A modular framework for the development of targeted Covid-19 blood transcript profiling panels

Darawan Rinchai[1#], Basirudeen Kabeer[1#], Mohammed Toufiq[1#], Zohreh Calderone[1], Sara Deola[1], Tobias Brummaier[2], Mathieu Garand[1], Ricardo Branco[1], Nicole Baldwin[3], Mohamed Alfaki[1], Matthew Altman[4,5], Alberto Ballestrero[6,7], Matteo Bassetti[8], Gabriele Zoppoli[6,7], Andrea De Maria[8], Benjamin Tang[9], Davide Bedognetti[1] and Damien Chaussabel[1]*

1 Sidra Medicine, Doha, Qatar

2 Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand

3 Baylor Institute for Immunology Research and Baylor Research Institute, Dallas, Texas, USA

4 Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington, USA

5 Systems Immunology, Benaroya Research Institute, Seattle, Washington, USA

6 Department of Internal Medicine, Università degli Studi di Genova, Genoa IT

7 IRCCS Ospedale Policlinico San Martino, Genoa IT

8 Division of Infectious and Tropical Diseases, IRCCS Ospedale Policlinico San Martino, Genoa, Italy, and Department of Health Sciences, University of Genoa, Italy

9 Nepean Clinical School, University of Sydney, Sydney, NSW, Australia

[#] Equal contributing authors

*To whom correspondence may be addressed

dchaussabel@sidra.org

SUMMARY

Covid-19 morbidity and mortality are associated with a dysregulated immune response. Tools are needed to enhance existing immune profiling capabilities in affected patients. Here we aimed to develop an approach to support the design of focused blood transcriptome panels for profiling the immune response to SARS-CoV-2 infection. We designed a pool of candidates based on a pre-existing and well-characterized repertoire of blood transcriptional modules. Available Covid-19 blood transcriptome data was also used to guide this process. Further selection steps relied on expert curation. Additionally, we developed several custom web applications to support the evaluation of candidates. As a proof of principle, we designed three targeted blood transcript panels, each with a different translational connotation: therapeutic development relevance, SARS biology relevance and immunological relevance. Altogether the work presented here may contribute to the future expansion of immune profiling capabilities via targeted profiling of blood transcript abundance in Covid-19 patients.

Keywords: Blood transcriptomics, SARS-CoV-2, Covid-19, Immune monitoring

INTRODUCTION

Covid-19 is an infectious, respiratory disease caused by a newly discovered coronavirus: SARS-CoV-2. The severity of symptoms and the course of infection vary widely, with most patients presenting mild symptoms. However, about 20% of patients develop severe disease and require hospitalization (1,2). The interaction between innate and adaptive immunity can lead to the development of neutralizing antibodies against SARS-CoV-2 antigens that might be associated with viral clearance and protection (3). But immune factors are also believed to play an important role in the rapid clinical deterioration observed in some Covid-19 patients (4). There is thus a need to develop new modalities that can improve the delineation of “immune trajectories” during SARS-CoV-2 infection.

Blood transcriptome profiling involves measuring the abundance of circulating leukocyte RNA on a genome-wide scale (5). Processing of the samples and the raw sequencing data however, is time consuming and requires access to sophisticated laboratory and computational infrastructure. Thus, the possibility of implementing this approach on large scales to ensure immediate translational potential is limited. Such unbiased omics profiling data might rather be leveraged to inform the development of more practical, scalable and targeted transcriptional profiling assays. These assays could in turn serve to significantly bolster existing immune profiling capacity.

Fixed sets of transcripts grouped based on co-expression observed in large collections of reference datasets provide a robust platform for transcriptional profiling data analyses (6). Here we leveraged a repertoire of 382 transcriptional modules previously developed by our team (7). The repertoire is based on a collection of reference patient cohorts encompassing 16 pathological or physiological states and 985 individual transcriptome profiles. In this proof of principle study, we used the available transcript profiling data from two separate studies to select Covid-19 relevant sets of modules (8,9). Next, we applied filters based on pre-specified selection criteria (e.g. immunologic relevance or therapeutic relevance). Finally, expert curation was used as the last selection step. For this we have developed custom web applications to consolidate the information necessary

for the evaluation of candidates. One of these applications provides access to module-level transcript abundance profiles for available Covid-19 blood transcriptome profiling datasets. Another web interface was implemented which serves as a scaffold for the juxtaposition of such transcriptional profiling data with extensive functional annotations.

RESULTS

Mapping Covid-19 blood transcriptome signatures against a pre-existing reference set of transcriptional modules

SARS-CoV-2 infection might not result in changes in transcript abundance across all of the 382 transcriptional modules constituting our repertoire [(encompassing 14,168 transcripts) see methods section and (7)]. Indeed, the 16 reference patient cohorts upon which our repertoire was based encompass a wide range of immune states, including infectious diseases but also autoimmune diseases, pregnancy and cancer. Thus, the first step involved identifying subsets of modules for which changes could be observed in Covid-19 patients.

We used two sets of Covid-19 patients for this proof of principle analysis. These datasets were contributed by Xiong *et al.* (9) (one control and three subjects) and Ong *et al.* (8) (nine controls and three subjects profiled at multiple time points). Their data were generated using RNA-seq and Nanostring technology, respectively. The generic 594 transcript panel used by Ong *et al.* did not give sufficient coverage across the 382-module set. We thus mapped the transcript changes at a lower resolution, using a framework formed by 38 module "aggregates". These 38 aggregates were constituted by grouping modules based on similarities in abundance patterns across the 16 reference datasets [see methods section and (7)].

We first assessed changes in transcript abundance resulting from SARS-CoV-2 infection across the 38 module aggregates (**Figure 1**). In general, we saw a decrease in aggregates associated with lymphocytic compartments (aggregates A1 & A5) and an increase in aggregates associated with myeloid compartments and inflammation (aggregates A33 & A35). As expected, we also saw increases over uninfected controls for

the module aggregate associated with interferon (IFN) responses (A28) and the module aggregate presumably associated with the effector humoral response (A27). We detected a wide spread of values for aggregate A11 for the Nanostring (Ong *et al.*) dataset. However, this aggregate comprises only one module, with only two of its transcripts measured in this Nanostring code set (the probe coverage across all aggregates is shown in **Supplementary Figure 1**).

Despite large differences between the two studies in terms of design, range of clinical severity, technology platforms and module coverage, the combined overall changes (detected at a high-level perspective) are consistent with those observed in known acute infections, such as those caused by influenza, respiratory syncytial virus (RSV) or *S. aureus.* This consistency is evidenced by the indicated by patterns of change observed for the reference fingerprints shown alongside those of Covid-19 patients (**Figure 1**).

Overall, such a high-level analysis allows us to identify module aggregates forming our repertoire that may be included in further selection steps. In this proof of principle analysis, we selected 17 aggregates for further analysis (**Table 1**).

Identification of coherent sets of Covid-19-relevant modules

The abundance patterns for modules comprised in a given aggregate are not always homogeneous (**Figure 2**). Thus, a next step would consist of identifying sets of modules within an aggregate that display coherent abundance patterns across modules forming a given aggregate.

To achieve this, we first mapped the changes in transcript abundance associated with Covid-19 disease using the RNAseq dataset from Xiong *et al*., as illustrated for A31 (**Figure 2A**) and A28 (**Figure 3A**). Similar plots can be generated for all other aggregates using the “COVID-19” web application (also compiled in **Supplementary File 1** and listed in **Table 2**). Next, we identified and assigned a module set ID for each aggregate the modules that formed homogeneous clusters. For example, we designated the first A28 set as A28/S1. Such module grouping is only based on patterns of transcript abundance observed in three

Covid-19 patients; however, the groupings were often consistent with those observed for the much larger reference cohorts that constitute the module repertoire (**Figure 2B** and **Figure 3B**). A28/S1, which is formed by M8.3 and M15.127, serves as a good example of this consistency (**Figure 3B**). Likewise, the segregation of the modules forming A31 based on differences observed in the three Covid-19 patients was also apparent in the reference patient cohorts (**Figure 2B**). Specifically, an increase in A31/S1 modules, which accompanied a decrease in A31/S2 modules, in these three patients was also characteristic of RSV patients.

We ultimately derived 28 homogeneous Covid-19 relevant module sets from the 17 aggregates selected in the earlier step (**Table 1**). These sets might be used as a basis for further selection.

Design of an illustrative targeted panel emphasizing immunological relevance

In the previous step, we used available Covid-19 data to guide the selection of 28 distinct "Covid-19 relevant module sets". In the next step, we selected the transcripts within each module set that warranted inclusion in one of three illustrative Covid-19 targeted panels. A first panel was formed using immunologic relevance as the primary criterion, a second was formed on the basis of relevance to coronavirus biology, a third was constituted on the basis of relevance to therapy.

For the first panel we matched transcripts comprised in each module set to a list of canonical immune genes (see methods for details). Expert curation also involved accessing transcript profiling data from the reference datasets, indicating for instance leukocyte restriction or patterns of response to a wide range of immune stimuli *in vitro*. We describe our approach for module and gene annotation in more detail below and provide access to our resources to support expert curation (**Table 2**).

For our illustrative case, we selected one representative transcript per module set to produce a panel comprised of 28 representative transcripts (**Table 3)**. Examples of signatures surveyed by such a panel include: **1) ISG15** in A28/S1 (interferon responses),

which encodes for a member of the ubiquitin family. ISG15 plays a central role in the host defense to viral infections (10). **2) GATA1** in A37/S1 (erythroid cells), which encodes for a master regulator of erythropoiesis (11). It is associated with a module signature (A37) that we recently reported as being associated with immunosuppressive states, such as late stage cancer and maintenance immunosuppressive therapy in solid organ transplant recipients (12). In the same report we also found an association between this signature and heightened severity in patients with RSV infection and established a putative link with a population of immunosuppressive circulating erythroid cells (13). **3) CD38** in A27/S1 (cell cycle), which encodes for the CD38 molecule expressed on different circulating leukocyte populations. In whole blood we find the abundance of its transcript to correlates with that of IGJ, TNFRSF17 (BCMA), TXNDC5 (M12.15). Such a signature was previously found to be increased in response to vaccination at day 7 post administration, to correlate with the prevalence of antibody producing cells, and the development of antibody titers at day 28 (14). **4) TLR8** in A35/S1 (inflammation), encodes toll-like receptor 8. Expression of transcripts comprising this aggregate is generally restricted to neutrophils and robustly increased during sepsis (e.g. as we have described in detail earlier for ACSL1, another transcript belonging to this aggregate (15)). **5) GZMB** in A2/S1 (Cytotoxic cells) encodes Granzyme B, a serine protease known to play a role in immune-mediated cytotoxicity. Other transcripts forming this panel are listed in **Table 3**.

Even with the limited amount of data available to guide the selection in the previous steps, it is reasonable to assume that such a panel (while not optimal) would already provide valid information for Covid-19 immune profiling. Additional Covid-19 blood transcriptome data that will become available in the coming weeks will allow us to refine the overall selection process.

Design of an illustrative targeted panel emphasizing therapeutic relevance

A different translational connotation was given for this second panel. Here, we based the selection on the same collection of 28 module sets. However, this time, whenever

possible, we included transcripts that could have value as targets for the treatment of Covid-19 patients. An initial screen identified 82 transcripts encoding molecules that are known targets for existing drugs (see Methods). We further prioritized these candidates based on an expert's evaluation of the compatibility of use of the drugs for treating Covid-19 patients. As an exception, module sets belonging to A28 (interferon response) were selected based on their suitability as markers of a response to interferon therapy. Sets for which no targets of clinical relevance were identified (16/28) were instead represented in the panel by immunologically-relevant transcripts (defined earlier).

