Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis

Authors: Hanif Esmail[1,2,3*], Rachel P. Lai[4], Maia Lesosky[2,5], Katalin A. Wilkinson[2,4], Christine M. Graham[6], Stuart Horswell[4], Anna K. Coussens[2, 7], Clifton E. Barry 3rd [2,8,9,10], Anne O'Garra[6,11], Robert J. Wilkinson[1,2,4*]

Affiliations:

[1] Department of Medicine, Imperial College London, W2 1PG, United Kingdom

[2] Wellcome Center for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine and Department of Medicine, University of Cape Town, Observatory 7925, Republic of South Africa

[3] Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

[4] The Francis Crick Institute, London, NW1 2AT United Kingdom

[5] Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine, University of Cape Town, Cape Town, Republic of South Africa

[6] The Francis Crick Institute, Laboratory of Immunoregulation and Infection, London, NW1 2AT United Kingdom

[7] Division of Medical Microbiology, Department of Pathology, University of Cape Town, Observatory 7925, Republic of South Africa

[8] Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Republic of South Africa

[9] Tuberculosis Research Section, NIAID, NIH, Bethesda, USA

[10] Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, Republic of South Africa

[11] National Heart and Lung Institute, Imperial College London, W2 1PJ, United Kingdom

*To whom correspondence should be addressed: hanif.esmail@ndcls.ox.ac.uk and r.j.wilkinson@imperial.ac.uk

Competing interests:

The authors have no competing interests to declare

Significance statement

Understanding the events in early tuberculosis disease will facilitate the development of novel tests to predict disease progression and interventions to prevent it. Blood-based transcriptomic approaches consistently identify several pathways relevant to clinical disease. Here we show in asymptomatic people with HIV infection and early subclinical tuberculosis defined by FDG-PET/CT, that transcripts relating to the classical complement pathway and Fcγ-receptor remain enriched, accompanied by rising levels of circulating antibody/antigen immune complexes. We confirm that transcripts relating to these pathways also rise in the 12 months prior to disease presentation in HIV-uninfected people. This supports observations that antigen may be present in early disease despite being paucibacillary, and demonstrates that modulation of the immune response could occur via immune complex formation.

Abstract

The transition between latent and active tuberculosis (TB) occurs prior to symptom onset. Better understanding of the early events in subclinical disease will facilitate the development of diagnostics and interventions that improve TB control. This is particularly relevant in the context of HIV-1 co-infection where progression of TB is more likely.

In a recent study using [^{18}F]-fluoro-2-deoxy-D-glucose positron emission/computed tomography (FDG-PET/CT) on 35 asymptomatic, HIV-1 infected adults we identified ten participants with radiographic evidence of subclinical disease, significantly more likely to progress than the 25 participants without. To gain insight into the biological events in early disease we performed blood-based whole genome transcriptomic analysis on these participants and 15 active TB patients. We found transcripts representing the classical complement pathway and Fcγ receptor-1 overabundant from subclinical stages of disease. Levels of circulating immune (antibody/antigen) complexes also increased in subclinical disease and were highly correlated with C1q transcript abundance. To validate our findings, we analyzed transcriptomic data from a publicly available dataset where samples were available in the 2 years prior to TB disease presentation. Transcripts representing the classical complement pathway and Fcγ receptor-1 were also differentially expressed in the 12 months prior to disease presentation.

Our results indicate that levels of antibody/antigen complexes increase early in disease, associated with increased gene expression of C1q and Fcγ receptors that bind them. Understanding the role this plays in disease progression may facilitate development of

interventions that prevent this leading to a more favorable outcome and may also be important to diagnostic development.

/body

Introduction

Conventionally, tuberculosis (TB) is divided into stages of asymptomatic latent infection, during which bacillary replication is effectively controlled in a healthy host, and active disease in which this has failed, resulting in symptomatic deterioration. Understanding the transition between these two states is important, although until recently this has been over-looked (1). Active disease is usually defined by a combination of symptoms, pathology (radiographically or histologically identified) and culturable bacilli. These features do not appear simultaneously but develop over time and may be intermittently present, hence the active disease processes may begin months before symptom onset i.e. as subclinical active disease (2). In some of those who initially reactivate, disease progression may be arrested and regress particularly when disease extent is limited(3). A greater understanding of the early events in disease is critical for the development of novel approaches to both identify people in early subclinical stages of disease and interventions to prevent progression. This is of particular importance in those with HIV-1 co-infection where TB disease progression is more likely. However, studying the natural history of TB in this group is further complicated by the imperative to provide preventive therapy, so novel approaches are needed.

In macaques the failure of the granuloma has been shown to be followed by cellular infiltration within bronchi (pneumonia)(4). In human autopsy studies pneumonic infiltration has also been

observed as the first pathological sign of pulmonary disease(5). Disease regression and self-healing of lesions is associated with fibrosis, however, it appears that disease risk following this is significantly increased(6). Medical imaging is a key approach to detecting evidence of early disease in asymptomatic persons, facilitated by a characteristic distribution of pathology. Chest radiography (CXR) has long been used for this purpose, however, CXR has limitations in both sensitivity and reproducibility(7). [^{18}F]-fluoro-2-deoxy-D-glucose positron emission/computed tomography (FDG-PET/CT) is a highly sensitive imaging modality, which combines cross-sectional anatomical detail from a CT scan with quantitation and localization of metabolic activity by quantifying uptake of FDG (a glucose analogue) by PET scan. FDG uptake in tuberculosis is related to activated macrophages and an inflammatory infiltrate at the site of disease(8). In a recent clinical study, we utilized FDG-PET/CT to identify evidence of early disease in asymptomatic people otherwise conventionally considered to have latent infection (QuantiFERON Gold in-tube (QF-GIT) positive, sputum culture negative, CXR no evidence of active disease) but at high risk of disease progression as HIV-1 infected, anti-retroviral therapy naïve. Of the 35 people scanned 10 (28.6%) had evidence of subclinical disease by FDG-PET/CT and they were significantly more likely to progress to clinical disease(2).

Transcriptomic profiling of peripheral blood has provided critical insight into active TB (9-12). Blankley *et al* have recently analyzed 16 such studies and shown overall findings in modular analysis of the transcriptome to be remarkably consistent. Meta-profiling identified 380 genes differentially abundant in active TB in at least 9 of the 16 studies. Based on pathway analysis they then determined the main functional groups of genes in active TB to relate to; Interferon signaling/inducible genes, the complement pathway, immunoglobulin receptors, recognition of

pathogen associated molecular patterns, inflammasome and pro-inflammatory pathways, (13, 14). However, it is not clear which of these processes occur at the earliest stages of disease. Understanding such host-pathogen interaction in early disease may provide insight that could facilitates novel approaches to diagnose and manage subclinical disease. Zak *et al* have recently demonstrated in HIV-1 uninfected adolescents that transcriptional differences are present in blood months prior to clinical patency suggesting that transcriptomic profiling of peripheral blood might be a useful approach to provide insight into subclinical disease.

In the present study, we utilized the cohort of HIV-1 infected adults described above in which presence or absence of subclinical TB was defined by FDG-PET/CT(2) and co-recruited control participants with active TB. We investigated which components of the active TB whole blood transcriptional profile were present in those with evidence of subclinical disease to provide insight into the biology underlying early progressive disease.

Results

Details of the clinical and radiographic features of the clinical cohort have been previously described(2). Thirty-five asymptomatic, healthy, HIV-1 infected, ART naïve participants with median CD4 count of 517/mm^3, a positive QFN-GIT and no previous history of TB, were screened for active TB by two sputum cultures (induced if needed) and CXR. All participants were of African ancestry resident in Khayelitsha (township of Cape Town, South Africa). Following 42-day negative sputum cultures, participants were confirmed as remaining symptom free, underwent FDG-PET/CT and were commenced on isoniazid preventive therapy. FDG-PET/CT scans were systematically analyzed, with abnormalities categorized and spatially

mapped, interpreted in the context of historical autopsy studies and then classified as being consistent with subclinical active TB or not. Ten of the 35 participants were classified as having subclinical active TB, due to either infiltrates and/or fibrotic scars (n = 9 – consistent with bronchogenic spread of disease) or multiple active nodules (n = 1 – consistent with haematogenous spread of disease) on FDG-PET/CT (referred to as "subclinical TB"); the remaining 25 participants had no evidence of disease activity on FDG-PET/CT scan (referred to as "latent TB"). Four out of 10 participants with evidence of subclinical TB were determined to have disease progression (defined as developing TB symptoms along with radiographic deterioration or culture positivity) requiring commencement of standard TB therapy in contrast to 0/25 participants with no evidence of subclinical TB. Participants were followed up for 6 months, 27 of 35 participants (six subclinical and 21 latent) then had repeat FDG-PET/CT. All six of the subclinical participants had improvement in baseline abnormalities in lymph nodes and lung parenchyma in contrast to only 1/21 of those without evidence of subclinical TB. Three participants (2 in subclinical group and 1 in the latent group) had been commenced on antiretroviral therapy during follow up as CD4 count fell below 350/mm^3, in accordance with local standard of care at the time of recruitment.

In addition, we recruited 15 participants with symptomatic, microbiologically confirmed pulmonary TB (referred to as "active TB") for blood sampling. There were no significant differences between those with active, subclinical and latent TB with regard to median age ($p=0.41$), sex ($p=0.52$) or median CD4 count ($p=0.09$). Conversely, median viral load (VL) and median C-reactive protein (CRP) were significantly different between active and latent TB ($p<0.0001$ and $p=0.0001$; respectively), but not between subclinical and latent TB ($p=0.97$ and $p=0.08$ respectively) (**SI Appendix, Figure S1**).

Eighty-two transcripts distinguish subclinical and active TB from latent TB

Transcript abundance in whole blood RNA was determined by Illumina HumanHT-12 v 4.0 Expression BeadChips microarray. We initially defined a set of transcripts that were differentially abundant between latent and active TB to determine a transcriptional profile of active TB. From the 47,231 probes on the microarray, 18,919 were present in at least 10% of samples (see methods). The 18919 transcripts were then statistically filtered (ANOVA with Benjamini Hochberg FDR correction ($p_{corr} < 0.05$) to identify 2573 transcripts showing significant differential abundance between the different stages of TB. From these transcripts, we identified 893 transcripts with at least a 1.5-fold difference in abundance between active and latent TB. To determine which of these transcripts might be relevant to early disease, we filtered the 893 transcripts by fold change to identify those differentially abundant between subclinical and latent TB. As disease pathology in subclinical TB is of minimal extent by comparison with active TB, we reasoned that the relevant transcripts may be expressed in blood at lower levels in comparison to active TB and hence both 1.25 and 1.5-fold change cut offs were used resulting in a 203 and 82-transcript list respectively (**SI Appendix, Figure S2 and Supp. Excel sheet**).

