

## **Systematic evaluation of transcriptomic disease risk and diagnostic biomarker overlap between COVID-19 and tuberculosis: a patient-level meta-analysis**

Dylan Sheerin<sup>1\*</sup>, Abhimanyu<sup>2\*</sup>, Xutao Wang<sup>3</sup>, W Evan Johnson<sup>3</sup>, Anna Coussens<sup>1,2,4</sup>

<sup>1</sup> Infectious Diseases and Immune Defence Division, The Walter & Eliza Hall Institute of Medical Research, Parkville 3279, VIC, Australia.

<sup>2</sup> Wellcome Centre for Infectious Diseases in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Anzio Rd, Observatory, 7925, Western Cape, South Africa.

<sup>3</sup> Division of Computational Biomedicine and Bioinformatics Program, Boston University, Boston, MA, USA; Department of Biostatistics, Boston University, Boston, MA, USA.

<sup>4</sup> Department of Medical Biology, University of Melbourne, Parkville 3010, VIC, Australia.

\* These authors contributed equally to the work.

## **Abstract**

### **Background**

The novel coronavirus, SARS-CoV-2, has increased the burden on healthcare systems already strained by a high incidence of tuberculosis (TB) as co-infection and dual presentation are occurring in syndemic settings. We aimed to understand the interaction between these diseases by profiling COVID-19 gene expression signatures on RNA-sequencing data from TB-infected individuals.

### **Methods**

We performed a systematic review and patient-level meta-analysis by querying PubMed and pre-print servers to derive eligible COVID-19 gene expression signatures from human whole blood (WB), PBMCs or BALF studies. A WB influenza dataset served as a control respiratory disease signature. Three large TB RNA-seq datasets, comprising multiple cohorts from the UK and Africa and consisting of TB patients across the disease spectrum, were chosen to profile these signatures. Putative “COVID-19 risk scores” were generated for each sample in the TB datasets using the TBSignatureProfiler package. Risk was stratified by time to TB diagnosis in progressors and contacts of pulmonary and extra-pulmonary TB. An integrative analysis between TB and COVID-19 single-cell RNA-seq data was performed and a population-level meta-analysis was conducted to identify shared gene ontologies between the diseases and their relative enrichment in COVID-19 disease severity states.

### **Results**

35 COVID-19 gene signatures from nine eligible studies comprising 98 samples were profiled on TB RNA-seq data from 1181 samples from 853 individuals. 25 signatures had significantly higher COVID-19 risk in active TB (ATB) compared with latent TB infection ( $p < 0.005$ ), 13 of which were validated in two independent datasets. *FCNI*- and *SPPI*-expressing macrophages enriched in BALF during severe COVID-19 were identified in circulation during ATB. Shared perturbed ontologies included antigen presentation, epigenetic regulation, platelet activation, and ROS/RNS production were enriched with increasing COVID-19 severity. Finally, we demonstrate that the overlapping transcriptional responses may complicate development of blood-based diagnostic signatures of co-infection.

### **Interpretation**

Our results identify shared dysregulation of immune responses in COVID-19 and TB as a dual risk posed by co-infection to COVID-19 severity and TB disease progression. These individuals should be followed up for TB in the months subsequent to SARS-CoV-2 diagnosis.

## Background

Three months after it was first detected, SARS-CoV-2 was declared to have caused the first global pandemic of the 21st Century. In the absence of an effective treatment or vaccine, mortality in the 35 million diagnosed thus far, currently stands at 2.9% (as of October 2020). By comparison tuberculosis (TB), a similar respiratory infection and humanity's longest continuing pandemic, causes 10 million annual cases, and has a mortality of 12–20%, the upper bound including those HIV-coinfected<sup>1</sup>. This high mortality, of roughly 4000 people a day, exists despite a vaccine that reduces infant mortality and antibiotics which have reduced mortality from the pre-antibiotic era by 50%. With an estimated quarter of the world's population infected with *Mycobacterium tuberculosis* (*Mtb*), it could be said that TB is a silent killer. Silent in that during the current pandemic, TB has killed roughly the same number of people a day as COVID-19 currently; with the socio-economic and health systems impact of COVID-19 lockdowns estimated to result in an additional 6.8 million TB cases and 1.4 million TB deaths over the next 5 years<sup>2,3</sup>. Countries where TB-HIV already causes high mortality were shielded from previous SARS and MERS outbreaks, and thus the interaction of these coronaviruses with concurrent TB co-infection, has not previously been experienced. COVID-19 and TB share a symptomatic presentation of productive cough, fever and shortness of breath, and clinical parameters of raised C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), D-Dimer, interleukin (IL)-6, leukopenia and neutrophilia. The similarity in clinical parameters and aspects of underlying immunological reactions suggests co-infection will not only complicate diagnostic algorithms, it indicates a potentially fatal convergence in immunopathogenesis. Emerging case studies and population level data indicate TB patients, and those latently infected or with a history of TB are at increased risk of severe COVID-19<sup>4-7</sup>. In order to develop contextually appropriate treatment and risk mitigation interventions in communities where the potential for *Mtb*-SARS-CoV-2 co-infection is high, we urgently need to understand how the immunopathogenesis of these two respiratory pathogens interact. We therefore conducted a systematic transcriptomic evaluation of whole blood (WB), peripheral blood mononuclear cell (PBMC) and bronchoalveolar lavage fluid (BALF) signatures associated with COVID-19 clinical severity and the spectrum of asymptomatic and symptomatic TB. Our findings suggest that subclinical and active TB (ATB) increase the risk of severe COVID-19 disease, due to increased abundance of circulating myeloid subpopulations also found in the lungs of severe COVID-19 patients. This shared pathway of immunopathogenesis also suggests that SARS-CoV-2 infection will trigger increased progression to TB disease. COVID-19 may therefore pose the biggest threat to ending the TB epidemic since HIV-1 and the modelling increase of 6.8 million extra TB cases in the next 5 years is significantly underestimated<sup>8</sup>.

## Methods

### Search strategy and selection criteria

The hypothesis that WB transcriptomic signatures present in those with existing TB infection will increase risk of severe COVID-19 and interfere in diagnostic biomarker selection was evaluated using a combination of transcriptomic data from COVID-19 patients and WB RNA-seq data from studies of TB disease progression on which the COVID-19 signatures were evaluated. A literature search of published and pre-print manuscripts was conducted on the NIH PubMed and bioRxiv, medRxiv, and SSRN servers uploaded/published between 01/02/20 and 20/09/20 (figure 1). The WB influenza dataset, GSE111368<sup>9</sup>, was used to generate an influenza virus control signature. The curatedTBData package,<sup>10</sup> which includes 48 publicly available TB RNA-seq datasets, was used to identify eligible TB datasets, that included individuals who progressed to TB during the duration of study follow-up, with RNA-seq data at baseline and time of diagnosis, and patient-level meta-data including time to TB progression. Selection criteria and cohort datasets are described in detail in appendix 1 p2 and p5 (table 1–2).

### Data Analysis

TB RNA-seq data were downloaded for eligible studies using the curatedTBData package,<sup>10</sup> as outlined in appendix 1 p2. Samples collected at all time points were included. The eligible COVID-19 and influenza control signatures (appendix 2 p1) were evaluated independently against the patient-level TB RNA-seq data, generating individual-sample putative “COVID-19 risk score” using gene set variation analysis (GSVA)<sup>11</sup> with the TBSignatureProfiler package<sup>12</sup> (appendix 1 p2), and score significance, as compared to latent TB infection (LTBI) controls, calculated by Bonferroni-corrected t-test (appendix 2 p2–4).

Single cell (sc)RNA-seq integrative comparison was conducted using the identified COVID-19 bronchoalveolar lavage fluid (BALF) dataset<sup>13</sup> downloaded from the NIH GEO database (GSE145926) and a TB PBMC dataset<sup>14</sup> downloaded from NCBI Short Read Archive (SRA, SRR11038989-SRR11038995). *filtered\_feature\_bc\_matrix.h5* files were read into RStudio using the Seurat *Read10X\_h5* function (v3.0).<sup>15</sup> Datasets were independently normalised before the Seurat *FindIntegrationAnchors* function was applied to generate a corrected data matrix for joint analysis. Functional enrichment analysis and cohort selection is described in detail in appendix 1 (p3). Briefly, selected datasets were evaluated using the Metascape<sup>16</sup> statistically enriched ontology terms, generating enriched pathways (appendix 1 p3–4). Pathway-associated gene clusters from the functional enrichment analysis were used to generate a protein-protein interaction (PPI) network and visualised with Cytoscape<sup>17</sup> (v4.01) using ranked gene set enrichment analysis (GSEA)<sup>18</sup>.

