

RESEARCH ARTICLE

Multiscale network analysis identifies potential receptors for SARS-CoV-2 and reveals their tissue-specific and age-dependent expression

Christian V. Forst^{1,2,3,4} , Lu Zeng¹, Qian Wang^{1,2,3}, Xianxiao Zhou^{1,2,3}, Sezen Vatansever^{1,2,3}, Peng Xu^{1,2,3}, Won-Min Song^{1,2,3}, Zhidong Tu^{1,3} and Bin Zhang^{1,2,3,5}

- 1 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
 2 Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, New York, NY, USA
 3 Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
 4 Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
 5 Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Correspondence

C. V. Forst, Department of Genetics & Genomic Sciences, Department of Microbiology, Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, 1470 Madison Avenue, New York, NY 10029, USA
 Tel: +1 212 824 8948
 E-mail: christian.forst@mssm.edu

The coronavirus disease 2019 (COVID-19) pandemic has affected tens of millions of individuals and caused hundreds of thousands of deaths worldwide. Here, we present a comprehensive, multiscale network analysis of the transcriptional response to the virus. In particular, we focused on key regulators, cell receptors, and host processes that were hijacked by the virus for its advantage. ACE2-controlled processes involved CD300e (a TYROBP receptor) as a key regulator and the activation of IL-2 pro-inflammatory cytokine signaling. We further investigated the age dependency of such receptors in different tissues. In summary, this study provides novel insights

Abbreviations

A549, adenocarcinomic human alveolar basal epithelial cells; ACAT, aggregated Cauchy association test; ACE2, angiotensin-converting enzyme 2; ACE2oe, ACE2 overexpression; AKT, AKT serine/threonine kinase; ARCHS⁴, All RNA-seq and ChIP-seq sample and signature search; avMx, average module expression; BALF, bronchoalveolar lavage fluid; BH, Benjamini and Hochberg method; BTK, Bruton tyrosine kinase; C14orf119, Chromosome 14 open reading frame 119; Calu3, non-small-cell lung cancer cell line; CCDC106, coiled-coil domain containing 106; CD300e, CD300e molecule; CD81, CD81 molecule; CEACAM1, CEA cell adhesion molecule 1; CENPBD1, CENPB DNA-binding domain containing 1; CLOCK, clock circadian regulator; CoV, Coronavirus; COVID-19, Coronavirus disease 2019; CRS, composite receptor score; CTA, cluster-trait association analysis; CTSL, Cathepsin L; CTSZ, Cathepsin Z; dbGaP, database of genotypes and phenotypes; DDX60, DEXD/H-box helicase 60; DDX60L, DEXD/H-box helicase 60 like; DEG, differentially expressed gene; DOHH, deoxyhypusine hydroxylase; DTX3, Deltex E3 ubiquitin ligase 3; EGF, epidermal growth factor; EGFR, EGF receptor; ENTPD3, ectonucleoside triphosphate diphosphohydrolase; EPHA6, EPH receptor A6; ERBB4, erb-b2 receptor tyrosine kinase 4; EXOC7, exocyst complex component 7; FC, fold change; FCGRT, Fc gamma receptor and transporter; FDR, false discovery rate; FE, fold enrichment; FET, Fisher's exact test; FOXN2, Forkhead box N2; PPFNC, fast planar filtered network construction; GABRP, γ -aminobutyric acid type A receptor subunit π ; GO, gene ontology; GTE_x, genotype-tissue expression; HBD, hemoglobin subunit δ ; HBG1, hemoglobin subunit γ 1; HBM, hemoglobin subunit μ ; IFI44, interferon-induced protein 44; IFITM1, interferon-induced transmembrane protein 1; IFN, interferon; IL-n, interleukin-n; IMPA2, inositol monophosphatase 2; IRF, interferon regulatory factor; ISG, interferon stimulated gene; ITGB2, integrin subunit beta 2; JAK, Janus kinase; LAMP3, lysosome-associated membrane glycoprotein 3; MAP2K5, mitogen-activated protein kinase kinase 5; MCA, multiscale clustering analysis; MEGENA, multiscale embedded gene co-expression network analysis; MERS, middle east respiratory syndrome; MFSD3, major facilitator superfamily domain containing 3; MHA, multiscale hub analysis; MSigDB, the molecular signatures database; NFKBIL1, NF- κ B inhibitor like 1; NF- κ B, nuclear factor κ B; NHBE, normal human bronchial epithelial (cells); NRP1, neuropilin 1; NYU, New York University; PASC, post-acute sequelae of COVID-19; PBMC, peripheral blood mononuclear cell; PC, principal component; PEER, probabilistic estimation of expression residuals; PI3K, phosphatidylinositol 3-kinase; PMI, post-mortem interval; PRR, pattern recognition receptor; PTPRM, protein tyrosine phosphatase receptor type M; RGS11, regulator of G protein signaling 11; RIN, RNA integrity number; RSV, respiratory syncytial virus; SARS, severe acute respiratory syndrome; STAT, signal transducer and activator of transcription; STSPR, SARS-CoV-2 triggered surface protein receptor; STSPR-AD, STSPR age dependence; STSPR-DEAD, STSPR differential expression and age dependence (score); TCIM, transcriptional and immune response regulator; THBS3, thrombospondin 3; THEMIS2, thymocyte selection associated family member 2; TLR, toll-like receptor; TMEM201, transmembrane protein 201; TMPRSS2, transmembrane serine protease 2; TNF, tumor necrosis factor; TNFRSF25, TNF receptor superfamily member 25; TYROBP, TYRO protein tyrosine kinase binding protein; Vero E6, African green monkey kidney cell line, clone E6; ZNFnnn, zinc finger protein nnn.

(Received 11 July 2022, revised 13 February 2023, accepted 27 February 2023, available online 4 April 2023)

doi:10.1002/1873-3468.14613

Edited by John Briggs

into the gene regulatory organization during the SARS-CoV-2 infection and the tissue-specific, age-dependent expression of the cell receptors involved in COVID-19.

Keywords: age dependency; CD300e; COVID-19; *in silico* validation; multiscale modeling

On December 31, 2019, the WHO was notified about a cluster of novel pneumonia cases in Wuhan City, Hubei Province of China. The causative agent was linked to a novel by Chinese authorities on January 7, 2020, inducing the activation of the R&D Blueprint as part of WHO's response to the outbreak. Coronaviruses (CoVs) belong to the group of enveloped, single, positive-stranded RNA viruses causing mild-to-severe respiratory illnesses in humans [1]. In the past two decades, two worldwide outbreaks have originated from CoVs (severe acute respiratory syndrome, SARS; Middle East respiratory syndrome, MERS) capable of infecting the lower respiratory tract, resulting in heightened pathogenicity and high mortality rates [2]. We are currently amid a third pandemic caused by a new CoV strain, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19). In the majority of cases, patients exhibit either no or mild symptoms, whereas in more severe cases, patients may develop severe lung injury and die from respiratory failure [2,3].

A viral infection generally triggers a physiological response at the cellular level after the initial replication of the virus [4]. The cellular system has an arsenal of pattern recognition receptors (PRRs) [5] at its disposal that guard against various microbes inside and outside of the cell. PRRs bind distinct structural features that are conserved among different pathogens [6]. In a viral infection, intracellular PRRs are detecting viral RNA defective particles that are often formed during virus replication [7]. Pathogen detection assembles the initial steps of a signaling cascade to activate downstream transcription factors, such as interferon regulator factors (IRFs) and nuclear factor κ B (NF- κ B) [6,8], which causes the activation of two general antiviral processes [6]. The first, predominantly intracellular, process initiates cellular defenses *via* transcriptional induction of type I and III interferons (IFN-I and IFN-III, respectively). Subsequently, IFN upregulates IFN-stimulated genes (ISGs) with antiviral properties [9]. The second, intercellular cascade of antiviral counteraction refers to the recruitment and coordination of a multitude of

leukocytes. Chemokine secretion [10,11] orchestrates this concerted action of immune-system countermeasures. This finding is confirmed when specifically comparing the host response between influenza virus (nine datasets), rhinovirus (two datasets), RSV (three datasets), SARS-CoV-1 (three datasets), and SARS-CoV-2 (24 datasets) in blood (whole blood; peripheral blood mononuclear cells, PBMCs), as well as respiratory samples and cells [nasal wash, bronchoalveolar lavage, Normal Human Bronchial Epithelial (NHBE) cells], as performed by Smith *et al.* [12]. Most conserved pathways across all five respiratory viruses are innate immune response signaling, such as the Janus kinase (JAK)—signal transducer and activator of transcription (STAT), tumor necrosis factor (TNF), and interleukin-17 (IL-17) signaling pathways. The top five upregulated host genes shared across all five respiratory viruses are DEXD/H-box helicase 60 (*DDX60*) and DEXD/H-box helicase 60 like (*DDX60L*), forkhead box N2 (*FOXN2*), interferon-induced protein 44 (*IFI44*), and Janus kinase 2 (*JAK2*). The top five common downregulated genes are coiled-coil domain containing 106 (*CCDC106*), Fc gamma receptor and transporter (*FCGRT*), inositol monophosphatase 2 (*IMPA2*), major facilitator superfamily domain containing 3 (*MFS3*), and zinc finger protein 219 (*ZNF219*). Upregulated genes of SARS-CoV-2 response that are downregulated in the response of the other four respiratory viruses are CENPB DNA-binding domain containing 1 (*CENPBD1*), deltex E3 ubiquitin ligase 3 (*DTX3*), mitogen-activated protein kinase kinase 5 (*MAP2K5*), NF- κ B inhibitor like 1 (*NFKBIL1*), regulator of G protein signaling 11 (*RGS11*), thrombospondin 3 (*THBS3*), and zinc finger protein 581 (*ZNF581*). *NFKBIL1* is involved in the regulation of innate immune response as negative regulator of TLR and IRF signaling pathways, thus quenching IFN response during SARS-CoV-2 infection. Another significantly upregulated, SARS-CoV-2-specific, gene, although shared with respiratory syncytial virus (RSV), is transcriptional and immune response regulator (*TCIM*). *TCIM* enhances NF- κ B activity inducing upregulation of cytokines involved in

inflammation. This commonality with RSV as well as high similarity in other differentially regulated genes is shared between these two viruses together with the resulting ‘cytokine storm’ and the caused damage to the respiratory tract of SARS-CoV-2 and RSV patients [12].

The selection pressure induced by such a broad antiviral response of the host and the evolvability of viruses has resulted in countless viral countermeasures [13]. Thus, the host response to a virus is generally not uniform. Viral infections can cause a spectrum of various degrees of morbidity and mortality.

Thair *et al.* [14] investigated host inflammation and compared the systemic immune response from blood transcriptomic data of patients suffering from a spectrum of viral infections, including COVID-19. Common pathways, which are all upregulated are related to neutrophil activation, response to virus, regulation of innate immune response, and type I IFN signaling. All these pathways are well-known inflammatory signatures [15,16]. However, downregulated pathways differentiate between SARS-CoV-2 and non-SARS-CoV-2-caused viral diseases. Ribosome-related processes are highly significantly downregulated in COVID-19 but not in non-COVID-19 viral infections. On the other contrary, cell–cell adhesion, cell activation, leukocyte activation, and immune response-activating cell surface receptor signaling are more significant in non-COVID-19 infections compared with COVID-19 infections [14].

Additional factors, such as sex, age, and other genetic factors, contribute to the diversity of immune response. Concerning COVID-19, age has been identified as the most significant risk factor in the mortality of patients. The overall symptomatic case fatality risk (the probability of dying after developing symptoms) of COVID-19 in Wuhan was 1.4% (0.9–2.1%) as of February 29, 2020. Compared with those aged 30–59 years, those aged below 30 and above 59 years were 0.6 (0.3–1.1) and 5.1 (4.2–6.1) times more likely to die after developing symptoms [17]. Similar data were reported for the United States. From February 12 to March 16, 2020, the Center for Disease Control (CDC) estimated a case fatality rate of patients 55–64 years old with 1.4–2%. This rate was 10.4–27.3% for patients 85 years or older [18].

