SCIENTIFIC DATA

Check for updates

OPEN

ANALYSIS

Consensus transcriptional regulatory networks of coronavirus-infected human cells

Scott A. Ochsner[1], Rudolf T. Pillich[2] & Neil J. McKenna [1]✉

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 MERS, SARS1 and SARS2 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 HCTs 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 transcriptional targets encoding their downstream effectors[1–3]. Placing the transcriptional events resulting from CoV infection in context with those associated with host signaling systems 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 regulatory relationships between signaling pathway nodes and their transcriptional targets 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

[1]The Signaling Pathways Project and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, 77030, USA. [2]Department of Medicine, University of California San Diego, La Jolla, CA, 92093, USA. ✉e-mail: nmckenna@bcm.edu

| Virus	Resource	Network type	DOI	Ref.
MERS-CoV	SPP	Consensome	https://doi.org/10.1621/jgxM527b8s.1	[129]
	NDEx	Consensome	https://doi.org/10.18119/N9QG7S	[133]
		Node family HCT intersection	https://doi.org/10.18119/N9PG63	[137]
		Node HCT intersection	https://doi.org/10.18119/N96G6R	[141]
SARS-CoV-1	SPP	Consensome	https://doi.org/10.1621/k9ygy4i49j.1	[130]
	NDEx	Consensome	https://doi.org/10.18119/N9KP4G	[134]
		Node family HCT intersection	https://doi.org/10.18119/N9JS46	[138]
		Node HCT intersection	https://doi.org/10.18119/N92P56	[142]
SARS-CoV-2	SPP	Consensome	https://doi.org/10.1621/vTiy8d4Iq7.1	[131]
	NDEx	Consensome	https://doi.org/10.18119/N9G02W	[135]
		Node family HCT intersection	https://doi.org/10.18119/N9F016	[139]
		Node HCT intersection	https://doi.org/10.18119/N9Z01V	[143]
IAV	SPP	Consensome	https://doi.org/10.1621/58AOyXDIAH.1	[132]
	NDEx	Consensome	https://doi.org/10.18119/N9B60Z	[136]
		Node family HCT intersection	https://doi.org/10.18119/N9989R	[140]
		Node HCT intersection	https://doi.org/10.18119/N9T609	[144]

Table 1. DOI-driven links to consensomes and HCT intersection networks. **SPP DOIs** point to the tabular web version of the consensome, which can be downloaded as an Excel file. **NDEx consensome DOIs** point to the full consensome network; for ease and clarity of display, only the top 5% of the consensome is shown in the initial graphic display; in addition, a subset of the data corresponding only to the top 5% of the consensome can be reached via a link in the "Description". **NDEx virus node family HCT intersection DOIs** point to networks containing all node families; **NDEx virus node HCT intersection DOIs** point to the full HCT intersection network; for ease and clarity of display, only the top 5% of the HCT intersection network is shown in the initial graphic display; in addition, a subset of the data corresponding only to the top 5% of the HCT intersection network can be reached via a link in the "Description".

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, numerous contemporaneous and independent *in vitro* and *in vivo* studies came to light that corroborate SARS2 network 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 (see figshare F1[12], 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. The full consensomes are available in figshare File F1[12] sections 2 (SARS1), 3 (SARS2), 4 (MERS) and 5 (IAV). To assist researchers in inferring viral 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 (figshare File F1[12], sections 2-5, columns A-C). In addition, to facilitate identification of transcripts that are selectively regulated in response to a specific CoV, or in response to CoV vs IAV infection, transcript percentile rankings for other consensomes are provided in each consensome tab in figshare FileF1[12]. Finally, Table 1 contains links to consensomes in the SPP knowledgebase and NDEx repository – see section "Visualization of the CoV transcriptional regulatory networks in the Signaling Pathways Project knowledgebase and Network Data Exchange repository" below and the instructional YouTube video (http://tiny.cc/2i56rz; strategies 2 (SPP) and 5 (NDEx)) for instructions on navigating these resources.

Ranking of interferon-stimulated genes (ISGs) in the CoV consensomes. As an initial benchmark for our CoV consensome analysis, we assembled a list of 20 canonical interferon-stimulated genes (ISGs), whose role in the anti-viral response is best characterized in the context of IAV infection[13]. As shown in Fig. 1, many
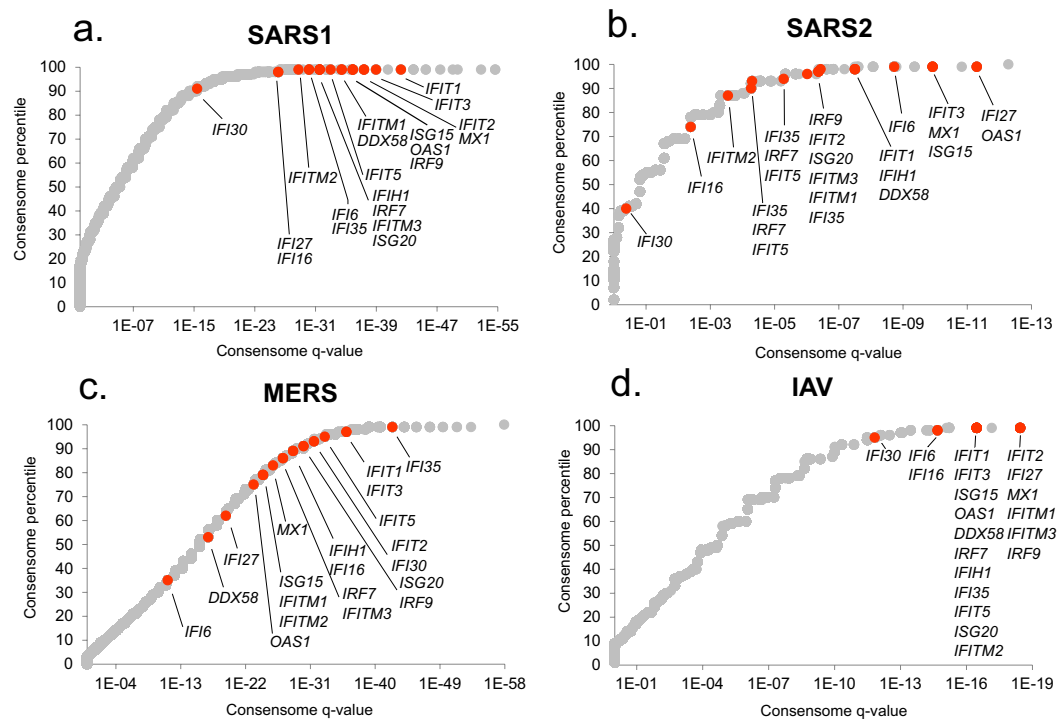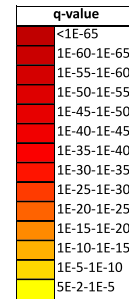
Fig. 1 Rankings of canonical interferon-stimulated genes (ISGs) in the viral consensomes. Shown are the percentile rankings of 20 ISGS[13] in the SARS1 **(a)**, SARS2 **(b)**, MERS **(c)** and IAV **(d)** consensomes. Note that numerous genes have identical *q*-value and percentile values and are therefore superimposed in the plots. Full consensome data are provided in figshare File 1[12] sections 2 (SARS1), 3 (SARS2), 4 (MERS) and 5 (IAV); see also Table 1 for links to consensomes in the SPP knowledgebase and NDEx repository. Please refer to the Methods section for a full description of the consensome algorithm.

