

1 **Multi-omics identify LRRC15 as a COVID-19 severity predictor and persistent pro-
2 thrombotic signals in convalescence**

3
4

5 Jack S. Gisby^{*1}, Norzawani B. Buang^{*1}, Artemis Papadaki¹, Candice L. Clarke^{1,2}, Talat
6 H. Malik¹, Nicholas Medjeral-Thomas^{1,2}, Damiola Pinheiro¹, Paige M. Mortimer¹,
7 Shanice Lewis¹, Eleanor Sandhu^{1,2}, Stephen P. McAdoo^{1,2}, Maria F. Prendecki^{1,2},
8 Michelle Willicombe^{1,2}, Matthew C. Pickering¹, Marina Botto¹, David C. Thomas^{†1,2},
9 James E. Peters^{†1}

10

11 *equal contributions

12 †jointly supervised the work

13

14

15 **Author Affiliations**

16 1) Centre for Inflammatory Disease, Dept of Immunology and Inflammation, Imperial College
17 London.

18 2) Renal and Transplant Centre, Hammersmith Hospital, Imperial College Healthcare NHS
19 Trust, London, United Kingdom.

20

21 Correspondence to: James E. Peters or David C. Thomas. Email: j.peters@imperial.ac.uk or
22 david.thomas1@imperial.ac.uk

23

24

25

26 **Abstract**

27 Patients with end-stage kidney disease (ESKD) are at high risk of severe COVID-19. Here,
28 we performed longitudinal blood sampling of ESKD haemodialysis patients with COVID-19,
29 collecting samples pre-infection, serially during infection, and after clinical recovery. Using
30 plasma proteomics, and RNA-sequencing and flow cytometry of immune cells, we identified
31 transcriptomic and proteomic signatures of COVID-19 severity, and found distinct temporal
32 molecular profiles in patients with severe disease. Supervised learning revealed that the
33 plasma proteome was a superior indicator of clinical severity than the PBMC transcriptome.
34 We showed that both the levels and trajectory of plasma LRRC15, a proposed co-receptor for
35 SARS-CoV-2, are the strongest predictors of clinical outcome. Strikingly, we observed that
36 two months after the acute infection, patients still display dysregulated gene expression
37 related to vascular, platelet and coagulation pathways, including *PF4* (platelet factor 4), which
38 may explain the prolonged thrombotic risk following COVID-19.

39

40

41 **Introduction**

42

43 COVID-19, caused by the SARS-CoV-2 virus, is a highly heterogenous disease. In most
44 individuals, it is a mild, self-limiting illness, but some individuals develop severe disease,
45 typically manifesting as respiratory failure with marked systemic inflammation and
46 immunopathology. Multiple studies have described immunological [1,2], transcriptomic [3–7],
47 and proteomic [8–16] correlates of severe disease. The importance of an aberrant host
48 immune response in tissue injury in severe COVID-19 is supported by the efficacy of anti-
49 inflammatory treatments. These include glucocorticoids [17], monoclonal antibodies blocking
50 the interleukin-6 receptor [18,19], and the Janus kinase (JAK) inhibitor baricitinib [20]. A wide
51 range of additional therapies directed at specific elements of the inflammatory response has
52 been developed for immuno-inflammatory diseases and present potential repurposing
53 opportunities for the treatment of severe COVID-19. Understanding the molecular basis for
54 severe COVID-19 is critical for the rational selection of such therapies.

55

56 Risk factors for severe COVID-19 include age, male sex, and the presence of comorbidities
57 such as chronic kidney disease (CKD). In CKD, the risk of severe COVID-19 is proportional to
58 the degree of renal impairment [21]. End-stage kidney disease (ESKD) confers particularly
59 high risk, with a population-based study estimating a hazards ratio for death of 3.69 [21] and
60 a European registry study reporting 23.9% 28-day mortality in dialysis patients with COVID-
61 19 [22]. In part, this is because ESKD patients are enriched for other risk factors for severe
62 COVID-19, including cardiometabolic disease. However, even after adjustment for these,

63 ESKD remains independently associated with the risk of severe COVID-19. In addition, ESKD
64 patients display impaired vaccine responses [23,24], and those on haemodialysis cannot
65 shield effectively during lockdowns as they need to access dialysis facilities regularly.

66

67 Here, we investigated the host response to SARS-CoV-2 in ESKD patients on haemodialysis
68 since study of such an at-risk group should enhance the probability of identifying severity
69 signals and might also point to either an exaggerated or even distinct immunological response
70 to the virus. Moreover, ESKD patients receiving haemodialysis present a unique opportunity
71 for serial blood sampling of both outpatients and inpatients with COVID-19, since patients
72 must attend medical facilities for regular dialysis regardless. This enabled us to perform
73 longitudinal analysis and avoid the selection bias that affects studies limited solely to
74 hospitalised patients.

75

76 The host response to SARS-CoV-2 is orchestrated by a complex network of cells and
77 mediators, including circulating proteins such as cytokines and soluble receptors. Soluble
78 proteins play key roles in multiple biological processes, including signaling, host defence and
79 repair, and are potential biomarkers and therapeutic targets. We therefore hypothesised that
80 a comprehensive analysis of both circulating proteins and immune cells should yield valuable
81 and complementary insights into the pathobiology of COVID-19. To this end, we used the
82 aptamer-based SomaScan platform that provides the broadest available coverage of the
83 plasma proteome (6,323 proteins), combined with RNA-sequencing and flow cytometry of
84 peripheral blood mononuclear cells (PBMCs). We integrated these data to provide a
85 comprehensive view of the COVID-19 multi-omic landscape, enabling us to link transcriptomic
86 and cellular changes with circulating proteins. Supervised learning identified plasma levels of
87 the LRRC15 protein, a recently proposed alternative receptor for SARS-CoV-2, as a key
88 marker of disease severity. Uniquely, by comparing pre-infection samples to samples
89 collected from the same individuals during COVID-19 and after clinical recovery, we revealed
90 persistent upregulation of gene expression signatures related to vascular and clotting
91 pathways several months after infection. These findings elucidate the biological underpinnings
92 of the prolonged pro-thrombotic state associated with COVID-19.

93

94 **Results**

95

96 Features of patient cohorts

97

98 We recruited two cohorts of ESKD patients on haemodialysis presenting with COVID-19
99 (**Figure 1A**). The Wave 1 cohort consisted of 53 patients recruited during the initial phase of

100 the COVID-19 pandemic (April-May 2020) (**Supplementary Table 1**). Serial blood sampling
101 was carried out where feasible (**Figure 1B**), given the pressure on hospital services and the
102 effects of national lockdown. We assessed disease severity using a WHO four-level ordinal
103 score, categorising it into mild, moderate, severe, and critical. Of the 53 patients, 25 had a
104 peak illness severity score of severe or critical (hereafter ‘severe/critical’) and 28 mild or
105 moderate (‘mild/moderate’). Nine died. The majority of patients were of non-European
106 ancestry. Further clinical and demographic details are provided in **Supplementary Table 1**.
107 We also contemporaneously recruited 59 non-infected haemodialysis patients to provide a
108 control group, selected to mirror the age, sex and ethnicity distribution of the COVID-19 cases
109 (**Supplementary Figure 1A-C**).

110

111 The Wave 2 cohort consisted of 17 ESKD patients with COVID-19 infected during the
112 resurgence of cases in January-March 2021 (**Supplementary Table 2**). All had been recruited
113 as part of the COVID-19 negative control group during Wave 1, thereby providing a pre-
114 infection sample collected 8-9 months earlier. For the Wave 2 cohort, we systematically
115 acquired serial samples for all patients at regular intervals (every 2-3 days over the course of
116 the acute illness) (**Figure 1C**). 9 patients had a peak illness severity of severe/critical (of whom
117 4 died), and 8 mild/moderate. For 12 of these patients, we acquired convalescent samples
118 approximately two months following infection.

119

120 The effect of COVID-19 on the PBMC transcriptome and plasma proteome in ESKD patients

121

122 We performed transcriptomic profiling using RNA-seq of PBMCs. Principal components
123 analysis (PCA) revealed a clear effect of COVID-19 in both Wave 1 (COVID-19 positive and
124 negative patient samples) and Wave 2 (pre-infection and subsequent COVID-19 positive
125 samples from the same individuals) (**Figure 2A**). In the Wave 1 cohort, differential gene
126 expression analysis between COVID-19 positive (n=179 samples from 51 patients) and
127 negative samples (n=55) using linear mixed models (LMM) identified 3,026 significantly up-
128 regulated and 3,329 down-regulated genes (1% false discovery rate, FDR) (**Supplementary**
129 **File 1A**). For the Wave 2 cohort, where we compared COVID-19 positive samples (n=90
130 samples from 17 individuals) with pre-infection samples from these same individuals, we
131 identified 2,871 up-regulated and 3,325 down-regulated genes (1% FDR, LMM)
132 (**Supplementary File 1A**). These findings demonstrate widespread transcriptomic changes
133 associated with COVID-19. The effect sizes for the differentially expressed genes between
134 the Wave 1 and 2 cohorts were highly concordant (Pearson’s r 0.80) (**Supplementary Figure**
135 **2A**), despite differences in the prevalent SARS-CoV-2 variant and developments in medical
136 management (8 of 17 patients in the Wave 2 cohort received glucocorticoids). To identify the

137 genes that were consistently differentially expressed across both cohorts, we used robust rank
138 aggregation (RRA) (**Supplementary File 1A, Supplementary Figure 3**).
139

140 To gain insight into the biological pathways underlying these changes, we used Gene Set
141 Variation Analysis (GSVA) [25] to compare COVID-19 positive and negative ESKD samples
142 (**Supplementary File 1B**). Enriched pathways included those related to cell cycle (e.g. “Polo-
143 like kinase mediated events”, which are involved in the cellular response to DNA damage) and
144 host defence (e.g. “Complement cascade”, “Fc-gamma receptor-dependent phagocytosis”,
145 and “Parasite infection”) (**Supplementary Figure 4**). This analysis also highlighted leukocyte-
146 endothelial interactions (“Cell surface interactions at the vascular wall”, which included *SELL*
147 and *CEACAM-1, -3, -6 and -8* genes). Examples of marked changes in gene expression
148 between the pre-infection and first acute infection sample in the Wave 2 cohort included
149 components of “Immunoregulatory interactions between a lymphoid and a non-lymphoid cell”
150 pathway term (e.g. *SIGLEC1, SIGLEC9, SELL*, all increased) and “Development and
151 heterogeneity of the ILC family” (e.g. *IFNG, GATA3, RORA*, all decreased) (**Figure 2B**).
152

153 We next assessed the circulating proteome, measuring 6,323 proteins using the SomaScan
154 platform (**Supplementary File 1C**). PCA showed clear differences between COVID-19
155 positive and negative samples (**Figure 2A**). We identified 1,273 differentially abundant
156 proteins between COVID-19 positive and negative samples in Wave 1 (86 samples from 37
157 COVID-19 positive ESKD patients versus 53 non-infected ESKD patient samples, LMM)
158 (**Supplementary File 1D, Supplementary Figure 5**). In Wave 2, comparison of COVID-19
159 positive samples (n=102 samples from 17 patients) with pre-infection samples from the same
160 individuals identified 5,265 differentially abundant proteins. The effect sizes were generally
161 concordant between the cohorts (Pearson’s r 0.57) (**Supplementary Figure 2B**). As for our
162 transcriptomic analysis, we used RRA to identify the differentially abundant proteins consistent
163 across both cohorts (**Supplementary File 1D**).
164

165 Enrichment analysis revealed upregulation of pathways, including “DDX58 IFIH1 mediated
166 induction of interferon-alpha/beta”, “Wilk *et al.*, 2021 IFN module” [26], “Host-pathogen
167 interaction of human coronaviruses interferon induction” and “SARS-CoV-2 innate immunity
168 evasion and cell-specific immune response”, reflecting host anti-viral responses and providing
169 validation of our analysis (**Figure 2C, Supplementary File 1E**). Highly up-regulated proteins
170 within these pathways included STAT1; DDX58 and ISG15, both crucial to the IFN-mediated
171 antiviral response in COVID-19 [27]; IFITM3, which is up-regulated in lung epithelial cells
172 during early SARS-CoV-2 infection [28]; and the chemokines CXCL11, CXCL1, CXCL6,
173 CXCL5 and CXCL10. Another significantly up-regulated pathway was “Senescence-

174 associated secretory phenotype”, which included up-regulated ubiquitin-conjugating enzymes
175 (UBE2S, UBE2E1), histones (H2BC21, H2BU1) and STAT3 (**Figure 2D**). Down-regulated
176 pathways included “Integrin cell surface interactions” and “Collagen biosynthesis and
177 modifying enzymes” which contained collagen proteins (e.g. COL11A2, COL13A1, COL15A1)
178 and related enzymes (e.g. P4HB, PCOLCE) (**Figure 2D**).

179

180 Transcriptomic and proteomic changes associated with COVID-19 severity

181

182 In both cohorts, the PCA of the PBMC transcriptomics revealed differences according to both
183 severity at time of sampling and overall clinical course (defined by peak severity score) (**Figure**
184 **3A**). There was a gradient of severity reflected in the molecular phenotype. We next assessed
185 molecular features associated with severity at time of blood sampling, encoded as an ordinal
186 variable. We identified 3,522 genes that were significantly associated with contemporaneous
187 severity in the Wave 1 cohort and 657 genes in the Wave 2 cohort (LMM, 1% FDR,
188 **Supplementary File 1F, Supplementary Figure 6**). We then applied GSVA to identify
189 pathways and used RRA to combine results from each cohort (**Supplementary File 1G**).
190

191

192 The up-regulated transcriptomic pathways in more severe disease included those involved in
193 oxidative stress (“Glutathione metabolism”, “Detoxification of reactive oxygen species”),
194 “Transcriptional regulation of granulopoiesis”, pathways containing numerous histone-
195 encoding genes (“HDACs deacetylate histones”, “Diseases of programmed cell death”, “RHO
196 GTPases activate PKNs”) and “Complement and coagulation cascades” (**Figure 3B-C,**
197 **Supplementary File 1G**). Down-regulated pathway terms included “TCRA pathway”,
198 “Pathogenesis of SARS-CoV-2 mediated by nsp9-nsp10 complex”, “TP53 activity”, and “PD1
199 signaling”, suggesting T cell activation in more severe COVID-19 (**Figure 3B-C,**
Supplementary File 1G).
200

201

202 PCA of the proteomic data revealed differences according to clinical severity (**Supplementary**
203 **Figure 7A**). We found 148 and 1,625 proteins associated with disease severity in the Wave 1
204 (86 COVID-19 positive samples) and Wave 2 (102 COVID-19 positive samples) datasets,
205 respectively (**Supplementary File 1H, Supplementary Figure 8**). Pathway analysis identified
206 15 severity-associated pathway terms that reached statistical significance (1% FDR) in both
207 cohorts (**Supplementary Figure 7B, Supplementary File 1I**). Among the most upregulated
208 pathways in more severe disease were “HDACs deacetylate histones”, pathways related to
209 transcriptional regulation (e.g. “mRNA splicing minor pathway”, “Spliceosome”, “RNA
210 polymerase II transcription termination”, “Processing of capped intron-containing pre mRNA”) and “RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function”,

211 while the most down-regulated pathways included “PD-1 signaling” and “T-cell receptor and
212 costimulatory signaling”. Example proteins from these pathways are shown in **Supplementary**
213 **Figure 7C.**

214

215 Severe COVID-19 is associated with dynamic multi-omic modular trajectories

216

217 We next examined the temporal trajectories of the transcriptome and the proteome during
218 COVID-19 by explicitly modelling molecular profiles with respect to time following symptom
219 onset (**Methods**). To aid biological interpretation, we first applied a dimension reduction
220 strategy using weighted gene correlation network analysis (WGCNA) [29]. WGCNA identified
221 23 modules of co-expressed genes (which we denote with the prefix ‘t’) (**Supplementary File**
222 **1J**), and 12 proteomic modules (denoted with ‘p’) (**Supplementary File 1K**). Longitudinal
223 modelling revealed 8 transcriptomic and 5 proteomic modules with significantly (5% FDR)
224 different temporal patterns in patients with mild/moderate versus severe/critical disease (LMM
225 time x clinical course (TxCC) interaction - **Methods**) (**Supplementary Tables 3-4**). Typically,
226 the modules displayed a flat temporal profile in mild/moderate COVID-19, whereas there was
227 a dynamic profile in severe/critical disease (**Figure 4A, Supplementary Figure 9**). Some
228 modules rose with time in severe/critical patients (e.g. tB, tL, p9 and p12), whilst others
229 dropped (e.g. tC, tP, tl, p7). Examples of individual genes from module tB exhibiting this
230 behaviour include *MMP9*, *ORM1*, *LRRN1* (**Figure 4B**).

231

232 We identified significant associations between modules, with transcriptomic and proteomic
233 modules clustered into larger positively or negatively correlated groupings (**Figure 4C**). The
234 inter-modular associations appeared to strongly reflect association with COVID-19 severity at
235 time of sampling (**Supplementary Tables 3-4**), implying that this is a strong underlying factor
236 in the -omics data. Consistent with this, integrated analysis of the transcriptomic and proteomic
237 datasets using MEFISTO [30] revealed a single factor that had a significantly different
238 trajectory in severe/critical versus mild/moderate disease ($p<0.0001$, LMM TxC)
239 (**Supplementary Figure 10**).

240

241 We characterised the modules by pathway analysis (**Figure 4A, Supplementary Tables 3-4**,
242 **Supplementary File 1L, Supplementary File 1M**). We also investigated whether disease
243 trajectory-associated transcriptomic modules might reflect a shift in cell-type proportions,
244 estimated using the CIBERSORTx algorithm (**Methods**) (**Supplementary Figure 11**,
245 **Supplementary File 1N**). The severity-associated modules tB and tJ were both strongly
246 positively associated with myeloid cell proportions, particularly neutrophils, and negatively
247 associated with lymphocyte subsets (**Supplementary Figure 11**). The presence of a

248 neutrophilic gene signature in the PBMC preparations may indicate the presence of low-
249 density granulocytes. Consistent with this, hub genes in Module tB (including *TECPR2*,
250 *CSF3R*, *STX3*; **Figure 4A**) are associated with granulocytes and autophagy, and pathway
251 analysis of the module genes revealed enrichment for pathways including “Neutrophil
252 degranulation” and “ROS and RNS production in phagocytes” (including genes encoding the
253 key cytosolic components of the phagocyte NADPH oxidase such as *NCF1*, *NCF2* and *NCF4*).
254 Module tB also contains genes encoding calcium-binding proteins (e.g. *S100A6*, *S100A9*,
255 *S100A11*, *S100A12*) that play important roles in regulating inflammatory pathways [31], as
256 well as integrins (e.g. *ITGA1*, *ITGAM*, *ITGB4*, *ITGAX*, *ITGAD*), adhesion molecules (e.g.
257 *CEACAM1*, *CEACAM3*, *CEACAM4*, *ICAM3*), *OSM* (encoding Oncostatin M) and *CSF1*
258 (encoding M-CSF). The tL module, which also displayed a rising trajectory in worse disease,
259 was strongly positively associated with imputed plasma cell proportion (**Supplementary**
260 **Figure 11**) and many of its members encoded immunoglobulins. The severity-associated
261 proteomic modules that strongly correlated with transcriptomic modules tB, tJ and tL were p8
262 and p9 (both enriched for pathways related to RNA splicing), and p12 (significantly enriched
263 for the pathway “HDACs deacetylate histones”) (**Supplementary Table 4**). The latter is
264 consistent with our earlier observations that a histone pathway signature was prominently
265 associated with COVID-19 severity in both the RNA-seq (**Figure 3C**, **Supplementary Figure**
266 **6**) and plasma proteomic data (**Supplementary Figure 7C**).
267

268 In contrast to tB, tJ and tL, the other transcriptomic modules (tP, tC, tF, tI, tN) all displayed a
269 decreasing trajectory in patients with worse disease (**Figure 4A**). These transcriptomic
270 modules tended to be positively associated with imputed lymphocyte subset proportions and
271 negatively associated with imputed myeloid proportions, implying that higher lymphocyte-
272 related gene signatures and lower myeloid-related ones is a favourable prognostic sign
273 (**Supplementary Figure 11**). While we cannot distinguish correlation from causation or indeed
274 reverse causation, it is possible that these modules represent genes that enable an
275 appropriate host response enabling viral clearance without an excessive inflammatory
276 response.
277

278 Flow cytometry identifies markers of enhanced interferon signaling early in severe disease

279
280 To understand whether transcriptional signatures in PBMCs reflected changes in blood cell
281 proportions, we performed flow cytometry on a subset of PBMC samples from the Wave 2
282 cohort. We found no major difference in the overall proportions of myeloid or lymphoid cells
283 within the PBMC fraction between pre-infection and COVID-19 positive samples, except for a
284 reduction in the proportion of type 2 dendritic cells (**Supplementary Figure 12**). Similarly,

285 there was little difference in the distribution of cells between mild/moderate and severe/critical
286 patients. We observed some severity-related differences within cell subsets. Within lymphoid
287 cells, we noted higher expression of the activation marker CD69 on CD4+ T cells at day 7 in
288 severe/critical disease compared to either pre-infection or mild/moderate disease
289 (**Supplementary Figure 13A**). At day 14, there was an increase in CD38^{hi} plasmablasts in
290 severe/critical disease compared to pre-infection or mild/moderate samples (**Supplementary**
291 **Figure 13B**). We also found that in severe/critical patients, there was a progressive drop in
292 the proportion of non-classical monocytes over the first 14 days of the illness that was more
293 marked than in mild/moderate patients (**Supplementary Figure 14A**). In severe/critical
294 patients there was a greater proportion of intermediate and non-classical monocyte subsets
295 expressing CD38 compared both to pre-infection samples and to mild/moderate patients
296 (**Supplementary Figure 14B**), likely reflecting enhanced activation [30]. In classical
297 monocytes there was a similar, but non-significant, trend. We found higher expression of
298 proliferation-associated Ki67 on classical monocytes in COVID-19 versus pre-infection
299 samples in both mild/moderate and severe/critical patients (**Supplementary Figure 14C**). In
300 our transcriptomic data we identified increased SIGLEC1 gene expression in COVID-19
301 (**Figure 2B**). SIGLEC-1 is exclusively expressed by CD14+ monocytes at the protein level.
302 SIGLEC-1 expression measured by flow cytometry correlated with GSVA enrichment score of
303 type I IFN signatures (**Supplementary Figure 14D**). We observed SIGLEC1 expression
304 increased at greater intensity as early as day 0-3 post infection in severe/critical versus
305 mild/moderate patients, suggesting stronger and a more immediate type I IFN response in
306 severe COVID-19 (**Supplementary Figure 14E**).
307

308 Longitudinal cytokine/chemokine analysis reveals distinct temporal profiles that distinguish
309 disease severity

310
311 Many plasma proteins associated with severe COVID-19 are canonically intra-cellular
312 proteins. Their elevation in severe COVID-19 may therefore be a readout of increased cell
313 turnover, death, stress, and viral hijacking of host cellular machinery. Consequently, we
314 performed a more focussed analysis examining proteins whose primary biological role is to
315 act extra-cellularly (e.g. cytokines, chemokines, growth factors and their receptors). These
316 classes of proteins are important therapeutic targets in inflammatory diseases [32].
317 Accordingly, we modelled the temporal profiles of 232 proteins that fell within the KEGG
318 pathway “Cytokine-cytokine receptor interaction”. Fifty proteins had significantly different
319 profiles in patients with a severe/critical clinical course versus those with mild/moderate ones
320 (TxC interaction effect, 5% FDR; **Supplementary File 1O**). Proteins exhibited distinct
321 patterns of divergence between severe/critical and mild/moderate disease over time (**Figure**

322 **5A**). Some (e.g. IL1 β , IL6, IL15RA, CCL2) showed a relatively stable temporal profile in
323 mild/moderate patients but rising trajectories in severe/critical patients (**Figure 5B**). Others
324 (e.g. CCL15, TNFSF13B (BAFF), PDGFRB, EDAR, IFNA10, IFNA13, IFNA16, IFNE and
325 IFNL3) were elevated early in the disease course and decreased over time, but displayed
326 more marked initial elevations in severe/critical patients (**Figure 5C**). Yet other proteins
327 displayed temporal profiles in mild/moderate patients that were inverted compared to
328 severe/critical. For example, CD40LG, TNFSF10 (TRAIL) and IL11 were reduced in the
329 severe/critical versus the mild/moderate group at early timepoints but increased in
330 severe/critical patients later (**Figure 5D**). Conversely, leptin, INHBA (inhibin A), and CCL22
331 were initially higher in severe/critical than mild/moderate patients but with the reverse pattern
332 later on (**Figure 5E**). These data illustrate the dynamic nature of the soluble protein response
333 and how this varies according to disease severity, highlighting the limitation of studies that use
334 a single snapshot.