## Results

### COVID-19 severity signatures are enriched during subclinical and active tuberculosis

Nine COVID-19 studies with 35 signatures were eligible for evaluation (figure 1, table 1). Three TB cohort studies encompassing 853 individuals and 1181 samples, at various time points, were eligible for comparison. COVID-19 signatures were first evaluated in the TB combined observational and prospective cohort (TCC), from the UK (London and Leicester) and South Africa, including 293 individuals<sup>19</sup>. Out of the 35 COVID-19 signatures profiled, 25 were significantly associated ( $p < 0.005$ ) with higher COVID-19 risk scores in TB progressors and ATB patients, compared with latently infected individuals. Conversely, the influenza signature centred on a score of zero, across the spectrum of TB infection in both countries (figure 2). Of the ten COVID-19 signatures that were not associated with a significantly higher COVID-19 risk score, across the TB spectrum, seven were enriched in those with mild COVID-19 in the original studies compared to severe COVID-19 patients. Of the 20 scRNA-seq immune cell population signatures profiled; innate immune cell signatures generated the highest COVID-19 risk scores in the ATB groups. The classical monocyte signature<sup>20</sup> was drastically increased in ICU cases of COVID-19, yielding the most significantly increased risk scores between LTBI and ATB samples ( $p < 0.0001$ ). The Silvin *et al.* COVID-19 WB neutrophil signature<sup>21</sup> was most significantly increased in ATB vs LTBI ( $p < 0.0001$ ). The lung macrophage sub-populations associated with severe disease in BALF from COVID-19 patients<sup>13</sup> showed high risk scores; in particular, the *FCN1*<sup>hi</sup> (monocyte-derived macrophages (MDMs), G1), *FCN1*<sup>lo</sup>*SPP1*<sup>hi</sup> (pre-fibrotic macrophages, G2), and intermediary G1/2 macrophages. Conversely, the majority of adaptive immune cell signatures (CD4 and CD8 T cells) were higher in LTBI and TB contacts who didn't progress to TB (figure 2). In general, T cell populations are depleted during severe COVID-19 infection<sup>13,20,22</sup> and a lower COVID-19 risk scores for these signatures in the active and progressive TB patients could reflect a similar depletion of these populations. In the Wei *et al.* study of patients recovering from COVID-19<sup>20</sup>, the monocyte signature was associated with the early recovery stage indicating persistence of hyperinflammatory response, whereas the NK cell, T cell and B cell signatures were enriched in the late recovery stage, explaining the low COVID-19 risk scores associated with these signatures in ATB cases.

The mild/moderate and severe COVID-19 WB signatures<sup>23</sup> (figure 2, Hadjadj *et al.*) showed significant differences in COVID-19 risk scores between ATB and LTBI ( $p < 0.00001$  and  $p < 0.005$ , respectively) while the critical signature generated a low COVID-19 risk score which was significantly higher in the Leicester ATB cases than the LTBI cases ( $p < 0.0001$ ). COVID-19 disease severity blood transcriptional module (BTM) signatures<sup>24</sup> included genes involved in IFN responses and antigen processing and presentation and clearly separated active disease samples from LTBI and QFT-negative TB contact samples ( $P < 0.0001$ ).

Assessing COVID-19 risk scores after further stratifying the Leicester TB contacts and ATB index cases by the location of index case TB (pulmonary [PTB] or extrapulmonary [EPTB]) revealed higher COVID-19 risk scores in contacts of PTB and PTB patients than those of EPTB, respectively (appendix 1 p6).

### **COVID-19 signatures separate latent tuberculosis cases from those who progress to active disease by COVID-19 risk score**

Thirteen signatures having significantly higher COVID-19 risk score (with an associated adjusted  $p < 0.005$ ) in at least two ATB groups, compared with LTBI, were selected for validation in the additional two TB progressor RNA-seq datasets: the ACS<sup>25</sup> and GC6 cohorts including 153 and 407 individuals, respectively.<sup>26</sup> Both datasets exhibited the same trend of increased COVID-19 risk score for all 13 signatures in progressors compared with LTBI ( $p \leq 0.01$ ), whilst scoring zero for the influenza signature. The COVID-19 IFN signatures<sup>23,27,28</sup> were associated with the greatest difference in COVID-19 risk score between LTBI and progressor groups,  $p < 0.001$  for both datasets). Plotting by days to TB diagnosis revealed an additional trend of higher risk of severe COVID-19 associated with proximity to ATB disease in the ACS cohort (figure 3A), but not the GC6 cohort (figure 3B).

### **High concordance between monocyte subpopulations identified in BALF of severe COVID-19 patients and those in circulation during active TB disease**

Reported similarities between BALF and WB scRNA-seq expression profiles observed in COVID-19 patients led us to investigate whether the BALF macrophage sub-lineages<sup>13</sup> that associated with COVID-19 risk in ATB patients (figure 2) could be detected in circulation during TB infection. An integrated scRNA-seq analysis was performed using publicly available TB scRNA-seq PBMC data, consisting of both active and LTBI samples with BALF scRNA-seq data<sup>13</sup>, from COVID-19 patients of varying disease severity (appendix 1 p7). A high concordance was observed between the immune cell populations present within ATB PBMC and severe COVID-19 BALF after tSNE dimensionality reduction (figure 4A, left panel). Canonical cell type marker genes were identified for each cluster and assigned to the tSNE plot to identify shared and unique subpopulations (figure 4A, right panel). Three major macrophage sub-lineage markers identified in the original BALF scRNA-seq analysis<sup>13</sup>, *FCNI*, *SPP1* and *FABP4*, were separately profiled on the macrophage clusters of the COVID-19 and TB samples. Of the three markers, *FCNI* had highest expression in the TB PBMCs, while *FABP4* was completely absent. The *FCNI*-expressing pro-inflammatory monocyte-derived macrophage population was the most abundant of the sub-lineages in severe COVID-19 patients<sup>13</sup>. Zooming in on the *FCNI*- and *SPP1*-expressing clusters, the additional inflammatory markers that were identified for these populations in the original analysis were profiled. Both ATB and severe COVID-19 samples

had high expression of these markers (figure 4C) indicating that these immune cell sub-lineages are active in the inflammatory response elicited by both diseases in the lungs and the blood.

### **Highly similar enriched ontologies are shared between COVID-19 and tuberculosis**

A meta-pathway enrichment analysis was performed using the transcriptomic data from the Leicester TB<sup>19</sup>, the COVID-19 WB scRNA-seq data<sup>21</sup>, and the influenza viral control cohort<sup>9</sup>. Among the top 1000 DEGs (selection explained in appendix 1 p3,8), DEGs from COVID-19, ATB and TB progressors were enriched for similar pathways compared to LTBI and influenza (figure 5A, appendix 1 p8, appendix 2 p5). Among the top 100 pathways across all groups, the highest percentage was associated with COVID-19 (96% of those considered), followed by ATB (93%), progressors (88%), influenza (64%), and LTBI (28%) (figure 5B). Comparing the twenty most significant ontologies, influenza and COVID-19 patients could be distinguished on the basis of an absence of IFN- $\gamma$  response, lack of TNF signalling, both of which were highly enriched in TB progressors and COVID-19 (figure 5A). Interestingly, LTBI had no enrichment of cytokine production and regulation of innate immune response compared with the other disease states, indicating a less active immunological response (figure 5A).

To identify some key effectors responsible for the activation of these pathways, ten significant protein-protein interaction (PPI) network clusters were identified and are represented as annotated networks in figure 5C. Among the 85 DEGs which were represented in the network analysis, only nine belonged to influenza, while there was greater alignment of PPI networks between COVID-19 and the TB groups, further confirming the similarity between COVID-19 and TB. Antigen processing and presentation was the largest cluster and was enriched mostly in COVID-19 and TB groups, with only one member gene representing influenza. The second most enriched cluster shared between COVID-19 and TB, annotated as epigenetic regulation of gene expression, consisted of many histone related proteins, suggesting both result in widespread changes in the epigenome. Other enriched clusters included platelet activation, cytokine production, and proteasome were exclusively enriched in COVID-19 and TB, indicating extensive commonality between perturbed pathways (figure 5C, p-values in appendix 2 p6).

### **Genes clustered based on similarity between COVID-19 and tuberculosis associate with COVID-19 disease severity**

Three clusters, based on similarity between the other diseases and COVID-19, were identified (figure 6A). Cluster 1: common to all TB disease states, but absent from influenza, consists of IFN- $\gamma$  response, apoptosis signalling and B cell activation pathway enrichment. Cluster 2: common across all disease groups, except for LTBI, consists of complement activation, inflammatory response, apoptotic signalling, T cell receptor signalling, and IFN- $\gamma$  production. Cluster 3: shared between ATB and TB progressors, but absent from LTBI and influenza, was enriched for platelet degranulation,



autophagy, antigen processing and presentation, TNFA signalling, IL6/JAK/STAT3 signalling, and cellular processes indicating higher metabolic requirements of cells in response to infection. Regulation of mTORC1 signalling was the only pathway exclusively enriched in COVID-19 from the top 100 enriched pathways (figure 6A). To further explore if the enriched pathway DEGs were associated with severity of COVID-19, GSEA was used to determine enrichment of the three identified clusters in a WB bulk RNA-seq dataset containing samples from different clinical stages of COVID-19<sup>21</sup>. Cluster 1, was found to be enriched in all three COVID-19 disease severity categories with moderate effect size, while cluster 2 and cluster 3 were enriched in more severe disease states with larger effect sizes as indicated by the *q*-values in figure 6B. DEGs common to COVID-19 and TB states were significantly enriched with increasing COVID-19 severity; 17% were enriched in moderate COVID-19, 18% were enriched in severe COVID-19 and the greatest proportion, 32%, were enriched in the ICU group (appendix 1 p9).