Hierarchical clustering was then performed on the 893, 203 and 82 transcripts which confirmed progressive improvement in ability to classify those with subclinical TB from those with latent TB while clustering them with active TB (**Figure 1a**). Two participants with subclinical TB did not cluster with active TB, one with multiple active nodules consistent with haematogeous spread of disease and the other who, of those with abnormalities consistent with bronchogenic spread, had the smallest infiltrate with amongst the lowest metabolic activity in lesion or lymph

node by FDG uptake (**SI Appendix, Table S1**). In 27 participants seen at 6-month follow-up after receiving either IPT or standard TB therapy the subclinical and latent groups became indistinguishable by transcriptional signature (**Figure 1b**).

To examine if abundance of these transcripts altered as subclinical disease activity increased we investigated the association of metabolic activity by FDG uptake with transcript abundance. Participants with subclinical TB were further divided into two categories; five participants showing intense FDG uptake (Visual score (VS) =3) within lung parenchyma or central lymph nodes and five participants with less intense FDG uptake (VS=0-2) (**SI Appendix, Table S1**). Hierarchical clustering using the group mean expression values for the 82-transcripts demonstrated that the subclinical participants with greater metabolic activity, clustered more closely with active TB than the subclinical TB group with low FDG uptake, which in turn clustered closer to the latent TB group (**Figure 1c**). This was then further demonstrated by disease risk score (determined for each participant by subtracting the summed normalized expression values of under-abundant transcripts from that of over-abundant transcripts as described in (12)) (**Figure 1d)**. There was no significant difference in mean C-reactive Protein (CRP) (2.62 mg/L vs 2.88 mg/L; p=0.85 – Normal range <5 mg/L), white cell count (WCC) (5.72 $x10^9$/L vs 5.77 $x10^9$/L; p=0.97 – Normal range 4.5-11 $x10^9$/L) or Log (HIV VL (copies/mL)) (4.29 vs 4.36; p=0.86) between the participants with subclinical TB as high or low metabolic activity on PET/CT.

To evaluate if HIV-1 viral load contributed the observed differential abundance between TB disease states in the 82-transcript signature, we performed multiple regression using TB status (latent, subclinical, active) and HIV viral load as explanatory variables for expression score. TB

status but not HIV VL was significantly associated (r^2=0.75, p<0.0001; TB status, coef=55.9, p<0.001; HIV VL, coef=9.18, p=0.08).

The 893, 203 and 82 transcripts mapped to 678, 181 and 72 genes respectively, of these 174, 39 and 17 genes respectively are present in the 380 meta-gene signature for active TB, identified by Blankley *et al* as being present in at least 9 other studies hence consistently associated with TB(13), (**SI Appendix, Table S2 and Supp. Excel sheet**).

Classical complement pathway transcripts are overabundant in subclinical disease and correlate with the concentration of circulating immune complexes

To identify biological processes common to subclinical and active TB, pathway analysis was performed on the three transcript signatures (893, 203 and 82). In the 893-transcript active TB signature showed complement and interferon signaling to be the most enriched canonical pathways (each, p=0.0002), followed by death receptor signaling and JAK family kinase in IL-6 signaling (**Figure 2a**). Of these, only complement remained over-represented in the 203 and 82-transcript signatures (p=0.0001 and p=0.006 respectively) suggesting that complement may be relevant in early active disease. Using only the 174, 39 and 17 genes common to the 380-meta signature this pattern was retained with complement remaining the most enriched pathway in all 3 gene lists (**SI Appendix, Figure S4**)

Predominantly components of the classical complement pathway (complement component 1q subcomponent B (C1QB), serpin peptidase inhibitor member 1 (SERPING1), complement component 2 (C2) and complement component 5 (C5)) were overabundant in the blood of subclinical and active TB patients compared to those with latent TB (**Figure 2b**). Of these

C1QB, SERPING1 and C2 are all in the 380-metasignature derived by Blankley *et al*, with C1QB and SERPING1 being overabundant in active TB for 15 of the 16 studies analyzed.

To further investigate the role of complement in subclinical TB, we next evaluated the relationship between serum concentration of complement proteins C1q, SERPING1 and C5 and their whole blood transcript abundance using paired samples. There was no correlation between serum concentration and transcript abundance for C1q (r_s=-0.08, p=0.59), SERPING1 (r_s=0.02, p=0.89) or C5 (r_s=0.16, p=0.29) (**Figure 3a, c, e**). Because the classical complement pathway is triggered by C1q binding of antibody/antigen immune complexes(15), we next evaluated the relationship between transcript abundance of C1QB and serum concentration of circulating immune complex (CIC – antibody/antigen complexes) as determined by a C1q binding assay. Transcript abundance of C1QB in whole blood was significantly correlated with CIC (r_s=0.48, p=0.0005) as was SERPING1 (r_s=0.46, p=0.0008) but not C5, where a trend was observed albeit not significant (r_s=0.26, p=0.08) (**Figure 3b, d, f**). As both TB and HIV can contribute to levels of CIC we performed multiple regression by TB status (latent, subclinical, active) and HIV viral load as explanatory variables for CIC. TB status but not HIV VL had a significant association (r^2=0.26, p=0.001; HIV VL, coef=0.10, p=0.25; TB status, coef=0.31, p=0.008).

We next evaluated the relationship between CIC and metabolic activity in subclinical TB classifying participants as having low or high metabolic activity as above. CIC was significantly increased across different TB disease states (p=0.0004), *post hoc* analysis controlling for false discovery demonstrated that those with active TB (p=0.0001, q=0.0002) and those with subclinical TB and high metabolic activity on PET/CT (p=0.038, q=0.04) had higher serum concentration of CIC than latent TB (**Figure 3g**). Furthermore serum CIC as a single marker to

discriminate the four individuals subsequently treated for TB during follow-up from the 31 individuals that were not, showed an AUC of 0.754. A cut off of 192 μg Eq/ml had sensitivity of 75% and specificity of 93.6% and correctly classified 91.4%. Of note the only participant of the four subsequently treated for active TB with a serum CIC less than 192 μg Eq/ml had a different radiographic presentation subclinical disease, with multiple nodules suggestive of haematogenous spread in contrast to the bronchogenic pattern of disease seen in the remainder.

In addition to binding C1q, antibody/antigen immune complexes also bind Fcγ receptors and Fcγ Binding Protein (FCGBP). Those with subclinical TB and active TB also had relative overabundance of transcripts relating to the Fc fragment of IgG, high affinity I receptor (CD64) (FCGR1) with FCGR1A/B/C all being present in the 82 and/or 203 transcript list, however there was a reduction in abundance of FCGBP (**Figure 2c and SI Appendix, Table S2**). FCGR1A, FCGR1B and FCGBP were all present in the 380 meta-signature. FCGBP is the most consistently underabundant transcript in active TB identified in 14 previous studies. By contrast FCGR1B and FCGR1A are amongst the most consistently overabundant transcripts in active TB being present in 16 and 14 studies respectively. RT-PCR was also conducted using primers for FCGR1C, FCGBP, C1QB and SERPING1 which confirmed the differences demonstrated by microarray between latent, subclinical and active TB (**SI Appendix, Figure S3**).

Classical complement and interferon signaling pathways are increased in progressive TB in HIV uninfected persons

To validate our findings and establish if classical complement and Fc gamma receptor gene activation were important in early TB disease, we reanalyzed the RNA sequencing data from the

277 publically available dataset from the study reported by Zak *et al*, in which, samples were
278 available in the 2 years prior to disease presentation with appropriate non-progressor
279 controls(16). Although no accurate subclinical disease phenotype was ascertained, we reasoned
280 that those in the 6 to 12 months prior to disease presentation may have had subclinical disease.
281 Furthermore, participants in Zak *et al* were all HIV uninfected allowing us to determine whether
282 our findings were generalizable. To investigate the biological processes relevant to progressing
283 disease, we first established the number of transcripts significantly differentially abundant (p_{corr}
284 <0.05) and with >1.5FC overabundance between progressors and non-progressors at <180 days,
285 181-360 days, 361-540 days and 541-720 days prior to disease presentation, identifying 362,
286 154, 3 and 3 genes respectively (**see methods and Supp. Excel sheet**). We then conducted IPA
287 analysis for the <180 day and 181-360 time points. The complement pathway was the second and
288 third most significantly enriched canonical pathway at 181-360 days ($p=5.80x10^{-5}$ (11.1% of
289 pathway, $p=5.80x10^{-5}$) and <180 days prior to presentation ($p=6.02x10^{-7}$ (19.4% of pathway,
290 $p=6.02x10^{-7}$) respectively, with the interferon signaling pathway the most significantly enriched
291 at these time points (Figure 4). In keeping with our initial results, the classical complement
292 pathway was largely responsible for this with C1QB/C, SERPING1 and C2 all significantly
293 differentially expressed at both time points and C5 and C1QA also significantly differentially
294 expressed <180 days before presentation. In addition, FCGR1A/B/CP were all amongst the most
295 significantly differentially expressed genes between progressors and non-progressors at both
296 time points (all in top 22 genes for both) (**Figure 4**).
297
298 *Interferon signaling transcripts are increased in both untreated HIV and active TB*

Enrichment of the complement pathway was observed in early TB disease states in both HIV-infected and un-infected persons, however, enrichment of interferon signaling was less prominent in HIV-associated TB. We further explored this by undertaking a modular analysis of the whole blood transcriptome in HIV and active TB. Modules are groups of co-regulated transcripts subsequently categorized into functional groups through unbiased literature review. The principles informing this approach have been previously published (17). Using this approach, the modular profile of active TB has been previously shown to be remarkably consistent across 16 studies (13).

