Cell Reports, Volume *30*

Supplemental Information

Dual RNA-Seq of Mtb-Infected Macrophages *In Vivo* Reveals Ontologically Distinct Host-Pathogen Interactions

Davide Pisu, Lu Huang, Jennifer K. Grenier, and David G. Russell

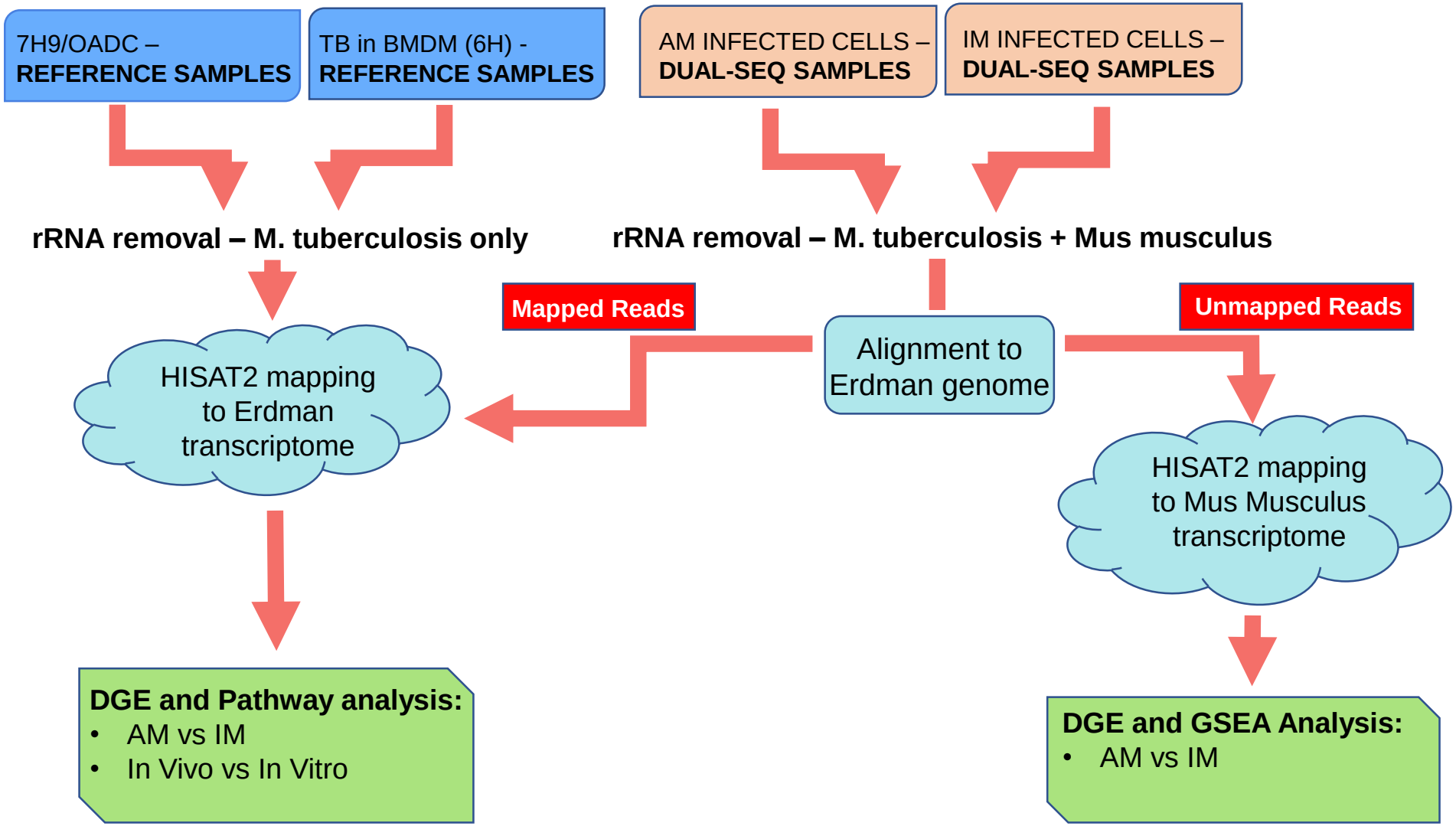

Figure S1. Diagram of the Data Analysis pipeline. Related to Figure 1 and Table 1.