The overlap between DEGs associated with COVID-19 disease severity and TB raised the possibility of an impediment to the discovery of a specific COVID-19 diagnostic WB biomarker.

TBSignatureProfiler was run on cluster-specific DEGs using TCC data. In all three clusters, groups ranging from TB progressors to ATB showed significant COVID-19 risk scores compared with the LTBI control group (figure 6C). Cluster 3, which contains the most shared DEGs between COVID-19, ATB and TB progressors, exhibited the most graded significant increase in COVID-19 risk scores among the three clusters, suggesting that shared DEGs between COVID-19 and TB would interfere with the derivation of a COVID-19-specific diagnostic signature (figure 6C, appendix 2 p9).

To determine whether a more specific COVID-19 diagnostic signature could be identified what would not be affected by TB co-infection, DEGs from the only COVID-19-specific pathway, mTORC1 signalling (figure 6C), were used to compare the COVID-19 risk scores between the Leicester TB cohort and the COVID-19 disease severity data. The mTORC1 signature demonstrated a clear increase in COVID-19 risk score with increasing COVID-19 disease severity (figure 6D); lower COVID-19 risk scores were generated from the ATB group, with the LTBI contacts and the healthy controls having similar scores. Using the full list of 820 exclusive pathways enriched in COVID-19 (appendix 2 p10) a 20-gene signature was derived by selecting genes from the top 10 significant pathways and excluding DEGs shared between any other disease group, enriched for significance using a ranked GSEA on the COVID-19 severity data (appendix 1 p10–11). This reduced signature yielded significant differences in risk scores between healthy controls and severe COVID-19 ( $p < 0.005$ ) and ICU patients ( $p < 0.0005$ ), and not TB groups (figure 6D, appendix 2 p13).



## Discussion

Emerging findings to aid rapid responses to the COVID-19 pandemic has led to an increased deposition of manuscripts on pre-print servers. Whilst they require rigorous peer-review, they have served as a valuable resource for further hypothesis generation and testing. We used several such studies to augment investigation of our hypothesis using these sequencing data in combination with new bioinformatics packages developed for assessing TB diagnostic signature performance.<sup>10,12</sup> Substituting TB signatures for COVID-19 signatures, within the TBSignatureProfiler package, we demonstrate the overlap between severe COVID-19 transcriptional signatures and WB TB disease phenotypes found in ATB patients and importantly those with asymptomatic TB who progress to symptomatic TB over 1–2 years.

The most compelling overlaps between COVID-19 signatures and the TB datasets reside in the circulating innate immune cells. Classical monocyte and neutrophil signatures derived from severe COVID-19 patients were associated with the highest COVID-19 risk scores when profiled across the TB spectrum. Circulating monocyte activation status is a determining factor for COVID-19 prognosis, with specific phenotypes leading to poorer outcomes<sup>29</sup>. Similar monocyte phenotypes have been detected during TB infection<sup>30</sup> and the presence of these in circulation prior to SARS-CoV-2 co-infection may be detrimental to the activation of key adaptive antiviral immune responses.

Conversely, adaptive immune cell populations enriched in milder COVID-19 cases were associated with lower risk scores in ATB and progressors. Both T and B cells were significantly reduced in COVID-19 ICU patient samples compared to pre and post ICU<sup>20</sup>. Impairment of functional CD4<sup>+</sup> and CD8<sup>+</sup> T cells has been associated with severe COVID-19<sup>31</sup> and may be exacerbated by the presence of exhausted T cell phenotypes indicative of chronic *Mtb* infection which display reduced capacity to produce effector cytokines<sup>32</sup>.

SARS-CoV-2-infected macrophages are known to accumulate in lungs of patients who died from COVID-19<sup>33</sup>. Activated macrophages play a major role in chronic inflammation in TB and HIV and increase the likelihood of severe COVID-19 infection (odds ratios: 1.7 and 2.3, respectively)<sup>34</sup>.

*FCNI*<sup>hi</sup> macrophages are abundant in the BALF of severe COVID-19 patients<sup>13</sup> and in BALF indicating their migration to the lungs from the blood, substantiated by an integrative analysis between COVID-19 blood and BALF scRNA-seq samples performed by Silvin *et al.*<sup>21</sup>. We show that both the *FCNI*<sup>hi</sup> and *FCNI*<sup>lo</sup>*SPPI*<sup>+</sup> sub-lineages associated with significantly higher COVID-19 risk scores in recent contacts of TB patients, progressors, and ATB patients. Our integrative analysis of TB PBMC and COVID-19 BALF scRNA-seq samples showed the presence of these phenotypes in circulation during ATB, indicating that the presence of these sub-lineages in the blood may predispose TB patients to more severe lung inflammation.

The importance of neutrophils has largely been underestimated in existing scRNA-seq workflows, due to technical issues associated with the preservation and sequencing of neutrophils, despite

evidence that neutrophils are correlated with poor prognosis and disease severity in COVID-19<sup>21,35,36</sup>, and out-of-control inflammation or a failure to establish adequate adaptive immune responses in severe TB<sup>37</sup>. The upregulation of calprotectin subunit genes *S100A8* and *S100A9* in these neutrophils may account for or trigger the cytokine storm that is characteristic of severe COVID-19<sup>21</sup> and *S100A8/A9*-dependent neutrophilic accumulation in the lungs of ATB patients leads to induction of inflammatory mediators and promotes lymphocyte trafficking<sup>38</sup>. These findings underscore the need for more WB single-cell studies rather than PBMCs in order to characterise the important contribution of neutrophils to severe COVID-19.

IFN-induced transcriptional signatures were among the most significantly upregulated in severe disease in the studies selected risk profiling<sup>20,24,27,28</sup> and generated high COVID-19 risk scores among ATB cases. Dysregulation of IFN production<sup>23,39</sup> and the nature of type I and III IFN responses (location, timing, and duration) in COVID-19<sup>40,41</sup> and TB<sup>42</sup> guide disease progression and outcomes. It is therefore plausible that dysregulation of type I IFN responses during SARS-CoV-2 co-infection may also have an impact on TB disease progression.

At a systems levels, shared biological pathways showed a graded enrichment of similar pathways between COVID-19 and TB, particularly ATB. A PPI-based network analysis identified common genes, highlighting that shared molecular determinants between COVID-19 and TB disease states can influence the clinical outcome if both diseases affect the same individual i.e., during co-infection. Common enriched pathways between COVID-19 and TB signify activation of innate immune responses directed against both *Mtb* and SARS-CoV-2 and include antigen presentation<sup>27,43</sup>, membrane trafficking<sup>44,45</sup>, ROS/RNS production<sup>46,47</sup>, activation of complement<sup>48,49</sup>, cytokine production<sup>50,51</sup>, and platelet activation<sup>52,53</sup>. This suggests that both hyperactivation of these responses or evasion by the pathogen may lead to severe clinical presentations for both of these infections and be synergistically exacerbated in co-infection. These mechanistic similarities would complicate efforts to derive a COVID-19-specific transcriptional signature. We demonstrated this with a 20-gene COVID-19-specific signature that gave rise to significantly higher scores among severe COVID-19 cases compared with healthy controls, but also produced slightly elevated scores in TB-infected individuals.

Although several of the pre-print manuscripts identified at the beginning of our study have since been published in high-impact journals, our COVID-19 signature search was complicated by the limited availability of comprehensive supplementary data detailing cluster markers or DEG lists.

Furthermore, the COVID-19 studies performed to date are confined to small numbers of patients in each group that exhibit inter-individual variability in immune cell phenotypes. We counteracted this limitation by including signatures from multiple studies for each immune cell population.

In summary, we show for the first time through large scale meta-analysis of the available transcriptomic data, that advanced COVID-19 and TB disease states overlap at the gene, cell, and systems levels. These shared disease mechanisms could prove to be hotspots for immune

exacerbation, inducing greater immunopathology, in the case of co-infection. We report a new 20-gene signature which distinguishes severe COVID-19 from active and LTBI that should be investigated further for its disease classification value in larger datasets as they become available. Taken together, the data presented here along with the emerging case reports identifying TB as a risk factor for severe COVID-19 suggesting that individuals with known previous TB history, recent TB exposures or LTBI with pre-existing lung pathology, are at increased risk of severe COVID-19 disease and, potentially, early progression to TB disease, following SARS-CoV-2 infection. Given the medical capacity to do so, we therefore propose that such individuals should 1) be closely followed to allow early detection of respiratory symptom onset, 2) be screened for SARS-CoV-2 and TB at symptom onset, and 3) be followed up for TB in the months subsequent to SARS-CoV-2 diagnosis. Diagnostic and clinical outcome data arising from early stages of the COVID-19 outbreak should be stratified by TB history in order to determine the effects of these co-infections as soon as possible.

