

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Infrared spectroscopy enables rapid, robust, portable COVID-19 saliva screening based on pathophysiological response to SARS-CoV-2

Short title: FTIR saliva analysis and SARS-CoV-2 infection

Seth T. Kazmer¹, Gunter Hartel^{2¶}, Harley Robinson^{1¶}, Renee S. Richards¹, Kexin Yan³, Sebastiaan J. Van Hal⁴, Raymond Chan⁴, Andrew Hind⁵, David Bradley⁵, Fabian Zieschang⁵, Daniel J. Rawle³, Thuy T. Le³, David W. Reid⁶, Andreas Suhrbier^{3,7}, Michelle M Hill^{1,8*}

¹Precision & Systems Biomedicine Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia 4006.

²Biostatistics Unit, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia 4006.

³Inflammation Biology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia 4006.

⁴New South Wales Health Pathology-Royal Prince Alfred Hospital, Camperdown, NSW, Australia 2050.

⁵Agilent Technologies Australia, Mulgrave, VIC, Australia 3170.

⁶Lung Inflammation & Infection Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia 4006, and The Prince Charles Hospital, Brisbane, Australia.

⁷Australian Infectious Disease Research Centre, GVN Centre of Excellence, Brisbane, Queensland, Australia, 4029 and 4072.

⁸UQ Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, Brisbane, QLD, Australia 4006.

*Corresponding author.

Email: michelle.hill@qimrberghofer.edu.au

¶ These authors contributed equally to this work

32 **Abstract**

33 Fourier-transform infrared (FTIR) spectroscopy provides a (bio)chemical snapshot of the sample,
34 and was recently proposed for COVID-19 saliva screening in proof-of-concept cohort studies. As
35 a step towards translation of this technology, we conducted controlled validation experiments in
36 multiple biological systems. SARS-CoV-2 or UV-inactivated SARS-CoV-2 were used to infect
37 Vero E6 cells *in vitro*, and K18-hACE2 mice *in vivo*. Potentially infectious culture supernatant or
38 mouse oral lavage samples were treated with ethanol or Trizol to 75% (v/v) for attenuated total
39 reflectance (ATR)-FTIR spectroscopy, or RT-PCR, respectively. The control condition, UV-
40 inactivated SARS-CoV-2 elicited strong biochemical changes in culture supernatant/oral lavage
41 despite lack of replication determined by RT-PCR or cell culture infectious dose 50%. Crucially,
42 we show that active SARS-CoV-2 infection induced additional FTIR signals over the UV-
43 inactivated SARS-CoV-2 infection, which correspond to innate immune response, aggregated
44 proteins, and RNA. For human patient cohort prediction, we achieved high sensitivity of 93.48%
45 on leave-on-out cross validation (n=104 participants) for predicting COVID-19 positivity using a
46 partial least squares discriminant analysis model, in agreement with recent studies. However,
47 COVID-19 patients negative on follow-up (RT-PCR on day of saliva sampling) were poorly
48 predicted in this model. Importantly, COVID-19 vaccination did not lead to mis-classification of
49 COVID-19 negatives. Meta-analysis revealed SARS-CoV-2 induced increase in Amide II band
50 in all arms of this study and recent studies, indicative of altered β -sheet structures in secreted
51 proteins. In conclusion, ATR-FTIR is a robust, simple, portable method for COVID-19 saliva
52 screening based on detection of pathophysiological responses to SARS-CoV-2.

53

54 **Keywords:** Fourier-transform infrared spectroscopy; FTIR; COVID-19 pandemic; kallikrein

55 Introduction

56

57 More than two years into the coronavirus disease 2019 (COVID-19) pandemic, there
58 remains a need for globally affordable rapid, field-deployable screening tests to detect active
59 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection. While the current
60 gold standard reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) test is
61 highly sensitive in detecting SARS-CoV-2 RNA, the technical requirements, time to result and
62 the accumulated testing costs are prohibitive in developing countries and for disadvantaged
63 communities. Equipment-free rapid antigen tests with immobilized anti-SARS-CoV-2 antibodies
64 in lateral flow devices have been developed to generate results in 5-20 minutes, but the reported
65 variable sensitivity of such diagnostics remain unaddressed (1, 2), and there is also the
66 imperative need to re-evaluate antibody sensitivity as each new variant emerges.

67

68 As an alternative to antibody-based rapid testing, Fourier transform infrared (FTIR)
69 spectroscopy was recently reported as a promising, point-of-care technology for COVID-19
70 detection using pharyngeal swab or saliva (3-5). FTIR provides a biochemical snapshot of the
71 sample by measuring the vibration of chemical bonds (6). FTIR spectra collected from saliva of
72 COVID-19 patients and healthy controls were used to develop prediction algorithms that
73 demonstrated high predictive accuracy in cross-validation of the same cohort (4, 5) or in an
74 independent cohort (3). FTIR sampling using either transfection (slide mount), or attenuated
75 total reflectance (ATR, directly deposited on highly reflective crystal) was able to distinguish
76 healthy controls from confirmed COVID-19 cases with high specificity and sensitivity (3-5).

77

78 These recent cross-sectional cohort studies provided promising proof-of-concept for the
79 use of FTIR in COVID-19 screening using saliva as a non-invasive sample that can be self-
80 collected. To further develop this technology towards point-of-care application, we generated
81 comprehensive data on the pathobiological basis underpinning the SARS-CoV-2/COVID-19
82 FTIR signal using a rapid and biosafe processing method. We utilized three different biological
83 systems: culture supernatant from *in vitro* cell infection, oral lavage of inoculated hACE-2 mice,
84 and human saliva from a limited cohort of COVID-19 patients and health controls. For the cell
85 and mouse models, UV-inactivated SARS-CoV-2 virus was used as control, and two post-
86 infection time points were examined. The potentially infectious biological samples were first
87 decontaminated by adding 100% ethanol (v/v) to 75% final (v/v), as per our recent study using
88 ATR-FTIR of plasma samples for prediction of COVID-19 disease severity (7). The high ethanol
89 percentage facilitated the rapid evaporation of the treated plasma (1 μ l) on the ATR-FTIR target
90 (~30 sec), which enables very rapid data acquisition (7). In the current study, we used the same
91 procedure to analysed cell culture secretome, mouse oral lavage and human saliva samples, and
92 additionally conducted proteomic analysis of the mouse oral lavage samples to elucidate the
93 pathobiology. Finally, a meta-analysis was conducted of all available COVID-19 FTIR spectra
94 data.

95

96

97 **Results**

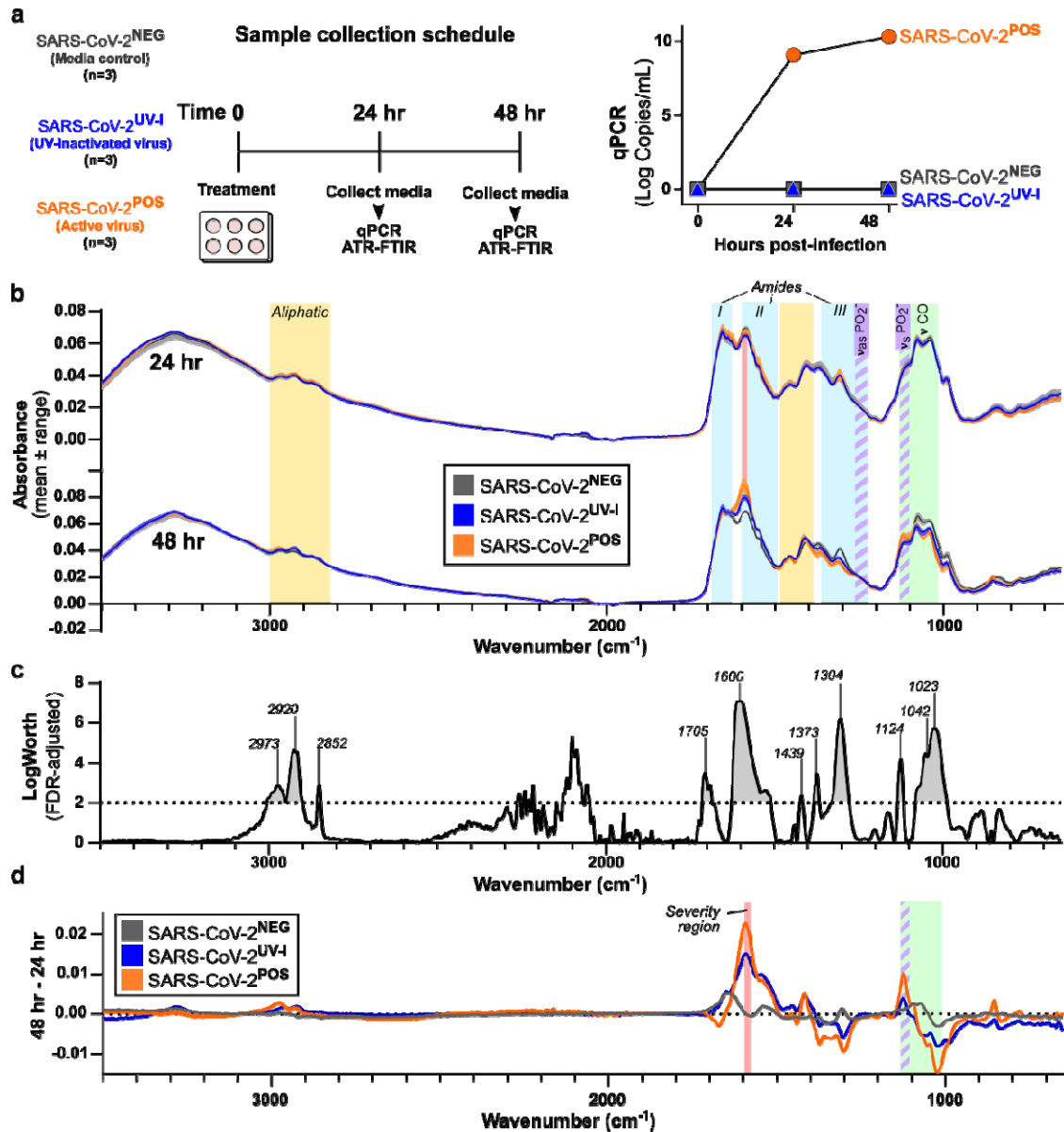
98 **Characterisation of *in vitro* SARS-CoV-2 infection-induced** 99 **secretome ATR-FTIR spectra**

100
101 As a first step, a standard Vero cell *in vitro* infection model was used to investigate the
102 secretory host response to SARS-CoV-2 infection. Two controls were used: a media control and
103 ultraviolet light (UV)-inactivated SARS-CoV-2, which cannot replicate as UV destroys RNA.
104 RT-qPCR of SARS-CoV-2 RNA of the culture supernatant confirmed the lack of infectivity for
105 both controls while the active infection demonstrated an increased SARS-CoV-2 RNA load at 24
106 and 48 hrs (Fig 1a).

