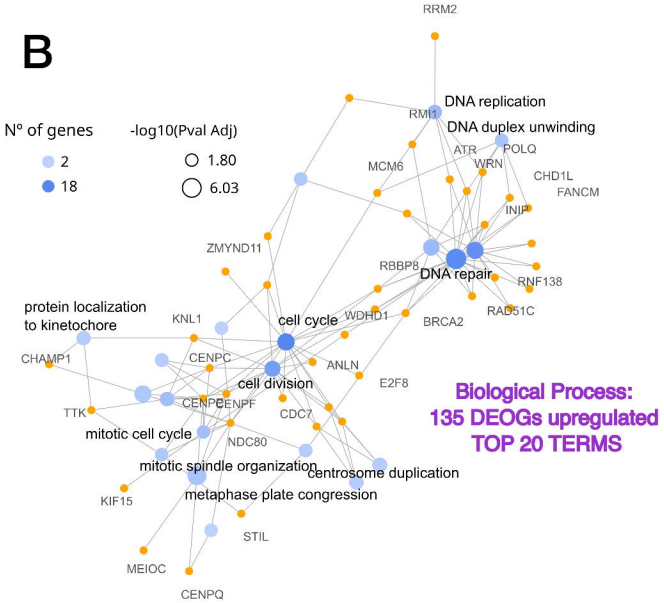
40. Pieters, J. Mycobacterium tuberculosis and the macrophage: maintaining a balance. *Cell Host Microbe* **3**, 399–407 (2008).
41. Amaral, E. P. *et al.* A major role for ferroptosis in Mycobacterium tuberculosis-induced cell death and tissue necrosis. *J. Exp. Med.* **216**, 556–570 (2019).
42. Gutierrez, M. G. *et al.* Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* **119**, 753–766 (2004).
43. DesJardin, L. E. *et al.* Mycobacterium tuberculosis-infected human macrophages exhibit enhanced cellular adhesion with increased expression of LFA-1 and ICAM-1 and reduced expression and/or function of complement receptors, FcgammaRII and the mannose receptor. *Microbiol. Read. Engl.* **148**, 3161–3171 (2002).
44. Marrero, J., Trujillo, C., Rhee, K. Y. & Ehrt, S. Glucose Phosphorylation Is Required for Mycobacterium tuberculosis Persistence in Mice. *PLOS Pathog.* **9**, e1003116 (2013).
45. Pellegrino, E. *et al.* Peroxisomal ROS control cytosolic Mycobacterium tuberculosis replication in human macrophages. *J. Cell Biol.* **222**, e202303066 (2023).
46. Vergne, I., Gilleron, M. & Nigou, J. Manipulation of the endocytic pathway and phagocyte functions by Mycobacterium tuberculosis lipoarabinomannan. *Front. Cell. Infect. Microbiol.* **4**, 187 (2014).
47. Lerner, T. R., Borel, S. & Gutierrez, M. G. The innate immune response in human tuberculosis. *Cell. Microbiol.* **17**, 1277–1285 (2015).
48. Kalam, H., Fontana, M. F. & Kumar, D. Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLOS Pathog.* **13**, e1006236 (2017).
49. Mathy, N. L. *et al.* Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. *Immunology* **100**, 63–69 (2000).
50. Gagneux, S. *et al.* Variable host–pathogen compatibility in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci.* **103**, 2869–2873 (2006).
51. Mvubu, N. E., Pillay, B., Gamieldien, J., Bishai, W. & Pillay, M. Mycobacterium tuberculosis strains exhibit differential and strain-specific molecular signatures in pulmonary epithelial cells. *Dev. Comp. Immunol.* **65**, 321–329 (2016).
52. Hiza, H. *et al.* Bacterial diversity dominates variable macrophage responses of tuberculosis patients in Tanzania. *Sci. Rep.* **14**, 9287 (2024).
53. Maqsood, M. I., Matin, M. M., Bahrami, A. R. & Ghasroldasht, M. M. Immortality of cell lines: challenges and advantages of establishment. *Cell Biol. Int.* **37**, 1038–1045 (2013).
54. Goig, G. A. *et al.* Whole-genome sequencing of Mycobacterium tuberculosis directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study. *Lancet Microbe* **1**, e175–e183 (2020).
55. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. (United States. Department of Energy. Office of Science, 2014).
56. Andrew, S. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
57. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
58. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).
59. Rau, A., Gallopin, M., Celeux, G. & Jaffrézic, F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146–2152 (2013).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

