Immunological dysfunction persists for 8 months following initial mild-moderate SARS-CoV-2 infection

Chansavath Phetsouphanh[1*], David Darley[2*], Daniel B Wilson[3], Annett Howe[1], C. Mee Ling Munier[1], Sheila K Patel[4], Jennifer A Juno[5], Louise M Burrell[4], Stephen J Kent[5, 6], Gregory J Dore[1, 2], Anthony D Kelleher[1, 2†], Gail V Matthews[1, 2†]

[1] The Kirby institute, University of New South Wales, NSW 2033, Australia

[2] St Vincent's Hospital, Darlinghurst, NSW 2010, Australia

[3] Department of Mathematics and Statistics, Boston University, MA 02215, United States of America

[4] Department of Medicine, Austin Health, University of Melbourne, Victoria 3084, Australia

[5] Department of Microbiology and Immunology, Peter Doherty Institute, University of Melbourne, Victoria 3000, Australia

[6] Melbourne Sexual health centre, Infectious Diseases Department, Alfred health, central Clinical School, Monash University, Melbourne, Victoria 3004, Australia

*, † These authors contributed equally

Correspondence

Dr Chansavath Phetsouphanh

cphetsouphanh@kirby.unsw.edu.au

Prof. Anthony Kelleher

akelleher@kirby.unsw.edu.au

Prof. Gail Matthews

gmatthews@kirby.unsw.edu.au

Abstract

A proportion of patients surviving acute COVID-19 infection develop post-COVID syndrome (long COVID) encompassing physical and neuropsychiatric symptoms lasting longer than 12 weeks. Here we studied a prospective cohort of individuals with long COVID compared to age/gender matched subjects without long COVID (from the ADAPT study), healthy donors and individuals infected with other non-SARS CoV2 human coronaviruses (the ADAPT-C study). We found highly activated innate immune cells and an absence of subsets of un-activated naïve T and B cells in peripheral blood of long COVID subjects, that did not reconstitute over time. These activated myeloid cells may contribute to the elevated levels of type I (IFN-β) and III interferon (IFN-λ1) that remained persistently high in long COVID subjects at 8 months post-infection. We found positive inter-analyte correlations that consisted of 18 inflammatory cytokines in symptomatic long COVID subjects that was not observed in asymptomatic COVID-19 survivors. A linear classification model was used to exhaustively search through all 20475 combinations of the 29 analytes measured, that had the strongest association with long COVID and found that the best 4 analytes were: IL-6, IFN-γ, MCP-1 (CCL2) and VCAM-1. These four inflammatory biomarkers gave an accuracy of 75.9%, and an F1 score of 0.759, and have also previously been associated with acute severe disease. In contrast, plasma ACE2 levels, while elevated in the serum of people previously infected with SARS-CoV-2 were not further elevated in subjects with long COVID symptoms. This work defines immunological parameters associated with long COVID and suggests future opportunities to prevention and treatment.

Introduction

Coronavirus disease 2019 (COVID-19), the illness caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is characterised by a wide spectrum of clinical severities in the acute phase, ranging from asymptomatic to severe fatal forms [1,2]. Immunological disturbance during acute illness has been well described and is likely to both play a role in host defence and be involved in the pathogenesis of severe COVID-19 [3]. Immunological abnormalities described in the acute phase include significant immune dysregulation with lymphopaenia and elevations in IL-2, IL-6, TGF-β, IL-10, GM-CSF and TNF [3,4].

A proportion of patients surviving acute COVID-19 develop post-COVID syndrome encompassing various physical and neuropsychiatric symptoms lasting longer than 12 weeks, also known as chronic COVID syndrome (CCS), post-acute sequelae of COVID-19 (PASC) and long-haul COVID (long COVID) [5-8]. Although well described following other viral infections including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-1)[9-11] and Middle East Respiratory Syndrome Coronavirus (MERS-CoV)[12], these syndromes appear commonly following SARS-CoV-2 infection including following initial mild-moderate illness [13-16]. Persistent symptoms 6 months after hospitalization were observed in 76% of discharged patients in a study by Huang *et al*, with the most frequent symptoms reported being muscle weakness and fatigue [17,18]. Amongst community managed COVID-19 cases, prevalence of persisting symptoms is lower, but remains higher than might be expected given the severity of the initial illness, affecting between 10-30% of individuals at 2-3 months post infection, [19,20] continuing out to at least 8 months after initial illness [21]. The spectrum of post-acute sequelae of COVID-19 (henceforth denoted as long COVID) is broad encompassing as many as 50 plus symptoms although characteristically manifests as severe relapsing fatigue often accompanied by dyspnoea, chest tightness, cough and headache [22]. The underlying pathophysiology of long COVID is poorly understood and there is currently little evidence defining the mechanisms of systemic inflammation following COVID-19 and any immune correlates of ongoing symptoms.

96 We examined a prospective cohort of individuals followed systematically post confirmed
97 COVID-19 infection (the ADAPT study) and compared them to healthy donors and individuals
98 contemporaneously infected with other non-SARS CoV2 human coronaviruses (HCoV, ADAPT-
99 C) to firstly characterise their immunological features up to 8 months post COVID-19 infection
100 and secondly to compare those with and without persisting symptoms at this time point.
101 Serum profiles of 29 analytes from 62 patients, 31 with long COVID and 31 matched
102 asymptomatic controls were evaluated at 5- and 8-months post SARS-CoV-2 infection.
103 Immune profiles identified in COVID-19 were compared with subjects who had been infected
104 with common cold coronaviruses (HCoV-NL63, O229E, OC43 or HKU1) and unexposed healthy
105 donors.
106
107 **Methods**
108
109 *Cohort Characteristics*
110
111 The ADAPT study is a prospective cohort study of post-COVID-19 recovery established in April
112 2020 [21]. 147 participants with confirmed SARS-CoV-2 infection were enrolled, the majority
113 following testing in community based clinics run by St Vincent's Hospital Sydney, with some
114 patients also enrolled with confirmed infection at external sites. Initial study follow-up was
115 planned for 12 months post-COVID-19, and subsequently extended to 2 years. Extensive
116 clinical data and a biorepository was systematically collected prospectively. The aims of
117 ADAPT are to evaluate a number of outcomes after COVID-19 relating to pathophysiology,
118 immunology and clinical sequalae. Laboratory testing for SARS-CoV-2 was performed using
119 nucleic acid detection from respiratory specimens with the EasyScreen™ Respiratory
120 Detection kit (Genetic Signatures, Sydney, Australia), and the EasyScreen™ SARS-CoV-2
121 Detection kit. Two ADAPT cohort sub-populations were defined based on initial severity of
122 COVID-19 illness; 1. patients managed in the community and 2. patients admitted to hospital
123 for acute infection (including those requiring intensive care support for acute respiratory
124 distress syndrome (ARDS). Patients were defined as 'long COVID' at 4-months based on the
125 presence of any ≥ 1 of the following; fatigue, dyspnoea or chest pain [21]. These patients were
126 gender and age (+/- 10 years) matched with ADAPT participants without long COVID (Matched
127 ADAPT controls) (Table 1). Samples for these analyses were collected at the 3-, 4- and 8-

128 month assessments. Our cohort of n=62 (31 long COVID and 31 matched controls); enrolment
129 visits were performed at median 76 (IQR 64-93) days after initial infection. Their 4 month
130 assessments were performed at median 128 (IQR 115-142) days after initial infection (4.2
131 months). Their 8 month assessments were performed at median 232 (IQR 226-253) days after
132 initial infection (7.7 months). 4 participants did not complete the 8-month assessment after
133 the 4-month assessment. The reasons for this include- did not attend (n=2) and Lost to
134 Follow-up (n=2). A further population of patients presenting to St Vincent's Hospital clinics
135 for COVID-19 testing on the multiplex respiratory panel who were PCR positive for any of the
136 4 human common cold coronaviruses (HCoV-NL63, O229E, OC43 or HKU1) and PCR negative
137 for SARS-CoV-2 were recruited into the ADAPT-C sub study and used as a comparator group.
138
139 *Ethics*
140
141 The ADAPT study was approved by the St Vincent's Hospital Research Ethics Committee
142 (2020/ETH00964) and is a registered trial (ACTRN12620000554965). ADAPT-C sub study was
143 approved by the same committee (2020/ETH01429). All data were stored using REDCap
144 electronic data capture tools. Unexposed healthy donors were recruited through St Vincent's
145 Hospital and was approved by St Vincent's Hospital Research Ethics Committee
146 (HREC/13/SVH/145). The University of Melbourne unexposed donors were approved by
147 Medicine and Dentistry HESC-Study ID 2056689. All participants gave written informed
148 consent.
149
150 *Sample processing and Flow cytometry*
151
152 Blood was collected for biomarker analysis (SST 8.5mLs x 1 (Serum) and EDTA 10mLs x 1
153 (Plasma)) and 36mLs was collected for PBMCs (ACD 9mLs x4). Phenotyping of PBMC was
154 performed as described previously [23]. Briefly, cryopreserved PBMCs were thawed using RPMI
155 (+L-glut) medium (ThermoFisher Scientific, USA) supplemented with Penicillin/Streptomycin
156 (Sigma-Aldrich, USW), and subsequently stained with antibodies binding to extracellular
157 markers. Extracellular panel included: Live/Dead dye Near InfraRed, CXCR5 (MU5UBEE), CD38
158 (HIT2) (ThermoFisher Scientific, USA); CD3 (UCHT1), CD8 (HIL-72021), PD-1 (EH12.1), TIM-3
159 (TD3), CD27 (L128), CD45RA (HI100), IgD (IA6-2), CD25 (2A3), and CD19 (HIB19) (Biolegend,

160 USA); CD4 (OKT4), CD127 (A019D5), HLA-DR (L234), GRP56 (191B8), CCR7 (G043H7) and CD57
161 (QA17A04) (BD Biosciences, USA). Perm Buffer II (BD Pharmingen) was used for intracellular
162 staining of granzyme B (GB11, BD Biosciences). Samples were acquired on an Cytek Aurora
163 (Biolegend, USA) using the Spectroflo software. Prior to each run, all samples were fixed in
164 0.5% Paraformaldehyde.
165
166 *Serum Analytes*
167
168 LEGENDplex™ Human Anti-Virus Response Panel (IL-1β, IL-6, IL-8, IL-10, IL-12p70, IFN-α2, IFN-
169 β, IFN-λ1, IFN-λ2/3, IFN-γ, TNF-α, IP-10, GM-CSF) and custom-made panel (IL-5, IL-9, IL-13, IL-
170 33, PD-1, sTIM-3, sCD25, CCL2(MCP-1), Pentraxin-3(PTX3), TGF-β1, CXCL9(MIG-1), MPO,
171 PECAM-1, ICAM-1, VCAM-1) were purchased from Biolegend (San Diego, CA, USA) and assays
172 performed as per manufacturer's instructions. Beads were acquired and analysed on BD
173 Fortessa X20 SORP (BD Biosciences, San Jose, CA, USA). Samples were run in duplicate, and
174 4000 beads were acquired per sample. Data analysis was performed using Qognit
175 LEGENDplex™ software (Biolegend, San Diego, CA, USA). Lower limit of detection (LOD) was
176 values were used for all analytes at the lower limit.
177
178 *Catalytic ACE2 detection in plasma*
179
180 Plasma ACE2 activity was measured using a validated, sensitive quenched fluorescent
181 substrate-based assay as previously described [24]. Briefly, plasma (0.25 ml) was diluted into
182 low-ionic-strength buffer (20 mmol/L Tris-HCl, pH 6.5) and added to 200 *m*l ANXSepharose 4
183 Fast-Flow resin (Amersham Biosciences, GE Healthcare, Uppsala, Sweden) that removed a
184 previously characterized endogenous inhibitor of ACE2 activity. After binding and washing,
185 the resulting eluate was assayed for ACE2 catalytic activity. Duplicate samples were incubated
186 with the ACE2-specific quenched fluorescent substrate, with or without 100 mM
187 ethylenediaminetetraacetic acid. The rate of substrate cleavage was determined by
188 comparison to a standard curve of the free fluorophore, 4-amino-methoxycoumarin (MCA;
189 Sigma, MO, USA) and expressed as *r*mole of substrate cleaved/mL of plasma/min. The intra-
190 assay and inter-assay coefficient of variation was 5.6% and 11.8% respectively. Samples
191 below the limit of detection were designated 0.02 (half the LOD – ie 50% x 0.04).

Linear Model

The four analytes most associated with long COVID were identified via Linear Classification. For an arbitrary set of four analytes, let the concentration of the ith analyte after four and eight months be denoted $\mathrm{a}_\mathrm{i}^{4\mathrm{M}}$ and $\mathrm{a}_\mathrm{i}^{8\mathrm{M}}$, respectively. Linear Classification assigns two weights $\mathrm{w}_\mathrm{i}^{4\mathrm{M}}$ and $\mathrm{w}_\mathrm{i}^{8\mathrm{M}}$ to each analyte. A linear function of these concentrations and weights takes the form $(\vec{a})$ is a threshold parameter. The weights $\mathrm{w}_\mathrm{i}^{4\mathrm{M}}$ and $\mathrm{w}_\mathrm{i}^{8\mathrm{M}}$ as well as the threshold parameter θ are selected to maximise the predictive power of the linear classifier by training on the analyte data.

$$f(\vec{a}) = w_\theta + \sum_{i=1}^{N} w_i{}^{4M} a_i{}^{4M} + \sum_{i=1}^{N} w_i{}^{8M} a_i{}^{8M}$$

The data are standardized prior to training, each analyte was rescaled to have zero mean and unit variance across all participants, thus removing the possibility of a single analyte dominating the classifier. Then for a given set of four analytes we stratified our data into a training set and a testing set. Due to the moderately small sample size of 58 participants we choose our training set to include all but one participant. The excluded participant does not influence the training of the classifier and is used to test the accuracy of the classifier. There are four possible outcomes; true positive (TP: both the classifier and data say the participant had long COVID), true negative (TN: both the classifier and data say the participant had asymptomatic COVID), false positive (FP: classifier predicts the participant will have long COVID but the data disagrees) and false negative (FN: classifier predicts the participant will have asymptomatic COVID but the data disagrees). Each participant is removed in turn to be the test data, and thus a total of 58 classifiers were trained and the number of TPs, TNs, FPs and FNs were recorded. This step is known as cross validation and removes the possibility of bias in the way the training and test data is selected. Thus for a given set of four analytes we measure how well they predict long COVID with two metrics.

Firstly, the accuracy is defined as TP+TN/n and measures the proportion of test participants that had their COVID status correctly predicted. The second measure is the F1 score and is defined as $\mathrm{TP}/(\mathrm{TP} + 0.5 \times (\mathrm{FP} + \mathrm{FN}))$, which is a measure that combines recall, how many long COVID cases were correctly predicted, and precision, of all the participants predicted to

220 have long COVID how many were correct. The accuracy and F1 score are used to select which
221 combination of four analytes are most associated with long COVID. For four analytes there
222 are a total of 20475 combinations. The best four analytes are selected by identifying the four
223 analytes that achieve the highest accuracy. If two or more sets of four analytes have the
224 highest accuracy we defer to the F1 score to identify the best set of four analytes. The training
225 of the many classifiers was performed in Python3 using the Scikit-learn machine learning
226 package.

227

228 *Dimensional Reduction and Clustering Analysis*

229

230 FCS3.0 files were compensated manually using acquisition-defined matrix as a guide, and
231 gating strategy was based on unstained or endogenous controls. Live singlets were gated from
232 long COVID and asymptomatic matched controls using Flowjo v.10.7.2, samples were
233 decoded and statistical analysis between groups and unsupervised analysis was performed,
234 with matched asymptomatic controls as the primary comparator group. For unsupervised
235 analysis, the following FlowJo plugins were used: DownSample (v.3), TriMap (v.0.2),
236 Phenograph (v.3.0) and ClusterExplorer (v.1.5.9) (all FlowJo LLC). First, 100 000 events per
237 sample were downsampled from the total live singlet gate (Supp. Fig. 1). The newly generated
238 FCS files were labelled according to control or patient group (long COVID or matched controls)
239 and concatenated per group. Subsequently, 20, 000 events were taken from each grouped
240 sample by downsampling. The two new FCS files corresponding to long COVID and matched
241 controls were then concatenated for dimensionality reduction analysis using TriMap (40, 000
242 events in total). TriMap was conducted using the following parameters- to include the
243 markers CD25, CD38, CCR7, CD19, IgD, CD45RA, PD-1, TIM-3, CD4, CD57, CD127, CD27, HLA-
244 DR, CD8, CXCR5,GPR56 and Granzyme B, and using the following conditions: metric =
245 euclidean, nearest neighbours = 15, and minimum distance = 0.5. Phenograph plugin was then
246 used to determine clusters of phenotypically related cells. The same markers as TriMap and
247 parameters k = 152 and Run ID = auto was used for analysis. Finally, ClusterExplorer plugin
248 was used to identify the phenotype of the clusters generated by phenograph.

