Comprehensive characterization of the transcriptional response to COVID-19 in multiple organs reveals shared signatures across tissues

The COVID Tissue Atlas Consortium*

Abstract

Infection by Coronavirus SARS-CoV2 is a severe and often deadly disease that has implications for the respiratory system and multiple organs across the human body. While the effects in the lung have been extensively studied, less is known about COVID-19's cellular impact across other organs. Here we contribute a single-nuclei RNA sequencing atlas comprising six human organs across 20 autopsies where we analyzed the transcriptional changes due to COVID-19 in multiple cell types. Computational cross-organ analysis for endothelial cells and macrophages identified systemic transcriptional changes in these cell types in COVID-19 samples. In addition, analysis of signaling pathways from multiple datasets showed several systemic dysregulations of signaling interaction in different cell types. Altogether, the COVID Tissue Atlas enables the investigation of both cell type-specific and cross-organ transcriptional responses to COVID-19, providing insights into the molecular networks affected by the disease and highlighting novel potential targets for therapies and drug development.

Introduction

One sentence summary: We applied single-nuclei transcriptomics to investigate the molecular response of human cells to SARS-Cov-2 across the body.

COVID-19 (Coronavirus Disease 2019) is the most devastating infectious disease in recent history. The pandemic has impacted all parts of the globe and resulted in nearly 500 million infections and over 6,000,000 deaths (Dong, Du, and Gardner 2020). Approximately 14% of infected unvaccinated individuals develop a severe clinical disease that requires hospitalization (Wu and McGoogan 2020). While the primary organ affected by severe COVID-19 is the lung, many other organs, including the heart, liver, and kidney, are also affected (Mokhtari et al. 2020; Xie et al. 2022; X. Wang et al. 2021). In addition, long-COVID has become an important and common sequela in those who recover from infection. Long-COVID often affects multiple organs and is more common in patients with a severe initial infection (Taquet et al. 2021).

The systemic effects of severe COVID-19 are largely mediated through the immune response to SARS-CoV-2 infection and subsequent inflammatory response. Viral infection stimulates macrophages to overproduce proinflammatory cytokines, including IL-6, leading to the “cytokine storm” that results in systemic inflammatory response syndrome (Hu, Huang, and Yin 2021). This heightened inflammatory state affects multiple organs, partly through effects on endothelial cells, which can be directly injured in response to pro-inflammatory cytokines and produce a procoagulant state leading to thrombosis (Fard et al. 2021). Improved understanding of the cellular and molecular mechanisms that drive severe COVID-19 and lead to damage in specific organs, as well as the development of long-COVID, requires a multi-organ approach.

We have previously shown that multi-organ, single-cell transcriptome-based approaches can yield significant insights into organ biology and cross-organ signaling (Tabula Muris Consortium et al. 2018; Tabula Sapiens Consortium* et al. 2022). In addition, several other studies have recently applied a single cell-based approach to autopsy samples from patients with severe COVID-19. These studies have yielded significant insights into how severe COVID-19 affects the lung (Delorey et al. 2021; Melms et al. 2021) and the brain (A. C. Yang et al. 2021), but have not described in detail the systemic and cross organ effects of severe COVID-19.

Here, we report a COVID single-nuclei RNA seq (snRNA-seq) atlas comprising six organs and approximately 86,000 cells. We showed that transcriptional changes in severe COVID19 infections were not restricted to the lung, the most severely affected organ upon SARS-CoV-2 infection, but to multiple organs, such as the liver and heart. In addition, we found significant changes in the transcriptional profiles of multiple cell types and identified a subset of recurrent molecular pathways commonly upregulated in multiple cell types across organs. The COVID Tissue Atlas (CTA) represents a comprehensive resource to investigate the transcriptional changes resulting from COVID-19 in different tissues. Moreover, the scope of the CTA dataset

enabled us to identify systemic transcriptional signatures that we would have missed by focusing on an individual organ. We anticipate our analysis and the CTA to be of significant value for future research, including identifying molecular targets for drug development and therapeutic applications.

Results

The COVID Tissue Atlas

We collected data from 20 different autopsies (17 males, 3 females) with an age range between 40 and 89 years old (median age = 68 years), of which 15 tested positive for COVID-19 (**Figure 1A**). The average time at which samples were collected was 63 hrs post-mortem. Ethnicities were distributed as Hispanic (n=5), African American (n=2), Asian (n=1), and White (n=12). For COVID-19 positive autopsies, the average positive test time before death was 20 days; however, not all donors died due to COVID-19 complications **(Supplementary Table 1)**. We optimized single-nuclei RNA extraction and sequencing from frozen tissue for Biosafety Level 2 work. All samples were sequenced at the Chan Zuckerberg Biohub using 10x genomics protocols. After quality control, 85,376 cells (60,946 cells from COVID-19 samples and 24,430 cells from healthy donors) were deemed high quality and used to form the CTA (**Figure 1B, C**). Single-nuclei RNA-seq is prone to high levels of ambient RNA contamination, which we corrected by applying an established correction algorithm (Fleming, Marioni, and Babadi 2019) along with filtering of doublets (**Methods**). The total numbers of single cells for each organ were as follows: heart (6,092 healthy; 13,999 COVID-19), lung (9,684 healthy; 11,790 COVID-19), liver (6,768 healthy; 8,889 COVID-19), prostate (1,886 healthy; 8,986 COVID-19), kidney (4,060 COVID-19) and testis (13,222 COVID-19) (**Figure 1D, E**). Additionally, small intestine, colon, and uninfected control kidney specimens were processed but did not yield sufficient high-quality nuclei for inclusion. We were not able to collect uninfected testis tissue.

We applied dimensionality reduction (PCA) and Leiden clustering for each organ while correcting batch effects across donors using scVI (Lopez et al. 2018). Finally, we visualized the resulting clustering using UMAP (McInnes, Healy, and Melville 2018). For each organ, we identified cell populations using the batch-corrected UMAP embedding by tissue experts based on the expression of known gene markers (**Methods**). We were able to identify most major cell types in each organ and verified that clusters with the same cell identity included both healthy and COVID+ cells, as an indication that batch effects were indeed removed (**Figure 1 - Supplement figure 1-2**). Additionally, we verified that our single-nuclei data was statistically comparable to whole-cell sequencing regarding the number of UMIs and detected genes (**Figure 1 - Supplement figure 3**).

Measurements of SARS-CoV-2 mRNA by RT-qPCR showed high to moderate expression in the lung samples from COVID-19 donors **(Figure 1 - Supplement figure 4A)**. While some of the

COVID-19 associated genes such as ACE2, TMPRSS2, and NRP1 were expressed in multiple organs **(Figure 1 - Supplement figure 4C)**, we did not detect significant viral mRNA load by RT-qPCR in the other organs processed (**Figure 1 - Supplement figure 4A)**. The low detection rate of viral mRNA could be attributed to the prolonged periods between initial infection and sample collection for some donors (Deinhardt-Emmer et al. 2021) **(Supplementary Table 1)**. Due to the balanced representation of both healthy and COVID-19 donors for lung, heart, and liver, we decided to focus our downstream analysis mainly on understanding the transcriptional responses of cell types in these organs. For the kidney, we integrated our data with a healthy single-nuclei atlas reference (Muto et al. 2021) and made the integrated object available (**Data Availability**). The results for differential expression analysis between COVID-19 and healthy samples for lung, heart, liver, kidney, and prostate are available as part of the CTA data release **(Supplementary Tables 2 and 3)**. Finally, our testis data, including only COVID-19 samples, is fully annotated and publicly available as part of the CTA release.

Cell type population changes in the COVID-19 lung

The CTA lung dataset comprised 21,474 cells, of which 11,790 were collected from COVID autopsies. After quality control and clustering (**Methods**), we identified ten distinct cell types, including primary epithelial and immune cells (**Figure 2 - Supplement Figure 1A, B**). Several lung single-nuclei and single-cell efforts have been published throughout the COVID-19 pandemic (Delorey et al. 2021; J. Xu et al. 2020; Hasan et al. 2021; Melms et al. 2021). To assess the quality and scope of the CTA, we compared our data to the comprehensive lung atlas generated by the Broad Institute (Delorey et al. 2021). We applied anchor-based integration (Stuart et al. 2019) to the lung samples from both datasets by including autopsies from the Broad atlas as additional donors in the CTA (**Methods**). After integration, the harmonized UMAP embedding showed that all major lung cell types integrated well across datasets (**Figure 2A**), with cells from the CTA and the Broad atlas contributing to most clusters (**Figure 2B**). The alignment between datasets showed that the CTA captured the expected diversity of cell types in the COVID-19 lung and that the gene expression profiles are similar for the same cell types across datasets.

We next focused on the effects that COVID-19 has on the different lung cell populations. In particular, significant epithelial cell damage resulting from COVID-19 is manifested as loss of alveolar type 1 (AT1) and alveolar type 2 (AT2) cells (Melms et al. 2021; Delorey et al. 2021). To investigate the changes in lung epithelial cells in COVID-19 autopsies in detail, we subset and re-clustered the AT1, AT2, and basal cells to obtain a new UMAP embedding (**Figure 2C**). All three cell types included healthy and COVID-19 cells (**Figure 2 - Supplement Figure 1C**) and expressed the corresponding canonical gene markers (**Figure - Supplement Figure 1D**). Consistent with previous studies (Delorey et al. 2021; Melms et al. 2021), we identified loss of AT1 and AT2 cells in COVID-19 lungs relative to healthy controls (**Figure 2D, Figure 2 - Supplement Figure 1A**), along with a significant expansion of basal cells (**Figure 2D, Figure 2 - Supplement Figure 1A**).

Figure 1. A human single-cell atlas enables identification of systemic responses to COVID-19

