Consensus transcriptional regulatory networks of coronavirus-infected human cells

Scott A Ochsner[1], Rudolf T Pillich[2] and Neil J McKenna[1]

[1] The Signaling Pathways Project and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030; [2] Department of Medicine, University of California San Diego, La Jolla, CA 92093.

Address Correspondence To:

Neil J. McKenna

Department of Molecular and Cellular Biology

Baylor College of Medicine

Houston, TX 77030

USA

e: nmckenna@bcm.edu

t: 713-798-8568

Abstract

Establishing consensus around the transcriptional interface between coronavirus (CoV) infection and human cellular signaling pathways can catalyze the development of novel anti-CoV therapeutics. Here, we used publicly archived transcriptomic datasets to compute consensus regulatory signatures, or consensomes, that rank human genes based on their rates of differential expression in MERS-CoV (MERS), SARS-CoV-1 (SARS1) and SARS-CoV-2 (SARS2)-infected cells. Validating the CoV consensomes, we show that high confidence transcriptional targets (HCTs) of CoV infection intersect with HCTs of signaling pathway nodes with known roles in CoV infection. Among a series of novel use cases, we gather evidence for hypotheses that SARS2 infection efficiently represses E2F family target genes encoding key drivers of DNA replication and the cell cycle; that progesterone receptor signaling antagonizes SARS2-induced inflammatory signaling in the airway epithelium; and that SARS2 HCTs are enriched for genes involved in epithelial to mesenchymal transition. The CoV infection consensomes and HCT intersection analyses are freely accessible through the Signaling Pathways Project knowledgebase, and as Cytoscape-style networks in the Network Data Exchange repository.

Introduction

Infection of humans by coronaviruses (CoV) represents a major current global public health concern. Signaling within and between airway epithelial and immune cells in response to infections by CoV and other viruses is coordinated by a complex network of signaling pathway nodes. These include chemokine and cytokine-activated receptors, signaling enzymes and transcription factors, and the genomic targets encoding their downstream effectors[1–3]. Placing the transcriptional events resulting from CoV infection in context with those associated with host signaling paradigms has the potential to catalyze the development of novel therapeutic approaches. The CoV research community has been active in generating and archiving transcriptomic datasets documenting the transcriptional response of human cells to infection by the three major CoV strains, namely, Middle East respiratory syndrome coronavirus (MERS-CoV, or MERS) and severe acute respiratory syndrome coronaviruses 1 (SARS-CoV-1, or SARS1) and 2 (SARS-CoV-2, or SARS2)[4–9]. To date however the field has lacked a resource that fully capitalizes on these datasets by, firstly, using them to identify human genes that are most consistently transcriptionally responsive to CoV infection and secondly, contextualizing these transcriptional responses by integrating them with ‘omics data points relevant to host cellular signaling pathways.

We recently described the Signaling Pathways Project (SPP)[10], an integrated ‘omics knowledgebase designed to assist bench researchers in leveraging publically archived transcriptomic and ChIP-Seq datasets to generate research hypotheses. A unique aspect of SPP is its collection of consensus regulatory signatures, or consensomes, which rank genes based on the frequency of their significant differential expression

across transcriptomic experiments mapped to a specific signaling pathway node or node family. By surveying across multiple independent datasets, we have shown that consensomes recapitulate pathway node-genomic target regulatory relationships to a high confidence level[10]. Here, as a service to the research community to catalyze the development of novel CoV therapeutics, we generated consensomes for infection of human cells by MERS, SARS1 and SARS2 CoVs. Computing the CoV consensomes against those for a broad range of cellular signaling pathway nodes, we discovered robust intersections between genes with high rankings in the CoV consensomes and those of nodes with known roles in the response to CoV infection. Integration of the CoV consensomes with the existing universes of SPP transcriptomic and ChIP-Seq data points in a series of use cases illuminates previously uncharacterized interfaces between CoV infection and human cellular signaling pathways. Moreover, while this paper was under review and revision, numerous contemporaneous and independent wet bench-based studies came to light that corroborate *in silico* predictions made using our analysis pipeline. To reach the broadest possible audience of experimentalists, the results of our analysis were made available in the SPP website, as well as in the Network Data Exchange (NDEx) repository. Collectively, these networks constitute a unique and freely accessible framework within which to generate mechanistic hypotheses around the transcriptional interface between human signaling pathways and CoV infection.

Results

Generation of the CoV consensomes

We first set out to generate a set of consensomes[10] ranking human genes based on statistical measures of the frequency of their significant differential expression in response to infection by MERS, SARS1 and SARS2 CoVs. To do this we searched the Gene Expression Omnibus (GEO) and ArrayExpress databases to identify datasets involving infection of human cells by these strains. Many of these datasets emerged from a broad-scale systematic multi-omics Pacific Northwest National Library analysis of the host cellular response to infection across a broad range of pathogens[11]. Since an important question in the development of CoV therapeutics is the extent to which CoVs have common transcriptional impacts on human cell signaling that are distinct from those of other viruses, we also searched for transcriptomic datasets involving infection by human influenza A virus (IAV). From this initial collection of datasets, we next carried out a three step quality control check as previously described[10], yielding a total of 3.3 million data points in 156 experiments from 38 independent viral infection transcriptomic datasets (Supplementary file 1, section 1). Using these curated datasets, we next used consensome analysis (see Methods and previous SPP publication[10]) to generate consensomes for each CoV strain. Supplementary file 1 contains the full human SARS1 (Section 2), SARS2 (Section 3), MERS (Section 4) and IAV (Section 5) infection transcriptomic consensomes. To assist researchers in inferring CoV infection-associated signaling networks, the consensomes are annotated using the previously described SPP convention[10] to indicate the identity of a gene as encoding a receptor,

protein ligand, enzyme, transcription factor, ion channel or co-node (Supplementary file 1, sections 2-5, columns A-C).

Consistent with the known effect of viral infection on the interferon response, the viral consensomes assign elevated rankings to many canonical interferon-stimulated genes (ISGs), which encode many of the key canonical viral response factors (Schneider et al., 2014). Moreover, other genes with known critical roles in the response to viral infection have high rankings in the CoV consensomes, including *NCOA7*[12] (percentiles: SARS1, 98th; SARS2, 97th; MERS, 89th; IAV, 99th), *STAT1*[13] (percentiles: SARS1, 99th; SARS2, 98th; MERS, 89th; IAV, 99th) and *TAP1*[14] (percentiles: SARS1, 99th; SARS2, 94th; MERS, 83rd; IAV, 99th). In addition to the appropriate elevated rankings for these known viral response effectors, the CoV consensomes assign similarly elevated rankings to transcripts that are largely or completely uncharacterized in the context of viral infection. Examples of such genes include *PSMB9*, encoding a proteasome 20S subunit (percentiles: SARS1, 98th; SARS2, 97th; MERS, 98th; IAV, 98th); *CSRNP1*, encoding a cysteine and serine rich nuclear protein (percentiles: SARS1, 99th; SARS2, 94th; MERS, 98th; IAV, 94th); and *CCNL1*, encoding a member of the cell cycle-regulatory cyclin family (percentiles: SARS1, 99th; SARS2, 94th; MERS, 99th; IAV, 97th),

Validation of the CoV consensomes

To illuminate the transcriptional interface between CoV infection and human cellular signaling pathways, we next computed genes with elevated rankings in the CoV consensomes against genes that have predicted high confidence regulatory relationships with cellular signaling pathway nodes. To do this we generated four lists of genes corresponding to the MERS, SARS1, SARS2 and IAV transcriptomic

consensome 95th percentiles. Next we retrieved genes in the 95th percentiles of available SPP human transcriptomic (n = 25) consensomes and ChIP-Seq (n = 864) pathway node consensomes[10]. For convenience we will refer from hereon to genes in the 95th percentile of a viral infection, node (ChIP-Seq) or node family (transcriptomic) consensome as high confidence transcriptional targets (HCTs). We then used the R GeneOverlap package[15] to compute the extent and significance of intersections between CoV HCTs and those of the pathway nodes or node families. We interpreted the extent and significance of intersections between HCTs for CoVs and pathway node or node families as evidence for a biological relationship between loss or gain of function of that node (or node family) and the transcriptional response to infection by a specific virus.

Results of viral infection and signaling node HCT intersection analyses are shown in Figure 1 (based on receptor and enzyme family transcriptomic consensomes), Figures 2 and 3 (based on ChIP-Seq consensomes for transcription factors and enzymes, respectively) and Supplementary file 2 (based on ChIP-Seq consensomes for selected co-nodes). Supplementary file 1, sections 6 (node family transcriptomic HCT intersection analysis) and 7 (node ChIP-Seq HCT intersection analysis) contain the full underlying numerical data. We surveyed $q < 0.05$ HCT intersections to identify (i) canonical inflammatory signaling pathway nodes with characterized roles in the response to CoV infection, thereby validating the consensome approach in this context; and (ii) evidence for nodes whose role in the transcriptional biology of CoV infection is previously uncharacterized, but consistent with their roles in the response to other viral

infections. In the following sections all *q*-values refer to those obtained using the GeneOverlap analysis package in R[15].

Receptors Reflecting their well-documented roles in the response to CoV infection[16–19], we observed appreciable significant intersections between CoV HCTs and those of the toll-like (TLRs; q-values: SARS1, 3e-85; SARS2, 5e-49; MERS, 2e-33), interferon (IFNR; q-values: SARS1, 1e-109; SARS2, 6e-53; MERS, 1e-24) and tumor necrosis factor (TNFR; q-values: SARS1, 1e-48; SARS2, 1e-35; MERS, 5e-32) receptor families (Fig. 1). HCT intersections between CoV infection and receptor systems with previously uncharacterized connections to CoV infection, including epidermal growth factor receptors (EGFR; q-values: SARS1, 4e-21; SARS2, 3e-48; MERS, 1e-35), and Notch receptor signaling (q-values: SARS1, 6e-24; SARS2, 2e-33; MERS, 2e-29; Fig. 1), are consistent with their known role in the context of other viral infections[20–24]. The Notch receptor HCT intersection points to a possible mechanistic basis for the potential of Notch pathway modulation in the treatment of SARS2[25]. The strong HCT intersection between CoV infection and xenobiotic receptors (q-values: SARS1, 1e-30; SARS2, 1e-44; MERS, 5e-32; Fig. 1) reflects work describing a role for pregnane X receptor in innate immunity[26] and points to a potential role for members of this family in the response to CoV infection. In addition, the robust intersection between HCTs for SARS2 infection and vitamin D receptor (q = 2e-35) is interesting in light of epidemiological studies suggesting a link between risk of SARS2 infection and vitamin D deficiency[27,28]. Consistent with a robust signature for the glucocorticoid receptor across all CoVs (GR; q-values: SARS1, 3e-35; SARS2, 1e-35; MERS, 7e-32), while this paper was under review, studies were published showing the GR agonist dexamethasone was a

successful therapeutic for SARS2 infection[29]. Finally, and also while this paper was under review, in vitro analyses confirmed our predictions of the modulation by SARS2 infection of ErbB/EGFR[18,30] and TGFBR[30] signaling systems (Fig. 1).

Transcription factors Not unexpectedly – and speaking again to validation of the consensomes - the strongest and most significant CoV HCT intersections were observed for HCTs for known transcription factor mediators of the transcriptional response to CoV infection, including members of the NFκB (q-value ranges: SARS1, 1e-7-1e-9; SARS2, 9e-3-2e-3; MERS, 1e-3-1e-4)[31–33], IRF (q-value ranges: SARS1, 2e-2-1e-31; SARS2, 2e-4-1e-17; MERS, 9e-4-7e-5)[34] and STAT (q-value ranges: SARS1, 1e-7-1e-55; SARS2, 2e-3-3e-29; MERS, 5e-2-3e-5)[35–37] transcription factor families (Fig. 2). Corroborating our finding of a strong intersection between STAT2 and SARS2 infection HCTs (*q* = 3e-29), a study that appeared while this manuscript was under review showed that STAT2 plays a prominent role in the response to SARS2 infection of Syrian hamsters[38]. HCT intersections for nodes originally characterized as having a general role in RNA Pol II transcription, including TBP (q-values: SARS1, 2e-10; SARS2, 6e-23; MERS, 3e-16), GTF2B/TFIIB (q-values: SARS1, 7e-10; SARS2, 3e-23; MERS, 9e-14) and GTF2F1 (q-values: SARS1, 2e-4; SARS2, 2e-13; MERS, 5e-5) were strong across all CoVs, and particularly noteworthy in the case of SARS2. In the case of GTF2B, these data are consistent with previous evidence identifying it as a specific target for orthomyxovirus[39], and the herpes simplex[40] and hepatitis B[41] viruses. Moreover, a proteomic analysis that appeared in BioRXiv while this paper was under review identified a high confidence interaction between GTF2F2 and the SARS2 NSP9 replicase[30].

In general, intersections between viral infection and ChIP-Seq enrichments for transcription factors and other nodes were more specific for individual CoV infection HCTs (compare Fig. 1 with Figs. 2 & 3 and Supplementary file 1, sections 6 and 7). This is likely due to the fact that ChIP-Seq consensomes are based on direct promoter binding by a specific node antigen, whereas transcriptomic consensomes encompass both direct and indirect targets of specific receptor and enzyme node families.

Enzymes Compared to the roles of receptors and transcription factors in the response to viral infection, the roles of signaling enzymes are less well illuminated – indeed, in the context of CoV infection, they are entirely unstudied. Through their regulation of cell cycle transitions, cyclin-dependent kinases (CDKs) play important roles in the orchestration of DNA replication and cell division, processes that are critical in the viral life cycle. CDK6, which has been suggested to be a critical G1 phase kinase[42,43], has been shown to be targeted by a number of viral infections, including Kaposi's sarcoma-associated herpesvirus[44] and HIV-1[45]. Consistent with this common role across distinct viral infections, we observed robust intersection between the CDK family HCTs (q-values: SARS1, 8e-23; SARS2, 2e-31; MERS, 1e-30; Fig. 1) and the CDK6 HCTs (q-values: SARS1, 1e-7; SARS2, 8e-8; MERS, 3e-4; Fig. 3) and those of all viral HCTs. As with the TLRs, IFNRs and TNFRs, which are known to signal through CDK6[46–48], intersection with the CDK6 HCTs was particularly strong in the case of the SARS2 HCTs (Fig. 3). Again, the subsequent proteomic analysis we alluded to earlier[30] independently corroborated our prediction of a role for CDK6 in the response to SARS2 infection.