ISGs were assigned elevated rankings across the four viral consensomes. The mean percentile of the ISGs was however appreciably higher in the IAV (98.7th percentile) and SARS1 (98.5th percentile; $p = 6\text{e-}1$, t-test IAV vs SARS1) consensomes than in the SARS2 (92nd percentile, $p = 5\text{e-}2$, t-test IAV v SARS2) and MERS (82nd percentile; $p = 7\text{e-}5$, t-test IAV v MERS) consensomes. This is consistent with previous reports of an appreciable divergence between IAV and SARS2 infection with respect to the interferon transcriptional response[8]. Other genes with known critical roles in the response to viral infection have high rankings in the CoV consensomes, including *NCOA7*[14] (percentiles: SARS1, 98th; SARS2, 97th; MERS, 89th; IAV, 99th), *STAT1*[15] (percentiles: SARS1, 99th; SARS2, 98th; MERS, 89th; IAV, 99th) and *TAP1*[16] (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). Finally, a preprint of a CRISPR/Cas9 study described validation of 27 human genes as critical modulators of the host response to SARS2 infection of human cells[17]. Corroborating our analysis, 16 of these genes have significant ($q < 0.05$) rankings in the SARS2 consensome, including *ACE2* and *DYRK1A* (both 97th percentile), *CTSL* (96th percentile), *KDM6A*, *ATRX*, *PIAS1* (all 94th percentile), *RAD54L2* and *SMAD3* (90th percentile).

To gather evidence for human signaling pathway nodes orchestrating the transcriptional response to CoV infection, we next compared transcripts with elevated rankings in the CoV consensomes with those that have predicted high confidence regulatory relationships with cellular signaling pathway nodes. We generated four lists of genes corresponding to the MERS, SARS1, SARS2 and IAV transcriptomic consensome 95th percentiles. We then 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[18] 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 Fig. 2 (based on receptor and enzyme family transcriptomic consensomes), Figs. 3 and 4 (based on ChIP-Seq consensomes for

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)				

q-value
| |
|---|
| <1E-65 |
| 1E-60-1E-65 |
| 1E-55-1E-60 |
| 1E-50-1E-55 |
| 1E-45-1E-50 |
| 1E-40-1E-45 |
| 1E-35-1E-40 |
| 1E-30-1E-35 |
| 1E-25-1E-30 |
| 1E-20-1E-25 |
| 1E-15-1E-20 |
| 1E-10-1E-15 |
| 1E-5-1E-10 |
| 5E-2-1E-5 |

Fig. 2 High confidence transcriptional target (HCT) intersection analysis of viral infection and human receptors or signaling enzymes based on transcriptomic consensomes. 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[18]. Full numerical data are provided in figshare File F1[12], section 6; see also Table 1 for links to virus-node family HCT intersection networks in the NDEx repository.

transcription factors and enzymes, respectively) and figshare File F2[12] (based on ChIP-Seq consensomes for selected co-nodes). figshare File F1[12], sections 6 (node family transcriptomic HCT intersection analysis) and 7 (node ChIP-Seq HCT intersection analysis) contain the full underlying numerical data. Please refer also to Table 1 for links to virus-node family and virus-node HCT intersection networks in the NDEx repository – see section "Visualization of the CoV transcriptional regulatory networks in the Signaling Pathways Project knowledgebase and Network Data Exchange repository" below and the instructional YouTube video (http://tiny. cc/2i56rz; strategy 6) for instructions on navigating these resources. 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[18].