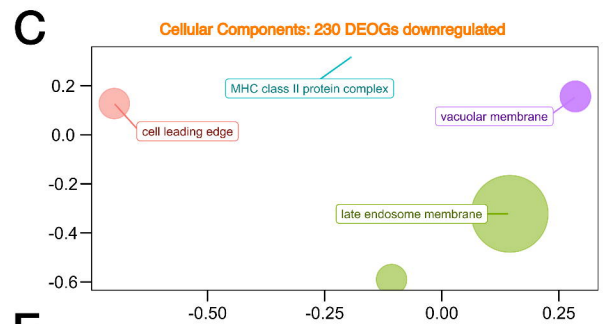
61. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
62. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. Camb. Mass* **2**, 100141 (2021).
63. Sayols, S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPublication Biol.* **2023**, 10.17912/micropub.biology.000811.
64. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
65. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
66. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).

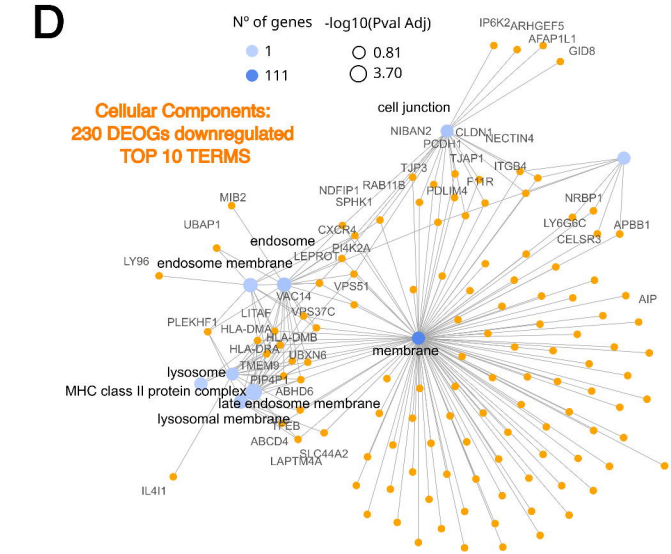

A
3 dpi
Infection ratio
Chimpanzee bacillus
M. bovis
L5
L6
B
4 hpi
1 dpi
3 dpi
High
Undetermined
Low
Cell
BoMac
THP-1
Compatibility
High
Undetermined
Low
Lineage
Chimpanzee bacillus
M. bovis
L5
L6
C
Day 6
live
apoptosis
necrosis
LogFoldChange
D
Log-fold Change