To better understand the molecular basis of the disease, we sought to characterize the transcriptional response to infection in both *in vitro* cell systems (tissue cultures and primary cells) and *in vivo* samples derived from COVID-19 patients. We employed an integrative network-based approach to identify host response co-expression networks in SARS-CoV-2 infection. In particular, we investigated functional

processes and key regulators affected by this specific virus, receptors, and processes hijacked for enabling the viral life cycle. We further studied the age dependence of targets, mainly receptors that the virus utilizes for entry and its life cycle.

Materials and methods

RNA-seq analysis

Raw reads were obtained from the Beijing Institute of Genomics (BIG) Data Center (<https://bigd.big.ac.cn/>) under the accession number CRA002390. BALF RNA-seq data from healthy subjects were obtained from NCBI/SRA (SIB028/SRR10571732, SIB030/SRR10571730, and SIB036/SRR10571724). The RNA-seq data were aligned to the Homo sapiens reference genome GRCh38/hg19 using the Star aligner v2.7.0f with modified ENCODE options, according to Xiong *et al.* [19]. Raw read counts were calculated using featureCounts v2.0.1. Raw read counts after Star alignment and featureCounts, as well as obtained from GSE147507, CEDGER/VOOM (v3.32.1 with R v4.0.0).

Single-cell RNA-seq analysis

Filtered bar-coded matrix files from GEO with accession number GSE145926 were integrated with SEURAT into a SEURAT object and subjected to the standard SEURAT pipeline (SEURAT v4.0.6) for scaling and clustering. BALF cell types were called according to the cell markers based on Liao *et al.* [20] (Fig. S2A). We used *AddModuleScore()* to calculate module scores on single-cell level aggregated by cell types. Function *DotPlot()* has been employed to calculate (scaled) average expression and percent expressed values of gene modules for each cell type (see Fig. S2B).

Identification of differentially expressed genes

We used the negative binomial models together with the empirical Bayes approach as implemented in the EDGER-package [21] to identify differentially expressed genes (DEGs). We considered an absolute fold change of 1.5 or higher and an FDR ≤ 0.05 as significant throughout the paper.

Gene co-expression network analysis

Multiscale Embedded Gene Co-Expression Network Analysis (MEGENA) [22] was performed to identify host modules of highly co-expressed genes in SARS-CoV-2 infection. The MEGENA workflow comprises four major steps: (a) Fast Planar Filtered Network construction (FPFNC), (b) Multiscale Clustering Analysis (MCA), (c) Multiscale Hub Analysis (MHA), (d) and Cluster-Trait Association

Analysis (CTA). The total relevance of each module to SARS-CoV-2 infection was calculated by using the Product of Ranks method with the combined enrichment of the differentially expressed gene (DEG) signatures as implemented: $G_j = \prod g_{ji}$, where g_{ji} is the relevance of a consensus j to a signature i ; and g_{ji} is defined as $(\max_j(r_{ji}) + 1 - r_{ji}) / \sum r_{ji}$, where r_{ji} is the ranking order of the significance level of the overlap between the module j and the signature.

Identification of enriched pathways and key regulators in the host modules

To functionally annotate gene signatures and gene modules identified in this study, we performed an enrichment analysis of the established pathways and signatures—including the gene ontology (GO) categories and the Molecular Signatures Database (MSigDB)—and the subject area-specific gene sets—including, Inflammation, Interferome, and InnateDB. The hub genes in each subnetwork were identified using the adopted Fisher's inverse chi-square approach in MEGENA; Bonferroni-corrected P -values smaller than 0.05 were set as the threshold to identify significant hubs.

Module ranking

Based on module enrichment for DEGs, we ranked modules using cumulative enrichment scores. In particular, we employed the 'Product of Ranks' (see also 'Gene co-expression network analysis'). The total relevance of each consensus module to influenza infection was calculated by summarizing the P -values of the DEG signature enrichments: $G_j = \prod g_{ji}$, where g_{ji} is the relevance of a consensus j to a signature i . g_{ji} is defined as $(\max_j(r_{ji}) + 1 - r_{ji}) / \sum r_{ji}$, where r_{ji} is the ranking order of the significance level of the overlap between the consensus module j and the signature i . We also explored alternative ranking methods, such as mean, mean of log, mean of log(log), or the Aggregated Cauchy Association Test (ACAT) with rather similar results in ranked modules.

Network enrichment

Fisher's exact test (FET) was performed to determine the overlap between network neighborhoods of potential key regulators (target) and an input DEG signature. For each target in the network in the 95 percentile of node strength (the sum of the edge weights for each node, divided between connected nodes) after MEGENA, the genes in the network neighborhoods between one and four steps away from the target were intersected with the DEG signature. MEGENA networks were tested with DEGs of all systems for further analysis (see the main text). Cumulative network enrichment scores $s = 1/n \cdot \sum_i -\log_{10} P_i$ based on individual FET P -values for each target were calculated. n

is the number of realizations (i.e., the number of different neighborhoods and systems used to calculate the particular score).

GTEX data preprocessing

We downloaded the Genotype-Tissue Expression (GTEx) v8 data [23] from the Database of Genotypes and Phenotypes (dbGaP) under accession phs000424.v8.p2. For all the available tissues, we selected those with at least 80 samples and samples with more than 20 million mapped reads and greater than a 40% mapping rate. Cell line data were removed from our analysis. Only genes with expression > 0.1 Transcripts Per Million (TPM) and aligned read count of 5 or more in more than 80% of samples within each tissue were used for aging gene identification. Expression measurements for each gene in each tissue were subsequently inverse-quantile normalized to the standard normal distribution to reduce the potential impact of outlier gene expression values. Our final dataset included samples from 46 tissue types. The sample size for each tissue ranged from 114 to 706, with an average of 315 samples.

Linear regression model for age and sex-associated gene detection

We implemented a linear regression model to identify age-associated gene expression (Eqn 1) [24].

$$Y_{ij} = \beta_j + \gamma_j \text{Age}_i + \delta_j \text{Sex}_i + \sum_{k=1}^5 \mu_{jk} \text{Genotype}_{ik} + \sum_{k=1}^N \alpha_{jk} \text{PC}_{ik} + \theta_j \text{RIN}_i + \delta_j \text{PMI}_i + \varepsilon_{ij}. \quad (1)$$

In this model, Y_{ij} is the expression level of gene j in sample i , Age_i denotes the donor age of sample i , Sex_i denotes the donor sex for sample i , Genotype_{ik} ($k \in (1, 2, 3, 4, 5)$) denotes the value of the k -th principal component value of the genotype profile from the GTEx data (see GTEx data preprocessing above) for the i -th sample, PC_{ik} ($k \in (1, \dots, N)$) denotes the value of the k -th principal component value of gene expression profile for the i -th sample, N is the total number of top principal components (PCs) under consideration, RIN_i denotes the RNA integrity number (RIN) score of sample i , PMI_i denotes the post-mortem interval (PMI) of sample i , ε_{ij} is the error term, γ_j , δ_j , μ_{jk} , α_{jk} , θ_j , δ_j are the regression coefficients for each covariate. The corresponding correlation coefficients and P -values (adjusted with the Benjamini and Hochberg (BH) [25] method) were then calculated for all genes; FDR values < 0.05 were considered as significant age-associated genes. Several covariates [such as genotype PCs and probabilistic estimation of expression residual (PEER) factors] we adjusted in the regression model were selected following

the method used by the GTEx consortium [23]. From the consortium's analysis, the top five genotype PCs were considered sufficient to capture the major population structure in the GTEx dataset and were used for the consortium paper.

Adjust gene expression for age analysis

We used a linear regression model to adjust gene expression (Eqn 2).

$$Y_{ij} = \beta_j + \delta_j \text{Sex}_i + \mu_j \text{Platform}_i + \theta_j \text{RIN}_i + \delta_j \text{PMI}_i + \varepsilon_{ij} \quad (2)$$

We regressed out the following confounding factors to obtain adjusted gene expression, which include Sex_i : the sex of the donor for sample i , Platform_i : the value of the platform for the i -th sample, RIN_i : the RIN score of sample i , and PMI_i : the PMI of sample i .

Expression measurements for each gene in each tissue were inverse-quantile normalized to follow the standard normal distribution to reduce the potential impact of outlier gene expression values. Composite receptor score (CRS) was then calculated for each receptor in each sample (Eqn 3).

$$\begin{aligned} \text{CRS}(Y_i) &= \text{sum}\{\text{sign}(X_{ij}, \tau)\} \text{ where } \text{sign}(X_{ij}, \tau) \\ &= \begin{cases} 0, & \text{if } X_{ij} < \tau \\ 1, & \text{if } X_{ij} \geq \tau \end{cases} \end{aligned} \quad (3)$$

In this equation, $\text{CRS}(Y_i)$ is the composite score of sample i , X_{ij} is the expression level of gene j in sample i , τ is the test score. We have tested τ with -0.25 , 0 , 0.25 , 0.5 , 0.75 , and 1 , Spearman correlation coefficients and P -values (adjusted with the BH method) were subsequently calculated between CRS score and age. $\tau = 0.25$ showed the overall best correlation and P -value between CRS and age (Table S7). We termed this correlation coefficient between SARS-CoV-2 surface protein receptors (STSPRs) CRS and age, STSPR differential expression, and age dependence (STSPR-DEAD) score.

Results

RNA-seq data from cell lines (NHBE, Normal Human Bronchial Epithelial cells, A549, adenocarcinomic human alveolar basal epithelial cells, and Calu-3, lung adenocarcinoma epithelial cells) and lung biopsies of two patients infected by SARS-CoV-2 were recently made available on NCBI/GEO (GSE147507) [6]. A second, clinical, transcriptomic dataset for a cohort of COVID-19 patients together with uninfected controls has recently been published [19]. Data were obtained from bronchoalveolar lavage fluid (BALF)

and PBMCs (10 samples total: 3 PBMC control, 2×2 BALF infected, 3 PBMC infected). RNA-seq data are available through the Beijing Institute of Genomics (BIG) Data Center (<https://bigd.big.ac.cn/>) under the accession number: CRA002390. We have combined the BALF with the lung biopsy datasets after correction for sample location, yielding datasets containing a total of 11 samples, that is, six infected and five control. In summary, four of the six infected (BALF) samples are from the CRA002390 dataset. The other two COVID-19 (lung biopsy) samples are from the GSE147507 dataset. The latter dataset also includes two samples from healthy controls. The study producing CRA002390 [19] did not generate control data. However, Xiong *et al.* utilized three publically available BALF data of healthy subjects from NCBI/SRA (SRR10571724, SRR10571730, and SRR10571732). We retrieved the identical three datasets and added them to the two control samples from GSE147507 for a total of five control samples. These datasets were processed by an integrative network analysis approach. Data from PBMCs and cell lines were excluded. For validation purposes, we have further secured data from a second cohort of 142 patients from the New York University (NYU) Langone Health Manhattan campus that required invasive mechanical ventilation [26].

Integrative network biology analysis of the β -coronavirus–host system

The basis of our prediction of SARS-CoV-2 processes and the host response is an integrative network analysis approach that combines network inference and network topological methods with molecular signatures. We first identified differentially expressed genes (DEGs) in each dataset that showed significant changes during SARS-CoV-2 infection. The biological functions of DEG signatures from each dataset were assessed by gene set enrichment methods. Given the particular interest in human patients' COVID-19 response, we used a corresponding subset of transcriptome data to infer multiscale gene co-expression MEGENA networks. We ranked MEGENA network modules based on their enrichment for SARS-CoV-2 infection responding DEGs. MEGENA modules were functionally assessed by GO, MSigDB, and blood cell-type-specific gene sets. We also investigated the underlying network topological structure by testing the network neighborhood of target genes for enrichment by SARS-CoV-2 DEGs and signatures responding to angiotensin-converting enzyme 2 (ACE2) overexpression (see Materials and methods). Finally, we analyzed

the age dependency of molecular processes during SARS-CoV-2 infection by employing a linear regression model on baseline gene expression using Genotype-Tissue Expression (GTEx) data.

