Open access • Posted Content • DOI:10.1101/2021.07.30.454441

A new paradigm for leprosy diagnosis based on host gene expression Insights from leprosy lesions transcriptomics — Source link

Thyago Leal-Calvo, Charlotte Avanzi, Mayara Abud Mendes, Andrej Benjak ...+6 more authors

Institutions: Oswaldo Cruz Foundation, École Polytechnique Fédérale de Lausanne, Pasteur Institute

Published on: 30 Jul 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Leprosy

A new paradigm for leprosy diagnosis based on host gene expression

Insights from leprosy lesions transcriptomics

Thyago Leal-Calvo[1], Charlotte Avanzi[2,#a], Mayara Abud Mendes[1], Andrej Benjak[2,#b], Philippe Busso[2], Roberta Olmo Pinheiro[1], Euzenir Nunes Sarno[1], Stewart T. Cole[2,3], Milton O. Moraes[1*]

Affiliations

[1] Laboratório de Hanseníase, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil

[2] Global Health Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[3] Institut Pasteur, Paris, France

[#a]Current address: Department of Microbiology, Immunology and Pathology, Mycobacteria Research Laboratories, Colorado State University, Fort Collins, Colorado, United States of America

[#b]Current address: Department for BioMedical Research, Oncogenomics Laboratory, University of Bern, Bern, Switzerland

* Corresponding author

E-mail: milton.moraes@fiocruz.br (MOM)

Abstract

Transcriptional profiling is a powerful tool to investigate and detect human diseases. In this study, we used bulk RNA-sequencing (RNA-Seq) to compare the transcriptomes in skin lesions of leprosy patients or matched controls affected by other dermal conditions such as granuloma annulare, a confounder for paucibacillary leprosy. We identified five genes capable of accurately distinguishing multibacillary and paucibacillary leprosy from other skin conditions. Indoleamine 2,3-dioxygenase 1 (*IDO1*) expression alone was highly discriminatory, followed by *TLR10*, *BLK*, *CD38*, and *SLAMF7*, whereas the *HS3ST2* and *CD40LG* mRNA separated multi- and paucibacillary leprosy. Finally, from the main differentially expressed genes (DEG) and enriched pathways, we conclude that paucibacillary disease is characterized by epithelioid transformation and granuloma formation, with an exacerbated cellular immune response, while multibacillary leprosy features epithelial-mesenchymal transition with phagocytic and lipid biogenesis patterns in the skin. These findings will help catalyze the development of better diagnostic tools and potential host-based therapeutic interventions. Finally, our data may help elucidate host-pathogen interplay driving disease clinical manifestations.

Author Summary

Despite effective treatment, leprosy is still a significant public health issue in more than 120 countries, with more than 200 000 new cases yearly. The disease is caused mainly by *Mycobacterium leprae*, a slow-growing bacillus still uncultivable in axenic media. This limitation has hampered basic research into host-pathogen

interaction and the development of new diagnostic assays. Currently, leprosy is diagnosed clinically, with no standalone diagnostic assay accurate enough for all clinical forms. Here, we use RNA-seq transcriptome profiling in leprosy lesions and granuloma annulare to identify mRNA biomarkers with potential diagnostic applications. Also, we explored new pathways that can be useful in further understanding the host-pathogen interaction and how the bacteria bypass host immune defenses. We found that *IDO1*, a gene involved with tryptophan catabolism, is an excellent candidate for distinguishing leprosy lesions from other dermatoses. Additionally, we observed that a previous signature of keratinocyte development and cornification negatively correlates with epithelial-mesenchymal transition genes in the skin, suggesting new ways in which the pathogen may subvert its host to survive and spread throughout the body. Our study identifies new mRNA biomarkers that can improve leprosy diagnostics and describe new insights about host-pathogen interactions in human skin.

Introduction

Leprosy is a chronic infectious disease caused mainly by the slow-growing intracellular pathogen *Mycobacterium leprae* that does not grow in axenic media. This bacterium resides preferentially in skin macrophages and Schwann cells in peripheral nerves, inducing dermatosis and/or neuritis. Patients can present several distinct clinical forms according to their immune response, histopathological characterization, and bacterial load. A localized tuberculoid form (TT) is characterized by low bacterial counts and a strong cellular immune response. Conversely, in the opposite lepromatous (LL) pole, a disseminated form, patients exhibit several lesions, a predominantly humoral response, and a high bacterial load in the tissues [1–3]. Borderline forms are classified according to their proximity to the poles. For operational and treatment purposes, leprosy is classified by the World Health Organization as paucibacillary (PB) or multibacillary (MB), based on the number of skin lesions, associated with nerve involvement or the bacilli detection in slit-skin smears [4].

Early and precise diagnosis is instrumental to leprosy control since delay in diagnosis leads to late multidrug therapy, higher disability risk, and continuing transmission, as highlighted by the 200,000 new cases consistently reported annually in the last 10 years [4,5]. However, bacteriological, immunological, genetics or molecular methods are not sufficient for specific diagnosis when used alone. Diagnosis most commonly relies on clinical evaluation, occasionally complemented with histopathological examination and bacterial counts, but these procedures are mostly performed in national reference centers [4,6].

Efforts have been deployed to improve leprosy diagnostics using cutting-edge technologies, such as molecular identification of *M. leprae*, serological tests for specific bacterial antigens, and quantification of host biomarkers in plasma or *in vitro* whole blood assays (WBA) [7–9]. Overall, all methods outperform standard clinical diagnosis and can compensate for the low accuracy in detecting PB patients [4,7,8,10–14]. Yet, until now such investigations involved comparing confirmed leprosy cases against healthy endemic controls, who are not representative of individuals with suspected leprosy. Here, other skin conditions represent a better comparator.

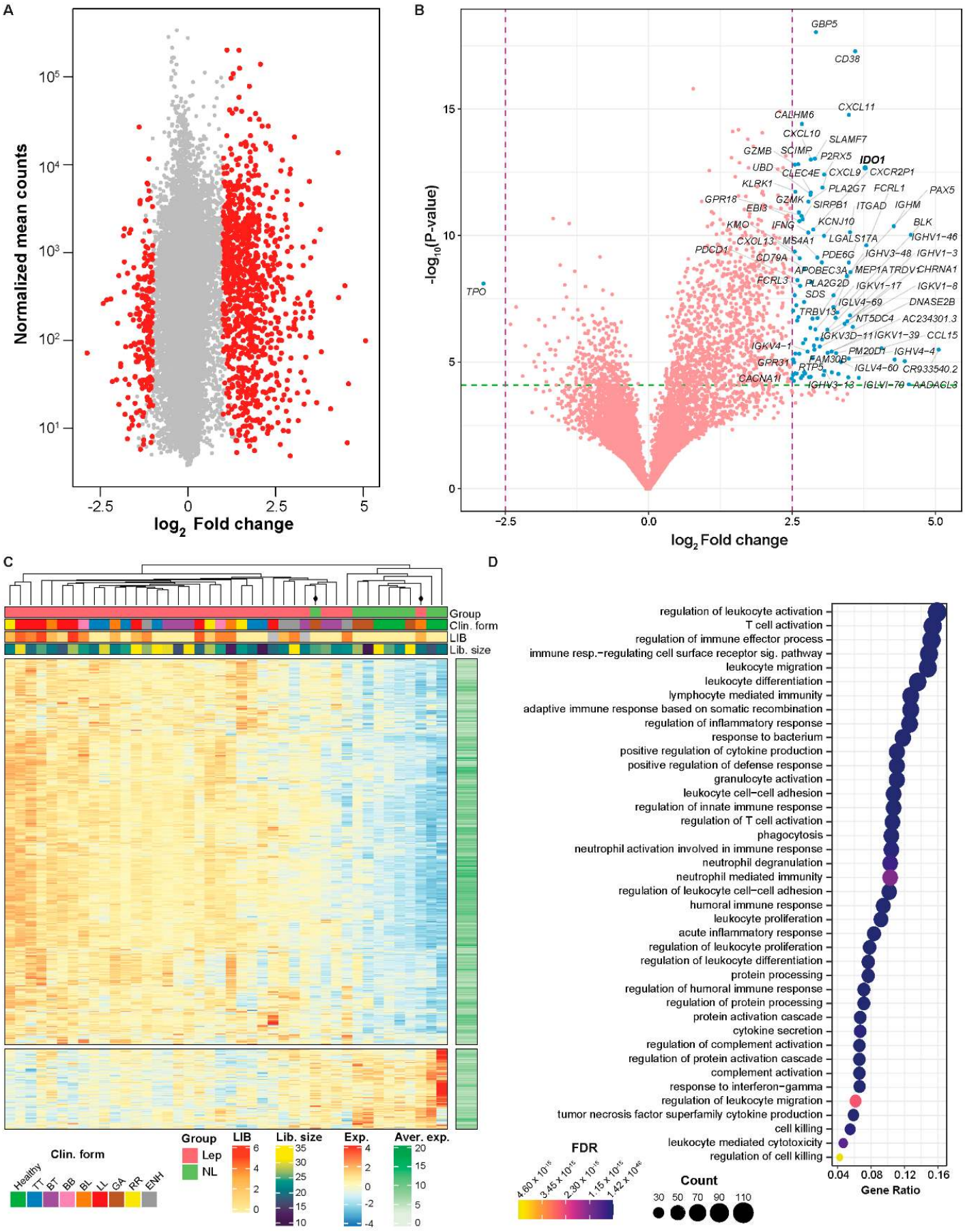
Identification of markers for early infection is hindered by our poor understanding of pathogenicity and the mechanism by which patients develop one or the other form of leprosy, and nerve injuries [15]. Gene expression signatures have been used as diagnostic tools for several illnesses, from infectious [10–12,14] and autoimmune diseases [16,17] to cancer [18–20]. Some signatures have already been approved for clinical use [12,21–23]. In leprosy, findings from past studies indicate the great potential of expression profiling for disease diagnosis [24–27]. Nonetheless, they were limited by the number of patients [28], or lacked proper epidemiological controls, such as differential diagnosis groups.

Here, we applied a combination of bulk RNA sequencing and quantitative validation by RT-qPCR on RNA extracted from skin biopsies of various leprosy forms and from non-leprosy patients to define a specific leprosy host signature applicable to diagnosis. Then, we explored gene expression patterns to improve our understanding of the immunopathogenic mechanisms towards leprosy polarization.

Results

Discrimination of leprosy *vs.* non-leprosy lesions based on mRNA expression

RNA sequencing was used for pinpointing host candidate genes capable of differentiating leprosy lesions from one of the commonest differential diagnoses of leprosy, granuloma annulare (GA), and from healthy skin. RNA from skin lesions of all leprosy clinical forms (n=33), plus GA (n=4) and healthy skin (n=5) were sequenced (S1 Table). Differentially expressed genes (DEG) in leprosy *vs.* non-leprosy (GA + healthy skin) samples resulted in 1160 DEG with a $|\log_2 FC| \geq 1$ and FDR ≤ 0.01, with 961 upregulated in leprosy forms compared to non-leprosy (Fig 1A-B and S2 Table). Exploratory hierarchical clustering of the DEG with $|\log_2 FC| \geq 1$ and FDR < 0.01 grouped all patients' samples into roughly two clusters, except for two: one BL leprosy and one GA that clustered apart from samples with the same diagnosis (Fig 1C). Gene Ontology enrichment analysis of up-regulated genes in leprosy compared to non-leprosy showed enrichment for biological processes associated with leukocyte activation, T-cell activation, immune response, response to the bacterium, neutrophil degranulation, cell killing, cytokine secretion, purinergic receptor signaling pathway, and regulation of defense response to viruses by the host (Fig 1D and S3 Table).



123 **Fig 1. Differentially expressed genes from RNA-seq in leprosy *vs.* GA and**

leprosy *vs.* non-leprosy. (A) MA plots showing the number of differentially expressed genes (red points) with FDR < 0.1 for the comparison of leprosy *vs.* non-leprosy. (B) Volcano plot depicting DEG from leprosy vs. non-leprosy, where violet dashed line marks $|\log_2 FC| = 2.5$ and green horizontal dashed line the FDR = 0.01. (C) Heatmap with hierarchical clustering of samples based on expression of the DEG from leprosy vs. non-leprosy comparison. Color scale ranges from lower expression (blue) to higher expression (red). Library size is given in millions. LIB, logarithmic index of bacilli. (D) Biological processes from GO enriched for up-regulated DEG from leprosy vs. non-leprosy comparison. FDR, false discovery rate; NL, non-leprosy; GA, granuloma annulare; non-leprosy: GA + healthy individuals.

A total of 15 genes with the largest effect size ($|\log_2 FC| \geq 1.5$, FDR < 0.001), highest area under the curve (AUC), and plausible involvement with leprosy pathogenesis (S4 Table) were then validated using a two-step RT-qPCR with a new, larger, and more heterogeneous dataset including skin lesion samples from leprosy patients (n=24), and other common dermatoses (n=30) (S1 Table). Other dermatological diseases (ODD) included dermatitis (n=11), eczema (n=1), erythema (n=5), GA (n=6), lichen planus (n=2), psoriasis (n=2) and pityriasis alba (n=2) (S1 Table). A total of 12 samples per group was estimated to be sufficient to attain a power of 85% based on the Welch t-test (PB *vs.* ODD, MB *vs.* ODD) with alpha set at 0.03 to replicate the standardized effect size ($\log_2 FC/SD$) estimated from RNA sequencing. Relative expression using the new sample set by RT-qPCR is shown in Fig 2A. Indeed, the validation data are in agreement with RNA sequencing, because most tested

147 genes were replicated by RT-qPCR in differentiating leprosy from other dermatoses

148 (Fig 2A) and by their mean differences (Fig 2B-C, S5 Table).

