

Supplemental Methods

RNA sequencing and raw data processing

We generated 100 RNA-seq libraries encompassing pairs of baseline and *Mtb*-challenged samples for 16 control (HC), 14 PrEP and 20 PLWH subjects. Indexed libraries were combined in four to a maximum of 12 libraries per pool, with library pair per subject on the same pool, and sequenced as 100 bp single-end reads with an expected throughput of 25 million reads per library. Raw fastq files were quality-checked with FastQC (v0.11.8) (1), RSeQC (v2.6.1) (2), and subjected to trimming with cutadapt (v2.6) (3) to remove leftover adaptors and low-quality bases (≤ 20 Phred score) allowing for a maximum read size of 99bp and a minimum of 75bp. Sequences were then aligned to the human reference genome GRCh38.p13 (ENSEMBL v99) (4) with STAR (v2.7.3a) (5). Aligned BAM files were input to Salmon version 1.1.0 (6) for expression quantification.

RNA expression matrix and gene filtering

Text files containing estimated counts and transcript per million (TPM) values for 60,676 annotated features, for each subject, were used to create a gene level expression matrix in R (v3.6.1) (7) using tximport (v1.12.3) (8) and biomaRt (2.40.5) (9). We adjusted the estimated counts by gene length and abundance (TPM) employing tximport's method "lengthScaledTPM" as discussed by Love *et al.* (10). Next, we filtered out genes with estimated counts ≤ 10 in more than 70 libraries and ran edgeR (11) function "calcNormFactors" to generate scaling normalization factors with the TMM method (12). Downstream analyses were focused on protein-coding genes with ≥ 100 estimated counts in at least 11 HC or 10 PrEP or 14 PLWH libraries, resulting in 10,362 genes for differential gene expression analyses. The gene matrix limited to the testable genes and the TMM scaling factors were input in limma (13) voom (14) (v3.40.6) to create log₂-CPM normalized expression matrices and associated weights.

ATACseq and ChIPseq data processing

We profiled chromatin accessibility and H3K27 acetylation for non-stimulated and *Mtb* challenged AMs from 44 subjects and 24 subjects, respectively. Libraries were sequenced as 100bp paired end reads for ATACseq and 100bp single end reads for ChIPseq. Nextera adaptors

and low quality reads were removed with TrimGalore v0.6.5 (3). Reads were then aligned to both human (hg38) and *Mtb* (H37Rv) genomes using BWA v0.7.17 default parameters (15). PICARD v2.18.9 was used to mark duplicates and to assess the fragment length distribution. Reads aligning to the mitochondrial genome were removed with samtools v1.9 (16). Next, alignmentSieve v3.3.2 was used to: (i) select unique paired-end reads (ATACseq) or unique single end reads (ChIPseq) using the --samFlagExclude 1804 and --samFlagExclude 1796 flags, respectively; (ii) remove ENCODE hg38 blacklisted regions (17); and (iii) to extract fragments between 40 – 2000bp in length (18). After quality control with MultiQC v1.8 (19), ATACseq libraries from two subjects yielded less than 20 million unique reads, and one library deviated from the periodic nucleosome pattern for ATACseq. Both the non-stimulated and *Mtb* challenged libraries of subjects failing QC were removed from further analysis.

Peak annotation

We integrated two approaches to annotate targeted genes for the 53,040 ATACseq tested peaks and 42,595 H3K27ac regions from ChIPseq. First, we assigned peaks based on distance < 5kb from a gene transcriptional start site (TSS) using ChIPSeeker v1.24 (20). Next, we intersected our peak set with the GeneHancer 2017 dataset using mergeByOverlaps from IRange package v2.22.2 (21, 22). GeneHancer includes a collection of regulatory regions assigned to genes based on the integration of multiple genome-wide regulatory databases (21). By combining the TSS and GeneHancer approaches ~80% of the tested peaks were assigned to at least one gene (Figure S5).

Chromatin accessibility and H3K27ac filtering, and library normalization

We performed a count-based quantification of accessible chromatin and H3K27 acetylation as shown in (23). Briefly, for each library MACS2 v2.2.6 callpeak was used with --call-summits in BAMPE mode (ATACseq) or with --shift -37 and --extsize 73 flags (ChIPseq) to identify regions of open chromatin or H3K27 acetylation. Next, the summit of each peak was extended 250bp in both directions resulting in fixed-width peaks of exactly 501bp. We merged fixed-width peaks overlapping in at least two samples with bedtools v2.26 (24). By restraining the overlap of accessible chromatin between samples to fixed-width peaks we avoid merging multiple

independent regions into a single large peak. We identified 116,547 regions of accessible chromatin and 120,833 of H3K27 acetylation present in at least two libraries of each approach, respectively. Next, for each library, featureCounts v1.6.3 was used to count the number of unique fragments (ATACseq) or unique reads (ChIP) overlapping the targeted regions (25). A matrix containing the individual counts for each approach was exported to R. Next, low count peaks were removed with edgeR's v3.30.3 filterByExpr using 25 counts as cut off and the quantification matrices were normalized with the upperquartile method implemented on calcNormFactors (11). Counts were then transformed to log2 count per million (CPM) using limma's v3.44.3 voom and the resulting matrix was used for downstream analysis (13, 14) .