## References

1. World Health Organization. Global tuberculosis report 2019, 2019.
2. Stop TB Partnership. We did a rapid assessment: THE TB RESPONSE IS HEAVILY IMPACTED BY THE COVID-19 PANDEMIC. 2020. [http://stoptb.org/news/stories/2020/ns20\\_014.html](http://stoptb.org/news/stories/2020/ns20_014.html) (accessed 5 October 2020).
3. Zumla A, Marais BJ, McHugh TD, et al. COVID-19 and tuberculosis-threats and opportunities. *Int J Tuberc Lung Dis* 2020; **24**(8): 757-60.
4. Boulle A, Davies MA, Hussey H, et al. Risk factors for COVID-19 death in a population cohort study from the Western Cape Province, South Africa. *Clin Infect Dis* 2020.
5. Motta I, Centis R, D'Ambrosio L, et al. Tuberculosis, COVID-19 and migrants: Preliminary analysis of deaths occurring in 69 patients from two cohorts. *Pulmonology* 2020; **26**(4): 233-40.
6. Tadolini M, Codecasa LR, García-García JM, et al. Active tuberculosis, sequelae and COVID-19 co-infection: first cohort of 49 cases. *Eur Respir J* 2020; **56**(1).
7. Liu Y, Bi L, Chen Y, et al. Active or latent tuberculosis increases susceptibility to COVID-19 and disease severity. *medRxiv* 2020: 2020.03.10.20033795.
8. Stop TB Partnership. THE POTENTIAL IMPACT OF THE COVID-19 RESPONSE ON TUBERCULOSIS IN HIGH-BURDEN COUNTRIES: A MODELLING ANALYSIS, 2020.
9. Dunning J, Blankley S, Hoang LT, et al. Progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza. *Nat Immunol* 2018; **19**(6): 625-35.
10. Wang X. curatedTBData: Curation of existing 42 tuberculosis transcriptomic studies. 2020.
11. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 2013; **14**(1): 7.
12. Jenkins D, Zhao Y, Johnson WE, Odom A, Love C. TBSignatureProfiler: Profile RA-Seq Data Using TB Pathway Signatures. 2020. <https://github.com/compbiomed/TBSignatureProfiler>.
13. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 2020; **26**(6): 842-4.
14. Cai Y, Dai Y, Wang Y, et al. Single-cell transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis. *EBioMedicine* 2020; **53**.
15. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell* 2019; **177**(7): 1888-902.e21.
16. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019; **10**(1): 1523.
17. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**(11): 2498-504.
18. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005; **102**(43): 15545.
19. Singhania A, Verma R, Graham CM, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat Commun* 2018; **9**(1): 2308.
20. Wei L, Ming S, Zou B, et al. Viral Invasion and Type I Interferon Response Characterize the Immunophenotypes during COVID-19 Infection. *SSRN* 2020.
21. Silvin A, Chapuis N, Dunsmore G, et al. Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. *Cell* 2020; **182**(6): 1401-18.e18.
22. Wen W, Su W, Tang H, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov* 2020; **6**: 31.
23. Hadjadj J, Yatim N, Barnabei L, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* 2020; **369**(6504): 718.
24. Arunachalam PS, Wimmers F, Mok CKP, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* 2020; **369**(6508): 1210-20.
25. Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016; **387**(10035): 2312-22.
26. Suliman S, Thompson EG, Sutherland J, et al. Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am J Respir Crit Care Med* 2018; **197**(9): 1198-208.

27. Wilk AJ, Rustagi A, Zhao NQ, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nature Medicine* 2020; **26**(7): 1070-6.
28. Huang L, Shi Y, Gong B, et al. Blood single cell immune profiling reveals the interferon-MAPK pathway mediated adaptive immune response for COVID-19. *medRxiv* 2020: 2020.03.15.20033472.
29. Zhang D, Guo R, Lei L, et al. COVID-19 infection induces readily detectable morphological and inflammation-related phenotypic changes in peripheral blood monocytes, the severity of which correlate with patient outcome. *medRxiv* 2020: 2020.03.24.20042655.
30. Lastrucci C, Bénard A, Balboa L, et al. Tuberculosis is associated with expansion of a motile, permissive and immunomodulatory CD16+ monocyte population via the IL-10/STAT3 axis. *Cell Research* 2015; **25**(12): 1333-51.
31. Zheng M, Gao Y, Wang G, et al. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol Immunol* 2020; **17**(5): 533-5.
32. Khan N, Vidyarthi A, Amir M, Mushtaq K, Agrewala JN. T-cell exhaustion in tuberculosis: pitfalls and prospects. *Crit Rev Microbiol* 2017; **43**(2): 133-41.
33. Yao XH, Li TY, He ZC, et al. [A pathological report of three COVID-19 cases by minimal invasive autopsies]. *Zhonghua Bing Li Xue Za Zhi* 2020; **49**(5): 411-7.
34. Martinez FO, Combes TW, Orsenigo F, Gordon S. Monocyte activation in systemic Covid-19 infection: Assay and rationale. *EBioMedicine* 2020; **59**.
35. Chen J, Fan H, Zhang L, et al. Retrospective Analysis of Clinical Features in 101 Death Cases with COVID-19. *medRxiv* 2020: 2020.03.09.20033068.
36. Barnes PF, Leedom JM, Chan LS, et al. Predictors of Short-Term Prognosis in Patients with Pulmonary Tuberculosis. *The Journal of Infectious Diseases* 1988; **158**(2): 366-71.
37. Lowe DM, Redford PS, Wilkinson RJ, O'Garra A, Martineau AR. Neutrophils in tuberculosis: friend or foe? *Trends Immunol* 2012; **33**(1): 14-25.
38. Gopal R, Monin L, Torres D, et al. S100A8/A9 proteins mediate neutrophilic inflammation and lung pathology during tuberculosis. *Am J Respir Crit Care Med* 2013; **188**(9): 1137-46.
39. Acharya D, Liu G, Gack MU. Dysregulation of type I interferon responses in COVID-19. *Nature Reviews Immunology* 2020; **20**(7): 397-8.
40. Major J, Crotta S, Llorian M, et al. Type I and III interferons disrupt lung epithelial repair during recovery from viral infection. *Science* 2020; **369**(6504): 712.
41. Broggi A, Ghosh S, Sposito B, et al. Type III interferons disrupt the lung epithelial barrier upon viral recognition. *Science* 2020; **369**(6504): 706.
42. Cliff JM, Kaufmann SHE, McShane H, van Helden P, O'Garra A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunological reviews* 2015; **264**(1): 88-102.
43. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology* 2017; **18**(1): 76.
44. Schnettger L, Rodgers A, Repnik U, et al. A Rab20-Dependent Membrane Trafficking Pathway Controls M. tuberculosis Replication by Regulating Phagosome Spaciousness and Integrity. *Cell Host & Microbe* 2017; **21**(5): 619-28.e5.
45. Bayati A, Kumar R, Francis V, McPherson PS. SARS-CoV-2 uses clathrin-mediated endocytosis to gain access into cells. *bioRxiv* 2020: 2020.07.13.201509.
46. Franchini AM, Hunt D, Melendez JA, Drake JR. FcγR-driven release of IL-6 by macrophages requires NOX2-dependent production of reactive oxygen species. *J Biol Chem* 2013; **288**(35): 25098-108.
47. Laforge M, Elbim C, Frère C, et al. Tissue damage from neutrophil-induced oxidative stress in COVID-19. *Nat Rev Immunol* 2020; **20**(9): 515-6.
48. Esmail H, Lai RP, Lesosky M, et al. Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* 2018; **115**(5): E964-E73.
49. Java A, Apicelli AJ, Liszewski MK, et al. The complement system in COVID-19: friend and foe? *JCI Insight* 2020; **5**(15).
50. Kaufmann SHE. Protection against tuberculosis: cytokines, T cells, and macrophages. *Ann Rheum Dis* 2002; **61** Suppl 2(Suppl 2): ii54-ii8.

51. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**(10223): 497-506.
52. Fox KA, Kirwan DE, Whittington AM, et al. Platelets Regulate Pulmonary Inflammation and Tissue Destruction in Tuberculosis. *Am J Respir Crit Care Med* 2018; **198**(2): 245-55.
53. Hottz ED, Azevedo-Quintanilha IG, Palhinha L, et al. Platelet activation and platelet-monocyte aggregate formation trigger tissue factor expression in patients with severe COVID-19. *Blood* 2020; **136**(11): 1330-41.



## **Declarations**

### **Acknowledgments**

The authors would like to thank colleagues at the Wellcome Centre for Infectious Diseases in Africa and the Walter and Eliza Hall Institute of Medical Research for their feedback throughout the study.

### **Competing interests**

The authors declare no competing interests.

### **Funding**

DS is supported by the Walter and Eliza Hall Institute of Medical Research. A is supported by an AXA Research Fund (Grant Ref: 25776), the National Research Foundation, South Africa (Free-standing Fellowship, UID: 8829), and SHIP-02-2013 granted to AC. XW is supported by the National Institutes of Health (R21AI154387). WEJ is funded by CRDF Global (DAA3-19-65672-1) and the National Institutes of Health (U19AI111276, U01CA220413, R01GM127430, R21AI154387). AC is supported by the Walter and Eliza Hall Institute of Medical Research, the Medical Research Council of South Africa (SHIP-02-2013), the National Institute of Health TB Research Unit (U19AI111276) and the NRF (UID109040).