107
108 Interestingly, despite the 9 log increase in SARS-CoV-2 RNA at 24 hr, the FTIR spectra
109 of the supernatant showed minimal change at this time point, apart from increased absorbance at
110 Amide I band ($1700\text{-}1600\text{ cm}^{-1}$) in the active SARS-CoV-2 infected sample (Fig 1b). At 48 hr,
111 the FTIR profiles of UV-inactivated SARS-CoV-2 and active SARS-CoV-2 infected
112 supernatants showed increased bands at 2970 cm^{-1} , 2924 cm^{-1} , 2874 cm^{-1} , 1590 cm^{-1} , 1415 cm^{-1} ,
113 and decreased at 1373 cm^{-1} , 1309 cm^{-1} , 1042 cm^{-1} , 988 cm^{-1} compared to media control (Fig 1b
114 and S1a Fig). However, at 48 hr active SARS-CoV-2 infected secretome displayed separation
115 from both controls in Amide I/II ($1700\text{-}1470\text{ cm}^{-1}$) and fingerprint (FP) region ($1450\text{-}600\text{ cm}^{-1}$)
116 (Fig 1b), as well as right shifting to a lower wavenumber at 1668 cm^{-1} to 1595 cm^{-1} (Fig 1d and
117 S1b Fig).

118
119 An FDR LogWorth analysis confirmed significance in a number of these wavelengths
120 from both controls, shown as regions above the dotted line in Fig. 1c ($p < 0.001$). To clarify the
121 spectral changes per each condition over time, averaged spectra of the 48 hr time point were
122 subtracted from those at 24 hr (Fig 1d and Fig S1). The greatest separations of spectra between
123 active SARS-CoV-2 and UV-inactivated SARS-CoV-2 occurred at 2977 cm^{-1} , 2920 cm^{-1} , 1668-
124 1665 cm^{-1} , 1595 cm^{-1} , 1418 cm^{-1} , 1298 cm^{-1} , 1122 cm^{-1} , 1021 cm^{-1} , 854 cm^{-1} (Fig 1d and S1
125 Fig). Active SARS-CoV-2 infection demonstrated separation from media control at 1600 cm^{-1} ,
126 1304 cm^{-1} , 1124 cm^{-1} , 1042 cm^{-1} , and 1023 cm^{-1} (Fig 1c,d). These features notably included
127 increased absorbance at 1124 cm^{-1} , a region considered to reflect symmetric stretching of
128 phosphodiester linkages of RNA ($(8, 9)(\nu_3\text{PO}_2^-)$).

129



130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144

Fig 1. ATR-FTIR spectral changes of culture supernatants, *in vitro* SARS-CoV-2 infection model. **a**) Vero-E6 cells (6×10^5) were treated with media alone (SARS-CoV-2^{NEG}), UV-inactivated (SARS-CoV-2^{UV-I}), or SARS-CoV-2 (SARS-CoV-2^{POS}) for 2 hours, after which cells were washed in PBS and media replaced. Aliquots of conditioned media were collected at 24 hr and 48 hr post-infection for qPCR and ATR-FTIR. Verification of viral load was accomplished via RT-qPCR ($p=0.0035$). **b**) Overlapping spectra of technical replicates for 24 hr and 48 hr timepoints. Colored bands indicate chemical components of interest: Aliphatic (yellow), Amide I/II/III (cyan), severity (red), Saccharide (green), phosphodiester asymmetric stretching ($v_{as}PO_2^-$) and symmetric stretching ($v_sPO_2^-$) (purple stripes). **c**) Significant features of SARS-CoV-2 infection compared to the two controls at 48 hr, using FDR-LogWorth analysis; dotted line represents FDR-LogWorth 2 ($p < 0.01$). **d**) Subtraction of supernatant spectra per each treatment from 24 hr to 48 hr.

145

146 **ATR-FTIR spectra of oral lavage from respiratory SARS-CoV-2-** 147 **infected mouse model**

148

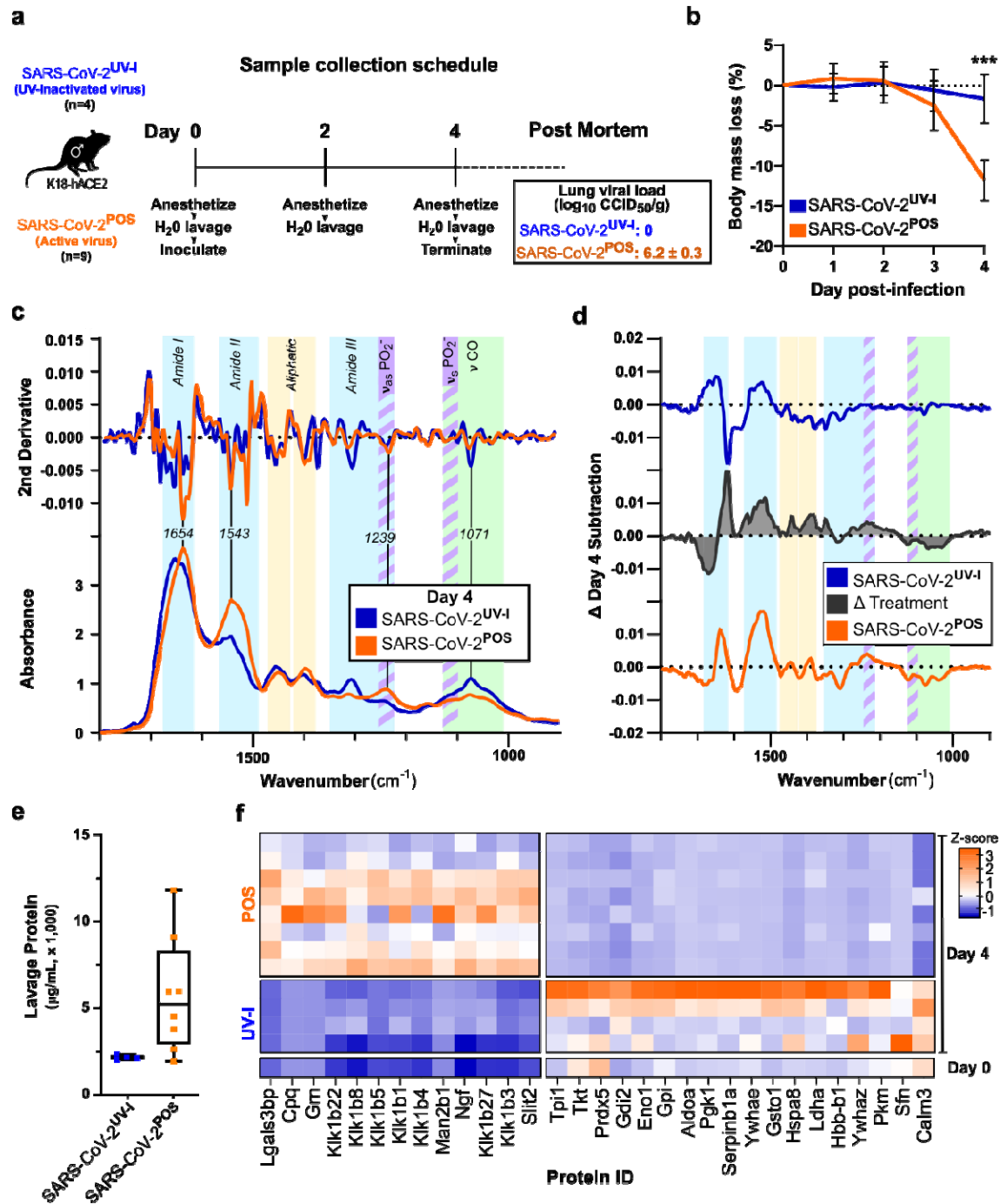
149 Next, transgenic ACE2 (K18-hACE2) mice were used to evaluate the ATR-FTIR spectra of
150 oral secretome following respiratory SARS-CoV-2 infection, again comparing to UV-inactivated
151 SARS-CoV-2. Past studies determined K18-hACE2 mice develop lung infection and a
152 respiratory disease resembling severe COVID-19 (10, 11), and have been widely used to
153 evaluate interventions against SARS-CoV-2 infection and disease (12-17).

154

155 Oral lavage was collected from anaesthetized mice prior to infection (day 0), and then on days 2
156 and 4 post inoculation (Fig 2a). Body mass of active SARS-CoV-2 infected mice started
157 declining on day 3, reaching minus 10-15% on day 4 (Fig 2b). Comparison of the average oral
158 lavage ATR-FTIR spectra showed more significant changes on day 4 compared to day 2 (S2
159 Fig). On day 2, aliphatic, fatty acids 1738 cm^{-1} , and ($\nu_{\text{as}}\text{PO}_2^-$) bands showed increased
160 absorbance with mild drop in saccharides between groups (S2a, S2b Fig), however, on day 4,
161 saccharides further dropped with pronounced increases in the amide bands (Fig 2c, 2d and S2a,
162 S2b Fig). Subtraction of spectra on day 4 from day 0 (baseline) allowed visualization of
163 respective time-course changes for the UV-inactivated SARS-CoV-2 infection group (SARS-
164 CoV-2^{UV-I}) (Fig 2d, top) and the active SARS-CoV-2 infection (SARS-CoV-2^{POS}) (Fig 2d,
165 bottom); additional subtractive analysis was performed between these two groups to observe the
166 unique changes attributed to active SARS-CoV-2 infection ($\Delta\text{Treatment}$, Fig 2d, middle). A
167 broad, rising Amide II peak at 1542 cm^{-1} was the most prominent feature resulting from SARS-
168 CoV-2 infection. Although all mice displayed an increase in Amide peaks between days 2 and 4,
169 those with active infection presented a significant Amide II peak and Amide I shift (Fig 2c, 2d
170 and S3 Fig, $p = 0.00001$).

