

# 1 Cell-Free DNA in Blood Reveals Significant Cell, Tissue and Organ 2 Specific injury and Predicts COVID-19 Severity

3  
4 **Authors:** Alexandre Pellan Cheng<sup>1,#</sup>, Matthew Pellan Cheng<sup>2,#</sup>, Wei Gu<sup>3,4,#</sup>, Joan Sasing Lenz<sup>1</sup>,  
5 Elaine Hsu<sup>3</sup>, Erwin Schurr<sup>5</sup>, Guillaume Bourque<sup>5</sup>, Mathieu Bourgey<sup>5</sup>, Jerome Ritz<sup>6,7</sup>, Francisco  
6 Marty<sup>6,8</sup>, Charles Y. Chiu<sup>3,4,9</sup>, Donald Cuong Vinh<sup>2</sup>, Iwijn De Vlaminc<sup>1,\*</sup>

## 7 **Affiliations:**

8 <sup>1</sup> Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA

9 <sup>2</sup> McGill University Health Center, Montreal, Quebec, Canada

10 <sup>3</sup> Department of Laboratory Medicine, University of California, San Francisco, CA, USA

11 <sup>4</sup> UCSF-Abbot Viral Diagnostics and Discovery Center, San Francisco, CA, USA

12 <sup>5</sup> Department of Human Genetics, McGill University, Montreal, Quebec, Canada

13 <sup>6</sup> Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

14 <sup>7</sup> Department of Medicine, Harvard Medical school, Boston, MA, USA

15 <sup>8</sup> Division of Infectious Disease, Brigham and Women's Hospital, Boston, MA, USA

16 <sup>9</sup> Department of Medicine, Division of Infectious Diseases, University of California, San  
17 Francisco, CA, USA

18  
19 # These authors contributed equally

20 \* Corresponding author ([vlaminc@cornell.edu](mailto:vlaminc@cornell.edu))

21  
22  
23  
24  
25  
26 **ABSTRACT:** COVID-19 primarily affects the lungs, but evidence of systemic disease with multi-  
27 organ involvement is emerging. Here, we developed a blood test to broadly quantify cell, tissue,  
28 and organ specific injury due to COVID-19, using genome-wide methylation profiling of circulating  
29 cell-free DNA in plasma. We assessed the utility of this test to identify subjects with severe  
30 disease in two independent, longitudinal cohorts of hospitalized patients. Cell-free DNA profiling  
31 was performed on 104 plasma samples from 33 COVID-19 patients and compared to samples  
32 from patients with other viral infections and healthy controls. We found evidence of injury to the  
33 lung and liver and involvement of red blood cell progenitors associated with severe COVID-19.  
34 The concentration of cfDNA correlated with the WHO ordinal scale for disease progression and  
35 was significantly increased in patients requiring intubation. This study points to the utility of cell-  
36 free DNA as an analyte to monitor and study COVID-19.

37

## 38 INTRODUCTION

39 The Coronavirus Disease-19 (COVID-19) pandemic is a major global health crisis. COVID-19 is  
40 a complex disease with diverse clinical features, ranging from asymptomatic infection to acute  
41 respiratory distress syndrome (ARDS) and multi-organ dysfunction. There is an urgent need for  
42 predictive biomarkers of COVID-19 severity detectable early in disease onset, and improved  
43 understanding of the pathogenesis of COVID-19. Here, we have investigated the utility of  
44 circulating cell-free DNA (cfDNA) in blood as an analyte *i)* to broadly monitor cell, tissue, and  
45 organ injury due to COVID-19, *ii)* to assess disease severity and predict disease outcomes, and  
46 *iii)* to elucidate the multi-organ involvement that characterizes COVID-19.

47 Autopsy studies indicate a broad organotropism for the SARS-CoV-2 virus beyond the lungs [1],  
48 [2]. Detection of the virus in the kidneys, heart, liver, brain and blood of many patients has been  
49 reported [2], [3]. The significant viral burden in the kidney seen in some patients may help explain  
50 the increased risk of acute kidney injury in patients with COVID-19. Damage to endothelial cells  
51 may contribute to COVID-19 coagulopathy and prothrombic state [4]–[8].

52 Initial reports have primarily described COVID-19 as a disease affecting tissues expressing ACE-  
53 2 [9]. However, there are emerging data that SARS-CoV-2 infection may also be accompanied  
54 by hematological derangements [10]–[13]. In addition, a dysregulated immune response to SARS-  
55 CoV-2 can occur, contributing to the development of ARDS, systemic tissue injury, and multi-  
56 organ failure [14]. A strong association between increased cytokine profiles and the severe  
57 deterioration of some patients has been observed [15]. In children, a multisystem inflammatory  
58 syndrome linked to recent SARS-CoV-2 infection is reported [16]. Given the disparate clinical  
59 manifestations and potential complications of COVID-19, there is an urgent need for tests that  
60 can quantify injury to multiple tissues simultaneously to monitor patients, analyze disease  
61 pathogenesis, predict clinical outcomes, and guide clinical management in patients with COVID-  
62 19.

63 Since the advent of cfDNA based noninvasive prenatal testing, myriad applications of cfDNA in  
64 diagnostic medicine have been established [17]–[19]. These short fragments of circulating DNA  
65 are the debris of dead cells from across the body. The value of cfDNA as a quantitative marker of  
66 tissue and organ injury was first recognized in solid-organ transplantation, where the level of  
67 transplant donor derived cfDNA in the blood is now widely used as a marker of transplant rejection  
68 [20]–[22]. More recently, several approaches have been developed to quantify the tissues-of-  
69 origin of cfDNA and thus monitor injury to any cell, tissue or organ type [23]–[27]. This is achieved  
70 by profiling epigenetic marks within cfDNA by quantitative molecular measurement technologies  
71 such as DNA sequencing. Here, we tested the hypothesis that cfDNA tissues-of-origin profiling  
72 enables the identification of specific tissue or cell types that are directly or indirectly targeted and  
73 injured throughout COVID-19 pathogenesis. We studied two independent patient cohorts, and  
74 found evidence of significant injury to the liver, lung, and kidney associated with COVID-19. We  
75 further observed a striking increase, both in terms of proportion and total abundance, of cfDNA  
76 derived from red blood cell precursors when compared to patients infected with other RNA viruses  
77 and healthy controls. Last, the total burden of cfDNA correlated with the WHO ordinal scale for  
78 disease progression, with an increase in cfDNA being strongly associated with admission to the  
79 intensive care unit and need for mechanical ventilation. Thus, cfDNA can provide a marker of  
80 disease severity as well as a prognostic tool that is straightforward to adopt.