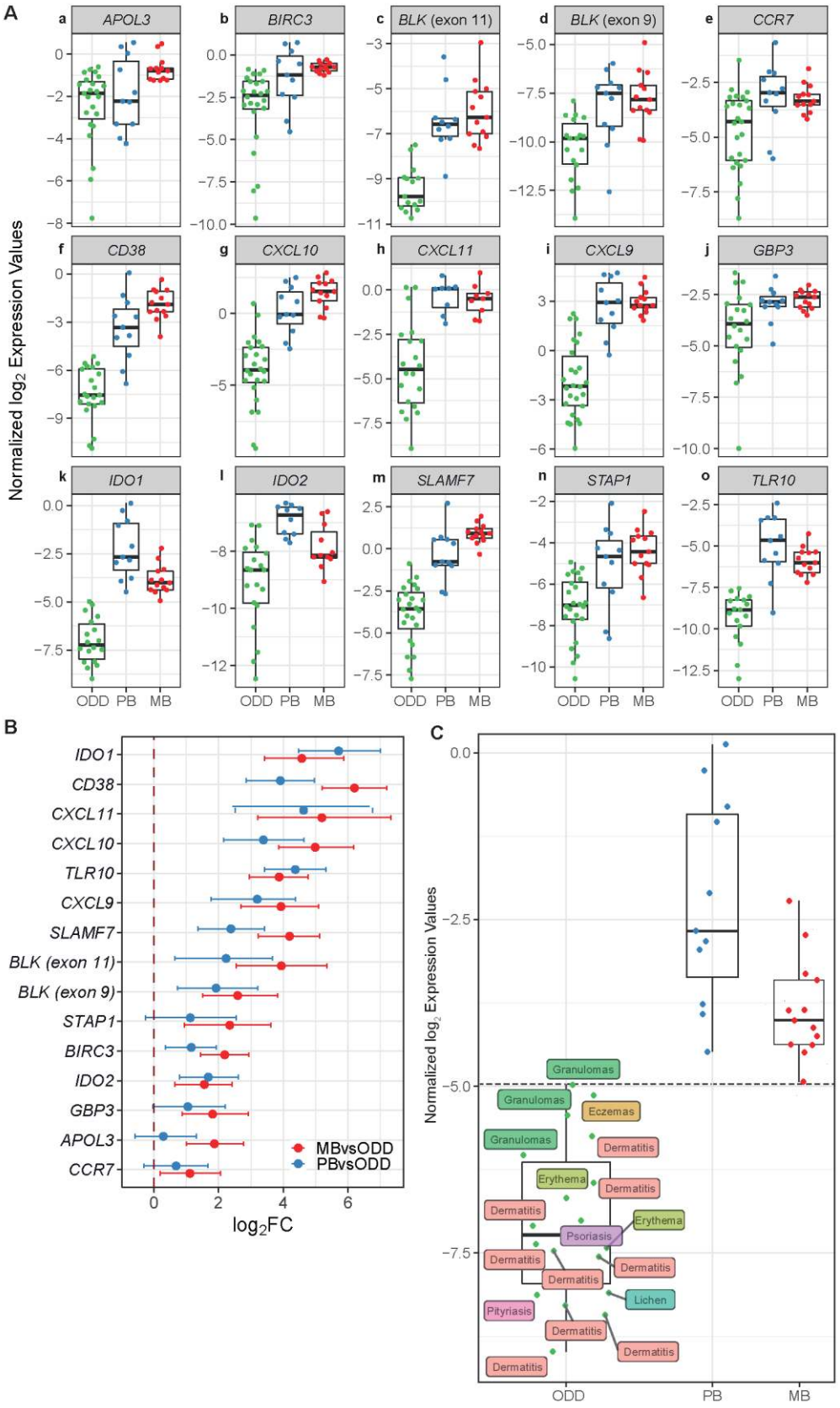


Fig 2. Technical and biological validation for selected DEG discovered from RNA sequencing. (A) Tukey boxplots with RT-qPCR normalized (2-3 reference genes) log_2 expression values (A.U) according to clinical and histopathological diagnosis. (B) log_2FC from MB-ODD and PB-ODD comparisons estimated from Bayesian linear mixed models and their 95% credible intervals. (C) Tukey boxplot highlighting *IDO1* RT-qPCR normalized log_2 expression values by final diagnosis grouped into ODD category. Missing values are omitted.

Next, hierarchical clustering with RT-qPCR data including missing values for some genes (no target gene amplification by RT-qPCR) was performed to examine all samples simultaneously. The analysis roughly revealed three major clusters (Fig 3A). At the highest tree subdivision, one small cluster (n=6) with the dendrogram grouped in light brown was composed of ODD samples with lower expression levels (Fig 3A). Due to several ODD having missing values, we confirmed that these samples had similar gene expression for the reference genes, thereby eliminating the possibility of insufficient cDNA input. Another cluster, grouped in the light purple dendrogram, included all MB and most PB samples (except three in light yellow dendrogram). GA samples displayed two patterns, the first with two samples showing undetectable *IDO1* expression (Fig 3A, bottom star symbols). The second set (n=4) is scattered among other ODD samples (Fig 3A). It can be seen that GA and PB samples show highly similar expression profiles for some genes (Fig 3A bottom diamond symbols), reinforcing the difficulty in clinically discriminating between these two conditions, and underlining the relevance of their inclusion in our comparisons [29–31].

Then, by applying principal component analysis (PCA) to the 15 gene signature obtained with the expanded sample panel tested by RT-qPCR, we uncovered two major patterns separating leprosy lesions from ODD (Fig 3B). As expected, MB samples appeared more homogeneous than PB and ODD samples, while the latter were more dispersed revealing heterogeneous expression patterns (Fig 3B).

Next, we quantified the individual classification potential of these genes in distinguishing leprosy from ODD using ROC analysis on RT-qPCR data. *IDO1* expression alone was found to be 100% accurate using an arbitrary threshold, followed by *BLK* (exon 11), *TLR10, CD38,* and *SLAMF7* (Fig 3C and S6 Table). Finally, to confirm the causal link between mycobacteria and our gene-set, we evaluated the mRNA profiles induced by other live-mycobacteria using a public RNA-seq dataset [32]. We observed that most gene expression signatures, including *IDO1,* could be successfully replicated as induced by either *M. leprae* and/or other mycobacteria (Fig 1 in Appendix S1 and S7 Table). By contrast, some of the tested genes such as *BLK*, *CXCL9*, *MS4A1,* and *TLR10* were not differentially expressed in any of the *in vitro* assays with mycobacteria (Fig 1 in Appendix S1 and S7 Table).

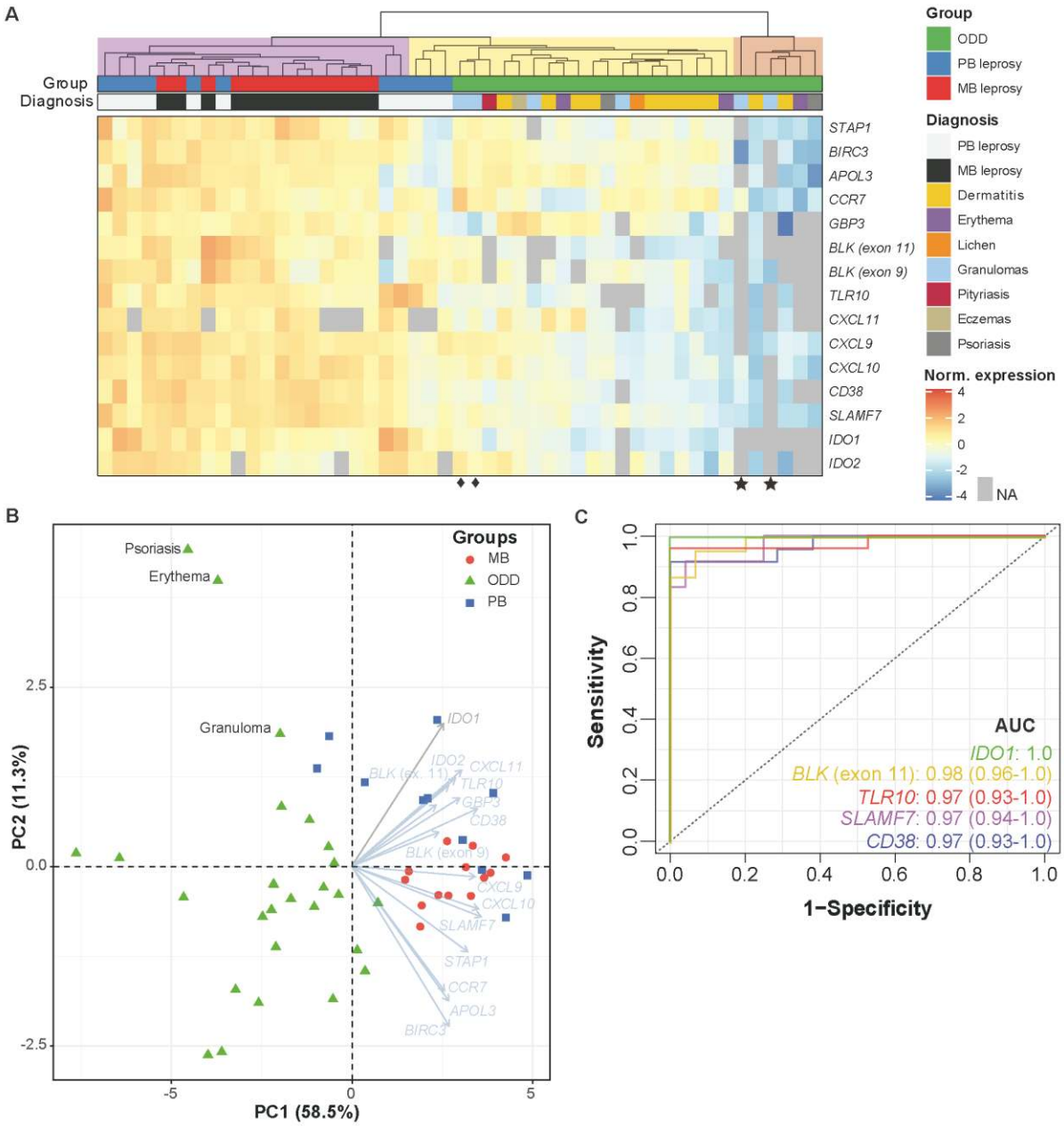


Fig 3. Hierarchical clustering of RT-qPCR replicated DEG and ROC analysis. (A) Hierarchical clustering with scaled and centered normalized log_2 RT-qPCR expression values (arbitrary units) and annotated according to group and specific diagnosis. Dendrogram tree was cut arbitrarily and cluster analysis is for hypothesis generating purposes only. (B) Principal component analysis (PCA) with 15 genes measured by RT-qPCR and using log_2 normalized scaled data. For PCA only, missing values were imputed by the gene arithmetic mean. NA, not amplified, i.e., Cp > 40. In this regard,

197 there were two outliers (psoriasis and erythema), which are samples with high
198 numbers of NA values and that were imputed using the gene arithmetic mean. (C)
199 Receiver operating characteristic analysis for genes with largest AUC (95%
200 confidence intervals) from RT-qPCR replication samples (complete data are shown in
201 S6 Table). See also S1 Appendix and S1 Fig.

202 MB and PB gene expression profiling and mRNA-based
203 classifier

204 To define a small subset of genes with high classificatory potential (i.e. with
205 non-overlapping expression values) to distinguish MB from PB lesions, we performed
206 a penalized logistic regression (LASSO) model with k-fold cross-validation trained on
207 the public microarray dataset [24]. This dataset was chosen because of the higher
208 number of PB/MB samples compared to our RNA-seq dataset. As a result, three genes
209 with non-zero coefficients were selected by the cross-validated LASSO model:
210 *HS3ST2, CD40LG,* and *CCR6*, but only the first two genes were most frequently
211 (~80%) selected across 10,000 bootstrapped samples within the training dataset (Fig
212 4A-B). The median misclassification error estimated by the resampling was about 4%
213 (±5.4% median absolute deviation), ranging from 0% to 32% (Fig 4C). Instability
214 assessment in the number of selected genes by LASSO (Fig 4D) showed that most
215 iterations resulted in four non-zero genes (range, 1-20). The final model containing the
216 three genes (*HS3ST2*, *CD40LG,* and *CCR6*) was evaluated on two test RNA-seq
217 datasets: our dataset, and the one from Montoya *et al.* including MB (n=9) and PB
218 (n=6) groups [28]. Penalized logistic regression demonstrated an accuracy of 100%
219 (lower 95% CIs: 86.8% and 78.2%, respectively) in classifying MB from PB samples

220 in both test RNA-seq datasets (Fig 4E-F). The *HS3ST2* gene was consistently more
221 expressed in MB leprosy lesions compared to PB, whereas the opposite was observed
222 for *CD40LG* (Fig 4E-H) and *CCR6* (S2 Fig). In both datasets, the combined expression
223 levels of *HS3ST2* and *CD40LG* showed good discrimination between the two groups
224 (Fig 4E-H). However, given the sample size and the bootstrapped estimates, it is not
225 currently possible to exclude *CCR6* from the model without additional replication.