171

172 To further elucidate the pathophysiology, we conducted untargeted proteomics on the lavage.
173 The protein concentrations of the SARS-CoV-2^{POS} lavage was elevated in comparison to the
174 SARS-CoV-2^{UV-I} group, indicating strong secretory response to SARS-CoV-2 infection (Fig 2e).
175 Proteomic analysis on equal amount of lavage protein revealed upregulation of several
176 kallikreins, and proteins involved in immune modulation such as lectin galactoside-binding
177 soluble 3 binding protein (Lgals3bp) and progranulin (Grn) (Fig 2f). Furthermore, a number of
178 proteins were comparatively down regulated, notably including Calmodulin-3 (Calm3).
179



180
181

182 **Fig 2. ATR-FTIR spectra and proteomic changes of oral lavage of SARS-CoV-2 mouse**
 183 **model.** a) Male K18-hACE2 mice were inoculated with intrapulmonary UV-inactivated (n=5) or
 184 active SARS-CoV-2 (n=9), and oral lavage was sampled on days 2 and 4. Viral load of mouse
 185 lung tissue was assessed via cell culture infectious dose 50% assay (CCID50) post-mortem,
 186 showing no active virus in the inactivated virus group. b) Body weight measurements were
 187 recorded daily. Error bars show standard error. *** on day 4, p= 0.0004. c) Day 4 oral lavage
 188 ATR-FTIR spectra of the Amide I/II and fingerprint regions with respective 2nd derivative
 189 (above). Colored bands indicate chemical components of interest: Amide (protein) bands I, II, III

190 (cyan), PO_2^- asymmetric (v_{as}) and symmetric (v_{s}) stretching (purple stripes), saccharides (green),
191 with identification of key peaks by wavenumber. **d**) Subtraction of Day 4 spectra from Day 0,
192 showing a time-course alteration for SARS-CoV-2^{UV-I} (blue) and SARS-CoV-2^{POS} (orange), as
193 well as the difference between the groups, $\Delta\text{Treatment}$ (black). Complete spectra (4000-600 cm^{-1})
194 as well as Day 2 data are available in Fig. S2. **e**) Protein concentration of Day 4 oral lavage
195 plotted per group. **f**) Proteomics was conducted on equal amounts of Day 4 oral lavage, and a
196 pooled sample of Day 0 oral lavage (3 samples) for comparison. Heatmap shows z-scores of
197 differential proteins ($p < 0.1$ adjusted) between SARS-CoV-2^{UV-I} and SARS-CoV-2^{POS} groups.
198

199

200

201 **ATR-FTIR spectra of human saliva distinguishes SARS-CoV-2** 202 **infection status**

203

204 To investigate the application of ATR-FTIR for COVID-19 screening in human saliva samples,
205 we collected saliva from 104 participants, healthy controls (COVID.NEG, $n=44$) and COVID-19
206 cases (COVID.POS $n=60$) (Fig 3a, S1 Table).

207

208 The acquired spectra ($n=3-6$ technical replicates per biological sample) were baseline corrected
209 and normalized, then the technical variance in the dataset was assessed using pairwise Euclidean
210 distancing (S4 Fig). The variance within replicates of a participant was significantly lower
211 compared to variance between participants (0.1925 ± 0.1941 vs 0.6089 ± 0.544 , $p < 0.0001$, S4
212 Fig), indicating acceptable technical variability relative to the observed biological variability.
213 The average spectra for COVID.POS and COVID.NEG groups showed visible differences in
214 aliphatic, amide I, II, III regions (Fig 3b). Discriminant analysis with canonical plot revealed
215 separation between COVID.NEG and COVID.POS groups on Canonical 1 (X-axis) but
216 interestingly showed separation within the COVID.POS groups along Canonical 2 (Y-axis)
217 correlating to PCR results on the day of saliva sampling; termed COVID.POS^{FU.POS} and
218 COVID.POS^{FU.NEG} (Fig 3c).

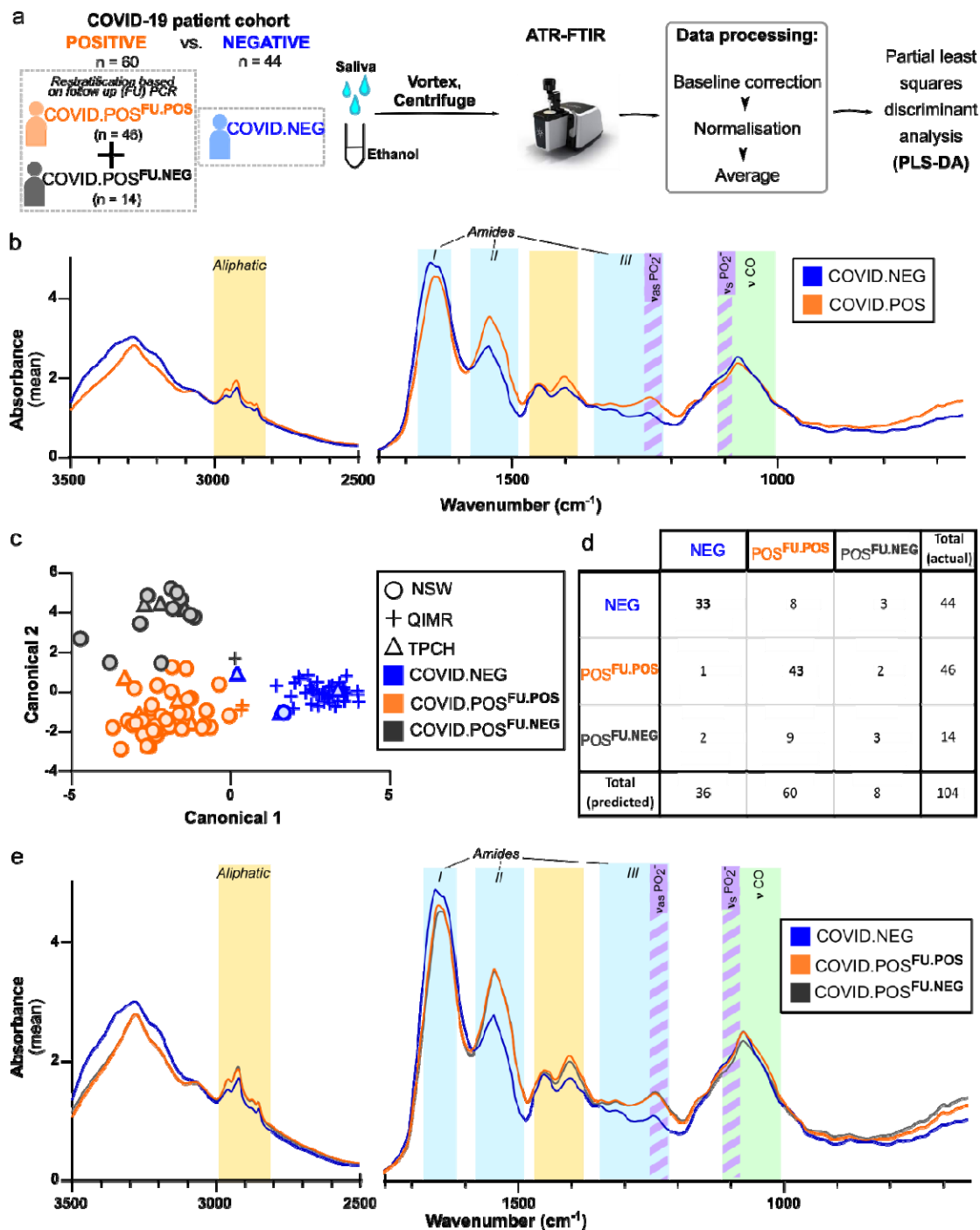
219

220 A Partial Least Squares Discriminant Analysis (PLS-DA) model using Non-linear Iterative
221 Partial Least Squares (NIPALS) algorithm was developed to predict the three clinical groups.
222 Seven factors explained 75.25% of the variation in the spectra. Based on Leave-One-Out Cross-
223 validation (LOO-CV) the PLS-DA correctly predicted 75% of COVID.NEG (specificity), and
224 93.48% of COVID.POS^{FU.POS} (sensitivity). In the COVID.NEG group, while the sensitivity was
225 in line with the previous cohort studies, the specificity was slightly lower in this cohort.
226 Therefore, we investigated if recency of COVID-19 vaccination may contribute to incorrect
227 prediction by saliva FTIR. Of the 44 COVID.NEG participants, 29 participants had received one
228 or both vaccine doses in the 8 -120 days prior to saliva collection. All 4 incorrectly predicted
229 samples had vaccination doses 22-67 days prior to saliva collection suggesting that vaccination
230 did not influence ATR-FTIR results.

231

232 Of the 14 COVID.POS^{FU.NEG} participants, only 2 (14.3%) were predicted correctly with the
233 remaining 10 (71.4%) and 2 (14.3%) predicted as COVID.NEG or COVID.POS (Fig 3d). This is
234 not surprising as viral clearance dynamics are highly variable in individual post infection onset.

