



## OPEN Specific human gene expression in response to infection is an effective marker for diagnosis of latent and active tuberculosis

Ritah Nakiboneka<sup>1,2,3,4</sup>, Natasha Walbaum<sup>1</sup>, Emmanuel Musisi<sup>1,5</sup>, Michael Nevels<sup>6</sup>, Tonney Nyirenda<sup>2</sup>, Marriott Nliwasa<sup>2,3</sup>, Chisomo L. Msefula<sup>2,3,4</sup>, Derek Sloan<sup>1</sup> & Wilber Sabiti<sup>1</sup>✉

RNA sequencing and microarray analysis revealed transcriptional markers expressed in whole blood can differentiate active pulmonary TB (ATB) from other respiratory diseases (ORDs), and latent TB infection (LTBI) from healthy controls (HC). Here we describe a streamlined reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) assay that could be applied at near point-of-care for diagnosing and distinguishing ATB from ORDs and LTBI from HC. A literature review was undertaken to identify the most plausible host-gene markers (HGM) of TB infection. Primers, and dual labelled hydrolysis probes were designed and analytically evaluated for accuracy in an in-vitro model of infection using a lung fibroblast cell-line. Best performing genes were multiplexed into panels of 2–4 targets and taken forward for clinical evaluation. Mycobacteria Growth Indicator Tube and QuantiFERON-TB Gold Plus were used as reference tests for ATB and LTBI respectively. A total of 16 HGM were selected and incorporated into five multiplex RT-qPCR panels. PCR assay efficiency of all evaluated targets was  $\geq 90\%$  with a median analytical sensitivity of 292 copies/ $\mu\text{l}$  [IQR: 215.0-358.3 copies/ $\mu\text{l}$ ], and a median limit of quantification of 61.7 copies/ $\mu\text{l}$  [IQR: 29.4-176.3 copies/ $\mu\text{l}$ ]. Clinically, ATB was characterised by higher gene expression than ORDs, while LTBI was associated with lower gene expression than HC, Kruskal-Wallis  $p < 0.0001$ . Crucially, *BATF2*, *CD64*, *GBP5*, *C1QB*, *GBP6*, *DUSP3*, and *GAS6* exhibited high differentiative ability for ATB from ORDs, LTBI or HC while *KLF2*, *PTPRC*, *NEMF*, *ASUN*, and *ZNF296* independently discriminated LTBI from HC. Our results show that different HGM maybe required for ATB and LTBI differentiation from ORDs or HC respectively and demonstrate the feasibility of host gene-based RT-qPCR to diagnose ATB and LTBI at near point-of-care.

**Keywords** Host gene expression, Diagnosis, Active tuberculosis, Latent tuberculosis, Reverse transcriptase-quantitative PCR

Early accurate detection of tuberculosis (TB) permits prompt treatment, breaks the chain of transmission, improves treatment outcome, and reduces related morbidity and mortality<sup>1</sup>. However, currently approved methods of TB diagnosis have posed several challenges to diagnosis and management of TB leading the World Health Organization (WHO) to call for non-sputum triage tests for TB diagnosis<sup>2</sup>. Transcriptional markers differentially expressed (DE) in active pulmonary TB (ATB) disease have been identified and are being evaluated as possible non-sputum alternatives for TB diagnosis<sup>3</sup>. Transcriptomic discovery studies used micro-bead arrays<sup>4–8</sup> and RNA sequencing<sup>9</sup> to identify DE genes in ATB compared to healthy controls (HC), latently infected TB (LTBI) or participants with other respiratory diseases (ORDs). These discovery studies reported hundreds<sup>8,10</sup> to thousands<sup>5,6</sup> of host gene markers (HGM) and required sophisticated programming and trained experts to

<sup>1</sup>Division of Infection and Global Health, School of Medicine, University of St Andrews, St Andrews KY16 9TF, UK.

<sup>2</sup>Department of Pathology, Kamuzu University of Health Sciences, Blantyre, Malawi. <sup>3</sup>Helse Nord Tuberculosis Initiative (HNTI), Pathology Department, Kamuzu University of Health Sciences, Blantyre, Malawi. <sup>4</sup>Africa Centre for Public Health and Herbal Medicines (ACEPHEM), Kamuzu University of Health Sciences, Blantyre, Malawi.

<sup>5</sup>Adroit Biomedical and Bio-entrepreneurship Research Services (ABBRS), Kampala, Uganda. <sup>6</sup>Biomedical Sciences Research Complex (BSRC), School of Biology, University of St Andrews, St Andrews, UK. ✉email: ws31@st-andrews.ac.uk

analyse the data. Complex analysis, however, makes clinical application of the findings difficult and conversion of the technology underpinning this approach to more straightforward near point-of-care (POC) tests with an easily interpretable read-out was required.

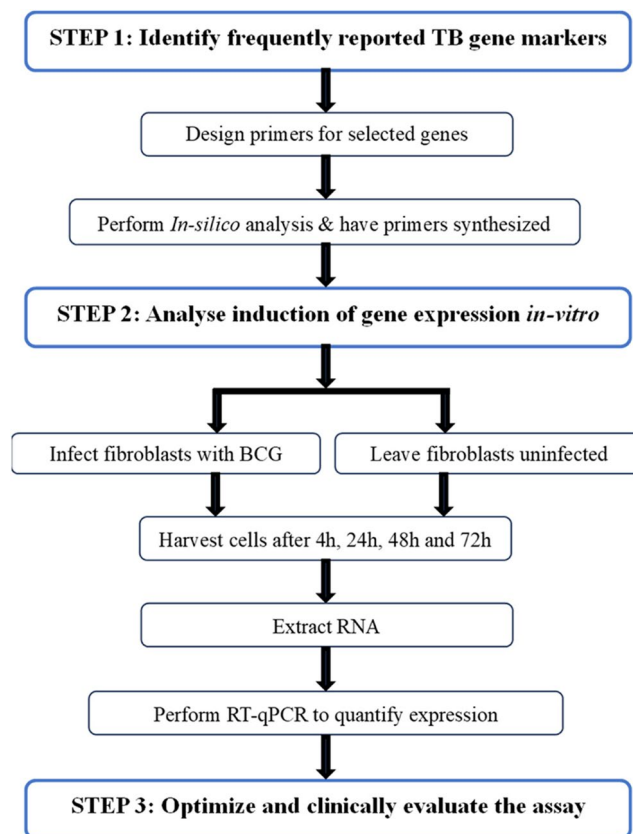
Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) assays allow for HGM quantification in whole blood after an initial RNA extraction step. This technology received great pertinency during the Coronavirus Infectious Disease 2019 (COVID-19) pandemic and was considered the gold standard for diagnosis of the infection<sup>11</sup>. Consequently, there is now increased worldwide availability of real-time qPCR machines<sup>12</sup> in clinical facilities, alongside skilled technical personnel to operate them. Several groups have used the RT-qPCR technology for ATB diagnosis through transcriptional marker quantification, although assay efficiency and quantitative read-outs are rarely reported<sup>12–15</sup>. Also, microarray quantifiable DE gene levels may not directly translate to measurable targets in RT-qPCR, as Darboe et al. reported some microarray discovered DE gene targets showing inconsistency in microfluidic RT-qPCR<sup>16</sup>. Here, we describe the steps taken to develop, optimise and evaluate an RT-qPCR assay to detect and quantify selected host transcriptional markers of TB infection (illustrated in Fig. 1) and report the assay quantitative read-out in well-characterised participant groups.

## Materials and methods

### Identification of host-gene markers for evaluation and design of qPCR primers, and probes

Published scientific literature was searched for plausible TB gene signatures using the Google Scholar search engine in March of 2021. The search term used was “tuberculosis transcriptional signatures”. Twenty signatures were selected as shown in Supplementary Table 1. Genes in the 20 signatures were listed, analysed, and scored for frequency of citation in the literature. Based on frequency of citation in the different selected signatures, as well as predicted network associations and biological function analysis performed on the STRING database<sup>17</sup> in the same period, genes were selected for inclusion into the customised RT-qPCR assay. Three reference genes namely *ACTB*, *GAPDH* and *B2M* reported highly expressed in fibroblasts by Vandesompele et al.<sup>18</sup> were also added to the panel and evaluated for invariance.

Primer design for the selected gene markers was performed using the Primer Quest program on the Integrated DNA Technologies (IDT) platform<sup>19</sup>. In-silico analysis of the selected primer pairs for hairpin formation, primer dimer formation and heterodimer formation were performed using the IDT Oligo Analyzer tool<sup>20</sup>. Primer sequences for *GBP5*, *DUSP3* and *KLF2* were identical to those published by Francisco et al.<sup>12</sup>, and suitable probes matching the primer sequences were designed for these genes using the same IDT platform.



**Figure 1.** Schematic flow diagram illustrating steps taken to design the RT-qPCR assay. BCG- bacillus Calmette–Guérin, RNA- ribonucleic acid and RT-qPCR- reverse transcriptase quantitative polymerase chain reaction assay.

### Assessing primer performance in gene expression quantification

Human embryonic lung fibroblast (MRC-5 cells, ATCC-CCL-171) were infected with *Bacillus Calmette–Guérin* (BCG) a surrogate of *Mycobacterium tuberculosis* (MTB) to induce gene expression in an in-vitro experiment. Cells used in the in-vitro experiment were at passages 24, 26, and 27 propagated in Dulbecco's Modified Eagle Medium supplemented with 10% foetal bovine serum (FBS). Each well of a six-well culture plate was seeded with  $1.2 \times 10^6$  MRC-5 cells. These were incubated at 37°C and in 5% CO<sub>2</sub> for 3 days to reach confluence before infection. Four wells of each culture plate were inoculated with 100 µl of BCG at optical density of 0.4 which is equivalent to 10<sup>5</sup> estimated colony forming units (eCFU) per ml and no inoculum was added to two control wells. Plates were then incubated under 5% CO<sub>2</sub> in a humidified incubator with sampling (cell harvesting) done at 4, 24, 48 and 72 h. Cell activity at each sampling point was arrested using a phenol-containing lysis buffer.