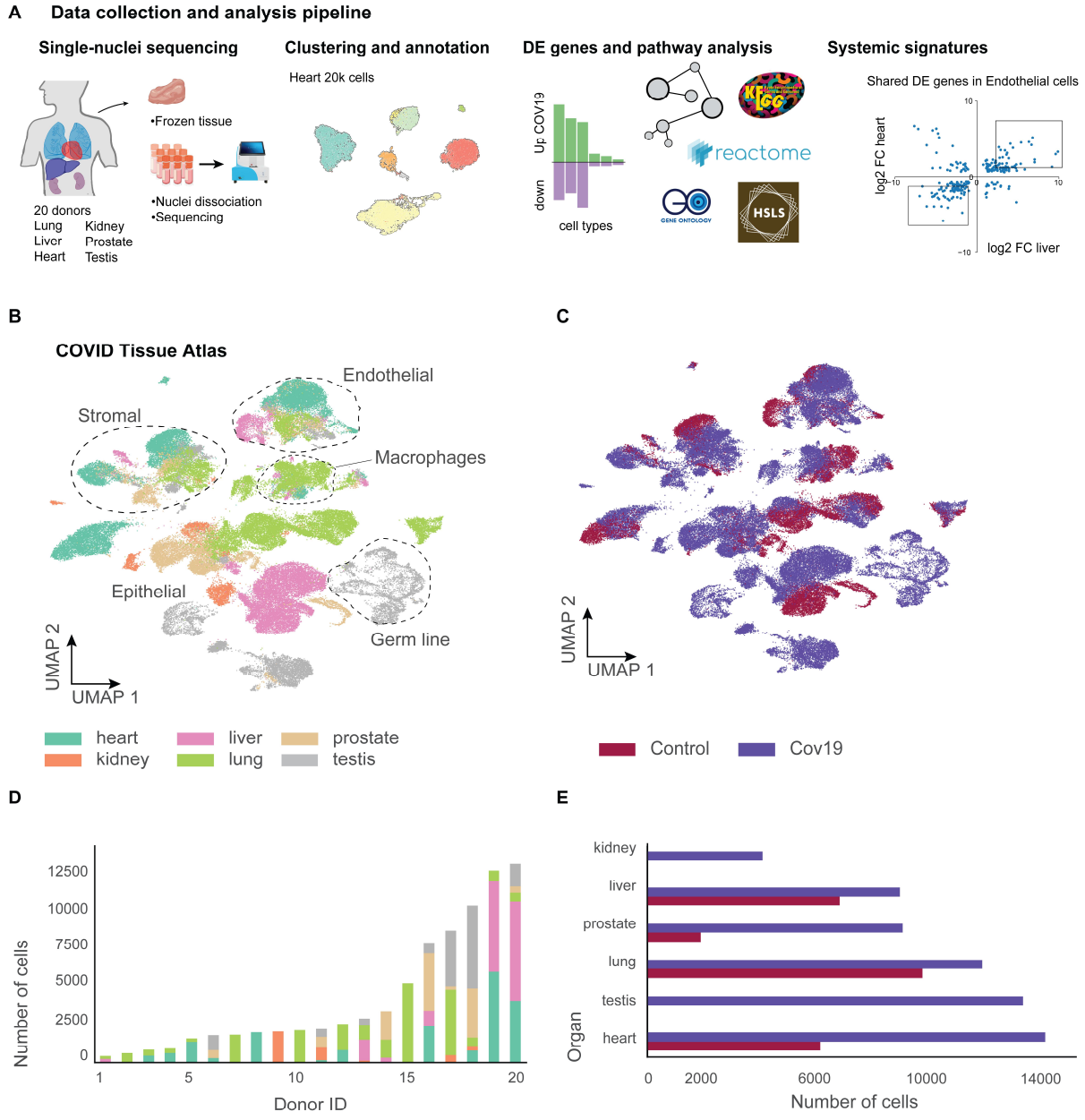


Figure 1

A human single-cell atlas enables the identification of systemic responses to COVID-19.
(**A**) Tissue samples were collected from different organs and frozen, then dissociated into single nuclei. Libraries for snRNA-seq were prepared using 10x Genomics Chromium Next GEM Single Cell 3' v.3.1kit, followed by sequencing on various Illumina platforms. After quality control and clustering, cell types for each organ were annotated by experts using literature gene markers. Differential gene expression and pathway enrichment analysis were performed between COVID-19 and healthy samples for all cell types. Finally, global transcriptional signatures were identified via a cross-organ analysis of differential expression. (**B**) The COVID tissue atlas comprises approximately 85,000 cells from 6 different organs. (**C**) Cells in the COVID tissue atlas cluster by cell identity rather than disease status. (**D**) Number of cells per donor grouped by the organ of origin. (**E**) Number of cells per organ grouped by COVID-19 status.

The increase in basal cells could be explained by trans-differentiation from AT2 cells, via alveolar-basal intermediates (ABIs), a phenomenon recently described *in vitro* that corresponds to cellular changes in fibrotic human lungs (Kathiriya et al. 2022). To investigate if a similar phenomenon occurs in the COVID-19 lungs, we integrated the CTA epithelial cells with a sc-RNA seq dataset of lungs with idiopathic pulmonary fibrosis (IPF) that includes a population of ABIs/Transitional AT2 (Habermann et al. 2020). After applying anchor-based integration (**Methods**), the harmonized UMAP showed that epithelial cells from both datasets generally integrated well (**Figure 2E**). Interestingly, a fraction of the ABIs/Transitional AT2s from (Habermann et al. 2020) mapped to a specific cluster between the AT2 and basal cell populations from the CTA (**Figure 2E - right**). To verify if this cluster from the CTA indeed corresponds to ABIs, we compared the expression profiles of single cells and clustered them by similarity (**Figure 2F**). As a result, we found that a fraction of CTA basal cells clustered together with ABIs/Transitional AT2s. Thus, our data suggest that ABIs (Kathiriya et al. 2022) are present in the COVID-19 lung, and the gain of abnormal basal cells in the alveoli could be accounted for by their trans-differentiation from endogenous AT2s, which are lost in COVID-19 lungs.

Alternatively, the Broad atlas identified a pre-alveolar Type 1 transitional cell state (PATS) population in COVID-19 lungs (Delorey et al. 2021) that bears similarities to what was previously described as ABIs/Transitional AT2s/aberrant basaloid cells from IPF lungs (Adams et al. 2020; Habermann et al. 2020; Kathiriya et al. 2022). We jointly analyzed the lung epithelial cells from the CTA and the Broad atlas to find out if the PATS population was present in our data. We used anchor-based integration (Stuart et al. 2019) and obtained a harmonized UMAP embedding which recapitulated the three populations across datasets (**Figure 2G**). The PATS population mostly overlapped with the principal AT1 cluster (**Figure 2G - right**), but no specific cluster from the CTA mapped directly to the PATS cells. This analysis indicates that the PATS population (Delorey et al. 2021) is likely to be attributed to patient-specific cellular heterogeneity (or sequencing method

Figure 2. Cell composition changes in the COVID-19 lung

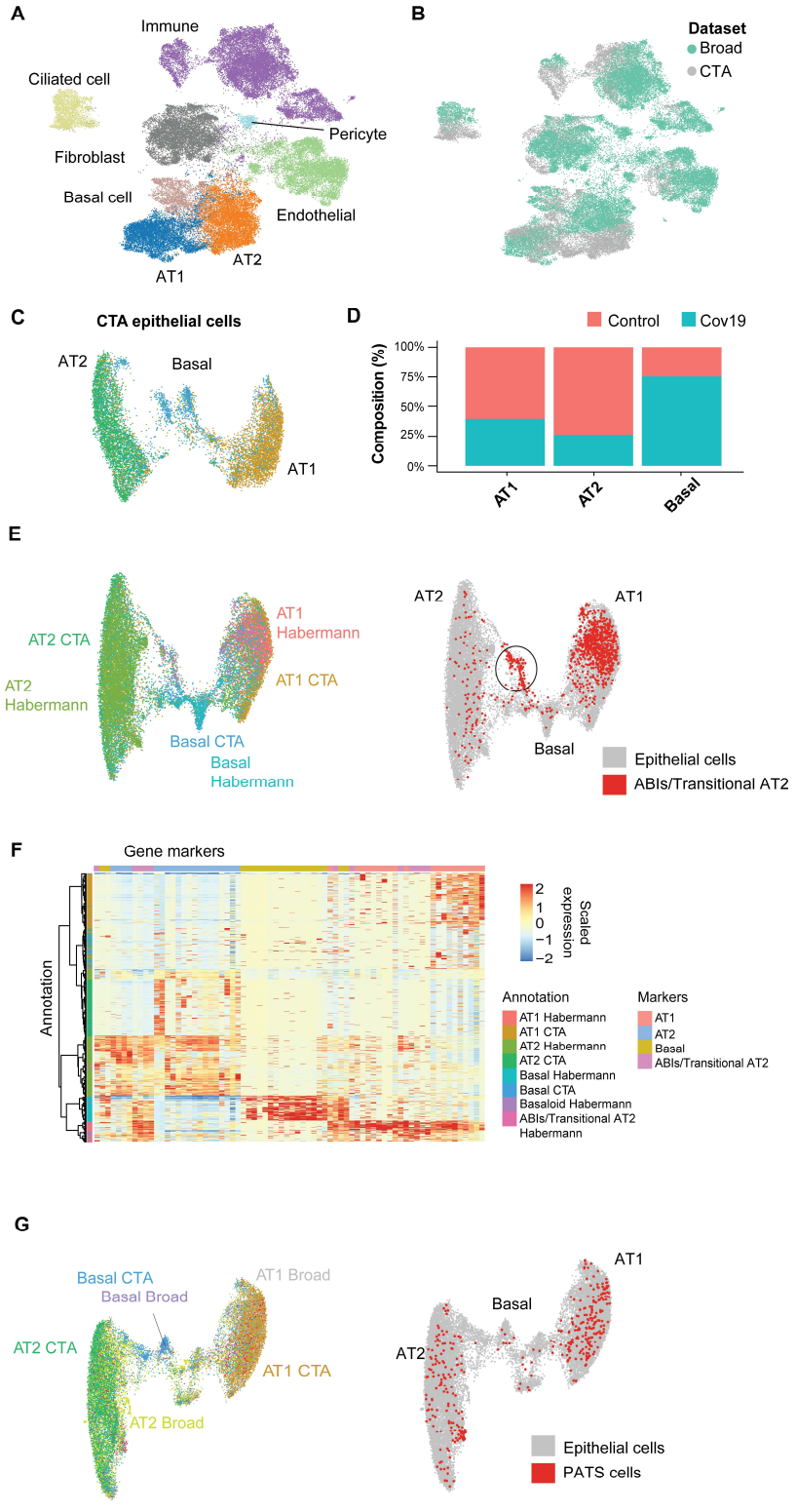


Figure 2

Cell type composition changes in the COVID lung

(**A**) Integration of the CTA lung with the lung COVID atlas by the Broad Institute. A harmonized UMAP shows that cells from both datasets integrate by their corresponding cell type annotation. (**B**) Integration of two lung COVID atlas colored by the dataset of origin. (**C**) Sub-clustering and UMAP projection of the CTA lung epithelial cells (AT1, AT2, and basal cells). (**D**) Relative cell composition in epithelial lung tissue from control and COVID-19 autopsies (CTA data only). (**E**) Integration of CTA epithelial cells and epithelial cells from (Habermann et al. 2020) (AT1, AT2, basal cells, and transitional ABIs/AT2 populations). ABIs/Transitional AT2 from (Habermann et al. 2020) are shown in red (right). (**F**) Heatmap of scaled gene expression of marker genes for all the different cell populations in E. (**G**) Joint embedding of CTA and (Delorey et al. 2021) (AT1, AT2, basal, and PATS cells). The PATS cells identified by (Delorey et al. 2021) are shown in red in the joint UMAP (right).

differences) and, therefore, was not detected in the CTA donors. Together, our results contribute to our understanding of the multiple regenerative strategies involved in re-establishing alveolar epithelial homeostasis in response to COVID-19 (Delorey et al. 2021).

Insulin signaling dysregulation in the liver

Across all six cell types identified in the liver (**Figure 3A**), hepatocytes comprised around 60% of cells in the healthy samples and more than 80% in the COVID-19 samples (**Figure 3B**). However, we observed an inverse trend for endothelial cells, where approximately 20% of the cells from healthy samples were annotated as endothelial as opposed to less than 10% in COVID donors, which may reflect recently reported endotheliopathy in COVID livers (McConnel et al. 2021, J Hep). COVID-19 livers also contained lower proportions of most immune cell populations than controls (**Figure 3B**).

To identify differentially expressed genes for each cell type in the liver, we applied a negative-binomial model implemented in MAST (Finak et al. 2015) that corrects for differences in sequencing depth across samples. Across all cell types, hepatocytes, endothelial cells, and macrophages showed the largest number of differentially expressed (DE) genes in COVID-19 donors (more than 200 upregulated genes with an average log2 Fold-Change >2 and adjusted-p< 1e-6; **Figure 3C**). In contrast, fibroblasts, intrahepatic cholangiocytes, and natural killer cells showed only a fraction of DE genes in comparison (fewer than 50 upregulated genes; **Figure 3C**). Samples from COVID livers generally comprised lower numbers of counts per cell **(Figure 3 - Supplement Figure 1)**; while we corrected for this difference when computing DE genes (**Methods**), we decided to focus on COVID-19 over-expressed genes to minimize potential artifacts in down-regulation resulting from lower sequencing depth.