CCNT2 is another member of the cyclin family that, along with CDK9, is a component of the viral-targeted p-TEFB complex[49]. Reflecting a potential general role in viral infection, appreciable intersections were observed between the CCNT2 HCTs and all viral HCTs (q-values: SARS1, 4e-4; SARS2, 6e-3; MERS, 7e-5; Fig. 3). Finally in the context of enzymes, the DNA topoisomerases have been shown to be required for efficient replication of simian virus 40[50] and Ebola[51] viruses. The prominent intersections between DNA topoisomerase-dependent HCTs and the CoV HCTs (q-values: SARS1, 3e-15; SARS2, 6e-21; MERS, 1e-26; Fig. 3) suggest that it may play a similar role in facilitating the replication of these CoVs.

Hypothesis generation use cases

We next wished to show how the CoV consensomes and HCT intersection networks, supported by existing canonical literature knowledge, enable the user to generate novel hypotheses around the transcriptional interface between CoV infection and human cellular signaling pathways. Given the current interest in SARS2, we have focused our use cases on that virus. In addition to these use cases, Supplementary file 2 contains a number of additional use cases omitted from the main text due to space constraints. Unless otherwise stated, all *q*-values below were obtained using the GeneOverlap analysis package in R[15]. We stress that all use cases represent preliminary *in silico* evidence only, and require rigorous pressure-testing at the bench for full validation.

Hypothesis generation use case 1: transcriptional regulation of the SARS2 receptor gene, *ACE2*

ACE2, encoding membrane-bound angiotensin converting enzyme 2, has gained prominence as the target for cellular entry by SARS1[52] and SARS2[53]. An important component in the development of ACE2-centric therapeutic responses is an understanding of its transcriptional responsiveness to CoV infection. Interestingly, based on our CoV consensome analysis, *ACE2* is more consistently transcriptionally responsive to infection by SARS CoVs (SARS1: 98th percentile, consensome *q* value (CQV)[10] = 1e-25; SARS2: 97th percentile, CQV = 4e-7) than by IAV (78th percentile, CQV = 3e-8) or MERS (49th percentile, CQV = 2e-16; Supplementary file 1, sections 2-5). The data points underlying the CoV consensomes indicate evidence for tissue-specific differences in the nature of the regulatory relationship between *ACE2* and viral infection. In response to SARS1 infection, for example, *ACE2* is induced in pulmonary cells but repressed in kidney cells (Fig. 4). On the other hand, in response to SARS2 infection, *ACE2* is repressed in pulmonary cells - a finding corroborated by other studies [54,55] - but inducible in gastrointestinal cells (Fig. 4). These data may relate to the selective transcriptional response of *ACE2* to signaling by IFNRs (92nd percentile; Supplementary file 1, section 8) rather than TLRs (48th percentile; Supplementary file 1, section 9) or TNFRs (13th percentile, Supplementary file 1, section 10). While this manuscript was under review, another study appeared confirming repression of induction of *ACE2* by interferon stimulation and by IAV infection[56]. Our data reflect a complex transcriptional relationship between *ACE2* and viral infection that may be illuminated in part by future single cell RNA-Seq analysis in the context of clinical or animal models of SARS2 infection.

Hypothesis generation use case 2: evidence for antagonistic cross-talk between progesterone receptor and interferon receptor signaling in the airway epithelium

A lack of clinical data has so far prevented a definitive evaluation of the connection between pregnancy and susceptibility to SARS2 infection in CoVID-19. That said, SARS2 infection is associated with an increased incidence of pre-term deliveries[57], and pregnancy has been previously associated with the incidence of viral infectious diseases, particularly respiratory infections[58,59]. We were therefore interested to observe consistent intersections between the progesterone receptor (PGR) HCTs and CoV infection HCTs (q-values: SARS1, 3e-35; SARS2, 5e-41; MERS 5e-28), with the intersection being particularly evident in the case of the SARS2 HCTs (Fig. 1; Supplementary file 1, section 6). To investigate the specific nature of the crosstalk implied by this transcriptional intersection in the context of the airway epithelium, we first identified a set of 12 genes that were HCTs for both SARS2 infection and PGR. Interestingly, many of these genes encode members of the classic interferon-stimulated gene (ISG) response pathway[60]. We then retrieved two SPP experiments involving treatment of A549 airway epithelial cells with the PGR full antagonist RU486 (RU), alone or in combination with the GR agonist dexamethasone (DEX). As shown in Figure 5, there was unanimous correlation in the direction of regulation of all 12 genes in response to CoV infection and PGR loss of function. These data are consistent with the reported pro-inflammatory effects of RU486 in a mouse model of allergic pulmonary inflammation[61]. Interestingly, SARS2-infected pregnant women are often asymptomatic[62,63]. Based on our data, it can be reasonably hypothesized that suppression of the interferon response to SARS2 infection by elevated circulating

progesterone during pregnancy may contribute to the asymptomatic clinical course. Indeed, crosstalk between progesterone and inflammatory signaling is well characterized in the reproductive system, most notably in the establishment of uterine receptivity[64] as well as in ovulation[65]. Consistent with our hypothesis, while this paper was under review, a clinical trial was launched to evaluate the potential of progesterone for treatment of COVID-19 in hospitalized men[66]. Interestingly, and also while this paper was under review, a paper appeared showing that progesterone inhibited SARS2 replication in African green monkey kidney Vero 6 cells[67]. These results indicate an additional mechanism, distinct from its potential crosstalk with the interferon response, by which progesterone signaling may impact SARS2 infection.

Hypothesis generation use case 3: association of an epithelial to mesenchymal transition transcriptional signature with SARS2 infection

Epithelial to mesenchymal transition (EMT) is the process by which epithelial cells lose their polarity and adhesive properties and acquire the migratory and invasive characteristics of mesenchymal stem cells[68]. EMT is known to contribute to pulmonary fibrosis[69], acute interstitial pneumonia[70] and acute respiratory distress syndrome (ARDs)[71], all of which have been reported in connection with SARS2 infection in COVID-19[72–74]. We were interested to note therefore that significant HCT intersections for three well characterized EMT-promoting transcription factors were specific to SARS2 infection (q-values: SNAI2/Slug[75], 2e-2; EPAS1/HIF2α[76], 9e-9; LEF1[77], 1e-3; Fig. 2, bold symbols; Supplementary file 1, section 7). Consistent with this, intersections between HCTs for TGFBRs, SMAD2 and SMAD3, known regulators of EMT transcriptional programs[78] – were stronger with HCTs for SARS2 (q-values: TGFBRs, 2e-31; SMAD2,

2e-7; SMAD3, 5e-17) than with those of SARS1 (q-values: TGFBRs, 6e-29; SMAD2, 2e-2; SMAD3, 3e-9) and MERS (q-values: TGFBRs, 1e-16; SMAD2, 3e-3; SMAD3, 2e-12) – see also Figs. 1 and 2 and Supplementary file 1, sections 6 and 7).

To investigate the connection between SARS2 infection and EMT implied by these HCT intersections, we then computed intersections between the individual viral HCTs and a list of 335 genes manually curated from the research literature as EMT markers[79] (Supplementary file 1, section 11). In agreement with the HCT intersection analysis, we observed significant enrichment of members of this gene set within the SARS2 HCTs (*q* = 4e-14), but not the SARS1 or MERS (both *q* = 2e-1) HCTs (Fig. 6A). Consistent with previous reports of a potential link between EMT and IAV infection[80], we observed significant intersection between the EMT signature and the IAV HCTs (*q* = 1e-04).

One possible explanation for the selective intersection between the literature EMT signature and the SARS2 HCTs relative to SARS1 and MERS was the fact that the SARS2 consensome was exclusively comprised of epithelial cell lines, whereas the SARS1 and MERS consensomes included non-epithelial cell biosamples (Supplementary file 1, section 1). To exclude this possibility therefore, we next calculated airway epithelial cell-specific consensomes for SARS1, SARS2 and MERS and computed intersections between their HCTs and the EMT signature. We found that significant intersection of the EMT signature with the CoV HCTs remained specific to SARS2 (q-values: SARS1, 2e-1; SARS2, 1e-8; MERS, 2e-1) in the lung epithelium-specific CoV consensomes.

We next retrieved the canonical EMT genes in the SARS2 HCTs and compared their percentile rankings with the other CoV consensomes. Although some EMT genes, such as *CXCL2* and *IRF9*, had elevated rankings across all four viral consensomes, the collective EMT gene signature had a significantly higher mean percentile value in the SARS2 consensome than in each of the other viral consensomes (Fig. 6B; SARS2 mean percentile = 97.5; SARS1 mean percentile = 86, p = 1e-5, t-test; MERS mean percentile = 63, p = 1e-9, t-test; IAV mean percentile = 76, p = 2e-7, t-test). A column named “EMT” in Supplementary file 1, sections 2 (SARS1), 3 (SARS2), 4 (MERS) and 5 (IAV) identifies the ranking of the EMT genes in each of the viral consensomes.

Given that EMT has been linked to ARDs[71], we speculated that the evidence connecting EMT and SARS2 acquired through our analysis might be reflected in the relatively strong intersection between ARDs markers in SARS2 HCTs compared to other viral HCTs. To test this hypothesis we carried out a PubMed search to identify a set of 88 expression biomarkers of ARDs or its associated pathology, acute lung injury (ALI). A column named “ALI/ARDs” in Supplementary file 1, sections 2 (SARS1), 3 (SARS2) 4 (MERS) and 5 (IAV) identifies the expression biomarker genes using the PubMed identifiers for the original studies in which they were identified. Consistent with our hypothesis, we observed appreciable intersections between this gene set and the HCTs of all four viruses (SARS1 odds ratio (OR) = 7, q = 5e-9; SARS2 OR = 10.4, q = 1e-9; MERS, OR = 4.2, q = 2e-5; IAV OR = 6.8; q = 9e-8) with a particularly strong intersection evident in the SARS2 HCTs.

Although EMT has been associated with infection by transmissible gastroenteritis virus[81] and IAV[80], this is to our knowledge the first evidence connecting CoV infection, and

specifically SARS2 infection, to an EMT signature. Interestingly, lipotoxin A4 has been shown to attenuate lipopolysaccharide-induced lung injury by reducing EMT[82]. Moreover, several members of the group of SARS2-induced EMT genes have been associated with signature pulmonary comorbidities of CoV infection, including *ADAR*[83], *CLDN1*[84] and *SOD2*[85]. Of note in the context of these data is the fact that signaling through two SARS2 cellular receptors, ACE2/AT2 and CD147/basigin, has been linked to EMT in the context of organ fibrosis[86–88]. Finally, while this manuscript was under review, a preprint was posted that described EMT-like transcriptional and metabolic changes in response to SARS2 infection[89]. Collectively, our data indicate that EMT warrants further investigation as a SARS2-specific pathological mechanism.

Hypothesis generation use case 4: SARS2 repression of E2F family HCTs encoding cell cycle regulators

Aside from EPAS1 and SNAI2, the only other transcription factors with significant HCT intersections that were specific to the SARS2 HCTs were the E2F/FOX class members E2F1 (q-values: SARS1, 5e-1; SARS2, 1e-2; MERS, 4e-1), E2F3 (q-values: SARS1, 6e-1; SARS2, 5e-2; MERS, 7e-1), E2F4 (q-values: SARS1, 1; SARS2, 9e-3; MERS, 1) and TFDP1/Dp-1 (q-values: SARS1, 1; SARS2, 3e-4; MERS, 1; Fig. 2, bold symbols; Supplementary file 1, section 7). These factors play well-documented interdependent roles in the promotion (E2F1, E2F3, TFDP1) and repression (E2F4) of cell cycle genes[90,91]. Moreover, E2F family members are targets of signaling through EGFRs[92] and CDK6[93], both of whose HCTs had SARS2 HCT intersections that were stronger those of the other CoVs (EGFRs: q-values: SARS1, 4e-21; SARS2, 3e-48; MERS, 1e-35; CDK6: q-values: SARS1, 1e-7; SARS2, 8e-8; MERS, 2e-4); Figs. 1 & 3). Based on

these data, we speculated that SARS2 infection might impact the expression of E2F-regulated cell cycle genes more efficiently than other CoVs. To investigate this we retrieved a set of SARS2 HCTs that were also HCTs for at least three of E2F1, E2F3, E2F4 and TFDP1 (Supplementary file 1, section 3, columns P-T). Consistent with the role of E2F/Dp-1 nodes in the regulation of the cell cycle, many of these genes – notably *CDK1*, *PCNA*, *CDC6*, *CENPF* and *NUSAP1* – are critical positive regulators of DNA replication and cell cycle progression[94–98] and are known to be transcriptionally induced by E2Fs[99–102]. Strikingly, with the exception of *E2F3*, all were consistently repressed in response to SARS2 infection (Fig. 7A). To gain insight into the relative efficiency with which the four viruses impacted expression of the E2F/Dp-1 HCT signature, we compared their mean percentile values across the viral consensomes. Consistent with efficient repression of the E2F/Dp-1 HCTs by SARS2 infection relative to other viruses, their mean percentile ranking was appreciably higher in the SARS2 consensome (97th percentile) than in the SARS1 (76th percentile; p = 6e-12, t-test), MERS (71.2 percentile; p = 9e-6, t-test) and IAV (71.2 percentile; p = 2e-5, t-test) consensomes (Fig. 7B). Although manipulation of the host cell cycle and evasion of detection through deregulation of cell cycle checkpoints has been described for other viruses[103–105], this represents the first evidence for the profound impact of SARS2 infection on host cell cycle regulatory genes, potentially through disruption of E2F mediated signaling pathways. Mechanistically, this may be related to the fact that single gene searches on the SPP website (see YouTube video (http://tiny.cc/2i56rz) for instructions on running a single gene search) indicate consistent transcriptional repression of *E2F1* and *E2F2*, and induction of *E2F4*, in response to SARS2 infection.