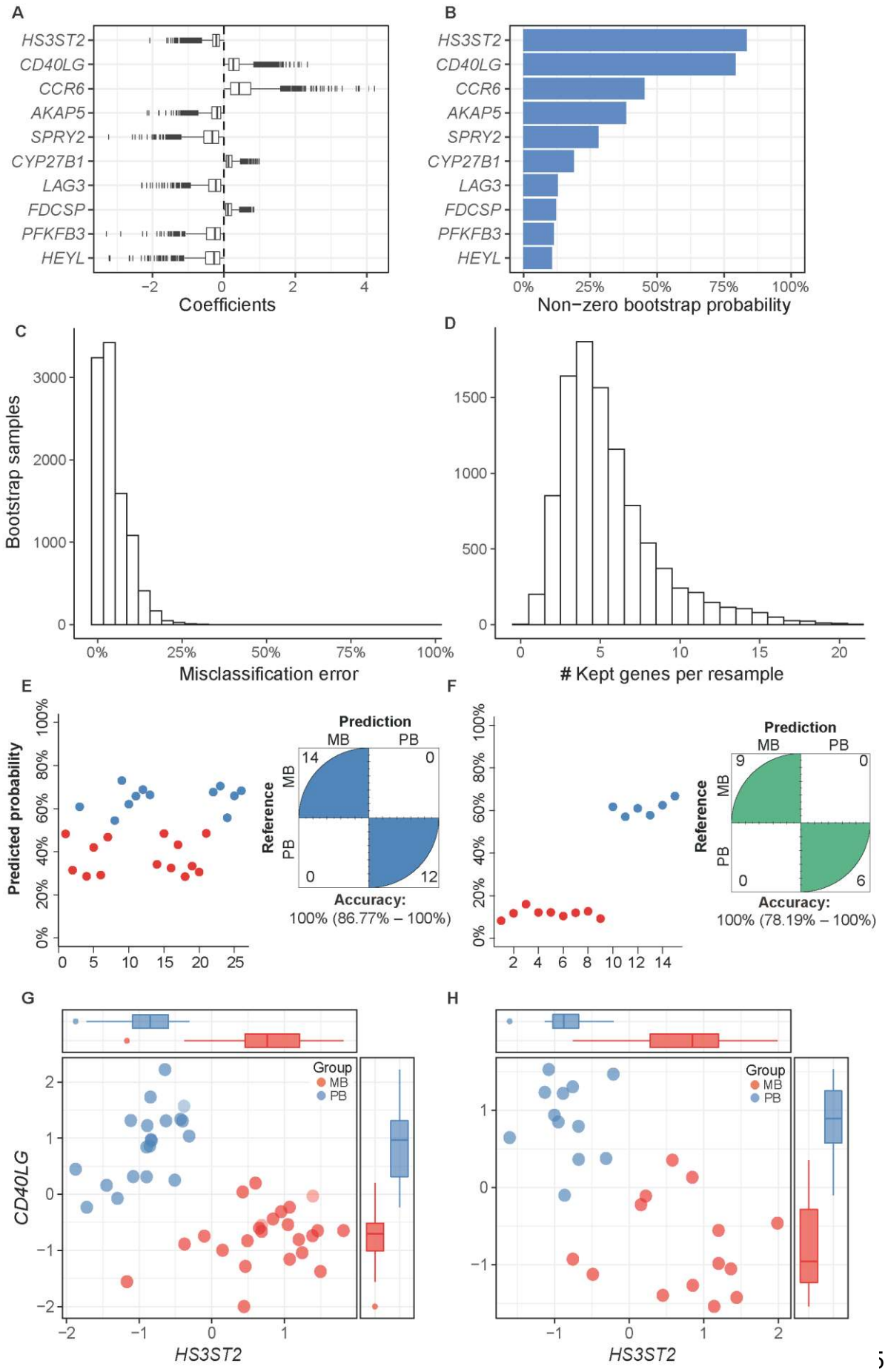


Fig 4. Gene candidates identified with the penalized logistic regression (LASSO) model as the most important to distinguish PB and MB leprosy lesions. (A) Coefficients (log odds) from the top 10 most selected genes (i.e., non-zero) across 10,000 bootstrap samples using the microarray from Belone *et al.* as training dataset. (B) Frequency of non-zero coefficients across all bootstrap samples. (C) Misclassification error distribution estimated from 4-fold cross-validation (k-) across 10,000 bootstrap samples, with median error of 3.70% (±5.4% median absolute deviation). (D) Number of genes kept across all resamples. Predicted probability from the final model performance on this study test RNA-seq (E) and Montoya *et al.* RNA-seq. (F) Normalized $\log_2$ gene expression (z-score) of the two most frequently selected variables for distinguishing MB from PB samples in the (G) microarray training dataset and (H) this study test RNA-seq. PB, paucibacillary leprosy; MB, multibacillary leprosy. Tukey box plots with 1st, 2nd and 3rd quartiles ± 1.5 × inter quartile range (IQR) whiskers. See also S2 Fig.

Next, to assess the dichotomy beyond cellular *vs.* humoral response in leprosy lesions [33,34], a comparison of gene expression in MB leprosy (LL+BL+BB) *vs.* PB (TT+BT) skin lesions was performed. Differential expression analysis with $|\log_2 FC| \geq 1$ and FDR ≤ 0.01 resulted in 112 DEGs; 69 up-regulated and 43 down-regulated (Fig 5A and S8 Table). In addition, we compared DEG to the public microarray data available in Gene Expression Omnibus (GEO) from Belone *et al.* [24,35] using only the FDR cutoff. With an FDR < 0.01, 161 DEGs were common to both studies, all except one showed concordant modulation characterized by an overall high correlation coefficient and concordance index, irrespective of the technology used, the

sample processing, and the data analysis methods (Fig 5B). Functional enrichment analysis of the RNA-seq up-regulated genes (i.e., more expressed in MB than PB) revealed processes involved with regulation of immune response, humoral immunity, phagocytosis, cholesterol metabolism, complement activation among others (Fig 5C and S9 Table). On the contrary, enrichment analysis of genes more expressed in PB revealed biological processes such as leukocyte differentiation, lymphocyte differentiation, lymphocyte-mediated immunity, B cell activation, STAT cascade activation/regulation, and JAK-STAT cascade activation (Fig 5D and S10 Table), which are consistent with exacerbated responses in granulomatous diseases. Localized clinical forms, i.e., BT and TT, show a gene expression pattern indicative of differentiation towards epithelioid transformation and granuloma assembly, which is also observed in cutaneous or pulmonary sarcoidosis [36,37].

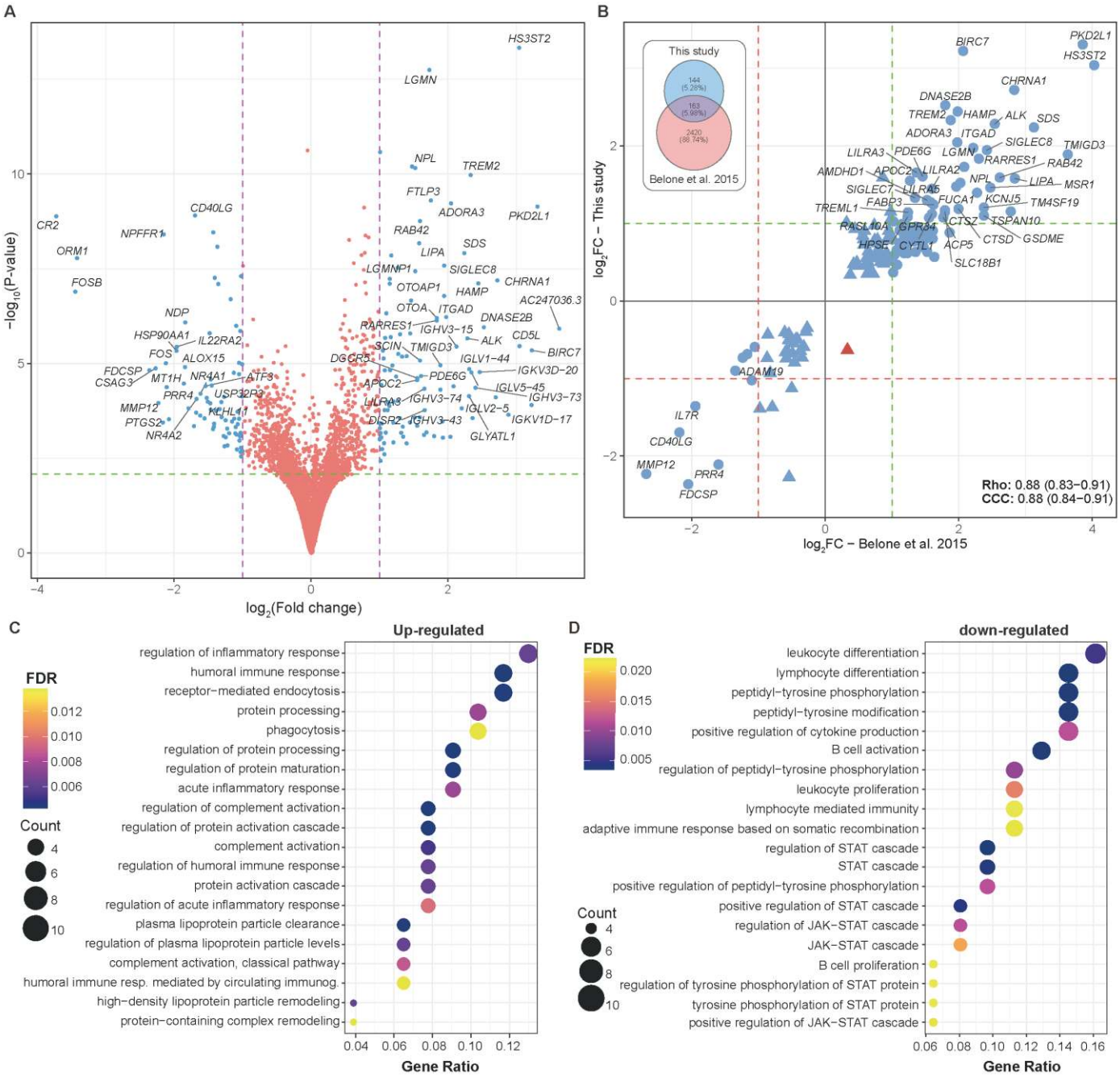


Fig 5. Differentially expressed genes from multibacillary (MB) vs. paucibacillary (PB) leprosy lesions. (A) Volcano plot showing DEG from the MB vs. PB comparison, where blue points are DE with $|\log_2FC| \geq 1$ and FDR < 0.1. (B) Scatter plots with the 161 DEG common between this study and Belone *et al.* (24) microarray for the same comparison. Red and green dashed lines indicate $\log_2FC$ of -1 and 1, respectively. Blue points are genes with the same modulation signal and red indicates discordancy. Rho, Spearman's rank correlation coefficient. CCC, Lin's concordance correlation

271 coefficient. Venn diagram on the right displays the number of DEG in each study
272 according to FDR < 0.01. (C) Biological processes from GO enriched from up-
273 regulated and (D) down-regulated DEG. FDR, false discovery rate.

274

275 Epithelial-mesenchymal transition (EMT) in the skin of
276 multibacillary leprosy patients

277 In a previous microarray meta-analysis, our group has identified a consistent
278 down-regulation of cornification, keratinocyte differentiation, and epidermal
279 development-related genes in leprosy lesions, predominantly in MB patients [35]. We
280 hypothesized that such regulation could result from *M. leprae* inducing
281 dedifferentiation of keratinocytes, similar to the phenomenon described previously in
282 infected Schwann cells [38], and also described in skin cancer by a process known as
283 epithelial-mesenchymal transition (EMT) [39,40]. To test the hypothesis that such
284 modulation was involved with EMT, we correlated the previously identified down-
285 regulated genes in leprosy [35] with the list of reported genes [38], together with
286 canonical genes involved with "stem-cell biology" (Reactome, R-HAS-452723) and
287 "TGF-beta in epithelial-mesenchymal transition" (R-HAS-2173791). We found a
288 consistent, moderate negative correlation mostly between
289 keratinization/cornification/epidermal development genes (*KRT5, KRT1, SCEL,*
290 *DMKN, GJA1, AQP3, TWIST1, VDR, IVL, PKPL*) with those involved with canonical
291 and alternative EMT and mesenchymal phenotypes (such as *TGFB1, TGIF2, RHOA,*
292 *ZEB2, PSAP, CTSZ, MMP9*) in our new RNA-seq dataset (Fig 6A-B, stars), and also
293 replicated these observations with an external microarray [24] and RNA-seq datasets

294 [28], respectively (Fig 6C, S3 and S4 Fig). Our results also showed a pattern in the
295 expression of the EMT-related genes as opposed to cytokeratins and epidermal
296 differentiation genes with low expression in healthy skin samples, and a linear
297 expression increase in PB and MB patients, especially with the microarray dataset,
298 except for *MMP9* (Fig 6C). From this result we hypothesized that in addition to TGFβ-
299 dependent immunosuppression in MB, activation of this pathway could be slowing or
300 arresting keratinocyte cornification processes in leprosy lesions thereby both
301 facilitating survival and/or spread of *M. leprae*.