335

336 Plasma LRRC15 as a predictor of COVID-19 severity

337

338 We next investigated whether clinical severity could be inferred from the transcriptomic and/or
339 proteomic data and which had the better predictive performance. For this analysis, we
340 combined the COVID-19 cases from both cohorts. For each COVID-19 patient, we selected
341 the first sample at the patient's peak severity score so that there was one sample per patient.
342 To predict COVID-19 severity at time of sampling, we employed two supervised learning
343 methods, lasso and random forests. We applied these separately on i) the plasma proteomic
344 data; ii) the PBMC transcriptomic data; and iii) the combination of both (the multi-omic data).

345

346 The proteomic-based models consistently outperformed the transcriptome-based ones, with
347 non-overlapping 95% confidence intervals (**Figure 6A, Supplementary Figure 15A**). The
348 lasso model generated on the proteome had an estimated area under the curve (AUC) of 0.93
349 (versus 0.86 for the transcriptome). The random forests model generated on the proteome
350 had an AUC of 0.88 (versus 0.83 for the transcriptome). The models based on the proteome
351 alone also had greater predictive performance than those trained on the multi-omic data,
352 although the confidence intervals for the AUC estimates overlapped (**Figure 6A,**
353 **Supplementary Figure 15A**).

354

355 We next examined the supervised learning models to identify the most important biomarkers
356 of severe/critical disease (**Methods**) (**Figure 6B, Supplementary Figure 15B,**
357 **Supplementary File 1P-R**). Although only a minority of the input features to the "multi-omic"
358 model were proteins (34%; 6,323/18,548), proteins made up the majority of the top 15 most

359 important predictors (10/15 for lasso and 9/15 for random forests). This, and our finding that
360 the plasma proteome was a superior predictor of severity than the PBMC transcriptome,
361 highlights that plasma proteins provide a valuable read-out of the pathophysiological
362 processes in severe COVID-19.

363

364 Importantly, both lasso and random forests identified plasma LRRC15 protein levels as the
365 most important predictor of COVID-19 severity. Interestingly, this protein was recently
366 identified by two pre-prints as a receptor for SARS-CoV-2 [33,34]. We next examined
367 LRRC15's longitudinal trajectory over the course of COVID-19 infection, finding that it
368 displayed a different temporal profile dependent on the disease course ($p<0.0001$, TxC
369 interaction, LMM). The concentration was stable in most individuals with mild/moderate
370 COVID-19 (**Figure 6C**), whereas it decreased over time in severe/critical patients. Thus, a
371 snapshot level of LRRC15 and its dynamic profile over time can convey information on the
372 current clinical state of the patient and the overall course of the disease, respectively.

373

374 Persistent deranged platelet and coagulation pathways in convalescence

375

376 For 12 of the 17 patients in the Wave 2 cohort, we obtained a sample after clinical recovery at
377 approximately two months following the acute infection. PCA analysis of the PBMC
378 transcriptome showed that while pre-COVID-19 and convalescent samples appeared more
379 similar than samples taken during COVID-19, there were differences between the
380 convalescent samples and their pre-infection counterparts (**Figure 7A**), indicating that they
381 have not fully returned to baseline. Comparison of the convalescent samples to their paired
382 pre-COVID-19 samples revealed 25 significantly differentially expressed genes (1% FDR), of
383 which 24 were up-regulated post-COVID-19 (**Figure 7B, Table 1, Supplementary File 1S**).
384 Up-regulated clotting-related genes included *PF4* (encoding platelet factor 4) and the related
385 gene *PF4V1* (platelet factor 4 variant 1). Of note, these genes are located in the same genomic
386 region on chromosome 4, along with the chemokine *CXCL5*, which was also significantly up-
387 regulated. Another nearby gene, *PPBP* (Pro-Platelet Basic Protein, aka *CXCL7*), was also up-
388 regulated in convalescent samples, although it did not quite reach significance at 1% FDR
389 (nominal $P = 3.33 \times 10^{-5}$, adjusted $P = 0.0159$). The upregulation of these neighbouring genes
390 suggests they are influenced by a shared genomic regulatory element. Overrepresentation
391 analysis of the 25 differentially expressed genes revealed significant enrichment of terms
392 including "Platelet activation, signaling and aggregation", "Formation of fibrin clot/clotting
393 cascade", "Chemokine signaling pathway", "SARS-CoV-2 innate immunity evasion and cell-
394 specific immune response" and "Smooth muscle contraction" (**Figure 7C, Supplementary
395 File 1T**). These data suggest persistent activation of abnormal processes for a considerable

396 time after clinical recovery. In particular, they implicate the vascular and clotting systems,
397 which may have implications for long-term risk of thrombosis.

398

399 Discussion

400

401 This study leverages a unique feature of COVID-19 in patients with ESKD on haemodialysis:
402 the possibility of obtaining serial blood samples throughout the disease course, irrespective of
403 disease severity. This allows a rare insight into the pathogenesis of COVID-19 through
404 examination of the temporal evolution of molecular and cellular changes. Moreover, ESKD
405 patients are an important group to study as they are at substantial risk of severe or fatal
406 disease [21,35]. Despite the remarkable success of vaccination programmes at the population
407 level, ESKD patients display impaired vaccine responses [23,24]. In addition, the majority of
408 patients in our study were of non-white ethnicity, which is also a risk factor for severe disease
409 [21].

410

411 Most studies of circulating proteins in COVID-19, including our previous work, have used Olink
412 immunoassay technology [8–10] or mass spectrometry [11,12]. The broadest Olink assay
413 system, used in the study of Filbin et al [8], measures 1,472 proteins, while mass spectrometry
414 is generally limited to reliable detection of less than 1,000 plasma proteins and lacks sensitivity
415 for low abundance proteins. A small number of studies have employed the aptamer-based
416 SomaScan v4 platform, that measures 4,665 unique proteins [8,13–16]. Here, we used the
417 SomaScan v4.1, which measures 6,323 unique proteins, and complemented this with RNA-
418 seq and flow cytometry. Our study is strengthened by data from two cohorts from different
419 waves of the pandemic, and the comparison of samples from before, during and after COVID-
420 19 from the same individuals.

421

422 Plasma proteomics identified several pathways upregulated in COVID-19 related to host
423 defence against viruses, including those previously described in SARS-CoV-2. Our PBMC
424 transcriptomic analysis identified numerous pathways that are up-regulated in COVID-19.
425 Many have been identified in previous studies of COVID-19 in other populations without
426 ESKD, indicating the presence of common patterns of COVID-19-related immunological
427 abnormalities. Examples include type 1 interferon signaling, the complement cascade, and
428 genes reflecting leukocyte-vascular interactions. Other up-regulated pathways included “Polo-
429 like kinase mediated events” and “Golgi-cisternae peri-centriolar stack re-organisation”. Both
430 are likely to reflect the extensive cell division of immunocytes that occurs in COVID-19. For
431 instance, the pericentriolar stacks of Golgi cisternae undergo extensive fragmentation and
432 reorganization in mitosis. Similarly, polo-like kinase is crucial for facilitating the G2/M

433 transition. These findings are consistent with the up-regulation of APC-Cdc20 mediated
434 degradation of Nek2A and other APC-Cdc20 related processes that we observed in the
435 proteomic data; Cdc20 is a protein that is key to the process of cell division.

436

437 Transcriptomic and proteomic associations with severe COVID-19 converged on some
438 unifying themes. For example, up-regulation of histone-encoding genes and elevated plasma
439 histone protein levels were both markers of COVID-19 severity. The increased expression of
440 histone-encoding transcripts may indicate increased immune cell proliferation. In each cell
441 cycle, sufficient histones are needed to package the newly replicated daughter DNA strands,
442 requiring tight coupling of histone synthesis to the cell cycle [36]. Excess histones within cells
443 can trigger chromatin aggregation and block transcription [37]. Thus, in severe COVID-19,
444 viral hijacking of cellular machinery may contribute to cellular damage through decoupling of
445 DNA synthesis and histone transcription. The preponderance of plasma histone proteins in
446 severe disease is likely to reflect the higher levels of cell damage and death. The presence of
447 histone proteins in plasma, however, is likely to represent more than just a marker of disease.
448 Histones are constituents of neutrophil extracellular traps (NETs) which contribute to tissue
449 injury in severe COVID-19. In addition, histones constitute powerful damage associated
450 molecular patterns (DAMPs) and can perpetuate inflammation via ligation of toll-like receptors
451 and direct damage to epithelial and endothelial cells [38]. Upregulation of pathways related to
452 control of transcription and translation was another feature of severe COVID-19
453 (**Supplementary Figure 7B**), perhaps reflecting subversion of normal cell biology by SARS-
454 CoV-2. In keeping with this, studies of cells infected with SARS-CoV-2 revealed “re-shaping”
455 of processes including translation, splicing and nucleic acid metabolism [39,40].

456

457 Modular analysis highlighted a rising neutrophilic gene signature as the illness progressed in
458 severe/critical patients, with enrichment of reactive oxygen and nitrogen species pathways.
459 This suggests prolonged activation of neutrophils and their key effector pathways including
460 NET formation. This neutrophilic gene signature likely indicates the presence of low-density
461 granulocytes within the PBMC fraction. Data from other infections suggest that phagocyte
462 NADPH oxidase-derived reactive oxygen species can be detrimental in acute viral infection;
463 mice lacking components of the NADPH oxidase have reduced disease severity and
464 inflammation in response to influenza and lymphocytic choriomeningitis virus infection
465 [41][42][43].

466

467 Cytokines and their receptors play a major role in the pathogenesis of inflammatory diseases
468 and are important targets of existing drugs [32]. Longitudinal examination of plasma
469 cytokines/chemokines revealed divergence temporal trajectories between disease severity

470 strata, manifesting in several patterns (**Figure 5**). For example, in patients with a
471 severe/critical disease course, IL11 was reduced early on but increased later relative to more
472 indolent disease (**Figure 5D**). IL11 is known to cause progressive fibrosis [44,45], and the
473 marked increases late in severe/critical disease may have implications for the development of
474 pulmonary sequelae. Leptin, INHBA (inhibin A), and CCL22 showed the opposite pattern
475 (**Figure 5E**). Leptin has roles in both cell metabolism and immunity with many immune cells
476 responding to leptin directly via the leptin receptor, resulting in a pro-inflammatory phenotype
477 [46]. It is produced by adipocytes, so its elevation early in severe/critical disease may be a
478 read-out of higher body mass index, which is a risk factor for severe COVID-19, or increased
479 cell metabolism/turnover. Its fall over time in severe/critical patients may reflect weight loss
480 and cell death. Whether leptin is also directly influencing risk of severe disease through its
481 immunological effects is unclear. Inhibin-A progressively increased over time in mild/moderate
482 patients but fell in severe/critical patients. Inhibin-A negatively regulates dendritic cell
483 maturation and promotes a tolerogenic phenotype [47]. Failure to upregulate it later in the
484 disease course may therefore contribute to deleterious inflammation. Similarly, CCL22 plays
485 an important role in switching off inflammation. CCL22 promotes dendritic cell-regulatory T
486 cell interactions and CCL22 deficiency is associated with excessive pathogenic inflammation
487 in mice [48].

488

489 Proteins in the type 1 interferon (IFN) pathway were higher in severe/critical than
490 mild/moderate patients early in disease (**Figure 5C**), suggesting a paradoxical role of this
491 pathway in COVID-19. While inherited or acquired deficiencies of IFN proteins predispose to
492 risk of severe COVID-19 [49,50], our data suggest that the picture may be more complex.
493 Thus, IFNs may act as a double-edged sword, with harm to the host from both insufficient
494 responses (leading to failure to control the virus) and from excessive responses (resulting in
495 immunopathology). While we cannot exclude the possibility that increased IFNs is a
496 consequence rather than a cause of severe disease, their elevation very early in disease
497 suggests this is less likely. Another consideration is that the greater IFN response in severe
498 disease might reflect higher viral burden.

499

500 Using two distinct supervised learning methods, we observed that the plasma proteome better
501 captures disease severity than the PBMC transcriptome. When supervised learning algorithms
502 were trained on both the proteomic and transcriptomic data simultaneously, plasma proteins
503 dominate the list of important biomarkers. There are several reasons why this might be the
504 case. Plasma is under strong homeostasis: derangement is a marker of loss of physiological
505 control. Plasma proteins may provide important read-outs of both pathogenesis and tissue

506 injury by reflecting the activity of cell types other than PBMCs, such as neutrophils,
507 endothelium and hepatocytes (a major source of coagulation and complement proteins).

508

509 A striking finding was the predictive value of plasma levels of LRRC15 in indicating COVID-
510 19 severity. Longitudinal profiling revealed that LRRC15 levels remain stable in those with a
511 mild/moderate clinical course but decrease over time in severe/critical illness (**Figure 6C**). Our
512 findings are particularly intriguing as two recent pre-prints have identified LRRC15 as an
513 accessory factor for SARS-CoV-2 entry to cells. Using arrayed transmembrane protein and
514 pooled genome-wide CRISPR activation screens, Shilts and colleagues demonstrated that
515 the SARS-CoV-2 spike protein interacts with LRRC15 [33]. Both screens identified the
516 interaction and the CRISPRa screen identified LRRC15 and the established SARS-CoV-2
517 binding partner, ACE2, as the two most prominent interactors. This work also showed that
518 ACE2 and LRRC15 bind the C-terminal domain of the spike protein, which contains the
519 receptor binding domain, suggesting that the two proteins may compete for spike protein
520 binding. Song and colleagues also used a CRISPRa approach to identify proteins that could
521 bind the SARS-CoV-2 spike protein to the A375 melanoma cell line [34]. The screen identified
522 ACE2 and LRRC15, and further showed that the interaction took place with the receptor
523 binding domain of the spike protein. Expression of LRRC15 on a HeLa cell line that expresses
524 ACE2 inhibited the entry of a SARS-CoV-2 spike pseudovirus. This paper notes, however,
525 that LRRC15 is expressed on different cells from those that express ACE2 and proposes that
526 LRRC15 inhibits virally entry *in trans*, acting as a decoy and binding virions that cannot then
527 enter cells via ACE2. Our data are consistent with a model in which a failure to up-regulate
528 LRRC15 increases risk of severe COVID-19 disease because of the lack of a receptor that
529 inhibits its entry to cells. Thus our study is the first human *in vivo* study to highlight the
530 importance of LRRC15 in the response to SARS-CoV-2.

531

532 Another unique strength of our study was the availability of baseline pre-infection samples for
533 the Wave 2 cohort, as well as samples taken two months after the acute COVID-19 episode.
534 Leveraging this, we demonstrate that there is chronic activation of vascular, platelet and
535 coagulation pathways for a prolonged period after clinical resolution of disease. The elevated
536 risk of thrombotic events during acute COVID-19 is well-documented. In a large study
537 encompassing both hospitalised and non-hospitalised patients [51], the risk of pulmonary
538 embolism (PE) and deep vein thrombosis (DVT) were 27-fold and 17-fold increased,
539 respectively, in the seven days following diagnosis. These risk ratios are much higher than
540 those previously associated with upper respiratory tract infections, suggesting unique features
541 specific to SARS-CoV-2 infection. The risk of arterial thrombosis was also significantly
542 increased, although smaller in magnitude than the risk of venous thromboembolism (VTE).

543 The pathophysiology underlying COVID-19 associated coagulopathy is complex and involves
544 the convergence of several pathways [52]. Invasion of ACE2-expressing epithelial cells by
545 SARS-CoV-2 results in down-regulation of ACE2 and increased angiotensin II levels. This in
546 turn leads to increased expression of PAI1 which impairs breakdown of fibrin and promotes
547 increased vascular tone, via smooth muscle contraction. Endothelial cell activation,
548 complement activation, NETosis, hypoxia and cytokine/chemokine secretion all promote
549 coagulopathy through increases in tissue factor and concomitant fibrin formation. Remarkably,
550 our data suggest that these pathways remain dysregulated months after acute infection has
551 resolved (**Figure 7, Table 1**). This is particularly important given emerging evidence indicating
552 that the risk of thrombo-embolism extends beyond the acute phase. Ho *et al* showed that risk
553 of a PE was 3.5-fold higher even in the time window 28 to 56 days after diagnosis of COVID-
554 19 [51]. A recent population-wide registry study revealed that following COVID-19 the risk of
555 DVT and PE was significantly elevated for 70 and 110 days, respectively [53]. Although VTE
556 risk was greatest for those with severe disease, even patients with mild disease had elevated
557 VTE risk. Our data provide a molecular basis that begins to explain this risk. Intriguingly,
558 among the genes up-regulated in convalescent samples compared to pre-infection was
559 platelet factor 4 (*PF4*). *PF4* is expressed in platelets and leucocytes. It is released from the
560 alpha granules of activated platelets, contributing to platelet aggregation. The prolonged up-
561 regulation of *PF4* after COVID-19 is therefore likely to contribute to a prothrombotic state. Of
562 note, autoantibodies to *PF4* are the pathogenic entity in both vaccine-induced thrombotic
563 thrombocytopenia (VITT) [54,55] and heparin-induced thrombocytopenia (HIT). *PF4* becomes
564 an autoantigen when it forms complexes with adenoviral vaccine components or heparin
565 respectively, unmasking epitopes to which autoantibodies bind [56]. It will therefore be
566 interesting for future studies to investigate whether autoantibodies to *PF4* might contribute to
567 post-COVID-19 thrombosis in some patients. Whether the molecular abnormalities found in
568 our study also apply to more general patient populations without background ESKD needs to
569 be determined. Ongoing studies focusing on the sequelae of COVID-19 are well placed to
570 address this.

571

572 Our study has several limitations. ESKD patients have considerable multi-morbidity and
573 deranged physiology, and our findings may not all be generalisable to other patient
574 populations. We lacked a comparator group of ESKD patients with another viral infection to
575 delineate COVID-19 specific features. We studied peripheral blood; while this can provide
576 valuable information, it does not always reflect processes at the site of tissue injury. We
577 performed bulk RNA-seq on PBMCs. Thus, transcriptomic signatures may reflect both
578 changes in gene expression and also alteration in the distribution of cell subtypes within
579 PBMCs. We mitigated this issue through use of deconvolution methods and flow cytometry,

580 but future studies using single cell RNA-seq and CITE-seq will provide further granularity. We
581 did not have measurements of viral load which would have aided interpretation of the
582 magnitude of host responses (e.g. interferon signaling). Finally, the convalescent samples
583 were taken relatively soon after clinical recovery: it will be important for future studies to
584 establish how long molecular abnormalities persist.

585

586 In summary, we demonstrate dynamic transcriptomic, proteomic and cellular signatures that
587 vary both with time and COVID-19 severity. We show that in patients with a severe clinical
588 course there is increased type 1 interferon signaling early in the illness, with increases in pro-
589 inflammatory cytokines later in disease. We identify plasma levels of the proposed alternative
590 SARS-CoV-2 receptor, LRRC15, as the strongest predictor of COVID-19 severity. Finally, we
591 show that immune cells display dysregulated gene expression two months following COVID-
592 19, with upregulation of clotting-related genes. This may contribute to the prolonged
593 thrombotic risk post-COVID-19.

594

595 **Methods**

596

597 **Patient cohorts and ethical approval**

598 All participants (patients and controls) were recruited from the Imperial College Renal and
599 Transplant Centre and its satellite dialysis units, London, United Kingdom, and provided
600 written informed consent prior to participation. Study ethics were reviewed by the UK National
601 Health Service (NHS) Health Research Authority (HRA) and Health and Care Research Wales
602 (HCRW) Research Ethics Committee (reference 20/WA/0123: The impact of COVID-19 on
603 patients with renal disease and immunosuppressed patients). Ethical approval was given.
604

605

606 We recruited two cohorts of ESKD patients with COVID-19 (**Figure 1A**). All patients were
607 receiving haemodialysis prior to acquiring COVID-19. The first cohort ('Wave 1') were recruited
608 during the initial phase of the COVID-19 pandemic (April-May 2020). Blood samples were
609 taken from 53 patients with COVID-19 (**Supplementary Table 1**). Serial blood sampling was
610 carried out where feasible (**Figure 1B**), given the pressure on hospital services and the effects
611 of national lockdown. We also contemporaneously recruited 59 non-infected haemodialysis
612 patients to provide a control group, selected to mirror the age, sex and ethnicity distribution of
613 the COVID-19 cases (**Supplementary Figure 1A-C**).
614

615

616 The Wave 2 cohort consisted of 17 ESKD patients with COVID-19 infected during the
617 resurgence of cases in January-March 2021 (**Supplementary Table 2**). These 17 individuals
618 had all been recruited as part of the COVID-19 negative control group during Wave 1, and so
619 a pre-infection sample collected in April/May 2020 (8-9 months preceding infection) was also
620 available. For the Wave 2 cohort, we systematically acquired serial samples for all patients at
621 regular intervals (every 2-3 days over the course of the acute illness) (**Figure 1C**). Additionally,
622 for 12 of these 17 patients, we acquired convalescent samples at approximately 2 months
623 post the acute COVID-19 episode (range 41-55 days from the initial sample). Convalescent
624 samples were unavailable for four patients who died and for one patient due to logistical
625 difficulties in sample collection.
626

627

628 To minimise variation related to the timing of dialysis, blood samples were taken prior to
629 commencing a haemodialysis session.
630

628 **Clinical severity scoring**

629 We assessed disease severity using a four-level ordinal score, categorising into mild,
630 moderate, severe, and critical, based on the WHO clinical management of COVID-19: Interim
631 guidance 27 May 2020. ‘Mild’ was defined as COVID-19 symptoms but no evidence of
632 pneumonia and no hypoxia. ‘Moderate’ was defined as symptoms of pneumonia or hypoxia
633 with oxygen saturation (SaO_2) greater than 92% on air, or an oxygen requirement no greater
634 than 4 L/min. ‘Severe’ was defined as SaO_2 less than 92% on air, or respiratory rate more
635 than 30 per minute, or oxygen requirement more than 4 L/min. ‘Critical’ was defined as organ
636 dysfunction or shock or need for high dependency or intensive care support (i.e. the need for
637 non-invasive ventilation or intubation). We recorded disease severity scores throughout the
638 illness, such that samples from the same individual could have differing severity scores
639 according to the temporal evolution of the disease. We defined the overall clinical course for
640 each patient as the peak severity score that occurred during the patient’s illness. Different
641 downstream analyses utilise either the severity at the time of sample (i.e. the sample-level
642 severity) or the overall clinical course (i.e. the patient-level severity), as described in the
643 relevant sections below.

644

645 **PBMC collection protocol**

646 Peripheral blood mononuclear cells (PBMCs) were obtained by density gradient centrifugation
647 using Lymphoprep (STEMCELL Technologies, Canada). Approximately 20 ml of blood were
648 diluted 1x with phosphate buffered saline (PBS) with addition of 2% FBS and layered on top
649 of 15 ml of Lymphoprep solution. The samples were then centrifuged at 800 g for 20 minutes
650 at room temperature without break. PBMCs were collected from the interface and washed
651 twice with PBS/2%FBS. 2 million PBMCs were centrifuged down to form a pellet and
652 resuspended in 350 ul RLT buffer + 1% β -Mercaptoethanol (from Qiagen RNAeasy kit) for
653 RNA extraction. Remaining PBMCs were cryopreserved in 1 ml freezing medium (FBS 10%
654 DMSO) and stored in –80 degrees C freezer.

655

656 **Plasma collection.**

657 5 ml of blood was collected in EDTA tubes and centrifuged at 1000 RPM for 15 mins. Plasma
658 was extracted and frozen at –80 degrees Celcius.

659

660 **RNA-seq of PBMCs**

661 RNA extraction and sequencing were done at GENEWIZ facilities (Leipzig, Germany). Total
662 RNA was extracted from using RNeasy Mini kits (Qiagen) as per the manufacturer’s
663 instructions, with an additional purification step by on-column DNase treatment using the
664 RNase-free DNase Kit (Qiagen) to remove any genomic DNA. Total RNA quality and
665 concentration was analysed using Agilent Tapestation (Agilent Tech Inc.). Samples with RIN
666 values ≥ 6.0 and ≥ 100 ng of total RNA were used to generate RNA-seq libraries. RNA-seq
667 libraries were made using NEBnext ultra II RNA directional kit per the manufacturer’s
668 instruction. Poly-A RNA was purified using poly-T oligo-attached magnetic beads followed by
669 hemoglobin mRNA depletion using QIAseq FastSelect Globin Kit to remove potential
670 contaminating RNA from red blood cell. Then, first and second cDNA strand synthesis was
671 performed. Next, cDNA 3' ends were adenylated and adapters ligated followed by library
672 amplification. The libraries were size selected using AMPure XP Beads (Beckman Coulter),
673 purified and their quality was checked using a short sequencing run on MiSeq Nano. Samples
674 were randomized to avoid confounding of batch effects with clinical status and multiplexed
675 libraries were run on 29 lanes of the Illumina HiSeq platform to generate approximately 30
676 million x 150bp paired-end reads per sample.

677

678 Initial quality control and alignment was performed using the nf-core RNA-seq v3.2 pipeline
679 [57] based on nextflow [58], a workflow management system. FastQC [59] was used to
680 evaluate and merge paired reads prior to adapter trimming using Trimgalore [60]. We used
681 STAR [61] to align reads to GRCh38 and htseq-count [62] to generate a counts matrix.