Receptors Reflecting their well-documented roles in the response to CoV infection[19–22], 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. 2). Indeed, anti-TNF therapy has been recently mooted as a potential clinical approach to SARS2 infection[23]. 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. 2), are consistent with their known role in the context of other viral infections[24–28]. The Notch receptor HCT intersection points to a possible mechanistic basis for the potential of Notch pathway modulation in the treatment of SARS2[29]. The strong HCT intersection between CoV infection and xenobiotic receptors (*q*-values: SARS1, 1e-30; SARS2, 1e-44; MERS, 5e-32; Fig. 2) reflects work describing a role for pregnane X receptor in innate immunity[30] 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[31,32]. Consistent with a robust signature for the glucocorticoid receptor across all CoVs (GR; *q*-values: SARS1, 3e-35; SARS2, 1e-35; MERS, 7e-32), recent studies have shown the GR agonist dexamethasone is a successful therapeutic for SARS2 infection[33]. Finally, independent *in vitro* analyses confirm our predictions of the modulation by SARS2 infection of ErbB/EGFR[22,34] and TGFBR[17,34] signaling systems (Fig. 2).
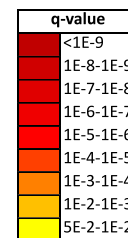
Transcription factors Not unexpectedly – and speaking again to validation of the consensomes - the strongest and most significant CoV HCT intersections were observed for HCTs for known transcription factor mediators of the transcriptional response to CoV infection, including members of the NFκB (*q*-value ranges: SARS1, 1e-7-1e-9; SARS2, 9e-3-2e-3; MERS, 1e-3-1e-4)[35–37], IRF (*q*-value ranges: SARS1, 2e-2-1e-31; SARS2, 2e-4-1e-17; MERS, 9e-4-7e-5)[38] and STAT (*q*-value ranges: SARS1, 1e-7-1e-55; SARS2, 2e-3-3e-29; MERS, 5e-2-3e-5)[39–41] transcription factor families (Fig. 3). Consistent with the similarity between SARS1 and IAV consensomes with respect to elevated rankings of ISGs (Fig. 2a,d), the IRF1 HCT intersection was strongest with the SARS1 (*q* = 2e-34) and IAV (*q* = 3e-49) HCTs. Corroborating our finding of a strong intersection between STAT2 and SARS2 infection HCTs (*q* = 3e-29), a recent preprint has shown that STAT2 plays a prominent role in the response to SARS2 infection of Syrian hamsters[42]. 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/
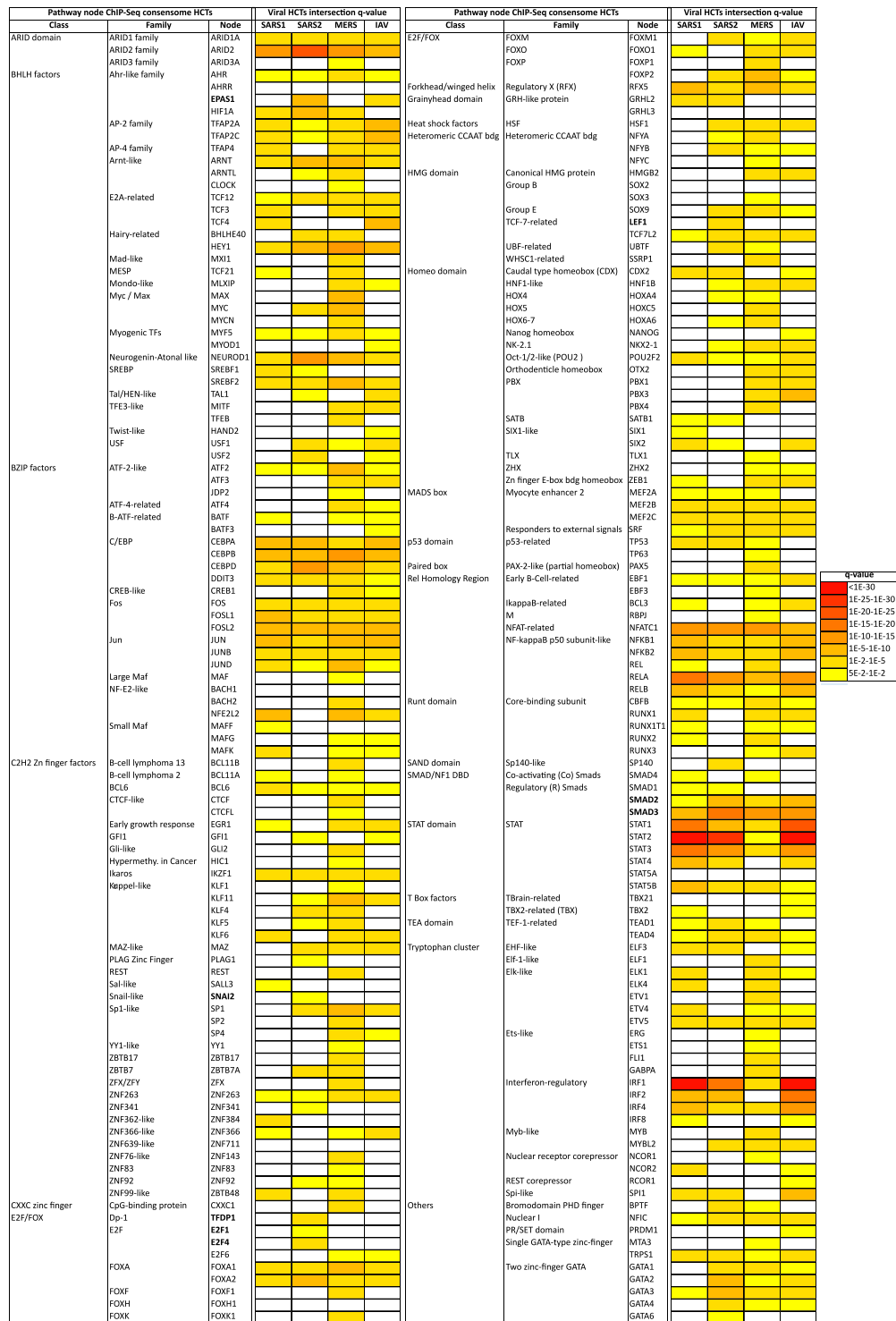
Fig. 3 High confidence transcriptional target (HCT) intersection analysis of viral infection and human transcription factors based on ChIP-Seq consensomes. White cells represent $q > 5e\text{-}2$ intersections. Due to space constraints some class and family names may differ slightly from those in the SPP knowledgebase. All q-values refer to those obtained using the GeneOverlap analysis package in R[18]. Full numerical data are provided in figshare File F1[12], section 7; see also Table 1 for links to virus-node HCT intersection networks in the NDEx repository.

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[43], and the herpes simplex[44] and hepatitis B[45] viruses. Moreover, a recent preprint has identified a high confidence interaction between GTF2F2 and the SARS2 NSP9 replicase[34].

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				

q-value

q-value
<1E-9
1E-8-1E-9
1E-7-1E-8
1E-6-1E-7
1E-5-1E-6
1E-4-1E-5
1E-3-1E-4
1E-2-1E-3
5E-2-1E-2

Fig. 4 High confidence transcriptional 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. 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[18]. Full numerical data are provided in figshare File F1[12], section 7; see also Table 1 for links to virus-node HCT intersection networks in the NDEx repository.

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. 2 with Figs. 3 and 4 and figshare File F1[12], 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[46,47], has been shown to be targeted by a number of viral infections, including Kaposi's sarcoma-associated herpesvirus[48] and HIV-1[49]. 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. 2) and the CDK6 HCTs (q-values: SARS1, 1e-7; SARS2, 8e-8; MERS, 3e-4; Fig. 4) and those of all viral HCTs. As with the TLRs, IFNRs and TNFRs, which are known to signal through CDK6[50,51], intersection with the CDK6 HCTs was particularly strong in the case of the SARS2 HCTs (Fig. 4). Again, the subsequent proteomic analysis we alluded to earlier[34] independently corroborated our prediction of a role for CDK6 in the response to SARS2 infection. Consistent with a recent study[52], the intersection of HCTs for the lysine demethylase KMT2A was much stronger with IAV (q-value = 2e-9) than with any of the CoVs (q-values: SARS1, 7e-2; SARS2, 2e-3; MERS, 1e-2).