249 *Statistical Analysis*

250

251 All column graphs are presented as medians with inter-quartile ranges. One-way ANOVA
252 utilising Kruskal-Wallis and Dunn's correction for multiple comparisons was used for serum
253 analyte analysis. Wilcoxon paired *t* test was used to analyse statistical data employing Prism
254 9.0 (GraphicPad, La Jolla, CA, USA) software. For unpaired samples Mann-Whitney U test was
255 used. RStudio version 1.2.1335 was used to generate correlograms utilizing corrplot package
256 with Spearman's method used to determine Rho values and cor.mtest with 95% confidence
257 level was used to determined *p* values. *p* values <0.05 were considered significant (*<0.05,
258 **<0.01, ***<0.001, and ****<0.0001).

259

260

261 **Results**

262

263 *The ADAPT cohort*

264

265 All participants were drawn from a single prospective observational cohort designed to follow
266 patients after acute COVID-19 infection. The ADAPT study[16] enrolled all adults with SARS-CoV-
267 2 infections confirmed by polymerase chain reaction (PCR) at St Vincent's Hospital community
268 based testing clinics in Sydney (Australia), who were able to be contacted and agreed to
269 participate (70.5% of those diagnosed). Participants were prospectively systematically
270 assessed according to a pre-defined schedule which included routine visits at enrolment,
271 month 4 and month 8. The study was commenced as soon as ethics approval was obtained
272 and for the majority of participants in the first wave, the first enrolment visit occurred
273 somewhere between month 2-3 post infection (median 79 days)[16,21] with 93.6% and 84.5% of
274 participants completing subsequent month 4 (median 128 days) and month 8 (median 232
275 days) visits timed from the date of diagnosis of their initial infection (Table 1). The drop-out
276 rate has been very low to date: approximately 9.4% at 12 months. Participants were
277 designated as Long COVID (LC) based on the presence of one of three major symptoms
278 (fatigue, dyspnoea or chest pain) at the month 4 timepoint. These participants were age and
279 gender matched with 31 participants from the same cohort who reported no ongoing
280 symptoms at the 4 month timepoint (designated as matched controls (MC)) (Suppl. Table 1).

281 The matched control group were symptomatic at acute initial infection (Suppl. Table 2), but
282 did not develop long COVID and were asymptomatic at month 4. In terms of clinical recovery,
283 there was a 10% trend towards some improvement over time, but this was not statistically
284 significant, Fischer Exact p=0.44. With regard to the external enrolments (29.5%), we also
285 allowed enrolment in the study for some participants who were diagnosed outside of the St
286 Vincent's testing clinic. In these cases participants were required to have documented
287 evidence of a positive swab and entered into the same schedule of assessments for follow-
288 up.
289
290 *Elevated levels of pro-inflammatory cytokines following SARS-CoV-2 infection that are*
291 *maintained at 8 months*
292
293 In order to assess biomarkers associated with recovery from symptomatic SARS-CoV-2
294 infection 28 analytes were analysed from serum of convalescent subjects with long COVID
295 (LC) or asymptomatic matched controls (MC) at 4 months post-infection. Individuals that were
296 PCR positive for prevalent common cold coronaviruses (HCoV) and sampled within the same
297 time frame following infection (3-6months) and unexposed healthy donors sampled prior to
298 December 2019, were used as comparator groups. Six pro-inflammatory cytokines were
299 highly elevated in the two COVID-19 convalescent ADAPT groups compared to control groups
300 (Fig. 1A), while no difference was observed with the 22 other analytes (Supp. Fig. 2A&B). It is
301 noteworthy that no difference was observed between long COVID and asymptomatic
302 matched controls for any of the individual analytes examined at 4 months post infection. Type
303 I interferon-β (IFN-β) was 7.92-fold (median values shown in Supp. Fig 2C) and 7.39-fold
304 higher in LC and MC compared to HCoV (both $p<0.0001$), and 7.32- and 6.83-fold higher
305 compared to unexposed healthy controls (both $p<0.0001$, respectively). Type III interferon-
306 lamdba1 (IFN-λ1) was 2.44-fold and 3.24-fold higher in LC and MC compared to HCoV (both
307 $p<0.0001$), and 2.42- and 3.21-fold higher compared to unexposed, ($p<0.01$ *and* $p<0.001$,
308 respectively). This was also the case for IL-8 (CXCL8) levels; LC (3.43-fold) and MC (3.56-fold)
309 higher compared to both HCoV and unexposed (all $p<0.0001$). Interferon gamma induced
310 chemokine CXCL10 (IP-10) was elevated in LC and compared to HCoV 2.15-fold ($p<0.001$) and
311 unexposed 3.20-fold ($p<0.001$), similarly MC group were also significantly higher than
312 comparator groups (1.7-fold, $p<0.05$ and 3.06-fold, $p<0.01$, respectively). CXCL9(MIG-1) was

313 1.69-fold higher than unexposed (*p<0.*05), and soluble TIM-3 levels were only elevated in LC
314 and not MC when compared to HCoV comparator group 1.46-fold (*p<0.05*).
315
316 To elucidate the persistence of these elevated analytes over a longer period of time, serum
317 from all available subjects at 8 months were examined and compared with levels at 4 months
318 (Fig. 1B). Decreases of both IFN-β and IFN-λ1 in MC group at 8 months were observed 4.4-
319 fold (9.36[6.25-21.35] pg/mL, *p<0.0001*) and 1.8-fold (10.51[7.72-16.30] pg/mL, *p<0.05*)
320 respectively. However this was not the case for LC group, where IFN-β slightly decreased by
321 1.5-fold (22.32[18.75-49.07] pg/mL) and IFN-λ1 slightly increased by 1.05-fold (16.87[9.60-
322 24.09] pg/mL) at 8 months, albeit not statistically significant. Overall, type I and type III
323 interferons remained high in LC subjects and was significantly higher than the MC group at 8
324 months post-infection (*p<0.05*, for both interferons) and was higher than HCoV and
325 unexposed groups (Suppl. Fig. 3A). Reduction of CXCL9 (MIG-1), CXCL10 (IP-10), IL-8 and sTIM-
326 3 were evident in both groups at 8 months. There were also decreases in the levels of some
327 of the 22 analytes not significantly different to control groups at month 4 by month 8 (Supp.
328 Fig. 3B&C).
329
330 *Elevated levels of ACE2 in SARS-CoV-2 convalescent subjects*
331
332 Plasma ACE2 activity levels have recently been shown to be elevated out to a median of 114
333 days following SARS-CoV-2 infection[25], and we sought to investigate whether this was the
334 case within the ADAPT longitudinal cohort at 3-, 4- and 8-months post-infection. At 3-months
335 post-infection, median plasma ACE2 activity was significantly higher in both LC and MC groups
336 compared to HCoV (LC=1.92 [0.02-7.36], MC=2.47 [0.63-5.94] *versus* HCoV =0.02 [0.02-1.99]
337 pmol/min/mL, *p<0.01* and *p<0.001* respectively). These levels remained elevated at 4 months
338 post-infection in both groups (LC=1.75 [0.02-5.01], MC=2.62 [0.36-6.3] *versus* HCoV =0.02
339 [0.02-1.99] pmol/min/mL, *p<0.05* and *p<0.001* respectively). Plasma ACE2 activity decreased
340 at 8 months for LC (0.02[0.02-1.36] pmol/min/mL) and MC (0.18[0.02-3.46] pmol/min/mL).
341 No significant difference was observed between these groups at 3, 4 and 8-months or when
342 compared to HCoV. It appears that increased plasma ACE2 activity is specific to SARS-CoV-2
343 infection and is not a common feature of other coronaviruses.
344

345 *High inter-analyte correlations in long COVIDs at 4 months*

346

347 Immune responses to viral infections involve communication of multiple specialised cell
348 populations through direct interactions, secreted cytokines and various mediators in a
349 concerted effort to combat the infectious agent. To obtain a broader scope of systems-level
350 interactions during convalescence, we generated correlation matrices from 29 analytes
351 measured at 4 months post-infection (Fig. 3). This analysis revealed multiple positive
352 correlations across a spectrum of 18 markers including: IFN-α2, IFN-λ1, IFN-λ2/3
353 (interferons), IL-1β, IL-33 (IL-1 family cytokines); IL-12p70, MCP-1 (CCL2), IL-6, TGF-β1, CXCL-
354 9 (myeloid activation), PD-1, sCD25 (T cell activation,) IFN-γ, TNF-α (Type 1 (Th1) cytokines);
355 IL-5, IL-9, IL-13, (Type 2 (Th2) cytokines) and GM-CS in the LC group most of which were not
356 present in the asymptomatic MC group. (Fig. 3 A&B). However at 8 months this profile had
357 altered in both groups (Supp. Fig. 4A&B). This profile was not observed in HCoV (Supp. Fig.
358 4C) or unexposed donors (Supp. Fig. 4D).

359

360 *Linear modelling defines 4 features most highly associated with long COVID*

361

362 To identify a subset of the 29 analytes that are most strongly associated with long COVID we
363 use linear classification. We trained our linear classifiers with data from 58 participants (29
364 with long COVID and 29 matched controls) with analyte concentrations after 4 and 8 months.
365 The optimal number of analytes giving the highest accuracy when using a linear classifier to
366 predict whether a participant suffered from long or asymptomatic COVID was 4. By
367 exhaustively searching through all 20475 combinations of 4 analytes we found that the 4
368 analytes IL-6, IFN-γ, MCP-1 (CCL2) and VCAM-1 gave an accuracy of 75.9%, and an F1 score
369 of 0.76 (Fig. 3C). Beyond identifying the four analytes that are most highly associated with
370 long COVID, linear classifiers offer a simple test to predict whether or not a patient will suffer
371 from long COVID. Linear classifiers define what is known as a decision boundary. A patient's
372 4 analyte concentrations at 4 and 8 months will lie on either side of this boundary, and which
373 side it is will determine whether or not the patient is predicted to experience long or
374 asymptomatic COVID. The decision boundary for the 4 best analytes is eight-dimensional due
375 to both the 4- and 8-month data. However, we can visualise the domain boundary with two-

376 dimensional projections of IFN-γ, against MCP-1, MCP-1 against VCAM-1 and VCAM-1 against
377 IL-6 as shown in Fig, 3D.
378
379 *Highly activated innate cells combined with absence of un-activated naïve T and B cell subsets-*
380 *a distinguishing feature of long COVID*
381
382 To investigate differences in immune cell phenotypes between LC and MC, a 19-parameter
383 phenotyping panel was developed and utilised for PBMC samples from LC and MC groups at
384 3 months and 8 months following infection, with MC group being the primary comparator.
385 Dimensional reduction via TriMap coupled with Phenograph clustering identified 24 distinct
386 cell clusters at month 3 and 21 clusters at month 8 (example of cell cluster identification
387 shown in Suppl. Fig. 5A); comprising T cells, B cells, NK cells and non-lymphoid (myeloid) cells
388 (Supp. Fig. 5B&C). Of the 24 subsets identified at month 3, five were absent in LC subjects
389 (Fig. 4A), these include:- 1) CD127^{low}GzmB-CCR7+CD45RA+CD27+ naïve CD8+ T cells, 2) CD57+
390 highly cytotoxic (GPR56+GzmB+) CD8+ T cells, 3) CD127^{low}TIM-3^{neg}CCR7+CD45RA+CD27+
391 naïve CD4+ T cells, 4) CD3+CD4-CD8- innate like T cells (may comprise NKT cells and γδ-T cells),
392 and 5) naïve CD38^{low}CD27-IgD+ B cells. At 8 months only 3 populations remined absent: 1)
393 CD127^{low}GzmB-CCR7+CD45RA+CD27+ Naïve CD8+ T cells, 2) CD127^{low}TIM-
394 3^{neg}CCR7+CD45RA+CD27+ Naïve CD4+ T cells, and 3) CD38^{low}CD27-Naïve IgD+ B cells (Fig. 4B).
395 These findings are indicative of ongoing inflammation through to 8 months in LC subjects.
396 There was no reconstitution of these un-activated naïve T and B cell subsets even at 8 months
397 post-infection in peripheral blood, however other naïve subsets that appear more activated
398 are present in LC. All concatenated patients sampled (7 LC, 7 MC and 7 unexposed)
399 contributed to all populations identified (Suppl. Fig 6A-D). These un-activated T and B cells
400 are evident in unexposed and MC comparator groups.
401
402 Highly activated non-lymphoid cells (defined by the co-expression of CD38+HLA-DR+)
403 remained elevated at 8 months in LC (Fig. 4C). In contrast, activation markers on non-
404 lymphoid cells decreased over time in asymptomatic matched controls (median 3 months=
405 11.2[3.43-12.5] and 8 months= 8.12[2.07-9.23], $p<0.05$). Activate monocytes (CD14+CD16+)
406 were higher in LC compared to MC at both 4 and 8 months (3.74-fold, $p<0.05$, and 4-fold,
407 $p<0.01$, respectively) (Fig. 4D). Plasmacytoid dendritic cells expressing activated markers

408 CD86+ CD38+ were also higher in LC at both timepoints (2.14-fold, *p<0.01* and 4.88-fold,
409 *p<0.001,* respectively) (Fig. 4E). No differences in activation of myeloid dendritic cells were
410 observed (Suppl. Fig. 7A). Notably, the immune checkpoint receptor commonly associated
411 with T cell activation/exhaustion PD-1 was highly expressed on CD8+ T cells at 3 and 8 months
412 in LC compared to MC (3.04-fold, *p<0.001* and 2.86-fold, *p<0.0001*)(Fig. 4F). T cell exhaustion
413 marker Tim-3 was also higher in LC at month 3 (1.6-fold, *p<0.05*), but decreased at month 8
414 (Fig. 4G). However, no difference was observed between LC and MC as judged by co-
415 expression of PD-1 and Tim-3 on CD4+ and CD8+ T cells (suppl. Fig. 7B).
416
417 **Discussion**
418
419 Our study demonstrates several unique and important findings in relation to understanding
420 long-term outcomes from COVID-19. We show that convalescent profiles post COVID-19 are
421 different from other coronaviruses. Several cytokines, mostly type I and III interferons, but
422 also chemokines down stream of IFN-γ, were highly elevated in people following resolution
423 of active SARS-CoV-2 infection compared to HCoV and unexposed controls at 4 months. IFN-
424 β and IFN-λ1 remained high in long COVID to 8 months after initial infection, while the levels
425 of these two cytokines had begun to resolve in the matched controls. This finding was
426 observed irrespective of persistent long COVID symptoms. We also demonstrated an
427 elevation of plasma ACE2 levels at 4 months post COVID infection with the degree of
428 elevation not being different between those with and without long COVID symptoms. ACE-2
429 levels trended towards normal by 8 months. Further, by employing a linear classification
430 model we identified 4 biomarkers highly associated with long COVID: IL-6, IFN-γ, MCP-1 (CCL2)
431 and VCAM-1. This indicates that components of the acute inflammatory response, T cell,
432 Myeloid cell and vascular endothelium activation are associated with long COVID.
433 Interestingly, immune cell phenotyping revealed chronic activation of: a subset of CD8 T cells
434 with expansion of PD-1 and Tim-3 positive subsets; plasmacytoid dendritic cells; and
435 monocytes (CD38 and HLA-DR co-expression) in long COVID patients. Levels of these subsets
436 are increased in long COVID patients relative to matched controls at 3 months with
437 perturbations persisting out to 8 months. Further, there is an apparent turnover of un-
438 activated naïve T and B cell subsets in peripheral blood in long COVID subjects over the same
439 time period. This suggests maintenance of long-lasting inflammation and immune activation

