

1 Proteome-wide autoantibody screening and holistic autoantigenomic analysis

2 unveil COVID-19 signature of autoantibody landscape

3

4 Kazuki M Matsuda, MD, PhD,^{1*} Yoshiaki Kawase, PhD,^{2*} Kazuhiro Iwadoh, MD,

5 PhD,¹ Makoto Kurano, MD, PhD,³ Yutaka Yatomi, MD, PhD,⁴ Koh Okamoto, MD, MS,

6 PhD,⁵ Kyoji Moriya, MD, PhD,^{5,6} Hirohito Kotani,¹ MD, PhD,¹ Teruyoshi Hisamoto,

7 MD, PhD,¹ Ai Kuzumi, MD, PhD¹, Takemichi Fukasawa MD, PhD^{1,7}, Asako

8 Yoshizaki-Ogawa, MD, PhD¹, Masanori Kono, MD, PhD,⁸ Tomohisa Okamura, MD,

9 PhD,⁸ Hirofumi Shoda, MD, PhD,⁸ Keishi Fujio, MD, PhD,⁸ Kei Yamaguchi, PhD,⁹

10 Taishi Okumura,⁹ Chihiro Ono,⁹ Yuki Kobayashi, PhD,⁹ Ayaka Sato,⁹ Ayako Miya,⁹

11 PhD, Naoki Goshima, PhD,⁹ Rikako Uchino,¹⁰ Yumi Murakami, PhD,¹⁰ Hiroshi

12 Matsunaka, PhD,¹⁰ Hiroshi Imai, PhD,² Shinichi Sato, MD, PhD,¹ Rudy Raymond,

13 PhD,² Ayumi Yoshizaki, MD, PhD^{1,7#}

14

15 1. Department of Dermatology, The University of Tokyo Graduate School of Medicine,

16 Tokyo, Japan

17 2. Department of Computer Science, The University of Tokyo Graduate School of

18 Information Science and Technology, Tokyo, Japan

- 19 3. Department of Clinical Laboratory Medicine, The University of Tokyo Graduate
20 School of Medicine, Tokyo, Japan
- 21 4. Graduate School, International University of Health and Welfare, Tokyo, Japan
- 22 5. Department of Infectious Diseases, The University of Tokyo Hospital, Tokyo, Japan
- 23 6. Division of Infection Prevention and Control, Postgraduate School of Healthcare,
24 Tokyo Healthcare University, Tokyo, Japan
- 25 7. Department of Clinical Cannabinoid Research, The University of Tokyo Graduate
26 School of Medicine, Tokyo, Japan
- 27 8. Department of Allergy and Rheumatology, The University of Tokyo Graduate
28 School of Medicine, Tokyo, Japan
- 29 9. ProteoBridge Corporation, Tokyo, Japan
- 30 10. NOV Academic Research, TOKIWA Pharmaceutical Co., Ltd., Tokyo, Japan

31

32 * Equally contributed.

33

34

35 **# Corresponding authors**

36 Ayumi Yoshizaki, MD, PhD

37 Department of Dermatology and Department of Clinical Cannabinoid Research, The

38 University of Tokyo Graduate School of Medicine, 7-3-1, Hongo, Bunkyo-ku, Tokyo,

39 Japan, 1138655

40 Phone: +81-3-3815-5411

41 ORCID: 0000-0002-8194-9140

42 E-mail: ayuyoshi@me.com

43 _____

44 **Abstract**

45 This study presents “aUToAntiBody Comprehensive Database (UT-ABCD)”, a
46 proteome-wide catalog of autoantibody profiles in 284 human individuals. The subjects
47 included patients diagnosed with Coronavirus disease 2019 (COVID-19), systemic
48 sclerosis (SSc), systemic lupus erythematosus (SLE), anti-neutrophil cytoplasmic
49 antibody-associated vasculitis (AAV), atopic dermatitis (AD), as well as healthy
50 controls (HC). Our investigation employed proteome-wide autoantibody screening
51 (PWAS) that utilizes 13,350 autoantigens displayed on wet protein arrays, covering
52 approximately 90% of the human transcriptome. Our findings demonstrated significant
53 elevation of autoantibody levels in COVID-19, SSc, and SLE patients. Unique sets of
54 disease-specific autoantibodies were identified, highlighting the role of autoantibodies
55 against proteins associated with cytokine signaling in immune systems and viral
56 infection pathways. Employing machine learning, we distinguished COVID-19 cases
57 with high accuracy based on autoantibody profiles, notably identifying antibodies
58 against proteins encoded by *BCORP1* and *KAT2A* as highly specific to COVID-19.
59 Longitudinal analysis revealed dynamic changes in autoantibody levels throughout the
60 course of COVID-19, independent of disease severity. Our research highlights the
61 effectiveness of integrating PWAS and autoantigenomics in exploring immune