CCNT2 is another member of the cyclin family that, along with CDK9, is a component of the viral-targeted p-TEFB complex[53]. 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. 4). Finally in the context of enzymes, the DNA topoisomerases have been shown to be required for efficient replication of simian virus 40[54] and Ebola[55] 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. 2) 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. figshare File F2[12] contains an additional use case linking the telomerase catalytic subunit TERT to CoV infection that was 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[18]. 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[56] and SARS2[57]. 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; figshare File F1[12], 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. 5). On the other hand, in response to SARS2 infection, *ACE2* is repressed in pulmonary cells - a finding corroborated by other studies[58,59] - but inducible in gastrointestinal cells (Fig. 5). These data may relate to the selective transcriptional response of *ACE2* to signaling by IFNRs (92nd percentile; figshare File F1[12], section 8) rather than TLRs (48th percentile; figshare File F1[12], section 9) or TNFRs (13th percentile, figshare File F1[12], section 10). These findings are consistent with a recent study confirming repression of induction of *ACE2* by interferon stimulation and by IAV infection[60]. 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[61], and pregnancy has been previously associated with the incidence of viral infectious diseases, particularly respiratory infections[62,63]. 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. 2; figshare File F1[12], 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[13]. 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 Fig. 6, 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[64]. Interestingly, SARS2-infected pregnant women are often asymptomatic[65,66]. 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[67] as well as in ovulation[68]. Consistent with our hypothesis, a recently launched clinical trial is evaluating the potential of progesterone for treatment of COVID-19 in hospitalized men[69]. Interestingly, the recently reported inhibition by progesterone of SARS2 replication in
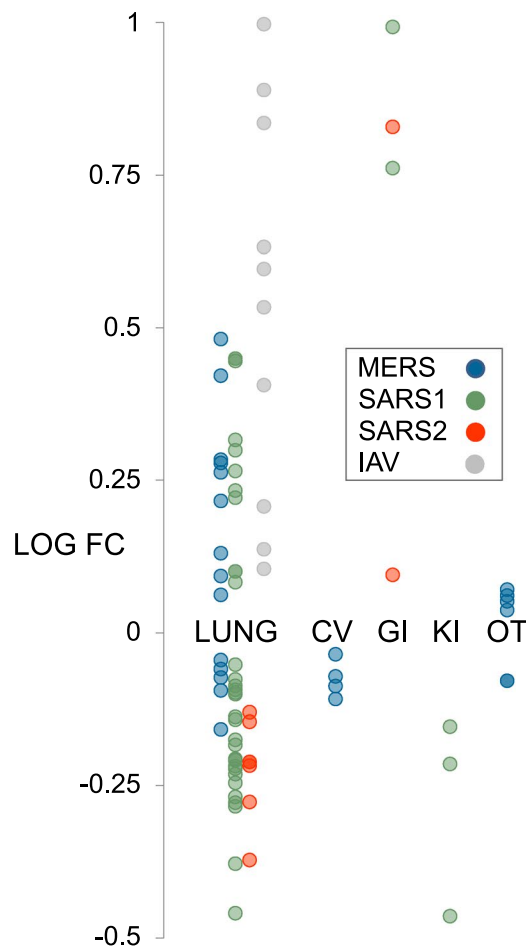
Fig. 5 Hypothesis generation use case 1: strain- and tissue-specific regulation of *ACE2* in response to CoV infection of human cells. All data points are p < 0.05. Refer to figshare File F1[12], section 1 for full details on the underlying datasets. Abbreviations: CV, cardiovascular; GI, gastrointestinal; KI, kidney; OT, others.

Vero 6 cells[70] indicates an additional mechanism, distinct from its potential crosstalk with the interferon response, by which progesterone signaling may impact SARS2 infection.
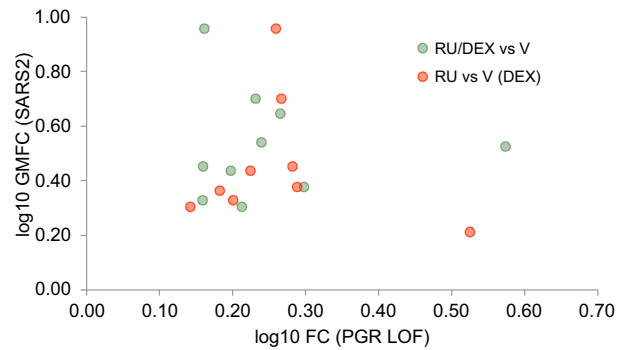
Hypothesis generation use case 3: association of an epithelial to mesenchymal transition transcriptional signature with SARS2 infection. Epithelial to mesenchymal transition (EMT) is the process by which epithelial cells lose their polarity and adhesive properties and acquire the migratory and invasive characteristics of mesenchymal stem cells[71]. EMT is known to contribute to pulmonary fibrosis[72], acute interstitial pneumonia[73] and acute respiratory distress syndrome (ARDs)[74], all of which have been reported in connection with SARS2 infection in COVID-19[75–77]. 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[78], 2e-2; EPAS1/HIF2α[79], 9e-9; LEF1[80], 1e-3; Fig. 3, bold symbols; figshare File F1[12], section 7). Consistent with this, intersections between HCTs for TGFBRs, SMAD2 and SMAD3, known regulators of EMT transcriptional programs[81] – 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. 2 and 3 and figshare File F1[12], sections 6 and 7). Moreover, a recent CRISPR/Cas9 screen identified a requirement for both TGFBR signaling and *SMAD3* in mediating SARS2 infection[17].

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[82], see also figshare File F1[12], 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. 7a). Consistent with previous reports of a potential link between EMT and IAV infection[83], we observed significant intersection between the EMT signature and the IAV HCTs (q = 1e-4).

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 (figshare

Symbol	Name	log10 GMFC SARS2	log10 FC PGR LOF	
			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	

Fig. 6 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 (LOF) experiments were retrieved from the SPP knowledgebase[145].
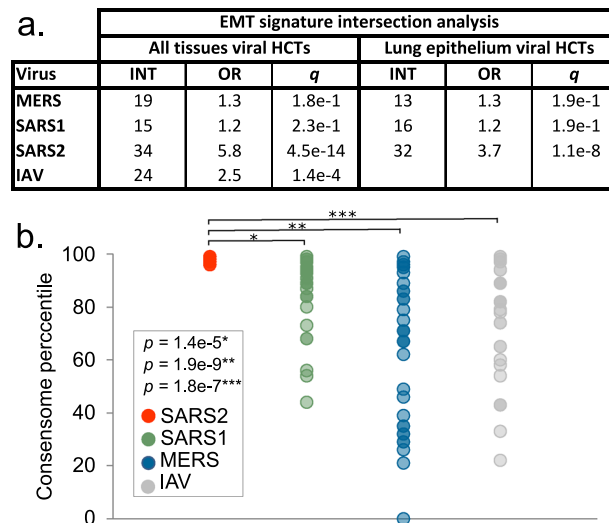
File F1[12], section 1). To exclude this possibility therefore, we next calculated airway epithelial cell-specific consensomes for SARS1, SARS2 and MERS and computed intersections between their HCTs and the EMT signature. We found that significant intersection of the EMT signature with the CoV HCTs remained specific to SARS2 (q-values: SARS1 & MERS, 2e-1; SARS2, 1e-8) 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. 7b; SARS2 mean percentile = 97.5; SARS1 mean percentile = 86, p = 1e-5, t-test SARS2 v SARS1; MERS mean percentile = 63, p = 1e-9, t-test SARS2 v MERS; IAV mean percentile = 76, p = 2e-7, t-test SARS2 v IAV)). A column named "EMT" in figshare File F1[12], 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[74], 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 figshare File F1[12], 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[84] and IAV[83], 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[85]. Moreover, several members of the group of SARS2-induced EMT genes have been associated with signature pulmonary comorbidities of CoV infection, including *ADAR*[86], *CLDN1*[87] and *SOD2*[88]. 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[89–91]. Finally, a recent preprint has described EMT-like transcriptional and metabolic changes in response to SARS2 infection[92]. 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. 3, bold symbols; figshare File F1[12], section 7). These factors play well-documented interdependent roles in the promotion (E2F1, E2F3, TFDP1)

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			

b.