302

303

304

305

306

307

308

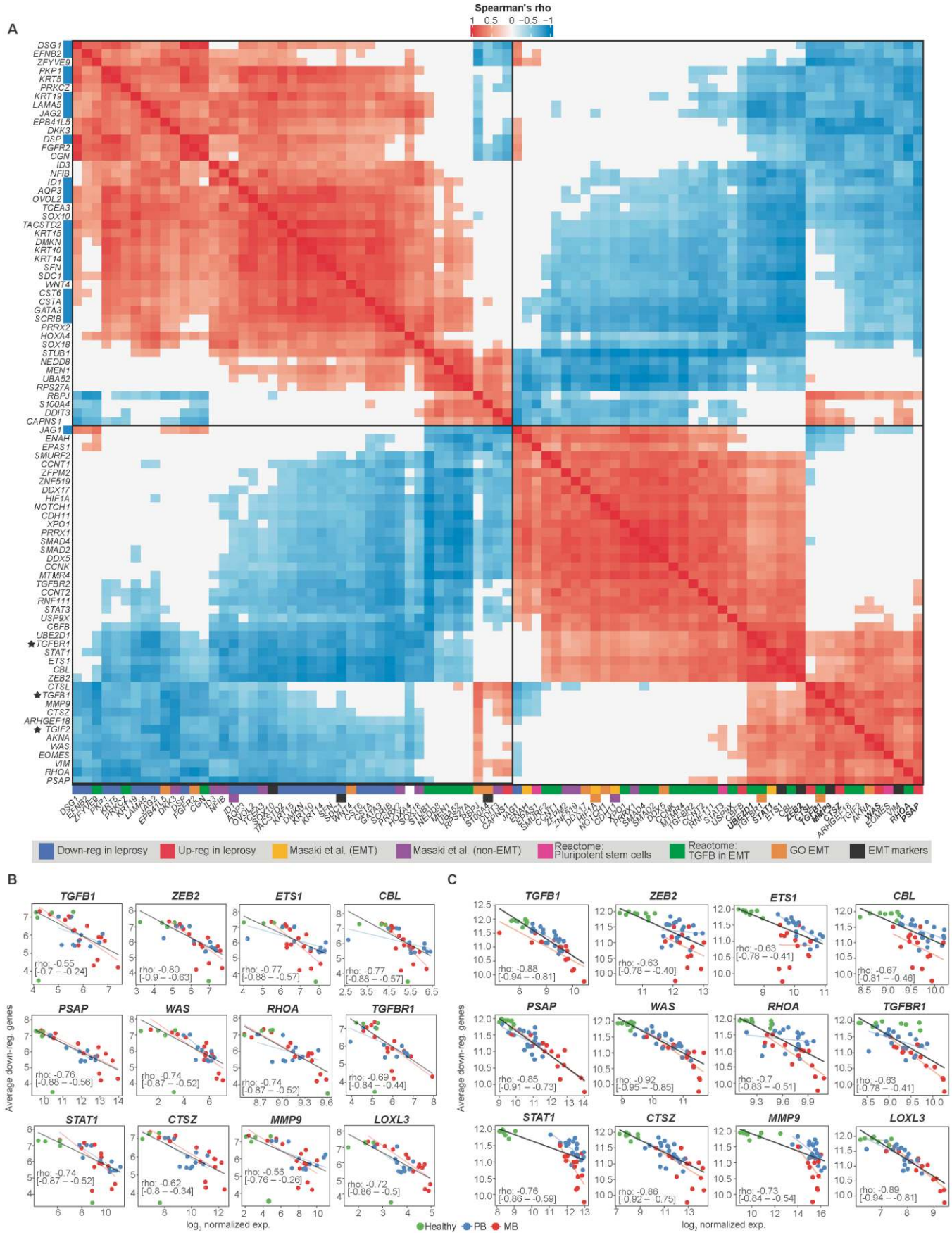
309

310

311

312

313



314 **Fig 6. Strongest correlations between keratinocyte and dedifferentiation genes**

in leprosy lesions. (A) Heat plot with Spearman's rho correlation coefficient of the strongest correlations after multiple testing adjustment (FDR ≤ 0.0001 and rho ≤ -0.8). Correlations with FDR > 0.1 are filled with white. Row and column colored squares identify gene categories (overlaps occur). Scatter plot of average $\log_2$ expression calculated with keratinocyte-related genes previously documented as down-regulated in leprosy skin against dedifferentiation-related genes using either (B) this study RNA-seq dataset or (C) Belone *et al.* microarray (GSE74481). Lines were drawn based on intercept and beta estimated from robust linear regression for all samples (black) or separately for PB (blue), and MB (red). Spearman's rho coefficient along with 95% nominal confidence intervals are shown inside scatter plots. See also S3 Fig and S4 Fig.

Discussion

One of the priorities in leprosy research is the development of reliable and accurate laboratory diagnosis tools for all leprosy forms to provide efficient treatment and prevent disability [41]. This goal includes diagnosing patients with early forms of the disease, those with low or mild apparent symptoms, thus assisting with ambiguous differential diagnoses, and even classifying the disease for treatment (MB *vs.* PB) [4].

Host response to infection as measured by gene expression in skin biopsies offers diagnostic, prognostic and predictive potential. By applying host transcriptomics to skin lesions from leprosy patients and other common confounding dermatoses that challenge clinicians and pathologists [9,30], we identified a small set of genes that provide a promising expression signature capable of distinguishing PB leprosy cases from other confounding dermatological diseases. The top candidate, *IDO1,* is a gene

involved in nutritional immunity and metabolism [42–45]. Alone, the expression of this gene was able to differentiate leprosy from non-leprosy lesions with high accuracy in our dataset and in others. According to the latest data from single-cell analysis [46], *IDO1* has been shown to be differentially expressed in Langerhans cells from leprosy lesions compared to healthy skin, corroborating our findings. However, *IDO1* expression is also increased in other mycobacterial diseases such as tuberculosis [47,48], which might decrease its specificity. The accuracy of classification could be improved by combining measurement of *IDO1* expression with that of four other biomarker genes *BLK, TLR10, CD38, and SLAMF7,* which also showed high classification accuracy in the replication dataset. In parallel, the penalized logistic regression model, evaluated on two independent datasets, demonstrated that *HS3ST2* and *CD40LG* hold great potential to differentiate between MB and PB lesions. Considering the functional evidence for *HS3ST2* [49], it is possible that this gene may be involved with granuloma disassembly, tissue permeability, and cellular migration in leprosy, which would explain its overexpression in MB lesions. On the contrary, *CD40LG* (also known as CD154) is more expressed in PB patients when compared to MB with a predominant role in the activation of the microbicidal *Th1* response associated with PB lesions [50]. After mechanistic validation of our findings, quantifying expression levels of *HS3ST2* and *CD40LG* from leprosy lesions could be useful to assess immune responsiveness against *M. leprae*, help patient stratification and/or provide a basis for host-based adjuvant treatment for leprosy lesions.

One of the challenges in translating gene expression signatures into medical diagnosis is the cost of measuring a large number of genes and transforming these values into a unique continuous or binary classifier. So far, we were able to reproduce

362 the finding using both bulk RNA-sequencing and relative RT-qPCR, with the latter
363 being more accessible to clinicians in reference centers or central hospitals. Although
364 there are successful RT-qPCR gene expression-based diagnostic tests currently
365 approved for diagnosing sepsis [12], clinical support for prostate [22], and breast
366 cancer [18], there is a need for alternatives to reduce the cost and complexity of such
367 assays. Quantification of mRNA based on isothermal amplification either with NASBA
368 [51,52], RT-LAMP [53,54] or CRISPR-Cas12 [55] is conceivable for less specialized
369 settings without high-end equipment. Besides, combining a multi-target expression-
370 based diagnostic test with qPCR detection of *M. leprae* DNA could increase the
371 specificity and sensitivity of leprosy diagnosis [56]. Alternatively, an ELISA assay
372 measuring the levels of IDO1 protein from skin interstitial fluid, for example, could be
373 proven useful [57].

374 In parallel with poor diagnosis, lack of fundamental understanding of leprosy
375 pathogenesis has misled scientists for centuries [5,6]. Herein, we also compared the
376 two leprosy poles, MB and PB, and identified several pathways already known to be
377 associated with leprosy, such as the humoral immune response, phagocytosis, and
378 complement activation. Genes involved with cholesterol and fatty acids were more
379 expressed in MB lesions, as already reported [58–60]. Interestingly, B-cell-related
380 genes were more expressed in PB than MB. In fact, it seems that both poles modulate
381 this pathway by a distinct set of genes. Involvement of B lymphocytes in PB leprosy
382 pathogenesis has been described by a few groups, which may indicate differential
383 involvement of such cells depending on the disease pole [61,62].

384 *M. leprae* subverts host cell metabolism [63] by inducing lipid biosynthesis,
385 while avoiding type II (IFN-gamma) responses through a type I IFNs mechanism,

386 following the phagolysosomal breach that releases DNA into the cytosol [64].
387 However, exactly how the bacilli spread throughout the body and bypass the
388 microbicidal immune response remains unknown. Here, we provide robust evidence
389 indicating that *M. leprae* may induce EMT in the skin within keratinocytes and
390 macrophages, as described in Schwann cells [38]. Indeed, *M. leprae* induced
391 dedifferentiation of infected Schwann cells into an immature stage resembling
392 progenitor/stem-like phenotype [38]. These reprogramming events induced by long-
393 term infection with *M. leprae* resulted in mesenchymal cells capable of migratory and
394 immune-permissive behavior, which in turn facilitated *M. leprae* spread to skeletal and
395 smooth muscles and furthered macrophage recruitment [38,65]. In our previous work,
396 we identified a down-regulated signature of keratinocyte differentiation and
397 cornification gene markers in MB skin lesions [35]. Here, we showed that such genes
398 are inversely correlated with genes involved with EMT, especially the members of the
399 TGFβ-EMT pathway, such as *TGFB1*, *TGFBR1*, *TGIF2, PSAP, ZEB2* [66,67]. Some
400 of these genes are directly or indirectly associated with EMT, such as a *PSAP* [68],
401 *WAS* [69], *RHOA* [70–73], *CTSZ* [74], *MMP9* [75], *LOXL3* [76], *HIF1A* [77,78] among
402 others.

403 Our hypothesis that *M. leprae* is inducing dedifferentiation or slowing the
404 cornification process in keratinocytes is plausible, given the evidence in Schwann cells
405 and a few reports of infection in this cell type (Fig 7) [79,80]. Nevertheless, other
406 phenomena could explain EMT's role in leprosy pathogenesis, such as wound healing
407 or loss of the epithelial barrier. Although, given its obligatory intracellular lifestyle, *M.*
408 *leprae* induces dedifferentiation in other cell types, either directly as in Schwann cells
409 or indirectly via chemokine and cytokine production in lesions. Besides inducing

keratinocyte dedifferentiation to mesenchymal cells, *M. leprae* might benefit from a decreased or alternative immune activation of these cells [81,82]. Further functional confirmatory experiments should elucidate the causality of this correlation and provide definitive evidence of the relationship between the bacilli and other cell types, such as keratinocytes, fibroblasts, and epithelial cells.

Our preliminary data also showed that the enriched pathways among PB skin lesions were consistent with profiles observed in other granulomatous diseases, such as noninfectious sarcoidosis and granuloma annulare, or chronic infectious diseases like tuberculosis [37,83–85]. Our findings revealed that PB (TT/BT) lesions have, among others, JAK-STAT cascade activation, which has been implicated in sarcoidosis and GA. Remarkably, the JAK-STAT specific biological inhibitor, tofacitinib, has a potent effect promoting rebalance of exacerbated immunity among sarcoidosis and granuloma annulare patients reestablishing homeostasis [83]. Another compound, everolimus, has been shown in experimental models to achieve the same response [37] suggesting that these drugs could be useful to treat PB, but not MB, leprosy.

To conclude, our combined findings provide highly discriminatory mRNA signatures from skin lesions that could distinguish leprosy from other dermatological diseases and allow disease classification by monitoring only a handful of genes. In addition, we report new genes and pathways that are likely informative regarding how *M. leprae* interacts with and subverts host cells to promote its spread within the body and subsequent transmission.

