

# Neutrophil degranulation, NETosis and platelet degranulation pathway genes are co-induced in whole blood up to six months before tuberculosis diagnosis

Stuart Meier<sup>1,2,3,4,5</sup>, James A. Seddon<sup>4,6,7</sup>, Elizna Maasdorp<sup>1,2,3,4,8</sup>, Léanie Kleynhans<sup>1,2,3</sup>, Nelita du Plessis<sup>1,2,3</sup>, Andre G. Loxton<sup>1,2,3</sup>, Stephanus T. Malherbe<sup>1,2,3</sup>, Daniel E. Zak<sup>9</sup>, Ethan Thompson<sup>9</sup>, Fergal J. Duffy<sup>9</sup>, Stefan H. E. Kaufmann<sup>10,11,12</sup>, Tom H. M. Ottenhoff<sup>13</sup>, Thomas J. Scriba<sup>14</sup>, Sara Suliman<sup>14</sup>, Jayne S. Sutherland<sup>15</sup>, Jill Winter<sup>16</sup>, Helena Kuivaniemi<sup>1,2,3</sup>, Gerhard Walzl<sup>1,2,3,4</sup>, Gerard Tromp<sup>1,2,3,4,8,\*</sup>, GC6-74 Consortium and Catalysis TB Biomarkers Consortium

<sup>1</sup>Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Stellenbosch University, Cape Town, South Africa

<sup>2</sup>DSI–NRF Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Cape Town, South Africa

<sup>3</sup>South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa

<sup>4</sup>South African Tuberculosis Bioinformatics Initiative, Stellenbosch University, Cape Town, South Africa

<sup>5</sup>Lung Infection and Immunity Unit, Division of Pulmonology, Department of Medicine, University of Cape Town, Cape Town, South Africa

<sup>6</sup>Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Stellenbosch University, Cape Town, South Africa

<sup>7</sup>Department of Infectious Diseases, Imperial College London, London, United Kingdom

<sup>8</sup>Centre for Bioinformatics and Computational Biology, Stellenbosch University, Cape Town, South Africa

<sup>9</sup>Seattle Children's Research Institute, Center for Global Infectious Disease Research, Seattle, WA, United States of America

<sup>10</sup>Max Planck Institute for Infection Biology, Berlin, Germany

<sup>11</sup>Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

<sup>12</sup>Hagler Institute for Advanced Study, Texas A&M University, College Station, TX, United States of America

<sup>13</sup>Department of Infectious Diseases, Leiden University Medical Center, Leiden, The

33 Netherlands

<sup>14</sup>South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town,

36 Cape Town, South Africa

<sup>15</sup>Vaccines & Immunity Theme, Medical Research Council Unit, The Gambia, at the London School of Hygiene and Tropical Medicine, Banjul, The Gambia

39 <sup>16</sup>Catalysis Foundation for Health, San Ramon, CA, United States of America

\*[gctromp@sun.ac.za](mailto:gctromp@sun.ac.za)

42

**Short title:** Neutrophil transcription activated prior to tuberculosis diagnosis

# Abstract

*Mycobacterium tuberculosis* (*M.tb*) causes tuberculosis (TB) and remains one of the leading causes of mortality due to an infectious pathogen. Host immune responses have been implicated in driving the progression from infection to severe lung disease. We analyzed longitudinal RNA sequencing (RNAseq) data from the whole blood of 74 TB progressors whose samples were grouped into four six-month intervals preceding diagnosis (the GC6-74 study). We additionally analyzed RNAseq data from an independent cohort of 90 TB patients with positron emission tomography-computed tomography (PET-CT) scan results which were used to categorize them into groups with high and low levels of lung damage (the Catalysis TB Biomarker study). These groups were compared to non-TB controls to obtain a complete whole blood transcriptional profile for individuals spanning from early stages of *M.tb* infection to TB diagnosis. The results revealed a steady increase in the number of genes that were differentially expressed in progressors at time points closer to diagnosis with 278 genes at 13–18 months, 742 at 7–12 months and 5,131 detected 1–6 months before diagnosis and 9,205 detected in TB patients. A total of 2,144 differentially expressed genes were detected when comparing TB patients with high and low levels of lung damage. There was a large overlap in the genes upregulated in progressors 1–6 months before diagnosis (86%) with those in TB patients. A comprehensive pathway analysis revealed a potent activation of neutrophil and platelet mediated defenses including neutrophil and platelet degranulation, and NET formation at both time points. These pathways were also enriched in TB patients with high levels of lung damage compared to those with low. These findings suggest that neutrophils and platelets play a critical role in TB pathogenesis, and provide details of the timing of specific effector mechanisms that may contribute to TB lung pathology.

## Author summary

*Mycobacterium tuberculosis* (*M.tb*) causes tuberculosis (TB) and remains one of the leading causes of mortality due to an infectious pathogen. Human immune responses must be balanced to inhibit disease progression while limiting self-damage, however in defense against *M.tb* host responses have been implicated in lung damage and in driving progression of *M.tb* infection to severe lung disease. The identification of immune responses that are activated during the development of TB could provide potential targets for interventions that might suppress disease progression, and possibly limit tissue damage. Here we identify a subset of genes that function in several biological processes that are strongly activated in the late stages of TB development and in TB patients. The proteins encoded by these genes are known to degrade lung tissue and contribute to severe lung disease.

# Introduction

*Mycobacterium tuberculosis* (*M.tb*), which causes tuberculosis (TB), remains one of the leading pathogens that is responsible for human death [1]. Although estimates suggest that 23% of the world's population are infected with *M.tb* [1], most individuals are able to eradicate or control the disease [2] and only 5-10% develop TB during their lifetime [1].

It is largely unknown why some *M.tb*-infected individuals progress to active TB, but an over-active inflammatory response is considered an important factor contributing to lung pathology [3,4]. Rapid necrosis, associated with a delayed-type hypersensitivity reaction against accumulated *M.tb* antigens [5], or an *M.tb*-mediated autoreactive response [4], is thought to cause an inflammatory response that drives lung extracellular matrix (ECM) destruction and cavity formation. This allows *M.tb* in the lung interstitium to access the airways and be transmitted [6].

Neutrophils are strongly activated in response to *M.tb* infection [7,8], although numerous studies have shown they are ineffective at killing or controlling *M.tb* replication (for review see [9]). Rather, the activation and infiltration of neutrophils at late stages of infection is associated with TB pathogenesis [9-13]. Indeed, neutrophils are the predominant immune cell type present in lung lesions and cavities of pulmonary TB patients and are associated with lung ECM destruction and cavity formation [12,14]. Neutrophils contain a diverse array of preformed proteases in their granules including neutrophil collagenase and matrix metalloproteinase 8 (MMP8), which digest ECM in human lung [12] and MMP8 is the most prevalent MMP present in the sputum of TB patients [15]. *M.tb*-induced MMP8 secretion is also associated with the secretion of neutrophil extracellular traps (NETs); a process by which neutrophils release their antimicrobial granule proteases, DNA and histones extracellularly in a type of programmed cell death named NETosis [16]. Collectively, these studies strongly implicate neutrophils in late stages of TB lung pathology including ECM destruction and cavity formation.

Whole blood transcriptomic studies have advanced our knowledge of the host response to many diseases, including TB [11,17-19]. Previous transcriptomics studies identified a

number of immune processes activated in TB patients including interferon (IFN)-  
signaling [11,17,20], myeloid cell inflammation [21], and the inflammasome and  
proinflammatory pathways [17]. Two large cohort studies that monitored individuals at  
high risk for developing TB for up to two years before diagnosis generated whole blood  
transcriptional profiles for TB progressors with the aim to develop predictive transcript-  
based risk signatures [22,23]. Subsequently, data from one of these studies, the South  
African adolescent cohort (aged 12-18 years) [23], was analyzed to identify biological  
pathways and processes that are active during TB progression [24,25]. These studies  
found that type I IFN signaling and the complement cascade were the main pathways  
activated at early stages of TB progression as observed in TB patients [24,25] and  
changes in functionally uncharacterized neutrophil and platelet gene modules occurred  
at times closer to TB diagnosis [25]. Here we extended the prior analyses by jointly  
analyzing two data sets and dissecting the neutrophil processes.

We performed a differential expression (DE) and comprehensive pathway analysis of  
whole-blood RNA sequencing (RNAseq) datasets obtained from the Gene Expression  
Omnibus (GEO) [26]. The first set, GEO series GSE94438 (GC6-74 study), was  
generated from individuals for a period of up to two years prior to TB diagnosis [22]. The  
second data set, GEO series GSE89403 (Catalysis TB Biomarker Study, hereafter the  
Catalysis study), compared TB patients to healthy controls. Collectively, the data sets  
provide a full spectrum of host transcriptional responses spanning from early infection  
stages to TB diagnosis. Our analysis focused on identifying host responses that may  
contribute to the development of TB.

## Results

### Study groups and data quality

We performed a DE and comprehensive pathway analysis of whole-blood RNAseq data  
sets obtained from the GEO.

The first data set included RNAseq data for TB progressors (GEO:GSE94438) and was  
generated as part of the Grand Challenges 6-74 (GC6-74) program. The program was a  
longitudinal study of household contacts of newly diagnosed, sputum smear-positive TB

cases in a high TB-prevalence settings and included samples from four different African populations. A detailed description of the study groups has been published previously [22,23,27,28]. Contacts diagnosed with TB within 3 months of recruitment were considered to have prior disease and were excluded from the study, while those developing disease after this point were considered to have true incident TB and were included as cases.

Since the aim of the present study was to identify genes that are DE at different stages of TB progression, the progressor samples were labelled according to the time the sample was collected before their TB diagnosis. As an example, if a participant was diagnosed with TB 8 months after enrollment and they had samples collected at enrollment (time 0) and 6 months, these samples would be classified as 8 and 2 months before diagnosis, respectively. The labelled samples were subsequently placed into groups that corresponded to 19–24, 13–18, 7–12 and 1–6 months before TB diagnosis (longitudinal data for subjects who were diagnosed with TB were aligned to the time of diagnosis and specimens taken before diagnosis were binned into the aforementioned groups). The number of samples in each study group, and mean time before TB diagnosis for each time point are listed in Table 1. For the GC6-74 study, all the TB progressor groups were compared to the same non-TB control samples that were collected at recruitment time. This ensured that any differences between the progressor groups were a result of changes in their respective expression rather than changes in the baseline group.

The second data set used included a subset of RNAseq data from the South African Catalysis study (GEO: GSE89403). Here, we only used RNAseq data from TB patients at diagnosis and the non-TB controls [27,28]. This data set also contained positron emission tomography-computed tomography (PET-CT) data for the TB patients which we used to categorise the patients (see Methods for details) into groups with high (high PET scores) and low (low PET scores) levels of lung damage. The primary analysis for the Catalysis study data set compared all TB patients to controls. We additionally performed a number of PET-score based sub-analyses where patients with high and

low PET scores were individually compared to controls, and patients with high and low PET scores were directly compared to identify differences between them.

**Table 1. Study groups.**

		GC6-74 Study <sup>a</sup>					Catalysis Study <sup>b</sup>	
		Time <sup>c</sup> before TB diagnosis in progressors				Control <sup>d</sup>	TB	Control <sup>e</sup>
		19–24	13–18	7–12	1–6			
N	Total	11	19	18	47	198	90	21
	Male	3	7	8	20	81	55	8
Age	Mean (y)	24.2	25.5	27.9	29.0	27.5	34.3	33.4
	SD	9.4	11.3	11.3	12.0	13.0	11.2	11.5
Time to diagnosis	Mean	21.5	16.1	9.2	4.2	NA <sup>f</sup>	NA	NA
	SD	0.6	0.4	0.4	0.2			
	95% CI	20.4–22.6	15.3–16.8	8.3–10.1	3.8–4.6			

<sup>a</sup>GEO series GSE94438

<sup>b</sup>GEO series GSE89403

<sup>c</sup>Time in months before diagnosis

<sup>d</sup>GC6-74 Controls: household contacts who did not develop TB

<sup>e</sup>Catalysis Controls: Other lung disease subjects

<sup>f</sup>NA, Not applicable

The FastQC analysis revealed that the read quality was good with no adapter contamination, therefore, no trimming was performed. Mapping statistics revealed that on average 86% of the reads mapped uniquely to the human genome and of these an average of 88% mapped uniquely to genes (data not shown). For each sample, a mean of 42.9 (42.4, 43.4, 95% CI; GC6-74 study) and 35.8 (35.4, 36.2, 95% CI; Catalysis study) million paired-end reads mapped to genes.

## Differential expression analysis

The edgeR DE analysis identified several significantly perturbed genes across the different time points (Table 2). For the progressors in the GC6-74 study [23], there was a steady increase in the number of DE genes detected at time points more proximal to TB diagnosis with only five genes at 19–24 months, 278 at 13–18 months, 742 at 7–12 months and 5,131 at 1–6 months before diagnosis. A total of 9,205 genes were DE in TB patients compared to controls (Catalysis study). In the Catalysis sub-group analysis, in comparison to controls, a similar number of genes were significantly DE in patients with high (9,211) and low (8,496) PET-scores (Table 2). Notably, a total of 2,144 significantly DE genes were identified when directly comparing patients with high and



low PET-scores confirming differences in transcriptional signatures between TB patients with high and low levels of lung damage.

The DE results for all genes and time points are presented in **S1 Table**. There was substantial overlap in the DE genes with 92% (445/492) of the genes upregulated 7–12 months before diagnosis also being upregulated in the samples taken 1–6 months prior to TB diagnosis. Similarly, 85% (2,365/2,791) of the genes upregulated 1–6 months before diagnosis in GC6-74 study subjects were also upregulated in Catalysis study TB patients (Fig 1A). The high level of overlap in genes upregulated 1–6 months before diagnosis with those upregulated in TB patients illustrates substantial homogeneity in the host response between the two cohorts and provides support for the validity in their comparisons.