Linear models

Differences in chromatin accessibility, H3K27 acetylation and transcriptomic between non-stimulated and *Mtb* challenged AMs were analysed in paired designs. To create the models we employed the function model.matrix from stats package (v4.0.2) using the factors: sample group (HC, PrEP and PLWH), *Mtb* status (baseline and *Mtb* challenge) and subject IDs (to enable paired design models) as denoted by the following equation:

$$\text{Log2}(CPM)_f \sim \beta_0 + \sum_{i=1}^{n-1} \beta_i \cdot \chi_i + \beta_{(HC.Mtb)} \cdot \chi_{(HC.Mtb)} + \beta_{(PLWH.Mtb)} \cdot \chi_{(PLWH.Mtb)} + \beta_{(PrEP.Mtb)} \cdot \chi_{(PrEP.Mtb)} + \varepsilon$$

Log2(CPM) represents the depth-normalized log-quantifications with f indicating each gene or peak tested. β_0 is the intercept term, representing log-quantification for an arbitrary non-challenged library acting as reference. The term β_i represents the individual-wise difference between the non-challenged sample from the i -th individual and the reference sample. χ_i is thus a dummy variable marking the identity of the $(n-1)$ individuals in the dataset, excluding the reference. The following three terms in the equation capture the group-specific *Mtb* challenge effects nested within each group. The β value for each of these terms corresponds to the logFC for the *Mtb* challenge per group after accounting for the inter-individual variance at the baseline present in our datasets. For ATAC-seq we included two additional standardized technical covariates in the linear model: the total number of unique reads per library and Bioanalyser average fragment length.

Differential gene expression, chromatin accessibility and H3K27 acetylation in response to Mtb.

Differences in RNA expression, chromatin accessibility and H3K27 acetylation between non-stimulated and *Mtb* challenged AMs were tested in parallel with limma. After defining the linear models, we used makeContrasts to guide the calculation of i) group-specific *Mtb* effects and ii) differential *Mtb* responses between groups (for RNA-seq only). For ii, a) $\beta_{(PLWH.Mtb)} - \beta_{(HC.Mtb)}$ was the difference in the response to *Mtb* of PLWH minus HC, b) $\beta_{(PrEP.Mtb)} - \beta_{(HC.Mtb)}$ the response difference of PrEP minus HC and c) $\beta_{(PLWH.Mtb)} - \beta_{(PrEP.Mtb)}$ the difference between PLWH minus PrEP. Of note, for ATACseq, we removed flowcell batch effects with ComBat as the paired nature of our approach prevented the inclusion of this covariate in a linear model (26, 27). We then applied voomMod from cbcSEQ to recalculate sample weights after batch correction. Next, we fitted the linear models with limma's function lmFit, derived coefficients for the contrasts defined via makeContrasts with eBayes (13) and extracted results with topTable. For ATAC and ChIP-seq we estimated false discovery rate (FDR) with the qvalue package v2.20.0 and considered peaks as significant if $FDR < 0.05$ and absolute log fold-change (logFC) > 0.2 . For RNA-seq we employed the FDR procedure implemented in stageR v 1.6.0 (28) and considered genes as differentially expressed if absolute logFC ≥ 0.2 and stageR *p*-val (stgr.p) ≤ 0.05 .

Gene enrichment analysis

Pathways and gene ontology (GO) enrichment analysis were performed with clusterProfiler package version 3.16.0 and ReactomePA version 1.32.0, using differentially expressed genes (DEGs), genes assigned to DOCs or differentially acetylated (DAc) regions. Three different databases and functions were used for the enrichment analysis, KEGG was tested with enrichKEGG function, Reactome with enrichPathway and GO biological process was tested via enrichGO (29). GO terms and pathways that had less than five assigned genes were excluded. Next, we merged the results from the three approaches and used the Benjamini-Hochberg's method to estimate the FDR. For DEGs enrichment we used $FDR \leq 0.1$ while the cutoff for ATAC and ChIP-seq was $FDR \leq 0.05$.