The SARS2 infection-mediated induction of *E2F3* (Fig. 7A) may represent a compensatory response to transcriptional repression of other E2F family members, as has been previously observed for this family in other contexts[106,107]. Consistent with our prediction in this use case, while this paper was in revision, a study appeared showing that infection by SARS2 results in cell cycle arrest[108]. Our results represent evidence that efficient modulation by SARS2 of E2F signaling, resulting in repression of cell cycle regulatory genes, may contribute to its unique pathological impact.

Visualization of the CoV transcriptional regulatory networks in the Signaling Pathways Project knowledgebase and Network Data Exchange repository

To enable researchers to routinely generate mechanistic hypotheses around the interface between CoV infection human cell signaling, we next made the consensomes and accompanying HCT intersection analyses freely available to the research community in the SPP knowledgebase and the Network Data Exchange (NDEx) repository. Table 1 contains digital object identifier (DOI)-driven links to the consensome networks in SPP and NDEx, and to the HCT intersection networks in NDEx.

We have previously described the SPP biocuration pipeline, database and web application interface[10]. Figure 8 shows the strategy for consensome data mining on the SPP website. The individual CoV consensomes can be accessed by configuring the SPP Ominer query form as shown, in this example for the SARS2 consensome (Fig. 8A). Figure 8B shows the layout of the consensomes, showing gene symbol, name, percentile ranking and other essential information. Genes in the 90th percentile of each consensome are accessible via the user interface, with the full consensomes available

for download in a tab delimited text file. Target gene symbols in the consensome link to the SPP Regulation Report, filtered to show only experimental data points that contributed to that specific consensome (Fig. 8C). This view gives insights into the influence of tissue and cell type context on the regulatory relationship. These filtered reports can be readily converted to default Reports that show evidence for regulation of a specific gene by other signaling pathway nodes. As previously described, pop-up windows in the Report provide experimental details, in addition to links to the parent dataset (Fig. 8D), curated accordingly to our previously described protocol[10]. Per FAIR data best practice, CoV infection datasets – like all SPP datasets – are associated with detailed descriptions, assigned a DOI, and linked to the associated article to place the dataset in its original experimental context (Fig. 8D). The full list of datasets is available for browsing in the SPP Dataset listing (https://www.signalingpathways.org/index.jsf).

The NDEx repository facilitates collaborative publication of biological networks, as well as visualization of these networks in web or desktop versions of the popular and intuitive Cytoscape platform[109–111]. Figure 9 shows examples of consensome and HCT intersection network visualizations within the NDEx user interface. For ease of viewing, the initial rendering of the full SARS2 (Fig. 10a) and other consensome networks shows a sample (Fig. 10a, red arrow 1) containing only the top 5% of regulated transcripts; the full data can be explored using the “Neighborhood Query” feature available at the bottom of the page (red arrow 2). The integration in NDEx of the popular Cytoscape desktop application enables any network to be seamlessly be imported in Cytoscape for additional analysis (red arrow 3). Zooming in on a subset of the SARS2 consensome (orange box) affords an appreciation of the diversity of molecular classes that are

transcriptionally regulated in response to SARS2 infection (Fig. 10b). Transcript size is proportional to rank percentile, and edge weight is proportional to the transcript geometric mean fold change (GMFC) value. Selecting a transcript allows the associated consensome data, such as rank, GMFC and family, to be examined in detail using the information panel (Fig. 10b, right panel). Highlighted to exemplify this feature is IL6, an inflammatory ligand that has been previously linked to SARS2 pathology[8,112]. Consensome GMFCs are signless with respect to direction of regulation[10]. Researchers can therefore follow the SPP link in the side panel (Fig. 10b, red arrow 4) to view the individual underlying experimental data points on the SPP site (Fig. 8c shows the example for *IFI27*). A network of the top 20 ranked transcripts in the SARS2 consensome (Fig. 10c) includes genes with known (*OAS1*, *MX1*[113]) and previously uncharacterized (*PDZKIP1*, *SAT1*, *TM4SF4*) transcriptional responses to SARS2 infection. Finally, to afford insight into pathway nodes whose gain or loss of function contributes to SARS2 infection-induced signaling, Figure 10d shows the top 5% ranked nodes in the SARS2 node HCT ChIP-Seq intersection network (see Supplementary file 1, section 7; see also Figs. 1 & 2 and accompanying discussion above). In this, as with all HCT intersection networks, node size is proportional to the q-value, such that the larger the circle, the lower the q-value, and the higher the confidence that a particular node or node family is involved in the transcriptional response to viral infection.

The visual organization of the NDEx interface offers insights into the impact of CoV infection on human cell signaling that are not readily appreciated in the current SPP interface. For example, it is readily apparent from the NDEx SARS2 consensome network (Fig. 10b; Table 1) that the single largest class of SARS2 HCTs encode

immunomodulatory ligands (OR = 4.6, *p* = 3.8 e-24, hypergeometric test), many of which are members of the cytokine and chemokine superfamilies. In contrast, although still overabundant (OR = 1.58, p = 6.8e-4, hypergeometric test), inflammatory ligands comprise a considerably smaller proportion of the SARS1 HCTs (Table 1). These data represent evidence that SARS2 infection is relatively efficient in modulating a transcriptional inflammatory response in host cells. Consistent with this hypothesis, and while this manuscript was under review, a study appeared showing induction of interferon-stimulated genes in COVID-19 patients was more robust than in response to SARS1 infection[114].

Discussion

An effective research community response to the impact of CoV infection on human health demands systematic exploration of the transcriptional interface between CoV infection and human cell signaling systems. It also demands routine access to computational analysis of existing datasets that is unhindered either by paywalls or by lack of the informatics training required to manipulate archived datasets in their unprocessed state. Moreover, the substantial logistical obstacles to high containment laboratory certification emphasize the need for fullest possible access to, and re-usability of, existing CoV infection datasets to focus and refine hypotheses prior to carrying out *in vivo* CoV infection experiments. Meta-analysis of existing datasets represents a powerful approach to establishing consensus transcriptional signatures – consensomes – which identify those human genes whose expression is most consistently and reproducibly impacted by CoV infection. Moreover, integrating these consensus transcriptional signatures with existing consensomes for cellular signaling pathway nodes can illuminate transcriptional convergence between CoV infection and human cell signaling nodes.

To this end, we generated a set of CoV infection consensomes that rank human genes by the reproducibility of their differential expression ($p < 0.05$) in response to infection of human cells by CoVs. Using HCT intersection analysis, we then computed the CoV consensomes against high confidence transcriptional signatures for a broad range of cellular signaling pathway nodes, affording investigators with a broad range of signaling interests an entrez into the study of CoV infection of human cells. Although other enrichment based pathway analysis tools exist[115], HCT intersection analysis differs from

these by computing against only genes that have the closest predicted regulatory relationships with upstream pathway nodes (i.e. HCTs). The use cases described here represent illustrative examples of the types of analyses that users are empowered to carry out in the CoV infection knowledgebase.

Previous network analyses across independent viral infection transcriptomic datasets, although valuable, have been limited to stand-alone studies[116,117]. Here, to enable access to the CoV consensomes and their >3,000,000 underlying data points by the broadest possible audience, we have integrated them into the SPP knowledgebase and NDEx repository to create a unique, federated environment for generating hypotheses around the impact of CoV infection on human cell signaling. NDEx provides users with the familiar look and feel of Cytoscape to reduce barriers of accessibility and provides for intuitive click-and-drag data mining strategies. Incorporation of the CoV data points into SPP places them in the context of millions more existing SPP data points documenting transcriptional regulatory relationships between human pathway nodes and their genomic targets. In doing so, we provide users with evidence for signaling nodes whose gain or loss of function in response to CoV infection gives rise to these transcriptional patterns. The transcriptional impact of viral infection is known to be an amalgam of host antiviral responses and co-option by viruses of the host signaling machinery in furtherance of its life cycle. It is hoped that dissection of these two distinct modalities in the context of CoV infection will be facilitated by the availability of the CoV consensomes in the SPP and NDEx knowledgebases.

The CoV consensomes have a number of limitations. Primarily, since they are predicated specifically on transcriptional regulatory technologies, they will assign low

rankings to transcripts that may not be transcriptionally responsive to CoV infection, but whose encoded proteins nevertheless play a role in the cellular response. For example, *MASP2*, which encodes an important node in the response to CoV infection[118], has either a very low consensome ranking (SARS1, MERS and IAV), or is absent entirely (SARS2), indicating that it is transcriptionally unresponsive to viral infection and likely activated at the protein level in response to upstream signals. This and similar instances therefore represent “false negatives” in the context of the impact of CoV infection on human cells. Another limitation of the transcriptional focus of the datasets is the absence of information on specific protein interactions and post-translational modifications, either viral-human or human-human, that give rise to the observed transcriptional responses. Although these can be inferred to some extent, the availability of existing[30,67,108] and future proteomic and kinomic datasets will facilitate modeling of the specific signal transduction events giving rise to the downstream transcriptional responses. Finally, although detailed metadata are readily available on the underlying data points, the consensomes do not directly reflect the impact of variables such as tissue context or duration of infection on differential gene expression. As additional suitable archived datasets become available, we will be better positioned to generate more specific consensomes of this nature.

The human CoV and IAV consensomes and their underlying datasets are intended as “living” resources in SPP and NDEx that will be updated and versioned with appropriate datasets as resources permit. This will be particularly important in the case of SARS2, given the expanded budget that worldwide funding agencies are likely to allocate to research into the impact of this virus on human health. Incorporation of future datasets

will allow for clarification of observations that are intriguing, but whose significance is currently unclear, such as the intersection between the CoV HCTs and those of the telomerase catalytic subunit (Supplementary file 2), as well as the enrichment of EMT genes among those with elevated rankings in the SARS2 consensome (Fig. 6). Although they are currently available on the SPP website, distribution of the CoV consensome data points via the SPP RESTful API[10] will be essential for the research community to fully capitalize on this work. For example, several co-morbidities of SARS2 infection, including renal and gastrointestinal disorders, are within the mission of the National Institute of Diabetes, Digestive and Kidney Diseases. In an ongoing collaboration with the NIDDK Information Network (DKNET)[119], the SPP API will make the CoV consensome data points available in a hypothesis generation environment that will enable users to model intersections of CoV infection-modulated host signaling with their own research areas of interest. We welcome feedback and suggestions from the research community for the future development of the CoV infection consensomes and HCT node intersection networks.

Methods

Consistent with emerging NIH mandates on rigor and reproducibility, we have used the Research Resource Identifier (RRID) standard[120] to identify key research resources of relevance to our study.

Dataset biocuration

Datasets from Gene Expression Omnibus (SCR_005012) and Array Express (SCR_002964) were biocurated as previously described, with the incorporation of an additional classification of peptide ligands[121] to supplement the existing mappings derived from the International Union of Pharmacology Guide To Pharmacology (SCR_013077).

Dataset processing and consensome analysis

Array data processing To process microarray expression data, we utilized the log2 summarized and normalized array feature expression intensities provided by the investigator and housed in GEO. These data are available in the corresponding "Series Matrix Files(s)". The full set of summarized and normalized sample expression values were extracted and processed in the statistical program R. To calculate differential gene expression for investigator-defined experimental contrasts, we used the linear modeling functions from the Bioconductor limma analysis package [122]. Initially, a linear model was fitted to a group-means parameterization design matrix defining each experimental variable. Subsequently, we fitted a contrast matrix that recapitulated the sample contrasts of interest, in this case viral infection vs mock infection, producing fold-change and significance values for each array feature present on the array. The current

BioConductor array annotation library was used for annotation of array identifiers. P values obtained from limma analysis were not corrected for multiple comparisons. *RNA-Seq data processing.* To process RNA-Seq expression data, we utilized the aligned, un-normalized, gene summarized read count data provided by the investigator and housed in GEO. These data are available in the corresponding “Supplementary file” section of the GEO record. The full set of raw aligned gene read count values were extracted and processed in the statistical program R using the limma[122] and edgeR analysis[123] packages. Read count values were initially filtered to remove genes with low read counts. Gene read count values were passed to downstream analysis if all replicate samples from at least one experimental condition had cpm > 1. Sequence library normalization factors were calculated to apply scale normalization to the raw aligned read counts using the TMM normalization method implemented in the edgeR package followed by the voom function[124] to convert the gene read count values to log2-cpm. The log2-cpm values were initially fit to a group-means parameterization design matrix defining each experimental variable. This was subsequently fit to a contrast design matrix that recapitulates the sample contrasts of interest, in this case viral infection vs mock infection, producing fold-change and significance values for each aligned sequenced gene. If necessary, the current BioConductor human organism annotation library was used for annotation of investigator-provided gene identifiers. P values obtained from limma analysis were not corrected for multiple comparisons.

Differential expression values were committed to the consensome analysis pipeline as previously described[10]. Briefly, the consensome algorithm surveys each experiment across all datasets and ranks genes according to the frequency with which they are

significantly differentially expressed. For each transcript, we counted the number of experiments where the significance for differential expression was ≤0.05, and then generated the binomial probability, referred to as the consensome p-value (CPV), of observing that many or more nominally significant experiments out of the number of experiments in which the transcript was assayed, given a true probability of 0.05. Genes were ranked firstly by CPV, then by geometric mean fold change (GMFC). A more detailed description of the transcriptomic consensome algorithm is available in a previous publication[10]. The consensomes and underlying datasets were loaded into an Oracle 13c database and made available on the SPP user interface as previously described[10].