Molecular signatures of SARS-CoV-2 infection

We have identified 572 up-, and 1338 downregulated DEGs from patient-derived lung biopsy, as well as 3573 up- and 1630 downregulated DEGs from human patient BALF expression data. 2382 DEGs are upregulated, and 2526 DEGs are downregulated in adenocarcinomic human alveolar basal epithelial (A549) cell lines [2017 up- and 2354 downregulated in non-small-cell lung cancer (Calu3) cell lines, resp.]. The exceptions are the NHBE and the first batch (Series 2) of the A549 data (GSE147507), which yielded a fraction of significant DEGs, with 144 genes up- and 55 genes downregulated in NHBE cells as well as 88 genes up- and 14 genes downregulated in A549 (Series 2). All datasets have comparable numbers of samples. DEGs were considered significant with $FDR \leq 0.05$ and a fold change of 1.5 or higher.

As others have already noted [6], there is a lack of *ACE2* expression in cell line data. A key protein relevant for SARS-CoV-2 entry as well as an ISG, *ACE2* is not significantly expressed in cell lines (S5_A549: 3.2-fold, $FDR = 0.15$; Calu3: 0.77-fold, $FDR = 0.12$; NHBE: 1.2-fold, $FDR = 0.52$). Only in the lung biopsy (27.6-fold, $FDR = 3.70 \times 10^{-6}$) and in BALF (50.5-fold, $FDR = 0.066$), we were able to identify significant expression fold change between healthy/Mock control and infection. According to GTEx data, *ACE2* baseline expression is observed in the small intestine (Terminal Ileum), female breast, thyroid, subcutaneous adipose tissue, testis, and coronary artery (Table S1). A detailed, single-cell-based study identified that *ACE2* and transmembrane serine protease 2 (*TMPRSS2*) are primarily expressed in bronchial transient secretory cells [27]. *TMPRSS2* expression is inconsistent in our datasets. It is highly upregulated in BALF (47.2-fold, $FDR = 2.98 \times 10^{-4}$) and upregulated in Calu3 cells (2.13-fold, $FDR = 2.71 \times 10^{-3}$), but downregulated in lung biopsy samples (0.16-fold, $FDR = 8.91 \times 10^{-7}$). The third known receptor and host factor neuropilin-1 (*NRP1*) [28,29] is downregulated in BALF (2.0-fold, $FDR = 5.52 \times 10^{-4}$) and marginally upregulated in A549 (1.37-fold, $FDR = 5.02 \times 10^{-3}$) and Calu3 (1.25-fold, $FDR = 0.096$) cells. As we are interested in an organismal response, our primary focus is on samples of human patients.

To validate our findings, we compared DEGs called during our analysis of human patient samples and

results from the NYU COVID-19 study [26]. For this purpose, we employed Super Exact Test [30], a generalization of Fisher's Exact Test to evaluate the set overlap of multiple sets. BALF and lung biopsy data show significant overlap with NYU COVID-19 data (Fig. S1).

Receptors, host factors, and biological processes required for the viral life cycle

Given that *ACE2* is essential for SARS-CoV-2 entry [31], and further, the viral life cycle, we hypothesize that *ACE2* expression may trigger other processes relevant to the viral life cycle. As we have established in the previous section that *ACE2* is indeed upregulated in human lung samples (both BALF and lung biopsy), we were interested in the effect of *ACE2* expression. To determine which receptors and targets are involved in such processes, we performed a network enrichment analysis using the *ACE2* overexpression (*ACE2oe*) signatures from the Blanco-Melo *et al.* dataset [6] and identified genes that potentially serve as novel host receptors and targets potentially facilitating the entry of the SARS-CoV-2 into the host cell and are required for the viral life cycle. For this purpose, we constructed a multiscale co-expression network to investigate co-expression and co-regulation relationships among genes underlying SARS-CoV-2 infection. In particular, we were interested in the organismal response from patients infected by SARS-CoV-2. Thus, we combined the available datasets from BALF and lung biopsies to construct a multiscale co-expression network of 13 398 genes and 35 483 interactions using MEGENA [22] (Fig. 1A). This co-expression network includes 900 modules. The majority of the top-ranked modules (using DEGs from both patient and cell data by excluding the *ACE2* overexpression dataset; see Materials and methods section) are enriched for well-known biological functions related to viral infection, including cell cycle, ribosome/translation, NF- κ B canonical pathway, or cytokine signaling. The 20 top-ranked modules are shown in Fig. 1B as a sector of a circus plot, together with information on enrichment for up- and downregulated DEGs and signature sets (MSigDB, blood cells, All RNA-seq and CHIP-seq sample and signature search (ARCHS [4]) tissues, and cell lines, SARS-CoV-2 life cycle genes, inflammasome, ISGs, transcription factors, miRNA targets). A few of these modules are enriched for MSigDB functions (Fig. 1C). As expected, we have identified a variety of cell types from the ARCHS [4] database accordant to the infection scenario, ranging from lung tissue and epithelial cells (Fig. 1D), alveolar

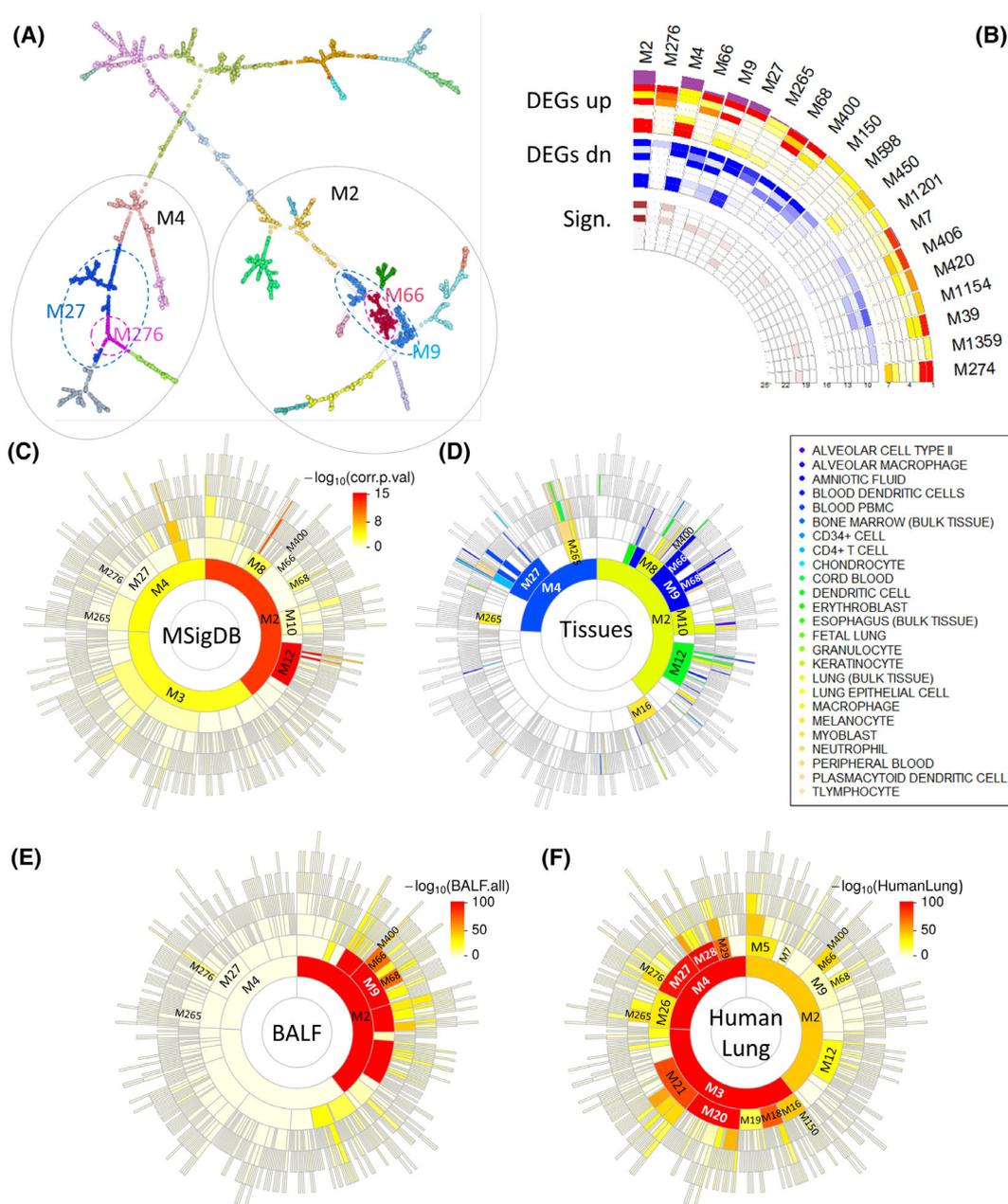


Fig. 1. Gene co-expression modules associated with SARS-CoV-2 infection. (A) A global MEGENA network. Different colors represent the modules at one particular compactness scale. (B) The top 20 MEGENA modules most enriched for the SARS-CoV-2 up- and downregulated DEG signatures are shown (outer rings: 'DEGs up' and 'DEGs dn', resp.). The center rings ('Sign.') show additional signatures, including biological processes, cells, and tissues, as well as SARS-CoV-2 host factors based on PPI. (C) A Sunburst plot of all 934 modules enriched for MSigDB canonical processes (C2.CP) is shown. (D) The module enrichment for 25 lung pathology-related tissue signatures after the 'ARCHS4' database [65] is depicted. (E, F) Sunburst plots of module enrichment for DEGs concerning (E) BALF and (F) lung biopsy tissues are displayed. The color bars in (C, E, and F) show the negative decadic logarithm of the adjusted *P*-values.

macrophages as well as lymphocytes (Fig. 1D). We have further validated the cell-type assignment with the single-cell data from BALF samples of health, moderately and severely ill COVID-19 patients [20] (see Fig. S2, Table S2 and Materials and methods).