RNA extraction from MRC-5 cells was performed using the phenol-chloroform procedure. Briefly, the lysed mixture was transferred to 300 µl of chloroform, vortexed and centrifuged. The upper aqueous phase, containing the nucleic acids was harvested into new tubes. Ice-cold 100% absolute ethanol was added to precipitate the nucleic acids. The pellet was further washed with 70% ethanol, dried in a heat hot block at 50°C for 30 min and resuspended in RNase-free water. Genomic DNA was removed from the RNA extract by treating the extract with RNase-free DNase set from Qiagen for 40 min at 30°C. The reaction was stopped by placing the tubes on ice.

Host gene RT-qPCR was performed on a Rotor-Gene 5plex platform (Qiagen, UK) using HGM specific primers and QuantiTect SYBR Green (Qiagen) reagent. Briefly, the SYBR green cycle included an initial step of reverse transcription (RT) at 50°C for 20 min followed by a polymerase activation step for 15 min at 95°C. This was followed by 40 cycles each having a step of denaturation at 94°C for 15 s, annealing at 60°C for 30 s and final extension at 72°C for 20 s. For each gene reaction, a RT negative tube was included to confirm complete removal of genomic DNA. Data were analysed using the Rotor-Gene Q pure detection software Version 2.3.5 (Build 1).

### Amplification efficiency of the designed primers

To determine primer amplification efficiency of the target genes in relation to the reference ACTB, two RNA extracts from 72 h-BCG-infected MRC-5 cells were used. Sample RNA concentration was determined using a Nanodrop machine. The samples were diluted in a 10-fold seven times. Each target dilution was normalized with a similar dilution of the reference gene. The ΔCq values were then plotted on a line graph on the basis that an absolute slope value close to zero would indicate similar amplification efficiencies of the target and reference genes and mean that the primer sets were an efficient pair<sup>21</sup>.

### Quantification of bacterial load in harvested infected media and cells

Lysis of both cell-free bacilli in the media and cell-associated bacilli was performed using the phenol-chloroform procedure. To each sample, 100 µl of a known concentration extraction control was added. The infected media was vortexed, spun down, and the pellet re-suspended in FASTRNA Pro blue solution (lysis) buffer. The mixture was transferred to lysing matrix made of 0.5 µm zirconium beads and homogenized at 6,000 rpm for 40 s using, Percellys 24 (Bertin Germany) homogeniser. After a step of centrifugation at 12,000 g for 5 min, the supernatant was harvested into 300 µl of chloroform followed by RNA purification steps as described in the RNA extraction step above. Quantification of bacilli was performed on a Rotor-Gene 5plex platform (Qiagen, UK) using primers, dual-labelled hydrolysis probes (TaqMan) targeting MTB complex 16 S rRNA, and the extraction control and QuantiTect Multiplex RT-PCR kit (Qiagen) as described by Sabiiti et al.<sup>22</sup>.

### Primer probe concentration determination and assay multiplexing

In-silico analysis for the selected probe sequences was performed to confirm that the primers and probe combinations would not anneal to each other. Only four colour detection channels were possible, thus designed probes were based on these. These included fluorescein amidites (FAM) acquired on the green channel, ATTO700 acquired on the crimson channel, ROX acquired on the orange channel and Hexachloro-fluorescein (HEX) which was acquired on the yellow channel.

The multiplex assay included HGM specific primers and TaqMan dual labelled hydrolysis probes specific for the target genes and was also performed on a Rotor-Gene 5plex platform. Concentrations of the primers and probes used were determined by (1) changing concentrations for each primer and its probe in a single target reaction to determine optimum concentration, (2) combining the primers and probes in a single, duplex, and finally multiplex reaction.

The RT-qPCR cycle published by Sabiiti et al.<sup>23</sup> was adopted. It included, an initial reverse transcription step at 50°C for 30 min, DNA polymerase activation step at 95°C for 15 min followed by 40 amplification cycles each having a step of denaturation at 94°C for 45 s and annealing at 60°C for 60 s.

### Target standard reconstitution, standard curve creation and PCR efficiency determination

Procured from Eurofins genomics (Germany) in a lyophilised form, primer-probe target region standards were diluted to a starting concentration of 100 pmol/µl. Using the supplied molecular weight (MW) for each target, conversion to ng/µl was performed. Final concentration conversion to copies/µl for all the targets was performed using Avogadro's constant ( $6.02E + 23$ )<sup>24</sup>, base pair length of the nucleic acid, conversion factor to ng ( $1.00E + 09$ ) and molecular weight of a DNA base pair (660). The formula below was used for the calculation:

$$\text{Copies}/\mu\text{l} = \frac{\text{Amount (ng)} * \text{Avogadro's constant}}{\text{Length (bp)} * 1.00E + 09 * 660}$$

Target standards were diluted to a working stock of 1:10 and mixed within their allocated panels described above. Fourteen further 10-fold dilutions were made to create the standard curve. Amplification efficiency curves

of the standards were drawn to determine the assay dynamic range and PCR efficiency. Neat samples to the 1:10,000 dilution for all targets did not yield good sigmoid amplification since the targets were too concentrated. Good sigmoid shape amplification curves were observed from the 1:100,000 dilution. To determine the limit of detection (LOD) and limit of quantification (LOQ), amplification efficiency assays of the target standard were performed in a 10-fold dilution 7 times, starting from a concentration of  $2.8E+08$  copies/ $\mu$ l ( $1 \times 10^{-5}$  dilution) to a  $2.8E-01$  copies/ $\mu$ l ( $1 \times 10^{-14}$  dilution).

### Clinical evaluation of the assay in human participants

Between 4th Jan 2022 and 30th September 2022, participants with Xpert MTB/RIF- and Mycobacteria Growth Indicator Tube (MGIT) culture- confirmed pulmonary ATB were prospectively enrolled into the study from three healthcare facilities (HCFs) in Blantyre Malawi. Participants with TB-like symptoms but testing negative for both Xpert MTB/RIF and MGIT culture were also enrolled into the study from the same HCFs, and these were denoted ORDs participants. ORDs were tested for LTBI using Gold plus QuantiFERON (QFT-Plus) test. Household contacts of confirmed ATB cases were enrolled and those who tested positive for TB Gold plus QuantiFERON (QFT-Plus) test were classified as LTBI. HC were recruited from the same Blantyre community based on HIV negative status, no known TB exposure, and a negative QFT-Plus test. Individuals were excluded from the study if they did not consent to join, and if they were less than 18 years old.

### Sputum liquid culture and xpert MTB/RIF tests

MTB detection in clinical samples was performed using MGIT culture and Xpert MTB/RIF Ultra test. To eliminate non-MTB bacteria before MGIT culture, sputum samples were decontaminated with 3% Sodium hydroxide + N-acetyl-L-cysteine (NaOH + Nalc) in a 1:1 sample-reagent ratio for 15 min. The action of NaOH + Nalc was neutralized by addition of sterile phosphate buffered saline. The mixture was centrifuged at 3000 g for 20 min to concentrate the bacteria and get rid of excess supernatant. Each MGIT tube (Becton Dickinson (BD) supplemented with BACTEC MGIT 960 OADC Growth Supplement (BD) and MGIT lyophilised mixture of Polymyxin B, Amphotericin B, Nalidixic Acid, Trimethoprim and Azlocilin (PANTA) Antibiotic Mixture (BD) was inoculated with 0.5mls of the concentrated decontaminated sample. Tubes were then incubated in the MGIT culture instrument at 37°C until they flagged positive in a 42-day period. Samples that did not flag positive within 42 days, were considered TB negative. Presence of MTB species in positive cultures was confirmed by acid fast bacilli (AFB) Ziehl-Neelsen (ZN) staining and detection of antigen MPT64, a protein that is only produced by *Mycobacterium tuberculosis* complex. For the Xpert MTB/RIF Ultra test, samples were mixed with sample reagent in a 1:2 volume ratio to liquefy the sputum and left to stand for 15 min at room temperature. The mixture was then added to the Xpert MTB/RIF Ultra cartridge and run on the GeneXpert instrument.

### RNA extraction, quantification and QFT-Plus assay

Procedures for RNA extraction, quantification, and QFT-Plus have previously been described<sup>25</sup>. Briefly, peripheral whole blood was collected from all study participants into Paxgene tubes and QFT-Plus tubes to perform these procedures. RNA extraction was performed using the PAXgene blood RNA kit (PreAnalytiX cat. no 762174) as per the manufacturer's recommendation. Extracted RNA yield was measured using Qubit RNA High Sensitivity reagent (Invitrogen Cat. No Q32855) on the Qubit machine version 3.0 (Life Technologies)<sup>26</sup> whilst the QFT-Plus test was performed following the Qiagen 2018 kit insert<sup>27</sup>.

### Statistical analysis

Data analysis was performed using Microsoft Excel and R statistical programming (version 4.2.1, Boston)<sup>28</sup>. For the cell infection assay, data normalisation was performed by subtracting the average Cycle quantification (Cq) values of the reference gene ACTB from the average Cq value of the target gene<sup>29</sup>. Whilst for clinical evaluation, Cq results were converted to concentration using target specific standard curves and recorded as copies/ $\mu$ l. Differences between groups were performed using Mann-Whitney U test, Kruskal-Wallis test and Dunn's Test with Bonferroni corrected p-values for multiple comparisons. A p-value of <0.05 was considered significant with \*=<0.05, \*\*=<0.01, \*\*\*=<0.001, and \*\*\*\*=<0.0001. Assay LOD and LOQ were determined as described by Bustin et al.<sup>30</sup> and Kralik et al.<sup>31</sup> respectively. Quantitative read-outs of the RT-qPCR assay data are reported as median with interquartile ranges [IQR]. Data visualisation was performed using GraphPad Prism (version 10.0.0 Boston, Massachusetts USA) and RStudio 2023.06.1 Build 524 (Posit Software, PBC)<sup>28</sup>.

### Ethical approval

This work was approved by the University of St Andrews Teaching and Research Ethics Committee (UTREC) under approval code MD15741 and the College of Medicine Research Ethics committee (COMREC) Malawi under Protocol number P.06/21/3342. All participants in the clinical evaluation step agreed to join the study and provided written informed consent. All study procedures and experiments were conducted in accordance with the relevant guidelines and regulations.