For this analysis, in addition to the HIV-1 infected, ART naïve participants with symptomatic active pulmonary TB (HIV+ART-TB+; n=15) and with no evidence of active TB (HIV+ART-TB-; n=25, i.e. excluding the 10 with subclinical active TB), we recruited HIV uninfected participants with symptomatic, active, pulmonary TB (HIV-TB+; n=14) and with no evidence of active TB (HIV-TB-; n=15). We also recruited HIV-1 infected participants established on ART fully with viral load fully suppressed and no evidence of active TB (HIV+ART+TB-; n=8). All these additional participants were also resident Khayelitsha and of African ancestry.

18,751 transcripts were "present" in at least 10% of samples and included in the modular analysis, using HIV-TB- participants as the primary control group. The average transcript abundance for each module was determined for each participant. Of the 260 modules previously characterized, 38 had functional roles determined and analysis was restricted to these modules, three of these modules relate to interferon signaling.

By comparison HIV-TB- participants, participants with untreated HIV only (HIV+ART-TB-) had significant differences (after correcting for false discovery) in modular expression in 8/38 modules (5 increased expression and 3 reduced expression), those with TB only (HIV-TB+) had significant differences in expression in 24/38 modules (15 increased expression and 9 reduced expression) (**SI Appendix, Table S3**). Both the HIV only group and TB only group had significantly increased expression in all three interferon modules. Those with HIV-associated TB (HIV+ART-TB+) had greater expression of the interferon modules in comparison to those with HIV or TB only (**Figure 5 and SI Appendix, Table 3**). Those with HIV but on ART (HIV+ART+TB-) with fully suppressed viral load showed normalization of expression within the interferon modules and had no significant differences in any of the 38 modules in comparison to HIV-TB- controls. Given the background increased abundance of interferon related transcripts in HIV infected persons not on ART it is likely that any increases expression within these pathways in early tuberculosis would be less discriminatory.

Recognizing that interferons are well known to enhance expression of Fcγ receptor we next sought to establish the effect of interferon on expression of complement components. We cross-referenced the 36 genes comprising the Ingenuity IPA complement canonical pathway against the Interferome database, a database of interferon regulated genes (IRG) from a variety of experimental settings (http://www.interferome.org/interferome/home.jspx). All human data on IRG were from *in vitro* experiments, we restricted analysis to studies on human blood and genes showing > 2-fold increase following interferon stimulation. Of the 36 genes, 9 (25%) had evidence of increased expression on interferon stimulation; C1QA/B/C, C2, CFB, CR1, CR2, ITGAX SERPING1. While abundance of some of these potentially interferon regulated

345 complement components were enriched in tuberculosis others were not, in addition C5 did not
346 have experimental evidence of interferon regulation. Furthermore overall abundance of
347 transcripts relating the to the TB relevant complement and FCGR1 components (C1QA/B/C,
348 SERPING1, C2, C5, FCGR1A/B/C) while significantly increased in TB, was not significantly
349 affected by HIV status (two-way ANOVA with TB and HIV as explanatory variables of disease
350 risk score; r^2 0.58, TB $p<0.0001$ and HIV $p=0.12$) with discriminatory ability of these transcripts
351 for TB in both HIV infected and uninfected persons excellent (disease risk score; HIV infected -
352 AUC 0.997, HIV uninfected – AUC 0.938) (**SI Appendix, Figure S5**). It is therefore unlikely
353 that the complement pathway member upregulated in TB merely reflect interferon responsive
354 genes.
355
356 To establish if this pattern of complement and FCGR transcript expression was seen in other
357 diseases we analysed the dataset (GEO dataset- GSE42834) of Bloom et al in which participants
358 with active TB, active sarcoid, non-active sarcoid, lung cancer, pneumonia and healthy controls
359 had whole blood transcriptional profiling undertaken(18). We profiled transcript abundance of
360 the 81 complement and FCGR transcripts (62 transcripts relating to the 36 complement genes in
361 the IPA database and all 19 FCGR transcripts on the platform). 19 of 81 transcripts showed
362 greater than 2-fold difference in transcript abundance in at least one of the disease categories
363 when compared to healthy controls. In initial hierarchical clustering of disease groups for these
364 19 transcripts showed active TB clustered with active sarcoid, and community-acquired
365 pneumonia clustered with lung cancer **(SI Appendix, Figure S6)**. We then performed K-means
366 clustering (3 clusters, 50 iterations, Euclidean distance metric) to provide an unbiased assessment
367 of the pattern of transcript abundance across the different diseases. One cluster contained

transcripts showing greatest abundance in tuberculosis over other diseases included C1QB, SERPING1, FCGR1A, FCGR1B and FCGR1C. The cluster containing transcripts showing greatest abundance in diseases other than tuberculosis (particularly pneumonia and lung cancer) included CR1, ITGAM, ITGAX, C3AR1, FCGR2A and FCGR3B. The final cluster contained FCGRBP as a single transcript that was under abundant in tuberculosis, community-pneumonia and lung cancer in comparison to control. These data support our findings of complement and FCGR expression in tuberculosis, demonstrated in other cohorts presented in the paper and also re-emphasis, that complement component expression is not simply a function of interferon stimulation or a generic inflammatory signal.

Discussion

We have previously shown that in a subgroup of asymptomatic, HIV-1 infected persons, conventionally diagnosed with latent TB, evidence of subclinical disease can be identified utilizing high resolution imaging (FDG-PET/CT) which is not apparent by routine clinical screening. Those with evidence of subclinical disease on FDG-PET/CT were also found to be more likely to progress clinically and to require treatment for active TB (2). In this study we have identified components of the active TB whole blood transcriptional response that are also present in subclinical disease and shown that circulating immune complexes rise during subclinical disease.

Previous transcriptional studies have consistently indicated several key pathways in active TB (interferon signaling/inducing genes, inflammasome/pro-inflammatory pathways, recognition of PAMP, the complement pathway and immunoglobulin receptors) but these studies have not been

able to dissect the biological processes that may occur at the outset of disease. We determined that a subset of transcripts in the whole blood transcriptional response to symptomatic active TB were relevant to subclinical disease and that these transcripts were enriched for the classical complement pathway and immunoglobulin receptors (FCγ receptor 1), suggesting a cellular response to antibody-antigen complexes may occur early in disease. This was supported by our finding that CIC in serum increased in relation to TB disease status and was highly correlated with transcript abundance of C1q. Furthermore, with increasing metabolic activity of subclinical disease the transcriptional response aligned more closely with active TB and levels of CIC increased. Importantly the abundance of these transcripts and levels of CIC did not show any significant relationship with HIV viral load.

We validated our findings utilizing the publicly available data reported by Zak *et al* (16). While subclinical disease was not characterized by imaging, regular samples were available up to 2 years prior to clinical disease presentation in HIV uninfected persons. Of note in contrast to our study, Zak *et al* also excluded participants who developed TB within 6 months of enrolment in the training set and 3 months of enrolment in the validation set. Our analysis confirmed that transcripts relating to the classical complement pathway and immunoglobulin receptors increased during the 12 months prior to disease presentation in HIV uninfected persons. However, in addition interferon signaling pathways also appear to increase in this population. Whilst interferon-signaling pathways were prominent in our 893-transcript signature differentiating active from latent TB in HIV infected persons they became less prominent as this signature was filtered to allow better distinction of subclinical TB from latent TB. We showed that untreated HIV infection itself leads to increases in interferon related transcripts in whole blood. Hence, the

finding that the interferon signaling pathway transcripts were less discriminatory in early disease in those with untreated HIV is likely due to the high background level of these transcripts in this population. We also showed that the pattern of complement and FCGR expression varied in different pulmonary diseases. While the pattern of expression in tuberculosis was also seen in active sarcoidosis though at a lower level of expression, a different pattern of expression was seen in community-acquired pneumonia and lung cancer where expression of C3 binding receptors was more prominent. Tuberculosis and sarcoidosis have previously been shown to have many similarities in whole blood transcriptional profile(18, 19).

Transcripts relating to the classical complement pathway (in particular, C1Q and SERPING1) and Fcγ Receptor 1 are amongst the most consistently overabundant in active TB and FCGBP as the most consistently underabundant supporting their role in early disease. Cliff *et al* also reported that transcripts relating to the complement pathway decrease within the first week of treatment for active TB(20). This suggests that the abundance of transcripts relating to the complement pathway may mark a rising antigen load rising early in disease, falling rapidly with treatment. CIC has frequently been reported to be elevated in active tuberculosis, increasing with extent of disease and reducing with treatment, CIC have been shown to contain secreted TB antigens (e.g. 38kDa and 30kDa (Ag85)) and non-secreted antigens (e.g. 16kDa (HspX)) (21, 22). In addition to opsonisation and complement activation, antibodies have an additional functional role in regulating the immune response through interaction with innate cells via a variety of Fc Receptors which may directly affect effector mechanisms and clearance of the pathogen (23-25).

Hunter *et al* investigated the histological appearances of early pulmonary TB in humans in several studies. They demonstrated that the earliest feature of disease is a lipid-rich pneumonia in which alveoli are filled with foamy macrophages with CD4 cells present at significantly lower frequency than in the granuloma. They also observed large clusters of B-cells bordering this region. Very few acid-fast bacilli (AFB) are visible at this stage but abundant mycobacterial antigen is evident on immunohistochemical staining. As necrosis develops within the pneumonia there is then a marked increase in numbers of AFB (26-28). Increased concentration of both 38kDa TB antigen and 38kDa specific IgG have been identified in bronchoalveolar lavage samples of TB patients(29).

Based on our findings we propose that at this early stage of disease, the increase in mycobacterial antigen concentration results in an increased concentration of antibody-antigen complexes increases at the site of disease. This may lead to increased expression of components of the classical complement pathway in recruited cells (to facilitate the targeted delivery of complement) and up-regulation of Fcγ receptor on monocytes and neutrophils to promote phagocytosis, as we have observed. By contrast expression of FCGBP is reduced. FCGBP has been shown to be the most consistently underabundant gene in active compared to latent TB but little is known about its function. Kobayashi *et al* have shown that Fcγ Binding Protein is present in sputum as well as other bodily secretions and that it inhibits activation of classical complement pathway(30). Its reduction in TB disease may therefore facilitate C1q binding of immune complexes promoting complement activation.