62 responses in COVID-19 and other diseases. It provides a deeper understanding of the
63 autoimmunity landscape in human disorders and introduces a new bioresource for
64 further investigation.
65

66 **Introduction**

67 Coronavirus disease 2019 (COVID-19), an infectious disease caused by severe
68 acute respiratory syndrome coronavirus 2 (SARS-CoV-2),¹ has brought a global
69 pandemic since early 2020 with threat on human health and public safety throughout the
70 world.² The pathophysiology of COVID-19 is characterized by multiple organ injuries
71 triggered by excessive immune response.^{3,4} Cytokine storm in the lung causes acute
72 respiratory distress syndrome, which leads to hypoxemia, respiratory failure,
73 requirement of ventilation, and even death. One of the biggest challenges in clinical
74 management of COVID-19 patients lies in accurately identifying and categorizing cases
75 at higher risk of such serious clinical course. Known risk factors include older age, male
76 gender, smoking, diabetes, obesity, hypertension, immunodeficiency, and
77 malignancies.⁵

78 Humoral immunity plays pivotal roles in COVID-19. Although dramatic
79 success of mRNA vaccines and SARS-CoV-2 neutralizing monoclonal antibodies in
80 preventing serious illnesses, accumulating evidence have suggested the vicious roles of
81 dysregulated humoral immunity. As well as earlier work linking anti-cytokine
82 antibodies to mycobacterial, staphylococcal and fungal diseases,^{6,7} autoantibodies
83 against cytokines have been described in COVID-19.⁸ Especially, anti-type I Interferon

antibodies distinguished ~10% of life-threatening pneumonia and ~20% of deaths from COVID-19.⁹⁻¹¹ A high-throughput screening by yeast display of the secretome further revealed the presence of autoantibodies against several immune factors, including chemokines.¹² In addition, autoantibodies characteristic of systemic autoimmune disorders, such as anti-phospholipid antibodies, anti-nuclear antibodies and rheumatoid factor, were reported in COVID-19.¹³ More recently, association between COVID-19 severity and autoantibodies targeting G protein-coupled receptors and renin-angiotensin system-related proteins has been reported.¹⁴

To comprehensively understand such clinical significance of autoantibodies in human diseases including COVID-19, high-precision autoantibody measurement with a proteome-wide scale is necessary. Herein, we engaged our unique technology for proteome-wide autoantibody screening (PWAS) that covers approximately 90% of the human transcriptome,¹⁵⁻¹⁷ in the serum samples derived from individual patients. Our pipeline integrates human cDNA library (HuPEX),¹⁸ a wheat germ cell-free system for high-throughput *in vitro* protein synthesis,¹⁹⁻²¹ and technology for manipulating protein arrays kept in moist conditions during the entire handling process,²² namely wet protein arrays (WPAs). We have applied this method in multiple inflammatory or malignant disorders for validating its potential for illustrating the “autoantibody landscape” of

102 human disorders, which revealed its usefulness for holistic evaluation of disease-related
 103 autoantibodies,¹⁷ developing novel biomarkers,¹⁵ and moreover, investigating unknown
 104 pathophysiology driven by autoantibodies.¹⁶

105 Our aim was to demonstrate the utility of our omics-based methodology for
 106 autoantibody evaluation and data interpretation procedure, so-called “autoantigenomics,”
 107 targeting COVID-19. In 2020, Moritz *et al.* defined autoantigenomics as a branch of
 108 systems immunology, which holistically analyze the repertoire of autoantibodies
 109 engaging omics-based bioinformatical approaches including hierarchical cluster analysis,
 110 enrichment analysis, and machine learning.²³ The concept of autoantigenomics stand on
 111 hypotheses that there might be differences in the sets of targeted antigens underlying
 112 intra-disease heterogeneity in human, which would be supported by our novel data
 113 shown below.

114

115 **Results**

116 *Overview*

117 We recruited 73 patients with COVID-19, 32 patients with systemic sclerosis
118 (SSc), 60 patients with systemic lupus erythematosus (SLE), 29 patients with
119 anti-neutrophil cytoplasmic antibody-associated vasculitis (AAV), 26 patients with
120 atopic dermatitis (AD), and 64 healthy controls (HC) for serum sample collection
121 (**Supplementary Table 1**). For each individual serum, PWAS was performed (**Fig. 1**).
122 The digest of the results will be available as “aUToAntiBody Comprehensive Database
123 (UT-ABCD)”. We found that sum of autoantibody levels (SAL) was significantly
124 elevated in patients with COVID-19, SSc, or SLE, compared to HCs, while there was
125 no statistically significant difference in SAL between AD or AAV patients and HCs
126 (**Fig. 2A**). This tendency was consistent across both gender and age groups.
127 (**Supplementary Fig. 1**).

