A Differential DNA Methylome Signature of Pulmonary Immune Cells from Individuals Converting to Latent Tuberculosis Infection

Lovisa Karlsson[∂1], Jyotirmoy Das[∂1], Moa Nilsson[1], Amanda Tyrén[1], Isabelle Pehrson[1], Nina Idh[1], Jakob Paues[1,2], Cesar Ugarte-Gil[3], Melissa Méndez-Aranda[•4], and Maria Lerm[•*1]

[1]Division of Inflammation and Infection, [2]Division of Infectious Diseases, Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences, Linköping University, SE-58185, Linköping, Sweden, [3]Facultad de Medicina, Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru, [4]Facultad de Ciencias y Filosofía, Laboratorio de Investigación en Enfermedades Infecciosas, Universidad Peruana Cayetano Heredia, Lima, Peru

∂) Shared first authorship

•) Shared last authorship

**) Corresponding author:*

Maria Lerm, Div. of Inflammation and Infection, Lab 1, floor 12
Dept. of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences
Linköping University, SE-58185 Linköping, Sweden
Phone: +46-732707786, E-mail: maria.lerm@liu.se

25 **Keywords:** DNA methylation, transcriptome, tuberculosis, BCG, induced sputum,

26 epigenetics, IGRA

27

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, spreads via aerosols and the first encounter with the immune system is with the pulmonary resident immune cells. The role of epigenetic regulations through DNA methylation in the immune cells is emerging. We have previously shown that capacity to kill *M. tuberculosis* is reflected in the DNA methylome. The aim of this study was to investigate epigenetic modifications in the pulmonary immune cells in a cohort of medical students with a previously documented increased risk of TB exposure, longitudinally. Sputum samples containing alveolar macrophages (AMs) and T cells were collected before and after study subjects worked in hospital departments with a high-risk of TB exposure. DNA methylome analysis revealed that a unique DNA methylation profile was present already at inclusion in subjects who developed latent TB during the study. The profile was both reflected in different overall DNA methylation distribution as well as more profound alterations in the methylation status of a unique set of CpG-sites. Over-representation analysis of the DMGs showed enrichment in pathways related to metabolic reprograming of macrophages and T cell migration and IFN-γ production. In conclusion, we identified a unique DNA methylation signature in individuals, while still IGRA-negative and who later developed latent TB. Epigenetic regulation was found in pathways that have previously been reported to be important in TB. Together the study suggests that DNA methylation status of pulmonary immune cells can predict IGRA conversion.

Introduction

Tuberculosis (TB) is a major global health concern, ranked as one of the top 10 causes of death worldwide and estimated to be responsible for 1.2 million deaths per year [1]. TB is caused by the facultative intracellular bacteria *Mycobacterium tuberculosis* and one-fourth of the world's population is estimated to be infected making *M. tuberculosis* is one of the most successful pathogens known. *M. tuberculosis* is air-born and enters the lung where the bacteria are internalized by alveolar macrophages (AMs) through phagocytosis. The "checkpoint model" can be used to describe the following immunological events post infection [2,3]. At the first checkpoint, to establish an infection, *M. tuberculosis* needs to evade elimination by AMs. *M. tuberculosis* has developed several strategies to manipulate the host immune response to extend its survival in the phagocytes. The pathogen can arrest maturation of the phagolysosome and direct phagocytes to necrosis, which is prerequisite for the bacterium to spread. In some individuals, referred to as *early clearers*, the pathogen is successfully cleared by innate immune mechanisms at this initial stage of infection [4–6]. If the pathogen is replicating and cannot be cleared by the innate immune system, the second checkpoint is reached; activation of the adaptive immune system. When the infection is controlled by the adaptive immune system, asymptomatic latent TB infection (LTBI) has developed. From this point, there is a 5-10% lifetime risk of progression to active TB (ATB), as a result of inadequate immune control, which is the third checkpoint for *M. tuberculosis* [7–10]. The Interferon-Gamma Release Assay (IGRA) is an immunological test used to confirm whether a person has been exposed to *M. tuberculosis* based on peripheral T cell release of interferon-γ (IFN-γ) in response to *M. tuberculosis* antigens [11]. Another method to confirm exposure is the purified protein derivative (PPD) or Tuberculin skin test, where PPD-antigen

is injected intradermally to the forearm and the skin reaction, based on T cell-mediated delayed hypersensitivity response against the antigen, is measured [12].

The role of epigenetics in TB immune responses is emerging. Several studies have described the concept of developing trained immunity through epigenetic reprogramming, leading to proper orchestration of gene expression upon re-exposure to a pathogen or pathogen-derived products [3,13,14]. DNA methylation, histone modifications, and regulation of small RNAs are considered as the important players in the regulation of epigenetic modifications [15]. DNA methylations in CpG-rich promotor regions are generally associated with gene silencing [16]. We and others have described the reprogramming of DNA methylation patterns in peripheral blood mononuclear cells (PBMCs) after exposure to live attenuated *Mycobacterium bovis* through the *Bacillus Calmette Guérin* (BCG) vaccination [17,18]. Further, we have demonstrated that differences in these DNA methylation patterns affect the efficacy of the macrophages to kill *M. tuberculosis in vitro* [19]. A wealth of literature focus on adaptive immune responses in peripheral blood, but since *M. tuberculosis* primarily infects the lung, it would be more relevant to address pulmonary immunity, including alveolar T cells and AMs [20,21].

Here, we have increased our focus on the pulmonary-resident immune cells and further investigated DNA methylomes of AMs and alveolar T cells in a cohort of medical students with a previously reported increased risk of *M. tuberculosis* exposure [22]. The aim of the study was to investigate epigenetic modifications in the pulmonary immune pre- and post- TB exposure in a natural setting. We hypothesized that this recent exposure to *M. tuberculosis* in the lung compartment would induce epigenetic alterations in AMs and alveolar T cells. By sputum induction AMs and alveolar T cells were isolated [23]. Using reduced representation of bisulfite sequencing (RRBS) the DNA methylome was investigated and we identified a different global DNA methylation profile in pulmonary immune cells of the study subjects

that developed latent TB infection during the study. Notably, this global DNA methylation profile was identified in the immune cells already before a latent TB infection could be detected with the IGRA test.

Results

Study design and cohort

Medical students were invited to participate in the study and donated sputum samples before (referred here as 0 months) and after (referred here as 6 months) clinical rotations in departments with a high-risk of *M. tuberculosis* exposure. A schematic overview of the study design is represented in Figure 1. Demographic information of the study subjects and IGRA results are shown in Table 1. One study subject was borderline IGRA-positive [24] ($IGRA_{pos}$) already at inclusion (0 months) (Table S1). Two study subjects developed latent TB infection during the study as demonstrated with a positive IGRA test at follow-up (6 months, referred to as IGRA converters in the following text) (Table S1).

DNA methylation patterns in alveolar macrophages and T cells before and after TB exposure distinguish IGRA converters

To investigate the epigenetic changes in the pulmonary immune cells over time, we isolated DNA from the AMs and alveolar T cells collected at 0 and 6 months and performed Genome-wide Reduced Representation Bisulfite sequencing (RRBS). After filtering the data as described in the method section, we identified a total of 1186 CpG-sites with ≥5 reads in the 19 samples from the AMs and 404 CpG-sites in the 19 samples from the alveolar T cells. To get an overview of the global DNA methylation distribution in the pulmonary immune cells, we made density plots of the M-values (the $\log_2$ ratio of the intensities of methylated versus unmethylated CpG-sites) obtained from CpG-sites in the AMs (1186 CpG-sites) and the

alveolar T cells (404 CpG-sites), shown in Figure 2A,B. The density plot showed a homogenous distribution of DNA methylation in all samples from the $IGRA_{neg}$ study subjects collected at both at 0 and 6 months. The sample from one study subject, that was $IGRA_{pos}$ at inclusion, followed the same global DNA methylation distribution. Whereas in the IGRA-converting study subjects, we identified a different global DNA methylation profile with more hyper- and hypomethylated CpG-sites as well as different peak densities. Notably, the sample collected at 6 months (after IGRA conversion) displayed a similar DNA methylation profile as the sample collected at inclusion.

The results coincided in both the AMs and the alveolar T cells. To further explore the data, we performed a principal component analysis (PCA) which revealed a distinct group formation from the IGRA converters (based on the PC1 (Dim1) with 70% C.I.), including both the samples collected before and after the IGRA conversion (Figure 2C,D). The $IGRA_{pos}$ individual on the other hand clustered with the data obtained from $IGRA_{neg}$ study subjects. We proceeded with a hierarchical clustering analysis by applying the Euclidean similarity/dissimilarity matrix calculation using the M-value (Figure 3A,B). In line with the PCA results, the data from the IGRA converters formed a separate cluster, including both the sample collected before and after IGRA conversion. The same cluster separation was found in both the samples from the AMs and from the alveolar T cells. To visualize the methylation status in the 1186 and 404 CpG-sites from the AMs and T cells respectively, we created heatmaps (Figure 4A,B), demonstrating a clear difference in the overall methylation between IGRA converters and $IGRA_{neg}$ study subjects.

High number of strong DMCs identified in IGRA-converting individuals

To further explore the epigenetic reprogramming that occurred between the two time points, 0 and 6 months, we identified the differentially methylated CpG-sites (DMCs) with the strict

147 filtering criteria (|$\log_2$ Fold Change ($\log_2$FC)| > 5 and Benjamini-Hochberg (BH) corrected *p*-
148 value < 0.01) in each study subject. The DMCs were divided into three different cutoff
149 levels: |$\log_2$FC| 10, 13 and 15 hyper- or hypomethylation in order to understand the
150 distribution of DMCs based on the level of change (Figure 5A,B). A high number of DMCs
151 with a |$\log_2$FC| >15 was identified in the IGRA converters, in both cell types. To normalize
152 the results to account for the different number of total DMCs (|$\log_2$FC| > 5, BH adj. *p*-value <
153 0.01) identified we looked at the percentage of DMCs with a |$\log_2$FC| > 15 (Figure 5C).
154
155 ***IGRA-converting individuals undergo similar epigenetic changes in DNA methylation***
156 The DMCs were annotated to the official Gene Symbols further analysis and is referred to as
157 DMGs in the following text. To identify the overlap in the DMGs with a |$\log_2$FC| >15
158 between the study subjects, a Venn analysis was performed and presented in an UpSet plot.
159 The IGRA converters shared the largest intersection in both cell types, 452 DMGs in the AMs
160 (Figure 6A) and 471 DMGs in the T cells (Figure S1). 32 DMGs overlapped between the
161 AMs and T cells. To filter out unspecific changes we selected the DMGs that changed in the
162 same direction in both IGRA converters. We identified 280 (128 DMGs were
163 hypermethylated and 152 hypomethylated) and 281 (159 DMGs were hypermethylated and
164 122 hypomethylated) DMGs that became hypo- or hypermethylated, in both IGRA converters
165 in the AMs and in the T cells respectively.
166
167 ***IGRA converters' DMGs are over-represented in pathways related to metabolic***
168 ***reprograming, T cell migration and IFN-γproduction.***
169 In an over-representation analysis (ORA) [25] using the PANTHER database [26] the identified
170 DMGs from the AMs unique to the IGRA converters were shown to be overrepresented in the
171 Pentose phosphate pathway and the Ras pathway (Table 2). In the ORA of the DMGs from

the T cells unique to the IGRA converters we identified muscarinic acetylcholine receptors 1-4 pathway and β1- and β2 adrenergic receptor signaling pathway (Table 3).