440 in these individuals. Together these findings suggest that SARS-CoV-2 infection exerts a
441 unique residual prolonged effect on aspects of both the innate and adaptive immune systems,
442 and that this may be driving the post COVID symptomology known as 'long COVID'.
443
444 Type I and III Interferons, namely IFN-β and IFN-λ1 were highly elevated in our COVID-19
445 convalescent subjects compared to HCOV and unexposed subjects,. While their levels
446 decreased over time in recovered subjects they remained high in those with long COVID. IFN-
447 β can be expressed by most cell types and can be induced early during viral infection, before
448 most IFN-α subtypes [26-28]. IFN-λ1 or IL-29 can also be induced in a broad range of cell types
449 including epithelial cells by stimulation of toll-like receptors (TLRs), Ku70, and RIG-1-like
450 receptor [29], however, as for type I interferons, the main producers were found to be type 2
451 myeloid dendritic cells (plasmacytoid DCs)[30-32]. Type I interferon receptors are widely
452 expressed on immune cells, and IFN-β responses during influenza and SARS-CoV-1 infection
453 can result in immunopathology [33,34]. In contrast, IFN-λ receptors are mainly expressed at
454 epithelial barriers and confer localised antiviral protection [35,36]. The morbidity of acute COVID-
455 19 infection appears to correlate with high expression of type I and III interferons in the lung
456 of patients [37]. Furthermore IFN-λ produced by dendritic cells in lungs of mice in response to
457 synthetic viral RNA is associated with damage to lung epithelium [38], and IFN-λ signalling can
458 hamper lung repair during influenza infection of mice [39]. Diminished type I interferon and
459 enhanced IL-6 and TNF-α responses in severe acute COVID-19 patients were reported by
460 Hadjadj *et al* [38]. Our cohort of long COVID subjects, however, mostly consisted of patients
461 with mild and moderate initial illness, and maintenance of elevated type I and III interferon
462 levels at 8 months demonstrates ongoing inflammation within these individuals and is
463 consistent with our observation of prolonged activation of plasmacytoid dendritic cells.
464
465 CXCL9 (MIG-1) and CXCL10 (IP-10) are pro-inflammatory T cell chemo-attractants that are
466 expressed by monocytes, endothelial cells, and fibroblasts downstream of interferon-gamma
467 signalling [40]. Upon engagement with their receptor CXCR3, mainly expressed on T cells, they
468 drive naïve CD4+ T cell differentiation into effector Th1 cells, as well as recruiting CD4+ and
469 CD8+ T cells in sites of tissue damage, which has been documented during vascular disease
470 and pulmonary fibrosis [41,42]. Both chemokines have been implicated in symptomatic acute

471 coronavirus infections, including COVID-19, where levels were higher in severe compared to
472 asymptomatic subjects [43]. Similarly, CXCL8 (IL-8) produced by lung epithelium and airway
473 resident macrophages, has previously been shown to be released by lung cells after activation
474 by spike proteins from SARS-CoV-1 [44]. CXCL8 acts as a chemoattractant that facilitates
475 migration of inflammatory T cells and neutrophils, that may contribute to alveolar damage
476 [45,46]. Further the elevation of CXCL9 (MIG-1) and CXCL10 (IP-10) is consistent with persistent
477 absence of un-activated naïve cell subsets in peripheral blood that we observed in the long
478 COVID group.
479
480 Increased shedding or cleavage of ectodomains of cell surface proteins during inflammation
481 has been well documented [47,48]. Increased levels of soluble T-cell immunoglobulin mucin
482 domain-3 (sTIM-3) a marker of T-cell activation and exhaustion has been reported in chronic
483 viral infections (i.e., HIV, hepatitis B and C)[49-51] where there is persistent antigenic stimulation,
484 and recently in COVID-19 infection, where patients admitted to ICU had high plasma levels of
485 sTIM-3 and myeloperoxidase (MPO) [52]. Elevated sTIM-3 levels were seen in those with long
486 COVID, but not recovered subjects or those with other HCoV, and is entirely consistent with
487 our observation of expanded subsets of memory CD8 T cells expressing Tim-3 and PD-1 in the
488 same group who appear to have chronic T cell activation and potentially T cell exhaustion
489 perhaps partially driven by the prolonged presence of activated antigen presenting cells and
490 the downstream drivers of T cell activation such as type I interferons and the IFN-γ driven
491 chemokines CXCL9 and CXCL10 . Another membrane bound protein found to be 'shed' during
492 COVID-19 is ACE2, which is also an important receptor for viral host cell entry [53]. Patel *et al*
493 showed that shedding of ACE2 from the cell membrane resulted in increased plasma ACE2
494 activity levels in convalescent COVID-19 subjects [25]. Our data corroborates with this finding,
495 where elevated plasma ACE2 activity occurred in COVID-19 infection regardless of symptom
496 severity, at a higher level than HCoV and unexposed, persisted at 4 months, and decreased at
497 8 months in most subjects.
498
499 We observed high inter-analyte correlations in long COVIDs at 4 months that was not
500 observed in MC, HCoV or unexposed comparator groups. These correlations reinforced the
501 idea that there was interrelated activation of acute phase inflammatory , innate and adaptive
502 immune responses. This profile was less prominent at 8 months. In order to assess across all

503 combinations of analytes those most strongly associated with long COVID, a linear
504 classification model was employed. A combination of 4 biomarkers, IL-6, IFNγ, MCP-1 (CCL2),
505 and VCAM-1 was identified as being the most accurate correlate of long COVID. These
506 analytes could be more prominent at the acute stage of infection. IL-6 is a pleiotropic
507 mediator that has an effect on inflammation and immune activation [54]. IFNγ produce by NK
508 and T cells have anti-viral activity and levels are associate with a heightened inflammatory
509 state. High IL-6/IFN-gamma ratio was found to be associated with severe disease in COVID-19
510 patients at acute infection [55]. MCP-1 (CCL2) is a chemokine that is produced by many cell types
511 (endothelial, fibroblasts, epithelial, monocytic etc [56]) and acts as a chemoattractant for
512 monocyte, NK, neutrophils, basophils, B and T cells and is involved in T cell activation [57].
513 VCAM-1 mediates adhesion of immune cells to the endothelium and has been implicated in
514 inflammatory processes ranging from atherosclerosis and rheumatoid arthritis [58]. Recent
515 meta-analysis of > 300,000 single-cell transcriptomic profiles from COVID-19-affected lungs
516 have found that IFNγ together with TNFα drive an drive a IP-10+CCL2+ macrophage
517 phenotype expanded in severe COVID-19 lungs and inflammatory diseases with tissue
518 inflammation [59]. These analytes could be more prominent at the acute stage of infection and
519 warrants further investigation. The observation that the best correlate of the presence of long
520 COVID is an eclectic combination of biomarkers reinforces breadth of host response pathways
521 that are activated during long COVID.
522
523 Finally, T cell activation via surrogate markers CD38 and HLA-DR, T cell exhaustion, together
524 with an increase in B cell plasmablasts during severe COVID-19 has been reported [60-62]. We
525 found no difference between activation levels as judged by cell surface expression of these
526 molecules on T cells between long COVID and asymptomatic (data not shown) but identified
527 highly activated myeloid cells comprising monocytes and dendritic cells. Elevated levels of
528 CD38+HLA-DR+ myeloid cells decreased over time within asymptomatic subjects but
529 remained unchanged in those with long COVID. It is known that type I and type III interferons
530 upregulate MHC-class expression including HLA-DR [63]. When an unbiased large-scale
531 dimensional reduction using triplets (TriMap) approach was used, we identified 5 population
532 clusters that were absent in those with long COVID. These consisted of mostly un-activated
533 naïve T and B cell subsets (CD4 (CD127^{low}TIM-3^{neg}CCR7+CD45RA+CD27+), CD8
534 (CD127^{low}GzmB-CCR7+CD45RA+CD27+) and CD127^{low}TIM-3^{neg}CD38^{low}CD27-IgD+ B cells), as

535 well as double negative (CD4-CD8-) Innate-like T cells and highly cytotoxic
536 CD57+GPR56+GzmB+ CD8+ T cells at 3 months. The depletion of these un-activated naïve
537 subsets persisted at the 8-month time-point. Other more activated naïve subsets were
538 present at 3 months in both groups and remained at 8 months. Naïve T and B cell expressed
539 low to mid-levels of CD127 in MC and unexposed (Suppl. Fig. 6E&F), but $CD127^{low}$Tim-3 ^{neg}
540 phenotype was absent in LC. Taken together these observations suggest persistent
541 recruitment of activated naïve T cells, potentially due to by-stander activation secondary to
542 underlying inflammation and/or antigen presentation by activated plasmacytoid dendritic
543 cells or monocytes, pushing naïve cells into an activated state, thereby depleting the un-
544 activated naïve T and B cell pools in those with long COVID. The ultimate result of this chronic
545 stimulation may be the observed expansion of exhausted PD-1 or Tim-3 positive CD8+ T cells.
546 Bystander activation of un-activated naïve subsets into more activated phenotypes is
547 consistent with observations in subsets of patients during acute severe COVID-19 infection
548 [64,65].
549
550 Our data does have some limitations. Firstly, although our long COVID 'cases' and controls
551 were matched for two major characteristics (age and gender) it is possible that the differences
552 observed reflect other key demographic or other factors between the groups. While long
553 COVID group had a higher representation of those with severe acute disease (8 LC and 2 MC),
554 sensitivity analyses excluding these patients did not substantially alter the statistical
555 significance any of our major associations. Also, due to timing of ethics approval and trial
556 setup we were not able to collect samples at acute infection. We are therefore unable to
557 differentiate whether the elevations in biomarkers seen during convalescence correlate with
558 levels during the acute infection. A review of the literature of biomarkers during acute
559 infection suggests that while some of the alterations observed here are potentially consistent
560 with a hypothesis that the major drivers of the levels of biomarkers in convalescence are
561 those in the acute infection, others are not. For instance diminished type I interferon and
562 enhanced IL-6 and TNF-α responses in severe acute COVID-19 patients have been reported
563 by Hadjadj *et al* [38]. Our cohort consisted mostly of patients with mild and moderate initial
564 illness and among our cohort those with long COVID do not have elevated levels of IL6 or TNF,
565 but do have elevated levels of both type I and III interferon. Although 7 of the 29 serum
566 analytes were abnormally elevated, key cytokines implicated in other fatigue syndromes, such

567 as Pentraxin-3(PTX3), TGF-β1, and IL-13[66,67], were not. This suggests long COVID may be
568 differentiated from other post viral fatigue syndromes, however, our results require
569 validation in other cohorts of long COVID to ensure their replicability. Finally, our definition
570 of 'long COVID' cases was internally set given the lack of international consensus regarding
571 this and thus is not definitive. Nevertheless, the inclusion of three of the commonest
572 persisting symptoms and the blinding of cases and controls from the laboratory scientists
573 helps to ensure the validity of our findings. Additional strengths of our study include the
574 prospective protocol defined collection of samples and data and the establishment of an
575 ADAPT–C cohort of individuals affected contemporaneously with other human coronaviruses
576 to act as comparators.
577
578 In summary, our data strongly suggest there is an ongoing, sustained inflammatory response
579 following even apparently mild to moderate acute COVID-19 infection, which is not found
580 following infections with other prevalent coronaviruses. This process may be asymptomatic
581 and therefore subclinical and takes more than 8 months to resolve. In those with persistent
582 long COVID symptoms as distinct from asymptomatic convalescence, there is a sustained
583 increase, out to at least 8 months, in highly activated myeloid cells, particularly monocytes
584 and plasmacytoid dendritic cells, that likely contribute to the sustained elevation of
585 inflammatory cytokines (particularly type I and III interferons), despite the majority of cohort
586 suffering initial mild-moderate illness. Further, there was evidence of sustained activation of
587 the adaptive immune system manifesting in perturbations of both naïve and
588 memory/effector cell compartments with absence of the un-activated naïve T and B cells and
589 increases in the expression of checkpoint molecules PD-1 and Tim-3 on CD8+ memory T cells.
590 This widespread activation signature is also reflected in the identification of serum levels of
591 IL-6, IFN-γ, MCP-1 (CCL2), and VCAM-1 as the 4 features that most strongly associated with
592 long COVID. The drivers of this activation require further investigation but possibilities include
593 persistence of antigen, an autoimmune phenomenon driven by unexpected antigenic cross
594 reactivity, or a reflection of damage repair. To our knowledge, this is the first demonstration
595 of an abnormal immune profile in COVID-19 patients at extended time points post infection
596 and provides clear support for the existence of a syndrome of 'long COVID'. While our findings
597 require validation in other similar cohorts, they provide an important foundation to

598 understanding further the pathophysiology of this syndrome and potential therapeutic
599 avenues for intervention.
600
601 **Acknowledgements**
602
603 The authors thank the staff at the St Vincent's Institute for Applied Medical Research-Clinical
604 Trials Unit for their expertise in specimen processing and bio-banking. We appreciate grant
605 support from the St Vincent's Clinic Foundation, the Curran Foundation, the Rapid Response
606 Research Fund (UNSW) and the Medical Research Futures Fund (Australia). SKP and LMB-
607 NHMRC programme grant APP 1055214 (LMB), Medical Research Future Fund award GNT
608 1175865, Austin Medical Research Foundation Grant. DBW- NSF-DMA grant 1902854, SJK and
609 JAJ- the Victorian Government, MRFF Award (2005544), NHMRC program grant (1149990
610 (SJK, ADK), NHMRC Fellowships (1136322-SJK, 1123673-JAJ).
611
612
613 **References**
614
615 1. Sudre, C.H.*, et al.* Attributes and predictors of long COVID. *Nat Med* **27**, 626-631
616 (2021).
617 2. Wiersinga, W.J., Rhodes, A., Cheng, A.C., Peacock, S.J. & Prescott, H.C.
618 Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease
619 2019 (COVID-19): A Review. *JAMA* **324**, 782-793 (2020).
620 3. Carvalho, T., Krammer, F. & Iwasaki, A. The first 12 months of COVID-19: a timeline of
621 immunological insights. *Nat Rev Immunol* **21**, 245-256 (2021).
622 4. Zhu, N.*, et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N*
623 *Engl J Med* **382**, 727-733 (2020).
624 5. Baig, A.M. Deleterious Outcomes in Long-Hauler COVID-19: The Effects of SARS-CoV-
625 2 on the CNS in Chronic COVID Syndrome. *ACS Chem Neurosci* **11**, 4017-4020 (2020).
626 6. Baig, A.M. Chronic COVID syndrome: Need for an appropriate medical terminology for
627 long-COVID and COVID long-haulers. *J Med Virol* **93**, 2555-2556 (2021).
628 7. Peluso, M.J.*, et al.* Rapid implementation of a cohort for the study of post-acute
629 sequelae of SARS-CoV-2 infection/COVID-19. *medRxiv* (2021).
630 8. Sykes, D.L.*, et al.* Post-COVID-19 Symptom Burden: What is Long-COVID and How
631 Should We Manage It? *Lung* **199**, 113-119 (2021).
632 9. Hui, D.S.*, et al.* Impact of severe acute respiratory syndrome (SARS) on pulmonary
633 function, functional capacity and quality of life in a cohort of survivors. *Thorax* **60**, 401-
634 409 (2005).
635 10. Hui, D.S. An overview on severe acute respiratory syndrome (SARS). *Monaldi Arch*
636 *Chest Dis* **63**, 149-157 (2005).

637 11. Lam, M.H.*, et al.* Mental morbidities and chronic fatigue in severe acute respiratory
638 syndrome survivors: long-term follow-up. *Arch Intern Med* **169**, 2142-2147 (2009).
639 12. Ahmed, H.*, et al.* Long-term clinical outcomes in survivors of severe acute respiratory
640 syndrome and Middle East respiratory syndrome coronavirus outbreaks after
641 hospitalisation or ICU admission: A systematic review and meta-analysis. *J Rehabil*
642 *Med* **52**, jrm00063 (2020).
643 13. Townsend, L.*, et al.* Persistent fatigue following SARS-CoV-2 infection is common and
644 independent of severity of initial infection. *PLoS One* **15**, e0240784 (2020).
645 14. Halpin, S.J.*, et al.* Postdischarge symptoms and rehabilitation needs in survivors of
646 COVID-19 infection: A cross-sectional evaluation. *J Med Virol* **93**, 1013-1022 (2021).
647 15. Carfi, A., Bernabei, R., Landi, F. & Gemelli Against, C.-P.-A.C.S.G. Persistent Symptoms
648 in Patients After Acute COVID-19. *JAMA* **324**, 603-605 (2020).
649 16. Darley, D.R.*, et al.* Persistent symptoms up to four months after community and
650 hospital-managed SARS-CoV-2 infection. *Med J Aust* **214**, 279-280 (2021).
651 17. Cortinovis, M., Perico, N. & Remuzzi, G. Long-term follow-up of recovered patients
652 with COVID-19. *Lancet* **397**, 173-175 (2021).
653 18. Huang, C.*, et al.* 6-month consequences of COVID-19 in patients discharged from
654 hospital: a cohort study. *Lancet* **397**, 220-232 (2021).
655 19. Sudre, C.H.*, et al.* Anosmia and other SARS-CoV-2 positive test-associated symptoms,
656 across three national, digital surveillance platforms as the COVID-19 pandemic and
657 response unfolded: an observation study. *medRxiv* (2020).
658 20. Tenforde, M.W.*, et al.* Characteristics of Adult Outpatients and Inpatients with COVID-
659 19 - 11 Academic Medical Centers, United States, March-May 2020. *MMWR Morb*
660 *Mortal Wkly Rep* **69**, 841-846 (2020).
661 21. Darley, D.*, et al.* Limited recovery from post-acute sequelae of SARS-CoV-2 (PASC) at
662 eight months in a prospective cohort. *medRxiv*, 2021.2003.2029.21254211 (2021).
663 22. Lopez-Leon, S.*, et al.* More Than 50 Long-Term Effects of COVID-19: A Systematic
664 Review and Meta-Analysis. *Res Sq* (2021).
665 23. Phetsouphanh, C.*, et al.* Maintenance of Functional CD57+ Cytolytic CD4+ T Cells in
666 HIV+ Elite Controllers. *Front Immunol* **10**, 1844 (2019).
667 24. Lew, R.A.*, et al.* Angiotensin-converting enzyme 2 catalytic activity in human plasma is
668 masked by an endogenous inhibitor. *Exp Physiol* **93**, 685-693 (2008).
669 25. Patel, S.K.*, et al.* Plasma ACE2 activity is persistently elevated following SARS-CoV-2
670 infection: implications for COVID-19 pathogenesis and consequences. *Eur Respir J*
671 (2021).
672 26. Juang, Y.T.*, et al.* Primary activation of interferon A and interferon B gene transcription
673 by interferon regulatory factor 3. *Proc Natl Acad Sci U S A* **95**, 9837-9842 (1998).
674 27. Katze, M.G. Regulation of the interferon-induced PKR: can viruses cope? *Trends*
675 *Microbiol* **3**, 75-78 (1995).
676 28. Katze, M.G., He, Y. & Gale, M., Jr. Viruses and interferon: a fight for supremacy. *Nat*
677 *Rev Immunol* **2**, 675-687 (2002).
678 29. Egli, A., Santer, D.M., O'Shea, D., Tyrrell, D.L. & Houghton, M. The impact of the
679 interferon-lambda family on the innate and adaptive immune response to viral
680 infections. *Emerg Microbes Infect* **3**, e51 (2014).
681 30. Zhang, S., Kodys, K., Li, K. & Szabo, G. Human type 2 myeloid dendritic cells produce
682 interferon-lambda and amplify interferon-alpha in response to hepatitis C virus
683 infection. *Gastroenterology* **144**, 414-425 e417 (2013).