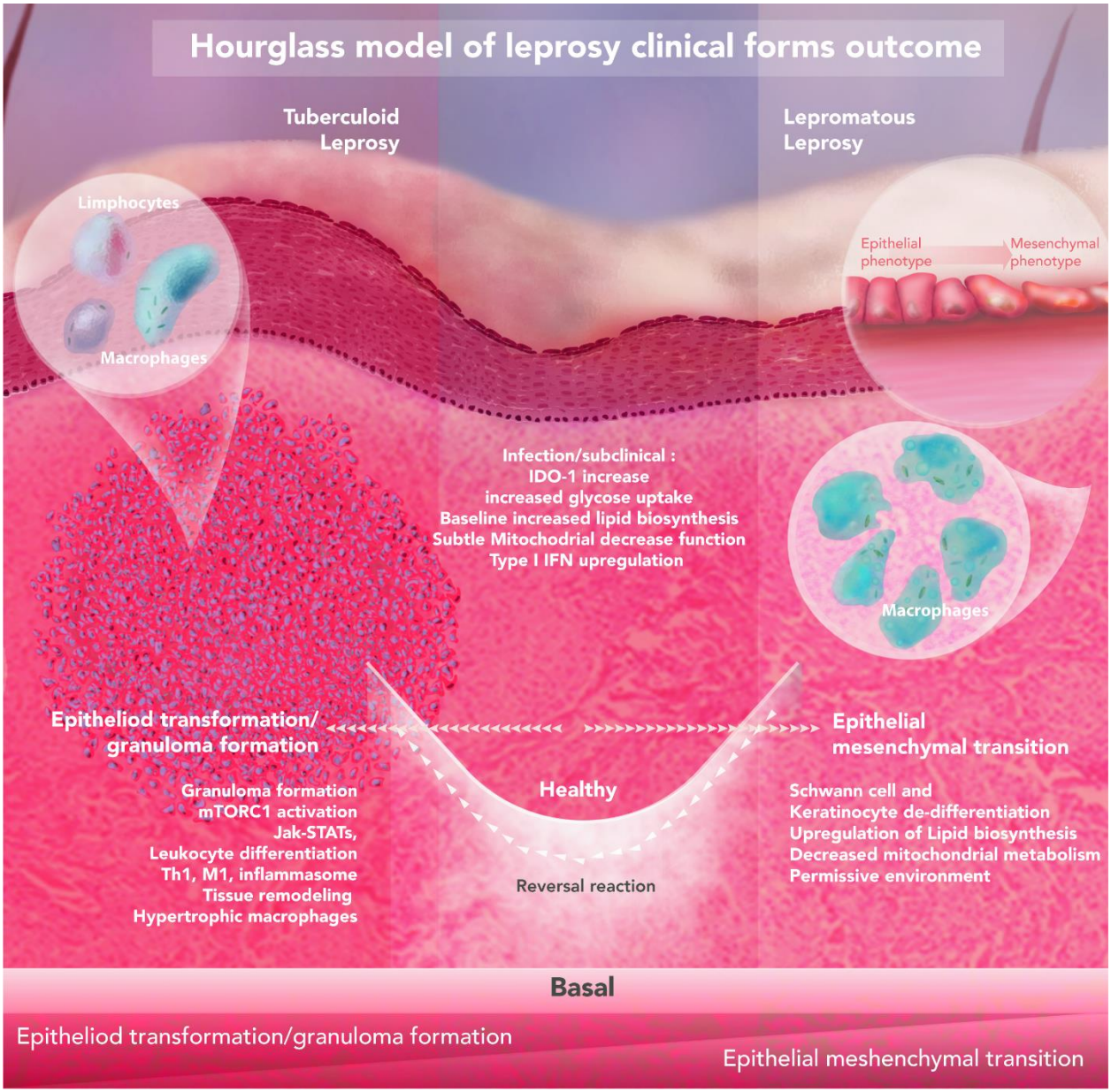


Fig 7. Hypothetical hourglass model contextualizing the observed findings for leprosy clinical outcomes. The host-pathogen interaction in the skin leads to opposing leprosy clinical forms. Upon infection, *M. leprae* induces baseline metabolic alterations such as an increase in glucose uptake, modulation of lipid biosynthesis, reduction of mitochondrial metabolism, and upregulation of IDO-1 and type I IFN. Eventually, progression towards an unspecified inflammatory state can be observed where three ways could be anticipated: I) self-healing; II) progression towards the tuberculoid pole; or III) progression to lepromatous pole. These outcomes are driven

by specific environmental and host genetic factors. It is expected that lower (or shorter) *M. leprae* exposure, food shortage, BCG vaccination, and polymorphisms in genes controlling autophagy/granuloma formation (*NOD2*, *LRRK2*, *PRKN*) all contribute to developing leprosy per se. Excessive inflammation is one phenotype observed, that is also seen in other granulomatous diseases (e.g., cutaneous sarcoidosis, granuloma annulare), especially in paucibacillary lesions. On the other pole, epithelial-mesenchymal transition and local immunosuppression are present due to a probably higher (and/or longer) *M. leprae* exposure, combined with host single-nucleotide polymorphisms (SNPs) at key genes, like lipid biogenesis (*APOE*) and central metabolism (*HIF1A, LACC1/FAMIN*), culminating in disease progression.

Materials and Methods

Patient cohort

All patients were enrolled after informed written consent was obtained with approval from the Ethics Committee of the Oswaldo Cruz Foundation, number 151/01. Leprosy clinical forms were classified according to the criteria of Ridley and Jopling [2]. Leprosy patients were treated according to the operational criteria established by the World Health Organization [4]. Leprosy and patients with other dermatological diseases were eligible if their diagnosis was confirmed by clinical and histopathological findings. Additionally, detection of *M. leprae* DNA by qPCR routinely performed in our laboratory could be employed to support diagnosis [56,86]. HIV and hepatitis B positive patients were not included in this study, in addition, we excluded individuals with a current or previous history of tuberculosis. No other comorbidities were used to

exclude patients and further individual information is available in S1 Table. Skin biopsy specimens containing both epidermis and dermis were obtained with 3 mm (diameter) sterile punches following local anesthesia from the lesion site. Skin biopsies were immediately stored in one milliliter of RNALater (Ambion, Thermo Fisher Scientific Inc., MA, USA) according to the manufacturer's instructions and stored in liquid nitrogen until RNA isolation. Normal skin biopsies were from lesion-free sites of patients diagnosed with indeterminate or pure neural leprosy.

Study Design

The main objective of this research was to identify host gene expression patterns capable of distinguishing leprosy (including the PB forms) from other differential diagnosis of skin lesions. Our working hypothesis was that leprosy lesions, despite their morphological and histopathological similarity to other skin diseases, may induce distinct patterns of gene expression in at a small subset. We predefined the comparison of leprosy (PB+MB) from non-leprosy including GA in addition to healthy patients for RNA sequencing experiment. In addition, we predetermined comparisons between leprosy poles: MB vs. PB. Our samples are representative of a population of individuals attending the Sousa Araujo Outpatient Clinic based in Rio de Janeiro, Brazil, which also receives patients from surrounding municipalities.

RNA isolation

Snap frozen skin biopsies were thawed in wet ice and processed using TRIzol Reagent (Ambion, Thermo Fisher Scientific Inc., MA, USA) according to the manufacturer's instructions with the help of Polytron Homogenizer PT3100

(Kinematica AG, Switzerland). RNA was treated with DNAse using the DNAfree kit (Thermo Fisher Scientific Inc., MA, USA) according to the standard manufacturer's protocol, prior to use for library preparation and RT-qPCR. RNA integrity was assessed in 1% agarose gel electrophoresis or TapeStation RNA ScreenTape (Agilent Technology, CA, USA). During RNA isolation, samples were randomly assigned to extraction batches and freeze-thaw cycles to minimize batch effects and the introduction of technical artifacts. All procedures applied to samples were carried out using reagents from the same lot. The first author conducted the experiments aware of each sample group during the entire process, therefore, no blinding scheme was used, although we do not rely on perceptual/abstract measurements or analyses nor did we purposefully exclude samples.

Library preparation and Illumina RNA sequencing

RNA-seq libraries were prepared with 1 µg of total RNA for each sample using the Illumina TruSeq mRNA kit (Illumina, USA) as recommended by the manufacturer using the Illumina CD RNA indexes (Illumina, USA). Libraries were quantified and qualified using a qPCR quantification protocol guide (KAPA Library Quantification Kits for Illumina Sequencing platforms) and TapeStation D1000 ScreenTape (Agilent Technologies, USA), respectively. The resulting libraries (fragment size 200-350bp) were multiplexed (17, 17, and 19 libraries, respectively) and sequenced using the NextSeq 500 platform (Illumina, USA), generating approximately 520 million single-end reads of 75 nucleotides in length.

RNA-sequencing analysis

RAW bcl files were converted into .fastq using Illumina's bcl2fastq script. Then, read quality was assessed using FastQC version 0.11.8 [87]. Next, transcript counts were estimated using Salmon (v.1.13.0) quasi-mapping (human transcriptome GRCh38_cdna sourced from Ensembl/RefGenie plus pre-computed salmon index, http://refgenomes.databio.org/#hg38_cdna) with default settings and --seqBias flag set [88]. Transcript counts were summarized into ENSEMBL gene counts using the R v.3.6.1 package tximport v.1.12.0 [89,90] and biomaRt v.2.40.5 [91]. The expression of sex-chromosome-specific genes, such as *UTY* and *XIST,* was used to rule out sample mislabeling. Differential expression was estimated using DESEq2 v.1.24.0, after filtering out weakly expressed genes with less than 10 counts per million and less than 15 total counts in 70% of samples [92–94]. In addition to the patient's biological sex, extraction batch and sequencing run, three surrogate variables estimated with RUVseq v.1.18.0 were included in DESeq2's generalized linear model [95,96]. Nominal P-values were inspected with histograms and adjusted for multiple testing according to the method [97] proposed for controlling the false discovery rate (FDR). All $\log_2$ fold-changes were shrunken prior to DE filtering with the apeglm [94] or normal algorithms. For visualization, counts per million (CPM) were computed with edgeR's cpm function v.3.26.1 and variance stabilized with the parametric method [92]. Then, surrogate variables and covariates were regressed out from the expression matrix using limma's removeBatchEffect [98–100] before being visualized with ggplot2 v.3.3.0 [101]. Hierarchical clustering, heatmaps, and ROC analysis were all performed with the previously processed expression matrix. Heatmap with hierarchical clustering was drawn with ComplexHeatmap v.2.0.0 [102] or pheatmap v.1.0.12 [103] using

gene-wise scaled and centered matrix with Euclidean distance and average agglomeration method. Overrepresentation analysis (ORA) was used to test for Gene Ontology Biological Process (GO BP) enrichment with clusterProfiler v.3.12.0 [104] and org.Hs.eg.db v.3.8.2 annotations [105]. Up and down-regulated lists were used as inputs and the background list was composed of all genes subjected to differential expression. P-values were adjusted for multiple testing using the Benjamini-Hochberg method [97]. Raw and normalized RNA sequencing data are available in EMBL-EBI's ENA and ArrayExpress under accessions ERP128243 and E-MTAB-10318, respectively.

RT-qPCR

A total of 2.5 µg of RNA was reversed transcribed into cDNA using 4 µL of Vilo Master Mix (Thermo Fisher Scientific Inc., USA) according to the manufacturer's instructions. Then, cDNA was diluted to a final concentration of 5 ng/µL using TE buffer (10 mM Tris-HCL and 0.1 mM EDTA in RNAse-free water). RT-qPCR was performed using Fast Sybr Master Mix (Thermo Fisher Scientific Inc., USA) in a final reaction volume of 10 µL. For each reaction, performed in duplicate, 5 µL of Fast Sybr Green were combined with 200 nM of each primer, 10 ng of cDNA, and q.s.p of injection-grade water. Thermal cycling and data acquisition were performed on Viia7 with 384 well block (Applied Biosystems, Thermo Fisher Scientific Inc., USA) following the master mix manufacturer cycling preset with a final melting curve analysis (65 °C to 95 °C, captured at every 0.5 °C). All primers were designed with NCBI Primer-Blast [106–109] to either flank intron(s) or span exon-exon junction(s) to avoid gDNA amplification (S11 Table). Further, primers were quality checked for specificity, dimers

and hairpin with MFEPrimer v.3.0 [110,111] and IDT's oligoAnalyzer (https://www.idtdna.com/calc/analyzer). Data were exported from QuantStudio software v.1.3 in RDML format, which was imported to LinRegPCR v.2020.0 for RT-qPCR efficiency determination and calculation of the N_0 value [112,113]. Finally, N_0 values were imported to R and normalized using as the denominator the normalization factor (NF) calculated from the geometric mean of at least three reference genes (*RPS16*, *RPL35* and *QRICH1*), which were previously tested for stability [114]. These N_0 normalized values were used for visualization in Fig 2A. For mean difference estimation between groups, RT-qPCR data were analyzed in a Bayesian framework (Markov Chain Monte Carlo sampling, MCMC) using generalized linear mixed effect models under lognormal-Poisson error with MCMC.qpcr v.1.2.4 [115,116]. Per-gene efficiency estimates from LinRegPCR were used in conjunction with Cp (crossing point) calculated in QuantStudio software v.1.3 to generate the counts table. Then, the generalized linear mixed-effect model was fitted using three reference genes (allowing up to 20% between-group variation) with 550,000 iterations, thin = 100, and burn-in of 50,000. The model specification included the sample (factor with 51 levels) as a random effect and the diagnosis group (factor with 3 levels) as a fixed effect. MCMC diagnostics were done by inspecting chain mixing plots and linear mixed model diagnostic plots. Ninety-five percent credible intervals were drawn around the posterior means and MCMC equivalent P-values were also computed.