Formation of immune complexes in disease has both beneficial and detrimental effects for the host and the complement system can both promote and mitigate immunopathology. Immune complex aggregates that precipitate and persist within the lung are likely to be damaging through C3a/C5a mediated neutrophil recruitment increasing inflammation and tissue necrosis which may promote bacillary multiplication(31). It has been suggested that solubilization and diffusion of immune complexes away from tissue via extracellular fluid and the systemic circulation for clearance by mononuclear phagocytic system is critical to minimize local inflammation and tissue damage (32, 33). This appears to be facilitated by rapid C1q initiation of the classical complement pathway and C3b deposition which results in reduction in the size of the immune complex aggregates to maintain solubility, deficiencies in classical complement components reduce immune complex solubility and increase risk of immune complex mediated autoimmune diseases (32, 33). It is possible, therefore, that the increased expression of classical complement components in tuberculosis is directly in response to increased production of immune complexes at the site of disease to allow localized delivery of C1q to inhibit the precipitation of immune complexes and minimize lung damage. Although, our insight has been informed by analysis of blood, Cai *et al* demonstrated that C1q mRNA and protein levels were greater at the site of disease (in pleural fluid and bronchoalveolar lavage fluid) compared to blood in active pleural and pulmonary tuberculosis (34).

We found that although abundance of classical complement components correlated with circulating immune complexes they did not correlate well with serum levels of the proteins they encode. There are several possible explanations of this. For blood cells migrating to the disease site, expression as well as consumption of complement may occur locally. In addition,

complement components may also be bound to immune complexes whereas the assay would only measure unbound complement proteins. Furthermore, whole the blood transcriptome predominantly reflects mRNA expression in leukocytes whereas circulating complement proteins are manufactured within the liver and by a number other cell types (e.g. epithelial and endothelial cells) affecting the correlation of whole blood RNA and serum protein(35).

There are limitations of this study. We used a broad definition of subclinical disease which grouped together several distinct abnormalities found to be consistent with early disease, with reference to historical autopsy studies. Our sample size was modest and participants often had a combination of abnormalities and hence we were unable to tease out transcriptional differences between participants with potentially different disease pathogenesis. For example, there may be differences between those presenting with multiple nodules suggestive of haematogeneous spread of disease in comparison to disease spreading bronchogenically which may be interesting the characterize. Larger studies will be needed explore this in greater detail and further validate findings. In addition, we used peripheral blood to investigate pathogenesis of disease, while this approach is widely adopted, confirmation of our findings is needed in samples from site of disease. Our study was conducted in ART naïve, HIV infected participants with well-preserved CD4 counts with validation in HIV uninfected adolescents, confirmatory studies in other populations are required. Furthermore, although we demonstrated that CIC rises in relation to disease state in tuberculosis. This may have a role as a biomarker for identification of subclinical disease, however further studies will be needed confirm these findings in independent cohorts.

Our results indicate that the abundance of transcripts relating to the classical complement pathway and Fcγ receptors is increased from early disease in both untreated HIV infected adults and HIV uninfected adolescents in contrast to transcripts related to interferon signaling which appear less discriminatory in untreated HIV infected people. This will be an important consideration in the development of universal transcript-based biomarkers that are predictive of disease. Our findings support the observation that in early disease, though paucibacillary, antigen may still be present. This may have the potential to modulate the immune response by forming antibody complexes that initiate the complement pathway and bind Fcγ receptors upregulated on cells. However further studies will be needed to clarify in TB, how immune complexes form, their composition and the role they play in early disease. Developing a greater understanding of the early events in TB disease will be important to future TB control efforts. Not all those who initially reactivate latent infection and develop evidence of early subclinical disease will go on to develop symptomatic clinical disease, as regression will occur in a proportion. Determining the factors that govern progression or regression of disease will allow us to develop rational interventions that may lead to more favorable outcomes and prevent disease.

Materials and Methods

Recruitment of participants

Ethical approval for this study was provided by the research ethics committees of the University of Cape Town (013/2011) and Stellenbosch University (N12/11/079). Recruitment and classification of participants whose samples were utilized for this study have been described in detail previously(2). Briefly all participants and controls were HIV-1 infected adults, who were

ART naïve and resident in Khayelitsha, a peri-urban township of Cape Town, South Africa where >95% of people are of Xhosa origin. Recruitment took place between 2011 and 2013, TB incidence at the time of the study was >1,000/100,000. All consent documents were provided to potential participants in English or Xhosa and read through with a member of the study team to ensure full understanding and capacity to consent before signing. Participants recruited to undergo PET/CT were recruited in a 2-stage process. At initial screening consent they were provided with an information leaflet in English or Xhosa explaining the study procedures and the risks and benefits of study involvement. At the end of the screening period this leaflet was explained in detail with eligible participants and any queries or concerns addressed prior to final consenting for study entry.

The 35 asymptomatic participants undergoing FDG-PET/CT had blood sampling for RNA (Tempus, Thermo Fisher Scientific, Waltham, MA) and serum on the day of scan prior to receiving any IPT. The 15 HIV-1 infected controls with symptomatic, microbiologically confirmed pulmonary tuberculosis that were age, sex and CD4 matched to the 35 asymptomatic participants undergoing FDG-PET/CT and had blood sampling for RNA and serum within 24 hours of commencing TB treatment but did not undergo FDG-PET/CT.

Additional participants were recruited as controls for modular analysis. All were resident in Khayelitsha. All participants were of African ancestry (almost entirely of Xhosa origin). Symptomatic HIV-1 infected and HIV-1 uninfected participants with active TB had their diagnosis confirmed by either a positive sputum culture for Mtb or a positive sputum GeneXpert TB-RIF (Cepheid, Sunnyvale, CA), sputum smear status was also established. Blood sampling of active TB controls was carried out before, or within 24 hours, of treatment commencing.

HIV-1 infected participants on ART were established on medication for > 6 months and had a suppressed viral load (<40 copies/mL or lower than detectable limit). Participants were diagnosed HIV-1 infected if they had documented evidence of a positive point of care (POC) test for HIV-1 in their medical notes and, either a positive HIV-1 viral load and/or a positive confirmatory HIV-1 ELISA. HIV-1 uninfected participants had a documented negative POC test for HIV-1. HIV-1 uninfected adults with no evidence of active TB were the main control group for comparison.

RNA extraction

Three mL of blood was drawn into Tempus Blood RNA tubes and shaken vigorously for 10-15 seconds. Tempus tubes were stored within 6 hours of sampling in a freezer between 20^{o}C and -80^{o}C for no longer than 21 months before RNA extraction. RNA was extracted using the PerfectPure™ Blood RNA Kit (5 PRIME) according to maufacturer's instructions. RNA integrity was assessed using the Bioanalyser Nano Assay (Agilent, Santa Clara, CA) with all samples having RNA integrity number (RIN) 7-9.5. RNA yield was assessed using a Nanodrop ND 1000 spectrophotometer (Thermoscientific, Waltham, MA) and RNA yield was suboptimal for a single sample (active TB control) that was not analysed.

Two thousand nanograms of the isolated total RNA was globin reduced using the Human GLOBINclear™ Kit (Ambion, Austin, TX) according to manufacturer's instructions. 200ng of globin-reduced RNA was then used to prepare and amplify biotinylated, antisense complementary RNA (cRNA) with the TotalPrep RNA Amplification Kit (Illumina, San Diego, CA). 750ng of labelled cRNA was then hybridized overnight to Illumina HumanHT-12 v 4.0 Expression BeadChips containing 47,231 probes. Samples were randomly distributed across

chips along with control samples to minimize chip related bias. The arrays were washed, blocked, stained and scanned on an Illumina HiScan, as per manufacturer's instructions. Signal intensity values were then generated on Genomestudio (Illumina).

Microarray data analysis

Raw background subtracted expression data was normalized using GeneSpring GX version 12.6 (Agilent). For each sample, probes were assigned a signal intensity detection p-value in comparison to background hybridization and probes were assigned a "present" flag using a 0.99 cut off. Raw signal values <1 were then set to a threshold of 1 and all values were $\log_2$ transformed. To minimize technical variation, each sample was then normalized using a 75^{th} percentile shift algorithm in which the $\log_2$ transformed intensity value corresponding to the 75^{th} percentile was subtracted from $\log_2$ transformed intensity value for each probe within a sample. Baseline transformation was then carried out to rescale intensity values to the median of all samples. Samples were visualized using principal component analysis (PCA) to confirm no outliers. Following this probes that were not "present" in at least 10% of samples were filtered out before further analysis.

Pathway analysis was conducted on gene lists using Ingenuity Pathway Analysis (IPA) software (Qiagen Bioinformatics, Hilden, Germany). The canonical pathways in IPA are well-characterized metabolic and cell signaling pathways based on the available scientific literature. The strength of association between gene lists and pathways was determined by Fisher's exact test. Where hierarchical clustering was used to visualize data, Pearson uncentered (cosine) distance metric and average linkage rule was used.

Modular analysis of the transcripts was undertaken. Modules are groups of co-regulated transcripts subsequently categorized into functional groups through unbiased literature review. The principles informing this approach have been previously published (17) as have the specific details of the modules used in this study along with complete transcript lists (36). Extensive functional analysis of these modules using a variety of analytic approaches are available through the following link http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules). The average transcript abundance for each module was determined for each participant. Of the 260 modules, 38 had functional roles determined and analysis was restricted to these modules. HIV-1 uninfected adults with no evidence of active TB were used as the primary control group. The *Simes-Benjamini-Hochberg* method was used to control for false discovery rate (FDR) in analysis of modules

The raw and normalized microarray data has been deposited in GEO (GSE69581). All data collected and analyzed in the experiments adhere to the Minimal Information About a Microarray Experiment (MIAME) guidelines.

