

1 Divergent inflammatory and neurology-related plasma protein profiles in individuals with
2 long COVID following primary and breakthrough SARS-CoV-2 infections.

3

4 Amit Bansal^{1,2*}, Sam W.Z. Olechnowicz,^{3,4*} Nicholas Kiernan-Walker^{4,5}, Jacob
5 Cumming^{5,6,7}, Ramin Mazhari^{4,5}, COVID PROFILE consortium^{5&}, Rebecca J. Cox^{1,8}, Ivo
6 Mueller^{4,5}, Rory Bowden^{3,4,5,\$#} and Emily M. Eriksson^{4,5\$#}

7

8 * Equal Contribution

9 \$ Co-senior authors

10

11 &COVID PROFILE Consortium Investigators: Siavash Foroughi^{1,2}, Jason A. Tye-
12 Din^{1,2,3,4}, Anna K. Coussens^{1,2,5} and Vanessa L. Bryant^{1,2,6}

13

14 1. Influenza Centre, Department of Clinical Science, University of Bergen, Bergen, Norway

15 2. Department of Infectious Diseases, The University of Melbourne, at the Peter Doherty
16 Institute for Infection and Immunity, Melbourne, Australia

17 3. The Walter and Eliza Hall Institute of Medical Research, Advanced Technology and
18 Biology Division, Melbourne, Australia

19 4. The University of Melbourne, Department of Medical Biology, Melbourne, Australia

20 5. The Walter and Eliza Hall Institute of Medical Research, Population Health and
21 Immunity Division, Melbourne, Australia

22 6. Disease Elimination Program, Burnet Institute, Melbourne, Australia

23 7. The University of Melbourne, Department of Mathematics and Statistics, Melbourne,
24 Australia

25 8. Department of Microbiology, Haukeland University Hospital, Bergen, Norway

26

27 # Corresponding Authors:

28 Emily M. Eriksson

29 The Walter and Eliza Hall Institute of Medical Research

30 1G Royal Parade, Parkville, VIC 3052 Australia

31 Tel: +61 3 93452870

32 email: eriksson@wehi.edu.au

33 ORCID # 0000-0002-7851-973X

34

35 Rory Bowden

36 The Walter and Eliza Hall Institute of Medical Research

37 1G Royal Parade, Parkville, VIC 3052 Australia

38 Tel: +61 3 93452762

39 email: bowden.r@wehi.edu.au

40 ORCID # 0000-0001-8596-0366

41

42

43

44

45 **Abstract**

46 Long COVID is a complex condition where symptoms persist for more than 3 months after
47 SARS-CoV-2 infection and affects an estimated 5-30% of individuals. While the
48 pathobiology of long COVID is still evolving, persistent inflammation has emerged as an
49 important feature of this condition. However, it is unclear if immune responses from
50 COVID-19 vaccination or SARS-CoV-2 re-infection exacerbate or mirror the initial
51 inflammatory responses. To address this question, we quantified 182 inflammatory and
52 neurology-related proteins in plasma using multiplexed affinity proteomics. Plasma
53 samples were collected 6-9 months after first infection, but before COVID-19 vaccination
54 from individuals who had recovered from COVID-19 (n=21) or from individuals with long
55 COVID (n=12). This was benchmarked against plasma from unvaccinated, SARS-CoV-2
56 naive individuals (n=24). In addition to this cross-sectional analysis, we performed
57 longitudinal analysis in a subset of individuals (n=34), where paired samples collected 2-
58 4 weeks after a third COVID-19 vaccine dose and after SARS-CoV-2 breakthrough
59 infection were available. Boruta feature selection and lasso regression models identified
60 a distinct plasma profile in long COVID individuals, characterised by elevated IL-20,
61 HAGH, NAAA, CLEC10A, LXN, and MCP-1 and reduced TRAIL, G-CSF, NBL1, and
62 CCL23 protein concentrations. Notably, longitudinal analysis demonstrated that neither
63 COVID-19 booster vaccination nor breakthrough infection replicated inflammatory and
64 neurology-related plasma protein profiles observed after primary infection suggesting an
65 altered immune response outcome in individuals with long COVID upon re-exposure.

66

67 **Introduction**

68 After initial SARS-CoV-2 infection, a significant proportion of individuals develop
69 persistent systemic symptoms that may continue for months. This condition is commonly
70 referred to as long COVID or post-acute sequelae of COVID-19 ^{1,2}. One of the major
71 challenges in the field is the insufficient understanding of the underlying pathophysiology
72 of this condition, coupled with the absence of reliable biomarkers that could be used in
73 standardised diagnostic tools. Currently, the diagnosis of long COVID is predominantly
74 based on clinical evaluation. The lack of standardised molecular and quantitative
75 diagnostic criteria has impeded accurate estimation of long COVID prevalence. However,
76 available estimates suggest that long COVID affects 10–30% of non-hospitalised
77 individuals and as many as 50–70% of those who were hospitalised during acute infection
78 ^{2,3}. Although COVID-19 vaccination may reduce the risk, the incidence of developing long
79 COVID post-vaccination is still estimated to be 10–12% ^{2,4,5}.

80

81 While the precise mechanisms underlying long COVID is poorly understood, distinct
82 patterns of immune activation have been observed in affected patients and persist several
83 months after SARS-CoV-2 infection. These patterns are evident both at the cellular level
84 and through changes in soluble factors in the bloodstream. The latter includes elevated
85 or altered concentrations of circulating inflammatory mediators⁶⁻⁹, complement proteins
86 ^{10,11}, and neuro-inflammatory molecules ^{9,12} and inflammatory markers have been
87 associated with a range of persistent long COVID symptoms ¹³. Collectively, these studies
88 have provided essential insights into dysregulated processes that may underlie long
89 COVID. However, many of these long COVID studies have been conducted in

90 populations with high community transmission of SARS-CoV-2 raising concerns about
91 the potential confounding effects of ongoing viral exposure(s) particularly on inflammatory
92 biomarker measurements. Moreover, studies that have monitored inflammatory and
93 neurology plasma protein profiles longitudinally in long COVID individuals remain limited.

94 In this study, we leveraged access to samples collected from an Australian population
95 with low community transmission during the two first years of the COVID-19 pandemic ¹⁴.
96 Using a multiplexed affinity proteomics approach, the Olink Target Proximity-Extension
97 Assay (PEA, Olink) and longitudinal plasma samples from initially SARS-CoV-2 naïve
98 individuals, COVID-19 recovered and long COVID individuals, we explored inflammatory
99 and neurology-related plasma protein profiles not only following a primary infection with
100 the SARS-CoV-2 ancestral strain, but also after COVID-19 booster vaccination and
101 breakthrough infections. Our study design provided a rare opportunity to understand how
102 circulating protein profiles after vaccination and re-exposure in the same individual
103 compare to plasma profiles after primary SARS-CoV-2 infections in long COVID
104 individuals.

105 We identified 14 top inflammatory and neurology-related proteins that were distinctively
106 associated with either long COVID or recovery. These included neuroblastoma
107 suppressor of tumorigenicity 1 (NBL1), interleukin (IL)-20, latexin (LXN), monocyte
108 chemoattractant protein 1 (MCP-1, CCL2), C-type lectin domain containing 10A
109 (CLEC10A), chemokine ligand 23 (CCL23), TNF-related apoptosis-inducing ligand
110 (TRAIL), hydroxyacylglutathione hydrolase (HAGH), granulocyte-colony stimulating
111 factor (G-CSF), N-acyl ethanolamine- hydrolysing acid amidase (NAAA), artemin (ARTN),

112 cystatin-D (CST5), fibroblast growth factor 19 (FGF-19), and delta and notch-like
113 epidermal growth factor-related receptor (DNER). Notably, long COVID patients exhibited
114 plasma profiles with atypical levels of these proteins, suggestive of ongoing dysregulated
115 processes 6-9 months post-infection. Longitudinal analysis revealed partial resolution of
116 elevated plasma protein levels following COVID-19 booster vaccination, but notably,
117 inflammatory and neurology-related protein patterns observed after primary infection
118 were not replicated following a subsequent breakthrough infection. Our analyses provide
119 pivotal understanding of vaccine-associated and subsequent SARS-CoV-2 infection-
120 induced responses in long COVID individuals and suggests that there is an altered
121 immune response outcome in these individuals upon re-infection without features of
122 dysregulation.

123

124 **Methods**

125 **COVID PROFILE study cohort and samples**

126 Plasma was collected as part of COVID PROFILE, a community-based longitudinal cohort
127 study conducted in Victoria, Australia from October 2020 to April 2023¹⁴. The study was
128 approved by the Walter and Eliza Hall Institute (#20/08) and Melbourne Health
129 (RMH69108) Human Research Ethics Committees. All subjects provided written informed
130 consent to participate in the study, in accordance with the Declaration of Helsinki.
131 Participants were enrolled based on SARS-CoV-2 infection status and initially categorised
132 as COVID-19 recovered or SARS-CoV-2 naïve (confirmed by SARS-CoV-2
133 PCR/serology and referred to as healthy). Demographic data and detailed clinical
134 information on COVID-19 symptoms, comorbidities, vaccination and breakthrough
135 infections were collected at enrolment and during follow-up visits. Primary vaccination
136 consisted of either two doses of the BNT162b2 mRNA vaccine (COMIRNATY,
137 Pfizer/BioNTech), or ChAdOx1 adenoviral vectored vaccine (Oxford-AstraZeneca),
138 followed by a third mRNA vaccine consisting of either BNT162b2 or mRNA-1273
139 (Moderna Tx Spikevax). Due to emergence of the long COVID condition, additional follow-
140 up surveys were conducted to identify participants that were experiencing long COVID.
141 These surveys collected comprehensive information on symptom duration and individuals
142 with long COVID were identified in accordance with WHO definition: symptoms persisting
143 or new symptoms 3 months after the initial SARS-CoV-2 infection, with these symptoms
144 lasting for at least 2 months with no other explanation¹. All other participants experienced
145 COVID-19 symptoms that lasted <90 days post-diagnosis and were classified as COVID-
146 19 recovered. Plasma collected at enrolment for healthy (n = 24) and 6-9 months after

147 primary COVID-19 diagnosis for COVID-19 recovered (n = 21) and long COVID
148 individuals (n = 12) were used for cross-sectional analyses. In a subset of individuals
149 (n=34), additional paired samples collected at 2-4 weeks after a third dose of COVID-19
150 vaccine and post-breakthrough SARS-CoV-2 infection were used for analysis and only
151 individuals with all three timepoints were included in the longitudinal analysis. Participant
152 demographic data for both analyses are summarised in Table 1.

153

154 **Luminex Multiplex assay**

155 A SARS-CoV-2 multi-antigen serology assay was used to detect SARS-CoV-2-specific
156 IgG antibodies against Spike and nucleoprotein (NP) antigens as described previously¹⁵.
157 Briefly, fluorescent and magnetic beads were coated with antigens of interest and
158 incubated with individual plasma samples to allow binding of antigen-specific antibodies.
159 Bound antibodies were detected using a secondary anti-human IgG fragment
160 crystallisable region (Fc) antibody conjugated to phycoerythrin (PE) and analysed using
161 MAGPIX[®] system. The levels of antigen-specific antibodies were reported as mean
162 fluorescent intensity (MFI).