We ultimately identified a preliminary set of 12 targets through this high stringency selection process (**Table 4**). Developing effective immune modulation therapies in critical care settings has proven challenging (16). Current efforts in the context of Covid-19 disease particularly aim at controlling runaway systemic immune responses or so called "cytokine storms" that have been associated with organ damage and clinical worsening. Targets of interest identified among our gene set include: **1) IL6R** in A35/S2 (inflammation), encoding the Interleukin-6 Receptor, which is a target for the biologic drug Tocilizumab. Several studies have tested this antagonist in open label single arm trials in Covid-19 patients with the intent of blocking the cytokine storm associated with severe Covid-19 infection (17,18). **2) CCR2** in A26/1 (monocytes), encoding the chemokine (C-C motif) receptor 2, is targeted along with CCR5 by the drug Cenicriviroc. This drug exerts potent anti-inflammatory activity (19). **3) TBXA2R** in A31/1 (platelets), encoding the Thromboxane receptor, is targeted by several drugs with anti-platelet aggregation properties (20). **4) PDE8A** in A33/S1 (inflammation), encoding Phosphodiesterase 8A, is targeted by Pentoxifylline, a non-selective phosphodiesterase inhibitor that increases perfusion and may reduce risk of acute kidney injury and attenuates LPS-induced inflammation (21). **5) NQO1** in A8/S1 (Complement), encoding NAD(P)H quinone dehydrogenase 1. The NQO1 antagonist Vatiquinone (EPI-743) has been found to inhibit ferroptosis (22), a process associated with tissue injury (23), including in sepsis (24). A complete list is provided in **Table 4**.

The fact that this transcript panel and the previous survey the same pre-defined 28 homogenous Covid-19 relevant module sets should make them largely synonymous (since modules are formed on the basis of co-expression). Nevertheless, this second panel may be more relevant for investigators interested in investigating new therapeutic approaches or measuring responses to treatment.

Design of a targeted panel of blood transcripts of relevance for SARS-CoV-2 biology

For the third panel designed in this proof of principle, we primarily selected transcripts based on their relevance to SARS biology. As a first step, we used a literature profiling tool to identify SARS, MERS, or Covid-19 literature articles that were associated with transcripts forming the 28 Covid-19 module sets. Next, the potential associations were subjected to expert curation (see Methods). Once again, to keep redundancies to a minimum, we only included one candidate per set in this panel (**Table 5**). Notable examples include: **1) LTF** in A38/S1 (neutrophil activation) encodes Lactotransferrin, that is known to block the binding of the SARS-CoV spike protein to host cells, thus exerting an inhibitory function at the viral attachment stage (25). **2) FURIN** in A37/S1 (Erythroid cells), encodes a proprotein convertase that preactivates SARS-CoV-2, thus reducing its dependence on target cell proteases for entry (26). **3) EGR1** in A7/S1 (Monocytes), encodes Early Growth Response 1, which upon induction by SARS Coronavirus Papain-Like Protease mediates up-Regulation of TGF-β1 (27). **4) STAT1** in A28/S3 (Interferon response), encodes a transcription factor known to play an important role in the induction of antiviral effector responses. It was reported that SARS ORF6 Antagonizes STAT1 function by preventing its translocation to the nucleus and acts as an interferon antagonist in the context of SARS-CoV infection (28).

This screen identified several molecules that may be of importance for SARS-CoV-2 entry and replication. It is expected that this knowledge will evolve rapidly over time and frequent updates may be necessary. And, as for the previous two panels, investigators may

also have an interest in including more than one candidate per module set. This of course would also be feasible, although at the expense of course of parsimony.

Development of an annotation framework in support of signatures curation efforts

A vast amount of information is available to support the work of expert curators who are responsible for finalizing the selection of candidates. This process often requires accessing a number of different resources (e.g. those listed in **Table 2**). Here we have built upon earlier efforts to aggregate this information in a manner that makes it seamlessly accessible by the curators.

As proof of principle, we created dedicated, interactive presentations in Prezi for module aggregates A28 ((24)) and A31 (https://prezi.com/view/zYCSLyo0nvJTwjfJkJqb/). These presentations are intended, on the one hand, to aggregate contextual information that can serve as a basis for data interpretation. On the other hand, they are intended to capture the results of the interpretative efforts of expert curators.

The interactive presentations are organized in sections, each showing aggregated information from a different level: module-sets, modules and transcripts (**Figure 5**). The information derived from multiple online sources, including both third party applications and custom applications developed by our team (**Table 2**). Among those is a web application developed specifically for this work, which was used to generate the Covid-19 plots from Ong *et al.* and Xiong *et al.* (**Figure 5A**). The interactive presentation itself permits to zoom in and out, determine spatial relationships and interactively browse the very large compendium of analysis reports and heatmaps generated as part of these annotation efforts. The last section that contains transcript-centric information, is also the area where interpretations from individual curators is aggregated.

We have annotated and interpreted some of the transcripts included in **A31/S1** in such a manner: **1) OXTR,** which encodes for the Oxytocin receptor through which anti-inflammatory and wound healing properties of Oxytocin are mediated (29). Among our reference cohort datasets, OXTR is most highly increased in patients with *S. aureus*

infection or active pulmonary tuberculosis (7). **2) CD9,** which encodes a member of the tetraspanin family, facilitates the condensation of receptors and proteases activating MERS-CoV and promoting its rapid and efficient entry into host cells (30). **3) TNFSF4**, which encodes for OX40L and is a member of the TNF superfamily. Although OX40L is best known as a T-cell co-stimulatory molecule, reports have also shown that it is present on the neutrophil surface (31). Furthermore, OX40L blockade improved outcomes of sepsis in an animal model.

Our interpretation efforts have been limited thus far by expediency. Certainly, interpretation will be the object of future, more targeted efforts. In the meantime, this annotation framework supports the selection of candidates forming the panels presented here. It may also serve as a resource for investigators who wish to design custom panels of their own.

DISCUSSION

Early reports point to profound immunological changes occurring in affected patients during the course of a SARS-CoV-2 infection (32,33). In particular, patterns of immune dysfunction have been associated with clinical deterioration and the onset of severe respiratory failure (34). However, disease outcomes remain highly heterogeneous and factors contributing to clinical deterioration are poorly understood. Among other modalities, means to establish comprehensive immune monitoring in cohorts of Covid-19 patients are needed.

Here we designed an approach select and curate targeted blood transcript panels relevant to Covid-19. When finalized, such panels could in turn serve as a basis for rapid implementation of focused transcript profiling assays. This process should become possible as more Covid-19 blood transcriptome profiling datasets become available in the coming weeks and months. These data could, for instance, be used to refine the delineation of Covid-19 module sets.

Because our selection strategy relies primarily on a pre-existing module repertoire framework, we anticipate that changes would only be relatively minor. Indeed, one

advantage of basing candidate selection on a repertoire of transcriptional modules is that it permits to derive non-synonymous transcripts sets. In other words, each transcript included in the panel could survey the abundance of a different module (signature). Basing selection on differential expression instead, for instance, would tend to select multiple transcripts from more dominant signatures (with highest significance / fold changes). But for more specific purposes, such as differential diagnosis or prediction, machine learning models would be more appropriate [for instance in sepsis studies: (35)]. Indeed, such panels have already been developed for Covid-19 (36), and we anticipate that more will emerge over the coming months. However, our intent here was different: our primary aim was to support the development of a solution that can monitor immune responses and functions.

Delineation of “immune trajectories” associated with clinical worsening of Covid-19 patients is one application to consider. Another application would be the measurement of responses to therapy (as part of standard of care or a trial). The immune profiling of asymptomatic or pre-symptomatic patients (e.g quarantined) would be another setting where implementation of such an assay could prove useful. For this, it would for instance be possible to use protocols that we have previously developed for home-based, self-sampling and blood RNA stabilization (37,38).

Different connotations were given for the three panels, which are presented here as a proof of principle. The panel consisting of immunologically relevant markers might have the highest general interest. However, measuring changes in the abundance of transcripts coding for molecules that are targetable by existing drugs could have higher translational potential. Another illustrative panel comprises transcripts coding for molecules that are of relevance to SARS-CoV-2 biology and might be of additional interest in investigations of host–pathogens interactions. The common denominator between these panels is that they comprise representative transcripts of each of the 28 module sets. Other transcript combinations following the same principle would be possible, as well as the inclusion of multiple transcripts from the same set for added robustness. The obvious disadvantage of the latter, however, is the increase to the size of the panel. Medium-throughput technology

platforms, such as the Nanostring Ncounter System, Fluidigm Biomark or ThermoFisher Openarray, would be appropriate for implementing custom profiling assays with the number of targets comprising the tentative panels presented here (or a combination thereof). Downsizing panels to comprise ±10 key markers might serve as a basis for implementation on more ubiquitous real-time PCR platforms.

Overall, this work lays the ground for a framework that could support the development of increasingly more refined and interpretable targeted panels for profiling immune responses to SARS-CoV-2 infection. This should be possible in part through the further development of environments providing investigators with seamless access to vast amounts of annotations aggregated from different sources.

ACKNOWLEDGEMENTS

Development of some of the bioinformatic resources and approaches employed here was supported by NPRP grant # 10-0205-170348 from the Qatar National Research Fund (a member of Qatar Foundation). The work reported herein is solely the responsibility of the authors.

AUTHOR CONTRIBUTIONS

DR, BK, MT, NB, Malt, MB, GZ, ADM, BT, DB, DC conceptualization. DR, BK, MT, ZC, SD, TB, MG: data curation and validation. DR, BK, MT and DC: visualization. DR, BK, MT, ZC, SD, TB, MG, DB, DC: analysis and interpretation. DB, DC: funding acquisition. DR and DC: methodology development. DR, BK, MT, DC: writing of the first draft. DR, MT, MAlf: software development and database maintenance. DR, BK, MT, ZC, SD, TB, MG, RB, NB, MAlf, Malt, AB, MB, GZ, ADM, BT, DB, DC writing–review and editing The contributor's roles listed above follow the Contributor Roles Taxonomy (CRediT) managed by The Consortia Advancing Standards in Research Administration Information (CASRAI) (https://casrai.org/credit/).

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1: Mapping Covid-19 blood transcriptome signatures at the module aggregate level. The columns on this heatmap represent samples (Xiong *et al.* & Ong *et al.*) or patient cohorts (Altman *et al.*). Module aggregates (A1-A38) are arranged as rows. The colored spots represent the proportion of transcripts comprising each transcriptional module aggregate found to be differentially expressed compared to control samples. The cutoffs vary from one study to another due to differences in the design and the profiling platforms used. Thus, module aggregate response values range from 100% (all transcripts comprised in the module aggregate increased) to -100% (all decreased). The Xiong *et al.* dataset comprised one control and three Covid-19 patients and transcript abundance was measured by RNA-seq. The Ong *et al.* dataset comprised three Covid-19 cases from whom samples were collected serially, and nine uninfected controls (8). Transcript abundance was measured using a 594 gene standard immune panel from Nanostring. Patterns are also shown for cohorts comprised in the Altman *et al.* dataset (7). The colored labels (right) indicate functional associations for some of the aggregates.

Figure 2. Delineation of sets of Covid-19 relevant A31 modules. A. Transcript abundance profiles of A31 modules in Covid-19 patients. This heatmap represents the abundance levels for transcripts forming modules belonging to aggregate A31 (rows), across three Covid-19 patients (P1-P3) relative to one uninfected control subject (columns). The data are expressed as the proportion of constitutive transcripts in each module being significantly increased (red circles) or decreased (blue circles) relative to N1. **B. Transcript abundance profiles of A31 modules in reference disease cohorts.** The top heatmap represents the abundance levels for transcripts forming modules belonging to aggregate

A31 (rows), across 16 reference patient cohorts (columns). The bottom heatmaps represent the changes in abundance across the individuals comprised in two relevant patient cohorts, including pediatric patients with severe influenza or RSV infection and adult patients with sepsis.