In the Catalysis study PET categorized sub-analysis, the TB high PET score group vs controls shared 91% (3,818/4,200) while the TB low PET score group vs controls shared 94% (3,957/4,200) of the upregulated genes identified in the complete analysis (Fig 1B). Of the 995 genes that were upregulated when directly comparing the TB high PET score group to the TB low PET score group, 95% (943/995) were upregulated when comparing the TB high PET score group to controls and 78% (773/995) were upregulated when comparing the TB low PET score group to controls. Thus, 773 genes that were upregulated when comparing the TB low PET score group to controls, were further significantly upregulated in TB patients with high PET scores compared to those with low PET scores.

**Table 2. Number of significantly differentially expressed (DE) genes (FDR<0.05) at each time point.**

Test group	GC6-74 Study <sup>a</sup>				Catalysis Study			
	19–24 <sup>b</sup>	13–18 <sup>b</sup>	7–12 <sup>b</sup>	1–6 <sup>b</sup>	TB <sup>c</sup>	PET high <sup>c</sup>	PET low <sup>c</sup>	PET high <sup>d</sup>
Case N	11	19	18	47	90	24 <sup>e</sup>	65 <sup>e</sup>	NA
<b>DE Gene Count</b>								
Up	5	247	492	2,791	4,200	4,284	4,031	995
Down	0	31	250	2,340	5,005	4,927	4,465	1,149
Total	5	278	742	5,131	9,205	9,211	8,496	2,144

<sup>a</sup> Reference group: GC6-74 non-TB controls

<sup>b</sup> Months before diagnosis

<sup>c</sup> Reference group: Catalysis non-TB controls

<sup>d</sup> Reference group: Catalysis PET-low

<sup>e</sup> One TB patient in the Catalysis Study group did not have PET-CT data

210

**Fig 1. Venn diagram illustrating the overlap in significantly upregulated genes. A)**

213 Comparison of upregulated genes detected 7–12 and 1–6 months before diagnosis and  
in TB patients. B) Comparison of upregulated genes detected in TB patients with high  
and low PET scores compared separately to controls and high and low PET scores  
216 compared to each other. The numbers in parenthesis indicate the total number of  
upregulated genes for the specific comparison.

219 Since the increased sample size at time points more proximal to TB diagnosis could be  
responsible for the increased number of DE genes observed, we performed  
subsampling to confirm that true biological differences were responsible for the  
222 increased number of genes. The subsampling of the 1–6 months before diagnosis  
group revealed that on average 1,454 genes were upregulated across 1,000 iterations  
of 19 samples compared to 2,791 upregulated genes detected with a sample size of 47  
225 (**S1 Fig**). For the TB group, a mean of 3,510 genes were upregulated across 1,000  
iterations of 19 samples compared to 4,200 upregulated genes detected with a sample  
size of 90 (**S2 Fig**). The higher mean number of upregulated genes detected when  
228 subsampling both groups compared to the 13–18 (249; n=19) and 7–12 (492; n=18)  
months before diagnosis groups (Table 2) indicated that the increased number of DE  
genes observed was predominately a consequence of the time point. This is also  
231 supported by almost double the number of genes detected in the 7–12 compared to the  
13–18 months before diagnosis groups which had similar sample sizes. The results also  
confirmed that larger sample numbers improve statistical power and facilitate the  
234 detection of additional DE genes.

**Gene Ontology (GO) and KEGG pathway analysis**

Since only five genes were DE 19–24 months before diagnosis (**S1 Table**), the pathway  
237 analysis was performed on the 13–18, 7–12 and 1–6 months before diagnosis time

points and on TB patients. The use of the weighted algorithm (weight01) in topGO identified the most specific and thus informative GO terms including several not previously reported during TB progression. In addition, only pathways that had a minimum of 10 DE genes were considered. In summary, the pathway analysis of the upregulated genes identified several defense-related pathways that were enriched early (13–18 months before diagnosis), remained elevated and intensified, based on increased numbers of annotated genes, at time points more proximal to TB diagnosis. Several additional processes that were specifically related to neutrophil and platelet defense responses were first enriched 1–6 months before diagnosis and remained enriched in TB patients.

A list of the GO and KEGG pathways significantly perturbed at different stages of TB progression is provided in **S2 and S3 Tables**. Here, in the main text, we focus on the induction of neutrophil-mediated defenses and related pathways and processes since we consider these the most likely drivers of disease progression.

## **Early (13–18 and 7–12 months before) pathway activation**

The most enriched GO terms detected 13–18 and 7–12 months before diagnosis were associated with IFN signaling including “type I interferon signaling pathway” and “interferon- $\gamma$ -mediated signaling pathway” (Fig 2A).

### **Fig 2. GO annotations and KEGG pathways first enriched at (A) 13–18 and 7–12 months before diagnosis and (B) 1–6 months before diagnosis.**

Radar plots illustrating proportions of DE genes in the gene sets at the indicated time points for selected pathways broadly related to neutrophil function. All four time points [13–18, 7–12, and 1–6 months before, and at diagnosis (TB)] are shown in each plot to demonstrate that annotations detected early persist and generally have more DE genes at later stages, although significance may decrease. The numbers in parenthesis represent the total numbers of genes in the annotation gene set. Asterisks indicate nominal significance: \*\*\*,  $< 5e^{-15}$ ; \*\*,  $< 5e^{-07}$ ; \*,  $< 5e^{-04}$ .

# One to six months before diagnosis

A total of 5,131 genes were significantly DE 1–6 months before TB diagnosis with respect to household controls with 2,791 (54%) up-regulated (Table 2). The marked increase in the number of DE genes was notably accompanied by a corresponding enrichment in several effector pathways that were not enriched at earlier time points documenting a strong and distinct activation of the host immune response at this time point. This response was dominated by a strong enrichment in up-regulated genes that function in neutrophil-mediated immunity with “neutrophil degranulation” [ $p < 10^{-30}$ , 222/464 (48%) annotated genes] being the most significant and specific GO biological process (BP) term detected (Fig 2B). Other enriched pathways related to neutrophil function included “neutrophil chemotaxis”, “Fc-gamma receptor signaling pathway involved in phagocytosis” and “platelet degranulation”. Several terms related to the production of reactive oxygen species (ROS) were also enriched including “positive regulation of superoxide anion generation”, “respiratory burst”, “mitochondrial electron transport” and “cellular oxidant detoxification”. The KEGG “NET formation” and “tuberculosis” disease pathways were also enriched at this time point (Fig 2B). An illustration of the KEGG “NET formation” pathway is presented in Figure 3A and 3B with fold-changes (log2) for significantly upregulated genes displayed for the different time points.

**Fig 3. KEGG NET formation pathway illustrating the fold changes (log2) for significantly upregulated genes at different time points.** A) Gene expression as fold-change is indicated by a color block in the background rectangle or each gene corresponding to the contrasts, 13–18, 7–12, 1–6 months before diagnosis and TB compared to their respective controls (from left to right). B) As above but illustrating the significantly increased expression of genes in TB patients with high PET-scores compared to those with low PET-scores. Figures rendered by the R Bioconductor pathview package [29], on finite resolution raster images from KEGG.

Annotated neutrophil degranulation genes upregulated 1–6 months before diagnosis included neutrophil *integrin subunit beta 2* (*ITGB2/LFA1*), *integrin subunit alpha M* (*ITGAM*) and *integrin subunit alpha X* (*ITGAX*) that form the transmembrane heterodimer integrin receptors ITGAM/ITGB2 (mac-1) and ITGAX/ITGB2 that facilitate the transmigration of neutrophils to infection sites [30] (Fig 4). The neutrophil granule transmembrane chemotactic *formyl peptide receptor* (*FPR*) 1 and 2 that bind formyl-peptides derived from bacteria and damaged host molecules and the *CXCR1* and 2 receptors that bind the neutrophil chemotactic and activation factors IL8 and CXCL7 (neutrophil-activating peptide-2, NAP2) were also induced. Ligation of the G protein-coupled receptors FPR1 and 2, and CXCR1 and 2 induces calcium signaling which regulates neutrophil migration, ROS synthesis and primes neutrophils for further activation and the release of subsequent granules [31]. The genes encoding the antimicrobial myeloid-related proteins (MRP), namely *S100A8*, *S100A9*, *S100A11*, and *S100A12*, were also upregulated [16,32].

#### **Fig 4. Heatmap of selected annotated neutrophil degranulation genes**

**significantly upregulated 1–6 months before diagnosis and in TB patients.** The fold changes presented at 1–6 months before diagnosis (6mBf) are in contrast to the non-TB time zero baseline group from the GC6-74 study. For the TB comparison, the 21 non-TB individuals from the Catalysis study served as controls (see Table 1). Fold-changes for Catalysis TB-patients, categorized with high and low PET scores, were also compared individually to controls and directly to each other to identify differences between the categorized groups.

Additional upregulated genes annotated to function in neutrophil degranulation include those encoding components of the superoxide generating NADPH oxidase 2 (*NOX2*). This included the transmembrane catalytic [cytochrome b-245 -alpha (*CYBA*) and -beta (*CYBB*)] and cytosolic regulatory subunits [neutrophil cytosolic factor (*NCF*)1, 2 and 4 and the small G-protein, *RAC2* [33,34]. The induction of genes encoding components of *NOX2* along with the enrichment of pathways involved in superoxide generation and the

respiratory burst, is consistent with the enrichment of FcγR mediated phagocytosis which activates NOX2 assembly [34,35]. NOX2 generated ROS also induce the synthesis of NETs [36,37].

The co-induction of “platelet degranulation” with “neutrophil degranulation”, first observed 1–6 months before diagnosis, is consistent with the physical interactions and costimulatory roles platelets and neutrophils share during immune responses [38,39].

## TB versus healthy controls

A total of 9,205 genes were DE in individuals with TB in contrast to the non-TB time zero baseline group from the GC6-74 study, of which 4,200 (46%) were up-regulated (Table 2). Similar to 1–6 months before diagnosis, there was a strong enrichment in pathways related to neutrophil-mediated immunity with “neutrophil degranulation” [ $p < 1 \times 10^{-30}$ , 305/464 (66%) annotated genes] being the most significant and specific GO BP term detected (Fig 2A). Approximately 95% (210/222) of the neutrophil degranulation (Fig 5A) and 97% (60/62) of NET formation (Fig 5B). annotated genes that were upregulated 1–6 months before diagnosis were also upregulated in TB patients. All the other enriched neutrophil related functions observed 1–6 months before diagnosis were also enriched in TB patients.

### **Fig 5. Venn diagram of the overlap in significantly upregulated genes that are annotated to function in: A) neutrophil degranulation, and B) NET formation.**

Differentially expressed genes at 7–12 and 1–6 months before diagnosis and in TB patients in contrast to the non-TB time zero baseline group from the GC6-74 study. The numbers in parenthesis indicate the total numbers of upregulated genes at the time point.

Several additional terms related to neutrophil function were enriched in TB patients, but not 1–6 months before diagnosis. This included the terms “leukocyte transendothelial migration”, “neutrophil extravasation”, “extracellular matrix disassembly”, “collagen catabolic process”, “antibacterial humoral response”, “antimicrobial humoral immune

response mediated by antimicrobial peptide”, “platelet activation” and “platelet aggregation” (Fig 6).

**Fig 6. Pathways first enriched in TB patients.** Radar plot with symbols as in Figure 2. The full name for “Antimicrobial Humoral Immune Response ...” is “Antimicrobial humoral immune response mediated by antimicrobial peptide (GO:0061844)”.

It is interesting to note that the expression of several genes that encode neutrophil granule antimicrobial peptides were only elevated in TB patients and not progressors (Fig 4). These included myeloperoxidase (*MPO*), neutrophil elastase (*ELANE*), cathepsin G (*CTSG*), cathelicidin antimicrobial peptide (*CAMP/LL37*), lactotransferrin (*LTF*), defensin alpha 4 (*DEFA4*), azurocidin 1 (*AZU1*), bactericidal/permeability-increasing protein (*BPI*), *MMP8*, *MMP9*, and neutrophil gelatinase-associated lipocalin (*NGAL*). A number of these genes are also annotated to function in other enriched pathways including NET formation and antimicrobial and antibacterial humoral defense responses.

## Catalysis PET-CT sub-analysis

Unsurprisingly, given the high level of overlap between the genes, the results of the pathway analysis were very similar to those observed when comparing the complete TB group to controls, with “neutrophil degranulation”, “platelet degranulation” and “NET formation” being significantly enriched in both the TB-high-PET score group and the TB-low-PET score group vs controls comparisons. Strikingly, these pathway were also enriched when directly comparing the TB-high-PET score group to the TB-low-PET score group with neutrophil degranulation [ $p = 8.1^{-19}$ , 86/464 (19%)], platelet degranulation [ $p = 4.5^{-6}$ , 21/105 (20%)] and NET formation [ $p = 5.8^{-12}$  34/142 (24%), Table 3], revealing that TB patients with increased lung damage have an enrichment in upregulated genes that function in these pathways. Annotated neutrophil degranulation genes that were significantly upregulated in TB patients with high PET scores compared



to those with low PET scores included *MMP8*, *S100A12*, *S100A8*, *S100A9*, *CAMP* and *PPBP* (Fig 4).

378

**Table 3. Pathways significantly enriched with upregulated genes when comparing TB patients with high PET-scores to TB patients with low PET-scores.**

Pathway	Total annotated genes	Number Up	% Up	P-value	Rank
Neutrophil degranulation <sup>a</sup>	464	86	18.5	1.2 <sup>-18c</sup>	1
Platelet degranulation <sup>a</sup>	105	21	20	4.5 <sup>-06c</sup>	8
NET formation <sup>b</sup>	142	34	23.9	5.8 <sup>-12</sup>	2

<sup>a</sup>GO biological process

<sup>b</sup>KEGG pathway

<sup>c</sup>weight01.algorithm p-value

## Discussion

381 Our results revealed an early induction of IFN-related signaling at 18 months before  
diagnosis. A strong induction of neutrophil and platelet degranulation and NETosis  
related genes was detected 6 months before TB diagnosis and persisted in TB patients  
384 supporting a pathogenic role of these responses in disease development. Although  
these processes have previously been detected in severe TB, here we document that  
they occur well before disease diagnosis, indicating that they may be critical in  
387 mediating lung tissue destruction and progression from non-infectious latent or primary  
TB to contagious TB.