Discussion

In this study we investigated the genome-wide DNA methylation in the pulmonary immune cells in individuals at high risk of exposure to *M. tuberculosis*. DNA methylation has been shown to play an important role in the innate immune systems response to mycobacteria by us and others [17,27–30], however, it has not yet been longitudinally studied in healthy individuals at risk of TB exposure. Two of the study subjects in this study developed latent TB in the course of sample collection as demonstrated by IGRA conversion and our analyses identified a distinct genome-wide DNA methylation profile separating these two individuals from those who remained $IGRA_{neg}$ throughout the study. This DNA methylation profile, which, contrary to our expectation, was present already at inclusion, was characterized by more hypo- and hypermethylated CpG-sites. Intriguingly, the subject diagnosed with latent TB at inclusion, did not cluster with the IGRA converters. From these observations, we propose the following hypotheses: i) the IGRA converters were exposed to TB via the airways already before inclusion and the exposure had not yet developed into circulating T cell memory (resulting in a negative IGRA test) but had caused reprogramming of the DNA methylome of AMs; ii) a different, *inherent* genomic DNA methylation profile as observed in the IGRA converters predispose for IGRA conversion in a high-endemic setting.

According the first hypothesis, the observed DNA methylation pattern represents a normal, protective response to TB exposure (as part of the first 'checkpoint'), which however was breached by the infection and therefore urged the host to induce an adaptive immune response (the second 'checkpoint') [2,3]. If the second hypothesis applies, a limited capacity for inducing a host-protective epigenetic reprogramming could cause susceptibility and predisposition for

197 latent TB infections. There is a large heterogeneity in the human susceptibility to develop
198 latent or active TB upon exposure and there is limited knowledge in the determining factors
199 [31], but larger studies could investigate the correlation between DNA methylation patterns and
200 TB susceptibility. Alternatively, the distinct DNA methylation pattern observed could also be
201 affected by *M. tuberculosis* manipulating the host cells epigenome [28,32,33].
202 The two IGRA converters underwent similar epigenetic changes between the timepoints,
203 during conversion to IGRA positivity. There was a large overlap in DMGs in the IGRA
204 converters in both cell types. The DMG intersect unique to the IGRA converters AMs showed
205 over-representation in the pentose phosphate pathway which has previously been described to
206 be critical in *M. tuberculosis* infected macrophages [34,35]. Infected macrophages become
207 dependent on both glycolysis and the pentose phosphate pathway to fulfill the metabolic and
208 bioenergetic requirements for production of cytokines, chemokines and reactive oxygen
209 species (ROS) [34–36]. The DMGs unique to the IGRA converters' T cells showed over-
210 representation in the pathway for β-adrenergic and muscarinic acetylcholine receptor
211 signaling. A functional role between β2 adrenergic receptor (β2AR) activation and IFN-γ and
212 Tumor necrosis factor (TNF) production in T helper 1 (Th1) cells have been established in
213 several studies, activation of β2AR by the neurotransmitter norepinephrine (NE) decreases the
214 Th1 cells IFN-γ expression [37–40]. Polymorphisms in the β2AR have also been reported to be
215 associated with TB [41]. The cholinergic system also has a functional role in the T cells. T cells
216 express both muscarinic and nicotinic acetylcholine (ACh) receptors and expression of the
217 enzyme choline acetyltransferase (ChAT) is induced in T cells during infection, T cell derived
218 ACh is involved in T cells migration to tissues [42–44].
219 One study subject was borderline IGRA positive at inclusion and this subject followed the
220 same DNA methylation distribution and was clustering with the $IGRA_{neg}$ study subjects. The
221 kinetics of the IFN-γ-reaction in relation to point of infection is not elucidated and after TB

222 treatment the peripheral T cell memory will persist for up to 15 months [45,46]. According to our
223 analysis, the epigenome of the pulmonary immune cells in this individual was more similar to
224 that of healthy individuals and this suggests the possibility that this individual became IGRA
225 positive a long time ago and that the epigenome profile in the pulmonary immune cells has
226 returned to baseline.
227
228 If the DNA methylation profile observed in the IGRA converters is an early result of *M.*
229 *tuberculosis* infection, the finding could have implications as a diagnostic tool for
230 identification of individuals that are developing latent TB infection before any of the currently
231 available diagnostic methods. However, if the patterns we identified is not a result of bacterial
232 infection but rather an epigenomic dysregulation predisposing individuals to convert upon
233 exposure, this global DNA methylation pattern could have implications in identifying risk-
234 groups who are more prone to convert upon exposure.
235
236 We recognize that this study was performed on a small number of study subjects and further
237 investigation in larger cohorts is needed to elucidate the results presented here. We also
238 acknowledge the limitations of the study design, we included a cohort of study subjects with
239 risk of *M. tuberculosis* exposure, but no definite measurement of exposure. The observations
240 made with IGRA converters were limited to two study subjects. However, we highlight the
241 significance of investigating pulmonary immune cells and the value of investigating DNA
242 methylomes in the same study subjects, longitudinally, before and after IGRA conversion.

Methods

Ethical Statement

Ethical approval was obtained from Universidad Peruana Cayetano Heredia (UPCH) Institutional Review Board Nº 103793. All participants signed an informed consent.

Study cohort and design

This was a prospective study aimed to investigate epigenetic DNA methylome alterations in pulmonary immune cells pre- and post- *M. tuberculosis* exposure in a natural setting. We enrolled medical students *(n=15)* in the fifth or sixth year of medical school at UPCH. The students donated sputum and blood samples before (0 months) and after (6 months) they had clinical rotations in high-risk departments of *M. tuberculosis* exposure. The department of infectious disease and internal medicine including the emergency department were defined as departments with high-risk of *M. tuberculosis* exposure. Sputum was used to isolate immune cells from the lung and the blood was used for IGRA. At inclusion, the participants filled in an individual case report form with demographic information. Participants also filled in an online questionnaire to collect background information before each session at 0 and 6 moths. The questionnaire was created in the tool Survey & Report, provided by Linköping University. Samples from 10 individuals (collected at 0 and 6 months for $IGRA_{neg}$ (*n=7*) and IGRA converters (*n=2*) and at 0 months for $IGRA_{pos}$ (*n=1*)) resulted in 19 samples that was selected for each cell type (AMs and alveolar T cells), for the sequencing analysis. The sample selection was based on results in the online questionnaire and on sample quality with regards to DNA concentration in the obtained in both AMs and T cells at the two sample collections at 0 and 6 months.

Interferon gamma releasing assay with QuantiFERON® TB-Gold Plus

At inclusion and follow-up, the participants donated blood samples that were used for IGRA with the QuantiFERON® TB-Gold Plus test (SSI Diagnostica, Hillerød, Denmark) according to the manufacturer's instructions. Tubes were filled and incubated at 37°C for maximum 24 hours and then analyzed with ELISA.

Sputum Induction

The sputum induction [23] was performed with an eFlow rapid nebulizer (PARI, Hamburg, Germany) filled with a hypertonic saline solution. The solution was prepared by mixing sterile water (Fresenius Kabi, Stockholm, Sweden) with 4% sodium chloride (B. Braun Medical AB, Stockholm, Sweden). The participants inhaled the solution for 9 minutes while simultaneously preforming breathing exercises in accordance with instructions from the lung clinic at Linköping University Hospital. The participants were asked to cough deeply to expectorate sputum from the lungs. The expectorates were collected into sterile 50 ml Falcon tubes (Thermo Fisher Scientific, Waltham, US) after each session. The Falcon tubes were kept on ice. The sputum inductions were performed in three replicates.

Sputum Processing and CD3 and HLA-DR Positive Cell Isolation

From the sputum samples, within two hours, plugs containing pulmonary immune cells [23] were picked and pooled. The plugs were dissolved by adding 0.1% dithiothreitol (DTT) (Thermo Fisher Scientific) mixed in phosphate-buffered saline (PBS) (Gibco, Cambridge, UK). This was added in a volume approximately 4 times the collected sputum volume and then vortexed and placed on a tilter with ice for 20 min. The dissolved sputum sample was then filtered through 50 µm cell strainers (Sigma-Aldrich, Saint Louise, US) into a new 50 ml Falcon tube and centrifuged for 5 min at 380g in 4°C. The supernatant was discarded, and the pellet was resuspended in 500 µl of an isolation buffer containing 500 mM

ethylenediaminetetraacetic acid (EDTA), 0.1% fetal bovine serum (FBS), Ca^{2+} and Mg^{2+} free PBS (pH 7.4). First, CD3 positive cells were isolated from the resuspended pellet. 25 µl of Dynabeads CD3 (Invitrogen Dynabeads®, Life Technologies AS, Oslo, Norway) were washed with 800 µl PBS and placed in a DynaMag-2 (Thermo Fisher Scientific) for 1 min. The supernatant was discarded and 800 µl isolation buffer were added two times for additional washing. The beads were then mixed with 500 µl of the cell sample and incubated for 30 min at 4°C while tilting. After incubation, the tube was placed in the DynaMag-2, supernatant was removed to be used to isolate HLA-DR positive cells next. The CD3 positive cells were resuspended in 200 µl PBS.