235 Although visualisation of the average spectra revealed little separation between these subgroups,
 236 a region we previously reported to correlate with COVID-19 disease severity (7) also was able to
 237 distinguish between COVID.POS^{FU.POS} and COVID.POS^{FU.NEG} groups (Fig 3e and S5b Fig).
 238
 239



240
 241 **Figure 3. ATR-FTIR spectral data of human cohort.** a) Workflow. Sublingual saliva were
 242 collected from human subjects with known COVID.POS and COVID.NEG status. Follow-up
 243 SARS-CoV-2 PCR was conducted for the COVID.POS group on the saliva or swab collected on
 244 day of saliva collection (COVID.POS^{FU.POS} or COVID.POS^{FU.NEG}). Clarified saliva adjusted to a

245 final concentration of 75% ethanol was used for ATR-FTIR on an Agilent Cary 630, with
246 samples dried (~30 sec) on the crystal. Data for each technical replicate were baseline corrected
247 then normalized to an AUC of 1. **b)** Average spectra (3500-650 cm^{-1}) of COVID.NEG and
248 COVID.POS groups. Colored bands indicate chemical components of interest: Aliphatic
249 (yellow), Amide I/II/III (cyan), Saccharide (green), phosphodiester (purple stripes). **c)** Canonical
250 plot with symbols indicating the location of sampling, NSW, New South Wales Health
251 Pathology; TPCH, The Prince Charles Hospital; QIMRB, QIMR Berghofer Medical Research
252 Institute. **d)** Contingency table for leave-one-out cross-validation of the partial least squares
253 discriminant analysis (PLS-DA) model. Columns represent actual designation while rows
254 represent predicted categorization. **e)** Average spectra (3500-650 cm^{-1}) of the three clinical
255 groups.

256
257

258 **Delineation of spectral signature for COVID.POS^{FU.POS} saliva**

259

260 To determine the significant spectral regions between groups, and the regions selected in the
261 PLS-DA model, we conducted Logworth FDR analysis (Fig 4a, 4b) and Variable Importance
262 Plot analysis (Fig 4c), respectively. Compared to COVID.NEG, COVID.POS^{FU.POS} saliva
263 demonstrated significant differences in all amide bands identified: increased absorbance in
264 Amide A and B (3500-3300 cm^{-1} , 3100 cm^{-1} , respectively), a narrowing of Amide I from a major
265 right shift (1710-1650 cm^{-1}) and minor left shift (1624-1596 cm^{-1}), a pronounced increase and
266 right shift of Amide II (1570-1470 cm^{-1}), and increase of Amide III (1320 cm^{-1}). Significantly
267 increased absorbances were also observed in aliphatic bands: 2956 cm^{-1} ($\nu_{\text{as}} \text{CH}_3$), 2870 cm^{-1} ($\nu_{\text{s}} \text{CH}_3$),
268 1464 cm^{-1} ($\delta_{\text{as}} \text{CH}_3$, asymmetric bending), 1420 cm^{-1} (δCH_2 and deformations), and 890
269 cm^{-1} (δCH_2). Bordering Amide III, the two most significant combined points, 1252 cm^{-1} and
270 1228 cm^{-1} , represent asymmetric phosphate stretching ($\nu_{\text{as}} \text{PO}_2^-$) among a diversity of
271 macromolecules such as phospholipids, phosphorylated proteins, and RNA (18).

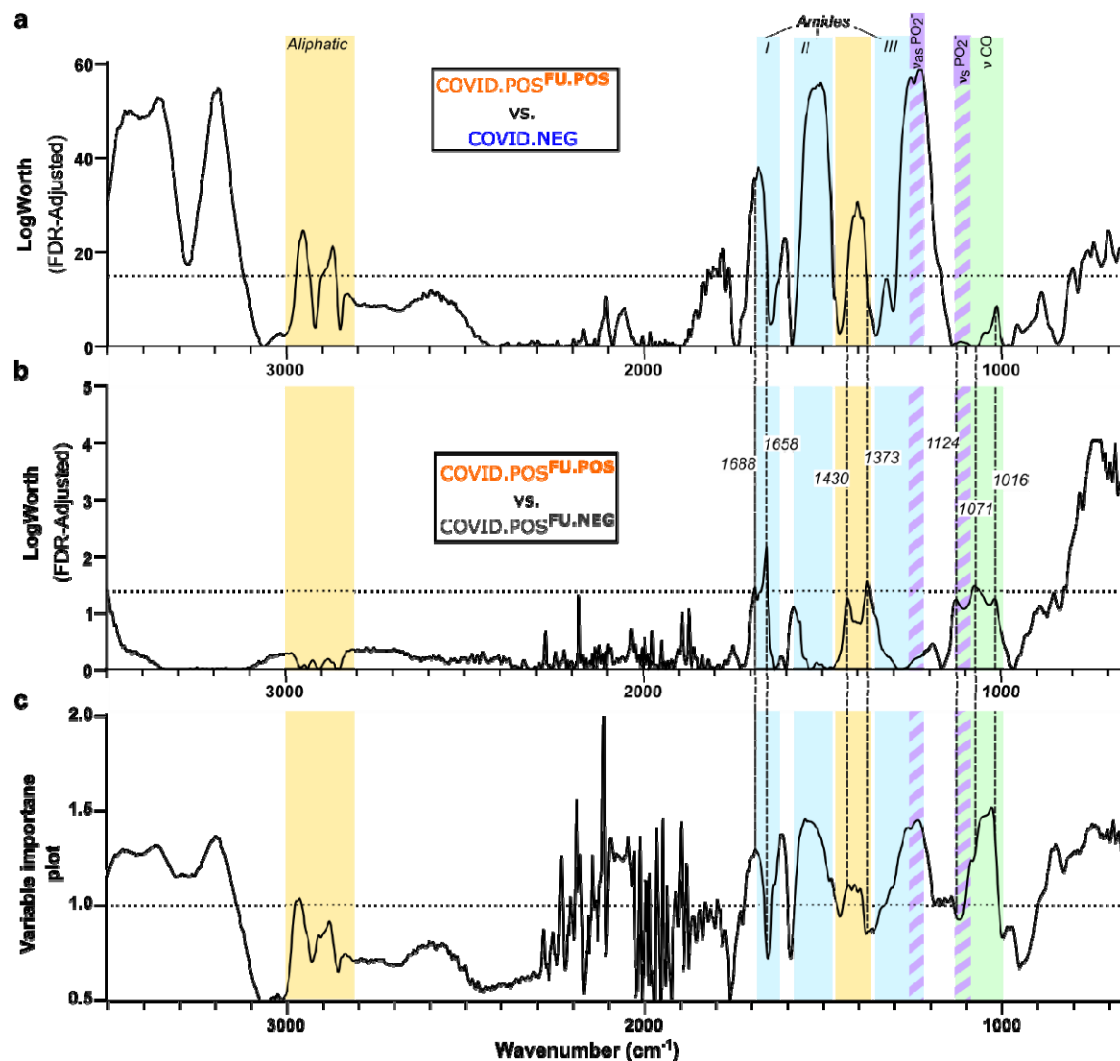
272

273 In addition, 7 points of varying significance ($p < 0.06-0.01$) correlated with the significant peaks
274 from COVID.POS^{FU.POS} vs COVID.NEG comparison: right shift of Amide I (1688/1658 cm^{-1}),
275 decreased aliphatic/RNA (1430 cm^{-1}), decreased δCH_3 - bending (1373 cm^{-1}), decreased $\nu_{\text{s}} \text{PO}_2^-$ -
276 RNA (1124 cm^{-1}), $\nu_{\text{s}} \text{PO}_2^-$, symmetric and C-O v ribose (1071 cm^{-1}), and decreased v C₄-OH -
277 glucose (1016 cm^{-1}). The right shifting Amide I peak in COVID.POS^{FU.POS} compared to both
278 COVID.NEG and COVID.POS^{FU.NEG} is in agreement with residual misfolded amyloid protein
279 fibrils and elevated IgA in COVID-19 patient saliva (S5 Fig)(4, 19, 20).

280

281 Comparison of the Logworth FDR analysis with the VIP analysis of the PLS-DA model (Fig 4c)
282 revealed that most of the predictive peaks overlap with the significant peaks from the
283 COVID.POS^{FU.POS} vs COVID.NEG analysis (Fig 4b).

284



285
 286 **Fig 4. Delineating FTIR spectral signature for COVID-19 saliva screening.** Saliva FTIR
 287 spectra from the cohort analysis in Fig. 3 was subjected to comparative FDR LogWorth analysis
 288 for a) COVID.POS^{FU.POS} and COVID.NEG saliva samples, and b) COVID.POS^{FU.POS} and
 289 COVID.POS^{FU.NEG} samples. Dotted lines visualizes the cut-off chosen at a level above noise for
 290 each comparison at LogWorth of 15 ($p = 1 \times 10^{-15}$) and 1.3 ($p = 0.05$), respectively. c) Variable
 291 importance plot for the final PLS-DA model shown in Fig. 3d, displaying spectral regions'
 292 contributions to the model. Dotted line indicates VIP of 1.0. Colored bands indicate chemical
 293 components of interest: Aliphatic (yellow), Amide I/II/III (cyan), Saccharide (green),
 294 phosphodiester (purple stripes).

295

296 **Delineation of COVID-19 spectral signature across diverse models**

297

298 Finally, we sought to establish the most characteristic COVID^{POS} spectral signature across
299 multiple models and studies by comparing the significant results from all three study arms, as
300 well as the three recent publications (Table 1), bearing in mind the differing nature of the models
301 and methodologies across studies. This analysis revealed several consistent spectral changes due
302 to SARS-CoV-2 infection across multiple models/studies (Table 1). Most strikingly, a change in
303 the structure of proteins was indicated by Amide II increase in all studies, indicative of β -sheet
304 structures. In all human cohorts (but not *in vitro* or mouse models), Amide III, aliphatic,
305 phosphodiester asymmetric stretching ($\nu_{as}PO_2^-$) and saccharide bands were also increased. In
306 contrast, saccharide bands were decreased in *in vitro* and mouse models, with VIP values 1.2-
307 1.43 over the range 1067-1006 cm^{-1} (Fig. 3b/e, Fig. S5b). This range of wavenumbers was
308 previously recognized for having significance in predicting severe COVID-19 outcomes when
309 evaluating blood plasma, including an elevated AUC at 1592-1588 cm^{-1} (7). Taken together,
310 these results suggest a change in the physiological response between end-point/severe COVID-
311 19 (*in vitro* cell, mouse model and severity study) and COVID-19 patients who are tolerating the
312 disease well.