Figure 3. Transcriptional changes and dysregulation of signaling pathways in the COVID-19 liver

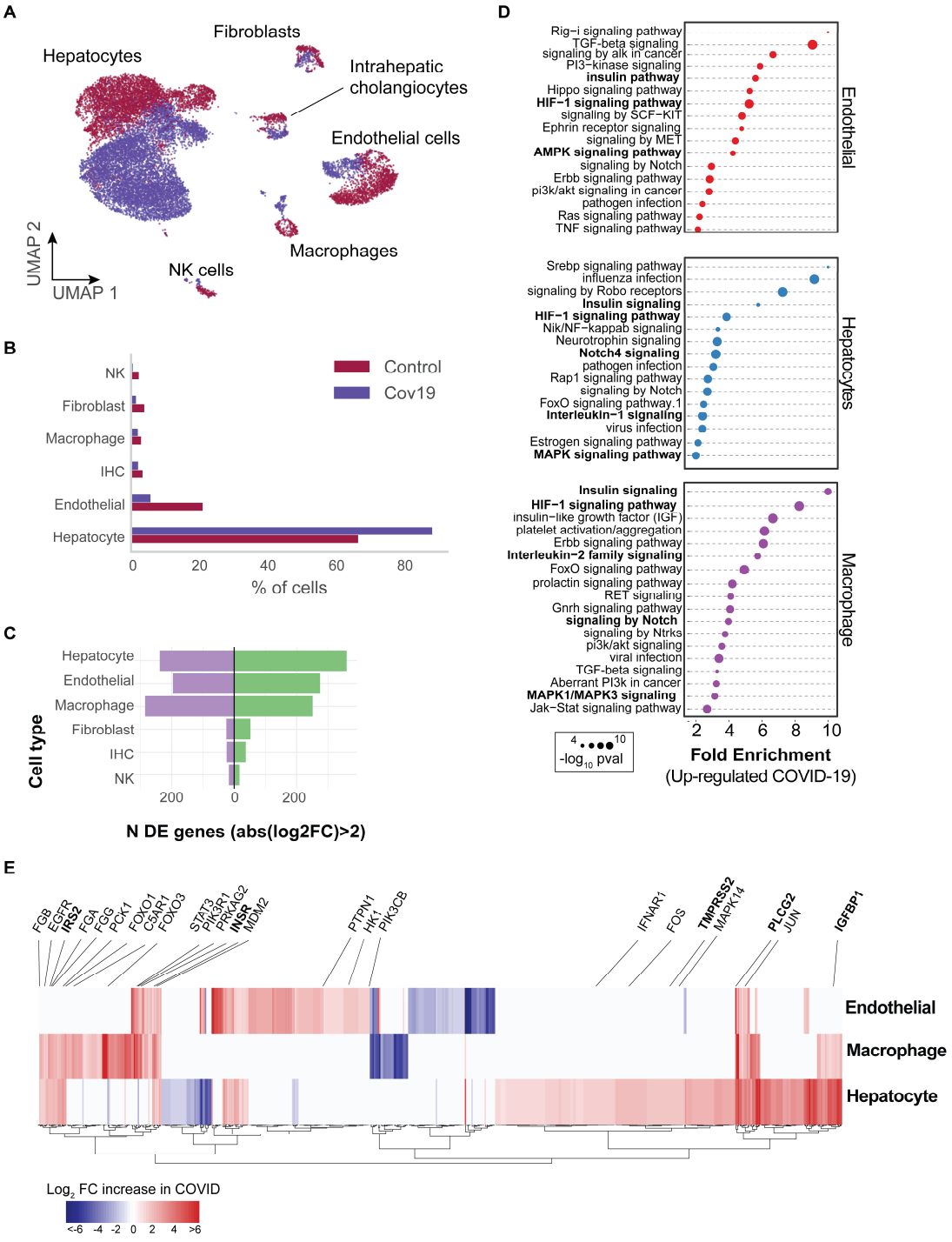


Figure 3

Transcriptional changes and dysregulation of cell signaling in the COVID liver
(**A**) UMAP plot showing all cells from liver samples (n = 6 donors) colored by COVID-19 status. Cell type annotations are indicated for each cluster. (**B**) Fraction of cells for each cell type grouped by COVID-19 status. (**C**) Number of differentially expressed genes found using MAST (Finak et al. 2015) (negative binomial model, correcting for the number of detected genes, p < 1e-6 and log2 FC >2). (**D**) The top enriched signaling pathways found for each cell type based on the DE genes shown in C. (**E**) Heatmap of log2 Fold-Change for the top differentially expressed genes. A few relevant genes are highlighted with a text legend.

Next, we applied pathway enrichment using PathFindR, an algorithm that identifies significant sets of genes based on both a reference pathway database and a protein-protein interaction network (Ulgen, Ozisik, and Sezerman 2019). We identified dysregulated signaling pathways in COVID-19 livers using four different reference pathway databases: KEGG (Kanehisa et al. 2016), BioCarta (Nishimura 2001), GO (Mi et al. 2019), and Reactome (Griss et al. 2020). We found known COVID-19 related gene sets in hepatocytes and macrophages ("*Coronavirus disease - COVID-19*" in the KEGG database), including TMPRSS2, EGFR, PLCG2, MAPK14, FOS, JUN, IFNAR1, C5AR1, CFB, C8G, MASP1, FGA, FGB, FGG in addition to multiple ribosomal-related transcripts, p < 1e-6) (**Supplementary Table 3**). The expression of known COVID-19 genes indicates general agreement between our data and previous studies (Harrison, Lin, and Wang 2020).

Several pathways were enriched in multiple cell types in the liver across all four databases, including Insulin, HIF-1, Notch, MAPK, and FoxO signaling (**Figure 3D, Supplementary Table 3**). We found dysregulation in the insulin signaling pathway in hepatocytes, macrophages, and endothelial cells from COVID-19 livers (**Supplementary Table 3;** p < 1e-6). Specifically, we observed upregulation of genes involved in insulin response, including INSR, PIK3R1, PIK3CB, GSK3B, PPP1CB, PHKA2, PRKAR1A, SORBS1, CBL, CBLB, ACACA, HK1, PRKAG2, RPS6, RHEB, PTPN1 (**Figure 3E, Supplementary Table 2**). Patients with type-2 diabetes have worse outcomes with severe COVID-19 infection (Xie and Al-Aly 2022) and clinical studies show aberrant glucose levels in SARS-Cov2 infected patients with type-2 diabetes (Reiterer et al. 2021). Thus, our data suggest that dysregulated insulin signaling —especially in hepatocytes, which play a critical role in maintaining glucose homeostasis (Klover PJ 2004)— might explain why SARS-Cov2 infected patients with type-2 diabetes have uncontrolled glucose homeostasis and are comorbid (Mishra and Dey 2021)) and why COVID-19 infection could lead to the development of type-2 diabetes (Barrett et al. 2022).

Signaling in the heart in response to COVID-19

COVID-19 can lead to cardiac involvement and injury via the following possible mechanisms: (1) indirect injury due to increased cytokines and immune-inflammatory response, (2) direct invasion of cardiomyocytes by SARS-CoV-2, and (3) respiratory damage from the virus causing hypoxia leading to oxidative stress and injury to cardiomyocytes (Tahir et al. 2020). To understand the

transcriptional changes induced by COVID-19 in the heart, we analyzed differential gene expression across cell types and identified the critical signaling pathways dysregulated as a response to COVID-19.

Heart samples yielded 20,091 cells after quality control (n = 11 donors) (**Figure 4A**). Across the eight cell types identified, the large majority of cells corresponded to endothelial cells (>40% in COVID-19 samples, 13% in healthy samples), cardiomyocytes (25% in COVID-19 samples, 28% in healthy samples), and fibroblasts (15% in COVID-19 samples, 45% in healthy samples) (**Figure 4B**). In addition, we found significant transcriptional changes in cardiomyocytes, endothelial cells, and macrophages based on the number of DE genes in COVID-19 samples (**Figure 4C**). Considering the top DE genes for each cell type, we then focused on understanding how COVID-19 affects heart cells in terms of gene regulatory pathways.

We first confirmed that our results agreed with current gene sets associated with COVID-19 (KEGG: *Coronavirus disease - COVID19* pathway in fibroblasts and macrophages, p <1e-5; Reactome: *Influenza infection* enriched in fibroblasts, p < 1e-5; **Supplementary Table 3**). In addition, multiple genes and GO pathways related to protein translation and ribosome activity (RNA polymerase II cis-regulatory region sequence-specific DNA binding) along with signaling and transcription factor activity (intracellular signal transduction, transcription cis-regulatory region binding, transcription factor binding), were enriched in multiple COVID-19 heart cell types (**Supplementary Table 3**).

Similar to the liver, we observed insulin pathway enrichment in cardiomyocytes from COVID-19 samples (**Figure 4C**). Heart failure is associated with generalized insulin resistance. Moreover, insulin-resistant states such as type 2 diabetes mellitus and obesity increase the risk of heart failure even after adjusting for traditional risk factors (Riehle and Dale Abel 2016). In agreement with our data, other studies found that COVID-19 triggers insulin resistance in patients, causing chronic metabolic disorders that were non-existent before infection (Govender et al. 2021). Additionally, we observed significant changes in Notch, Hippo, and MAPK signaling pathways in cardiomyocytes from COVID-19 samples (**Figure 4D**). Conversely, the BMP and TGFβ signaling pathways showed specific down-regulation in endothelial cells from COVID-19 hearts, including down-regulation of BMPR1A, BMPR1B, SMAD6, and BMP6 **(Supplementary Table 3)**.

Interestingly, Notch signaling has been proposed as a target to prevent SARS-CoV-2 infection and interfere with the progression of COVID-19- associated heart and lung disease (Rizzo et al. 2020). Hippo signaling also appeared as one of the top signaling terms for cardiomyocytes (**Figure 4D**). Recent studies indicate that Hippo signaling is involved in the development of many diseases caused by viruses. Whether virus-induced diseases, specifically COVID-19, can be ameliorated by modulating the Hippo signaling pathway is worth pursuing (Z. Wang et al. 2019). Finally, TGFβ signaling is linked to the response of endothelial cells to inflammation in COVID-19 (Yoshimatsu and Watabe 2022). Together, these results build on previously reported evidence to show that multiple signaling pathways in the heart undergo both cell type-specific and systemic changes in response to COVID-19.

Figure 4. Transcriptional changes and dysregulation of signaling pathways in the COVID-19 heart