Fig. 7 Hypothesis generation use case 3: evidence for a SARS2 infection-associated EMT transcriptional signature. (**a**) CoV HCT intersection (INT) 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 figshare File F1[12], section 3 for the list of EMT SARS2 HCTs. q-values refer to those obtained using the GeneOverlap analysis package in R[18]. (**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.

and repression (E2F4) of cell cycle genes[93,94]. Moreover, E2F family members are targets of signaling through EGFRs[95] and CDK6[96], 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. 2 and 4). 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 (figshare File F1[12], section 3, columns E2F1, E2F3, E2F4, TFDP1 & 95th 3/4). 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[97–101] and are known to be transcriptionally induced by E2Fs[102–105]. Strikingly, with the exception of *E2F3*, all were consistently repressed in response to SARS2 infection (Fig. 8a). 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 SARS2 v SARS1), MERS (71.2 percentile; $p = 9e-6$, t-test SARS2 v MERS) and IAV (71.2 percentile; $p = 2e-5$, t-test SARS2 v IAV) consensomes (Fig. 8b). Although manipulation of the host cell cycle and evasion of detection through deregulation of cell cycle checkpoints has been described for other viruses[106–108], 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. The SARS2 infection-mediated induction of *E2F3* (Fig. 8a) may represent a compensatory response to transcriptional repression of other E2F family members, as has been previously observed for this family in other contexts[109,110]. 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[111]. 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 virus-node and virus-node family HCT intersection networks in NDEx.

We have previously described the SPP biocuration pipeline, database and web application interface[10]. Figure 9 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. 9A). Figure 9B shows the layout of the consensomes, showing gene symbol, name, percentile ranking and
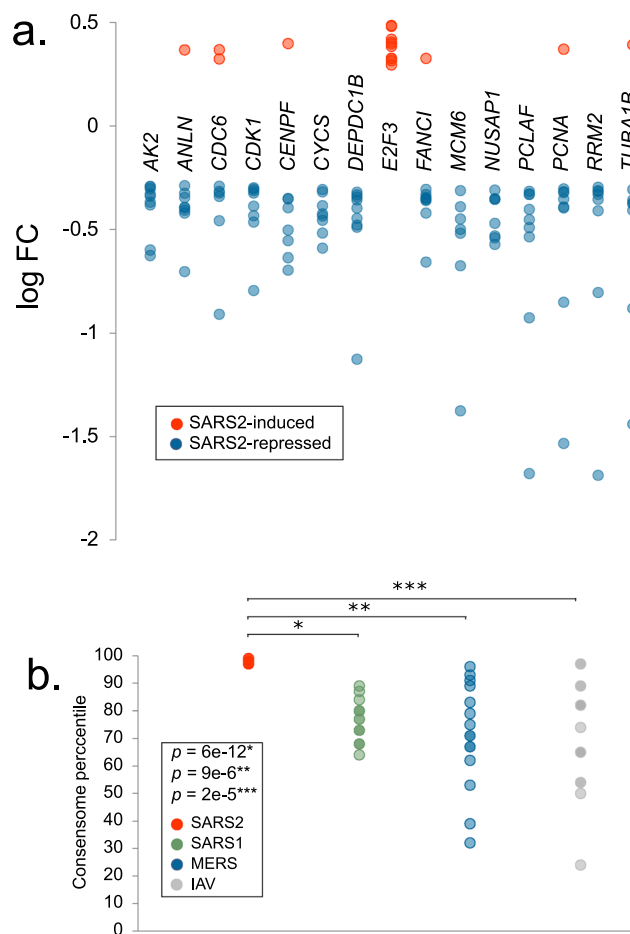
Fig. 8 Hypothesis generation use case 4: efficient SARS2 repression of E2F family HCTs encoding key cell cycle regulators. **(a)** Relative abundance of E2F HCT cell cycle regulators in response to SARS2 infection. **(b)** Comparison of SARS2, SARS1, MERS and IAV consensome percentiles of the E2F HCT 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.

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. 9C). 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. 9D), 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. 9D). 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[112–114]. Figure 10 shows examples of consensome and HCT intersection network visualizations within the NDEx user interface, in which transcripts or nodes, respectively, are organized using SPP's Category-Class-Family classification[10]. 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 info table (Fig. 10b, right panel). Highlighted to exemplify this feature is IL6, an inflammatory ligand that has been previously linked to SARS2 infection-related
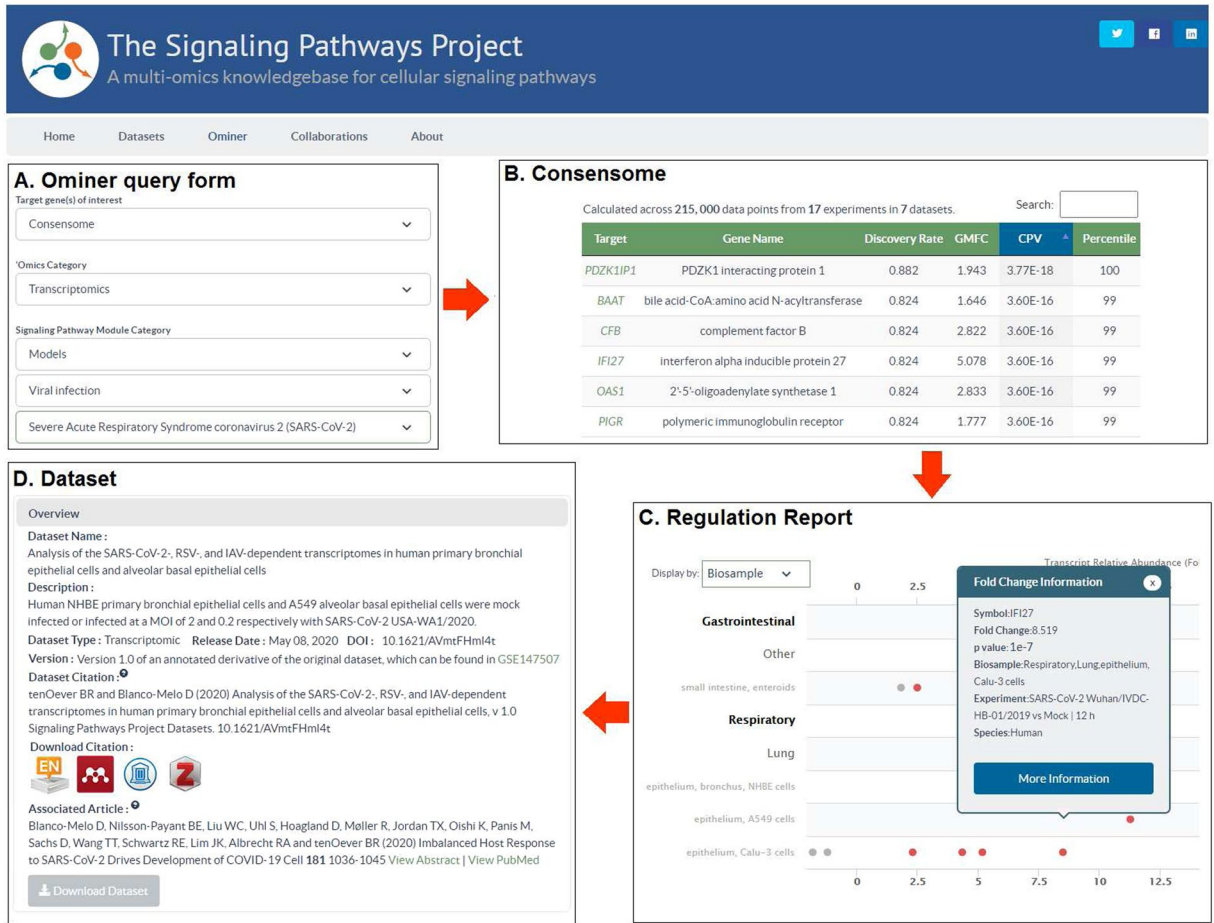
Fig. 9 Mining of CoV consensomes and underlying data points in the SPP knowledgebase. (**A**) The Ominer query form can be configured as shown to access the CoV infection consensomes. In the example shown, the user wishes to view the SARS2 consensome. (**B**) Consensomes are displayed in a tabular format. Target transcript symbols in the consensomes link to SPP transcriptomic Regulation Reports (**C**) 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**). 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).