163

164 **Affinity proteomics assays**

165 Protein biomarkers were quantified using the Proximity Extension Assay at the WEHI
166 Genomics facility, on the Olink Target platform using Inflammation and Neurology panels
167 (Olink Proteomics, Sweden) for a total of 182 protein targets. Briefly, 1µL of each plasma
168 sample per panel was incubated with a multiplexed set of two oligo-labelled detection
169 antibodies per protein target, which after proximity extension can form a specific DNA

170 barcode¹⁶. The barcodes were pre-amplified then detected by qPCR using 96.96 Protein
171 Expression integrated fluidic circuits (IFCs) on a BioMark HD with HX controller (Fluidigm
172 / Standard BioTools, USA). Detection cycle threshold or C(t) values were then processed
173 to NPX values by normalising to the internal Extension Control and the external Inter-
174 plate Control using Olink[®] Signature software, and batch correction was applied by
175 intensity normalisation and the OlinkAnalyse R package¹⁷. The values are expressed in
176 arbitrary NPX units that are specific to Olink Proteomics platform and are on a log₂x scale.

177

178 **Statistical analyses**

179 Participant were stratified according to SARS-CoV-2 infection and COVID-19 booster
180 status. Statistical analyses were conducted in R version 4.3.2 (2023-10-31) for macOS
181 using the lme4, ggplot2, ggeffects, ggthemes, patchwork, OlinkAnalyze, performance,
182 Boruta, glmnet, glmnetUtils rpart, rattle, e1071, corrplot and corr packages.

183

184 Descriptive statistics were used to compare participant demographic, clinical
185 characteristics, and exposure factors (Kruskal–Wallis rank-sum test for continuous or
186 Fisher's exact test for categorical variables). Correlation matrices were plotted by
187 Spearman's rank correlation test with hierarchical clustering using corrplot package.

188

189 To identify potential protein biomarkers, we used a combination of penalised regression
190 (lasso) and feature selection algorithm, using glmnet and Boruta packages in R. The
191 glmnet fitted a generalised linear model via penalised maximum likelihood (lasso, least
192 absolute shrinkage and selection operator) to enhance the prediction accuracy and

193 interpretability of the resulting statistical model^{18,19}. The incorporated model fitted a
194 traditional logistic regression model for the log-odds with function `glmnet(x, y,`
195 `standardise=TRUE, alpha = 1, family = "binomial", intercept = T)`; where x variable was
196 standardised prior to fitting the model sequence and `alpha=1` for the lasso penalty. The
197 model demonstrated superior performance when employing lasso penalty (`alpha=1`)
198 compared to a ridge penalty (`alpha=0`) within the elastic-net regularisation path for
199 regression. Analyte selection also involved the Boruta package in R, which uses wrapper
200 algorithms coupled with random forest classifiers²⁰. The overall accuracy rate is
201 computed along with a 95 percent confidence interval (95% CI) for this rate (using exact
202 binomial test; `binom.test` function) and unweighted Kappa statistics are also calculated.
203 Both accuracy and kappa range from -1 to 1. A value of 1 indicates perfect agreement,
204 while a value of -1 indicates complete disagreement.

205

206 $F1\ score = (1 + \beta^2) * precision * recall / ((\beta^2 * precision) + recall)$, where $precision =$
207 $A / (A + B)$ and $recall = A / (A + C)$.

208 $Kappa = (Observed\ agreement - Chance\ agreement) / (1 - Chance\ agreement)$ or

209 $Kappa = 2 * (TP * TN - FN * FP) / (TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)$

210 $Sensitivity = A / (A + C)$

211 $Specificity = D / (B + D)$

212 $Accuracy = (TP + TN) / (TP + TN + FP + FN)$.

213

214 Boxplots were plotted displaying the distribution of data; with minimum value, the first
215 quartile (Q1), the median, the third quartile (Q3), and the maximum value. The

216 olink_anova_posthoc performs a post-hoc ANOVA test using the function emmeans from
217 the R library emmeans with Tukey p-value adjustment per assay (by OlinkID) at
218 confidence level 0.95, p-value annotations: "****"=0.001, "***"=0.01, "**"=0.05.

219
220 We employed recursive partitioning and regression trees (rpart) to identify protein
221 analytes predictive of COVID-19 status (long COVID and COVID-19 recovered, long
222 COVID and healthy, and COVID-19 recovered and healthy) before vaccination (n = 57).
223 Models were constructed using rpart and fancyRpartPlot functions in R to understand the
224 hierarchical relationships and thresholds of protein analytes, extracted by lasso and
225 Boruta random forest, that differentiated COVID-19 status (long COVID and COVID-19
226 recovered, long COVID and healthy, and COVID-19 recovered and healthy). Decision
227 boundaries, generated using support vector machines (SVM) algorithms from the e1071
228 package in R²¹, were plotted to distinguish between the groups (healthy, recovered, and
229 long COVID) based on the selected protein concentrations.

230
231 Longitudinal data (n = 34) was analysed by olink_lmer function, which fitted a linear
232 mixed-effects model for each protein analyte or antibodies as outcome measure. It uses
233 the lmer function from the R library lmerTest and the anova function from the R library
234 stats. Adjusted p-values are calculated using the p.adjust function from the R library stats
235 with the Benjamini & Hochberg method. The threshold for significance is $p < 0.05$. The
236 olink_lmer_posthoc function is used to extract and plot significant proteins. It performs a
237 post-hoc analysis based on a linear mixed-effects model using the lmer function. Immune
238 dynamics were plotted using linear mixed-effects models fitted by the lmer function of the

239 lme4 package. The resulting linear equation of the model was $y = \alpha + \beta * \text{Group} * \text{Condition}$
240 $+ (1 | \text{Participant})$, where group is disease status and condition are exposure factors.
241 Randomisation was not utilised, as the study aimed to compare individuals with long
242 COVID individuals with those recovered or uninfected or healthy controls. Data collection
243 and analysis were not conducted in a blinded manner with respect to the experimental
244 conditions. Both unsupervised and supervised machine learning algorithms for protein
245 analyte selection were applied, selecting models that demonstrated relatively good fit.

246

247 **Results**

248 **Cohort characteristics**

249 A total of 57 participants (aged 20 to 68 years; 38 females, 67%) enrolled in a longitudinal
250 cohort conducted from October 2020 to April 2023 were included in this study. The
251 participants were categorised into three groups: healthy SARS-CoV-2 PCR/serology
252 confirmed uninfected individuals (n = 24; age 20-68 years, median age 37; 14 (58%)
253 females), COVID-19 recovered (n = 21; age 21-66 years, median age 43; 17 (81%)
254 females) and long COVID (n = 12; age 30-67 years, median age 56; 7 (58%) females)
255 with no significant sex difference between groups but median age higher in the long
256 COVID group (p=0.035) (Fig. 1 and Table 1). Most participants experienced mild to
257 moderate primary COVID-19 disease with only 9.5% of individuals of the COVID-19
258 recovered group requiring hospitalisation due to severe disease (Table 1). A range of co-
259 morbidities (including well-controlled asthma, hypothyroidism, cancer or other chronic
260 illnesses) was observed among the healthy (13%), COVID-19 recovered (48%) and long
261 COVID (48%) participants. Half (50%, 6/12) of individuals diagnosed with long COVID

262 reported an alleviation of symptoms during follow-up (4 became asymptomatic before
263 COVID-19 boosters and 2 before re-infection). Conversely, the remaining cohort (50%)
264 continued to experience long COVID symptomatology both after receiving a booster
265 vaccination and breakthrough SARS-CoV-2 infection. The most common symptom
266 reported in this study cohort was fatigue, brain fog and loss of smell and taste (Fig. 1). Of
267 the 57 participants, longitudinal analysis was performed on 34 participants with paired
268 blood samples at three timepoints: pre-vaccination; 2-4 weeks after receiving a booster
269 COVID-19 vaccine dose; and either after a new SARS-CoV-2 infection (healthy group) or
270 after re-infection (COVID-19 recovered or long COVID groups).

271

272 **Distinct patterns of inflammatory and neurology-related proteins in plasma from**
273 **healthy, COVID-19 recovered individuals and long COVID sufferers.**

274 Persistent inflammation has been proposed as a key mechanism underlying long COVID,
275 particularly because it may explain the varied manifestations associated with this
276 condition. To determine whether long COVID individuals in our low transmission cohort
277 align with previous observations of lasting inflammation and whether inflammatory and
278 neurology-related proteins can be predictive of long COVID, we profiled the levels of 182
279 plasma proteins using multiplexed affinity proteomics. These plasma samples were
280 collected at enrolment for healthy (n = 24) and 6-9 months after primary COVID-19
281 diagnosis for COVID-19 recovered (n = 21) and long COVID individuals (n = 12). All
282 plasma was collected prior to COVID-19 vaccination for this cross-sectional analysis.
283 Initial analysis mapped the overall plasma protein profiles of each individual
284 (supplementary Fig. S1) and correlations between analytes. as Correlation of proteins

285 associated with long COVID disease status were explored using Spearman's rank
286 correlation analyses (supplementary Fig. S2 and S3). In this first analysis, only TRAIL
287 was significantly associated with long COVID ($r = -0.32$, $p = 0.014$) (supplementary Fig.
288 S3). Correlations of analytes between COVID-19 recovered and long COVID groups
289 (supplementary Fig. S4), healthy and long COVID groups (supplementary Fig. S5) and
290 healthy and COVID-19 recovered groups (supplementary Fig. S6) were visualised using
291 network analyses plots of the 182 analytes, where proteins are grouped into clusters
292 suggesting functional or biological relationships, but not causal interactions. Furthermore,
293 we used gene set enrichment analysis (GSEA) to associate disease status with specific
294 groups of proteins (supplementary Fig. S7). While these GSEA results did not reach
295 statistical significance, the observed pathways highlight the potentially ongoing and
296 altered processes in long COVID at 6-9 months following primary SARS-CoV-2 infection.
297 This also extended to some degree to the COVID-19 recovered group where individuals
298 in the COVID-19 recovered group exhibited upregulated neuroinflammatory pathway
299 responses compared to the healthy group. However, individuals with long COVID
300 displayed a more complex neuroinflammatory response profile compared to the COVID-
301 19 recovered group.

302 To further understand which analytes differed significantly between healthy, COVID-19
303 recovered and long COVID groups, lasso regression analysis was also performed. First,
304 we investigated plasma profile differences between long COVID and COVID-19
305 recovered groups. Lasso regression identified 4 top protein analytes associated with long
306 COVID (NBL1, IL-20, LXN, MCP-1), with an accuracy of 0.55 (95%CI: 0.24-0.83) (Fig 2B
307 and C, Table 2). In addition, Boruta random forest and recursive partitioning and

308 regression trees were conducted to identify proteins predictive of COVID-19 status.
309 Consistent with lasso regression analysis, NBL1 and IL-20 were identified as top features
310 with both Boruta random forest (Fig. 2A and C) with similar accuracy of 0.55 (0.23 - 0.83;
311 Table 2) and recursive partitioning and regression trees (Accuracy = 0.88, 95% CI: 0.72-
312 0.97; and F1 score = 0.91) (Fig 3A and Table 2). IL-20 concentrations were particularly
313 lower in individuals who recovered compared to long COVID individuals (estimate or β =
314 -0.29, 95% CI: -0.49 - -0.09, adj p = 0.007) (Fig. 2C). Conversely, long COVID individuals
315 were more likely to have lower levels of NBL1 (< 4.7) than those who recovered (Figs.
316 2C, 3A and B).