### Results

#### Selection of a plausible panel of genes to convert into a diagnostic test for TB

The 20 selected signatures (details in Supplementary Table 1) comprised of 549 genes (excluding repeats). Of these, 490/549 appeared in only one signature, 41/549 appeared in 2 signatures and 11/549 appeared in 3 signatures. *GBP5* was the most frequently reported gene appearing in 8/20 signatures with a 40% occurrence followed by *GBP6* at a 35% occurrence (7/20 signatures). *CIQB*, *FCGR1B*, *SEPT4*, *FCGR1A*, and *GAS6* followed with a 25% or 20% frequency of occurrence in published literature. Therefore, all these HGM were included in the panel. Table 1 provides details of reported gene markers with the highest frequency of occurrence.

No.	Gene	No. of times mentioned	% Occurrence
1	<i>GBP5</i>	8	40
2	<i>GBP6</i>	7	35
3	<i>CIQB</i>	5	25
4	<i>FCGR1B / CD64B</i>	5	25
5	<i>SEPT4</i>	5	25
6	<i>FCGR1A / CD64A</i>	4	20
7	<i>GAS6</i>	4	20
8	<i>ANKRD22</i>	3	15
9	<i>BATF2</i>	3	15
10	<i>CIQC</i>	3	15
11	<i>DHRS9</i>	3	15
12	<i>DUSP3</i>	3	15
13	<i>FER1L3 (MYOF)</i>	3	15
14	<i>GBP1</i>	3	15
15	<i>IFITM3</i>	3	15
16	<i>SERPING1</i>	3	15
17	<i>SMARCD3</i>	3	15
18	<i>STAT1</i>	3	15

**Table 1.** Gene occurrence in 20 selected TB reported signatures.

Next, the selected genes and those that occurred in at least 3 signatures (11 genes) were submitted to the STRING database<sup>17</sup> to ascertain protein-protein interaction clustering and biological function. Selection among the eleven genes reported in 3 different signatures was based on a protein having unique biological function different from the above already selected frequently occurring (No. 1–7). Proteins clustering with the already selected most frequently occurring markers, and having similar biological function as those prior selected markers were excluded. Figure 2a shows the modelling performed on the STRING website and the clustering. Consequently, *BATF2* and *DUSP3* were additionally selected and added to the panel since they had unique biological function compared to *GBP5* with only two predicted network interactions. Furthermore, since the Sweeney3 signature<sup>32</sup> now had 2 of its genes selected, *KLF2* was added to this panel to make the 3 gene Sweeney3 panel complete. Since *GBP1*, *IFITM3* and *STAT1* were in the same cluster of biological functions as *GBP5*, they were excluded. Similarly, *CIQC* and *SERPING1* were clustering with the complement marker *CIQB* and were therefore excluded. The biological functions of *ANKRD22*, *DHRS9* and *SMARCD3* were not given on the STRING website and were thus also excluded. Lastly, *MYOF* was excluded because it had similar function as *DUSP3*. Primer design and in-silico analysis proceeded for the selected panel. *CD64* primers were designed to span both the *FCGR1B* and *FCGR1A* regions. Primer pairs for reference genes *ACTB*, *GAPDH* and *B2M* were also designed. Further research was performed to identify plausible markers highly discriminatory of LTBI from HC and a panel of *ASUN*, *NEMF*, *DHX29* and *PTPRC* published by Lee et al.<sup>15</sup> was added. HGM *ZNF296* and *ARG1* reported downregulated in ATB disease compared to LTBI and ORDs, respectively<sup>33</sup>, were also added to the panel. Figure 2b shows the additional marker panel. A protein network showing the selected 15 markers for evaluation is shown in Fig. 2c. Annotation of the selected gene biological functions as indicated in literature are shown in Supplementary Table 2.

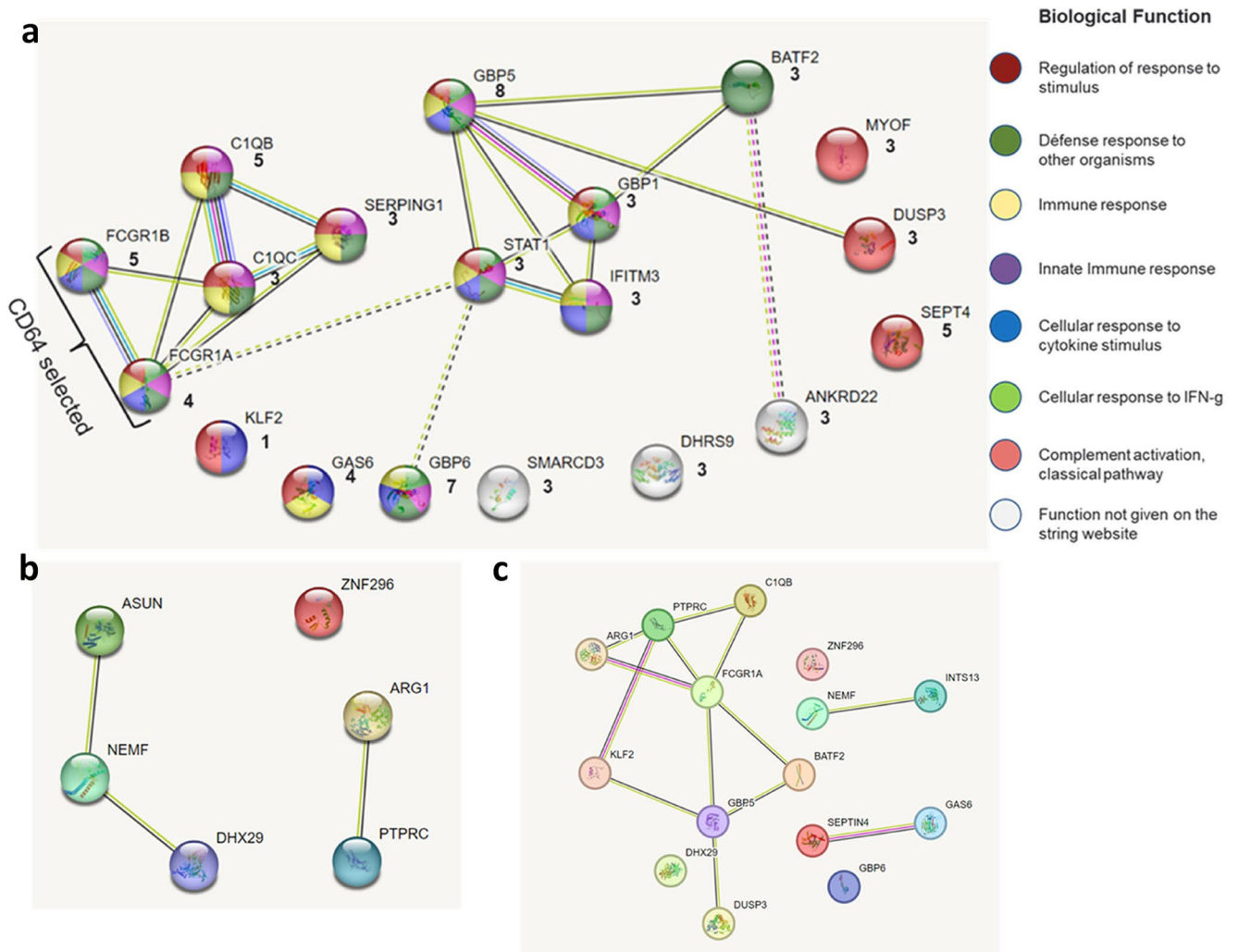
### Designed primers and probes were optimal for accurate amplification

The designed primer length was between 18 and 25mer. Guanine-cytosine (GC) content was 40 to 66.7%. All hairpins formed by the primers had a melting temperature ( $T_m$ ) below the  $T_m$  of the primers, all the primers had a Gibbs free energy ( $\Delta G$ ) for formation of heterodimers of greater than  $-10$  and none formed strong heterodimer bonds with each other. Supplementary Table 3 summarises key characteristics for each primer. Similarly, designed probe properties are included in Supplementary Table 4. These did not form heterodimers with the primers, were 0 to 6 base pairs longer than the primers and had slightly higher  $T_m$  than the primers.

### Successful infection of MRC-5 cells and characterisation of primer performance in HGM quantification

The in-vitro experiment performed to induce gene expression in lung fibroblast cells is illustrated in Fig. 3a. MRC-5 cells were successfully infected with BCG bacilli as shown by the progressive shift in the molecular bacterial load (estimated colony forming units per ml (eCFU/ml) from extracellular media to cells from 4 to 72 h of incubation (Fig. 3b). The reference genes *ACTB*, *B2M* and *GAPDH* were initially assessed for invariance in their expression within lung fibroblasts irrespective of experimental conditions and reproducibility between repeats. *ACTB* showed the least variation in its expression for all hours of incubation thus chosen as the appropriate control gene to normalise target expression (Supplementary Fig. 1). *B2M* expression was highly variable. For validity of data normalisation and calculation of the  $\Delta Cq$  method, primer amplification efficiency of the target genes must be equal to the reference gene<sup>21</sup> and this was achieved as shown by the slope values in Fig. 3c. Gel