pathology[8,115]. Consensome GMFCs are signless with respect to direction of regulation[10]. Researchers can therefore follow the SPP link in the info table (Fig. 10b, red arrow 4) to view the individual underlying viral infection data points on the SPP site (Fig. 9C 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*[116]) 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, Fig. 10d shows the top 5% ranked nodes in the SARS2 node HCT ChIP-Seq intersection network (see figshare File F1[12], section 7; see also Figs. 2 and 3 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 NDEx interface leverages the SPP classification system to provide visual 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. 10c; Table 1) that the single largest class of SARS2 HCTs encodes 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 compared to SARS1, SARS2 infection may be relatively efficient in modulating a transcriptional inflammatory response in host cells.
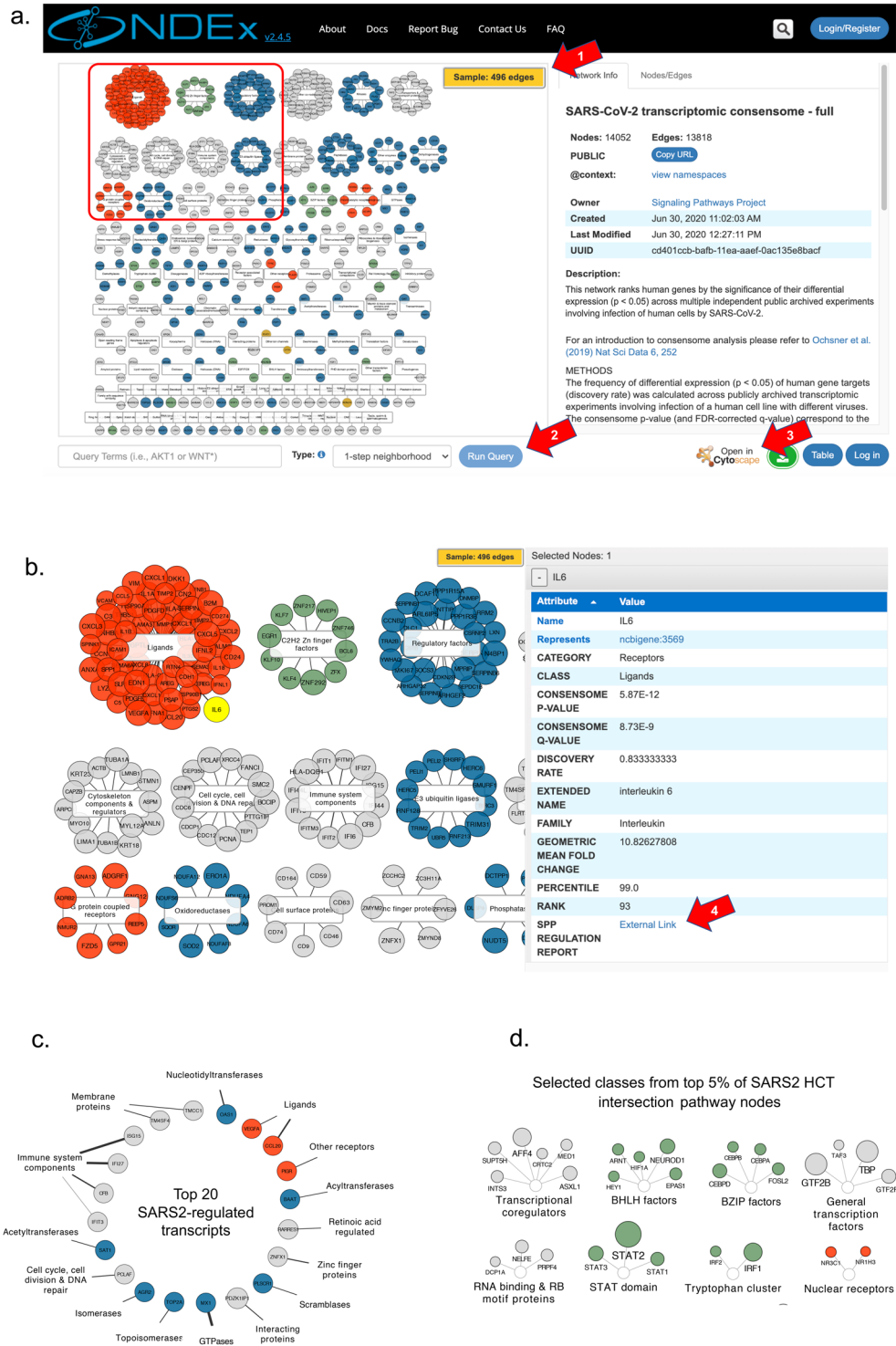
Fig. 10 Visualization of viral consensomes and HCT intersection networks in the NDEx repository. In all panels, transcripts (consensome networks; panels a–c) and nodes (HCT intersection network; panel d) are color-coded according to their category as follows: receptors (orange), enzymes (blue), transcription 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 inset refers to the zoomed-in view in panel (**b**). The IL6 transcript is highlighted to show the contextual information available in the info table to the right. (**c**) Top 20 ranked transcripts in the SARS2 consensome. Edge thickness is 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.

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[117], 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 HCT-based analyses that users are empowered to carry out to illuminate principles of CoV infection signaling.

Previous network analyses across independent viral infection transcriptomic datasets, although valuable, have been limited to stand-alone studies[118,119]. 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 transcriptional 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[120], 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, integration of existing[34,70,111] 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 (figshare File F2[12]), as well as the enrichment of EMT genes among those with elevated rankings in the SARS2 consensome (Fig. 7). 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)[121], 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[122] to identify key research resources of relevance to our study.

Dataset biocuration. Datasets from Gene Expression Omnibus (RRID: SCR_005012) and Array Express (RRID: SCR_002964) were biocurated as previously described, with the incorporation of an additional classification of peptide ligands[123] to supplement the existing mappings derived from the International Union of Pharmacology Guide To Pharmacology (RRID: 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[124]. 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[124] and edgeR analysis[125] 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[126] 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 15c database and made available on the SPP user interface as previously described[10].