682

683 For the Wave 1 cohort, following quality control (QC), transcriptomic data were available for
684 179 samples from 51 COVID-19 positive ESKD patients (median 3 samples per patient, range
685 1-8) (**Supplementary Figure 1D**), plus 55 non-infected ESKD patient samples. For the Wave
686 2 cohort (17 patients), following QC, transcriptomic data were available for 90 samples
687 collected during acute COVID-19 infection (median of 6 samples per patient, range 3-7), plus
688 17 pre-infection samples and 12 convalescent samples.
689

690 Prior to further analysis, genes with insufficient counts were removed using edgeR's
691 filterByExpr function [63]; for differential expression analyses, the "group" argument was set
692 to the main group of interest. For all analyses, gene expression was TMM normalised [64],
693 converted to counts per million (CPM) and log-transformed. We primarily used ENSEMBL
694 identifiers [65], however for plots we report the HGNC gene ID [66] where available. For
695 analyses that considered multiple proteins simultaneously (PCA, WGCNA, MEFISTO,
696 supervised learning), we additionally: i) removed genes with low variance (33% of genes with
697 the lowest maximum absolute deviation) [67]; ii) centered and scaled the data.
698

699 **Plasma proteomics**

700 We performed proteomics on EDTA plasma samples using the aptamer-based SomaScan
701 platform (SomaLogic, Boulder, Colorado, USA). The SomaScan v4.1 assay contains 7,288
702 modified-aptamers (Somamers) that target human proteins. Since more than one aptamer
703 may target the same protein, these 7,288 aptamers map to 6,347 unique proteins. 48
704 Somamers were removed due to QC failure, so the final dataset contains 7,240 Somamers
705 representing 6,323 unique proteins. We annotated these proteins using the Human Protein
706 Atlas [68]; 4,980 proteins were labelled as intracellular, 1,586 were annotated as membrane
707 proteins and 1,160 as secreted (**Supplementary Figure 16A**). Many proteins were labelled
708 as both intracellular and as membrane or secreted, reflecting the biology of protein storage
709 and extra-cellular secretion/excretion (**Supplementary Figure 16B**).
710

711 We report proteins by their corresponding HGNC gene ID [66], which provides a more
712 standardised nomenclature compared to protein names and allows direct comparison with the
713 transcriptomic data.
714

715 Where multiple Somamers related to the same protein, we retained these Somamers for
716 univariate analyses such as differential abundance analyses. However, for analyses that
717 considered multiple proteins simultaneously (PCA, WGCNA, MEFISTO, supervised learning),
718 we selected one Somamer at random to represent each protein. One COVID-19 positive
719 sample in the wave 2 cohort failed QC and was excluded from the analyses. The raw
720 SomaScan data was separated by cohort. The expression values for each Somamer were
721 inverse-rank normalised prior to downstream analyses.
722

723 For the Wave 1 cohort, following QC, proteomic data were available for 86 samples from 37
724 COVID-19 positive ESKD patients (median 3 samples per patient, range 1-3), plus 53 non-
725 infected ESKD patients. For the Wave 2 cohort (n=17 patients), following QC, proteomic data
726 were available for 102 samples collected serially during acute COVID-19 infection (median of
727 6 samples per patient, range 5-7) and 16 pre-infection samples.
728

729 **Differential expression analyses: COVID-19 positive versus negative**

730 We compared COVID-19 positive and negative patients using linear mixed models (LMM),
731 which account for serial samples from the same individual [69]. Age, sex and ethnicity were
732 included as covariates. A random intercept term was used to estimate the variability between
733 individuals in the study and thus account for repeated measures. We performed differential
734 expression analyses for the transcriptomic data and the proteomic data. The regression model
735 for these analyses in Wilkinson-style notation was:
736

$$737 E \sim covid_status + sex + age + ethnicity + (1 | individual)$$

738
739 where, E represents expression (gene or protein, depending on the data type being analysed)
740 and “covid_status” was a categorical variable (infected/non-infected).
741

742 For differential expression of proteins, we applied LMM using the ImeRTest package [70].
743 Differential gene expression analysis was performed using the same model formula, applied
744 using the differential expression for repeated measures (dream) pipeline [71] in the
745 variancePartition package [72]. For all data types, we fitted LMM using restricted maximum
746 likelihood (REML) and calculated P-values using a type 3 F-test, in conjunction with
747 Satterthwaite’s method for estimating the degrees of freedom for fixed effects [70]. Multiple
748 testing correction was performed using the Benjamini-Hochberg method and a 1% FDR used
749 for the significance threshold.
750

751 The Wave 1 cohort was analysed separately to the Wave 2 cohort. For Wave 1, we compared
752 samples from COVID-19 positive ESKD patients to COVID-19 negative ESKD patients. For
753 Wave 2, we compared samples from COVID-19 positive ESKD patients to samples from these
754 patients taken approximately 8 months prior to infection.
755

756 When reporting the number of differentially expressed proteins in the text we refer to the
757 number of unique proteins rather than the number of significant Somamers.
758

759 **Testing transcriptomic and proteomic features for association with COVID-19 severity**

760 We performed a within-cases analysis, testing for the association of gene expression with
761 COVID-19 severity at time of sampling. We used the four-level WHO severity rating (mild,
762 moderate, severe, critical), which could vary between samples from the same individual
763 reflecting the clinical status at the time the sample was taken. We again used a linear mixed
764 model to account for samples from the same individual. The regression model was:
765

$$766 \quad E \sim \text{covid_severity} + \text{sex} + \text{age} + \text{ethnicity} + (1 | \text{individual})$$

767
768 The “covid_severity” variable represents severity at the time of the sample and was encoded
769 using orthogonal polynomial contrasts to account for ordinal nature of severity levels.
770

771 COVID-19 positive samples from the Wave 1 cohort were analysed separately to those from
772 the Wave 2 cohort.
773

774 The same approach was used for the proteomics data.
775

776 **Gene set variation analysis**

777 To identify pathways that were up- or down-regulated in COVID-19 positive versus negative
778 samples, we applied gene set variation analysis (GSVA) [25]. To define gene sets, we used
779 the MSigDb C2 canonical pathways [73]; we discarded sets with less than ten genes. We
780 additionally included a gene set for the peripheral immune response defined for patients with
781 severe COVID-19 [26] and a set of type 1 interferons active in patients with systemic lupus
782 erythematosus (SLE) [74]. After reduction of genes into gene sets, we then performed testing
783 for dysregulated pathways using the same linear mixed modelling approach as for the
784 differential gene and protein expression analyses. P-values were adjusted by Benjamini-
785 Hochberg, with a significance threshold of 1% FDR.
786

787 To dissect out the key molecules underpinning enriched pathways, we examined the genes
788 that comprise these pathway terms and identified which of these featured most prominently in
789 the differential gene expression analysis.
790

791 We repeated this procedure for testing of association of pathways with severity at the time of
792 sample using the 4-level ordinal score.
20

793

794 We then applied the same approach to the proteomics data for the COVID-19 positive versus
795 negative analysis, and for testing associations with COVID-19 severity at the time of sample.

796

797 **Robust rank aggregation**

798 The Wave 1 and Wave 2 cohorts were analysed separately for both the differential expression
799 analyses between COVID-19 positive and negative samples and for the within-cases severity
800 analyses. To identify the associations that were most consistent between the Wave 1 and
801 Wave 2 cohorts, for each analysis, we integrated the P-values for each cohort using robust
802 rank aggregation (RRA) [75]. This method identifies features that are ranked higher than
803 expected across multiple lists. RRA generates a significance score analogous to a P-value;
804 we -log₁₀ transform these values such that a larger score indicates more consistent
805 associations between the Wave 1 cohort and the Wave 2 cohort. RRA was applied to the
806 results of the transcriptomic, proteomic and GSVA analyses comparing COVID-19 positive
807 versus negative samples from Wave 1 and Wave 2. Similarly, it was applied to the analyses
808 testing for association of molecular features with COVID-19 severity at the time of sampling.

809

810 **Modelling modular longitudinal trajectories**

811 We examined the temporal trajectories of the transcriptome following infection, by explicitly
812 modelling molecular markers with respect to time following COVID-19 symptom onset. We
813 used a two-step approach.

814

815 Step 1. To aid biological interpretation, we first applied a dimension reduction strategy using
816 weighted gene correlation network analysis (WGCNA) [29] to identify modules of correlated
817 molecular features. For this analysis, we combined samples from the Wave 1 and Wave 2
818 cohorts. Additionally, since our goal was to perform longitudinal analysis, we only selected
819 patients who had been sampled at least three times prior to 21 days following COVID-19
820 symptom onset. The default implementation of WGCNA is not designed for use with non-
821 independent samples [76], so we modified the analysis pipeline by generating a correlation
822 matrix using a repeated measures correlation metric (rmcorr) that is appropriate for repeated
823 measures [77]. We used WGCNA's pickSoftThreshold.fromSimilarity function to pick the
824 minimum soft-thresholding power that satisfied the minimum scale free topology fitting index
825 ($R^2 > 0.85$) and maximum mean connectivity (100). We subsequently defined signed adjacency
826 and topological overlap matrices before applying average-linkage hierarchical clustering. We
827 cut this tree with a hybrid dynamic tree cutting algorithm, with the parameters deepSplit = 4
828 and minClusterSize = 30 [78]. Finally, we defined eigengenes for each module and merged
829 those with a distance less than 0.25. The eigen-genes provide a numerical representation for
830 each module of co-expressed genes.

831

832 We used the same approach to analyse the proteomic data.

833

834 Step 2. To examine the trajectory of each module over time, we fitted a linear mixed model
835 with time from symptom onset as an independent variable and the eigengene (or eigenprotein
836 in the case of proteomic modules) as the dependent variable. Time was defined for each
837 sample as time from first symptoms; where date of first symptoms was not available, we
838 instead used date of first positive swab. Samples that were taken more than 21 days from
839 each individual's baseline date were excluded. We used R's bs function to fit a polynomial
840 spline of degree two to model the expression of modules with respect to time from baseline
841 [79]. To test whether modules displayed different temporal patterns according to the overall
842 clinical course of COVID-19 (defined as a binary variable indicating whether the peak WHO
843 severity score was mild/moderate or severe/critical), we included clinical course as a covariate
844 in the model, and an interaction term between time from symptom onset and clinical course
845 (TxC).

846

847 The regression model used is displayed using Wilkinson-style notation below.

848
849 eigen-expression ~ clinical_course * time + sex + age + ethnicity + wave + (1 | individual)
850

851 We extracted the P-values for the TxC term in this model and applied Benjamini-Hochberg
852 adjustment, using 5% FDR as the significance threshold. A significant interaction effect for the
853 TxC term indicates that the module has a different temporal profile in mild/moderate versus
854 severe/critical disease.
855

856 **Additional WGCNA module annotation and association testing**
857 To better understand the biological information reflected in the transcriptomic and proteomic
858 modules, we further characterised them through a multi-pronged analytical strategy. We
859 tested association of eigen-genes and eigen-proteins with other variables. First, we tested for
860 the association of the modules with WHO severity at the time of the sample using the LMM
861 approach described above in subsection ‘Testing transcriptomic and proteomic features for
862 association with COVID-19 severity’ i.e.:
863

864 E ~ covid_severity + sex + age + ethnicity + wave + (1 | individual)
865

866 Second, since PBMCs represent a mixed population of immune cells, we investigated whether
867 disease trajectory-associated transcriptomic modules might reflect shift in cell type
868 proportions. To this end, we applied CIBERSORTx, a computational algorithm to impute
869 immune cell fractions from RNA-seq data (see subsection ‘Cell fraction imputation’ below).
870 We then tested for correlations between these imputed immune cell proportions and module
871 eigengenes using LMM:
872

873 eigen-expression ~ cell_fraction + sex + age + ethnicity + wave + (1 | individual)
874

875 Both these models included an additional fixed effect (“wave”) to reflect the cohort.
876

877 Third, we performed pathway enrichment analysis on the modules using the R package
878 clusterProfiler’s “enricher” function [80]. Gene sets were defined using MSigDB C2 canonical
879 pathways [73].
880

881 Lastly, to understand the relationship between the transcriptomic and proteomic modules, we
882 performed correlation analysis using LMM.
883

884 5% FDR was used for statistical significance for these analyses.
885

886 **Cell fraction imputation**

887 We used CIBERSORTx [81] to impute cell fractions from the normalised bulk RNA-seq
888 dataset. The program was run with default parameters. We inferred the cell fractions of 22
889 immune cell types in the isolated PBMCs of each sample using the LM22 signature matrix file
890 [82].
891

892 **Multi-omic longitudinal factor analysis with MEFISTO**

893 MEFISTO [83] is an extension of Multi-Omics Factor Analysis (MOFA) that can exploit
894 temporal relationships between samples to find factors that change over time (from baseline).
895 We used this method to find joint factors of variation in the transcriptomic and proteomic
896 datasets. For the MEFISTO analysis, we used the same set of samples as in the network
897 analysis and applied the same pre-processing steps to the data (**see Methods – network**
898 **analysis**). Additionally, we removed genes with the lowest maximum absolute deviation [67]
899 such that the number of genes retained were equal to the number of unique proteins measured
900 (6,323) to avoid imbalance numbers of features between the transcriptomic and proteomic
901 data which can impact the MEFISTO algorithm. Using the “slow” convergence criterion,

902 MEFISTO identified 8 factors that had a minimal variance explained of 1% in at least one data
903 modality.

904
905 We then applied the longitudinal model described earlier to test for an interaction effect
906 between time from first symptoms and clinical course, with a latent factor identified by
907 MEFISTO as the dependent variable. The regression model used is displayed using
908 Wilkinson-style notation below:

909
910 Latent_factor ~ clinical_course * time + sex + age + ethnicity + wave + (1 | individual)
911

912 **Longitudinal modelling of cytokines and cytokine receptors**

913 We modelled the temporal profiles of 232 plasma proteins that fell within the KEGG pathway
914 “Cytokine-cytokine receptor interaction”. As for the longitudinal analyses described earlier, we
915 used a linear mixed model with a time x clinical course interaction term.

916
917 P ~ clinical_course * time + sex + age + ethnicity + wave + (1 | individual)

918
919 P values for the time x clinical course interaction were extracted and adjusted for multiple
920 testing with the Benjamini-Hochberg procedure, with significance threshold of 5% FDR.

921 **Supervised learning**

922 The goal of this analysis was to predict clinical severity from the molecular features
923 (transcriptomic, proteomic or both). We performed supervised learning using the R caret
924 framework [84]; caret uses the randomForest package to fit random forest models and glmnet
925 [85] to fit lasso models. For this analysis, we only included samples on which both
926 transcriptomics and proteomics had been performed. We then selected the earliest sample for
927 each individual at which they had reached their peak COVID-19 WHO severity score, so that
928 there was one sample per patient. We then categorised the clinical severity score
929 corresponding to each sample into a binary variable such that patients with a WHO severity
930 score of mild or moderate were considered “mild/moderate” and those with a WHO score of
931 severe or critical were considered “severe/critical”. This resulted in n=37 mild/moderate
932 samples and n=14 severe/critical samples.

933
934 We trained models using Monte Carlo cross-validation for: i) the plasma proteomic data alone
935 (6,323 features); ii) the PBMC RNA-seq data alone (12,225 features); and iii) the combined
936 proteomic and RNA-seq datasets. The first step in this training process was to create 200
937 random partitions of the data, such that 80% of the data was used to train the model in each
938 resample and 20% was retained as a validation set. In each resample, we calculated the area
939 under the curve (AUC) of the receiver operating characteristic (ROC) curve. We then
940 calculated confidence intervals for the 200 AUC-ROC values generated for each model and
941 feature type.

942
943 The random forest model’s parameters were kept constant at 500 trees and the mtry value
944 (number of proteins randomly sampled as candidates at each node) was calculated as the
945 square root of the number of features. After cross-validation, we fitted a final random forest
946 model using the entirety of the dataset. We extracted important features from this model using
947 the R randomForestExplainer package, based on the accuracy decrease metric (the average
948 decrease in prediction accuracy upon swapping out a feature). For the lasso model, the
949 lambda value that maximised the mean AUC-ROC during cross-validation was selected. We
950 recorded the features selected by the lasso model in each data resample; feature importance
951 was subsequently defined as the number of models in which each feature had a non-zero
952 coefficient. The feature importance metrics from both models were scaled by dividing their
953 values by the maximum value, such that the most important feature has an importance metric
954 of 1.

956

957 **Differential gene expression analysis: pre-infection versus recovery samples**

958 For the 12 individuals in the Wave 2 cohort for whom we collected a convalescent sample
959 (approximately 2 months post-infection; range 41-55 days from the initial sample), we
960 performed a differential gene expression analysis comparing these samples to the paired pre-
961 infection samples using LMM, implemented with the R dream package [71]. Age, sex and
962 ethnicity were included as covariates and a random intercept term used to account for the
963 paired nature of the samples. Statistical significance was defined as 1% FDR. To identify
964 enriched pathways in the list of differentially expressed genes, we performed
965 overrepresentation analyses using the same approach as described above for annotating the
966 WGCNA modules.

967

968 **Flow cytometry**

969 Flow cytometry analysis was performed on a subset of the Wave 2 PBMC samples. We
970 examined samples taken during acute COVID-19 from 17 patients (of whom 9 patients had a
971 mild/moderate clinical course and 8 patients with severe/critical course), and pre-infection
972 samples from 15 of these same patients.

973

974 Cryopreserved PBMCs were thawed in humidified 37°C, 5% CO₂ incubator and resuspended
975 in thawing medium (RPMI, 20% FBS). PBMCs were washed twice with PBS and stained with
976 Zombie Yellow LIVE/DEAD (Biolegend) following the manufacturer's protocol to exclude dead
977 cells. Then, PBMCs were washed twice with FACS buffer (1% BSA, 0.09% Azide, 1 mM
978 EDTA), and Fc receptors were blocked with Human TruStain Fc Receptor Blocking Solution
979 (Biolegend). Then, surface staining were performed using the selected
980 fluorochrome-conjugated monoclonal antibodies detailed in **Supplementary Table 5** for 20
981 minutes at 4°C. Following incubation, cells were fixed and permeabilized using the
982 eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set (Invitrogen) for intracellular
983 staining. Cells were incubated with selected antibodies or isotype controls for 30 minutes at
984 4°C and resuspended in FACS buffer for analysis. Aurora Spectral Flow Cytometry (Cytek®)
985 and FlowJo software, version 10 (Tree Star Inc. Ashland, OR, USA) were used for analysis of
986 all samples.

987

988 **Flow cytometry statistical analysis**

989 To evaluate decomposition performance by CIBERSORTx analysis, cell proportion estimates
990 were compared to cell percentages from Flow Cytometry analysis using Pearson's correlation
991 analysis (n=68 samples). We were unable to examine for the presence of LDGs using our flow
992 cytometry data since this was performed on cryopreserved PBMCs and LDGs do not survive
993 the freeze-thaw process (whereas we performed transcriptomics on RNA extracted from fresh
994 PBMCs). We observed significant correlation of estimated cell proportions from CIBERSORTx
995 analysis compared to proportions measured by flow cytometry for all other cell types (Pearson
996 r > 0.4045, p-value < 0.0001).

997

998 For severity analysis, one sample per patient was selected at a time that coincided with the
999 expected spike in the inflammatory response (nearest sample to day 7 after symptom onset;
1000 no more than +/- 72 hours). Patients were classified according to the overall peak illness
1001 severity into two groups (mild/moderate = 9, severe/critical = 8). Change of cell proportion
1002 across time were accessed by grouping samples into 4 days interval post COVID-19-positive
1003 test. One-way ANOVA was used to calculate significant differences between multiple groups
1004 with Dunnet's correction for multiple-way comparisons. Significance is based upon p-value <
1005 0.05.

1006

1007

1008 **Data and code availability**

- 1009 • Individual-level data for transcriptomics, proteomics and flow cytometry are available
1010 without restriction from Zenodo (doi: 10.5281/zenodo.6497251)
1011 • Code is available at: <https://github.com/jackgisby/covid-longitudinal-multi-omics>

1012
1013 **Acknowledgements**

1014 The authors thank the patients who volunteered for this study and the staff at Imperial College
1015 Healthcare NHS Trust (the Imperial College Healthcare NHS Trust renal COVID-19 group and
1016 dialysis staff):

1017 Appelbe M, Ashby DR, Brown EA, Cairns T, Charif R, Condon M, Corbett RW, Duncan N,
1018 Edwards C, Frankel A, Griffith M, Harris S, Hill P, Kousios A, Levy JB, Loucaidou M, Lightstone
1019 L, Liu L, Lucisano G, Lynch K, Mclean A, Moabi D, Muthusamy A, Nevin M, Palmer A, Parsons
1020 D, Prout V, Salisbury E, Smith C, Tam F, Tanna A, Tansey K, Tomlinson J, Webster P.

1021
1022 We also acknowledge the efforts of renal specialist doctors in training for assistance with
1023 recruiting patients to this study.

1024
1025 We acknowledge the Imperial College Research Computing Service (DOI:
1026 10.14469/hpc/2232).

1027
1028 **Funding statement**

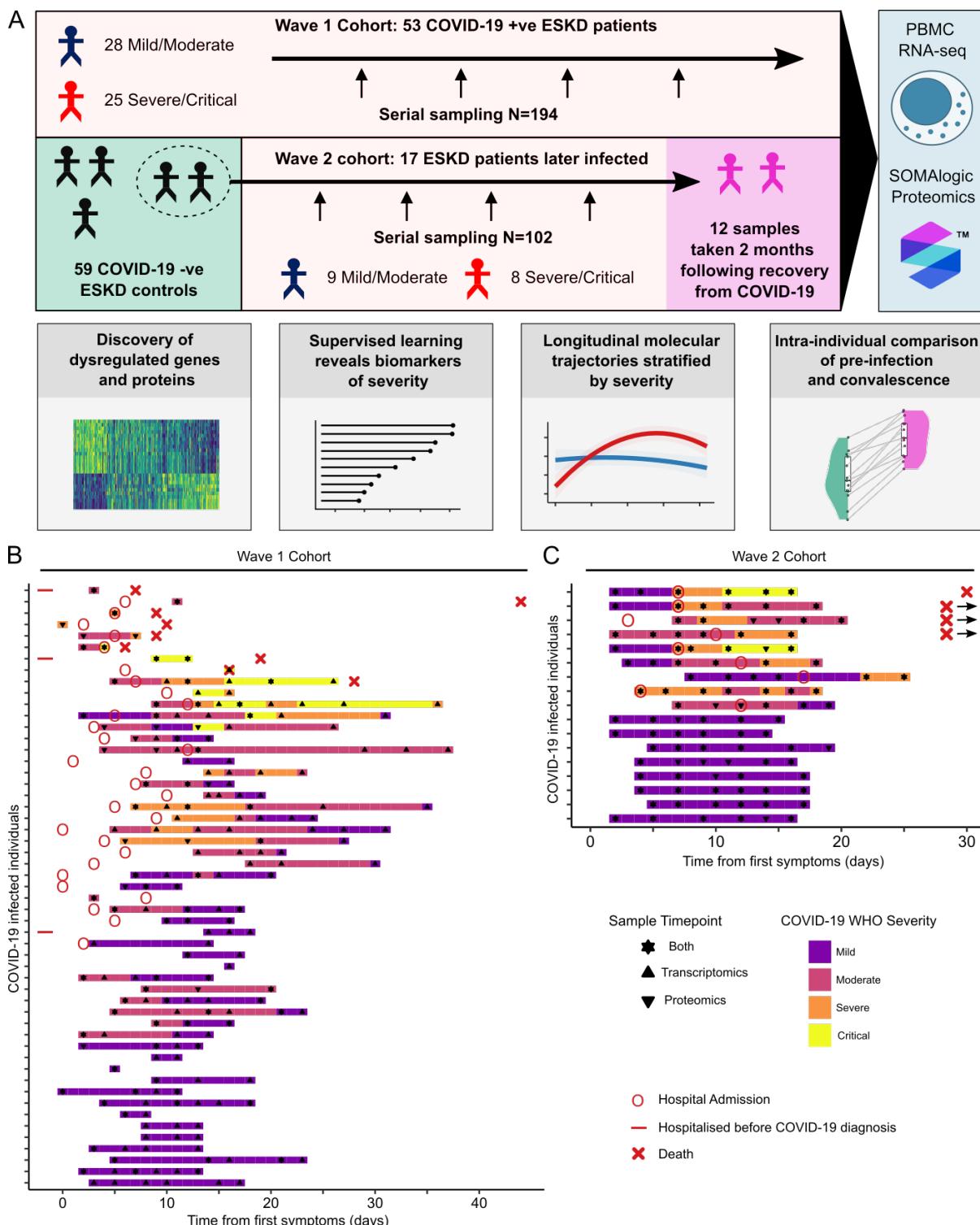
1029 This research was partly funded by Community Jameel and the Imperial President's
1030 Excellence Fund and by a UKRI-DHSC COVID-19 Rapid Response Rolling Call
1031 (MR/V027638/1) (to JEP), and by funding from UKRI/NIHR through the UK Coronavirus
1032 Immunology Consortium (UK-CIC) (to MB). We also acknowledge the National Institute for
1033 Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare
1034 NHS Trust and Imperial College London. The views expressed are those of the author(s) and
1035 not necessarily those of the NHS, the NIHR or the Department of Health. JEP was supported
1036 by UKRI Innovation Fellowship at Health Data Research UK (MR/S004068/2). DCT is
1037 supported by a Stage 2 Wellcome-Beit Prize Clinical Research Career Development
1038 Fellowship (20661206617/A/17/Z and 206617/A/17/A) and the Sidharth Burman endowment.
1039 MCP is a Wellcome Trust Senior Fellow in Clinical Science (212252/Z/18/Z). NM-T and ES
1040 are supported by Wellcome Trust and Imperial College London Research Fellowships, and
1041 CLC by an Auchi Clinical Research Fellowship.