(A) Mtb-only samples were aligned against a custom GTF file to remove rRNA reads and then processed with Hisat2 as illustrated in methods. Dual-seq samples were processed for rRNA removal as well and the reads aligned to the Erdman genome. Mapped reads were then aligned to the Mtb transcriptome, while unmapped reads were aligned to the Mus musculus transcriptome. Raw read counts were obtained for both species using the pipeline described in the methods section of the paper.

For *M.tuberculosis* transcriptome analysis the two bacterial datasets from IM- and AM-isolated Mtb were compared against each other as well as against transcriptional profiles generated from cultures of mid-log growing bacteria in 7H9/OADC bacterial medium and early stage (6 hour) infected bone marrow-derived murine (BMDM) macrophages (Reference samples). For *Mus musculus* transcriptome analysis we compared infected and uninfected samples to reveal ontology and infection-specific transcriptional signatures.

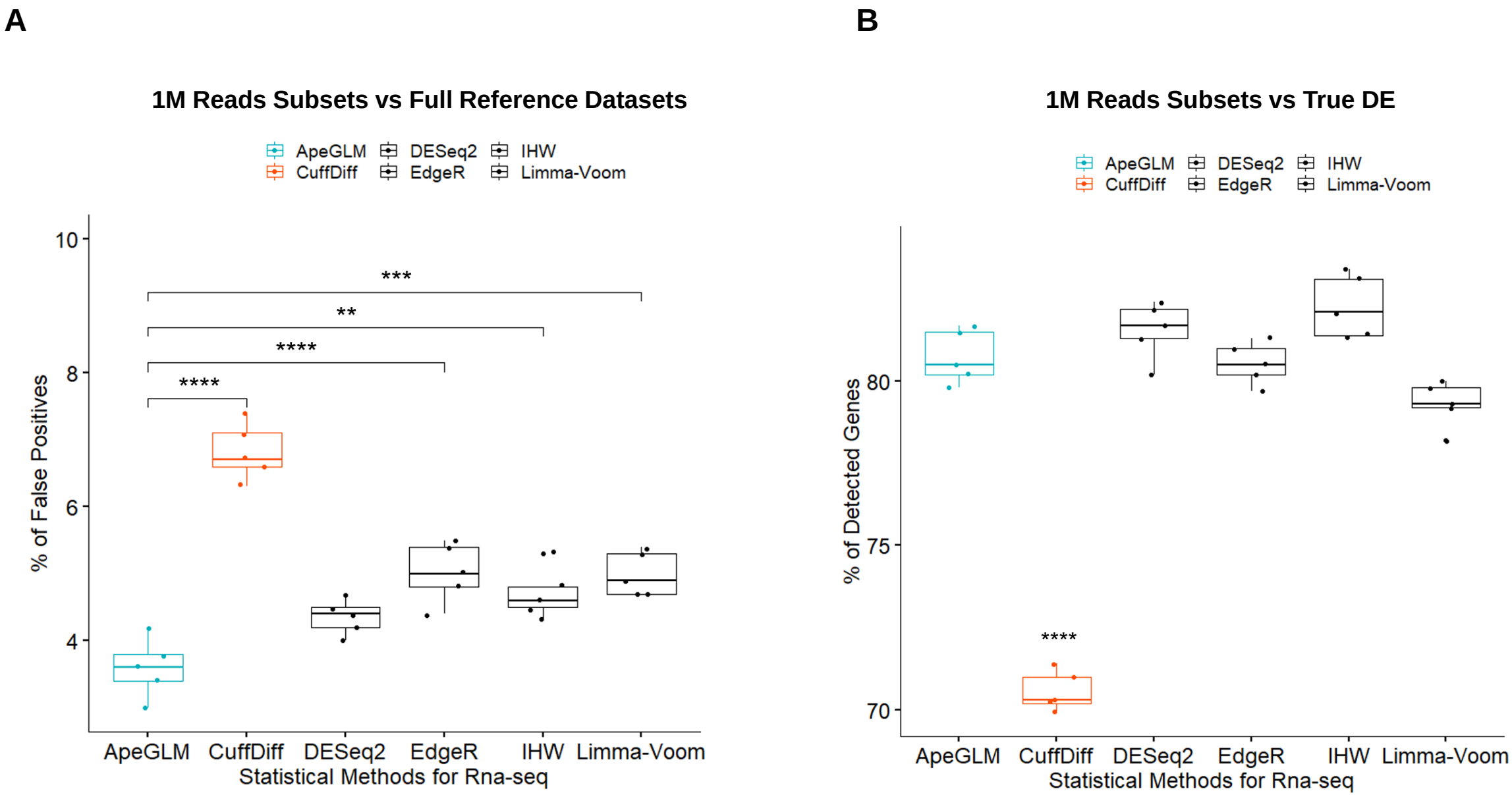

Figure S2. APEGLM improves the performance of the differential expression analysis for samples with low sequencing depth. Related to Figure 1.

(A) Percentage of false positives (genes detected as differentially expressed (*abs*($\log_2$FC >1, FDR < 0.05) in the subsets of 1M reads but not called DE in the analysis of the full datasets) for the most common used statistical approaches in Rna-seq analysis. APEGLM outperformed all other methods with a false positive percentage of < 4%. Cuffdiff resulted in very unreliable estimates both in terms of false positives and detection power (~70% vs an average of 80%, data not shown). **(B)** All methods were able to detect more than 80% of the "TRUE DE" genes in the DGE analysis of the 1M reads subsets, excluding CuffDiff for which the detection power was ~70%. P-values were calculated using one-way ANOVA followed by Tukey's post hoc test.