Secondly, HLA-DR positive cells were isolated. 25 µl Magnetic Pan Mouse IgG Dynabeads™ (cat no: 11041, Thermo Fisher) were washed with 800 µl PBS and placed in a DynaMag-2 (Thermo Fisher Scientific) for 1 min. The supernatant was discarded and 800 µl isolation buffer were added two times for additional washing. 5 µl monoclonal HLA-DR antibodies (cat no: 14-9956-82, Thermo Fisher) were added to conjugate the Dynabeads. Tubes were incubated for 40 min, then placed in the DynaMag-2 for 30 seconds and supernatant removed. 800 µl isolation buffer were added two times for washing. The tube was placed in DynaMag-2 for 30 seconds and supernatant removed. The beads were then mixed with 500 µl of the supernatant from the CD3 isolation and incubated for 30 min at 4°C while tilting. After incubation, the tube was placed in the DynaMag-2, supernatant was discarded, and HLA-DR positive cells were resuspended in 200 µl PBS. CD3 positive cells are referred to as alveolar T cells (T cells) and HLA-DR positive cells are referred to as alveolar macrophages (AMs).

DNA Extraction and Quantification

317 DNA was extracted from the AMs and T cells with the AllPrep® DNA/RNA Mini Kit
318 (Qiagen, Hilden, Germany) within 4 hours from cell isolation. Concentration of DNA was
319 quantified with a Qubit® 4.0 Fluorometer (Thermo Fisher Scientific), using dsDNA High
320 Sensitivity (HS) Assay Kit (Thermo Fisher Scientific). The measurement was performed
321 according to the manufacturer's instructions.
322
323 *Reduced Representation Bisulfite Sequencing of DNA from AMs and T cells*
324 DNA samples were sequenced with Reduced Representation Bisulfide Sequencing (RRBS) at
325 the Bioinformatics and Expression Analysis (BEA) core facility at Karolinska Institute (KI)
326 with Diagenode´s RRBS. The DNA was enzymatically digested, bisulfite converted, and PCR
327 amplified before ready for Illumina´s HiSeq 2000. The use of the restriction enzyme MspI,
328 which cleaves CCGG from the 5´end, results in shorter sequences to analyze and is therefore
329 cost-effective.
330
331 *Data processing and computational analysis*
332 The RAW files (in fastq format) were generated from the RRBS analysis, were quality
333 checked using the fastQC [47] (v0.11.9). The sequences were trimmed to remove artificially
334 filled-in cytosines at the 3′ end using the TrimGalore (v.0.6.5) [48] with a phred score cutoff of
335 20 and quality checked again after trimming. The trimmed sequences were aligned with the
336 human reference genome (hg38.13) using Bowtie2 [49] and removed the deduplicates using the
337 Bismark v.0.22.3 [50]. The methylation extractor from Bismark was used to extract the CpG
338 methylation data from the sequences. The SAMtools (v1.7) [51] package was used to sort the
339 bam files on CpG-site chromosomal location and converted to SAM files. The methylated and
340 unmethylated CpG counts were extracted and combined using the DMRfinder (v0.3) [52]
341 package in R (v4.0.2) [53].

To read the Bismark coverage files, the *edgeR* (v3.32.1) [54,55] package was used. The CpG-sites located in the X and Y chromosome as well as CpG-sites from mitochondrial DNA were filtered out. CpG-sites with a read coverage < 5 with both methylated and unmethylated reads were removed from the analysis. The M-values were calculated using the $\log_2$ ratio of the intensities of methylated verses unmethylated CpG-sites (with addition of +2 to each count to avoid logarithms of zeros) [56].

$$M_i = log_2\left(\frac{\max(y_{i,methylation,}0)+\alpha}{\max(y_{i,unmethylation,}0)+\alpha}\right),\ \alpha=2$$

The CpG-sites were annotated using *org.Hs.eg.db* (v3.12) [57] and AnnotationDbi (v1.52) [58] packages. After filtering and annotating the data, we identified a total of 1186 CpG-sites in the AMs and 404 CpG-sites from T cells that were covered in all samples.

Statistical analysis

The Principal Component Analysis (PCA) was calculated using the *factoExtra* (v1.0.7) [59] and *factoMineR* (v2.4)[60]. The hierarchical clusters was estimated using the *ape* (v5.4-1) [61] and *dendextend* (v1.14.0)[62] by calculating the Euclidian similarity/dissimilarity matrix. For the identification of the individual differentially methylated CpG-sites (DMCs) the EdgeR *lmFit* function [63] was used to estimate the counts of the unmethylated (Un) and methylated (Me) reads in the conditions 0 months (A) and 6 months (B) in each individual, calculated as:

$$\beta_A = log_2\left(\frac{MeA+\alpha}{UnA+\alpha}\right),\ \alpha = 0.125$$

$$\beta_B = log_2\left(\frac{MeB+\alpha}{UnB+\alpha}\right),\ \alpha = 0.125$$

DMCs were defined as CpG-sites with a $|\log_2 FC| > 5$ and significant with the Benjamini-Hochberg (BH) corrected *p*-value < 0.01. The lollipop plots and the upset plots were made using the *ggpubr* (v0.4.0) [64] and the *UpSetR* (v1.4.0) [65,66] packages. Two $IGRA_{neg}$ study subjects and 1 study subject had no DMCs with a $|\log_2 FC| > 15$ in the AMs and T cells

366 respectively and were therefore excluded from the analysis. The pathway analysis was
367 performed with the DMGs with a $|\log_2 FC| > 15$ and Benjamini-Hochberg (BH) corrected *p*-
368 value < 0.01 unique to the IGRA converters AMs and T cells respectively from the
369 PANTHER database (v16.0) [26] using the WEB-based Gene SeT AnaLysis Toolkit
370 (WebGestalt) webserver (v2019) [25]. The false discovery rate (FDR) in the pathway analysis is
371 BH corrected *p*-values. The significant pathways identified in the AMs and the top 4 most
372 significant pathways identified in the T cells are presented in Table 2.

373 **Data availability**

374 The datasets will be made available at the final publication.

375

376 **Acknowledgements**

377 We thank the Bioinformatics and Expression Analysis core facility (BEA) at Karolinska
378 Institute that preformed the sequencing analysis. We thank the Swedish National
379 Infrastructure for Computing (SNIC) at National Supercomputing Centre (NSC), Linköping
380 University for the computing systems enabling the data handling, partially funded by the
381 Swedish Research Council through grant agreement Nº 2018-05973. The work was supported
382 by grants from the Swedish Research Council (Vetenskapsrådet) Nº 2018-05973 and Nº 2018-
383 04246 and the Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica
384 CONCYTEC and Cienciactiva Nº 106-2018-FONDECYT. J.D is a postdoctoral fellow
385 supported through the Medical Infection and Inflammation Center (MIIC) at Linköping
386 University.

387

388 **Author Contributions**

389 N.I, J.P, I.P, C.U-G, M.M-A and M.L wrote the ethical application and conceived the study.
390 I.P, M.N and A.T performed all sample collections. J.D and L.K preformed the bioinformatic
391 analyses. M.L, L.K, J.D and I.P wrote the manuscript. M.M-A and C.U-G supervised the
392 project in Peru.

393

394 **Consent for publication**

395 All authors have approved the final version of the manuscript.

396

397 **Competing Interests**

398 The authors declare no competing interests.

References

1. World Health Organization. *Global Tuberculosis Report*. https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf (2019).
2. Schön, T., Lerm, M. & Stendahl, O. Shortening the 'short-course' therapy- insights into host immunity may contribute to new treatment strategies for tuberculosis. *J. Intern. Med.* **273**, 368–382 (2013).
3. Lerm, M. & Netea, M. G. Trained immunity: A new avenue for tuberculosis vaccine development. *Journal of Internal Medicine* (2016) doi:10.1111/joim.12449.
4. Verrall, A. J. *et al.* Early clearance of Mycobacterium tuberculosis is associated with increased innate immune responses . *J. Infect. Dis.* (2019) doi:10.1093/infdis/jiz147.
5. Verrall, A. J. *et al.* Early Clearance of Mycobacterium tuberculosis: The INFECT Case Contact Cohort Study in Indonesia. *J. Infect. Dis.* (2019) doi:10.1093/infdis/jiz168.
6. Verrall, A. J., G. Netea, M., Alisjahbana, B., Hill, P. C. & van Crevel, R. Early clearance of Mycobacterium tuberculosis: A new frontier in prevention. *Immunology* (2014) doi:10.1111/imm.12223.
7. O'Garra, A. *et al.* The Immune Response in Tuberculosis. *Annu. Rev. Immunol.* (2013) doi:10.1146/annurev-immunol-032712-095939.
8. Cooper, A. M. Cell-Mediated Immune Responses in Tuberculosis. *Annu. Rev. Immunol.* **27**, 393–422 (2009).
9. Lerner, T. R., Borel, S. & Gutierrez, M. G. The innate immune response in human tuberculosis. *Cellular Microbiology* (2015) doi:10.1111/cmi.12480.
10. Pieters, J. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. *Cell Host and Microbe* vol. 3 399–407 (2008).
11. Pai, M. *et al.* Gamma interferon release assays for detection of Mycobacterium

tuberculosis infection. *Clin. Microbiol. Rev.* (2014) doi:10.1128/CMR.00034-13.

12. LENDRUM, J. D. Tuberculin skin testing. *Med. World (New. York).* (1953) doi:10.1378/chest.76.6_supplement.764.

13. Netea, M. G. *et al.* Trained immunity: A program of innate immune memory in health and disease. *Science* (2016) doi:10.1126/science.aaf1098.

14. Seeley, J. J. *et al.* Induction of innate immune memory via microRNA targeting of chromatin remodelling factors. *Nature* (2018) doi:10.1038/s41586-018-0253-5.

15. Villota-Salazar, N. A., Mendoza-Mendoza, A. & González-Prieto, J. M. Epigenetics: from the past to the present. *Front. Life Sci.* (2016) doi:10.1080/21553769.2016.1249033.

16. Bock, C. *et al.* DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Mol. Cell* **47**, 633–647 (2012).

17. Verma, D. *et al.* Anti-mycobacterial activity correlates with altered DNA methylation pattern in immune cells from BCG-vaccinated subjects. *Sci. Rep.* **7**, 1–10 (2017).