81

82

## 83 RESULTS

84 We tested the utility of cfDNA to quantify cell, tissue, and organ specific injury associated with  
85 COVID-19 in two independent patient cohorts from two different hospitals in North America (**Fig.**  
86 **1A, supplementary table 1,2**). We assayed a total of 104 plasma samples from 33 patients  
87 across these cohorts. We performed shotgun DNA sequencing after bisulfite treatment to  
88 determine the tissues-of-origin of cfDNA isolated from all plasma samples by methylation profiling.  
89 We obtained  $62 \pm 35$  million (mean  $\pm$  standard deviation) paired-end reads per sample, leading  
90 to a per-base genome coverage of  $1.3 \pm 0.8$ . We verified that we achieved a high bisulfite  
91 conversion efficiency for all samples ( $0.996 \pm 0.005$ , Methods). To determine the cell, tissue, and  
92 organ types that contribute cfDNA to the mixture in blood (Methods), we analyzed plasma cfDNA  
93 methylation profiles against a reference set of 147 cell, tissue, and organ types using previously  
94 described bioinformatic approaches (**Fig. 1B,C, supplementary data 1, Methods**) [25].

### 95 Temporal dynamics of cell-free DNA tissues-of-origin in plasma of COVID-19 patients

96 We first assayed 52 serial samples collected at short time intervals from five adult patients with  
97 COVID-19 that were treated at University of California, San Francisco (UCSF) Medical Center  
98 (median of 8 samples per patient [range 6-18]). These plasma samples were residual from clinical  
99 testing and were collected from this group of patients over a treatment time-period of up to 14  
100 days with up to four samples collected within 24 hours (median time between consecutive  
101 collections of 13 hours [range 5-64]). These samples allowed us to study dynamic changes in  
102 cfDNA profiles in patients diagnosed with and treated for COVID-19 (**Fig. 2A**). Treatments  
103 included standard of care (n=2), remdesivir (n=1), hydroxychloroquine (n=1), or a combination of  
104 remdesivir, hydroxychloroquine, azithromycin and tocilizumab (n=1). In addition to plasma from  
105 COVID-19 patients, we performed cfDNA tissues-of-origin profiling for six samples collected from  
106 patients with other respiratory viral infection treated at the same hospital, including influenza B  
107 (n=2), Metapneumovirus (n=1), Coronavirus HKU1 (n=1), Coronavirus NL63 (n=1) and  
108 Respiratory syncytial virus B (n=1) (**Fig. 2B**).

109 We plotted the relative abundance of cfDNA derived from different cell, tissue, and organ types  
110 and found that differences in cfDNA profiles between individuals were larger than differences  
111 within individuals over the sampling period. For subjects Z1, Z5, Z6, and Z42 but not Z12, we  
112 observed gradual changes in the tissues-of-origin profiles over sampling periods of six to seven  
113 days. We used the Bray Curtis dissimilarity to quantify the inter and intra-individual differences in  
114 cfDNA profiles (**Fig. 2C-D**). This analysis confirmed the visual appearance of the tissues-of-origin  
115 profiles in Figure 2A and demonstrated that the largest differences in cfDNA were found for  
116 samples collected from different individuals. Within subjects, smaller differences were observed  
117 for samples collected on the same day (**Fig. 2D**). Last, the Bray Curtis dissimilarity increased with  
118 time interval between samples for patients Z1, Z5, Z6, and Z42 but not for Z12. Together these  
119 analyses indicate that cfDNA profiles are subject specific, and that changes in cfDNA tissues-or-  
120 origin profiles occur gradually over days and not hours, therefore adequate longitudinal data can  
121 be collected every few days.

122 We next compared the cfDNA tissues-of-origin profiles associated with COVID-19 versus those  
123 associated with respiratory infection with other viruses (**Fig. 2E, supplementary table 3**). We  
124 found significant increases in the relative proportion of lung specific cfDNA in the blood of COVID-  
125 19 patients, which was likely related to COVID-19 associated tissue injury (2.5% vs 0.6%, p-value  
126 = 0.019, Wilcoxon). We found a similar association with liver-derived cfDNA (5.0% vs 0.9%, p-  
127 value = 0.025, Wilcoxon), and this was validated by the elevated liver function tests in 4 of 5  
128 COVID-19 patients. Strikingly, we also observed an increase in the relative proportion of cfDNA

129 derived from erythroblasts in the blood of COVID-19 patients compared to the control group (75%  
130 vs 17% samples with an erythroblast fraction greater than 0, p-value = 0.003, 2-sample  
131 proportions test, **Fig. 2E, supplementary figure 1**). Erythroblasts are nucleated cells typically in  
132 the adult bone marrow from which red blood cells develop. The increase in cfDNA derived from  
133 red blood progenitor cells seen here may be an indirect consequence of the hypoxemia and/or  
134 cytokine-mediated anemia that characterize severe COVID-19, or may indicate a more direct  
135 involvement of coronavirus with red blood cell precursors. We note that erythroblast cfDNA was  
136 elevated in a single patient in the control group, who was being treated for recurrent stage IV  
137 diffuse large B-cell lymphoma (**Fig. 2 B,E**).

### 138 **Randomized clinical trial cohort**

139 To test the robustness of these initial observations, we assayed an additional 52 samples  
140 collected from 28 patients that were recruited into a randomized control trial at the McGill  
141 University Health Centre in Montreal, Canada. Patients were assigned to either an experimental  
142 antiviral therapy consisting of a combination of Lopinavir and Ritonavir (brand name Kaletra) or  
143 to the standard of care. Of these patients, 14 were treated with the Lopinavir/Ritonavir, and 14  
144 were treated with the standard of care. Of the 28 patients, 21 were discharged after treatment,  
145 one patient remains hospitalized as of July 19<sup>th</sup>, 2020, and six patients died. Serial samples were  
146 collected from these patients at three predetermined timepoints: days 1, 5, and 15 after  
147 enrollment in the clinical trial, provided they remained hospitalized on the days of collection (**Fig.**  
148 **3A**). We determined the relative abundance of tissue-specific cfDNA using the approaches  
149 described above. In addition, we quantified the absolute concentration of tissue-specific cfDNA  
150 by multiplying the proportion of tissue-specific cfDNA with the concentration of total cfDNA  
151 (Methods).