The enrichment for the two main DEG signature sets, BALF and human lung biopsy are shown in Fig. 1E, F. Although there are differences in the DEGs between these two DEG sets, we have identified common DEG enrichment in modules M2, M9, M12,

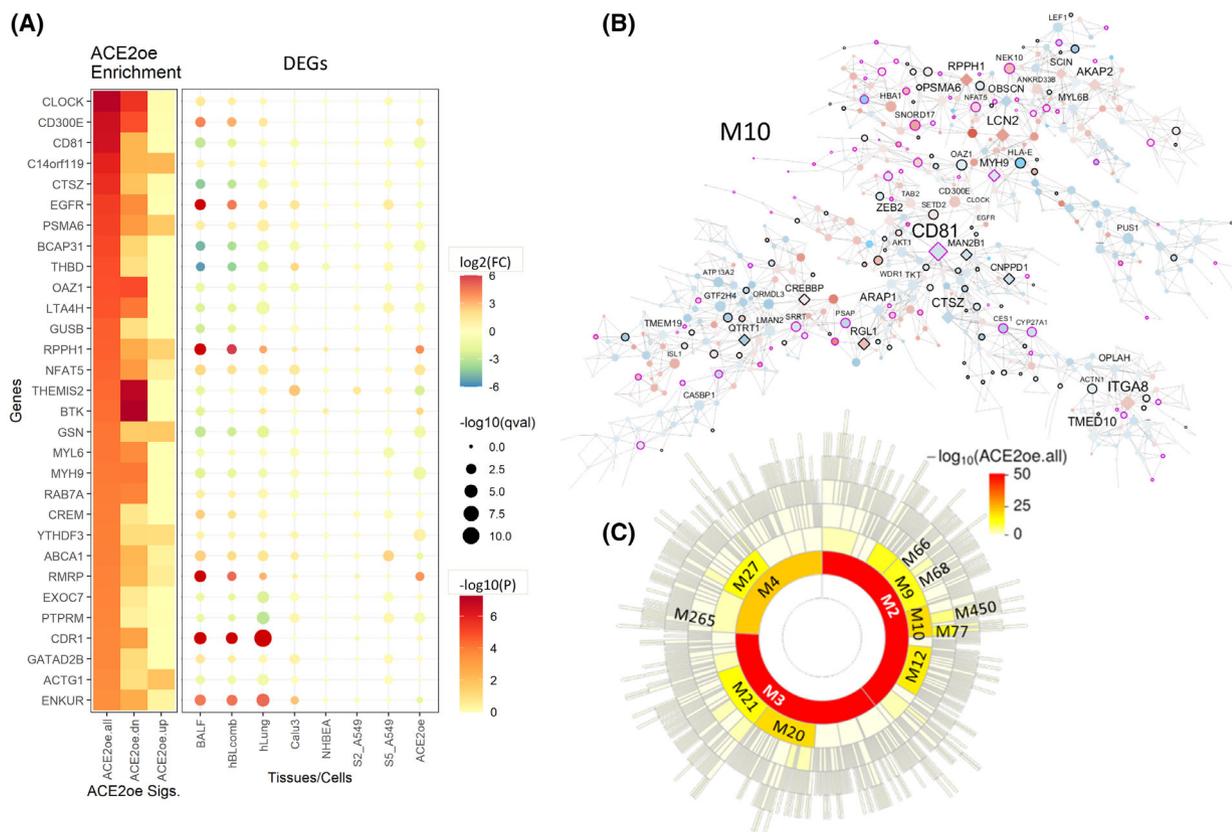


Fig. 2. Network neighborhood and network enrichment for gene signatures and key regulators. (A) Top-scored targets after network enrichment by *ACE2* overexpression signatures together with their directional response are shown. Many of these targets are members of M10. (B) The number 35 ranked module M10 is depicted, which is significantly enriched for *ACE2oe* signatures. The node color indicates a directional response. Red nodes are upregulated, and blue nodes are downregulated after infection. Diamond-shaped nodes indicate key regulators. The nodes with a black border denote genes significantly responding to *ACE2* overexpression with fold change (FC) of 1.5 or higher. Purple borders indicate *ACE2oe* responding genes with $FC \geq 2$. (C) The color tiles refer to network enrichment scores. The ‘ $-\log_{10}(P)$ ’ color scale on the right refers to the cumulative P -value used for ranking. Dark red color denotes a higher rank. The bubble plot denotes up- (red) and downregulated (blue) genes. The color of the circles refers to the fold change of expression between virus-infected and mock-infected samples. The size indicates the FDR as $-\log_{10}(qval)$. (E) A sunburst plot of the modules with *ACE2oe* enrichment is shown.

M66, M68, and M400. Most of these modules are related to translation and the ribosome.

Based on network enrichment (see [Materials and methods](#)), Fig. 2A shows a heat map of the 30 best-ranked targets, along with fold change (FC) of expression during SARS-CoV-2 infection in lung samples and cell lines. All the targets are members of the M2-M10-M77 branch, except for Bruton tyrosine kinase (*BTK*) and thymocyte selection associated family member 2 (*THEMIS2*, M2-M8-M59 branch) and exocyst complex component 7 (*EXOC7*) and protein tyrosine phosphatase receptor type M (*PTPRM*, M3-M20-M203 branch). Module M10, together with *ACE2oe* signature genes, is shown in Fig. 2B (Fig. S3 depicts parent module M2). As shown in Fig. 2C, M2, M10, and M77 are highly enriched for the *ACE2oe* signature

with FET P -value = $1.20e-95$ [1.7-fold enrichment (FE)], $1.54e-20$ (2.1FE) and $7.88e-13$ (2.7 FE). All three modules are further enriched for lung tissue signatures after ARCHS [4] tissues. According to the single-cell BALF data, the genes in these modules are abundant and highly expressed in lung epithelial cells. However, we observed the highest expression of M2 genes in alveolar macrophage cells (Fig. S2 and Table S2). Other modules such as M4, M9, M66, M69, M265, and M450 are also significantly enriched for *ACE2oe* signature (Fig. 2C). Although M3 is significantly enriched for *ACE2oe* signatures, we do not consider M3 for further analysis. M3 is ranked number 248 based on SARS-CoV-2 infection signatures. Thus, this module is not relevant to the infection process. M2 (module rank 1) and M4 (module rank 3) are the two

largest modules associated with SARS-CoV-2 infection (Fig. 1B shows the ranked modules after DEG enrichment). They are associated with different biological functions such as ribosome (M2) and transcription (M4). Table S3 shows the most significantly enriched functions, and Table S4 includes additional enrichment results. M2 and M4 are the parents of several daughter modules. For example, in addition to the modules mentioned above, M10 (rank 35, Fig. 3A) and M77 (rank 38, Fig. 3B), M2's daughter modules include highly ranked M7 (rank 14), M9 (rank 5, Fig. 3C), M66 (rank 4, Fig. 3D), M68 (rank 8), M400 (rank 9), M450 (rank 12), and M1201 (rank 13). A few of these modules are enriched for MSigDB functions (Fig. 1C). Module M7 is enriched for phenylalanine metabolism, M9 for epithelium development and IL-2 signaling, M10 for developmental biology, M68 meiotic recombination, and nucleosome assembly. Although M66, M400, M450, and M1201 are best-ranked and enriched for SARS-CoV-2 signatures, they are not significantly enriched for any known biological functions. Thus, these modules potentially indicate novel biological processes relevant to COVID-19. For example, the fourth-ranked M66 is driven by downregulated key regulators deoxyhypusine hydroxylase (*DOHH*), transmembrane protein 201 (*TMEM201* or *SAMPI*), TNF receptor superfamily member 25 (*TNFRSF25*), and *ZNF419*, as well as upregulated ectonucleoside triphosphate diphosphohydrolase (*ENTPD3*) and interferon-induced transmembrane protein 1 (*IFITM1*, Fig. 3D). *TMEM201* is required for mitotic spindle assembly and γ -tubulin localization. The depletion of *TMEM201* results in aneuploidy phenotypes, that is, the presence of an abnormal number of chromosomes in a cell, yielding bi-nucleated cells, and failed cytokinesis [32]. *TNFRSF25* is a member of the TNF receptor family. This receptor has been shown to stimulate NF- κ B activity and regulate cell apoptosis. *TNFRSF25* is further thought to be involved in controlling lymphocyte proliferation induced by T-cell activation. Thus, M66 likely plays a role in cytokinesis and cell proliferation. Concerning M4, highly ranked submodules (children) are M27 (rank 6, Fig. 3E), M265 (rank 7), and M276 (rank 2, Fig. 3F). M276, with 81 genes, includes upregulated hemoglobin subunits δ , γ 1, and μ (*HBD*, *HBG1*, and *HBM*), which form part of the hemoglobin complex (FET *P*-value = 0.05, 62.1 FE). M276 is potentially responsible for oxygen transport (FET *P*-value = 0.089, 49.7 FE). M27 and M265 are not significantly enriched for any biological function (Fig. 1A shows the M4-M27-M276 branch).

The best-ranked ACE2oe network enriched targets are clock circadian regulator (CLOCK), CD300e,

CD81, chromosome 14 open reading frame 119 (*C14orf119*), and cathepsin Z (*CTSZ*). All but *C14orf119* are in the immediate network neighborhood of *CD81* (see Fig. 3B). Clock circadian regulator (CLOCK) plays a central role in the regulation of circadian rhythms. CLOCK, a transcription factor, is upregulated in BALF and A549 samples. CD300e is a member of the CD300 glycoprotein family of transmembrane cell surface proteins expressed on myeloid cells. It is upregulated in lung samples. The protein interacts with the TYRO protein tyrosine kinase binding protein (TYROBP) and is thought to act as an activating receptor. Activation *via* CD300e provided survival signals that prevented monocyte and Myeloid dendritic cell apoptosis, triggered the production of pro-inflammatory cytokines, and upregulated the expression of cell surface co-stimulatory molecules in both cell types [33]. The expression and function of human CD300 receptors on blood circulating mononuclear cells are distinct in neonates and adults [34], potentially contributing to the difference in clinical outcome after COVID-19 infection. Zenarruzabeitia *et al.* [34] reported a stark downregulation of CD300e on monocytes in patients with severe disease. However, we cannot confirm this finding in our BALF validation data. In the NYU COVID-19 study, CD300e is upregulated 1.6-fold in patients with severe diseases compared with patients with a mild outcome. Another ACE2oe network enriched target is *CD81*, with downregulation in lung samples and cell lines. CD81 is a tetraspanin cell surface protein instrumental for B-cell activation, as a result of SARS-CoV-2 infection [35]. By regulating *CD19* expression in B-cells, CD81 enables Hepatitis C virus infection of human cells [36]. Thus, CD81 is an entry co-receptor for the Hepatitis C virus [37]. CD81 is also a host factor of the Chikungunya virus and is required for viral replication [38]. *CD81* is the only ACE2oe target which network neighborhood is significantly enriched for SARS-CoV-2 signatures, yielding a rank of 79 based on NWes. Furthermore, *CD81* is a key regulator in the M2-M10-M77 branch (Fig. S2 and Fig. 3A,B). Although speculative, CD81 may modulate SARS-CoV-2 replication. *CTSZ* is a lysosomal cysteine proteinase and a member of the peptidase C1 family. It is downregulated in lung samples and slightly upregulated in A549. Similar to *CD81*, *CTSZ* is a key regulator in M2-M10-M77. Singh *et al.* [39], hypothesized that cathepsins are among other factors facilitating SARS-CoV-2 entry into the host cell. In particular, cathepsin L (CTSL) plays a key role in SARS-CoV-2 infection by cleaving the spike protein and enhancing virus entry [40]. The epidermal growth factor receptor EGFR is a