317
318 Using the same approach, analysis of plasma protein profiles between the long COVID
319 and healthy groups identified 4 differential proteins (IL-20, CLEC10A, CCL23, TRAIL)
320 using lasso regression analysis (Fig. 2E) with accuracy of 0.67 (0.35 - 0.91; Table 2), and
321 collectively 5 top predictive proteins (HAGH, G-CSF, NAAA) with Boruta random forest
322 (Fig. 2D, F and G) with accuracy of 0.75 (0.43 - 0.95; Table 2) and (HAGH, CLEC10A
323 and NAAA) using decision tree (Accuracy = 0.89, 95% CI: 0.74-0.97; and F1 score = 0.92;
324 Fig. 3C). The healthy group had lower levels of HAGH (estimate= -0.88, 95% CI: -1.58 -
325 -0.18, adj p = 0.02) and CLEC10A concentrations (β = -0.31, 95% CI: -0.62 - -0.005, adj
326 p = 0.046), but higher TRAIL concentration (β = 0.31, 95% CI = 0.08 - 0.55, adj p = 0.01)
327 compared to individuals with long COVID (Fig 2F and G). Specifically, the decision tree
328 analysis revealed that individuals with HAGH NPX levels > 5.4 and/or CLEC10A >4.3 and
329 NAAA < 2.9 were more likely to be suffering from long COVID compared to the healthy
330 group (Fig. 3C and D).

331
332 Finally, we assessed plasma protein profile differences between COVID-19 recovered
333 and healthy groups. In this comparison, a distinct set of differential proteins was identified
334 (ARTN, CST5, FGF-19, MCP-1) using lasso regression (Fig. 2I, J and K) with accuracy
335 of 0.56 (0.31- 0.79) (Table 2). Additionally, predictive proteins consisted of HAGH, DNER
336 as determined using Boruta random forest (Fig. 2H) with accuracy of 0.72 (0.47 - 0.90)
337 and HAGH, MCP-1 and ARTN identified using a decision tree (Accuracy = 0.8, 95% CI:
338 0.65-0.90; and F1 score = 0.82; Fig 3 E and F). The healthy group had higher
339 concentrations of DNER than COVID-19 recovered (β = 0.28, 95% CI: 0.05 - 0.51, adj p
340 = 0.018) and individuals with HAGH NPX levels of > 5.3 and/or MCP-1 <12 and ARTN >
341 -0.15 were more likely to be COVID-19 recovered than healthy (Fig. 3E and F).

342 Collectively these findings demonstrate that distinct plasma protein profiles can be
343 detected in groups of individuals with different levels of SARS-CoV-2 exposure (healthy
344 vs. previously SARS-CoV-2-infected) and varying disease outcomes (COVID-19
345 recovered vs. long COVID), even long after SARS-CoV-2 acute infection and diagnosis.

346
347 **Lower SARS-CoV-2-specific IgG antibody levels following breakthrough infection**
348 **in long COVID and COVID-19 recovered individuals compared to healthy**
349 **individuals.**

350 Humoral responses to COVID-19 vaccines and SARS-CoV-2 infections after vaccination
351 (breakthrough infections) result in high levels of antigen-specific antibodies²²⁻²⁴.
352 However, this response is poorly documented in individuals with long COVID. To gain
353 insight into the overall responsiveness to COVID-19 booster vaccination and SARS-CoV-

354 2 breakthrough infection in our cohort, we measured SARS-CoV-2-specific anti-Spike and
355 anti-NP antibody responses in all individuals (n=57) prior to COVID-19 vaccination, but
356 6-9 months after infection in COVID-19 recovered and long COVID individuals (Fig. 4; left
357 panels). Longitudinal analysis of antibody levels was also performed (Fig. 4; right panels)
358 using paired samples from individuals (n=34) at three timepoints collected at pre-
359 vaccination (pre-vax), 2-4 weeks after the third COVID-19 vaccine dose (post-booster),
360 and 2-4 weeks after either new SARS-CoV-2 infection (healthy group) or being re-
361 infection (COVID-19 recovered or long COVID groups; post(re)infection). As expected
362 prior to vaccination (pre-vax), individuals previously infected with SARS-CoV-2 had
363 significantly higher Spike and NP-specific antibodies than healthy individuals (Fig. 4, p
364 <0.001). However, individuals with long COVID had significantly lower Spike-specific and
365 NP-specific IgG antibodies compared to the COVID-19 recovered group (Fig. 4 left
366 panels, p <0.001) before vaccination. Following booster vaccination all three groups
367 showed high levels of Spike-specific antibodies (Fig. 4; top right panel, p <0.001). At this
368 time point, as expected, NP-specific antibodies remained higher in the COVID-19
369 recovered and long COVID groups compared to the healthy group (Fig. 4; bottom right
370 panel). After breakthrough infection, all groups had increased NP-specific antibody levels
371 (Fig. 4; bottom right panel). However, vaccinees with long COVID showed substantially
372 higher NP-specific antibodies than newly infected vaccinees (healthy). In contrast, Spike-
373 specific antibodies in the latter group were higher compared to antibody levels measured
374 following re-infection in the COVID-19 recovered and long COVID groups (Fig. 4 top right
375 panel).

376

377 **Differential plasma profiles after COVID-19 booster vaccination, primary and**
378 **breakthrough infections in long COVID individuals.**

379 Vaccines are designed to induce short term inflammatory responses in order to achieve
380 desired immune responses that ultimately provide protection from invading pathogens.
381 However, the impact of vaccine-induced inflammatory responses in individuals that are
382 indicated to experience dysregulated inflammatory activity, such as those with long
383 COVID, is not well-established. Similarly, understanding of inflammatory reactions
384 following breakthrough SARS-CoV-2 infection in long COVID individuals is lacking. To
385 investigate these specific aspects, we analysed longitudinal and paired samples from
386 individuals (n=34) across three timepoints as previously (pre-vaccination, post-booster
387 and post-infection/re-infection). In this mixed-effects models analyses, we extracted 7
388 significant proteins (LAT, CASP-8, SIRT2, 4E-BP1, CCL4, STAMBP, and HAGH) from
389 the 182 measured inflammatory and neurology-related protein levels using the Olink
390 proteomics assays. These proteins were either different between groups or across time
391 points within a group (Fig. 5). We found that SIRT2 concentrations were higher in the
392 healthy group compared to COVID-19 recovered pre-vaccination ($p=0.03$; Fig. 5A), and
393 these group differences were maintained following the booster COVID-19 dose ($p = 0.04$).
394 Similar changes were also observed with 4E-BP1 when comparing healthy and COVID-
395 19 recovered groups ($p = 0.02$) at pre-vaccination and following booster ($p = 0.21$; Fig.
396 5B). Notably, SIRT2 and 4E-BP1 NPX values were similar between the long COVID and
397 healthy groups, both at pre-vaccination and post-booster. However, after breakthrough
398 infection, the levels of these two proteins in both long COVID and healthy groups were
399 significantly lower than at pre-vaccination and aligned with the COVID-19 recovered

400 group. Similar patterns were observed with LAT and STAMPB (Fig 5B and C) although
401 substantial differences were only noted between time points within groups. The protein
402 HAGH, previously identified as a differential and predictive marker at pre-vaccination
403 using Boruta random forest and lasso regression analysis, was not elevated following
404 breakthrough infections in this longitudinal analysis (Fig. 5C). In healthy individuals, CCL4
405 levels decreased after breakthrough infection, but remained similar for the three sample
406 events in the COVID-19 recovered and long COVID groups (Fig. 5C). Finally, caspase-8
407 (CASP-8) was significantly higher in the long COVID group at pre-vaccination but
408 decreased to levels similar to those in the COVID-19 recovered and healthy groups.
409 CASP-8 was not induced by either COVID-19 vaccination or a breakthrough infection in
410 any of the three groups (Fig. 5A). Overall, these findings suggest that primary infections
411 and SARS-CoV-2 breakthrough infections result in dissimilar plasma protein profiles,
412 particularly in individuals with long COVID. While COVID-19 booster vaccination is
413 expected to induce a burst of new onset inflammatory responses, this does not appear to
414 exacerbate or persist divergently in long COVID individuals.

415

416 **Discussion**

417 Long COVID is a complex condition that can affect multiple organ systems although its
418 pathophysiology remains elusive. Emerging concepts suggest that prolonged
419 inflammation, viral antigen persistence, human herpesvirus reactivation, dysbiosis,
420 microvascular dysfunction and autoreactive immune responses are probable contributing
421 factors^{25,26}. In this study we focused on establishing the persistent inflammatory and
422 neurology-related plasma protein profiles after initial disease-related effects subside. We

423 also assessed, primary infections, alongside longitudinal samples to investigate if
424 dysregulated inflammatory reactions are mirrored or exacerbated post-COVID-19 booster
425 vaccination and re-infection. Our results suggest that modified levels of neurology-related
426 and inflammation proteins in plasma can persist up to 6-9 months after primary infection
427 in long COVID individuals. Notably, these primary infection responses are not repeated
428 early after breakthrough infection in long COVID individuals.

429

430 We used affinity proteomics to establish inflammatory and neurology-related plasma
431 protein profiles in our cohort sample set. This highly sensitive multiplexed Proximity-
432 Extension Assay (PEA, Olink Proteomics) has increasingly been utilised to both
433 investigate plasma and cerebrospinal fluid (CSF) profiles following COVID-19 in the
434 search for potential biomarkers for long COVID ^{12,27-30}. These studies have collectively
435 identified proteins that can be used to classify long COVID, but also show that different
436 markers can be related to specific long COVID subtypes based on symptoms. Similarly,
437 using the same methodology, we identified a panel of plasma proteins from our cross-
438 sectional analysis (NBL1, IL-20, LXN, MCP-1, CLEC10A, CCL23, TRAIL, HAGH, G-CSF,
439 and NAAA) that were either biomarkers of long COVID or observed at differential levels
440 from either healthy or COVID-19 recovered individuals. Of note, decreased levels of
441 TRAIL, a cytokine with anti-inflammatory properties, in CSF was previously associated
442 with long COVID and align with our finding in plasma ¹². However, while some of our
443 proteins have been reported to be differentially expressed across these specific studies,
444 the remaining proteins in our panel, were not among the proteins previously identified in
445 plasma as predictive of long COVID. It is likely that the divergent panel of proteins

446 reported between the studies is due to the different COVID-19 sub-groups included for
447 comparison. These previous studies focused on differences between acutely ill and/or
448 hospitalised individuals with long COVID sufferers, whereas our study focused on
449 individuals after recovery from mild to moderate disease and long COVID individuals.
450 Importantly our samples were collected within a similar timeframe after diagnosis for both
451 groups. Nevertheless, each study underscores the complexity of long COVID and
452 highlights the importance of considering various comparisons to comprehensively unravel
453 serological features of this condition.

454 We found that IL-20 was a prominent feature of long COVID in our cohort. IL-20 is a
455 pleiotropic cytokine with potent inflammatory properties and has been positively
456 associated to the severity of SARS-CoV-2 infection ³¹. The observation of increased IL-
457 20 in long COVID individuals is intriguing as IL-20 exerts its effects through the activation
458 of signal transducer and activator of transcription (STAT) proteins and specifically via the
459 type I IL-20 receptor complex, which is composed of IL20RA and IL20RB subunits.
460 Cytokines that use IL-20 receptor complex are believed to be linked to the development
461 of chronic inflammation and autoimmune diseases ³² with increased IL-20 concentrations
462 linked to autoantibody production in rheumatoid arthritis patients ³³. This raises the
463 intriguing possibility that dysregulation of the IL-20 pathway, characterised by enhanced
464 IL-20 levels, might contribute to the development of autoimmune features known to occur
465 in a subset of long COVID individuals ^{34,35}. IL-20 also induces synovial fibroblasts and
466 kidney mesangial cells to produce pro-inflammatory molecules, including MCP-1 ^{36,37} a
467 protein that was also found in our study to be a feature of long COVID.