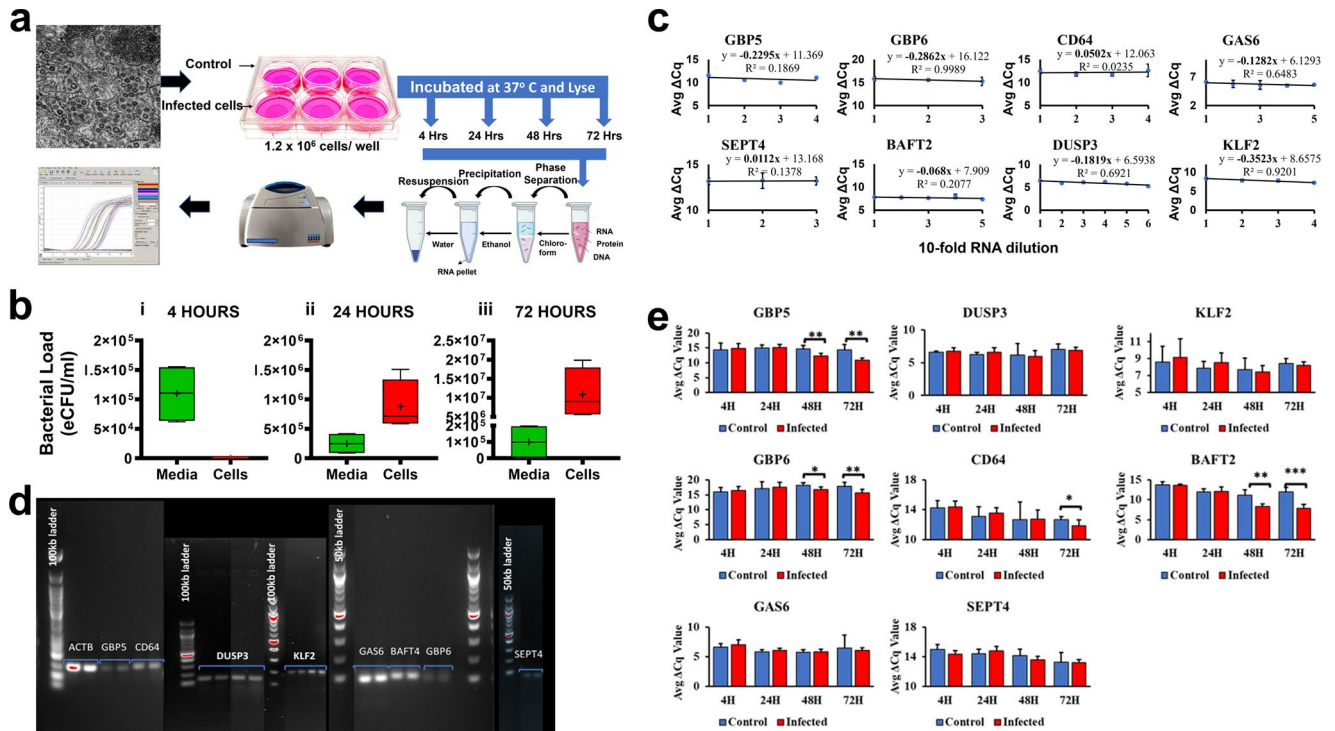


**Figure 2.** Frequently reported TB gene markers and functional relationships between them. (A) Selected gene biomarker protein interaction network created on the STRING website. Genes were grouped according to their associated pathways and biological functions. The colour within the circles represents the biological function and the key summarises the functions. The evidence mode was used to summarize the network predicted associations. Lines indicate the available evidence used in predicting the functional associations, the colour of the lines denotes the type of interaction evidence. The number of interaction lines illustrate more evidence of modules available for that interaction. Seven types of interaction evidence were used. Interaction analysis on the STRING website was performed in March of 2021 and may change with new evidence submitted to the database. Numbers underneath the gene symbol indicate the number of times gene was mentioned in plausible signatures. (B) Network interaction of the additional markers for LTBI. (C) Final interaction network of all the selected gene markers.

electrophoresis of all the RT-qPCR products confirmed there were no primer dimers formed by the designed primers (Fig. 3d). Subsequently, the model revealed that expression of HGM *GBP5* ( $p=0.0080$  and  $p=0.0012$ ), *GBP6* ( $p=0.0232$  and  $p=0.0042$ ) and *BATF2* ( $p=0.0024$  and  $p=0.0009$ ) was significantly higher in infected-than uninfected- cells by 48 h and 72 h of incubation, respectively (Fig. 3e). *CD64* achieved differential expression in infected cells by 72 h of incubation ( $p=0.022$ ), suggesting a delayed response to infection compared to other genes. Expression of the other gene markers, *DUSP3*, *KLF2*, *GAS6* and *SEPT4*, did not differ between infected and uninfected cells at all incubation points. These results indicated that the designed primers could optimally detect targets and were able to show variation in amounts of targets between infected and uninfected cells. The additional gene marker panel described in Fig. 2b and HGM *C1QB* primers were optimised with only in-silico analysis and cell-free medium and not with RNA from in-vitro infection model because they were adopted at the clinical validation stage.

#### Assay multiplexing and analysis of precision

For all the gene targets, there was no difference in average Cq value detected in the single-target detection reaction (single plex -SP) assay compared to the three- or four- (multiplex - MP) target detection reaction assay except for



**Figure 3.** Assessing primer performance in measuring gene expression in a tissue culture model of infection. **a**, Illustration of the lung fibroblast infection assay, extraction procedure and PCR amplification cycle. **b**, BCG association with lung fibroblasts increases with time of incubation. Shown are box plots of bacterial load in eCFU/ml in infected media (green) and in infected lung fibroblast cells (red) after (i) 4 h, (ii) 24 h, and (iii) 72 h of incubation. **c**, Amplification efficiency of the target and reference primers compared. Line graph showing average  $\Delta Cq$  values between target and reference primers at serial dilutions of mRNA. A slope of close to 0 indicated similar amplification efficiency of the target genes and the reference gene ACTB. **d**, Agarose gel electrophoresis for amplified targets. **e**, Upregulated expression of genes in infected human lung fibroblast cells. Bar graphs showing average (Avg.) normalised Cycle quantification (Cq) values as a ratio to Avg. Cq values for the reference gene, ACTB. The lower the Avg.  $\Delta Cq$  value the higher the gene expression. Mann-Whitney test was used to evaluate for statistical differences. Data are reported as mean values from three repeats of each experiment consisting of 2 uninfected and 4 infected wells, \* indicates p-value < 0.05, \*\* indicates a p-value < 0.01 and \*\*\* indicates a p-value < 0.001. Error bars are indicative of the standard deviation of the  $\Delta Cq$  values.

*DUSP3* and *BATF2* which had lower Cq values in MP and DP respectively (Fig. 4a). The final multiplex reactions consisted of 5 panels namely: Panel 1: *GBP5*, *KLF2*, *DUSP3* and *ACTB*; Panel 2: *PTPRC*, *BATF2*, *SEPT4* and *GBP6*; Panel 3: *GAS6*, *ASUN*, *NEMF*, and *CIQB*; Panel 4: *ZNF296* and *ARG1*; Panel 5: *CD64* and *DHX29*. Median Cq coefficient of variation (CV) was 1.5% [IQR: 1.1–1.7%] for all gene targets in the multiplex assay indicating adequate precision of the assay even when measured after another freeze thaw cycle (Fig. 4b). Median standard deviation (SD) of these Cq values was 0.27 [range: 0.11–0.48] for all gene target except *SEPT4* [SD = 3.3].

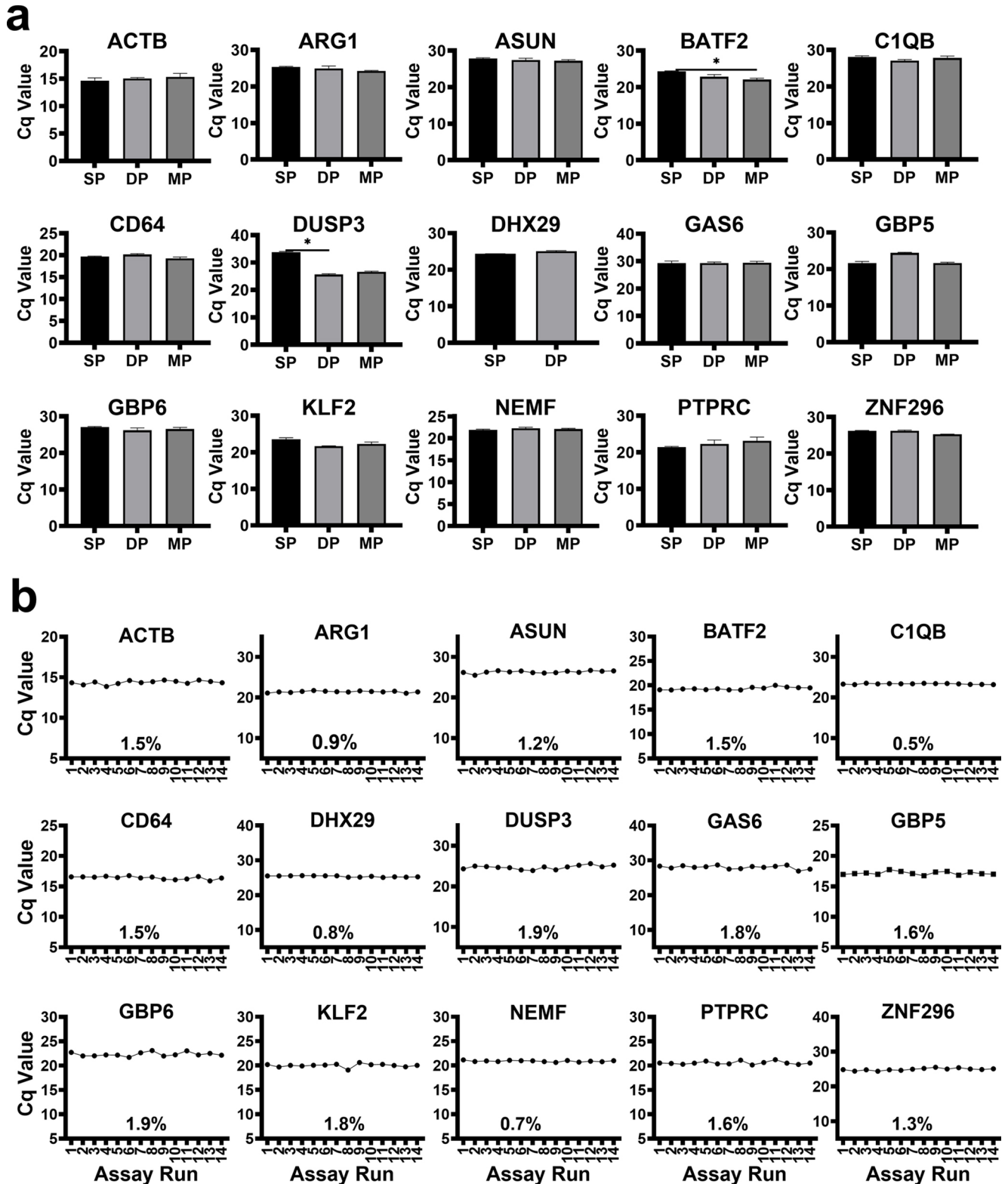
#### Assay efficiency, limit of detection (LOD) and limit of quantification (LOQ)

The assay had acceptable dynamic range for target nucleic sequence amplification with linear regression equation Pearson correlation coefficient (R) and coefficient of determination ( $R^2$ ) values above 0.98. This indicated that the designed primers and probes reliably detected their specific targets (Fig. 5). The average slope of the standard curve was between the acceptable range of -3.1 and -3.6. Additionally, all primer sets had good efficiency (90 to 110%) with true statistical doubling for amplification of their specific targets in a multiplex assay (Fig. 5).