18. Hasso-Agopsowicz, M., Scriba, T. J., Hanekom, W. A., Dockrell, H. M. & Smith, S. G. Differential DNA methylation of potassium channel KCa3.1 and immune signalling pathways is associated with infant immune responses following BCG vaccination. *Sci. Rep.* (2018) doi:10.1038/s41598-018-31537-9.

19. Das, J., Verma, D., Gustafsson, M. & Lerm, M. Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naïve subjects. *Epigenetics* **14**, 589–601 (2019).

20. Yona, S., Viukov, S., Guilliams, M. & Misharin, A. Tissue Macrophages Under Homeostasis. *Immunity* **38**, 79–91 (2013).

21. Yao, Y. *et al.* Induction of Autonomous Memory Alveolar Macrophages Requires T Cell Help and Is Critical to Trained Immunity. *Cell* (2018)

449 doi:10.1016/j.cell.2018.09.042.

450 22. Pérez-Lu, J. E., Cárcamo, C. P., García, P. J., Bussalleu, A. & Bernabé-Ortiz, A.
451 Tuberculin skin test conversion among health sciences students: A retrospective cohort
452 study. *Tuberculosis* (2013) doi:10.1016/j.tube.2012.10.001.

453 23. Das, J., Idh, N., Sikkeland, L. I. B., Paues, J. & Lerm, M. DNA methylome-based
454 validation of induced sputum as an effective protocol to study lung immunity:
455 construction of a classifier of pulmonary cell types. *bioRxiv* (2021)
456 doi:10.1101/2021.03.12.435086.

457 24. Jonsson, J. *et al.* A borderline range for Quantiferon Gold In-Tube results. *PLoS One*
458 (2017) doi:10.1371/journal.pone.0187313.

459 25. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set
460 analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* (2019)
461 doi:10.1093/nar/gkz401.

462 26. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed
463 by function. *Genome Res.* (2003) doi:10.1101/gr.772403.

464 27. Wei, M. *et al.* NLRP3 Activation Was Regulated by DNA Methylation Modification
465 during Mycobacterium tuberculosis Infection. *Biomed Res. Int.* (2016)
466 doi:10.1155/2016/4323281.

467 28. Sharma, G. *et al.* Genome-wide non-CpG methylation of the host genome during M.
468 tuberculosis infection. *Sci. Rep.* (2016) doi:10.1038/srep25006.

469 29. Zheng, L. *et al.* Unraveling methylation changes of host macrophages in
470 Mycobacterium tuberculosis infection. *Tuberculosis* (2016)
471 doi:10.1016/j.tube.2016.03.003.

472 30. DiNardo, A. R. *et al.* DNA hypermethylation during tuberculosis dampens host
473 immune responsiveness. *J. Clin. Invest.* (2020) doi:10.1172/JCI134622.

474 31. Cadena, A. M., Fortune, S. M. & Flynn, J. L. Heterogeneity in tuberculosis. *Nature*
475 *Reviews Immunology* (2017) doi:10.1038/nri.2017.69.
476 32. Sharma, G., Upadhyay, S., Srilalitha, M., Nandicoori, V. K. & Khosla, S. The
477 interaction of mycobacterial protein Rv2966c with host chromatin is mediated through
478 non-CpG methylation and histone H3/H4 binding. *Nucleic Acids Res.* (2015)
479 doi:10.1093/nar/gkv261.
480 33. Yaseen, I., Kaur, P., Nandicoori, V. K. & Khosla, S. Mycobacteria modulate host
481 epigenetic machinery by Rv1988 methylation of a non-tail arginine of histone H3. *Nat.*
482 *Commun.* (2015) doi:10.1038/ncomms9922.
483 34. Kumar, R. *et al.* Immunometabolism of Phagocytes During Mycobacterium
484 tuberculosis Infection. *Frontiers in Molecular Biosciences* (2019)
485 doi:10.3389/fmolb.2019.00105.
486 35. Cumming, B. M., Addicott, K. W., Adamson, J. H. & Steyn, A. J. C. Mycobacterium
487 tuberculosis induces decelerated bioenergetic metabolism in human macrophages. *Elife*
488 (2018) doi:10.7554/eLife.39169.
489 36. Giri, P. K., Kruh, N. A., Dobos, K. M. & Schorey, J. S. Proteomic analysis identifies
490 highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate
491 protein-treated macrophages. *Proteomics* (2010) doi:10.1002/pmic.200900840.
492 37. Sanders, V. M. *et al.* Differential expression of the beta2-adrenergic receptor by Th1
493 and Th2 clones: implications for cytokine production and B cell help. *J. Immunol.*
494 (1997).
495 38. Borger, P. *et al.* β-Adrenoceptor-mediated inhibition of IFN-γ, IL-3, and GM-CSF
496 mRNA accumulation in activated human T lymphocytes is solely mediated by the β2-
497 Adrenoceptor subtype. *Am. J. Respir. Cell Mol. Biol.* (1998)
498 doi:10.1165/ajrcmb.19.3.2765.

499 39. Estrada, L. D., Ağaç, D. & Farrar, J. D. Sympathetic neural signaling via the β2-
500 adrenergic receptor suppresses T-cell receptor-mediated human and mouse CD8+ T-
501 cell effector function. *Eur. J. Immunol.* (2016) doi:10.1002/eji.201646395.
502 40. Swanson, M. A., Lee, W. T. & Sanders, V. M. IFN-γ Production by Th1 Cells
503 Generated from Naive CD4 + T Cells Exposed to Norepinephrine . *J. Immunol.* (2001)
504 doi:10.4049/jimmunol.166.1.232.
505 41. Azar, A. F., Jazani, N. H., Bazmani, A., Vahhabi, A. & Shahabi, S. Polymorphisms in
506 Beta-2 Adrenergic Receptor Gene and Association with Tuberculosis. *Lung* (2017)
507 doi:10.1007/s00408-016-9968-y.
508 42. Fujii, T. *et al.* Expression and function of the cholinergic system in immune cells.
509 *Frontiers in Immunology* (2017) doi:10.3389/fimmu.2017.01085.
510 43. Murray, K. *et al.* Neuroanatomy of the spleen: Mapping the relationship between
511 sympathetic neurons and lymphocytes. *PLoS One* (2017)
512 doi:10.1371/journal.pone.0182416.
513 44. Cox, M. A. *et al.* Choline acetyltransferase–expressing T cells are required to control
514 chronic viral infection. *Science (80-.).* (2019) doi:10.1126/science.aau9072.
515 45. Pollock, N. R. *et al.* Evaluation of the Effect of Treatment of Latent Tuberculosis
516 Infection on QuantiFERON-TB Gold Assay Results. *Infect. Control Hosp. Epidemiol.*
517 (2009) doi:10.1086/596606.
518 46. Dyrhol-Riise, A. M. *et al.* Diagnosis and follow-up of treatment of latent tuberculosis;
519 the utility of the QuantiFERON-TB Gold In-tube assay in outpatients from a
520 tuberculosis low-endemic country. *BMC Infect. Dis.* (2010) doi:10.1186/1471-2334-10-
521 57.
522 47. Andrews, S. FASTQC A Quality Control tool for High Throughput Sequence Data.
523 *Babraham Inst.* (2015).

524 48. Krueger, F. Trim Galore: a wrapper tool around gutadapt and FastQC to consistently
525 apply quality and adapter trimming to FastQ files. Available from
526 https://www.bioinformatics.babraham.ac.uk/projects. *Babraham Institute* (2015).
527 49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*
528 *Methods* (2012) doi:10.1038/nmeth.1923.
529 50. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for
530 Bisulfite-Seq applications. *Bioinformatics* (2011) doi:10.1093/bioinformatics/btr167.
531 51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
532 (2009) doi:10.1093/bioinformatics/btp352.
533 52. Gaspar, J. M. & Hart, R. P. DMRfinder: Efficiently identifying differentially
534 methylated regions from MethylC-seq data. *BMC Bioinformatics* (2017)
535 doi:10.1186/s12859-017-1909-0.
536 53. Rstudio, T. RStudio: Integrated Development for R. *Rstudio Team, PBC, Boston, MA*
537 *URL http://www.rstudio.com/* (2020) doi:10.1145/3132847.3132886.
538 54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for
539 differential expression analysis of digital gene expression data. *Bioinformatics* (2009)
540 doi:10.1093/bioinformatics/btp616.
541 55. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of
542 multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids*
543 *Res.* (2012) doi:10.1093/nar/gks042.
544 56. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying
545 methylation levels by microarray analysis. *BMC Bioinformatics* (2010)
546 doi:10.1186/1471-2105-11-587.
547 57. Marc Carlson. org.Mm.eg.db: Genome wide annotation for Mouse. R package version
548 3.8.2. *Bioconductor* (2019).

549 58. Pagès, H., Carlson, M., Falcon, S. & Maintainer, N. L. Package 'AnnotationDbi'.
550 *Bioconductor Packag. Maint.* (2017).
551 59. Kassambara, A. & Mundt, F. factoextra: Extract and Visualize the Results of
552 Multivariate Data Analyses. Package Version 1.0.7. *R Packag. version* (2020).
553 60. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J.*
554 *Stat. Softw.* (2008) doi:10.18637/jss.v025.i01.
555 61. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and
556 evolutionary analyses in R. *Bioinformatics* (2019) doi:10.1093/bioinformatics/bty633.
557 62. Galili, T. dendextend: An R package for visualizing, adjusting and comparing trees of
558 hierarchical clustering. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv428.
559 63. Smyth, G. K., Chen, Y., Pal, B. & Visvader, J. E. Differential methylation analysis of
560 reduced representation bisulfite sequencing experiments using edgeR. *F1000Research*
561 (2017) doi:10.12688/f1000research.13196.1.
562 64. Kassambara, A. Package 'ggpubr': 'ggplot2' Based Publication Ready Plots. *R Packag.*
563 *version 0.4.0* (2020).
564 65. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization
565 of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* (2014)
566 doi:10.1109/TVCG.2014.2346248.
567 66. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: An R package for the visualization
568 of intersecting sets and their properties. *Bioinformatics* (2017)
569 doi:10.1093/bioinformatics/btx364.
570

571 **Figure and Table legends**

572 Figure 1. **Flow chart of the study design.** Study subjects donated sputum and blood samples
573 at 0 and 6 months, which corresponded to before and after clinical rotations at departments
574 with high-risk of *M. tuberculosis* exposure at hospitals in Lima, Peru. IGRA tests were taken
575 to confirm TB-infection. AMs and alveolar T cells were isolated from sputum samples. DNA
576 was extracted and the samples were sequenced with reduced representation bisulfite
577 sequencing (RRBS) for methylation analysis.