Reanalysis of public gene expression datasets

Belone and collaborators GSE74481 [24] and de Toledo-Pinto and cols. GSE35423 [64] microarray datasets were reanalyzed as described elsewhere [35]. Blischak and cols. [32] RNA-seq dataset (GSE67427) was reanalyzed from counts per

sample file from the author's Bitbucket repository (https://bitbucket.org/jdblischak/tb-data/src/master/). Briefly, a normalized log_2 expression matrix was regressed out for RNA integrity number and extraction batch variables. Then, differences in gene expression (48h post-infection) for specific genes and treatments were tested using a gene-wise linear mixed model with a random intercept per sample (replicate) followed by Dunnet comparison against a "mock" group using emmeans v.1.5.3. Montoya and collaborators' dataset was retrieved from GEO (GSE125943) already normalized (DESeq2 median ratio method) and transformed with base 2 logarithm with no further processing [28].

Correlation analyses

For RNA-seq datasets, normalized log_2 counts-per-million values were used and log_2 normalized intensities for microarray. Spearman's rank correlation method was chosen because it is robust against outliers, does not rely on normality assumption, and also identifies monotonic but non-linear relationships. Initially, a list of keratinocyte/cornification/epidermal development genes that were DE in the meta-analysis was assembled [35]. Then, lists of target genes were compiled from results of Masaki *et al.* [38]: EMT and non-EMT; and from Reactome: R-HSA-452723 (Transcriptional regulation of pluripotent stem cells), R-HAS-5619507.3 (Activation of HOX genes during differentiation), R-HAS-2173791 (TGFβ receptor signaling in EMT); as well as Gene Ontology GO:0001837 (EMT), and literature for EMT canonical markers. Pairwise Spearman correlation was calculated using the Hmisc's rcorr function v.4.2-0 for each of the gene sets assembled. P-values were adjusted for multiple testing using the BH method for FDR control [97]. Additionally, 95% nominal

confidence intervals were calculated using the Fieller method implemented by correlation R package v.0.5.0 [117,118]. Next, a full correlation was computed using the union of all gene sets against the keratinocyte/cornification signature. Scatter plots were drawn with ggplot2 v.3.3.3 showing lines from coefficients estimated using default robust regression (MASS::rlm v.7.3-51.4). No outliers were omitted.

Regularized (LASSO) logistic regression classification

Normalized $\log_2$ expression matrices regressed out for covariates and batches were used as input predictors. The model was trained using the microarray dataset from Belone et al. [24] with penalized regression (L1-norm, LASSO) and 4-fold cross-validation (k-fold CV) with the negative binomial log-likelihood link function, glmnet v.4.1 [119–121]. Predictors were standardized to have mean zero and unit variance inside the cv.glmnet function. We opted for L1-norm because it results in a smaller number of genes (#features ≤ n) with non-zero coefficients, as compared to elastic-net or ridge regression counterparts. In addition, this model is suitable for high-dimensional data as it combines feature selection during model tuning and training, mitigating the effects of predictors' collinearity and reducing overfitting. To assess the coefficients' error, misclassification error rate, feature stability and model size we used non-parametric bootstrap (boot v.1.3.25) with 10,000 samples, with 4-fold cross-validation inside each loop [122,123]. The final LASSO model selected by 4-fold cross-validation contained three non-zero genes. Finally, independent RNA-seq test datasets were used to compute the accuracy of the final model. Alternatively, the whole process was repeated with leave-one-out cross-validation instead of k-fold. The

623 results were practically indistinguishable, especially regarding the feature stability
624 (data not shown).

625 Sample sizes

626 The sample size for RNA sequencing was selected based on previous leprosy
627 work with microarrays, aiming at detecting genes with at least a differential fold-change
628 of two. For RT-qPCR validation, sample size calculation was performed using the per-
629 gene standardized effect size estimated from the RNA-seq data, aiming at a power of
630 85% and alpha = 0.03. No samples were discarded after successful data collection
631 (i.e. outliers). In the end, the sample sizes per group for RT-qPCR were: MB = 13,
632 PB=11, ODD = 30. All RT-qPCR reactions were conducted in duplicate for each
633 biological unit (here, a fragment of a skin biopsy derived from an individual).

634 RT-qPCR and ROC statistical analyses

635 Normalized RT-qPCR gene expression data were $\log_2$ transformed before use
636 in data visualization. Additionally, we checked if the Bayesian results remained
637 consistent using a more common procedure (data not shown). For this, the mean
638 normalized expression (from N_0) was compared pairwise for the prior stipulated groups
639 using Welch's t-test implemented in R language, using the predetermined alpha of
640 0.03. Normality assumption was verified with normal quantile-quantile plots (qqplots,
641 car v. 3.0-2). In cases where quantile-quantile plots showed huge deviation from
642 theoretical normal distribution, the Wilcoxon Rank Sum was used to verify results.

643 Receiver Operating Curve (ROC) analysis was used to determine the accuracy
644 (measured by the area under the curve, AUC) and its respective best classification

threshold, aiming at maximizing AUC with equal importance for sensitivity and specificity. Confidence intervals (95%) for AUC were calculated using the Delong non-parametric method as implemented in pROC v.1.15.3 [124–126].

Data and code reporting

Raw .fastq data are available in EMBL-EBI European Nucleotide Archive (ENA) database (ERP128243). Raw Salmon counts and normalized batch cleaned expression matrices are available in EMBL-EBI ArrayExpress, under E-MTAB-10318, along with experimental and phenotypic metadata. R source code and accompanying intermediate data used in all analyses in this manuscript are also readily available through Zenodo, doi.org/10.5281/zenodo.4682010.

Acknowledgements

The authors wish to acknowledge Suelen Justo Moreira (MSc) and Rhana Prata (PhD) for assistance with skin biopsy RNA isolation. Helen Ferreira (MSc) and José Augusto for their technical and logistic support. The Gene Expression Core Facility (GECF) at EPFL, Lausanne, Switzerland, especially Drs. Elisa Cora and Bastien Mangeat for sequencing assistance. All patients and staff (physicians, nurses and technicians) from Sousa Araujo Outpatient clinic at FIOCRUZ, Rio de Janeiro, Brazil.

References

1. Britton WJ, Lockwood DN. Leprosy. The Lancet. 2004;363: 1209–1219. doi:10.1016/S0140-6736(04)15952-7

665 2. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-
666 group system. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1966;34: 255–73.

667 3. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman W, Williams DL.
668 The Continuing Challenges of Leprosy The Continuing Challenges of Leprosy. Clin
669 Microbiol Rev. 2006;19: 338–381. doi:10.1128/CMR.19.2.338

670 4. WHO. Guidelines for the Diagnosis, Treatment and Prevention of Leprosy.
671 Geneva: World Health Organization; 2018 p. 106.

672 5. WHO. Global leprosy (Hansen disease) update, 2019: time to step-up
673 prevention initiatives. Wkly Epidemiol Rec. 2020;95: 417–440.

674 6. Nath I, Saini C, Valluri VL. Immunology of leprosy and diagnostic challenges.
675 Clin Dermatol. 2015;33: 90–98. doi:10.1016/j.clindermatol.2014.07.005

676 7. van Hooij A, Tjon Kon Fat EM, Batista da Silva M, Carvalho Bouth R, Cunha
677 Messias AC, Gobbo AR, et al. Evaluation of Immunodiagnostic Tests for Leprosy in
678 Brazil, China and Ethiopia. Sci Rep. 2018;8: 1–9. doi:10.1038/s41598-018-36323-1

679 8. van Hooij A, van den Eeden S, Richardus R, Tjon Kon Fat E, Wilson L,
680 Franken KLMC, et al. Application of new host biomarker profiles in quantitative point-
681 of-care tests facilitates leprosy diagnosis in the field. EBioMedicine. 2019;47: 301–
682 308. doi:10.1016/j.ebiom.2019.08.009

683 9. Manta FS de N, Leal-Calvo T, Moreira SJM, Marques BLC, Ribeiro-Alves M,
684 Rosa PS, et al. Ultra-sensitive detection of Mycobacterium leprae: DNA extraction
685 and PCR assays. Poonawala H, editor. PLoS Negl Trop Dis. 2020;14: e0008325.
686 doi:10.1371/journal.pntd.0008325

687 10. Gliddon HD, Herberg JA, Levin M, Kaforou M. Genome-wide host RNA
688 signatures of infectious diseases: discovery and clinical translation. Immunology.
689 2018;153: 171–178. doi:10.1111/imm.12841

690 11. Ko ER, Yang WE, McClain MT, Woods CW, Ginsburg GS, Tsalik EL. What
691 was old is new again: Using the host response to diagnose infectious disease.
692 Expert Rev Mol Diagn. 2015;15: 1143–1158. doi:10.1586/14737159.2015.1059278

693 12. Miller RR, Lopansri BK, Burke JP, Levy M, Opal S, Rothman RE, et al.
694 Validation of a host response assay, SeptiCyte LAB, for discriminating sepsis from
695 systemic inflammatory response syndrome in the ICU. Am J Respir Crit Care Med.
696 2018;198: 903–913. doi:10.1164/rccm.201712-2472OC

697 13. Van Hooij A, Fat EMTK, Van Den Eeden SJF, Wilson L, Da Silva MB,
698 Salgado CG, et al. Field-friendly serological tests for determination of M. Leprae-
699 specific antibodies. Sci Rep. 2017;7: 1–8. doi:10.1038/s41598-017-07803-7

700 14. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for
701 tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med.
702 2019;16. doi:10.1371/journal.pmed.1002786

703 15. Röltgen K, Pluschke G, Spencer JS, Brennan PJ, Avanzi C. The immunology
704 of other mycobacteria: M. ulcerans, M. leprae. Semin Immunopathol. 2020;42: 333–
705 353. doi:10.1007/s00281-020-00790-4

706 16. Mesko B, Poliska S, Nagy L. Gene expression profiles in peripheral blood for
707 the diagnosis of autoimmune diseases. Trends Mol Med. 2011;17: 223–233.
708 doi:10.1016/j.molmed.2010.12.004

709 17. Wang B, Chen S, Zheng Q, Gao Z, Chen R, Xuan J, et al. Development and
710 initial validation of diagnostic gene signatures for systemic lupus erythematosus. Ann
711 Rheum Dis. 2019. doi:10.1136/annrheumdis-2019-216695

712 18. Carlson JJ, Roth JA. The impact of the Oncotype Dx breast cancer assay in
713 clinical practice: A systematic review and meta-analysis. Breast Cancer Res Treat.
714 2013;141: 13–22. doi:10.1007/s10549-013-2666-z

715 19. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy
716 S, et al. Translation of microarray data into clinically relevant cancer diagnostic tests
717 using gene expression ratios in lung cancer and mesothelioma. Cancer Res.
718 2002;62: 4963–4967.

719 20. Narrandes S, Xu W. Gene expression detection assay for cancer clinical use.
720 J Cancer. 2018;9: 2249–2265. doi:10.7150/jca.24744

721 21. Clark-Langone KM, Sangli C, Krishnakumar J, Watson D. Translating tumor
722 biology into personalized treatment planning: analytical performance characteristics
723 of the Oncotype DX®Colon Cancer Assay. BMC Cancer. 2010;10: 691.
724 doi:10.1186/1471-2407-10-691

725 22. Knezevic D, Goddard AD, Natraj N, Cherbavaz DB, Clark-Langone KM,
726 Snable J, et al. Analytical validation of the Oncotype DX prostate cancer assay - a
727 clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics.
728 2013;14: 1–12. doi:10.1186/1471-2164-14-690

729 23. Laible M, Schlombs K, Kaiser K, Veltrup E, Herlein S, Lakis S, et al. Technical
730 validation of an RT-qPCR in vitro diagnostic test system for the determination of
731 breast cancer molecular subtypes by quantification of ERBB2 , ESR1 , PGR and

732 MKI67 mRNA levels from formalin- fixed paraffin-embedded breast tumor
733 specimens. BMC Cancer. 2016; 1–14. doi:10.1186/s12885-016-2476-x

734 24. Belone A de FF, Rosa PS, Trombone APF, Fachin LRV, Guidella CC, Ura S,
735 et al. Genome-wide screening of mRNA expression in leprosy patients. Front Genet.
736 2015;6: 1–12. doi:10.3389/fgene.2015.00334

737 25. Jorge KTOS, Souza RP, Assis MTA, Araújo MG, Locati M, Jesus AMR, et al.
738 Characterization of MicroRNA Expression Profiles and Identification of Potential
739 Biomarkers in Leprosy. J Clin Microbiol. 2017;55: 1516–1525.
740 doi:10.1128/JCM.02408-16

741 26. Tió-Coma M, van Hooij A, Bobosha K, van der Ploeg-van Schip JJ, Banu S,
742 Khadge S, et al. Whole blood RNA signatures in leprosy patients identify reversal
743 reactions before clinical onset: a prospective, multicenter study. Sci Rep. 2019;9:
744 17931. doi:10.1038/s41598-019-54213-y

745 27. Tió-Coma M, Kiełbasa SM, van den Eeden SJF, Mei H, Roy JC, Wallinga J, et
746 al. Blood RNA signature RISK4LEP predicts leprosy years before clinical onset.
747 EBioMedicine. 2021;68: 103379. doi:10.1016/j.ebiom.2021.103379