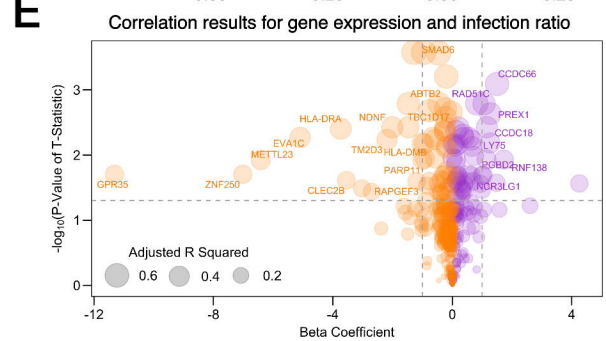

A
THP-1
Dim2 (20%)
Dim1 (28.3%)
B
BoMac
Dim2 (23.3%)
Dim1 (41.7%)
Infection Day
1
3
Lineage
Control
Chimpanzee
M.bovis
L5
L6
Virulence
Control
High
Undetermined
Low
C
High Upregulated THP-1
High Upregulated BoMac
Low Upregulated BoMac
Low Upregulated THP-1
2,995
452
441
3,479
135
230
240
172
D
THP-1: High vs Low
5,399 DEGs not shared
230 DEOGs
4,412 DEGs not shared
135 DEOGs
MIR210HG
ANGPTL4
PLOD2
BNIP3
HIF1A-AS3
AK4
IL23A
ZNF395
ADM
SULF2
SLC5A8
SLC6A11
EDNRA
CEACAM6
F13A1
PRAC2
TMSB4Y
ZEY
PRSS30P
DDX3Y
EIF1AY
RPS4Y1
−Log10 padj
Log2 Fold Change
E
BoMac: High vs Low
5,399 DEGs not shared
230 DEOGs
4,412 DEGs not shared
135 DEOGs
HECW2
TSPAN12
GYPB
ODAD2
ADGRV1
RADX
CFH
ISM1
CENPE
CLGN
bta-mir-2887-1
MEIOC
CIITA
HCK
PARVB
TNFSF18
APOL3
IFI47
ID1
JPH3
−Log10 padj
Log2 Fold Change