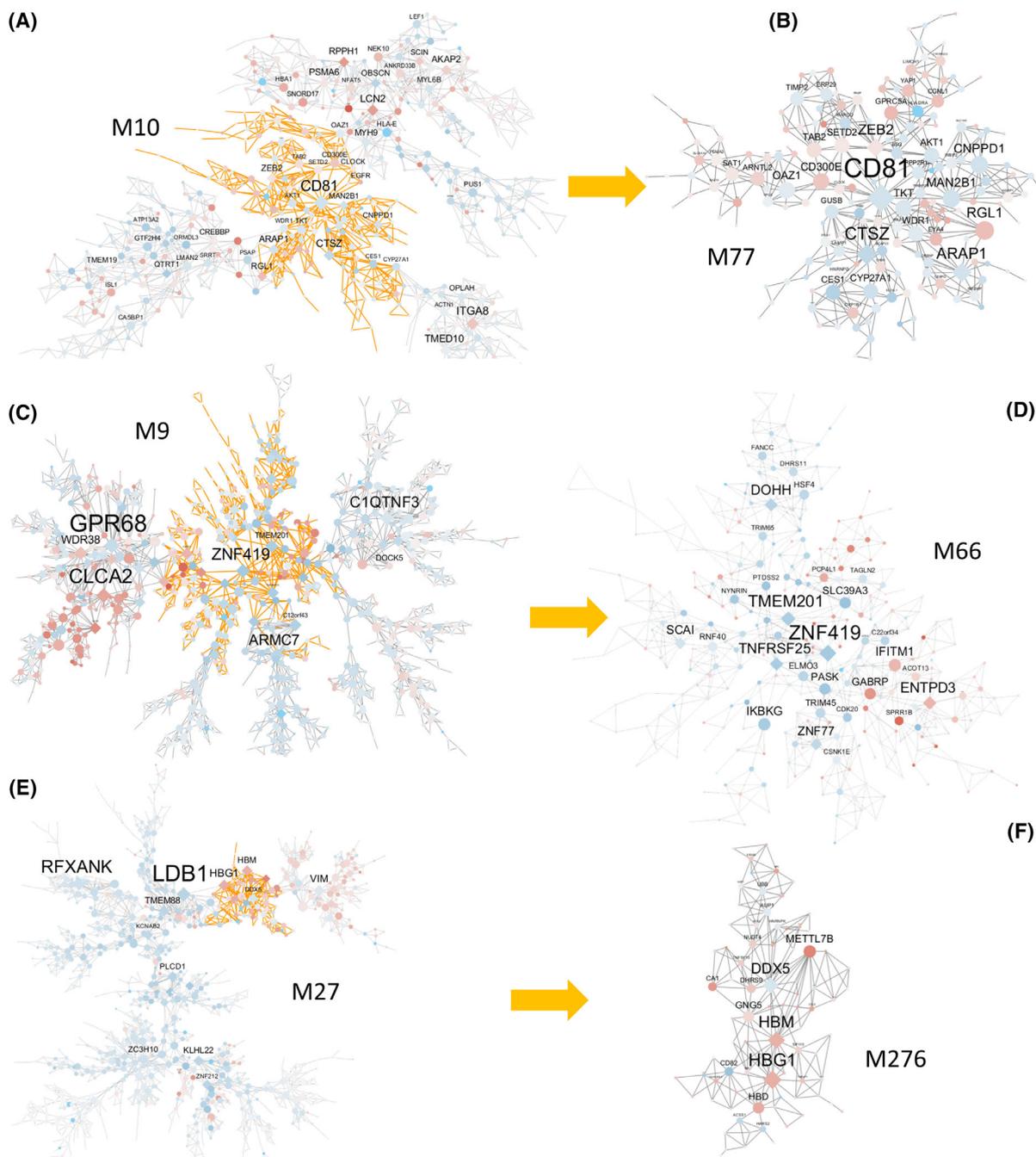


Fig. 3. Gene co-expression modules associated with SARS-CoV-2 infection. (A) With a rank of 35, M10 is not among the best 20 ranked modules. It is potentially responsible for cellular stress response/Golgi apparatus/antigen processing and presentation and is enriched for DEGs, ACE2oe, and bulk lung tissue signatures. (B) Number 38 ranked module M77 is a daughter module of M10. M77 potentially functions for the regulation of cell adhesion. Like its parent module M10, M777 is enriched for DEGs, ACE2oe, and bulk lung tissue signatures. (C) M9 is the parent of M66 and ranked number 5, and is enriched for DEGs and ACE2oe signatures. Similar to M66, it is enriched for macrophages/neutrophils tissue signature. (D) Ranked fourth and second-ranked module with less than 100 genes is M66, which is enriched for DEGs and ACE2oe signatures. M66 is enriched for macrophages/neutrophils ARCHS⁴ signature. (E) M27 is the parent of M276 and is ranked sixth. It is enriched for DEGs, ACE2oe, and blood PBMC signatures. (F) The top-ranked module with less than 1000 genes, M276, is highly enriched for upregulated DEGs. M276 is among the smallest top-ranked modules with 81 genes. – Node colors refer to the direction of regulation. Upregulated genes are red, and downregulated genes are blue. Diamond-shaped nodes denote key regulators. The size of the nodes refers to the connectivity in the network. (A, C, E) The subnetworks with orange edges refer to the corresponding daughter modules shown in (B, D, F).

transmembrane glycoprotein and is present on the cell surface of epithelial cells. It is significantly upregulated in lung samples, A549, and Calu3 cells. EGFR is a host factor for hepatitis C virus entry [41]. Respiratory viruses induce EGFR activation, suppressing IFN regulatory factor (IRF) 1-induced IFN- λ , and antiviral defense in airway epithelium [42]. *Via* EGFR signaling, EGF induces *ACE2* expression and pro-inflammatory changes in lung cells during COVID-19 [43]. Thus, EGFR may not be required for SARS-CoV-2 entry, but it may be a potential host factor for the viral life cycle.

We validated our findings with results derived from the NYU COVID-19 cohort. Fig. S4A shows a heatmap of the 20 best-ranked modules enriched for DEG signatures identified in this manuscript and deduced from the NYU COVID-19 cohort. Although the majority of modules are enriched for the combined lung and BALF dataset, we can identify significant enrichment for best-ranked modules, in particular, for NYU COVID-19 DL and HL signatures. We further evaluated the similarity in gene content between modules from this study and modules derived from the NYU COVID-19 cohort (Fig. S4B). In particular, the best-ranked modules show significant overlap, validating the findings.

We have further investigated other cell surface proteins, in particular cell surface receptors. For this purpose, we use data on experimentally verified high-confidence cell surface receptors from the cell surface protein atlas [44] and data from the *in silico* human surfaceome [45]—an extension from the protein atlas by using the measured protein data as a learning set for *in silico* prediction. From 2800 surface proteins, 1199 are classified as receptors by Surfaceome [45], capable of transducing signals triggered by binding ligands or, hypothetically, surface proteins of the SARS-CoV-2 virion. However, the prediction of such binding events or, further, the prediction that such binding events are required for viral entry is beyond our predictive methods and not intended here. But, similar to the behavior of *ACE2*, we hypothesize that the expression of genes coding for such surface proteins can be triggered by the infection. We further hypothesize that such surface proteins mediate the transcriptomic response of downstream genes. Thus, we expect upregulation of the surface protein-coding genes and enrichment of DEGs in such receptors' network neighborhood. Out of the 1199 receptors from the Surfaceome, 413 are in the MEGENA network. We identified further candidates in addition to the above-discussed surface receptors and key regulators *CD81*, *CD300E*, and *EGFR*. We expanded our criteria

and included surface proteins that are significantly expressed across all datasets (employing ACAT, an aggregated Cauchy association test [46]). Surface proteins with the lowest aggregated *P*-value that are upregulated in most datasets were chosen. The highest-ranked candidate is lysosome-associated membrane glycoprotein 3 (*LAMP3*), followed by *EGFR*, as discussed above. *LAMP3* family plays a critical role in the autolysosome fusion process. *LAMP3* is expressed explicitly in lung tissues and is involved in influenza A virus replication in A549 cells [47]. It activates the phosphatidylinositol 3-kinase (PI3K)/AKT serine/threonine kinase pathway required for the influenza life cycle and is necessary for SARS-CoV to establish infection, as demonstrated in African green monkey kidney (Vero E6) cells [48]. Third-best-ranked surface protein is CEA cell adhesion molecule 1 (*CEACAM1*). Multiple cellular activities have been attributed to the encoded protein, including roles in the differentiation and arrangement of three-dimensional tissue structure, angiogenesis, apoptosis, tumor suppression, metastasis, and the modulation of innate and adaptive immune responses. Both *CEACAM1* and *LAMP3* are members of the M4-M27 branch.

We were further interested whether and how cell-surface proteins show co-expression patterns, indicating a possible (indirect) interactive relationship between members. For this purpose, we constructed a co-expression network of the 411 cell surface receptors yielding 33 modules. In particular, we have identified a receptor module (RM6, Fig. 4A) that includes both known (*NRP1* and *TMPRSS2*) receptors and receptors predicted in this paper (*CD300e* and *CD81*). Key regulator of RM6 is downregulated integrin subunit beta 2 (*ITGB2* or *CD18*, *LFA-1*), which plays an important role in immune response, for example, T-cell migration [49] and tuning of the T-cell program [50]. *ACE2* is a member of a second module, RM13 (Fig. 4B). The co-expression of these receptors provides further evidence of their potential role during COVID-19.

We also explored complementary approaches using knowledge-based interaction data of protein-protein interactions (PPI), as co-expression networks not necessarily reflect physical interactions between molecules. For this purpose, we used information from the STRING database [51]. Figure S5 shows a protein-protein interaction network, including the targets *ACE2*, *CD300e*, *CD81*, *CTSZ*, and *EGFR*, based on experimentally determined and database-annotated interactions. The hubs of this PPI network are growth factor receptor-bound protein 2 (*GRB2*) and phosphoinositide-3-kinase regulatory subunit 1 (*PIK3R1*).

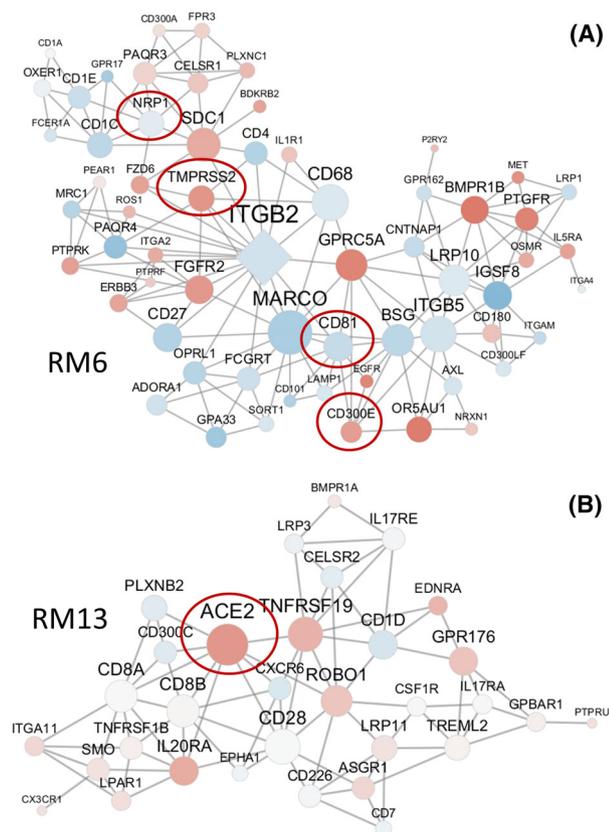


Fig. 4. Receptor-only MEGENA modules. Two MEGENA modules inferred from genes coding for cell receptors are shown. (A) Receptor module RM6 includes known (*NRP1*, *TMPRSS1*) receptors as well as novel, predicted targets (*CD81*, *CD300e*). (B) RM6 is a separate module that includes *ACE2*. These five genes are highlighted by red circles. The node color indicates a directional response. Red nodes are upregulated hub genes.

SARS-CoV-2-triggered surface protein receptor expression shows clear tissue-specific age dependency

We were also interested in the age dependency of the molecular processes and receptors involved in SARS-CoV-2 infection (see also the Appendix S1). A significant age disparity for severe cases, often causing death, has been widely reported for COVID-19. Being highly disproportional, more elderly patients experience severe symptoms and die due to this particular disease. A recent study by Muus *et al.* [52] discovered that the expression of the entry factors *ACE2*, *TMPRSS2*, and *CTSL* increased with age and in males, including in airway secretory cells and alveolar type 2 cells. We hypothesize that not only these three targets but also many host factors required for the virus life cycle have an age-dependent expression. By filtering in the genes upregulated in at least two of the SARS-CoV-2

studies, we obtained 216 genes encoding cell surface proteins. These surface proteins are involved in transmembrane transport of small molecules (MSigDB c2.cp enrichment: $P = 3.38e-08$, 4.3-fold), erb-b2 receptor tyrosine kinase 4 (*ERBB4*) pathway ($P = 7.46e-08$, 5.5-fold), neuroactive ligand–receptor interaction ($P = 8.51e-05$, 4.7-fold) or cytokine–cytokine receptor interaction ($P = 1.22e-04$, 4.2-fold). The tissue-specific age dependency of these genes' baseline expression was calculated by a linear model using data from GTEx (see [Materials and methods](#)). We examined correlations between the expression of these SARS-CoV-2 triggered surface protein receptors (STSPRs) with chronological age using GTEx v8 data covering 46 tissues (Table S5). A large number of these surface protein receptors have their gene expression levels associated with age in many tissues, especially in the tibial artery, tibial nerve, and visceral fat. More than 70 receptors were significantly correlated with age. In contrast, very few receptors were associated with age in the liver, coronary artery, and brain substantia nigra (< 5 receptors). Moreover, in most cases, the gene expression levels of these receptors increased with age (Table S6).