Figure 3. Delineation of sets of Covid-19 relevant A28 modules. A. Transcript abundance profiles of A28 modules in Covid-19 patients. This heatmap shows the abundance levels for transcripts forming modules belonging to aggregate A28 (rows), across three Covid-19 patients (P1-P3) relative to one uninfected control subject (columns). The data are expressed as the proportion of constitutive transcripts in each module being significantly increased (red circles) or decreased (blue circles) relative to N1 **B. Transcript abundance profiles of A28 modules in reference disease cohorts.** The top heatmap shows the abundance levels for transcripts forming modules belonging to aggregate A28 (rows), across 16 reference patient cohorts (columns). The bottom heatmaps show changes in abundance across individuals constituting the two relevant patient cohorts, including pediatric patients with severe influenza or RSV infection and adult patients with sepsis.

Figure 4 Changes in abundance of transcripts comprising aggregate A28 in response to SARS-CoV-2 infection. The heatmaps display the changes in transcript abundance in three Covid-19 patients comprising the Xiong *et al.* RNA-seq transcriptome datasets. The top heatmap summarizes the module-level values for the six modules forming aggregate A28. The color code indicates membership to one of the three Covid-19 module sets that were defined earlier. The bottom heatmap shows patterns of abundance for the same six modules, but at the individual gene level. The line graphs on the right show changes in abundance for three transcripts from the “therapeutic relevance panel” in three Covid-19 patients profiled by Ong *et al*. using a generic Nanostring immune set comprising 594 transcripts.

Figure 5: High resolution annotation framework supporting the curation and interpretation of Covid-19 module sets. This series of screenshots shows the content of the interactive presentations that have been established to provide curators with access to detailed annotations regarding modules forming a given aggregate, its constitutive modules and targets that have been selected for inclusion in transcript panels. Links to interactive presentations and resources mentioned below are available in **Table 2**. **A. Module aggregate-level information.** This section displays patterns of transcript abundance across the modules forming a given aggregate, as well as the degree of association of this aggregate with the severity of RSV disease. Plots used to populate this section were generated using three web applications, including one that was developed in support of this work that compiles the Covid-19 blood transcriptional data available to date. The other two applications were developed as part of a previous study to generate plots for the reference disease cohorts and RSV severity association plots (45). **B. Module-level information.** This section includes, for a given module, reports from functional profiling tools as well as patterns of transcript abundance across the genes forming the module. Drug targeting profiles were added to provide another level of information. **C. Gene-centric information.** The information includes curated pathways from the literature, articles and reports from public resources. Gene-centric transcriptional profiles that are available via gene expression browsing applications deployed by our group are also captured and used for context (GXB). A synthesis of the information gathered by expert curation and potential relevance to SARS-Cov-2 infection can also be captured and presented here.

Supplementary Figure 1: Coverage of the pre-established 38 transcriptional module aggregate repertoire by the Nanostring immunology panel 2. The bar graphs show the distribution of the 579 transcript constituting the standard Nanostring immunology panel used by Ong *et al.* across the 38 module aggregates forming this repertoire. The Venn

diagram shows the degree of overlap between the Nanostring panel and the transcripts forming this modular repertoire.

Table 1: List of Covid-19 relevant aggregates and module sets

Module Aggregate	Module Set	Modules	Functional annotations
A1	**A1/S1**	M14.42, M15.38, M12.6, M13.27,	T cells
	A1/S2	M14.23, M15.87, M14.5, M14.49, M12.1, M14.20	Gene transcription
	A1/S3	M12.8, M15.29, M14.58, M15.51, M14.64, M16.78, M14.75, M15.82, M14.80	B cells
A2	**A2/S1**	M13.21, M9.1	Cytotoxic lymphocytes
	A2/S2	M14.13, M14.72, M13.13, M13.14, M14.45, M13.10, M15.91	TBD
A4	**A4/S1**	M16.69, M16.72, M16.50, M16.77	Antigen presentation,
A5	**A5/S1**	M16.95, M16.36	B cells
	A5/S2	M16.57, M16.18, M16.65, M16.111, M16.99	B cells
A7	**A7/S1**	M15.61	Monocytes
A8	**A8/S1**	M16.30	Complement
	A8/S2	M16.106	TBD
A10	**A10/S1**	M15.102	Prostanoids
A26	**A26/S1**	M12.2	Monocytes
A27	**A27/S1**	M13.32, M12.15, M16.92, M15.110, M16.60	Antibody producing cells
A28	**A28/S1**	M15.127, M8.3	Interferon response
	A28/S2	M15.64	Interferon response
	A28/S3	M15.86, M10.1, M13.17	Interferon response
A31	**A31/S1**	M16.64	Platelet/Prostaglandin
	A31/S2	M15.58	Monocytes
A33	**A33/S1**	M15.104, M14.82, M14.24, M15.108	Cytokines/chemokines, Inflammation
	A33/S2	M14.19, M14.76, M14.50, M14.26, M16.101, M16.100, M16.80	Inflammation
A34	**A34/S1**	M14.59, M10.3, M16.109, M8.2	Platelets, Prostanoids
A35	**A35/S1**	M14.65, M14.28, M15.81, M16.79, M13.3, M14.7,	Monocytes, Neutrophils
	A35/S2	M15.26, M12.10, M13.22, M15.109, M15.78, M13.16,	Neutrophils, Inflammation
A36	**A36/S1**	M16.34, M16.82, M15.97, M14.51, M15.118, M16.88	Gene transcription

A37	**A37/S1**	M9.2, M14.53, M11.3, M12.11, M15.100 M15.74, M13.26, M13.30, M15.53,	Erythroid cells
A38	**A38/S1**	M10.4	Neutrophil activation
	A38/S2	M16.96, M12.9, M14.68	Erythroid cells

Table 2: Resources used for annotation and interpretation

Platform / Use	Name	Source	Notes	Link	Demo video	Ref.
Interactive presentations (Prezi) Exploration of aggregated annotations / curation / capture interpretation of functional relevance in the context of Covid-19 disease	**Covid-19 Module Sets annotation framework**	In house / open	See Figure 5 for details	A31: https://prezi.com/view/zYCSLyo0nvJTwjfJkJqb/ A28: https://prezi.com/view/7lbgGwfiNflffqQzvL14/)	Part 1: https://youtu.be/_7sNE3e5W5g Part 2: https://youtu.be/l6wHrrmbet4 Part 3: https://youtu.be/iHnM7OH_nw8	Present work
R Shiny Web Applications/ Exploration of module-level and gene-level blood transcriptome profiling data. Design and export of custom plots to populate the annotation framework.	**Covid-19 app**	In house / open	Provides access to two Covid-19 blood transcript profiling datasets. More will be added as they become available.	https://drinchai.shinyapps.io/COVID_19_project/	https://youtu.be/XhQZj9mm2ME	Present work
	Gen3 app	In house / open	Provides access to 16 reference patient cohorts datasets	https://drinchai.shinyapps.io/dc_gen3_module_analysis/#	https://youtu.be/y_7xKJo5e4	Altman et al. (7)
	RSV app	In house / open	Provides access to six public RSV blood transcriptome datasets.	https://drinchai.shinyapps.io/RSV_Meta_Module_analysis/	https://youtu.be/htNSMreM8es	Rinchai et al. (12)
Gene Expression Browser (GXB). Interactive browsing of expression profiles for individual transcripts. Themed curated dataset collections have been created.	**GXB sepsis collection**	In house / open	Makes 93 curated datasets relevant to sepsis	http://sepsis.gxbsidra.org/dm3/geneBrowser/list	https://youtu.be/D1rGYfVSAoM	Toufiq et al. (in preparation), and Speake et al. (46)
			A reference dataset presenting transcript abundance profiles across purified leukocyte populations	http://sepsis.gxbsidra.org/dm3/geneBrowser/show/4000098		Linsley et al. (47)
			Reference dataset presenting the response to *in vitro* blood stimulations	http://sepsis.gxbsidra.org/dm3/geneBrowser/show/4000152		Obermoser et al. (14)
	GXB Acute Respiratory Infection	In house / open	34 curated datasets relevant to acute respiratory infections	http://vri1.gxbsidra.org/dm3/geneBrowser/list		Bougarn et al. (48)
			Reference dataset presenting changes in blood transcript abundance in patients with pneumonia	http://vri1.gxbsidra.org/dm3/miniURL/view/Mh		Parnell et al. (49)
Gene Set Annotation tool.	**GSAN**	Third party / open		https://gsan.labri.fr/		Allion-Benitez et al. (50)

Pathway Analysis tool	**Ingenuity Pathway Analysis**	Third party / commercial	Pathway enrichment. Used for expert curation of candidates.	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/		
Accummenta Biotech LiteratureLab Literature profiling and keyword enrichment tool	**LitLab Gene Retreiver**	Collaboration with Third party / commercial	Retrieves genes from a collection of literature records provided by the user	https://www.acumenta.com/generetriever		
Drug target identification	**Open targets**	Third party / open		https://www.targetvalidation.org/		
Retrieval of lists of immune-relevant genes	**Immport**	Third party / open		https://immport.org/shared/home		Bhattacharya et al. (44)

Table 3: Illustrative targeted panel – Immunology relevance focus

Module set	Module ID	NCBI Entrez ID	Symbol	Name	Module set functional annotation
A1/S1	M15.38	916	CD3E	CD3e molecule	T cells
A1/S2	M14.49	974	CD79B	CD79b molecule	Gene transcription
A1/S3	M14.80	3122	HLA-DRA	major histocompatibility complex, class II, DR alpha	B cells
A2/S1	M9.1	3002	GZMB	granzyme B	Cytotoxic lymphocytes
A2/S2	M13.13	4282	MIF	macrophage migration inhibitory factor	TBD
A4/S1	M16.77	3811	KIR3DL1	killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 1	Antigen presentation
A5/S1	M16.95	972	CD74	CD74 molecule	B cells
A5/S2	M16.111	27242	TNFRSF21	TNF receptor superfamily member 21	B cells
A7/S1	M15.61	23166	STAB1	stabilin 1	Monocytes
A8/S1	M16.30	3600	IL15	interleukin 15	Complement
A8/S2	M16.106	57823	SLAMF7	SLAM family member 7	TBD
A10/S1	M15.102	246	ALOX15	arachidonate 15-lipoxygenase	Prostanoids
A26/S1	M12.2	942	CD86	CD86 molecule	Monocytes
A27/S1	M12.15	608	CD38	CD38 molecule	Cell Cycle
A28/S1	M8.3	9636	ISG15	ISG15 ubiquitin like modifier	Interferon response
A28/S2	M15.64	10475	TRIM38	tripartite motif containing 38	Interferon response
A28/S3	M10.1	115362	GBP5	guanylate binding protein 5	Interferon response
A31/S1	M16.64	1950	EGF	epidermal growth factor	Platelet/Prostaglandin
A31/S2	M15.58	2214	FCGR3A	Fc fragment of IgG receptor IIIa	Monocytes
A33/S1	M14.24	91	ACVR1B	activin A receptor type 1B	Cytokines/chemokines, Inflammation
A33/S2	M14.19	23765	IL17RA	interleukin 17 receptor A	Inflammation
A34/S1	M8.2	3674	ITGA2B	integrin subunit alpha 2b	Platelets, Prostanoids
A35/S1	M13.3	1241	LTB4R	leukotriene B4 receptor	Inflammation
A35/S2	M12.10	51311	TLR8	toll like receptor 8	Neutrophils, Inflammation
A36/S1	M16.34	2993	GYPA	glycophorin A (MNS blood group)	Gene transcription
A37/S1	M11.3	2623	GATA1	GATA binding protein 1	Erythroid cells
A38/S1	M10.4	4057	LTF	lactotransferrin	Neutrophil activation
A38/S2	M16.96	56729	RETN	resistin	Erythroid cells