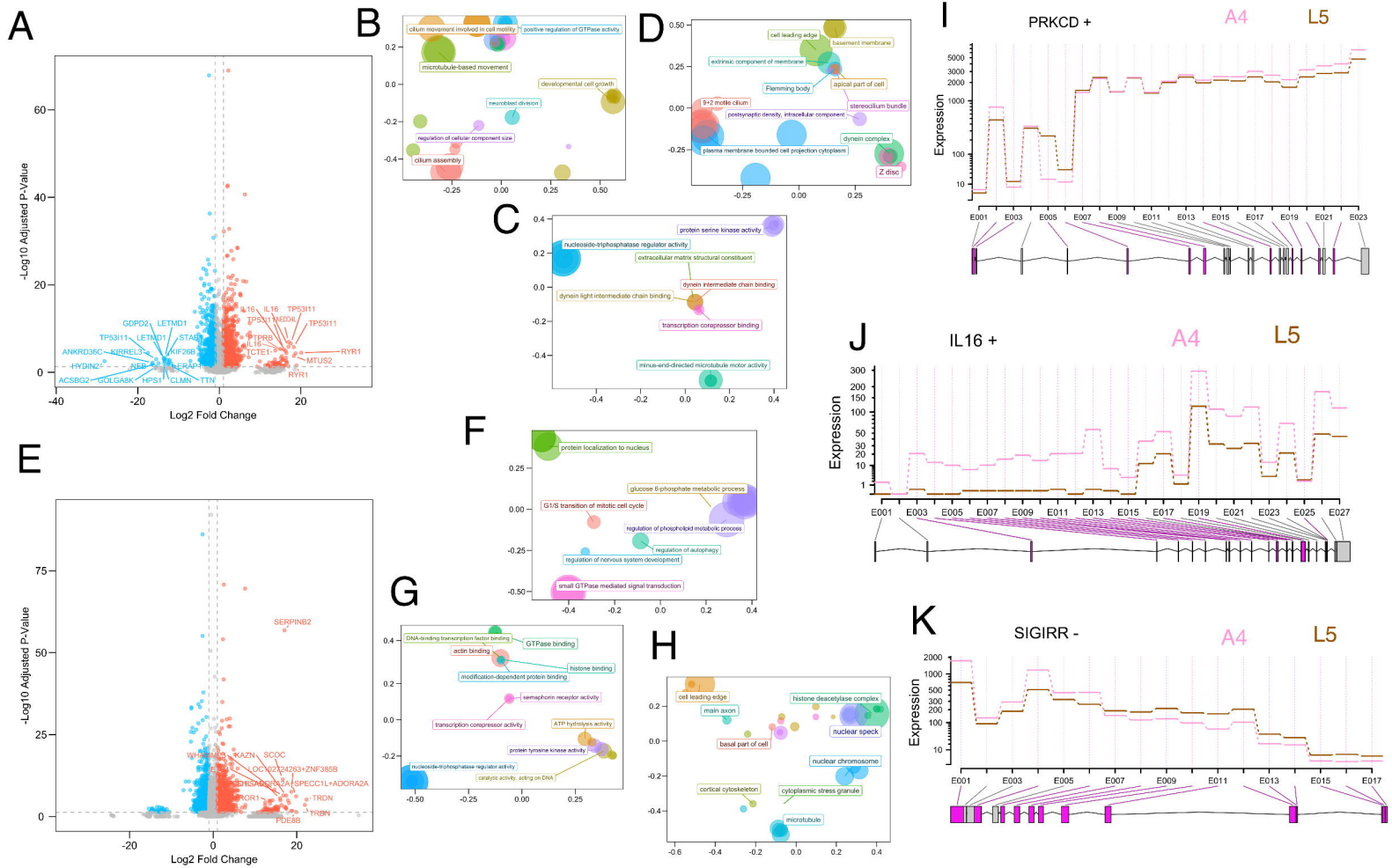

A
Biological Process: 135 DEOGs upregulated
nuclear division
protein localization to condensed chromosome
ether lipid metabolic process
blastocyst development
response to mineralocorticoid
2'-deoxyribonucleotide metabolic process
metaphase chromosome alignment
double-strand break repair
regulation of chromosome segmentation
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.4
-0.2
0.0
0.2