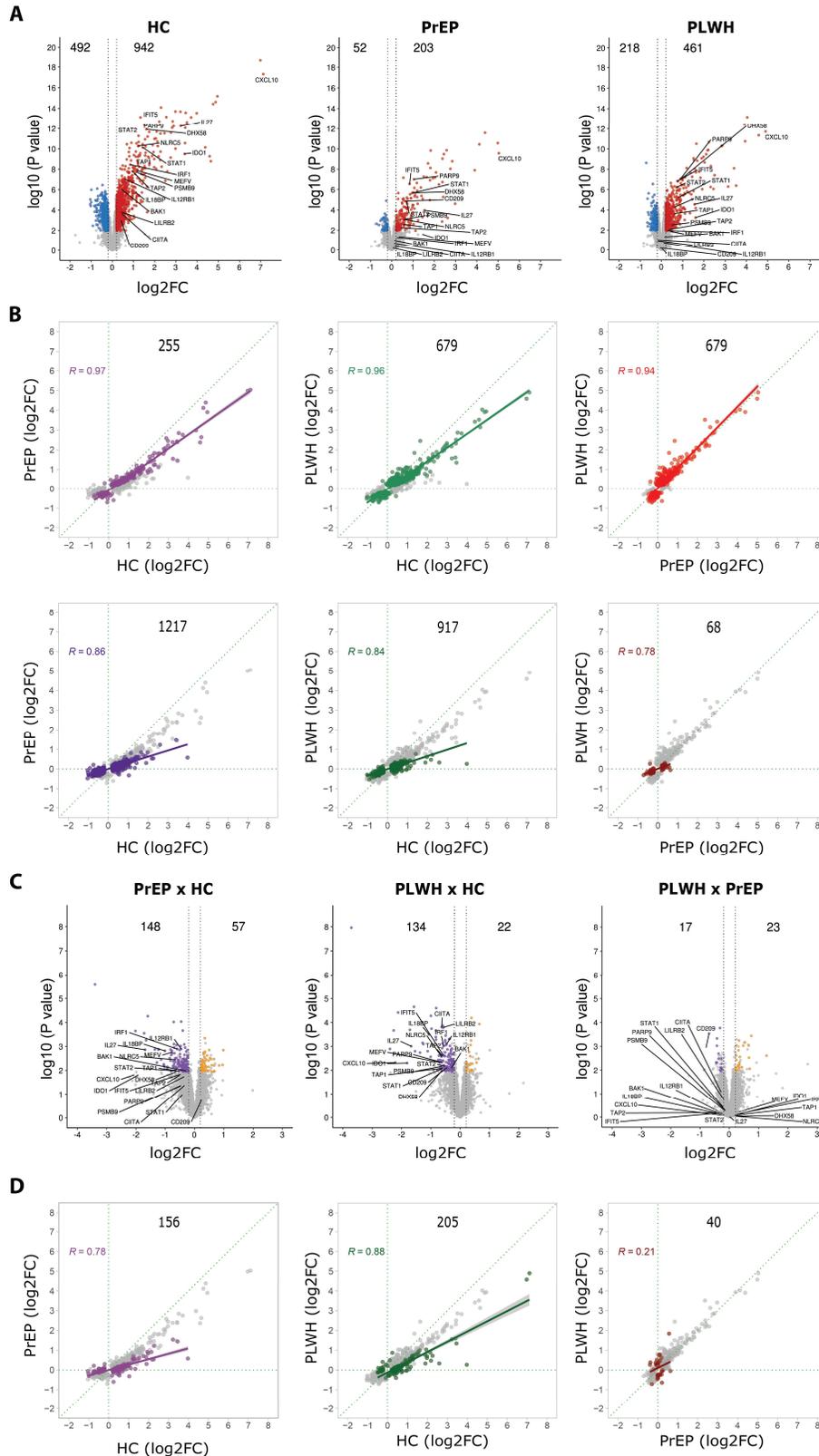
Data visualization

To visualize region of accessible chromatin as genomic tracks we applied a scaling factor based on edgeR's upperquartile normalization method. Of note, as the ATACseq and ChIPseq quantifications considered only fragments/reads in peaks (FrIP), normalizing genomic tracks by total library depth does not correspond to the quantification analysed by the linear models. To address that, a scaling factor per sample was calculated for genomic tracks as the reciprocal of $(\frac{\text{Upperquartile norm} \times \text{Total FrIP}}{1000000})$. DeepTool's bamCoverage was used to produce scaled BigWig and bedGraphs from analysed BAM files (30). We used SparK to calculate the mean coverage and standard deviation per base-pair per group for the genomic track plots (31). Volcano and rankPlots were produced by adapting tools from the MAGeCKFlute package (32). For the boxplots we used the residual log2CPM after regressing out covariates included in the linear models with removeBatchEffect from limma (13). Density, Manhattan, bar and dot plots were produced using ggplot2 (33).