Table 4: Illustrative targeted panel – Therapeutic relevance focus

Module set	Module ID	NCBI Entrez ID	Symbol	Name	Relevance	Notes
A1/S1	M15.38	916	CD3E	CD3e molecule	Immunological	Not suitable for targeting (adaptive immunity)
A1/S2	M14.49	974	CD79B	CD79b molecule	Immunological	Not suitable for targeting (adaptive immunity)
A1/S3	M14.80	3122	HLA-DRA	major histocompatibility complex, class II, DR alpha	Immunological	Not suitable for targeting (adaptive immunity)
A2/S1	M9.1	3002	GZMB	Granzyme B	Immunological	Not suitable for targeting (adaptive immunity)
A2/S2	M13.13	4282	MIF	Macrophage migration inhibitory factor	Immunological	Not suitable for targeting (adaptive immunity -presumed)
A4/S1	M16.77	3811	KIR3DL1	Killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 1	Immunological	Not suitable for targeting (adaptive immunity)
A5/S1	M16.95	972	CD74	CD74 molecule	Immunological	Not suitable for targeting (adaptive immunity)
A5/S2	M16.111	27242	TNFRSF21	TNF receptor superfamily member 21	Immunological	Not suitable for targeting (adaptive immunity)
A7/S1	M15.61	23166	STAB1	stabilin 1	Immunological	No suitable candidates identified
A8/S1	M16.30	1728	NQO1	NAD(P)H quinone dehydrogenase 1	**Therapeutic**	**Vatiquinone (EPI-743) has been found to inhibit ferroptosis (22), a process associated with tissue injury (23), including in sepsis (24).**
A8/S2	M16.106	57823	SLAMF7	SLAM family member 7	Immunological	No suitable candidates identified
A10/S1	M15.102	246	ALOX15	Arachidonate 15-lipoxygenase	Immunological	No suitable candidates identified
A26/S1	M12.2	729230	CCR2	C-C motif chemokine receptor 2	**Therapeutic**	**Anti-inflammatory properties have been attributed to the CCR2/CCR5 blocker Cenicriviroc (51)**
A27/S1	M12.15	608	TNFRSF17	TNF receptor superfamily member 17	Immunological	Not suitable for targeting (adaptive immunity)
A28/S1	M8.3	4599	MX1	MX dynamin like GTPase 1	**Therapeutic**	**Inducible by Interferon-beta treatment**
A28/S2	M15.64	1230	CCR1	C-C motif chemokine receptor 1	**Therapeutic**	**Inducible by Interferon-beta treatment**
A28/S3	M10.1	3433	IFIT2	interferon induced protein with tetratricopeptide repeats 2	**Therapeutic**	**Inducible by Interferon-beta treatment**
A31/S1	M16.64	6915	TBXA2R	Thromboxane A2 receptor	**Therapeutic**	**Thromboxane A2 synthase inhibitors have antiplatelet aggregation activities and anti-inflammatory activities (drugs include: Defibrotide / Seratrodast, Ozagrel)**
A31/S2	M15.58	5743	PTGS2	Prostaglandin-endoperoxide	**Therapeutic**	**PTGS2 encodes COX-2. Several specific inhibitors**

				synthase 2		**are available which possess anti-inflammatory properties (e.g. celecoxib, rofecoxib, valdecoxib).**
A33/S1	M15.104	5151	PDE8A	Phosphodiesterase 8A	**Therapeutic**	**PDE8A, is targeted by Pentoxifylline, a non-selective phosphodiesterase inhibitor that increases perfusion and may reduce risk of acute kidney injury and attenuates LPS-induced inflammation**
A33/S2	M14.19	23765	IL17RA	Interleukin 17 receptor A	**Therapeutic**	**Brodalumab may be beneficial in reducing the viral illness exacerbation. But current recommendation is discontinuation of use in COVID 19**
A34/S1	M16.109	5742	PTGS1	Prostaglandin-endoperoxide synthase 1	**Therapeutic**	**Encodes for Cox-1. COX inhibitors including Aspirin, Indomethacin, Naproxen have direct antiviral properties as well as anti-inflammatory and antithrombotic properties**
A35/S1	M15.81	5293	PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta	**Therapeutic**	**PI3K-δ and PI3K-γ Inhibition by IPI-145 Abrogates Immune Responses and Suppresses Activity in Autoimmune and Inflammatory Disease Models (52)**
A35/S2	M15.109	3570	IL6R	Interleukin 6 receptor	**Therapeutic**	**IL6R is a target for the biologic drug Tocilizumab. Several studies have tested this antagonist in open label single arm trials in Covid-19 patients with the intent of blocking the cytokine storm associated with Covid-19 disease (15,16)**
A36/S1	M16.34	2993	GYPA	glycophorin A (MNS blood group)	Immunological	Not suitable for targeting (erythropoiesis)
A37/S1	M11.3	2623	GATA1	GATA binding protein 1	Immunological	Not suitable for targeting (erythropoiesis)
A38/S1	M10.4	4057	LTF	Lactotransferrin	Immunological	No suitable candidates identified
A38/S2	M16.96	56729	RETN	Resistin	Immunological	Not suitable for targeting (erythropoiesis)

Table 5: Illustrative targeted panel – SARS biology relevance focus

Module set	Module ID	NCBI Entrez ID	Symbol	Name	Relevance	Notes
A1/S1	M15.38	916	CD3E	CD3e molecule	Immunological	
A1/S2	M12.1	60489	APOBEC3G	apolipoprotein B mRNA editing enzyme catalytic subunit 3G	CoV Biology	APOBEC3G associates with SARS viral structural proteins (53), with a possible role in restriction of RNA virus replication (54)
A1/S3	M14.64	51284	TLR7	toll like receptor 7	CoV Biology	TLR7 Signaling Pathway is inhibited by SARS Coronavirus Papain-Like Protease (27)
A2/S1	M13.21	3458	IFNG	interferon gamma	CoV Biology	Interferon-gamma and interleukin-4 Downregulate Expression of the SARS Coronavirus Receptor ACE2 (55)
A2/S2	M13.10	25	ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase	CoV Biology	Abl Kinase inhibitors block SARS-Cov fusion (56)
A4/S1	M16.77	3811	KIR3DL1	killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 1	Immunological	
A5/S1	M16.65	4092	SMAD7	SMAD family member 7	CoV Biology	MERS Coronavirus Induces Apoptosis in Kidney and Lung by Upregulating Smad7 and FGF2 (57)
A5/S2	M16.111	27242	TNFRSF21	TNF receptor superfamily member 21	Immunological	
A7/S1	M15.61	1958	EGR1	early growth response 1	CoV Biology	SARS Coronavirus Papain-Like Protease Induces Egr-1-dependent Up-Regulation of TGF-β1 (58)
A8/S1	M16.30	857	CAV1	caveolin 1	CoV Biology	Severe Acute Respiratory Syndrome Coronavirus Orf3a Protein Interacts with Caveolin (59)
A8/S2	M16.106	57823	SLAMF7	SLAM family member 7	Immunological	
A10/S1	M15.102	246	ALOX15	arachidonate 15-lipoxygenase	Immunological	
A26/S1	M12.2	942	CD86	CD86 molecule	Immunological	
A27/S1	M12.15	608	TNFRSF17	TNF receptor superfamily member 17	Immunological	
A28/S1	M8.3	9636	ISG15	ISG15 ubiquitin like modifier	CoV Biology	SARS-CoV PLpro exhibits ISG15 precursor processing activities (60)
A28/S2	M15.64	1230	CCR1	C-C motif chemokine receptor 1	CoV Biology	MLN-3897, a CCR1 antagonist inhibits replication of SARS-CoV-2 replication (61)
A28/S3	M10.1	6772	STAT1	signal transducer and activator of transcription 1	CoV Biology	SARS ORF6 Antagonizes STAT1 Function (28)
A31/S1	M16.64	1950	EGF	epidermal growth factor	Immunological	
A31/S2	M15.58	5743	PTGS2	prostaglandin-endoperoxide synthase 2	CoV Biology	Encodes COX2, which expression is stimulated by SARS Spike protein (62)
A33/S1	M14.24	114548	NLRP3	NLR family pyrin domain containing 3	CoV Biology	Multiple SARS-Coronavirus protein have been reported to activates

						NLRP3 inflammasomes (63,64)
A33/S2	M14.19	23765	IL17RA	interleukin 17 receptor A	Immunological	
A34/S1	M8.2	3674	ITGA2B	Integrin subunit alpha 2b	Immunological	
A35/S1	M13.3	1241	LTB4R	Leukotriene B4 receptor	Immunological	
A35/S2	M15.78	290	ANPEP	alanyl aminopeptidase, membrane	CoV Biology	A potential receptor for human CoVs (65)
A36/S1	M16.88	6352	CCL5	C-C motif chemokine ligand 5	CoV Biology	CCL5/RANTES is associated with the replication of SARS in THP-1 Cells (66)
A37/S1	M13.26	5045	FURIN	Furin, paired basic amino acid cleaving enzyme	CoV Biology	Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion (67)
A38/S1	M10.4	4057	LTF	Lactotransferrin	CoV Biology	Lactotransferrin blocks the binding of the SARS-CoV spike protein to host cells, thus exerting an inhibitory function at the viral attachment stage (25)
A38/S2	M12.9	1508	CTSB	Cathepsin B	CoV Biology	Activation of SARS- and MERS-coronavirus is mediated cathepsin L (CTSL) and cathepsin B (CTSB) (68)

Table 6: List of housekeeping genes that may be suitable for blood transcript profiling applications

Housekeeping Genes	NCBI Entrez ID	Symbol	Name
Housekeeping Gene	1794	DOCK2	dedicator of cytokinesis 2
Housekeeping Gene	1915	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
Housekeeping Gene	90268	FAM105B/ OTULIN	OTU deubiquitinase with linear linkage specificity
Housekeeping Gene	2512	FTL	ferritin light chain
Housekeeping Gene	103910	MYL12B/MRLC2	myosin light chain 12B
Housekeeping Gene	4637	MYL6	myosin light chain 6
Housekeeping Gene	6204	RPS10	ribosomal protein S10
Housekeeping Gene	6230	RPS25	ribosomal protein S25

METHODS

Datasets

Two Covid-19 blood transcriptional datasets available at the time this work was conducted were used: **1)** Xiong *et al.* (9) obtained peripheral mononuclear cell samples obtained from one uninfected control individual and three patients with Covid-19. RNA abundance was profiled via RNAseq. The data were deposited in the Genome Sequence Archive of the Beijing Institute of Genomics, Chinese Academy of Sciences, under the accession number CRA002390. FASTQ passed QC and were aligned to reference genome GRChg38/hg19 using Hisat2 (v2.05). BAM files were converted to a raw count expression matrix using subreads (v1.6.2). Raw expression data was corrected for within lane and between lane effects using R package EDASeq (v2.12.0) and quantile normalized using preprocessCore (v1.36.0). The modular analysis was performed by using 10,617 RNA-seq genes which overlapped with transcripts from the 3rd generation module construction (7). details of the analysis as described below section.

2) Ong *et al.* (8) collected whole blood stabilized in RNA buffer from uninfected controls and three Covid-19 patients at multiple time points. RNA abundance was profiled using a standard immunology panel from Nanostring comprising 594 transcripts. The data were deposited in the arrayexpress public repository with accession ID E-MTAB-8871. The normalized data were downloaded, and modular analysis was performed by using 403 NanoString genes which overlapped with transcripts from the 3nd generation module construction details of the analysis as described below section.

3) We also used a reference dataset generated by our group that was previously used to construct the 382 blood transcriptional module repertoire (7). Briefly, this repertoire consists of the following cohorts of patients and respective control subjects: *S. aureus* infection (99 cases, 44 controls), sepsis (35 cases, 12 controls), tuberculosis (23 cases, 11 controls), Influenza (25 cases, 14 controls), RSV infection (70 cases, 14 controls), HIV infection (28 cases, 35 controls), systemic lupus erythematosus (55 cases, 14 controls), multiple sclerosis (34 cases, 22 controls), juvenile dermatomyositis (40 cases, 9 controls),

Kawasaki disease (21 cases, 23 controls), systemic onset idiopathic arthritis (62 cases, 23 controls), COPD (19 cases, 24 controls), melanoma (22 cases, 5 controls), pregnancy (25 cases, 20 controls), liver transplant recipients (94 cases, 30 controls), and B-cell deficiency (20 cases, 13 controls). All samples were run at the same facility on Illumina HumanHT-12 v3.0 Gene Expression BeadChips. The data have been deposited in NCBI GEO with accession number GSE100150.