578

579 Figure 2. **Unique distribution of genomic DNA methylation discriminates IGRA-**
580 **converting individuals.** Density plots of the distribution of M-values from CpG-sites
581 identified in **A.** AMs and **B.** alveolar T cells. Full line represents samples collected at 0
582 months and dotted lines samples collected at 6 months**.** The Principal Component Analysis
583 (PCA) of the methylation data from **C.** AMs and **D.** alveolar T cells. The ellipses represent the
584 70% confidence interval (C.I.) in the dataset.

585

586 Figure 3. **Hierarchical clustering analysis separating IGRA-converting individuals. A,B.**
587 Unsupervised hierarchical clustering dendrogram applying the Euclidian distance matrix
588 calculation and Ward D2 method. The dendrograms show the clustering of methylome data
589 from **A.** AMs. and **B.** alveolar T cells. The scale represents the Euclidean distance.

590

591 Figure 4. **Heatmaps reveal a different overall DNA methylation pattern in IGRA-**
592 **converting study subjects. A.** Heatmap of all CpG-sites (1186) identified in the AMs of each
593 study subject. **B.** Heatmap of all CpG-sites (404) identified from the alveolar T cells. The
594 color bar represents the M-value scale ranging from -10 (blue) to 10 (red).

595

Figure 5. **Profound DNA methylome alterations identified in IGRA-converting study subjects.** Lollipop plots showing the number of differentially methylated CpG-sites (DMCs) identified in the **A.** AMs and **B**. alveolar T cells of each study subject. C. The percentage of DMCs with a $|\log_2 FC| > 15$ compared to the total number of DMCs ($|\log_2 FC| > 5$, adj *p*-value > 0.01) identified in each study subject AMs and T cells.

Figure 6. **IGRA converters undergo unique DNA methylation changes during IGRA conversion.** UpSet plot showing the intersects of the DMGs identified in each study subjects AMs.

Figure S1. **UpSet plot showing the intersects of the DMGs identified in each study subject's alveolar T cells.** UpSet plot showing the intersects of the DMGs identified in each study subjects T cells.

Table 1. **Demographics of study subjects**. A total of 15 participants were included in the study. IGRA test confirmed IGRA conversion in two study subjects. shows the range of the data.

Characteristics	**Participants (*N=15*)**
Sex (male and female)	8 / 7
Age (years)	22.7 (21-29)
Weight (kg)	70.5 (48.5-101)
Height (cm)	168.8 (159-182)
BMI (kg/m^2)	24.6 (17.2-34.5)
IGRA result 0 months (negative and positive)	14 / 1
IGRA result 6 months (negative and positive)	12 / 3

Table 2. **ORA of DMG intersect from IGRA converter's AMs**

ORA using Panther pathways of the DMG intersect identified in the AMs of the IGRA converters.

Gene Set	**Description**	**Ratio**	***p*-value**	**FDR**
P02762	Pentose Phosphate Pathway	20.981	0.004	0.41
P04393	Ras Pathway	3.596	0.049	1

Table 3.**ORA of DMG intersect from IGRA converter's alveolar T cells**

ORA of the DMG intersect identified in the T cells of the IGRA converters.

Gene Set	**Description**	**Ratio**	***p*-value**	**FDR**

P00042	Muscarinic acetylcholine receptor 1 and 3 signaling pathway	6.34	0.003	0.35
P04377	Beta 1 adrenergic receptor signaling pathway	6.08	0.012	0.44
P04378	Beta 2 adrenergic receptor signaling pathway	6.08	0.012	0.44
P00043	Muscarinic acetylcholine receptor 2 and 4 signaling pathway	4.8	0.022	0.44