Statistical analysis

High confidence transcript intersection analysis was performed using the Bioconductor GeneOverlap analysis package[15] (SCR_018419) implemented in R. Briefly, given a whole set I of IDs and two sets $A \in I$ and $B \in I$, and $S = A \cap B$, GeneOverlap calculates the significance of obtaining S. The problem is formulated as a hypergeometric distribution or contingency table, which is solved by Fisher's exact test. *p*-values were adjusted for multiple testing by using the method of Benjamini & Hochberg to control the false discovery rate as implemented with the p.adjust function in R, to generate q-values. The universe for the intersection was set at a conservative estimate of the total number of transcribed (protein and non protein-coding) genes in the human genome (25,000)[125]. R code for analyzing the intersection between an investigator gene set and CoV consensome HCTs has been deposited in the SPP Github account. A two tailed two sample t-test assuming equal variance was used to compare the mean percentile

ranking of the EMT (12 degrees of freedom) and E2F (14 degrees of freedom) signatures in the MERS, SARS1, SARS2 and IAV consensomes using the PRISM software package (SCR_005375).

Consensome generation

The procedure for generating transcriptomic consensomes has been previously described[10]. To generate the ChIP-Seq consensomes, we first retrieved processed gene lists from ChIP-Atlas[126] (SCR_015511), in which human genes are ranked based upon their average MACS2 occupancy across all publically archived datasets in which a given pathway node is the IP antigen. Of the three stringency levels available (10, 5 and 1 kb from the transcription start site), we selected the most stringent (1 kb). According to SPP convention[10], we then mapped the IP antigen to its pathway node category, class and family, and the experimental cell line to its appropriate biosample physiological system and organ. We then organized the ranked lists into percentiles to generate the node ChIP-Seq consensomes. The 95th percentiles of all consensomes (HCTs, high confidence targets) was used as the input for the HCT intersection analysis.

SPP web application

The SPP knowledgebase (SCR_018412) is a gene-centric Java Enterprise Edition 6, web-based application around which other gene, mRNA, protein and BSM data from external databases such as NCBI are collected. After undergoing semiautomated processing and biocuration as described above, the data and annotations are stored in SPP's Oracle 13c database. RESTful web services exposing SPP data, which are served to responsively designed views in the user interface, were created using a Flat

UI Toolkit with a combination of JavaScript, D3.JS, AJAX, HTML5, and CSS3. JavaServer Faces and PrimeFaces are the primary technologies behind the user interface. SPP has been optimized for Firefox 24+, Chrome 30+, Safari 5.1.9+, and Internet Explorer 9+, with validations performed in BrowserStack and load testing in LoadUIWeb. XML describing each dataset and experiment is generated and submitted to CrossRef (SCR_003217) to mint DOIs[10].

Data availability

Important note on data availability: this paper refers to the first versions of the consensomes and HCT intersection networks based on the datasets available at the time of publication. As additional CoV infection datasets are archived over time, we will make updated versions of the consensomes and HCT intersection analyses accessible in future releases. The entire set of experimental metadata is available in Supplementary file 1, section 1. Consensome data points are in Supplementary file 1, sections 2-5.

SPP SPP consensomes, their underlying data points and metadata, as well as original datasets, are freely accessible at https://ww.signalingpathways.org. Programmatic access to all underlying data points and their associated metadata are supported by a RESTful API at https://www.signalingpathways.org/docs/. All SPP datasets are biocurated versions of publically archived datasets, are formatted according to the recommendations of the FORCE11 Joint Declaration on Data Citation Principles, and are made available under a Creative Commons CC 4.0 BY license. The original datasets are available are linked to from the corresponding SPP datasets using the original repository accession identifiers. These identifiers are for transcriptomic datasets, the Gene Expression Omnibus (GEO) Series (GSE); and for cistromic/ChIP-Seq datasets, the NCBI Sequence Read Archive (SRA) study identifier (SRP). SPP consensomes are assigned DOIs as shown in Table 1.

NDEx NDEx is a recommended repository for Scientific Data, Springer Nature and the PLOS family of journals and is registered on FAIRsharing.org; for additional info and

documentation, please visit http://ndexbio.org. The official SPP account in NDEx is available at: https://bit.ly/30nN129. NDEx versions of consensomes and HCT intersection networks are assigned DOIs as shown in Table 1.

Code availability

SPP source code is available in the SPP GitHub account under a Creative Commons CC 4.0 BY license at https://github.com/signaling-pathways-project.

Acknowledgements

This work was supported by the National Institute of Diabetes, Digestive and Kidney Diseases NIDDK Information Network (DK097748), the National Cancer Institute (CA125123, CA184427) and by the Brockman Medical Research Foundation. We acknowledge the assistance of Apollo McOwiti, Shijing Qu and Michael Dehart in making the datasets and consensomes available in the SPP knowledgebase. We thank all investigators who archive their datasets, without whom SPP would not be possible.

Author contributions

Dataset biocuration: SO

Data analysis: SO, RP, NM

Manuscript drafting: NM

Manuscript editing: NM, RP, SO

Competing interests

The authors declare no competing interests.

References

1. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21**, 335–376 (2003).

2. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–64 (1998).

3. Darnell, J. E., Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science (80-.).* **264**, 1415–1421 (1994).

4. DeDiego, M. L. *et al.* Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis. *PLoS Pathog.* **7**, e1002315 (2011).

5. Josset, L. *et al.* Cell host response to infection with novel human coronavirus EMC predicts potential antivirals and important differences with SARS coronavirus. *MBio* **4**, e00165-13 (2013).

6. Sims, A. C. *et al.* Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. *J. Virol.* **87**, 3885–3902 (2013).

7. Yoshikawa, T. *et al.* Dynamic innate immune responses of human bronchial epithelial cells to severe acute respiratory syndrome-associated coronavirus infection. *PLoS One* **5**, e8729 (2010).

8. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).

9. Lamers, M. M. *et al.* SARS-CoV-2 productively infects human gut enterocytes. *Science* (2020) doi:10.1126/science.abc1669.

10. Ochsner, S. A. *et al.* The Signaling Pathways Project, an integrated ’omics knowledgebase for mammalian cellular signaling pathways. *Sci. data* **6**, 252 (2019).

11. Aevermann, B. D. *et al.* A comprehensive collection of systems biology data characterizing the host response to viral infection. *Sci. data* **1**, 140033 (2014).

12. Doyle, T. *et al.* The interferon-inducible isoform of NCOA7 inhibits endosome-mediated viral entry. *Nat. Microbiol.* **3**, 1369–1376 (2018).

13. Chapgier, A. *et al.* A partial form of recessive STAT1 deficiency in humans. *J. Clin. Invest.* **119**, 1502–1514 (2009).

14. Gruter, P. *et al.* TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* **1**, 649–659 (1998).

15. Shen, L. GeneOverlap. (2013).

16. Totura, A. L. *et al.* Toll-Like Receptor 3 Signaling via TRIF Contributes to a Protective Innate Immune Response to Severe Acute Respiratory Syndrome Coronavirus Infection. *MBio* **6**, e00638-15 (2015).

17. Hensley, L. E. *et al.* Interferon-beta 1a and SARS coronavirus replication. *Emerg. Infect. Dis.* **10**, 317–319 (2004).

18. Appelberg, S. *et al.* Dysregulation in mTOR/HIF-1 signaling identified by proteo-transcriptomics of SARS-CoV-2 infected cells. *bioRxiv* (2020) doi:10.1101/2020.04.30.070383.

19. Wang, W. *et al.* Up-regulation of IL-6 and TNF-alpha induced by SARS-coronavirus spike protein in murine macrophages via NF-kappaB pathway. *Virus Res.* **128**, 1–8 (2007).

20. Zheng, K., Kitazato, K. & Wang, Y. Viruses exploit the function of epidermal growth factor receptor. *Rev. Med. Virol.* **24**, 274–286 (2014).

21. Ng, S. S. M., Li, A., Pavlakis, G. N., Ozato, K. & Kino, T. Viral infection increases glucocorticoid-induced interleukin-10 production through ERK-mediated phosphorylation of the glucocorticoid receptor in dendritic cells: potential clinical implications. *PLoS One* **8**, e63587 (2013).

22. Ostler, J. B., Harrison, K. S., Schroeder, K., Thunuguntla, P. & Jones, C. The Glucocorticoid Receptor (GR) Stimulates Herpes Simplex Virus 1 Productive Infection, in Part Because the Infected Cell Protein 0 (ICP0) Promoter Is Cooperatively Transactivated by the GR and Kruppel-Like Transcription Factor 15. *J. Virol.* **93**, (2019).

23. Hayward, S. D. Viral interactions with the Notch pathway. *Semin. Cancer Biol.* **14**, 387–396 (2004).

24. Ito, T. *et al.* The critical role of Notch ligand Delta-like 1 in the pathogenesis of influenza A virus (H1N1) infection. *PLoS Pathog.* **7**, e1002341 (2011).

25. Rizzo, P. *et al.* COVID-19 in the heart and the lungs: could we ‘Notch’ the inflammatory storm? *Basic research in cardiology* vol. 115 31 (2020).

26. Wang, S. *et al.* Xenobiotic pregnane X receptor (PXR) regulates innate immunity via activation of NLRP3 inflammasome in vascular endothelial cells. *J. Biol.*

Chem. **289**, 30075–30081 (2014).

27. Marik, P. E., Kory, P. & Varon, J. Does vitamin D status impact mortality from SARS-CoV-2 infection? *Medicine in drug discovery* 100041 (2020) doi:10.1016/j.medidd.2020.100041.

28. Rhodes, J. M., Subramanian, S., Laird, E. & Kenny, R. A. Editorial: low population mortality from COVID-19 in countries south of latitude 35 degrees North supports vitamin D as a factor determining severity. *Alimentary pharmacology & therapeutics* (2020) doi:10.1111/apt.15777.

29. Ledford, H. Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature* (2020) doi:10.1038/d41586-020-01824-5.

30. Stukalov, A. *et al.* Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv* 2020.06.17.156455 (2020) doi:10.1101/2020.06.17.156455.

31. Poppe, M. *et al.* The NF-kappaB-dependent and -independent transcriptome and chromatin landscapes of human coronavirus 229E-infected cells. *PLoS Pathog.* **13**, e1006286 (2017).

32. Ludwig, S. & Planz, O. Influenza viruses and the NF-kappaB signaling pathway - towards a novel concept of antiviral therapy. *Biol. Chem.* **389**, 1307–1312 (2008).

33. Ruckle, A. *et al.* The NS1 protein of influenza A virus blocks RIG-I-mediated activation of the noncanonical NF-kappaB pathway and p52/RelB-dependent gene expression in lung epithelial cells. *J. Virol.* **86**, 10211–10217 (2012).

34. Chiang, H.-S. & Liu, H. M. The Molecular Basis of Viral Inhibition of IRF- and

STAT-Dependent Immune Responses. *Front. Immunol.* **9**, 3086 (2018).

35. Frieman, M. *et al.* Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J. Virol.* **81**, 9812–9824 (2007).

36. Garcia-Sastre, A. *et al.* The role of interferon in influenza virus tissue tropism. *J. Virol.* **72**, 8550–8558 (1998).

37. Blaszczyk, K. *et al.* The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. *Cytokine Growth Factor Rev.* **29**, 71–81 (2016).

38. Boudewijns, R. *et al.* STAT2 signaling as double-edged sword restricting viral dissemination but driving severe pneumonia in SARS-CoV-2 infected hamsters. *bioRxiv* (2020) doi:10.1101/2020.04.23.056838.

39. Haas, D. A. *et al.* Viral targeting of TFIIB impairs de novo polymerase II recruitment and affects antiviral immunity. *PLoS Pathog.* **14**, e1006980 (2018).

40. Gelev, V. *et al.* A new paradigm for transcription factor TFIIB functionality. *Sci. Rep.* **4**, 3664 (2014).

41. Haviv, I., Shamay, M., Doitsh, G. & Shaul, Y. Hepatitis B virus pX targets TFIIB in transcription coactivation. *Mol. Cell. Biol.* **18**, 1562–1569 (1998).

42. Bellail, A. C., Olson, J. J. & Hao, C. SUMO1 modification stabilizes CDK6 protein and drives the cell cycle and glioblastoma progression. *Nat. Commun.* **5**, 4234 (2014).

43. Grossel, M. J. & Hinds, P. W. Beyond the cell cycle: a new role for Cdk6 in

differentiation. *J. Cell. Biochem.* **97**, 485–493 (2006).

44. Kaldis, P., Ojala, P. M., Tong, L., Makela, T. P. & Solomon, M. J. CAK-independent activation of CDK6 by a viral cyclin. *Mol. Biol. Cell* **12**, 3987–3999 (2001).

45. Pauls, E. *et al.* Cell cycle control and HIV-1 susceptibility are linked by CDK6-dependent CDK2 phosphorylation of SAMHD1 in myeloid and lymphoid cells. *J. Immunol.* **193**, 1988–1997 (2014).

46. Cingoz, O. & Goff, S. P. Cyclin-dependent kinase activity is required for type I interferon production. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2950–E2959 (2018).

47. Hennessy, E. J., Sheedy, F. J., Santamaria, D., Barbacid, M. & O'Neill, L. A. J. Toll-like receptor-4 (TLR4) down-regulates microRNA-107, increasing macrophage adhesion via cyclin-dependent kinase 6. *J. Biol. Chem.* **286**, 25531–25539 (2011).

48. Handschick, K. *et al.* Cyclin-dependent kinase 6 is a chromatin-bound cofactor for NF-kappaB-dependent gene expression. *Mol. Cell* **53**, 193–208 (2014).

49. Zaborowska, J., Isa, N. F. & Murphy, S. P-TEFb goes viral. *Bioessays* **38 Suppl 1**, S75-85 (2016).

50. Wobbe, C. R. *et al.* In vitro replication of DNA containing either the SV40 or the polyoma origin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **317**, 439–453 (1987).

51. Takahashi, K. *et al.* DNA topoisomerase 1 facilitates the transcription and replication of the Ebola virus genome. *J. Virol.* **87**, 8862–8869 (2013).

52. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the

SARS coronavirus. *Nature* **426**, 450–454 (2003).

53. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

54. Verdecchia, P., Cavallini, C., Spanevello, A. & Angeli, F. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *Eur. J. Intern. Med.* (2020) doi:10.1016/j.ejim.2020.04.037.

55. Fadason, T. *et al.* A transcription regulatory network within the ACE2 locus may promote a pro-viral environment for SARS-CoV-2 by modulating expression of host factors. *bioRxiv* 2020.04.14.042002 (2020) doi:10.1101/2020.04.14.042002.

56. Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* (2020) doi:10.1016/j.cell.2020.04.035.

57. Zhu, H. *et al.* Clinical analysis of 10 neonates born to mothers with 2019-nCoV pneumonia. *Transl. Pediatr.* **9**, 51–60 (2020).

58. Sappenfield, E., Jamieson, D. J. & Kourtis, A. P. Pregnancy and susceptibility to infectious diseases. *Infect. Dis. Obstet. Gynecol.* **2013**, 752852 (2013).

59. Siston, A. M. *et al.* Pandemic 2009 influenza A(H1N1) virus illness among pregnant women in the United States. *JAMA* **303**, 1517–1525 (2010).

60. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* **32**, 513–545 (2014).

61. Pastva, A., Estell, K., Schoeb, T. R. & Schwiebert, L. M. RU486 blocks the anti-

inflammatory effects of exercise in a murine model of allergen-induced pulmonary inflammation. *Brain. Behav. Immun.* **19**, 413–422 (2005).

62. Breslin, N. *et al.* COVID-19 infection among asymptomatic and symptomatic pregnant women: Two weeks of confirmed presentations to an affiliated pair of New York City hospitals. *Am. J. Obstet. Gynecol. MFM* 100118 (2020) doi:10.1016/j.ajogmf.2020.100118.

63. Sutton, D., Fuchs, K., D'Alton, M. & Goffman, D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *The New England journal of medicine* (2020) doi:10.1056/NEJMc2009316.

64. Bazer, F. W., Burghardt, R. C., Johnson, G. A., Spencer, T. E. & Wu, G. Interferons and progesterone for establishment and maintenance of pregnancy: interactions among novel cell signaling pathways. *Reprod. Biol.* **8**, 179–211 (2008).

65. Richards, J. S., Russell, D. L., Ochsner, S. & Espey, L. L. *Ovulation: New dimensions and new regulators of the inflammatory-like response*. *Annual Review of Physiology* vol. 64 (2002).

66. Ghandehari, S. Progesterone for the Treatment of COVID-19 in Hospitalized Men. (2020).

67. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020) doi:10.1038/s41586-020-2286-9.

68. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).

69. Hill, C., Jones, M. G., Davies, D. E. & Wang, Y. Epithelial-mesenchymal transition contributes to pulmonary fibrosis via aberrant epithelial/fibroblastic cross-talk. *J. lung Heal. Dis.* **3**, 31–35 (2019).

70. Li, H. *et al.* Alveolar epithelial cells undergo epithelial-mesenchymal transition in acute interstitial pneumonia: a case report. *BMC Pulm. Med.* **14**, 67 (2014).

71. Gouda, M. M., Shaikh, S. B. & Bhandary, Y. P. Inflammatory and Fibrinolytic System in Acute Respiratory Distress Syndrome. *Lung* **196**, 609–616 (2018).

72. George, P., Wells, A. & Jenkins, G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet* (2020) doi:https://doi.org/10.1016/S2213-2600(20)30225-3.

73. Adair, L. B. 2nd & Ledermann, E. J. Chest CT Findings of Early and Progressive Phase COVID-19 Infection from a US Patient. *Radiology case reports* (2020) doi:10.1016/j.radcr.2020.04.031.

74. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

75. Bolós, V. *et al.* The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *J. Cell Sci.* **116**, 499–511 (2003).

76. Yang, J. *et al.* HIF-2α promotes epithelial-mesenchymal transition through regulating Twist2 binding to the promoter of E-cadherin in pancreatic cancer. *J. Exp. Clin. Cancer Res.* **35**, 26 (2016).

77. Kobayashi, W. & Ozawa, M. The transcription factor LEF-1 induces an epithelial-

mesenchymal transition in MDCK cells independent of β-catenin. *Biochem. Biophys. Res. Commun.* **442**, 133–138 (2013).

78. Xu, J., Lamouille, S. & Derynck, R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* **19**, 156–172 (2009).

79. Zhao, M., Kong, L., Liu, Y. & Qu, H. dbEMT: an epithelial-mesenchymal transition associated gene resource. *Sci. Rep.* **5**, 11459 (2015).

80. Moheimani, F. *et al.* Influenza A virus infection dysregulates the expression of microRNA-22 and its targets; CD147 and HDAC4, in epithelium of asthmatics. *Respir. Res.* **19**, 145 (2018).

81. Xia, L., Dai, L., Yu, Q. & Yang, Q. Persistent Transmissible Gastroenteritis Virus Infection Enhances Enterotoxigenic Escherichia coli K88 Adhesion by Promoting Epithelial-Mesenchymal Transition in Intestinal Epithelial Cells. *J. Virol.* **91**, (2017).

82. Yang, J.-X. *et al.* Lipoxin A(4) ameliorates lipopolysaccharide-induced lung injury through stimulating epithelial proliferation, reducing epithelial cell apoptosis and inhibits epithelial-mesenchymal transition. *Respir. Res.* **20**, 192 (2019).

83. Diaz-Pina, G. *et al.* The Role of ADAR1 and ADAR2 in the Regulation of miRNA-21 in Idiopathic Pulmonary Fibrosis. *Lung* **196**, 393–400 (2018).

84. Vukmirovic, M. *et al.* Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis. *BMC Pulm. Med.* **17**, 15 (2017).

85. Gao, F., Kinnula, V. L., Myllarniemi, M. & Oury, T. D. Extracellular superoxide dismutase in pulmonary fibrosis. *Antioxid. Redox Signal.* **10**, 343–354 (2008).

86. Kato, N. *et al.* Basigin/CD147 promotes renal fibrosis after unilateral ureteral obstruction. *Am. J. Pathol.* **178**, 572–579 (2011).

87. Ruster, C. & Wolf, G. Angiotensin II as a morphogenic cytokine stimulating renal fibrogenesis. *J. Am. Soc. Nephrol.* **22**, 1189–1199 (2011).

88. Wang, C. *et al.* CD147 Induces Epithelial-to-Mesenchymal Transition by Disassembling Cellular Apoptosis Susceptibility Protein/E-Cadherin/beta-Catenin Complex in Human Endometriosis. *Am. J. Pathol.* **188**, 1597–1607 (2018).

89. Stewart, C. A. *et al.* SARS-CoV-2 infection induces EMT-like molecular changes, including ZEB1-mediated repression of the viral receptor ACE2, in lung cancer models. *bioRxiv* 2020.05.28.122291 (2020) doi:10.1101/2020.05.28.122291.

90. Shan, B., Farmer, A. A. & Lee, W. H. The molecular basis of E2F-1/DP-1-induced S-phase entry and apoptosis. *Cell growth Differ. Mol. Biol. J. Am. Assoc. Cancer Res.* **7**, 689–697 (1996).

91. Harbour, J. W. & Dean, D. C. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev.* **14**, 2393–2409 (2000).

92. Reimer, D. *et al.* E2F3a is critically involved in epidermal growth factor receptor-directed proliferation in ovarian cancer. *Cancer Res.* **70**, 4613–4623 (2010).

93. Classon, M. & Harlow, E. The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* **2**, 910–917 (2002).

94. Diril, M. K. *et al.* Cyclin-dependent kinase 1 (Cdk1) is essential for cell division

and suppression of DNA re-replication but not for liver regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3826–3831 (2012).

95. Shivji, K. K., Kenny, M. K. & Wood, R. D. Proliferating cell nuclear antigen is required for DNA excision repair. *Cell* **69**, 367–374 (1992).

96. Borlado, L. R. & Méndez, J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* **29**, 237–243 (2008).

97. Holt, S. V *et al.* Silencing Cenp-F weakens centromeric cohesion, prevents chromosome alignment and activates the spindle checkpoint. *J. Cell Sci.* **118**, 4889–4900 (2005).

98. Vanden Bosch, A. *et al.* NuSAP is essential for chromatin-induced spindle formation during early embryogenesis. *J. Cell Sci.* **123**, 3244–3255 (2010).

99. Engeland, K. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ.* **25**, 114–132 (2018).

100. Wang, N. *et al.* Increased expression of RRM2 by human papillomavirus E7 oncoprotein promotes angiogenesis in cervical cancer. *Br. J. Cancer* **110**, 1034–1044 (2014).

101. Li, Y.-Y., Wang, L. & Lu, C.-D. An E2F site in the 5'-promoter region contributes to serum-dependent up-regulation of the human proliferating cell nuclear antigen gene. *FEBS Lett.* **544**, 112–118 (2003).

102. Furukawa, Y., Terui, Y., Sakoe, K., Ohta, M. & Saito, M. The role of cellular transcription factor E2F in the regulation of cdc2 mRNA expression and cell cycle control of human hematopoietic cells. *J. Biol. Chem.* **269**, 26249–26258 (1994).

103. Yuan, X. *et al.* SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. *Virology* **346**, 74–85 (2006).

104. Fan, Y., Sanyal, S. & Bruzzone, R. Breaking Bad: How Viruses Subvert the Cell Cycle. *Front. Cell. Infect. Microbiol.* **8**, 396 (2018).

105. Gasser, S. & Raulet, D. H. The DNA damage response arouses the immune system. *Cancer Res.* **66**, 3959–3962 (2006).

106. Kong, L.-J., Chang, J. T., Bild, A. H. & Nevins, J. R. Compensation and specificity of function within the E2F family. *Oncogene* **26**, 321–327 (2007).

107. To, B. & Andrechek, E. R. Transcription factor compensation during mammary gland development in E2F knockout mice. *PLoS One* **13**, e0194937 (2018).

108. M. Bouhaddou & Krogan, N. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* (2020) doi:https://doi.org/10.1016/j.cell.2020.06.034.

109. Pillich, R. T., Chen, J., Rynkov, V., Welker, D. & Pratt, D. NDEx: A Community Resource for Sharing and Publishing of Biological Networks. *Methods Mol. Biol.* **1558**, 271–301 (2017).

110. Pratt, D. *et al.* NDEx 2.0: A Clearinghouse for Research on Cancer Pathways. *Cancer Res.* **77**, e58–e61 (2017).

111. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

112. Chen, X. *et al.* Detectable serum SARS-CoV-2 viral load (RNAaemia) is closely correlated with drastically elevated interleukin 6 (IL-6) level in critically ill COVID-19 patients. *Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am.* (2020)

doi:10.1093/cid/ciaa449.

113. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).

114. Zhou, Z. *et al.* Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. *Cell Host Microbe* **27**, 883-890.e2 (2020).

115. Cirillo, E., Parnell, L. D. & Evelo, C. T. A Review of Pathway-Based Analysis Tools That Visualize Genetic Variants. *Front. Genet.* **8**, 174 (2017).

116. Mitchell, H. D. *et al.* A network integration approach to predict conserved regulators related to pathogenicity of influenza and SARS-CoV respiratory viruses. *PLoS One* **8**, e69374 (2013).

117. Ackerman, E. E., Alcorn, J. F., Hase, T. & Shoemaker, J. E. A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery. *BMC Bioinformatics* **20**, 297 (2019).

118. Wilk, C. M. Coronaviruses hijack the complement system. *Nat. Rev. Immunol.* (2020) doi:10.1038/s41577-020-0314-5.

119. Whetzel, P. L., Grethe, J. S., Banks, D. E. & Martone, M. E. The NIDDK Information Network: A Community Portal for Finding Data, Materials, and Tools for Researchers Studying Diabetes, Digestive, and Kidney Diseases. *PLoS One* **10**, e0136206 (2015).

120. Bandrowski, A. E. & Martone, M. E. RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods.

Neuron **90**, 434–436 (2016).

121. Ramilowski, J. A. *et al.* A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).

122. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

123. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

124. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

125. Pertea, M. *et al.* CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 208 (2018).

126. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, (2018).

127. Muzikar, K., Nickols, N. & Dervan, P. Analysis of the dexamethasone (Dex)-dependent transcriptome in A549 lung adenocarcinoma cells *Signaling Pathways Project Datasets*. doi:10.1621/xigKzGn1se.

Figure Titles and Legends

Figure 1. High confidence target (HCT) intersection analysis of viral infection and human receptors or signaling enzymes based on transcriptomic consensomes. Full numerical data are provided in Supplementary file 1, section 6. Due to space constraints some class and family names may differ slightly from those in the SPP knowledgebase. All q-values refer to those obtained using the GeneOverlap analysis package in R[15].

Figure 2. High confidence target (HCT) intersection analysis of viral infection and human transcription factors based on ChIP-Seq consensomes. White cells represent $q > 5e\text{-}2$ intersections. Full numerical data are provided in Supplementary file 1, section 7. Due to space constraints some class and family names may differ slightly from those in the SPP knowledgebase. All q-values refer to those obtained using the GeneOverlap analysis package in R[15].

Figure 3. High confidence target (HCT) intersection analysis of viral infection and human signaling enzymes based on ChIP-Seq consensomes. White cells represent non-significant ($q > 5e\text{-}2$) intersections. Full numerical data are provided in Supplementary file 1, section 7. Due to space constraints some class and family names may differ slightly from those in the SPP knowledgebase. All q-values refer to those obtained using the GeneOverlap analysis package in R[15].

Figure 4. Hypothesis generation use case 1: strain- and tissue-specific regulation of *ACE2* in response to CoV infection of human cells. All data points are $p < 0.05$. Refer to Supplementary file 1, section 1 for full details on the underlying datasets.

Figure 5. Hypothesis generation use case 2: antagonism between PGR and SARS2 inflammatory signaling in the regulation of viral response genes in the airway epithelium. GMFC: geometric mean fold change. PGR loss of function experiments were retrieved from the SPP knowledgebase[127].