### **Authors' contributions**

The study was conceived and designed by AC, DS, and A. DS and A contributed equally to the data analysis. AC supervised the analysis and assisted with the study design. Systematic review of COVID-19 manuscripts and identification of TB datasets was performed by DS and A. The curatedTBData package used to obtain the TB transcriptomic data and associated patient metadata was created by XW and WEJ. COVID-19 signature risk profiling and scRNA-seq analysis was performed by DS with input from WEJ, who developed the TBSignatureProfiler package with colleagues. Functional enrichment analysis and GSEA were performed by A. DS, A, and AC prepared the manuscript, with editorial input from all authors.

## Tables

**Table 1. Characteristics of the datasets used to profile COVID-19 risk from COVID-19 immune cell and pathway signatures.**

	Setting	Samples included	Study design	Population	Sampling	TB case definition	Follow-up duration and method
TB contacts cohort (TCC)	London, UK	54 (21 active TB, 33 LTBI non-progressors)	Cohort	HIV-negative adults (18–78 yrs.) TB cases and TB contacts	Baseline	TB: Culture confirmed or clinically diagnosed	1.9*
	Cape Town, South Africa	47 (16 active TB, 31 LTBI non-progressors I)				LTBI: QFT positive	
	Leicester, UK	313 (53 active TB, 23 progressors, 118 LTBI non-progressors, 119 healthy non-progressors)	Cohort	HIV-negative individuals (16–84 yrs.) TB cases and TB contacts	Baseline plus serial for a subset	TB: Culture-confirmed or Xpert MTB/RIF  LTBI: QFT positive	2 years, active
Adolescent cohort study (ACS)	Cape Town, South Africa	355 (110 progressors, 245 matched non-progressors)	Nested case-control	HIV-negative adolescents (12–18 yrs.) with latent TB infection	Serial (0, 6, 12, and 24 months)	TB: Intrathoracic disease with 2 positive smears or 1 positive culture  LTBI: QFT positive	2 years, active

**Table 2. Characteristics of studies used to derive signatures for risk profiling on tuberculosis datasets.**

	Tissue	Sequencing platform	Cohort	Severity	Eligible signatures
Wilk <i>et al.</i> <sup>26</sup>	PBMCs	Seq-Well	7 patients hospitalized with confirmed COVID-19 (all male, aged 20-80+ yrs.), 6 healthy controls	3 COVID-19 patients were on ventilation and diagnosed with ARDS; 4 were less severely ill patients. Samples collected 2-16 days following symptom onset	4
Huang <i>et al.</i> <sup>27</sup>	PBMCs	10X Chromium	8 active disease patients, 2 cured patients, 3 healthy controls, and non-COVID-19 patients	Active disease patients: 1 critical case, 1 severe case, 6 moderate cases. Non-COVID-19 patients: 1 case of influenza A, 1 case of acute pharyngitis, and 1 case of cerebral infarction	1
Wen <i>et al.</i> <sup>20</sup>	PBMCs	10X Chromium	10 recovering patients (5 male, 5 female, aged 40-70 yrs.), plus healthy controls	5 early-recovery stage (ERS) and 5 late-recovery stage (LRS) patients, classified by days between blood sampling date and negative qPCR	5
Liao <i>et al.</i> <sup>13</sup>	BALF	10X Chromium	6 COVID-19 patients (5 male, 1 female, median age: 49.5), 8 previously reported healthy lung controls	3 severe, 3 mild	6
Xiong <i>et al.</i> <sup>x</sup>	BALF and PBMCs	MGISEQ-2000	3 COVID-19 patients and 3 healthy controls	No severity information provided	2
Hadjadj <i>et al.</i> <sup>22</sup>	Whole blood	nanoString nCounter	50 COVID-19 patients with a spectrum of disease severity	15 mild/moderate, 17 severe, and 18 critical	5
Wei <i>et al.</i> <sup>21</sup>	PBMCs	10X Chromium	4 COVID-19 patients	Patients sampled before, during, and after ICU care	4

Silvin <i>et al.</i> <sup>28</sup>	Whole blood	10X Chromium	3 COVID-19 patients and 3 healthy controls	1 mild, 2 severe patients sampled at day 0 and day 10	3
Arunachalam <i>et al.</i> <sup>23</sup>	PBMCs	10X Chromium	7 COVID-19 patients and 5 healthy controls	No severity information provided for scRNA-seq samples	5
Dunning <i>et al.</i> <sup>9</sup>	Whole blood	Illumina GenomeStudio	131 influenza patients, 155 healthy controls	Not applicable	1

## Figure captions

**Figure 1. PRISMA flow chart of study selection.** PBMCs, peripheral blood mononuclear cells; BALF, bronchoalveolar lavage fluid.

**Figure 2. Profiling immune cell signatures from COVID-19 patients highlights increasing risk of severe disease associated with progression to active tuberculosis.** COVID-19 immune cell signatures were derived from bulk and single-cell RNA-sequencing (RNA-seq) studies and used to generate putative “COVID-19 risk scores” from a tuberculosis (TB) whole blood bulk RNA-seq dataset using the TBSignatureProfiler package. TB samples are grouped according to disease state and COVID-19 signatures were categorised by immune cell or signature type. Scores for each signature were compared by contrasting each group with the London Latent group using a t-test adjusted for multiple testing using a Bonferroni correction. All signatures were derived from peripheral blood mononuclear cells (PBMCs) unless otherwise stated in the boxplot title. WB, whole blood; M $\phi$ , macrophage; BALF, bronchoalveolar lavage fluid; ISG, interferon (IFN)-stimulated gene; NK, natural killer; ICU, intensive care unit; BTM, blood transcriptional module. \* $<0.05$ , \*\* $<0.005$ , \*\*\* $<0.0005$ , \*\*\*\* $<0.00005$ .

**Figure 3. Risk of developing severe COVID-19 is significantly elevated in patients that progress from latent to active tuberculosis disease.** COVID-19 immune cell signatures that were associated with significant differences in COVID-19 risk score between controls and progressor/active tuberculosis (TB) cases were validated on two additional whole blood RNA-sequencing TB datasets - (A) the Adolescent Cohort Study and (B) the Grand Challenges 6 study. TB samples were classified as latent or progressors and COVID-19 signatures were categorised by immune cell or signature type. Samples from patients that progressed to active TB disease during the study follow-up period are coloured and scaled according to time to TB diagnosis, measured in days and plotted on a log<sub>2</sub> scale. Scores for each signature were compared by contrasting progressor with latent cases using a t-test adjusted for multiple testing using a Bonferroni correction. All signatures were derived from whole blood unless otherwise stated in the boxplot title. \* $<0.05$ , \*\* $<0.005$ , \*\*\* $<0.0005$ , \*\*\*\* $<0.00005$ .