Motif enrichment and footprint analysis

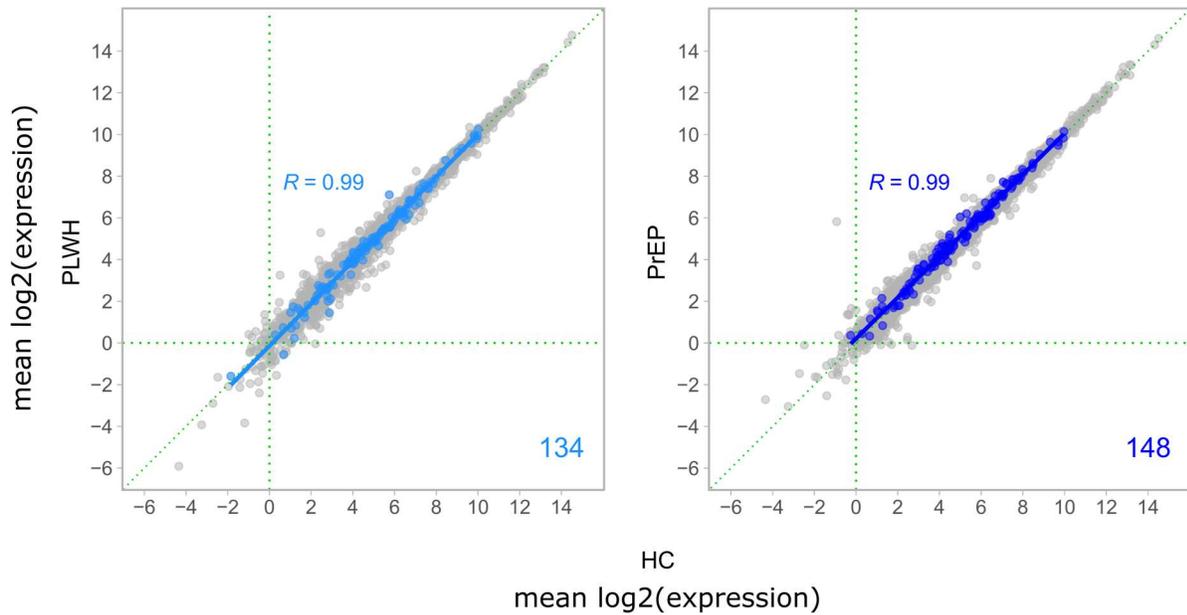
The findMotifsGenome function (HOMER v4.11) was used to assess enrichment of TF motifs in DOCs and DAc regions versus all other tested peaks as background (34). We used HINT-ATAC from RGT v0.13.0 to identify TF footprints and to estimate TF activity over footprints located in DOCs and DAc regions (35). The analysis for differential TF activity over footprint compared the average depth of footprints encompassing motifs of a given TF between paired non-stimulated and *Mtb* challenged AM. This approach produced an activity score per TF with a corresponding *p* value for each subject t (35). TF containing an average of less than 50 footprints across samples were excluded from the analysis. We used a meta-analysis to combine *p* values from each paired comparison using the sum of Z Stouffer's method from metap v1.4 package and applied a q-value FDR correction over meta-analysed *p* values. TF with FDR < 0.05 and activity score > 0.005. We used a hypergeometric test to estimate the enrichment of IRF9 and ZNF684 active footprints in the TSS of DEGs and GO/Pathway analysis.

Supplemental Figures



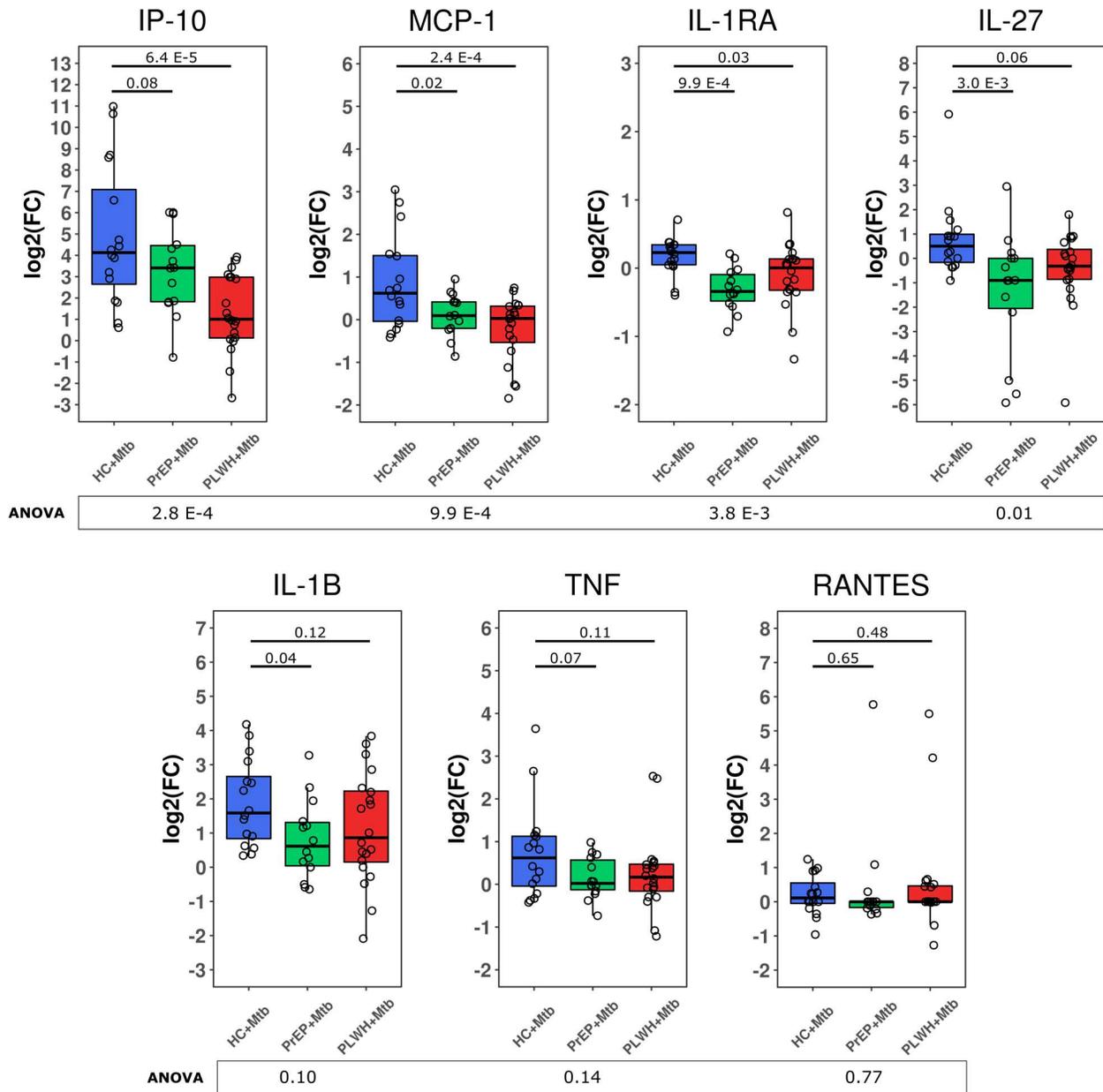
Supplemental Figure 1. mRNA profile in response to *Mtb*.