152 We first compared the cfDNA tissues-of-origin profiles measured for these patients with the  
153 tissues-of-origin profiles for four healthy subjects (**Fig. 3B, supplementary figure 2**). We found  
154 that 62% of samples from patients with COVID-19 had a higher concentration of lung cfDNA than  
155 the highest concentration measured for a healthy individual (p-value = 0.017, 2-sample  
156 proportions test). In addition, hospitalized patients with COVID-19 had both an elevated relative  
157 and absolute burden of cfDNA derived from the liver (liver fraction 9.1 vs 1.6%, p-value = 0.054,  
158 and 0.051 ng/ $\mu$ L vs 0.00029 ng/ $\mu$ L, p-value = 0.010, Wilcoxon). In addition to these tissue-specific  
159 features, we again observed a significant increase in cfDNA derived from erythroblast cells for  
160 COVID-19 patients compared to healthy controls (7.7% vs 0%, p-value = 0.027, Wilcoxon; 65%  
161 vs 0% of samples showing erythroblast fraction greater than 0, p-value = 0.0099, 2-sample  
162 proportions test, **Fig. 3B**). We evaluated the temporal dynamics of the contribution of different cell  
163 and tissue types to the mixture in plasma of COVID-19 patients and observed a slow recovery in  
164 tissue injury and a slow increase in the contribution of cfDNA derived from erythroblasts  
165 (**supplementary figure 3**).

166 We then compared cfDNA signatures for COVID-19 patients as function of disease severity, and  
167 found that erythroblast cfDNA proportions at any timepoint are predictive of in-hospital mortality  
168 (19.6% vs 4.1%, p-value = 0.0004, Wilcoxon). Receiver operating characteristic (ROC) analysis  
169 of the performance of the relative proportion of Erythroblast derived DNA to predict COVID-19  
170 mortality yielded an area under the curve (AUC) of 0.83 (95% CI 0.69-0.98, [deceased n = 12;  
171 hospitalized or discharged n = 40]). Additionally, our analysis revealed that kidney cfDNA was  
172 significantly elevated in COVID-19 patients who eventually died (1.8% vs 0.5% vs 0.005%  
173 between deceased, non-deceased and healthy controls, p-value = 0.0018 between deceased and  
174 non-deceased COVID-19 patients).

175 We then compared the cfDNA tissues-of-origin profiles to the WHO clinical progression scale for  
176 COVID-19 [28] (**Fig. 3C**). We found a strong association between the total cfDNA concentrations  
177 isolated from plasma and the WHO clinical progression scores (**Fig. 3C,D**). Notably, a clinical  
178 score of 7 or greater (indicating the need for admission to the intensive care unit and invasive  
179 mechanical ventilation), was associated with a sharp increase in the total burden of cfDNA (**Fig.**  
180 **3C,D**, mean 1.5 ng/ $\mu$ L vs 0.16 ng/ $\mu$ L, between clinical scores from 7 to 9 and 4 to 6, respectively;  
181 p-value =  $1.5 \times 10^{-6}$ , Wilcoxon). ROC analysis of cfDNA concentrations to predict ordinal scores  
182 revealed AUCs of 0.89 (95% CI 0.80-0.99), 0.84 (95% CI 0.72-0.97) and 0.56 (95% CI 0.37-0.76)  
183 for total, erythroblast and lung cfDNA, respectively. Furthermore, samples taken from patients  
184 with a clinical score of 9 (use of extracorporeal membrane oxygenation [ECMO]) had significantly  
185 higher erythroblast-derived cfDNA than patients with a clinical score of 7-8 (1.23 ng/ $\mu$ L vs 0.06  
186 ng/ $\mu$ L, p-value = 0.006, Wilcoxon). Patients on ECMO tend to bleed and require additional blood  
187 volumes, which may contribute to the increased erythroblast signal. However, erythroblast-  
188 derived cfDNA was significantly increased in patients with a clinical score of 7 or higher as well  
189 (**Fig. 3C,D**, mean 0.43 ng/ $\mu$ L vs 0.003 ng/ $\mu$ L, p-value =  $1.83 \times 10^{-5}$ , Wilcoxon).

190 Erythroblast and liver cfDNA contributions correlated with clinical metrics for anemia and liver  
191 damage, respectively (**Fig3. E-G**). We observed significant negative correlations between the  
192 proportion of erythroblast cfDNA and hematocrit and hemoglobin (Pearson's R (R) = -0.51,  
193 Spearman's  $\rho$  ( $\rho$ ) = -0.37 and R = -0.52,  $\rho$  = -0.49, respectively). Similarly, we found positive  
194 correlations between the proportion of liver-derived cfDNA and alanine aminotransferase (ALT)  
195 and aspartate transaminase (R = 0.63,  $\rho$  = 0.47 and R = 0.76,  $\rho$  = 0.24, respectively). We did not  
196 observe a correlation between kidney-derived cfDNA and serum creatinine (R = 0.05,  $\rho$  = 0.09).  
197 We found similar results when comparing the tissue-derived cfDNA concentration to these clinical  
198 markers (erythroblast cfDNA concentration vs hematocrit and hemoglobin: R = -0.42,  $\rho$  = -0.32  
199 and R = -0.38,  $\rho$  = -0.45, respectively. Liver cfDNA concentration vs ALT and AST: R = 0.84,  $\rho$  =  
200 0.52 and R = 0.20,  $\rho$  = 0.23, respectively. Kidney cfDNA concentration vs creatinine: R = 0.56,  $\rho$   
201 = 0.20).

202 Recent papers from Yan *et al.* and Zhou *et al.* identified lactate dehydrogenase (LDH) as a strong  
203 predictor of COVID-19 outcome [29], [30]. LDH is found in virtually all cells and is a commonly  
204 used biomarker for tissue damage and hemolysis [31]–[33]. We found significant correlation  
205 between LDH and the proportion of erythroblast-derived cfDNA (R = 0.64,  $\rho$  = 0.65), and between  
206 LDH and total cfDNA (R = 0.67,  $\rho$  = 0.76). Together, these data suggest that cfDNA tissues-of-  
207 origin can be applied to resolve the specific tissues contributing to non-specific detection of LDH  
208 in blood.