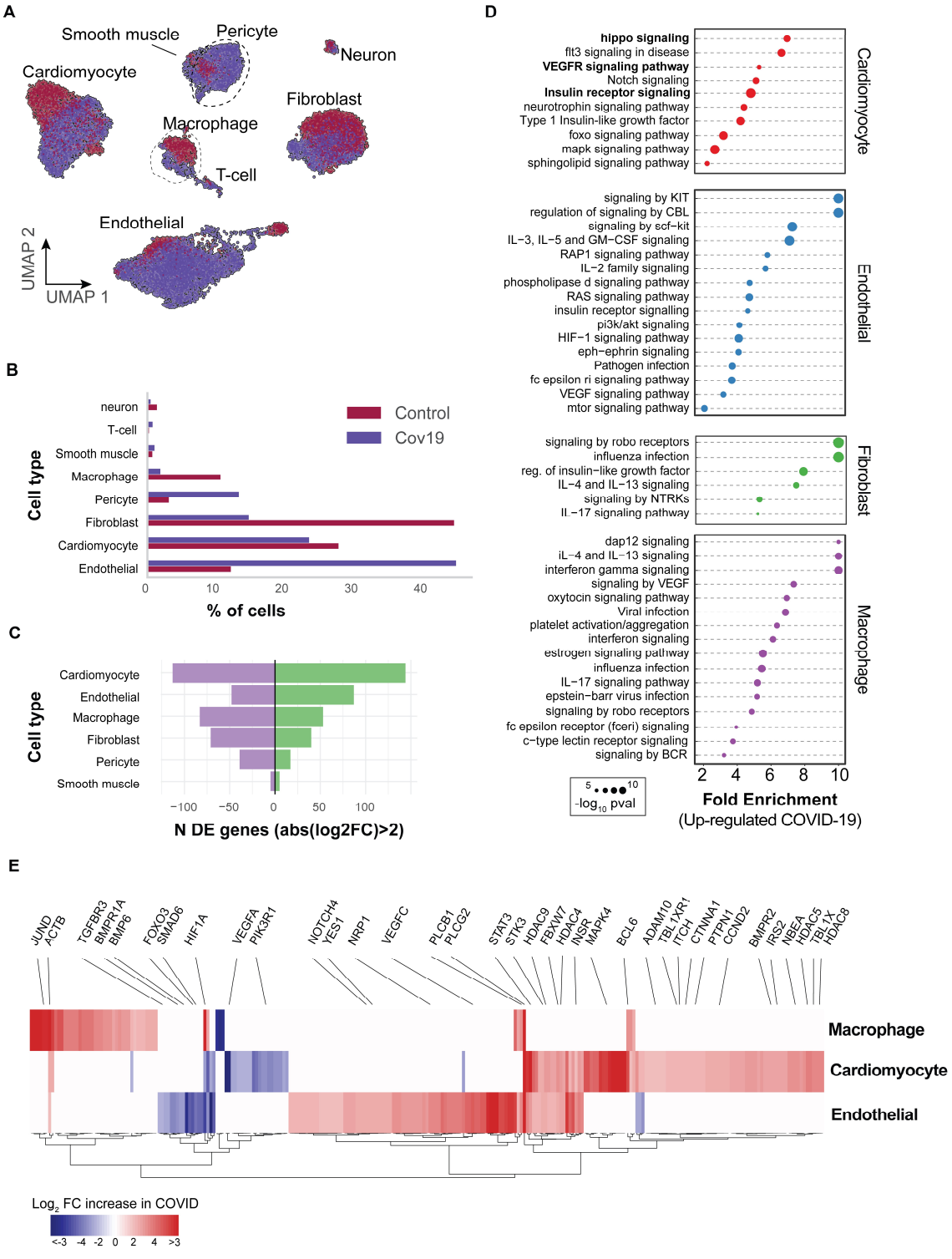


Figure 4

Transcriptional changes and dysregulation of cell signaling in the COVID heart
(**A**) UMAP plot showing all cells from heart samples (n = 11 donors) colored by COVID-19 status. Cell type annotations are indicated for each cluster. (**B**) Fraction of cells for each cell type grouped by COVID-19 status. (**C**) Number of differentially expressed genes found using MAST (Finak et al. 2015) (negative binomial model, correcting for the number of detected genes, p < 1e-6 and log2 FC >2). (**D**) The top signaling pathways found for each cell type using the genes in C. (**E**) Heatmap of log2 Fold-Change for the top differentially expressed genes. A few relevant genes are highlighted with a text legend.

Shared transcriptional responses across organs

The CTA provides a unique opportunity to identify systemic transcriptional responses across organs. As an indication of a systemic response to COVID-19, we found enrichment of the same signaling pathways in multiple cell types and across organs, including HIF-1, insulin, and Notch signaling (**Figure 3D, 4D**). Therefore, we decided to quantify the cross-organ transcriptional changes in COVID-19 autopsies by finding overlapping sets of differentially expressed genes and signaling pathways across organs.

In macrophages, we found a significant overlap in DE genes across organs compared to random sampling expectations (**Figure 5A**). Specifically, we found a set of 89 DE genes in COVID-19 macrophages from all three organs, including PLCG2, HIF1A, ACTB, and JUND. There were also many overlapping DE genes in pairs of organs, with macrophages from the liver and lung showing the highest overlap with 124 shared DE genes (**Figure 5A, Supplementary Table 4**). We performed the same analysis for endothelial cells and similarly found sets of overlapping DE genes; the highest overlap occurring between endothelial cells from the liver and lung (**Figure 5 - Supplement figure 1** and **Supplementary Table 4**).

To further analyze these data, we defined the shared transcriptional response (STR) for a cell type as the set of genes that show differential expression in at least three organs from COVID-19 donors (p<1e-4 and log2 FC >1). We restricted the analysis to macrophages, endothelial cells, and stromal cells, which appear in multiple organs, and calculated the correlation between the log-FC values of all genes in the STR across pairs of organs (**Figure 5 B-E**). Generally, we saw high correlation coefficients, indicating coordination in the COVID-19 induced STR across organs. For example, among the genes with the highest log-FC across organs, we found HIF1A (in macrophages from liver, lung, and heart; **Figure 5B-C**), JUND (in macrophages from liver and heart; **Figure 5C**), and PLCG2 (in endothelial cells from liver and heart; **Figure 5D**).

To identify the cell types with high coordination in their STR across organs, we defined a coordination score by considering pairs of organs and the fraction of genes in the STR that showed the same direction in DE (up-regulated in both or down-regulated in both; **Figure 5F**). Finally, we generated a null distribution for the expected coordination by shuffling the log-FC

across genes (**Figure 5F; Methods**) and computed a z-score between the null distribution and the observed coordination for each cell type and pairs of organs (**Figure 5G**).

The STR of endothelial cells from COVID-19 samples showed the highest coordination across multiple pairs of organs (**Figure 5G**). These results are consistent with previous studies focused on the effect of COVID-19 on endothelial tissues (Ruhl et al. 2021; Huertas et al. 2020). We also found significant coordination in macrophages across the liver, lung, and heart (z-score > 5; **Figure 5G**). Macrophages from the lung showed lower coordination scores compared to the heart and liver, an indication of lung-specific transcriptional regulation (**Figure 5G** and **Figure 5B** off-diagonal quadrants). In contrast, the STR of fibroblasts and stromal cells showed no significant coordination compared to the randomized control, possibly due to the low number of overlapping DE genes (**Figure 5G** and **Figures 3C, 4C**). Together, these results indicate that COVID-19 infection induces coordinated transcriptional regulation in macrophages and endothelial cells across multiple organs.

Systemic transcriptional responses in endothelial cells and macrophages

To investigate the relevance of the COVID-19 STR in macrophages and endothelial cells, we identified enriched pathways considering the sets of genes that showed coordinated DE in at least three organs (diagonal quadrants in **Figure 5B-E**, **Supplementary Table 4**). We visualized the results as a matrix of pathways vs. organs, including the top pathway terms (p < 1e-3, Fold Enrichment > 3) that appeared in at least two organs for macrophages (**Figure 6A**) and endothelial cells (**Figure 6B**). Multiple signaling pathways were enriched in the STR of macrophages across organs (**Figure 6A**). The HIF-1 pathway showed high Fold Enrichment across all organs (**Figure 6A**), suggesting a pivotal role of macrophages in the systemic response to oxygen homeostasis in COVID-19. The Notch pathway was also enriched in macrophages from all three organs (**Figure 6A**), confirming that Notch signaling has a crucial role in the systemic response to COVID-19 (Breikaa and Lilly 2021; Farahani et al. 2022).

The STR consists of shared DE genes across multiple organs, however, the magnitude of the differential expression of a given gene, in terms of log-FC and p-value, can vary across organs **(Figure 5B-E, Supplementary Table 4)**. Therefore, when performing pathway analysis, some signaling pathways showed statistically significant enrichment only in subsets of tissues. For example, in macrophages, Interleukin-4/13 showed significant enrichment only in the liver and heart) and the adherens junction pathway only in the liver and the lung **(Figure 6A)**. Similarly, a few gene pathways showed organ-specific enrichment (**Supplementary Table 5**), indicating that genes in the SRT, while simultaneously differentially expressed across organs, might also modulate some cellular processes in organ-specific ways, due to quantitative DE differences.

In the coordinated STR of endothelial cells, we found multiple enriched pathways, including Notch and Ephrin signaling in the lung, liver, and heart (**Figure 6B**). Specifically, several Notch-related genes were up-regulated in COVID-19 samples for all three organs, including HDAC9, a selective regulator of Notch, FBXW7, a regulator of angiogenesis through Notch (Izumi et al. 2012), and TBLR1, an indirect Notch regulator through degradation (Perissi et al. 2008). Additionally, we

Figure 5. A shared transcriptional response in macrophages and endothelial cells across organs

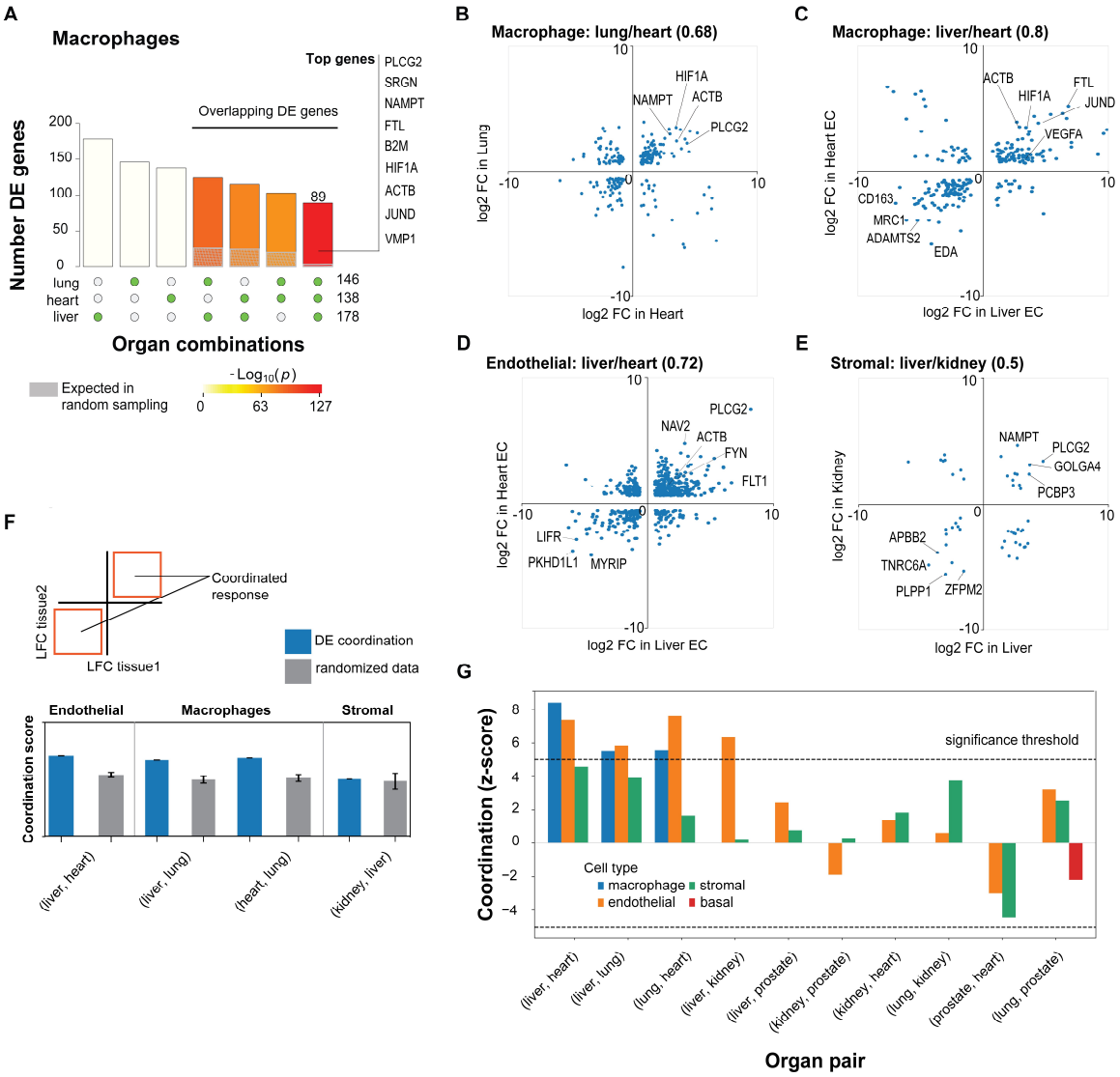


Figure 5

A shared transcriptional response in macrophages and endothelial cells across organs

(**A**) Overlap of differentially expressed genes in COVID-19 macrophages across organs. The gray shaded area indicates the expected overlap for each organ combination (green circles) under a null hypothesis of random sampling (we computed the *p* values against this null model). The white bars indicate the number of genes that showed DE in a single organ. The names of the top genes DE in all three organs are shown based on their log2 Fold Enrichment. (**B**) Scatter plot comparing the log2 FC for DE genes in COVID-19 macrophages from lung and heart. (**C**) Same as B but comparing DE genes in COVID-19 macrophages from the liver and heart. (**D**) log2 Fold-Change for COVID-19 endothelial cells from the liver and heart. (**E**)

log2 Fold-Change for COVID-19 stromal cells from the liver and kidney. (**F**) A fully coordinated transcriptional signature would imply that all genes lie in the bottom-left and top-right quadrants (red squares). We define the coordination score as the number of DE genes that show the same direction (up-up, down-down) for the two organs, divided by the number of shared DE genes (A). Gray bars show the score expectation when sampling DE genes randomly from each organ. (**G**) Coordination scores for different cell types across all pairs of organs. The dotted line indicates a significance threshold of z-score > 5 standard deviations compared to the expectation by chance.