We further examined the overall correlation between STSPRs expression and age in a tissue-specific manner. Specifically, we first computed a composite receptor score (CRS) for each tissue of each sample in GTEx by summarizing the normalized expression values of the STSPRs and then assessed the correlation between CRS and age (see [Materials and methods](#) for details; Fig. 5). The CRS is a discretized, aggregated expression value across all samples for each receptor and tissue. Three tissues, including the tibial artery, skeletal muscle, and subcutaneous fat, show the strongest positive correlations between their respective CRS and age. On the contrary, the whole blood, the frontal cortex (BA9), the ovary, and the cerebellum have the strongest negative correlations. Interestingly, the lung is ranked 31 out of 46 tissues, indicating that COVID-19 may impact far more tissues in different age populations than what we observed. As expected, the top-ranked tissues have a large number of significantly age-correlated receptors, consistent with the direction of the overall correlation. For example, in the tibial artery, which has a significant positive CRS-age correlation, 94 STSPRs are significantly positive, and nine STSPRs are significantly negatively correlated with age. Whereas in the frontal cortex, 56 STSPRs are significantly negative, and two STSPRs are significantly positively correlated with age, respectively (Fig. 5A). The age effect on various disease pathologies is known for some of these tissues, with significant correlations

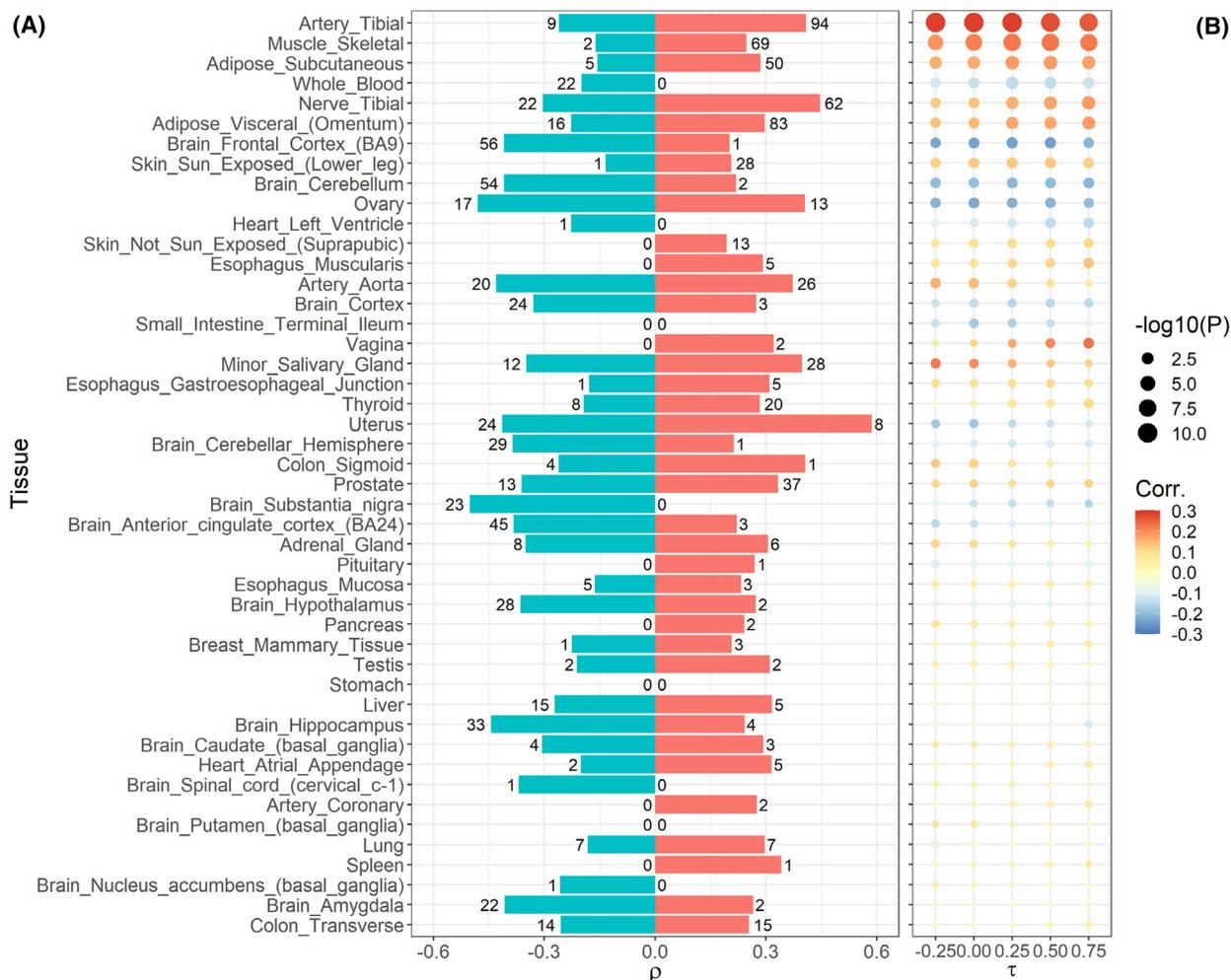


Fig. 5. Number of receptors significantly correlated with age in the GTEx data. (A) The range of significant individual receptor/age correlation ρ is shown for each tissue. Numbers next to the bars denote the number of receptors that are significantly positively (red bars) or negatively (green bars) associated with age, respectively. Missing bars indicate the absence of a significant correlation. (B) The age dependency of gene expression between tissues and composite receptor score (CRS) based on the genes coding for cell surface proteins (rows) are shown. Tissues are ranked based on correlation significance with parameter $\tau = 0.25$. Colors refer to the positive (red) and the negative (blue) correlation between age and CRS. The size denotes the FDR in $-\log_{10}(\text{adj. } P\text{-value})$.

between CRS and age. For example, age is a known risk factor for adverse outcomes in peripheral artery disease. The risk of severe limb ischemia, the sudden loss of blood flow to a limb caused by embolism or thrombosis, significantly increases with age [53]. Thrombosis and microvascular injury have been identified as an implication of severe COVID-19 infection [54]. Another example is skeletal muscles with well-studied age-related wasting and weakness. Cellular and molecular mechanisms contributing to a decline in muscular function involve neuromuscular factors, hormones, testosterone or growth hormone, insulin, myogenic regulatory factors (MRFs), the Notch signaling pathway, as well as cytokines and inflammatory

pathways [55]. A cytokine storm and robust production of cytokines [6] are known to contribute to the severity of COVID-19 infections [56], potentially inducing systemic effects across many tissues and organs.

Age dependency of a systemic SARS-CoV-2 response

Network neighborhoods of several STSPRs such as *ENTPD3*, γ -aminobutyric acid type A receptor subunit π (*GABRP*), and EPH receptor A6 (*EPHA6*) are enriched for the SARS-CoV-2 induced DEG signatures from human patient lung samples. The *GABRP*

mRNA level is positively correlated with age in three tissues (subcutaneous fat, lung, and minor salivary gland) and negatively correlated with age in three other tissues (tibial nerve, not sun-exposed skin, and small intestine terminal ileum). *EPHA6*, a member of the M2-M9 branch (Figs 1A and 3C), promotes angiogenesis [57] and regulates neuronal and spine morphology [58]. The network neighborhood of *EPHA6* is enriched for pentose and glucuronate interconversion, glucuronidation, and systemic lupus erythematosus (FET P -values $< 7.5e-03$). *EPHA6* mRNA level increases with age in six tissues (artery aorta, cerebellar brain hemisphere, brain cerebellum, esophagus gastroesophageal junction, esophagus mucosa, and ovary). It decreases in four tissues (brain amygdala, brain cortex, brain hippocampus, and brain hypothalamus). Interestingly, *ACE2* mRNA level increases with age in five tissues (adrenal gland, lung, ovary, stomach, and uterus tissue) and decreases in three tissues (aorta artery, minor salivary gland, and tibial nerve; Table S5).

We further investigated the potential age dependencies of STSPRs in biological processes realized by MEGENA co-expression modules. For this purpose, we have identified network modules enriched for tissue-specific age-correlated STSPR. The 3227-strong generic transcription module M4 is enriched for both positive and negative correlated STSPRs. M4 is enriched for positive age-correlated STSPR in the prostate (FET P -value = 0.015, 1.85 FE) and for negative age-correlated STSPR in the liver (FET P -value = 0.0015, 2.77 FE). We have identified the M4-M27 branch with signaling functions underlying COVID-19 (Fig. 1A shows the M4-M27-M276 branch). Using blood cell-type signatures, we found that M4 is enriched for neutrophils (FET P -value = 0.037, 3.0 FE; average module expression (avMx) = 1.37 and expressed in 96.4% neutrophils according to the Liao *et al.* [20] BALF single-cell data, see Fig. S2 and Table S2). Neutrophil-mediated innate immune responses against pathogens in the lungs determine the outcome of infection; insufficient neutrophil recruitment can lead to life-threatening infection, although an extreme accumulation of neutrophils can result in excessive lung injury associated with inflammation [59]. Such a massive intra-alveolar neutrophilic infiltration has been observed in COVID-19 patients with a longer clinical course, likely due to superimposed bacterial pneumonia [60]. M4 genes are also expressed in M2 macrophages (avMx = 1.76, 97.2% cells) and NK cells (avMx = 0.9, 91.8% cells).

Other enriched modules involve number 66 ranked M26 (positive age-correlated STSPRs in adrenal gland:

FET P -value = $1.32e-04$, 6.88 FE), and number 35 ranked M10 (negative age-correlated STSPRs in mammary breast tissue: FET P -value = 0.069, 6.10 FE). M26 is another child of M4 with cell cycle (M/G1 transition) function.