209 Finally, we found no differences between lung, liver, kidney or erythroblast-derived cfDNA for  
210 patients receiving standard of care, or the experimental lopinavir/ritonavir treatment  
211 (**supplementary figure 4**). These data are in line with the results of recent clinical trials that  
212 treatment with lopinavir/ritonavir is not significantly different from standard of care treatment for  
213 COVID-19 [34], [35].

## 214 **DISCUSSION**

215 We find significant support for the utility of cfDNA profiling as a prognostic tool for the early  
216 detection and monitoring of cell and tissue injury associated with COVID-19. A minimally invasive  
217 molecular blood test that can inform cell, tissue and organ specific injury due to COVID-19 has  
218 the potential to alleviate the impact of the COVID crisis **(i)** by providing quantifiable prognostic  
219 parameters and a more granular assessment of clinical severity at the time of presentation; and

220 *ii*) by providing a surrogate biomarker that can be included in clinical trials of candidate COVID-  
221 19 treatments.

222 In line with the diverse clinical manifestations of COVID-19, we find evidence for lung, liver and  
223 kidney injury in hospitalized patients with COVID-19. While lung-derived cfDNA was elevated in  
224 COVID-19 patients, we did not find it to be a major contributor to plasma cfDNA. The level of lung  
225 specific cfDNA in plasma was similar to the levels observed in lung transplant patients that suffer  
226 acute lung transplant rejection [20] and lung cancer patients [36], [37]. We observed a striking  
227 correlation between the total abundance of circulating cfDNA in plasma and the WHO ordinal  
228 scale for disease progression. We propose that the total abundance of cfDNA, which can be  
229 measured within one hour at a low cost, can be used in the context of clinical trials and patient  
230 management in the near term.

231 In addition to the practical application of cfDNA profiling to patient monitoring and COVID-19 risk  
232 stratification, the cfDNA methylation assay and data reported may help elucidate aspects of  
233 COVID-19 pathogenesis. The most significant cfDNA signature observed in the two cohorts  
234 relative to controls was an increase in cfDNA derived from erythroid or red blood progenitor cells.  
235 Given that cfDNA is estimated to have a half-life of about 1 hour [38] and that the proportion of  
236 the erythroid lineage was relatively stable over several days, the elevated erythroid cfDNA is likely  
237 due to a continuous increased erythroid turnover. In support of elevated erythroid turnover and  
238 production, two recent studies have identified red blood cell distribution width (RDW), a measure  
239 of the variation in size of red blood cells (RBCs), as an important prognostic predictor for severe  
240 COVID-19 [15], [16]. The increased RDW was speculated to be associated with increased  
241 turnover of RBCs since increased reticulocytes or newly formed RBCs have a wider diameter  
242 [16]. However, our analysis demonstrated that there was no association with RDW and patient  
243 outcomes (mean 15.4 vs 14.0 between deceased and discharged or hospitalized,  $p$ -value = 0.2,  
244 Wilcoxon) and that erythroblast cfDNA was not strongly correlated with RDW ( $R = 0.26$ ,  $\rho = 0.13$   
245 [with data from UCSF and MUHC]).

246 Increased erythroid turnover may be due to erythroid destruction as the primary driver, followed  
247 by compensatory production, and is supported by anemia (Hgb <13.5 g/dL for men and Hgb < 12  
248 for women) found in 26 of 33 COVID-19 patients across both studies. Possible mechanisms  
249 include: *i*) excessive inflammation and cytokine storm [39], [40], *ii*) hemophagocytosis in relation  
250 to inflammation [41], and *iii*) consumption in microthrombi [6]–[8], [10]. We note that 18 of 33  
251 patients in all studies, C-reactive protein (CRP) was elevated (> 10 mg/L). It is notable however  
252 that megakaryocytes proportions were not increased in either cohort and would not support  
253 microthrombi as the predominant reason for increased erythroid turnover. Alternatively, past work  
254 has shown that angiotensin II regulates normal erythropoiesis and stimulates early erythroid  
255 proliferation through unclear downstream mechanisms [42]–[44]. The binding of SARS-CoV-2 to  
256 the host ACE2 may dysregulate erythropoiesis through the downstream angiotensin II pathway.  
257 The significant increase in cfDNA derived from red blood progenitor cells, may alternatively be  
258 due to injury to red cell precursors [45], through direct or indirect processes. These hypotheses  
259 are further testable through various routes, including comprehensive evaluation of erythrocytosis  
260 in patients with COVID-19, for example through evaluation of circulating reticulocytes and  
261 evaluation of the bone marrow; these measures were not systematically in place during the initial  
262 rapid wave of the pandemic and were not implemented in this study.

263 This study has several limitations. First, we assayed samples from only hospitalized patients, and  
264 we have not evaluated cfDNA profiles for mild COVID-19 cases. Second, while this study spans  
265 two independent cohorts, with patient groups that are genetically and geographically unrelated,  
266 the overall sample size and patient numbers may not be sufficient to generalize our findings to

267 the entire spectrum of COVID-19 cases. Nonetheless, our analysis of cfDNA tissues of origin can  
268 provide immediate insights into the dynamics and pathogenesis of COVID-19. Last, the resolution  
269 of our measurements is limited by the availability of isolated cells and tissue methylation patterns.  
270 Our current reference dataset does not include all known human cell types and tissue types.  
271 Therefore, we are not sensitive to those rarer tissues that may play a role in the pathogenesis of  
272 COVID-19. More comprehensive investigations are therefore needed to confirm and further refine  
273 the observations reported here.

274 In summary, we report the application of cfDNA profiling to quantify cellular and tissue specific  
275 injury due to COVID-19.

## 276 **MATERIALS AND METHODS**

277 **High frequency sampling.** Clinical samples from UCSF were processed through protocols  
278 approved by the UCSF Institutional Review Board (protocol number 10-00476, 18-25287).  
279 Residual plasma was collected as part of routine clinical testing and stored at 4 °C for up to 5  
280 days and subsequently stored at -80 °C until batched extraction. Plasma was initially isolated from  
281 blood by the clinical laboratory after centrifugation at approximately 800g for 10 minutes. After  
282 storage, the plasma was centrifuged at 16,000g for 10 minutes. cfDNA extraction was performed  
283 according to manufacturer recommendations (Qiagen MinElute Circulating Nucleic Acid Kit,  
284 reference #55204 or Qiagen EZ1 Virus Mini Kit v2.0 955134) at 0.4-1 mL plasma input.