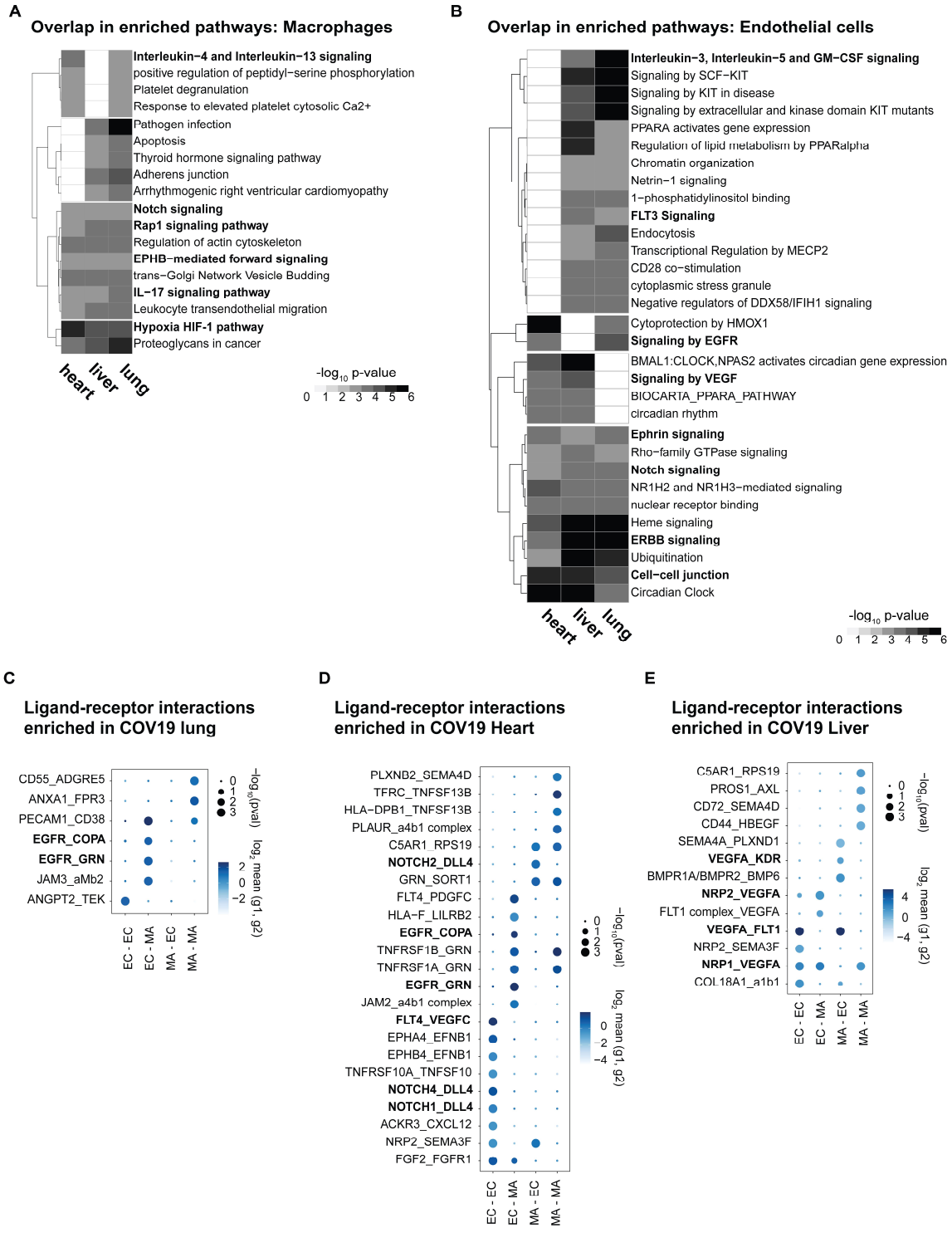
found enrichment for VEGF signaling in liver and heart (p < 1e-3, and lung p < 1e-2; **Supplementary Table 6**). Interestingly, despite up-regulation of the VEGF signaling pathway in multiple organs, some pathway genes showed organ-specific regulation. For example, AKT3 was enriched in the liver and lung, whereas PXN contributed to VEGF signaling enrichment only in the heart and lung (**Supplementary Table 6**). A recent study using measurements of growth factors and cytokines in serum identified VEGF-D as the most predictive indicator for the severity of COVID-19 (Kong et al. 2020). Similarly, VEGF was proposed as a promising therapeutic target for suppressing inflammation during SARS-CoV-2 infection (Yin et al. 2020). Our results indicate that changes in VEGF signaling in COVID-19 donors are not necessarily organ-specific but rather part of a systemic response of endothelial cells and, therefore, of relevance for the development of treatments and as potential drug targets.

Macrophage-Endothelial signaling interactions in COVID tissues

The enrichment of key cell-to-cell pathways such as Notch and Ephrin in the STR of endothelial cells and macrophages due to COVID-19 suggests that these two cell types may be signaling to each other. Therefore, we used CellPhoneDB (Efremova et al. 2020) to investigate potential signaling interactions between these two cell types by finding over-represented expression ligand-receptor pairs in COVID-19 samples compared to healthy donors (**Methods**).

Multiple enriched ligand-receptor pairs were identified between macrophages and endothelial cells in all three organs from COVID-19 autopsies (23 ligand-receptor pairs in the heart, 13 in the liver, and 7 in the lung, p < 1e-2; **Figure 6C-E**). Among the top signaling interactions, we found expression of VEGF ligand-receptor pairs in the liver and heart (FLT4:VEGFC; VEGFA: KDR; NRP2:VEGFA; VEGFA:FLT1). In the lung, we found expression of EGFR in endothelial cells and expression of COPA and GRN in macrophages, suggesting another mechanism of cell-cell signaling between these two cell types. In the heart, we found multiple Notch ligand-receptor enriched pairs involving the expression of the Dll4 ligand in endothelial cells (**Figure 6D**). Interestingly, the expression of Notch receptors was cell-type dependent: endothelial cells expressed Notch4 and Notch1, whereas macrophages expressed Notch2 (**Figure 6D**). A Dll4-dependent signaling mechanism involving endothelial cells and macrophages in the COVID-19 heart is potentially related to HIF-1 signaling since these pathways are known to cross-talk through multiple mechanisms (Breikaa and Lilly 2021; Zheng et al. 2008).

Figure 6. Identifying signaling interactions between endothelial cells and macrophages



448 Figure 6

449 **Identifying interactions between endothelial cells and macrophages**
450 (**A**) From the shared DE genes across organs, we identified the top enriched signaling pathways for COVID-
451 19 macrophages across the lung, liver, and heart. The value in the heatmap is the log10 p-value for the
452 gene pathway. Only pathways with Fold Enrichment > 3 and adjusted p-value < 1e-3 in at least two organs
453 are shown. (**B**) Enriched pathways in the shared transcriptional response of endothelial cells across lung,
454 liver, and heart (using the same significance thresholds as A). (**C-E**) Enriched expression of ligand-receptor
455 components in COVID-19 macrophages and endothelial cells in the lung (**C**), heart (**D**) and liver (**E**). The
456 x-axis indicates the pair of cell types considered (EC endothelial cells, MA macrophages). The y-axis
457 indicates all the enriched signaling interactions found, and the circles indicate the significance and
458 magnitude of enrichment. We calculated enrichment using CellPhoneDB on the raw sequencing counts.
459 Only ligand-receptor pairs with adjusted p-value < 1e-3 are shown.

460

461 Discussion

462 We generated the CTA, a single-cell atlas of six organs from autopsies of COVID patients. Our
463 analyses highlight that multiple organs are damaged by COVID-19 infection and allow for
464 assessing transcriptomic changes in multiple cell types across these organs. While the lung is the
465 primary organ affected by COVID infection, our data identified broad signaling changes across
466 multiple organs and cell types. Notably, we localize signaling changes in two affected organs, the
467 liver and heart, where we identified dysregulated insulin and HIF signaling and prominent
468 macrophage-endothelial interactions.

469 Through analysis of the CTA, we identified a shared transcriptional signature (STR) in COVID-19
470 autopsy specimens across tissues. This transcriptional signature was evident in macrophages
471 and endothelial cells in hearts and liver from COVID-19 tissue specimens compared to control
472 specimens. These shared signatures between macrophages and endothelial cells may be
473 mediated by the known effects of the dysregulated immune system in the context and sequelae
474 of COVID infection.

475 The effects of COVID-19 on the human body are yet to be fully understood, and we need
476 comprehensive maps of the changes at the transcriptional and proteomic levels. The CTA and
477 the corresponding analyses represent an integrated effort toward understanding the effects of this
478 disease from an organism-wide point of view. More generally, we expect some of the
479 computational analysis presented in this study to be generalized to other cell atlas datasets to
480 reveal systemic transcriptional signatures of disease by analyzing the responses of individual
481 cells while considering the global context of the human body. Our results may also have
482 implications for understanding the sequelae of COVID-19 across organs and increased risk for
483 diseases associated with COVID-19 infection. For example, insulin signaling dysregulation may
484 contribute to the development of diabetes in COVID-19 patients. Long COVID, which appears to

485 be a complex set of symptoms with variable organ dysfunction, may also be informed by our
486 understanding of cellular changes across multiple tissues.

487 Overall, the CTA contributes to our molecular understanding of the effects of severe SARS-CoV2
488 infection across multiple organs and cell types.

References

Adams, Taylor S., Jonas C. Schupp, Sergio Poli, Ehab A. Ayaub, Nir Neumark, Farida Ahangari, Sarah G. Chu, et al. 2020. “Single-Cell RNA-Seq Reveals Ectopic and Aberrant Lung-Resident Cell Populations in Idiopathic Pulmonary Fibrosis.” *Science Advances* 6 (28): eaba1983.

Barrett, Catherine E., Alain K. Koyama, Pablo Alvarez, Wilson Chow, Elizabeth A. Lundeen, Cria G. Perrine, Meda E. Pavkov, et al. 2022. “Risk for Newly Diagnosed Diabetes >30 Days After SARS-CoV-2 Infection Among Persons Aged <18 Years - United States, March 1, 2020-June 28, 2021.” *MMWR. Morbidity and Mortality Weekly Report* 71 (2): 59–65.

Breikaa, Randa M., and Brenda Lilly. 2021. “The Notch Pathway: A Link Between COVID-19 Pathophysiology and Its Cardiovascular Complications.” *Frontiers in Cardiovascular Medicine* 8 (May): 681948.

Deinhardt-Emmer, Stefanie, Daniel Wittschieber, Juliane Sanft, Sandra Kleemann, Stefan Elschner, Karoline Frieda Haupt, Vanessa Vau, et al. 2021. “Early Postmortem Mapping of SARS-CoV-2 RNA in Patients with COVID-19 and the Correlation with Tissue Damage.” *eLife* 10 (March). https://doi.org/10.7554/eLife.60361.

Delorey, Toni M., Carly G. K. Ziegler, Graham Heimberg, Rachelly Normand, Yiming Yang, Åsa Segerstolpe, Domenic Abbondanza, et al. 2021. “COVID-19 Tissue Atlases Reveal SARS-CoV-2 Pathology and Cellular Targets.” *Nature* 595 (7865): 107–13.