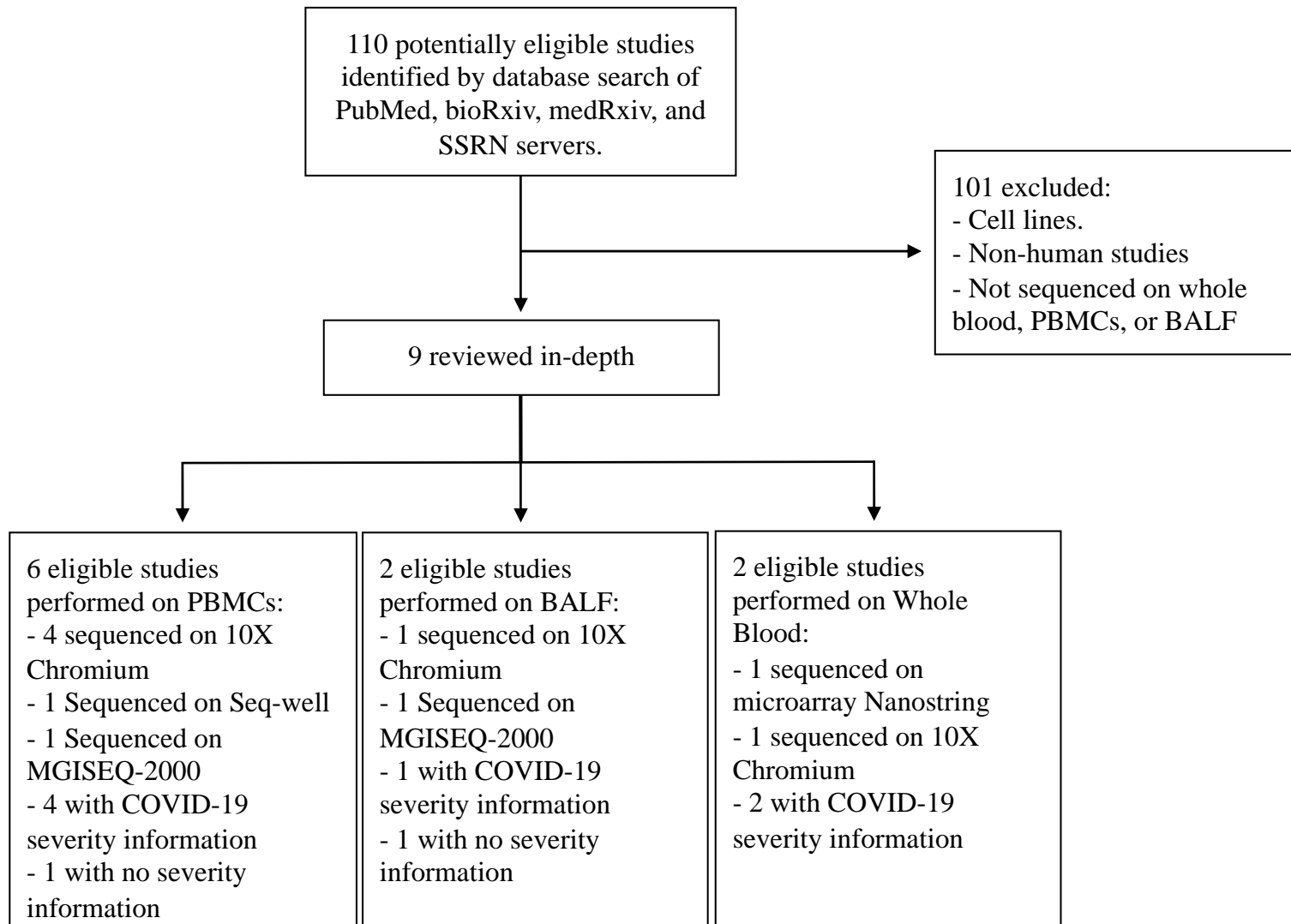
**Figure 4. Macrophage subpopulations upregulated in the lungs of severe COVID-19 patients can also be found in the circulation during active tuberculosis disease.** Single-cell RNA-sequencing (scRNA-seq) data from bronchoalveolar lavage fluid (BALF) of severe COVID-19 patients (n = 6) and peripheral blood mononuclear cells (PBMCs) from patients with active tuberculosis (TB) disease (n = 3) were integrated. (A) t-distributed stochastic neighbour embedding (tSNE) plot of integrated scRNA-seq data in the left panel, with cells from severe COVID-19 patients in guava and those from active TB patients in cyan, with corresponding cell clusters annotated based on identified markers in the right panel. (B) Macrophage clusters from the severe COVID-19 patients (left column) and active TB patients (right column), with the expression of major macrophage subpopulation markers identified in the original COVID-19 study - *FCN1*, high in G1, low in G2; *SPP1*, G2 and G3; *FABP4*, G4 - highlighted in red. (C) Violin plots depicting the expression levels of additional inflammatory marker genes associated with the macrophage subpopulations also present in the TB PBMC data, for each macrophage cluster.

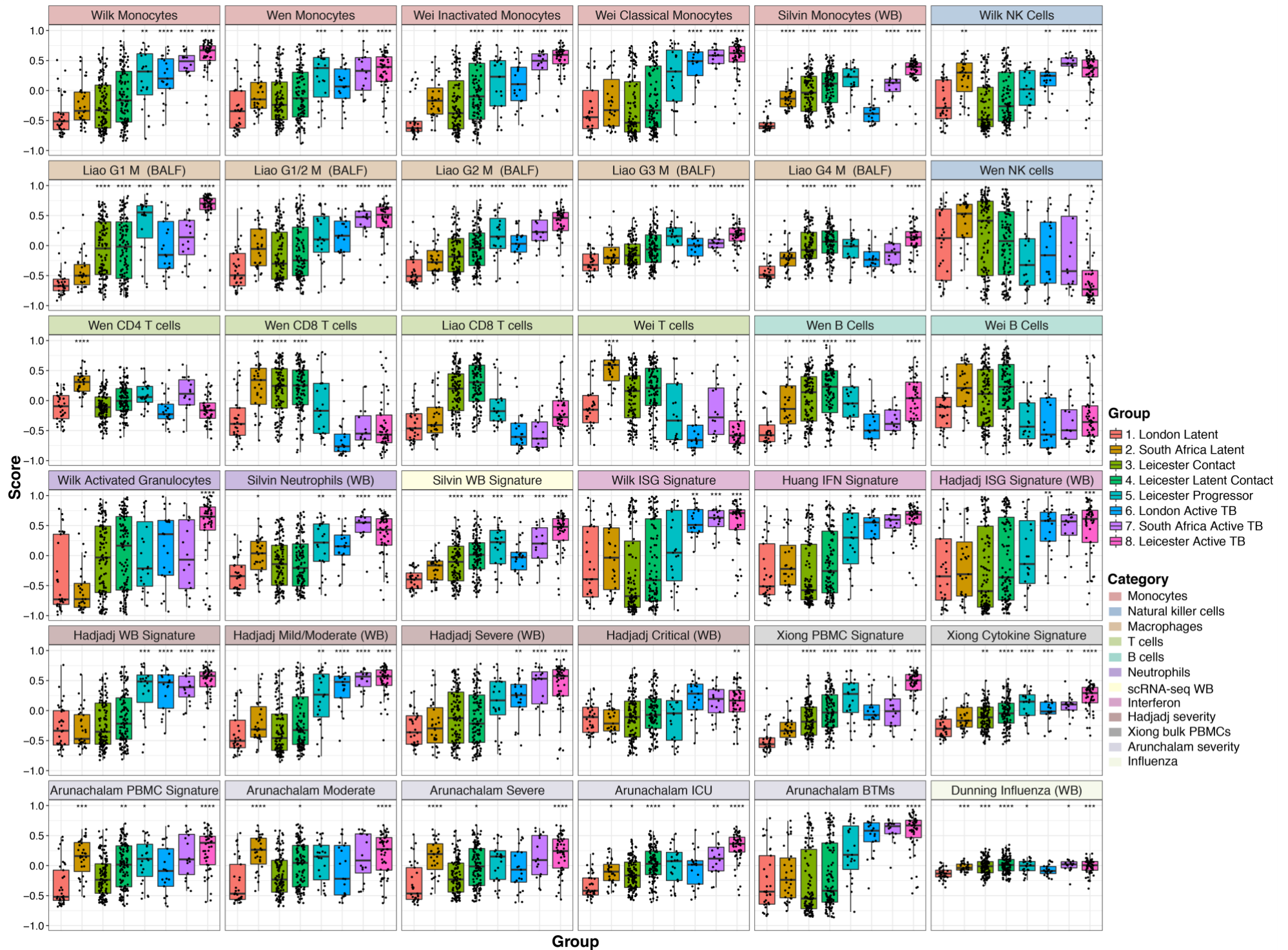
**Figure 5. Active TB disease and COVID-19 have similar perturbed pathways and gene networks.** (A) Heatmap of the top 20 enriched pathways from the meta-analysis across the groups with the best p-values. Cells are coloured by their  $-\log_{10}(p\text{-values})$ , grey cells indicate a lack of significant enrichment for that term/ontology in the corresponding DEGs list. (B) Upset intersection plot showing the number of pathways shared between COVID-19 and the other groups compared. Wherever the genes are shared the specific coloured dots appear below the column bar graph which show the number of shared pathways, which are connected by vertical lines, denoting shared categories. The horizontal bars represent the category gene count and the numbers on the horizontal axes represent the common pathways between the represented category and COVID-19. (C) All MCODE components from a protein-protein interaction (PPI) network analysis of all merged gene lists are displayed as networks, nodes are displayed as pie charts, colour coded by disease group. The labels are added manually and are derived from functional labels based on the top three enriched terms for that cluster.