We also analyzed the dependence of the STSPRs on age in each tissue in the GTEx by computing correlations between differential expression of the STSPRs in COVID-19 and correlations between the STSPRs and age in each tissue in the GTEx (termed STSPR differential expression and age dependence (STSPR-DEAD) score; see details in Materials and methods and Table S7). The subcutaneous fat, tibial artery, substantia nigra, esophagus gastroesophageal junction, and liver show the strongest STSPR-DEAD score. A heatmap of STSPR-DEAD scores between 46 tissues and seven sample types is shown in Fig. 6A. Many tissues have negative STSPR-DEAD scores. Examples are tibial artery ($\rho = 0.32$, $P = 0.029$; Fig. 6B), liver ($\rho = 0.38$, $P = 4.4e-05$; Fig. 6C) and esophagus gastroesophageal junction ($\rho = -0.39$, $P = 1.4e-03$; Fig. 6D). The substantia nigra has the strongest positive STSPR-DEAD score and possesses the highest correlation coefficient in absolute terms with DEGs (DEGs from combined BALF and lung biopsies, $\rho = -0.32$, not shown). Our STSPR-AD signature genes (Table S5) are enriched for the signatures of 19 different cell types identified from the single-cell lung tissue data [61] (B-cells, monocytes, CD¹⁶⁺ monocytes, macrophages, NK cells, plasma cells, CD⁴⁺ and CD⁸⁺ T-cells, T_{reg} cells, epithelial AT1 and AT2 cells, ciliated cells, club cells, goblet cells, mast cells, pericytes, smooth muscle cells, lymphatic and vascular endothelial cells; Table S8). Noteworthy is the significant enrichment of STSPR-AD signatures in the monocyte and macrophage signatures, indicating monocyte mediated inflammatory processes in the age-related pathology of COVID-19. For example, Pence [62] concluded that pathological monocyte response in COVID-19 showed a similar pattern to those in aging, suggesting that monocytes may contribute to the disproportionate severity of COVID-19 in older adults. Tizazu *et al.* [63] seconded this finding of pro-inflammatory monocytes both in aging and in severe COVID-19 cases. The lack of enrichment of STSPR-AD tissues for the heart and kidney signatures in the study by Delorey *et al.* [61] emphasizes the role of monocytes in the age-related pathology of COVID-19. With the COVID-19 pathology data from the tissues studied by Delorey *et al.* [61], we can further identify the COVID-19 vulnerable tissues in old patients. Comparing the results of the STSPR-DEAD scores (Fig. 6 and Table S7) with the STSPR-AD tissues with

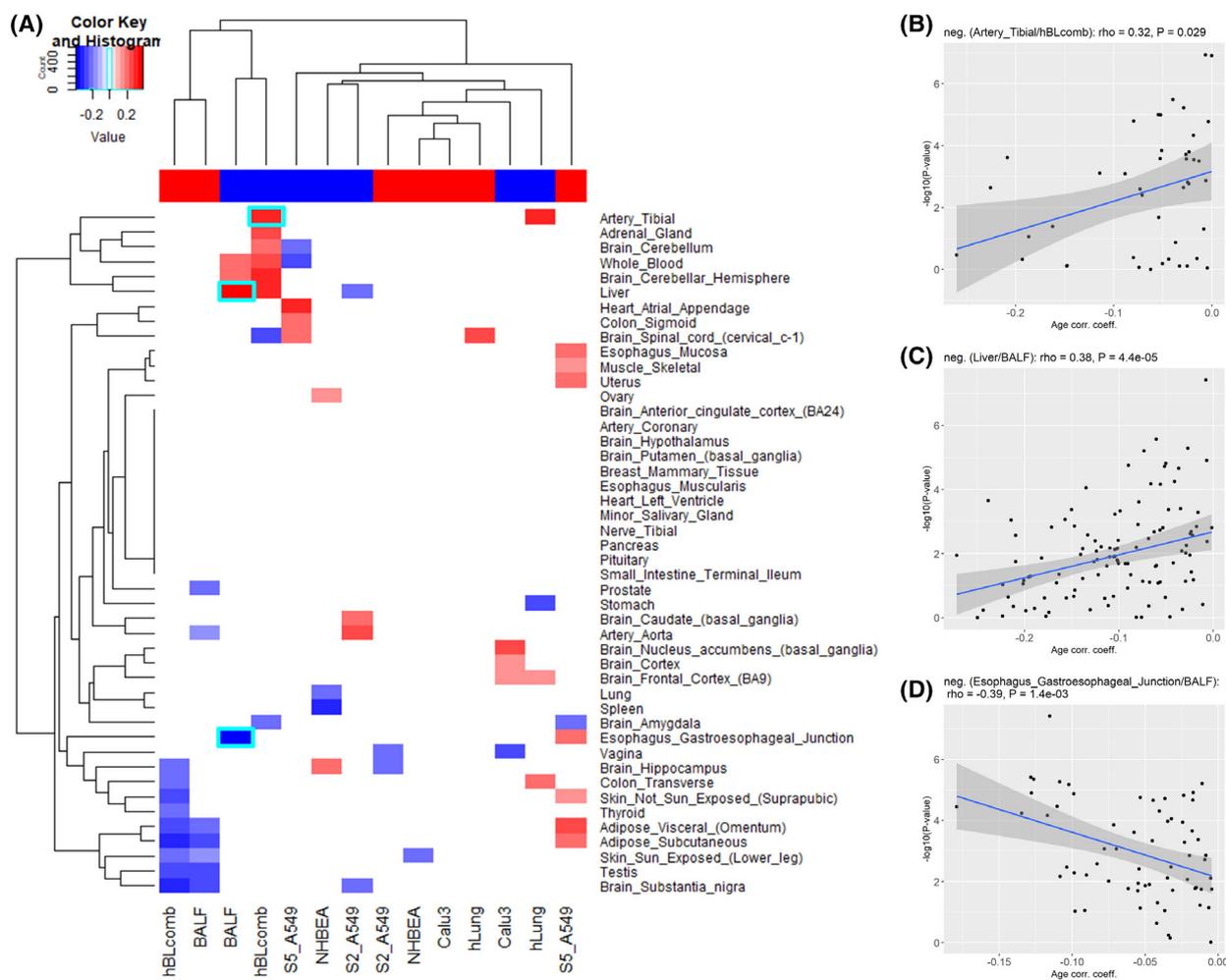


Fig. 6. Correlation between the surface receptors' differential expression in SARS-CoV-2 infection and their tissue-specific age dependence. (A) A heatmap of correlation coefficients after tissue age effect (STSPR-DEAD, see text) and DEGs correlation is shown. Only the correlation coefficients with nominal $P \leq 0.05$ are shown. The top color bar indicates the direction of the STSPR-DEADs, with red denoting positive STSPR-DEADs and blue referring to negative STSPR-DEADs. Tiles with cyan boundaries indicate select tissue/DEG pairs. (B–D) Dot plots between STSPR-DEAD and DEGs of select tissues with the best correlation coefficients are shown: (B) artery tibia against combined BALF/lung biopsy DEGs, (C) liver against BALF, and (D) esophagus gastroesophageal junction against BALF.

COVID-19 pathology identified by Delorey *et al.* [61] (Table S8) highlights tibial artery and subcutaneous fat as tissues with the most severe effect of age-related COVID-19 pathology. However, investigation of the results shown in Table S8 shows that some brain tissues, such as the amygdala, hippocampus, anterior cingulate cortex, frontal cortex, and cortex, are also significantly enriched for monocyte signatures. Pro-inflammatory monocytes and their inflammatory signatures were found to be correlated with the development of postacute sequelae of COVID-19 (PASC) and the neurological manifestation [64].

We have further validated the dependence of the STSPRs on age in GTEx tissues with data from the

NYU COVID-19 cohort. Figure S6A shows the heatmap between 46 tissues, three sample types, and one combined dataset (Xsq), corresponding to Fig. 6A. Figure S6B depicts a plot between STSPR-DEAD and DEGs of esophagus gastroesophageal junction against HL ($\rho = -0.49, P = 3.2e-05$) corresponding to Fig. 6D. The findings from the NYU COVID-19 cohort are in accordance with the corresponding STSPRs results reported above.

To explore the gene expression changes of STSPRs with age, we have separated GTEx donors into two cohorts: a young (≤ 45 years) cohort and an old cohort (≥ 60 years). Gene expression was then adjusted to compare the difference between these two

cohorts (see [Materials and methods](#)). In subcutaneous fat and tibial artery, the young cohort showed a lower gene expression level, while a higher level of gene expression was observed in the elder cohort. This pattern can also be seen in the esophagus gastroesophageal junction and skeletal muscle (Fig. S7).

Overall, we found a clear age effect of genes coding for cell surface proteins and receptors that are potentially utilized by SARS-CoV-2. In particular, we have identified that STSPRs showed stronger age dependency in the tibial artery, skeletal muscle, adipose, and brain tissues. Such an age-dependent effect could potentially contribute to the elevated severity of COVID-19 in the elderly.

Discussion

In the present study, we focus on the biological processes and key regulators modulating the host response to SARS-CoV-2 infections. Our multiscale network analysis of the gene expression data from both patient samples and cell lines has revealed network structures and key regulators underlying the host response to SARS-CoV-2 infection.

Essential aspects of COVID-19 pathology are the biological processes hijacked by the virus for its advantage. Expression of the ACE2 receptor on the host cell and binding of the viral Spike protein for cell entry are among the first steps. Other processes beneficial for the virus may be staged by ACE2 expression and triggered by the binding process. CD300e and its interacting partner TYROBP trigger pro-inflammatory cytokines and prevent apoptosis, an essential process controlled by many viruses. On the contrary, severe inflammation significantly contributes to the pathology of COVID-19 disease. Other potential surface protein host factors are CD81 and EGFR. Additional surface proteins are CEACAM1 and LAMP3. Multiple cellular activities have been attributed to CEACAM1, including differentiation and arrangement of three-dimensional tissue structure, angiogenesis, apoptosis, tumor suppression, metastasis, and the modulation of innate and adaptive immune responses. LAMP3, however, plays a critical role in the autolysosome fusion process. It activates the PI3K/AKT pathway, which is necessary for SARS-CoV to establish infection.

We have further investigated the age dependence of receptors' expression as clinicians have observed a severe disparity in survival between old and young COVID-19 patients. We have identified a strong correlation between tissue age dependency and SARS-CoV-2 infection-induced receptor expression in subcutaneous fat, tibial artery, brain substantia nigra, esophagus

gastroesophageal junction, and liver. In particular, we found significant enrichment for macrophages and monocytes using COVID-19 pathology data from the tissues studied by Delorey *et al.* This finding concurs with the conclusion of Pence and Tizazu *et al.* that monocytes may contribute to the disproportionate severity of COVID-19 in older adults. Comparing our results with the tissue-specific COVID-19 pathology data by Delorey *et al.* highlights the tibial artery and subcutaneous fat as tissues with the most severe effect of age-related COVID-19 pathology. We may further hypothesize that the significant enrichment of some brain tissues for monocyte signatures may contribute to the development and the neurological manifestation of postacute sequelae of COVID-19 (PASC). However, the exact contribution of specific receptors' age dependency on the disease's pathology requires additional investigation.

In conclusion, our analyses presented here suggest that SARS-CoV-2 utilizes multiple novel receptors to facilitate its life cycle and spawns a unique response in the host system. Novel hypotheses involving the utilization of cell surface receptors and their age-dependent expression offer new insights into the molecular mechanisms of SARS-CoV-2 infection and pave the way for developing new therapeutic interventions against COVID-19.

Acknowledgements

This work was supported in part through the computational resources provided by Scientific Computing at the Cahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant U1TR004419 from the National Center for Advancing Translational Sciences. The Genotype-Tissue Expression (GTEx) Project was supported by the [Common Fund](#) of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. This work was funded by grants to CVF from the National Institutes of Health (R21 AI149013).

Author contributions

CVF and BZ conceived and designed the study. CVF, LZ, QW, XZ, SV, PX, and W-MS analyzed the data. ZT provided insights into age dependency. CVF and BZ wrote the paper.

Data accessibility

Publicly available datasets were used for the current study. These datasets are available at NCBI/GEO with

accession number [GSE147507](#), and [GSE145926](#) at NCBI/SRA with accession numbers PRJNA688510, SRR10571724, SRR10571730, and SRR10571732 and at Beijing Institute of Genomics (BIG) Data Center (<https://bigd.big.ac.cn/>) under the accession number CRA002390. The derived datasets generated after analysis during the current study are available at Synapse, [10.7303/syn26485918](https://doi.org/10.7303/syn26485918).