Statistical analysis. High confidence transcriptional target intersection analysis was performed using the Bioconductor GeneOverlap analysis package[18] (RRID: 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)[127]. 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 v 7.0 (RRID: 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[128] (RRID: 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 transcriptional targets) was used as the input for the HCT intersection analysis.

SPP web application. The SPP knowledgebase (RRID: 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 15c 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 (RRID: 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 figshare File F1[12], section 1. Consensome data points are in figshare File F1[12], sections 2–5.

SPP SPP MERS[129], SARS1[130], SARS2[131] and IAV[132] 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 BY 4.0 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 versions of consensomes (MERS[133], SARS1[134], SARS2[135] and IAV[136]) and node family (MERS[137], SARS1[138], SARS2[139] and IAV[140]) and node (MERS[141], SARS1[142], SARS2[143] and IAV[144]) HCT intersection networks are freely available in the NDEx repository and assigned DOIs as shown in Table 1. 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.

Code availability

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

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* **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* **369**, 50–54 (2020).
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. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. A transcriptional regulatory atlas of coronavirus infection of human cells. *figshare* https://doi.org/10.6084/m9.figshare.c.4949427 (2020).
13. 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).
14. Doyle, T. *et al.* The interferon-inducible isoform of NCOA7 inhibits endosome-mediated viral entry. *Nat. Microbiol.* **3**, 1369–1376 (2018).
15. Chapgier, A. *et al.* A partial form of recessive STAT1 deficiency in humans. *J. Clin. Invest.* **119**, 1502–1514 (2009).
16. Gruter, P. *et al.* TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* **1**, 649–659 (1998).
17. Wei, J. *et al.* Genome-wide CRISPR screen reveals host genes that regulate SARS-CoV-2 infection. Preprint at, https://doi.org/10.1101/2020.06.16.155101 (2020).
18. Shen, L. GeneOverlap: Test and visualize gene overlaps. R package version 1.24.0, http://shenlab-sinai.github.io/shenlab-sinai/ (2020).
19. 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).
20. Hensley, L. E. *et al.* Interferon-beta 1a and SARS coronavirus replication. *Emerg. Infect. Dis.* **10**, 317–319 (2004).
21. Appelberg, S. *et al.* Dysregulation in mTOR/HIF-1 signaling identified by proteo-transcriptomics of SARS-CoV-2 infected cells. Preprint at, https://doi.org/10.1101/2020.04.30.070383 (2020).
22. 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).

23. Feldman, M. *et al.* Trials of anti-tumour necrosis factor therapy for COVID-19 are urgently needed. *Lancet* **395**, 1407–1409 (2020).
24. Zheng, K., Kitazato, K. & Wang, Y. Viruses exploit the function of epidermal growth factor receptor. *Rev. Med. Virol.* **24**, 274–286 (2014).
25. 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).
26. 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).
27. Hayward, S. D. Viral interactions with the Notch pathway. *Semin. Cancer Biol.* **14**, 387–396 (2004).
28. 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).
29. Rizzo, P. *et al.* COVID-19 in the heart and the lungs: could we 'Notch' the inflammatory storm? *Basic Res. Cardiol.* **115**, 31 (2020).
30. 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).
31. Marik, P. E., Kory, P. & Varon, J. Does vitamin D status impact mortality from SARS-CoV-2 infection? *Med. Drug Discov.*, https://doi.org/10.1016/j.medidd.2020.100041 (2020).
32. 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. *Aliment Pharmacol Ther* **51**, 1434–1437 (2020).
33. Ledford, H. Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature* **469**, https://doi.org/10.1038/d41586-020-01824-5 (2020).
34. Stukalov, A. *et al.* Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. Preprint at, https://doi.org/10.1101/2020.06.17.156455 (2020).
35. 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).
36. 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).
37. 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).
38. Chiang, H.-S. & Liu, H. M. The Molecular Basis of Viral Inhibition of IRF- and STAT-Dependent Immune Responses. *Front. Immunol.* **9**, 3086 (2018).
39. 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).
40. Garcia-Sastre, A. *et al.* The role of interferon in influenza virus tissue tropism. *J. Virol.* **72**, 8550–8558 (1998).
41. 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).
42. Boudewijns, R. *et al.* STAT2 signaling as double-edged sword restricting viral dissemination but driving severe pneumonia in SARS-CoV-2 infected hamsters. Preprint at, https://doi.org/10.1101/2020.04.23.056838 (2020).
43. Haas, D. A. *et al.* Viral targeting of TFIIB impairs de novo polymerase II recruitment and affects antiviral immunity. *Plos Pathog.* **14**, e1006980 (2018).
44. Gelev, V. *et al.* A new paradigm for transcription factor TFIIB functionality. *Sci. Rep.* **4**, 3664 (2014).
45. Haviv, I., Shamay, M., Doitsh, G. & Shaul, Y. Hepatitis B virus pX targets TFIIB in transcription coactivation. *Mol. Cell. Biol.* **18**, 1562–1569 (1998).
46. 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).
47. Grossel, M. J. & Hinds, P. W. Beyond the cell cycle: a new role for Cdk6 in differentiation. *J. Cell. Biochem.* **97**, 485–493 (2006).
48. 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).
49. 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).
50. 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).
51. 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).
52. Aso, H. *et al.* Multiomics investigation revealing the characteristics of HIV-1-infected cells *in vivo*. *Cell Rep.* **32**, 107887 (2020).
53. Zaborowska, J., Isa, N. F. & Murphy, S. P-TEFb goes viral. *Bioessays* **38**(Suppl 1), S75–85 (2016).
54. 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).
55. Takahashi, K. *et al.* DNA topoisomerase 1 facilitates the transcription and replication of the Ebola virus genome. *J. Virol.* **87**, 8862–8869 (2013).
56. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
57. 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).
58. Verdecchia, P., Cavallini, C., Spanevello, A. & Angeli, F. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *Eur. J. Intern. Med.* **76**, 14–20 (2020).
59. 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. Preprint at, https://doi.org/10.1101/2020.04.14.042002 (2020).
60. 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* **181**, 1016–1035 (2020).
61. Zhu, H. *et al.* Clinical analysis of 10 neonates born to mothers with 2019-nCoV pneumonia. *Transl. Pediatr.* **9**, 51–60 (2020).
62. Sappenfield, E., Jamieson, D. J. & Kourtis, A. P. Pregnancy and susceptibility to infectious diseases. *Infect. Dis. Obstet. Gynecol.* **2013**, 752852 (2013).
63. Siston, A. M. *et al.* Pandemic 2009 influenza A(H1N1) virus illness among pregnant women in the United States. *JAMA* **303**, 1517–1525 (2010).
64. 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).
65. 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.* https://doi.org/10.1016/j.ajogmf.2020.100118 (2020).
66. Sutton, D., Fuchs, K., D'Alton, M. & Goffman, D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *N Engl J Med* **382**, 2163–2164 (2020).