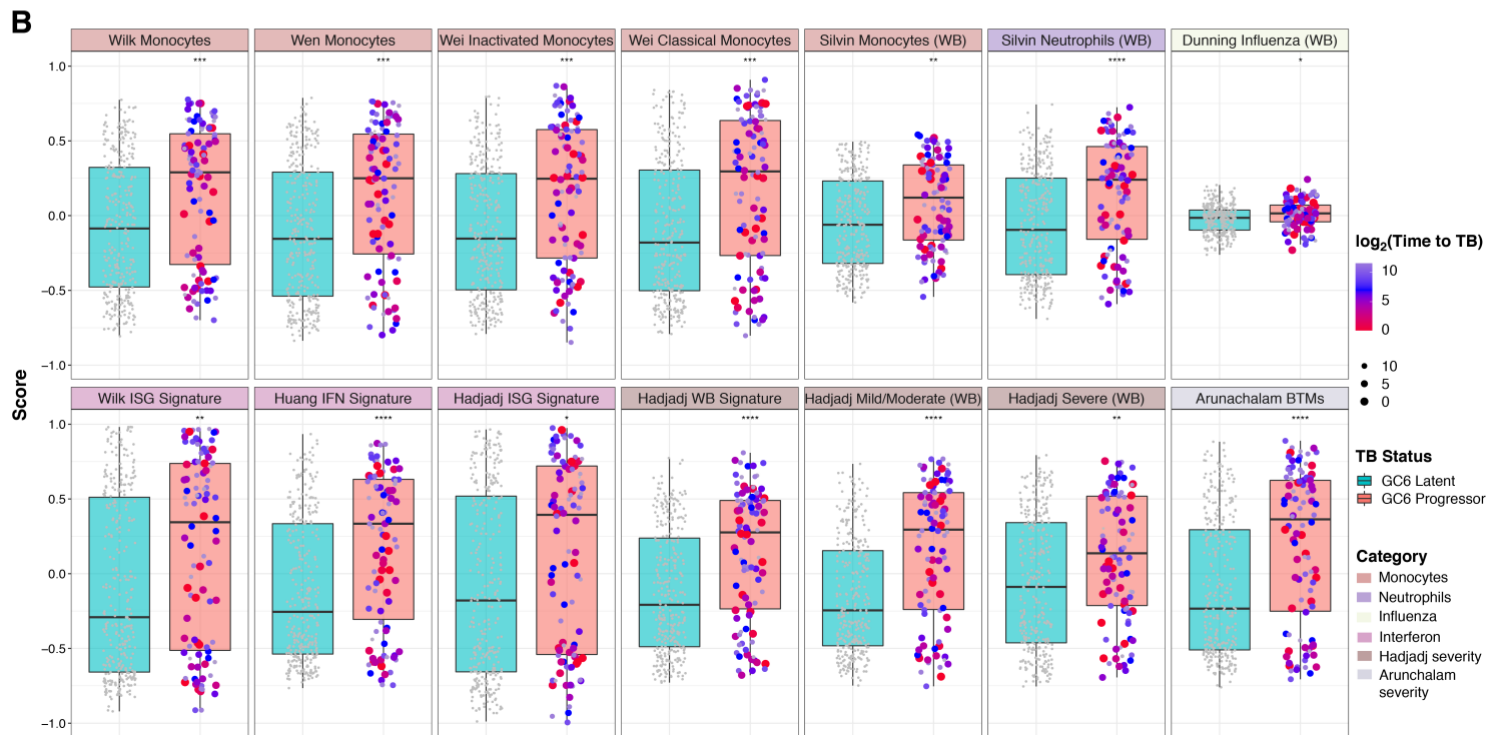
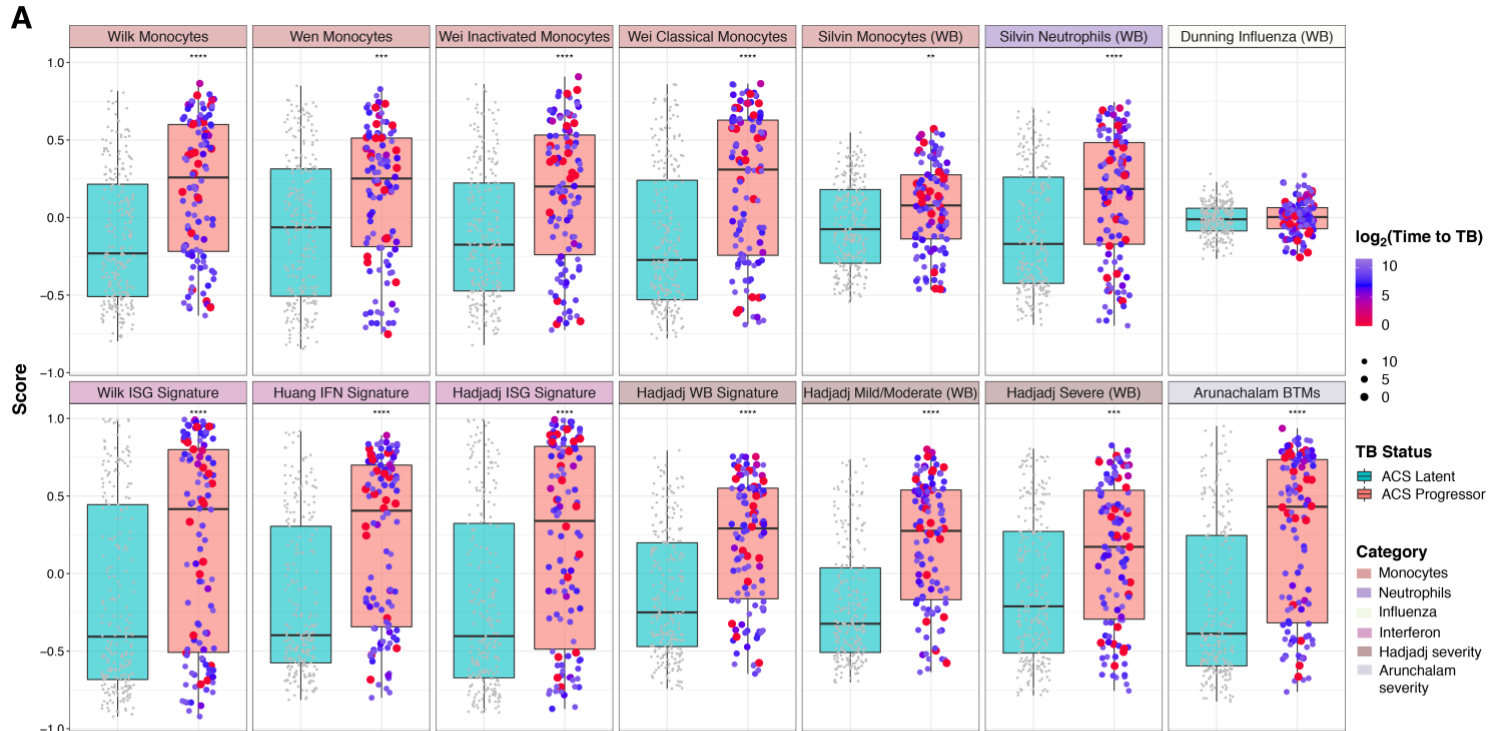
**Figure 6. Shared COVID-19 and TB cluster differentially expressed genes correlate with disease severity and interfere with COVID-19-specific gene signature detection.** (A) Heatmap of the top 100 enriched pathways and three identified gene clusters, the circle colours correspond to the groups that formed part of the cluster, the letters in the circle indicate the groups: C, COVID-19; A, active TB; P, TB progressors; L, latent TB; F, influenza. (B) Bubble plot depicting the enrichment of identified cluster differentially expressed genes (DEGs) by COVID-19 disease severity (appendix 2 p8). The size and shade of the circle correspond to the  $-\log_{10} q\text{-value}$ , represented as a scale on the right. (C) Box plots of the gene set variation analysis (GSVA) of COVID-19 scores based on the DEGs of each of the three identified clusters. Membership of the groups in the cluster is indicated in the circle appearing in the title of each boxplot. Scores for each signature were compared between each group at the “London Latent” control group using a t-test adjusted for multiple testing using a Bonferroni correction. (D) All pathways identified in the pathway meta-analysis were considered and selected based on COVID-19 exclusivity; bar plot of GSVA for mTORC1 signalling pathway genes and the reduced COVID-19 20-gene signature, when profiled on the Leicester cohort and COVID-19 disease severity data. Scores for each signature were compared between each group at the “Healthy (Arunachalam *et al.*)” control group using a t-test adjusted for multiple testing using a Bonferroni correction.  $* < 0.05$ ,  $** < 0.005$ ,  $*** < 0.0005$ ,  $**** < 0.00005$ .