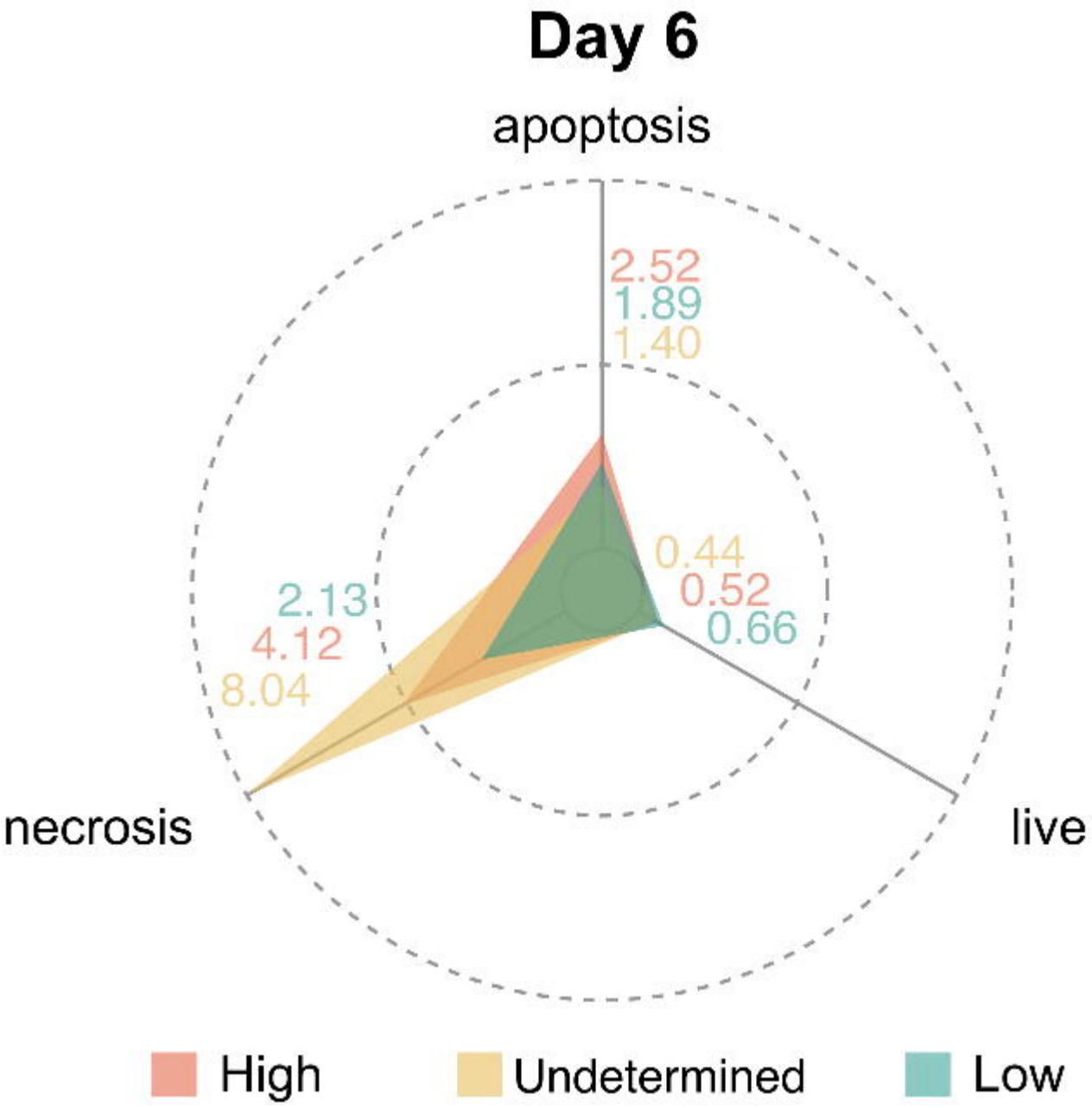

B
Nº of genes
2
18
-log10(Pval Adj)
1.80
6.03
RRM2
DNA replication
RMI1
DNA duplex unwinding
ATR
POLQ
WRN
MCM6
CHD1L
FANCM
INIP
ZMYND11
RBBP8
DNA repair
RNF138
RAD51C
BRCA2
WDHD1
protein localization
to kinetochore
KNL1
cell cycle
CHAMP1
CENPC
ANLN
cell division
E2F8
CENPE
CENPF
CDC7
TTK
mitotic cell cycle
NDC80
mitotic spindle organization
centrosome duplication
metaphase plate congression
KIF15
STIL
MEIOC
CENPQ
Biological Process:
135 DEOGs upregulated
TOP 20 TERMS


C
Cellular Components: 230 DEOGs downregulated
MHC class II protein complex
cell leading edge
vacuolar membrane
late endosome membrane
0.2
0.0
-0.2
-0.4
-0.6
-0.50
-0.25
0.00
0.25


D
Nº of genes
1
111
-log10(Pval Adj)
0.81
3.70
IP6K2
ARHGEF5
AFAP1L1
GID8
Cellular Components:
230 DEOGs downregulated
TOP 10 TERMS
cell junction
NIBAN2
CLDN1
NECTIN4
PCDH1
TJAP1
ITGB4
TJP3
F11R
RAB11B
PDLIM4
NDFIP1
SPHK1
MIB2
NRBP1
UBAP1
CXCR4
LY6G6C
APBB1
CELSR3
endosome
PI4K2A
LEPROT
LY96
endosome membrane
VPS51
VAC14
AIP
LITAF
VPS37C
PLEKHF1
HLA-DMA
HLA-DMB
HLA-DRA
membrane
UBXN6
TMEM9
lysosome
PIP4P1
MHC class II protein complex
ABHD6
late endosome membrane
lysosomal membrane
TPCB
ABCD4
SLC44A2
LAPTM4A
IL4I1