Figure 6. Hypothesis generation use case 3: evidence for a SARS2 infection-associated EMT transcriptional signature. A. CoV HCT intersection with the literature-curated EMT signature for all-biosample and lung epithelium-specific consensomes. The IAV consensome is comprised of lung epithelial cell lines and was therefore omitted from the lung epithelium-only consensome analysis. Refer to the column "EMT" in Supplementary file 1, section 3 for the list of EMT SARS2 HCTs. q-values refer to those obtained using the GeneOverlap analysis package in R[15]. B. Comparison of mean percentile ranking of the EMT-associated SARS2 HCTs across viral consensomes. Note that SARS2 HCTs are all in the 97-99th percentile and are therefore superimposed in the scatterplot. Indicated are the results of the two-tailed two sample t-test assuming equal variance comparing the percentile rankings of the SARS2 EMT HCTs across the four viral consensomes.

Figure 7. Hypothesis generation use case 4: efficient SARS2 repression of E2F family HCTs encoding key cell cycle regulators. (A) Relative abundance of E2F HCT cell cycle regulators in response to SARS2 infection. (B) Comparison of SARS2, SARS1, MERS and IAV consensome percentiles of the E2F HCT cell cycle regulators. Indicated are the results of the two-tailed two sample t-test assuming equal variance comparing the percentile rankings of the SARS2 EMT HCTs across the four viral consensomes.

Figure 8. Mining of CoV consensomes and underlying data points in the SPP knowledgebase. A. The Ominer query form can be configured as shown to access the CoV infection consensomes. In the example shown, the user wishes to view the SARS2 consensome. **B**. Consensomes are displayed in a tabular format. Target transcript symbols in the consensomes link to SPP transcriptomic Regulation Reports (**C**) **C.** Regulation Reports for consensome transcripts are filtered to show only data points that contributed to their consensome ranking. Clicking on a data point opens a Fold Change Information window that links to the SPP curated version of the original archived dataset (D). **D.** Like all SPP datasets, CoV infection datasets are comprehensively aligned with FAIR data best practice and feature human-readable names and descriptions, a DOI, one-click addition to citation managers, and machine-readable downloadable data files. For a walk-through of CoV consensome data mining options in SPP, please refer to the accompanying YouTube video (http://tiny.cc/2i56rz).

Figure 9. Visualization of viral consensomes and HCT intersection networks in the NDEx repository. In all panels, transcripts (consensome networks A, B & C) and nodes (HCT intersection network, D) are color-coded according to their category as follows: receptors (orange), enzymes (blue), transcription factors (green), ion channels (mustard) and co-nodes (grey). Additional contextual information is available in the description of each network on the NDEx site. Red arrows are explained in the text. **A.** Sample view of SARS2 consensome showing top 5% of transcripts. White rectangles represent classes to which transcripts have been mapped in the SPP biocuration pipeline[10]. Orange rectangle refers to the view in panel B**. B.** Zoomed-in view of orange rectangle in panel A. IL6 transcript is highlighted to show the contextual information

available in the side panel. **C.** Top 20 ranked transcripts in the SARS2 consensome. Edge widths are proportional to the GMFC. **D.** Selected classes represented in the top 5% of nodes in the SARS2 ChIP-Seq HCT intersection network. Node circle size is inversely proportional to the intersection q-value.

Tables and Table legends

Table 1. DOI-driven links to consensomes and HCT intersection networks. SPP DOIs point to the web browser version of the consensome, which contains a downloadable version of the full consensome. For clarity of visualization, NDEx consensome DOIs point to networks containing transcripts in the top 5% of each consensome (i.e. HCTs for each viral infection); the full consensome network can be reached from this page. Similarly, NDEx HCT intersection DOIs point to networks containing nodes in the top 5% of each HCT intersection network; the full HCT intersection network can be reached from this page. TX, transcriptomic node family intersection; CX, ChIP-Seq node intersection.

1118

Table 1

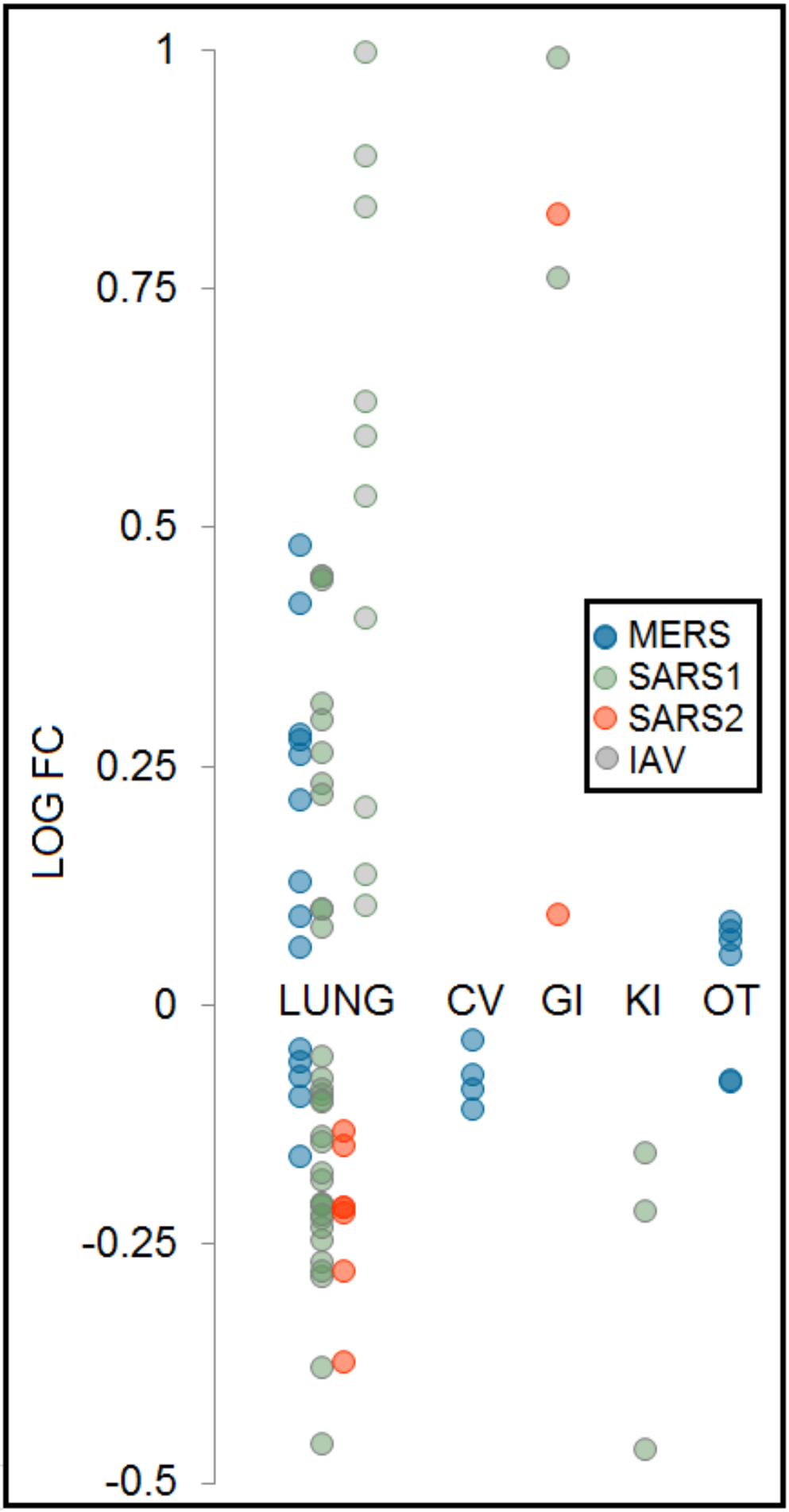
Virus	Resource	Network type	DOI
MERS-CoV	SPP	Consensome	10.1621/jgxM527b8s.1
	NDEx	Consensome	10.18119/N9QG7S
		HCT intersection (TX)	10.18119/N9PG63
		HCT intersection (CX)	10.18119/N96G6R
SARS-CoV-1	SPP	Consensome	10.1621/jgxM527b8s.1
	NDEx	Consensome	10.18119/N9KP4G
		HCT intersection (TX)	10.18119/N9JS46
		HCT intersection (CX)	10.18119/N92P56
SARS-CoV-2	SPP	Consensome	10.1621/k9ygy4i49j.1
	NDEx	Consensome	10.18119/N9G02W
		HCT intersection (TX)	10.18119/N9F016
		HCT intersection (CX)	10.18119/N9Z01V
IAV	SPP	Consensome	10.1621/58AOyXDIAH.1
	NDEx	Consensome	10.18119/N9B60Z
		HCT intersection (TX)	10.18119/N9989R
		HCT intersection (CX)	10.18119/N9T609

Pathway node family transcriptomic HCTs			Viral HCTs intersection q-value			
Category	**Class**	**Family**	**SARS1**	**SARS2**	**MERS**	**IAV**
Receptors	Catalytic receptors	Collagen receptors				
		Epidermal growth factor receptors				
		Fibroblast growth factor receptors				
		Insulin receptor family				
		Interferon receptor family				
		Notch receptors				
		Toll-like receptors				
		Transforming growth factor-β receptor family				
		Tumor necrosis factor receptors				
	G protein coupled receptors	Class Frizzled GPCRs				
	Nuclear receptors	Androgen receptor				
		Estrogen receptors				
		Estrogen-related receptors				
		Glucocorticoid receptor				
		Peroxisome proliferator-activated receptors				
		Progesterone receptor				
		Retinoic acid receptors				
		Retinoid X receptors				
		Vitamin D receptor				
		Xenobiotic receptors				
Enzymes	Kinases	Abl kinases (ABL)				
		Cyclin-dependent kinases (CDK)				
		Src kinases				
	Nucleotidyltransferases	Telomerase reverse transcriptase				
		DNA topoisomerases (TOP)				

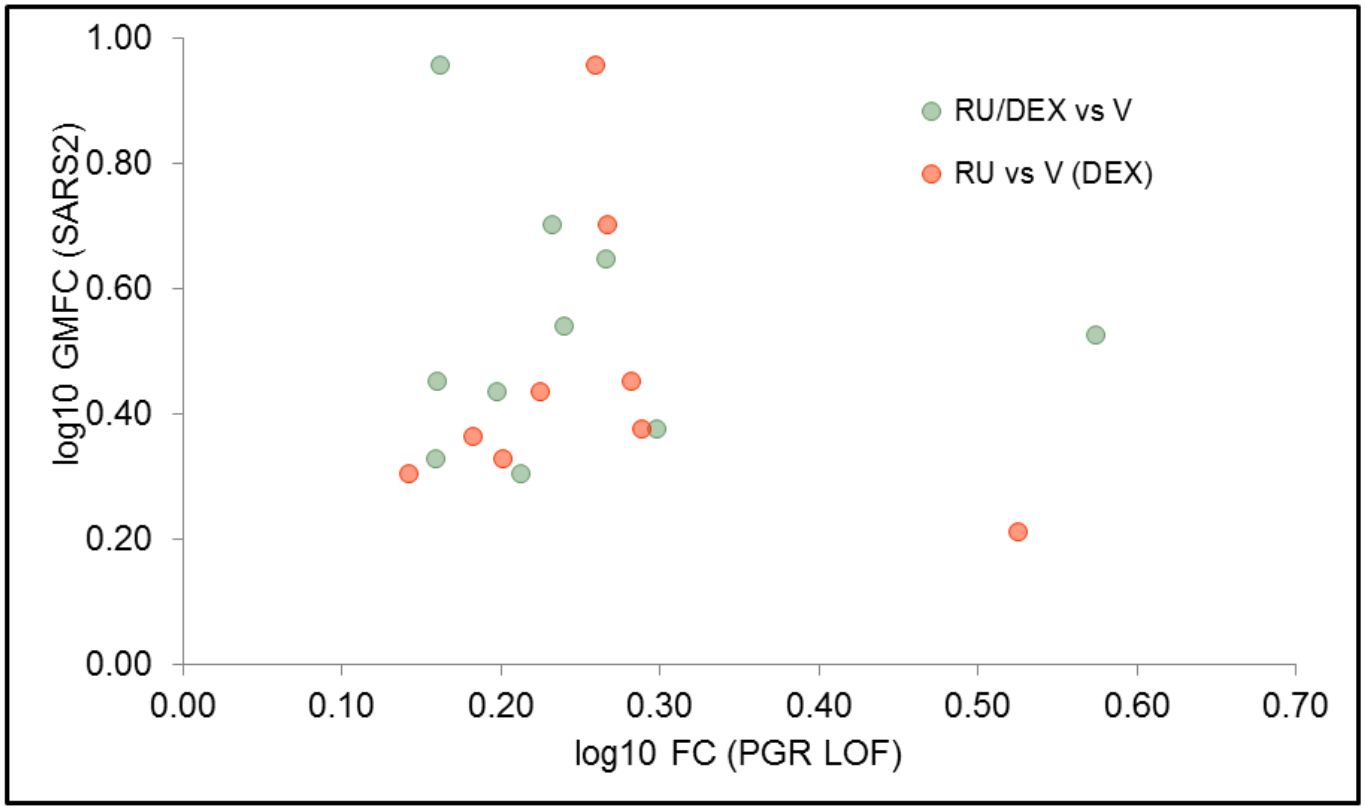
Pathway node ChIP-Seq consensome HCTs			Viral HCTs intersection q-value			
Class	Family	Node	SARS1	SARS2	MERS	IAV
ARID domain	ARID1 family	ARID1A				
	ARID2 family	ARID2				
	ARID3 family	ARID3A				
BHLH factors	Ahr-like family	AHR				
		AHRR				
		EPAS1				
		HIF1A				
	AP-2 family	TFAP2A				
		TFAP2C				
	AP-4 family	TFAP4				
	Arnt-like	ARNT				
		ARNTL				
		CLOCK				
	E2A-related	TCF12				
		TCF3				
		TCF4				
	Hairy-related	BHLHE40				
		HEY1				
	Mad-like	MXI1				
	MESP	TCF21				
	Mondo-like	MLXIP				
	Myc / Max	MAX				
		MYC				
		MYCN				
	Myogenic TFs	MYF5				
		MYOD1				
	Neurogenin-Atonal like	NEUROD1				
	SREBP	SREBF1				
		SREBF2				
	Tal/HEN-like	TAL1				
	TFE3-like	MITF				
		TFEB				
	Twist-like	HAND2				
	USF	USF1				
		USF2				
BZIP factors	ATF-2-like	ATF2				
		ATF3				
		JDP2				
	ATF-4-related	ATF4				
	B-ATF-related	BATF				
		BATF3				
	C/EBP	CEBPA				
		CEBPB				
		CEBPD				
		DDIT3				
	CREB-like	CREB1				
	Fos	FOS				
		FOSL1				
		FOSL2				
	Jun	JUN				
		JUNB				
		JUND				
	Large Maf	MAF				
	NF-E2-like	BACH1				
		BACH2				
		NFE2L2				
	Small Maf	MAFF				
		MAFG				
		MAFK				
C2H2 Zn finger factors	B-cell lymphoma 13	BCL11B				
	B-cell lymphoma 2	BCL11A				
	BCL6	BCL6				
	CTCF-like	CTCF				
		CTCFL				
	Early growth response	EGR1				
	GFI1	GFI1				
	Gli-like	GLI2				
	Hypermethy. in Cancer	HIC1				
	Ikaros	IKZF1				
	Kruppel-like	KLF1				
		KLF11				
		KLF4				
		KLF5				
		KLF6				
	MAZ-like	MAZ				
	PLAG Zinc Finger	PLAG1				
	REST	REST				
	Sal-like	SALL3				
	Snail-like	**SNAI2**				
	Sp1-like	SP1				
		SP2				
		SP4				
	YY1-like	YY1				
	ZBTB17	ZBTB17				
	ZBTB7	ZBTB7A				
	ZFX/ZFY	ZFX				
	ZNF263	ZNF263				
	ZNF341	ZNF341				
	ZNF362-like	ZNF384				
	ZNF366-like	ZNF366				
	ZNF639-like	ZNF711				
	ZNF76-like	ZNF143				
	ZNF83	ZNF83				
	ZNF92	ZNF92				
	ZNF99-like	ZBTB48				
CXXC zinc finger	CpG-binding protein	CXXC1				
E2F/FOX	Dp-1	**TFDP1**				
	E2F	**E2F1**				
		E2F4				
		E2F6				
	FOXA	FOXA1				
		FOXA2				
	FOXF	FOXF1				
	FOXH	FOXH1				
	FOXK	FOXK1				