LOD was defined as the lowest concentration at which the analyte could be accurately measured in both replicates for a gene marker target with 95% confidence (probability)<sup>30</sup>. All gene targets were accurately detected at a median LOD concentration of 292 copies/ $\mu$ l [IQR: 215–358.3 copies/ $\mu$ l] (Table 2). LOQ was defined as the lowest average concentration of RNA for a gene marker that was accurately measured between 2 replicates with a CV of less than 25%<sup>31</sup>. The median LOQ for all the target standards was 61.7 copies/ $\mu$ l [IQR: 29.4–176.3 copies/ $\mu$ l]. Table 2 shows the LOQs for optimal gene target detection and the corresponding Cq and CV values.

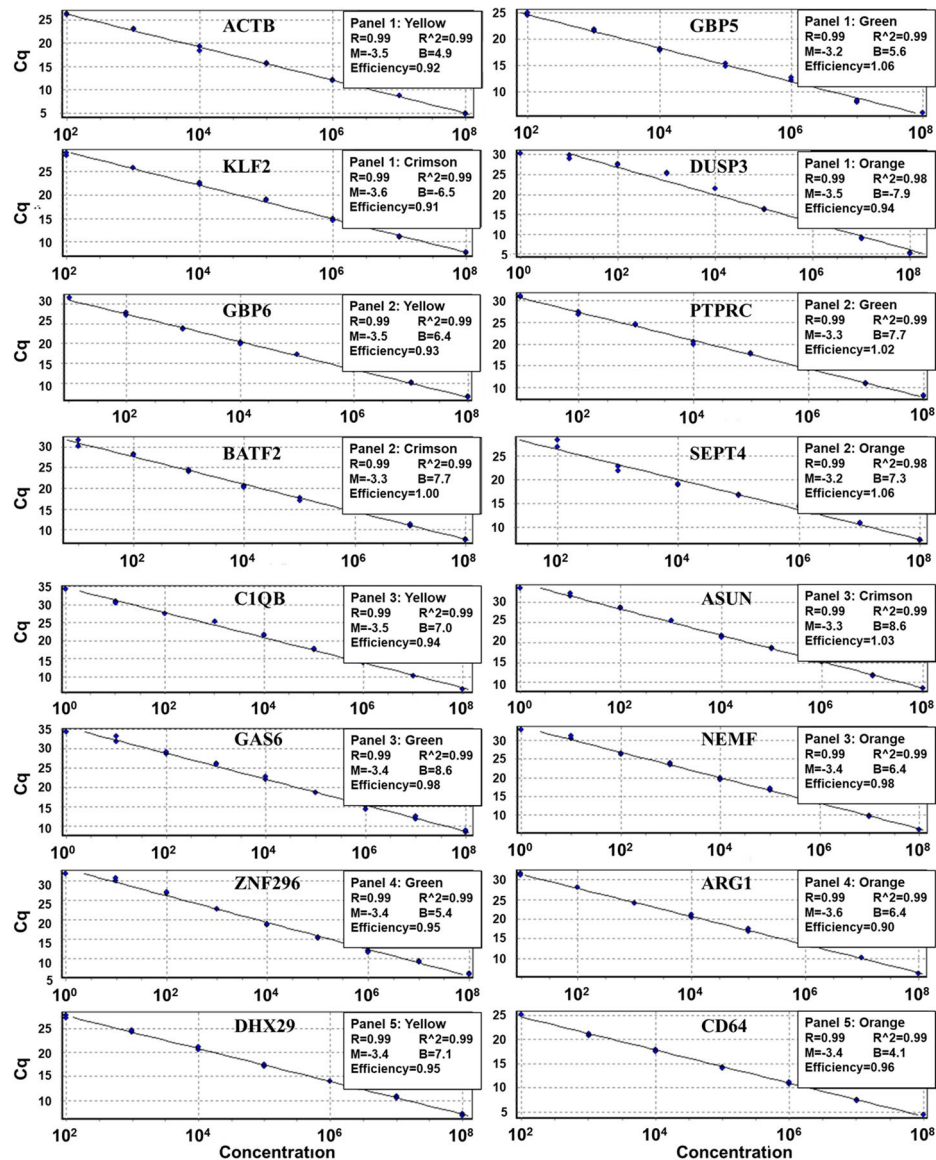
#### Assay quantitative read-out in well characterised participants

A total of 204 participants (61 ATB, 82 ORDs, 24 LTBI and 37 HC) were included in assay clinical evaluation. Supplementary Fig. 2 summaries the procedures involved in the application of the assay to clinical samples. Participant demographic, clinical and laboratory data is shown in Table 3. ATB participants were MGIT culture



**Figure 4.** Assay Optimization. Shown are Cq values measured from an ATB participant sample. **a**, Assay multiplexing: SP - Single-target reaction, DP - Double-target reaction and MP - Multiplex, mean expression levels. Comparison was performed using the Kruskal-Wallis test and Dunn’s Test with Bonferroni corrected p-values for multiple comparisons. A p-value of 0.05 or less was considered significant and \* is equivalent to  $p < 0.05$ . **b**, Assay precision testing line graphs showing assay read-out Cq values. Assay technical repeat runs 1–6 were performed on 4th of November 2022 and assay technical repeat runs 7 to 14 were performed on 6th February 2023. Percentage values within the graph represent the Cq CVs between runs for target genes.





**Figure 5.** Assay efficiency assessment. PCR efficiency represented by the best of fit regression line of the standard curve for the HGM targets. Cq – quantification cycles, the x-axis shows the concentration of standard target sequences in copies/ $\mu$ l. Blue dots in the curves represents standard target sequences quantified at different concentrations. Assay threshold value for all targets was set at 0.01 except *GBP6* whose threshold was set at 0.005 because it had low expression in clinical samples. Cycling A - the amplification cycle in the Yellow (HEX), Green (FIM), Crimson (ATTO700) and Orange (ROX) channel. Panel 1/2/3/4/5 are the multiplex reaction panels; Panel 1 to 3 each consisted of 4 HGM and panel 4 and 5 consisted of 2 HGM.  $R^2$  value - coefficient of determination and the R value - Pearson correlation coefficient is the square root of  $R^2$ . M and B - represent the slope (M) and the intercept (B) of the standard curve. Efficiency – reaction efficiency by which the independent variable correlates with the dependent variable.

and Xpert MTB/RIF positive while ORDs participants were MGIT culture and Xpert MTB/RIF negative. LTBI participants were QFT-Plus positive while HC participants were QFT-Plus negative. 38% (31/82) of ORDs participants were also QFT-Plus positive. Extracted median RNA yield was 98.2 ng/ $\mu$ l [IQR: 73.4–130] for ATB, 66.8 ng/ $\mu$ l [IQR: 51.9–92.5] for ORDs, 59 ng/ $\mu$ l [IQR: 44.1–84.6] for LTBI and 91.2 ng/ $\mu$ l [IQR: 72.2–118] for HC participants. RNA yield from ATB participants was significantly more than extracted yield from LTBI and ORDs participants ( $p < 0.001$ ) and RNA yield from LTBI participant was significantly lower than yield for HC participants ( $p = 0.03$ ). No difference in extracted RNA yield was observed between ATB and HC participants.

The host-gene RT-qPCR assay efficiently detected the target markers with minimal dispersion among same group participants. Median expression for ATB was above the assay LOD and LOQ for all gene markers except *SEPT4*. Expression of 7 evaluated gene markers namely *BATF2*, *CD64*, *GBP5*, *CIQB*, *GBP6*, *DUSP3*, and *GAS6* showed distinct upregulation of the genes in ATB compared to either HC, LTBI or ORDs (Fig. 6). Notably, expression levels for LTBI individuals were lower than expression levels of either HC, ORDs, or ATB for all

No.	Gene Target	LOD	LOQ		
		Copies/ $\mu$ l	Copies/ $\mu$ l	Cq	CV
1	ACTB	355	28.7	31.0	13.6%
2	ASUN	221	221	31.2	19.8%
3	ARG1	739	29.4	31.4	24.1%
4	BATF2	197	82.1	31.7	20.6%
5	CIQB	382	23.6	34.3	4.8%
6	CD64	178	64.6	28.8	4.4%
7	DHX29	354	29.4	33.6	18.8%
8	DUSP3	159	34.1	32.9	20.6%
9	GAS6	298	40.5	33.1	1.1%
10	GBP5	441	162	29.6	2.6%
11	GBP6	275	59.2	28.7	16.9%
12	KLF2	286	219	31.4	14.2%
13	NEMF	368	244	29.3	7.0%
14	PTPRC	328	23.2	31.0	11.3%
15	SEPT	282	282	31.0	1.8%
16	ZNF296	192	64.2	30.7	1.5%

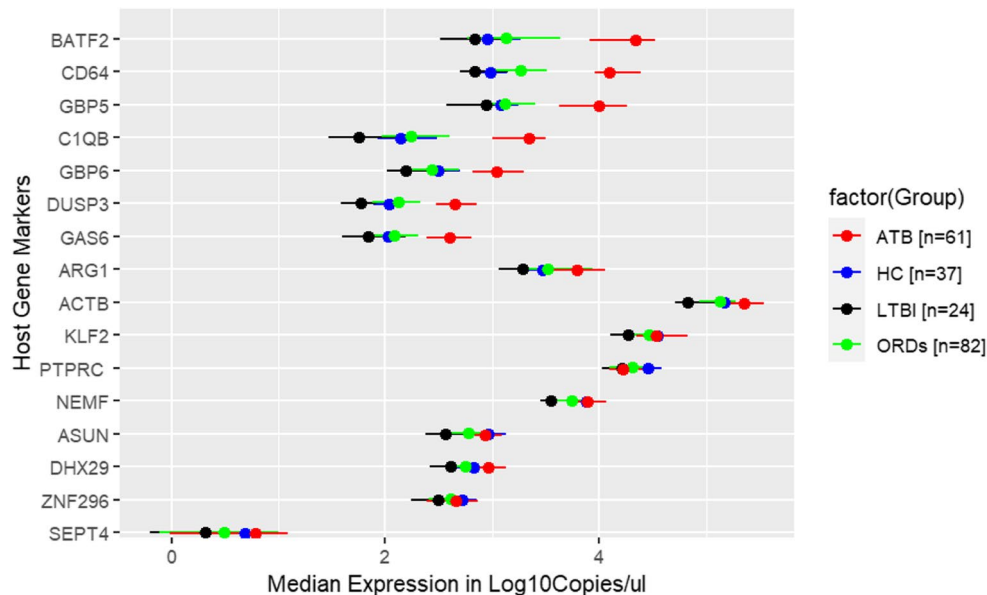
**Table 2.** Assay limit of detection and limit of quantification.