(A) Volcano plots for healthy controls (HC), subjects on pre-exposure prophylaxis (PrEP), and persons living with HIV receiving antiretroviral treatment (PLWH). Differentially expressed genes (DEG) in response to *Mtb* are displayed as function of \log_2FC (x-axis) and negative \log_{10} unadjusted *P*-value (y-axis). Each dot corresponds to a single gene, vertical dashed lines indicate the \log_2FC -0.2 and 0.2 thresholds. Red and blue dots represent genes significant at stageR FDR \leq 5% that were, respectively, up- or down-regulated. Grey dots are genes that did not meet the \log_2FC thresholds (inside dashed lines) or that were not significantly differentially expressed (below coloured dots). Genes identified by symbols were significant DE for the HC response to *Mtb*, were significant for the PLWH vs HC differential gene response (panel C, middle), were in significant GO-terms/pathway for the PLWH vs HC contrast, and had significant DOCs. These genes are prototypical examples for the diminished response by alveolar macrophages (AM) from PLWH to *Mtb*. (B) Scatterplots for \log_2FC from *Mtb* challenge effect. \log_2FC for two groups are plotted against each other with the group on the x-axis considered the reference group. Grey dots represent the union of DEG from both groups. Top panel coloured dots indicate DEGs for the y-axis group (counts for DEG y-axis group are shown). Pearson correlations (*R* value indicated) were calculated based on the coloured dots irrespective if the corresponding gene in the x-axis group reached significance for DEG. Bottom panels contain the same data and order as the top ones, however coloured dots now represent DEGs for the x-axis group. As above, correlations were calculated for coloured \log_2FC pairings. These plots illustrate the overall blunted transcriptional response to *Mtb* by AM from PLWH and PrEP subjects. (C) Volcano plots for *Mtb* response differences between groups (PrEP vs HC; PLWH vs HC; PLWH vs PrEP). Purple dots represent significant genes with diminished \log_2FC (lower response) for PLWH or PrEP in contrast to HC subjects. Yellow dots represent significant genes with increased \log_2FC (higher response) for the non-reference groups. Grey dots depict non-significant genes for the response differences presented. Annotated genes are the same as described in panel A. (D) Genes differentially triggered between groups. As in panel B, grey dots represent the union of DEG from both groups. Coloured dots indicate DEG detected by the interaction terms for the groups indicated on the x and y-axes. The counts of DEG for the group contrasts is indicated on top of each panel. The x-axis depicts the *Mtb* effect (\log_2FC) for the indicated reference group while the y-axis shows the \log_2FC of the *Mtb* effect for the contrasted group. The correlations show that the majority of PLWH vs HC and PrEP vs HC DE genes were more strongly induced in HC subjects.



Supplemental Figure 2. Scatterplots contrasting baseline mean log₂ expression between groups.

Baseline analysis was done for non-challenged libraries of phenotypic groups adjusting on the covariates: sex, age, smoking status (cigarette or cannabis), sequencing batch and date of BAL sampling. For these plots, depth normalized expression values were adjusted by regressing out the effect of covariates listed. Baseline gene expression values for the HC group are plotted on the x-axes while corresponding gene expression levels for the PLWH and PrEP groups are plotted on the y-axes. Grey dots represent the mean expression for all testable genes. Light blue dots depict 134 genes reported with lower expression induction in PLWH contrasted with HC after *Mtb* challenge (Figure 2A, first column). Likewise, dark blue dots represent 148 genes with reduced response to *Mtb* for PrEP in comparison to HC (Figure 2A, second column). Pearson correlations (R value indicated) were calculated based on the coloured dots. The plots indicate very similar baseline gene expression profiles of groups in absence of *Mtb* challenge.



Supplemental Figure 3. *Mtb*-triggered secretion of seven cytokines by AM cells.

After performing the challenge of AM with *Mtb* for RNA-seq, we kept the 20 hrs supernatants for all 100 samples (50 pairs of challenged and non-challenged). Estimates for levels of secreted cytokines were obtained with Milliplex Map (EMD Millipore, St. Charles, MO, USA) multiplex magnetic bead-based antibody detection kits and the MAGPIX platform (Luminex). Concentration values for each sample were converted to \log_2 scale and the pair-wise \log_2 fold change ($\log_2\text{FC}$) was derived for each sample. We assessed group differences using standard linear regression. Box-plots represent $\log_2\text{FC}$ for each subject on the y-axis, while x-axes display group assignments. ANOVA results for each cytokine are shown in the boxes below the plots.

Straight lines on top display the significance of *Mtb*-triggered cytokine secretion for PrEP or PLWH AM against AM from HC subjects. Blue, green and red boxes represent HC, PrEP and PLWH groups respectively.