623

Figure 1

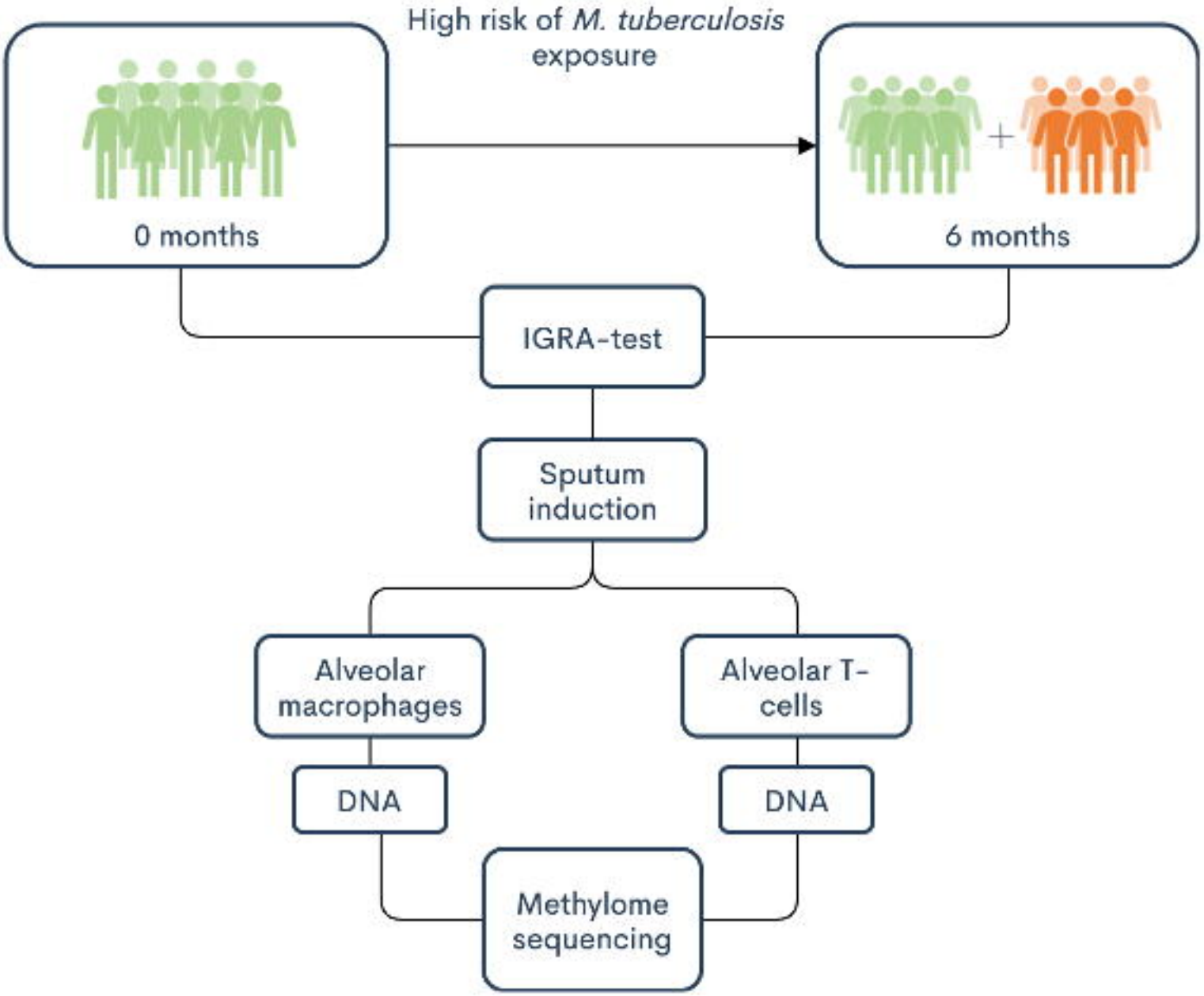
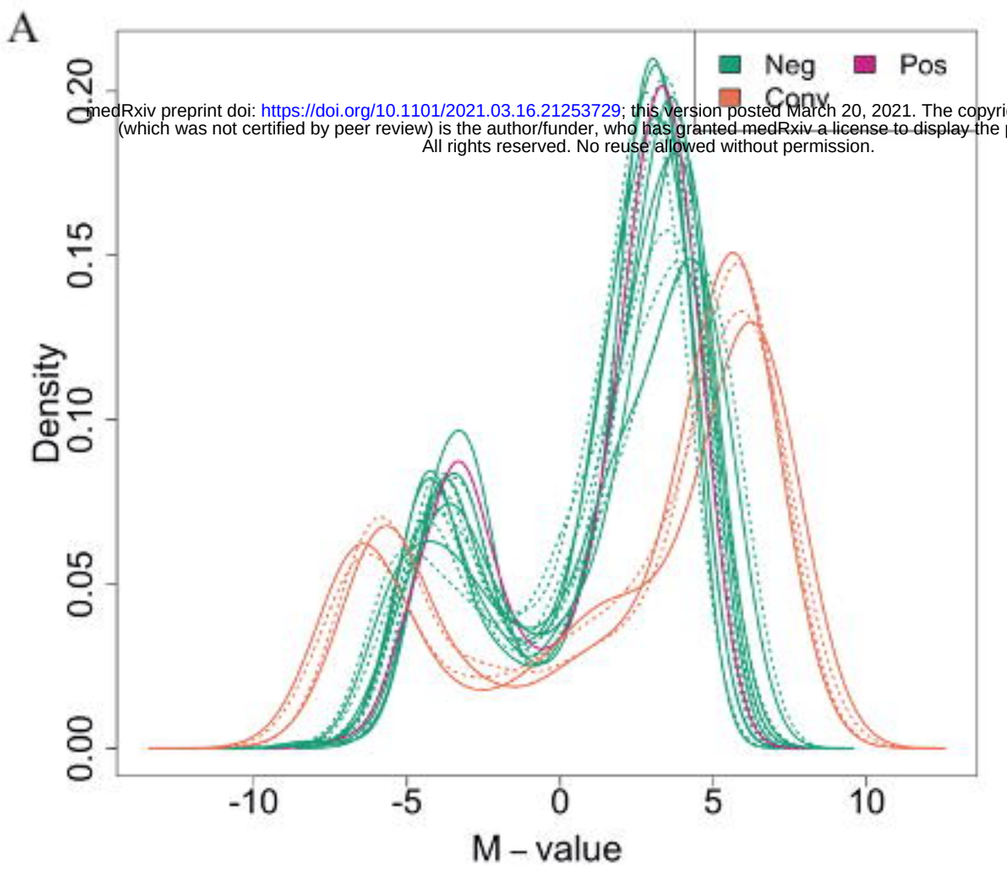
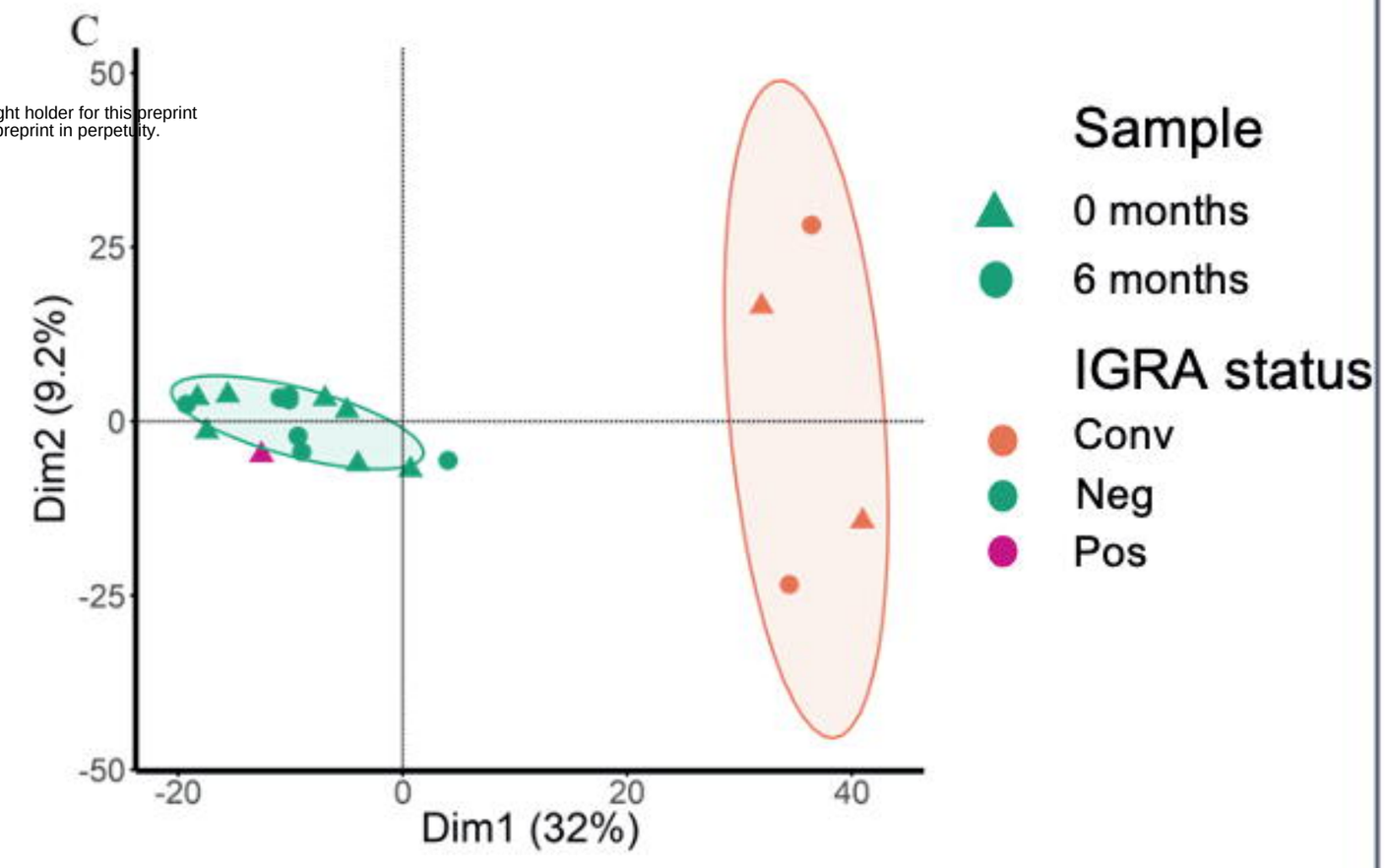
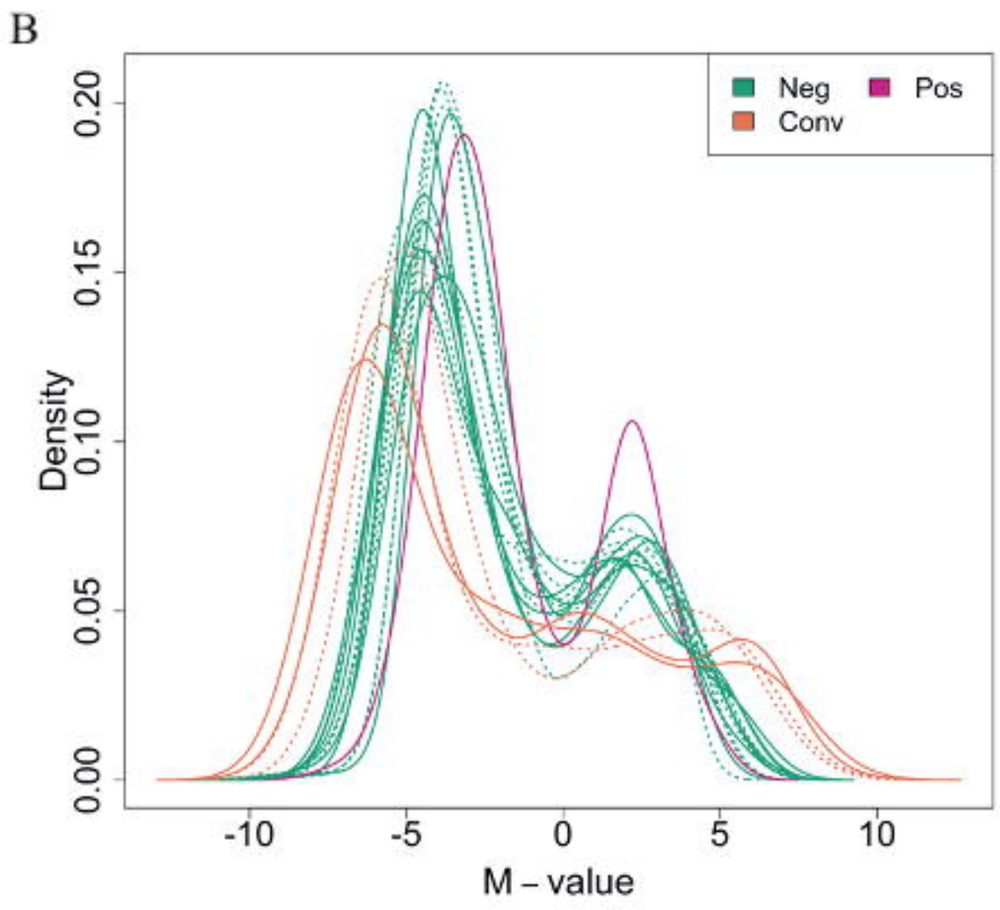

High risk of *M. tuberculosis* exposure
0 months
6 months
IGRA-test
Sputum induction
Alveolar macrophages
Alveolar T-cells
DNA
DNA
Methylome sequencing