285 **Randomized clinical trial.** Individuals diagnosed with COVID-19 were recruited to a randomized,  
286 controlled clinical trial at the McGill University Health Center, where they received either  
287 Lopinavir/ritonavir, or standard-of-care (<https://clinicaltrials.gov/ct2/show/NCT04330690>). Blood  
288 samples were collected under MUHC Research Ethics Board protocol 10-256 through standard  
289 venipuncture in standard blood collection tubes and immediately centrifuged at 850g for 10  
290 minutes. The supernatant is then transferred to new tubes, and centrifuged at 16,000g for 10  
291 minutes. Plasma-containing supernatant is collected and stored in DNA cryostorage vials  
292 (Eppendorf, reference #0030079400) at -80 °C. Plasma was shipped overnight on dry ice from  
293 the McGill University Health Center (Montreal, Canada) to Cornell University (Ithaca, United-  
294 States). Plasma was stored at -80 °C until used for cfDNA extraction. cfDNA extraction was  
295 performed according to manufacturer recommendations (Qiagen Circulating Nucleic Acid Kit,  
296 reference #55114).

297 **Healthy controls.** Volunteers were recruited for blood donations through a protocol approved by  
298 the Cornell Institutional Review Board (protocol number 1910009101). Blood was collected in K2  
299 EDTA tubes (BD, reference #366643) and immediately centrifuged at 1600g for 10 minutes. The  
300 supernatant was transferred to new tubes, and centrifuged at 16,000g for 10 minutes.  
301 Supernatant is then stored in DNA cryostorage vials (Thermo Scientific #363401) at -80 °C until  
302 cfDNA extraction. cfDNA extraction was performed according to manufacturer recommendations  
303 (Qiagen Circulating Nucleic Acid Kit, reference #55114).

304 **Whole genome bisulfite sequencing.** Bisulfite treatment of DNA converts cytosine residues to  
305 uracil but leaves methylated cytosines unaffected [46]. DNA sequencing of bisulfite-treated cfDNA  
306 can be used to reveal methylation patterns with single nucleotide resolution. Because these  
307 patterns are cell, tissue, and organ types specific, they can inform the origins of cfDNA. Following  
308 treatment with bisulfite, whole-genome sequencing (WGS) libraries were prepared according to  
309 manufacturer's protocols (Zymo EZ Methylation-Gold kit, #D5005 and Swift Biosciences Accel-  
310 NGS Methyl-Seq DNA Library Kit #30024) using a dual indexing barcode strategy (Swift  
311

316 biosciences #38096, NEBNext Multiplex Oligos for Illumina E7500L, or custom primers). Paired-  
317 end DNA sequencing was performed on the Illumina NextSeq 500 (2x75bp) at Cornell University  
318 or the Illumina NovaSeq (2x150bp) at University of California San Francisco. Resulting paired-  
319 end fastq files were trimmed to 75bp for downstream analysis.

320  
321 **Human genome alignment.** Adapter sequences were trimmed using BBDUK (BBTools software  
322 suite [47]). Resulting sequences were aligned to the human genome (version hg19) and  
323 deduplicated using Bismark [48]. Alignment files were filtered with a minimum mapping quality of  
324 10 using SAMtools [49].

325  
326 **Reference methylomes and tissues of origin.** Reference methylation profiles were obtained  
327 from publicly available datasets and international epigenetic consortium projects (**supplementary**  
328 **data 1**) and processed as previously described [25]. Briefly, files were downloaded and  
329 normalized to a standard 4 column BED format at single nucleotide resolution using hg19  
330 coordinates. Differentially methylated regions (DMRs) were found using Metilene [50].  
331 Methylation densities within these DMRs were averaged. Tissues with methylation profiles highly  
332 dissimilar from the same tissues were removed. cfDNA methylation densities were extracted  
333 using Bismark [48] and averaged over the DMRs. Tissues of origin were deconvoluted using a  
334 non-negative least squares approach.

335  
336 **cfDNA concentration measurement - MUHC patients.** Plasma samples were processed in  
337 batches of 4 to 10 alongside a control containing 8  $\mu$ L of approximately 150 ng/ $\mu$ L of synthetic  
338 oligos. DNA concentration measurements were performed after cfDNA extraction (Qubit  
339 Fluorometer 3.0) and the normalized concentration was calculated by multiplying the sample's  
340 concentration by the input/output ratio of the control.

341  
342 **Depth of coverage.** The depth of DNA sequencing coverage was calculated by dividing the  
343 number of mapped nucleotides to the autosomal chromosomes to the size of the non-N hg19  
344 autosomal genome.

345  
346 **Bisulfite conversion efficiency.** The bisulfite conversion efficiency achieved in experiments was  
347 estimated using MethPipe [51] by calculating the reported methylation density of cytosines  
348 present at C[A/T/G] dinucleotides, which are rarely methylated in mammalian genomes.

349  
350 **Quality control filtering.** Samples from the high frequency sampling cohort were selected for  
351 analysis if 10 or more spike-in molecules were identified after sequencing and were also filtered  
352 for sufficient depth of sequencing ( $>0.2x$  human genome). Samples from the randomized control  
353 trial cohort were sequenced to a minimum depth of  $0.7x$  human genome coverage. All samples  
354 had a minimum bisulfite conversion efficiency of 96%.

355  
356 **Statistical analysis.** All statistical analyses were performed in R, version 3.5.0. Groups were  
357 compared using the two-sided, nonparametric Wilcoxon test. If the data distributions were zero-  
358 skewed, a two-sided, 2-sample proportions test without continuity correction was performed.  
359 Boxplots span from the 25th and 75th percentiles. The band in the box indicates the median,  
360 lower and higher whiskers extend to the smallest and largest values at most  $1.5 \times$  IQR of the  
361 hinge, respectively.

362  
363 **Data availability.**  
364 Genomic data will be hosted on the Sequence Read Archive. The code used to generate figures  
365 and analyze primary data is available at [www.github.com/alexpcheng/cfDNAme](https://www.github.com/alexpcheng/cfDNAme).



366 **CONFLICTS OF INTEREST**

367 A.P.C., M.P.C., W.G., C.Y.C., D.C.V. and I.D.V. are inventors on a patent application submitted  
368 by Cornell University Center for Technology Licensing.