Supplemental Tables

Supplemental Table 1. Summary statistics for study subjects.

| Covariates | HC (n=16) | PrEP (n=14) | PLWH (n=20) |
|---|---------------------|-----------------------|-----------------------|
| Sex (n, %) | | | |
| Male | 11 (69%) | 14 (100%) | 19 (95%) |
| Female | 5 (31%) | 0 | 1 (1%) |
| Age | | | |
| Mean (SD) | 40.56 (11.64) | 36.86 (12.18) | 52.85 (7.68) |
| Median | 40 | 36 | 56 |
| Range | 25 - 59 | 25 - 65 | 33 - 62 |
| IQR | 19.5 | 11.25 | 8.25 |
| Ethnicity | | | |
| Caucasian | 15 (94%) | 11 (79%) | 20 (100%) |
| Asian | 1 (6%) | 1 (7%) | 0 |
| Hispanic | 0 | 1 (7%) | 0 |
| North African | 0 | 1 (7%) | 0 |
| Smoker or recreational cannabis (n, %) | | | |
| Yes | 4 (25%) | 7 (50%) | 12 (60%) |
| No | 12 (75%) | 7 (50%) | 8 (40%) |
| CD4 count (cells/mm³) | | | |
| Mean (SD) | 936.8 (337.44) | 907 (303.55) | 690.9 (274.57) |
| Median | 947 | 904 | 619.5 |
| Range | 439 - 1536 | 208 - 1454 | 269 - 1267 |
| IQR | 593 | 231.75 | 431.75 |

Supplemental Table 2, entitled “Differential gene expression testing results”, is available as a separated excel file.

Supplemental Table 3. Differential gene expression and term enrichment test summary.

| Test | Contrast | DEGs | # Genes FDR ≤ 0.05 | % Genes in combined-terms ^A | # combined-terms FDR ≤ 0.05 ^B | KEGG | | Reactome | | GO BP | |
|------------------------------------|---------------------------|------------------------------|-----------------------|---|---|--------|--------|----------|--------|--------|--------|
| | | | | | | FDR 5% | Tested | FDR 5% | Tested | FDR 5% | Tested |
| <i>Mtb</i> challenge effect | HC + <i>Mtb</i> vs HC | Combined | 1434 | 76.56 | 1445 | 77 | 219 | 209 | 477 | 1159 | 2885 |
| | | Up-regulated | 942 | 81.1 | 1452 | 83 | 168 | 223 | 373 | 1146 | 2361 |
| | | Down-regulated | 492 | 67.88 | 90 | 7 | 48 | 7 | 68 | 76 | 989 |
| | PLWH + <i>Mtb</i> vs PLWH | Combined | 679 | 74.22 | 845 | 59 | 132 | 82 | 191 | 704 | 1776 |
| | | Up-regulated | 461 | 80.91 | 851 | 57 | 99 | 64 | 132 | 730 | 1446 |
| | | Down-regulated | 218 | 60.09 | 16 | 0 | 8 | 2 | 20 | 14 | 390 |
| | PrEP + <i>Mtb</i> vs PrEP | Combined | 255 | 68.23 | 254 | 25 | 32 | 17 | 48 | 212 | 623 |
| | | Up-regulated | 203 | 72.41 | 286 | 26 | 29 | 17 | 32 | 243 | 515 |
| | | Down-regulated | 52 | 51.92 | 17 | 1 | 1 | 1 | 1 | 15 | 19 |
| <i>Mtb</i> response differences | PLWH vs HC | Combined | 156 | 71.79 | 262 | 29 | 33 | 28 | 36 | 205 | 355 |
| | | Higher response ^C | 22 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lower response ^D | 134 | 75.37 | 261 | 31 | 33 | 28 | 36 | 205 | 355 |
| | PrEP vs HC | Combined | 205 | 70.24 | 373 | 30 | 41 | 30 | 41 | 313 | 573 |
| | | Higher response | 57 | 56.14 | 6 | 0 | 0 | 0 | 0 | 6 | 6 |
| | | Lower response | 148 | 71 | 363 | 29 | 34 | 30 | 34 | 304 | 434 |
| | PLWH vs PrEP | Combined | 40 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Higher response | 23 | 26.08 | 5 | 0 | 0 | 0 | 0 | 5 | 5 |
| | | Lower response | 17 | 23.52 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

^A To derive gene percentage in significant GO terms/pathways, we considered the results from "Combined DEGs" enrichment testing.

^B Only terms with at least 5 assigned genes were considered.

^{C,D} Refers to positive and negative log₂FC results, respectively, from each interaction test.

Supplemental Table 4, entitled “Differential gene expression results subset for all genes significant in interaction terms”, is available as a separated excel file.

