Dendritic cells isolated from individuals susceptible to tuberculosis have an altered gene expression profile

John D. Blischak1,2, Ludovic Tailleux3, Marsha Myrthil1, Luis B. Barreiro4,5\*, Yoav Gilad1\*

1Department of Human Genetics, University of Chicago, Chicago, Illinois, USA

2Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, Illinois, USA

3Mycobacterial Genetics Unit, Institut Pasteur, Paris, France

4Department of Genetics, CHU Sainte-Justine Research Center, Montreal, Québec, Canada

5Department of Pediatrics, University of Montreal, Montreal, Québec, Canada

\*Correspondence should be addressed to YG (gilad@uchicago.edu) and LBB (luis.barreiro@umontreal.ca).

Write abstract here…

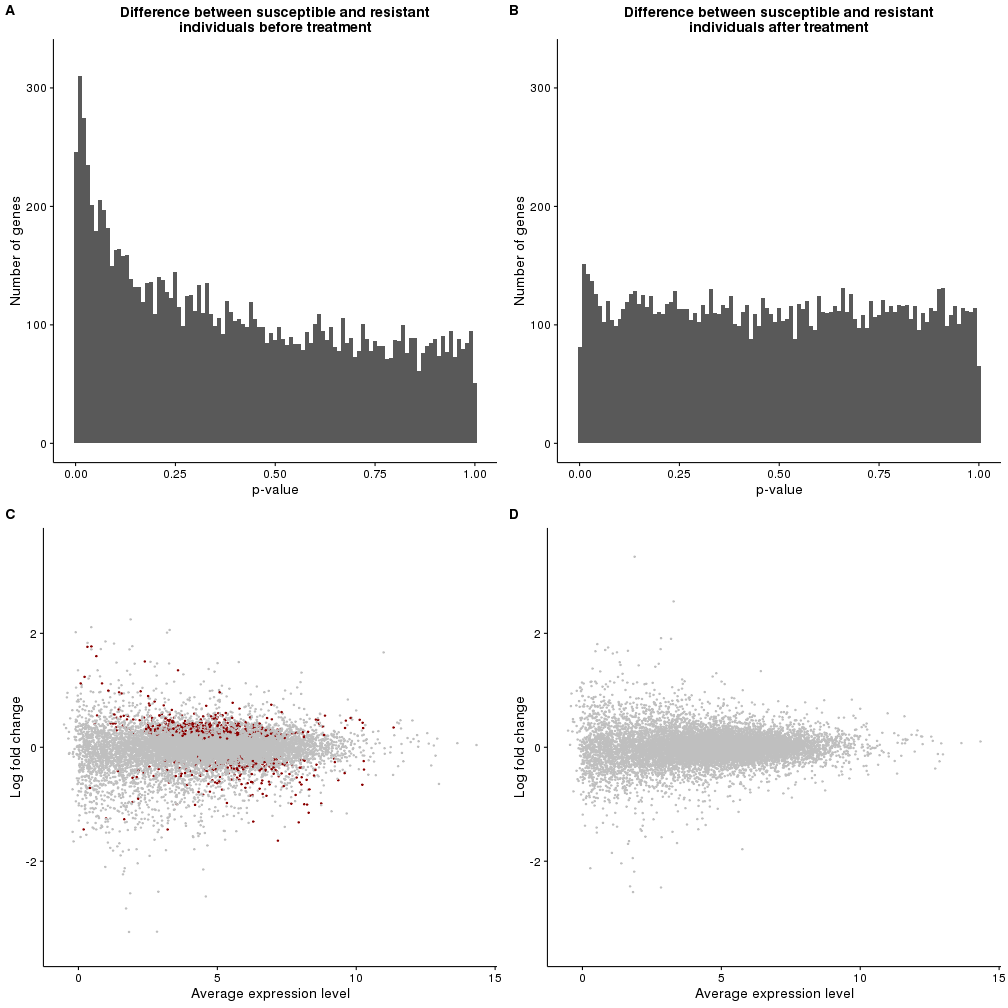
# Introduction

# Results

## Susceptible individuals have an altered transcriptome in the noninfected state

We obtained whole blood samples from 25 healthy individuals. Six of the donors had recovered from a previous active TB infection, and are thus susceptible. The remaining 19 tested positive for a latent TB infection without ever experiencing symptoms of active TB, and are thus resistant. We isolated dendritic cells (DCs) and treated them with *Mycobacterium tuberculosis* (MTB) or a mock control. To measure genome-wide gene expression levels, we sequenced the RNA, using a processing pipeline designed to minimize the introduction of unwanted technical variation, and obtained a mean of X $\pm$ X million raw reads per sample. We performed quality control analyses to remove non-expressed genes (Supplementary Fig. \ref{fig:genes}), identify and remove outliers (Supplementary Fig. \ref{fig:outliers}), and check for confounding batch effects (Supplementary Fig. \ref{fig:batch}). Ultimately 6 samples were removed from all downstream analyses (Supplementary Table \ref{tab:outliers}).