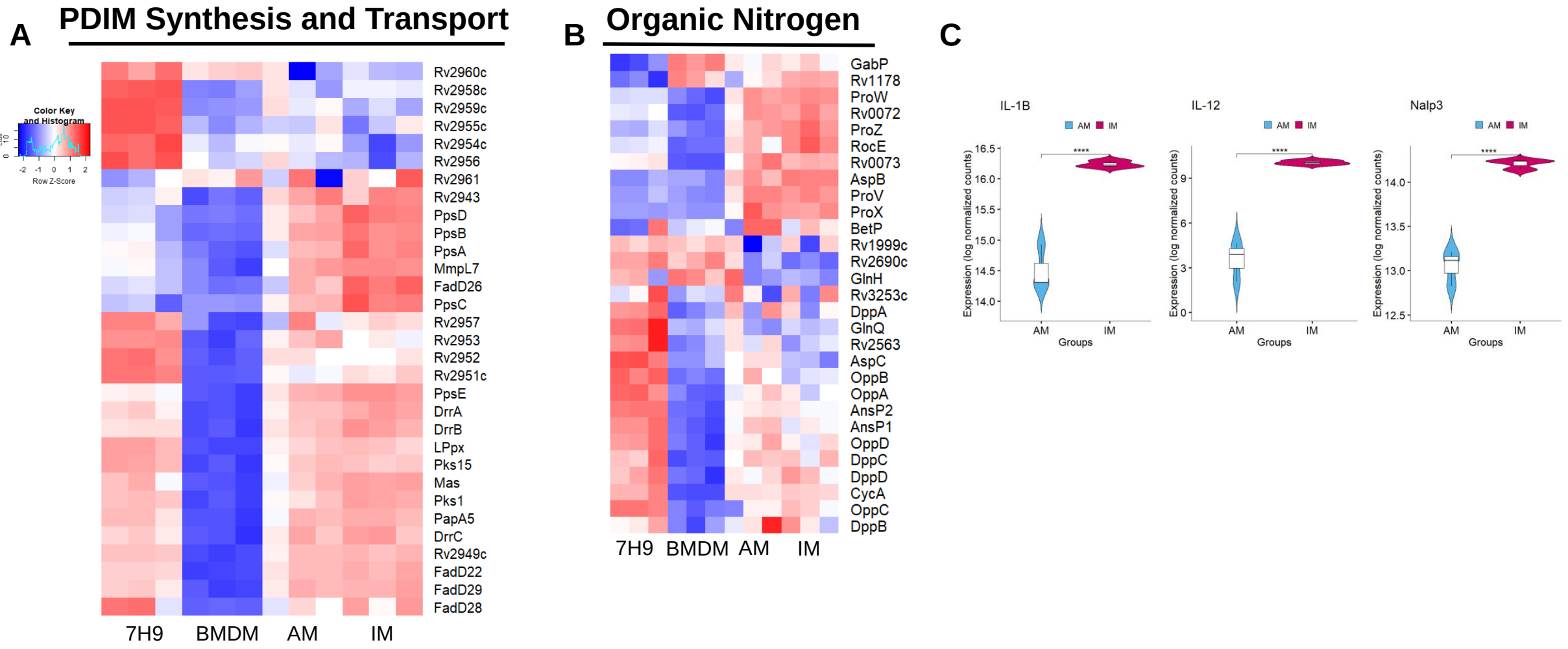

Figure S3. Relative expression levels for Mtb gene-sets specifically upregulated *in vivo*. Related to Figure 2.

(A-B) Heatmaps showing relative expression levels for genes in the PDIM and organic nitrogen pathways in Mtb. **(C)** Violin plots showing expression levels for the *Il-1B*, *Il-12*, *Nalp3* host genes in the AM- and IM-infected macrophages. Statistical significance is shown on the picture. (* p-adj. <0.05, ** p-adj. <0.01, *** p-adj. <0.001, ****p-adj. <0.0001). Q-values for comparisons among the groups were calculated using the Wald test as implemented in the DESeq2 pipeline.