313

314 **Table 1. Spectral features for saliva/secretion in COVID^{POS} cohorts/models.**

Band Designation	In Vitro ^a	Mouse ^a	Human ^a	Chemical components (19, 21-26)	Barauna (3)	Wood(5)	Martinez □ Cuazitl (4)
Amide A		3246	3358 3518-3280 ^b	N-H, O-H stretching		x	
Amide B		3067	3190 3248-3110 ^b	Amide II overtone, aromatic amino acids		x	
Aliphatic	2973	2931-2880 2858 2837	2954 2870 2968-2944 ^b	-CH ₃ /-CH ₂ ; C-H symmetric (v _s) & asymmetric (v _{as}) stretching		x	
Fatty Acids	1705	1702	1722-1704 1714-1690 ^b	-COOH, C=O v; and ketones		x	O
*Amide I	1690		1680 1680 ^b	Protein β-sheets; C=O guanine			
		1638		Protein β-sheets			
	1600		1625-1594 1632-1585 ^b	Protein aggregates; amyloid fibrils			x
Amide II	1524	1578-31	1572-1470	N-H Mostly β-sheet		x	x
Aliphatic		1468	1464	-CH ₂ δ, bending vibrations		x	
Fingerprint	1439	1431	1416 1420^b	-CH ₂ δ, symmetric stretching band of carboxyl group, CH ₂ ω, wagging; RNA	x	x	x
			1402 1400^b	C-H deformation; CH ₂ ω; C-N stretching; In-plane C2'OH in RNA		x	x
	1373	1370	1375 1388-1376^b	-CH ₃ δ, C-H v; methyl bending/stretching		x	x
Amide III	1302	1302	1319	Amino acid side-chains; terminal oxygen (PO ₃)		x	x
		1378-1354 1335-1280	1340-1285 1330-1177^b	-CH ₂ ω; -CH ₃ δ, amyloid contribution		x	x
			1250 1250^b	PO ₂ v _{as} ; C-N v		x	x
			1243-1218 1226^b	PO ₂ v _{as} ; amyloid fibrils		x	x
RNA	1124		1129 ^b	PO ₂ v _s , phosphodiester stretching		x	
Saccharide		1094		-C-O-C, ether linkages; -O-Ca ²⁺ c	x		
	1072	1064	1077 ^b	PO ₂ - v _s , symmetric and C-O v	x	x	x
	1050			-C-O v, C-OH group; C-C v (sugars)		x	x
	1023		1034-1003 1095-997^b	C-O v; P-O v; C-OH δ			x
	1008	1012	1012	C ₄ -OH, Glucose		x	x
	988	988-974		PO ₂ ⁻ v _s ; -C-O-, ribose		x	
			940	P-O v, phosphorylation; -C-C- v		x	
		887-866	889	P-O-P v _{as} ; -C-C- v; aromatics		O	x
		830	836^b	P-O-C v; =C-H δ; aromatics		O	x

315 ^a Significance by FDR LogWorth analysis

316 ^b Also significant by Variable Importance Plot analysis

317 ^c Long Fasting collection; analysis of truncated spectra

318 * Some significance found in Amide I region were due to shift, not proportion change of specific chemical species alone

319 **Bold type** indicates where SARS-CoV-2 spectra was ↑ higher

320 X = Separation of spectra from control

321 O = Noticeable change in spectra, value not legible/not reported

Discussion

322
323 This study provides controlled validation and the underpinning pathobiology data required to
324 support the translation of ATR-FTIR for COVID-19 saliva screening. The consistency of SARS-
325 CoV-2^{POS} FTIR signature across *in vitro*, mouse and human arms of this study, and three existing
326 independent reports (3-5) provides confidence that this robust technology is ready for clinical
327 development and deployment. Compared to the single output of current PCR and antigen tests,
328 the broad biochemical information from the saliva FTIR spectra provides both diagnostic and
329 prognostic information supporting the use of saliva as a non-invasive, self-collectable bio-sample
330 reflective of the physiological response (27-29).

331
332 The COVID-19 saliva FTIR signature shares many biochemical features with amyloid deposits
333 (aliphatic, amide, and phosphodiester ν_{as}) and lipofuscin (19, 30). Enhanced amyloid formation
334 in SARS-CoV-2 has been a recent area of focused research (31-33). The consistent COVID-19-
335 associated amide absorbance shifting from α -helix ($\sim 1652\text{ cm}^{-1}$) composition to β -sheet (~ 1636
336 cm^{-1}) is likely related to SARS-CoV-2 induction of protein aggregates through its spike protein
337 (34-36). Strikingly, our proteomics data revealed extensive upregulation of kallikrein proteins in
338 oral lavage of SARS-CoV-2^{POS} mice. Savitt et al. recently reported direct interaction and
339 activation of the kallikrein/kinin system (KKS) by recombinant SARS-CoV-2 proteins S, M, N,
340 and E (37). High molecular weight kininogen (HK) and plasma prekallikrein (PK) bring about
341 the sequelae of bradykinin, and complexing of HK/PK with Blood Coagulation Factor XII (FXII)
342 initiate the intrinsic clotting cascade with the aid of misfolded proteins and polyphosphate (38).
343 Polyphosphate may serve as a natural defense blocking the receptor binding domain for SARS-
344 CoV-2 (39). Whether derived from platelets or commensal bacteria, extensive utilization of
345 polyphosphate in the KKS/FXII pathogenesis of SARS-CoV-2 offers another substantial
346 explanation for the phosphate and Amide III profiles among these studies (40-42). Another
347 upregulated protein contributing to the Amide I/II bands, Lgals3bp, is potentially upregulated as
348 a compensatory defensive mechanism to the prolonged innate immune response by day 4 (43).
349 Further investigations should be carried out to establish protein disaggregated in human saliva
350 samples and the involvement of KKS-FXII-polyphosphate.

351
352 While the cell culture and mouse model data were generally consistent with our human cohort
353 data, the reduced saccharide band in cell and mouse SARS-CoV-2 treated samples contrasted
354 with the increases observed in human COVID.POS samples. This discrepancy was most
355 probably due to different stages in the evolution of infection, as the mouse and cell experiments
356 represent models of severe illness, while the human cohorts consisted of individuals where
357 SARS-CoV-2 infection was generally well tolerated and many subjects were entering the
358 recovery phase. Studies have provided mechanistic evidence for metabolic dysregulation in
359 COVID-19, notably through insulin resistance involving adiponectin/leptin and proinflammatory
360 alterations, (44, 45) which fits with our observations, along with the decreased food intake
361 observed during the acute phase of human disease, i.e. as evidenced by the infected mice in our
362 studies (45).

363
364 A novel finding from our patient cohort is the separation of the COVID.POS patients into a sub-
365 group with low/undetectable viral load on the day of saliva collection based on their saliva FTIR
366 spectra. Saliva FTIR spectra of this COVID.POS^{FU.NEG} group displayed reduced saccharide
367 (ribose) bands (1038 cm^{-1} , 1074 cm^{-1}) compared with COVID.POS^{FU.POS}. These bands coincided

368 with the ATR-FTIR bands for extracted SARS-CoV-2 RNA(46), in agreement with the PCR
369 results; however, this difference in saccharide absorbance may also indicate recovery from the
370 previously described, hyperglycemic state This group also showed reduced signal in the finger
371 print region, proposed by Martinez-Cuazitl et al.(4) to represent immunoglobulins IgG, IgM and
372 IgA. As these COVID.POS^{FU.NEG} patients are likely to have continued immunoglobulin
373 expression/secretion (47), our results suggest that the Amide I/II and fingerprint regions more
374 likely correlates with clearance of protein aggregates (β -sheet) and aliphatic amino acids as seen
375 by significant decreases at 1688 cm^{-1} and 1373 cm^{-1} , respectively (Fig. 4b).

376
377 While the predictive model using saliva ATR-FTIR spectra showed high sensitivity in predicting
378 COVID.NEG and COVID.POS^{FU.POS} cases, it was unable to differentiate COVID.POS^{FU.NEG}
379 cases accurately. Nevertheless, when thinking how one may utilize this assay, as a possible
380 point-of-care screening test, the high sensitivity makes it able to rule out infected individuals
381 who are likely to transmit SARS-CoV-2, which is extremely important from an epidemiological
382 and social perspective, particularly in the context of family groups when a child or one parent is
383 infected, but other family members are not. This feature differentiates the FTIR saliva test from
384 other point-of-care test with high specificity testing characteristics, making FTIR preferable
385 especially in settings of large gatherings thereby potentially circumventing a super spreader
386 event(s). One additional application for this assay is the ability to monitor physiological
387 responses to COVID-19 which in turn may inform patient infection responses and predictions on
388 prognosis and need for therapeutic interventions.

389
390 In contrast to previous fasting requirements of >8-hr prior to saliva collection (4), we took a
391 pragmatic approach of only 20-30 minutes abstinence from food prior to testing. Our results
392 support this time interval between sample collection and testing making a point-of-care rapid
393 testing application more feasible. It is unlikely that this time interval can be shortened further as
394 saliva is likely to be “contaminated” with food particles interfering with FTIR signals. We did
395 notice, however, excessive precipitation while mixing saliva with ethanol, secondary to the
396 initial high postprandial cephalic secretion. Adding a low-speed centrifugation of raw saliva
397 prior to inactivation with ethanol circumvented this problem. Our simple, inactivation procedure
398 with ethanol removes any possible biosafety concerns. All these features make
399 future development for point-of-care application feasible. However, saliva collection and
400 processing methods would require additional refinement (e.g. use of a capillary action sampling
401 cartridge).

402
403 In conclusion, ATR-FTIR technology with saliva self-collection provides a simple, rapid and
404 biosafe sample processing, which has high potential as a non-invasive, low-resource method for
405 COVID-19 screening. The simplicity of the method means that only basic skills are required to
406 conduct the test, which would satisfy the global need for rapid COVID-19 screening at diverse
407 locations such as airports and public venues. Further evaluation may also establish utility for
408 COVID-19 prognosis. As the method requires only generic laboratory equipment, ethanol, an
409 ATR-FTIR with implemented predictive algorithm, and a power source, it offers promise as a
410 global tool in the management of COVID-19 pandemic.