Dong, Ensheng, Hongru Du, and Lauren Gardner. 2020. “An Interactive Web-Based Dashboard to Track COVID-19 in Real Time.” *The Lancet Infectious Diseases* 20 (5): 533–34.

Efremova, Mirjana, Miquel Vento-Tormo, Sarah A. Teichmann, and Roser Vento-Tormo. 2020. “CellPhoneDB: Inferring Cell-Cell Communication from Combined Expression of Multi-Subunit Ligand-Receptor Complexes.” *Nature Protocols* 15 (4): 1484–1506.

Farahani, Masoumeh, Zahra Niknam, Leila Mohammadi Amirabad, Nasrin Amiri-Dashatan, Mehdi Koushki, Mohadeseh Nemati, Fahima Danesh Pouya, Mostafa Rezaei-Tavirani, Yousef Rasmi, and Lobat Tayebi. 2022. “Molecular Pathways Involved in COVID-19 and Potential Pathway-Based Therapeutic Targets.” *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie* 145 (January): 112420.

Fard, Mahin Behzadi, Samaneh Behzadi Fard, Shahin Ramazi, Amir Atashi, and Zahra Eslamifar. 2021. “Thrombosis in COVID-19 Infection: Role of Platelet Activation-Mediated Immunity.” *Thrombosis Journal* 19 (1): 59.

Finak, Greg, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, et al. 2015. “MAST: A Flexible Statistical Framework for Assessing Transcriptional Changes and Characterizing Heterogeneity in Single-Cell RNA Sequencing Data.” *Genome Biology* 16 (December): 278.

Fleming, Stephen J., John C. Marioni, and Mehrtash Babadi. 2019. “CellBender Remove-Background: A Deep Generative Model for Unsupervised Removal of Background Noise from scRNA-Seq Datasets.” *bioRxiv*. https://doi.org/10.1101/791699.

Govender, Nalini, Olive P. Khaliq, Jagidesa Moodley, and Thajasvarie Naicker. 2021. “Insulin Resistance in COVID-19 and Diabetes.” *Primary Care Diabetes*. https://doi.org/10.1016/j.pcd.2021.04.004.

Griss, Johannes, Guilherme Viteri, Konstantinos Sidiropoulos, Vy Nguyen, Antonio Fabregat, and Henning Hermjakob. 2020. “ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis.” *Molecular & Cellular Proteomics: MCP* 19 (12): 2115–25.

Habermann, Arun C., Austin J. Gutierrez, Linh T. Bui, Stephanie L. Yahn, Nichelle I. Winters, Carla L. Calvi, Lance Peter, et al. 2020. “Single-Cell RNA Sequencing Reveals Profibrotic Roles of Distinct Epithelial and Mesenchymal Lineages in Pulmonary Fibrosis.” *Science Advances* 6 (28): eaba1972.

Harrison, Andrew G., Tao Lin, and Penghua Wang. 2020. “Mechanisms of SARS-CoV-2 Transmission and Pathogenesis.” *Trends in Immunology* 41 (12): 1100–1115.

Hasan, Md Zobaer, Syful Islam, Kenichi Matsumoto, and Taro Kawai. 2021. “Meta-Analysis of Single-Cell RNA-Seq Data Reveals Phenotypic Switching of Immune Cells in Severe COVID-19 Patients.” *Computers in Biology and Medicine* 137 (October): 104792.

Hu, Biying, Shaoying Huang, and Lianghong Yin. 2021. “The Cytokine Storm and COVID-19.” *Journal of Medical Virology* 93 (1): 250–56.

Huertas, Alice, David Montani, Laurent Savale, Jérémie Pichon, Ly Tu, Florence Parent, Christophe Guignabert, and Marc Humbert. 2020. “Endothelial Cell Dysfunction: A Major Player in SARS-CoV-2 Infection (COVID-19)?” *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology*. https://doi.org/10.1183/13993003.01634-2020.

Izumi, Nanae, Christian Helker, Manuel Ehling, Axel Behrens, Wiebke Herzog, and Ralf H. Adams. 2012. “Fbxw7 Controls Angiogenesis by Regulating Endothelial Notch Activity.” *PloS One* 7 (7): e41116.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. “KEGG as a Reference Resource for Gene and Protein Annotation.” *Nucleic Acids Research* 44 (D1): D457–62.

Kathiriya, Jaymin J., Chaoqun Wang, Minqi Zhou, Alexis Brumwell, Monica Cassandras, Claude Jourdan Le Saux, Max Cohen, et al. 2022. “Human Alveolar Type 2 Epithelium Transdifferentiates into Metaplastic KRT5 Basal Cells.” *Nature Cell Biology* 24 (1): 10–23.

Kong, Yaxian, Junyan Han, Xueying Wu, Hui Zeng, Jingyuan Liu, and Henghui Zhang. 2020. “VEGF-D: A Novel Biomarker for Detection of COVID-19 Progression.” *Critical Care / the Society of Critical Care Medicine* 24 (1): 373.

Lopez, Romain, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. 2018. “Deep Generative Modeling for Single-Cell Transcriptomics.” *Nature Methods* 15 (12): 1053–58.

McInnes, Leland, John Healy, and James Melville. 2018. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.” https://doi.org/10.48550/ARXIV.1802.03426.

Melms, Johannes C., Jana Biermann, Huachao Huang, Yiping Wang, Ajay Nair, Somnath Tagore, Igor Katsyv, et al. 2021. “A Molecular Single-Cell Lung Atlas of Lethal COVID-19.” *Nature* 595 (7865): 114–19.

Mi, Huaiyu, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D. Thomas. 2019. “PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools.” *Nucleic Acids Research* 47 (D1): D419–26.

Mishra, Devanshi, and Chinmoy Sankar Dey. 2021. “Type-2 Diabetes, a Co-Morbidity in Covid-19: Does Insulin Signaling Matter?” *Biochemical Society Transactions* 49 (2): 987–95.

Mokhtari, Tahmineh, Fatemeh Hassani, Neda Ghaffari, Babak Ebrahimi, Atousa Yarahmadi, and Ghomareza Hassanzadeh. 2020. “COVID-19 and Multiorgan Failure: A Narrative Review on Potential Mechanisms.” *Journal of Molecular Histology* 51 (6): 613–28.

Muto, Yoshiharu, Parker C. Wilson, Nicolas Ledru, Haojia Wu, Henrik Dimke, Sushrut S. Waikar, and Benjamin D. Humphreys. 2021. “Single Cell Transcriptional and Chromatin Accessibility Profiling Redefine Cellular Heterogeneity in the Adult Human Kidney.” *Nature Communications* 12 (1): 2190.

Nishimura, Darryl. 2001. “BioCarta.” *Biotech Software & Internet Report* 2 (3): 117–20.

Perissi, Valentina, Claudio Scafoglio, Jie Zhang, Kenneth A. Ohgi, David W. Rose, Christopher K. Glass, and Michael G. Rosenfeld. 2008. “TBL1 and TBLR1 Phosphorylation on Regulated Gene Promoters Overcomes Dual CtBP and NCoR/SMRT Transcriptional Repression Checkpoints.” *Molecular Cell* 29 (6): 755–66.

Reiterer, Moritz, Mangala Rajan, Nicolás Gómez-Banoy, Jennifer D. Lau, Luis G. Gomez-Escobar, Lunkun Ma, Ankit Gilani, et al. 2021. “Hyperglycemia in Acute COVID-19 Is

589 Characterized by Insulin Resistance and Adipose Tissue Infectivity by SARS-CoV-2." *Cell*
590 *Metabolism* 33 (11): 2174–88.e5.
591 Riehle, Christian, and E. Dale Abel. 2016. "Insulin Signaling and Heart Failure." *Circulation*
592 *Research*. https://doi.org/10.1161/circresaha.116.306206.
593 Rizzo, Paola, Francesco Vieceli Dalla Sega, Francesca Fortini, Luisa Marracino, Claudio
594 Rapezzi, and Roberto Ferrari. 2020. "COVID-19 in the Heart and the Lungs: Could We
595 'Notch' the Inflammatory Storm?" *Basic Research in Cardiology* 115 (3): 31.
596 Rodriguez-Lanetty, Mauricio, Wendy S. Phillips, and Virginia M. Weis. 2006. "Transcriptome
597 Analysis of a Cnidarian-Dinoflagellate Mutualism Reveals Complex Modulation of Host
598 Gene Expression." *BMC Genomics* 7 (February): 23.
599 Ruhl, Louisa, Isabell Pink, Jenny F. Kühne, Kerstin Beushausen, Jana Keil, Stella Christoph,
600 Andrea Sauer, et al. 2021. "Endothelial Dysfunction Contributes to Severe COVID-19 in
601 Combination with Dysregulated Lymphocyte Responses and Cytokine Networks." *Signal*
602 *Transduction and Targeted Therapy* 6 (1): 418.
603 Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William
604 M. Mauck 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019.
605 "Comprehensive Integration of Single-Cell Data." *Cell* 177 (7): 1888–1902.e21.
606 Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and
607 processing, Library preparation and sequencing, Computational data analysis, Cell type
608 annotation, Writing group, Supplemental text writing group, and Principal investigators.
609 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris." *Nature*
610 562 (7727): 367–72.
611 Tabula Sapiens Consortium*, Robert C. Jones, Jim Karkanias, Mark A. Krasnow, Angela
612 Oliveira Pisco, Stephen R. Quake, Julia Salzman, et al. 2022. "The Tabula Sapiens: A
613 Multiple-Organ, Single-Cell Transcriptomic Atlas of Humans." *Science* 376 (6594):
614 eabl4896.
615 Tahir, Faryal, Taha Bin Arif, Jawad Ahmed, Farheen Malik, and Muhammad Khalid. 2020.
616 "Cardiac Manifestations of Coronavirus Disease 2019 (COVID-19): A Comprehensive
617 Review." *Cureus* 12 (5): e8021.
618 Taquet, Maxime, Quentin Dercon, Sierra Luciano, John R. Geddes, Masud Husain, and Paul J.
619 Harrison. 2021. "Incidence, Co-Occurrence, and Evolution of Long-COVID Features: A 6-
620 Month Retrospective Cohort Study of 273,618 Survivors of COVID-19." *PLoS Medicine* 18
621 (9): e1003773.
622 The Tabula Sapiens Consortium, and Stephen R. Quake. 2021. "The Tabula Sapiens: A
623 Multiple Organ Single Cell Transcriptomic Atlas of Humans." *bioRxiv*. bioRxiv.
624 https://doi.org/10.1101/2021.07.19.452956.
625 Ulgen, Ege, Ozan Ozisik, and Osman Ugur Sezerman. 2019. "pathfindR: An R Package for
626 Comprehensive Identification of Enriched Pathways in Omics Data Through Active
627 Subnetworks." *Frontiers in Genetics* 10 (September): 858.
628 Wang, Xinyi, Jianyong Lei, Zhihui Li, and Lunan Yan. 2021. "Potential Effects of Coronaviruses
629 on the Liver: An Update." *Frontiers of Medicine* 8 (September): 651658.
630 Wang, Zhilong, Wanhang Lu, Yiling Zhang, Feng Zou, Zhigang Jin, and Tiejun Zhao. 2019. "The
631 Hippo Pathway and Viral Infections." *Frontiers in Microbiology* 10: 3033.
632 Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis. 2018. "SCANPY: Large-Scale Single-
633 Cell Gene Expression Data Analysis." *Genome Biology* 19 (1): 15.
634 Wolock, Samuel L., Romain Lopez, and Allon M. Klein. 2019. "Scrublet: Computational
635 Identification of Cell Doublets in Single-Cell Transcriptomic Data." *Cell Systems* 8 (4): 281–
636 91.e9.
637 Wu, Zunyou, and Jennifer M. McGoogan. 2020. "Characteristics of and Important Lessons From
638 the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72
639 314 Cases From the Chinese Center for Disease Control and Prevention." *JAMA: The*

Journal of the American Medical Association 323 (13): 1239–42.