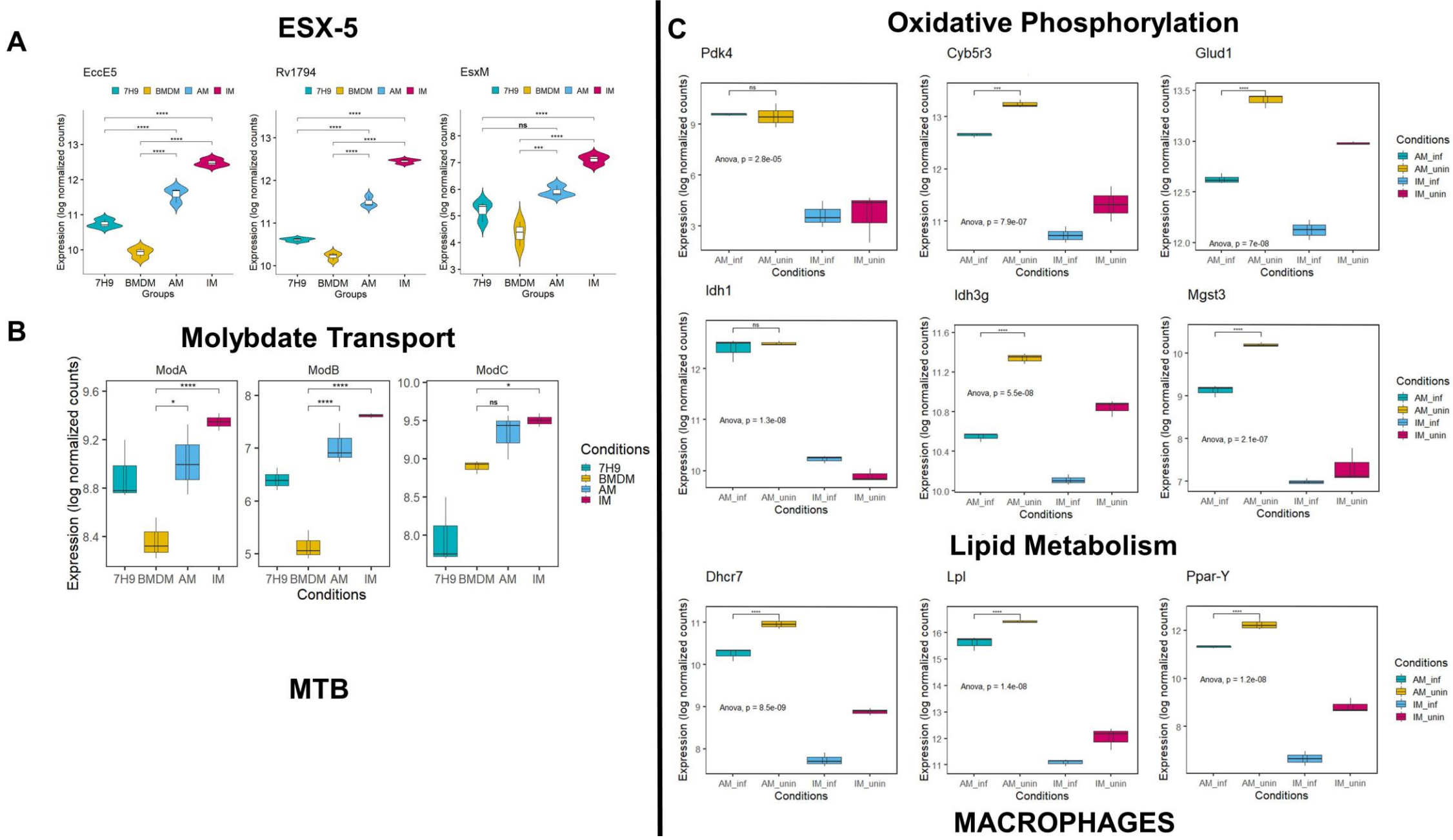

Figure S4. Relative expression levels for cluster of genes in both Mtb and the host cells. Related to Figure 2 and 4.

(A) Violin-plots showing expression levels (in log-normalized counts) of genes involved in the ESX-5 pathway. **(B)** Boxplots showing expression levels (in log-normalized counts) of genes involved in Molybdate transport in Mtb. **(C)** Boxplots showing expression levels (in log-normalized counts) for genes involved in oxidative phosphorylation and lipid metabolism in different host cell conditions. Statistical significance is shown on the picture. (* p-adj. <0.05, ** p-adj. <0.01, *** p-adj. <0.001, ****p-adj. <0.0001). Q-values for comparisons among the groups were calculated using the Wald test as implemented in the DESeq2 pipeline.