468

469 Given that neurological symptoms, including fatigue and cognitive dysfunction such as
470 memory difficulties, attention deficits and sleep disturbances, are common in long COVID
471 and were frequently reported by our patients, we also quantified protein biomarkers with
472 known relation to neurological disease. Interestingly HAGH, LXN, Linker For Activation
473 Of T Cells (LAT) and NAAA have all been related to cognitive performance in older adults
474 with metabolic syndrome ³⁸. HAGH is also a potential biomarker of Alzheimer disease ³⁹.
475 Increased levels of HAGH have been linked to low cognitive function whereas an inverse
476 relationship of lower levels of LXN, LAT and NAAA were related to higher cognitive
477 function ³⁸. Apart from LAT, these were all elevated in the long COVID group in our cross-
478 sectional analysis. However, HAGH does not appear to be an indicator of long COVID as
479 increased HAGH was reported in mild neuro-COVID cases ¹², but was not a predictor of
480 long COVID in that study. Similarly, we also observed that HAGH was not specifically a
481 feature of long COVID, but rather a feature of SARS-CoV-2 infection as HAGH was a top
482 protein for both COVID-19 recovered and long COVID individuals. Nevertheless, taken
483 together this further verifies a signature of neurological alteration after SARS-CoV-2
484 infection. Features within this signature may prove useful in monitoring cognitive functions
485 in affected individuals.

486 Long COVID has also been associated with specific neuro-inflammatory biomarkers
487 including IL-6, TNF- α , CCL23, G-CSF ^{40,41}. More recently, upregulated G-CSF and MCP-
488 1 have been linked to brain fog ⁴². While our study corroborates prior research connecting
489 long COVID to specific neurological and inflammatory biomarkers such as CCL23, G-
490 CSF, and MCP-1, we found that only MCP-1 was noticeably elevated aligning with
491 previous report ³⁰. In contrast reduced levels of CCL23 and G-CSF were detected in long

492 COVID individuals compared to healthy individuals. The role of both CCL23 and G-CSF
493 in COVID-19 immunity remains perplexing as these proteins have been linked to both
494 severe disease and/or COVID-19 related mortality in COVID-19 patients ⁴³⁻⁴⁸ as well as
495 being linked to asymptomatic disease and survival ^{49,50}. Similarly, unchanged or
496 downregulated MCP-1 has also been associated with long COVID ^{51,52}. Thus, the
497 biological relevance of the association with long COVID for these proteins are likely
498 multifaceted.

499
500 A noteworthy, and to our knowledge, unique aspect of our study was the access to paired
501 samples that allowed analysis of inflammatory profiles after COVID-19 booster
502 vaccination and breakthrough infections in groups of healthy, COVID-19 recovered and
503 long COVID individuals. Using mixed-effects model, we found that before vaccination,
504 protein levels, particularly CASP8 and HAGH, were elevated in plasma of long COVID
505 individuals. Interestingly, the atypical protein levels of CASP8, HAGH and to some extent
506 LAT normalised to ranges similar to those observed in the healthy and COVID-19
507 recovered groups after the COVID-19 booster vaccine. Whether this normalisation is a
508 vaccine-induced effect in long COVID individuals remains to be established and could be
509 achieved by comparing plasma protein profiles immediately before and after vaccination.
510 Alternatively, it is possible that the underlying factors driving atypical protein expression
511 may gradually diminish over time, as suggested by previous research ⁵³. It is also possible
512 that the observed changes in certain proteins (such as CASP-8, HAGH and LAT)
513 correlate with the resolution of cognitive symptoms within our cohort. However, due to the
514 limited sample size, we were unable to thoroughly investigate this potential association.

515 Importantly, we established that the atypical levels of HAGH and CASP8 were not
516 exacerbated nor mirrored by re-infection. Furthermore, we showed that protein levels for
517 SIRT2, 4E-BP1, CCL4 and STAMBP in long COVID individuals and healthy SARS-CoV-
518 2 naïve controls were similar before vaccination. In contrast, COVID-19 recovered
519 individuals had lower levels of these specific molecules especially for SIRT2 and 4E-BP1.
520 The similarity in this profile between long COVID individuals and healthy individuals, and
521 their difference from COVID-19 recovered individuals, suggests a potential lack of down
522 regulation of these proteins in response to primary infection in long COVID individuals.
523 This is particularly interesting to observe with 4E-BP1 and SIRT2 which both have been
524 linked to viral replication. 4E-BP1 is associated with the PI3K/AKT/mTOR pathway and
525 blocking this pathway can have antiviral effects. For instance, modulation of this pathway
526 has previously demonstrated inhibition of Middle East respiratory syndrome coronavirus
527 (MERS-CoV) replication⁵⁴. Similarly, SIRT2 inhibition can provide broad-spectrum
528 antiviral effects including inhibition including inhibition of coronavirus replication^{55,56}.
529 Unaltered levels of these proteins could therefore indicate a potential dysregulation of
530 antiviral processes typically induced to inhibit SARS-CoV-2 replication. Intriguingly the
531 same pattern is not observed upon re-infection in long COVID individuals. Furthermore,
532 decreased levels of both 4E-BP1 and SIRT2 in response to infection are also reflected in
533 previously SARS-CoV-2 naïve healthy individuals. However, whether 4E-BP1 and SIRT2
534 detected in plasma correlates to intracellular levels and activity remains to be determined.
535 Finally, we assessed SARS-CoV-2-specific Spike and NP IgG antibody levels in all study
536 participants. Elevated SARS-CoV-2-specific antibody titres have previously been
537 reported in long COVID individuals^{7,57,58}. Our group had associated these elevated titres

538 with long COVID symptoms such as fatigue ⁵⁹ in low transmission setting, and some
539 studies suggest a potential role of viral persistence in driving these responses ^{60,61}. Other
540 studies have found no significant difference in antibody titres between individuals with
541 long COVID and those who have recovered ^{30,62}. It has been proposed that this may
542 reflect long COVID individuals that follow two main patterns of antiviral immune
543 responses. In some individuals, the response is robust, with strong cellular and humoral
544 responses similar to those seen in people who have fully recovered and in others, there
545 are lower antiviral responses, but still with detectable antigen-specific CD4+ T cells and/or
546 antibodies ⁶³. We observed that both Spike and NP-specific IgG antibodies were lower in
547 the long COVID group than in the recovered group 6-9 months after primary infection.
548 Although this observation is likely the result of multiple factors, future studies could
549 specifically investigate antibody responses in parallel to 4E-BP1 and SIRT2 plasma
550 concentrations in a larger cohort to further decipher the potential divergent antiviral
551 activities in long COVID.

552

553 There are some differences and limitations in our study which may be reflected in the
554 protein biomarkers discovered. Samples were selected based on the availability of
555 specimens that met our long COVID criteria and had been collected at a specific timepoint
556 after diagnosis. Additionally, for the longitudinal analysis, we were restricted by samples
557 available for individuals that had also received a COVID-19 booster vaccination and had
558 had a breakthrough infection while enrolled in the study. For these reasons, our sample
559 set was also heterogeneous in terms of long COVID symptoms. Consequently, this limited
560 the statistical power of our analyses, particularly for subgroup analyses and examining

561 differing long COVID trajectories. This limitation means that non-significant findings may
562 mask true associations between variables, necessitating cautious interpretation of
563 regression estimates following booster and (re-)infection. Larger cohort sizes are needed
564 in future studies to validate these findings. Furthermore, we lack information on the
565 specific variant responsible for breakthrough infections. Although the timing of these
566 breakthrough SARS-CoV-2 infections in our cohort aligns with epidemiological data
567 indicating that most community infections at the time were due to the Omicron variant
568 ^{64,65}, definitive conclusion on breakthrough infective strains cannot be drawn. Finally, our
569 analyses does not reveal how or whether our putative biomarker proteins are
570 mechanistically linked to long COVID and further studies need to be conducted to explore
571 their functional role in this context.

572
573 In conclusion, we identified several biomarkers that were substantially associated with
574 long COVID which can be further explored to discern long COVID disease subsets. Our
575 longitudinal analysis demonstrated that COVID-19 booster vaccination efficiently induced
576 high levels of antigen-specific antibodies without resulting in exacerbation of atypical
577 inflammatory and neurology-related plasma protein levels. Notably, we established that
578 plasma profiles following SARS-CoV-2 breakthrough infection in long COVID individuals
579 diverged from the primary infection that lacked features of previously noted dysregulation.
580 These findings provide intriguing and valuable insight into immune response outcomes,
581 especially to SARS-CoV-2 re-infections, in individuals with long COVID which may
582 facilitate long COVID management strategies for this poorly understood condition.

583 **Figure legends**

584 **Figure 1: Study overview.**

585 Overview of the study including: (Top panel) Cohort participant description, sample
586 collection time points and associated long COVID symptoms. (Middle panel) visual
587 representation of methodology and (Bottom panel) summary of predictive and
588 differentially expressed proteins in long COVID after primary SARS-CoV-2 infection and
589 antibody levels pre-and post-COVID-19 booster vaccination and (re)infection. Created
590 with BioRender.com (2024)

591

592 **Figure 2: Atypical levels of neurology-related and inflammatory biomarkers in**
593 **patients with long COVID detected 6-9 months after primary SARS-CoV-2 infection.**

594 Data representing plasma levels of 182 proteins from 57 individuals was analysed using
595 Boruta feature selection²⁰ and lasso regression models (glmnet function^{18,19}) to identify
596 group differences. Boruta feature selection subplots representing differences between (a)
597 recovered vs. long COVID, (d) healthy vs. long COVID and (h) healthy vs. recovered.
598 Blue box-plots correspond to the shadow attributes (with minimum, mean and maximum
599 attribute values), green box-plots represent important variables, yellow box-plots are
600 tentative, and red box-plots are not significant. Subplots represent the coefficient paths
601 of lasso regression models to identify group differences between (b) COVID-19 recovered
602 vs. long COVID, (e) healthy vs. long COVID and (i) healthy vs. COVID-19 recovered. The
603 x-axis shows the fraction of deviance explained against a coefficient profile plot. Subplots
604 representing boxplots for identified differential proteins between (c) COVID-19 recovered
605 vs. long COVID, (f and g) healthy vs. long COVID and (j and k) healthy vs. COVID-19

606 recovered individuals. Each circle represents an individual. The lower and upper hinges
607 correspond to the first and third quartiles (25th and 75th percentiles), with the median at
608 the midpoint. Boxplots were created using `olink_anova_posthoc` and `olink_boxplot`
609 function ¹⁷ with Tukey p-value adjustment for multiple testing within an assay at
610 confidence level 0.95; two-tailed p <0.05 #; 0.01-0.001 ##; <0.001 ### were considered
611 significant for group differences.