References

- Cui J, Li F and Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181–192.
- Daamen AR, Bachali P, Owen KA, Kingsmore KM, Hubbard EL, Labonte AC, Robl R, Shrotri S, Grammer AC and Lipsky PE (2020) Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. *bioRxiv* doi: [10.1101/2020.05.28.121889](https://doi.org/10.1101/2020.05.28.121889)
- Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y *et al.* (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507–513.
- tenOever BR (2016) The evolution of antiviral defense systems. *Cell Host Microbe* **19**, 142–149.
- Janeway CA Jr and Medzhitov R (2002) Innate immune recognition. *Annu Rev Immunol* **20**, 197–216.
- Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D *et al.* (2020) Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9.
- Wang C, Forst CV, Chou TW, Geber A, Wang M, Hamou W, Smith M, Sebra R, Zhang B, Zhou B *et al.* (2020) Cell-to-cell variation in defective virus expression and effects on host responses during influenza virus infection. *MBio* **11**, e02880-19.
- Hur S (2019) Double-stranded RNA sensors and modulators in innate immunity. *Annu Rev Immunol* **37**, 349–375.
- Lazear HM, Schoggins JW and Diamond MS (2019) Shared and distinct functions of type I and type III interferons. *Immunity* **50**, 907–923.
- Proudfoot AE (2002) Chemokine receptors: multifaceted therapeutic targets. *Nat Rev Immunol* **2**, 106–115.
- Sokol CL and Luster AD (2015) The chemokine system in innate immunity. *Cold Spring Harb Perspect Biol* **7**, a016303.
- Smith T, Rohaim MA and Munir M (2022) Mapping molecular gene signatures mediated by SARS-COV-2 and large-scale and genome-wide transcriptomics comparative analysis among respiratory viruses of medical importance. *Mol Cell Probes* **64**, 101820.
- Ten Garcia-Sastre A (2017) Strategies of interferon evasion by viruses. *Cell Host Microbe* **22**, 176–184.
- Thair SA, He YD, Hasin-Brumshtein Y, Sakaram S, Pandya R, Toh J, Rawling D, Rimmel M, Coyle S, Dalekos GN *et al.* (2021) Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections. *iScience* **24**, 101947.
- Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, Meng Q, Peters M, He Y, Ni C *et al.* (2012) Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol Syst Biol* **8**, 594.
- Zhao Y, Forst CV, Sayegh CE, Wang IM, Yang X and Zhang B (2016) Molecular and genetic inflammation networks in major human diseases. *Mol Biosyst* **12**, 2318–2341.
- Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, Cowling BJ, Lipsitch M and Leung GM (2020) Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* **26**, 506–510.
- CDC COVID-19 Response Team (2020) Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* **69**, 343–346.
- Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, Guo D, Hu W, Yang J, Tang Z *et al.* (2020) Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect* **9**, 761–770.
- Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F *et al.* (2020) Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* **26**, 842–844.
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3.
- Song WM and Zhang B (2015) Multiscale embedded gene Co-expression network analysis. *PLoS Comput Biol* **11**, e1004574.
- GTE Consortium (2017) Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213.
- Zeng L, Yang J, Peng S, Zhu J, Zhang B, Suh Y and Tu Z (2020) Transcriptome analysis reveals the difference between “healthy” and “common” aging and their connection with age-related diseases. *Aging Cell* **19**, e13121.
- Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol* **57**, 289–300.
- Sulaiman I, Chung M, Angel L, Tsay JJ, Wu BG, Yeung ST, Krolikowski K, Li Y, Duerr R, Schluger R *et al.* (2021) Microbial signatures in the lower airways

- of mechanically ventilated COVID19 patients associated with poor clinical outcome. *medRxiv* doi: [10.1101/2021.02.23.21252221](https://doi.org/10.1101/2021.02.23.21252221)
- 27 Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Winter H, Meister M, Veith C, Boots AW *et al.* (2020) SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* **39**, e105114.
- 28 Daly JL, Simonetti B, Klein K, Chen KE, Williamson MK, Antón-Plágaro C, Shoemark DK, Simón-Gracia L, Bauer M, Hollandi R *et al.* (2020) Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861–865.
- 29 Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, van der Meer F, Kallio K, Kaya T, Anastasina M *et al.* (2020) Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856–860.
- 30 Wang M, Zhao Y and Zhang B (2015) Efficient test and visualization of multi-set intersections. *Sci Rep* **5**, 16923.
- 31 Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A *et al.* (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8.
- 32 Larsson VJ, Jafferali MH, Vijayaraghavan B, Figueroa RA and Hallberg E (2018) Mitotic spindle assembly and gamma-tubulin localisation depend on the integral nuclear membrane protein Samp1. *J Cell Sci* **131**, jcs211664.
- 33 Brekalo T, Calzetti F, Pérez-Cabezas B, Borràs FE, Cassatella MA and López-Botet M (2010) Functional analysis of the CD300e receptor in human monocytes and myeloid dendritic cells. *Eur J Immunol* **40**, 722–732.
- 34 Zenarruzabeitia O, Vitallé J, García-Obregón S, Astigarraga I, Eguizabal C, Santos S, Simhadri VR and Borrego F (2016) The expression and function of human CD300 receptors on blood circulating mononuclear cells are distinct in neonates and adults. *Sci Rep* **6**, 32693.
- 35 Zhu R, Yan T, Feng Y, Liu Y, Cao H, Peng G, Yang Y, Xu Z, Liu J, Hou W *et al.* (2021) Mesenchymal stem cell treatment improves outcome of COVID-19 patients via multiple immunomodulatory mechanisms. *Cell Res* **31**, 1244–1262.
- 36 Luo RF, Zhao S, Tibshirani R, Myklebust JH, Sanyal M, Fernandez R, Gratzinger D, Marinelli RJ, Lu ZS, Wong A *et al.* (2010) CD81 protein is expressed at high levels in normal germinal center B cells and in subtypes of human lymphomas. *Hum Pathol* **41**, 271–280.
- 37 Cormier EG, Tsamis F, Kajumo F, Durso RJ, Gardner JP and Dragic T (2004) CD81 is an entry coreceptor for hepatitis C virus. *Proc Natl Acad Sci USA* **101**, 7270–7274.
- 38 Lasswitz L, Zapatero-Belinchón FJ, Moeller R, Hülskötter K, Laurent T, Carlson LA, Goffinet C, Simmons G, Baumgärtner W and Gerold G (2022) The Tetraspanin CD81 is a host factor for chikungunya virus replication. *MBio* **13**, e0073122.
- 39 Singh M, Bansal V and Feschotte C (2020) A single-cell expression map of human coronavirus entry factors. *bioRxiv* doi: [10.1101/2020.05.08.084806](https://doi.org/10.1101/2020.05.08.084806)
- 40 Zhao MM, Yang WL, Yang FY, Zhang L, Huang WJ, Hou W, Fan CF, Jin RH, Feng YM, Wang YC *et al.* (2021) Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduct Target Ther* **6**, 134.
- 41 Lupberger J, Zeisel MB, Xiao F, Thumann C, Fofana I, Zona L, Davis C, Mee CJ, Turek M, Gorke S *et al.* (2011) EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat Med* **17**, 589–595.
- 42 Ueki IF, Min-Oo G, Kalinowski A, Ballon-Landa E, Lanier LL, Nadel JA and Koff JL (2013) Respiratory virus-induced EGFR activation suppresses IRF1-dependent interferon λ and antiviral defense in airway epithelium. *J Exp Med* **210**, 1929–1936.
- 43 Vagapova ER, Lebedev TD and Prassolov VS (2021) Viral fibrotic scoring and drug screen based on MAPK activity uncovers EGFR as a key regulator of COVID-19 fibrosis. *Sci Rep* **11**, 11234.
- 44 Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, Moest H, Omasits U, Gundry RL, Yoon C *et al.* (2015) A mass spectrometric-derived cell surface protein atlas. *PLoS ONE* **10**, e0121314.
- 45 Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT and Wollscheid B (2018) The in silico human surfaceome. *Proc Natl Acad Sci USA* **115**, E10988–E10997.
- 46 Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E and Lin X (2019) ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am J Hum Genet* **104**, 410–421.
- 47 Zhou Z, Xue Q, Wan Y, Yang Y, Wang J and Hung T (2011) Lysosome-associated membrane glycoprotein 3 is involved in influenza A virus replication in human lung epithelial (A549) cells. *Virol J* **8**, 384.
- 48 Mizutani T, Fukushi S, Saijo M, Kurane I and Morikawa S (2005) JNK and PI3k/Akt signaling pathways are required for establishing persistent SARS-CoV infection in Vero E6 cells. *Biochim Biophys Acta* **1741**, 4–10.
- 49 Walling BL and Kim M (2018) LFA-1 in T cell migration and differentiation. *Front Immunol* **9**, 952.
- 50 Verma NK and Kelleher D (2017) Not just an adhesion molecule: LFA-1 contact tunes the T lymphocyte program. *J Immunol* **199**, 1213–1221.

- 51 Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al.* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362–D368.
- 52 Muus C, Luecken MD, Eraslan G, Sikkema L, Waghray A, Heimberg G, Kobayashi Y, Vaishnav ED, Subramanian A, Smillie C *et al.* (2021) Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat Med* **27**, 546–559.
- 53 Domenick N, Saqib NU, Marone LK, Rhee RY, Makaroun MS and Chaer RA (2012) Impact of gender and age on outcomes of tibial artery endovascular interventions in critical limb ischemia. *Ann Vasc Surg* **26**, 937–945.
- 54 Magro C, Mulvey JJ, Berlin D, Nuovo G, Salvatore S, Harp J, Baxter-Stoltzfus A and Laurence J (2020) Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases. *Transl Res* **220**, 1–13.
- 55 Ryall JG, Schertzer JD and Lynch GS (2008) Cellular and molecular mechanisms underlying age-related skeletal muscle wasting and weakness. *Biogerontology* **9**, 213–228.
- 56 Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G and Melino G (2020) COVID-19 infection: the perspectives on immune responses. *Cell Death Differ* **27**, 1451–1454.
- 57 Li S, Ma Y, Xie C, Wu Z, Kang Z, Fang Z, Su B and Guan M (2015) EphA6 promotes angiogenesis and prostate cancer metastasis and is associated with human prostate cancer progression. *Oncotarget* **6**, 22587–22597.
- 58 Das G, Yu Q, Hui R, Reuhl K, Gale NW and Zhou R (2016) EphA5 and EphA6: regulation of neuronal and spine morphology. *Cell Biosci* **6**, 48.
- 59 Craig A, Mai J, Cai S and Jeyaseelan S (2009) Neutrophil recruitment to the lungs during bacterial pneumonia. *Infect Immun* **77**, 568–575.
- 60 Tian S, Xiong Y, Liu H, Niu L, Guo J, Liao M and Xiao S-Y (2020) Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. *Mod Pathol* **33**, 1007–1014.
- 61 Delorey TM, Ziegler CGK, Heimberg G, Normand R, Yang Y, Segerstolpe Å, Abbondanza D, Fleming SJ, Subramanian A, Montoro DT *et al.* (2021) COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113.
- 62 Pence BD (2020) Severe COVID-19 and aging: are monocytes the key? *Geroscience* **42**, 1051–1061.
- 63 Tizazu AM, Mengist HM and Demeke G (2022) Aging, inflammaging and immunosenescence as risk factors of severe COVID-19. *Immun Ageing* **19**, 1–18.
- 64 Vanderheiden A and Klein RS (2022) Neuroinflammation and COVID-19. *Curr Opin Neurobiol* **76**, 102608.
- 65 Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC and Ma'ayan A (2018) Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun* **9**, 1–10.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Network enrichment.

Fig. S1. Super Exact Test between different patient-derived DEGs.

Fig. S2. Single-cell data derived from BALF samples.

Fig. S3. Gene co-expression modules associated with SARS-CoV-2 infection.

Fig. S4. Module-module overlap and enrichment.

Fig. S5. Protein-protein interaction network.

Fig. S6. Validation of the correlation between the surface receptors' differential expression in SARS-Cov-2 infection and their tissue-specific age dependence using data from the NYU COVID-19 cohort.

Fig. S7. The age-dependency of cell surface proteins.

Table S1. GTEx Baseline expression of ACE2 in different tissues, adjusted for age and sex.

Table S2. MEGENA modules, single cell based average expression and cell percentage.

Table S3. MEGENA modules, ranks and functional enrichments.

Table S4. MEGENA modules, and additional functional enrichments.

Table S5. Statistics of COVID-19 gene expression changes associated with age (FDR \leq 0.05) in GTEx v8.

Table S6. Frequency of tissues that contain COVID-19 age-associated genes.

Table S7. Correlation between CRS (τ) and age in GTEx tissues.

Table S8. SARS-CoV-2 tissue enrichment.

Table S9. Frequency of COVID-19 age-associated genes in GTEx tissues.