748 28. Montoya DJ, Andrade P, Silva BJA, Teles RMB, Ma F, Bryson B, et al. Dual
749 RNA-Seq of Human Leprosy Lesions Identifies Bacterial Determinants Linked to
750 Host Immune Response. Cell Rep. 2019;26: 3574-3585.e3.
751 doi:10.1016/j.celrep.2019.02.109

752 29. Bhatia S, Shenoi SD, Pai K, Srilatha PS. Granuloma multiforme: an
753 uncommon differential for leprosy. Trop Doct. 2019;49: 55–58.
754 doi:10.1177/0049475518803191

755 30. Kundakci N, Erdem C. Leprosy: A great imitator. Clin Dermatol. 2019;37:
756 200–212. doi:10.1016/j.clindermatol.2019.01.002

757 31. Zhu TH, Kamangar F, Silverstein M, Fung MA. Borderline Tuberculoid
758 Leprosy Masquerading as Granuloma Annulare: A Clinical and Histological Pitfall.
759 Am J Dermatopathol. 2017;39: 296–299. doi:10.1097/DAD.0000000000000698

760 32. Blischak JD, Tailleux L, Mitrano A, Barreiro LB, Gilad Y. Mycobacterial
761 infection induces a specific human innate immune response. Sci Rep. 2015;5: 1–16.
762 doi:10.1038/srep16882

763 33. Modlin RL. Th1-Th2 paradigm: insights from leprosy. J Invest Dermatol.
764 1994;102: 828–832. doi:10.1111/1523-1747.ep12381958

765 34. Yamamura M, Uyemura K, Deans RJ, Weinberg K, Rea TH, Bloom BR, et al.
766 Defining protective responses to pathogens: Cytokine profiles in leprosy lesions.
767 Science. 1991;254: 277–279. doi:10.1126/science.1925582

768 35. Leal-Calvo T, Moraes MO. Reanalysis and integration of public microarray
769 datasets reveals novel host genes modulated in leprosy. Mol Genet Genomics.
770 2020;295: 1355–1368. doi:10.1007/s00438-020-01705-6

771 36. Judson MA, Marchell RM, Mascelli M, Piantone A, Barnathan ES, Petty KJ, et
772 al. Molecular profiling and gene expression analysis in cutaneous sarcoidosis: the
773 role of interleukin-12, interleukin-23, and the T-helper 17 pathway. J Am Acad
774 Dermatol. 2012;66: 901–910, 910.e1–2. doi:10.1016/j.jaad.2011.06.017

775 37. Linke M, Pham HTT, Katholnig K, Schnöller T, Miller A, Demel F, et al.
776 Chronic signaling via the metabolic checkpoint kinase mTORC1 induces

777 macrophage granuloma formation and marks sarcoidosis progression. Nat Immunol.
778 2017;18: 293–302. doi:10.1038/ni.3655

779 38. Masaki T, Qu J, Cholewa-Waclaw J, Burr K, Raaum R, Rambukkana A.
780 Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli
781 promotes dissemination of infection. Cell. 2013;152: 51–67.
782 doi:10.1016/j.cell.2012.12.014

783 39. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev
784 Cancer. 2018;18: 128–134. doi:10.1038/nrc.2017.118

785 40. Pastushenko I, Blanpain C. EMT Transition States during Tumor Progression
786 and Metastasis. Trends Cell Biol. 2019;29: 212–226. doi:10.1016/j.tcb.2018.12.001

787 41. Khazai Z, Van Brakel W, Essink D, Gillis T, Kasang C, Kuipers P, et al.
788 Reviewing Research Priorities of the Leprosy Research Initiative (LRI): a
789 stakeholder's consultation. Lepr Rev. 2019;90: 3–30. doi:10.47276/lr.90.1.3

790 42. Chen W. IDO: more than an enzyme. Nat Immunol. 2011;12: 809–811.
791 doi:10.1038/ni.2088

792 43. Greco FA, Coletti A, Camaioni E, Carotti A, Marinozzi M, Gioiello A, et al. The
793 Janus-faced nature of IDO1 in infectious diseases: challenges and therapeutic
794 opportunities. Future Med Chem. 2016;8: 39–54. doi:10.4155/fmc.15.165

795 44. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al.
796 The human transcriptome across tissues and individuals. Science. 2015;348: 660–
797 665. doi:10.1126/science.aaa0355

798 45. Yamazaki F, Kuroiwa T, Takikawa O, Kido R. Human indolylamine 2,3-
799 dioxygenase. Its tissue distribution, and characterization of the placental enzyme.
800 Biochem J. 1985;230: 635–638. doi:10.1042/bj2300635

801 46. Hughes TK, Wadsworth MH, Gierahn TM, Do T, Weiss D, Andrade PR, et al.
802 Second-Strand Synthesis-Based Massively Parallel scRNA-Seq Reveals Cellular
803 States and Molecular Features of Human Inflammatory Skin Pathologies. Immunity.
804 2020;53: 878-894.e7. doi:10.1016/j.immuni.2020.09.015

805 47. Gautam US, Foreman TW, Bucsan AN, Veatch AV, Alvarez X, Adekambi T, et
806 al. In vivo inhibition of tryptophan catabolism reorganizes the tuberculoma and
807 augments immune-mediated control of Mycobacterium tuberculosis. Proc Natl Acad
808 Sci U S A. 2018;115: E62–E71. doi:10.1073/pnas.1711373114

809 48. Yeung AWS, Terentis AC, King NJC, Thomas SR. Role of indoleamine 2,3-
810 dioxygenase in health and disease. Clin Sci. 2015;129: 601–672.
811 doi:10.1042/CS20140392

812 49. Denys A, Allain F. The emerging roles of heparan sulfate 3-O-
813 sulfotransferases in cancer. Front Oncol. 2019;9. doi:10.3389/fonc.2019.00507

814 50. Yamauchi PS, Bleharski JR, Uyemura K, Kim J, Sieling PA, Miller A, et al. A
815 Role for CD40-CD40 Ligand Interactions in the Generation of Type 1 Cytokine
816 Responses in Human Leprosy. J Immunol. 2000;165: 1506–1512.
817 doi:10.4049/jimmunol.165.3.1506

818 51. Heim A. Highly sensitive detection of gene expression of an intronless gene:
819 amplification of mRNA, but not genomic DNA by nucleic acid sequence based

820 amplification (NASBA). Nucleic Acids Res. 1998;26: 2250–2251.
821 doi:10.1093/nar/26.9.2250

822 52. Patterson SS, Casper ET, Garcia-Rubio L, Smith MC, Paul JH. Increased
823 precision of microbial RNA quantification using NASBA with an internal control. J
824 Microbiol Methods. 2005;60: 343–352. doi:10.1016/j.mimet.2004.10.011

825 53. Ganguli A, Ornob A, Spegazzini N, Liu Y, Damhorst G, Ghonge T, et al.
826 Pixelated spatial gene expression analysis from tissue. Nat Commun. 2018;9.
827 doi:10.1038/s41467-017-02623-9

828 54. Pandey M, Singh D, Onteru SK. Reverse transcription loop-mediated
829 isothermal amplification (RT-LAMP), a light for mammalian transcript analysis in low-
830 input laboratories. J Cell Biochem. 2018;119: 4334–4338. doi:10.1002/jcb.26624

831 55. Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, et al.
832 CRISPR–Cas12-based detection of SARS-CoV-2. Nat Biotechnol. 2020;38: 870–
833 874. doi:10.1038/s41587-020-0513-4

834 56. Barbieri RR, Manta FSN, Moreira SJM, Sales AM, Nery JAC, Nascimento
835 LPR, et al. Quantitative polymerase chain reaction in paucibacillary leprosy
836 diagnosis: A follow-up study. PLoS Negl Trop Dis. 2019;13: e0007147.
837 doi:10.1371/journal.pntd.0007147

838 57. Strassner JP, Rashighi M, Ahmed Refat M, Richmond JM, Harris JE. Suction
839 blistering the lesional skin of vitiligo patients reveals useful biomarkers of disease
840 activity. J Am Acad Dermatol. 2017;76: 847-855.e5. doi:10.1016/j.jaad.2016.12.021

841 58. Elamin AA, Stehr M, Singh M. Lipid Droplets and Mycobacterium leprae
842 Infection. J Pathog. 2012;10. doi:10.1155/2012/361374

843 59. Lobato LS, Rosa PS, Ferreira J da S, Neumann A da S, da Silva MG, do
844 Nascimento DC, et al. Statins increase rifampin mycobactericidal effect. Antimicrob
845 Agents Chemother. 2014;58: 5766–74. doi:10.1128/AAC.01826-13

846 60. Wang D, Zhang D-F, Li G-D, Bi R, Fan Y, Wu Y, et al. A pleiotropic effect of
847 the APOE gene: association of APOE polymorphisms with multibacillary leprosy in
848 Han Chinese from Southwest China. Br J Dermatol. 2018;178: 931–939.
849 doi:10.1111/bjd.16020

850 61. Fabel A, Giovanna Brunasso AM, Schettini AP, Cota C, Puntoni M, Nunzi E,
851 et al. Pathogenesis of Leprosy. Am J Dermatopathol. 2019;41: 422–427.
852 doi:10.1097/DAD.0000000000001310

853 62. Iyer AM, Mohanty KK, van Egmond D, Katoch K, Faber WR, Das PK, et al.
854 Leprosy-specific B-cells within cellular infiltrates in active leprosy lesions. Hum
855 Pathol. 2007;38: 1065–1073. doi:10.1016/j.humpath.2006.12.017

856 63. Medeiros RCA, Girardi K do C de V, Cardoso FKL, Mietto B de S, Pinto TG
857 de T, Gomez LS, et al. Subversion of Schwann Cell Glucose Metabolism by
858 Mycobacterium leprae. J Biol Chem. 2016;291: 21375–21387.
859 doi:10.1074/jbc.M116.725283

860 64. de Toledo-Pinto TG, Ferreira ABR, Ribeiro-Alves M, Rodrigues LS, Batista-
861 Silva LR, Silva BJ de A, et al. STING-Dependent 2′-5′ Oligoadenylate Synthetase–
862 Like Production Is Required for Intracellular Mycobacterium leprae Survival. J Infect
863 Dis. 2016;214: 311–320. doi:10.1093/infdis/jiw144

864 65. Hess S, Rambukkana A. Bacterial-induced cell reprogramming to stem cell-
865 like cells: new premise in host–pathogen interactions. Curr Opin Microbiol. 2015;23:
866 179–188. doi:10.1016/j.mib.2014.11.021

867 66. Vandewalle C, Comijn J, De Craene B, Vermassen P, Bruyneel E, Andersen
868 H, et al. SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell
869 junctions. Nucleic Acids Res. 2005;33: 6566–6578. doi:10.1093/nar/gki965

870 67. DaSilva-Arnold SC, Kuo CY, Davra V, Remache Y, Kim PCW, Fisher JP, et
871 al. ZEB2, a master regulator of the epithelial-mesenchymal transition, mediates
872 trophoblast differentiation. Mol Hum Reprod. 2018;25: 61–75.
873 doi:10.1093/molehr/gay053

874 68. Jiang Y, Zhou J, Hou D, Luo P, Gao H, Ma Y, et al. Prosaposin is a biomarker
875 of mesenchymal glioblastoma and regulates mesenchymal transition through the
876 TGF-β1/Smad signaling pathway. J Pathol. 2019;249: 26–38. doi:10.1002/path.5278

877 69. Frugtniet BA, Martin TA, Zhang L, Jiang WG. Neural Wiskott-Aldrich
878 syndrome protein (nWASP) is implicated in human lung cancer invasion. BMC
879 Cancer. 2017;17. doi:10.1186/s12885-017-3219-3

880 70. Bendris N, Arsic N, Lemmers B, Blanchard JM. Cyclin A2, Rho GTPases and
881 EMT. Small GTPases. 2012;3: 225–228. doi:10.4161/sgtp.20791

882 71. Bhowmick NA, Ghiassi M, Bakin A, Aakre M, Lundquist CA, Engel ME, et al.
883 Transforming growth factor-β1 mediates epithelial to mesenchymal
884 transdifferentiation through a RhoA-dependent mechanism. Mol Biol Cell. 2001;12:
885 27–36. doi:10.1091/mbc.12.1.27

886 72. Salvi A, Thanabalu T. WIP promotes in-vitro invasion ability, anchorage
887 independent growth and EMT progression of A549 lung adenocarcinoma cells by
888 regulating RhoA levels. Biochem Biophys Res Commun. 2017;482: 1353–1359.
889 doi:10.1016/j.bbrc.2016.12.040

890 73. Wang Q, Yang X, Xu Y, Shen Z, Cheng H, Cheng F, et al. RhoA/Rho-kinase
891 triggers epithelial-mesenchymal transition in mesothelial cells and contributes to the
892 pathogenesis of dialysis-related peritoneal fibrosis. Oncotarget. 2018;9: 14397–
893 14412. doi:10.18632/oncotarget.24208

894 74. Wang J, Chen L, Li Y, Guan XY. Overexpression of cathepsin Z contributes to
895 tumor metastasis by inducing epithelial-mesenchymal transition in hepatocellular
896 carcinoma. PLoS ONE. 2011;6. doi:10.1371/journal.pone.0024967

897 75. Lin CY, Tsai PH, Kandaswami CC, Lee PP, Huang CJ, Hwang JJ, et al.
898 Matrix metalloproteinase-9 cooperates with transcription factor Snail to induce
899 epithelial-mesenchymal transition. Cancer Sci. 2011;102: 815–827.
900 doi:10.1111/j.1349-7006.2011.01861.x

901 76. Peinado H, del Carmen Iglesias-de la Cruz M, Olmeda D, Csiszar K, Fong
902 KSK, Vega S, et al. A molecular role for lysyl oxidase-like 2 enzyme in Snail
903 regulation and tumor progression. EMBO J. 2005;24: 3446–3458.
904 doi:10.1038/sj.emboj.7600781

905 77. Tam SY, Wu VWC, Law HKW. Hypoxia-Induced Epithelial-Mesenchymal
906 Transition in Cancers: HIF-1α and Beyond. Front Oncol. 2020;10.
907 doi:10.3389/fonc.2020.00486

908 78. Zhu Y, Tan J, Xie H, Wang J, Meng X, Wang R. HIF-1α regulates EMT via the
909 Snail and β-catenin pathways in paraquat poisoning-induced early pulmonary
910 fibrosis. J Cell Mol Med. 2016;20: 688–697. doi:10.1111/jcmm.12769

911 79. Lyrio ECD, Campos-Souza IC, Corrêa LCD, Lechuga GC, Verícimo M, Castro
912 HC, et al. Interaction of Mycobacterium leprae with the HaCaT human keratinocyte
913 cell line: new frontiers in the cellular immunology of leprosy. Exp Dermatol. 2015;24:
914 536–542. doi:10.1111/exd.12714

915 80. Okada S, Komura J, Nishiura M. Mycobacterium leprae found in epidermal
916 cells by electron microscopy. IntJLeprOther MycobactDis. 1978;46: 30–34.