RNA sequence data analysis from Zak et al

Sequencing data was downloaded from NCBI GEO repository with project number PRJNA315611. Metadata was provided in GEO repository and supplementary data of associated publication(16). 93 samples were available from progressors. Per protocol time to TB was used to categorise each sample as being taken <180 days, 181-360 days, 361-540 days, 541-720 days, >720 days prior to diagnosis or being taken after diagnosis. Samples taken after diagnosis or more than 720 days before diagnosis were not included in analysis leaving 72 samples from

progressors across four timepoints. In order to allow for determination of differentially expressed genes at each of the four timepoints, four groups of non-progressors were assigned in a 1:1 ratio ensuring matching by time from recruitment (**see Supp Excel sheet**). All sequence data were quality checked with FastQC (version 0.11.5) before being aligned to the human genome (NCBI GRCh38.p10 build) using STAR (version 2.5.2b) with default parameters. Gene counting of the position-sorted aligned reads was done using HTSeq-count (version 0.6.1p1) with the default union mode. All genes with <1 reads on average across all samples were filtered out from expression analysis. Read count normalisation and differential gene expression analysis was done using DESeq2 in R (37) and the fold change and false discovery rate corrected significance (adjusted p-value) for each gene between progressors and non-progressors are reported.

RT-PCR arrays

RNA extracted from whole blood used for the microarrays was reverse transcribed using the RT^2 First Strand Kit (Qiagen, Hilden, Germany) which includes a genomic DNA (gDNA) elimination step. cDNA was analysed using 384-well RT^2 Profiler Custom PCR Arrays (Qiagen, Hilden, Germany) on the Roche 480 platform. The arrays included 4 house keeping (HK) genes (*HPRT1*, *B2M*, *RPLP0*, *HSP90AB1*), three control reactions to monitor gDNA contamination and reverse transcription efficiency, and the genes of interest (including *C1QB*, *FCGBP*, *FCGR1*, *SERPING1*). Delta Ct (ΔCt) was calculated according to the average Ct of the 4 HK genes. Delta delta Ct (ΔΔCt) was calculated as the sample ΔCT-Median ΔCT of all samples, and fold change was calculated as 2^(-ΔΔCt). Log2 normalised fold change values (mean=0, variance=1) were plotted as box plots in Qlucore Omics explorer.

643 *Serum analysis*

644 Blood was drawn into 5mL SST^{TM} II Advance tubes (BD Diagnostics, Franklin Lakes, NJ),
645 transported to the laboratory and centrifuged at 1200g for 10 minutes. The serum was aliquoted
646 and stored at -80^{o}C. Two of 15 active TB participants had no serum available. All 35
647 participants that underwent FDG-PET/CT had serum available. Analytes were measured by
648 ELISA: SerpinG1 (Cusabio, Wuhan, China), Circulating Immune complex (Bühlmann, Bremen,
649 Germany), C1q (Abcam, Cambridge, UK), C5 (assay measures uncleaved C5 and C5a) (Abcam,
650 Cambridge, UK). ELISA plates were read on a Bio-Rad iMark microplate reader (Bio-Rad,
651 Hercules, CA, USA) with standard curves generated for each analyte. Values that were below
652 manufacturer determined minimal detectable limit or the minimum asymptote of the standard
653 curve were assigned a zero value.
654

655 *Statistical Analysis*

656 Apart from microarray data where analysis was conducted primarily in GeneSpring GX version
657 12.6 (Agilent) and by Ingenuity Pathway Analysis software (Qiagen) as described above,
658 statistical analysis of continuous and categorical variables, correlation analysis and data
659 visualization was conducted in Stata ver. 12.1 (StataCorp) and Prism ver. 7.0c (GraphPad
660 software). The normality of data was assessed by the Shapiro-Wilk test and variance compared
661 by F-test or Bartlett's test. Non-parametric data was compared using the Mann-Whitney *U* test
662 or the Kruskal Wallis test and parametric data compared using t-test or ANOVA. Two-stage
663 linear step-up procedure of Benjamini, Krieger and Yekutieli was to control for false discovery
664 rate following Kruskal Wallis test if needed. Correlation was performed by Spearman rho or

Pearson's test depending on distribution of data. Proportions were compared by χ^2 test or Fisher's exact test (if the contingency included a number ≤ 5).

Author Contributions

R.J.W., H.E., C.E.B. and A.O'G. designed the study; H.E., R.P.L. and C.M.G. processed whole blood for microarray experiments; H.E., R.P.L, S.H. and M.L. analyzed transcriptomic data, with advice and input from C.M.G., A.K.C. and A.O'G.; H.E. and K.A.W. performed serum studies; A.K.C performed and analyzed RT-PCR; M.L. conducted additional statically analyses; R.J.W. supervised data analysis; H.E. and R.J.W wrote the manuscript and subsequently all authors provided advice and approved the final manuscript.

Acknowledgements

We thank all participants, clinic and laboratory staff for invaluable assistance. We also thank Douglas B Young of the Francis Crick Institute for early discussion around study design and analysis. This work was funded by the Wellcome Trust (090170, 203135, 104803), the Bill and Melinda Gates Foundation/Wellcome Trust Grand challenges in Global Health (37822), the intramural research program of NIH/NIAID and the National Institutes of Health (R01 HL106804). RJW is supported by the Francis Crick Institute that receives its core funding from Cancer Research UK (FC00110218), the UK Medical Research Council (FC00110218), and the Wellcome Trust (FC00110218). R.J.W. also received support from the European Union (FP7 HEALTH F3-2012-305578), National Research Foundation of South Africa (96841) and Medical Research Council of South Africa (SHIP-02-2013). A.O'G is also supported by the Francis Crick Institute. A.K.C is supported by the Medical Research Council of South Africa

687 (SHIP-02-2013) and the National Institutes of Health TB Research Unit 1U19AI111276 (TBRU-
688 BURU).

689

690 **References**

691 1. Esmail H, Barry CE, 3rd, Young DB, Wilkinson RJ (2014) The ongoing challenge of
692 latent tuberculosis. *Philos Trans R Soc Lond B Biol Sci* 369(1645):20130437.

693 2. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM*, et al.* (2016)
694 Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-D-
695 glucose positron emission and computed tomography. *Nat Med* 22(10):1090-3.

696 3. Loveday M, Ramjee A, Osburn G, Master I, Kabera G*, et al.* (2017) Drug-resistant
697 tuberculosis in patients with minimal symptoms: favourable outcomes in the absence of
698 treatment. *Int J Tuberc Lung Dis* 21(5):556-63.

699 4. Lin PL, Ford CB, Coleman MT, Myers AJ, Gawande R*, et al.* (2014) Sterilization of
700 granulomas is common in active and latent tuberculosis despite within-host variability in
701 bacterial killing. *Nat Med* 20(1):75-9.

702 5. Hunter RL (2011) Pathology of post primary tuberculosis of the lung: An illustrated
703 critical review. *Tuberculosis* 91(6):497-509.

704 6. Steinbruck P, Dankova D, Edwards LB, Doster B, Livesay VT (1972) Tuberculosis risk
705 in persons with "fibrotic" x-ray lesions. *Bull Int Union Tuberc* 47:135-59.

706 7. Koppaka R, Bock N. How reliable is chest radiography? In: Frieden T, editor. Toman's
707 tuberculosis case detection, treatment, and monitoring: questions and answers. Geneva: World
708 Health Organization; 2004. p. 51-60.

709 8. Mamede M, Higashi T, Kitaichi M, Ishizu K, Ishimori T*, et al.* (2005) [18F]FDG uptake
710 and PCNA, Glut-1, and Hexokinase-II expressions in cancers and inflammatory lesions of the
711 lung. *Neoplasia* 7(4):369-79.

712 9. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K*, et al.* (2007) Candidate
713 biomarkers for discrimination between infection and disease caused by *Mycobacterium*
714 *tuberculosis*. *J Mol Med* 85(6):613-21.

715 10. Maertzdorf J, Repsilber D, Parida SK, Stanley K, Roberts T*, et al.* (2011) Human gene
716 expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun* 12(1):15-22.

717 11. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA*, et al.* (2010) An interferon-
718 inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*
719 466(7309):973-7.

720 12. Kaforou M, Wright VJ, Oni T, French N, Anderson ST*, et al.* (2013) Detection of
721 tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression
722 signatures: a case-control study. *PLoS Med* 10(10):e1001538.

723 13. Blankley S, Graham CM, Levin J, Turner J, Berry MP*, et al.* (2016) A 380-gene meta-
724 signature of active tuberculosis compared with healthy controls. *Eur Respir J* 47(6):1873-6.

14. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI*, et al.* (2016) The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. *PLoS ONE* 11(10):e0162220.

15. Walport MJ (2001) Complement. Second of two parts. *N Engl J Med* 344(15):1140-4.

16. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S*, et al.* (2016) A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 387:2312-22.

17. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C*, et al.* (2008) A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 29(1):150-64.

18. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS*, et al.* (2013) Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS ONE* 8(8):e70630.

19. Maertzdorf J, Weiner J, 3rd, Mollenkopf HJ, Bauer T, Prasse A*, et al.* (2012) Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci U S A* 109(20):7853-8.

20. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG*, et al.* (2013) Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 207(1):18-29.

743 21. Raja A, Ranganathan UD, Ramalingam B (2006) Clinical value of specific detection of
744 immune complex-bound antibodies in pulmonary tuberculosis. *Diagn Microbiol Infect Dis*
745 56(3):281-7.

746 22. Raja A, Narayanan PR, Mathew R, Prabhakar R (1995) Characterization of
747 mycobacterial antigens and antibodies in circulating immune complexes from pulmonary
748 tuberculosis. *J Lab Clin Med* 125(5):581-7.

749 23. Lu LL, Chung AW, Rosebrock TR, Ghebremichael M, Yu WH*, et al.* (2016) A
750 Functional Role for Antibodies in Tuberculosis. *Cell* 167(2):433-43.

751 24. Kozakiewicz L, Phuah J, Flynn J, Chan J (2013) The role of B cells and humoral
752 immunity in Mycobacterium tuberculosis infection. *Adv Exp Med Biol* 783:225-50.

753 25. Nimmerjahn F, Ravetch JV (2008) Fcgamma receptors as regulators of immune
754 responses. *Nat Rev Immunol* 8(1):34-47.

755 26. Welsh KJ, Risin SA, Actor JK, Hunter RL (2011) Immunopathology of postprimary
756 tuberculosis: increased T-regulatory cells and DEC-205-positive foamy macrophages in cavitary
757 lesions. *Clin Dev Immunol* 2011:307631.