612

613 **Figure 3: Decision tree and hyperplane boundaries displaying hierarchical**
614 **relationships and thresholds of protein features detected at 6-9 months after**
615 **primary SARS-CoV-2 infection to predict long COVID or COVID-19 recovery.**

616 Models were constructed using `rpart` and `fancyRpartPlot` functions in R to understand the
617 hierarchical relationships and thresholds among protein analytes, extracted by lasso
618 regression models and Boruta random forest models that differentiated between (a) long
619 COVID and COVID-19 recovered, (b) long COVID and healthy, and (c) COVID-19
620 recovered and healthy ^{66,67}. Two-dimensional scatter plots of NPX values for the 2 top
621 protein analytes separating three groups (b) NBL1 and IL-20 for COVID-19 recovered vs.
622 long COVID with 57% accuracy (d) CLEC10A and HAGH for healthy vs. long COVID with
623 65% accuracy; and (f) HAGH and MCP-1 for healthy vs. COVID-19 recovered with 56%
624 accuracy). A decision boundary, created using support vector machines (SVM) algorithms
625 from the `e1071` package in R ²¹, separates the groups based on the protein
626 concentrations.

627

628 **Figure 4. High levels of SARS-CoV-2-specific IgG antibodies post-COVID-19**
629 **booster vaccination and post (re)infection.** Left panels: Levels of Spike and anti-
630 nucleoprotein (NP) IgG antibodies in healthy (n=24; black), COVID-19 recovered (n=21;
631 blue) and long COVID individuals (n=12; red) 6-9 months after SARS-CoV-2 infection.
632 Each circle represents an individual. The lower and upper hinges correspond to the first
633 and third quartiles (25th and 75th percentiles), with the median at the midpoint. Right
634 panels: Levels of Spike and anti-nucleoprotein (NP) IgG antibodies in healthy (n=13;
635 black), COVID-19 recovered (n=13; blue) and long COVID individuals (n=8; red) 6-9
636 months after SARS-CoV-2 infection (pre-vax), 2-4 weeks after COVID-19 booster
637 vaccination (post-booster) and 2-4 weeks after SARS-CoV-2 (re)infection (post-
638 (re)infection) The y-axes show antibody levels in median fluorescence intensities (MFI).
639 (Left panels) Wilcoxon Rank Sum and Signed Rank Tests were performed with Tukey p-
640 value adjustment for multiple testing within an assay at confidence level 0.95; two-tailed
641 $p < 0.05$ #; 0.01-0.001 ##; < 0.001 ### were considered significant for group differences.
642 (Right panels) Mixed-effects models with the subject as a random effect was performed
643 to analyse longitudinal data from 34 individuals categorised as HC, COVID-19 recovered
644 and long COVID groups with lmer function and Benjamini-Hochberg procedure in R.
645 Estimated marginal means with 95% confidence intervals are depicted as vertical lines
646 with error bars. Individual data points are represented as circles. Two-tailed $p < 0.05$ #;
647 0.01-0.001 ##; < 0.001 ### were considered significant for group differences (recovered
648 vs. long COVID, HC vs. long COVID and HC vs. recovered) and $p < 0.05$ *; 0.01-0.001
649 **, < 0.001 *** for longitudinal changes.
650

651 **Figure 5. Longitudinal analysis of neuro-inflammatory profiles using mixed-effects**
652 **models at pre-vaccination, post-COVID-19 booster vaccination and post**
653 **(re)infections.**

654 Mixed-effects models with the subject as a random effect were performed to analyse
655 longitudinal data from 34 individuals categorised as healthy, COVID-19 recovered and
656 long COVID. Paired samples collected at pre-vaccination, 2-4 weeks post COVID-19
657 booster vaccination and 2-4 weeks post-(re)infection was included for each individual.
658 Linear modelling was performed to identify important proteins and the Benjamini-
659 Hochberg procedure used to control for false discoveries among the 182 assays used¹⁷.
660 Within-assay multiple testing correction was applied using the Benjamini-Hochberg
661 procedure with Kenward-Roger degrees-of-freedom method for the 15 selected
662 contrasts. Estimated marginal means with 95% confidence intervals are depicted as
663 vertical lines with error bars. Individual data points are represented as circles. Two-tailed
664 test, $p < 0.05$ #; 0.01-0.001 ##; < 0.001 ### were considered significant for group
665 differences (COVID-19 recovered vs. long COVID, healthy vs. long COVID and healthy
666 vs. COVID-19 recovered) and $p < 0.05$ *; 0.01-0.001 **; < 0.001 *** for longitudinal
667 changes⁶⁸

668

669 Table 1: Baseline demographic and clinical characteristics of the study cohort

Variable	Baseline (N = 57)				Longitudinal (n = 34)			
	Healthy* N = 24 ¹	Recovered N = 21 ¹	Long COVID N = 12 ¹	p- value ²	Healthy N = 13 ¹	Recovered N = 13 ¹	Long COVID N = 8 ¹	p- value ²
Sex: Female	14 / 24 (58%)	17 / 21 (81%)	7 / 12 (58%)	0.2	9 / 13 (69%)	12 / 13 (92%)	4 / 8 (50%)	0.11
Age: Median (IQR)	37 (30, 46)	43 (36, 57)	56 (43, 60)	0.035	44 (37, 47)	49 (41, 57)	57 (53, 60)	0.068
Age range (min-max)	20, 68	21, 66	30, 67		20, 63	21, 66	30, 66	
Comorbidity[#]	3 / 24 (13%)	10 / 21 (48%)	7 / 12 (58%)	0.008	3 / 13 (23%)	6 / 13 (46%)	7 / 8 (88%)	0.019
COVID-19 disease status				<0.001				<0.001
Healthy or uninfected*	24 / 24 (100%)	0 / 21 (0%)	0 / 12 (0%)		13 / 13 (100%)	0 / 13 (0%)	0 / 8 (0%)	
Asymptomatic	-	1 / 21 (4.8%)	0 / 12 (0%)		-	1 / 13 (7.7%)	0 / 8 (0%)	
Mild- Moderate	-	18 / 21 (86%)	12 / 12 (100%)		-	11 / 13 (85%)	8 / 8 (100%)	
Severe	-	2 / 21 (9.5%)	0 / 12 (0%)		-	1 / 13 (7.7%)	0 / 8 (0%)	
Vaccine: Two-dose schedule				0.2				0.3
AstraZeneca	5 / 24 (21%)	3 / 21 (14%)	5 / 12 (42%)		4 / 13 (31%)	2 / 13 (15%)	4 / 8 (50%)	
Pfizer	19 / 24 (79%)	18 / 21 (86%)	7 / 12 (58%)		9 / 13 (69%)	11 / 13 (85%)	4 / 8 (50%)	
Vaccine: Third dose								>0.9
Moderna	-	-	-	-	3 / 13 (23%)	4 / 13 (31%)	2 / 8 (25%)	

Pfizer	-	-	-	-	10 / 13 (77%)	9 / 13 (69%)	6 / 8 (75%)	
--------	---	---	---	---	------------------	-----------------	----------------	--

670 ¹n / N (%)

671 ²Fisher's exact test; Kruskal-Wallis rank sum test

672 - Not relevant

673 # Comorbidities include well-controlled asthma, hypothyroidism, cancer or other chronic illnesses

674 * During follow-up after booster COVID-19 doses, individuals in healthy group were infected, whereas individuals
675 who were enrolled as COVID-19 recovered or long COVID were re-infected

676 Table 2: Best features identified by selected models at pre-vaccination time point, and models accuracy.

Condition	No. of features	Best features	Model	F1 score	Kappa	Sensitivity	Specificity	Accuracy(95%CI)
Recovered vs Long COVID	4	NBL1, IL-20, LXN, MCP-1	LASSO (Least Absolute Shrinkage and Selection Operator)	0.62	0.07	0.67	0.40	0.55 (0.24- 0.83)
	2	NBL1, IL-20	Boruta Random Forest	0.62	0.07	0.67	0.40	0.55 (0.23 - 0.83)
Healthy vs Long COVID	4	IL-20, CLEC10A, CCL23, TRAIL	LASSO	0.71	0.31	0.71	0.60	0.67 (0.35 - 0.91)
	3	HAGH, G-CSF, NAAA	Boruta Random Forest	0.82	0.44	0.99	0.40	0.75 (0.43 - 0.95)
Healthy vs Recovered	4	ARTN, CST5, FGF-19, MCP-1	LASSO	0.67	0.18	0.99	0.20	0.56 (0.31- 0.79)
	2	HAGH, DNER	Boruta Random Forest	0.74	0.46	0.88	0.60	0.72 (0.47 - 0.90)

677

678 References

- 679 1 WHO Long COVID Definition, <[https://www.who.int/europe/news-room/fact-sheets/item/post-](https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition)
680 [covid-19-condition](https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition)> (
- 681 2 Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and
682 recommendations. *Nat Rev Microbiol* **21**, 133-146 (2023). [https://doi.org/10.1038/s41579-022-](https://doi.org/10.1038/s41579-022-00846-2)
683 [00846-2](https://doi.org/10.1038/s41579-022-00846-2)
- 684 3 Ceban, F. *et al.* Fatigue and cognitive impairment in Post-COVID-19 Syndrome: A systematic
685 review and meta-analysis. *Brain Behav Immun* **101**, 93-135 (2022).
686 <https://doi.org/10.1016/j.bbi.2021.12.020>
- 687 4 Al-Aly, Z., Bowe, B. & Xie, Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med* **28**,
688 1461-1467 (2022). <https://doi.org/10.1038/s41591-022-01840-0>
- 689 5 Ayoubkhani, D. *et al.* Risk of Long COVID in People Infected With Severe Acute Respiratory
690 Syndrome Coronavirus 2 After 2 Doses of a Coronavirus Disease 2019 Vaccine: Community-Based,
691 Matched Cohort Study. *Open Forum Infect Dis* **9**, ofac464 (2022).
692 <https://doi.org/10.1093/ofid/ofac464>
- 693 6 Group, P.-C. C. Clinical characteristics with inflammation profiling of long COVID and association
694 with 1-year recovery following hospitalisation in the UK: a prospective observational study. *Lancet*
695 *Respir Med* **10**, 761-775 (2022). [https://doi.org/10.1016/S2213-2600\(22\)00127-8](https://doi.org/10.1016/S2213-2600(22)00127-8)
- 696 7 Klein, J. *et al.* Distinguishing features of long COVID identified through immune profiling. *Nature*
697 **623**, 139-148 (2023). <https://doi.org/10.1038/s41586-023-06651-y>
- 698 8 Peluso, M. J. *et al.* Plasma Markers of Neurologic Injury and Inflammation in People With Self-
699 Reported Neurologic Postacute Sequelae of SARS-CoV-2 Infection. *Neurol Neuroimmunol*
700 *Neuroinflamm* **9** (2022). <https://doi.org/10.1212/NXI.000000000200003>
- 701 9 Phetsouphanh, C. *et al.* Immunological dysfunction persists for 8 months following initial mild-to-
702 moderate SARS-CoV-2 infection. *Nat Immunol* **23**, 210-216 (2022).
703 <https://doi.org/10.1038/s41590-021-01113-x>
- 704 10 Cervia-Hasler, C. *et al.* Persistent complement dysregulation with signs of thromboinflammation
705 in active Long Covid. *Science* **383**, eadg7942 (2024). <https://doi.org/10.1126/science.adg7942>
- 706 11 Gu, X. *et al.* Probing long COVID through a proteomic lens: a comprehensive two-year longitudinal
707 cohort study of hospitalised survivors. *EBioMedicine* **98**, 104851 (2023).
708 <https://doi.org/10.1016/j.ebiom.2023.104851>
- 709 12 Etter, M. M. *et al.* Severe Neuro-COVID is associated with peripheral immune signatures,
710 autoimmunity and neurodegeneration: a prospective cross-sectional study. *Nat Commun* **13**,
711 6777 (2022). <https://doi.org/10.1038/s41467-022-34068-0>
- 712 13 Tandon, P. *et al.* Unraveling Links between Chronic Inflammation and Long COVID: Workshop
713 Report. *J Immunol* **212**, 505-512 (2024). <https://doi.org/10.4049/jimmunol.2300804>
- 714 14 Eriksson, E. M. *et al.* Cohort Profile: A longitudinal Victorian COVID-19 cohort (COVID PROFILE).
715 *medRxiv*, 2023.2004.2027.23289157 (2023). <https://doi.org/10.1101/2023.04.27.23289157>
- 716 15 Mazhari, R. *et al.* SARS-CoV-2 Multi-Antigen Serology Assay. *Methods Protoc* **4** (2021).
717 <https://doi.org/10.3390/mps4040072>
- 718 16 Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity,
719 and excellent scalability. *PLoS One* **9**, e95192 (2014).
720 <https://doi.org/10.1371/journal.pone.0095192>
- 721 17 Nevola K *et al.* `_OlinkAnalyze`: Facilitate Analysis of Proteomic Data from Olink_. R package version
722 3.6.2,. (2024).
- 723 18 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via
724 Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