DE and pathway enrichment analysis revealed that an induction of the host defense  
390 response, that is dominated by type I and II IFN signaling, was first detectable in whole  
blood of TB progressors as early as 13–18 months before diagnosis. A progressive  
induction in the intensity of the IFN response was observed as time points approached  
393 TB diagnosis with a distinct induction of neutrophil-mediated defense observed 1–6  
months before diagnosis that persisted and intensified in patients diagnosed with TB.  
Given that we examined transcriptional responses in whole blood, changes observed  
396 are likely to result from signaling molecules released from infected cells, most likely in  
the lungs, rather than through direct contact with the pathogen. The signaling is likely to  
prime circulating leukocytes for activation in preparation for their recruitment to infection  
399 sites.



A unique and striking finding was the strong induction of specific neutrophil-mediated defenses, including neutrophil degranulation, and NET formation, that are first observed 1–6 months before diagnosis (Fig 2B), maintained in TB patients and significantly enriched in TB patients with high levels of lung damage compared to those with low level damage assessed by PET-CT (**S4 Table**). The induction of neutrophil-mediated defenses is consistent with literature that associates neutrophil activation and infiltration with TB as well as pulmonary destruction [9,11,13]. Neutrophils have been reported to be the predominant immune cell type present in the sputum and at the site of infection in the lung [40], and are also the main cell type infected with *M.tb* in sputum and lung cavities [14]. In addition, neutrophil markers are associated with necrotic areas in granulomas [41] and excessive neutrophil infiltration is associated with the softening of caseous lesions in the lung [5].

Although the neutrophil degranulation and NET formation processes are both enriched 1–6 months before diagnosis and in TB patients, the additional increase in the expression level and induction of additional genes that function in these pathways in TB patients suggests a sequential induction of neutrophil activation during the late stages of disease progression. Genes induced 1–6 months before diagnosis encode neutrophil membrane integrins and chemotactic receptors that mediate their adhesion and transmigration to infection sites as well as components of NOX2 [42]. These genes are generally involved in the priming of neutrophils in preparation for full activation at infection sites. In TB patients, there is a distinct induction of genes that encode neutrophil granule proteins that function in ECM degradation including *MMP8* and *MMP9* and those that encode numerous neutrophil azurophilic granule proteins including *MPO*, *ELANE*, *DEFA4* and *BPI*. The *MPO*, *ELANE* and *PADI4* peptides play a critical role in NET formation, driving chromatin decondensation, cell rupture and the extracellular release of DNA [43–45]. The significantly higher induction of genes in patients with high-PET scores compared to those with low-PET scores, including *MMP8*, links their increased expression to increased lung damage.

The induction of genes that encode neutrophil granule proteins, including bactericidal granule enzymes, is intriguing since expression of these genes is typically high in

neutrophils during maturation when granule proteins are synthesized, and declines in mature cells [46,47]. A highly similar set of genes including *AZU1*, *CAMP*, *CTSG*,  
 432 *DEFA4*, *ELANE*, *LTF*, and *MPO* is activated and forms part of a co-expression module in isolated low-density granulocytes (LDG) from systemic lupus erythematosus (SLE) patients and it is thought that LDGs are immature neutrophils that have been released  
 435 into the circulation during granulopoiesis [45,48]. Elevated LDG levels are correlated with disease severity in TB patients, but *in vitro* studies indicated LDGs were generated from normal neutrophils after degranulation or NET formation [49,50]. These studies,  
 438 however, did not investigate whether the expression of genes encoding the granule proteins were elevated in these cells. Given our analysis is on whole blood which includes multiple cell types, it is possible that the observed elevated expression of  
 441 neutrophil related genes is a result of an increase in the proportion of neutrophils in the blood. This, however, could only partially account for the observed increases since the expression of a number of these genes including *ELANE*, *DEFA4*, *BPI* and *LTF* (Fig 4)  
 444 are elevated > 4-fold ( $\log_2$  FC > 2). Since neutrophils already constitute 50% to 70% of all circulating leukocytes, an increase in their abundance could not possibly account for the increased transcription observed. Nevertheless, the increased expression of  
 447 neutrophil granule genes observed in this study may contribute to tissue destruction and TB pathogenesis since pro-inflammatory LDGs have an enhanced capacity to secrete NETs and granule peptides in TB and SLE patients [45,50].

450 While *M.tb* has been shown to stimulate NET synthesis [51], NETs have a limited ability to kill *M.tb* [52] and it has been suggested they rather provide a platform for extracellular  
 453 *M.tb* replication that facilitates pulmonary lesion growth and drives the transition to infectious TB [9,53,54]. In TB patients' plasma NET, MPO and ELANE levels are correlated with TB severity [55], while elevated serum levels of citrullinated histone H3, a NET biomarker, are associated with lung cavitation and poor treatment outcome [56].  
 456 A number of NET components, including dsDNA, mitochondrial DNA, and granule proteinases function as immune-stimulatory molecules when released extracellularly [45] and have been identified as important drivers of immune-pathogenesis in both  
 459 infectious and non-infectious human diseases [57]. Given the early and sustained level of type I IFN signaling observed during TB progression, it is interesting to note that IFN-

$\alpha$  can prime neutrophils for NETosis [58]. In turn, NETs can stimulate type I IFN  
production in plasmacytoid dendritic cells (pDCs), prolonging the inflammatory response  
[59].

The co-induction of platelet and neutrophil-mediated defense responses starting as  
early as 13–18 months before diagnosis is consistent with their established dependent  
roles during immune responses [38,39,60]. The activation of platelet and neutrophil  
degranulation, including the induction of platelet and neutrophil derived granule  
chemokines and membrane proteins that mediate their co-migration and physical  
adhesion including *P-selectin (SELP)* and *CXCL7 (neutrophil-activating peptide-2  
NAP2)* from platelets and *selectin P ligand (SELPLG)*, *CXCR1* and *2*, and *ITGB2/LFA1*  
from neutrophils is consistent with studies that document their physical interactions  
during defense responses [38,60]. Platelet-neutrophil adhesion induces intracellular  
signaling cascades that activate many neutrophil antimicrobial functions observed in this  
study including ROS production, phagocytosis and NETosis [39,61]. Indeed, the  
neutrophil membrane integrin ITGB2/LFA1 is required to mediate neutrophil-platelet  
adhesion that drives NET release in human sepsis [61]. It has been suggested that  
platelets function as a barometer, stimulating NET synthesis when bacterial levels  
exceed the neutrophils' capacity to control infection through alternative mechanisms  
[61,62].

Platelet-neutrophil complexes are implicated in the pathogenesis of pulmonary  
inflammation and acute lung injury [39]. In TB patients, platelet numbers and activity are  
increased [63,64] and the concentration of numerous platelet-derived mediators  
including P-selectin, RANTES and PDGF was increased and correlated with levels of  
tissue-degrading MMPs 1, 7, 8, and 9, in bronchoalveolar lavage samples [65]. The co-  
induction of platelet and neutrophil functions that are first observed 1–6 months before  
diagnosis and further activated in TB patients suggests a pathological role for these  
interactions in the late stages of TB development. This is further supported by the  
increased induction of these pathways in TB patients with high PET scores compared to  
those with low PET scores.

The current findings confirm and elaborate the findings of a previous study by Scriba et al. [25] that identified the induction of uncharacterized neutrophil and platelet gene modules 6 months before TB diagnosis. Here, we identified specific processes that are mediated by these cells, including neutrophil and platelet degranulation as well as NET formation that are activated 6 months before TB diagnosis. Our analyses additionally discovered that these processes are further activated in individuals with TB and at an elevated level in TB patients with increased lung damage. Collectively, these results support that the activation of these pathways is linked to TB progression and increased lung damage in TB patients.

## Conclusion

The distinct co-induction of neutrophil and platelet degranulation as early as 13–18 months before diagnosis, NET formation 1–6 months before diagnosis, as well as their further activation in TB patients is consistent with these processes playing a critical role in the late stages of disease progression. This is further supported by the enrichment in upregulated genes that function in these pathways in TB patients with increased levels of lung damage. Platelet-neutrophil interactions are required for mediating their chemotaxis and recruitment to infection sites and for NET synthesis [66]. NETs are associated with TB pathogenesis and are thought to provide an extracellular platform for *M.tb* growth [53] while neutrophil granule enzymes degrade the ECM of the lung [12]. Collectively, these responses can lead to rapid lesion growth and tissue destruction that allows *M.tb* to disseminate into the airways [53]. The detectable activation of these specific processes in whole blood around 6 months before TB diagnosis makes them promising candidates for targeted therapeutic interventions that may limit lung damage and prevent progression to active TB.

## Methods

### Study outline and data sources

The whole blood RNAseq data analyzed was obtained from two independent data sets.

The data for TB progressors were generated as part of the the Bill and Melinda Gates Foundation GC6-74 program that was a longitudinal study of household members of newly diagnosed TB cases which was conducted across four African sites including South Africa, The Gambia, Uganda and Ethiopia. In brief, when a newly diagnosed TB case was identified, individuals with whom they shared a house for a minimum period of three months were recruited with the expectation that they would have been exposed (likely repeatedly) and infected with *M.tb* and were, therefore, at high risk of developing TB. A total of 4,466 household contacts were followed for two years and whole blood was collected at recruitment (time zero), 6 and 18 months thereafter for RNAseq. Further details of the study, including ethical approval details have been described previously [22,23]. The samples have been used in several previous publications for validation and discovery of TB biomarkers [22,23]. The raw RNAseq FASTQ files for the study were downloaded from the GEO public database (accession number GSE94438).

The second data set used included a subset of RNAseq data from the South African Catalysis study (GEO: GSE89403) which was a longitudinal study of resolution of lung inflammation in TB cases. Here, we only used RNAseq data from TB patients at diagnosis and the non-TB controls [27,28] to enlarge the size of the sample of TB at diagnosis patients. The study recruited adult HIV-negative TB patients, whose diagnosis was confirmed with a sputum culture [27,28]. Asymptomatic individuals that were recruited from the same community and tested negative for TB on sputum and chest X-ray were used as controls. The results of 18-F fluorodeoxyglucose (FDG) PET-CT scans from TB patients at enrollment were used to categorize patients into groups with high and low levels of lung inflammation activity based on PET. The metric used is a sum of total glycolytic activity index (TGAI) [27,67] of all metabolically active lesions and the products of the mean lesion intensities with the cavity volumes, abbreviated ComTGAI, and is correlated with levels of lung inflammation [27,68] Patients were dichotomized into low and high PET scores based on a threshold of 4,000 units at baseline.

# RNAseq and quality control

The data used in the current study had been obtained with RNA extracted from whole blood and sequenced on the Illumina HiSeq-4000 (GC6-74) and HiSeq-2000 (Catalysis) platforms generating 50 bp stranded paired-end reads. We used the FastQC program (version 0.11.5) [69] to assess the quality of the reads.

# Read mapping

The Spliced Transcripts Alignment to a Reference (STAR) software (version STAR\_2.5.3a) [70] was used to map reads to the Ensembl [71] human genome primary assembly (version GRCh38.89). The quantMode GeneCounts option was selected to generate raw genewise read counts for each sample.

# Differential expression analysis

The DE analysis was performed in R [72] using the edgeR (version 3.26.8) [73] Bioconductor [74] package. Briefly, raw counts were filtered to remove genes with low expression, normalized, and negative binomial generalized linear models were fitted. In addition to time before diagnosis, sex and site were included as factors in the model matrix since they were observed to be responsible for the separation of principal component (PC)1 and PC2 based on multidimensional scaling plots (data not shown).

The linear model applied was:

$$Gene = 0 + Time + sex + site$$

In R syntax as: `model.matrix(~ 0 +TimeBf+ sex + site)`, i.e., a model without an intercept

The quasi-likelihood F-test (QLF-Test) was used to determine significance and identify DE genes. For the GC6-74 data set, all TB progressor groups were compared to the non-TB baseline group (time 0). For the Catalysis data set, the TB patients were compared to non-TB controls. TB patients categorized with high and low PET scores were also separately compared to controls and directly to each other to identify

differences. Significantly DE genes were selected that had a false discovery rate (FDR) <0.05.

To determine the influence that the larger sample sizes [1–6 months before (n=47) and TB (n=90) compared to 13–18 months before (n=18) and 7–12 months before (n=19), see Table 1] had on the number of DE genes detected, the 1–6 months before diagnosis and TB groups with larger samples were subsampled. These groups were subsampled 1000 times at a sample size of 19 and a DE analysis was performed for each iteration.

## Gene Ontology and KEGG pathway analysis

The topGO [75] Bioconductor package was used to test for enrichments in any GO terms [76] associated with the DE genes. The GO graph structure was generated using both the “classic” and “weight01”, algorithms using the Fisher’s exact test to identify enriched terms. In brief, the classic algorithm tests each GO category independently. The “weight01” is the default algorithm used by the topGO package and essentially penalizes scores for more general terms that share genes with more specific neighboring terms, weighting the analysis for the identification of more specific and therefore informative ontologies. “Weight01” is a mixture of two algorithms, “elim” which eliminates genes shared between a node and its ancestor, and “weight” which weights genes locally as a function of the ratio between child and ancestor nodes [75]. The weight01 mixture of the two algorithms moderates the extremes of the two individual algorithms.

The kegg function in edgeR was used to perform a KEGG [77] pathway enrichment analysis on the DE gene sets. The Bioconductor [74] pathview package [29] was subsequently used to visualize gene expression on enriched KEGG pathway graphs.

Due to the extensive overlap of genes in the hierarchical GO categories, neither the topGO or kegg programs correct for multiple hypothesis testing and recommend against it. We therefore applied a stringent p-value cut-off of  $<5 \times 10^{-4}$  to identify enriched pathways.



# Acknowledgements

The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project.