Pathway node ChIP-Seq consensome			Viral HCTs intersection q-value			
Class	Family	Node	SARS1	SARS2	MERS	IAV
E2F/FOX	FOXM	FOXM1				
	FOXO	FOXO1				
	FOXP	FOXP1				
		FOXP2				
Forkhead/winged helix	Regulatory X (RFX)	RFX5				
Grainyhead domain	GRH-like protein	GRHL2				
		GRHL3				
Heat shock factors	HSF	HSF1				
Heteromeric CCAAT bdg	Heteromeric CCAAT bdg	NFYA				
		NFYB				
		NFYC				
HMG domain	Canonical HMG protein	HMGB2				
	Group B	SOX2				
		SOX3				
	Group E	SOX9				
	TCF-7-related	**LEF1**				
		TCF7L2				
	UBF-related	UBTF				
	WHSC1-related	SSRP1				
Homeo domain	Caudal type homeobox (CDX)	CDX2				
	HNF1-like	HNF1B				
	HOX4	HOXA4				
	HOX5	HOXC5				
	HOX6-7	HOXA6				
	Nanog homeobox	NANOG				
	NK-2.1	NKX2-1				
	Oct-1/2-like (POU2)	POU2F2				
	Orthodenticle homeobox	OTX2				
	PBX	PBX1				
		PBX3				
		PBX4				
	SATB	SATB1				
	SIX1-like	SIX1				
		SIX2				
	TLX	TLX1				
	ZHX	ZHX2				
	Zn finger E-box bdg homeobox	ZEB1				
MADS box	Myocyte enhancer 2	MEF2A				
		MEF2B				
		MEF2C				
	Responders to external signals	SRF				
p53 domain	p53-related	TP53				
		TP63				
Paired box	PAX-2-like (partial homeobox)	PAX5				
Rel Homology Region	Early B-Cell-related	EBF1				
		EBF3				
	IkappaB-related	BCL3				
	M	RBPJ				
	NFAT-related	NFATC1				
	NF-kappaB p50 subunit-like	NFKB1				
		NFKB2				
		REL				
		RELA				
		RELB				
Runt domain	Core-binding subunit	CBFB				
		RUNX1				
		RUNX1T1				
		RUNX2				
		RUNX3				
SAND domain	Sp140-like	SP140				
SMAD/NF1 DBD	Co-activating (Co) Smads	SMAD4				
	Regulatory (R) Smads	SMAD1				
		SMAD2				
		SMAD3				
STAT domain	STAT	STAT1				
		STAT2				
		STAT3				
		STAT4				
		STAT5A				
		STAT5B				
T Box factors	TBrain-related	TBX21				
	TBX2-related (TBX)	TBX2				
TEA domain	TEF-1-related	TEAD1				
		TEAD4				
Tryptophan cluster	EHF-like	ELF3				
	Elf-1-like	ELF1				
	Elk-like	ELK1				
		ELK4				
		ETV1				
		ETV4				
		ETV5				
	Ets-like	ERG				
		ETS1				
		FLI1				
		GABPA				
	Interferon-regulatory	IRF1				
		IRF2				
		IRF4				
		IRF8				
	Myb-like	MYB				
		MYBL2				
	Nuclear receptor corepressor	NCOR1				
		NCOR2				
	REST corepressor	RCOR1				
	Spi-like	SPI1				
Others	Bromodomain PHD finger	BPTF				
	Nuclear I	NFIC				
	PR/SET domain	PRDM1				
	Single GATA-type zinc-finger	MTA3				
		TRPS1				
	Two zinc-finger GATA	GATA1				
		GATA2				
		GATA3				
		GATA4				
		GATA6				

Pathway node ChIP-Seq consensome HCTs			Viral HCTs intersection q-value			
Class	Family	Node	SARS1	SARS2	MERS	IAV
Acetyltransferases	CBP/p300	CREBBP				
		EP300				
	Lysine acetyltransferases (KAT)	KAT7				
	Nuclear receptor coactivator (NCOA)	NCOA1				
ATPases	EP400	EP400				
Deacetylases	Histone deacetylases (HDAC)	HDAC1				
		HDAC2				
		HDAC6				
Demethylases	Histone-H3-lysine-36 demethylases (KDM)	KDM1A				
		KDM2B				
		KDM4A				
		KDM4C				
		KDM5A				
		KDM5B				
		KDM5D				
		KDM6A				
		KDM6B				
	Jumonji domain containing	JMJD1C				
		JMJD6				
Dioxygenases	Ten-eleven translocation (TET)	TET2				
E3 ubiquitin ligases	BRCA1	BRCA1				
	Protein inhibitor of activated STAT (PIAS)	PIAS1				
	Tripartite motif-containing (TRIM)	TRIM24				
		TRIM25				
		TRIM28				
Helicases (DNA)	ATRX chromatin remodeler	ATRX				
	Chromodomain-helicases-DNA binding (CHD)	CHD1				
		CHD2				
	ERCC excision repair (ERCC)	ERCC3				
Kinases	Cyclin-dependent kinases (CDK)	CDK6				
		CDK7				
		CDK8				
		CDK9				
	Extracellular signal-regulated kinases (ERK)	MAPK1				
	Mammalian target of rapamycin (MTOR)	MTOR				
	Mitogen-activated protein kinases (MAPK)	MAPK14				
	Protein kinase C (PKC)	PRKCQ				
Methyltransferases	ASH like methyltransferase	ASH2L				
	DNA (cytosine-5-)-methyltransferases (DNMT)	DNMT3A				
	Euchromatic histone-lysine N-methyltransferases	EHMT2				
	Histone-lysine N-methyltransferases (KMT)	KMT2A				
		KMT2B				
		KMT2C				
		KMT2D				
	Protein arginine methyltransferases (PRMT)	PRMT1				
	SET domain-containing (SETD)	SETD1A				
Peptidases	Proteasome 20S subunit	PSMB5				
Regulatory factors	Cyclin-dependent kinase inhibitors (CDKN)	CDKN1B				
	Cyclins (CCN)	CCND2				
		CCNT2				
	Elongation factors for RNA polymerase II	ELL2				
	PAF1 homolog	PAF1				
	Protein phosphatase 1 regulatory subunits	NONO				
		SFPQ				
Ribonucleases	Argonaute	AGO2				
Topoisomerases	DNA topoisomerases (TOP)	TOP1				
Other enzymes	Recombination activating	RAG1				
		RAG2				