- 725 19 Tay, J. K., Narasimhan, B. & Hastie, T. Elastic Net Regularization Paths for All Generalized Linear
726 Models. *J Stat Softw* **106** (2023). <https://doi.org/10.18637/jss.v106.i01>
- 727 20 Kursu, M. B., & Rudnicki, W. R. Feature Selection with the Boruta Package. *Journal of Statistical*
728 *Software* **36**, 1-13 (2010).
- 729 21 Meyer D, Dimitriadou E, Hornik K, Weingessel A & F, L. _e1071: Misc Functions of the Department
730 of Statistics, Probability Theory Group (Formerly: E1071), TU Wien_. R package version 1.7-14.
731 (2023).
- 732 22 Gianfagna, F. *et al.* Anti-SARS-CoV-2 antibody levels and kinetics of vaccine response: potential
733 role for unresolved inflammation following recovery from SARS-CoV-2 infection. *Sci Rep* **12**, 385
734 (2022). <https://doi.org/10.1038/s41598-021-04344-y>
- 735 23 Koutsakos, M. *et al.* The magnitude and timing of recalled immunity after breakthrough infection
736 is shaped by SARS-CoV-2 variants. *Immunity* **55**, 1316-1326 e1314 (2022).
737 <https://doi.org/10.1016/j.immuni.2022.05.018>
- 738 24 Zhang, Z. *et al.* Humoral and cellular immune memory to four COVID-19 vaccines. *Cell* **185**, 2434-
739 2451 e2417 (2022). <https://doi.org/10.1016/j.cell.2022.05.022>
- 740 25 Mehandru, S. & Merad, M. Pathological sequelae of long-haul COVID. *Nat Immunol* **23**, 194-202
741 (2022). <https://doi.org/10.1038/s41590-021-01104-y>
- 742 26 Peluso, M. J. & Deeks, S. G. Early clues regarding the pathogenesis of long-COVID. *Trends Immunol*
743 **43**, 268-270 (2022). <https://doi.org/10.1016/j.it.2022.02.008>
- 744 27 Iosef, C. *et al.* Plasma proteome of Long-COVID patients indicates HIF-mediated vasculo-
745 proliferative disease with impact on brain and heart function. *J Transl Med* **21**, 377 (2023).
746 <https://doi.org/10.1186/s12967-023-04149-9>
- 747 28 Liew, F. *et al.* Large-scale phenotyping of patients with long COVID post-hospitalization reveals
748 mechanistic subtypes of disease. *Nat Immunol* **25**, 607-621 (2024).
749 <https://doi.org/10.1038/s41590-024-01778-0>
- 750 29 Patel, M. A. *et al.* Organ and cell-specific biomarkers of Long-COVID identified with targeted
751 proteomics and machine learning. *Mol Med* **29**, 26 (2023). <https://doi.org/10.1186/s10020-023-00610-z>
- 752
- 753 30 Zhao, J. *et al.* Plasma biomarkers for systemic inflammation in COVID-19 survivors. *Proteomics Clin*
754 *Appl* **16**, e2200031 (2022). <https://doi.org/10.1002/prca.202200031>
- 755 31 Rajamanickam, A. *et al.* Characterization of IL-10 Family of Cytokines in Acute and Convalescent
756 COVID-19 Individuals. *J Interferon Cytokine Res* **43**, 469-477 (2023).
757 <https://doi.org/10.1089/jir.2023.0075>
- 758 32 Chen, J., Caspi, R. R. & Chong, W. P. IL-20 receptor cytokines in autoimmune diseases. *J Leukoc*
759 *Biol* **104**, 953-959 (2018). <https://doi.org/10.1002/JLB.MR1117-471R>
- 760 33 Kragstrup, T. W. *et al.* The interleukin-20 receptor axis in early rheumatoid arthritis: novel links
761 between disease-associated autoantibodies and radiographic progression. *Arthritis Res Ther* **18**,
762 61 (2016). <https://doi.org/10.1186/s13075-016-0964-7>
- 763 34 Chen, Y. *et al.* New-onset autoimmune phenomena post-COVID-19 vaccination. *Immunology* **165**,
764 386-401 (2022). <https://doi.org/10.1111/imm.13443>
- 765 35 Wang, E. Y. *et al.* Diverse functional autoantibodies in patients with COVID-19. *Nature* **595**, 283-
766 288 (2021). <https://doi.org/10.1038/s41586-021-03631-y>
- 767 36 Hsu, Y. H. *et al.* Anti-IL-20 monoclonal antibody inhibited inflammation and protected against
768 cartilage destruction in murine models of osteoarthritis. *PLoS One* **12**, e0175802 (2017).
769 <https://doi.org/10.1371/journal.pone.0175802>
- 770 37 Kragstrup, T. W. *et al.* Increased interleukin (IL)-20 and IL-24 target osteoblasts and synovial
771 monocytes in spondyloarthritis. *Clin Exp Immunol* **189**, 342-351 (2017).
772 <https://doi.org/10.1111/cei.12973>

- 773 38 Llaurador-Coll, M. *et al.* Plasma levels of neurology-related proteins are associated with cognitive
774 performance in an older population with overweight/obesity and metabolic syndrome.
775 *Geroscience* **45**, 2457-2470 (2023). <https://doi.org/10.1007/s11357-023-00764-y>
- 776 39 Ahmad, S. *et al.* CDH6 and HAGH protein levels in plasma associate with Alzheimer's disease in
777 APOE epsilon4 carriers. *Sci Rep* **10**, 8233 (2020). <https://doi.org/10.1038/s41598-020-65038-5>
- 778 40 Lai, Y. J. *et al.* Biomarkers in long COVID-19: A systematic review. *Front Med (Lausanne)* **10**,
779 1085988 (2023). <https://doi.org/10.3389/fmed.2023.1085988>
- 780 41 Thomas, C. *et al.* Blood Biomarkers of Long COVID: A Systematic Review. *Mol Diagn Ther* (2024).
781 <https://doi.org/10.1007/s40291-024-00731-z>
- 782 42 Greene, C. *et al.* Blood-brain barrier disruption and sustained systemic inflammation in individuals
783 with long COVID-associated cognitive impairment. *Nat Neurosci* **27**, 421-432 (2024).
784 <https://doi.org/10.1038/s41593-024-01576-9>
- 785 43 Bello, S. *et al.* IL-6 and cfDNA monitoring throughout COVID-19 hospitalization are accurate
786 markers of its outcomes. *Respir Res* **24**, 125 (2023). [https://doi.org/10.1186/s12931-023-02426-](https://doi.org/10.1186/s12931-023-02426-1)
787 [1](https://doi.org/10.1186/s12931-023-02426-1)
- 788 44 deKay, J. T. *et al.* DEspR(high) neutrophils are associated with critical illness in COVID-19. *Sci Rep*
789 **11**, 22463 (2021). <https://doi.org/10.1038/s41598-021-01943-7>
- 790 45 Hawerkamp, H. C. *et al.* Characterisation of the pro-inflammatory cytokine signature in severe
791 COVID-19. *Front Immunol* **14**, 1170012 (2023). <https://doi.org/10.3389/fimmu.2023.1170012>
- 792 46 Kalinina, O. *et al.* Cytokine Storm Signature in Patients with Moderate and Severe COVID-19. *Int J*
793 *Mol Sci* **23** (2022). <https://doi.org/10.3390/ijms23168879>
- 794 47 Metzemaekers, M. *et al.* Kinetics of peripheral blood neutrophils in severe coronavirus disease
795 2019. *Clin Transl Immunology* **10**, e1271 (2021). <https://doi.org/10.1002/cti2.1271>
- 796 48 Noren Hooten, N. *et al.* Association of extracellular vesicle inflammatory proteins and mortality.
797 *Sci Rep* **12**, 14049 (2022). <https://doi.org/10.1038/s41598-022-17944-z>
- 798 49 Monserrat, J. *et al.* Role of Innate and Adaptive Cytokines in the Survival of COVID-19 Patients. *Int*
799 *J Mol Sci* **23** (2022). <https://doi.org/10.3390/ijms231810344>
- 800 50 Soares-Schanoski, A. *et al.* Asymptomatic SARS-CoV-2 Infection Is Associated With Higher Levels
801 of Serum IL-17C, Matrix Metalloproteinase 10 and Fibroblast Growth Factors Than Mild
802 Symptomatic COVID-19. *Front Immunol* **13**, 821730 (2022).
803 <https://doi.org/10.3389/fimmu.2022.821730>
- 804 51 Besteher, B. *et al.* Cortical thickness alterations and systemic inflammation define long-COVID
805 patients with cognitive impairment. *Brain Behav Immun* **116**, 175-184 (2024).
806 <https://doi.org/10.1016/j.bbi.2023.11.028>
- 807 52 Kovarik, J. J. *et al.* A multi-omics based anti-inflammatory immune signature characterizes long
808 COVID-19 syndrome. *iScience* **26**, 105717 (2023). <https://doi.org/10.1016/j.isci.2022.105717>
- 809 53 Phetsouphanh, C. *et al.* Improvement of immune dysregulation in individuals with long COVID at
810 24-months following SARS-CoV-2 infection. *Nat Commun* **15**, 3315 (2024).
811 <https://doi.org/10.1038/s41467-024-47720-8>
- 812 54 Kindrachuk, J. *et al.* Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation
813 for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome
814 analysis. *Antimicrob Agents Chemother* **59**, 1088-1099 (2015).
815 <https://doi.org/10.1128/AAC.03659-14>
- 816 55 Kong, F. *et al.* Sirtuins as Potential Therapeutic Targets for Hepatitis B Virus Infection. *Front Med*
817 *(Lausanne)* **8**, 751516 (2021). <https://doi.org/10.3389/fmed.2021.751516>
- 818 56 Roche, K. L. *et al.* An allosteric inhibitor of sirtuin 2 deacetylase activity exhibits broad-spectrum
819 antiviral activity. *J Clin Invest* **133** (2023). <https://doi.org/10.1172/JCI158978>