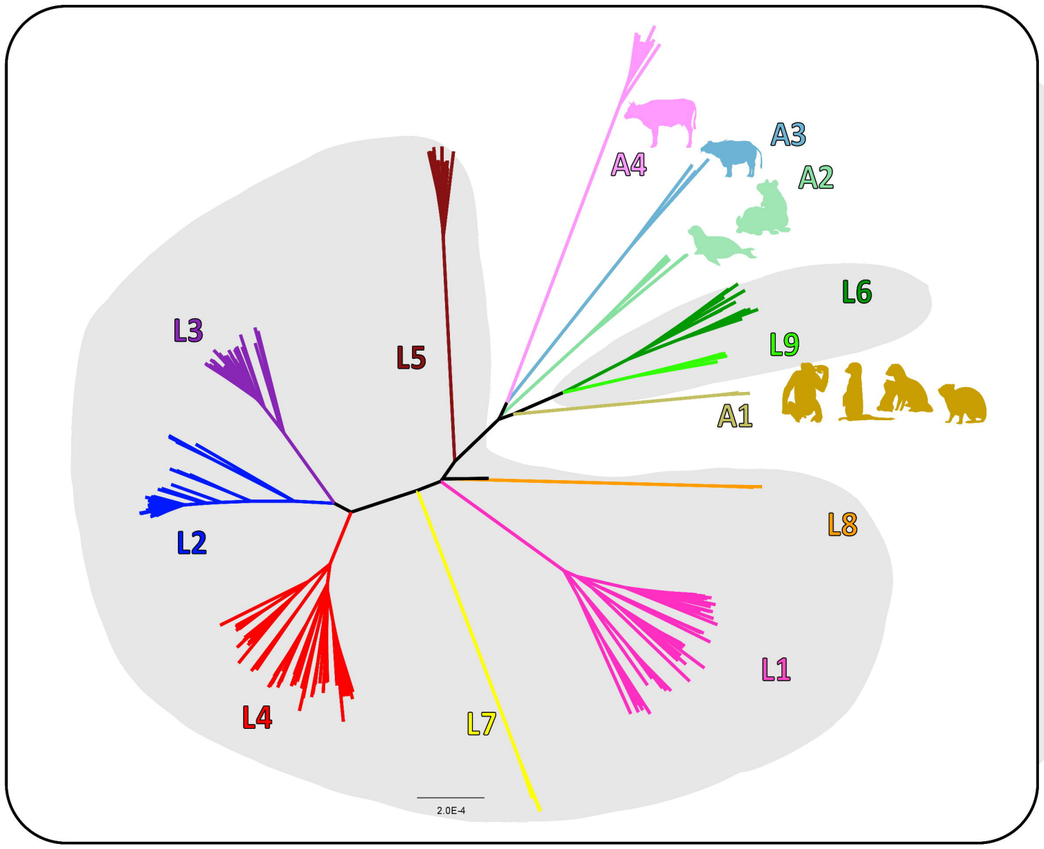

E
Correlation results for gene expression and infection ratio
SMAD6
CCDC66
ABTB2
RAD51C
HLA-DRA
NDNF
TBC1D17
PREX1
CCDC18
EVA1C
METTL23
TM2D3
HLA-DMB
LY75
PARP11
PGBD4
RNF138
GPR35
ZNF250
CLEC2B
RAPGEF3
NCR3LG1
Adjusted R Squared
0.6
0.4
0.2
-log10(P-Value of T-Statistic)
Beta Coefficient
-12
-8
-4
0
4
3
2
1
0


A
-Log10 Adjusted P-Value
Log2 Fold Change
GDPD2
LETMD1
TP53I11
LETMD1
STAB1
ANKRD36C
KIRREL3
KIF26B
HYDIN2
NEB
ERAP1
ACSBG2
GOLGA8K
HPS1
CLMN
TTN
IL16
IL16
TP53I11
TP53I11
NEDD4L
TP53I11
PTPRB
IL16
TCTE1
RYR1
MTUS2
RYR1
B
cilium movement involved in cell motility
positive regulation of GTPase activity
microtubule-based movement
developmental cell growth
neuroblast division
regulation of cellular component size
cilium assembly
C
protein serine kinase activity
nucleoside-triphosphatase regulator activity
extracellular matrix structural constituent
dynein intermediate chain binding
dynein light intermediate chain binding
transcription corepressor binding
minus-end-directed microtubule motor activity
D
cell leading edge
basement membrane
extrinsic component of membrane
Flemming body
apical part of cell
9+2 motile cilium
stereocilium bundle
postsynaptic density, intracellular component
dynein complex
plasma membrane bounded cell projection cytoplasm
Z disc
E
-Log10 Adjusted P-Value
Log2 Fold Change
SERPINB2
WHAMMP3
KAZN
SCOC
LOC102724263+ZNF385B
SPECC1L+ADORA2A
TRDN
TRDN
PDE8B
F
protein localization to nucleus
glucose 6-phosphate metabolic process
G1/S transition of mitotic cell cycle
regulation of phospholipid metabolic process
regulation of autophagy
regulation of nervous system development
small GTPase mediated signal transduction
G
DNA-binding transcription factor binding
GTPase binding
actin binding
histone binding
modification-dependent protein binding
semaphorin receptor activity
transcription corepressor activity
ATP hydrolysis activity
protein tyrosine kinase activity
nucleoside-triphosphatase regulator activity
catalytic activity, acting on DNA
H
cell leading edge
histone deacetylase complex
main axon
nuclear speck
basal part of cell
nuclear chromosome
cortical cytoskeleton
cytoplasmic stress granule
microtubule
I
PRKCD +
A4
L5
Expression
E001
E003
E005
E007
E009
E011
E013
E015
E017
E019
E021
E023
J
IL16 +
A4
L5
Expression
E001
E003
E005
E007
E009
E011
E013
E015
E017
E019
E021
E023
E025
E027
K
SIGIRR -
A4
L5
Expression
E001
E003
E005
E007
E009
E011
E013
E015
E017