917 81. Pivarcsi A, Kemény L, Dobozy A. Innate Immune Functions of the
918 Keratinocytes. Acta Microbiol Immunol Hung. 2004;51: 303–310.
919 doi:10.1556/AMicr.51.2004.3.8

920 82. Pivarcsi A, Nagy I, Lajos K. Innate Immunity in the Skin: How Keratinocytes
921 Fight Against Pathogens. Curr Immunol Rev. 2005;1: 29–43.
922 doi:10.2174/1573395052952941

923 83. Damsky W, Thakral D, McGeary MK, Leventhal J, Galan A, King B. Janus
924 kinase inhibition induces disease remission in cutaneous sarcoidosis and granuloma
925 annulare. J Am Acad Dermatol. 2020;82: 612–621. doi:10.1016/j.jaad.2019.05.098

926 84. Flynn JL, Chan J, Lin PL. Macrophages and control of granulomatous
927 inflammation in tuberculosis. Mucosal Immunol. 2011;4: 271–278.
928 doi:10.1038/mi.2011.14

929 85. Locke LW, Crouser ED, White P, Julian MW, Caceres EG, Papp AC, et al. IL-
930 13–regulated Macrophage Polarization during Granuloma Formation in an In Vitro

931 Human Sarcoidosis Model. Am J Respir Cell Mol Biol. 2019;60: 84–95.
932 doi:10.1165/rcmb.2018-0053OC

933 86. Manta FSN, Barbieri RR, Moreira SJM, Santos PTS, Nery JAC, Duppre NC,
934 et al. Quantitative PCR for leprosy diagnosis and monitoring in household contacts:
935 A follow-up study, 2011–2018. Sci Rep. 2019;9. doi:10.1038/s41598-019-52640-5

936 87. Brabaham Bioinformatics. FastQC: A Quality Control Tool for High
937 Throughput Sequence Data [Online]. 2015. Available:
938 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

939 88. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast
940 and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–
941 419. doi:10.1038/nmeth.4197

942 89. R Core Team. R: A language and environment for statistical computing.
943 Vienna, Austria; 2017. Available: https://www.r-project.org/

944 90. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq:
945 transcript-level estimates improve gene-level inferences. F1000Research. 2016;4:
946 1521. doi:10.12688/f1000research.7563.2

947 91. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al.
948 BioMart and Bioconductor: a powerful link between biological databases and
949 microarray data analysis. Bioinformatics. 2005;21: 3439–3440.
950 doi:10.1093/bioinformatics/bti525

951 92. Anders S, Huber W. Differential expression analysis for sequence count data.
952 Genome Biol. 2010;11: R106. doi:10.1186/gb-2010-11-10-r106

953 93. Love MI, Huber W, Anders S. Moderated estimation of fold change and
954 dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.
955 doi:10.1186/s13059-014-0550-8

956 94. Zhu A, Ibrahim JG, Love MI. Heavy-Tailed prior distributions for sequence
957 count data: Removing the noise and preserving large differences. Bioinformatics.
958 2019;35: 2084–2092. doi:10.1093/bioinformatics/bty895

959 95. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted
960 variation in microarray data. Biostatistics. 2012;13: 539–552.
961 doi:10.1093/biostatistics/kxr034

962 96. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using
963 factor analysis of control genes or samples. Nat Biotechnol. 2014;32: 896–902.
964 doi:10.1038/nbt.2931

965 97. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical
966 and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society.
967 Series B (Methodological). WileyRoyal Statistical Society; 1995.
968 doi:10.2307/2346101

969 98. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust
970 hyperparameter estimation protects against hypervariable genes and improves
971 power to detect differential expression. Ann Appl Stat. 2016;10: 946–963.
972 doi:10.1214/16-AOAS920

973 99. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
974 differential expression analyses for RNA-sequencing and microarray studies. Nucleic
975 Acids Res. 2015;43: e47. doi:10.1093/nar/gkv007

976 100. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing
977 Differential Expression in Microarray Experiments Linear Models and Empirical
978 Bayes Methods for Assessing Differential Expression in Microarray Experiments.
979 Stat Appl Genet Mol Biol. 2004;3: 1–26. doi:10.2202/1544-6115.1027

980 101. Wickham H. ggplot2-Elegant Graphics for Data Analysis. 1st ed. New York,
981 NY: Springer New York; 2009. doi:10.1007/978-0-387-98141-3

982 102. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and
983 correlations in multidimensional genomic data. Bioinformatics. 2016;32: 2847–2849.
984 doi:10.1093/bioinformatics/btw313

985 103. Kolde R. pheatmap: Pretty Heatmaps. 2015. Available: https://cran.r-
986 project.org/package=pheatmap

987 104. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for
988 Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. 2012;16:
989 284–287. doi:10.1089/omi.2011.0118

990 105. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. 2019.
991 Available: 10.18129/B9.bioc.org.Hs.eg.db

992 106. Koressaar T, Remm M. Enhancements and modifications of primer design
993 program Primer3. Bioinformatics. 2007;23: 1289–1291.
994 doi:10.1093/bioinformatics/btm091

995 107. Kõressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M.
996 Primer3_masker: integrating masking of template sequence with primer design
997 software. Bioinformatics. 2018;34: 1937–1938. doi:10.1093/bioinformatics/bty036

998 108. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al.
999 Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40: e115–e115.
1000 doi:10.1093/nar/gks596

1001 109. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-
1002 BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC
1003 Bioinformatics. 2012;13: 134. doi:10.1186/1471-2105-13-134

1004 110. Qu W, Shen Z, Zhao D, Yang Y, Zhang C. MFEprimer: Multiple factor
1005 evaluation of the specificity of PCR primers. Bioinformatics. 2009;25: 276–278.
1006 doi:10.1093/bioinformatics/btn614

1007 111. Wang K, Li H, Xu Y, Shao Q, Yi J, Wang R, et al. MFEprimer-3.0: Quality
1008 control for PCR primers. Nucleic Acids Res. 2019;47: W610–W613.
1009 doi:10.1093/nar/gkz351

1010 112. Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM. Assumption-
1011 free analysis of quantitative real-time polymerase chain reaction (PCR) data.
1012 Neurosci Lett. 2003;339: 62–66. doi:10.1016/S0304-3940(02)01423-4

1013 113. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, Van den hoff
1014 MJB, et al. Amplification efficiency: Linking baseline and bias in the analysis of
1015 quantitative PCR data. Nucleic Acids Res. 2009;37. doi:10.1093/nar/gkp045

1016 114. Vandesompele J, De Preter K, Pattyn ilip, Poppe B, Van Roy N, De Paepe A,
1017 et al. Accurate normalization of real-time quantitative RT-PCR data by geometric
1018 averaging of multiple internal control genes. Genome Biol. 2002;3: 34–1.
1019 doi:10.1186/gb-2002-3-7-research0034

115. Matz MV, Wright RM, Scott JG. No control genes required: Bayesian analysis of qRT-PCR data. PloS One. 2013;8: 1–12. doi:10.1371/journal.pone.0071448

116. Steibel JP, Poletto R, Coussens PM, Rosa GJM. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics. 2009;94: 146–152. doi:10.1016/j.ygeno.2009.04.008

117. Fieller EC, Hartley HO, Pearson ES. TESTS FOR RANK CORRELATION COEFFICIENTS I. Biometrika. 1957;44: 470–481. doi:10.1093/biomet/44.3-4.470

118. Makowski D, Ben-Shachar MS, Patil I, Lüdecke D. Methods and Algorithms for Correlation Analysis in R. J Open Source Softw. 2020;5: 2306. doi:10.21105/joss.02306

119. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33: 1–22. doi:10.18637/jss.v033.i01

120. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw. 2011;39: 1–13. doi:10.18637/jss.v039.i05

121. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B Methodol. 1996;58: 267–288.

122. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity. 1st ed. Chapman and Hall/CRC; 2015.

123. Davison AC, Hinley DV. Bootstrap Methods and Their Application. Cambrige University Press; 1997. Available: http://statwww.epfl.ch/davison/BMA/

124. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44: 837. doi:10.2307/2531595

125. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12: 77. doi:10.1186/1471-2105-12-77

126. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. IEEE Signal Process Lett. 2014;21: 1389–1393. doi:10.1109/LSP.2014.2337313

Supporting Information

S1 Appendix. Linking expression profiles to mycobacteria species.

S1 Fig. Gene expression in MB and PB groups from test and training datasets. Normalized $\log_2$ expression values per group from (A) this study RNA-seq dataset or (B) Belone *et al.* (GSE74481) [24]. The genes shown were selected in 25%–50% of the LASSO models (Fig 4B) according to the bootstrap. MB, multibacillary leprosy; PB, paucibacillary leprosy; TT, tuberculoid leprosy; BT, borderline-tuberculoid; BB, borderline-borderline; BL, borderline-lepromatous; LL, lepromatous. Each point represents an independent skin biopsy from a patient. Y-axis values are not comparable between panels A and B.

S2 Fig. Strongest correlations between the average expression of genes associated with keratinocyte/cornification against dedifferentiation-related genes using Montoya *et al.* RNA-seq dataset [28]**.** Scatter plots of scores (average

normalized log_2 expression) calculated from genes with previously documented down-regulation in leprosy skin lesions against dedifferentiation-related genes with Montoya *et al.* RNA-seq dataset (GSE125943) [28]. Lines were drawn based on intercept and beta estimates from robust linear regression for all samples (black) or separately for TL (tuberculoid leprosy, blue), and LL (lepromatous leprosy, red). X-axis shows log_2 normalized expression values. Spearman's rho are shown along with nominal 95% confidence intervals inside the plots. Most genes shown have FDR < 0.1 and rho ≤ -0.6. Related to figure 6.

S3 Fig. Strongest correlations between modulated genes from keratinocyte/cornification and dedifferentiation-related genes using Belone et al. microarray dataset (GSE74481) [24]**.** Heat plot with Spearman's rho correlation coefficient of the strongest correlations from all ontologies screened after multiple testing adjustment (BH-FDR). Most genes shown have FDR ≤ 0.0001 and rho ≤ -0.7. Bottom colored rectangles indicate which category the gene was present (some genes co-occur). Related to figure 6.

S1 Table. Demographic and clinical metadata from human participants.

S2 Table. Genes differentially expressed from leprosy vs. non-leprosy with $|log_2FC| \geq 1$ and FDR ≤ 0.01.

S3 Table. Over-representation analysis (ORA) for leprosy vs. non-leprosy (up-regulated) differentially expressed genes.

S4 Table. ROC analysis from RNA-seq dataset using leprosy vs. non-leprosy samples.

S5 Table. Posterior log_2FC estimates, 95% credible intervals and MCMC P-values from PB-OD and MB-OD comparisons.

S6 Table. ROC analysis results using RT-qPCR with the validation dataset (Related to Fig 3). 95% confidence intervals are shown, except for AUCs of 1.0. The table is sorted from highest to lowest AUC.

S7 Table. Log_2FC estimates, confidence intervals, and Dunnet *P*-values from distinct mycobacterial stimuli in human macrophages *in vitro*.

S8 Table. Genes differentially expressed from multibacillary paucibacillary leprosy with $|log_2FC| \geq 1$ and FDR ≤ 0.01.

S9 Table. Over-representation analysis (ORA) for MB vs. PB (up-regulated) differentially expressed genes.

S10 Table. Over-representation analysis (ORA) for MB vs. PB (down-regulated) differentially expressed genes.

S11 Table. Oligonucleotide sequences.