758 27. Hunter RL, Actor JK, Hwang SA, Karev V, Jagannath C (2014) Pathogenesis of post
759 primary tuberculosis: immunity and hypersensitivity in the development of cavities. *Ann Clin*
760 *Lab Sci* 44(4):365-87.

761 28. Hunter RL (2016) Tuberculosis as a three-act play: A new paradigm for the pathogenesis
762 of pulmonary tuberculosis. *Tuberculosis* 97:8-17.

763 29. Raja A, Baughman RP, Daniel TM (1988) The Detection by Immunoassay of Antibody
764 to Mycobacterial Antigens and Mycobacterial Antigens in Bronchoalveolar Lavage Fluid from
765 Patients with Tuberculosis and Control Subjects. *Chest* 94(1):133-7.

766 30. Kobayashi K, Ogata H, Morikawa M, Iijima S, Harada N*, et al.* (2002) Distribution and
767 partial characterisation of IgG Fc binding protein in various mucin producing cells and body
768 fluids. *Gut* 51(2):169-76.

769 31. Larsen GL, Mitchell BC, Henson PM (1981) The pulmonary response of C5 sufficient
770 and deficient mice to immune complexes. *Am Rev Respir Dis* 123(4 Pt 1):434-9.

771 32. Schifferli JA, Ng YC, Peters DK (1986) The role of complement and its receptor in the
772 elimination of immune complexes. *N Engl J Med* 315(8):488-95.

773 33. Schifferli JA, Peters DK (1983) Complement, the immune-complex lattice, and the
774 pathophysiology of complement-deficiency syndromes. *Lancet* 2(8356):957-9.

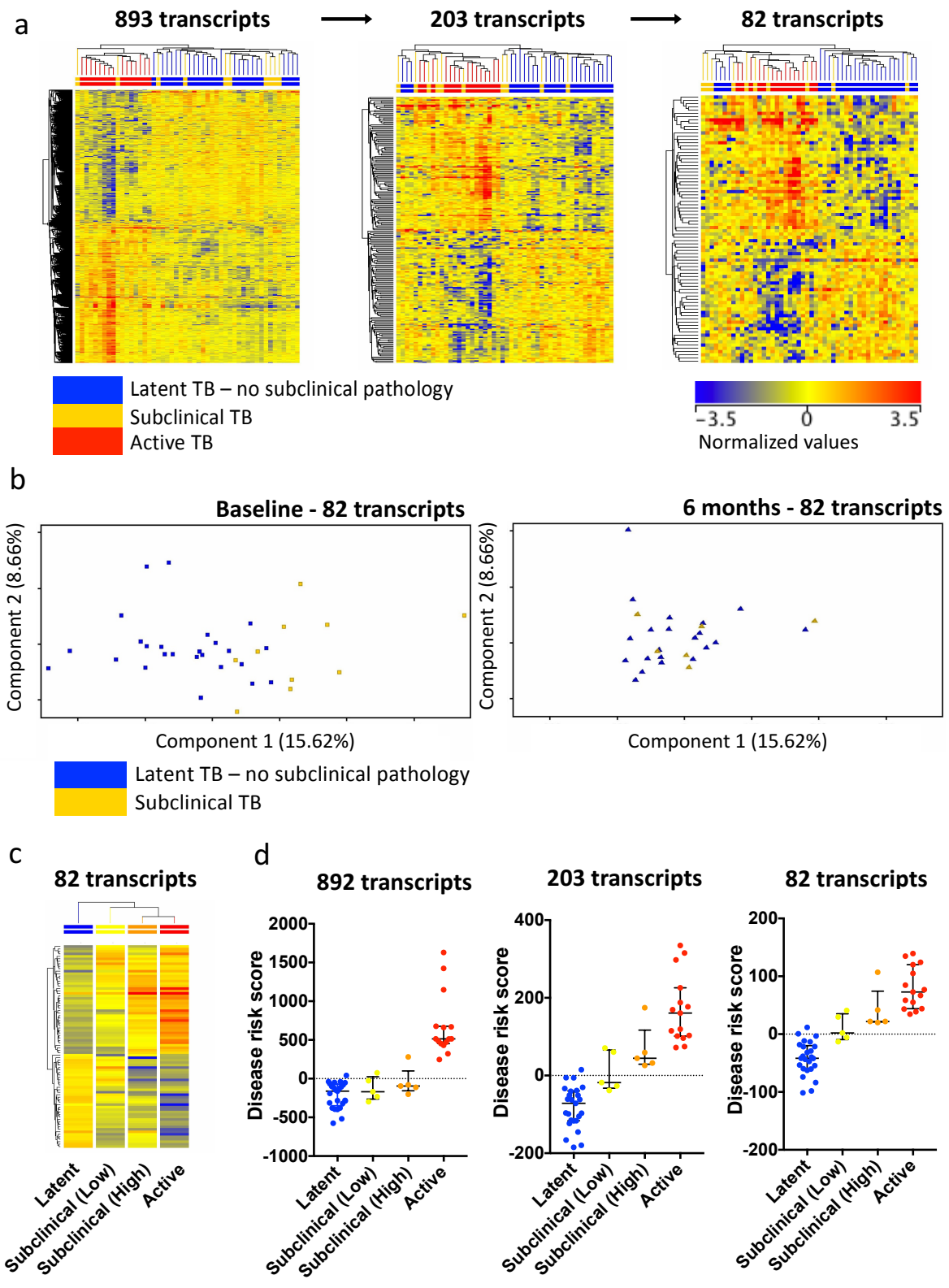
775 34. Cai Y, Yang Q, Tang Y, Zhang M, Liu H*, et al.* (2014) Increased complement C1q level
776 marks active disease in human tuberculosis. *PLoS ONE* 9(3):e92340.

777 35. Morgan BP, Gasque P (1997) Extrahepatic complement biosynthesis: where, when and
778 why? *Clin Exp Immunol* 107(1):1-7.

779 36. Obermoser G, Presnell S, Domico K, Xu H, Wang Y*, et al.* (2013) Systems scale
780 interactive exploration reveals quantitative and qualitative differences in response to influenza
781 and pneumococcal vaccines. *Immunity* 38(4):831-44.

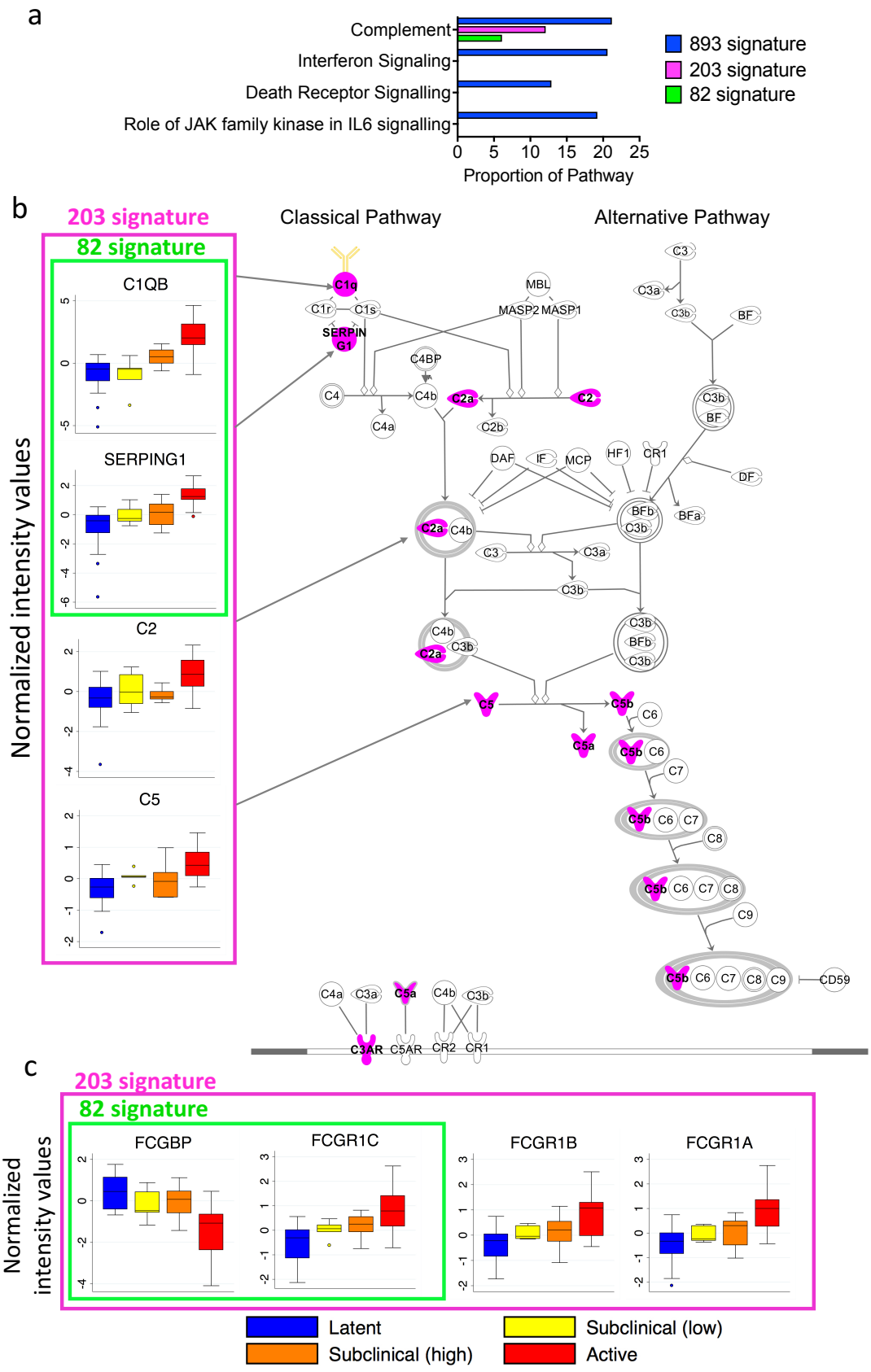
37. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

Figure 1


a
893 transcripts
203 transcripts
82 transcripts
Latent TB – no subclinical pathology
Subclinical TB
Active TB
−3.5
0
3.5
Normalized values
b
Baseline - 82 transcripts
6 months - 82 transcripts
Component 2 (8.66%)
Component 1 (15.62%)
Latent TB – no subclinical pathology
Subclinical TB
c
82 transcripts
Latent
Subclinical (Low)
Subclinical (High)
Active
d
892 transcripts
203 transcripts
82 transcripts
Disease risk score
2000
1500
1000
500
0
-500
-1000
400
200
-200
100
-100