684 31. Wolk, K.*, et al.* Maturing dendritic cells are an important source of IL-29 and IL-20 that
685 may cooperatively increase the innate immunity of keratinocytes. *J Leukoc Biol* **83**,
686 1181-1193 (2008).
687 32. Yin, Z.*, et al.* Type III IFNs are produced by and stimulate human plasmacytoid dendritic
688 cells. *J Immunol* **189**, 2735-2745 (2012).
689 33. Channappanavar, R.*, et al.* Dysregulated Type I Interferon and Inflammatory
690 Monocyte-Macrophage Responses Cause Lethal Pneumonia in SARS-CoV-Infected
691 Mice. *Cell Host Microbe* **19**, 181-193 (2016).
692 34. Davidson, S., Crotta, S., McCabe, T.M. & Wack, A. Pathogenic potential of interferon
693 alphabeta in acute influenza infection. *Nat Commun* **5**, 3864 (2014).
694 35. Parker, B.S., Rautela, J. & Hertzog, P.J. Antitumour actions of interferons: implications
695 for cancer therapy. *Nat Rev Cancer* **16**, 131-144 (2016).
696 36. Broggi, A., Tan, Y., Granucci, F. & Zanoni, I. IFN-lambda suppresses intestinal
697 inflammation by non-translational regulation of neutrophil function. *Nat Immunol* **18**,
698 1084-1093 (2017).
699 37. Broggi, A.*, et al.* Type III interferons disrupt the lung epithelial barrier upon viral
700 recognition. *Science* **369**, 706-712 (2020).
701 38. Hadjadj, J.*, et al.* Impaired type I interferon activity and inflammatory responses in
702 severe COVID-19 patients. *Science* **369**, 718-724 (2020).
703 39. Major, J.*, et al.* Type I and III interferons disrupt lung epithelial repair during recovery
704 from viral infection. *Science* **369**, 712-717 (2020).
705 40. Tokunaga, R.*, et al.* CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A
706 target for novel cancer therapy. *Cancer Treat Rev* **63**, 40-47 (2018).
707 41. Kameda, M.*, et al.* CXCL9, CXCL10, and CXCL11; biomarkers of pulmonary
708 inflammation associated with autoimmunity in patients with collagen vascular
709 diseases-associated interstitial lung disease and interstitial pneumonia with
710 autoimmune features. *PLoS One* **15**, e0241719 (2020).
711 42. Metzemaekers, M., Vanheule, V., Janssens, R., Struyf, S. & Proost, P. Overview of the
712 Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-
713 Inducible CXC Chemokine Receptor 3 Ligands. *Front Immunol* **8**, 1970 (2017).
714 43. Chi, Y.*, et al.* Serum Cytokine and Chemokine Profile in Relation to the Severity of
715 Coronavirus Disease 2019 in China. *J Infect Dis* **222**, 746-754 (2020).
716 44. Chang, Y.J., Liu, C.Y., Chiang, B.L., Chao, Y.C. & Chen, C.C. Induction of IL-8 release in
717 lung cells via activator protein-1 by recombinant baculovirus displaying severe acute
718 respiratory syndrome-coronavirus spike proteins: identification of two functional
719 regions. *J Immunol* **173**, 7602-7614 (2004).
720 45. Barnes, P.J. Chronic obstructive pulmonary disease. *N Engl J Med* **343**, 269-280 (2000).
721 46. Folkerts, G., Kraneveld, A.D. & Nijkamp, F.P. New endogenous CXC chemokine ligands
722 as potential targets in lung emphysema. *Trends Pharmacol Sci* **29**, 181-185 (2008).
723 47. Lichtenthaler, S.F., Lemberg, M.K. & Fluhrer, R. Proteolytic ectodomain shedding of
724 membrane proteins in mammals-hardware, concepts, and recent developments.
725 *EMBO J* **37**(2018).
726 48. Rose-John, S. The Soluble Interleukin 6 Receptor: Advanced Therapeutic Options in
727 Inflammation. *Clin Pharmacol Ther* **102**, 591-598 (2017).
728 49. Zilber, E.*, et al.* Soluble plasma programmed death 1 (PD-1) and Tim-3 in primary HIV
729 infection. *AIDS* **33**, 1253-1256 (2019).

730 50. Hoel, H.*, et al.* Soluble T-Cell Immunoglobulin Mucin Domain-3 Is Associated With
731 Hepatitis C Virus Coinfection and Low-Grade Inflammation During Chronic Human
732 Immunodeficiency Virus Infection. *Open Forum Infect Dis* **7**, ofaa033 (2020).
733 51. Mohammadizad, H., Shahbazi, M., Hasanjani Roushan, M.R., Soltanzadeh-Yamchi, M.
734 & Mohammadnia-Afrouzi, M. TIM-3 as a marker of exhaustion in CD8(+) T cells of
735 active chronic hepatitis B patients. *Microb Pathog* **128**, 323-328 (2019).
736 52. Ueland, T.*, et al.* Elevated plasma sTIM-3 levels in patients with severe COVID-19. *J*
737 *Allergy Clin Immunol* **147**, 92-98 (2021).
738 53. Hoffmann, M.*, et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is
739 Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278 (2020).
740 54. Tanaka, T., Narazaki, M. & Kishimoto, T. IL-6 in inflammation, immunity, and disease.
741 *Cold Spring Harb Perspect Biol* **6**, a016295 (2014).
742 55. Lagunas-Rangel, F.A. & Chavez-Valencia, V. High IL-6/IFN-gamma ratio could be
743 associated with severe disease in COVID-19 patients. *J Med Virol* **92**, 1789-1790
744 (2020).
745 56. Deshmane, S.L., Kremlev, S., Amini, S. & Sawaya, B.E. Monocyte chemoattractant
746 protein-1 (MCP-1): an overview. *J Interferon Cytokine Res* **29**, 313-326 (2009).
747 57. Gschwandtner, M., Derler, R. & Midwood, K.S. More Than Just Attractive: How CCL2
748 Influences Myeloid Cell Behavior Beyond Chemotaxis. *Front Immunol* **10**, 2759 (2019).
749 58. Varona, J.F.*, et al.* Soluble ICAM 1 and VCAM 1 Blood Levels Alert on Subclinical
750 Atherosclerosis in Non Smokers with Asymptomatic Metabolic Syndrome. *Arch Med*
751 *Res* **50**, 20-28 (2019).
752 59. Zhang, F.*, et al.* IFN-gamma and TNF-alpha drive a CXCL10+ CCL2+ macrophage
753 phenotype expanded in severe COVID-19 lungs and inflammatory diseases with tissue
754 inflammation. *Genome Med* **13**, 64 (2021).
755 60. Wang, Z.*, et al.* COVID-19 Severity Correlates with Weaker T-Cell Immunity,
756 Hypercytokinemia, and Lung Epithelium Injury. *Am J Respir Crit Care Med* **202**, 606-
757 610 (2020).
758 61. Mathew, D.*, et al.* Deep immune profiling of COVID-19 patients reveals patient
759 heterogeneity and distinct immunotypes with implications for therapeutic
760 interventions. *bioRxiv* (2020).
761 62. Koutsakos, M.*, et al.* Integrated immune dynamics define correlates of COVID-19
762 severity and antibody responses. *Cell Rep Med* **2**, 100208 (2021).
763 63. Keskinen, P., Ronni, T., Matikainen, S., Lehtonen, A. & Julkunen, I. Regulation of HLA
764 class I and II expression by interferons and influenza A virus in human peripheral blood
765 mononuclear cells. *Immunology* **91**, 421-429 (1997).
766 64. Gregorova, M.*, et al.* Post-acute COVID-19 associated with evidence of bystander T-
767 cell activation and a recurring antibiotic-resistant bacterial pneumonia. *Elife* **9**(2020).
768 65. Habel, J.R.*, et al.* Suboptimal SARS-CoV-2-specific CD8(+) T cell response associated
769 with the prominent HLA-A*02:01 phenotype. *Proc Natl Acad Sci U S A* **117**, 24384-
770 24391 (2020).
771 66. Tomoda, A.*, et al.* Cytokine production and modulation: comparison of patients with
772 chronic fatigue syndrome and normal controls. *Psychiatry Res* **134**, 101-104 (2005).
773 67. Montoya, J.G.*, et al.* Cytokine signature associated with disease severity in chronic
774 fatigue syndrome patients. *Proc Natl Acad Sci U S A* **114**, E7150-E7158 (2017).
775

776

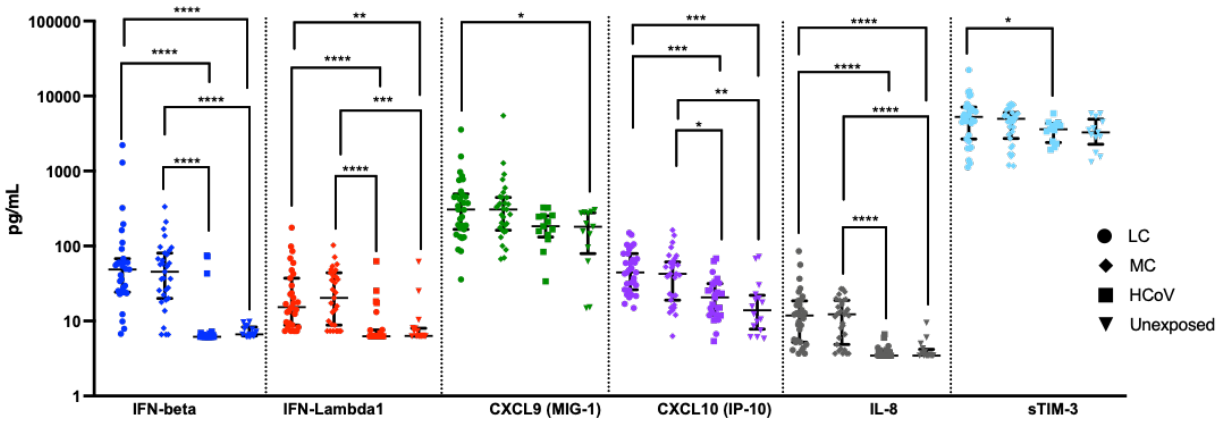
Table 1

Characteristics	ADAPT participants		HCoV (ADAPT-C)	Unexposed (St Vincent's)	Unexposed (Uni Melbourne)
	'LONG COVID	Matched ADAPT controls			
n	31	31	25	16	30
Mean age (y)	49.6 ± 14.9	48.9 ± 12.8	47.4 ±16.9	37.13 ±10.02	48.13±11.89
Male gender, n (%)	15 (48%)	15 (48 %)	14 (54%)	8 (50%)	15(50%)
Median days after positive SARS-CoV-2 PCR (IQR)	128 (115-142)	120 (115-142)	N/A	N/A	N/A
Ethnicity (n, %-total) - Caucasian	28 (90%)	26 (84%)		12 (75%)	N/A
Severity (n, %-total) - Hospitalised - Community	8 (26%) 23 (74%%)	2/31 (6%) 29/31 (94%)	N/A	N/A	N/A
Co-morbidities (n, %-total) - Yes	12 (39%)	12 (39%)	N/A	N/A	N/A
Transmission - Overseas-acquired	13 (42%)	15/31(48%)	N/A	N/A	N/A
Phenotyping (n)	14	14	N/A	14	-
HCoV positive (n) HKU-1 229E NL63 OC43	-	-	 1 14 14 10	-	-
ACE2 assay (n) 3-month 5-month 8-month	 26 25 27	 29 24 29	26 - - -	-	30 - - -

Table 1- *Patient characteristics.* Age, gender , ethnicity and co-morbidities within the cohorts sampled. HCoV patients who were PCR positive for 229E were also positive for NL63.

Figure 1

A



B

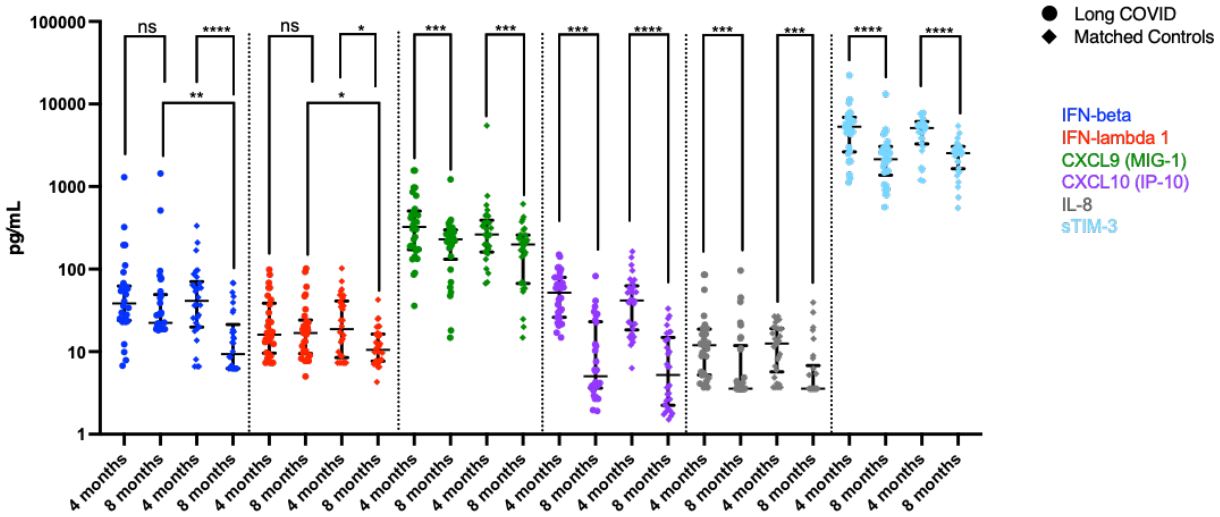


Figure 1- *Elevated levels of pro-inflammatory cytokines that persists after month 8 of convalescence.* A) Higher levels of IFN-β, IFN-λ1, CXCL9 (MIG-1), CXCL10 (IP-10), IL-8 and sTIM-3 at 4 months in long COVIDs and asymptomatic matched controls compared to subjects infected with common cold coronavirus (HCoV) and unexposed healthy donors. B) Reduction of cytokine levels at 8 months in MC. IFN-β and IFN-λ1 level in LC were higher than MC at 8 months. Data shown as medians with interquartile ranges. Two-tailed *p* values <0.05 (*) were considered significant. Wilcoxon t-test was used for paired samples at 8 months.