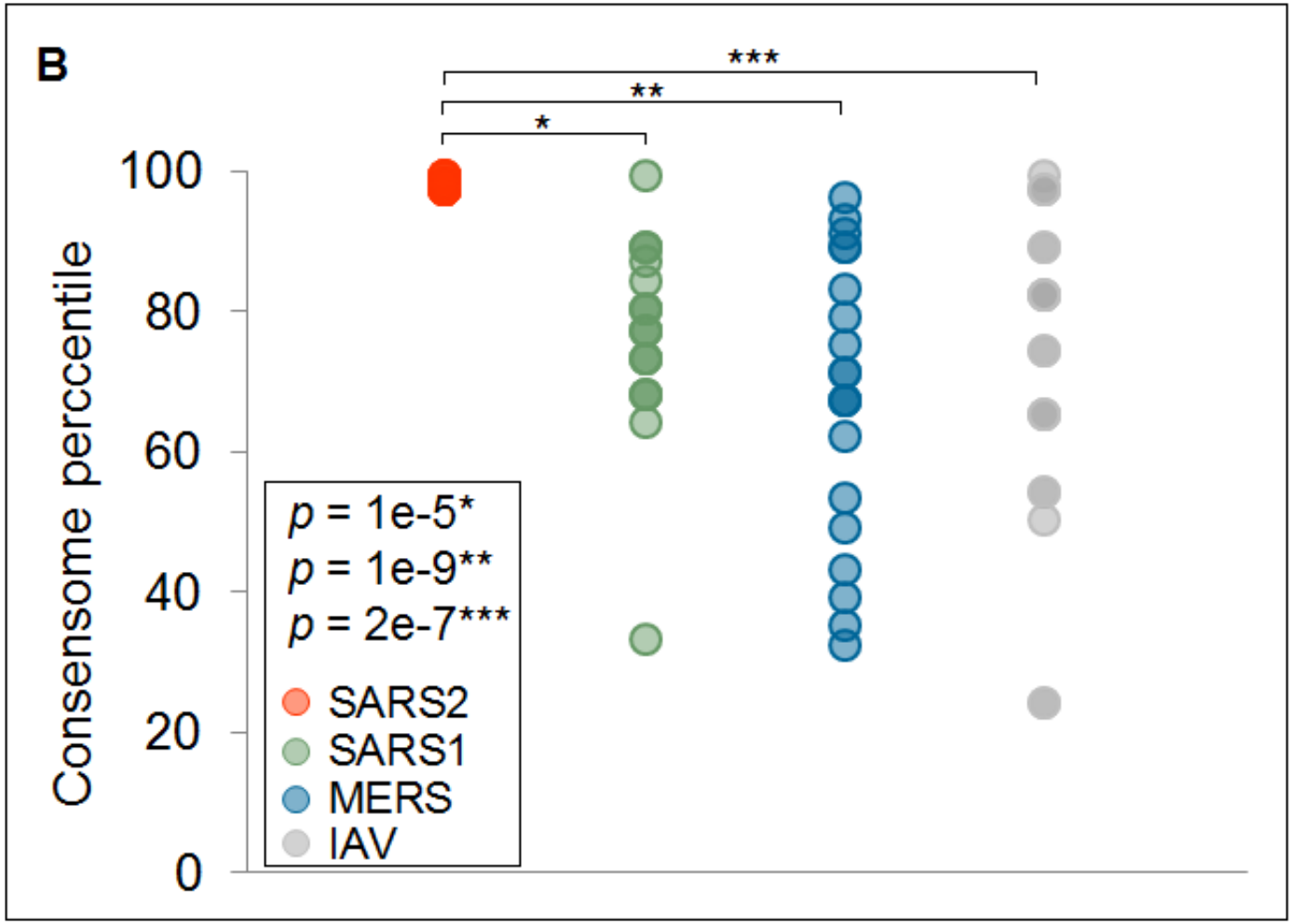

1
0.75
0.5
0.25
0
-0.25
-0.5
LOG FC
LUNG
CV
GI
KI
OT
MERS
SARS1
SARS2
IAV

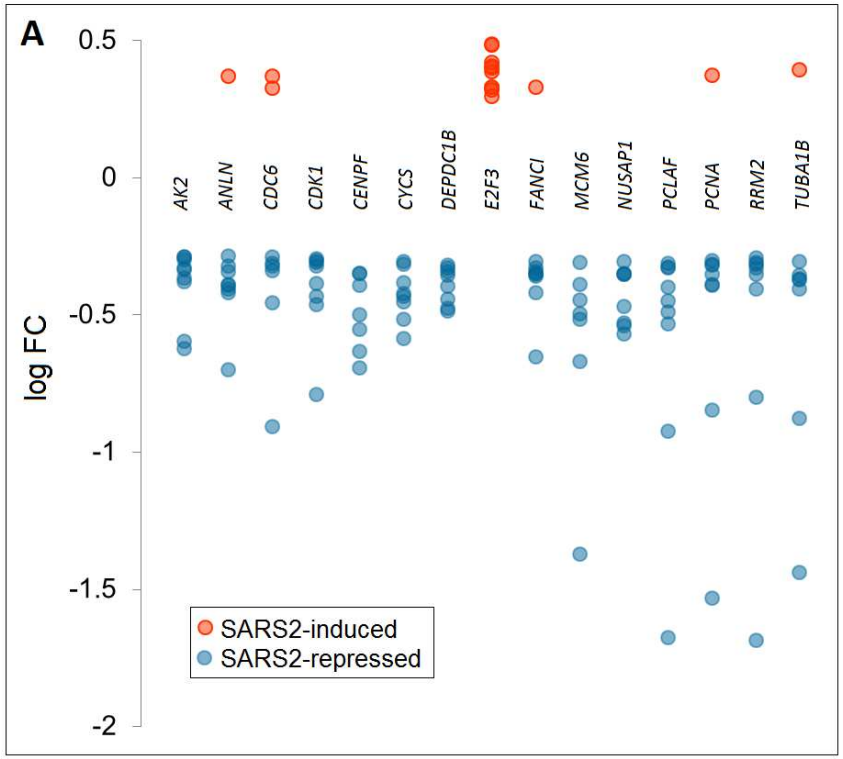
			log10 FC PGR LOF	
Symbol	**Name**	**log10 GMFC SARS2**	**RU/DEX vs V**	**RU vs V (DEX)**
CXCL1	C-X-C motif chemokine ligand 1	0.36		0.18
CXCL2	C-X-C motif chemokine ligand 2	0.53	0.57	
IER3	immediate early response 3	0.21		0.53
IFIT3	interferon induced protein with tetratricopeptide repeats 3	0.96	0.16	0.26
IFITM3	interferon induced transmembrane protein 3	0.33	0.16	0.20
IL1B	interleukin 1 beta	0.38	0.30	0.29
ISG15	ISG15 ubiquitin like modifier	0.70	0.23	0.27
ISG20	interferon stimulated exonuclease gene 20	0.44	0.20	0.22
NFKBIA	NFKB inhibitor alpha	0.54	0.24	
OAS1	2'-5'-oligoadenylate synthetase 1	0.45	0.16	0.28
STAT1	signal transducer and activator of transcription 1	0.30	0.21	0.14
TNFAIP3	TNF alpha induced protein 3	0.65	0.27	

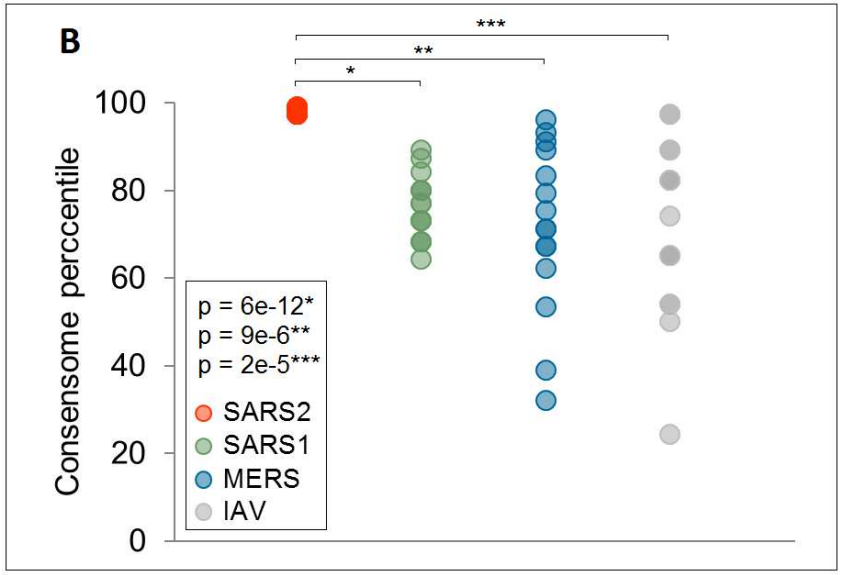


A

	EMT signature intersection analysis					
	All tissues viral HCTs			Lung epithelium viral HCTs		
Virus	**INT**	**OR**	***q***	**INT**	**OR**	***q***
MERS	19	1.3	1.8e-1	13	1.3	1.9e-1
SARS1	15	1.2	2.3e-1	16	1.2	1.9e-1
SARS2	34	5.8	4.5e-14	32	3.7	1.1e-8
IAV	24	2.5	1.4e-4			




A
0.5
0
-0.5
-1
-1.5
-2
log FC
AK2
ANLN
CDC6
CDK1
CENPF
CYCS
DEPDC1B
E2F3
FANCI
MCM6
NUSAP1
PCLAF
PCNA
RRM2
TUBA1B
SARS2-induced
SARS2-repressed


B
***
**
*
100
80
60
40
20
0
Consensome perccentile
p = 6e-12*
p = 9e-6**
p = 2e-5***
SARS2
SARS1
MERS
IAV

The Signaling Pathways Project

A multi-omics knowledgebase for cellular signaling pathways

Home | Datasets | Ominer | Collaborations | About

A. Ominer query form

Target gene(s) of interest

Consensome

'Omics Category

Transcriptomics

Signaling Pathway Module Category

Models

Viral infection

Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2)

B. Consensome

Calculated across **215,000** data points from **17** experiments in **7** datasets.

Search:

Target	Gene Name	Discovery Rate	GMFC	CPV	Percentile
PDZK1IP1	PDZK1 interacting protein 1	0.882	1.943	3.77E-18	100
BAAT	bile acid-CoA:amino acid N-acyltransferase	0.824	1.646	3.60E-16	99
CFB	complement factor B	0.824	2.822	3.60E-16	99
IFI27	interferon alpha inducible protein 27	0.824	5.078	3.60E-16	99
OAS1	2'-5'-oligoadenylate synthetase 1	0.824	2.833	3.60E-16	99
PIGR	polymeric immunoglobulin receptor	0.824	1.777	3.60E-16	99

C. Regulation Report

Display by: Biosample

Transcript Relative Abundance (Fo

Gastrointestinal

Other

small intestine, enteroids

Respiratory

Lung

epithelium, bronchus, NHBE cells

epithelium, A549 cells

epithelium, Calu-3 cells

0 2.5 5 7.5 10 12.5

Fold Change Information

Symbol:IFI27
Fold Change:8.519
p value: 1e-7
Biosample:Respiratory,Lung,epithelium, Calu-3 cells
Experiment:SARS-CoV-2 Wuhan/IVDC-HB-01/2019 vs Mock | 12 h
Species:Human

More Information

D. Dataset

Overview

Dataset Name :
Analysis of the SARS-CoV-2-, RSV-, and IAV-dependent transcriptomes in human primary bronchial epithelial cells and alveolar basal epithelial cells

Description :
Human NHBE primary bronchial epithelial cells and A549 alveolar basal epithelial cells were mock infected or infected at a MOI of 2 and 0.2 respectively with SARS-CoV-2 USA-WA1/2020.

Dataset Type : Transcriptomic **Release Date :** May 08, 2020 **DOI :** 10.1621/AVmtFHml4t

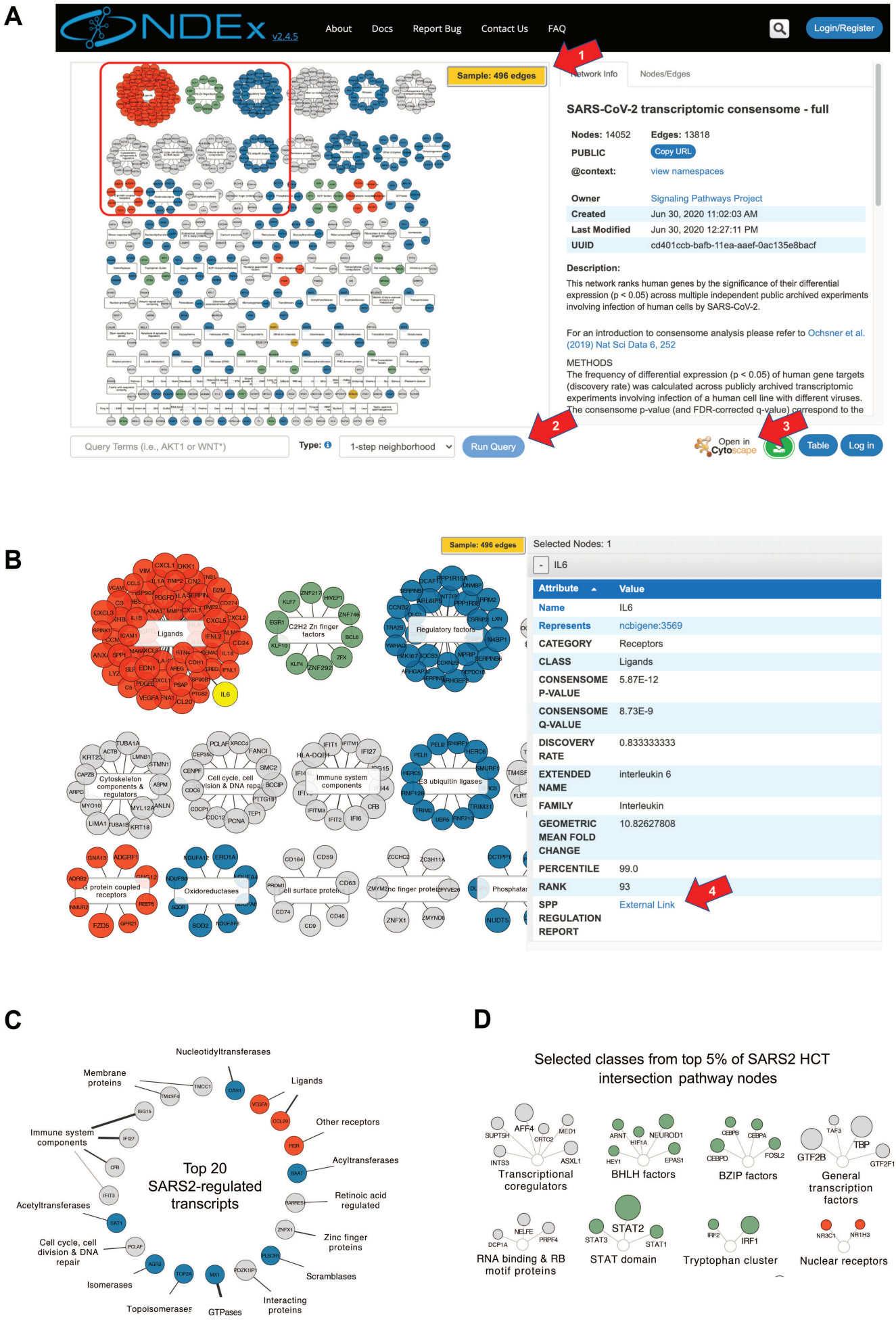
Version : Version 1.0 of an annotated derivative of the original dataset, which can be found in GSE147507

Dataset Citation :
tenOever BR and Blanco-Melo D (2020) Analysis of the SARS-CoV-2-, RSV-, and IAV-dependent transcriptomes in human primary bronchial epithelial cells and alveolar basal epithelial cells, v 1.0 Signaling Pathways Project Datasets. 10.1621/AVmtFHml4t

Download Citation :

Associated Article :
Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA and tenOever BR (2020) Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19 Cell **181** 1036-1045 View Abstract | View PubMed

Download Dataset


A
NDEx v2.4.5
About
Docs
Report Bug
Contact Us
FAQ
Login/Register
Sample: 496 edges
Network Info
Nodes/Edges
SARS-CoV-2 transcriptomic consensome - full
Nodes: 14052
Edges: 13818
PUBLIC
Copy URL
@context:
view namespaces
Owner
Signaling Pathways Project
Created
Jun 30, 2020 11:02:03 AM
Last Modified
Jun 30, 2020 12:27:11 PM
UUID
cd401ccb-bafb-11ea-aaef-0ac135e8bacf
Description:
This network ranks human genes by the significance of their differential expression (p < 0.05) across multiple independent public archived experiments involving infection of human cells by SARS-CoV-2.
For an introduction to consensome analysis please refer to Ochsner et al. (2019) Nat Sci Data 6, 252
METHODS
The frequency of differential expression (p < 0.05) of human gene targets (discovery rate) was calculated across publicly archived transcriptomic experiments involving infection of a human cell line with different viruses. The consensome p-value (and FDR-corrected q-value) correspond to the
Query Terms (i.e., AKT1 or WNT*)
Type:
1-step neighborhood
Run Query
Open in Cytoscape
Table
Log in
1
2
3
B
Sample: 496 edges
Selected Nodes: 1
IL6
Ligands
C2H2 Zn finger factors
Regulatory factors
Cytoskeleton components & regulators
Cell cycle, cell division & DNA repair
Immune system components
E3 ubiquitin ligases
G protein coupled receptors
Oxidoreductases
Cell surface proteins
Zinc finger proteins
Phosphatases
Attribute
Value
Name
IL6
Represents
ncbigene:3569
CATEGORY
Receptors
CLASS
Ligands
CONSENSOME P-VALUE
5.87E-12
CONSENSOME Q-VALUE
8.73E-9
DISCOVERY RATE
0.833333333
EXTENDED NAME
interleukin 6
FAMILY
Interleukin
GEOMETRIC MEAN FOLD CHANGE
10.82627808
PERCENTILE
99.0
RANK
93
SPP REGULATION REPORT
External Link
4
C
Top 20 SARS2-regulated transcripts
Nucleotidyltransferases
Membrane proteins
Ligands
Other receptors
Immune system components
Acyltransferases
Retinoic acid regulated
Acetyltransferases
Zinc finger proteins
Cell cycle, cell division & DNA repair
Scramblases
Isomerases
Interacting proteins
Topoisomerases
GTPases
D
Selected classes from top 5% of SARS2 HCT intersection pathway nodes
Transcriptional coregulators
BHLH factors
BZIP factors
General transcription factors
RNA binding & RB motif proteins
STAT domain
Tryptophan cluster
Nuclear receptors