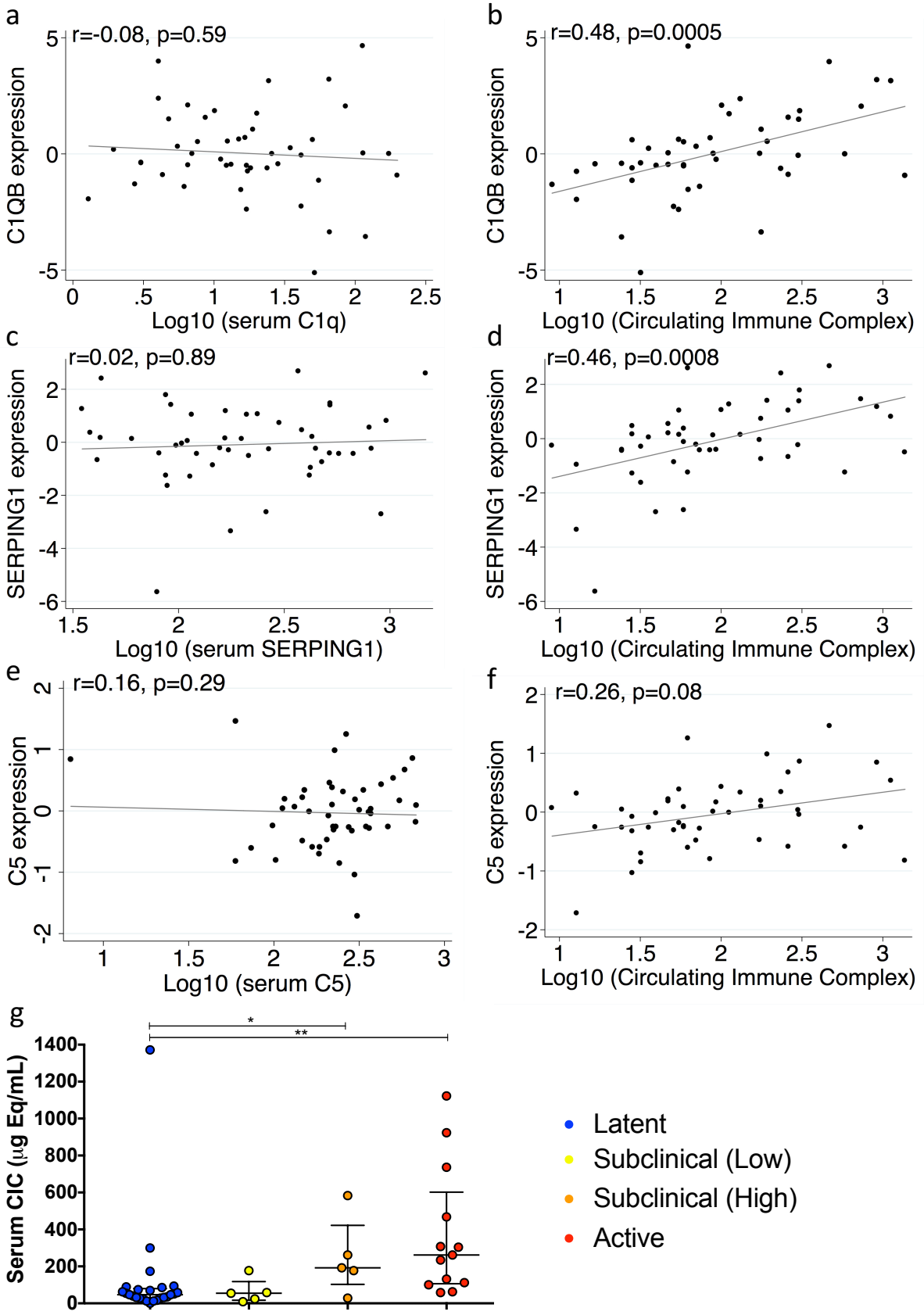
831

Figure 2


a
Complement
Interferon Signaling
Death Receptor Signalling
Role of JAK family kinase in IL6 signalling
0 5 10 15 20 25
Proportion of Pathway
893 signature
203 signature
82 signature
b
203 signature
82 signature
Classical Pathway
Alternative Pathway
C1QB
SERPING1
C2
C5
Normalized intensity values
C1q
C1r
C1s
SERPIN G1
MBL
MASP2 MASP1
C4BP
C4
C4b
C2a
C2
C4a
C2b
C3
C3a
C3b
BF
Bb
Ba
DAF
IF
MCP
HF1
CR1
DF
BFb
BFa
C3b
C5
C5b
C6
C5a
C7
C8
C9
CD59
C3AR
C5AR
CR2
CR1
c
203 signature
82 signature
FCGBP
FCGR1C
FCGR1B
FCGR1A
Normalized intensity values
Latent
Subclinical (high)
Subclinical (low)
Active

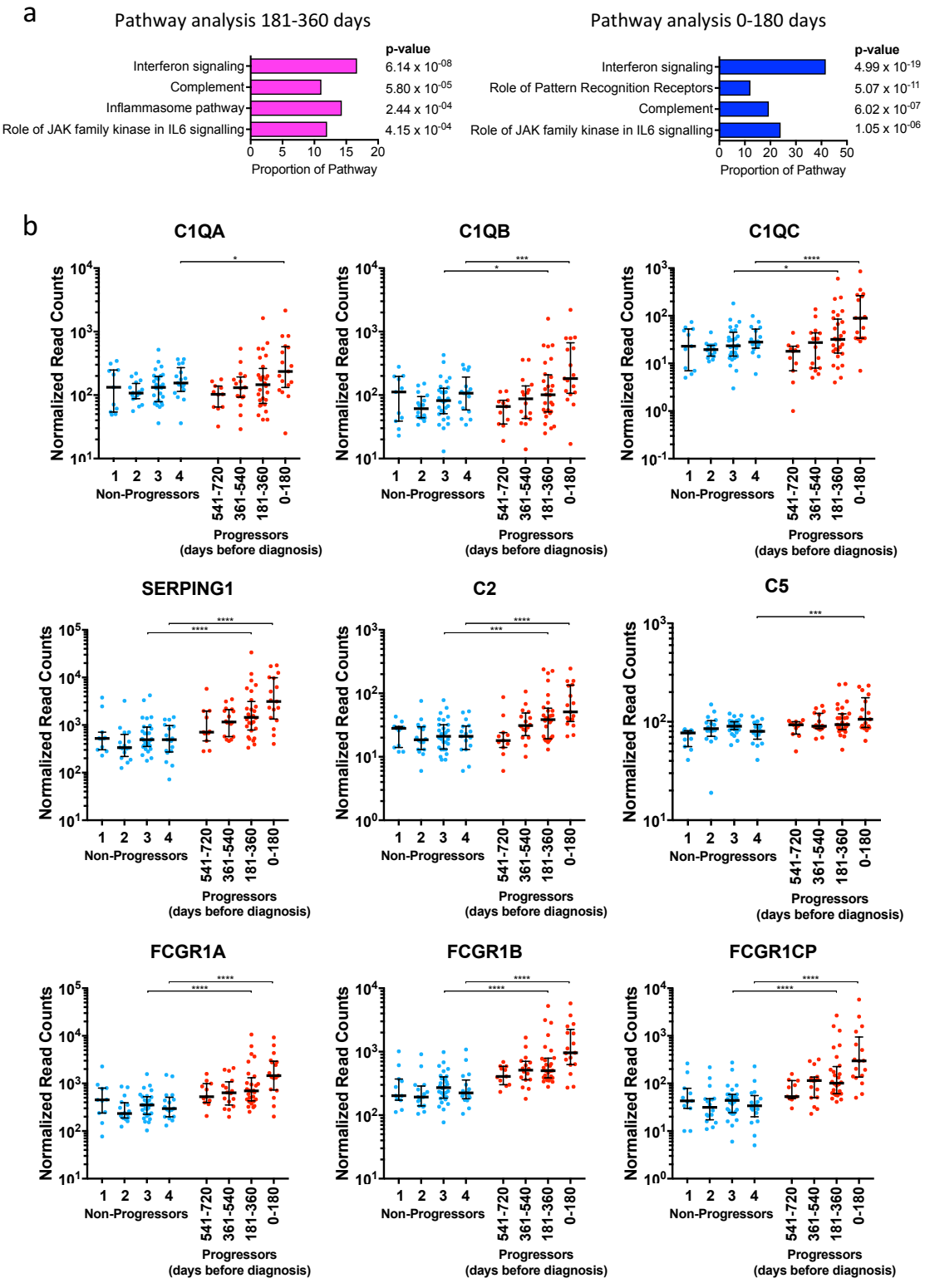
832

Figure 3


a
r=-0.08, p=0.59
C1QB expression
Log10 (serum C1q)
b
r=0.48, p=0.0005
C1QB expression
Log10 (Circulating Immune Complex)
c
r=0.02, p=0.89
SERPING1 expression
Log10 (serum SERPING1)
d
r=0.46, p=0.0008
SERPING1 expression
Log10 (Circulating Immune Complex)
e
r=0.16, p=0.29
C5 expression
Log10 (serum C5)
f
r=0.26, p=0.08
C5 expression
Log10 (Circulating Immune Complex)
g
Serum CIC (µg Eq/mL)
*
**
Latent
Subclinical (Low)
Subclinical (High)
Active