Participant characteristics	ATB [ <i>n</i> = 61]	ORDs [ <i>n</i> = 82]	LTBI [ <i>n</i> = 24]	HCS [ <i>n</i> = 37]
Sex (male); <i>n</i> [%]	48 [79]	61 [74]	7 [34.7]	19 [52.9]
Age (years); median [IQR]	32 [24–37]	42 [31–52]	36 [30–42]	26 [25–31]
Weight (kgs); median [IQR]	50 [46.5–55]	55.6 [48–61]	62.5 [53–71]	62 [58–69]
Having Cough; <i>n</i> [%]	61 [100]	80 [98]	1 [4]	2 [5]
Night sweats; <i>n</i> [%]	42 [69]	61 [74]	2 [8]	1 [3]
Loss of Weight; <i>n</i> [%]	59 [97]	60 [73]	3 [12.5]	0
Smoking; <i>n</i> [%]	19 [31]	23 [28]	1 [4]	5 [13.5]
Firewood cooking; <i>n</i> [%]	5 [8]	18 [22]	2 [8]	2 [5.4]
Alcohol Use; <i>n</i> [%]	23 [38]	33 [40]	4 [16.7]	10 [27]
BCG Vaccine; <i>n</i> [%]	59 [97]	80 [98]	24 [100]	35 [94.6]
Treated for TB before; <i>n</i> [%]	11 [18]	16 [20]	1 [4]	0
Can read; <i>n</i> [%]	54 [89]	76 [93]	23 [95.8]	37 [100]
Employed; <i>n</i> [%]	36 [59]	54 [66]	13 [54.2]	19 [51.4]
Married; <i>n</i> [%]	22 [36]	51 [62]	18 [75]	25 [67.6]
HIV Status; <i>n</i> [%] Positive	22 [36]	37 [45]	5 [20.8]	0
Negative	39 [64]	45 [55]	19 [79.2]	37 [100]
CD4 Count (cells/mm3) median [IQR]	220 [140.5–367.5]	323 [218–423.5]	501 [459–563]	NA
IGRA test; <i>n</i> [%] Positive	49 [80]	31 [38]	24 [100]	0
Negative	3 [5]	44 [54]	0	37 [100]
Indeterminate	9 [15]	7 [9]	0	0
RNA Yield (ng/ $\mu$ l); median [IQR]	98.2 [73.4–130]	66.8 [51.9–92.5]	59 [44–84.6]	91.2 [72.4–117]
Xpert MTB/RIF; <i>n</i> [%] Positive	61 [100]	0	ND	ND
Negative	0	82 [100]		
MGIT (positive); <i>n</i> [%] Positive	61 [100]	0	ND	ND
Negative	0	82 [100]		

**Table 3.** Participant demographic and clinical characteristics. ND = Not done, NA = Not applicable.

markers, indicating gene marker suppression in LTBI compared to other groups. Table 4 shows the median with IQR of the assay Cq values and the concentration in copies/ $\mu$ l of the measured targets in whole blood.

When the absolute measurements of the gene levels were compared, HGM expression of *BATF2*, *CD64*, *GBP5*, *CIQB*, *GBP6*, *DUSP3*, *GAS6*, *ARG1*, and *DHX29* was significantly upregulated in ATB individuals compared to expression in ORDs, LTBI or HC (Fig. 7). Whilst expression levels of HGM *KLF2*, *PTPRC*, *NEMF*, *ASUN*, and *ZNF296* were significantly downregulated in LTBI compared to HC. Notably, expression levels of *KLF2*, *NEMF*, *ASUN*, and *ZNF296* were akin in ATB and HC participants.



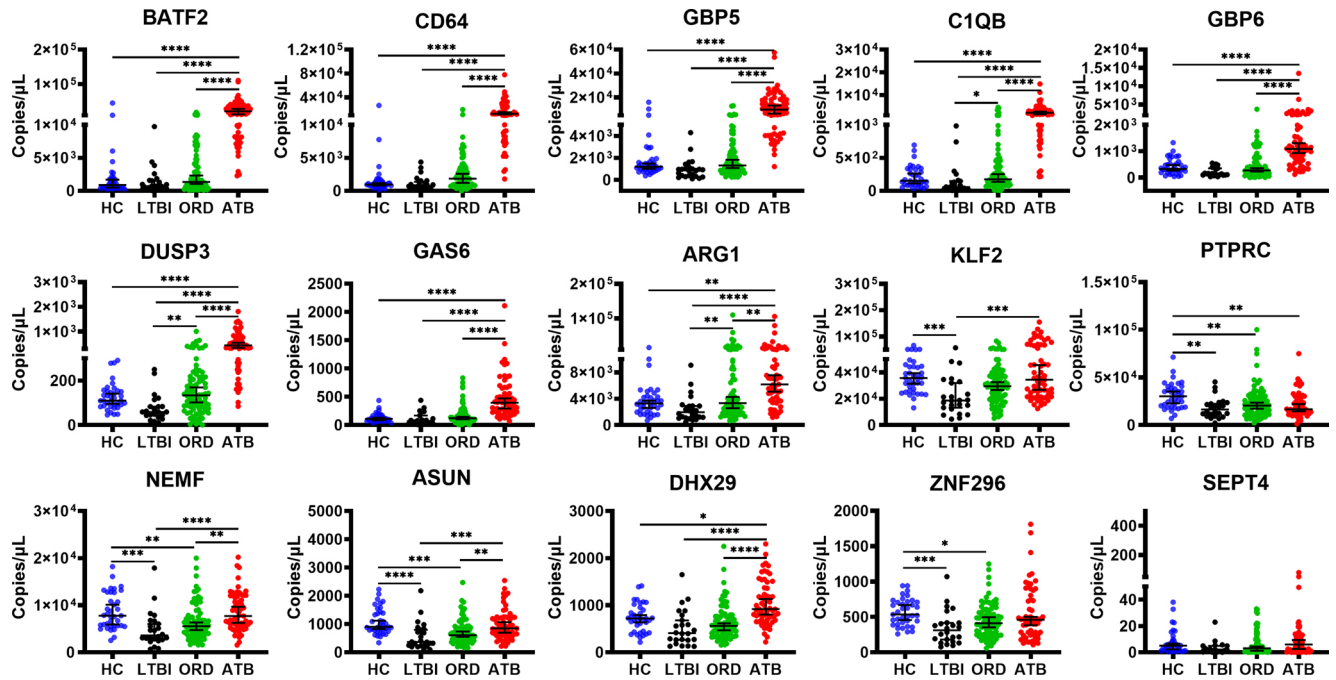
**Figure 6.** Clinical assay evaluation. Stratified median log<sub>10</sub> copies/μl expression with IQRs of the evaluated HGM.

Gene Marker	ATB		ORDS		LTBI		HC	
	Cq	Copies/μl	Cq	Copies/μl	Cq	Copies/μl	Cq	Copies/μl
ACTB	15.9 [15.3–16.8]	226,500 [129000.0–349500.0]	16.7 [16.2–17.4]	135250.0 [86612.5–192250.0]	17.8 [16.9–18.2]	67,050 [51100–120750]	16.5 [16.1–17.0]	152500.0 [116500.0–199000.0]
ARG1	23.1 [22.1–23.9]	6180.0 [3710.0–11400.0]	24.0 [22.5–24.9]	3325.0 [1926.3–8793.8]	24.9 [24.2–25.7]	1952.5 [1146.3–2893.8]	24.1 [23.6–24.8]	3265.0 [2050.0–4260.0]
ASUN	26.7 [26.1–27.1]	846.0 [622.5–1242.5]	27.1 [26.6–27.7]	603.8 [419.5–869.3]	27.9 [26.7–28.5]	364.8 [237.4–803.6]	26.6 [26.0–26.8]	902.5 [772.0–1390.0]
BATF2	21.3 [20.7–22.8]	21950.0 [8210.0–33850.0]	25.4 [23.7–26.6]	1335.0 [585.0–4340.0]	26.4 [25.2–27.4]	688.3 [321.1–1525.0]	26.0 [24.9–27.0]	886.8 [439.1–1778.8]
C1QB	24.7 [24.2–25.9]	2190.0 [992.5–3190.0]	28.5 [27.3–29.5]	175.0 [94.0–396.3]	30.3 [28.8–31.2]	56.0 [29.3–144]	28.8 [27.6–29.5]	149.5 [91.9–325.0]
CD64	19.0 [18.0–19.5]	12700.0 [9020.0–24600.0]	21.9 [21.0–23.0]	1842.5 [871.8–3203.8]	23.3 [22.3–23.8]	691.8 [502.6–1402.5]	22.9 [22.3–23.2]	949.5 [774.0–1355.0]
DHX29	26.0 [25.4–26.4]	918.0 [694.0–1355.0]	26.7 [26.3–27.4]	560.0 [360.9–760.9]	27.2 [26.4–27.9]	407.8 [263.5–706.6]	26.4 [26.0–27.1]	719.0 [448.0–898.0]
DUSP3	26.1 [25.4–26.7]	442.5 [294.0–730.5]	28.0 [27.2–28.8]	133.3 [75.3–215.3]	29.2 [28.7–29.8]	58.7 [38.9–83.2]	28.2 [27.8–28.6]	110.5 [86.5–151.5]
GAS6	28.3 [27.6–29.1]	394.5 [245.0–652.0]	30.2 [29.3–30.9]	119.9 [78.0–204.0]	30.9 [29.8–31.7]	69.4 [40.1–154.9]	30.3 [29.8–30.8]	108.8 [76.8–153.9]
GBP5	19.7 [18.9–20.9]	9920 [4255.0–18150.0]	22.6 [21.7–23.2]	1312.5 [825.0–2493.8]	23.1 [22.9–24.3]	878 [371.3–1052]	22.7 [22.2–23.0]	1195.5 [928.5–1635.0]
GBP6	25.4 [24.5–26.3]	1085.0 [656.5–1990.0]	27.5 [26.7–28.5]	276.8 [144.7–484.6]	28.4 [27.2–29.1]	155.5 [104.8–372.4]	27.4 [26.5–27.8]	332.0 [239.0–533.5]
KLF2	21.7 [20.7–22.4]	34350.0 [22500.0–67700.0]	21.9 [21.5–22.6]	29,500 [19387.5–38175.0]	22.7 [21.8–23.2]	18,625 [12907.5–32112.5]	21.7 [21.3–22.1]	35650.0 [26050.0–44250.0]
NEMF	21.9 [21.2–22.2]	7710.0 [5950.0–11750.0]	22.4 [21.9–22.9]	5537.5 [3765.0–7436.3]	23.0 [22.2–23.4]	3577.5 [2806.3–6186.3]	21.8 [21.1–22.3]	7780.0 [5650.0–12650.0]
PTPRC	21.6 [20.9–22.0]	16600.0 [12600.0–28450.0]	21.3 [20.8–22.0]	20375.0 [12500.0–30037.5]	21.7 [21.2–22.2]	16,350 [10830–22600]	20.8 [20.4–21.4]	30100.0 [19000.0–38450.0]
SEPT4	33.2 [31.9–36.0]	6.0 [1.0–12.4]	34.8 [32.9–35.9]	3.1 [0.8–10.0]	34.3 [33.3–35.9]	2.1 [0.6–4.9]	33.3 [32.3–35.0]	5.0 [1.7–8.9]
ZNF296	25.4 [24.7–26.3]	460.0 [247.0–740.0]	25.6 [25.1–26.3]	411.0 [255.5–565.6]	26.0 [25.5–26.9]	310.5 [172.5–416.0]	25.2 [24.7–25.5]	535.0 [423.0–709.0]