1042
1043 The funders had no role in study design, data collection and interpretation, or the decision to
1044 submit the work for publication.

1045
1046 **Competing interests**

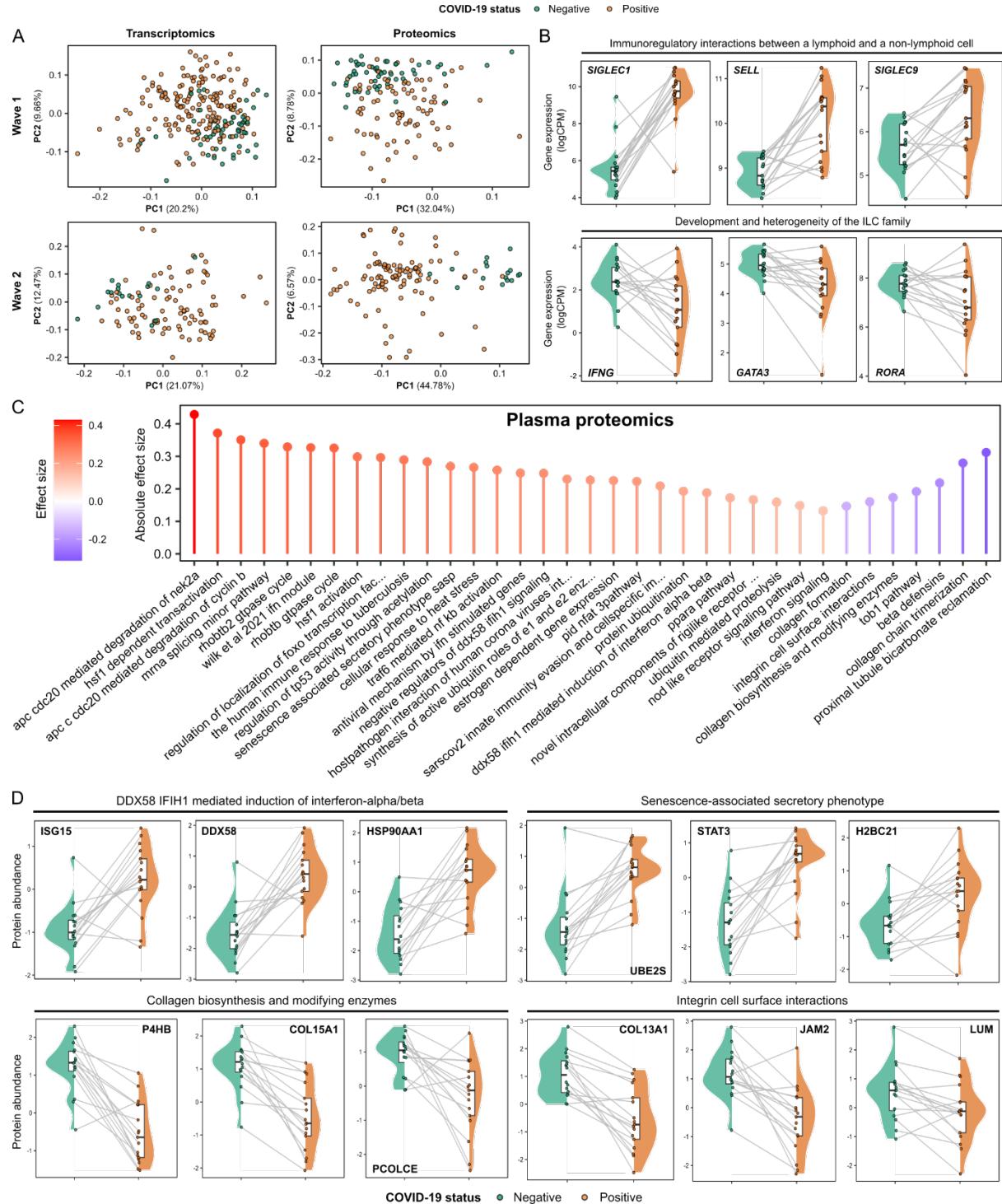
1047 None of the authors have any patents (planned, pending or issued) or competing interests
1048 relevant to this work. Other interests unrelated to this work: SPM reports personal fees from
1049 Celltrion, Rigel, GSK and Cello; MCP reports consulting honoraria with Alexion, Apellis,
1050 Achillion, Novartis and Gyroscope; DCT reports speaker and consultancy fees from Astra-
1051 Zeneca and Novartis; JEP has received travel and accommodation expenses and hospitality
1052 from Olink to speak at Olink-sponsored academic meetings (none within the past 5 years).
1053
1054 None of the other authors have any interests to declare.

1057 **FIGURES**
 1058



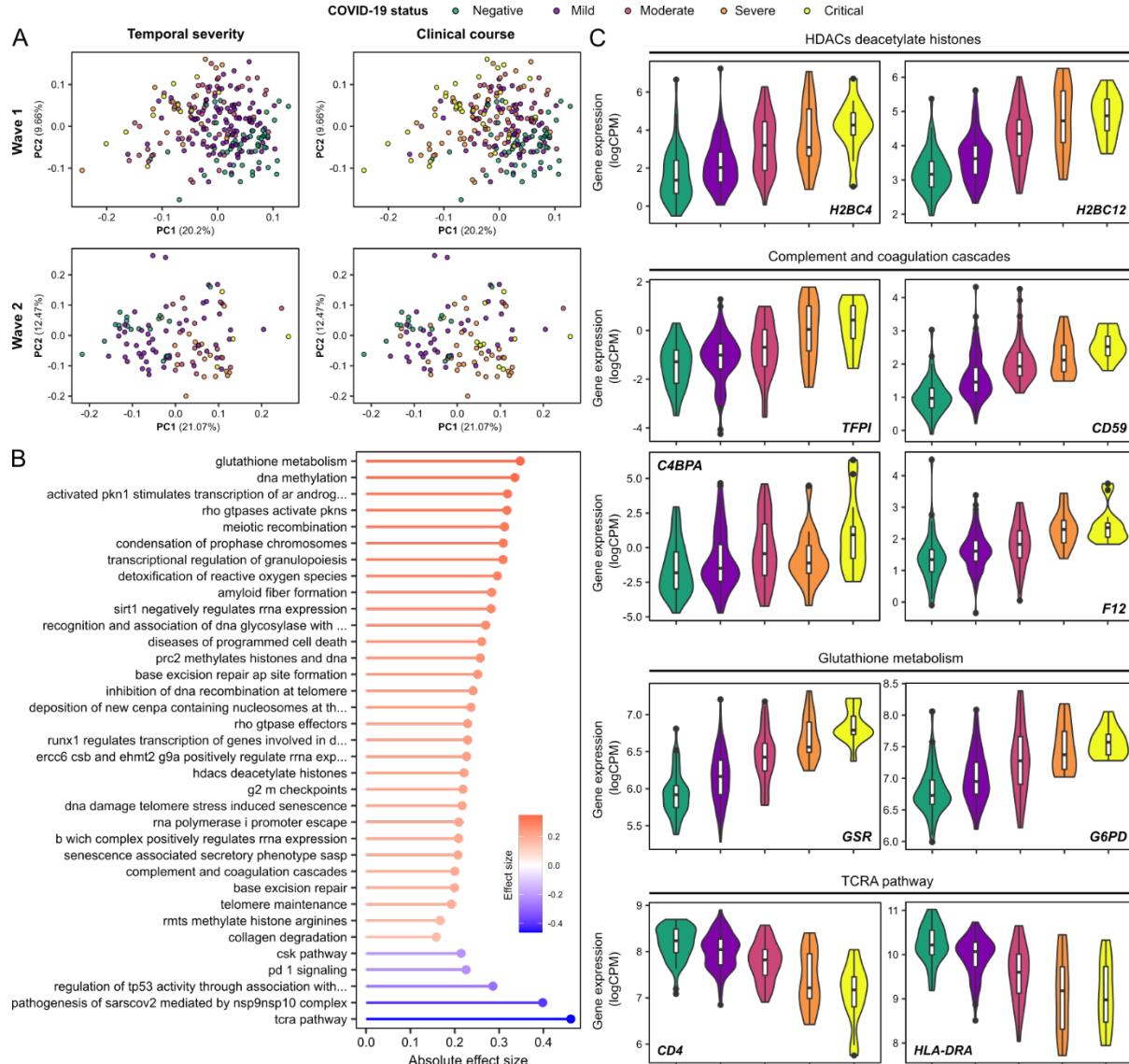
1059
 1060
 1061
 1062
 1063
 1064
 1065

Figure 1: Study design and cohort summary. **A)** Graphical summary of the patient cohorts, sampling, and major analyses. **B/C)** For each cohort, the timing of the serial blood sampling is shown by triangles and the temporal COVID-19 severity by coloured bars. Three patients were hospitalised prior to COVID-19 diagnosis in the Wave 1 cohort. Three of the four patients in the Wave 2 cohort with fatal outcomes died >30 days from first positive swab.



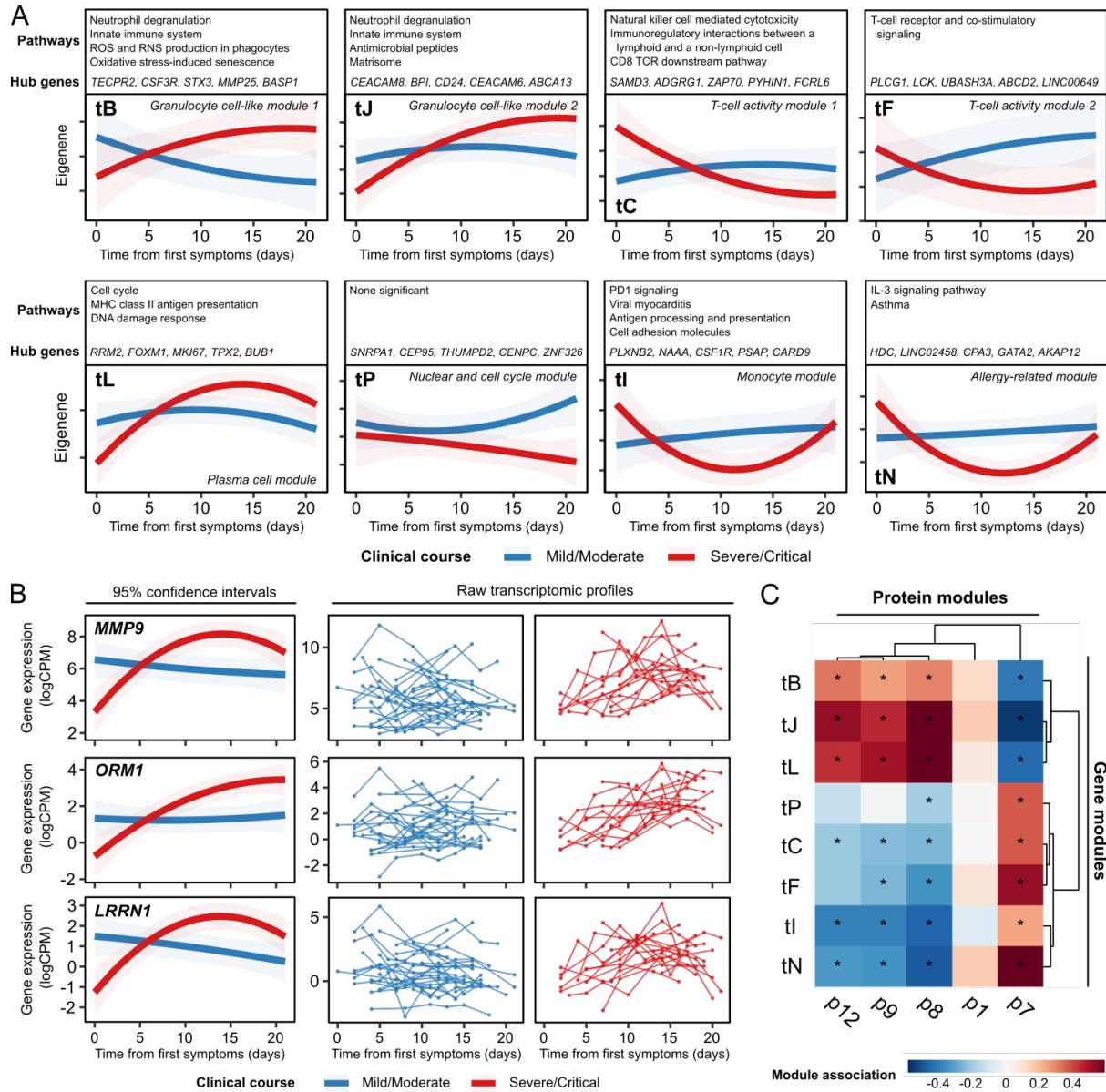
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078

Figure 2: Signatures of COVID-19 in ESKD. **A)** PCA of the PBMC transcriptome (left) and plasma proteome (right). Each point represents a sample and is coloured by COVID-19 status. **B)** Paired violin plots showing intra-individual comparisons of pre-infection and most severe sample (Wave 2 cohort) during COVID-19 for selected genes. Grey lines link each individual's pre-infection and infection samples. All genes shown were significantly differentially expressed (1% FDR) in both cohorts. Genes are grouped by membership to pathways that were significantly enriched (1% FDR) in GSVA. **C)** The 30 protein pathway enrichment terms with the greatest RRA scores (indicating consistent dysregulation in both the Wave 1 and Wave 2 proteomic datasets), ordered by effect size. All pathway terms shown were significantly enriched in the individual cohort analyses (1% FDR). Red= up-regulated in COVID-19 versus controls; blue= down-regulated. **D)** As for **B**, but displaying selected plasma proteins (significant at 1% FDR).



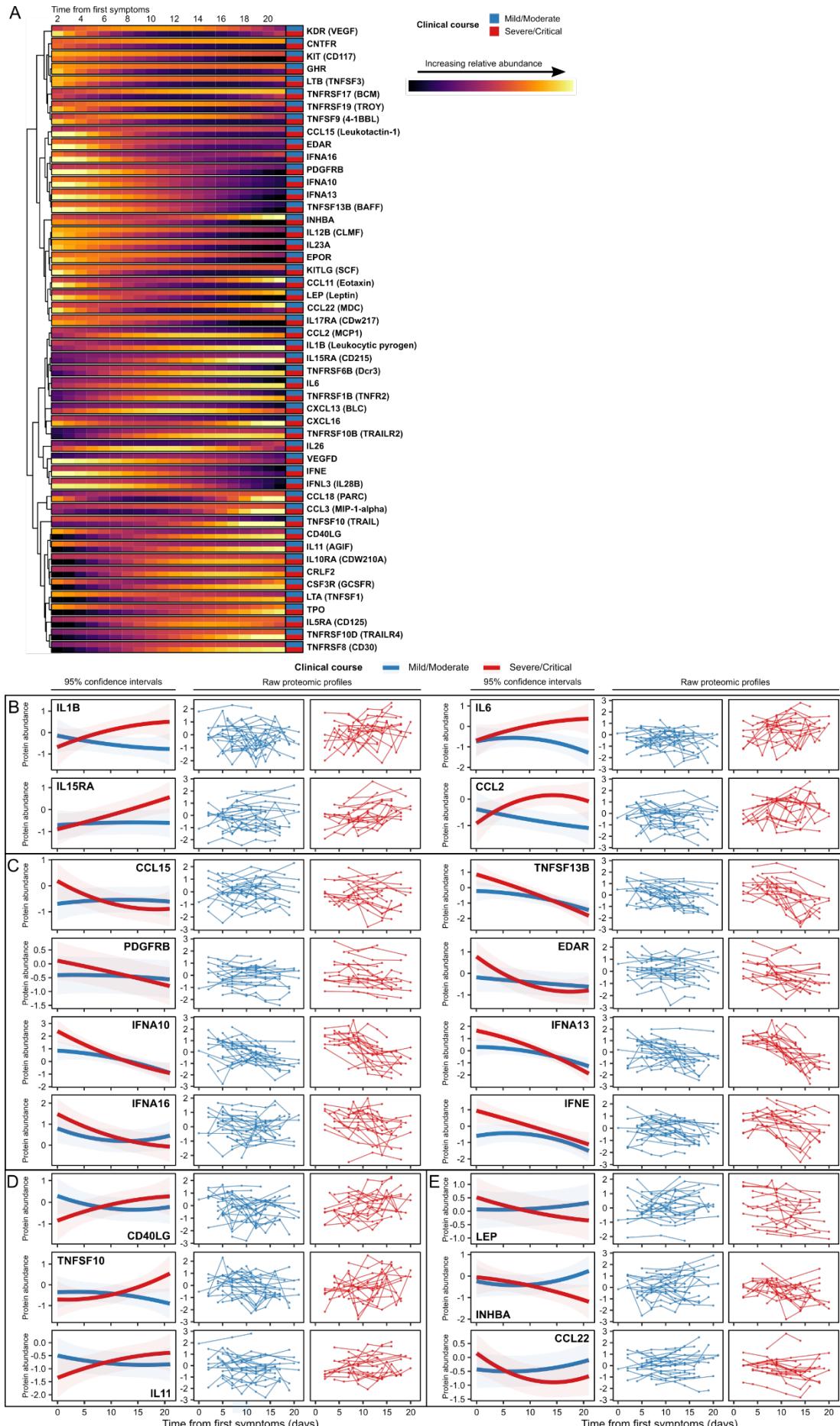
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089

Figure 3: Association of the PBMC transcriptome and COVID-19 severity. **A)** PCA of the PBMC transcriptome. Each point represents a sample and is coloured by contemporaneous COVID-19 WHO severity (left) and overall clinical course (right). **B)** The 30 GSVA transcriptomic pathway enrichment terms with the greatest RRA scores. All were significantly enriched in both Wave 1 and 2 cohorts (1% FDR). Terms are ordered and coloured by their effect size. Red= up-regulated in more severe COVID-19; blue= down-regulated. **C)** Violin plots show gene expression values (Wave 1 cohort) stratified by COVID-19 status and severity (at time of sample) for selected genes. All genes shown were significantly associated (1% FDR) with severity in both the Wave 1 and 2 cohorts. Genes are grouped by membership to pathways that were significantly enriched (1% FDR) in GSVA.



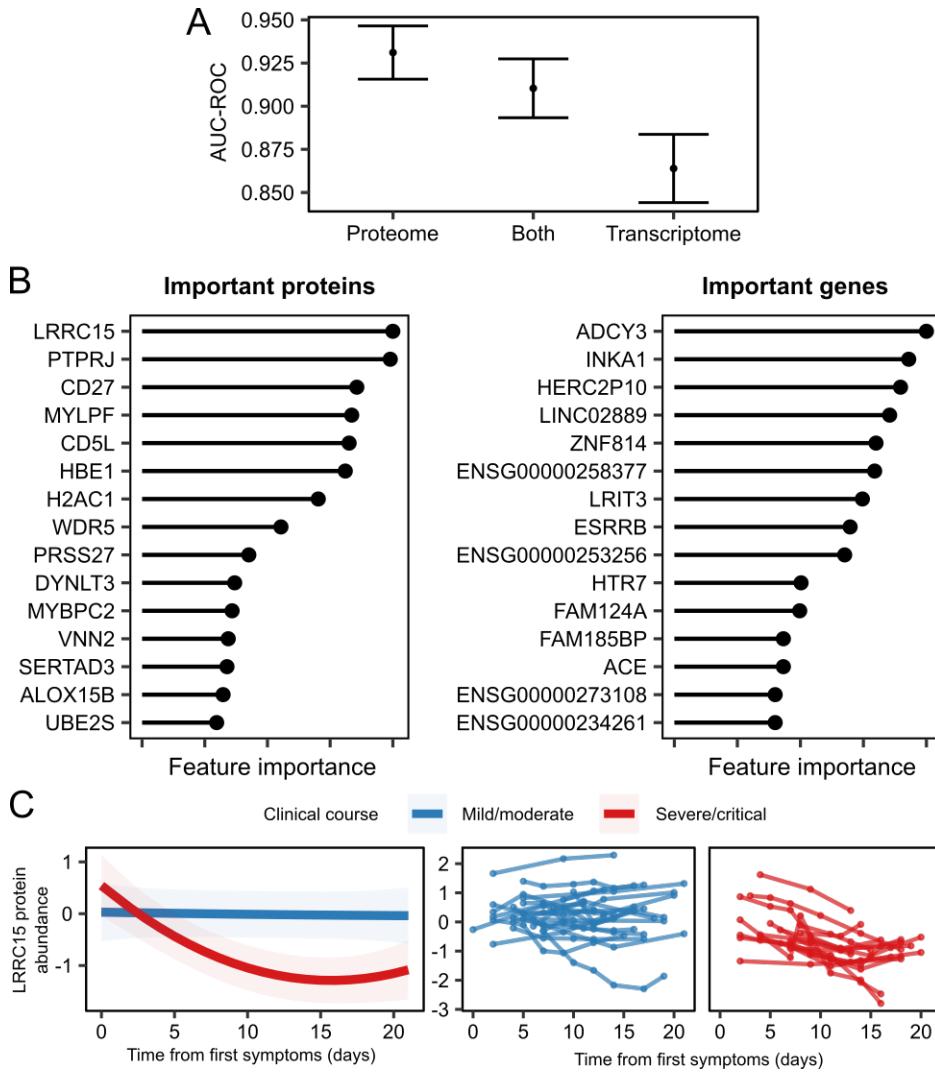
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099

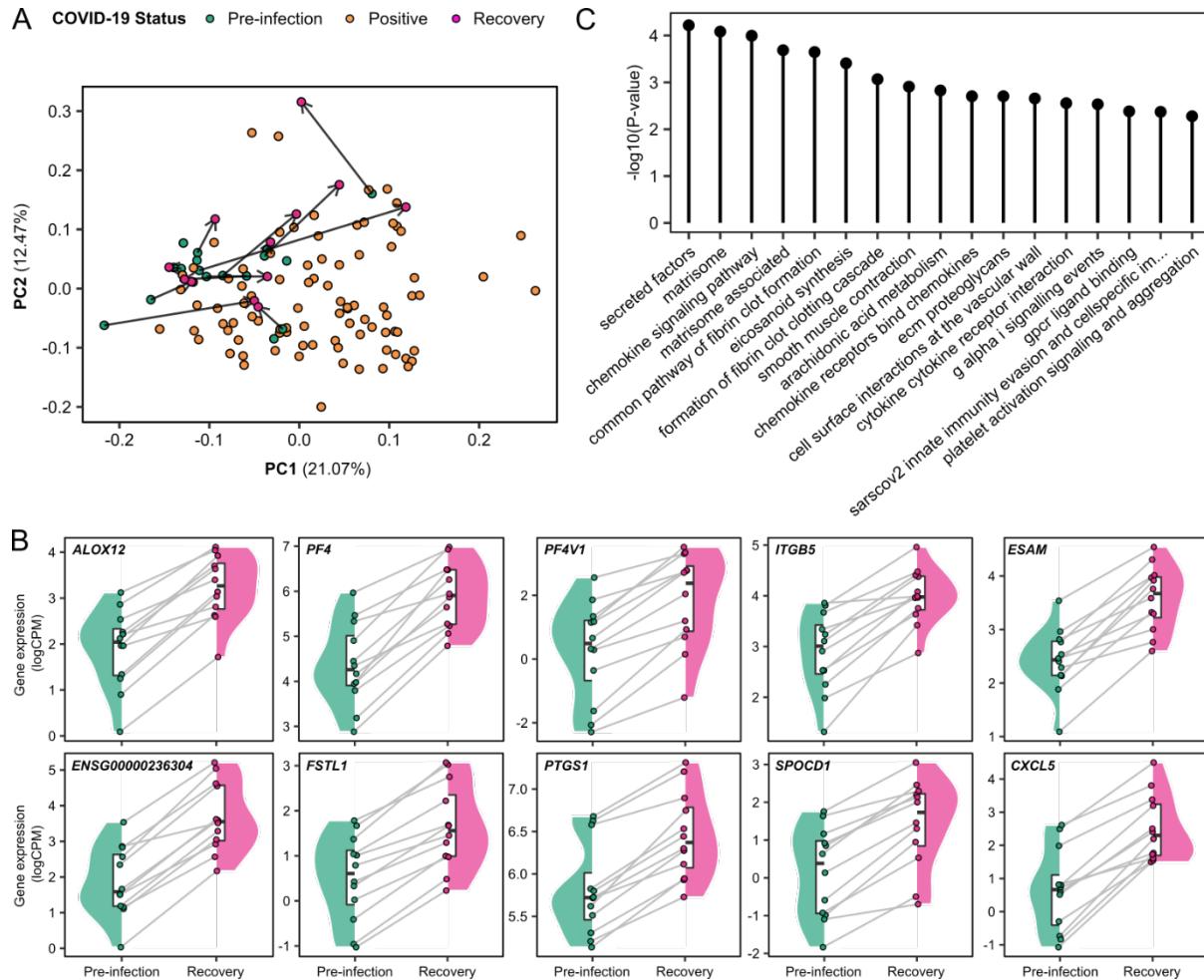
Figure 4: Longitudinal profiles of transcriptomic modules. A) The longitudinal profiles of significant (5% FDR) gene modules, stratified by clinical course. Lines represent model estimates and shaded areas represent 95% confidence intervals. **B)** Modelled longitudinal profiles of genes within module B with the most significant TxCC interaction effects. Left: model estimates and 95% confidence intervals. Right: individual-level data. **C)** Heatmap displaying associations (LMM) between transcriptomic and proteomic modules (right). Red= positive correlation, blue= negative correlation. Significant associations (5% FDR) are represented by an asterisk.