Members of The Bill and Melinda Gates Foundation Grand Challenges 6, 74 (GC6-74) “Biomarkers of Protective Immunity Against TB in the Context of HIV/AIDS in Africa” Consortium: Gerhard Walzl<sup>1,2,3</sup>, Gillian F. Black<sup>1,2,3</sup>, Gian van der Spuy<sup>1,2,3,4</sup>, Kim Stanley<sup>1,2,3</sup>, Magdalena Kriel<sup>1,2,3</sup>, Nelita Du Plessis<sup>1,2,3</sup>, Nonhlanhla Nene<sup>1,2,3</sup>, Teri Roberts<sup>1,2,3</sup>, Léanie Kleynhans<sup>1,2,3</sup>, Andrea Gutschmidt<sup>1,2,3</sup>, Bronwyn Smith<sup>1,2,3</sup>, Andre G. Loxton<sup>1,2,3</sup>, Novel N. Chegou<sup>1,2,3</sup>, Gerard Tromp<sup>1,2,3,4,5</sup>, David Tabb<sup>1,2,3,4,5</sup>, Tom H. M. Ottenhoff<sup>6</sup>, Michel R. Klein<sup>6</sup>, Marielle C. Haks<sup>6</sup>, Kees L. M. C. Franken<sup>6</sup>, Annemieke Geluk<sup>6</sup>, Krista E. van Meijgaarden<sup>6</sup>, Simone A. Joosten<sup>6</sup>, W. Henry Boom<sup>7</sup>, Bonnie Thiel<sup>7</sup>, Harriet Mayanja-Kizza<sup>8</sup>, Moses Joloba<sup>8</sup>, Sarah Zalwango<sup>8</sup>, Mary Nsereko<sup>8</sup>, Brenda Okwera<sup>8</sup>, Hussein Kisingo<sup>8</sup>, Stefan H. E. Kaufmann<sup>9</sup> (GC6-74 principal investigator), Shreemanta K. Parida<sup>9</sup>, Robert Golinski<sup>9</sup>, Jeroen Maertzdorf<sup>9</sup>, January Weiner 3rd<sup>9</sup>, Marc Jacobson<sup>9</sup>, Hazel M. Dockrell<sup>10</sup>, Maeve Lalor<sup>10</sup>, Steven Smith<sup>10</sup>, Patricia Gorak-Stolinska<sup>10</sup>, Yun-Gyoung Hur<sup>10</sup>, Ji-Sook Lee<sup>10</sup>, Amelia C. Crampin<sup>11</sup>, Neil French<sup>11</sup>, Bagrey Ngwira<sup>11</sup>, Anne Ben-Smith<sup>11</sup>, Kate Watkins<sup>11</sup>, Lyn Ambrose<sup>11</sup>, Felanji Simukonda<sup>11</sup>, Hazzie Mvula<sup>11</sup>, Femia Chilongo<sup>11</sup>, Jacky Saul<sup>11</sup>, Keith Branson<sup>11</sup>, Sara Suliman<sup>12</sup>, Thomas J. Scriba<sup>12</sup>, Hassan Mahomed<sup>12</sup>, E. Jane Hughes<sup>12</sup>, Nicole Bilek<sup>12</sup>, Mzwandile Erasmus<sup>12</sup>, Onke Xasa<sup>12</sup>, Ashley Veldsman<sup>12</sup>, Katrina Downing<sup>12</sup>, Michelle Fisher<sup>12</sup>, Adam Penn-Nicholson<sup>12</sup>, Humphrey Mulenga<sup>12</sup>, Brian Abel<sup>12</sup>, Mark Bowmaker<sup>12</sup>, Benjamin Kagina<sup>12</sup>, William Kwong Chung<sup>12</sup>, Willem A. Hanekom<sup>12</sup>, Jerry Sadoff<sup>13</sup>, Donata Sizemore<sup>13</sup>, S. Ramachandran<sup>13</sup>, Lew Barker<sup>13</sup>, Michael Brennan<sup>13</sup>, Frank Weichold<sup>13</sup>, Stefanie Muller<sup>13</sup>, Larry Geiter<sup>13</sup>, Desta Kassa<sup>14</sup>, Almaz Abebe<sup>14</sup>, Tsehayenesh Mesele<sup>14</sup>, Belete Tegbaru<sup>14</sup>, Debbie van Baarle<sup>15</sup>, Frank Miedema<sup>15</sup>, Rawleigh Howe<sup>16</sup>, Adane Mihret<sup>16</sup>, Abraham Aseffa<sup>16</sup>, Yonas Bekele<sup>16</sup>, Rachel Iwnetu<sup>16</sup>, Mesfin Tafesse<sup>16</sup>, Lawrence Yamuah<sup>16</sup>, Martin Ota<sup>17</sup>, Jayne Sutherland<sup>17</sup>, Philip Hill<sup>17</sup>, Richard Adegbola<sup>17</sup>, Tumani Corrah<sup>17</sup>, Martin Antonio<sup>17</sup>, Toyin Togun<sup>17</sup>, Ifedayo Adetifa<sup>17</sup>, Simon Donkor<sup>17</sup>, Peter Andersen<sup>18</sup>, Ida Rosenkrands<sup>18</sup>, Mark Doherty<sup>18</sup>, Karin Weldingh<sup>18</sup>, Gary Schoolnik<sup>19</sup>, Gregory Dolganov<sup>19</sup>, and Tran Van<sup>19</sup>

1. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Stellenbosch University, Cape Town, South Africa
2. Department of Science and Innovation/National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Cape Town, South Africa
3. South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa
4. South African Tuberculosis Bioinformatics Initiative, Stellenbosch University, Cape Town, South Africa
5. Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa
6. Department of Infectious Diseases, Leiden University Medical Centre, Leiden, the Netherlands
7. Tuberculosis Research Unit, Department of Medicine, Case Western Reserve University School of Medicine and University Hospitals Case Medical Center, Cleveland, Ohio, USA



8. Department of Medicine and Department of Microbiology, College of Health Sciences, Faculty of Medicine, Makerere University, Kampala, Uganda
9. Department of Immunology, Max Planck Institute for Infection Biology, Berlin, Germany
10. Department of Immunology and Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
11. Karonga Prevention Study, Chilumba, Malawi
12. South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa
13. Aeras, Rockville, Maryland, USA
14. Ethiopian Health and Nutrition Research Institute, Addis Ababa, Ethiopia
15. University Medical Centre, Utrecht, the Netherlands
16. Armauer Hansen Research Institute, Addis Ababa, Ethiopia
17. Vaccines and Immunity Theme, Medical Research Council Unit, Fajara, the Gambia
18. Department of Infectious Disease Immunology, Statens Serum Institute, Copenhagen, Denmark
19. Department of Microbiology and Immunology, Stanford University, Stanford, California, USA

Members of The Catalysis Biomarker Consortium: Stephanus T. Malherbe<sup>1,2</sup>, Patrick Dupont<sup>3,4</sup>, Ilse Kant<sup>4</sup>, Katharina Ronacher<sup>1,2</sup>, Magdalena Kriel<sup>1,2</sup>, André G. Loxton<sup>1,2</sup>, Ray Y. Chen<sup>5</sup>, Laura E. Via<sup>5,6</sup>, Friedrich Thienemann<sup>6,7</sup>, Robert J. Wilkinson<sup>6,7,8,9</sup>, Clifton E. Barry III<sup>1,2,5,6</sup>, Stephanie Griffith-Richards<sup>10</sup>, Annare Ellman<sup>4</sup>, Jill Winter<sup>11</sup>, Gerhard Walzl<sup>1,2</sup>, Nelita Du Plessis<sup>1,2</sup>, Caroline G.G. Beltran<sup>1,2</sup>, Lani Thiar<sup>1,2</sup>, Gerard Tromp<sup>1,2</sup>, Lance A. Lucas<sup>1,2</sup>, Bronwyn Smith<sup>1,2</sup>, Kim Stanley<sup>1,2</sup>, David Alland<sup>12</sup>, Shubhada Shenai<sup>12</sup>, Lori E. Dodd<sup>13</sup>, and James M. Warwick<sup>4</sup>.

1. Department of Science and Innovation/National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research and South African Medical Research Council Centre for Tuberculosis Research.
2. Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
3. Laboratory for Cognitive Neurology, Department of Neurosciences, Katolieke Universiteit, Leuven, Belgium.
4. Division of Nuclear Medicine, Department of Medical Imaging and Clinical Oncology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
5. Tuberculosis Research Section, Laboratory of Clinical Infectious Diseases, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA.
6. Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Science, University of Cape Town, Cape Town, South Africa
7. Department of Medicine, Groote Schuur Hospital, Faculty of Health Science, University of Cape Town, Cape Town, South Africa.
8. The Francis Crick Institute, London, United Kingdom
9. Department of Medicine, Imperial College London, London, United Kingdom
10. Division of Radiodiagnosis, Department of Medical Imaging and Clinical Oncology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
11. Catalysis Foundation for Health, Emeryville, California, USA.
12. Center for Emerging Pathogens, Department of Medicine, Rutgers-New Jersey Medical School, Rutgers Biomedical and Health Sciences, Newark, New Jersey, USA.

13. Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health Bethesda, Maryland, USA.

# References

1. World Health Organization. Global tuberculosis report 2020. 2020. 208 p.
2. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. *Annu Rev Immunol*. 2013;31:475-527. Epub 2013/03/23. doi: 10.1146/annurev-immunol-032712-095939. PubMed PMID: 23516984.
3. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. *Nat Rev Immunol*. 2017;17(11):691-702. Epub 2017/07/25. doi: 10.1038/nri.2017.69. PubMed PMID: 28736436; PubMed Central ID: PMC6247113.
4. Elkington P, Tebruegge M, Mansour S. Tuberculosis: An Infection-Initiated Autoimmune Disease? *Trends Immunol*. 2016;37(12):815-8. Epub 2016/10/25. doi: 10.1016/j.it.2016.09.007. PubMed PMID: 27773684; PubMed Central ID: PMC5135577.
5. Hunter RL. Tuberculosis as a three-act play: A new paradigm for the pathogenesis of pulmonary tuberculosis. *Tuberculosis (Edinb)*. 2016;97:8-17. Epub 2016/03/17. doi: 10.1016/j.tube.2015.11.010. PubMed PMID: 26980490; PubMed Central ID: PMC4795183.
6. Ong CW, Elkington PT, Friedland JS. Tuberculosis, pulmonary cavitation, and matrix metalloproteinases. *Am J Respir Crit Care Med*. 2014;190(1):9-18. Epub 2014/04/10. doi: 10.1164/rccm.201311-2106PP. PubMed PMID: 24713029; PubMed Central ID: PMC4226026.
7. Alemán M, Beigier-Bompadre M, Borghetti C, de la Barrera S, Abbate E, Isturiz M, et al. Activation of peripheral blood neutrophils from patients with active advanced tuberculosis. *Clin Immunol*. 2001;100(1):87-95. Epub 2001/06/21. doi: 10.1006/clim.2001.5044. PubMed PMID: 11414749.
8. Fiorenza G, Farroni MA, Bogue C, Selenscig D, Lamas DM, Dlugovitzky D. Functional characteristics of neutrophils and mononuclear cells from tuberculosis patients stimulated in vitro with heat killed M. tuberculosis. *Arch Med Res*.

2007;38(5):526-33. Epub 2007/06/15. doi: 10.1016/j.arcmed.2007.01.011. PubMed  
PMID: 17560458.

9. Lyadova IV. Neutrophils in Tuberculosis: Heterogeneity Shapes the Way? Mediators  
Inflamm. 2017;2017:8619307. Epub 2017/06/20. doi: 10.1155/2017/8619307.

PubMed PMID: 28626346; PubMed Central ID: PMC5463159.

10. Barry CE, 3rd, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, et al. The spectrum of  
latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev  
Microbiol. 2009;7(12):845-55. Epub 2009/10/27. doi: 10.1038/nrmicro2236. PubMed  
PMID: 19855401; PubMed Central ID: PMC4144869.

11. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-  
inducible neutrophil-driven blood transcriptional signature in human tuberculosis.  
Nature. 2010;466(7309):973-7. Epub 2010/08/21. doi: 10.1038/nature09247.  
PubMed PMID: 20725040; PubMed Central ID: PMC3492754.

12. Ong CW, Elkington PT, Brilha S, Ugarte-Gil C, Tome-Esteban MT, Tezera LB, et al.  
Neutrophil-Derived MMP-8 Drives AMPK-Dependent Matrix Destruction in Human  
Pulmonary Tuberculosis. PLoS Pathog. 2015;11(5):e1004917. Epub 2015/05/23.  
doi: 10.1371/journal.ppat.1004917. PubMed PMID: 25996154; PubMed Central ID:  
PMC4440706.

13. Sutherland JS, Jeffries DJ, Donkor S, Walther B, Hill PC, Adetifa IM, et al. High  
granulocyte/lymphocyte ratio and paucity of NKT cells defines TB disease in a TB-  
endemic setting. Tuberculosis (Edinb). 2009;89(6):398-404. Epub 2009/08/18. doi:  
10.1016/j.tube.2009.07.004. PubMed PMID: 19683473.

14. Eum SY, Kong JH, Hong MS, Lee YJ, Kim JH, Hwang SH, et al. Neutrophils are the  
predominant infected phagocytic cells in the airways of patients with active  
pulmonary TB. Chest. 2010;137(1):122-8. Epub 2009/09/15. doi: 10.1378/chest.09-  
0903. PubMed PMID: 19749004; PubMed Central ID: PMC2803122.

15. Ugarte-Gil CA, Elkington P, Gilman RH, Coronel J, Tezera LB, Bernabe-Ortiz A, et  
al. Induced sputum MMP-1, -3 & -8 concentrations during treatment of tuberculosis.  
PLoS One. 2013;8(4):e61333. Epub 2013/04/25. doi:  
10.1371/journal.pone.0061333. PubMed PMID: 23613834; PubMed Central ID:  
PMC3632571.