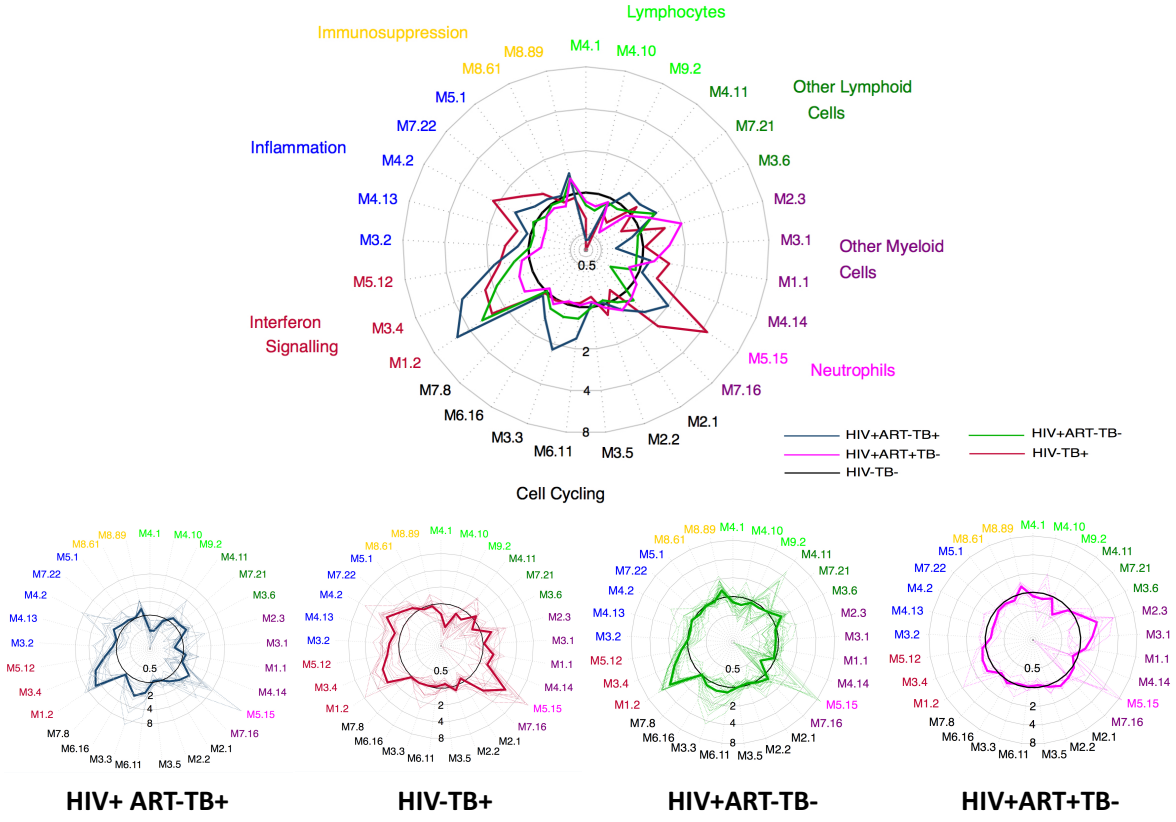
833

Figure 4


a
Pathway analysis 181-360 days
p-value
Interferon signaling
Complement
Inflammasome pathway
Role of JAK family kinase in IL6 signalling
6.14 x 10-08
5.80 x 10-05
2.44 x 10-04
4.15 x 10-04
Proportion of Pathway
Pathway analysis 0-180 days
p-value
Interferon signaling
Role of Pattern Recognition Receptors
Complement
Role of JAK family kinase in IL6 signalling
4.99 x 10-19
5.07 x 10-11
6.02 x 10-07
1.05 x 10-06
Proportion of Pathway
b
C1QA
C1QB
C1QC
SERPING1
C2
C5
FCGR1A
FCGR1B
FCGR1CP
Normalized Read Counts
Non-Progressors
Progressors
(days before diagnosis)
541-720
361-540
181-360
0-180

834

Figure 5


Lymphocytes
Other Lymphoid Cells
Other Myeloid Cells
Neutrophils
Cell Cycling
Interferon Signalling
Inflammation
Immunosuppression
HIV+ ART-TB+
HIV-TB+
HIV+ART-TB-
HIV+ART+TB-

836 **Figure Legends**

837

838 **Figure 1**

839 **Statistical and fold change filtering reveals 82 transcripts that separate active and**

840 **subclinical TB from latent TB**

841 **a**. Heatmaps showing hierarchical clustering (Pearson Uncentered (cosine) distance metric and

842 average linkage rule) for the 893, 203 and 82 transcript signatures. Active TB (n=15, red),

843 Subclinical TB (n=10, orange), Latent TB (n=25, blue), showing increasing clustering of

844 subclinical TB with active TB and increasing separation of subclinical TB from latent TB (with

845 no subclinical pathology).

846

847 **b**. 2D-PCA plot showing effect of treatment on 82 transcripts. Left panel shows pre-treatment,

848 with separation seen between subclinical TB (n=10) represented as orange squares and latent TB

849 (with no evidence of subclinical pathology) (n=25) as blue squares. In the right panel post

850 treatment (for 27 of 35 participants) those with previous subclinical TB (n=6) (orange triangle)

851 cluster with latent TB (n=21) (blue triangles).

852

853 **c**. Subclinical TB cases with higher metabolic activity on FDG-PET/CT clustered more closely

854 with active TB. Participants with subclinical TB (n=10) were subclassified into those with low

855 intensity FDG uptake within lung parenchymal or mediastinal lymph nodes (Visual Score 0-1 –

856 FDG uptake less than or equal to mediastinal blood pool, Yellow – n=5) or high intensity uptake

857 (Visual Score 3 – FDG uptake greater than mediastinal blood pool, Orange – n=5). Participants

858 with active TB (n=15) are shown in Red and with latent TB (n=25) in Blue. Heatmap (c)

represents average expression within the 4 groups for each of the 82-transcripts and shows hierarchical clustering (Pearson Uncentered (cosine) distance metric and average linkage rule). Participants with subclinical TB and high intensity FDG uptake were more closely related to Active TB.

d. A Disease risk score was determined for each participant using the 893, 203 and 82 transcripts by subtracting the summed normalized expression values of under-abundant transcripts from that of over-abundant transcripts (see (12)). A dot plot was used to visualize differences between the groups

Figure 2

Transcripts relating to the classical complement pathway are overabundant in subclinical and active TB.

a. Top 4 enriched canonical pathways listed by significance ($p<0.05$, Fisher's exact) for the 893-transcript signature of active TB shown. The bar chart shows the proportion of these pathways represented in the 893, 203 and 82-transcript signatures.

b. Schematic of the complement pathway, created using Ingenuity Pathway Analysis software, showing components of the pathway with overabundant transcripts in active (n=15) and subclinical TB (n=10) compared to latent TB (n=25) (in pink). Box and whisker plots show normalized expression of complement components SERPING1 and C1QB (part of 82-transcript signature) and C2 and C5 (part of 203- transcript signature) by TB disease state.

882 **c**. Box and whisker plots show normalized expression of immunoglobulin related transcripts
883 FCGBP and FCGR1C (part of 82-transcript signature) and FCGR1A and FCGR1B (part of 203-
884 transcript signature) by TB disease state.
885
886 **Figure 3**
887 **Correlation of transcript abundance with serum concentration of complement components**
888 **and Circulating Immune Complex**
889 **a-f**. Scatter plots showing correlation of transcript abundance against serum concentration of
890 protein product with Spearman rho and p-value for; (a) C1QB transcript abundance vs C1q
891 serum concentration, (b) C1QB transcript abundance vs serum circulating immune complex
892 (CIC) concentration, (c) SERPING1 transcript abundance vs. SERPING1 serum concentration,
893 (d) SERPING1 transcript abundance vs serum CIC concentration, (e) C5 transcript abundance vs
894 C5 serum concentration, (f) C5 transcript abundance vs serum CIC concentration. Correlation
895 was calculated for 48 participants with serum samples (13 active TB (as 2 of 15 participants did
896 not have serum available), 10 subclinical TB and 25 latent TB). For graphs b, d and f, a single
897 outlier with serum circulating immune complex concentration below detectable limit is not
898 shown.
899
900 **g**. Dot plot with median and IQR showing the CIC level for participants with latent, subclinical
901 (low activity), subclinical (high activity) and active TB. CIC was significantly increased across
902 different TB disease states (p=0.0004 - Kruskal Wallis), *post hoc* analysis controlling for false
903 discovery by Two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli
904 demonstrated that those with active TB (p=0.0001, q=0.0002) and those with subclinical TB and

high metabolic activity on PET/CT (p=0.038, q=0.04) had higher serum concentration of CIC than latent TB

Figure 4

Complement and Fc Gamma receptor genes significantly differentially expressed between progressors and non-progressors from Zak *et al*

Analysis of RNAseq data for HIV uninfected persons who eventually progressed to TB in comparison to matched non-progressors from Zak *et al*. Genes with <1 read on average across all samples were filtered out with expression analysis therefore conducted on 25,518 genes (see methods). 892 genes were significantly differentially abundant (p_{corr}<0.05) between progressors 0-180 days prior to diagnosis and matched non-progressors, of which 362 were also >1.5 fold overabundant. 610 genes were significantly differentially abundant (p_{corr}<0.05) between progressors 181-360 days prior to diagnosis and matched non-progressors, of which 154 were also >1.5 fold overabundant. IPA pathway analysis was conducted on these gene lists.

a. Top 4 enriched canonical pathways listed by significance (p<0.05, Fisher's exact) for the 362 genes and 154 genes differentially abundant at 0-180 and 181-360 days prior to diagnosis. The bar chart shows the proportion of these pathways represented.

b. Dot plots showing gene abundance (as normalized read counts) in progressors in 180 day blocks prior to disease presentation in comparison to 1:1 matched non-progressors (labeled 1-4 see methods and Supp Excel sheet)). All components of the complement pathway or Fc gamma receptor within the group of significantly differentially abundant at either <180 days or 181-360

days before diagnosis are shown. Significance shown by asterisk; p_{corr} <0.05(*), <0.01(**), <0.001(***), <0.0001(****) (see Supp Excel sheet for exact values)

Figure 5

Comparison of modular transcript abundance by TB and HIV status

a – Radar plots depicting fold change in transcript abundance in comparison to median modular transcript abundance in control group (HIV- TB-) for 29 modules. Plot on left shows fold change of median modular transcript abundance for HIV+ART-TB- (green), HIV- TB+ (red), HIV+ART- TB+ (blue) and HIV+ART+TB- (pink) in comparison to HIV-TB- controls (black) overlaid on same graph Plots on right show fold change for each individual participant in the four groups in comparison to control (HIV- TB-) with the median value in bold. Log2 scale.. See also Supp table 3.