Supplemental Table 5. Listing of 40 differentially expressed genes for the contrasts *Mtb* challenge response in PLWH vs PrEP.

| gene ID | log2FC_HC.Mtb | log2FC_PLWH.Mtb | log2FC_PrEP.Mtb | log2FC_PLWH.x.PrEP.Mtb | stgr.pval_PLWH.x.PrEP.Mtb |
|--------------------------|---------------|-----------------|-----------------|------------------------|---------------------------|
| ENSG00000151065_DCP1B | -0.0297385 | -0.2131193 | 0.105109 | -0.3182283 | 0.000746622 |
| ENSG00000090659_CD209 | 0.399909624 | -0.055992475 | 0.636002069 | -0.691994543 | 0.001275205 |
| ENSG00000169914_OTUD3 | 0.100345334 | 0.383729024 | -0.366055307 | 0.74978433 | 0.001604151 |
| ENSG00000134028_ADAMDEC1 | -0.08724319 | -0.407451817 | 0.013367624 | -0.420819442 | 0.002051195 |
| ENSG00000099814_CEP170B | 0.161882237 | 0.296493444 | -0.034829235 | 0.331322679 | 0.00250621 |
| ENSG00000165474_GJB2 | 0.11323178 | 0.492506382 | -0.040968291 | 0.533474674 | 0.002978432 |
| ENSG00000181409_AATK | -0.250470465 | 0.262928245 | -0.388665432 | 0.651593677 | 0.003842469 |
| ENSG00000164970_FAM219A | 0.150372282 | 0.252021393 | 0.00563563 | 0.246385762 | 0.004020677 |
| ENSG00000080573_COL5A3 | 0.054225795 | 0.361640229 | -0.109644228 | 0.471284457 | 0.004116781 |
| ENSG00000100575_TIMM9 | -0.078791616 | -0.23261684 | 0.190858244 | -0.423475084 | 0.004161273 |
| ENSG00000169372_CRADD | 0.090868434 | 0.312550005 | -0.000531985 | 0.31308199 | 0.006435933 |
| ENSG00000109911_ELP4 | -0.087075207 | -0.372950471 | -0.006755399 | -0.366195072 | 0.008842803 |
| ENSG00000198853_RUSC2 | 0.148752775 | 0.219624982 | -0.01036035 | 0.229985332 | 0.010720881 |
| ENSG00000276085_CCL3L3 | 1.349552792 | 1.838615114 | 0.568556344 | 1.270058769 | 0.010784191 |
| ENSG00000116017_ARID3A | 0.001652497 | 0.2709595 | -0.08700081 | 0.3579603 | 0.0121509 |
| ENSG00000169908_TM4SF1 | 0.522607122 | 0.784147053 | 0.112638217 | 0.671508836 | 0.01314304 |
| ENSG00000130749_ZC3H4 | 0.124040059 | -0.074926932 | 0.201265596 | -0.276192528 | 0.013997389 |
| ENSG00000213762_ZNF134 | -0.165176365 | -0.25222791 | 0.114434031 | -0.366661941 | 0.014541229 |
| ENSG00000153395_LPCAT1 | -0.117304166 | 0.032786337 | -0.20719563 | 0.239981967 | 0.016228528 |
| ENSG00000155016_CYP2U1 | -0.364015213 | -0.469287971 | -0.017903114 | -0.451384857 | 0.020733113 |
| ENSG00000125657_TNFSF9 | 0.167324937 | 0.464526467 | 0.118153763 | 0.346372704 | 0.020769268 |
| ENSG00000258227_CLEC5A | -0.43931682 | -0.720769374 | -0.296489123 | -0.424280252 | 0.020852452 |
| ENSG00000204472_AIF1 | -0.152933816 | -0.26461196 | -0.003301537 | -0.261310423 | 0.023518391 |
| ENSG00000135525_MAP7 | 0.17936453 | 0.398728233 | 0.052592309 | 0.346135924 | 0.025665002 |
| ENSG00000174945_AMZ1 | -0.081063766 | 0.272892048 | -0.025987325 | 0.298879373 | 0.026709912 |
| ENSG00000178385_PLEKHM3 | 0.231095435 | 0.223904251 | -0.059670061 | 0.283574311 | 0.031884813 |
| ENSG00000175592_FOSL1 | 0.371820009 | 0.61401958 | 0.186402862 | 0.427616718 | 0.032638114 |
| ENSG00000147650_LRP12 | 0.19993291 | 0.278175504 | -0.06188825 | 0.340063754 | 0.033767647 |
| ENSG00000136044_APPL2 | -0.155729182 | -0.27661718 | 0.025124669 | -0.301741849 | 0.034031608 |
| ENSG00000096070_BRPF3 | -0.002116972 | 0.2006202 | -0.01217041 | 0.2127906 | 0.03444748 |
| ENSG00000143382_ADAMTSL4 | 0.164140239 | -0.039382159 | 0.235020213 | -0.274402372 | 0.035210832 |
| ENSG00000124882_EREG | 0.461644342 | 0.752636844 | 0.058858299 | 0.693778545 | 0.038062197 |
| ENSG00000149591_TAGLN | 0.253166034 | -0.006543408 | 0.245722247 | -0.252265654 | 0.039225944 |
| ENSG00000140694_PARN | -0.043986145 | -0.263740111 | -0.014815284 | -0.248924827 | 0.039320844 |
| ENSG00000139725_RHOF | 0.034960015 | 0.249947015 | 0.015588476 | 0.234358538 | 0.041604923 |
| ENSG00000255833_TIFAB | -0.745451313 | -0.646911336 | 0.210915331 | -0.857826667 | 0.041924324 |
| ENSG00000111145_ELK3 | -0.223104036 | -0.2777801 | 0.032460383 | -0.310240483 | 0.043315909 |
| ENSG00000132002_DNAJB1 | 0.170473231 | 0.245512063 | -0.021413914 | 0.266925977 | 0.044270801 |
| ENSG00000182134_TDRKH | -0.283150058 | -0.260641339 | 0.028269881 | -0.28891122 | 0.048552949 |
| ENSG00000222009_BTBD19 | 0.168243536 | 0.267965297 | 0.020564661 | 0.247400636 | 0.048664156 |

log2FC_{contrast} represents the average log2 fold change for a contrast at hand. For instance, “log2FC_PLWH.x.PrEP.Mtb” is the average response difference between PLWH *Mtb* response against PrEP *Mtb* response (interaction test).