785

786

Figure 2

A

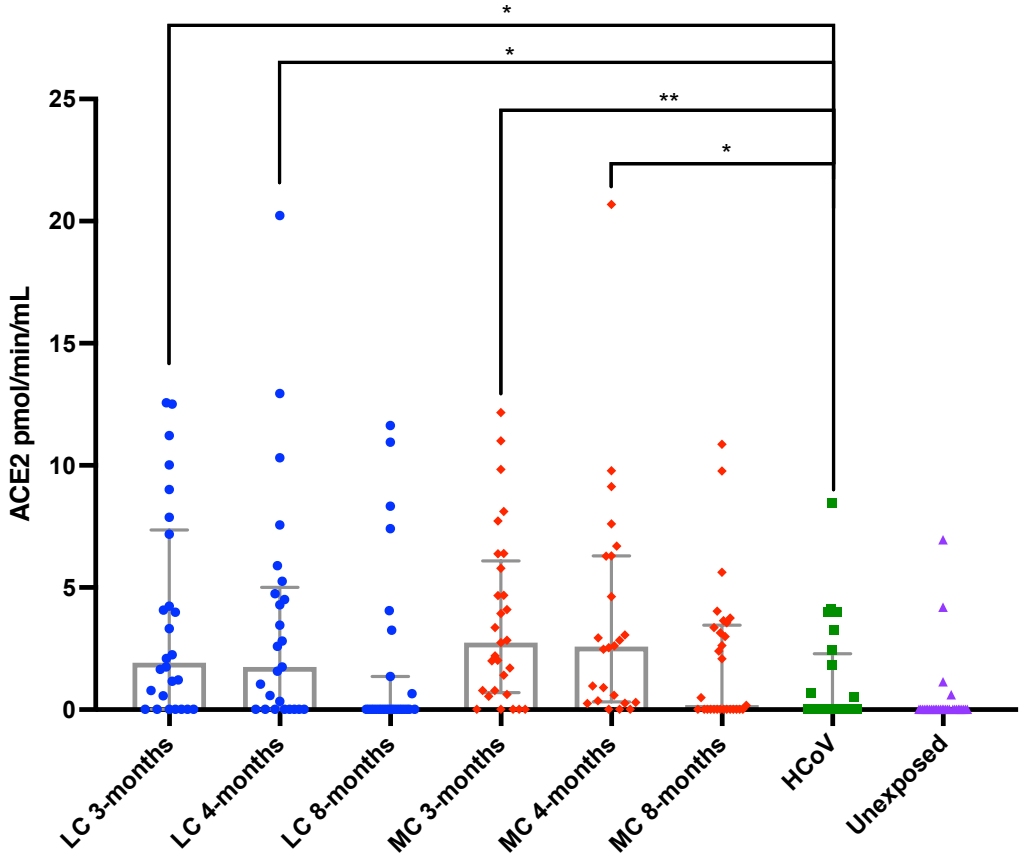


Figure 2- Elevated plasma ACE2 activity in SARS-CoV-2 infected subjects that decreases over time. A) Increased ACE2 activity at 3 months and 4 months post SARS-CoV-2 infection, that decreased at 8 months in both groups. Minimal ACE2 in HCoV and unexposed donors from University of Melbourne (Unexposed). LC (Long COVID), MC (asymptomatic matched controls). Data shown as column graphs showing medians with interquartile ranges. Two-tailed p values <0.05 (*) were considered significant.

787

788

789

790

791

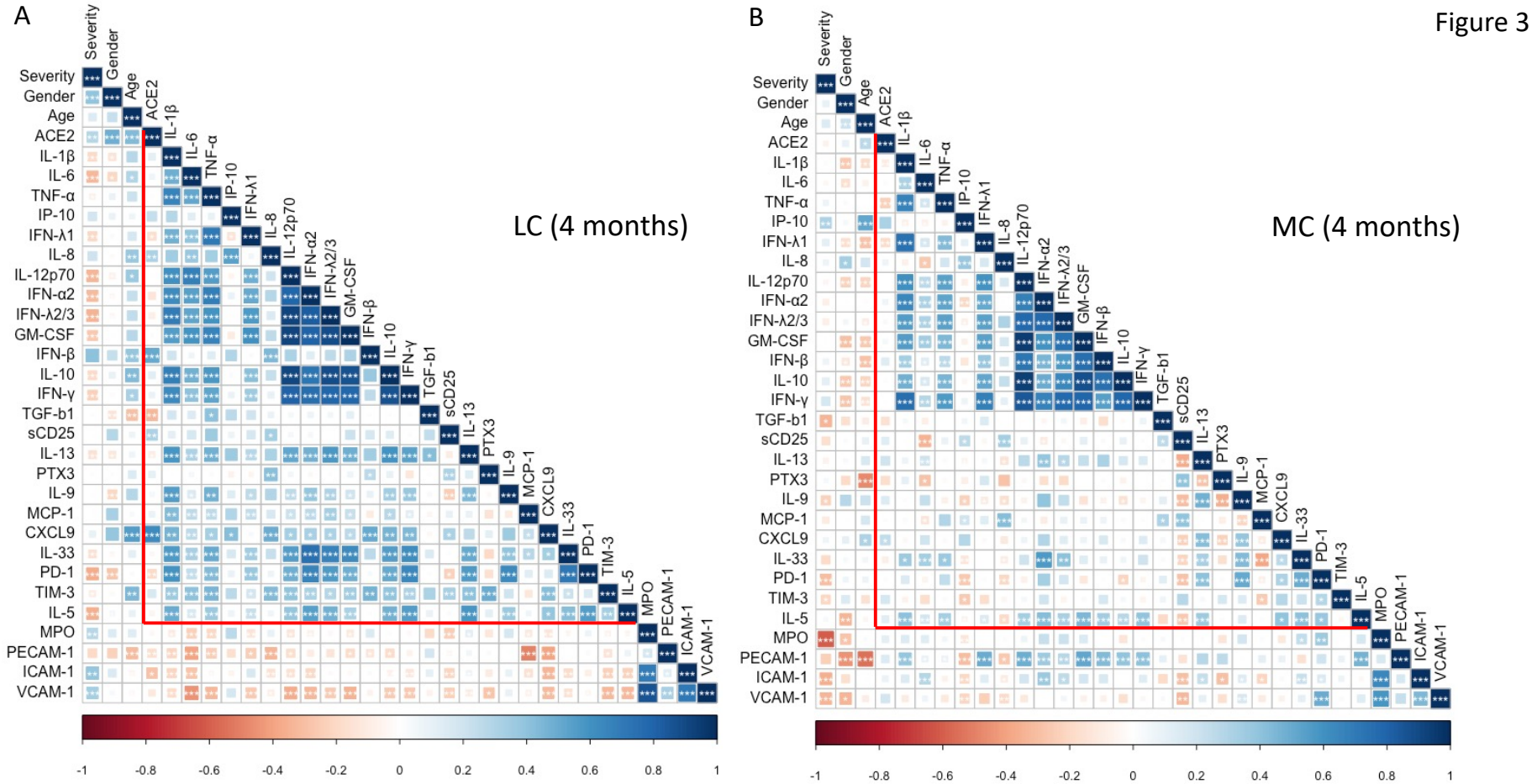


Figure 3

Figure 3- *high inter-analyte correlations and the 4 best features in long COVID.* A) Spearman correlation and hierarchical clustering comparing Severity, Gender, Ethnicity and Age with ACE2 and 28 other analytes at 4 months. Diffuse elevation of inflammatory analytes in Long COVIDs (highlighted in red triangle). B) No diffuse inflammatory profile in matched controls at 4 months. *p* values <0.05 were considered significant (*<0.05, **<0.01, ***<0.001). C) Best 4 analytes highly associated with LC eere: IL-6, IFN-γ, MCP-1 (CCL2) and VCAM-1. These four analytes gave an accuracy of 75.9%, and an F1 score of 0.759. TP= true positive, TN= true negative, FP= false positive and FN= false negative D) Visualisation of the domain boundary with two-dimensional projections. Red dots are long COVIDs, Blue dots are matched controls. Axes represent concentrations pg/mL.

792

Figure 3

C

Best 4 analytes	Accuracy	F1 Score	TP	TN	FP	FN
IL-6, IFN-γ , MCP-1 , VCAM-1	0.759	0.759	22	22	7	7

D

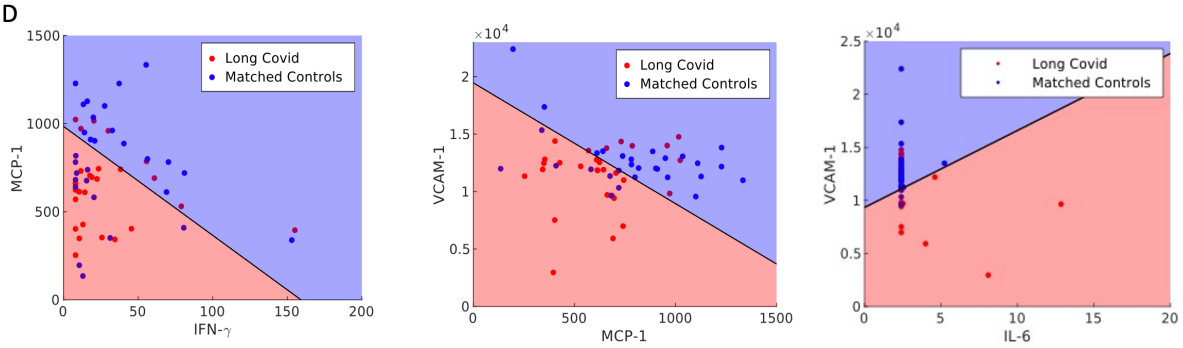
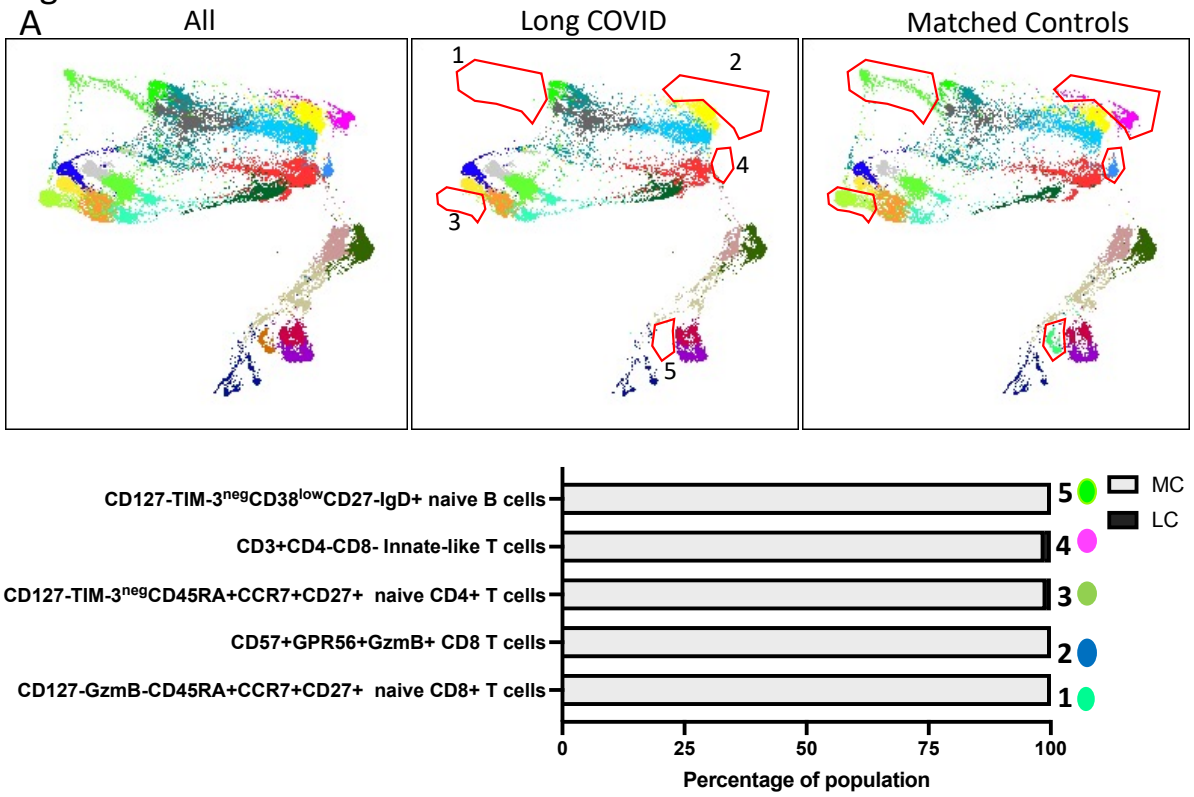


Figure 4


A
All
Long COVID
Matched Controls
1
2
3
4
5
CD127-TIM-3negCD38lowCD27-IgD+ naive B cells
CD3+CD4-CD8- Innate-like T cells
CD127-TIM-3negCD45RA+CCR7+CD27+ naive CD4+ T cells
CD57+GPR56+GzmB+ CD8 T cells
CD127-GzmB-CD45RA+CCR7+CD27+ naive CD8+ T cells
MC
LC
0
25
50
75
100
Percentage of population

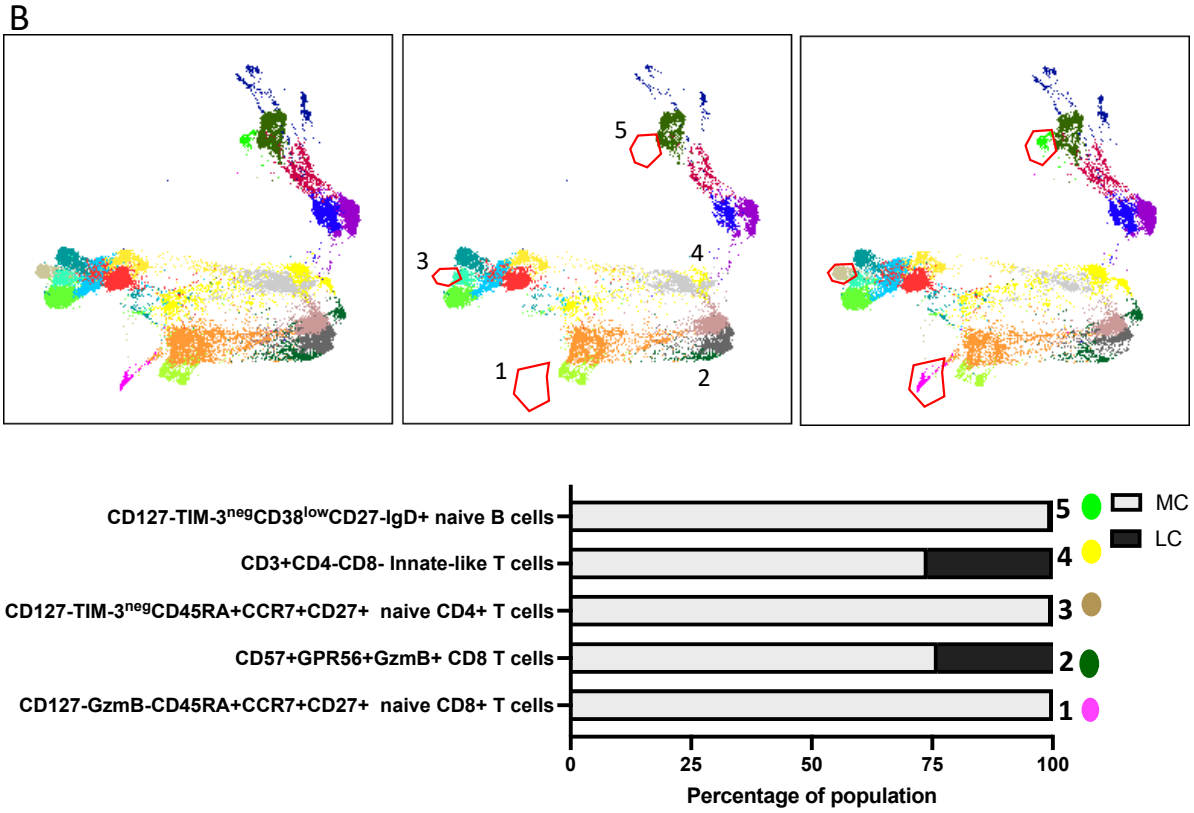

B
1
2
3
4
5
CD127-TIM-3negCD38lowCD27-IgD+ naive B cells
CD3+CD4-CD8- Innate-like T cells
CD127-TIM-3negCD45RA+CCR7+CD27+ naive CD4+ T cells
CD57+GPR56+GzmB+ CD8 T cells
CD127-GzmB-CD45RA+CCR7+CD27+ naive CD8+ T cells
MC
LC
0
25
50
75
100
Percentage of population