1101 **Figure 5: Dynamic temporal changes in circulating cytokines and receptors vary between severe**
1102 **and mild COVID-19. A)** Heatmap displaying proteins with a significantly different temporal profile in
1103 mild vs severe disease (TxC, LMM, FDR <0.05). Colour indicates model estimates over time, stratified
1104 by patient group. Proteins are clustered based on the temporal profile of the discordance between
1105 mild/moderate and severe/critical disease. Proteins are annotated using gene symbols, with alternative
1106 common protein identifiers in parentheses. **B-E)** Examples of proteins with differing patterns of
1107 discordance over time in severe/critical versus mild/moderate patients.

1108





1119
1120
1121 **Figure 7: Persistent dysregulation of immune cell gene expression two months following COVID-**
1122 **19. A)** PCA of the Wave 2 PBMC transcriptomic data, including pre-infection, infection and recovery
1123 samples (taken 2 months after the acute illness). Each point represents a sample. Arrows link recovery
1124 samples to the pre-infection sample from the same individual. **B)** Paired violin plots for differentially
1125 expressed genes in recovery versus pre-infection samples. Grey lines link each individual's pre-
1126 infection sample to their recovery sample. **C)** All significantly enriched (5% FDR) pathway terms for the
1127 differentially expressed genes in recovery versus pre-infection samples.
1128

1129 **TABLES**

1130

1131 **Table 1: Genes that do not return to baseline 2 months after recovery from COVID-19.**

1132 Genes that are significantly differentially expressed (1% FDR) in recovery versus pre-infection
1133 samples. Estimate represents the average difference in log CPM (counts per million). A
1134 complete table including all genes tested can be found in **Supplementary Table**
1135 **recovery_results.**

1136

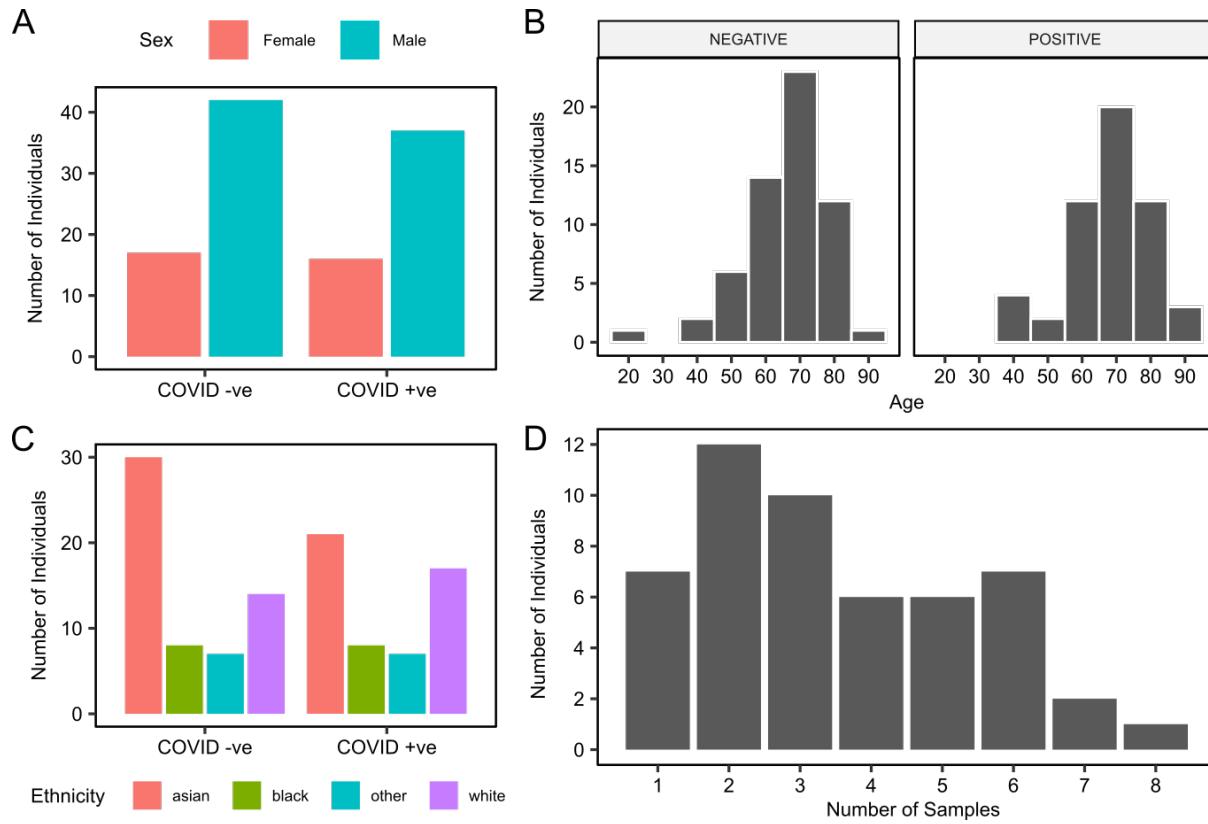
Gene ID	Gene name	Estimate (recovery - pre-infection)	P-value	Adjusted P-value
ENSG00000236304	IncRNA	1.84	5.37E-08	9.20E-04
FSTL1	Follistatin Like 1	1.17	1.60E-07	1.37E-03
PTGS1	Prostaglandin-Endoperoxide Synthase 1	0.64	3.69E-07	2.11E-03
SPOCD1	SPOC Domain Containing 1	1.31	6.41E-07	2.53E-03
CXCL5	C-X-C Motif Chemokine Ligand 5	1.84	7.39E-07	2.53E-03
ALOX12	Arachidonate 12-Lipoxygenase, 12S Type	1.33	1.24E-06	3.53E-03
PF4	Platelet Factor 4	1.53	1.71E-06	3.91E-03
ESAM	Endothelial Cell Adhesion Molecule	1.18	1.83E-06	3.91E-03
MT-RNR1	Mitochondrially Encoded 12S RRNA	1.00	2.61E-06	4.13E-03
MMD	Monocyte To Macrophage Differentiation Associated	0.77	2.80E-06	4.13E-03
ENSG00000240093	IncRNA	-0.71	2.82E-06	4.13E-03
MTURN	Maturin, Neural Progenitor Differentiation Regulator Homolog	0.54	3.11E-06	4.13E-03
GNG11	G Protein Subunit Gamma 11	1.53	3.13E-06	4.13E-03
CAVIN2	Caveolae Associated Protein 2	1.15	4.51E-06	5.52E-03
DOK6	Docking Protein 6	1.59	5.00E-06	5.71E-03
LINC00989	IncRNA	1.25	5.76E-06	5.81E-03
SPARC	Secreted Protein Acidic And Cysteine Rich	1.65	6.07E-06	5.81E-03
PF4V1	Platelet Factor 4 Variant 1	1.60	6.11E-06	5.81E-03
ABLIM3	Actin Binding LIM Protein Family Member 3	1.41	6.88E-06	6.07E-03
MFAP3L	Microfibril Associated Protein 3 Like	0.74	7.09E-06	6.07E-03
CALD1	Caldesmon 1	1.78	7.79E-06	6.23E-03
ITGB5	Integrin Subunit Beta 5	1.10	8.00E-06	6.23E-03
LINC01750	IncRNA	2.00	1.13E-05	8.22E-03
PCSK6	Proprotein Convertase Subtilisin/Kexin Type 6	1.22	1.15E-05	8.22E-03
PVALB	Parvalbumin	2.04	1.46E-05	9.97E-03

1137

1138 **Supplementary Figures**

1139

1140



1141

1142

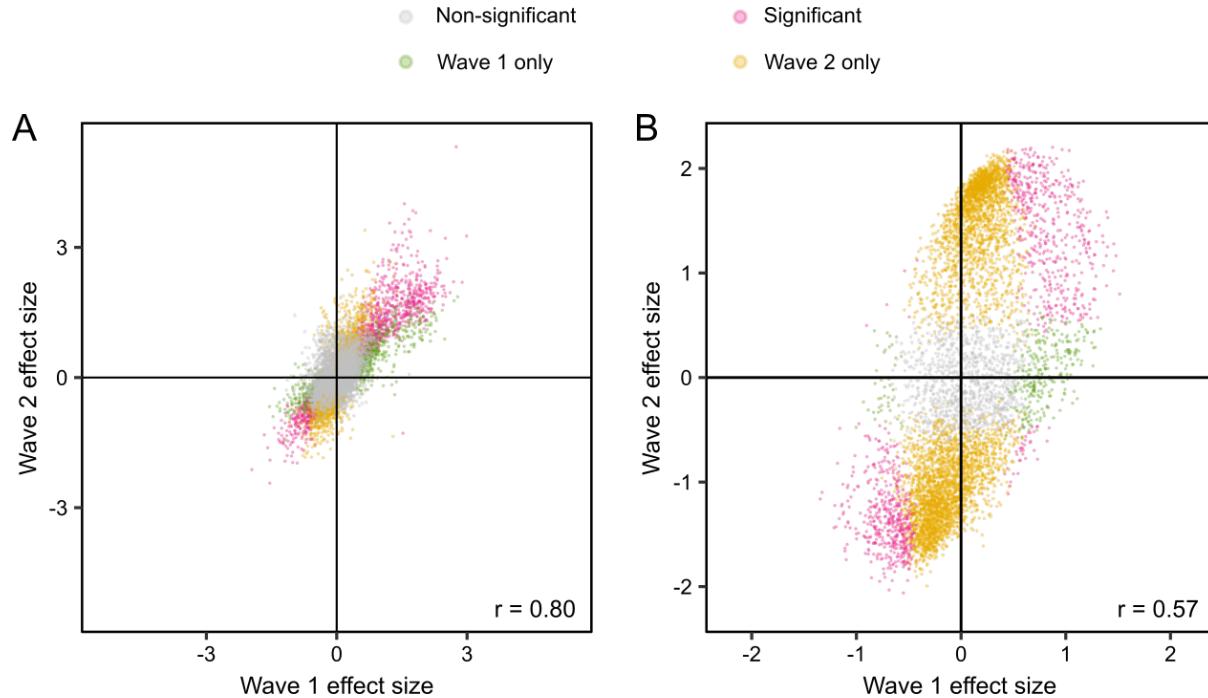
1143 **Supplementary Figure 1: Characteristics of the Wave 1 cohort.** The number of COVID-19 positive and negative patients, stratified by A), sex B), age C) ethnicity. D) Number of serial PBMC samples with post-QC RNA-seq data available for COVID-19 patients.

1144

1145

1146

1147



1148

1149

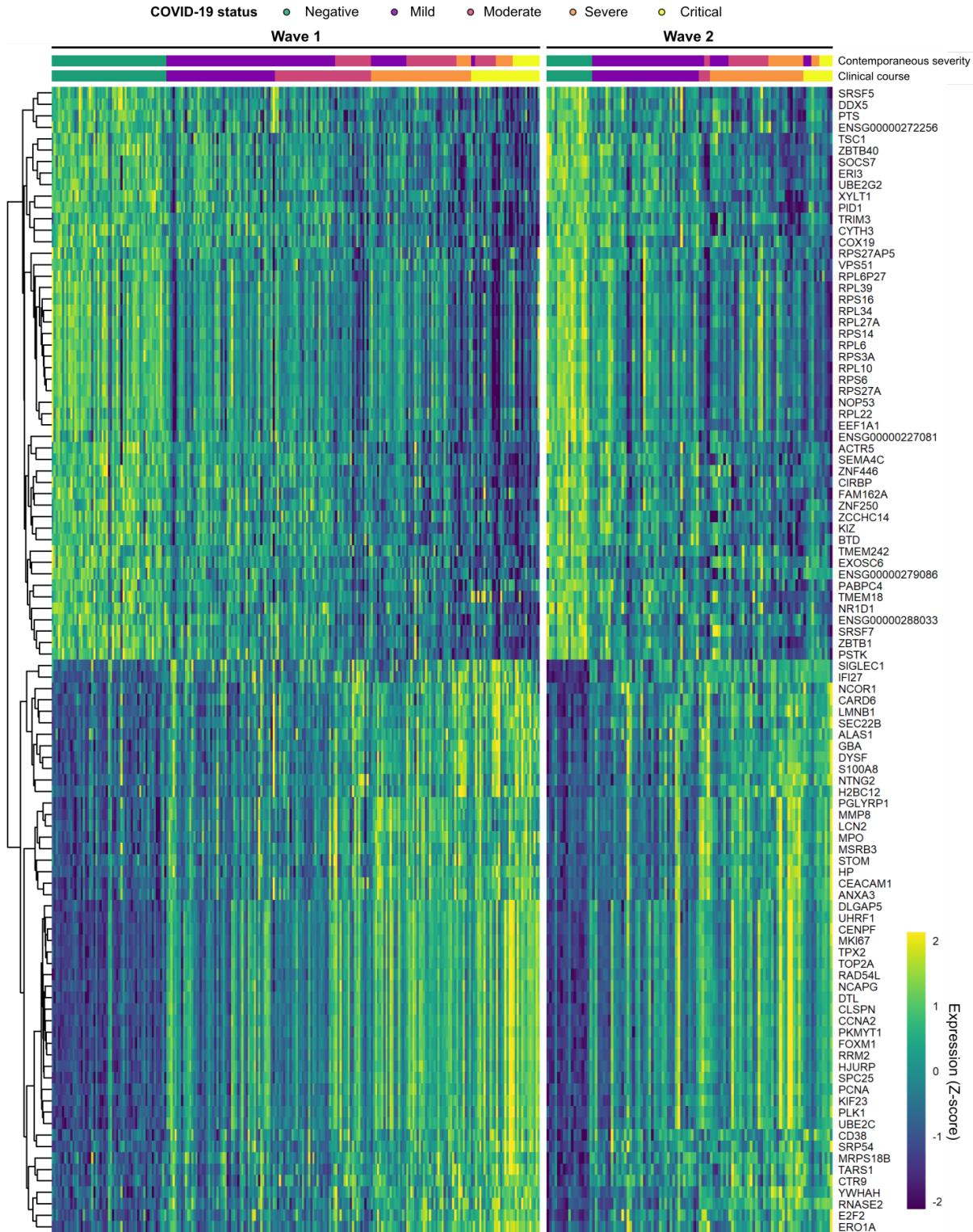
1150

1151

1152

1153

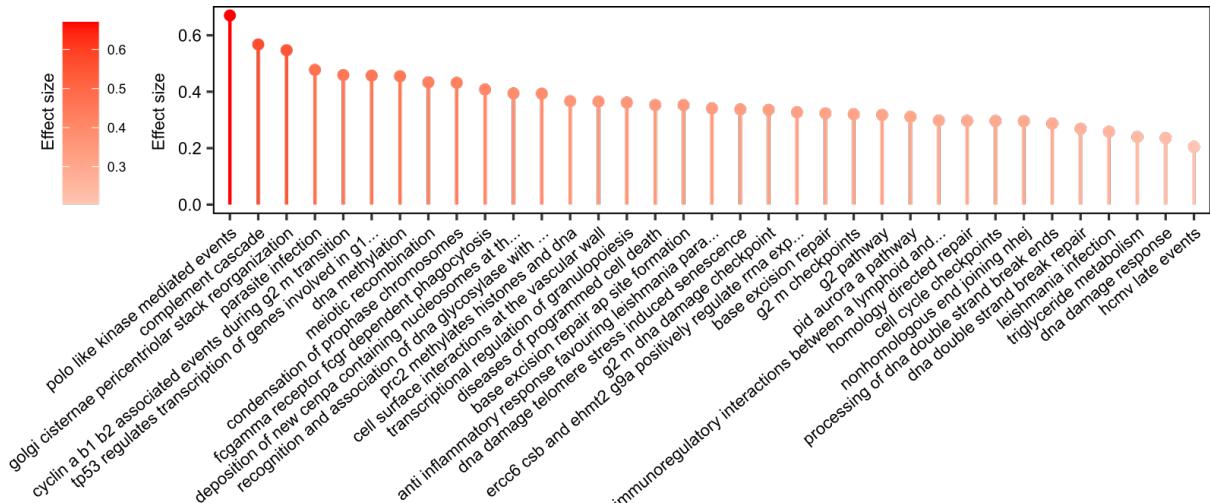
Supplementary Figure 2: A) Comparison of effect sizes (coefficients from LMM) for the Wave 1 and Wave 2 infected vs non-infected differential expression analyses for the transcriptome. Each point is a gene, coloured according to its significance in the Wave 1 and 2 analyses. r = Pearson's correlation coefficient. **B)** As A, for the proteome.



1154

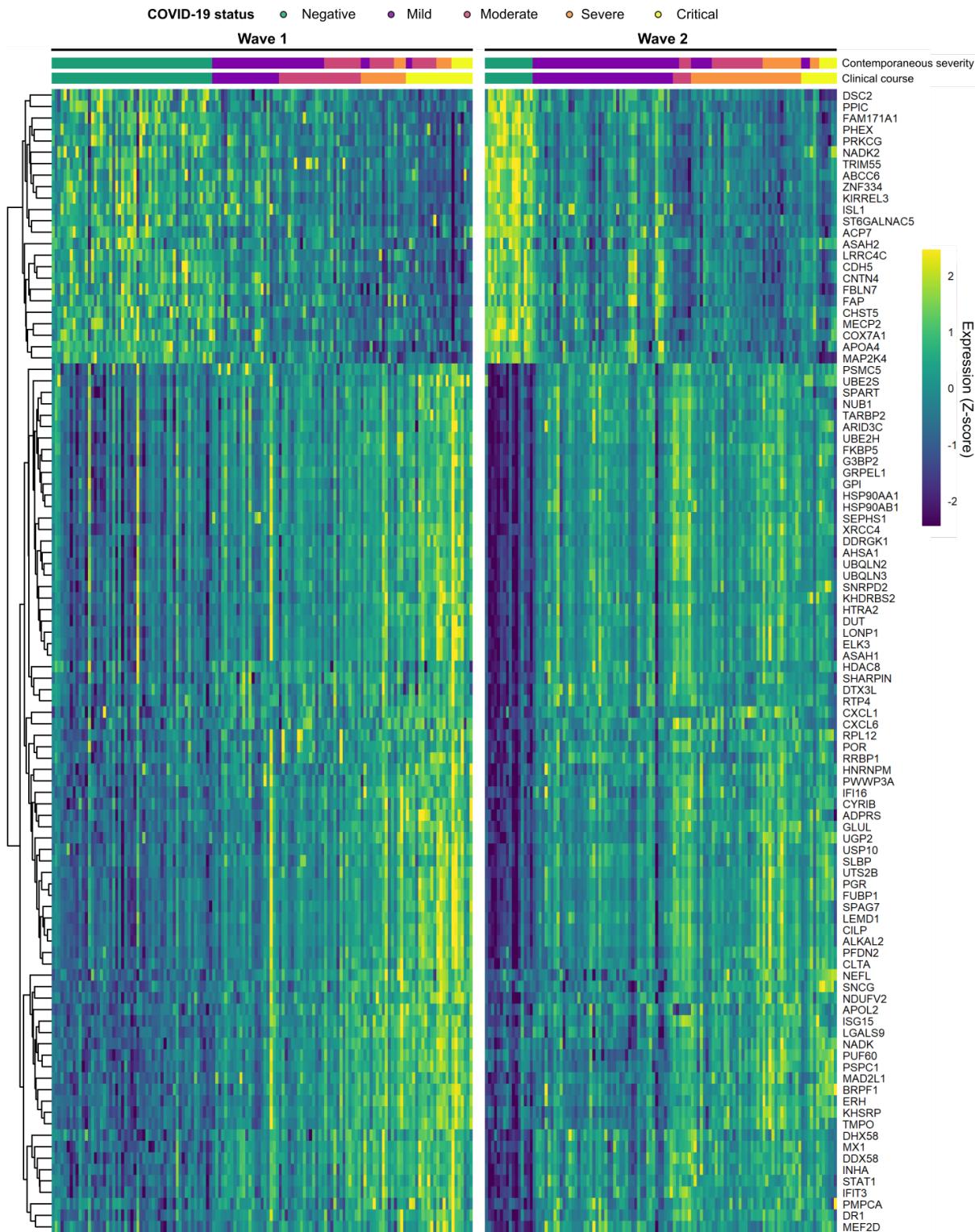
Supplementary Figure 3: Genes differentially expressed in COVID-19. Heatmap for the 100 transcriptomic features most significantly differentially expressed between COVID-19 cases and controls, according to robust rank aggregation (RRA). These genes were significant in both cohorts (1% FDR). Columns are ordered by COVID-19 status and severity. For the Wave 1 heatmap, genes are ordered by hierarchical clustering; the Wave 2 heatmap is ordered to match this. Each feature was scaled and centred separately in each dataset.

1161



1162
1163
1164
1165
1166
1167
1168
1169
1170

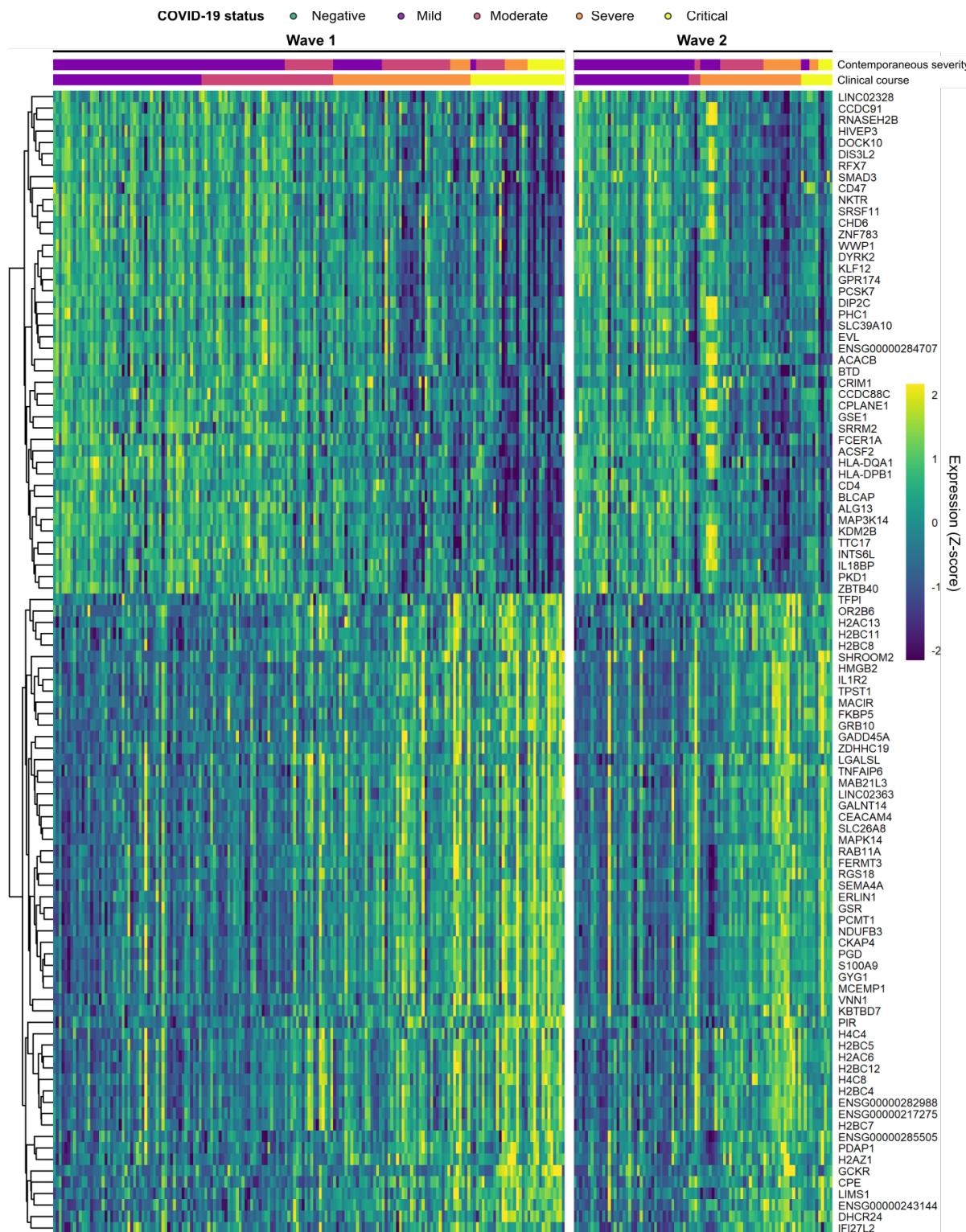
Supplementary Figure 4: Transcriptomic enrichment analysis comparing infected and non-infected patients. The 30 GSVA gene pathway enrichment terms with the greatest RRA scores (indicating consistent dysregulation in both the Wave 1 and Wave 2 transcriptomic datasets). All pathway terms shown were significantly enriched in the individual cohort analyses (1% FDR). Terms are ordered and coloured by their effect size. All terms were up-regulated in COVID-19 cases, so they are all coloured red.