stgr.pval_{contrast} is the stageR adjusted-p values (FDR).

Supplemental Table 6. ATAC-seq and H3K27ac differential accessibility/mark and term enrichment test summary.

| Approach | Contrast | Peaks | # Regions FDR < 0.05 | # genes | KEGG | | Reactome | | GO BP | |
|----------|---------------------------|-----------------|-------------------------|---------|--------|--------|----------|--------|--------|--------|
| | | | | | FDR 5% | Tested | FDR 5% | Tested | FDR 5% | Tested |
| ATACseq | HC + <i>Mtb</i> vs HC | Combined | 12369 | 16085 | 63 | 323 | 41 | 1395 | 620 | 6348 |
| | | Open | 8389 | 11562 | 42 | 316 | 42 | 1338 | 461 | 6001 |
| | | Close | 3971 | 6430 | 3 | 309 | 5 | 1047 | 23 | 4549 |
| H3K27ac | HC + <i>Mtb</i> vs HC | More acetylated | 1293 | 5046 | 51 | 294 | 35 | 961 | 532 | 4367 |
| | | Less acetylated | 1218 | 3957 | 0 | 303 | 2 | 1021 | 39 | 4655 |
| | PLWH + <i>Mtb</i> vs PLWH | More acetylated | 389 | 1735 | 10 | 294 | 25 | 961 | 84 | 4367 |
| | | Less acetylated | 56 | 238 | . | . | . | . | . | . |
| | PrEP + <i>Mtb</i> vs PrEP | More acetylated | 98 | 497 | 8 | 294 | 15 | 961 | 134 | 4367 |
| | | Less acetylated | 0 | 0 | . | . | . | . | . | . |

Supplemental Table 7, entitled “Differential chromatin accessibility testing results”, is available as a separated excel file.

Supplemental Table 8, entitled “Differential H3K27ac mark testing results”, is available as a separated excel file.

Supplemental Table 9. Enrichment of IRF9 and ZNF684 footprints in the promoter of DEGs.

| | | DEG in HC + <i>Mtb</i> | | Total | |
|---|-----|------------------------|------|-------|---------------------------|
| | | Yes | No | | |
| IRF9 active footprint in the gene TSS | Yes | 65 | 70 | 135 | $p = 7.9 \times 10^{-23}$ |
| | No | 1369 | 8858 | 10227 | |
| ZNF684 active footprint in the gene TSS | Yes | 258 | 1047 | 1305 | $p = 5.3 \times 10^{-11}$ |
| | No | 1176 | 7881 | 9057 | |
| Total | | 1434 | 8928 | 10362 | |

Supplemental Table 10, entitled “Expressed genes with IRF9 footprint at the transcription starting site”, is available as a separated excel file.

Supplemental References

1. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
2. Wang L, et al. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184-5.
3. Martin M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnetjournal*. 2011.
4. Zerbino DR, et al. Ensembl 2018. *Nucleic acids research*. 2018;46(D1):D754-d61.
5. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
6. Patro R, et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-9.
7. R Core Team (2018). Vienna, Austria.: R Foundation for Statistical Computing.; 2018.
8. Sonesson C, et al. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521.
9. Durinck S, et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols*. 2009;4(8):1184-91.
10. Love MI, et al. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research*. 2018;7:952.
11. Robinson MD, et al. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40.
12. Robinson MD, and Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11(3):R25.
13. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015;43(7):e47.
14. Law CW, et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.
15. Li H, and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
16. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
17. Amemiya HM, et al. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*. 2019;9(1):9354.
18. Ramirez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*. 2016;44(W1):W160-5.
19. Ewels P, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-8.
20. Yu G, et al. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31(14):2382-3.
21. Fishilevich S, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017;2017.
22. Lawrence M, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118.
23. Corces MR, et al. The chromatin accessibility landscape of primary human cancers. *Science*. 2018;362(6413).

24. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-2.
25. Liao Y, et al. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30.
26. Nedelec Y, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016;167(3):657-69 e21.
27. Leek JT, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-3.
28. Van den Berge K, et al. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biol*. 2017;18(1):151.
29. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-7.
30. Ramirez F, et al. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research*. 2014;42(Web Server issue):W187-91.
31. Kurtenbach S, and William Harbour J. SparK: A Publication-quality NGS Visualization Tool. *bioRxiv*. 2019:845529.
32. Wang B, et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nature protocols*. 2019;14(3):756-80.
33. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated; 2009.
34. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-89.
35. Li Z, et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol*. 2019;20(1):45.