369

370 **AUTHOR CONTRIBUTIONS**

371 A.P.C., M.P.C., W.G., C.Y.C., D.C.V. and I.D.V. designed the study. M.P.C., W.G., C.Y.C., D.C.V.  
372 consented patients and obtained clinical data. J.S.L., E.H, W.G. performed experiments. A.P.C.,  
373 M.P.C., W.G. and I.D.V. analyzed data. A.P.C., W.G., M.P.C., D.V. and I.D.V. wrote manuscript.  
374 All authors provided edits and comments.

375

376

377 **ACKNOWLEDGEMENTS**

378 We thank Dr. Peter Schweitzer and colleagues at the Cornell Genomics Center for help with  
379 sequencing assays. We thank Dr. Lynn Johnson and the Cornell Statistical Consulting Unit for  
380 help with statistical analysis. We thank Dr. Lucie Roussel, Dr. Marianna Orlova and Pauline  
381 Cassart for technical assistance. This work was supported by NIH Grant 1DP2AI138242 (to  
382 I.D.V.), R01AI146165 (to I.D.V. and M.P.C), 1R01AI151059 (to I.D.V.), K08-CA230156 (to W.G),  
383 R33-AI129455 to C.Y.C, a Synergy award from the Rainin Foundation (to I.D.V.), a SARS-CoV-  
384 2 seed grant at Cornell (to I.D.V.), a National Sciences and Engineering Research Council of  
385 Canada fellowship PGS-D3 (to A.P.C.), and Burroughs-Wellcome CAMS Award (to W.G.). D.C.V.  
386 is supported by a Fonds de la recherche en sante du Quebec Clinical Research Scholar Junior 2  
387 award. C.Y.C. is supported by the California Initiative to Advance Precision Medicine, and the  
388 Charles and Helen Schwab Foundation (C.Y.C.).

389

390

## REFERENCES

- [1] S. E. Fox, A. Akmatbekov, J. L. Harbert, G. Li, J. Quincy Brown, and R. S. Vander Heide, "Pulmonary and cardiac pathology in African American patients with COVID-19: an autopsy series from New Orleans," *Lancet Respir. Med.*, Jul. 2020.
- [2] V. G. Puelles *et al.*, "Multiorgan and Renal Tropism of SARS-CoV-2," *N. Engl. J. Med.*, vol. 0, no. 0, p. null, 2020.
- [3] A. Gupta *et al.*, "Extrapulmonary manifestations of COVID-19," *Nat. Med.*, vol. 26, no. 7, pp. 1017–1032, Jul. 2020.
- [4] Z. Varga *et al.*, "Endothelial cell infection and endotheliitis in COVID-19," *Lancet*, vol. 395, no. 10234, pp. 1417–1418, May 2020.
- [5] P. Julien *et al.*, "Pulmonary Embolism in COVID-19 Patients: Awareness of an Increased Prevalence," *Circulation*, vol. 0, no. 0, May 2020.
- [6] G. Goshua *et al.*, "Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study," *Lancet Haematol.*, Jul. 2020.
- [7] F. A. Klok *et al.*, "Incidence of thrombotic complications in critically ill ICU patients with COVID-19," *Thromb. Res.*, vol. 191, pp. 145–147, 2020.
- [8] M. Kashi *et al.*, "Severe arterial thrombosis associated with Covid-19 infection," *Thromb. Res.*, vol. 192, pp. 75–77, Aug. 2020.
- [9] W. Sungnak *et al.*, "SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes," *Nat. Med.*, vol. 26, no. 5, pp. 681–687, May 2020.
- [10] Y. Zhang *et al.*, "Mechanisms involved in the development of thrombocytopenia in patients with COVID-19," *Thromb. Res.*, vol. 193, pp. 110–115, Jun. 2020.
- [11] A. Mitra *et al.*, "Leukoerythroblastic reaction in a patient with COVID-19 infection," *Am. J. Hematol.*, vol. n/a, no. n/a, Mar. 2020.
- [12] B. H. Foy *et al.*, "Elevated RDW is Associated with Increased Mortality Risk in COVID-19," *medRxiv*, p. 2020.05.05.20091702, May 2020.
- [13] J. Gong *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China," *Clin. Infect. Dis.*, p. ciae443, Apr. 2020.
- [14] L. Pirofski and A. Casadevall, "Pathogenesis of COVID-19 from the Perspective of the Damage-Response Framework," *MBio*, vol. 11, no. 4, pp. e01175-20, Aug. 2020.
- [15] Q. Ye, B. Wang, and J. Mao, "The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19," *J. Infect.*, vol. 80, no. 6, pp. 607–613, Jun. 2020.
- [16] E. M. Dufort *et al.*, "Multisystem Inflammatory Syndrome in Children in New York State," *N. Engl. J. Med.*, Jun. 2020.
- [17] P. Burnham, K. Khush, and I. De Vlaminc, "Myriad applications of circulating cell-free DNA in precision organ transplant monitoring," *Ann. Am. Thorac. Soc.*, vol. 14, 2017.
- [18] H. C. Fan, Y. J. Blumenfeld, U. Chitkara, L. Hudgins, and S. R. Quake, "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 16266–16271, 2008.
- [19] R. W. Chiu *et al.*, "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma," *Proc Natl Acad Sci USA*, vol. 105, 2008.
- [20] I. De Vlaminc *et al.*, "Noninvasive monitoring of infection and rejection after lung transplantation," *Proc. Natl. Acad. Sci.*, vol. 112, no. 43, pp. 13336–13341, Oct. 2015.
- [21] I. De Vlaminc *et al.*, "Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection," *Sci. Transl. Med.*, vol. 6, no. 241, 2014.
- [22] T. M. Snyder, K. K. Khush, H. A. Valentine, and S. R. Quake, "Universal noninvasive