411

412

413 **Materials and methods**

414 **SARS-CoV-2 virus**

415 The SARS-CoV-2 isolate (hCoV-19/Australia/QLD02/2020) was kindly provided by Dr Alyssa
416 Pyke (Queensland Health Forensic & Scientific Services, Queensland Department of Health,
417 Brisbane, Australia). Virus stocks were prepared in Vero E6 cells as described (48) with all
418 infectious SARS-CoV-2 work conducted in a dedicated suite in a biosafety level-3 (PC3) facility
419 at the QIMR Berghofer MRI (Australian Department of Agriculture, Water and the Environment
420 certification Q2326 and Office of the Gene Technology Regulator certification 3445). An aliquot
421 of the viral stock was extracted and sequenced using illumina technology and uploaded to
422 GISAID (<https://www.gisaid.org/>) under Accession ID (EPL_ISL_407896).

423 ***In vitro* cell model**

424 Vero E6 (C1008, ECACC, Wiltshire, England; obtained via Sigma Aldrich) were maintained in
425 RPMI 1640 (Thermo Fisher Scientific), supplemented with endotoxin free 10% heat-inactivated
426 fetal bovine serum (FBS; Sigma-Aldrich), at 37°C and 5% CO₂. Cells were checked for
427 mycoplasma using MycoAlert Mycoplasma Detection Kit (Lonza, Basel, Switzerland). FBS was
428 checked for endotoxin contamination before purchase as described (49).

429
430 Vero E6 cells (6x10⁵) were plated onto 6-well plates in 2ml RPMI + 10% FBS without phenol
431 red (to limit background spectra). Following 24h of growth, media was removed and cells were
432 subjected to control, mock-infection or infection regimes, each in triplicate. Control (untreated)
433 cells were rinsed 2xPBS and placed in 3ml phenol-red RPMI + 2% FBS. Mock-infected cells
434 were incubated with 500 µL of UV-inactivated SARS-CoV-2 stock for 30 min, while infected
435 cells were incubated with 500 µL SARS-CoV-2 viral stock MOI 0.01 for 30 min. Mock-infected
436 and infected cells were then rinsed 2xPBS and placed in 3ml phenol-red RPMI + 2% FBS.
437 Aliquots of conditioned media were collected at 24h and 48 h post-infection. At each time point,
438 100 µL of conditioned media was mixed with 300 µL ice cold 100% ethanol for FTIR, while 200
439 µL conditioned media was mixed with 600 µL Trizol-LS for PCR. At the 48h time point,
440 remaining supernatant was discarded and cells were harvested in 400 µL Trizol-LS for PCR.

441 **Nucleic acid extraction and RT-qPCR**

442 RNA was purified from tissue culture supernatants and saliva (TPCH and QIMRB cohorts) using
443 Direct-zol RNA microprep kits (Zymo Research) and cDNA was generated using iScript™
444 Reverse Transcription Supermix (BioRad). For qPCR, SsoAdvanced™ Universal SYBR® Green
445 Supermix (Bio-Rad) was used with two previously published primer sets targeting different
446 regions of SARS-CoV-2: 1) Forward (5'-CAATGCTGCAATCGTGCTAC-3') and reverse (5'-
447 GTTGCGACTACGTGATGAGG-3') primers targeting the N-gene; 2) Forward (5'-
448 ACCTTCCCAGGTAACAAACCA-3') and reverse (5'-TTACCTTTCGGTCACACCCG-3')
449 primers targeting the 5'UTR. Cycling was carried out in a CFX384 Touch™ Real-Time PCR
450 Detection System (Bio-Rad) under the following conditions: 95°C 30s; 95°C 10s, 60°C 30s
451 (40x); melt curve 65°C-95°C. Viral copy number in experimental samples was estimated relative
452 to a reference cDNA standard, using primer set 1. The reference cDNA was generated from a
453 pool of SARS-CoV-2 infected VERO-E6 cell supernatant RNA and the viral copy number of
454 reference cDNA was estimated relative to a plasmid containing the 5'UTR of SARS-CoV-2 (gift
455 from Dongsheng Li, QIMR Berghofer MRI), using primer set 2. Plasmid copy number was

456 determined using the URI Genomics and Sequencing Centre online calculator
457 (<http://cels.uri.edu/gsc/cndna.html>). Saliva RNA quality was confirmed by amplification of
458 housekeeping gene, β 2-microglobulin, using forward (5'-ACTCTCTCTCTTTCTGGCCTGG-3')
459 and reverse (5'-CATTCTCTGCTGGATGACGTG-3') primers.

460 **Mouse model**

461 All mouse work was conducted in accordance with the “Australian code for the care and use of
462 animals for scientific purposes” as defined by the National Health and Medical Research Council
463 of Australia. Mouse work was approved by the QIMR Berghofer Medical Research Institute
464 animal ethics committee (P3600, A2003-607). K18-hACE2^{+/-} mice were purchased from
465 Jackson laboratories and were maintained in-house as heterozygotes by backcrossing to
466 C57BL6/J mice (17, 48). Mice were typed as described (48) using hACE2 Primers: Forward: 5'-
467 CTT GGT GAT ATG TGG GGT AGA -3'; Reverse: 5'-CGC TTC ATC TCC CAC CAC TT -3'
468 (recommended by NIOBIOHN, Osaka, Japan).

469
470 Prior to oral lavage, 4-5 months old K18-hACE2^{+/-} mice were placed in static micro isolator
471 cages (Techniplast Static Micro-isolator # 1264) with gridfloor accessory, allowing feces to pass
472 through for 1 hour without food or water, to avoid fecal contamination in oral cavity. Oral
473 lavage was conducted with the mice under light anesthesia: 3% isoflurane (Piramal Enterprises
474 Ltd., Andhra Pradesh, India) delivered using The Stinger, Rodent Anesthesia System (Advanced
475 Anaesthesia Specialists/Darvall, Gladesville, NSW, Australia). With the mouse lying on its back,
476 25 μ L of milliQ water was placed into the side of the mouth just behind the teeth of the lower
477 mandible. The water was pipetted up and down 4 times to wash the mouth without injury or
478 abrasion of gums or lips. The lavage was recovered, and 15 μ L was added to 45 μ L of 100%
479 ethanol to obtain 75% v/v ethanol, then stored at -80°C.

480
481 Mice were infected intrapulmonary via the nasal route with 5×10^4 CCID₅₀ SARS-CoV-2 in 50
482 μ L medium while under light anesthesia. Saliva samples were collected before infection and on
483 the indicated days after infection. Mice body weights were measured each day. Mice were
484 euthanized using CO₂ on day 4 post infection and lung titers determined by CCID₅₀ assay of
485 serial dilution of supernatants from homogenized lung tissues (48).

486

487 **Mouse Lavage Proteomics**

488 Protein was extracted from ethanol-containing lavage samples by centrifugation at 16,000x g for
489 25 minutes, at 4°C. Supernatant was discarded, and protein pellet washed twice with 50 mM
490 triethylammonium bicarbonate (TEAB, Sigma-Aldrich) buffer. Proteins were resuspended in 50
491 mM TEAB and underwent protein estimation by BCA assay, per manufacturer's instruction
492 (Thermo Fisher Scientific). Outliers were defined as samples with a protein abundance 3
493 standard deviations above the mean for that condition, and were excluded from further
494 processing. The resulting day 0 lavage protein samples were pooled due to their low abundance
495 resulting in a single proteomics sample. Day 4 lavage samples contained adequate protein
496 abundance to continue as individual replicates, n = 4 for SARS-CoV-2^{UV-I} and n = 8 for SARS-
497 CoV-2^{POS} conditions. The BRAVO AssayMap platform (Agilent Technologies) was used for in-
498 solution digest and C18 desalting procedures. 1% sodium deoxycholate was added to each
499 sample for increased protein solubility. A standard automated trypsin digest method was
500 followed using 5 mM dithiothreitol and 20 mM 2-iodoacetamide. Samples were diluted 1:10

501 with 50 mM TEAB and porcine trypsin (Promega) added (final 1:30 trypsin to sample protein
502 ratio). Digests incubated overnight at 37°C and acidified using trifluoroacetic acid (TFA) to a
503 final concentration of 0.5%. Sodium deoxycholate was pelleted by centrifugation for 30 min, at
504 5,000x g, room temperature. Peptides were subsequently desalted by AssayMAP C18 cartridge,
505 following manufacture's instruction. Eluted peptide was dried and resuspended in 0.5% TFA.
506 Peptides were resolved on a Thermo U3000 nanoHPLC system and analysed on a Thermo Q
507 Exactive Plus Orbitrap mass spectrometer. The HPLC setup used a C18 trap column and a 50 cm
508 EasySpray C-18 analytical column (Thermo Fisher, catalogue: 160454, ES803A). Mobile phases
509 were A: 0.1% formic acid, and B: 80% acetonitrile with 0.1% formic acid. The loading pump ran
510 on 3% B at 10 µL per minute. 1 µg peptide were loaded in 3% B. The nano-capillary pump ran at
511 250 nL per minute, starting at 3% B. The multi-step gradient was 3% to 6% B over 1 minute, 6%
512 to 30% B over the following 60 minutes, 30% to 50% B over the following 12 minutes, then
513 50% to 95% B over 1 minute. After maintaining 95% B for 12 minutes, the system was re-
514 equilibrated to 3% B. The mass spectrometer ran an EasySpray source in positive ion DDA
515 mode, using settings typical for high complexity peptide analyses. Mass lock was set to "Best".
516 Full MS scans from 350 m/z to 1400 m/z were acquired at 70k resolution, with an AGC target of
517 3E6 and 100 ms maximum injection time. MS2 fragmentation was carried out on the Top 10
518 precursors, excluding 1+ and > 7+ charged precursors. The dynamic exclusion window was 30
519 seconds. Precursor isolation width was 1.4 m/z and NCE was 27. MS2 resolution was 17,500,
520 with an AGC target of 5E5 and a maximum injection time of 50 ms. Protein identification was
521 completed by MaxQuant using Swiss-Prot mouse proteome (version 2021_04) and default
522 parameters. Label-free quantitation intensities were analysed by the LFQ-Analyst pipeline to
523 determine differentially abundant proteins based on p-values < 0.1 (Benjamini Hochberg
524 adjusted p-value). Intensities were Z-score normalized and expressed as a heat map.