16. Petretto A, Bruschi M, Pratesi F, Croia C, Candiano G, Ghiggeri G, et al. Neutrophil extracellular traps (NET) induced by different stimuli: A comparative proteomic analysis. *PLoS One*. 2019;14(7):e0218946. Epub 2019/07/10. doi: 10.1371/journal.pone.0218946. PubMed PMID: 31283757; PubMed Central ID: PMC6613696.
17. Blankley S, Berry MP, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos Trans R Soc Lond B Biol Sci*. 2014;369(1645):20130427. Epub 2014/05/14. doi: 10.1098/rstb.2013.0427. PubMed PMID: 24821914; PubMed Central ID: PMC4024221.
18. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med*. 2013;10(10):e1001538. Epub 2013/10/30. doi: 10.1371/journal.pmed.1001538. PubMed PMID: 24167453; PubMed Central ID: PMC3805485 Risk score (GB1201766.1) and TB/LTBI and TB/OD signatures (GB1213636.2).
19. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med*. 2016;4(3):213-24. Epub 2016/02/26. doi: 10.1016/s2213-2600(16)00048-5. PubMed PMID: 26907218; PubMed Central ID: PMC4838193.
20. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One*. 2012;7(9):e45839. Epub 2012/10/03. doi: 10.1371/journal.pone.0045839. PubMed PMID: 23029268; PubMed Central ID: PMC3448682.
21. Joosten SA, Fletcher HA, Ottenhoff TH. A helicopter perspective on TB biomarkers: pathway and process based analysis of gene expression data provides new insight into TB pathogenesis. *PLoS One*. 2013;8(9):e73230. Epub 2013/09/26. doi: 10.1371/journal.pone.0073230. PubMed PMID: 24066041; PubMed Central ID: PMC3774688.

- 783 22. Suliman S, Thompson EG, Sutherland J, Weiner J, 3rd, Ota MOC, Shankar S, et al.  
Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am J*  
Respir Crit Care Med. 2018;197(9):1198-208. Epub 2018/04/07. doi:  
786 10.1164/rccm.201711-2340OC. PubMed PMID: 29624071; PubMed Central ID:  
PMC6019933.
23. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A  
789 blood RNA signature for tuberculosis disease risk: a prospective cohort study.  
*Lancet*. 2016;387(10035):2312-22. Epub 2016/03/28. doi: 10.1016/s0140-  
6736(15)01316-1. PubMed PMID: 27017310; PubMed Central ID: PMC5392204.
- 792 24. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Horswell S, et al.  
Complement pathway gene activation and rising circulating immune complexes  
characterize early disease in HIV-associated tuberculosis. *Proc Natl Acad Sci U S*  
795 *A*. 2018;115(5):E964-e73. Epub 2018/01/18. doi: 10.1073/pnas.1711853115.  
PubMed PMID: 29339504; PubMed Central ID: PMC5798330.
25. Scriba TJ, Penn-Nicholson A, Shankar S, Hraha T, Thompson EG, Sterling D, et al.  
798 Sequential inflammatory processes define human progression from *M. tuberculosis*  
infection to tuberculosis disease. *PLoS Pathog*. 2017;13(11):e1006687. Epub  
20171116. doi: 10.1371/journal.ppat.1006687. PubMed PMID: 29145483; PubMed  
801 Central ID: PMC5689825.
26. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI  
GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*.  
804 2013;41(Database issue):D991-5. Epub 2012/11/30. doi: 10.1093/nar/gks1193.  
PubMed PMID: 23193258; PubMed Central ID: PMC3531084.
27. Malherbe ST, Shenai S, Ronacher K, Loxton AG, Dolganov G, Kriel M, et al.  
807 Persisting positron emission tomography lesion activity and *Mycobacterium*  
*tuberculosis* mRNA after tuberculosis cure. *Nat Med*. 2016;22(10):1094-100. Epub  
2016/09/07. doi: 10.1038/nm.4177. PubMed PMID: 27595324; PubMed Central ID:  
810 PMC5053881.
28. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, et al. Host blood  
RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinb)*.

- 2017;107:48-58. Epub 2017/10/21. doi: 10.1016/j.tube.2017.08.004. PubMed PMID: 29050771; PubMed Central ID: PMC5658513.
29. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. 2013;29(14):1830-1. Epub 2013/06/07. doi: 10.1093/bioinformatics/btt285. PubMed PMID: 23740750; PubMed Central ID: PMC3702256.
30. Ostermann G, Weber KS, Zerneck A, Schröder A, Weber C. JAM-1 is a ligand of the beta(2) integrin LFA-1 involved in transendothelial migration of leukocytes. *Nat Immunol*. 2002;3(2):151-8. Epub 2002/01/29. doi: 10.1038/ni755. PubMed PMID: 11812992.
31. Futosi K, Fodor S, Mócsai A. Neutrophil cell surface receptors and their intracellular signal transduction pathways. *Int Immunopharmacol*. 2013;17(3):638-50. Epub 2013/09/03. doi: 10.1016/j.intimp.2013.06.034. PubMed PMID: 23994464; PubMed Central ID: PMC3827506.
32. Simard JC, Girard D, Tessier PA. Induction of neutrophil degranulation by S100A9 via a MAPK-dependent mechanism. *J Leukoc Biol*. 2010;87(5):905-14. doi: 10.1189/jlb.1009676. PubMed PMID: 20103766.
33. El-Benna J, Hurtado-Nedelec M, Marzaioli V, Marie JC, Gougerot-Pocidalo MA, Dang PM. Priming of the neutrophil respiratory burst: role in host defense and inflammation. *Immunol Rev*. 2016;273(1):180-93. Epub 2016/08/26. doi: 10.1111/imr.12447. PubMed PMID: 27558335.
34. Nguyen GT, Green ER, Meccas J. Neutrophils to the ROScue: Mechanisms of NADPH Oxidase Activation and Bacterial Resistance. *Front Cell Infect Microbiol*. 2017;7:373. Epub 2017/09/12. doi: 10.3389/fcimb.2017.00373. PubMed PMID: 28890882; PubMed Central ID: PMC5574878.
35. Nunes P, Demarex N, Dinanier MC. Regulation of the NADPH oxidase and associated ion fluxes during phagocytosis. *Traffic*. 2013;14(11):1118-31. Epub 2013/08/29. doi: 10.1111/tra.12115. PubMed PMID: 23980663.
36. Fuchs TA, Abed U, Goosmann C, Hurwitz R, Schulze I, Wahn V, et al. Novel cell death program leads to neutrophil extracellular traps. *J Cell Biol*. 2007;176(2):231-



41. Epub 2007/01/11. doi: 10.1083/jcb.200606027. PubMed PMID: 17210947; PubMed Central ID: PMC2063942.
37. Papayannopoulos V, Metzler KD, Hakkim A, Zychlinsky A. Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *J Cell Biol.* 2010;191(3):677-91. Epub 2010/10/27. doi: 10.1083/jcb.201006052. PubMed PMID: 20974816; PubMed Central ID: PMC3003309.
38. Manfredi AA, Ramirez GA, Rovere-Querini P, Maugeri N. The Neutrophil's Choice: Phagocytosis vs Make Neutrophil Extracellular Traps. *Front Immunol.* 2018;9:288. Epub 2018/03/09. doi: 10.3389/fimmu.2018.00288. PubMed PMID: 29515586; PubMed Central ID: PMC5826238.
39. Rossaint J, Margraf A, Zarbock A. Role of Platelets in Leukocyte Recruitment and Resolution of Inflammation. *Front Immunol.* 2018;9:2712. Epub 2018/12/06. doi: 10.3389/fimmu.2018.02712. PubMed PMID: 30515177; PubMed Central ID: PMC6255980.
40. Young C, Ahlers P, Hiemstra AM, Loxton AG, Gutschmidt A, Malherbe ST, et al. Performance and immune characteristics of bronchoalveolar lavage by research bronchoscopy in pulmonary tuberculosis and other lung diseases in the Western Cape, South Africa. *Translational Medicine Communications.* 2019;4(1):7. doi: 10.1186/s41231-019-0039-2.
41. Diedrich CR, O'Hern J, Gutierrez MG, Allie N, Papier P, Meintjes G, et al. Relationship Between HIV Coinfection, Interleukin 10 Production, and *Mycobacterium tuberculosis* in Human Lymph Node Granulomas. *J Infect Dis.* 2016;214(9):1309-18. Epub 2016/07/28. doi: 10.1093/infdis/jiw313. PubMed PMID: 27462092; PubMed Central ID: PMC5079364.
42. Miralda I, Uriarte SM, McLeish KR. Multiple Phenotypic Changes Define Neutrophil Priming. *Front Cell Infect Microbiol.* 2017;7:217. Epub 2017/06/15. doi: 10.3389/fcimb.2017.00217. PubMed PMID: 28611952; PubMed Central ID: PMC5447094.
43. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. *Science.* 2004;303(5663):1532-5. Epub 2004/03/06. doi: 10.1126/science.1092385. PubMed PMID: 15001782.

44. Jorch SK, Kubes P. An emerging role for neutrophil extracellular traps in noninfectious disease. *Nat Med.* 2017;23(3):279-87. Epub 2017/03/08. doi: 10.1038/nm.4294. PubMed PMID: 28267716.
45. Kegerreis BJ, Catalina MD, Geraci NS, Bachali P, Lipsky PE, Grammer AC. Genomic Identification of Low-Density Granulocytes and Analysis of Their Role in the Pathogenesis of Systemic Lupus Erythematosus. *J Immunol.* 2019;202(11):3309-17. Epub 2019/04/26. doi: 10.4049/jimmunol.1801512. PubMed PMID: 31019061.
46. Borregaard N, Sørensen OE, Theilgaard-Mönch K. Neutrophil granules: a library of innate immunity proteins. *Trends Immunol.* 2007;28(8):340-5. Epub 2007/07/14. doi: 10.1016/j.it.2007.06.002. PubMed PMID: 17627888.
47. Lawrence SM, Corriden R, Nizet V. The Ontogeny of a Neutrophil: Mechanisms of Granulopoiesis and Homeostasis. *Microbiol Mol Biol Rev.* 2018;82(1). Epub 2018/02/14. doi: 10.1128/mmbr.00057-17. PubMed PMID: 29436479; PubMed Central ID: PMC5813886.
48. Jourde-Chiche N, Whalen E, Gondouin B, Speake C, Gersuk V, Dussol B, et al. Modular transcriptional repertoire analyses identify a blood neutrophil signature as a candidate biomarker for lupus nephritis. *Rheumatology (Oxford).* 2017;56(3):477-87. Epub 2016/12/30. doi: 10.1093/rheumatology/kew439. PubMed PMID: 28031441.
49. Deng Y, Ye J, Luo Q, Huang Z, Peng Y, Xiong G, et al. Low-Density Granulocytes Are Elevated in Mycobacterial Infection and Associated with the Severity of Tuberculosis. *PLoS One.* 2016;11(4):e0153567. Epub 2016/04/14. doi: 10.1371/journal.pone.0153567. PubMed PMID: 27073889; PubMed Central ID: PMC4830625.
50. Su R, Peng YP, Deng Z, Deng YT, Ye JQ, Guo Y, et al. Mycobacterium tuberculosis Infection Induces Low-Density Granulocyte Generation by Promoting Neutrophil Extracellular Trap Formation via ROS Pathway. *Front Microbiol.* 2019;10:1468. Epub 2019/07/30. doi: 10.3389/fmicb.2019.01468. PubMed PMID: 31354639; PubMed Central ID: PMC6637951.
51. Francis RJ, Butler RE, Stewart GR. Mycobacterium tuberculosis ESAT-6 is a leukocidin causing Ca<sup>2+</sup> influx, necrosis and neutrophil extracellular trap formation.



Cell Death Dis. 2014;5(10):e1474. Epub 2014/10/17. doi: 10.1038/cddis.2014.394.

PubMed PMID: 25321481; PubMed Central ID: PMC4237235.

52. Ramos-Kichik V, Mondragón-Flores R, Mondragón-Castelán M, Gonzalez-Pozos S, Muñiz-Hernandez S, Rojas-Espinosa O, et al. Neutrophil extracellular traps are induced by Mycobacterium tuberculosis. Tuberculosis (Edinb). 2009;89(1):29-37.

Epub 2008/12/06. doi: 10.1016/j.tube.2008.09.009. PubMed PMID: 19056316.

53. Cardona PJ. The Progress of Therapeutic Vaccination with Regard to Tuberculosis.

Front Microbiol. 2016;7:1536. Epub 2016/10/14. doi: 10.3389/fmicb.2016.01536.

PubMed PMID: 27733848; PubMed Central ID: PMC5039189.

54. Ehlers S, Schaible UE. The granuloma in tuberculosis: dynamics of a host-pathogen collusion. Front Immunol. 2012;3:411. Epub 2013/01/12. doi:

10.3389/fimmu.2012.00411. PubMed PMID: 23308075; PubMed Central ID: PMC3538277.

55. Schechter MC, Buac K, Adekambi T, Cagle S, Celli J, Ray SM, et al. Neutrophil extracellular trap (NET) levels in human plasma are associated with active TB.

PLoS One. 2017;12(8):e0182587. Epub 2017/08/05. doi:

10.1371/journal.pone.0182587. PubMed PMID: 28777804; PubMed Central ID: PMC5544211.

56. de Melo MGM, Mesquita EDD, Oliveira MM, da Silva-Monteiro C, Silveira AKA, Malaquias TS, et al. Imbalance of NET and Alpha-1-Antitrypsin in Tuberculosis Patients Is Related With Hyper Inflammation and Severe Lung Tissue Damage.

Front Immunol. 2018;9:3147. Epub 2019/01/29. doi: 10.3389/fimmu.2018.03147.

PubMed PMID: 30687336; PubMed Central ID: PMC6335334.

57. Kaplan MJ, Radic M. Neutrophil extracellular traps: double-edged swords of innate immunity. J Immunol. 2012;189(6):2689-95. Epub 2012/09/08. doi:

10.4049/jimmunol.1201719. PubMed PMID: 22956760; PubMed Central ID: PMC3439169.

58. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. Nat

Rev Immunol. 2018;18(2):134-47. Epub 2017/10/11. doi: 10.1038/nri.2017.105.

PubMed PMID: 28990587.