67. 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).
68. Richards, J. S., Russell, D. L., Ochsner, S. & Espey, L. L. Ovulation: New dimensions and new regulators of the inflammatory-like response. *Annu Rev Physiol* **64**, 69–92 (2002).
69. Ghandehari, S. Progesterone for the Treatment of COVID-19 in Hospitalized Men. https://clinicaltrials.gov/ct2/show/NCT04365127 (2020).
70. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
71. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).
72. 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 Health Dis* **3**, 31–35 (2019).
73. Li, H. *et al.* Alveolar epithelial cells undergo epithelial-mesenchymal transition in acute interstitial pneumonia: a case report. *BMC Pulm. Med.* **14**, 67 (2014).
74. Gouda, M. M., Shaikh, S. B. & Bhandary, Y. P. Inflammatory and Fibrinolytic System in Acute Respiratory Distress Syndrome. *Lung* **196**, 609–616 (2018).
75. George, P., Wells, A. & Jenkins, G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir Med* **8**, 807–815 (2020).
76. Adair, L. B. II & Ledermann, E. J. Chest CT Findings of Early and Progressive Phase COVID-19 Infection from a US Patient. *Radiol Case Rep* **15**, 819–824 (2020).
77. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
78. 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).
79. 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).
80. 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).
81. Xu, J., Lamouille, S. & Derynck, R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* **19**, 156–172 (2009).
82. Zhao, M., Kong, L., Liu, Y. & Qu, H. dbEMT: an epithelial-mesenchymal transition associated gene resource. *Sci. Rep.* **5**, 11459 (2015).
83. 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).
84. 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**, e01256–17 (2017).
85. 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).
86. 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).
87. 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).
88. Gao, F., Kinnula, V. L., Myllarniemi, M. & Oury, T. D. Extracellular superoxide dismutase in pulmonary fibrosis. *Antioxid. Redox Signal.* **10**, 343–354 (2008).
89. Kato, N. *et al.* Basigin/CD147 promotes renal fibrosis after unilateral ureteral obstruction. *Am. J. Pathol.* **178**, 572–579 (2011).
90. Ruster, C. & Wolf, G. Angiotensin II as a morphogenic cytokine stimulating renal fibrogenesis. *J. Am. Soc. Nephrol.* **22**, 1189–1199 (2011).
91. 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).
92. 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. Preprint at, https://doi.org/10.1101/2020.05.28.122291 (2020).
93. 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* **7**, 689–697 (1996).
94. Harbour, J. W. & Dean, D. C. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev* **14**, 2393–2409 (2000).
95. Reimer, D. *et al.* E2F3a is critically involved in epidermal growth factor receptor-directed proliferation in ovarian cancer. *Cancer Res.* **70**, 4613–4623 (2010).
96. Classon, M. & Harlow, E. The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* **2**, 910–917 (2002).
97. 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).
98. Shivji, K. K., Kenny, M. K. & Wood, R. D. Proliferating cell nuclear antigen is required for DNA excision repair. *Cell* **69**, 367–374 (1992).
99. Borlado, L. R. & Méndez, J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* **29**, 237–243 (2008).
100. 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).
101. Vanden Bosch, A. *et al.* NuSAP is essential for chromatin-induced spindle formation during early embryogenesis. *J. Cell Sci.* **123**, 3244–3255 (2010).
102. Engeland, K. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ* **25**, 114–132 (2018).
103. 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).
104. 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).
105. 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).
106. 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).
107. Fan, Y., Sanyal, S. & Bruzzone, R. Breaking Bad: How Viruses Subvert the Cell Cycle. *Front. Cell. Infect. Microbiol.* **8**, 396 (2018).
108. Gasser, S. & Raulet, D. H. The DNA damage response arouses the immune system. *Cancer Res.* **66**, 3959–3962 (2006).
109. 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).
110. To, B. & Andrechek, E. R. Transcription factor compensation during mammary gland development in E2F knockout mice. *Plos One* **13**, e0194937 (2018).
111. Bouhaddou, D. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182**, 1–28 (2020).

112. 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).

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

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

115. 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.*, https://doi.org/10.1093/cid/ciaa449 (2020).

116. 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).

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

118. 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).

119. 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).

120. Wilk, C. M. Coronaviruses hijack the complement system. *Nat. Rev. Immunol.* **20**, 350.

121. 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).

122. 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).

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

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

125. 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).

126. 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).

127. 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).

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

129. Ochsner, S. A. & McKenna N. J. Middle East respiratory syndrome coronavirus (MERS-CoV) transcriptomic consensome, version 1. *Signaling Pathways Project*, https://doi.org/10.1621/jgxM527b8s.1 (2020).

130. Ochsner, S. A. & McKenna N. J. Sudden acute respiratory syndrome coronavirus 1 (SARS-CoV-1) transcriptomic consensome, version 1. *Signaling Pathways Project*, https://doi.org/10.1621/k9ygy4i49j.1 (2020).

131. Ochsner, S. A. & McKenna N. J. Sudden acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transcriptomic consensome, version 1. *Signaling Pathways Project*, https://doi.org/10.1621/vTiy8d4Iq7.1 (2020).

132. Ochsner, S. A. & McKenna N. J. Influenza A virus (IAV) transcriptomic consensome, version 1. *Signaling Pathways Project*, https://doi.org/10.1621/58AOyXDIAH.1 (2020).

133. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. MERS-CoV transcriptomic consensome - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9QG7S (2020).

134. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. SARS-CoV-1 transcriptomic consensome - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9KP4G (2020).

135. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. SARS-CoV-2 transcriptomic consensome - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9G02W (2020).

136. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. IAV transcriptomic consensome - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9B60Z (2020).

137. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: MERS-CoV vs pathway node family (transcriptomic) – all families. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9PG63 (2020).

138. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: SARS-CoV-1 vs pathway node family (transcriptomic) – all families. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9JS46 (2020).

139. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: SARS-CoV-2 vs pathway node family (transcriptomic) – all families. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9F016 (2020).

140. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: IAV vs pathway node family (transcriptomic) – all families. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9989R (2020).

141. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: MERS-CoV vs pathway node (ChIP-Seq) - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N96G6R (2020).

142. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: SARS-CoV-1 vs pathway node (ChIP-Seq) - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N92P56 (2020).

143. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: SARS-CoV-2 vs pathway node (ChIP-Seq) - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9Z01V (2020).

144. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. HCT overlap: IAV vs pathway node (ChIP-Seq) - full. *The Network Data Exchange – NDEx* https://doi.org/10.18119/N9T609 (2020).

145. Muzikar, K., Nickols, N. & Dervan, P. Analysis of the dexamethasone (Dex)-dependent transcriptome in A549 lung adenocarcinoma cells. *Signaling Pathways Project*, https://doi.org/10.1621/xigKzGn1se (2015).

Acknowledgements

Author contributions

Dataset biocuration: S.O. Data analysis: S.O., R.P. and N.M. Manuscript drafting: N.M. Manuscript editing: N.M., R.P. and S.O.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.J.M.

Reprints and permissions information is available at www.nature.com/reprints.

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