525 **Cohort study**

526 The project was approved by Human Research Ethics Committees of QIMR Berghofer Medical
527 Research Institute (QIMRB, P3675), New South Wales Health Pathology (NSWHP-RPAH
528 2020/ETH02630) and The Prince Charles Hospital (TPCH, ID 63003). All participants provided
529 written informed consent.

530
531 The cohort originated from three sites and included i) asymptomatic healthy volunteers (QIMRB
532 and TPCH) not in contact with COVID-19 cases for past 14 days (COVID.NEG); ii) COVID-19
533 positive (COVID.POS) hospitalised patients at TPCH, and iii) COVID.POS individuals in hotel
534 quarantine in New South Wales, Sydney, Australia. For cohorts ii) and iii) saliva sampling was
535 performed within 14 days after the initial PCR diagnostic test.

536
537 Fasting prior to saliva collection was considered but dropped to align with the real-world
538 screening scenario. Participants were requested to rinse mouth with water and refrain from eating
539 and drinking for 20 minutes prior to collecting 1.2 to 3 ml saliva as sublingual drool into a clean
540 receptacle.

541
542 Samples from cohorts i) and ii) were stored on ice and processed within 30 minutes. After brief
543 vortex, an aliquot of raw saliva was transferred to a 1.5 ml Eppendorf tube and centrifuged for 10
544 minutes at 500x g at 4 C to remove particulates. Clarified saliva was transferred to a cryotube
545 containing ethanol to obtain 75% v/v ethanol and incubated at room temperature for 30 minutes.

546 Inactivated saliva samples were stored at -80°C . For the TPCOVID.POS samples, another
547 tube was prepared to 75% v/v Trizol for RT-PCR using the protocol described above.

548
549 Samples from cohort iii) were initially transported to the laboratory at room temperature.
550 Aliquots of raw saliva were frozen at -80°C , subsequently thawed on ice, inactivated with 75%
551 v/v ethanol and shipped on dry-ice to QIMR Berghofer for FTIR analysis. A nasal pharyngeal
552 swab was collected on the same day as saliva, and was analysed by RT-PCR using TaqPath
553 COVID-19 Combo Kit (Thermo Fisher Scientific) according to manufacturer instructions.

554
555 For COVID.POS individuals, a subset tested PCR negative and were classified as
556 COVID.POS^{FU.NEG} (Table S1).

557

558 **ATR-FTIR spectra acquisition and processing**

559 Samples in 75% ethanol were thawed on ice and homogenised by high speed vortexing. An
560 aliquot of 2 μL was applied to the crystal of an ATR-FTIR instrument (Agilent Cary 630), and
561 allowed to air dry (~ 30 sec) before spectral acquisition occurred over the wavenumber range,
562 $4000\text{-}650\text{ cm}^{-1}$. Background was collected without sample, i.e. ambient room air at 21 \square between
563 each measurement following cleaning of the crystal with 80% ethanol. Settings included 64
564 scans (Sample/Background) with a resolution of 8 cm^{-1} . All spectra were baseline adjusted with
565 baseline estimated using regions $2031\text{-}1865\text{ cm}^{-1}$ and $3971\text{-}3799\text{ cm}^{-1}$. Spectra were then
566 normalised by adjusting to area under the curve (AUC) as 1.

567

568 **Statistical analysis**

569 Euclidean distance was calculated for each pairwise comparison of normalised spectra to
570 determine intra- and inter-sample variability. Each comparison was grouped into a “intra-
571 sample” (spectra from same biological replicate, 1,970 comparisons) or “inter-sample” (spectra
572 from different biological replicate, 89,253 comparisons) category and represented as a violin
573 plot.

574

575 Clustering of the samples was explored using discriminant analysis to create a canonical plot to
576 display clustering of clinical groups. LogWorth statistic was applied to identify spectral regions
577 that significantly deviate between two sample groups. The false discovery rate p-value cut-off for
578 each comparison was chosen in a data-dependent manner accounting for the differences from
579 baseline.

580

581 A predictive model was developed using Partial Least Squares Discriminant analysis (PLS-DA)
582 to predict clinical group based on the spectra. A six factors solution was chosen to account for at
583 least 70% of the variation in the spectrum. A variable importance plot (VIP) was generated to
584 indicate what areas of the spectrum most contributed to the predictive model. The fit of the
585 model was evaluated using Leave-One-Out Cross-validation (LOO-CV), generating a receiver
586 operating characteristic (ROC) curve and a confusion matrix giving the cross validated
587 sensitivity and specificity. The cutoff for predicting positive or negative were chosen to
588 maximize Youden’s Index (ie. the sum of sensitivity and specificity).

589 **Data availability**

590 Raw FTIR spectra and associated clinical data are available at DOI: 10.5281/zenodo.5703689,
591 <https://zenodo.org/record/5703689#.YZLxWGBBxaQ>
592 Proteomics data are available via ProteomeXchange with identifier PXD030012.
593

594 **Acknowledgements**

595 We thank all participants for donating their time and saliva samples to this research. From
596 QIMR Berghofer Medical Research Institute we thank Dr Itaru Anraku for managing the PC3
597 (BSL3) facility and animal house staff for mouse breeding and agistment. We thank Dr Alyssa
598 Pyke and Mr Fredrick Moore (Queensland Health, Brisbane) for providing the SARS-CoV-2
599 isolates. We thank Clive Berghofer and the Brazil Family Foundation (and many others) for
600 their generous philanthropic donations to support SARS-CoV-2 research at QIMR Berghofer.
601 We thank Peter Vardon, research nurse for coordinating the biological sample collection at The
602 Prince Charles Hospital. A.S. holds an Investigator grant from the National Health and Medical
603 Research Council (NHMRC) of Australia (APP1173880).

604 **Funding**

605 This project was supported by Agilent Technologies Applications & Core Technology -
606 University Research Grant (MMH), QIMR Berghofer Medical Research Institute COVID
607 Research (AS, DR) and The Prince Charles Hospital Research Foundation (DR).

608 **Conflict of interest statement**

609 AH, DB, FZ are employees of Agilent Technologies. All other authors declare that they have no
610 competing interests.

611 **Supporting information**

612 S1 Fig. Culture supernatant ATR-FTIR spectra and subtractive analysis of Vero cell SARS-CoV-
613 2 infection model.
614 S2 Fig. Full ATR-FTIR spectra and subtraction analysis of mouse oral lavage.
615 S3 Fig. Significant features of in vivo infection mouse model using LogWorth FDR analysis.
616 S4 Fig. Acceptable technical variance between replicates using pairwise Euclidean distancing
617 analysis of human samples.
618 S5 Fig. Human subject ATR-FTIR close-up of averaged spectra per three groups.
619 S1 Table. Summary of cohorts.

620 **AUTHOR CONTRIBUTIONS:**

621 Conceptualization: MMH, RSR, DB, AS
622 Data curation: GH, HR
623 Formal analysis: GH
624 Funding acquisition: DWR, AS, MMH
625 Investigation: STK, HR, RSR, KY, TTL
626 Methodology: HR, RSR, AH, FZ, DJR, AS, MMH
627 Project administration: RSR, DWR, MMH
628 Resources: SVH, RC, DWR
629 Supervision: AS, MMH
630 Visualization: HR

631 Writing – original draft: MMH, STK, HR
632 Writing – review & editing: GH, RSR, DJR, SVH, DWR, AH, DB

633 **ORCID:**

634 Gunter Hartel 0000-0002-5454-6450
635 Harley Robinson 0000-0002-9374-5532
636 Renee S. Richards 0000-0002-8911-8966
637 David Bradly 0000-0001-6480-2545
638 David W. Reid 0000-0001-9517-8212
639 Andreas Suhrbier 0000-0001-8986-9025
640 Michelle M Hill 0000-0003-1134-0951

641

642 **References**

- 643 1. Toropov N, Osborne E, Joshi LT, Davidson J, Morgan C, Page J, et al. SARS-CoV-2 Tests: Bridging the
644 Gap between Laboratory Sensors and Clinical Applications. *ACS Sens.* 2021;6(8):2815-37.
- 645 2. Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and
646 molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev.* 2021;3:CD013705.
- 647 3. Barauna VG, Singh MN, Barbosa LL, Marcarini WD, Vassallo PF, Mill JG, et al. Ultrarapid On-Site
648 Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High
649 Sensitivity and Specificity. *Anal Chem.* 2021;93(5):2950-8.
- 650 4. Martinez-Cuazitl A, Vazquez-Zapien GJ, Sanchez-Brito M, Limon-Pacheco JH, Guerrero-Ruiz M,
651 Garibay-Gonzalez F, et al. ATR-FTIR spectrum analysis of saliva samples from COVID-19 positive patients.
652 *Scientific Reports.* 2021;11(1):19980.
- 653 5. Wood BR, Kochan K, Bedolla DE, Salazar-Quiroz N, Grimley SL, Perez-Guaita D, et al. Infrared Based
654 Saliva Screening Test for COVID-19. *Angew Chem Int Ed Engl.* 2021;60(31):17102-7.
- 655 6. Naseer K, Ali S, Qazi J. ATR-FTIR spectroscopy as the future of diagnostics: a systematic review of the
656 approach using bio-fluids. *Appl Spectrosc Rev.* 2021;56(2):85-97.
- 657 7. Banerjee A, Gokhale A, Bankar R, Palanivel V, Salkar A, Robinson H, et al. Rapid Classification of
658 COVID-19 Severity by ATR-FTIR Spectroscopy of Plasma Samples. *Anal Chem.* 2021;93(30):10391-6.
- 659 8. Simonova D, Karamancheva I. Application of Fourier Transform Infrared Spectroscopy for Tumor
660 Diagnosis. *Biotechnol Biotec Eq.* 2013;27(6):4200-7.
- 661 9. Mateus T, Almeida I, Costa A, Viegas D, Magalhaes S, Martins F, et al. Fourier-Transform Infrared
662 Spectroscopy as a Discriminatory Tool for Myotonic Dystrophy Type 1 Metabolism: A Pilot Study. *Int J Env Res*
663 *Pub He.* 2021;18(7).
- 664 10. Yinda CK, Port JR, Bushmaker T, Offei Owusu I, Purushotham JN, Avanzato VA, et al. K18-hACE2 mice
665 develop respiratory disease resembling severe COVID-19. *PLoS Pathog.* 2021;17(1):e1009195.
- 666 11. Arce VM, Costoya JA. SARS-CoV-2 infection in K18-ACE2 transgenic mice replicates human pulmonary
667 disease in COVID-19. *Cell Mol Immunol.* 2021;18(3):513-4.
- 668 12. Rosenfeld R, Noy-Porat T, Mechaly A, Makdasi E, Levy Y, Alcalay R, et al. Post-exposure protection of
669 SARS-CoV-2 lethal infected K18-hACE2 transgenic mice by neutralizing human monoclonal antibody. *Nat*
670 *Commun.* 2021;12(1):944.
- 671 13. Zheng J, Wong LR, Li K, Verma AK, Ortiz ME, Wohlford-Lenane C, et al. COVID-19 treatments and
672 pathogenesis including anosmia in K18-hACE2 mice. *Nature.* 2021;589(7843):603-7.
- 673 14. Garcia-Arriaza J, Garaigorta U, Perez P, Lazaro-Frias A, Zamora C, Gastaminza P, et al. COVID-19
674 vaccine candidates based on modified vaccinia virus Ankara expressing the SARS-CoV-2 spike induce robust T-
675 and B-cell immune responses and full efficacy in mice. *J Virol.* 2021.
- 676 15. Alsoussi WB, Turner JS, Case JB, Zhao H, Schmitz AJ, Zhou JQ, et al. A Potently Neutralizing Antibody
677 Protects Mice against SARS-CoV-2 Infection. *J Immunol.* 2020;205(4):915-22.
- 678 16. Hassan AO, Case JB, Winkler ES, Thackray LB, Kafai NM, Bailey AL, et al. A SARS-CoV-2 Infection
679 Model in Mice Demonstrates Protection by Neutralizing Antibodies. *Cell.* 2020;182(3):744-53 e4.
- 680 17. Mills RJ, Humphrey SJ, Fortuna PRJ, Lor M, Foster SR, Quaipe-Ryan GA, et al. BET inhibition blocks
681 inflammation-induced cardiac dysfunction and SARS-CoV-2 infection. *Cell.* 2021;184(8):2167-82 e22.