Figure 4

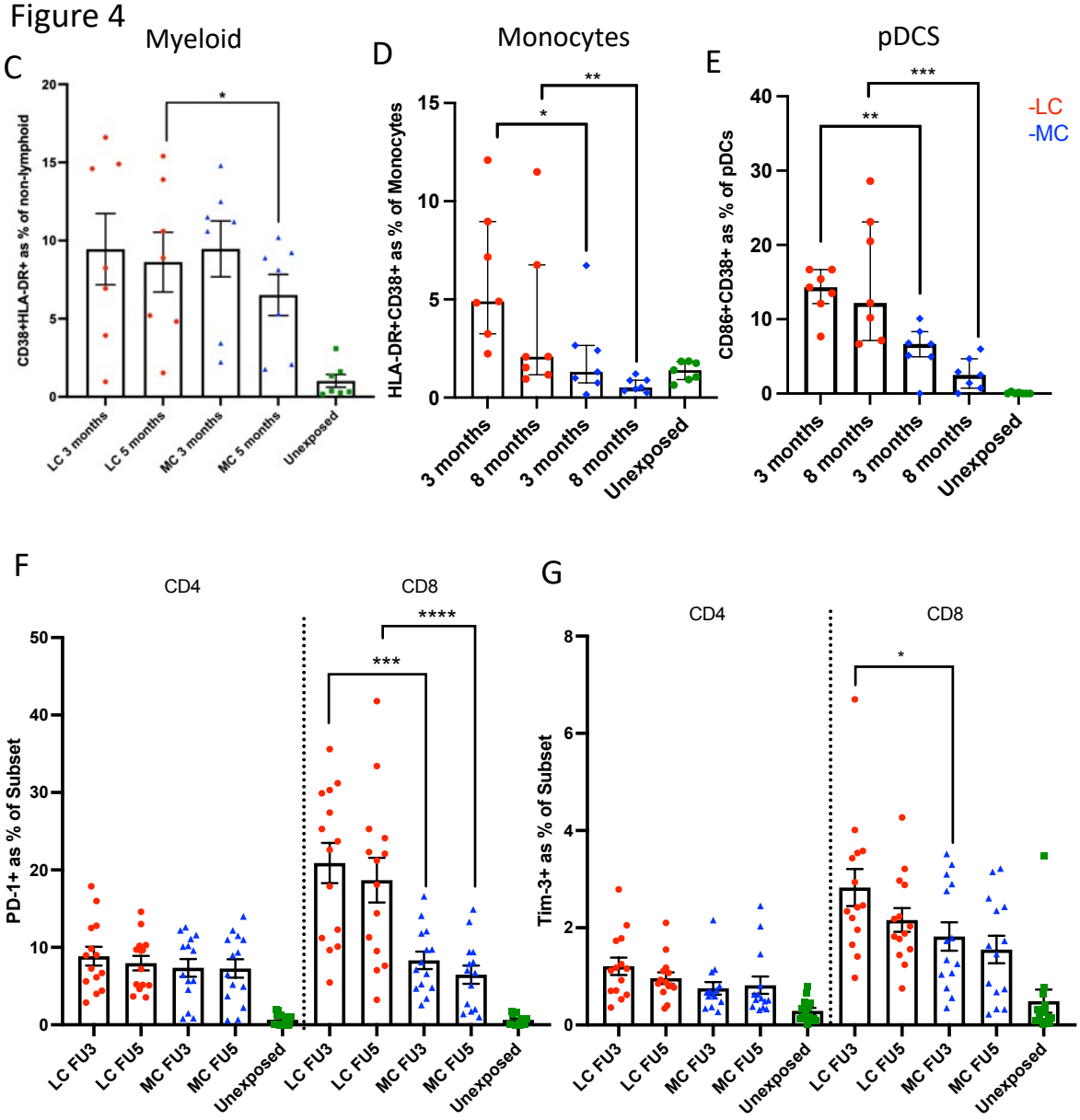


Figure 4- *Distinct activation phenotype in non-lymphoid cells and absence of naïve T and B cells found in long COVID.* A) Dimensional reduction utilising TriMap and clustering with Phenograph was used to visualise immune cell phenotypes. 5 populations consisting of un-activated naïve and cytotoxic phenotypes were absent in LC compared to MC at 3 months. Absent populations are outlined in red (middle panel), with frequency of population shown in bar graph. B) 3 naive populations remained absent in peripheral blood at 8 months in Long COVIDs. C) Activated non-lymphoid (myeloid) cells with combined expression of CD38 and HLA-DR in LC and MC at 3 months but was not reduce at 8 months in LCs. D) More activated monocytes and (E) plasmacytoid dendritic cells (pDCS) in LC compared to MC at 3 and 8 months. F) No difference in PD-1 levels on CD4 T cells, but higher expression by CD8+ T cells in LC at both timepoints. Data shown as medians with interquartile ranges. G) Higher Tim-3 expressed onCD8+ T cells in LC at 3months. Two-tailed *p* values <0.05 (*) were considered significant.

801

802

Supplementary Table 1

Participant	Symptom Severity at acute initial infection (Mild=1, Moderate=2, Severe=3)	Gender (0=F, 1=M)	Ethnicity	Long-COVIDs Significant Symptom Criteria (1=Long COVID)	Symptoms at 8 months	Long-COVID at 8 months (1=Long COVID)
1	1	0	Caucasian/White	1		
2	3	1	Caucasian/White	1		
3	3	1	Caucasian/White	1	Yes	1
4	1	0	Caucasian/White	1	Yes	1
5	3	1	Asian	0	No	0
6	1	1	Caucasian/White	1	Yes	1
7	3	1	Caucasian/White	1	Yes	1
8	3	1	Caucasian/White	1	Yes	1
9	2	1	Caucasian/White	0	No	0
10	2	1	Caucasian/White	0	No	0
11	1	1	Caucasian/White	0	No	0
12	3	1	Caucasian/White	1	Yes	1
13	1	0	Caucasian/White	0	No	0
14	2	0	Caucasian/White	1	Yes	1
15	1	1	Caucasian/White	0	No	0
16	2	0	Caucasian/White	0	No	0
17	2	1	Caucasian/White	0	No	0
18	2	0	Hispanic or Latino	1	Yes	1
19	1	1	Caucasian/White	0	No	0
20	2	0	Caucasian/White	0		
21	2	1	Caucasian/White	0	No	0
22	2	0	Caucasian/White	1	Yes	1
23	2	0	Asian	0	No	0
24	2	1	Caucasian/White	0	No	0
25	2	0	Caucasian/White	1	No	0
26	2	1	Caucasian/White	0	No	0
27	2	1	Caucasian/White	1	No	0
28	1	0	Other	0	No	0
29	2	0	Caucasian/White	0	No	0
30	2	0	Caucasian/White	0	No	0
31	2	1	Caucasian/White	1	No	0
32	1	1	Caucasian/White	0	No	0
33	1	0	Caucasian/White	0	No	0
34	1	0	Caucasian/White	0	No	0
35	2	0	Caucasian/White	1	Yes	1
36	2	0	Caucasian/White	0	No	0
37	2	0	Hispanic or Latino	0	No	0
38	1	0	Caucasian/White	0	No	0
39	1	0	Caucasian/White	0	No	0
40	1	1	Caucasian/White	1	No	0
41	3	1	Caucasian/White	0		
42	1	1	Caucasian/White	0	No	0
43	1	0	Caucasian/White	0	No	0
44	2	0	Caucasian/White	1	Yes	1
45	3	0	Caucasian/White	1	Yes	1
46	1	0	Caucasian/White	1	Yes	0
47	2	1	Caucasian/White	1	Yes	1
48	1	1	Caucasian/White	0	No	0
49	2	0	Caucasian/White	1	Yes	1
50	2	0	Caucasian/White	0	No	0
51	1	0	Caucasian/White	1	Yes	1
52	2	0	Caucasian/White	1	Yes	1
53	2	0	Caucasian/White	1	Yes	1
54	3	1	Caucasian/White	1	Yes	1
55	2	1	Caucasian/White	1	Yes	1
56	2	0	Asian	1	Yes	1
57	2	1	Caucasian/White	1	Yes	1
58	1	0	Asian	0	No	0
59	2	0	Caucasian/White	1	Yes	1
60	2	1	Caucasian/White	1	Yes	1
61	3	1	Hispanic or Latino	1	Yes	1
62	1	1	Caucasian/White	0	No	0

Supplementary Table 1- *Clinical data of Long COVID and Matched controls.* A) Symptom severity, Gender, Ethnicity of 31 long COVID and 31 asymptomatic matched controls. Subject 15 was on Everilomus and Mycophenolate due to prior renal transplant, and subject 45 was on Plaquenil for treatment of SLE.

803

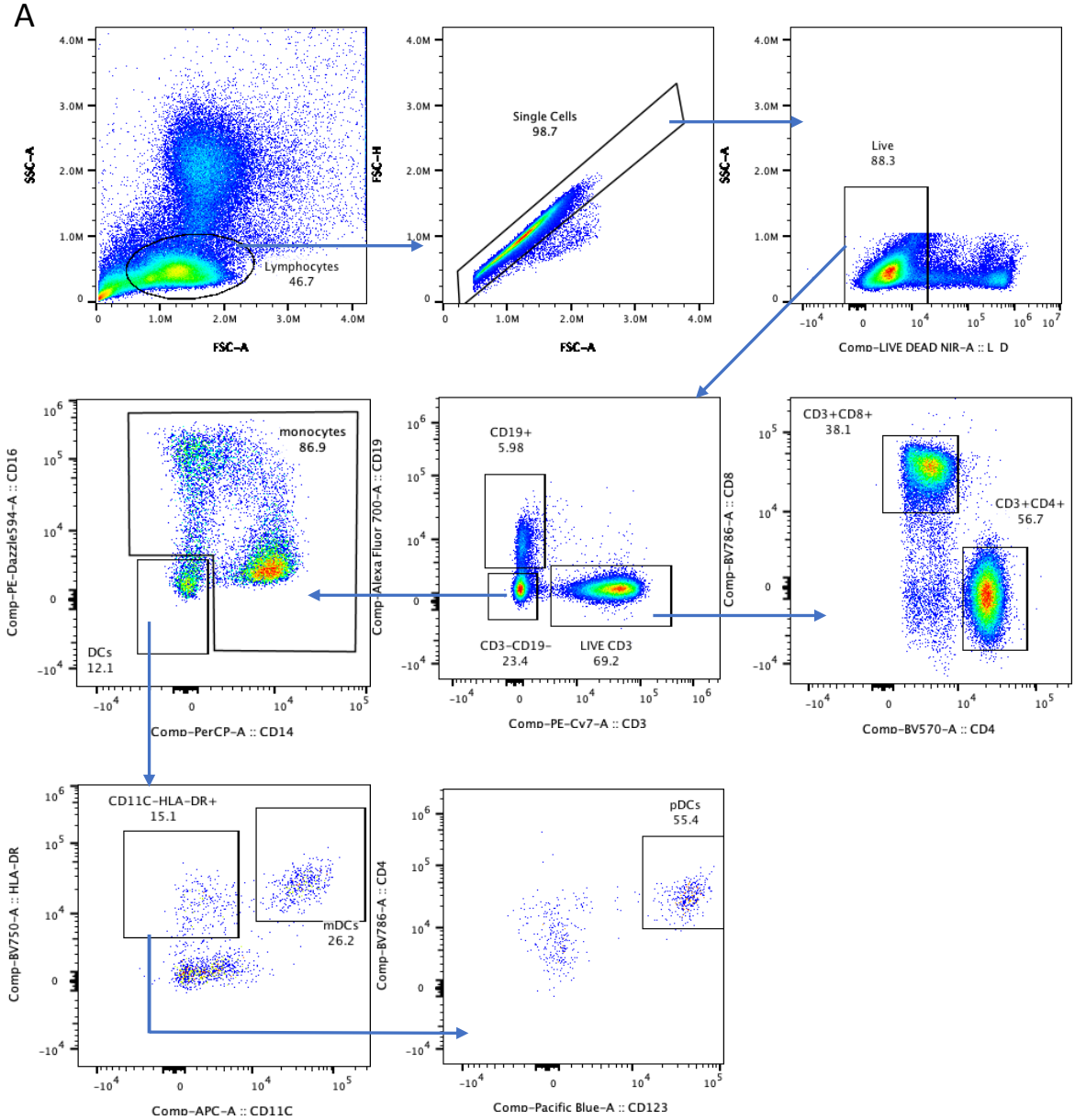
Supplementary Table 2

	Median date of symptom onset	Fever / chills	Cough	Rhinorrhoea / runny nose	Anosmia / loss of smell	Ageusia / loss of taste	Sore throat	Shortness of breath	Chest pain	Nausea / Vomiting	Diarrhoea	Headache	Myalgia / muscle aches	Arthralgia / joint aches	Malaise / weakness	Fatigue	Other
Long COVID (n=31)	19/03/2020	17	19	13	13	12	14	15	21	10	9	10	24	22	14	19	25
Matched Controls (n=31)	18/03/2020	15	15	14	10	9	12	8	4	3	8	18	17	14	16	26	7

Supplementary Table 2- *Symptoms at acute initial infection.* A) Median date of symptom onset for two cohort groups (long COVID and matached control). Numbers represent subjects with reported symptom at acute initial infection.

Supplementary Figure 1

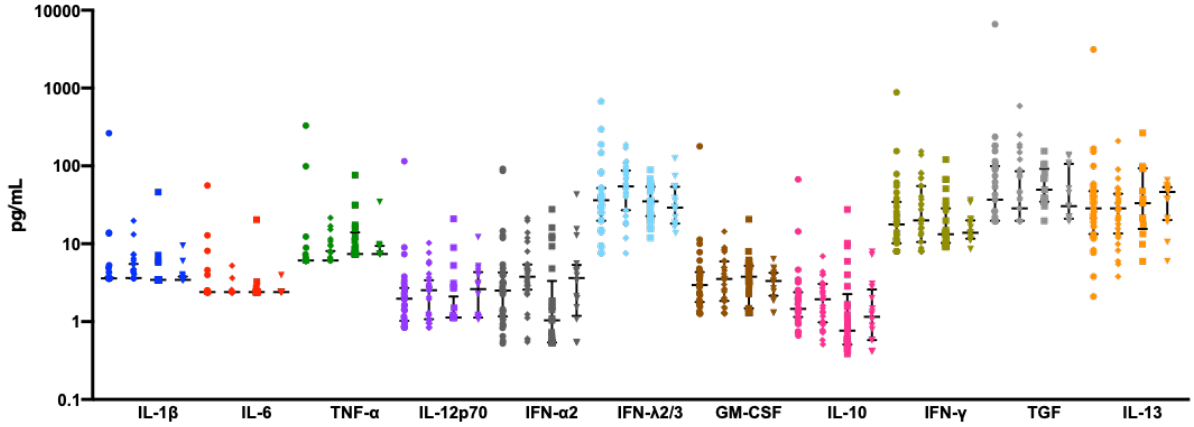
A



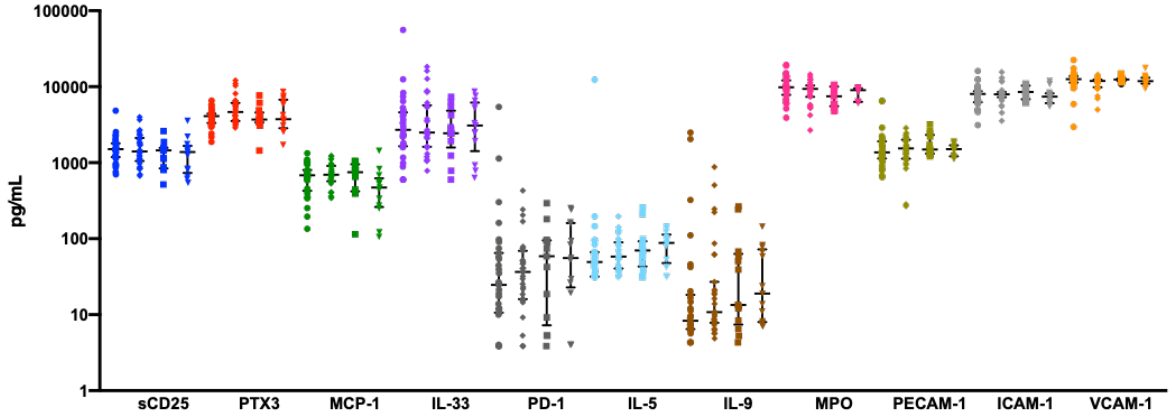
Supplementary Figure 1- *Gating strategy.* A) Representative dot plot and gating strategy for flow cytometry analysis. Cells use for downstream analysis were live singlet expressing lineage markers, mDC (myeloid dendritic cells), pDC (plasmacytoid dendritic cells).