**Table 4.** Assay quantitative read-out among well characterised groups of participants. \*All data is reported as median quantified level with IQR for each gene marker within each characterized group of participants.

### Discussion

The aim of this work was to identify, design, optimize and clinically evaluate a host gene-based RT-qPCR assay for TB diagnosis. A total of 16 gene targets were selected and converted into 5 multiplex RT-qPCR panels of two to four detection targets per panel and evaluated for best accuracy to diagnose ATB, LTBI and distinguish them



**Figure 7.** Host-gene marker expression levels among the study participants. Shown are scatter plots of measured expression levels among the participants namely HC ( $n = 37$ ), LTBI ( $n = 24$ ), ORD ( $n = 82$ ), ATB ( $n = 61$ ). Lines show the median and 95% confidence intervals. Statistical comparison was performed using the Kruskal-Wallis test and Dunn's Test with Bonferroni corrected p-values for multiple comparisons. \* =  $<0.05$ , \*\* =  $<0.01$ , \*\*\* =  $<0.001$ , and \*\*\*\* =  $<0.0001$ .

from ORDs and HC. While other studies attempted the RT-qPCR approach using SYBR green and single targets per reaction<sup>12,14,15,34,35</sup>, our RT-qPCR integrates dual-labelled hydrolysis probes to detect multiple targets in a single reaction minimising sample wastage, and reducing labour and material costs<sup>36</sup>. Furthermore, the SYBR green dye binds all double stranded DNA and is thus reported nonspecific, generating false positive signals<sup>36</sup>. In contrast, we have shown that our assay is specific and distinguishes ATB from LTBI, ORDs and HC individuals in a near point-of-care (POC) context.

In an in-vitro infection model, the designed primers detected and amplified HGM *GBP5*, *DUSP3*, *KLF2*, *GBP6*, *BATF2*, *CD64*, *SEPT4*, and *GAS6* in both uninfected and infected lung fibroblast cells. Differential expression between infected and uninfected lung fibroblasts was observed at 48 h and/or 72 h for *GBP5*, *GBP6*, *BATF2* and *CD64* genes an indication that the designed primers could optimally measure variation in amounts of expressed gene targets between infected and uninfected cells. Absence of activated gene expression at earlier times (4 h and 24 h) may signify the slow rate of endocytosis, hence delaying the signal transduction to induce gene expression. In the same context, it could be that genes *DUSP3*, *KLF2*, *SEPT4*, and *GAS6* needed longer time of pathogen exposure to exhibit differential gene expression. In-vivo, lung fibroblasts are constituent cells of the TB granuloma, which make them ideal in-vitro tissue culture model of infection for both MTB and BCG<sup>37,38</sup>. Furthermore, ex-vivo analysis of fibroblasts from normal lung tissue of LTBI individuals demonstrated that these cells were infected with MTB in-vivo<sup>39</sup>.

The multiplex RT-qPCR assay demonstrated excellent performance for target specific amplification with high efficiency and large dynamic range with true statistical doubling in each cycle. This indicated that optimal primer and probe pairs were designed and that reaction conditions were well optimized which ensured accurate detection of the targets<sup>30</sup>. The assay also displayed high precision, indicating consistent results with minimal variability. Consequently, the assay LOD and LOQ were determined demonstrating the assay's ability to quantify low concentration target genes in a multiplex reaction<sup>36</sup>. All this indicated that the detection signals of the assay were within the linear range making the assay robust, highly precise, and specific.

Clinical evaluation of the assay in a group of well-characterized clinical participants demonstrated clear differential expression of the genes among different groups supporting its potential for diagnostic application. All gene targets were detected below the LOQ Cq value with minimal dispersion among participants within the same group except for *SEPT4*. The absolute method used to quantify expression profiles is accurate since it accounts for the amplicon length<sup>36</sup>. Other researchers have also found the RT-qPCR assay a valuable tool for accurate quantification of transcriptional markers in whole blood for TB diagnosis although the quantification levels were rarely reported<sup>12-14</sup>. Assay quantitative read-outs are important for comparability and diagnostic cut-off determination which is crucial for sensitivity and assessing response to treatment in the context of managing ATB.

Our results are comparable with previous reports of upregulated expression of genes in ATB participants compared to ORDs, LTBI or HC<sup>12,14,33,40</sup> and HGM suppression in LTBI compared ATB<sup>4,15</sup>. This observed



suppressed expression profile in LTBI compared to HC individuals has been fully deciphered and reported by our team elsewhere<sup>25</sup>. Whilst a detailed evaluation of the expression profiles among ATB participants compared to ORDs participants and how the profiles in ATB change with treatment is being undertaken and will be reported separately. The clinical significance of the expressed expression profiles in ATB compared to ORDs with robust modelling to identify models highly discriminative of the two conditions will be included in that report. Nonetheless, the results reported here show that LTBI is associated with a suppressed expression profile while ATB leads to an upregulation of HGM an indication that different gene markers may be required for diagnosis of these two physiologically distinct states of TB disease. Additionally, immune regulatory genes namely *KLF2*, *NEMF*, *ASUN*, and *ZNF296* showed comparable expression in HC and ATB participants. Others have also reported similar expression levels of some of these HGM between HC and ATB individuals<sup>15</sup>. This implies that MTB may not necessarily be subverting host regulatory system in ATB as seen in LTBI. Future research studies to understand the normal level of expression exhibited by these immune regulators in ATB acute immune activation are warranted.

SEPT4 was reported as part of the Bloom144<sup>10</sup>, and Kaforou44<sup>4</sup> signatures for distinguishing ATB from ORDs, the Zak 16<sup>9</sup> and RISK4<sup>41</sup> signatures for ATB progression from LTBI and the Dawany251<sup>42</sup> signature for TB detection among HIV/TB infected and HIV mono-infected participants, although its quantitative read-outs were never reported. This SEPT4 performance was not replicated in our study. Nevertheless, since this target standard region was adequately amplified with good efficiency, the designed primer-probe pairs were optimal for its quantification and the possibility of this gene expression level being low in participant samples was more likely. Darboe et al. also reported poor reproducibility with high redundant signal yield for SEPT4 in a microfluidic RT-qPCR assay<sup>16</sup> which further indicates low levels of target gene in clinical samples for RT-qPCR measurements. This illustrates the need to translate microarray measured findings of DE genes into simple near POC test like the RT-qPCR to confirm reproducibility.

Composite biological functions of the evaluated HGM include immune response signalling and regulation<sup>43,44</sup>, defence mechanism against pathogens<sup>45–47</sup>, and biological interaction between organisms<sup>48</sup>, transcription initiators<sup>49</sup>, translation initiators<sup>50</sup>, ribosomal quality control gene<sup>51</sup>, and lastly genes involved in cell cycle mitosis<sup>52</sup>, proliferation<sup>53,54</sup>, and differentiation regulators<sup>55</sup> (refer to Supplementary Table 2). These results point to the need to translate microarray measured findings into simple near POC test like the RT-qPCR to confirm reproducibility. Combining RT-qPCR with Whole Genome Sequencing (WGS) to confirm size of the targets in participant samples after RNA extraction might also be beneficial.

We acknowledge limitations of this work which include use of BCG- an attenuated form of MTB complex to induce gene expression. BCG induced gene expression may not be the same as that induced by wildtype MTB, and in-vitro host-pathogen interaction may not recapitulate the in-vivo situation. However, this step was intended to verify whether the designed primers and probes could detect and amplify the target regions with differential HGM quantification between differing experimental conditions. Secondly, underlying cause of disease among ORDs participants was never confirmed. Clinical evaluation of the assay was performed in the same local Blantyre region and study findings may lack generalizability although the evaluated gene targets were discovered in different geographical regions and those frequently reported in different signatures were chosen. Also, LTBI and HC sample sizes were small, and this may impact statistical power of the analysis. Lastly, while our HGM panel targets several key genes associated with LTBI and ATB, there may be other genes that are also relevant to TB diagnosis not included. Nonetheless, this assay can easily be adopted for newly identified HGM thus continuous updating and refining of gene marker selection based on emerging research and clinical findings will provide a more comprehensive assay.

In conclusion, our study has provided a method to make HGM-based diagnosis of LTBI and ATB forms of TB at near point-of-care. This method is robust, highly sensitive and specific for gene expression quantification enabling prompt participant clinical classification. Further studies are needed to ascertain robustness and reproducibility of the gene marker expression in different genetic backgrounds.

## Data availability

All data generated and analysed during this study are available on the University of St Andrews OneDrive and is accessible upon request and meeting ethical requirements as per study participants consent from Dr Wilber Sabiiti ws31@st-andrews.ac.uk.