- 682 18. Zucchiatti P, Mitri E, Kenig S, Bille F, Kourousias G, Bedolla DE, et al. Contribution of Ribonucleic Acid
683 (RNA) to the Fourier Transform Infrared (FTIR) Spectrum of Eukaryotic Cells. *Anal Chem*. 2016;88(24):12090-8.
- 684 19. Ami D, Lavatelli F, Rognoni P, Palladini G, Raimondi S, Giorgetti S, et al. In situ characterization of
685 protein aggregates in human tissues affected by light chain amyloidosis: a FTIR microspectroscopy study. *Sci Rep*.
686 2016;6:29096.
- 687 20. Milosevic J, Prodanovic R, Polovic N. On the Protein Fibrillation Pathway: Oligomer Intermediates
688 Detection Using ATR-FTIR Spectroscopy. *Molecules*. 2021;26(4).
- 689 21. Wood BR. The importance of hydration and DNA conformation in interpreting infrared spectra of cells and
690 tissues. *Chem Soc Rev*. 2016;45(7):1980-98.
- 691 22. Movasaghi Z, Rehman S, Rehman IU. Fourier transform infrared (FTIR) spectroscopy of biological tissues.
692 *Appl Spectrosc Rev*. 2008;43(2):134-79.
- 693 23. Geinguenaud F, Militello V, Arluison V. Application of FTIR Spectroscopy to Analyze RNA Structure.
694 *Methods Mol Biol*. 2020;2113:119-33.
- 695 24. Zhuang JS, Li M, Pu YQ, Ragauskas AJ, Yoo CG. Observation of Potential Contaminants in Processed
696 Biomass Using Fourier Transform Infrared Spectroscopy. *Appl Sci-Basel*. 2020;10(12).
- 697 25. Banyay M, Sarkar M, Graslund A. A library of IR bands of nucleic acids in solution. *Biophys Chem*.
698 2003;104(2):477-88.
- 699 26. Mehrotra R. *Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis*. 2006.
- 700 27. Williams E, Bond K, Zhang BW, Putland M, Williamson DA. Saliva as a Noninvasive Specimen for
701 Detection of SARS-CoV-2. *Journal of Clinical Microbiology*. 2020;58(8).
- 702 28. Manabe YC, Reuland C, Yu T, Azamfirei R, Hardick JP, Church T, et al. Self-Collected Oral Fluid Saliva
703 Is Insensitive Compared With Nasal-Oropharyngeal Swabs in the Detection of Severe Acute Respiratory Syndrome
704 Coronavirus 2 in Outpatients. *Open Forum Infect Di*. 2021;8(2).
- 705 29. Butler-Laporte G, Lawandi A, Schiller I. Comparison of Saliva and Nasopharyngeal Swab Nucleic Acid
706 Amplification Testing for Detection of SARS-CoV-2: A Systematic Review and Meta-analysis (Jan,
707 10.1001/jamainternmed.2020.8876, 2021). *Jama Intern Med*. 2021;181(3):409-.
- 708 30. Cai S, Yang H, Zeng K, Zhang J, Zhong N, Wang Y, et al. EGCG Inhibited Lipofuscin Formation Based
709 on Intercepting Amyloidogenic beta-Sheet-Rich Structure Conversion. *PLoS One*. 2016;11(3):e0152064.
- 710 31. Sinha N, Thakur AK. Likelihood of amyloid formation in COVID-19-induced ARDS. *Trends Microbiol*.
711 2021.
- 712 32. Jana AK, Greenwood AB, Hansmann UHE. Presence of a SARS-CoV-2 Protein Enhances Amyloid
713 Formation of Serum Amyloid A. *J Phys Chem B*. 2021;125(32):9155-67.
- 714 33. Zinellu A, Paliogiannis P, Carru C, Mangoni AA. Serum amyloid A concentrations, COVID-19 severity
715 and mortality: An updated systematic review and meta-analysis. *Int J Infect Dis*. 2021;105:668-74.
- 716 34. Tavassoly O, Safavi F, Tavassoly I. Seeding Brain Protein Aggregation by SARS-CoV-2 as a Possible
717 Long-Term Complication of COVID-19 Infection. *Acs Chem Neurosci*. 2020;11(22):3704-6.
- 718 35. Idrees D, Kumar V. SARS-CoV-2 spike protein interactions with amyloidogenic proteins: Potential clues to
719 neurodegeneration. *Biochem Bioph Res Co*. 2021;554:94-8.
- 720 36. Geng H, Subramanian S, Wu L, Bu HF, Wang X, Du C, et al. SARS-CoV-2 ORF8 Forms Intracellular
721 Aggregates and Inhibits IFNgamma-Induced Antiviral Gene Expression in Human Lung Epithelial Cells. *Front*
722 *Immunol*. 2021;12:679482.
- 723 37. Savitt AG, Manimala S, White T, Fandaros M, Yin W, Duan H, et al. SARS-CoV-2 Exacerbates COVID-
724 19 Pathology Through Activation of the Complement and Kinin Systems. *Front Immunol*. 2021;12:767347.
- 725 38. Renne T, Stavrou EX. Roles of Factor XII in Innate Immunity. *Front Immunol*. 2019;10:2011.
- 726 39. Muller WEG, Neufurth M, Wang SF, Tan RW, Schroder HC, Wang XH. Morphogenetic (Mucin
727 Expression) as Well as Potential Anti-Corona Viral Activity of the Marine Secondary Metabolite Polyphosphate on
728 A549 Cells. *Mar Drugs*. 2020;18(12).
- 729 40. Maas C, Renne T. Coagulation factor XII in thrombosis and inflammation. *Blood*. 2018;131(17):1903-9.
- 730 41. Muller F, Mutch NJ, Schenk WA, Smith SA, Esterl L, Spronk HM, et al. Platelet polyphosphates are
731 proinflammatory and procoagulant mediators in vivo. *Cell*. 2009;139(6):1143-56.
- 732 42. Mailer RK, Hanel L, Allende M, Renne T. Polyphosphate as a Target for Interference With Inflammation
733 and Thrombosis. *Front Med-Lausanne*. 2019;6.
- 734 43. Laubli H, Alisson-Silva F, Stanczak MA, Siddiqui SS, Deng LW, Verhagen A, et al. Lectin Galactoside-
735 binding Soluble 3 Binding Protein (LGALS3BP) Is a Tumor-associated Immunomodulatory Ligand for CD33-
736 related Siglecs. *Journal of Biological Chemistry*. 2014;289(48):33481-91.

- 737 44. Montefusco L, Ben Nasr M, D'Addio F, Loretelli C, Rossi A, Pastore I, et al. Acute and long-term
738 disruption of glycometabolic control after SARS-CoV-2 infection. *Nat Metab.* 2021;3(6):774-85.
- 739 45. Reiterer M, Rajan M, Gomez-Banoy N, Lau JD, Gomez-Escobar LG, Ma L, et al. Hyperglycemia in acute
740 COVID-19 is characterized by insulin resistance and adipose tissue infectivity by SARS-CoV-2. *Cell Metab.* 2021.
- 741 46. Kitane DL, Loukman S, Marchoudi N, Fernandez-Galiana A, El Ansari FZ, Jouali F, et al. A simple and
742 fast spectroscopy-based technique for Covid-19 diagnosis. *Scientific Reports.* 2021;11(1).
- 743 47. Lucas C, Klein J, Sundaram ME, Liu F, Wong P, Silva J, et al. Delayed production of neutralizing
744 antibodies correlates with fatal COVID-19. *Nat Med.* 2021;27(7):1178-86.
- 745 48. Amarilla AA, Sng JDJ, Parry R, Deerain JM, Potter JR, Setoh YX, et al. A versatile reverse genetics
746 platform for SARS-CoV-2 and other positive-strand RNA viruses. *Nat Commun.* 2021;12(1):3431.
- 747 49. Johnson BJ, Le TT, Dobbin CA, Banovic T, Howard CB, Flores Fde M, et al. Heat shock protein 10
748 inhibits lipopolysaccharide-induced inflammatory mediator production. *J Biol Chem.* 2005;280(6):4037-47.
- 749