Supplementary Figure 2

A



B



C

Groups	IFN-β	IFN-λ1	IL-8	CXCL9	CXCL10	sTIM-3
LC Median [IQR]	48.72 [24.2-67.77]	15.31 [8.83-37.21]	13.9 [7.78-21.97]	306.33 [167.19-494.19]	44.5 [26.21-79.29]	Log_{10} 3.72 [3.42-3.85])
MC	45.47 [20.04-80.39]	20.29 [8.83-20.29]	11.82 [5.23-18.55]	307.25 [162.25-443.33]	42.47 [18.96-61.25]	Log_{10} 3.69 [3.42-3.77]
HCoV	6.15 [6.15-6.72]	6.26 [6.26-7.59]	12.28 [4.86-18.83]	183.56 [131.82-253.55]	20.69 [12.03-61.25]	Log_{10} 3.55 [3.38-3.63]
Unexposed	6.65 [6.15-8.30]	6.32 [6.26-8.02]	3.45 [3.45-4.19]	180.9 [79.06-277.12]	13.9 [7.78-21.97]	Log_{10} 3.515 [3.35-3.69]

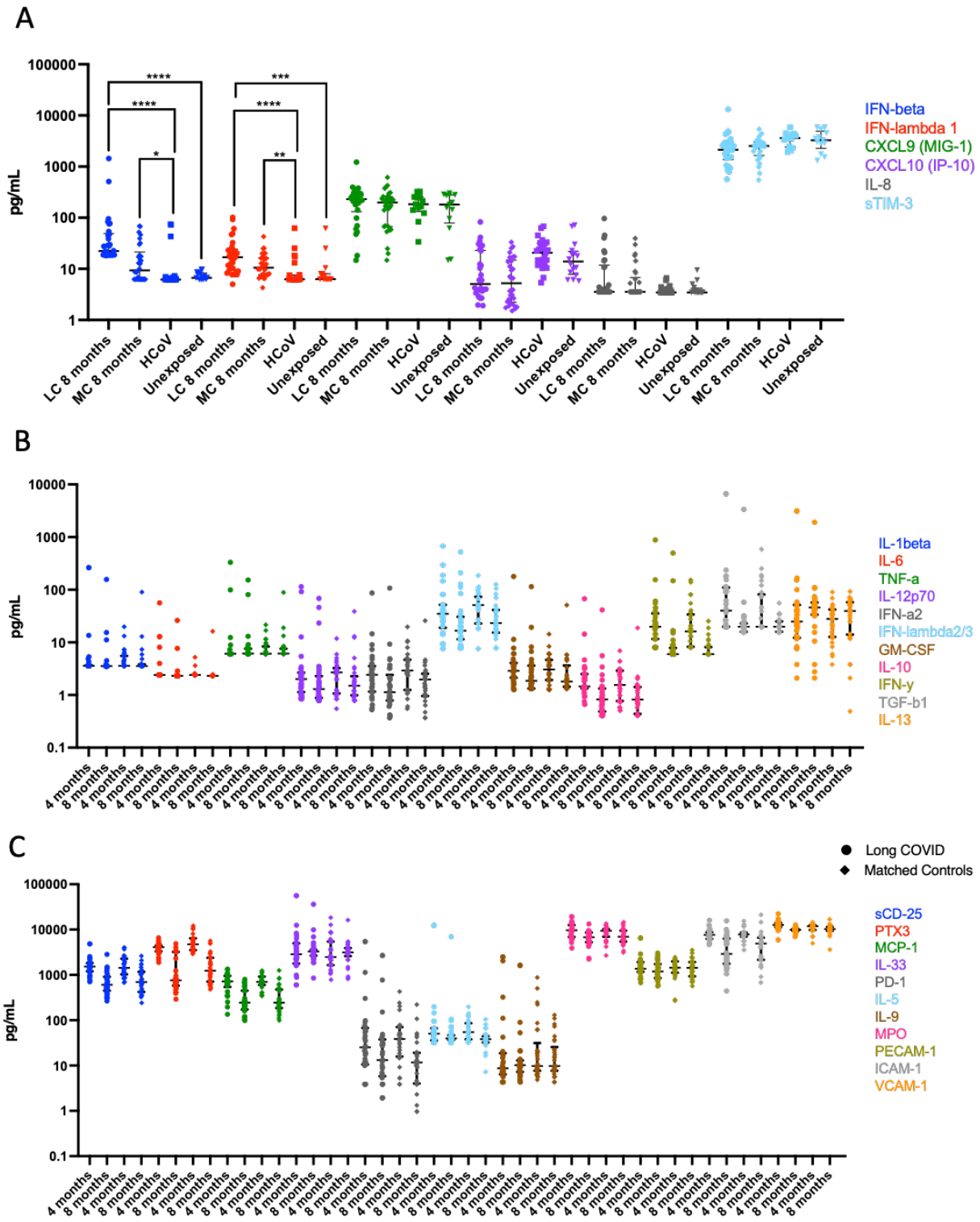
Supplementary Figure 2- *No difference in levels of 22 analytes following SARS-CoV-2 infection.* A) No differences in IL-1b, IL-6, TNF-α, IL-12p70, IFN-α2, IFN-λ2/3, GM-CSF, IL-10, IFN-γ, TGF-β1 or IL-13 levels between the 4 groups at 4 months. B) No differences in sCD25, PTX3, CCL2(MCP-1), IL-33, PD-1, IL-5, IL-9, MPO, PECAM-1, ICAM-1 or VCAM-1 levels between the 4 groups. Data shown as medians with interquartile ranges. C) Median and IQR values for 4 groups at 4 months.

813

814

815

Supplementary Figure 3



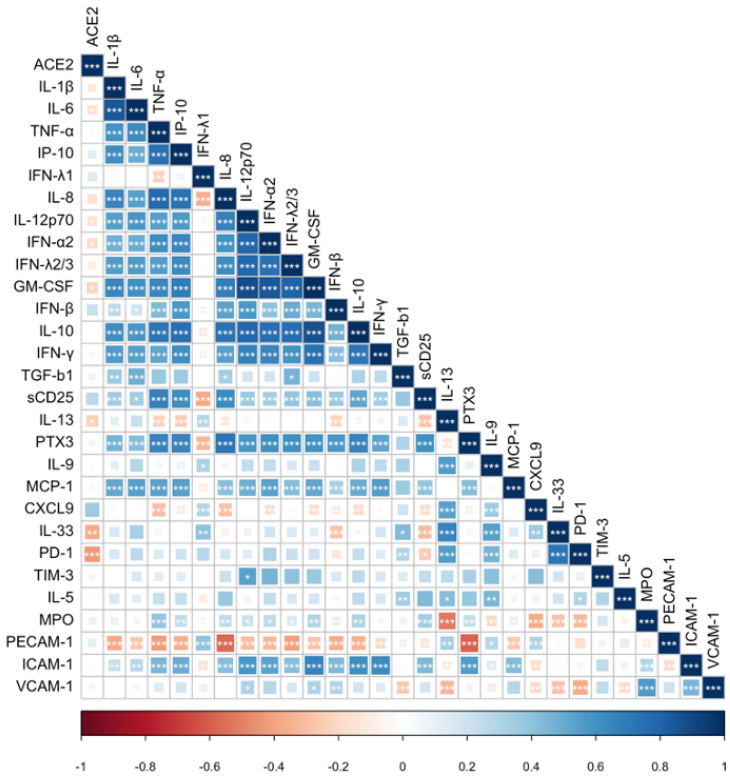
Supplementary Figure 3- *Reduction of some cytokines at 8 months.* . A) IFN-β and IFN-λ1 levels remained higher in LC at 8 months compared to HCoV and unexposed. B) Significant reduction of TGF-β1 at 8 months in LC. C) Decrease in most analytes at 8 months. Data shown as medians with interquartile ranges.

816

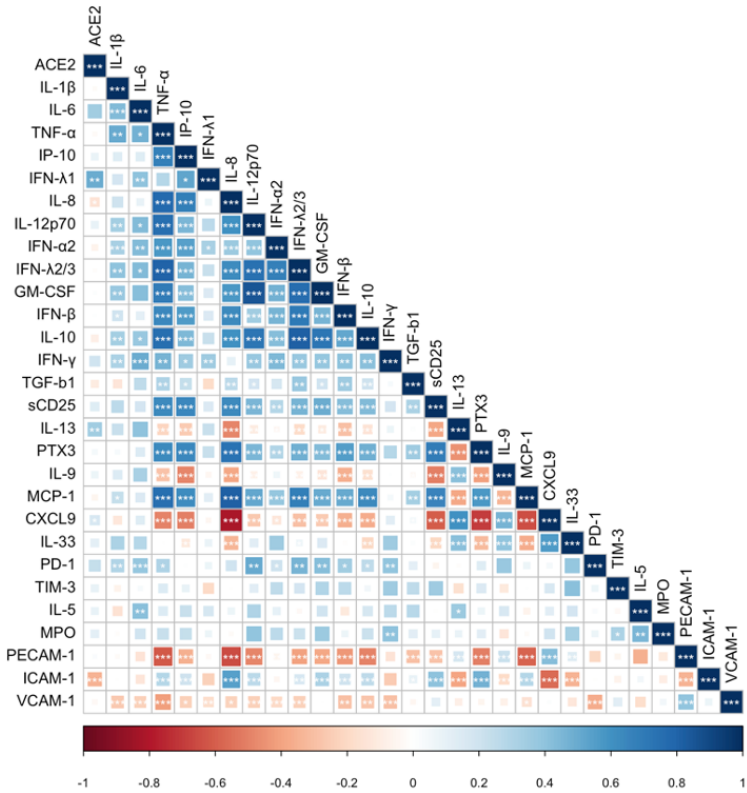
817

A

Supplementary Figure 4



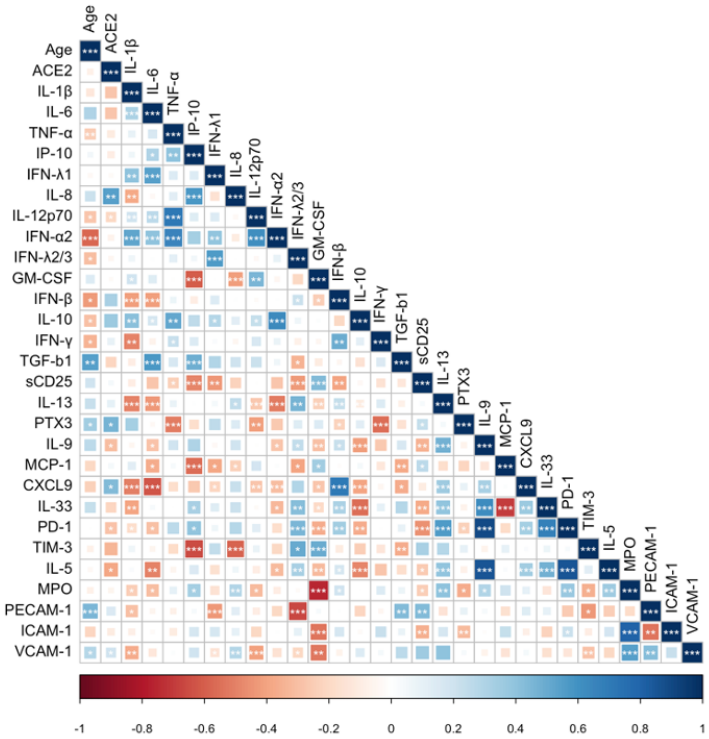
B



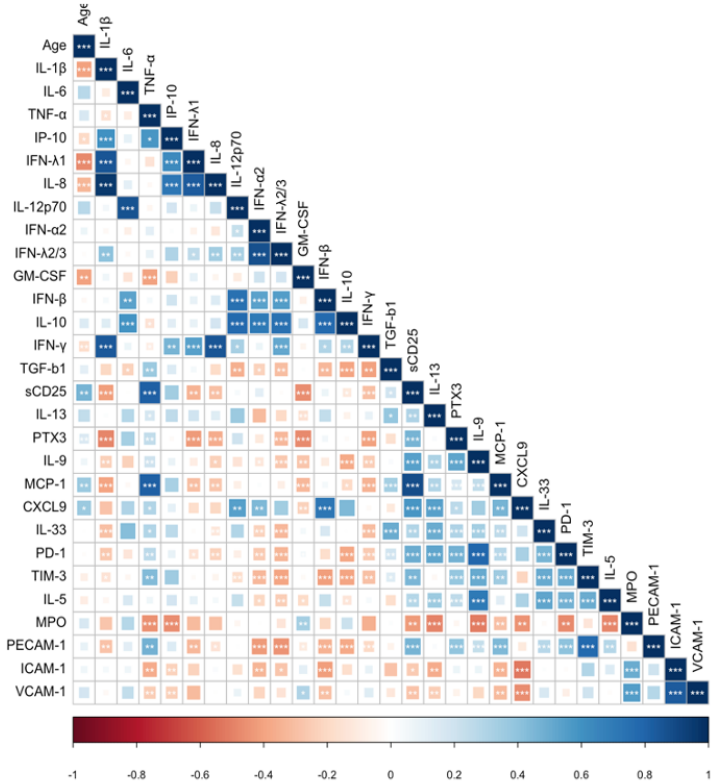
Supplementary Figure 4- *Correlation matrices from HCoV and unexposed donors.* Spearman correlation comparing indicated features in LC (A) and MC (B) at 8months and common cold coronavirus infected donors-HCoV (C) and unexposed healthy donors (D).