Transcriptional module repertoire

The method used to construct the transcriptional module repertoire has been described elsewhere (39,40). The version used here is the third and last to have been developed by our group over a period of 12 years. It is the object of a separate publication (available on a pre-print server (7)).

Briefly, the approach consists of identifying sets of co-expressed transcripts in a wide range of pathological or physiological states, focusing in this case on the blood transcriptome as the biological system. We determined co-expression based on patterns of co-clustering observed for all gene pairs across the collection of 16 reference datasets listed in the previous section and that encompassed viral and bacterial infectious diseases as well as several inflammatory or autoimmune diseases, B-cell deficiency, liver transplantation, stage IV melanoma and pregnancy. Overall, this collection comprised 985 blood transcriptome profiles. A weighted, co-expression network was built with the weight of the nodes connecting a gene pair being based on the number of times co-clustering was observed for the pair among the 16 reference datasets. Thus, the weights ranged from 1 (where co-clustering occurs in one of 16 datasets) to 16 (where co-clustering occurs in all 16 datasets). Next, this network was mined using a graph theory algorithm to define subsets of densely connected gene sets that constituted our module repertoire (“Cliques” and “Paracliques”).

Overall, 382 transcriptional modules were identified, encompassing 14,168 transcripts. A supplemental file including the definition of this module repertoire along with

the functional annotations is available elsewhere (7). To provide another level of granularity and facilitate data interpretation, a second round of clustering was performed to group the modules into “aggregates”. This process was achieved by grouping the set of 382 modules according to the patterns of transcript abundance across the 16 reference datasets that were used for module construction. This segregation resulted in the formation of 38 aggregates, each comprising between one and 42 modules.

Module repertoire analyses

The modular analyses were performed using the core set of 14,168 transcripts forming the module repertoire. For group-level comparisons (cases vs controls), a paired t-test was performed on the log2-transformed data [Fold change (FC) cut off = 1.5; FDR cut off = 0.1]. For individual-level analyses, each sample was compared to the mean value of the corresponding control samples (or individual sample in the case of the Xiong *et al.* dataset). The cut off comprised an absolute FC >1.5 and a difference in counts >10. The results for each module are reported as the percentage of its constitutive transcripts that increased or decreased in abundance. Because the genes comprised in a module are selected based on the co-expression observed in blood, the changes in abundance within a given module tend to be coordinated and the dominant trend is therefore selected (the greater value of the percentage increased vs. percentage decreased). Thus, the values range from -100% (all constitutive modules are decreased) to +100% (all constitutive modules are increased). A module was considered to be “responsive” when the proportion of transcripts found to be increased was >15%, or when the proportion of transcripts found to be decreased was ≤15%. At the aggregate-level, the percent values of the constitutive modules were averaged.

Data visualization

Changes in transcript abundance reduced at the module or module aggregate-level were visualized using a custom fingerprint heatmap format. For each module, the percentage of increased transcripts is represented by a red spot and the percentage of decreased

transcripts is represented by a blue spot. The fingerprint grid plots were generated using "ComplexHeatmap" (41). A web application was developed to generate the plots and browse modules and module aggregates (https://drinchai.shinyapps.io/COVID_19_project/). A detailed description and source code will be available as part of a separate publication BioRxiv deposition on GitHub and BioRxiv (in preparation).

Selection of transcripts for inclusion in targeted panels

Therapeutic relevance: Covid-19 module sets belonging to aggregates comprising module annotations relating to inflammation, monocytes, neutrophils or coagulation pathway were selected for screening (A7, A8, A26, A31, A33, A34, A35). In turn, transcripts from each of the corresponding module sets were selected on the basis of their status as a known therapeutic target of a drug for which clinical precedence exists (source: targetvalidation.org). Next, candidates were prioritized via expert curation on the basis of compatibility and a potential benefit as a Covid-19 treatment. Curators were tasked with prioritizing candidates within each of the Covid-19 module sets. Only the top ranked candidate from each set was selected for inclusion in the panel. Module sets from aggregate A28 (interferon response) may also be of clinical relevance, may also be of clinical relevance, as indicators of a treatment response since interferon administration has been shown to increase the activity of anti-viral drugs in Covid-19 patients (42). The selection of candidates for aggregate A28 sets was thus based on the amplitude of the response to beta-interferon therapy measured in patients with multiple sclerosis [fold-change over pre-treatment baseline (43) & NCBI GEO accession GSE26104]. The remaining nine aggregates, which tended to associate preferentially with adaptive immune responses and for which targeting by therapies might prove detrimental, were not included in this screen. For these, representative transcripts from the default panel of immune relevant transcripts were included.

Relevance to Coronavirus biology: for the second panel, transcripts were primarily selected based on their relevance to SARS biology. As a first step, a literature profiling tool

was used to identify among the SARS, MERS, or Covid-19 literature articles that were associated with transcripts forming the 28 Covid-19 module sets (Literature Lab, Gene Profiler module; Accumenta Biotech, Boston, MA). Next, the potential associations were assessed by manual curation. The curators prioritized the transcripts for which the associations could be confirmed based on importance and robustness.

Immunological relevance: Lists of immunologically relevant genes were retrieved from Immport (44), and were used along with membership to IPA pathways (Ingenuity Pathway Analysis, QIAGEN, Germantown MD) to annotate transcripts comprising Covid-19 module sets. The curators prioritized annotated transcripts on the basis of their relevance to the functional annotations of the module set (e.g. interferon, inflammation, cytotoxic cells). The transcript with the highest priority rank was included in the assay.

Housekeeping genes: A recommended set of housekeeping genes is provided in **Table 6**. These were selected on the basis of low variance observed across the 985 transcriptome profiles generated for our reference cohorts.

Annotation framework

Links to the resources described in this section and to video demonstrations are available in **Table 2.** Interactive presentations were created via the Prezi web application. For this we have built and expanded upon an annotation framework established as part of the characterization of our reference blood transcriptome repertoire (7). Several bioinformatic resources were used to populate interactive presentations that served as a framework for annotation of Covid-19 relevant module sets. These resources include web applications deployed using Shiny R, which permit to plot transcript abundance patterns at the module and aggregate levels. Two of these applications were developed as part of a previous work establishing the blood transcriptome repertoire and applying it in the context of a meta-analysis of six public RSV datasets (12). A third application was developed as part of this work and can generate profiles at the transcript, module and module-aggregate levels for the Xiong *et al.* and Ong *et al.* datasets.

SUPPLEMENTAL INFORMATION

Supplemental File 1: Delineation of Covid-19 relevant modules sets in all 17 aggregates retained in the first step of the selection process.

REFERENCES

1. Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, et al. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19):A multi-center study in Wenzhou city, Zhejiang, China. J Infect. 2020;80(4):388–93.

2. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020 30;382(18):1708–20.

3. Long Q-X, Liu B-Z, Deng H-J, Wu G-C, Deng K, Chen Y-K, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med. 2020 Apr 29;

4. Henderson LA, Canna SW, Schulert GS, Volpi S, Lee PY, Kernan KF, et al. On the Alert for Cytokine Storm: Immunopathology in COVID-19. Arthritis Rheumatol Hoboken NJ. 2020 Apr 15;

5. Chaussabel D. Assessment of immune status using blood transcriptomics and potential implications for global health. Semin Immunol. 2015 Feb;27(1):58–66.

6. Zhou W, Altman RB. Data-driven human transcriptomic modules determined by independent component analysis. BMC Bioinformatics. 2018 Sep 17;19(1):327.

7. Altman MC, Rinchai D, Baldwin N, Toufiq M, Whalen E, Garand M, et al. Development and Characterization of a Fixed Repertoire of Blood Transcriptome Modules Based on Co-expression Patterns Across Immunological States. bioRxiv. 2020 Apr 28;525709.

8. Ong EZ, Chan YFZ, Leong WY, Lee NMY, Kalimuddin S, Haja Mohideen SM, et al. A Dynamic Immune Response Shapes COVID-19 Progression. Cell Host Microbe. 2020 Apr 30;

9. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect. 2020 Jan 1;9(1):761–70.

10. Perng Y-C, Lenschow DJ. ISG15 in antiviral immunity and beyond. Nat Rev Microbiol. 2018;16(7):423–39.

11. Gutiérrez L, Caballero N, Fernández-Calleja L, Karkoulia E, Strouboulis J. Regulation of GATA1 levels in erythropoiesis. IUBMB Life. 2020;72(1):89–105.

12. Rinchai D, Altman MB, Konza O, Hässler S, Martina F, Toufiq M, et al. Identification of erythroid cell positive blood transcriptome phenotypes associated with severe respiratory syncytial virus infection. bioRxiv. 2020 May 10;527812.

13. Elahi S. Neglected Cells: Immunomodulatory Roles of CD71+ Erythroid Cells. Trends Immunol. 2019 Mar;40(3):181–5.

14. Obermoser G, Presnell S, Domico K, Xu H, Wang Y, Anguiano E, et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. Immunity. 2013 Apr 18;38(4):831–44.

15. Roelands J, Garand M, Hinchcliff E, Ma Y, Shah P, Toufiq M, et al. Long-Chain Acyl-CoA Synthetase 1 Role in Sepsis and Immunity: Perspectives From a Parallel Review of Public Transcriptome Datasets and of the Literature. Front Immunol. 2019;10:2410.

16. Remy KE, Brakenridge SC, Francois B, Daix T, Deutschman CS, Monneret G, et al. Immunotherapies for COVID-19: lessons learned from sepsis. Lancet Respir Med [Internet]. 2020 Apr 28 [cited 2020 May 18];0(0). Available from: https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30217-4/abstract

17. Guo C, Li B, Ma H, Wang X, Cai P, Yu Q, et al. Tocilizumab treatment in severe COVID-19 patients attenuates the inflammatory storm incited by monocyte centric immune interactions revealed by single-cell analysis. bioRxiv. 2020 Apr 9;2020.04.08.029769.

18. Roumier M, Paule R, Groh M, Vallee A, Ackermann F. Interleukin-6 blockade for severe COVID-19. medRxiv. 2020 Apr 22;2020.04.20.20061861.

19. Friedman S, Sanyal A, Goodman Z, Lefebvre E, Gottwald M, Fischer L, et al. Efficacy and safety study of cenicriviroc for the treatment of non-alcoholic steatohepatitis in adult subjects with liver fibrosis: CENTAUR Phase 2b study design. Contemp Clin Trials. 2016 Mar 1;47:356–65.

20. Dogné J-M, Hanson J, de Leval X, Pratico D, Pace-Asciak CR, Drion P, et al. From the design to the clinical application of thromboxane modulators. Curr Pharm Des. 2006;12(8):903–23.

21. Goicoechea M, García de Vinuesa S, Quiroga B, Verdalles U, Barraca D, Yuste C, et al. Effects of pentoxifylline on inflammatory parameters in chronic kidney disease patients: a randomized trial. J Nephrol. 2012 Dec;25(6):969–75.

22. Kahn-Kirby AH, Amagata A, Maeder CI, Mei JJ, Sideris S, Kosaka Y, et al. Targeting ferroptosis: A novel therapeutic strategy for the treatment of mitochondrial disease-related epilepsy. PloS One. 2019;14(3):e0214250.

23. Stockwell BR, Friedmann Angeli JP, Bayir H, Bush AI, Conrad M, Dixon SJ, et al. Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. Cell. 2017 Oct 5;171(2):273–85.

24. Wang C, Yuan W, Hu A, Lin J, Xia Z, Yang CF, et al. Dexmedetomidine alleviated sepsis-induced myocardial ferroptosis and septic heart injury. Mol Med Rep. 2020 Jul;22(1):175–84.