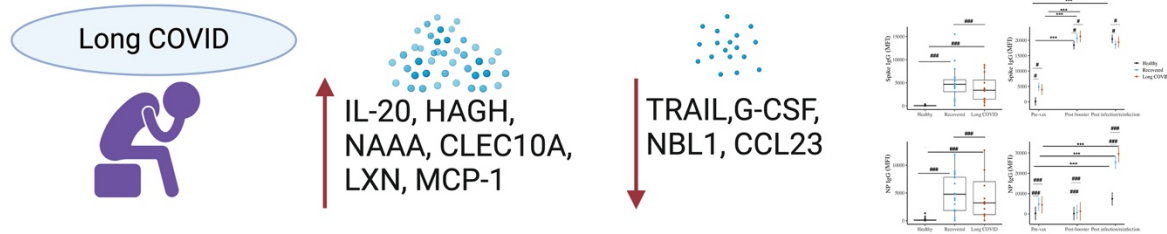
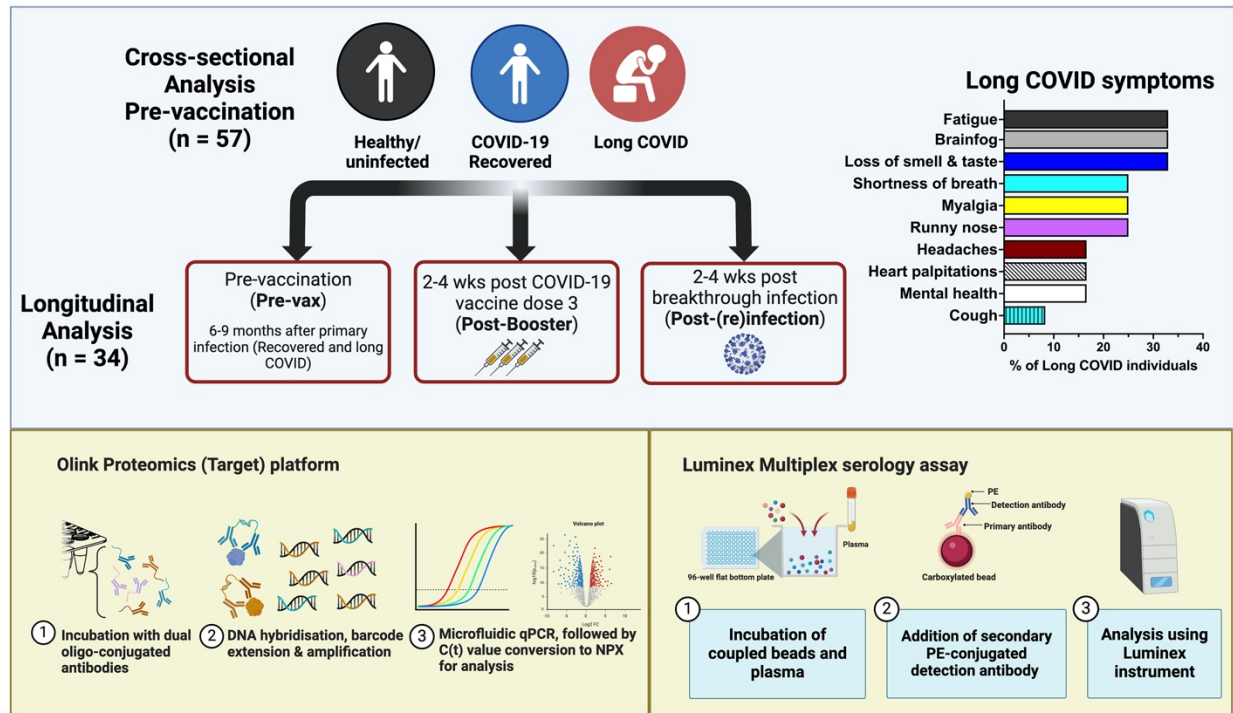
- 820 57 Acosta-Ampudia, Y. *et al.* Persistent Autoimmune Activation and Proinflammatory State in Post-
821 Coronavirus Disease 2019 Syndrome. *J Infect Dis* **225**, 2155-2162 (2022).
822 <https://doi.org/10.1093/infdis/jiac017>
- 823 58 Yin, K. *et al.* Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated
824 adaptive immune response to SARS-CoV-2. *Nat Immunol* **25**, 218-225 (2024).
825 <https://doi.org/10.1038/s41590-023-01724-6>
- 826 59 Blomberg, B. *et al.* Long COVID in a prospective cohort of home-isolated patients. *Nat Med* **27**,
827 1607-1613 (2021). <https://doi.org/10.1038/s41591-021-01433-3>
- 828 60 Ghafari, M. *et al.* Prevalence of persistent SARS-CoV-2 in a large community surveillance study.
829 *Nature* **626**, 1094-1101 (2024). <https://doi.org/10.1038/s41586-024-07029-4>
- 830 61 Zollner, A. *et al.* Postacute COVID-19 is Characterized by Gut Viral Antigen Persistence in
831 Inflammatory Bowel Diseases. *Gastroenterology* **163**, 495-506 e498 (2022).
832 <https://doi.org/10.1053/j.gastro.2022.04.037>
- 833 62 Augustin, M. *et al.* Immunological fingerprint in coronavirus disease-19 convalescents with and
834 without post-COVID syndrome. *Front Med (Lausanne)* **10**, 1129288 (2023).
835 <https://doi.org/10.3389/fmed.2023.1129288>
- 836 63 Kervecan, J. *et al.* Divergent adaptive immune responses define two types of long COVID. *Front*
837 *Immunol* **14**, 1221961 (2023). <https://doi.org/10.3389/fimmu.2023.1221961>
- 838 64 COVID-19 Australia: Epidemiology Report 57
839 <[https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA25](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_57_reporting_period_ending_16_january_2022.pdf)
840 [87320081F7BF/\\$File/covid_19_australia_epidemiology_report_57_reporting_period_ending_16](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_57_reporting_period_ending_16_january_2022.pdf)
841 [_january_2022.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_57_reporting_period_ending_16_january_2022.pdf)> (
- 842 65 COVID-19 Australia: Epidemiology Report60,
843 <[https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA25](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_60_reporting_period_ending_10_april_2022.pdf)
844 [87320081F7BF/\\$File/covid_19_australia_epidemiology_report_60_reporting_period_ending_10](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_60_reporting_period_ending_10_april_2022.pdf)
845 [_april_2022.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/Content/C50CAE02452A48A7CA2587320081F7BF/$File/covid_19_australia_epidemiology_report_60_reporting_period_ending_10_april_2022.pdf)> (
- 846 66 Therneau T, A. B. *rpart: Recursive Partitioning and Regression Trees_*. R package version 4.1.23,
847 <<https://CRAN.R-project.org/package=rpart>> (2023).
- 848 67 Williams, G. J. Data mining with rattle and R: the art of excavating data for knowledge discovery.
849 Series Use R!, Springer. (2011).
- 850 68 Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means_*. R package version
851 1.10.0, (2024).