C

Supplementary Figure 4



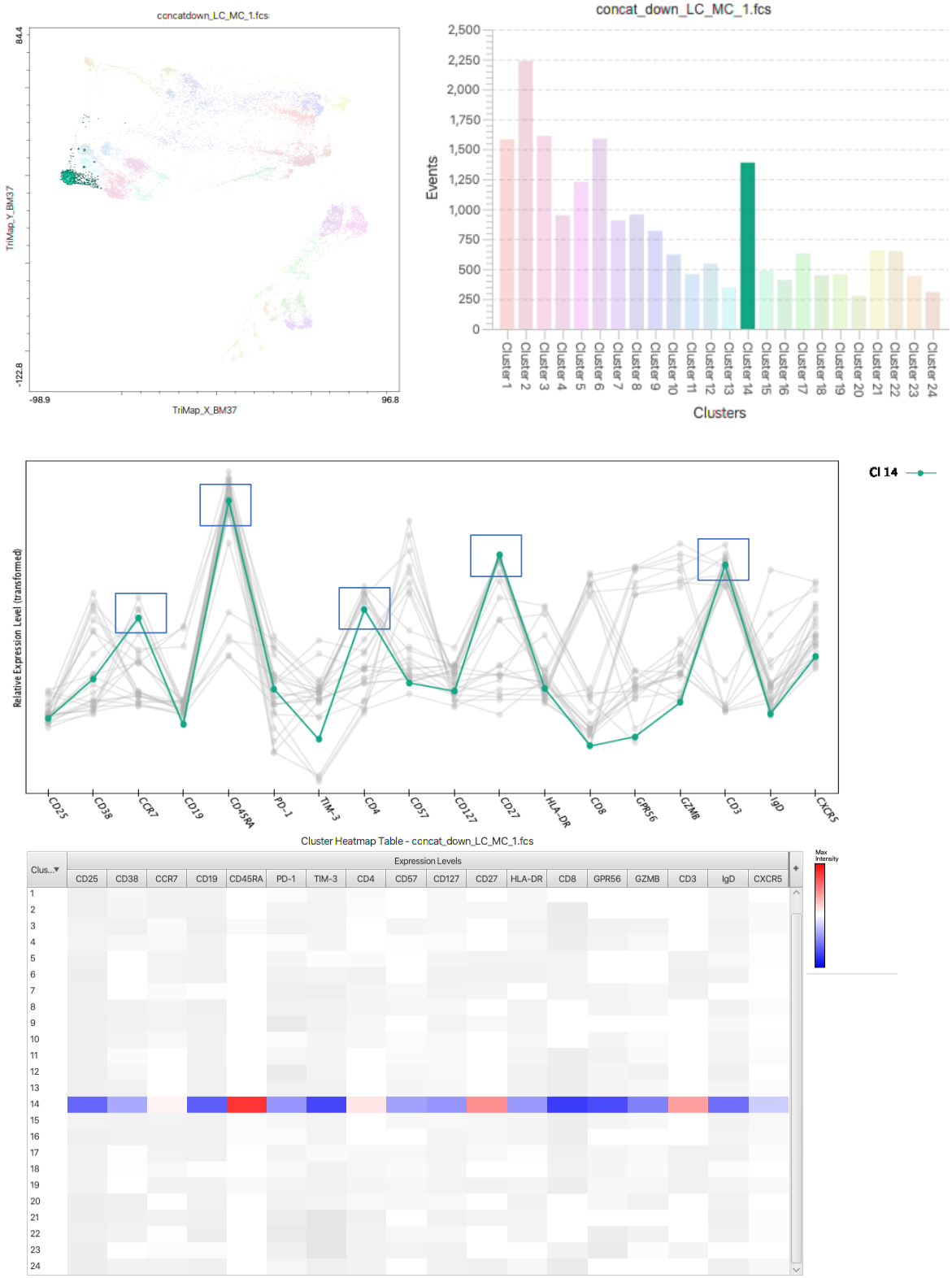
D



820

Supplementary Figure 5

A

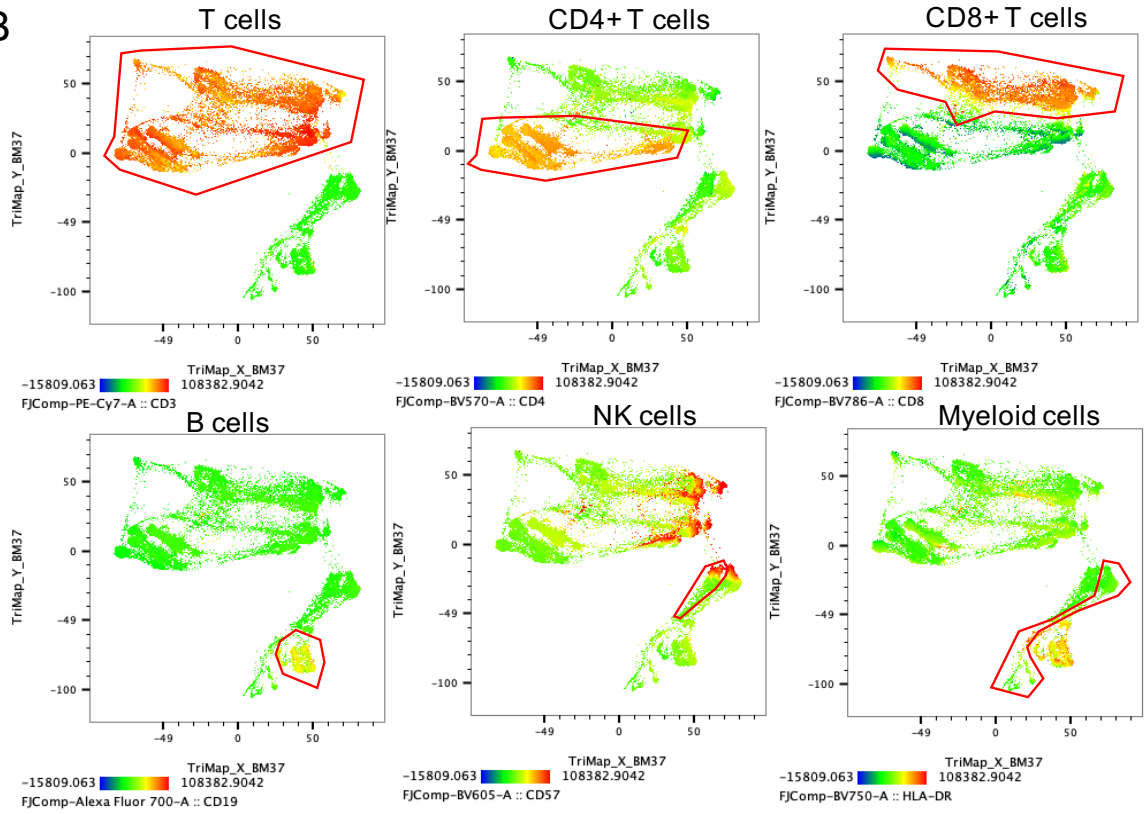


Supplementary Figure 5- *Clustering of cell phenotypes. A*) Example of cell cluster phenotype identification using CellExplorer plugin on Flowjo. Cluster 14 is equivalent to population 3, identified as CCR7+CD45RA+CD27+ naïve CD4+ T cells. TriMap of T cell, B cell, NK Cell and myeloid cell clusters (outlined in red) at 3 months (B) and 8 months (C).

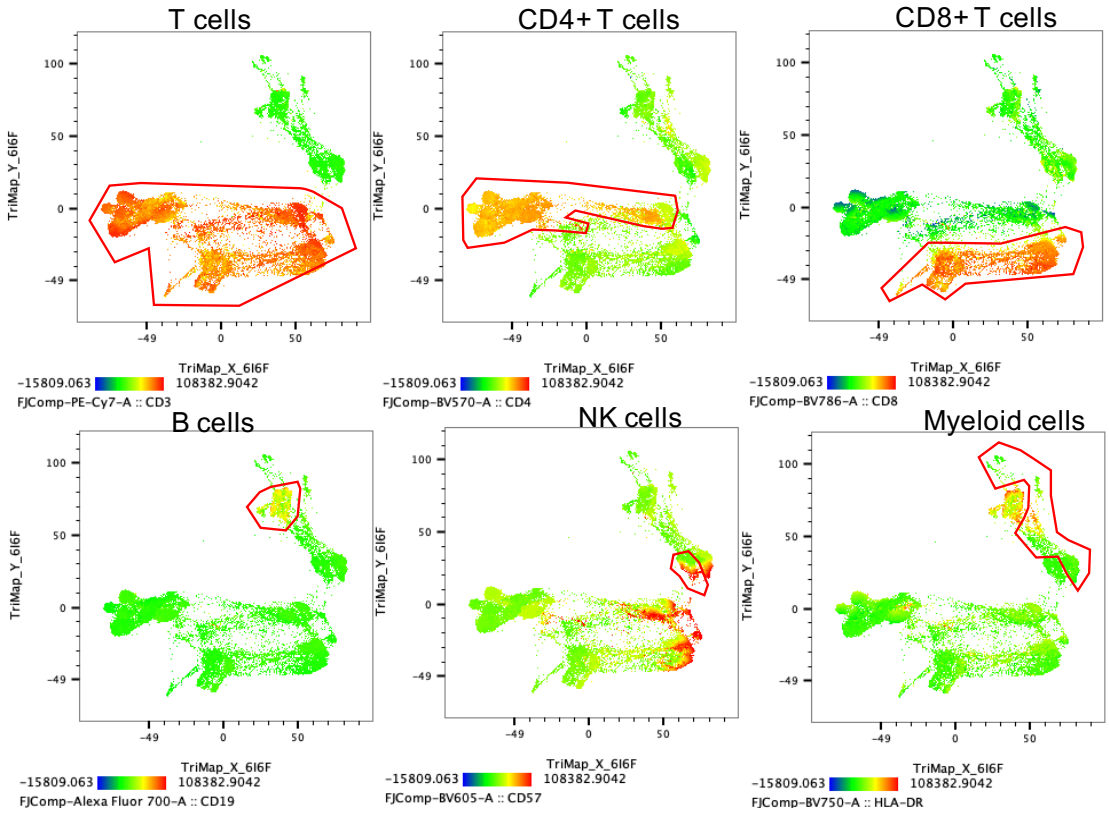
821

Supplementary Figure 5

B



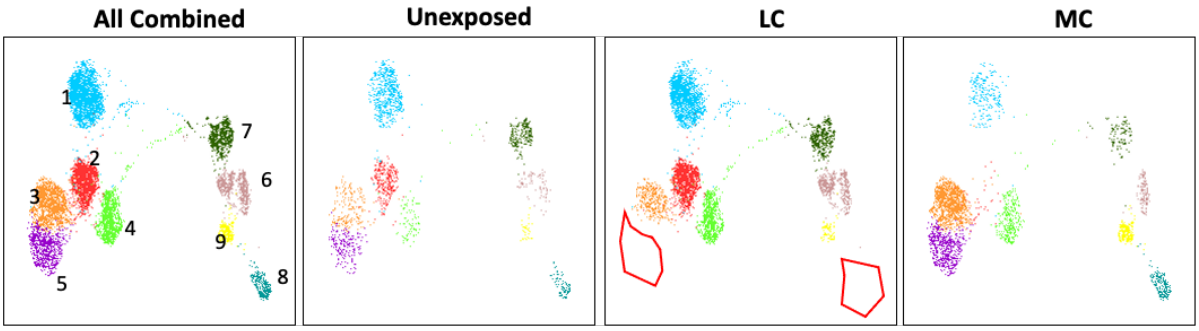
C



822

Supplementary Figure 6

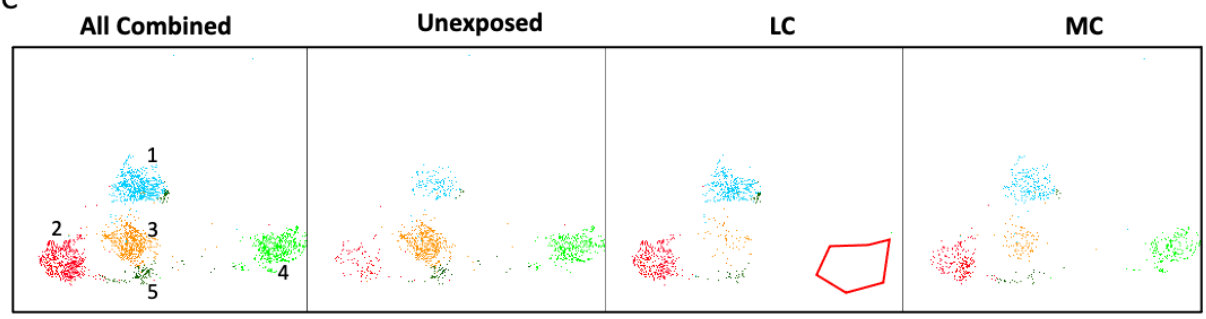
A



B

Population	Phenotype	All	Unexposed	LC	MC
1	CD4+CD127^{high}TIM-3^{mid}CD57+CD38^{low}HLA-DR^{mid}	2770	609	1937	224
2	CD4+CD127^{mid}TIM-3^{mid}CD38^{mid}HLA-DR^{low}	1973	314	1625	34
3	CD4+CD127^{low}TIM-3^{low}CD38^{low}HLA-DR^{mid}	1843	208	408	1227
4	CD4+CD127^{mid}TIM-3^{high}CD38^{mid}HLA-DR^{low}	1379	101	923	355
5	CD4+CD127^{low} TIM-3^{neg}CD38^{low}HLA-DR^{low}	818	194	0	624
6	CD8+CD127^{high}GzmB+CD57+CD38^{mid}HLA-DR^{mid}	920	260	549	111
7	CD8+CD127^{mid}GzmB+CD57-CD38^{low}HLA-DR^{low}	876	97	629	150
8	CD8+CD127^{low}GzmB-CD57-CD38^{low}HLA-DR^{low}	323	101	0	222
9	CD8+CD127^{low}GzmB-CD57-CD38^{mid}HLA-DR^{low}	361	32	89	240

C

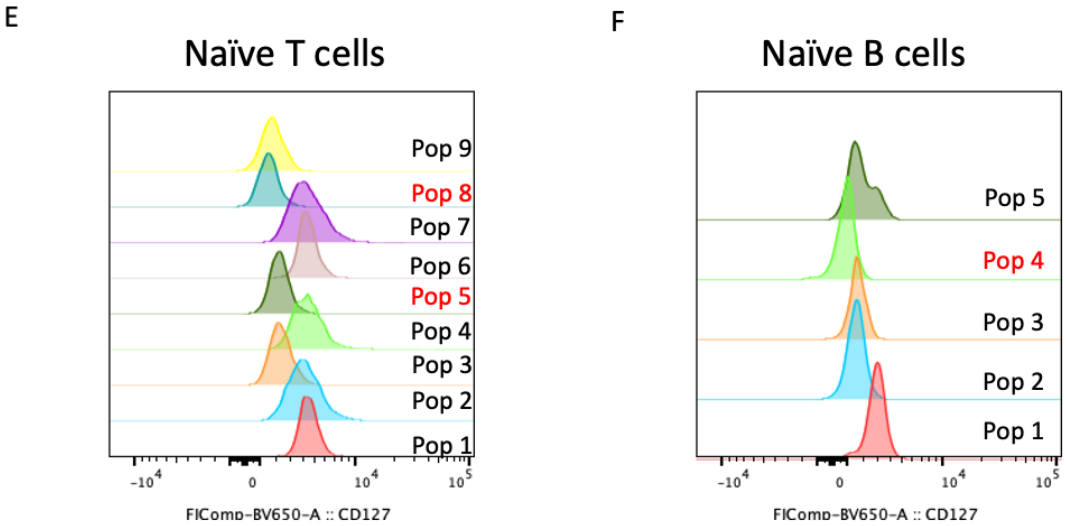


D

Population	Phenotype	All	Unexposed	LC	MC
1	CD127^{high}TIM-3^{mid}CD38^{mid}HLA-DR^{low}	725	142	362	221
2	CD127^{mid}TIM-3^{low}CD38^{mid}HLA-DR^{mid}	686	105	371	210
3	CD127^{mid}TIM-3^{low}CD38^{low}HLA-DR^{low}	676	516	59	101
4	CD127^{low}TIM-3^{neg}CD38^{low}HLA-DR^{mid}	567	358	0	209
5	CD127^{mid}TIM-3^{low}CD38^{low}HLA-DR^{mid}	153	67	53	33

823

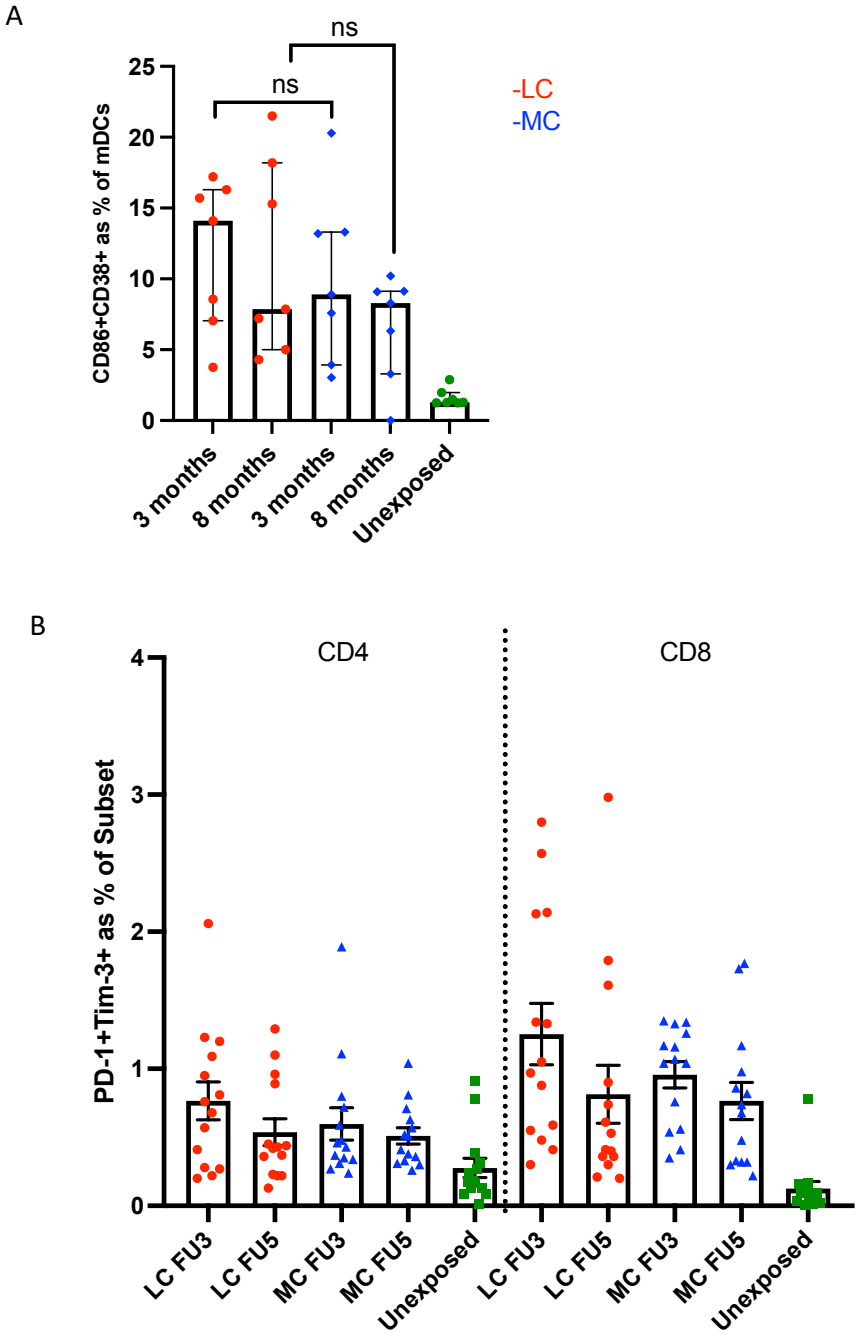
Supplementary Figure 6



Supplementary Figure 6- *Deciphering naïve phenotypes.* Representative TriMap and phenograph clustering of CCR7+CD45RA+ naïve T cells (A) and IgD+CD27- naïve B cells (C). Red gates and text represent un-activated naïve subsets absent in long COVID subjects. Phenotype and cell frequencies from naïve T cell s (B) and naïve B cells (D). CD127 expression on differing naïve T cell (E) and naïve B cell (F) populations (all data combined). Red text highlights subsets absent in long COVID. Unexposed, LC (long COVID0, and MC (asymptomatic matched controls).

824

Supplementary Figure 7



Supplementary Figure 7- *Activated phenotypes.* (A) No difference in activated state of Myeloid dendritic cells (mDCs) at 3 months and 8 months. (B) No difference between LC and MC at 3 and 8 months when co-expression of PD-1 and Tim-3 were examined (long COVID, and MC (asymptomatic matched controls).

825