25. Lang J, Yang N, Deng J, Liu K, Yang P, Zhang G, et al. Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. PloS One. 2011;6(8):e23710.

26. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci U S A. 2020 May 6;

27. Li S-W, Wang C-Y, Jou Y-J, Huang S-H, Hsiao L-H, Wan L, et al. SARS Coronavirus Papain-Like Protease Inhibits the TLR7 Signaling Pathway through Removing Lys63-Linked Polyubiquitination of TRAF3 and TRAF6. Int J Mol Sci. 2016 May 5;17(5).

28. Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P, Baric RS. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. J Virol. 2007 Sep;81(18):9812–24.

29. Li T, Wang P, Wang SC, Wang Y-F. Approaches Mediating Oxytocin Regulation of the Immune System. Front Immunol. 2016;7:693.

30. Earnest JT, Hantak MP, Li K, McCray PB, Perlman S, Gallagher T. The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. PLoS Pathog [Internet]. 2017 Jul 31 [cited 2020 May 20];13(7). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552337/

31. Karulf M, Kelly A, Weinberg AD, Gold JA. OX40 ligand regulates inflammation and mortality in the innate immune response to sepsis. J Immunol Baltim Md 1950. 2010 Oct 15;185(8):4856–62.

32. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020 Apr 28;1–12.

33. Azkur AK, Akdis M, Azkur D, Sokolowska M, van de Veen W, Brüggen M-C, et al. Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. Allergy. 2020 May 12;

34. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe. 2020 Apr 17;

35. Sweeney TE, Perumal TM, Henao R, Nichols M, Howrylak JA, Choi AM, et al. A community approach to mortality prediction in sepsis via gene expression analysis. Nat Commun. 2018 15;9(1):694.

36. Yan Q, Li P, Ye X, Huang X, Mo X, Wang Q, et al. Longitudinal peripheral blood transcriptional analysis of COVID-19 patients captures disease progression and reveals potential biomarkers. medRxiv. 2020 May 8;2020.05.05.20091355.

37. Speake C, Whalen E, Gersuk VH, Chaussabel D, Odegard JM, Greenbaum CJ. Longitudinal monitoring of gene expression in ultra-low-volume blood samples self-collected at home. Clin Exp Immunol. 2017;188(2):226–33.

38. Rinchai D, Anguiano E, Nguyen P, Chaussabel D. Finger stick blood collection for gene expression profiling and storage of tempus blood RNA tubes. F1000Research. 2016;5:1385.

39. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity. 2008 Jul 18;29(1):150–64.

40. Chaussabel D, Baldwin N. Democratizing systems immunology with modular transcriptional repertoire analyses. Nat Rev Immunol. 2014;14(4):271–80.

41. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinforma Oxf Engl. 2016 15;32(18):2847–9.

42. Hung IF-N, Lung K-C, Tso EY-K, Liu R, Chung TW-H, Chu M-Y, et al. Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. Lancet Lond Engl. 2020 May 8;

43. Malhotra S, Bustamante MF, Pérez-Miralles F, Rio J, Ruiz de Villa MC, Vegas E, et al. Search for specific biomarkers of IFNβ bioactivity in patients with multiple sclerosis. PloS One. 2011;6(8):e23634.

44. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data. 2018 27;5:180015.

45. Rinchai D, Konza O, Hassler S, Martina F, Mejias A, Ramilo O, et al. Characterizing blood modular transcriptional repertoire perturbations in patients with RSV infection: a hands-on workshop using public datasets as a source of training material. bioRxiv. 2019 Jan 22;527812.

46. Speake C, Presnell S, Domico K, Zeitner B, Bjork A, Anderson D, et al. An interactive web application for the dissemination of human systems immunology data. J Transl Med. 2015 Jun 19;13:196.

47. Linsley PS, Speake C, Whalen E, Chaussabel D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. PloS One. 2014;9(10):e109760.

48. Bougarn S, Boughorbel S, Chaussabel D, Marr N. A curated transcriptome dataset collection to investigate the blood transcriptional response to viral respiratory tract infection and vaccination. F1000Research. 2019;8:284.

49. Parnell G, McLean A, Booth D, Huang S, Nalos M, Tang B. Aberrant cell cycle and apoptotic changes characterise severe influenza A infection--a meta-analysis of genomic signatures in circulating leukocytes. PloS One. 2011 Mar 8;6(3):e17186.

50. Ayllon-Benitez A, Bourqui R, Thébault P, Mougin F. GSAn: an alternative to enrichment analysis for annotating gene sets. NAR Genomics Bioinforma [Internet]. 2020 Jun 1 [cited 2020 Apr 1];2(2). Available from: https://academic.oup.com/nargab/article/2/2/lqaa017/5805305

51. Ambade A, Lowe P, Kodys K, Catalano D, Gyongyosi B, Cho Y, et al. Pharmacological Inhibition of CCR2/5 Signaling Prevents and Reverses Alcohol-Induced Liver Damage, Steatosis, and Inflammation in Mice. Hepatol Baltim Md. 2019;69(3):1105–21.

52. Winkler DG, Faia KL, DiNitto JP, Ali JA, White KF, Brophy EE, et al. PI3K-δ and PI3K-γ inhibition by IPI-145 abrogates immune responses and suppresses activity in autoimmune and inflammatory disease models. Chem Biol. 2013 Nov 21;20(11):1364–74.

53. Wang S-M, Wang C-T. APOBEC3G cytidine deaminase association with coronavirus nucleocapsid protein. Virology. 2009 May 25;388(1):112–20.

54. Milewska A, Kindler E, Vkovski P, Zeglen S, Ochman M, Thiel V, et al. APOBEC3-mediated restriction of RNA virus replication. Sci Rep. 2018 13;8(1):5960.

55. de Lang A, Osterhaus ADME, Haagmans BL. Interferon-gamma and interleukin-4 downregulate expression of the SARS coronavirus receptor ACE2 in Vero E6 cells. Virology. 2006 Sep 30;353(2):474–81.

56. Sisk JM, Frieman MB, Machamer CE. Coronavirus S protein-induced fusion is blocked prior to hemifusion by Abl kinase inhibitors. J Gen Virol. 2018;99(5):619–30.

57. Yeung M-L, Yao Y, Jia L, Chan JFW, Chan K-H, Cheung K-F, et al. MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. Nat Microbiol. 2016 Feb 22;1:16004.

58. Li S-W, Wang C-Y, Jou Y-J, Yang T-C, Huang S-H, Wan L, et al. SARS coronavirus papain-like protease induces Egr-1-dependent up-regulation of TGF-β1 via ROS/p38 MAPK/STAT3 pathway. Sci Rep. 2016 13;6:25754.

59. Padhan K, Tanwar C, Hussain A, Hui PY, Lee MY, Cheung CY, et al. Severe acute respiratory syndrome coronavirus Orf3a protein interacts with caveolin. J Gen Virol. 2007 Nov;88(Pt 11):3067–77.

60. Lindner HA, Lytvyn V, Qi H, Lachance P, Ziomek E, Ménard R. Selectivity in ISG15 and ubiquitin recognition by the SARS coronavirus papain-like protease. Arch Biochem Biophys. 2007 Oct 1;466(1):8–14.

61. Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Burgstaller-Muehlbacher S, et al. A Large-scale Drug Repositioning Survey for SARS-CoV-2 Antivirals. bioRxiv. 2020 Apr 17;2020.04.16.044016.

62. Liu M, Yang Y, Gu C, Yue Y, Wu KK, Wu J, et al. Spike protein of SARS-CoV stimulates cyclooxygenase-2 expression via both calcium-dependent and calcium-independent protein kinase C pathways. FASEB J Off Publ Fed Am Soc Exp Biol. 2007 May;21(7):1586–96.

63. Shi C-S, Nabar NR, Huang N-N, Kehrl JH. SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. Cell Death Discov. 2019;5:101.

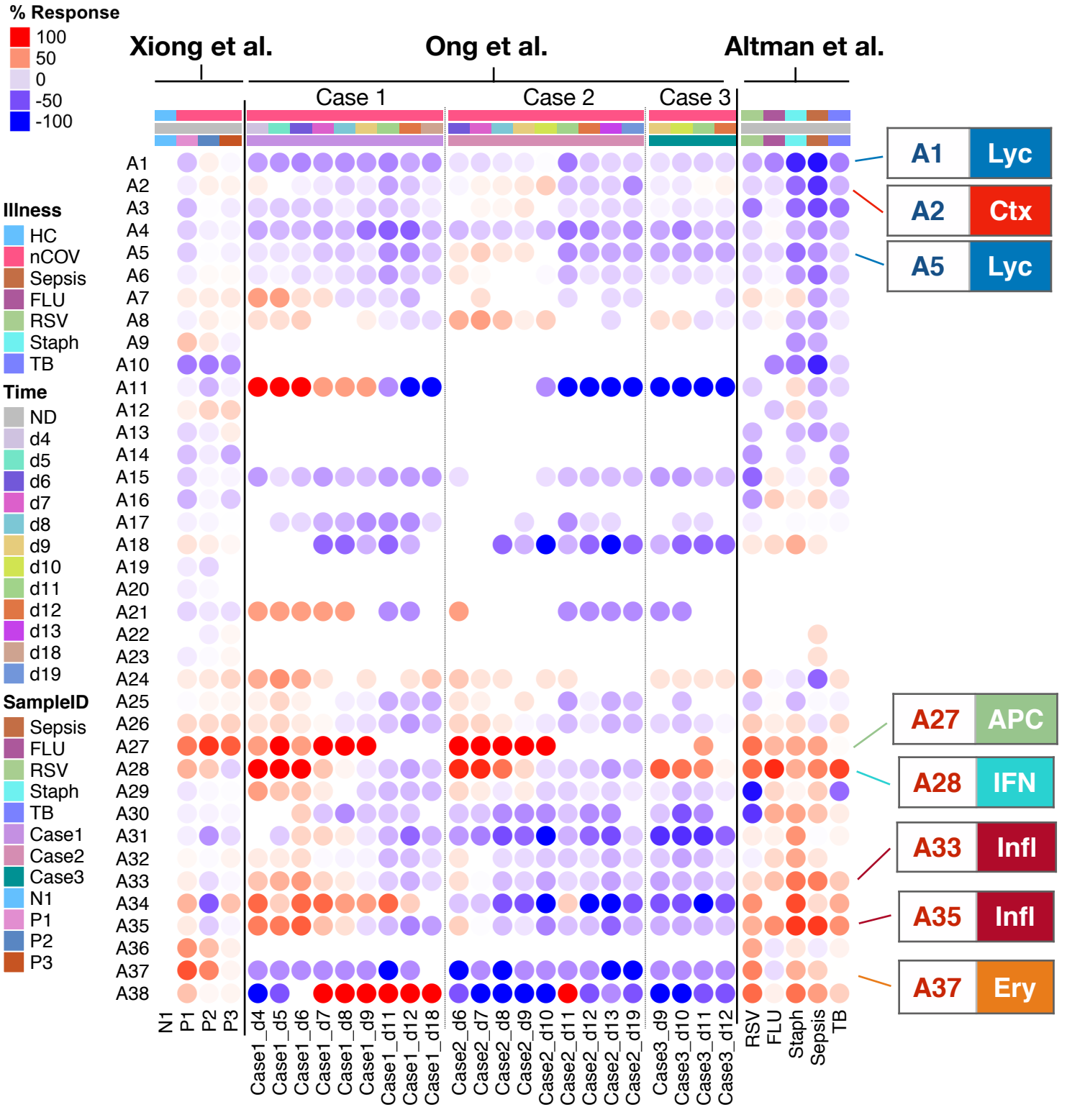
64. Siu K-L, Yuen K-S, Castaño-Rodriguez C, Ye Z-W, Yeung M-L, Fung S-Y, et al. Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. FASEB J Off Publ Fed Am Soc Exp Biol. 2019;33(8):8865–77.