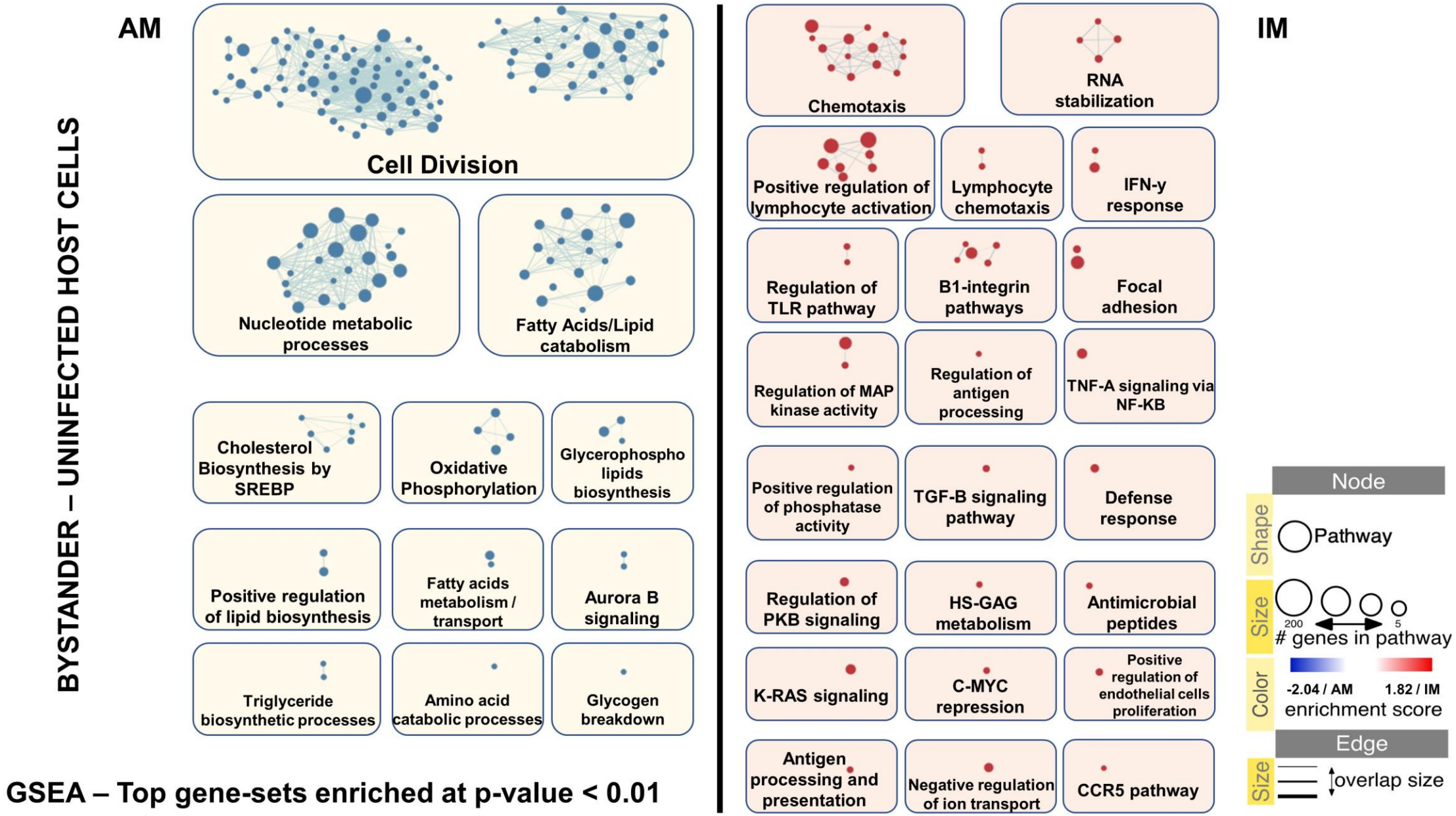

Figure S5. Enrichment map comparing pathways upregulated in uninfected-AM vs uninfected-IM. Related to Figure 3.