- detection of solid organ transplant rejection.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, pp. 6229–6234, 2011.
- [23] K. Sun *et al.*, "Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 40, pp. E5503-12, Oct. 2015.
- [24] M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, and J. Shendure, "Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin," *Cell*, vol. 164, no. 1, pp. 57–68, Jan. 2016.
- [25] A. P. Cheng *et al.*, "A cell-free DNA metagenomic sequencing assay that integrates the host injury response to infection," *Proc. Natl. Acad. Sci.*, vol. 116, no. 37, p. 18738 LP-18744, Sep. 2019.
- [26] R. Lehmann-Werman *et al.*, "Identification of tissue-specific cell death using methylation patterns of circulating DNA," *Proc. Natl. Acad. Sci.*, vol. 113, no. 13, pp. E1826–E1834, Mar. 2016.
- [27] A. P. Cheng *et al.*, "Cell-free DNA Tissues-of-Origin Profiling to Predict Graft versus Host Disease and Detect Infection after Hematopoietic Cell Transplantation," *bioRxiv*, p. 2020.04.25.061580, Jan. 2020.
- [28] J. C. Marshall *et al.*, "A minimal common outcome measure set for COVID-19 clinical research," *The Lancet Infectious Diseases*, vol. 0, no. 0. Lancet Publishing Group, 2020.
- [29] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020.
- [30] L. Yan *et al.*, "An interpretable mortality prediction model for COVID-19 patients," *Nat. Mach. Intell.*, vol. 2, no. 5, pp. 283–288, 2020.
- [31] G. J. Kato *et al.*, "Lactate dehydrogenase as a biomarker of hemolysis-associated nitric oxide resistance, priapism, leg ulceration, pulmonary hypertension, and death in patients with sickle cell disease," *Blood*, vol. 107, no. 6, pp. 2279–2285, Mar. 2006.
- [32] R. F. Henderson, E. G. Damon, and T. R. Henderson, "Early damage indicators in the lung I. Lactate dehydrogenase activity in the airways," *Toxicol. Appl. Pharmacol.*, vol. 44, no. 2, pp. 291–297, May 1978.
- [33] E. Fernandez-Cruz, P. Escartin, A. Bootello, M. Kreisler, and J. M. Segovia de Arana, "Hepatocyte damage induced by lymphocytes from patients with chronic liver diseases, as detected by LDH release.," *Clin. Exp. Immunol.*, vol. 31, no. 3, pp. 436–42, Mar. 1978.
- [34] "WHO discontinues hydroxychloroquine and lopinavir/ritonavir treatment arms for COVID-19." [Online]. Available: <https://www.who.int/news-room/detail/04-07-2020-who-discontinues-hydroxychloroquine-and-lopinavir-ritonavir-treatment-arms-for-covid-19>. [Accessed: 20-Jul-2020].
- [35] "No clinical benefit from use of lopinavir-ritonavir in hospitalised COVID-19 patients studied in RECOVERY — RECOVERY Trial." [Online]. Available: <https://www.recoverytrial.net/news/no-clinical-benefit-from-use-of-lopinavir-ritonavir-in-hospitalised-covid-19-patients-studied-in-recovery>. [Accessed: 20-Jul-2020].
- [36] J. Moss *et al.*, "Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease," *Nat. Commun.*, vol. 9, no. 1, pp. 1–12, Dec. 2018.
- [37] K. Sun *et al.*, "Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin," *Genome Res.*, vol. 29, no. 3, pp. 418–427, Mar. 2019.
- [38] S. C. Y. Yu *et al.*, "High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing," *Clin. Chem.*, vol. 59, no. 8, pp. 1228–1237, Aug. 2013.
- [39] N. B. A. Roy and C. Babbs, "The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I," *Br. J. Haematol.*, vol. 185, no. 3, pp. 436–449, May

- 2019.
- [40] M. P. Rodero *et al.*, “Type I interferon-mediated autoinflammation due to DNase II deficiency,” *Nat. Commun.*, vol. 8, no. 1, p. 2176, 2017.
  - [41] D. A. Dorward *et al.*, “Tissue-specific tolerance in fatal Covid-19,” *medRxiv*, p. 2020.07.02.20145003, Jul. 2020.
  - [42] M. Mrug, T. Stopka, B. A. Julian, J. F. Prchal, and J. T. Prchal, “Angiotensin II stimulates proliferation of normal early erythroid progenitors,” *J. Clin. Invest.*, vol. 100, no. 9, pp. 2310–2314, Nov. 1997.
  - [43] Y. C. Kim, O. Mungunsukh, E. A. McCart, P. J. Roehrich, D. K. Yee, and R. M. Day, “Mechanism of erythropoietin regulation by angiotensin II,” *Mol. Pharmacol.*, vol. 85, no. 6, pp. 898–908, 2014.
  - [44] D. V. Vlahakos, K. P. Marathias, and N. E. Madias, “The role of the renin-angiotensin system in the regulation of erythropoiesis,” *American Journal of Kidney Diseases*, vol. 56, no. 3. Am J Kidney Dis, pp. 558–565, Sep-2010.
  - [45] A. Cavezzi, E. Troiani, and S. Corrao, “COVID-19: hemoglobin, iron, and hypoxia beyond inflammation. A narrative review.,” *Clin. Pract.*, vol. 10, no. 2, p. 1271, May 2020.
  - [46] M. Frommer *et al.*, “A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 5, pp. 1827–1831, Mar. 1992.
  - [47] B. Bushnell, J. Rood, and E. Singer, “BBMerge – Accurate paired shotgun read merging via overlap,” *PLoS One*, vol. 12, no. 10, p. e0185056, Oct. 2017.
  - [48] F. Krueger and S. R. Andrews, “Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications,” *Bioinforma. Appl. NOTE*, vol. 27, no. 11, pp. 1571–1572, 2011.
  - [49] H. Li *et al.*, “The Sequence Alignment/Map format and SAMtools,” *Bioinforma. Appl. NOTE*, vol. 25, no. 16, pp. 2078–2079, 2009.
  - [50] F. Jühling, H. Kretzmer, S. H. Bernhart, C. Otto, P. F. Stadler, and S. Hoffmann, “Metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data,” *Genome Res.*, vol. 26, no. 2, pp. 256–262, Feb. 2016.
  - [51] Q. Song *et al.*, “A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics,” *PLoS One*, vol. 8, no. 12, p. e81148, Dec. 2013.