65. Qi F, Qian S, Zhang S, Zhang Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. Biochem Biophys Res Commun. 2020 May 21;526(1):135–40.

66. Li D, Wu N, Yao H, Bader A, Brockmeyer NH, Altmeyer P. Association of RANTES with the replication of severe acute respiratory syndrome coronavirus in THP-1 cells. Eur J Med Res. 2005 Mar 29;10(3):117–20.

67. Follis KE, York J, Nunberg JH. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. Virology. 2006 Jul 5;350(2):358–69.

68. Zhou Y, Vedantham P, Lu K, Agudelo J, Carrion R, Nunneley JW, et al. Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. 2015 Apr;116:76–84.



Lyc Lymphocytic

Ctx Cytotoxic activity

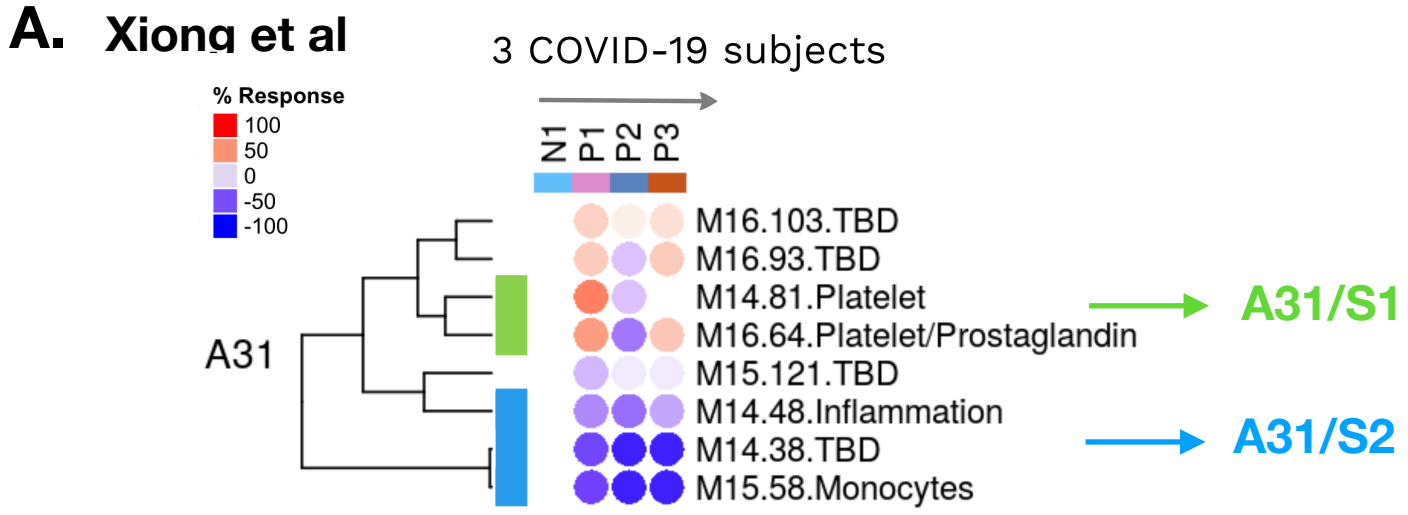
APC Antibody Producing Cells

IFN Interferon

Infl Inflammation

Ery Erythroid

Figure 1



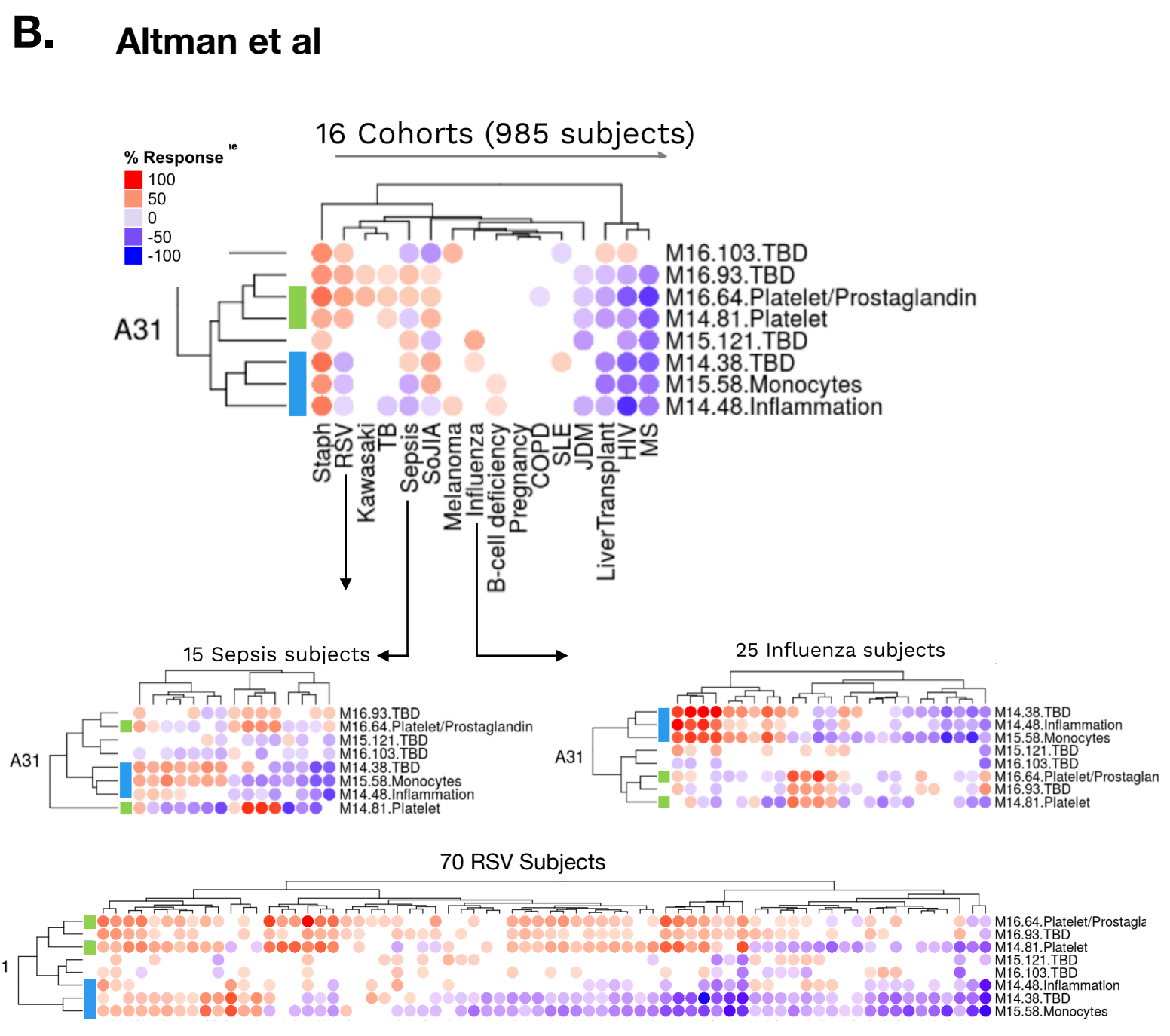


Figure 2

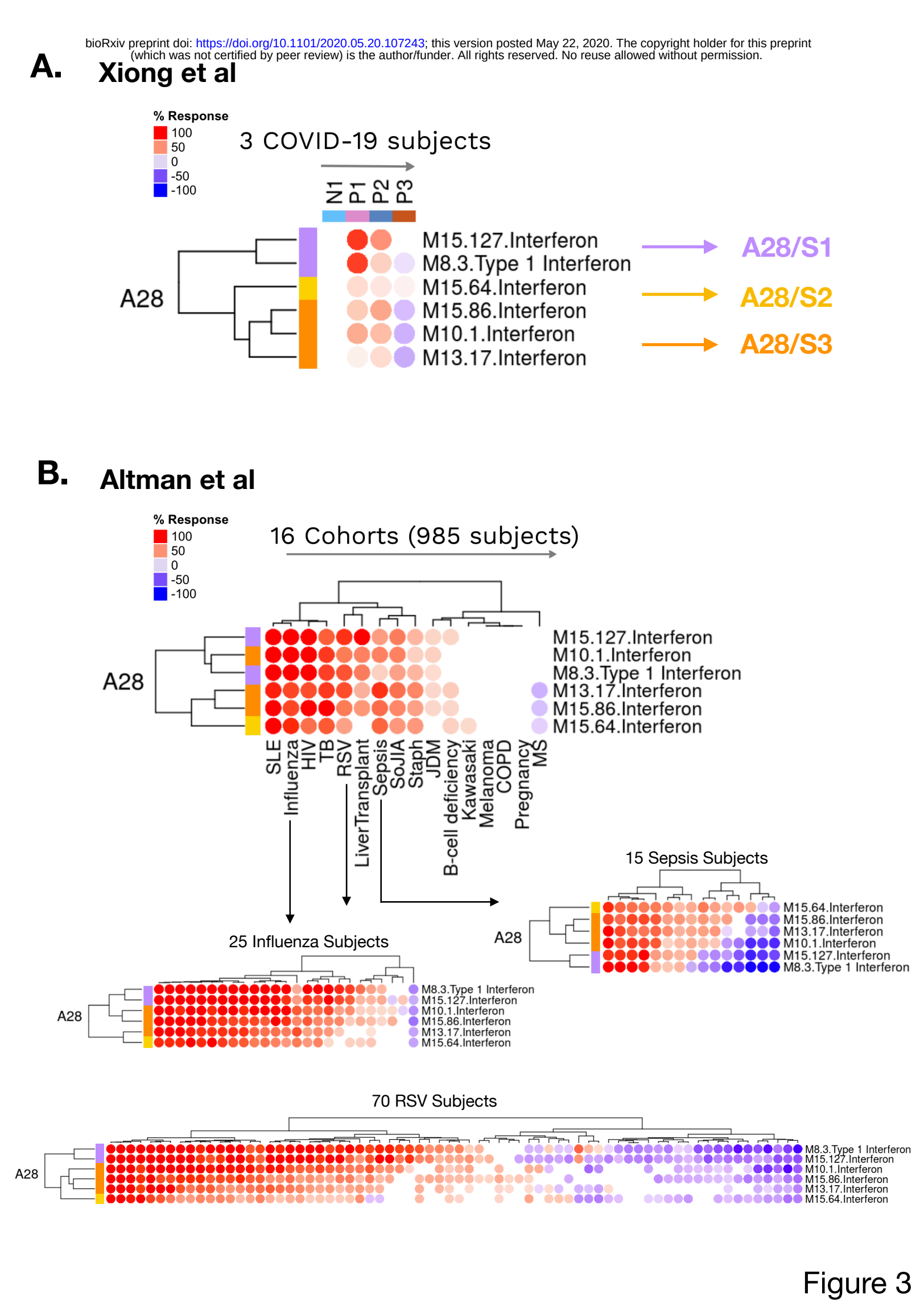


Figure 3

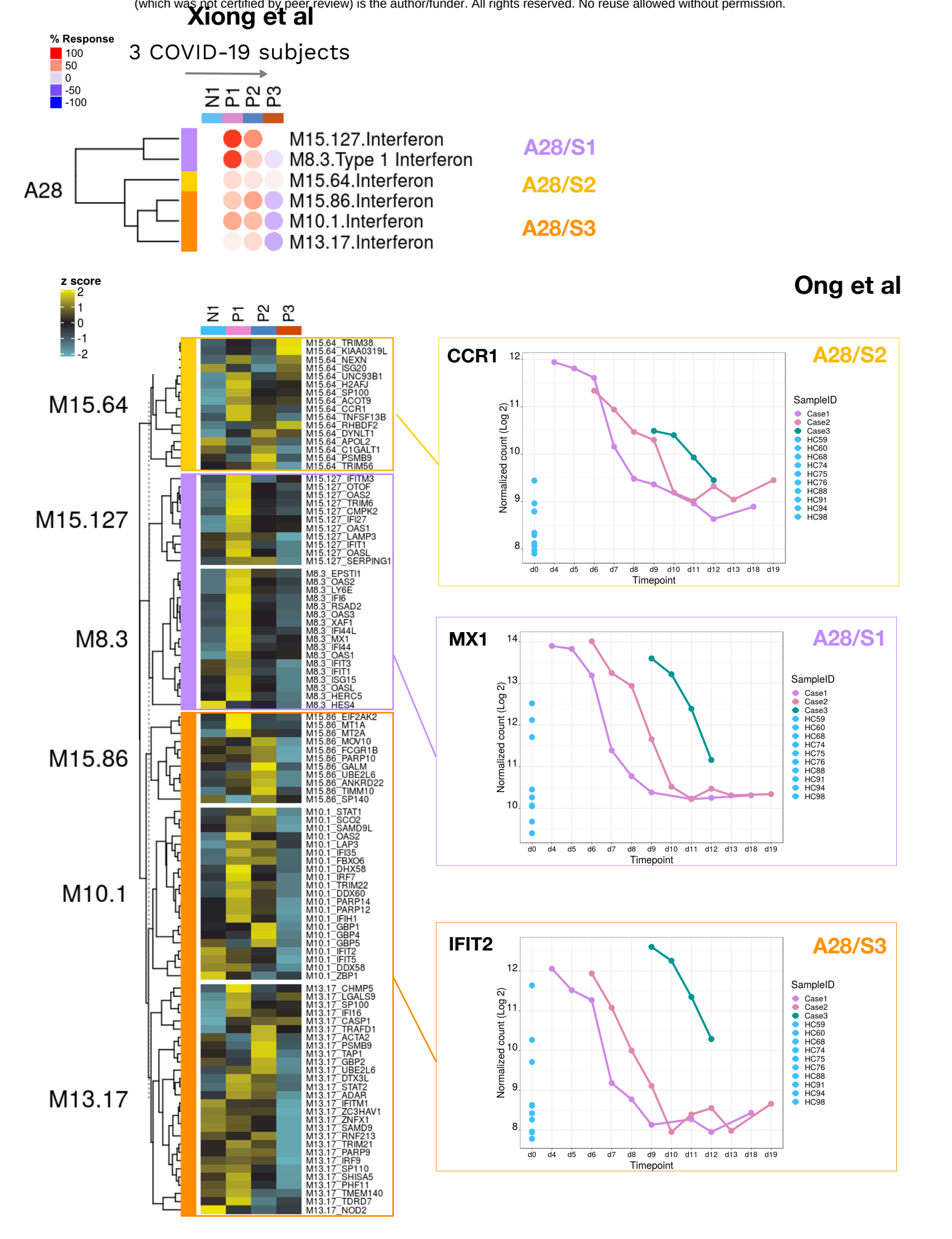
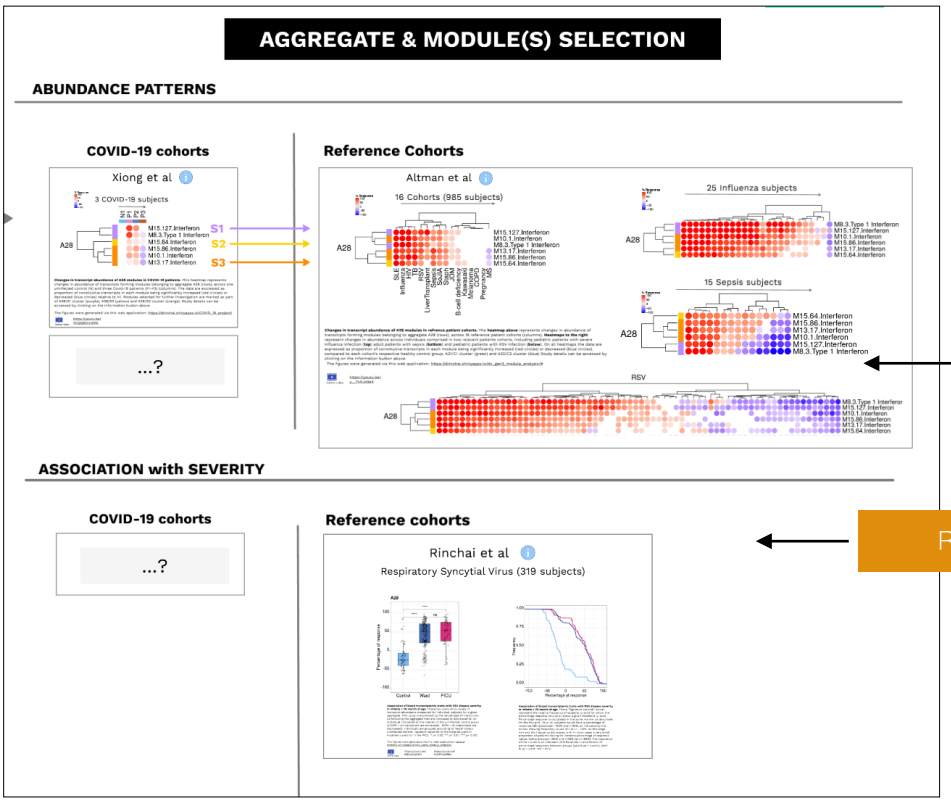


Figure 4

A.

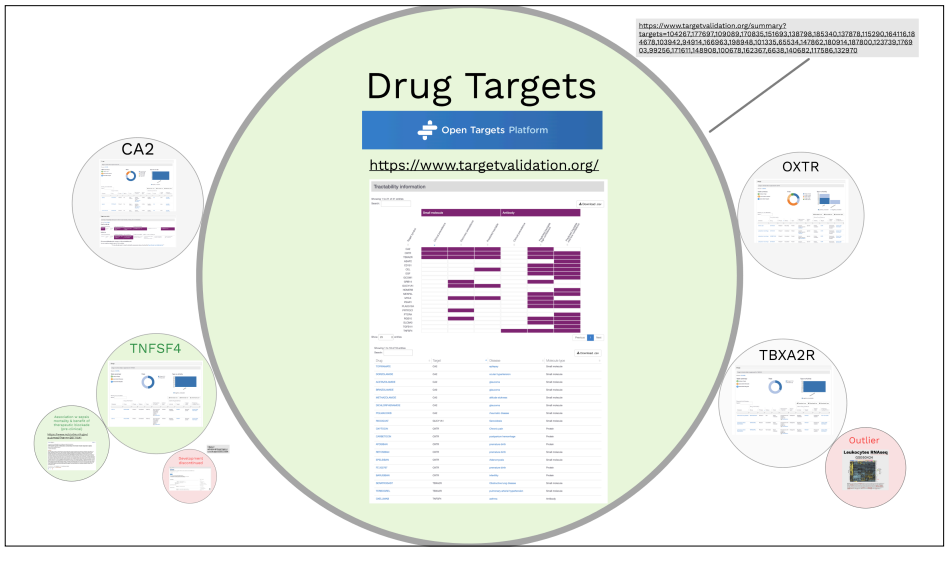


COVID-19

Modular fingerprinting assay

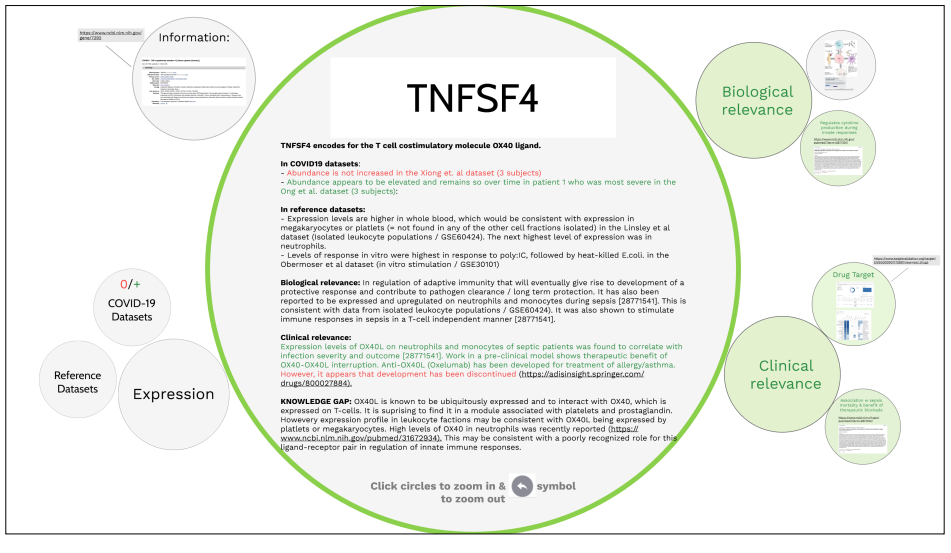
RSV Blood Transcriptome Fingerprints

B.



Open Targets Platform

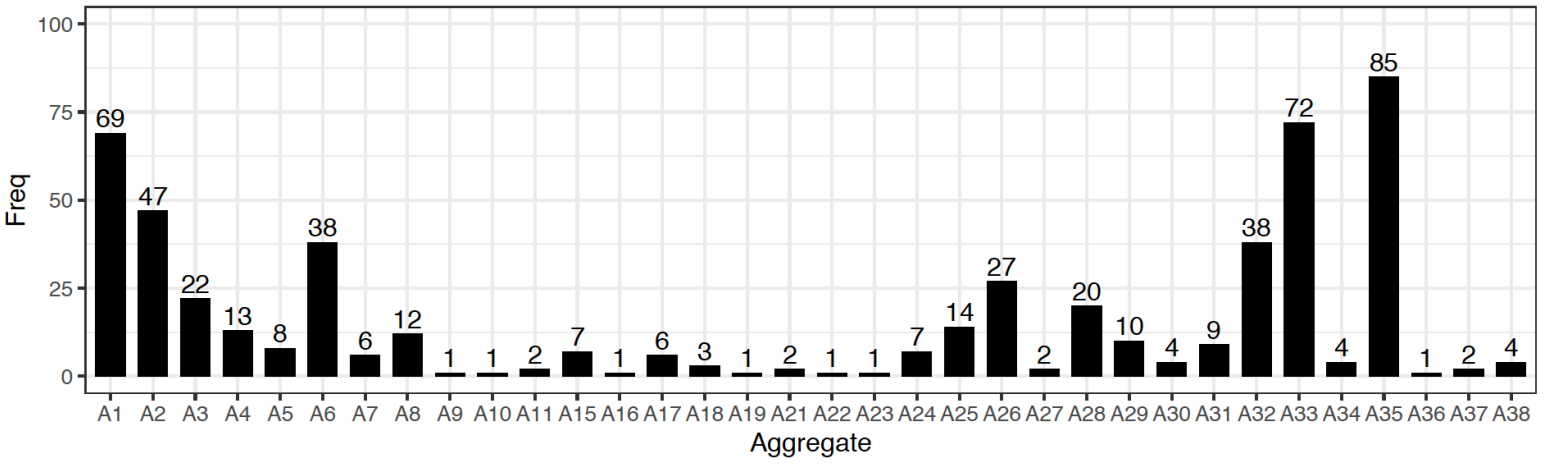
C.

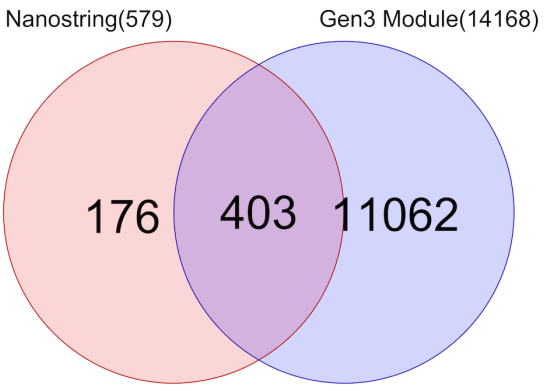


PubMed.gov

SIDRA GXB

Figure 5





Supplementary Figure 1