1171

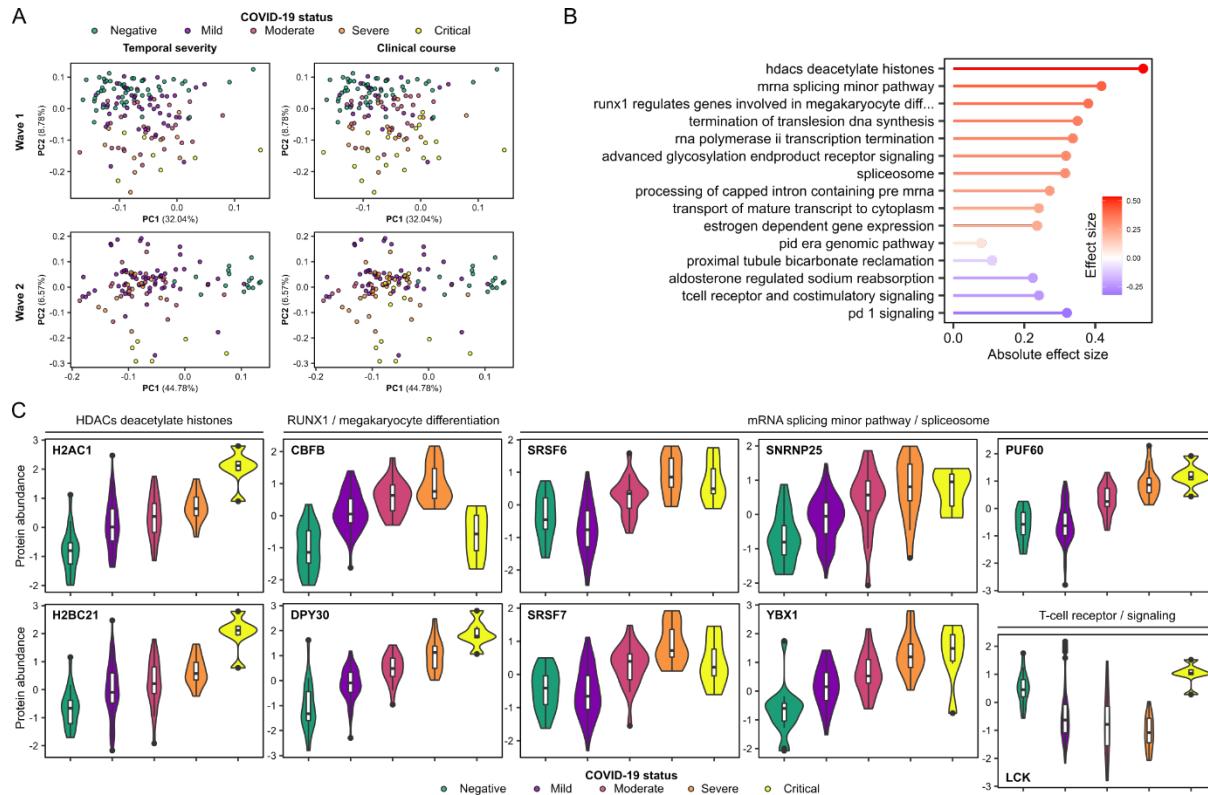
1172 **Supplementary Figure 5: Most significant differentially abundant plasma proteins in**
1173 **COVID-19 positive versus negative ESKD samples.** Heatmap for the 100 proteomic
1174 features most significantly differentially abundant between COVID-19 cases and controls,
1175 according to robust rank aggregation (RRA). These proteins were significant in both cohorts
1176 (1% FDR, LMM). Columns are ordered by COVID-19 status and severity. For the Wave 1
1177 heatmap, proteins are ordered by hierarchical clustering; the Wave 2 heatmap is ordered to
1178 match this. Each feature was scaled and centred separately in each dataset.

1179



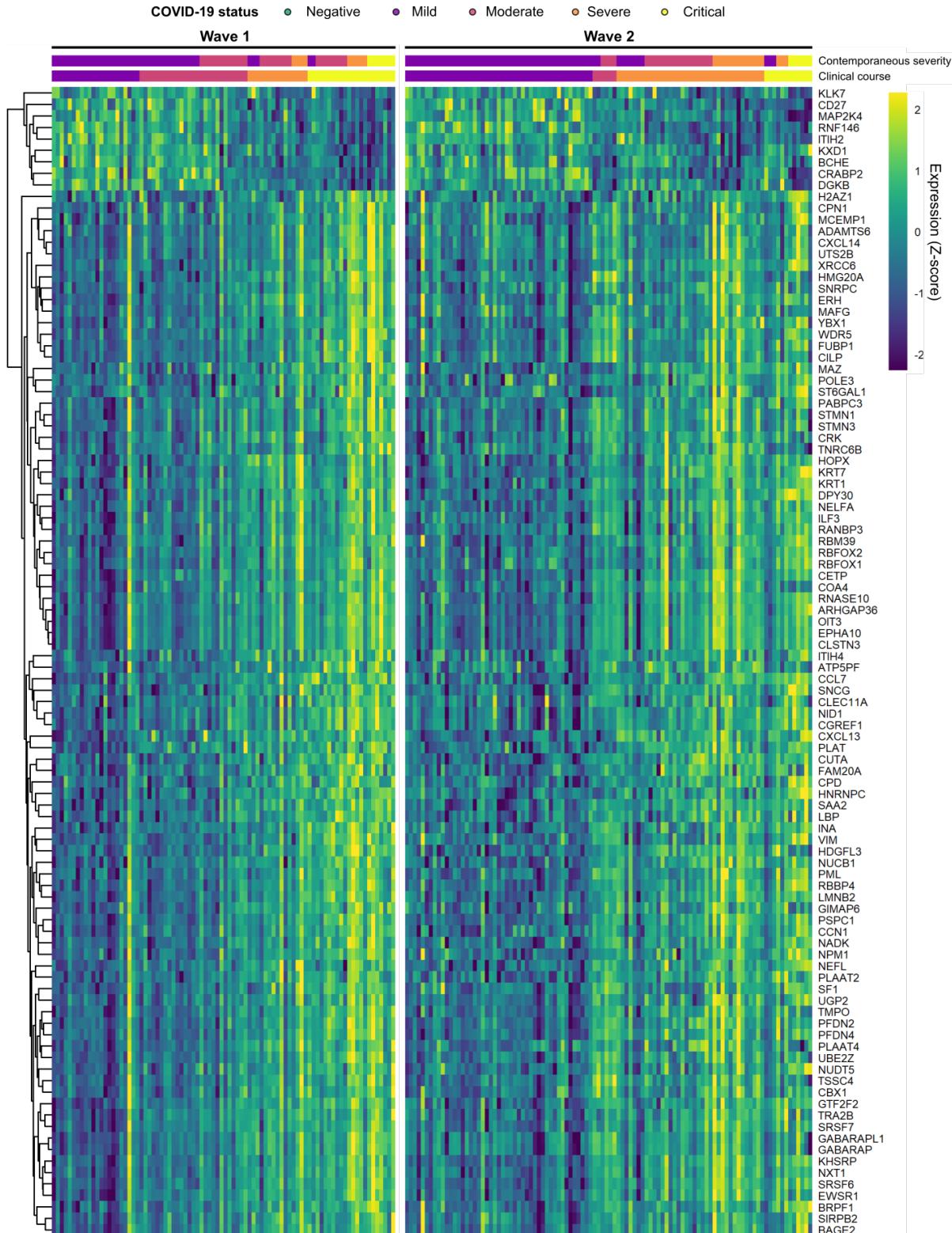
1180

1181 **Supplementary Figure 6: Genes with the most significant association with COVID-19**
1182 **severity.** Heatmap for the 100 transcriptomic features most significantly associated with
1183 contemporaneous severity, according to robust rank aggregation (RRA). These genes were
1184 significant in both cohorts (1% FDR). Columns are ordered by COVID-19 status and severity.
1185 For the Wave 1 heatmap, genes are ordered by hierarchical clustering; the Wave 2 heatmap
1186 is ordered to match this. Each feature was scaled and centred separately in each dataset.



1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200

Supplementary Figure 7: Plasma proteomic associations with COVID-19 severity. A) PCA of the proteome. Each point represents a sample and is coloured by contemporaneous COVID-19 severity (left) and overall clinical course of the patient (right). **B)** The GSVA protein pathway enrichment terms significantly associated (1% FDR) with contemporaneous severity in both the Wave 1 and Wave 2 proteomic datasets. All pathway terms shown were significantly enriched in the individual cohort analyses (1% FDR). Terms are ordered and coloured by their effect size. Terms up-regulated in more severe COVID-19 are coloured red while down-regulated terms are blue. **C)** Violin plots show gene expression values (Wave 2 cohort) stratified by COVID-19 status and severity (at time of sample) for selected genes. All genes shown were significantly associated (1% FDR) with contemporaneous severity in the Wave 2 cohort.

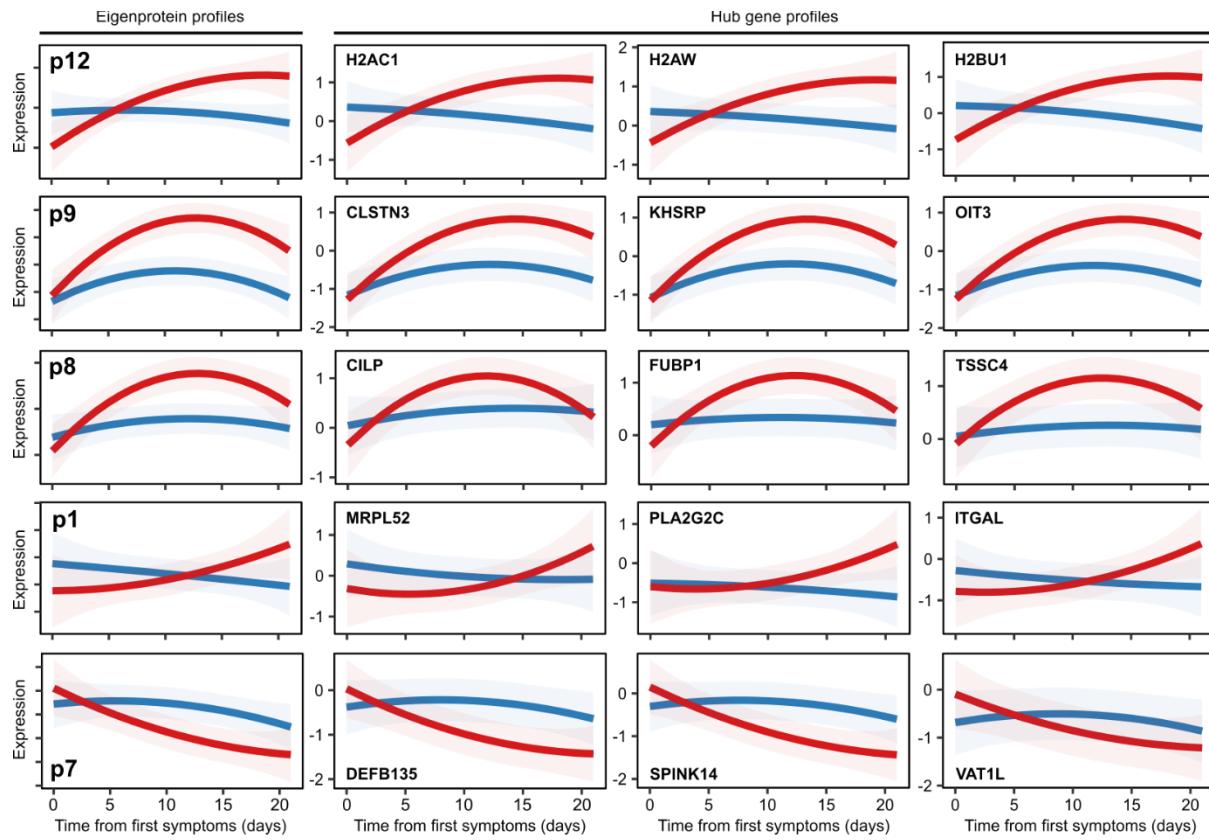


1201

1202 **Supplementary Figure 8: Selected proteins associated with COVID-19 severity.** Heatmap
1203 for the 100 proteomic features most significantly associated with contemporaneous WHO
1204 severity, according to robust rank aggregation (RRA). These proteins were significant in both
1205 cohorts (1% FDR). Columns are ordered by COVID-19 status and severity. For the Wave 1
1206 heatmap, proteins are ordered by hierarchical clustering; the Wave 2 heatmap is ordered to
1207 match this. Each feature was scaled and centred separately in each dataset.

1208

1209



1210

1211

Supplementary Figure 9: Longitudinal profiles of protein modules and hub proteins.

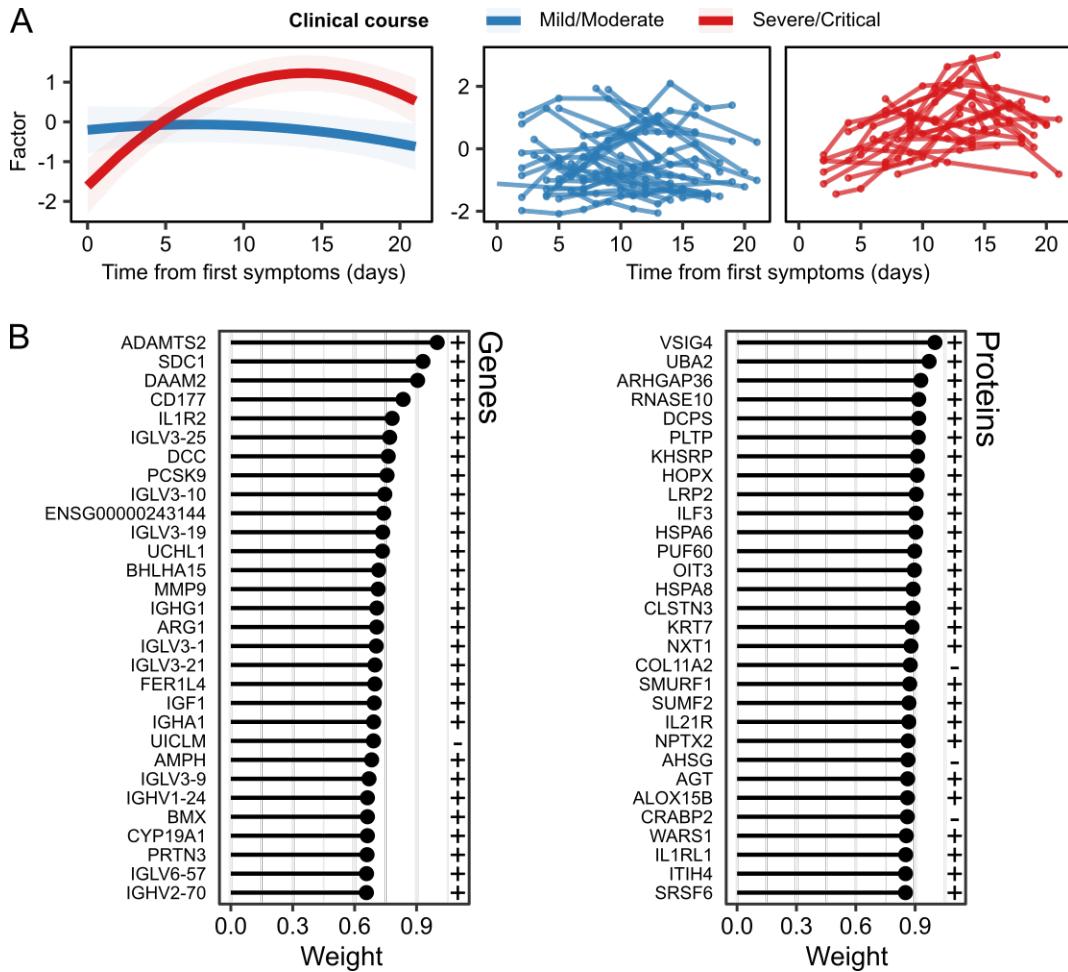
1212

1213

1214

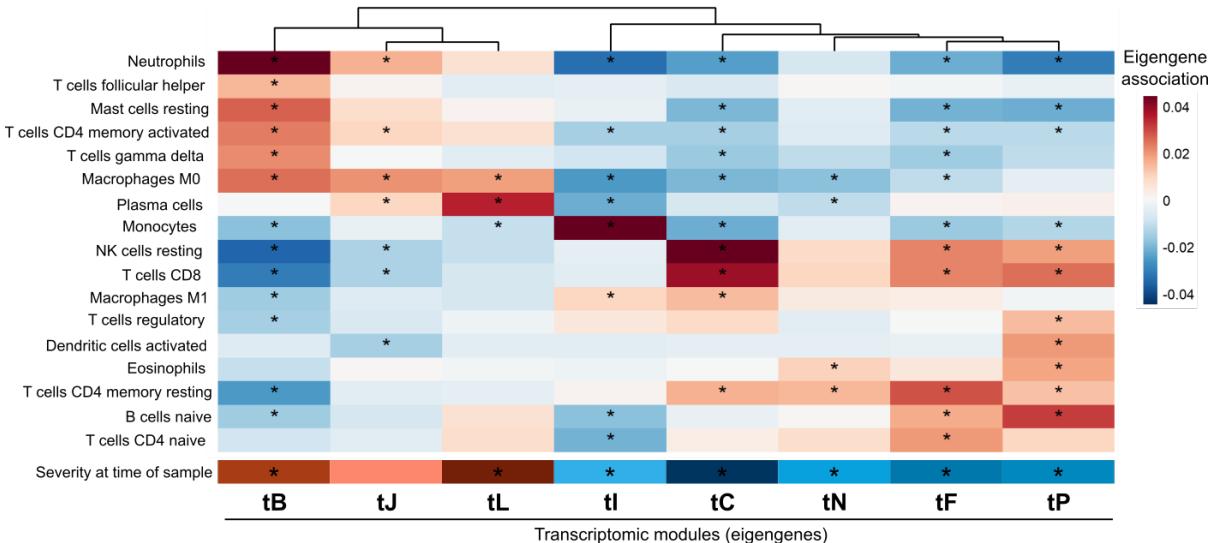
1215

Modelled longitudinal profiles of protein modules with significant (5% FDR) TxCc interactions (left) and the profiles of their top three most central hub proteins (right). Red lines= patients with a severe/critical clinical course; blue lines= mild/moderate clinical course. Shaded areas represent 95% confidence intervals.



1216
1217

1218 **Supplementary Figure 10: A multi-modal factor representing COVID-19 severity. A)** The
1219 longitudinal profile of a factor identified by MEFISTO with a significant (5% FDR) TxC
1220 C interaction. Model estimates and 95% confidence interval (left) and the trajectory of the factor
1221 for each individual (right). **B)** The weights of individual genes and proteins with respect to the
1222 factor identified by MEFISTO. A plus sign indicates molecules positively associated with the
1223 factor while molecules negatively associated with the factor have a minus sign.
1224

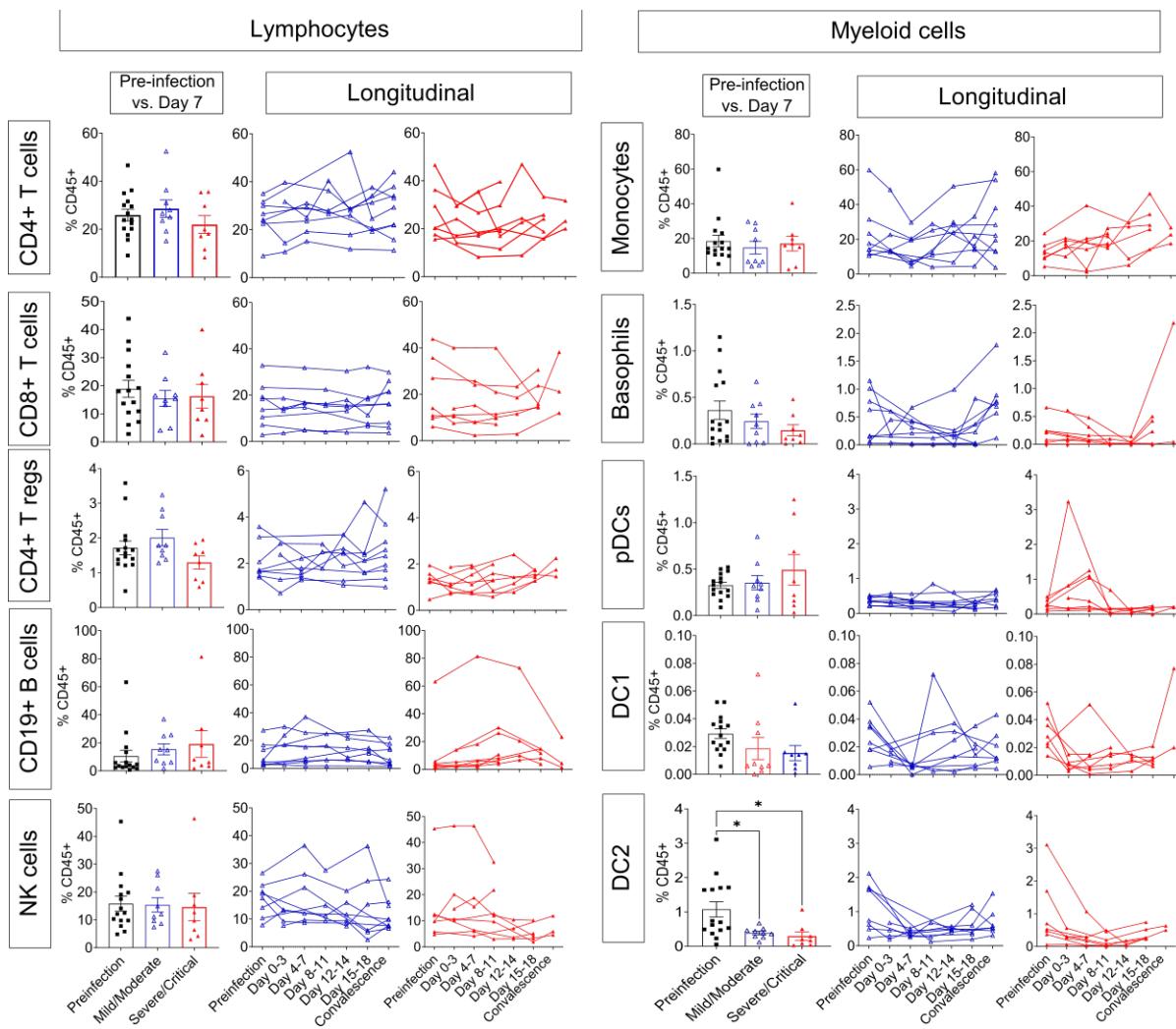


1225

1226

1227 **Supplementary Figure 11: Association of transcriptomic modules with imputed cell**
 1228 **abundances.** Association of modules (columns) with cell abundances (rows) imputed using
 1229 CIBERSORTx. Analysis with LMM. All modules shown have significant TxC interaction
 1230 effects. The bottom row shows the association with severity score at time of sample. Asterisks
 1231 represent significant (5% FDR) associations.
 1232

1233
1234

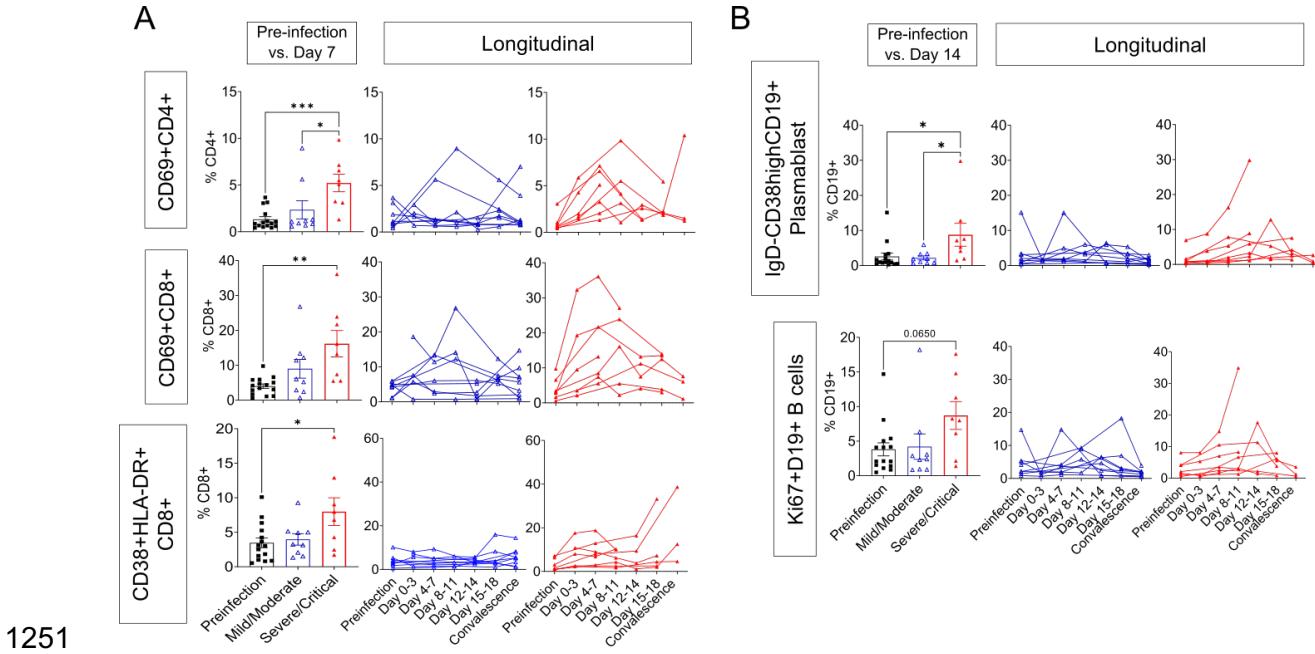


1235
1236 **Supplementary Figure 12: Proportion of major lymphocytes and myeloid cells in**
1237 **PBMCs (Wave 2 cohort).**

1238 Panels marked "Pre-infection vs. Day 7": frequencies of major immune subsets gated on
1239 CD45+ leukocytes in samples taken prior to infection (Pre-infection n=15) and at closest
1240 sample to day 7 from symptom onset (Mild/moderate n=9; Severe/critical n=8). Data presented
1241 as mean (box) \pm S.E.M (whiskers). Each symbol represents an individual.