Day 6
apoptosis
2.52
1.89
1.40
0.44
0.52
0.66
2.13
4.12
8.04
necrosis
live
High
Undetermined
Low

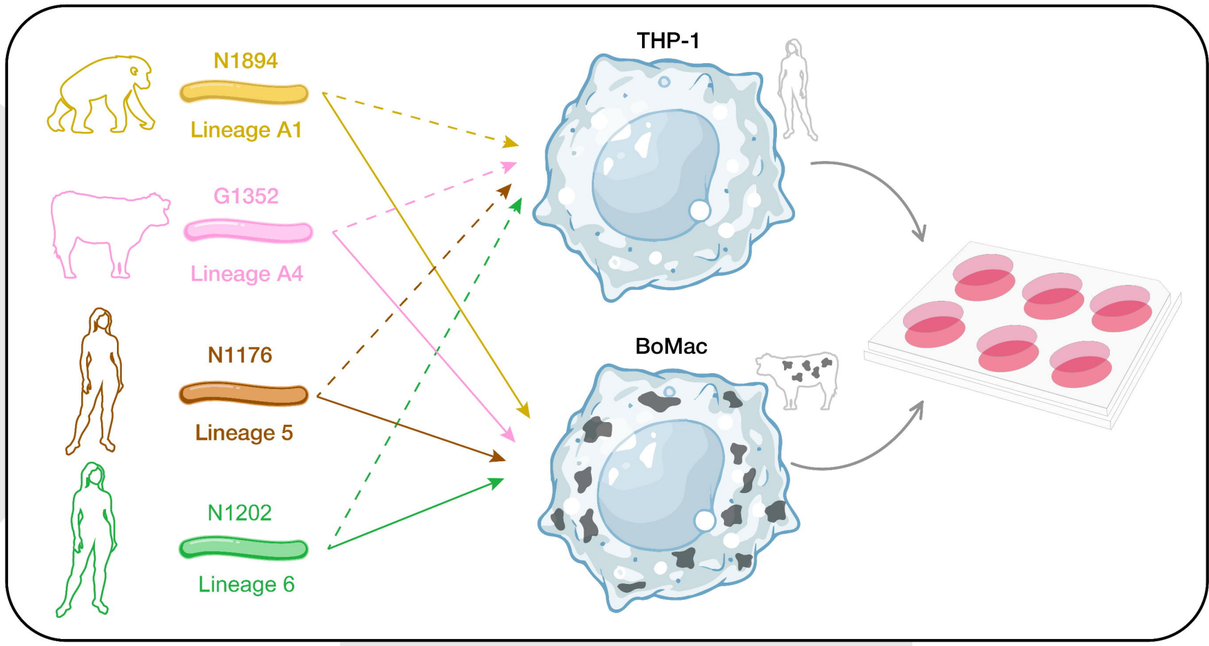
A

## PHYLOGENETIC COMPATIBILITIES