59. Lande R, Ganguly D, Facchinetti V, Frasca L, Conrad C, Gregorio J, et al.  
936 Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide  
complexes in systemic lupus erythematosus. *Sci Transl Med*. 2011;3(73):73ra19.  
Epub 2011/03/11. doi: 10.1126/scitranslmed.3001180. PubMed PMID: 21389263;  
939 PubMed Central ID: PMC3399524.
60. Kim SJ, Jenne CN. Role of platelets in neutrophil extracellular trap (NET) production  
and tissue injury. *Semin Immunol*. 2016;28(6):546-54. Epub 2016/11/24. doi:  
942 10.1016/j.smim.2016.10.013. PubMed PMID: 27876233.
61. McDonald B, Urrutia R, Yipp BG, Jenne CN, Kubes P. Intravascular neutrophil  
extracellular traps capture bacteria from the bloodstream during sepsis. *Cell Host*  
945 *Microbe*. 2012;12(3):324-33. Epub 2012/09/18. doi: 10.1016/j.chom.2012.06.011.  
PubMed PMID: 22980329.
62. Clark SR, Ma AC, Tavener SA, McDonald B, Goodarzi Z, Kelly MM, et al. Platelet  
948 TLR4 activates neutrophil extracellular traps to ensnare bacteria in septic blood. *Nat*  
*Med*. 2007;13(4):463-9. Epub 2007/03/27. doi: 10.1038/nm1565. PubMed PMID:  
17384648.
- 951 63. Büyükaşık Y, Soylu B, Soylu AR, Özcebe OI, Canbakan S, Haznedaroğlu IC, et al.  
In vivo platelet and T-lymphocyte activities during pulmonary tuberculosis. *Eur*  
*Respir J*. 1998;12(6):1375-9. Epub 1999/01/07. doi:  
954 10.1183/09031936.98.12061375. PubMed PMID: 9877494.
64. Lee MY, Kim YJ, Lee HJ, Cho SY, Park TS. Mean Platelet Volume in  
Mycobacterium tuberculosis Infection. *Biomed Res Int*. 2016;2016:7508763. Epub  
957 2016/07/16. doi: 10.1155/2016/7508763. PubMed PMID: 27419136; PubMed  
Central ID: PMC4933846.
65. Fox KA, Kirwan DE, Whittington AM, Krishnan N, Robertson BD, Gilman RH, et al.  
960 Platelets Regulate Pulmonary Inflammation and Tissue Destruction in Tuberculosis.  
*Am J Respir Crit Care Med*. 2018;198(2):245-55. Epub 2018/02/09. doi:  
10.1164/rccm.201710-2102OC. PubMed PMID: 29420060; PubMed Central ID:  
963 PMC6058979.

66. Deniset JF, Kubes P. Recent advances in understanding neutrophils. *F1000Res*. 2016;5:2912. Epub 2017/01/21. doi: 10.12688/f1000research.9691.1. PubMed PMID: 28105328; PubMed Central ID: PMC5225409.
67. Malherbe ST, Dupont P, Kant I, Ahlers P, Kriel M, Loxton AG, et al. A semi-automatic technique to quantify complex tuberculous lung lesions on (18)F-fluorodeoxyglucose positron emission tomography/computerised tomography images. *EJNMMI Res*. 2018;8(1):55. Epub 20180625. doi: 10.1186/s13550-018-0411-7. PubMed PMID: 29943161; PubMed Central ID: PMC6020088.
68. Malherbe ST, Chen RY, Dupont P, Kant I, Kriel M, Loxton AG, et al. Quantitative 18F-FDG PET-CT scan characteristics correlate with tuberculosis treatment response. *EJNMMI Res*. 2020;10(1):8. Epub 20200210. doi: 10.1186/s13550-020-0591-9. PubMed PMID: 32040770; PubMed Central ID: PMC7010890.
69. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010.
70. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. Epub 2012/10/30. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PubMed Central ID: PMC3530905.
71. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49(D1):D884-d91. doi: 10.1093/nar/gkaa942. PubMed PMID: 33137190; PubMed Central ID: PMC7778975.
72. R Development Core Team. R: A Language and Environment for Statistical Computing. 3.3.0 ed. Vienna, Austria: R Foundation for Statistical Computing; 2016.
73. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40. Epub 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central ID: PMC2796818.
74. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*.

2015;12(2):115-21. Epub 2015/01/31. doi: 10.1038/nmeth.3252. PubMed PMID: 25633503; PubMed Central ID: PMC4509590.

75. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006;22(13):1600-7. Epub 2006/04/12. doi: 10.1093/bioinformatics/btl140. PubMed PMID: 16606683.

76. The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*. 2021;49(D1):D325-d34. doi: 10.1093/nar/gkaa1113. PubMed PMID: 33290552; PubMed Central ID: PMC7779012.

77. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi: 10.1093/nar/28.1.27. PubMed PMID: 10592173; PubMed Central ID: PMC102409.

## Supporting information

**S1 Table. List of differentially expressed genes at all time points.** The statistical metrics presented for each comparison include: log2 fold change (log2FC), average log2 counts per million (logCPM), quasi-likelihood F-statistic (F), p-value (PValue) and false discovery rate (FDR). The non-TB time zero baseline group from the GC6 study were the control group for all comparisons prior to TB diagnosis (i.e., 24, 18, 12 and 6 months before diagnosis) while the Catalysis study healthy controls served as the baseline for the TB comparison.

**S2 Table. Gene ontology (GO) enrichment analysis results for significantly upregulated genes at all time points.** The analysis was performed using the topGO R Bioconductor package. The statistical metrics presented for each process include: the total number of genes annotated to the process (Total annotated), the number of genes that were significantly up (N up) -regulated, the percent of total annotated that were up-regulated (% Up) and the number that were expected by chance (Expected). The uncorrected Fisher's exact test p-value (Pvalue) and overall rank for over-representation of the GO term in the set using both the classic and weight01 (W1) algorithms are presented.

### **S3 Table. KEGG pathway enrichment analysis results for significantly**

**differentially expressed genes at all time points.** The analysis was performed using

the kegg function from the edgeR R Bioconductor package. The statistical metrics presented for each pathway include: the total number of genes annotated to the pathway (Total annotated), the number of genes that were significantly up (N Up) or down (N Down) -regulated, the % of the total annotated that were up (% Up) or down (% Down) -regulated and the uncorrected Fisher's exact test p-value (Pvalue) for up and down comparisons (P Up, P Down) for over-representation of the KEGG pathway in the set.

### **S4 Table. KEGG and gene ontology (GO) enrichment analyses for significantly differentially expressed genes between low and high PET score groups for the**

**terms and pathways identified in the main analyses.** The analyses were performed using the topGO and kegg function from the edgeR R Bioconductor packages. The statistical metrics presented are as for S2 Table and S3 Table respectively.

### **S1 Fig. Density plot of subsampling the 1–6 months before TB diagnosis sample numbers.**

Subsampling was performed on the 1–6 month-before TB group to investigate the effect the larger sample size (n=47) had on the number of significantly up-regulated genes identified. A total of 1000 differential expression analysis tests were performed using a sample size of 19 which is similar to the earlier 12 month before (n=19) and 18-month before time points (n=18). The mean number of significantly up-regulated genes identified in the subsampling was 1,436 (95% CI:120, 3,100). This number was lower than the 2,791 identified with the full 47 samples but far higher than the number identified in the 18 (249; n=19) and 12 (492; n=18) months before diagnosis groups.

### **S2 Fig. Density plot of subsampling the TB group sample numbers.**

Subsampling was performed on the Catalysis study TB group to investigate the effect the larger sample size (n=90) had on the number of significantly up-regulated genes identified. A total of 1000 differential expression analysis tests were performed using a sample size of 19 which is similar to the earlier 12 month before (n=19) and 18-month before time points (n=18). The mean number of significantly upregulated genes identified in the

subsampling was 3,485 (95% CI:2,719, 4,190). This was lower than the 4,391 identified with the full 90 samples but far higher than the number identified in the 18 (249; n=19) and 12 (492; n=18) months before diagnosis groups.

## Author Contributions

**Conceptualization and design:** Stuart Meier, Gerhard Walzl, Gerard Tromp

**Acquisition and curation of data:** Stuart Meier, Gerard Tromp, Elizna Maasdorp, James A. Seddon, Gerhard Walzl, Helena Kuivaniemi, Thomas Scriba, Jill Winter, Nelita du Plessis, Léanie Kleynhans, Andre G. Loxton, Stephanus T. Malherbe, Daniel E. Zak, Ethan Thompson, Fergal J. Duffy, Stefan H.E. Kaufmann, Tom H.M. Ottenhoff, Sara Suliman, Jayne S. Sutherland

**Analysis and interpretation of data:** Stuart Meier, Gerard Tromp, Elizna Maasdorp, James A. Seddon, Gerhard Walzl, Helena Kuivaniemi, Thomas Scriba, Jill Winter, Nelita du Plessis, Léanie Kleynhans, Andre G. Loxton, Stephanus T. Malherbe, Daniel E. Zak, Ethan Thompson, Fergal J. Duffy, Stefan H.E. Kaufmann, Tom H.M. Ottenhoff, Sara Suliman, Jayne S. Sutherland

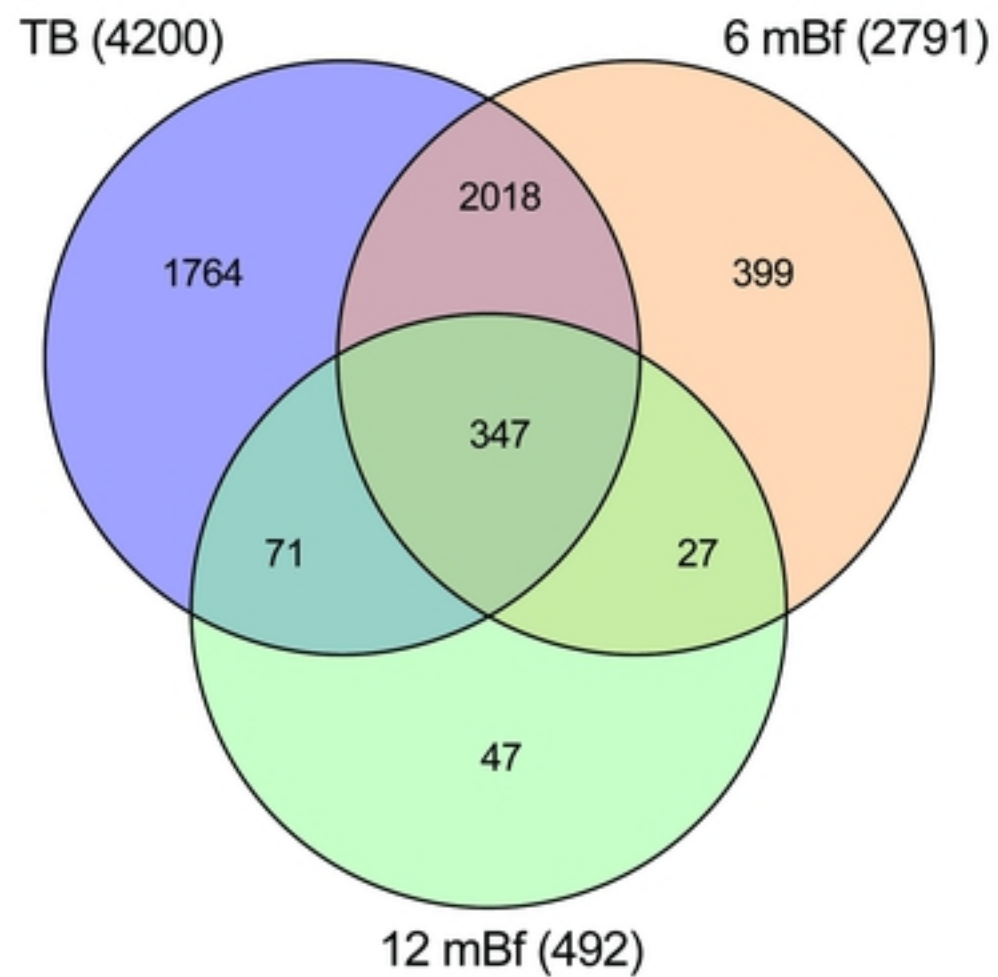
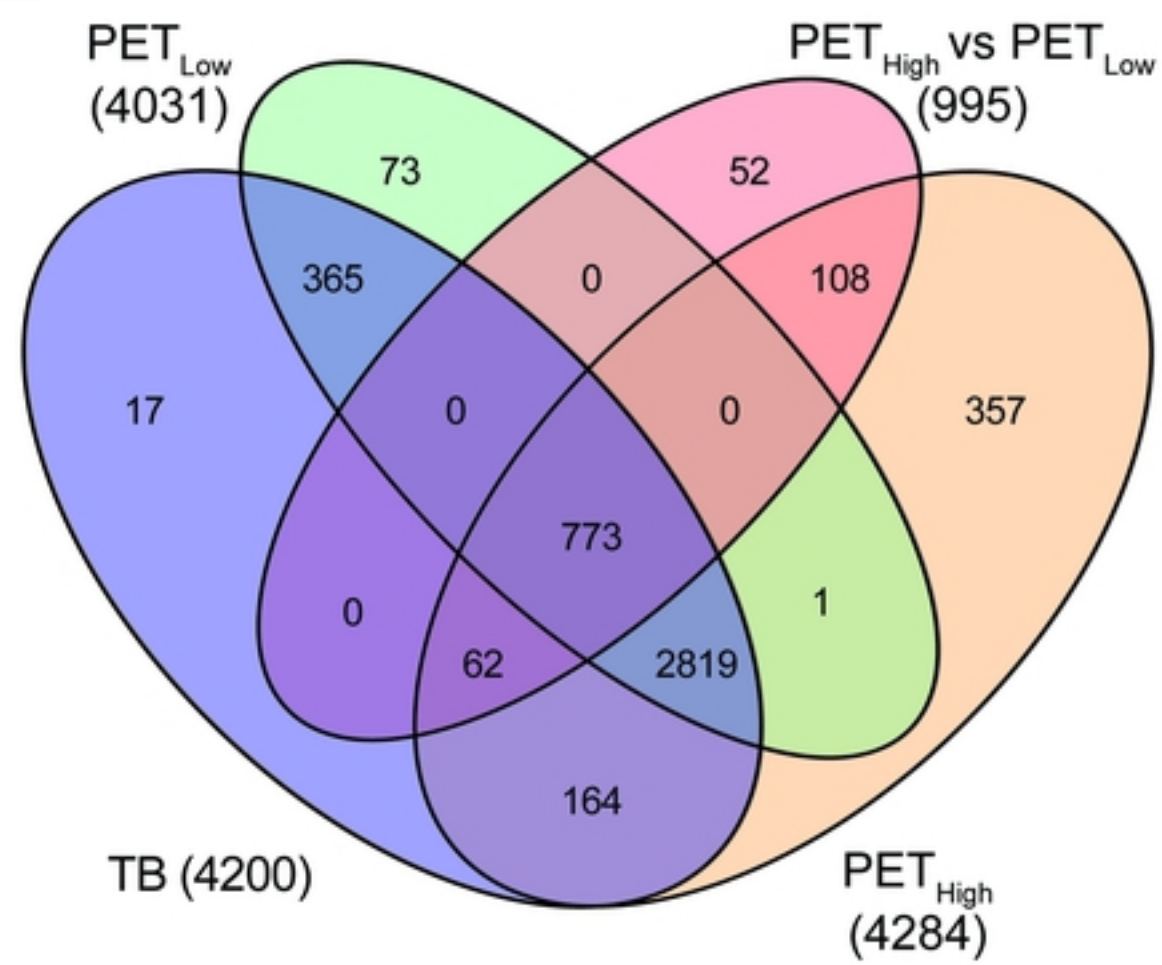
**Funding acquisition:** Gerhard Walzl