Figure 2







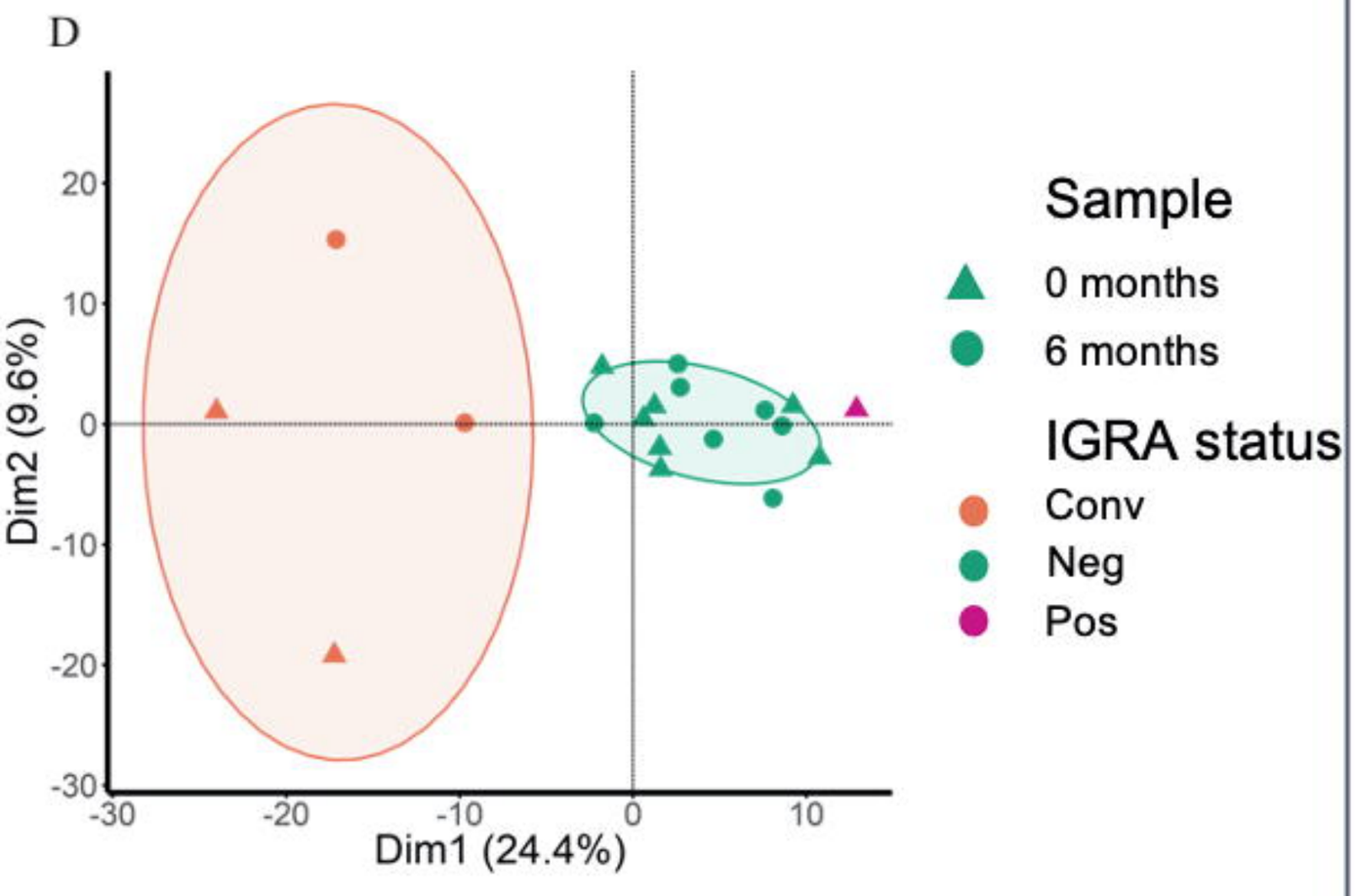


Figure 3

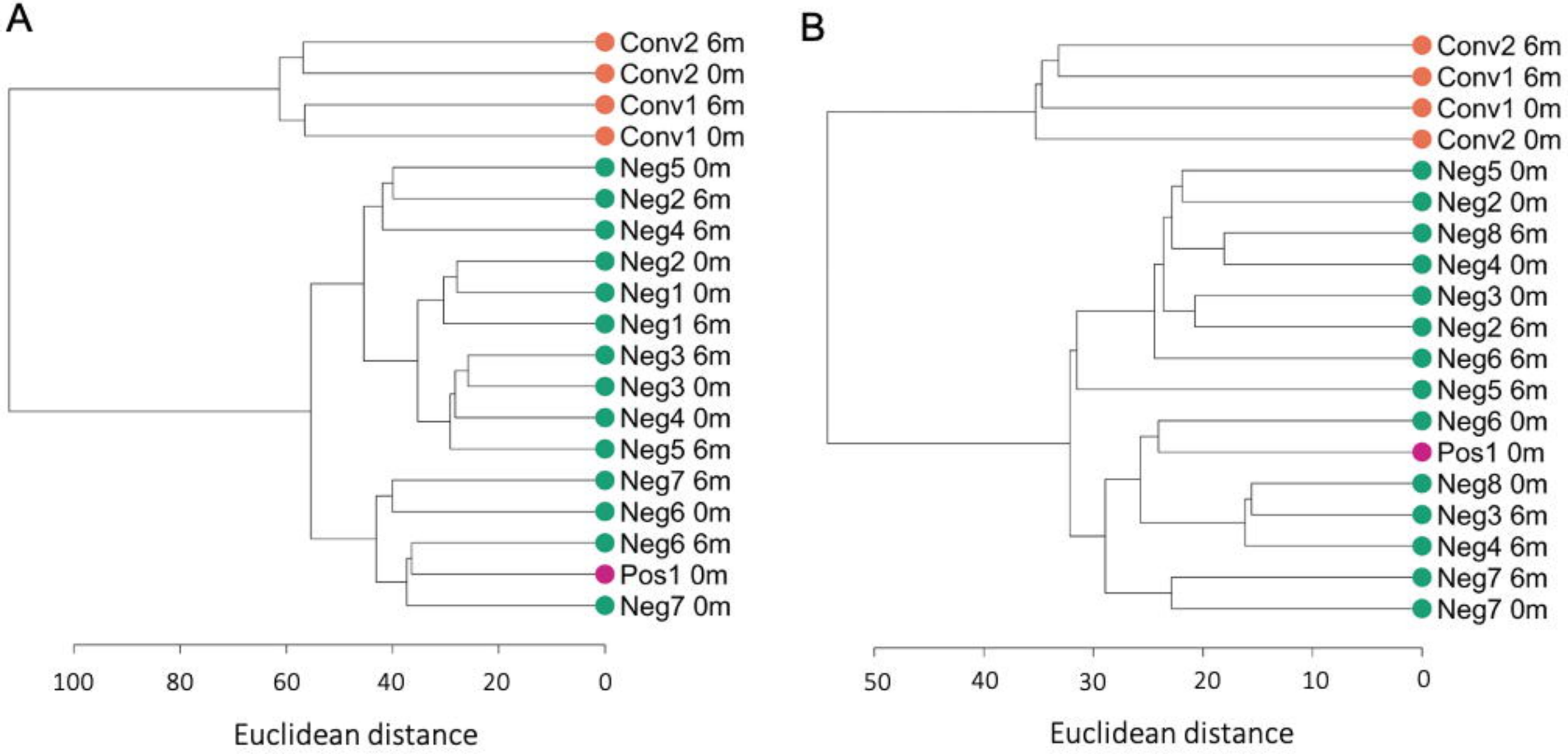

A
Conv2 6m
Conv2 0m
Conv1 6m
Conv1 0m
Neg5 0m
Neg2 6m
Neg4 6m
Neg2 0m
Neg1 0m
Neg1 6m
Neg3 6m
Neg3 0m
Neg4 0m
Neg5 6m
Neg7 6m
Neg6 0m
Neg6 6m
Pos1 0m
Neg7 0m
100
80
60
40
20
0
Euclidean distance
B
Conv2 6m
Conv1 6m
Conv1 0m
Conv2 0m
Neg5 0m
Neg2 0m
Neg8 6m
Neg4 0m
Neg3 0m
Neg2 6m
Neg6 6m
Neg5 6m
Neg6 0m
Pos1 0m
Neg8 0m
Neg3 6m
Neg4 6m
Neg7 6m
Neg7 0m
50
40
30
20
10
0
Euclidean distance