1242 Panels marked "Longitudinal": frequencies of major immune subsets before, during and after
1243 infection as a line plot. Each line depicts one individual. Blue = patients with a mild/moderate
1244 clinical course; sample numbers: Pre-infection n=8, Day 0-3 n=4, Day 4-7 n=6, Day 8-11 n=4,
1245 Day 12-14 n=6, Day 15-18 n=5, Convalescence n=9. Red= severe/critical clinical course;
1246 sample numbers: Pre-infection n=7, Day 0-3 n=5, Day 4-7 n=6, Day 8-11 n=6, Day 12-14 n=4,
1247 Day 15-18 n=5, Convalescence n=3. One-way ANOVA with Dunnet's for multiple comparisons
1248 correction was used for statistical analysis. Only significant differences are indicated. *p <0.05.
1249 NK= natural killer; pDCs= plasmacytoid dendritic cells. DC= dendritic cells.

1250



1251

1252 **Supplementary Figure 13: Increased activated T cells, plasmablast and proliferating B**
1253 **cells is associated with disease severity in COVID-19 ESKD patients.**

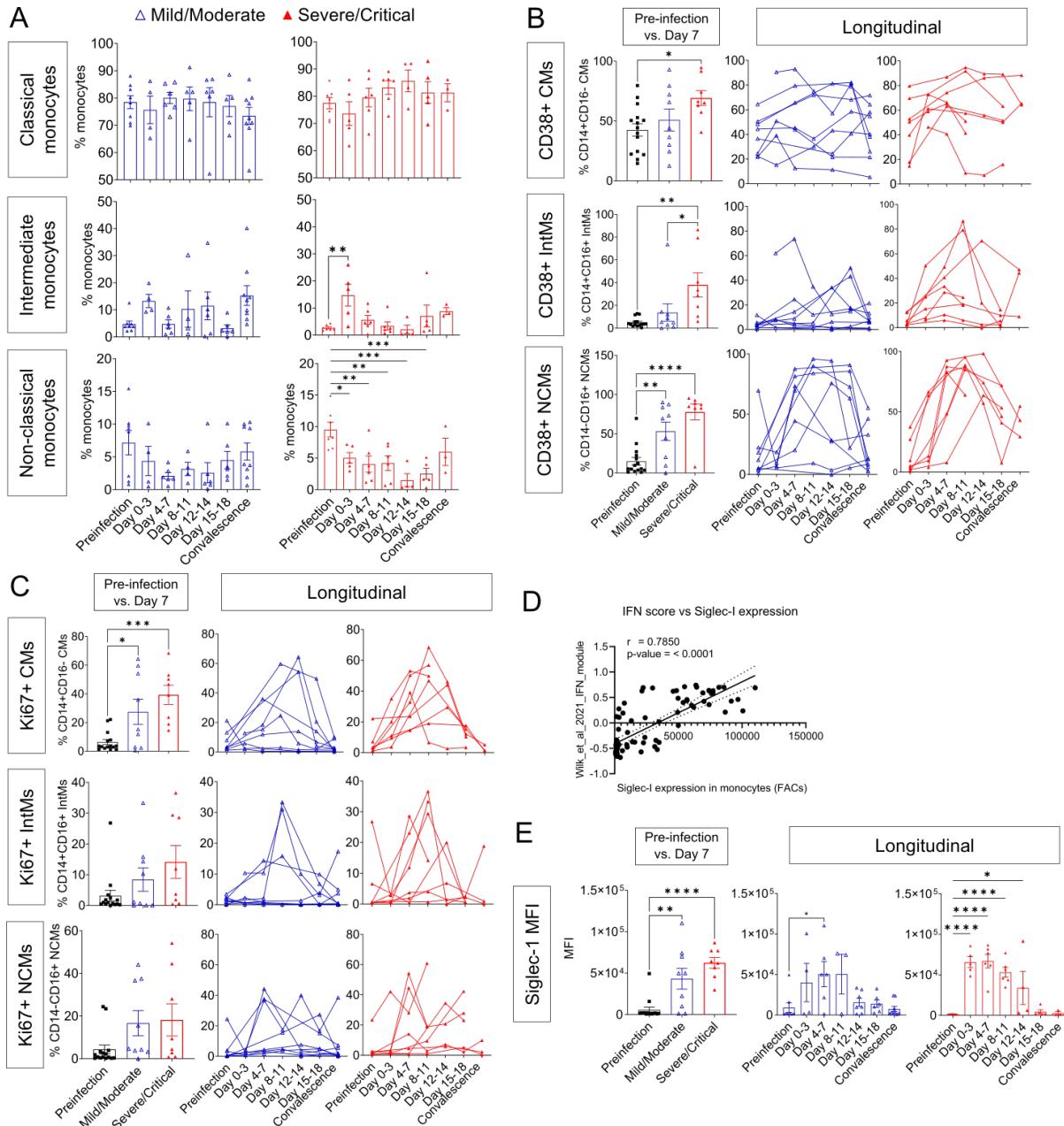
1254 **A) Percentage of cells expressing activation markers CD69, CD38 and HLA-DR gated**
1255 **on CD4+ T cells or CD8+ T cells.** Panel marked “Pre-infection vs. Day 7”: data plotted for
1256 samples taken prior to infection (Pre-infection n=15) and for closest sample to day 7 from
1257 symptom onset (Mild/Moderate n=9; Severe/Critical n=8).

1258 **B) Frequencies of IgD-CD38^{hi}CD19+ plasmablast and Ki67+ cells gated on CD19+ B**
1259 **cells.** Panel marked “Pre-infection vs. Day 14”: data plotted for samples taken prior to infection
1260 (Pre-infection n=15) and for closest sample to day 14 from symptom onset (Mild/Moderate
1261 n=9; Severe/Critical n=8).

1262 (A-B) Left panels: data presented as mean (bar) \pm S.E.M (whiskers). Each symbol represents
1263 an individual. Right panels: frequencies of cells over the course of infection as a line plot. Each
1264 continuous line depicts one individual. Mild/moderate patients in blue (Pre-infection n=8, Day
1265 0-3 n=4, Day 4-7 n=6, Day 8-11 n=4, Day 12-14 n=6, Day 15-18 n=5, Convalescence n=9)
1266 and severe/critical in red (Pre-infection n=7, Day 0-3 n=5, Day 4-7 n=6, Day 8-11 n=6, Day
1267 12-14 n=4, Day 15-18 n=5, Convalescence n=3). One-way ANOVA with Dunnet's for multiple
1268 comparisons correction used for statistical analysis. Only significant differences are indicated.
1269 *p <0.05; **p <0.01; ***p <0.001.

1270

1271



1272

1273 **Supplementary Figure 14: Increased activated and proliferating circulatory monocytes**
1274 **is associated with disease severity in COVID-19 ESKD patients.**

1275 **A)** Frequencies of monocyte subsets as defined with CD14+ and CD16+ expression gated on
1276 total monocytes. Percentage of cells expressing CD38 (**B**) and Ki67 (**C**) gated on
1277 CD14+CD16- CMs, CD14+CD16+ IntMs, CD14-CD16+ NCMs. **D)** Correlation (Pearson's r)
1278 of SIGLEC-1 protein expression by MFI to RNAseq-derived GSVA enrichment score for type
1279 I IFN signatures. **E)** Siglec-1 expression in MFI gated on total monocytes.

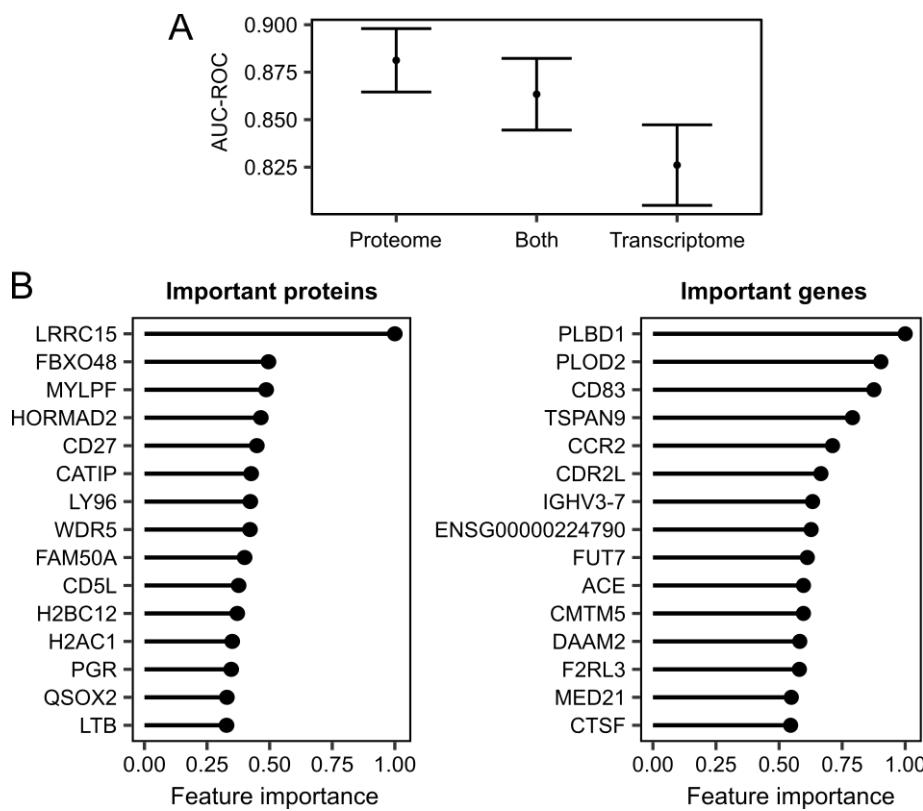
1280 (B, C, E): Data plotted for samples taken prior to infection (Pre-infection n=15) and for closest
1281 sample to day 7 from symptom onset (Mild/Moderate n=9; Severe/Critical n=8). Data
1282 presented as mean \pm S.E.M. Each symbol represents an individual. (A, B, C, E) Frequencies
1283 of cells over the course of infection as a line or bar plot. Mild/moderate patients in blue (Pre-
1284 infection n=8, Day 0-3 n=4, Day 4-7 n=6, Day 8-11 n=4, Day 12-14 n=6, Day 15-18 n=5,
1285 Convalescence n=9) and severe/critical in red (Pre-infection n=7, Day 0-3 n=5, Day 4-7 n=6,
1286 Day 8-11 n=6 Day 12-14 n=4, Day 15-18 n=5, Convalescence n=3).

1287 (A, B, C, E): One-way ANOVA with Dunnet's for multiple comparisons correction used for
1288 statistical analysis.

1289 Only significant differences are indicated. *p <0.05; **p <0.01; ***p <0.001. CM=Classical
1290 Monocytes; IntM=Intermediate Monocytes; NCM=Non-classical Monocytes; MFI=Median
1291 Fluorescence Intensity.

1292

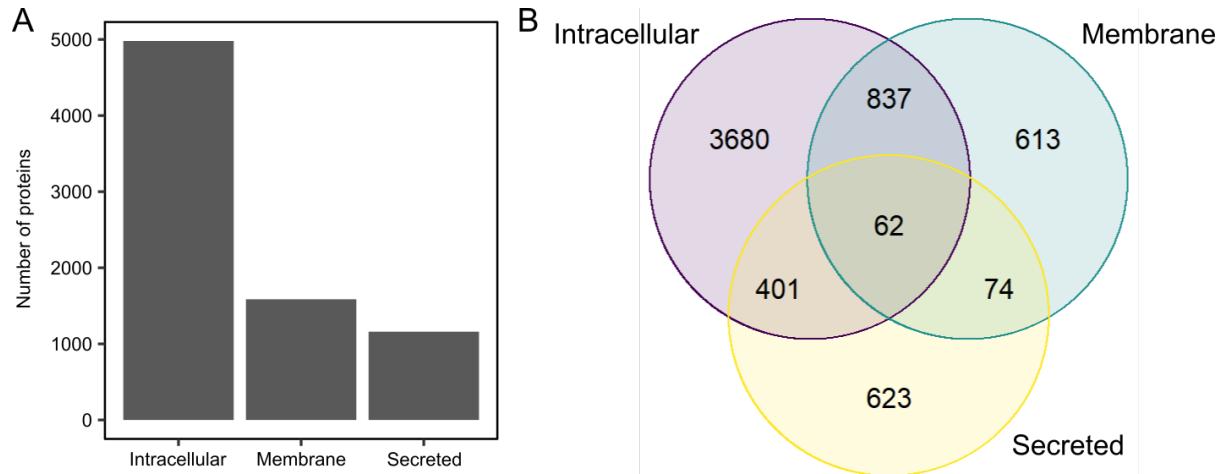
1293



1294

1295

1296 **Supplementary Figure 15: Predicting COVID-19 severity using random forests. A)** Point
1297 estimates of area under the curve from receiver operator analysis (AUC-ROC) for predicting
1298 COVID-19 severity with 95% confidence intervals using random forests. “Both” = supervised
1299 learning on the combined proteomic and transcriptomic data. **B)** Important proteins (left) and
1300 genes (right) for the random forests model. Feature importance is scaled between 0 and 1,
1301 where 1 represents the most important feature.
1302



1303
1304 **Supplementary Figure 16: Human Protein Atlas classification of proteins measured by**
1305 **the SomaScan v4.1 assay.** A) The number of unique proteins measured that were labelled
1306 as intracellular, membrane and secreted. B) Venn diagram illustrating overlap between the
1307 annotations.
1308

1309
1310
1311

Supplementary Tables

Supplementary Table 1: Characteristics of the Wave 1 cohort.

	Overall (n = 53)	Peak severity mild or moderate (n = 28)	Peak severity severe or critical (n = 25)
Age Median (IQR)	72.0 62.0-76.0	73.0 64.8-76.2	68.0 62.0-76.0
Sex M F	37 (69.8%) 16 (30.2%)	18 (64.3%) 10 (35.7%)	19 (76.0%) 6 (24.0%)
Ethnicity Asian White Black Other	21 (39.6%) 17 (32.1%) 8 (15.1%) 7 (13.2%)	11 (39.3%) 6 (21.4%) 5 (17.9%) 6 (21.4%)	10 (40.0%) 11 (44.0%) 3 (12.0%) 1 (4.0%)
Diabetes	32* (60.4%)	16* (57.1%)	16 (64.0%)
Current smoker	1 (1.9%)	1 (3.6%)	0 (0.0%)
ESKD cause DN HTN/vascular GN/autoimmune Genetic Other/unknown	27 (50.9%) 5 (9.4%) 3 (5.7%) 1 (1.9%) 17 (32.1%)	14 (50.0%) 3 (10.7%) 1 (3.6%) 1 (3.6%) 9 (32.1%)	13 (52.0%) 2 (8.0%) 2 (8.0%) 0 (0.0%) 9 (36.0%)
Hospitalisation due to COVID-19†	32 (60.4%)	7 (25.0%)	25 (100%)
Fatal COVID-19	9 (17.0%)	0 (0.0%)	9 (36.0%)

1312 DN = diabetic nephropathy. GN = glomerulonephritis. HTN = hypertension. IQR = inter-
1313 quartile range. Subsets defined according to peak WHO severity over the course of the illness.
1314

1315 *One patient had type 1 diabetes, the remainder type 2. †3 patients were hospitalised prior to
1316 COVID-19 diagnosis. 8 patients diagnosed with COVID-19 as outpatients subsequently
1317 deteriorated were hospitalised.
1318

1319

Supplementary Table 2: Characteristics of the Wave 2 cohort.

	Overall (n = 17)	Peak severity mild or moderate (n = 9)	Peak severity severe or critical (n = 8)
Age Median (IQR)	72.0 61.0-77.0	72.0 60.0-75.0	70.5 64.0-80.5
Sex M F	10 (58.8%) 7 (41.2%)	4 (44.4%) 5 (55.6%)	6 (75.0%) 2 (25.0%)
Ethnicity Asian White Black Other	11 (64.7%) 4 (23.5%) 1 (5.9%) 1 (5.9%)	6 (66.7%) 2 (22.2%) 0 (0.0%) 1 (11.1%)	5 (62.5%) 2 (25.0%) 1 (12.5%) 0 (0.0%)
Diabetes	11 (64.7%)	6 (66.7%)	5 (62.5%)
Current smoker	0 (0.0%)	0 (0.0%)	0 (0.0%)
ESKD cause DN HTN/vascular GN/autoimmune Genetic Other/unknown	10 (58.8%) 0 (0.0%) 1 (5.9%) 0 (0.0%) 6 (35.3%)	6 (66.7%) 0 (0.0%) 0 (0.0%) 0 (0.0%) 3 (33.3%)	4 (50.0%) 0 (0.0%) 1 (12.5%) 0 (0.0%) 3 (37.5%)
Hospitalisation due to COVID-19†	9 (52.9%)	1 (11.1%)	8 (100%)
Fatal COVID-19	4 (23.5%)	0 (0.0%)	4 (50.0%)

1320

DN = diabetic nephropathy. GN = glomerulonephritis. HTN = hypertension. IQR = interquartile range. Subsets defined according to peak WHO severity over the course of the illness.

1321

1322

1323

1324
1325

Supplementary Table 3: Transcriptomic modules associated with disease trajectory.

Module code	Module name	Size	Selected enrichments	Primary cell type	Hub genes	Severity association
tB	Granulocyte cell-like module 1	1,309	R: Neutrophil degranulation R: Innate immune system R: ROS and RNS production in phagocytes R: Oxidative stress-induced senescence	Neutrophils	<i>TECPR2, CSF3R, STX3, MMP25, BASP1, MBOAT7, NCF4, GLT1D1, RNF24, DHX34</i>	↑
tJ	Granulocyte cell-like module 2	169	R: Neutrophil degranulation R: Innate immune system R: Antimicrobial peptides NABA: Matrisome	Macrophages (M0)	<i>CEACAM8, BPI, CD24, CEACAM6, ABCA13, DEFA4, LTF, AZU1, ELANE, LCN2</i>	↑*
tL	Plasma cell module	655	R: Cell cycle R: MHC class II antigen presentation WP: DNA damage response R: Factors involved in megakaryocyte development and platelet production	Plasma cells	<i>RRM2, FOXM1, MKI67, TPX2, BUB1, KIFC1, TK1, MZB1, CDK1, CCNA2</i>	↑
tP	Nuclear and cell cycle module	464	None significant	B cells naive	<i>SNRPA1, CEP95, THUMPD2, CENPC, ZNF326, NDUFAF5, ILKAP, SUPV3L1, USP36, GOLT1B</i>	↓
tC	T-cell activity module 1	545	KEGG: Natural killer cell mediated cytotoxicity R: Immunoregulatory interactions between a lymphoid and a non-lymphoid cell PID: CD8 TCR downstream pathway	NK cells resting	<i>SAMD3, ADGRG1, ZAP70, PYHIN1, FCRL6, PRSS23, FGFBP2, NFATC2, PTCH1, LLGL2</i>	↓
tF	T-cell activity module 2	761	WP: T-cell receptor and co-stimulatory signaling	T cells CD4 memory resting	<i>PLCG1, LCK, UBASH3A, ABCD2, LINC00649, PRKCQ-AS1, TC2N, LINC01550, SEPTIN1, NLRC3</i>	↓
tl	Monocyte module	299	R: PD1 signaling KEGG: Viral myocarditis KEGG: Antigen processing and presentation KEGG: Cell adhesion molecules R: Interferon gamma signaling	Monocytes	<i>PLXNB2, NAAA, CSF1R, PSAP, CARD9, SLC7A7, PEA15, ARHGEF10L, ZNF385A, ARRB1</i>	↓
tN	Allergy-related module	54	WP: IL-3 signaling pathway KEGG: Asthma	T cells CD4 memory resting	<i>HDC, LINC02458, CPA3, GATA2, AKAP12, MS4A2, ENPP3, FCER1A, TRIM51EP, SLC45A3</i>	↓

1326 Transcriptomic modules associated with disease trajectory (i.e. with a significant TxC
1327 interaction) are tabulated and assigned names that are representative of their members. Size
1328 = the number of genes assigned to the module. Primary cell type indicates the CIBERSORTx
1329 cell type with the greatest positive correlation with the module's eigengene. Severity
1330 association indicates whether the module's eigengene is positively (↑) or negatively (↓)
1331 associated with contemporaneous WHO severity (5% FDR).

1332 * Module tJ was positively correlated with severity, but this association was not significant at
1333 5% FDR.

1334 R = Reactome

1335 WP = WikiPathways

1336 PID = Pathway Interaction Database

1337

1338 **Supplementary Table 4: Proteomic modules associated with disease trajectory.**

1339

Module code	Module name	Size	Selected enrichments	Hub genes	Severity association
p12	Histone-associated module	40	R: HDACs deacetylate histones KEGG: Systemic lupus erythematosus R: Chromatin modifying enzymes R: HCMV late events	H2AC1, H2AW, H2BU1, H2BC21, H2BC12, H2AZ1, CELF2, H2AC11, EEF1B2, MMP17	↑
p9	Splicing and nuclear module 1	241	PID: FRA pathway WP: Striated muscle contraction R: Complement cascade KEGG: Spliceosome	CLSTN3, KHSRP, OIT3, NELFA, EWSR1, CETP, PUF60, SRSF6, NXT1, ARHGAP36	↑
p8	Splicing and nuclear module 2	107	R: Processing of capped intron containing pre-mRNA R: mRNA splicing KEGG: Spliceosome	CILP, FUBP1, TSSC4, ALKAL2, STMN3, STMN1, STMN2, UBE2Z, MAFG, LEMD1	↑
p1	Mixed immune module 1	281	WP: Development of pulmonary dendritic cells and macrophage subsets	MRPL52, PLA2G2C, ITGAL, NLRP4, GDF3, VSTM4, ITGB1, SVBP, ANXA8, UBD	↑
p7	Mixed immune module 2	66	None significant	DEFB135, SPINK14, VAT1L, NT5E, PAXIP1, ZHX3, ODC1, PLEKHM2, DEFB112, F3	↓

1340 Proteomic modules associated with disease trajectory (i.e. with a significant TxC C interaction)
 1341 are tabulated and assigned names that are representative of their members. Size = the
 1342 number of proteins assigned to the module. Severity association indicates whether the
 1343 module's eigenprotein is positively (↑) or negatively (↓) associated with contemporaneous
 1344 WHO severity (5% FDR).