## FIGURE CAPTIONS

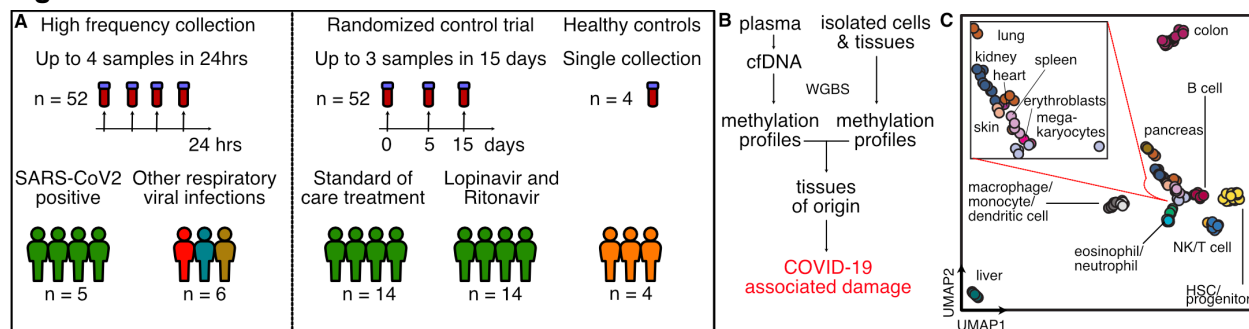
**Figure 1.** Study design. **A)** Two independent cohorts were used in our study: First, a high frequency collection cohort with 5 SARS-CoV-2 patients (n = 52 samples) and 6 SARS-CoV-2 negative, RNA-virus positive patients (n = 6 samples). Second, a randomized control trial of 28 SARS-CoV-2 patients with plasma at serial time points (n = 52 samples). 4 healthy individuals volunteered plasma for cell-free DNA analysis. **B)** Experimental workflow. cfDNA is extracted from plasma and whole-genome bisulfite sequencing is performed. In parallel, methylation profiles of cell and tissue genomes are obtained from publicly-available databases. cfDNA methylation profiles are compared to those of cell and tissue references to infer relative contributions of tissues to the cfDNA mixtures. **C)** UMAP of differentially methylated regions for isolated cell and tissue types used as a reference.

**Figure 2.** High frequency sample collection cohort at UCSF. **A-B)** Patient-specific relative tissue contributions for SARS-CoV-2 patients (**A**) and other RNA-virus infection patients (**B**). Triangles indicate sampling times. **C)** Heatmaps of Bray-Curtis dissimilarity. **D)** Scatterplot of patient-specific Bray-Curtis dissimilarity (left) and boxplot of Bray-Curtis dissimilarity between cfDNA tissue proportions from samples collected from either the same day (within 24 hours), the same person (but not within 24 hours), or from all patients (right). **E)** Comparison of tissue fraction of four cell and tissue types (neutrophil, erythroblast, lung and liver) between SARS-CoV-2 positive patients and other RNA-virus positive patients. \* : p-value < 0.05; \*\* : p-value < 0.01; \*\*\* : p-value < 0.001 (p-values calculated using a Wilcoxon test)

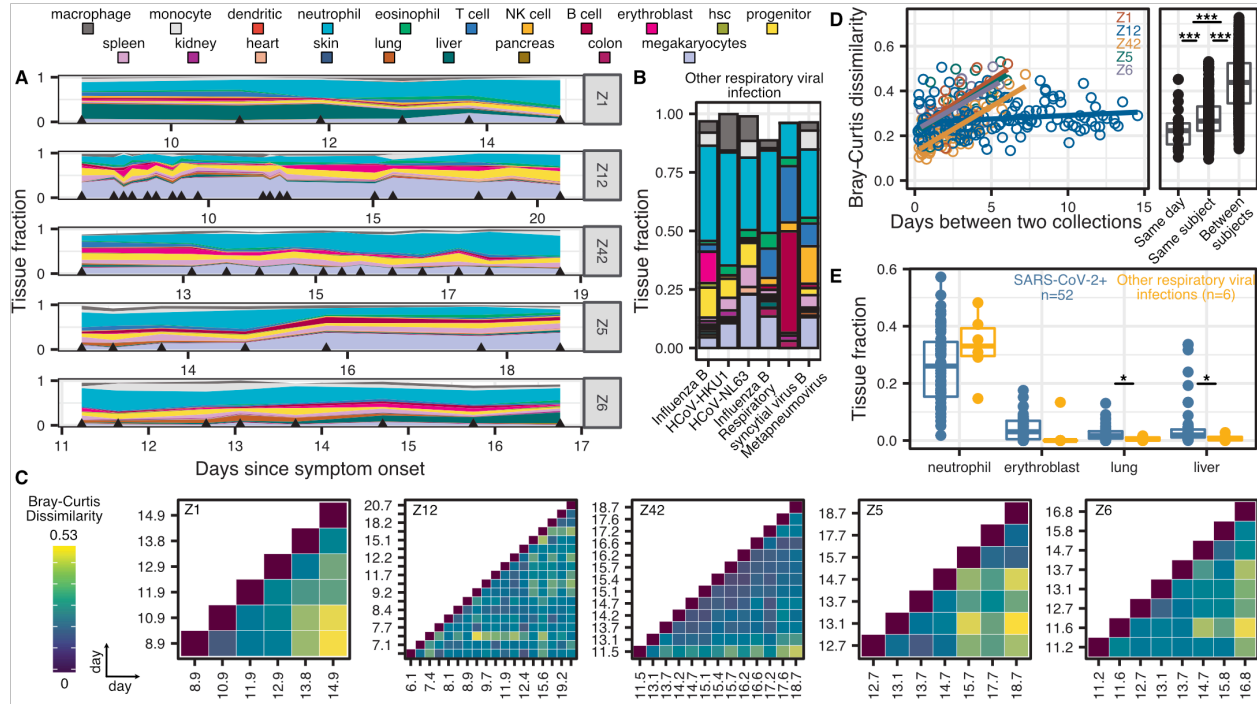
**Figure 3.** Randomized control trial cohort from MUHC. **A)** Patient sample-collection map by day of enrollment into the study. **B)** Relative proportion of cfDNA derived from four cell and tissue types (neutrophil, erythroblast, lung, liver) by hospitalization status (p-values calculated using a Wilcoxon test). **C)** Absolute cfDNA concentrations compared to the WHO ordinal scale for COVID progression. Blue shading indicates ordinal scores requiring admittance to the intensive care unit (ICU) **D)** Receiver operating characteristic analysis of the performance of absolute cfDNA concentration of different tissues (lung, erythroblast and total) in distinguishing patients presenting with ordinal scales from 4-6 (hospitalized) and 7-9 (hospitalized in the ICU). **E-G)** Scatterplot comparisons between relative proportions of erythroblast cfDNA fraction and hemoglobin (**E**), liver cfDNA fraction and alanine aminotransferase (ALT) (**F**) and total cfDNA concentration and lactate dehydrogenase (LDH) (**H**). Green shading indicates normal levels. \* : p-value < 0.05; \*\* : p-value < 0.01; \*\*\* : p-value < 0.001.

## FIGURES

Figure 1



**Figure 2**



**Figure 3**