852

853

854

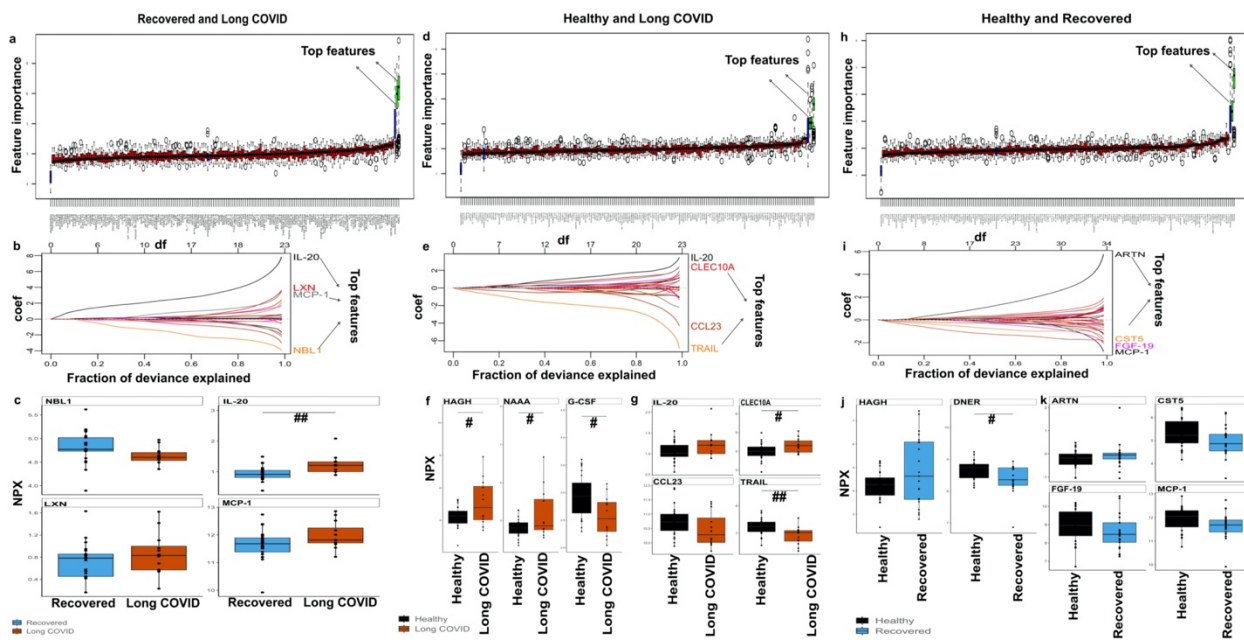
855 **Figure 1**



856

857

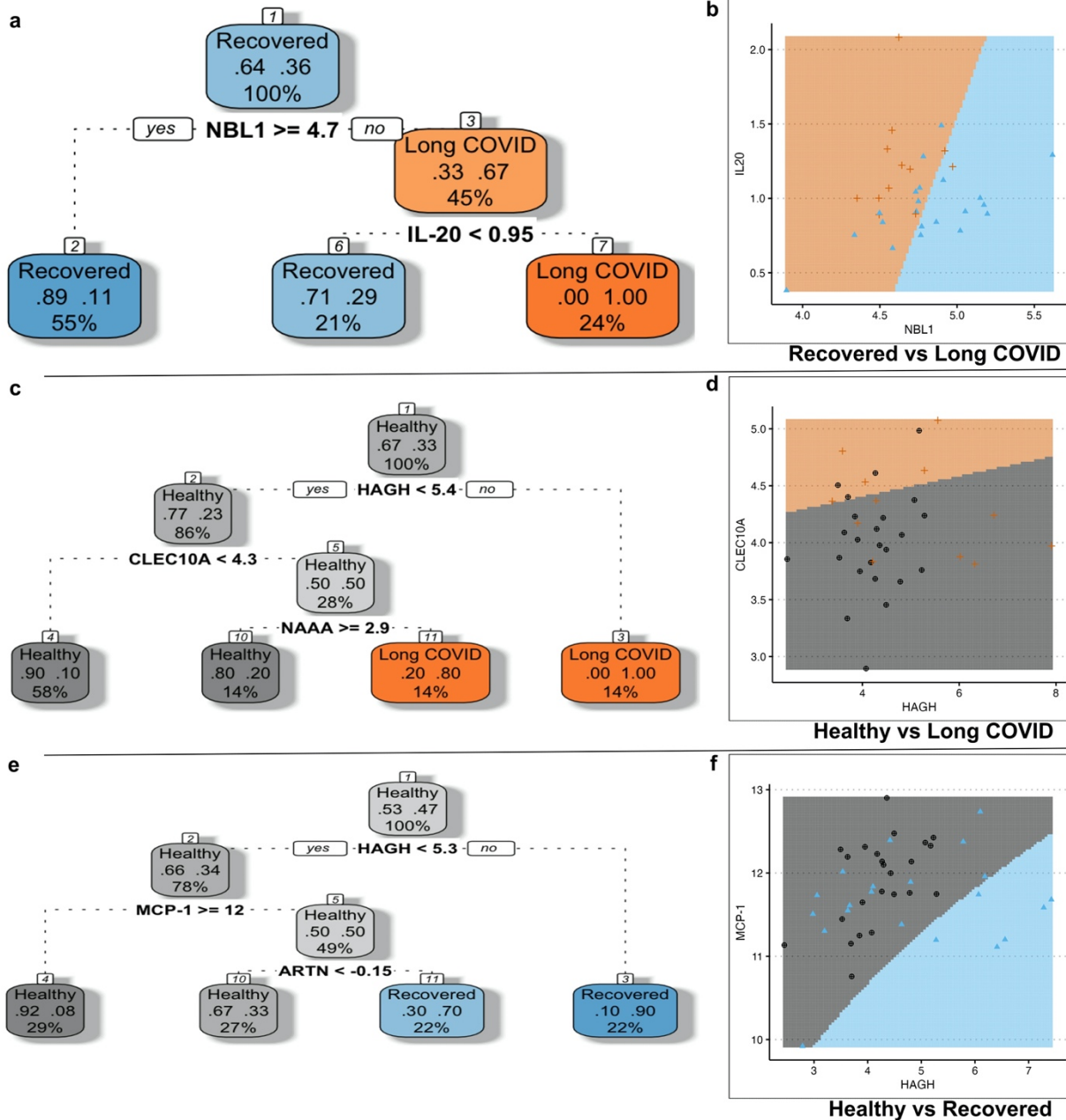
858 **Figure 2**



859

860

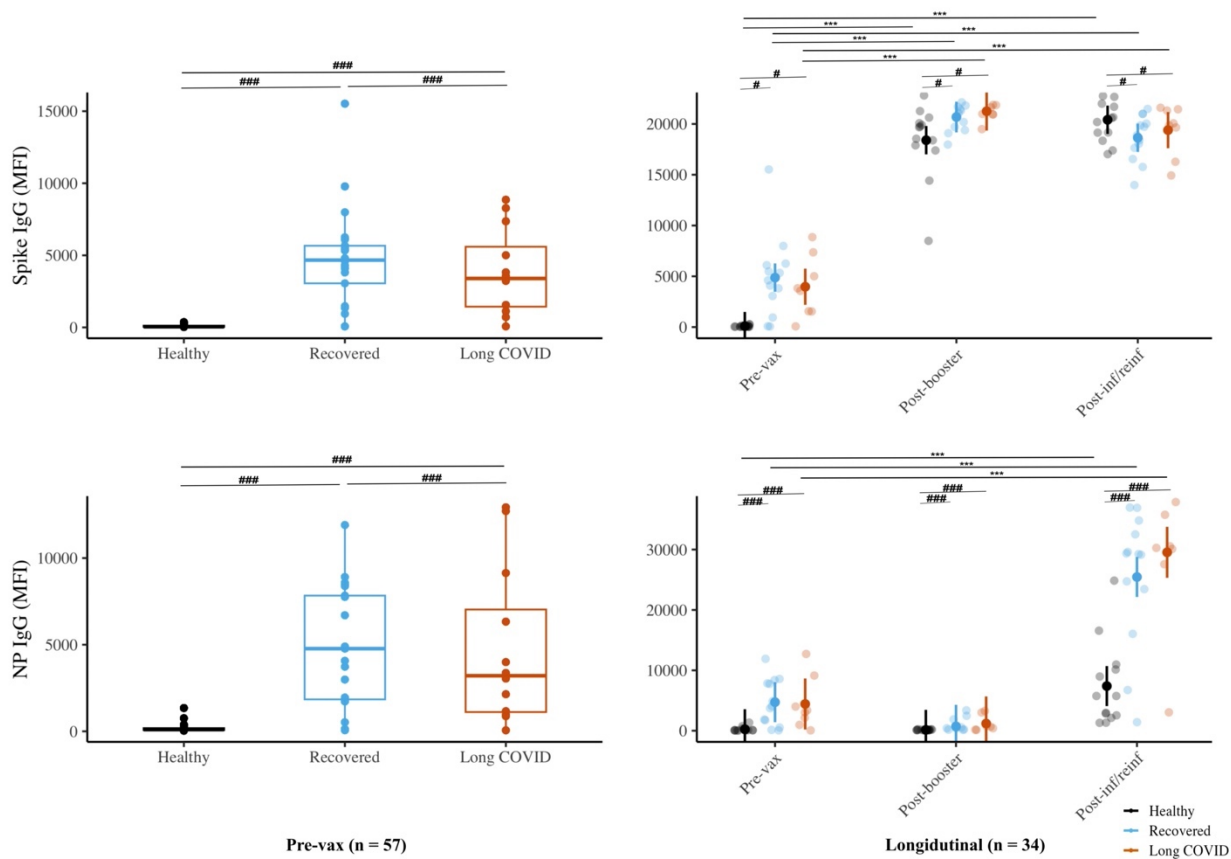
861 **Figure 3**



862

863

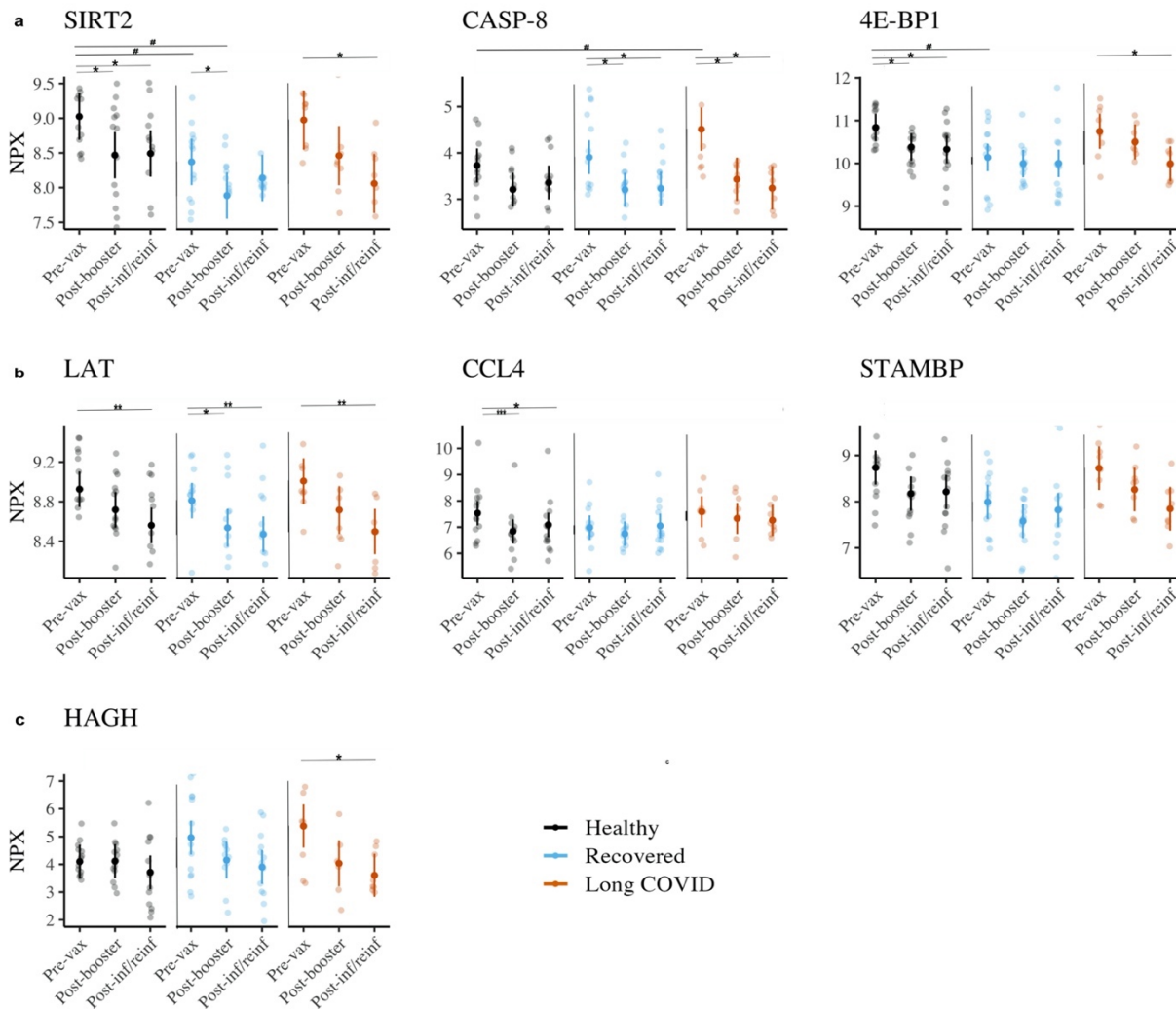
864 **Figure 4**



865

866

867 **Figure 5**



868

869

870

871 **Acknowledgements**

872 We wish to thank Anne Hart, Maureen Ford and all members of the COVID PROFILE
873 consortium and the study participants in the COVID PROFILE study. We also wish to
874 thank Imadh Azeez and Dr. Lauren Smith for valuable discussion and input around data
875 analysis. Dr. Bansal received funding from the University of Bergen, The National
876 Graduate School in Infection Biology and Antimicrobials (or IBA) and Pasteur legatet &
877 Thjøtta's legat, University of Oslo, Norway [101563]. The COVID PROFILE study was
878 supported by WHO Unity funds (2020/1085469-0), and WEHI philanthropic funds. I.M. is
879 supported by an NHMRC Senior Research Fellowship (#1043345). This work was made
880 possible through Victorian State Government Operational Infrastructure Support and
881 Australian Government NHMRC IRIISS.

882

883 **Author contribution:**

884 SO and EME conceptualised the study. AB, SO, JC and EME performed the data
885 analysis. SO, NWK and RM generated the data. AB, SO, NWK, RB and EME designed
886 experimental work, AB, SO and EME wrote the first draft of the manuscript. RB, IM and
887 RJC provided guidance and discussion and edited the manuscript.

888

889 **COVID PROFILE consortium members:** Emily M. Eriksson, Anne Hart, Maureen
890 Forde, Siavash Foroughi, Nicholas Kiernan-Walker, Ramin Mazhari, Erin C. Lucas, Mai
891 Margetts, Anthony Farchione, Dylan Sheerin, George Ashdown, Rachel Evans,
892 Catherine Chen, Shazia Ruybal-Pesántez, Eamon Conway, Marilou H. Barrios, Jasper

893 Cornish, Maria Edmonds, Lee M. Henneken, Lisa J. Ioannidis, Sam W. Z. Olechnowicz,
894 Ryan B. Munnings, Joanna R. Groom, Diana S. Hansen, Rory Bowden, Anna K.
895 Coussens, Jason A. Tye-Din, Vanessa L. Bryant, Ivo Mueller

896

897 **Declaration of interests:** The authors declare no competing interests.

898

899 **Data and code availability:** To protect patient privacy, this study uses anonymised data
900 from the COVID PROFILE study. Access to more data requires committee approval in
901 Melbourne and will be provided by the corresponding author upon request. This paper
902 does not report original code. This project leverages open-source R code. We have
903 documented the specific R packages and functions used in relevant sections. Any
904 additional information required to reanalyse the data reported in this paper is available
905 from the lead contact upon reasonable request.

906