1345 R = Reactome

1346 WP = WikiPathways

1347 PID = Pathway Interaction Database

1348

1349 **Supplementary Table 5. List of antibodies used in flow cytometry.**

Antibodies	Source	Clone	Identifier
BV421-CD4	BioLegend	A161A1	357424
BV421-CD14	BioLegend	63D3	367144
PB-CD57	BioLegend	QA17A04	393316
PB-Foxp3 (Intracellular)	BioLegend	206D	320116
BV510-CD95	BioLegend	DX2	305640
BV510-IgD	BioLegend	IA6-2	348219
BV605-CCR7	BioLegend	G043H7	353224
BV605-CD4	BioLegend	SK3	344645
BV650-HLA-DR	BioLegend	L243	307650
BV650-CD123	BioLegend	6H6	306032
BV711-KLRG1	BioLegend	2F1/KLRG1	138427
BV711-CD11c	BioLegend	3.9	301630
BV750-CD45RA	BioLegend	HI100	304166
BV750-HLA-DR	BioLegend	L243	307672
BV785-CD19	BioLegend	HIB19	302240
AF488-CD3	BioLegend	HIT3a	300320
BB515-CD152/CTLA-4 (Intracellular)	BD Bioscience	BNI3	566917
BB515-CD8	BD Bioscience	RPA-T8	564526
PE-CD56	BioLegend	39D5	355504
PE-CD335	BioLegend	9^E2	331908
PE- CD16	BioLegend	3G8	302008
PE- CD14	BioLegend	63D3	367104
PE- CD19	BioLegend	HIB19	302208
PE-Siglec-1	BioLegend	7-239	346004
PEDAZ594-CD69	BioLegend	FN50	310941
PEDAZ594-NKG2D	BioLegend	1D11	320828
PE-Cy5-CD27	Thermofisher	O323	15-0279-42
PE-Cy5-CD25	Biolegend	BC96	302608
Percp-cy5.5-Ki67 (Intracellular)	BioLegend	Ki-67	350520
Percp-ef710-gdTCR	Thermofisher	B1.1	46-9959-42
Percp-ef710-CD16	Thermofisher	CB16	46-0168-42
PE-Cy7-PD1	BioLegend	A17188B	621616
PE-Cy7-CD141	Biolegend	M80	344110
AF647-CD45	BioLegend	HI30	304056
AF700-CD8	BioLegend	HIT8a	300920
AF700-CD66b	BioLegend	VI MA81	305114
APC/Fire750-CD38	BioLegend	HB-7	356626

1350

1351

1352 **Titles for Supplementary File - (in Excel file)**

1353 **Supplementary File 1A. Differential gene expression analysis comparing COVID-19**
1354 **positive versus negative PBMC samples.** Contains the linear mixed model estimates and
1355 corresponding P-values for each gene, for both the Wave 1 and Wave 2 cohorts. The column
1356 "Aggregated Score" represents the RRA score for the P-values from both cohorts.

1357 **Supplementary File 1B. Gene set analysis comparing COVID-19 positive versus**
1358 **negative samples.** Contains the linear mixed model estimates and corresponding P-values
1359 for each GSVA gene set, for both the Wave 1 and Wave 2 cohorts. The column "Aggregated
1360 Score" represents the RRA score for the P-values from both cohorts.

1361 **Supplementary File 1C. Protein annotations.** List of proteins measured by the SomaScan
1362 v4.1 assay, their corresponding UniProt and GenIDs, and their annotations in the human
1363 protein atlas.

1364 **Supplementary File 1D. Differential plasma protein abundance analysis comparing**
1365 **COVID-19 positive versus negative samples.** Contains the linear mixed model estimates
1366 and corresponding P-values for each protein, for both the Wave 1 and Wave 2 cohorts. The
1367 column "Aggregated Score" represents the RRA score for the P-values from both cohorts.

1368 **Supplementary File 1E. Protein set analysis comparing COVID-19 positive versus**
1369 **negative samples.** Contains the linear mixed model estimates and corresponding P-values
1370 for each GSVA protein set, for both the Wave 1 and Wave 2 cohorts. The column "Aggregated
1371 Score" represents the RRA score for the P-values from both cohorts.

1372 **Supplementary File 1F. Transcriptomic associations with contemporaneous COVID-19**
1373 **severity.** Associations with 4-level ordinal WHO severity score at the time of the sample.
1374 Contains the linear mixed model estimates and corresponding P-values for each gene, for
1375 both the Wave 1 and Wave 2 cohorts. The column "Aggregated Score" represents the RRA
1376 score for the P-values from both cohorts.

1377 **Supplementary File 1G. Associations of gene sets with contemporaneous COVID-19**
1378 **severity.** Associations with 4-level ordinal WHO severity score at the time of the sample.
1379 Contains the linear mixed model estimates and corresponding P-values for each GSVA gene
1380 set, for both the Wave 1 and Wave 2 cohorts. The column "Aggregated Score" represents the
1381 RRA score for the P-values from both cohorts.

1382 **Supplementary File 1H. Proteomic associations with contemporaneous COVID-19**
1383 **severity.** Associations with 4-level ordinal WHO severity score at the time of the sample.
1384 Contains the linear mixed model estimates and corresponding P-values for each protein, for
1385 both the Wave 1 and Wave 2 cohorts. The column "Aggregated Score" represents the RRA
1386 score for the P-values from both cohorts.

1387 **Supplementary File 1I. Associations of protein sets with contemporaneous COVID-19**
1388 **severity.** Associations with 4-level ordinal WHO severity score at the time of the sample.
1389 Contains the linear mixed model estimates and corresponding P-values for each GSVA gene
1390 set, for both the Wave 1 and Wave 2 cohorts. The column "Aggregated Score" represents the
1391 RRA score for the P-values from both cohorts.

1392 **Supplementary File 1J. The membership of genes to WGCNA transcriptomic modules.**

1393 **Supplementary File 1K. The membership of proteins to WGCNA proteomic modules.**

- 1394 **Supplementary File 1L. Gene set enrichment of transcriptomic WGCNA modules.**
1395 Overrepresentation analysis of gene sets for each module.
- 1396 **Supplementary File 1M. Protein set enrichment of proteomic WGCNA modules.**
1397 Overrepresentation analysis of protein sets for each module.
- 1398 **Supplementary File 1N. Associations of imputed cell proportions with transcriptomic**
1399 **WGCNA modules.** Contains the linear mixed model estimates and corresponding P-values
1400 for each WGCNA module – imputed cell type pair.
- 1401 **Supplementary File 1O. Longitudinal profiles of cytokine proteins in plasma.** P-values
1402 for the linear mixed modelling of cytokines and related proteins. P-values are included for the
1403 time, clinical course, and time * clinical course (TxC) terms.
- 1404 **Supplementary File 1P. Importance metrics for supervised learning of the**
1405 **transcriptome.** The relative importance of genes according to the random forests (accuracy
1406 decrease) and lasso (number of models in which each gene had a non-zero coefficient during
1407 cross-validation) models. The metrics are normalised such that the most important gene has
1408 a value of 1.
- 1409 **Supplementary File 1Q. Importance metrics for supervised learning of the proteome.**
1410 The relative importance of proteins according to the random forests (accuracy decrease) and
1411 lasso (number of models in which each protein had a non-zero coefficient during cross-
1412 validation) models. The metrics are normalised such that the most important gene has a value
1413 of 1.
- 1414 **Supplementary File 1R. Multi-omic supervised learning importance metrics.** The relative
1415 importance of features (genes or proteins) according to the random forests (accuracy
1416 decrease) and lasso (number of models in which each feature had a non-zero coefficient
1417 during cross-validation) models. The metrics are normalised such that the most important
1418 gene has a value of 1.
- 1419 **Supplementary File 1S. Paired differential expression analysis of pre-infection versus**
1420 **convalescent samples.** Contains the linear mixed model estimates and corresponding P-
1421 values for each gene.
- 1422 **Supplementary File 1T. Gene set enrichment of convalescence analysis.** Contains the
1423 linear mixed model estimates and corresponding P-values for each GSVA gene set, for both
1424 the Wave 1 and Wave 2 cohorts. The column “Aggregated Score” represents the RRA score
1425 for the P-values from both cohorts.
- 1426

1427 REFERENCES

- 1428 1. Mann, E. R. *et al.* Longitudinal immune profiling reveals key myeloid signatures
1429 associated with COVID-19. *Sci. Immunol.* **5**, eabd6197 (2020)
1430 doi:10.1126/sciimmunol.abd6197.
- 1431 2. Laing, A. G. *et al.* A dynamic COVID-19 immune signature includes associations with
1432 poor prognosis. *Nat. Med.* (2020) doi:10.1038/s41591-020-1038-6.
- 1433 3. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in
1434 COVID-19. *Nat. Med.* **27**, 904–916 (2021) doi:10.1038/s41591-021-01329-2.
- 1435 4. Bernardes, J. P. *et al.* Longitudinal Multi-omics Analyses Identify Responses of
1436 Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19.
1437 *Immunity* **53**, 1296-1314.e9 (2020) doi:<https://doi.org/10.1016/j.jimmuni.2020.11.017>.
- 1438 5. Szabo, P. A. *et al.* Longitudinal profiling of respiratory and systemic immune responses
1439 reveals myeloid cell-driven lung inflammation in severe COVID-19. *Immunity* **54**, 797-
1440 814.e6 (2021) doi:<https://doi.org/10.1016/j.jimmuni.2021.03.005>.
- 1441 6. Bergamaschi, L. *et al.* Longitudinal analysis reveals that delayed bystander CD8+ T cell
1442 activation and early immune pathology distinguish severe COVID-19 from mild disease.
1443 *Immunity* **54**, 1257-1275.e8 (2021) doi:<https://doi.org/10.1016/j.jimmuni.2021.05.010>.
- 1444 7. Ahern, D. J. *et al.* A blood atlas of COVID-19 defines hallmarks of disease severity and
1445 specificity. *Cell* **185**, 916-938.e58 (2022) doi:10.1016/j.cell.2022.01.012.
- 1446 8. Filbin, M. R. *et al.* Longitudinal proteomic analysis of severe COVID-19 reveals survival-
1447 associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Reports
1448 Med.* **2**, 100287 (2021) doi:10.1016/j.xcrm.2021.100287.
- 1449 9. Gisby, J. *et al.* Longitudinal proteomic profiling of dialysis patients with COVID-19
1450 reveals markers of severity and predictors of death. *eLife* **10**, 2020.11.05.20223289
1451 (2021) doi:10.7554/eLife.64827.
- 1452 10. Rodriguez, L. *et al.* Systems-Level Immunomonitoring from Acute to Recovery Phase
1453 of Severe COVID-19. *Cell reports. Med.* **1**, 100078 (2020)
1454 doi:10.1016/j.xcrm.2020.100078.
- 1455 11. Demichev, V. *et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell
1456 Syst.* **12**, 780-794.e7 (2021) doi:<https://doi.org/10.1016/j.cels.2021.05.005>.
- 1457 12. Gutmann, C. *et al.* SARS-CoV-2 RNAemia and proteomic trajectories inform
1458 prognostication in COVID-19 patients admitted to intensive care. *Nat. Commun.* **12**,
1459 3406 (2021) doi:10.1038/s41467-021-23494-1.
- 1460 13. Galbraith, M. D. *et al.* Seroconversion stages COVID19 into distinct pathophysiological
1461 states. *eLife* **10**, e65508 (2021) doi:10.7554/eLife.65508.
- 1462 14. Paranjpe, I. *et al.* Proteomic Characterization of Acute Kidney Injury in Patients
1463 Hospitalized with SARS-CoV2 Infection. *medRxiv* 2021.12.09.21267548 (2021)
1464 doi:10.1101/2021.12.09.21267548.
- 1465 15. Su, C.-Y. *et al.* Circulating proteins to predict adverse COVID-19 outcomes. *medRxiv*
1466 2021.10.04.21264015 (2021) doi:10.1101/2021.10.04.21264015.
- 1467 16. Sullivan, K. D. *et al.* The COVIDome Explorer researcher portal. *Cell Rep.* **36**, (2021)
1468 doi:10.1016/j.celrep.2021.109527.
- 1469 17. Horby, P. *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *N. Engl. J. Med.*
1470 **384**, 693–704 (2021) doi:10.1056/NEJMoa2021436.

- 1471 18. Gordon, A. C. *et al.* Interleukin-6 Receptor Antagonists in Critically Ill Patients with
1472 Covid-19. *N. Engl. J. Med.* **384**, 1491–1502 (2021) doi:10.1056/NEJMoa2100433.
- 1473 19. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a
1474 randomised, controlled, open-label, platform trial. *Lancet (London, England)* **397**,
1475 1637–1645 (2021) doi:10.1016/S0140-6736(21)00676-0.
- 1476 20. Kalil, A. C. *et al.* Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N.*
1477 *Engl. J. Med.* **384**, 795–807 (2021) doi:10.1056/NEJMoa2031994.
- 1478 21. Williamson, E. J. *et al.* Factors associated with COVID-19-related death using
1479 OpenSAFELY. *Nature* **584**, 430–436 (2020) doi:10.1038/s41586-020-2521-4.
- 1480 22. Goffin, E. *et al.* COVID-19-related mortality in kidney transplant and haemodialysis
1481 patients: a comparative, prospective registry-based study. *Nephrol. Dial. Transplant.*
1482 *Off. Publ. Eur. Dial. Transpl. Assoc. - Eur. Ren. Assoc.* **36**, 2094–2105 (2021)
1483 doi:10.1093/ndt/gfab200.
- 1484 23. Chen, J.-J. *et al.* Immunogenicity Rates After SARS-CoV-2 Vaccination in People With
1485 End-stage Kidney Disease: A Systematic Review and Meta-analysis. *JAMA Netw.*
1486 *open* **4**, e2131749 (2021) doi:10.1001/jamanetworkopen.2021.31749.
- 1487 24. Anand, S. *et al.* Antibody Response to COVID-19 Vaccination in Patients Receiving
1488 Dialysis. *J. Am. Soc. Nephrol.* **32**, 2435–2438 (2021) doi:10.1681/ASN.2021050611.
- 1489 25. Hänelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for
1490 microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013) doi:10.1186/1471-
1491 2105-14-7.
- 1492 26. Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with
1493 severe COVID-19. *Nat. Med.* **26**, 1070–1076 (2020) doi:10.1038/s41591-020-0944-y.
- 1494 27. Liu, G. *et al.* ISG15-dependent activation of the sensor MDA5 is antagonized by the
1495 SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat. Microbiol.* **6**,
1496 467–478 (2021) doi:10.1038/s41564-021-00884-1.
- 1497 28. Hachim, M. Y. *et al.* Interferon-Induced Transmembrane Protein (IFITM3) Is
1498 Upregulated Explicitly in SARS-CoV-2 Infected Lung Epithelial Cells. *Front. Immunol.*
1499 **11**, 1372 (2020) doi:10.3389/fimmu.2020.01372.
- 1500 29. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network
1501 analysis. *BMC Bioinformatics* **9**, (2008) doi:10.1186/1471-2105-9-559.
- 1502 30. Velten, B., Braunger, J. M., Arnol, D., Argelaguet, R. & Stegle, O. Identifying temporal
1503 and spatial patterns of variation from multi-modal data using MEFISTO. *bioRxiv*
1504 2020.11.03.366674 (2020) doi:10.1101/2020.11.03.366674.
- 1505 31. Bagheri-Hosseinabadi, Z., Abbasi, M., Kahnooji, M., Ghorbani, Z. & Abbasifard, M. The
1506 prognostic value of S100A calcium binding protein family members in predicting severe
1507 forms of COVID-19. *Inflamm. Res.* **71**, 369–376 (2022) doi:10.1007/s00011-022-
1508 01545-7.
- 1509 32. Schett, G., McInnes, I. B. & Neurath, M. F. Reframing Immune-Mediated Inflammatory
1510 Diseases through Signature Cytokine Hubs. *N. Engl. J. Med.* **385**, 628–639 (2021)
1511 doi:10.1056/NEJMra1909094.
- 1512 33. Shilts, J. *et al.* LRRC15 mediates an accessory interaction with the SARS-CoV-2 spike
1513 protein. *bioRxiv* 2021.09.25.461776 (2021) doi:10.1101/2021.09.25.461776.
- 1514 34. Song, J. *et al.* LRRC15 is an inhibitory receptor blocking SARS-CoV-2 spike-mediated
1515 entry in trans. *bioRxiv: the preprint server for biology* (2021)

- 1516 doi:10.1101/2021.11.23.469714.
- 1517 35. Ng, J. H. *et al.* Outcomes of patients with end-stage kidney disease hospitalized with
1518 COVID-19. *Kidney Int.* (2020) doi:10.1016/j.kint.2020.07.030.
- 1519 36. Mei, Q. *et al.* Regulation of DNA replication-coupled histone gene expression.
1520 *Oncotarget* **8**, 95005–95022 (2017) doi:10.18632/oncotarget.21887.
- 1521 37. Singh, R. K., Kabbaj, M.-H. M., Paik, J. & Gunjan, A. Histone levels are regulated by
1522 phosphorylation and ubiquitylation-dependent proteolysis. *Nat. Cell Biol.* **11**, 925–933
1523 (2009) doi:10.1038/ncb1903.
- 1524 38. Silk, E., Zhao, H., Weng, H. & Ma, D. The role of extracellular histone in organ injury.
1525 *Cell Death Dis.* **8**, e2812 (2017) doi:10.1038/cddis.2017.52.
- 1526 39. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy
1527 targets. *Nature* **583**, 469–472 (2020) doi:10.1038/s41586-020-2332-7.
- 1528 40. Finkel, Y. *et al.* SARS-CoV-2 uses a multipronged strategy to impede host protein
1529 synthesis. *Nature* **594**, 240–245 (2021) doi:10.1038/s41586-021-03610-3.
- 1530 41. Vlahos, R. *et al.* Inhibition of Nox2 Oxidase Activity Ameliorates Influenza A Virus-
1531 Induced Lung Inflammation. *PLOS Pathog.* **7**, e1001271 (2011).
- 1532 42. Lang, P. A. *et al.* Reactive oxygen species delay control of lymphocytic choriomeningitis
1533 virus. *Cell Death Differ.* **20**, 649–658 (2013) doi:10.1038/cdd.2012.167.
- 1534 43. Randzavola, L. O. *et al.* EROS-mediated control of NOX2 and P2X7 biosynthesis.
1535 *bioRxiv* 2021.09.14.460103 (2021) doi:10.1101/2021.09.14.460103.
- 1536 44. Ng, B., Cook, S. A. & Schafer, S. Interleukin-11 signaling underlies fibrosis,
1537 parenchymal dysfunction, and chronic inflammation of the airway. *Exp. Mol. Med.* **52**,
1538 1871–1878 (2020) doi:10.1038/s12276-020-00531-5.
- 1539 45. Schafer, S. *et al.* IL-11 is a crucial determinant of cardiovascular fibrosis. *Nature* **552**,
1540 110–115 (2017) doi:10.1038/nature24676.
- 1541 46. Kiernan, K. & MacIver, N. J. The Role of the Adipokine Leptin in Immune Cell Function
1542 in Health and Disease. *Front. Immunol.* **11**, 622468 (2020)
1543 doi:10.3389/fimmu.2020.622468.
- 1544 47. Segerer, S. E. *et al.* The glycoprotein-hormones activin A and inhibin A interfere with
1545 dendritic cell maturation. *Reprod. Biol. Endocrinol.* **6**, 17 (2008) doi:10.1186/1477-
1546 7827-6-17.
- 1547 48. Rapp, M. *et al.* CCL22 controls immunity by promoting regulatory T cell communication
1548 with dendritic cells in lymph nodes. *J. Exp. Med.* **216**, 1170–1181 (2019)
1549 doi:10.1084/jem.20170277.
- 1550 49. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening
1551 COVID-19. *Science* **370**, (2020) doi:10.1126/science.abd4570.
- 1552 50. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with life-threatening
1553 COVID-19. *Science* **370**, (2020) doi:10.1126/science.abd4585.
- 1554 51. Ho, F. K. *et al.* Thromboembolic Risk in Hospitalized and Nonhospitalized COVID-19
1555 Patients: A Self-Controlled Case Series Analysis of a Nationwide Cohort. *Mayo Clin. Proc.* **96**, 2587–2597 (2021) doi:10.1016/j.mayocp.2021.07.002.
- 1557 52. Gorog, D. A. *et al.* Current and novel biomarkers of thrombotic risk in COVID-19: a
1558 Consensus Statement from the International COVID-19 Thrombosis Biomarkers

- 1559 Colloquium. *Nat. Rev. Cardiol.* 1–21 (2022) doi:10.1038/s41569-021-00665-7.
- 1560 53. Katsoularis, I. et al. Risks of deep vein thrombosis, pulmonary embolism, and bleeding
1561 after covid-19: nationwide self-controlled cases series and matched cohort study. *BMJ*
1562 **377**, (2022) doi:10.1136/bmj-2021-069590.
- 1563 54. Greinacher, A. et al. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19
1564 Vaccination. *N. Engl. J. Med.* **384**, 2092–2101 (2021) doi:10.1056/NEJMoa2104840.
- 1565 55. Schultz, N. H. et al. Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19
1566 Vaccination. *N. Engl. J. Med.* **384**, 2124–2130 (2021) doi:10.1056/NEJMoa2104882.
- 1567 56. Greinacher, A. et al. Insights in ChAdOx1 nCoV-19 vaccine-induced immune thrombotic
1568 thrombocytopenia. *Blood* **138**, 2256–2268 (2021) doi:10.1182/blood.2021013231.
- 1569 57. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics
1570 pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020) doi:10.1038/s41587-020-0439-x.
- 1571 58. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat.*
1572 *Biotechnol.* **35**, 316–319 (2017) doi:10.1038/nbt.3820.
- 1573 59. Andrews, S. et al. FastQC: a quality control tool for high throughput sequence data.
1574 (2012).
- 1575 60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1576 reads. *EMBnet.journal*; Vol 17, No 1 Next Gener. Seq. Data Anal. -
1577 10.14806/ej.17.1.200 (2011).
- 1578 61. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
1579 (2013) doi:10.1093/bioinformatics/bts635.
- 1580 62. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
1581 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015)
1582 doi:10.1093/bioinformatics/btu638.
- 1583 63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1584 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–
1585 140 (2010) doi:10.1093/bioinformatics/btp616.
- 1586 64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential
1587 expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010) doi:10.1186/gb-
1588 2010-11-3-r25.
- 1589 65. Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021)
1590 doi:10.1093/nar/gkaa942.
- 1591 66. Tweedie, S. et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic*
1592 *Acids Res.* **49**, D939–D946 (2021) doi:10.1093/nar/gkaa980.
- 1593 67. John, C. R. et al. M3C: Monte Carlo reference-based consensus clustering. *Sci. Rep.*
1594 **10**, 1–14 (2020) doi:10.1038/s41598-020-58766-1.
- 1595 68. Uhlen, M. et al. Tissue-based map of the human proteome. *Science* (80-.). **347**,
1596 1260419–1260419 (2015) doi:10.1126/science.1260419.
- 1597 69. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models
1598 Using lme4. *J. Stat. Softw.* **67**, (2015) doi:10.18637/jss.v067.i01.
- 1599 70. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in
1600 Linear Mixed Effects Models. *J. Stat. Softw.* **82**, (2017) doi:10.18637/jss.v082.i13.
- 1601 71. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for

- 1602 repeated measures designs. *Bioinformatics* **37**, 192–201 (2021)
1603 doi:10.1093/bioinformatics/btaa687.
- 1604 72. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in
1605 complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016)
1606 doi:10.1186/s12859-016-1323-z.
- 1607 73. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**,
1608 1739–1740 (2011) doi:10.1093/bioinformatics/btr260.
- 1609 74. Buang, N. et al. Type I interferons affect the metabolic fitness of CD8+ T cells from
1610 patients with systemic lupus erythematosus. *Nat. Commun.* **12**, 1980 (2021)
1611 doi:10.1038/s41467-021-22312-y.
- 1612 75. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration
1613 and meta-analysis. *Bioinformatics* **28**, 573–580 (2012)
1614 doi:10.1093/bioinformatics/btr709.
- 1615 76. Li, J. et al. Application of Weighted Gene Co-expression Network Analysis for Data from
1616 Paired Design. *Sci. Rep.* **8**, 622 (2018) doi:10.1038/s41598-017-18705-z.
- 1617 77. Bakdash, J. Z. & Marusich, L. R. Repeated measures correlation. *Front. Psychol.* **8**, 1–
1618 13 (2017) doi:10.3389/fpsyg.2017.00456.
- 1619 78. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster
1620 tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008)
1621 doi:10.1093/bioinformatics/btm563.
- 1622 79. Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. & Schmid, M. A review of spline
1623 function procedures in R. *BMC Med. Res. Methodol.* **19**, 1–16 (2019)
1624 doi:10.1186/s12874-019-0666-3.
- 1625 80. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
1626 *Innov.* **2**, 100141 (2021) doi:<https://doi.org/10.1016/j.xinn.2021.100141>.
- 1627 81. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues
1628 with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019) doi:10.1038/s41587-019-
1629 0114-2.
- 1630 82. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression
1631 profiles. *Nat. Methods* **12**, 453–457 (2015) doi:10.1038/nmeth.3337.
- 1632 83. Velten, B. et al. Identifying temporal and spatial patterns of variation from multimodal
1633 data using MEFISTO. *Nat. Methods* **19**, (2022) doi:10.1038/s41592-021-01343-9.
- 1634 84. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **28**,
1635 1–26 (2008) doi:10.18637/jss.v028.i05.
- 1636 85. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear
1637 Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010)
1638 doi:10.18637/jss.v033.i01.
- 1639