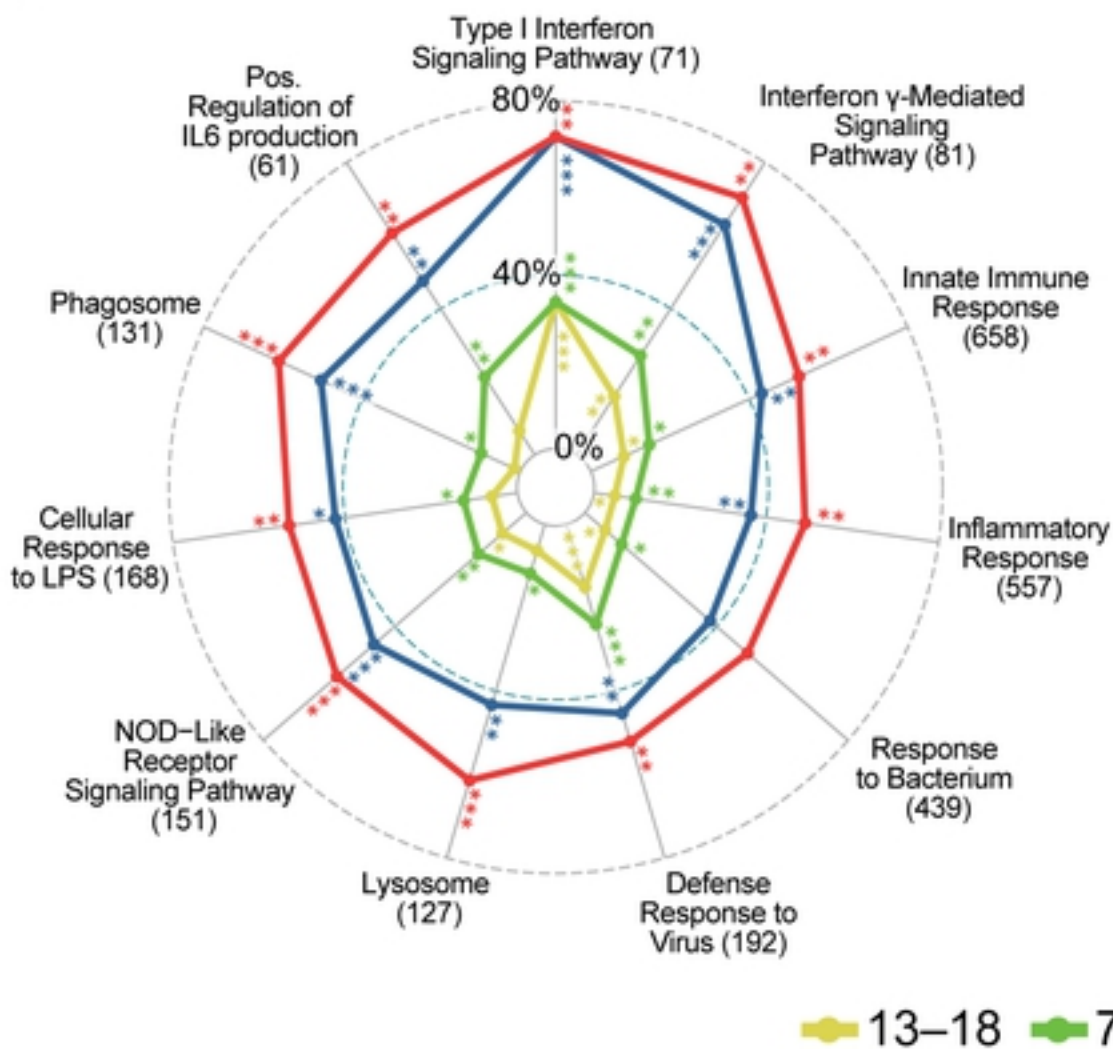
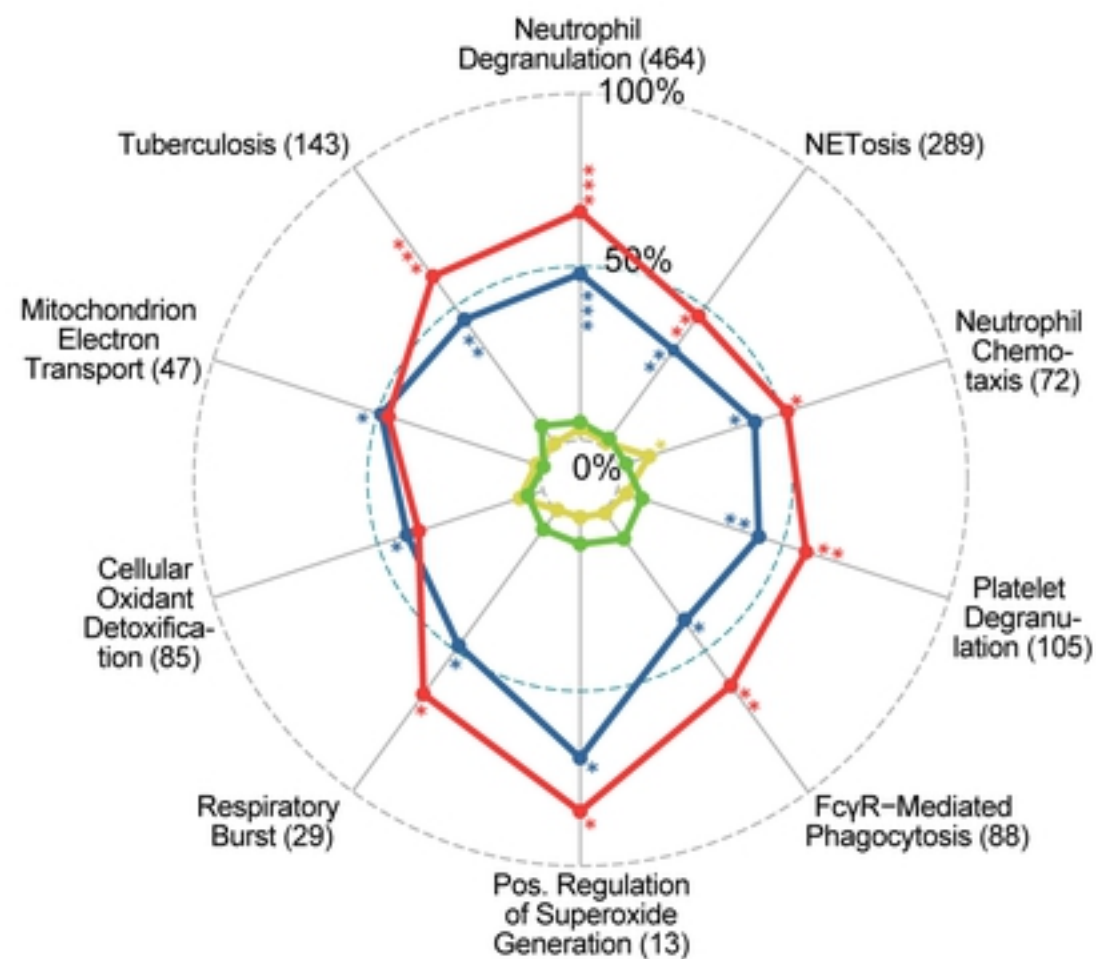
**Writing – original draft:** Stuart Meier, Gerard Tromp, James A. Seddon, Helena Kuivaniemi

**Writing – review & editing:** Stuart Meier, James A. Seddon, Gerhard Walzl, Gerard Tromp, Helena Kuivaniemi, Thomas Scriba, Jill Winter, Nelita du Plessis, Elizna Maasdorp, Léanie Kleynhans, Nelita du Plessis, Andre G. Loxton, Stephanus T. Malherbe, Daniel E. Zak, Ethan Thompson, Fergal J. Duffy, Stefan H.E. Kaufmann, Tom H.M. Ottenhoff, Sara Suliman, Jayne S. Sutherland

**Final approval of the manuscript:** Stuart Meier, James A. Seddon, Elizna Maasdorp, Léanie Kleynhans, Nelita du Plessis, Andre G. Loxton, Stephanus T. Malherbe, Daniel E. Zak, Ethan Thompson, Fergal J. Duffy, Stefan H.E. Kaufmann, Tom H.M. Ottenhoff, Thomas J. Scriba, Sara Suliman, Jayne S. Sutherland, Jill Winter, Helena Kuivaniemi, Gerhard Walzl, Gerard Tromp