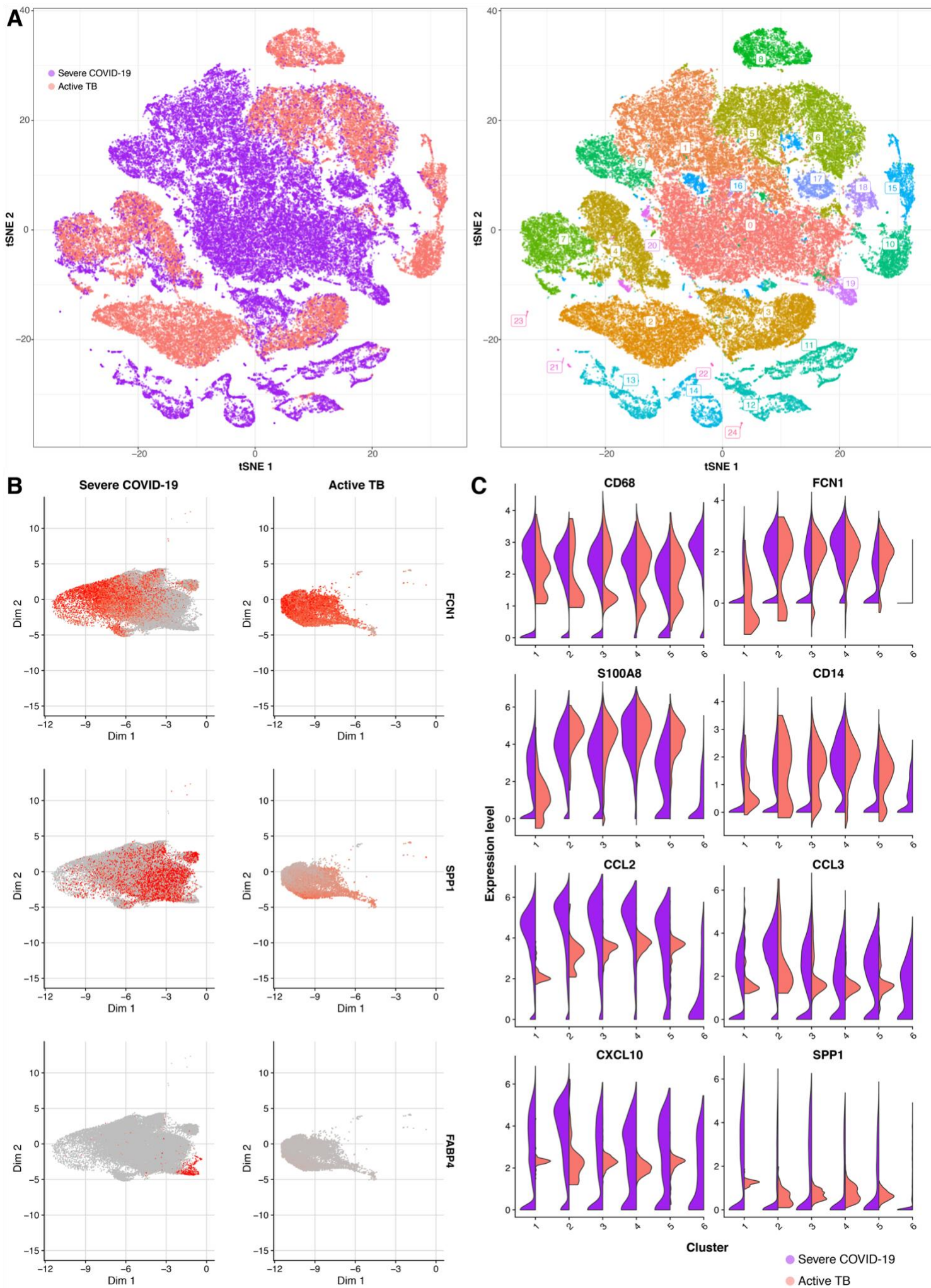
## Figures

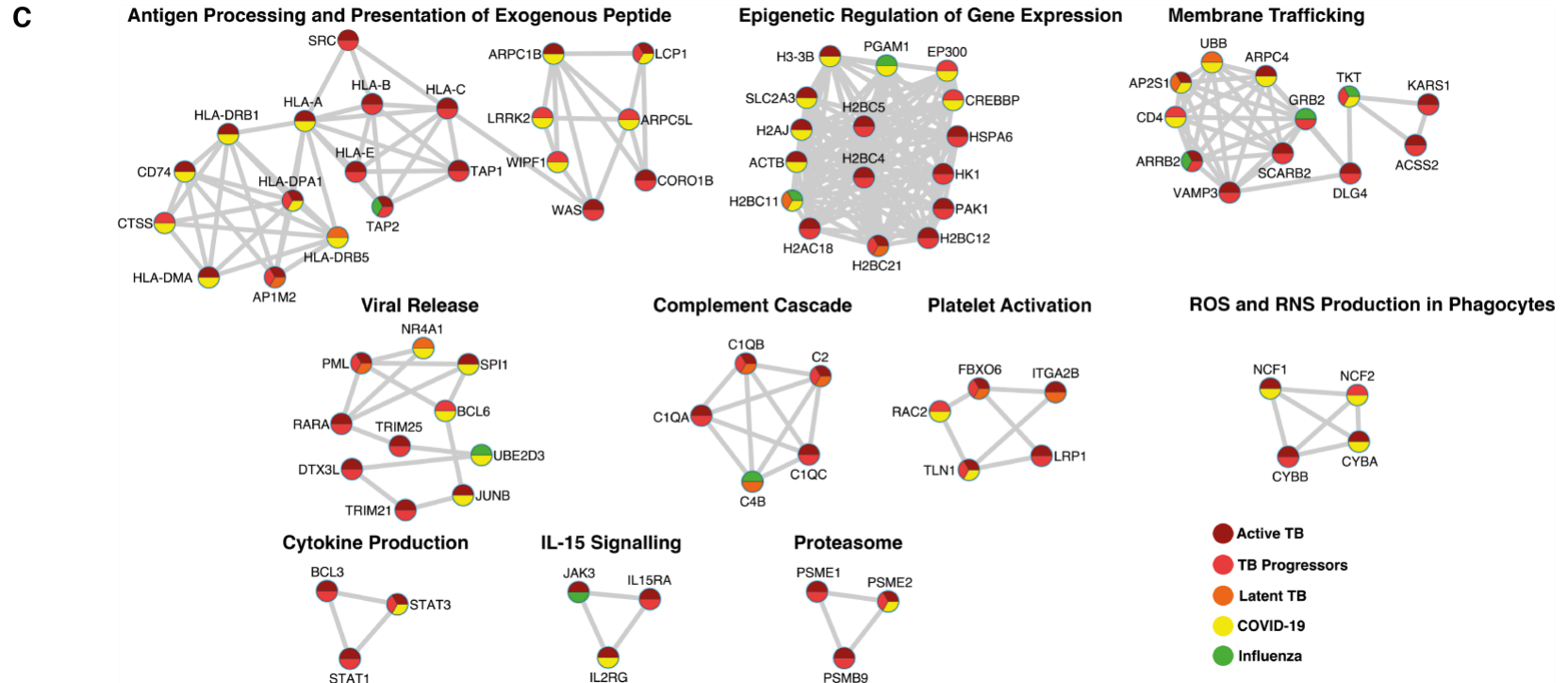
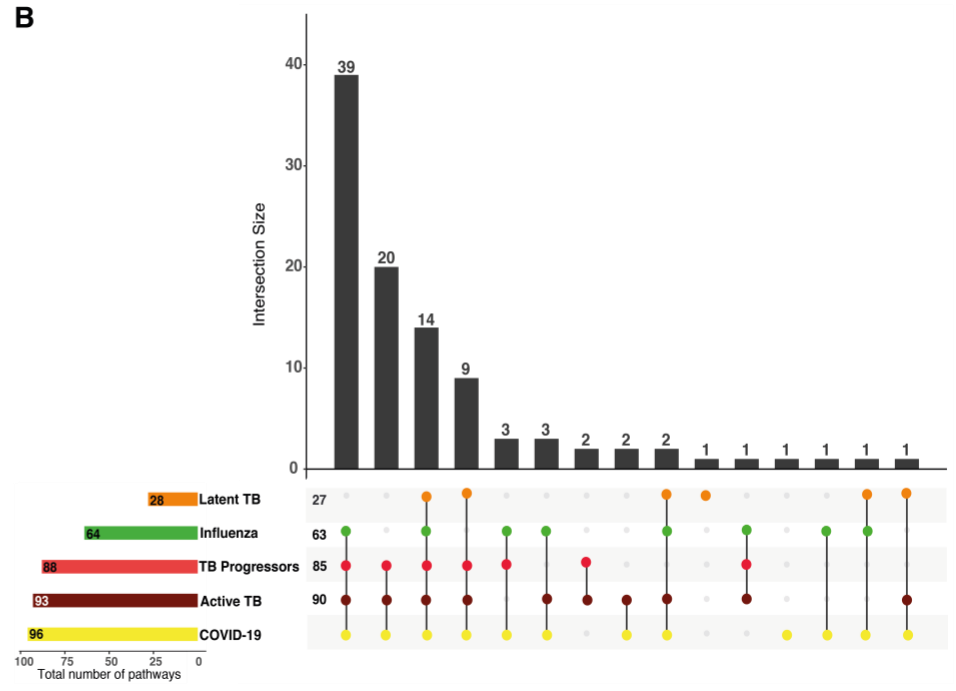
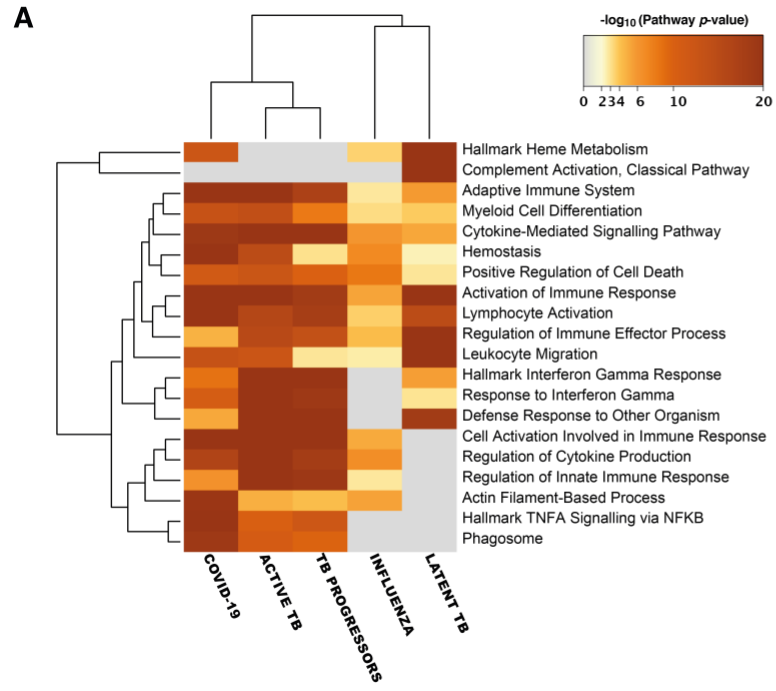












It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