Figure 4

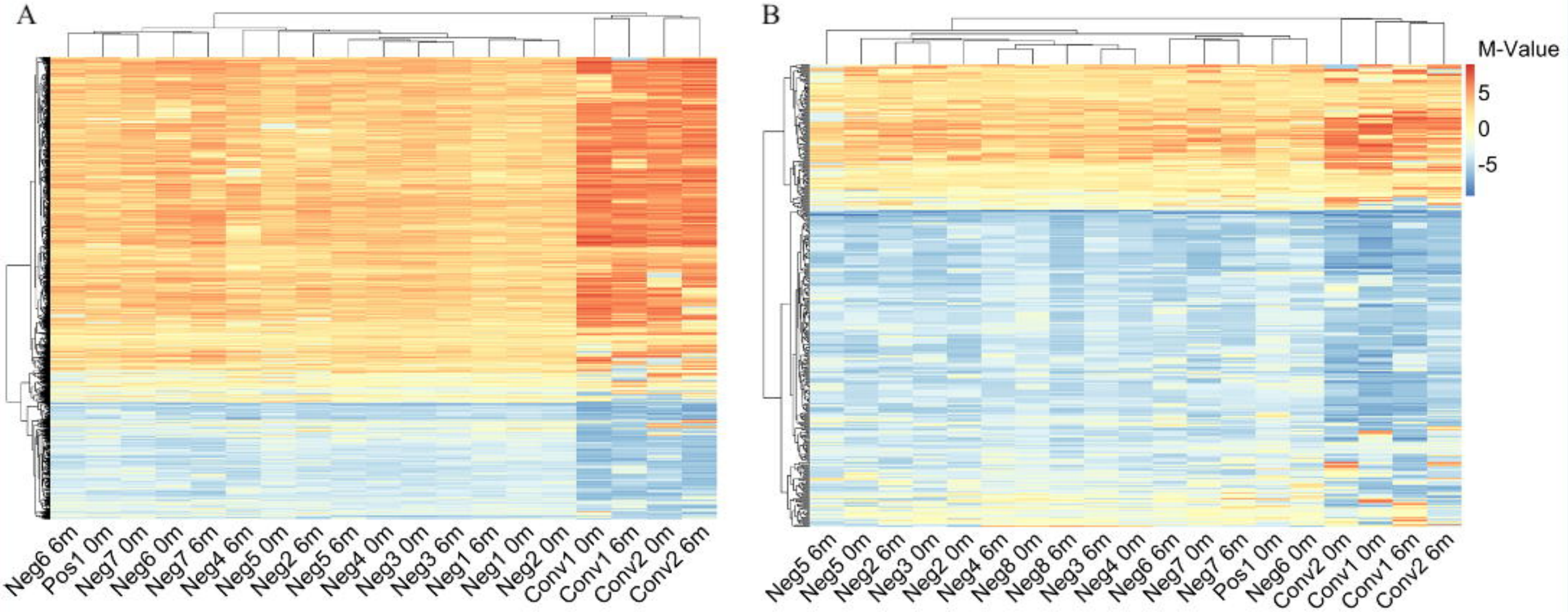
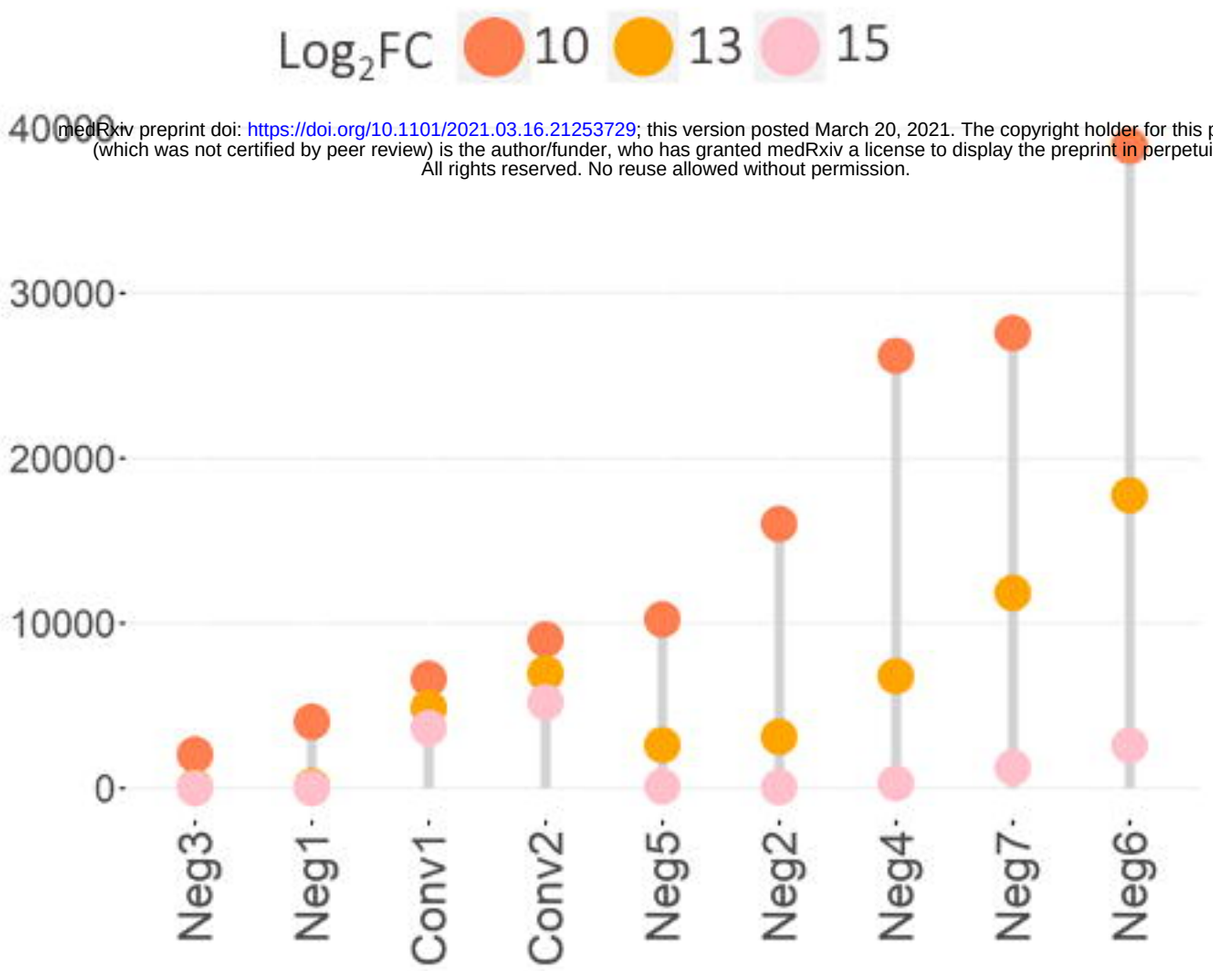
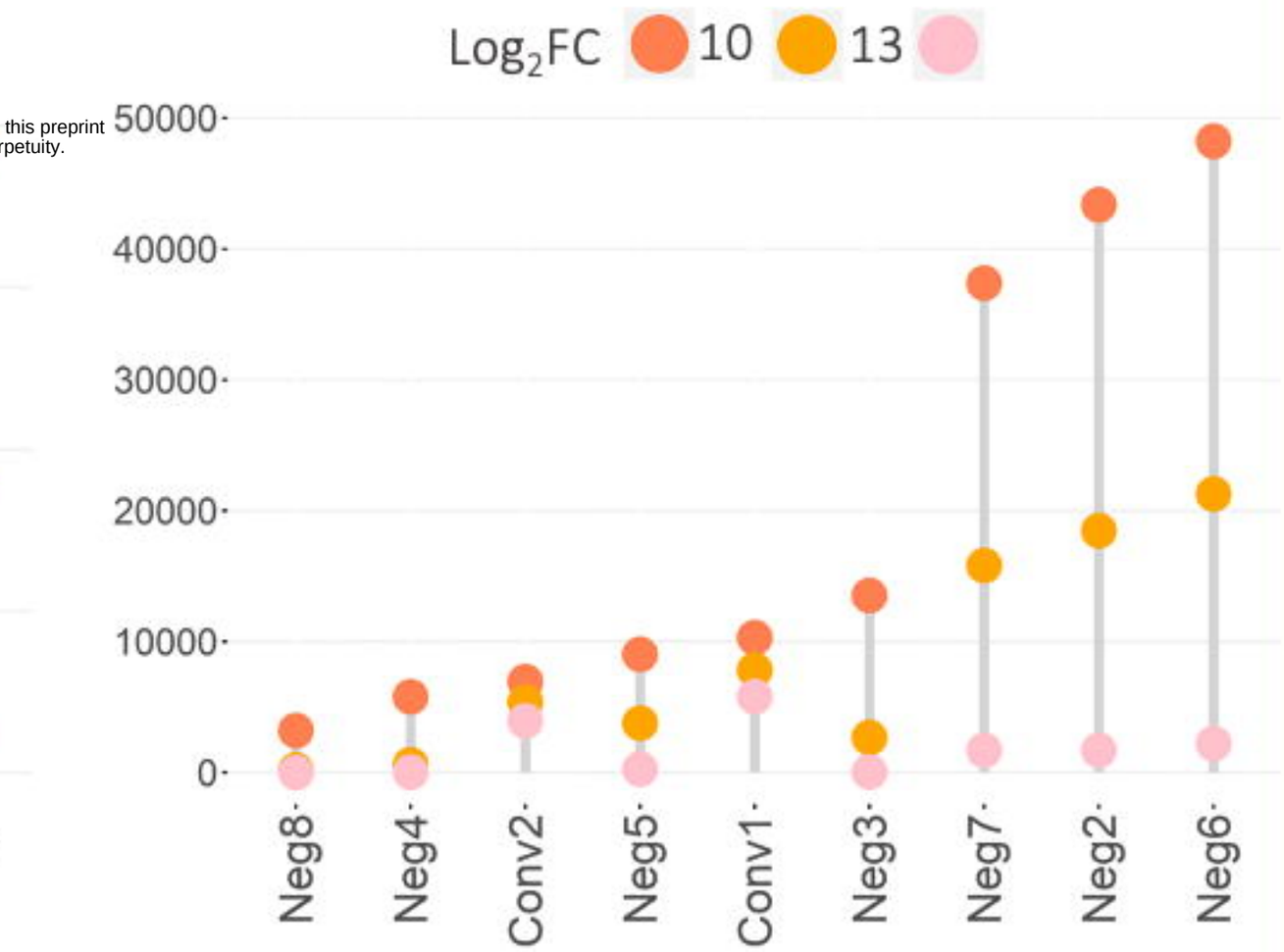

A
Neg6 6m
Pos1 0m
Neg7 0m
Neg6 0m
Neg7 6m
Neg4 6m
Neg5 0m
Neg2 6m
Neg5 6m
Neg4 0m
Neg3 0m
Neg3 6m
Neg1 6m
Neg1 0m
Neg2 0m
Conv1 0m
Conv1 6m
Conv2 0m
Conv2 6m
B
M-Value
5
0
-5
Neg5 6m
Neg5 0m
Neg2 6m
Neg3 0m
Neg2 0m
Neg4 6m
Neg8 0m
Neg8 6m
Neg3 6m
Neg4 0m
Neg6 6m
Neg7 0m
Neg7 6m
Pos1 0m
Neg6 0m
Conv2 0m
Conv1 0m
Conv1 6m
Conv2 6m

Figure 5





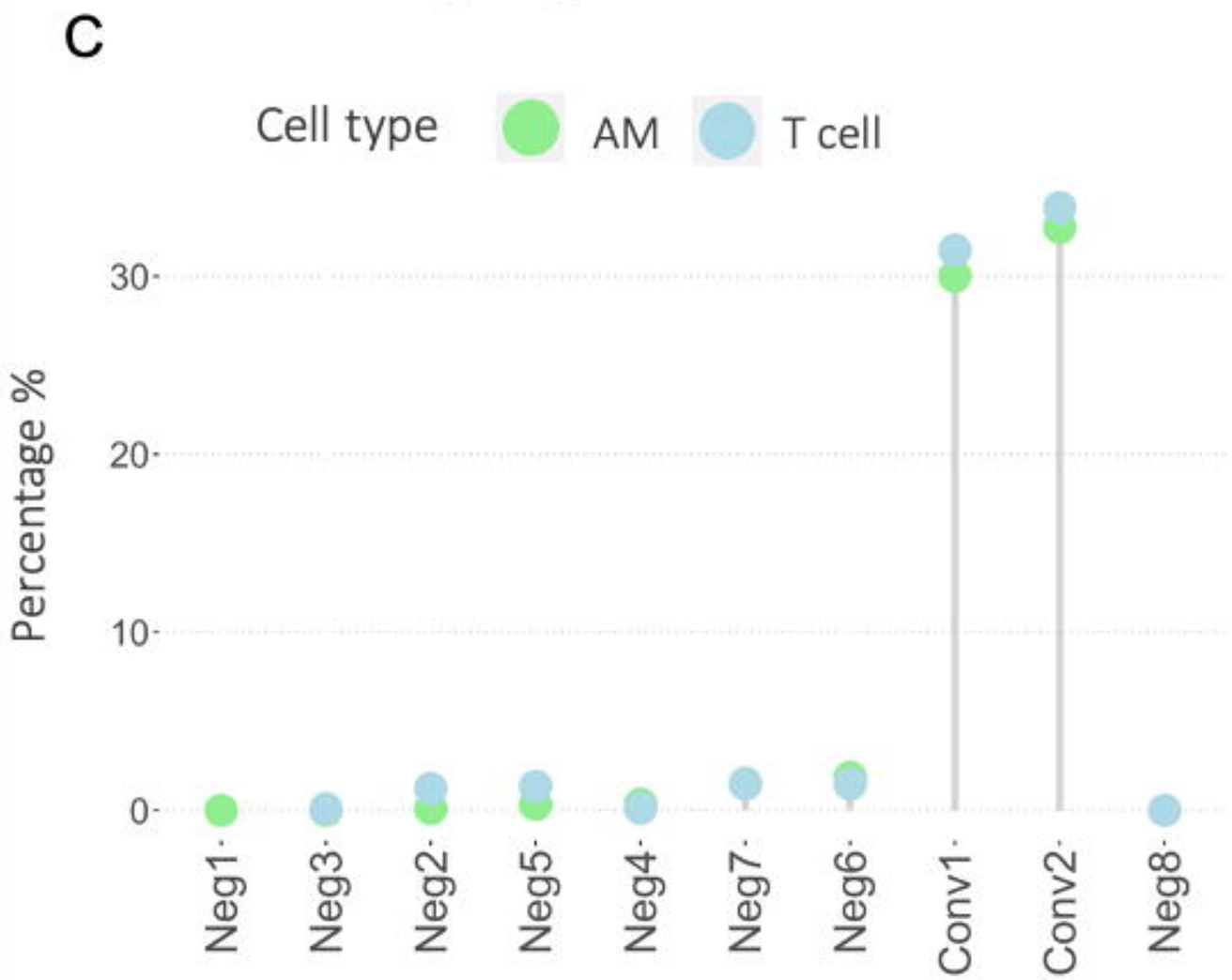


Figure 6