Xie, Yan, and Ziyad Al-Aly. 2022. "Risks and Burdens of Incident Diabetes in Long COVID: A Cohort Study." *The Lancet. Diabetes & Endocrinology*, March. https://doi.org/10.1016/S2213-8587(22)00044-4.

Xie, Yan, Evan Xu, Benjamin Bowe, and Ziyad Al-Aly. 2022. "Long-Term Cardiovascular Outcomes of COVID-19." *Nature Medicine* 28 (3): 583–90.

Xu, Chenling, Romain Lopez, Edouard Mehlman, Jeffrey Regier, Michael I. Jordan, and Nir Yosef. 2021. "Probabilistic Harmonization and Annotation of Single-Cell Transcriptomics Data with Deep Generative Models." *Molecular Systems Biology* 17 (1): e9620.

Xu, Jincheng, Xiaoyue Xu, Lina Jiang, Kamal Dua, Philip M. Hansbro, and Gang Liu. 2020. "SARS-CoV-2 Induces Transcriptional Signatures in Human Lung Epithelial Cells That Promote Lung Fibrosis." *Respiratory Research* 21 (1): 182.

Yang, Andrew C., Fabian Kern, Patricia M. Losada, Maayan R. Agam, Christina A. Maat, Georges P. Schmartz, Tobias Fehlmann, et al. 2021. "Dysregulation of Brain and Choroid Plexus Cell Types in Severe COVID-19." *Nature* 595 (7868): 565–71.

Yang, Shiyi, Sean E. Corbett, Yusuke Koga, Zhe Wang, W. Evan Johnson, Masanao Yajima, and Joshua D. Campbell. 2020. "Decontamination of Ambient RNA in Single-Cell RNA-Seq with DecontX." *Genome Biology* 21 (1): 57.

Yin, Xi-Xi, Xiang-Rong Zheng, Wang Peng, Mao-Lan Wu, and Xiao-Yuan Mao. 2020. "Vascular Endothelial Growth Factor (VEGF) as a Vital Target for Brain Inflammation during the COVID-19 Outbreak." *ACS Chemical Neuroscience* 11 (12): 1704–5.

Yoshimatsu, Yasuhiro, and Tetsuro Watabe. 2022. "Emerging Roles of Inflammation-Mediated Endothelial-Mesenchymal Transition in Health and Disease." *Inflammation and Regeneration* 42 (1): 9.

Zheng, Xiaofeng, Sarah Linke, José M. Dias, Xiaowei Zheng, Katarina Gradin, Tristan P. Wallis, Brett R. Hamilton, et al. 2008. "Interaction with Factor Inhibiting HIF-1 Defines an Additional Mode of Cross-Coupling between the Notch and Hypoxia Signaling Pathways." *Proceedings of the National Academy of Sciences of the United States of America* 105 (9): 3368–73.

The COVID Tissue Atlas Consortium Author List

Overall Project Direction and Coordination

Alejandro Granados[1], Franklin W. Huang[2,3], Guo N. Huang[4], Michael G. Kattah[2], Tien Peng[2], Angela Oliveira Pisco[1], Norma Neff[1], Bruce Wang[2]

Writing group

Alejandro Granados[1], Simon Bucher[2], Aditi Agrawal[1], Hanbing Song[3], Ann T. Chen[4], Angela Oliveira Pisco[1], Norma Neff[1], Franklin W. Huang[2,3], Bruce Wang[2]

Organ Processing and Library Preparation

Aditi Agrawal[1], Nancy Allen[2], Benjamin Hyams[2] Simon Bucher[2], Deviana Burhan, Angela Detweiler[1], Shelly Huynh[1], Maurizio Morri[1], Michelle Tan[1], Hannah N.W. Weinstein, Rose Yan[1]

Sequencing

Norma Neff[1], Michelle Tan[1], Angela Detweiler[1], Honey Mekonen[1], Rose (Jia) Yan[1]

Data processing

Aaron McGeever[1], Angela Oliveira Pisco[1], Alejandro Granados[1]

Cell Type Annotation Expert Group

Nancy Allen[2], Xiaoxin Chen[4], Francisco Galdos, Alejandro Granados[1], Guo N. Huang[4], Michael G. Kattah[2], Elvira Mennillo, Abhishek Murti, Poorvi Rao, Iulia Rusu, Hanbing Song[3], Tien Peng[2,] Bruce Wang[2], Jamie Xie

Data Analysis

Alejandro Granados[1], Ann T. Chen[4], Hanbing Song[3], Jonathan Liu[1]

Data release and portal

Alejandro Granados[1], Sharon S. Huang[1], Alexander Tarashansky[1], Angela Oliveira Pisco[1], Kyle Awayan[1]

696 **Affiliations**

697 [1]Chan Zuckerberg Biohub; San Francisco, CA, USA.

698 [2]Department of Medicine and Liver Center, University of California San Francisco; San Francisco,

699 CA, USA.

700 [3]Department of Medicine, San Francisco Veterans Affairs Medical Center, University of California

701 San Francisco; San Francisco, CA, USA

702 [4]Department of Physiology and Cardiovascular Research Institute, University of California San

703 Francisco; San Francisco, CA, USA.

Figure captions

Figure 1

A human single-cell atlas enables the identification of systemic responses to COVID-19.
(**A**) Tissue samples were collected from different organs and frozen, then dissociated into single nuclei. Libraries for snRNA-seq were prepared using 10x Genomics Chromium Next GEM Single Cell 3' v.3.1kit, followed by sequencing on various Illumina platforms. After quality control and clustering, cell types for each organ were annotated by experts using literature gene markers. Differential gene expression and pathway enrichment analysis were performed between COVID-19 and healthy samples for all cell types. Finally, global transcriptional signatures were identified via a cross-organ analysis of differential expression. (**B**) The COVID tissue atlas comprises approximately 85,000 cells from 6 different organs. (**C**) Cells in the COVID tissue atlas cluster by cell identity rather than disease status. (**D**) Number of cells per donor grouped by the organ of origin. (**E**) Number of cells per organ grouped by COVID-19 status.

Figure 2

Cell type composition changes in the COVID lung
(**A**) Integration of the CTA lung with the lung COVID atlas by the Broad Institute. A harmonized UMAP shows that cells from both datasets integrate by their corresponding cell type annotation. (**B**) Integration of two lung COVID atlas colored by the dataset of origin. (**C**) Sub-clustering and UMAP projection of the CTA lung epithelial cells (AT1, AT2, and basal cells). (**D**) Relative cell composition in epithelial lung tissue from control and COVID-19 autopsies (CTA data only). (**E**) Integration of CTA epithelial cells and epithelial cells from (Habermann et al. 2020) (AT1, AT2, basal cells, and transitional ABIs/AT2 populations). ABIs/Transitional AT2 from (Habermann et al. 2020) are shown in red (right). (**F**) Heatmap of scaled gene expression of marker genes for all the different cell populations in E. (**G**) Joint embedding of CTA and (Delorey et al. 2021) (AT1, AT2, basal, and PATS cells). The PATS cells identified by (Delorey et al. 2021) are shown in red in the joint UMAP (right).

Figure 3

Transcriptional changes and dysregulation of cell signaling in the COVID liver
(**A**) UMAP plot showing all cells from liver samples (n = 6 donors) colored by COVID-19 status. Cell type annotations are indicated for each cluster. (**B**) Fraction of cells for each cell type grouped by COVID-19 status. (**C**) Number of differentially expressed genes found using MAST (Finak et al. 2015) (negative binomial model, correcting for the number of detected genes, p < 1e-6 and log2 FC >2). (**D**) The top enriched signaling pathways found for each cell type based on the DE genes shown in C. (**E**) Heatmap of log2 Fold-Change for the top differentially expressed genes. A few relevant genes are highlighted with a text legend.

Figure 4

Transcriptional changes and dysregulation of cell signaling in the COVID heart

(**A**) UMAP plot showing all cells from heart samples (n = 11 donors) colored by COVID-19 status. Cell type annotations are indicated for each cluster. (**B**) Fraction of cells for each cell type grouped by COVID-19 status. (**C**) Number of differentially expressed genes found using MAST (Finak et al. 2015) (negative binomial model, correcting for the number of detected genes, p < 1e-6 and log2 FC >2). (**D**) The top signaling pathways found for each cell type using the genes in C. (**E**) Heatmap of log2 Fold-Change for the top differentially expressed genes. A few relevant genes are highlighted with a text legend.

Figure 5

A shared transcriptional response in macrophages and endothelial cells across organs

(**A**) Overlap of differentially expressed genes in COVID-19 macrophages across organs. The gray shaded area indicates the expected overlap for each organ combination (green circles) under a null hypothesis of random sampling (we computed the p values against this null model). The white bars indicate the number of genes that showed DE in a single organ. The names of the top genes DE in all three organs are shown based on their log2 Fold Enrichment. (**B**) Scatter plot comparing the log2 FC for DE genes in COVID-19 macrophages from lung and heart. (**C**) Same as B but comparing DE genes in COVID-19 macrophages from the liver and heart. (**D**) log2 Fold-Change for COVID-19 endothelial cells from the liver and heart. (**E**) log 2 Fold-Change for COVID-19 stromal cells from the liver and kidney. (**F**) A fully coordinated transcriptional signature would imply that all genes lie in the bottom-left and top-right quadrants (red squares). We define the coordination score as the number of DE genes that show the same direction (up-up, down-down) for the two organs, divided by the number of shared DE genes (A). Gray bars show the score expectation when sampling DE genes randomly from each organ. (**G**) Coordination scores for different cell types across all pairs of organs. The dotted line indicates a significance threshold of z-score > 5 standard deviations compared to the expectation by chance.

Figure 6

Identifying interactions between endothelial cells and macrophages

(**A**) From the shared DE genes across organs, we identified the top enriched signaling pathways for COVID-19 macrophages across the lung, liver, and heart. The value in the heatmap is the log10 p-value for the gene pathway. Only pathways with Fold Enrichment > 3 and adjusted p-value < 1e-3 in at least two organs are shown. (**B**) Enriched pathways in the shared transcriptional response of endothelial cells across lung, liver, and heart (using the same significance thresholds as A). (**C-E**) Enriched expression of ligand-receptor components in COVID-19 macrophages and endothelial cells in the lung (**C**), heart (**D**) and liver (**E**). The x-axis indicates the pair of cell types considered (EC endothelial cells, MA macrophages). The y-axis indicates all the enriched signaling interactions found, and the circles indicate the significance and magnitude of enrichment. We calculated enrichment using CellPhoneDB on the raw sequencing counts. Only ligand-receptor pairs with adjusted p-value < 1e-3 are shown.

Methods

Sample collection

Organs from post-mortem control individuals and patients with COVID-19 were obtained from the University of California, San Francisco Medical Center, and the Saarland University Hospital Institute for Neuropathology, with approval from local ethics committees. Supplementary Table 1 presents all group characteristics.

Tissue processing

During the autopsy, tissue samples were stored in ice-cold Wisconsin solution for transportation, then immediately processed as follows: tissues were rinsed twice with ice-cold PBS, then wiped off. Next, tissues were pre-cut into 1-2 mm3 cubes, flash-frozen in dry ice, and then stored at -80C for single-nuclei extraction and total RNA extraction.

COVID testing

COVID testing was performed on patients according to the testing procedure of host hospitals. For sample testing, total RNA was extracted using a hybrid TRIzol (Life Technologies #15596026) and RNeasy Mini kit (Qiagen #74104) protocol (Wolock, Lopez, and Klein 2019; Rodriguez-Lanetty, Phillips, and Weis 2006). RT- qPCR test for SARS-CoV2 mRNA detection was performed starting from 100 ng of total RNA using a one-Step RT-qPCR enzyme mix (QuantaBio, 94134-500), with primers and probes specific for the SARS-CoV-2 Nucleocapsid N1 and N2 genes, and for human gene ribonuclease PP30 which was used as an internal control (Integrated DNA Technologies, 10006713). The absolute number of transcripts was calculated using a standard curve generated with a positive control for the SARS-CoV2 Nucleocapsid sequence (Integrated DNA Technologies, 10006625).