Received: 12 August 2024; Accepted: 21 October 2024

Published online: 06 November 2024

## References

1. WHO. WHO. Early detection of Tuberculosis: An overview of approaches, guidelines and tools. WHO/HTM/STB/PSI/2011.21. 1–32 at. (2011).
2. WHO. *High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting*. WHO Meeting Report (2014). <https://iris.who.int/handle/10665/135617>
3. Burel, J. G. & Peters, B. Discovering transcriptional signatures of disease for diagnosis versus mechanism. *Nature Reviews Immunology* at (2018). <https://doi.org/10.1038/nri.2018.26>
4. Kaforou, M. et al. Detection of tuberculosis in HIV-Infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med.* **10** (10), e1001538. <https://doi.org/10.1371/journal.pmed.1001538> (2013).
5. Bloom, C. I. et al. Detectable changes in the blood transcriptome are Present after two weeks of Antituberculosis Therapy. *PLoS One.* **7** (10), e46191. <https://doi.org/10.1371/journal.pone.0046191> (2012).
6. Walter, N. D. et al. Blood transcriptional biomarkers for active tuberculosis among patients in the United States: a case-control study with systematic cross-classifier evaluation. *J. Clin. Microbiol.* **54**, 274–282 (2016).
7. Roe, J. K. et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight.* **1**, 0–14 (2016).

8. Berry, M. P. R. et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. **466**, 973–977 (2010).
9. Zak, D. E. et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet*. **387** (10035), 2312–2322. [https://doi.org/10.1016/S0140-6736\(15\)01316-1](https://doi.org/10.1016/S0140-6736(15)01316-1) (2016).
10. Bloom, C. I. et al. Transcriptional blood signatures distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung cancers. *PLoS One*. <https://doi.org/10.1371/journal.pone.0070630> (2013).
11. Boger, B. B. et al. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. <https://doi.org/10.1016/j.ajic.2020.07.011> (2021).
12. Francisco, N. M. et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. *J. Infect.* <https://doi.org/10.1016/j.jinf.2017.09.012> (2017).
13. Maertzdorf, J. et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol. Med.* **8**, 86–95 (2016).
14. Costa, L. D. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis*. <https://doi.org/10.1016/j.tube.2015.04.008> (2015).
15. Lee, S. W. et al. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. *BMC Bioinform.* **17** (Suppl 1), S3. <https://doi.org/10.1186/s12859-015-0848-x> (2016).
16. Darboe, F. et al. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. *Tuberculosis* **108**, 124–126 (2018).
17. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
18. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–12 (2002).
19. Prediger, E. How to design primers and probes for PCR and qPCR | IDT. 2021 (2021).
20. IDT. OligoAnalyzer Tool - Primer analysis and Tm Calculator | IDT. <https://eu.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fanalyze>
21. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**, 402–408 (2001).
22. Sabiiti, W. et al. Tuberculosis bacillary load, an early marker of disease severity: the utility of Tuberculosis Molecular Bacterial load assay. *Thorax*. <https://doi.org/10.1136/thoraxjnl-2019-214238> (2020).
23. Sabiiti, W. et al. A Tuberculosis Molecular Bacterial Load Assay (TB-MBLA). *JoVE (Journal Vis. Exp.)*. **2020**, e60460 (2020).
24. Becker, P. History and progress in the accurate determination of the Avogadro constant. *Rep. Prog. Phys.* **64**, 1945 (2001).
25. Nakiboneka, R. et al. Suppression of host gene expression is associated with latent TB infection: a possible diagnostic biomarker. *Sci. Rep.* **14** (1), 15621. <https://doi.org/10.1038/s41598-024-66486-z> (2024).
26. QftXna, M. Qubit<sup>®</sup> RNA HS Assay Kits (Q32852, Q32855; Rev A.0). *Manual* 1–10 at (2015). [https://assets.thermofisher.com/TF5-Assets/LSG/manuals/Qubit\\_RNA\\_HS\\_Assay\\_UG.pdf](https://assets.thermofisher.com/TF5-Assets/LSG/manuals/Qubit_RNA_HS_Assay_UG.pdf)
27. QIAGEN. QuantiFERON<sup>®</sup>-TB Gold (QFT<sup>®</sup>) Elisa package insert. Qiagen www.QuantiFERON.com. (2018).
28. Team, P. & RStudio Integrated Development Environment for R. *Posit Software, PBC, Boston, MA.* (2023). <http://www.posit.co/>
29. Sample and Assay Technologies. Critical factors for successful real-time PCR. **1** (1), 1–63. Qiagen. (2010).
30. Bustin, S. A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* <https://doi.org/10.1373/clinchem.2008.112797> (2009).
31. Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Frontiers in Microbiology* vol. 8 at (2017). <https://doi.org/10.3389/fmicb.2017.00108>
32. Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med.* [https://doi.org/10.1016/S2213-2600\(16\)00048-5](https://doi.org/10.1016/S2213-2600(16)00048-5) (2016).
33. Gliddon, H. D. et al. Identification of reduced host transcriptomic signatures for tuberculosis and digital PCR-based validation and quantification. *bioRxiv* <https://doi.org/10.1101/583674> (2019).
34. Maertzdorf, J. et al. Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun.* **12**, 15–22 (2011).
35. Cai, Y. et al. Increased complement C1q level marks active disease in human tuberculosis. *PLoS ONE*. **9** (3), e92340. <https://doi.org/10.1371/journal.pone.0092340> (2014).
36. Bustin, S., Shipley, G. & Sethi, M. The future of qPCR Webinar Q&A: best practices, standardization and the MIQE guidelines. *Thermo Sci.* **9** (2011).
37. Barry, C. E. et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/nrmicro2236> (2009).
38. Mariotti, S. et al. Mycobacterium tuberculosis may escape helper T cell recognition by infecting human fibroblasts. *Hum. Immunol.* <https://doi.org/10.1016/j.humimm.2013.02.005> (2013).
39. Hernández-Pando, R. et al. Persistence of DNA from Mycobacterium tuberculosis in superficially normal lung tissue during latent infection. *Lancet*. [https://doi.org/10.1016/S0140-6736\(00\)03493-0](https://doi.org/10.1016/S0140-6736(00)03493-0) (2000).
40. Turner, C. T. et al. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. *Lancet Respir Med.* **8**, 407–419 (2020).
41. Suliman, S. et al. Four-gene pan-african blood signature predicts progression to tuberculosis. *Am. J. Respir Crit. Care Med.* **197**, 1198–1208 (2018).
42. Dawany, N. et al. Identification of a 251 gene expression signature that can accurately detect M. Tuberculosis in patients with and without HIV Co-infection. <https://doi.org/10.1371/journal.pone.0089925>
43. Perussia, B., Dayton, E. T., Lazarus, R., Fanning, V. & Trinchieri, G. Immune interferon induces the receptor for monomeric IgG1 on human monocytic and myeloid cells. *J. Exp. Med.* **158**, 1092–1113 (1983).
44. Jha, P. & Das, H. KLF2 in regulation of NF- $\kappa$ B-mediated immune cell function and inflammation. *Int. J. Mol. Sci.* **18** (11), 2383. <https://doi.org/10.3390/ijms18112383> (2017).
45. Shenoy, A. R. et al. GBP5 Promotes NLRP3 Inflammasome Assembly and Immunity in Mammals. *Science (80-.)*. **336**, 481 LP – 485 (2012).
46. Janeway, C. A., Travers, P. & Walport, M. The complement system and innate immunity - Immunobiology - NCBI Bookshelf. *Immunobiology: The Immune System in Health and Disease*. vol. 5th Editio 1–13 at (2001). <https://www.ncbi.nlm.nih.gov/books/NBK27100/>
47. Guler, R., Roy, S., Suzuki, H. & Brombacher, F. Targeting Batf2 for infectious diseases and cancer. *Oncotarget* **6**, 26575–26582 (2015).
48. Liu, P. F. et al. Guanylate-binding protein 6 is a novel biomarker for tumorigenesis and prognosis in tongue squamous cell carcinoma. *Clin. Oral Investig.* **24**, 2673–2682 (2020).
49. Matsuura, T., Miyazaki, S., Miyazaki, T., Tashiro, F. & Miyazaki, J. Ichi. Zfp296 negatively regulates H3K9 methylation in embryonic development as a component of heterochromatin. *Sci. Rep.* **7**, 1–12 (2017).
50. Parsyan, A. et al. The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc. Natl. Acad. Sci. U S A.* **106**, 22217–22222 (2009).
51. Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and assembly pathway of the ribosome quality control complex. *Mol. Cell.* **57**, 433–444 (2015).

52. Jodoin, J. N. et al. Human asunder promotes dynein recruitment and centrosomal tethering to the nucleus at mitotic entry. *Mol. Biol. Cell.* **23**, 4713–4724 (2012).
53. Wu, G. et al. Molecular insights of Gas6/TAM in cancer development and therapy. *Cell. Death Dis.* **8**, 1–10 (2017).
54. Keyse, S. M. Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell. Biol.* **12**, 186–192 (2000).
55. Du, Y. & Grandis, J. R. Receptor-type protein tyrosine phosphatases in cancer. *Chin. J. Cancer.* **34**, 61–69 (2015).

### Acknowledgements

We acknowledge the assistance of Jingyi Chen, who provided the MRC-5 lung fibroblast cell line. We would like to pay special tribute to the late Dr Christina Paulus, who provided critical technical support in the development of this methodology using MRC-5 cells. We are indebted to the study participants in Malawi who provided clinical samples to evaluate the assay.

### Author contributions

RN., WS., Conceived the study and interpreted the results; RN., NW., and EM., Performed the RT-qPCR assays, MNe., Provided the MRC5 cells, RN., Data analysis and interpretation, TN., CM., MNL., DS., WS., Provided intellectual expertise and supervised the study, RN., NW., MNe., DS., WS., Wrote the manuscript. All authors read and approved submission of the manuscript.

### Funding

This work was supported by the Wellcome Trust Institutional Strategic Support Fund of the University of St Andrews, 204821/Z/16/Z and the Global St Andrews Doctoral Scholarship. Additional funding was obtained from Helse Nord Tuberculosis Initiative (HNTI), Pathology Department, Kamuzu University of Health Sciences.

### Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-77164-5>.

**Correspondence** and requests for materials should be addressed to W.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024