Next we performed a standard differential expression analysis. Of most interest are genes which were differentially expressed (DE) between susceptible and resistant individuals in the noninfected and infected states (Fig. \ref{fig:limma}). After correcting for multiple testing, we identified X DE genes in the noninfected state, and none in the infected state. <Insert GO results and a few examples>



\begin{figure}[ht]

\centering

\includegraphics[width=\linewidth]{../figure/limma}

\caption{

Differential expression analysis. The top panel contains the distribution of unadjusted p-values after testing for differential expression between susceptible and resistant individuals in the (A) noninfected or (B) infected state. The bottom panel contains the corresponding MA plots for the (C) noninfected and (D) infected states. The x-axis is the average gene expression level and the y-axis is the log fold change in expression level between susceptible and resistant individuals. Red dots indicate genes which are significant differentially expressed after applying a multiple testing correction.

}

\label{fig:limma}

\end{figure}

## Differentially expressed genes are enriched for TB susceptibility loci

## Gene expression levels in the noninfected state can predict susceptibility status

# Discussion

To do…

# Methods

## Sample collection

We collected whole blood samples from healthy Caucasian male individuals living in France. The putatively resistant individuals tested positive for a latent TB infection in an interferon-$\gamma$ release assay, but had never developed active TB. The putatively sensitive individuals had developed active TB in the past, but were currently healthy. Our protocol was approved by the Institutional Review Boards of the University of Chicago (10-504-B) and the Institut Pasteur (IRB00006966).

## Isolation and infection of dendritic cells

## RNA extraction and sequencing

We extracted RNA using the Qiagen miRNeasy Kit and prepared sequencing libraries using the Illumina TruSeq Kit. We sent the master mixes to the University of Chicago Functional Genomics Facility to be sequenced on an Illumina HiSeq 4000. We designed the batches for RNA extraction, library preparation, and sequencing to balance the experimental factors of interest and thus avoid potential technical confounders.

## Read mapping

We mapped reads to human genome hg38 (GRCh38) using Subread and discarded non-uniquely mapping reads. We assigned mapped reads to 19,800 Ensembl protein-coding genes (Ensembl 83, Dec 2015, GRCh38.p5) using featureCounts.

## Quality control

First we filtered genes by their expression level by removing all genes with a median log2 counts per million (cpm) less than zero. This resulted in a final set of 11,336 genes for downstream analysis (Supplementary Fig. Sx). Next we used principal components analysis (PCA) and hierarchical clustering to identify and remove 6 outliers (Supplementary Fig. Sx, Sx, Sx; Supplementary Table \ref{tab:outliers}).

## Differential expression analysis

We used limma+voom \cite{Smyth2004, Law2014, Ritchie2015} to implement the following linear model to test for differential expression:

$$ Y\ \sim \beta\_0 + X\_{treat}\beta\_{treat} + X\_{status}\beta\_{status} + X\_{treat,status}\beta\_{treat,status} + I + \epsilon $$

## Comparison to GWAS results

## Classifier

# Acknowledgements

We thank T. Thye for sharing the GWAS data with us. This study was funded by National Institutes of Health (NIH) Grant AI087658 to YG and LT. JDB was supported by NIH T32GM007197. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

# Author Contributions

YG, LT, and LBB conceived of the study and designed the experiments. LT performed the infection experiments. MM extracted the RNA and prepared the sequencing libraries. JDB analyzed the results. LBB and YG supervised the project. JDB and YG wrote the original draft. All authors reviewed the manuscript.

\bibliography{references}

# Supplementary Information

## Software implementation

## Supplementary Tables

|  |  |  |
| --- | --- | --- |
| ID | Description | PCs |
| 18-contact-infected | Resistant individual after infection | 1 |
| 06-contact-none | Resistant individual before infection | 2, 4, 5 |
| 06-contact-infected | Resistant individual after infection | 2 |
| 04-tb-none | Susceptible individual before infection | 3, 5, 6 |
| 02-contact-none | Resistant individual before infection | 3, 5, 6 |
| 01-tb-infected | Susceptible individual after infection | 5 |

\begin{table}[ht]

\centering

\caption{

The following outliers were removed from all analyses based on their discordant clustering observed with PCA and hierarchical clustering.

}

\begin{tabular}{|l|l|l|}

\hline

ID & Description & PCs \\

\hline

18-contact-infected & Resistant individual after infection & 1 \\

\hline

06-contact-none & Resistant individual before infection & 2, 4, 5 \\

\hline

06-contact-infected & Resistant individual after infection & 2 \\

\hline

04-tb-none & Susceptible individual before infection & 3, 5, 6 \\

\hline

02-contact-none & Resistant individual before infection & 3, 5, 6 \\

\hline

01-tb-infected & Susceptible individual after infection & 5 \\

\hline

\end{tabular}

\label{tab:outliers}

\end{table}