Nuclei dissociation

The protocol for nuclei isolation was performed in a BSL2+ biosafety cabinet for the lung and in a BSL2 biosafety cabinet for all other organs wearing personal protective equipment (PPE). We carried out all procedures on ice or at 4 °C. Single nuclei were generated from around 50 mg of flash-frozen tissues using the SingulatorTM machine (S2Genomics, Livermore, CA), following the manufacturer's recommendations. The extended protocol was used for the ileum and colon, and the regular protocol was used for all other organs. After isolation, nuclei preparations were cleaned as follows: nuclei were centrifuged at 500 g for 5 min and resuspended in 2 ml of cold Storage Buffer (S2Genomics), then centrifuged again at 500g for 5 min, resuspended in 2 ml of Storage Buffer, and filtered through a 40 µm Flowmi Tip Strainer filter. After centrifugation, nuclei were resuspended in 50 to 500 ul of Storage Buffer supplemented with 1 U/µl of RNAse inhibitor (Sigma Aldrich, cat: 3335402001) and counted using a LUNA-FL™ Dual Fluorescence Cell Counter (Logo Biosystems, Anyang-si, South Korea).

10x Genomics protocol

For droplet-based snRNA-seq, libraries were prepared using the Chromium Next GEM Single Cell 3' v.3.1 according to the manufacturer's protocol (10x Genomics), targeting 10,000 nuclei per sample after counting with a TC20 Automated Cell Counter (Bio-Rad). We performed 12 cycles for cDNA amplification for all of the samples. To generate the final dual or single indexed 10X libraries, 13 cycles were performed.

Library pooling and quality control

After library preparation, individual libraries were quality checked on an Agilent 4200 Tapestation using D5000 screen tape. These libraries were pooled equal molar into a total of 7 pools ranging from 4-15 nM final concentration and quality checked again on an Agilent 4200 Tapestation using a D5000 screen tape, followed by qPCR on a BioRad CFX96 RT PCR thermal cycler using the KAPA library quantification kit (# KK4923).

Sequencing

Individual pools of 10x 3' gene expression libraries were sequenced on Illumina's Nextseq 2000 P3, Novaseq S2 and/or NovaSeq S4 flow cells with a targeted sequencing read depth of 20,000 reads per cell. Sequencing parameters were as follows: 1.) for dual indexed libraries: Read 1= 28 cycles, Index 1= 10 cycles, Index 2= 10 cycles, Read 2= 90 cycles; 2.) for single indexed libraries: Read 1= 28 cycles, Index 1= 8 cycles, Index 2= 0 cycles, Read 2= 91 cycles.

Alignment

Sequences were de-multiplexed using bcl2fastq version 2.20.0.4.22. Reads were aligned to an extended Gencode Reference 30 (GRCh38) genome containing SARS-Cov2 genes (kindly provided by Aviv Regev and Carly Ziegler) using CellRanger version 5.0.1, available from 10x Genomics, with default parameters.

snRNA-seq quality control

The count matrices generated by CellRanger were pre-processed by removing contamination of ambient RNA. We noticed high levels of contamination in single-nuclei data, which has been reported before (S. Yang et al. 2020), and applied Cellbender version 0.1 (Fleming, Marioni, and Babadi 2019) to generate decontaminated count matrices (FDR = 0.01 and default parameters). For quality control, pre-processing, and clustering we used Scanpy (Wolf, Angerer, and Theis 2018). We applied quality control filters directly on the count matrices generated by Cellbender. The minimum number of counts per cell we applied as a cut-off varied depending on the sample and ranged between 300 - 800 counts per cell. We observed high mitochondrial content in some of the samples and filtered out cells that exceeded the cut-off threshold (10-20% depending on the sample). We also applied Scrublet for automated identification of potential doublets (Wolock, Lopez, and Klein 2019).

Data clustering

For each organ, we first integrated the samples from different donors into a harmonized UMAP embedding using scVI (Lopez et al. 2018) release 0.11.0. For training the scVI's variational autoencoder neural network, we used default parameters except for *n_latent=64* and *n_layers=2*. We allowed each gene to have its own variance parameter by setting *dispersion="gene"*. We then used the UMAP algorithm to visualize the resulting embedding in 2 dimensions. All UMAPs for each organ shown in the manuscript were generated in the same way. The UMAPs generated using scVI's latent space showed minimal batch effect and allowed for the identification of cell populations based on known markers for each organ. For each organ, we first verified that individual clusters expressed known gene markers for the expected cell types. Some clusters, however, co-expressed multiple mutually exclusive markers, an indication of ambient RNA contamination, so we labeled these cells as doublets. Clusters that either expressed gene markers for multiple cell types (doublets) or did not express any markers for the cell types expected in the organ (unidentifiable cells) were systematically removed from the dataset. Finally, for each organ, we generated h5ad files with the cell type annotations and the harmonized UMAP.

Cell type annotation

We used the batch corrected UMAPs for cell-type annotation. In brief, tissue experts at either UCSF or Stanford (from research labs focused on specific human tissues) analyzed the expression of cell-type specific markers and assigned identities to the clusters. Confident annotations for some clusters, however, were not possible due to high levels of RNA contamination or low expression of marker genes. We therefore only considered clusters for which a cell type identity was clearly defined. The second round of quality control was applied based on feedback from tissue experts and their annotations. We increased the cut-off values for mitochondrial genes and filtered out putative doublets (cells co-expressing gene markers for mutually exclusive cell types). After the second round of review with the tissue experts, we finalized the cell type annotations for all organs and used them for all downstream analyses. We use the cell type label annotations as ground truth for Differential Expression (DE) analysis, Pathway enrichment, and ligand-receptor enrichment analysis (see Signaling interactions between cell types).

Integration with external datasets

For annotation of cell types in the kidney, we integrated our COVID samples with a single-nuclei atlas of the kidney (Muto et al. 2021). We applied scANVI (C. Xu et al. 2021) for integration and label transfer and confirmed that cell types from COVID donors integrated well with the kidney atlas by inspection of cell-type specific markers (**Figure 1 - Supplement figure 2C-E**). We used the integrated kidney object to compute DE genes and gene pathway enrichment. Additionally, to increase the statistical significance of the identified DE genes, we integrated the COVID and healthy lung single-nuclei samples with the lung data from the Tabula Sapiens dataset (The Tabula Sapiens Consortium and Quake 2021). This integration allowed us to increase the number of healthy cells in endothelial cells and macrophages for which we had not enough large

populations in our healthy single-nuclei samples. We used scVI to integrate samples from the COVID Tissue Atlas and Tabula Sapiens and verified that cell types independently identified on each dataset clustered together in the harmonized embedding.

For the sub-clustering and analysis of lung epithelial cells, we independently integrated the CTA lung samples with the COVID lung atlas published by the Broad Institute (Delorey et al. 2021) and with the lung dataset (Kathiriya et al. 2022). For each data source, we considered only epithelial cells (basal, AT1, and AT2) and performed integration using Seurat 3 (Stuart et al. 2019) (correcting batch effect by donor). We kept the original annotations from each dataset to perform comparisons. Within the integrated dataset, we set the default assay parameter to "RNA" to compute the top ten differentially expressed genes. To investigate the transcriptomic differences and similarities between (Kathiriya et al. 2022) and the CTA dataset, we generated a hierarchical clustering heatmap by down sampling the datasets to 500 cells per population, using the top 20 genes in the signature gene sets developed in the control dataset. Heatmaps were generated using the R package *pheatmap* v1.0.12 with the clustering algorithm set to *ward.D2*.

Differential gene expression

To identify differentially expressed (DE) genes between healthy and COVID samples, we used a negative-binomial model using the *zlm* method as implemented by the MAST R package v1.20 (Finak et al. 2015). Following standard practices in single-cell DE, we corrected for the number of detected genes as a potential confounding variable (Finak et al. 2015). Finally, to correct the p-values for multiple testing, we applied Bonferroni correction and defined significant DE using an adjusted p-value cut-off of 0.05 and a minimum absolute log2 fold-change of 1.

Gene set enrichment analysis

To identify gene sets enriched in COVID donors, we selected the top DE genes for each cell type (COVID vs healthy) and used them as input for pathfindR (Ulgen, Ozisik, and Sezerman 2019), a gene-set enrichment algorithm that includes the fold-change along with potential interactions using a protein-protein interaction network. For selecting significant DE genes, we applied a threshold of log2-FC > *abs*(1) & adjusted p-value < 0.001.

We used 4 different pathway databases as references for our analysis to be comprehensive, KEGG, Reactome, GO, and BioCarta. We then manually curated the enriched pathways, discussed them with tissue experts, and cross-validated them with existing literature to identify the signatures enriched in COVID donors for each cell type and organ. We only considered enriched pathways with a p-value < 0.001.

Coordination in transcriptional responses

To identify transcriptional coordination in COVID samples, we developed a custom analysis method to quantify shared responses across organs. First, we examined the set of genes that appear DE (adjusted p-value <0.001 & log2FC > *abs*(1)) in at least two-thirds of the organs. Some cell types appear in all organs whereas some only appear in two or three. We, therefore,

applied the coordination analysis only for cell types that appear in at least 3 organs (macrophages, fibroblasts, and endothelial cells).

For each cell type, we calculated a custom coordination score, which was defined as follows. For a pair of organs, we took the set of shared DE genes common to both organs and computed the sign of change for each gene in each organ (i.e., positive/negative for up/down-regulation, respectively). For genes that possessed the same sign in both organs, we assigned a value of 1; genes that possessed opposing signs were assigned a value of 0. The coordination score for the pair of organs was then defined as the average value across shared DE genes (i.e., sum of values divided by the number of genes). Thus, a coordination score of 1 indicates that all shared DE genes are jointly up or down-regulated (i.e., perfect coordination), whereas a score of 0 indicates that they are oppositely up or down-regulated (i.e., perfect anti-coordination). For each cell type and for each pair of organs, we thus computed the coordination score.

As a negative control, we repeated this analysis with a computationally shuffled dataset. Here, for each pair of organs for a particular cell type, we held the log2FC values per gene in one organ fixed and randomly shuffled the log2FC values per gene in the second organ. We reasoned that this shuffled dataset should possess near-zero coordination (i.e., a score of 0.5), with some small random deviation due to the finite size of the shared gene list. For each pair of organs, we generated N=1000 computationally shuffled datasets and calculated the resulting coordination scores for each instance, producing a distribution of coordination scores as a negative control. We then averaged the results and retained the mean and standard error, to be compared with the coordination scores from the actual data.

We then used the shared responses as input for pathway enrichment (see above) considering only the cell types that showed significant coordination compared to the random control (macrophages and endothelial cells).

Signaling interactions between cell types

We applied CellPhoneDB (Efremova et al. 2020) and identified significant pairs of ligands and receptors between macrophages and endothelial cells in COVID-19 tissues (adjusted *p*-value < 0.05). We first identified the significant ligand-receptor interactions in healthy and COVID samples independently and considered only those that were enriched in COVID but not in healthy samples.

Data and code availability

Processed and annotated h5ad files for each organ, as used in this study along with links to raw data, are available at the COVID Tissue Atlas portal [https://covid-tissue-atlas.ds.czbiohub.org]. All code used in this study including Jupyter notebooks for pre-processing, analysis, and visualization is available on the COVID tissue atlas GitHub repository czbiohub/CovidTissueAtlas: UCSF Covid Tissue Atlas (github.com).

Supplementary tables

Supplementary Table 1: Patient characteristics

Supplementary Table 2: Differential gene expression COVID vs Healthy across all cell types in the CTA

Supplementary Table 3: Pathway enrichment analysis for all cell types in CTA

Supplementary Table 4: Shared DGE signatures in macrophages and endothelial

Supplementary Table 5: Up-regulated pathways in the shared transcriptional response of Macrophages

Supplementary Table 6: Up-regulated pathways in the shared transcriptional response of Endothelial cells