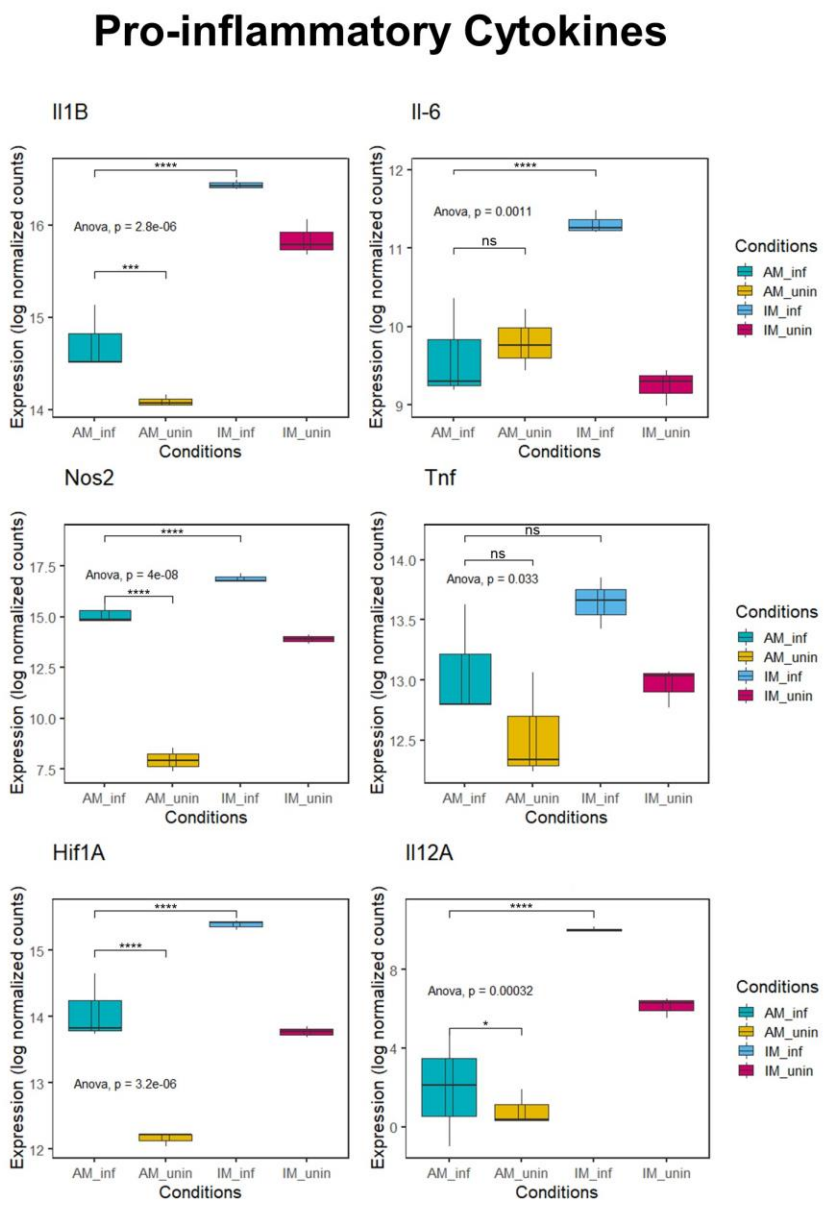

Figure S6. Relative expression levels of pro-inflammatory cytokines in different populations of host macrophages. Related to Figure 4.
(A) Boxplots showing relative expression levels (in log-normalized counts) for pro-inflammatory cytokines in different host cell conditions. Statistical significance is shown on the picture. (* p-adj. <0.05, ** p-adj. <0.01, *** p-adj. <0.001, ****p-adj. <0.0001). Q-values for comparisons among the groups were calculated using the Wald test as implemented in the DESeq2 pipeline.

Sample	Raw Reads	Mouse				*M.tuberculosis*			
		rRNA	% of Raw Reads	Mouse Mapped	%	rRNA	% of Raw Reads	Mtb Mapped	% of Raw Reads
No Enrichment 1	27.7 M	4.6 M	16.8%	9.4 M	41%	0.09 M	0.3%	0.03M	0.14%
No Enrichment 2	20.8 M	3.5 M	16.9%	6.2 M	37%	0.03 M	0.2%	0.04 M	0.2%
BMDM – 50% host RNA removal	54.2 M	14.1 M	26%	21.9 M	40%	0.3 M	0.55%	4.9 M	9 %
BMDM – 30% host RNA removal	55.6 M	4.7 M	8%	35.1 M	63.1%	0.15M	0.27%	2.9 M	5.2%
Control – TB only RNA extraction from BMDM	29.1 M	1.9 M	6.5%	0.01 M	0%	2.1 M	7.2%	18.6 M	64 %

Table S1. Number of reads for each of the test samples matching to either *M.tuberculosis* or Mus Musculus genomes. Related to Figure 1.

Comparison	Number of DGE genes, p.adj < 0.05	Breakdown
7H9-OADC vs IM	2486 LFC > 0	1207 up in 7H9-OADC 1279 up in IM
7H9-OADC vs AM	2019 LFC > 0	942 up in 7H9-OADC 1077 up in AM
7H9-OADC vs BMDM	2934 LFC > 0	1474 up in 7H9-OADC 1460 up in BMDM
BMDM vs IM	2610 LFC > 0	1346 up in BMDM 1264 up in IM
BMDM vs AM	2657 LFC > 0	1306 up in BMDM 1351 up in AM
AM vs IM	319 LFC > 0	96 up in AM 223 up in IM

Table S2. Number of genes differentially expressed by *M.tuberculosis* in each of the different combinations of environmental conditions analyzed in this study. Related to Figure 3.