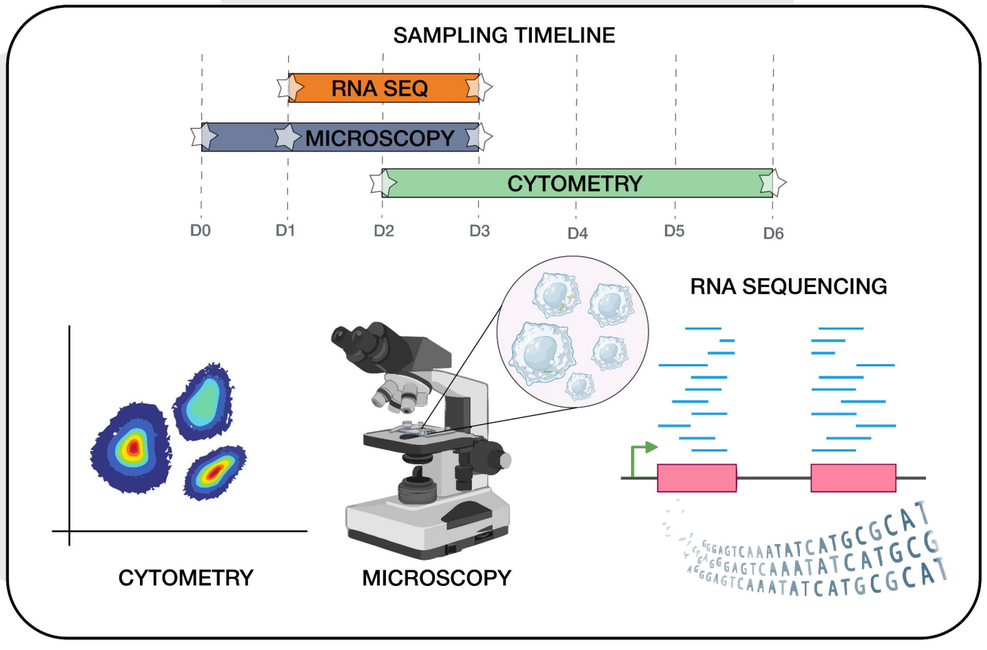



B

## INFECTION COMBINATIONS

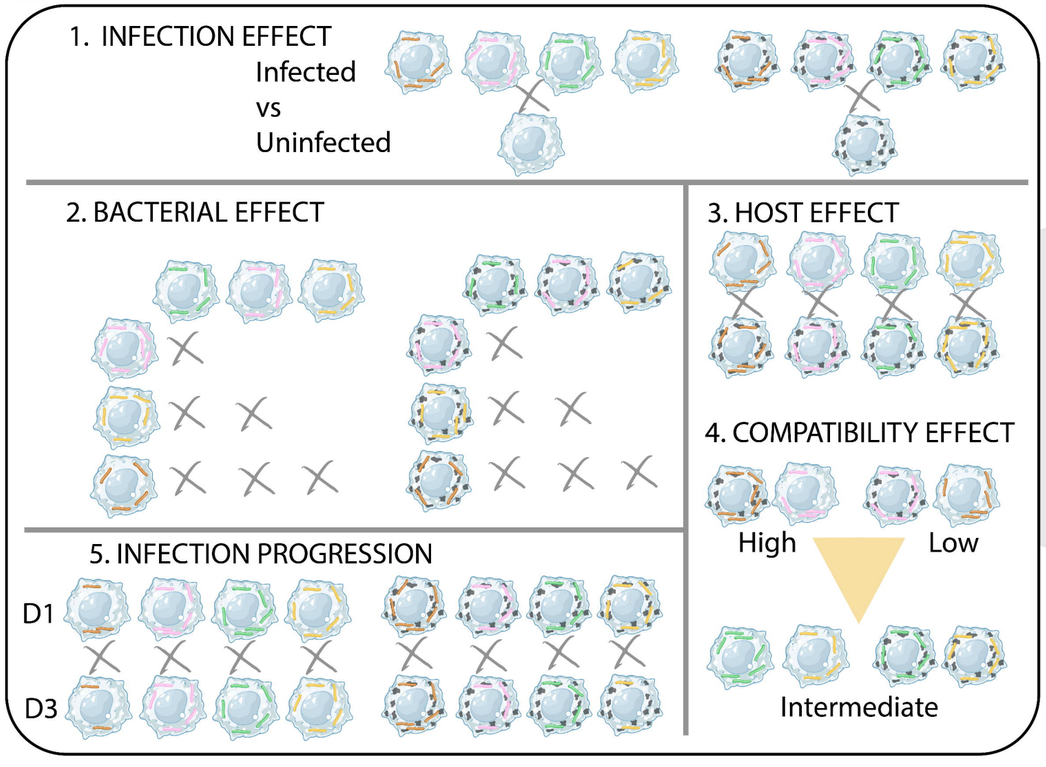



C




## INFECTION ANALYSIS

D




## INFECTION COMPARISONS