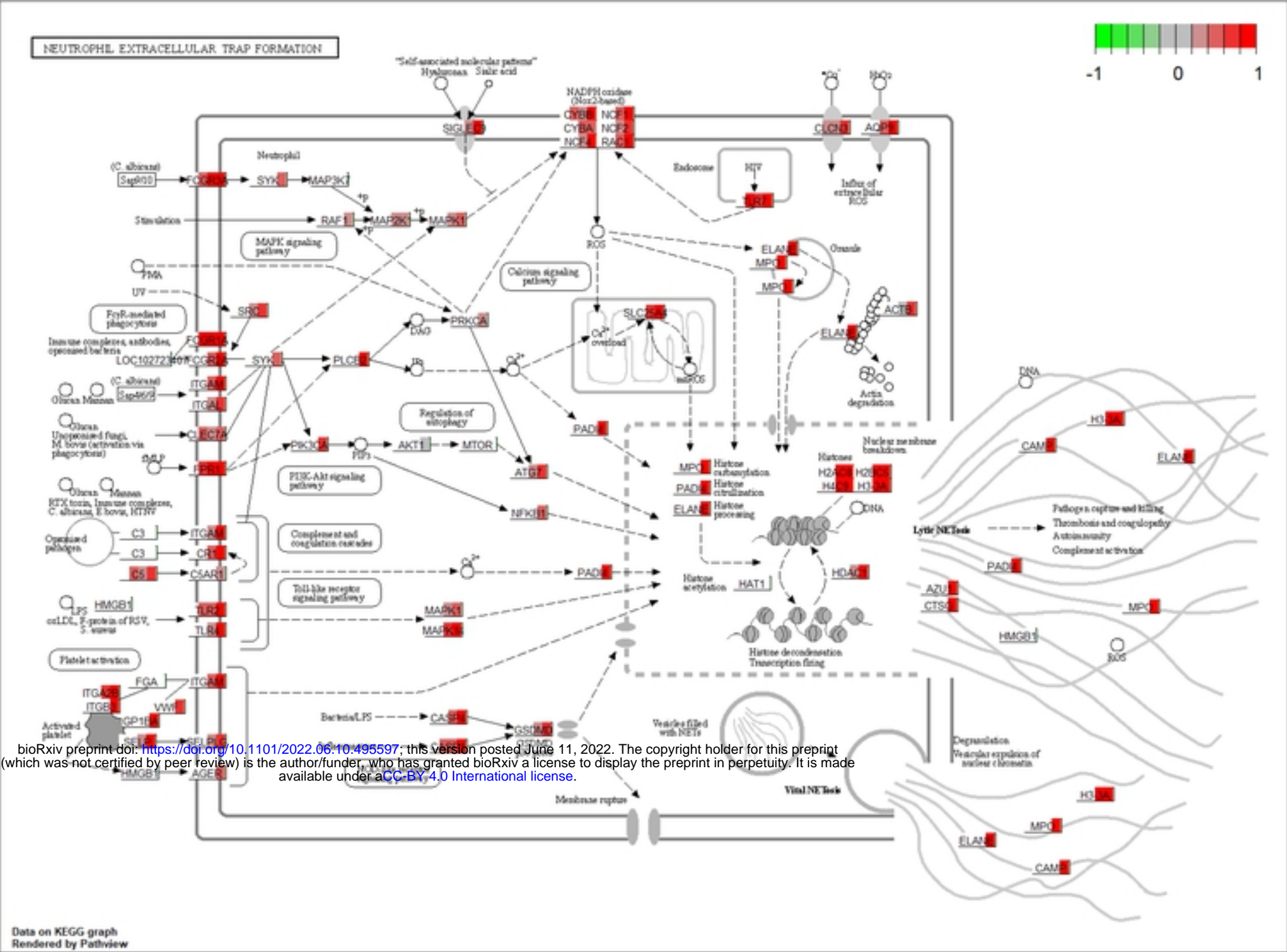


**A****B****Fig 1**

**A****B****Fig 2**



A





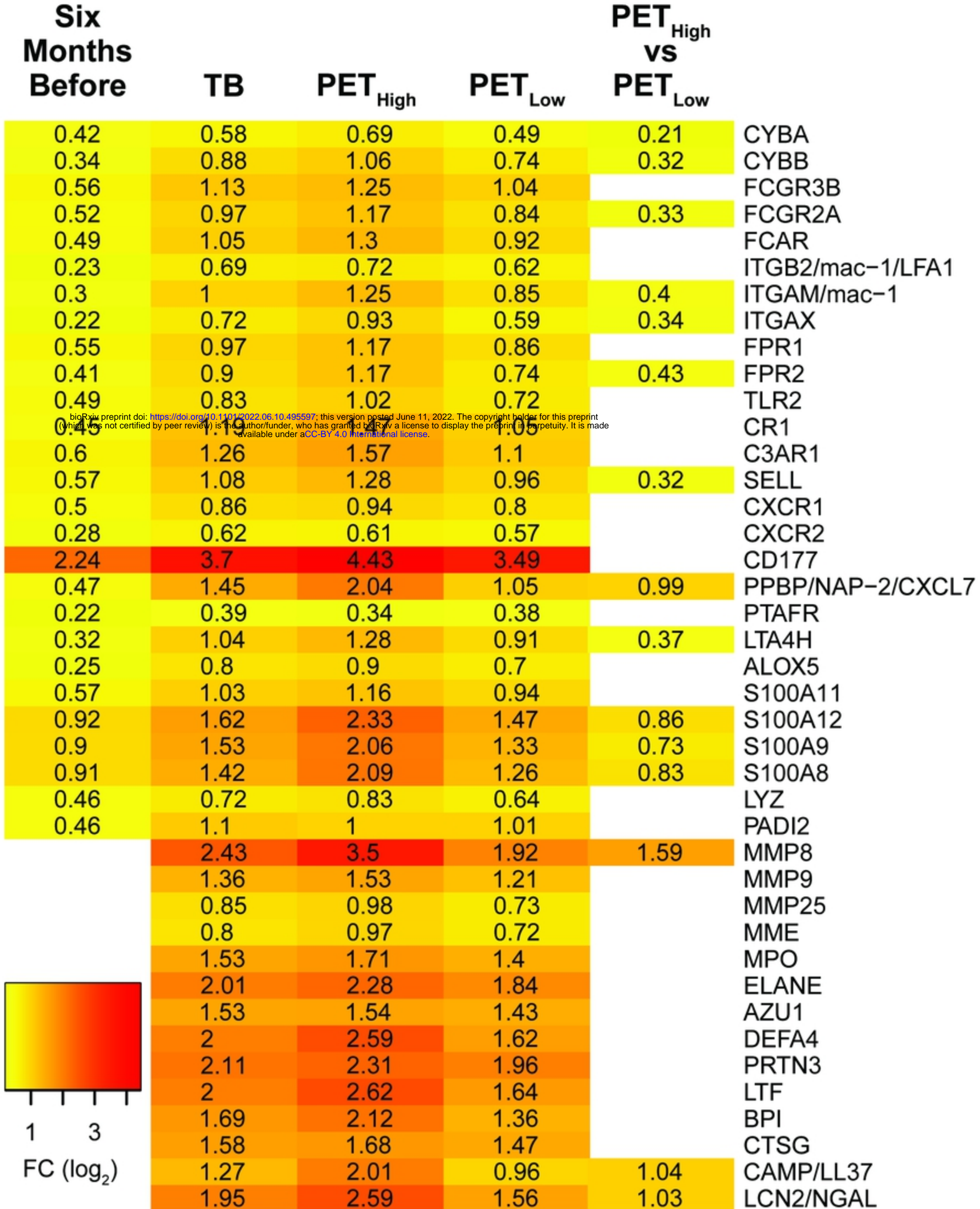


Fig 4

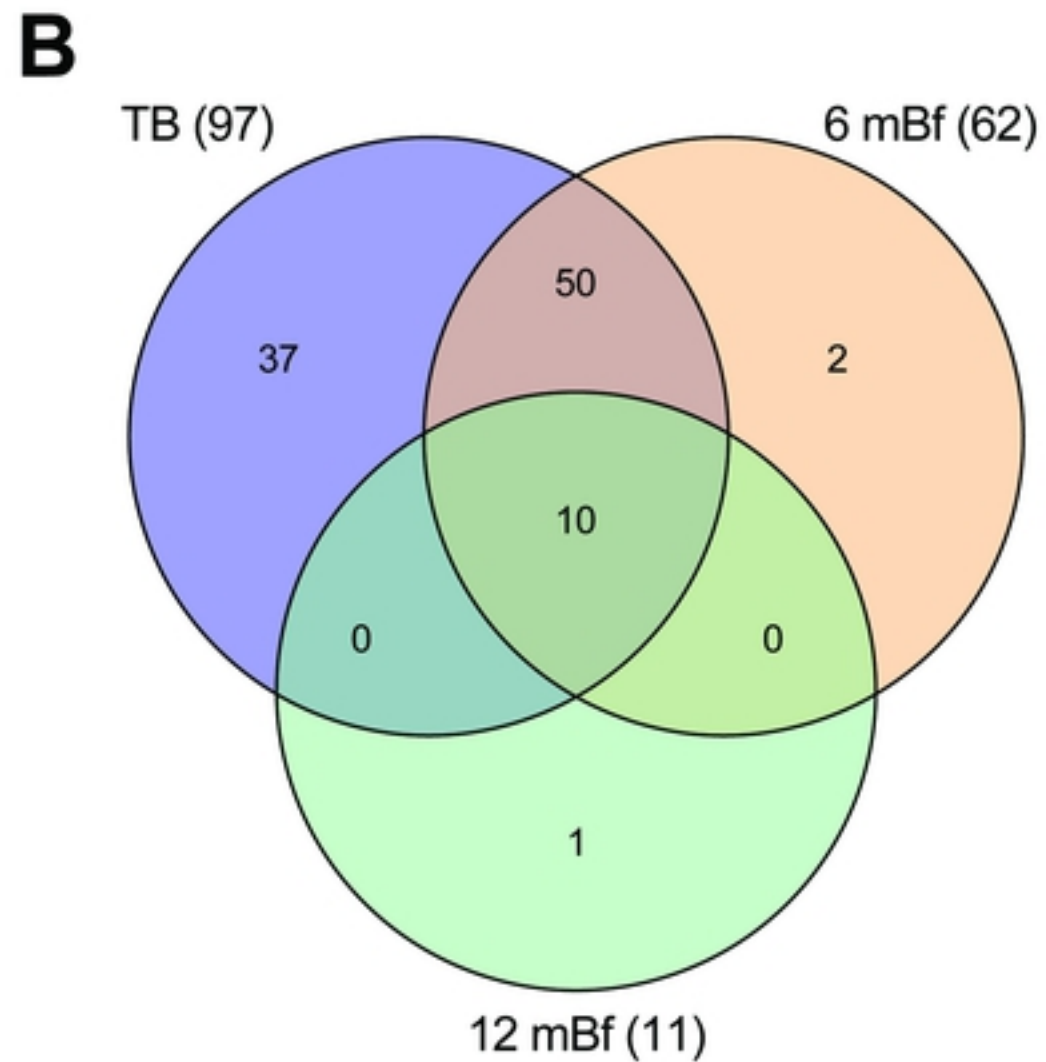
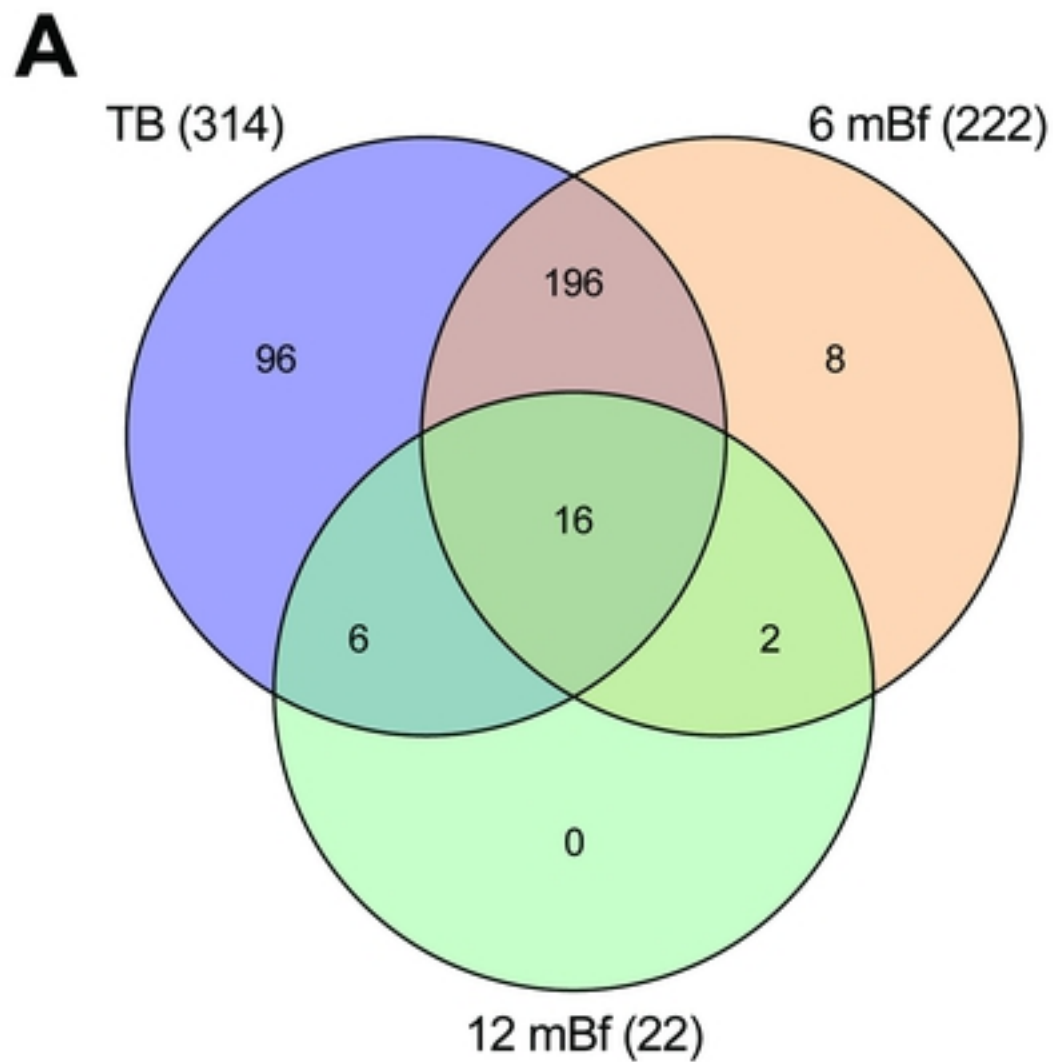


Fig 5



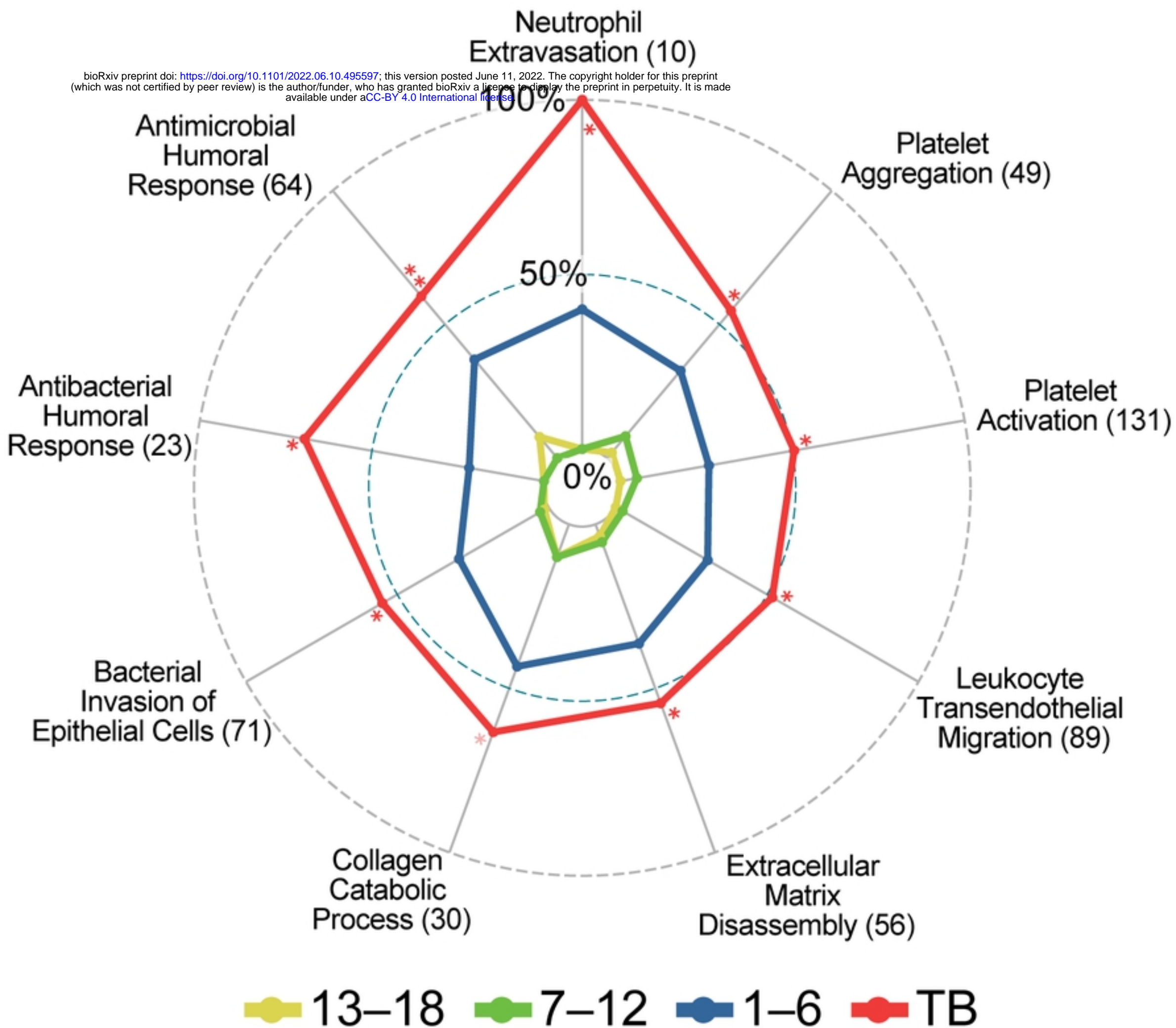


Fig 6