128

129 *Identification of disease-specific autoantibodies*

130 We identified distinct sets of autoantibodies that showed a more than twofold
131 significant increase in each disease condition relative to HCs (**Fig. 2B**). Notably, certain
132 autoantibodies were unique to each disease (**Fig. 2C**). To illustrate the variability in

133 serum levels of these disease-specific autoantibodies across individuals, we utilized
 134 uniform manifold approximation and projection (UMAP), resulting in distinctive
 135 patterns on the UMAP plots for each condition (**Fig. 2D**). Gene ontology analysis
 136 linked to the genes responsible for the proteins targeted by such disease-specific
 137 autoantibodies pointed to shared biological functions, with a focus on viral infection
 138 pathways and cytokine signaling in immune system, in COVID-19, SLE, and AAV (**Fig.**
 139 **2E and 2F**). Holistic analysis of autoantibodies targeting cytokines, or their receptors
 140 displayed on our WPAs revealed that strong positivity for autoantibodies targeting type
 141 1 interferon was specifically observed in COVID-19 patients, while weak positivity was
 142 seen in SLE patients (**Supplementary Fig. 2**).

143

144 *Selection of machine learning frameworks*

145 To further investigate the association between autoantibody profiles and
 146 COVID-19, we adopted a machine learning approach. We tested nine different methods
 147 to differentiate COVID-19 cases from the others: simple linear regression, Ridge
 148 regression, logistic regression with data normalization, logistic regression with data
 149 standardization, support vector machine with data normalization, support vector
 150 machine with data standardization, light gradient boosting machine (LightGBM), and

151 extremely gradient boosting decision trees (XGBoost). As a result, XGBoost showed
152 the highest value of the area under the receiver-operator characteristics curve for
153 distinction of COVID-19 cases from the others (**Supplementary Table 2**).
154 Consequently, we opted to focus on this method for our subsequent analysis.

155

156 *Performance of XGBoost*

157 In our subsequent analysis using the entire dataset, we experimented with
158 binary (COVID-19 vs. others), ternary (mild COVID-19 vs. moderate to severe
159 COVID-19 vs. others), and multiclass (mild COVID-19 vs. moderate to severe
160 COVID-19 vs. AAV vs. AD vs. SSc vs. SLE vs. HCs) classifications through XGBoost.
161 The most significant autoantibodies identified across all models are depicted in **Fig. 3A**,
162 **3B, and 3C**, with autoantibodies against translational products from *BCL6 Corepressor*
163 *Pseudogene 1 (BCORP1)* emerging as a top feature in every model. Similarly,
164 antibodies against K-Acetyltransferase 2A (KAT2A) were consistently prominent.
165 Notably, Anti-BCORP1 and anti-KAT2A Abs were highlighted as important items in
166 all the candidate machine learning methods tested (**Supplementary Fig. 3**). There was a
167 correlation between anti-BCORP Abs and anti-KAT2A Abs as illustrated in **Fig. 3D**,
168 **3E, and 3F**. Remarkably, established serum markers for SSc and SLE, such as

169 anti-topoisomerase 1 (TOP1) Abs, anti-centromere protein-B (CENPB) Abs,
 170 anti-tripartite motif-containing protein 21 (TRIM21) Abs, anti-small nuclear
 171 ribonucleoprotein polypeptide (SNRP)-A Abs, and anti-SNRPB Abs, were also
 172 identified. The visualization of mean serum levels of these prominent markers through
 173 spider charts revealed distinctive patterns across the different conditions (**Fig. 3G, 3H,**
 174 **and 3I**). The models incorporating these markers as features achieved the highest
 175 accuracy in both binary and ternary classifications and showed significantly better
 176 outcomes than chance in the complex seven-class classification (**Supplementary Table**
 177 **3**).

178

179 *Clinical relevance of autoantibodies*

180 The serum levels of the top 20 autoantibodies highlighted through multi-class
 181 classification for each participant were depicted in a heatmap (**Fig. 4A**). Hierarchical
 182 clustering identified three unique groups of autoantibodies: cluster I, which included
 183 two autoantibodies highly specific to COVID-19 (anti-BCORP1 and anti-KAT2A Abs);
 184 cluster II, comprising autoantibodies that are commonly elevated across various
 185 conditions; and cluster III, involving well-established biomarkers for SSc and SLE.
 186 Principal component analysis (PCA) effectively distinguished between seven categories

187 (Fig. 4B), particularly using principal component (PC) 2 as an indicator for COVID-19
 188 (Fig. 4C). Correspondingly, antibodies against BCORP1 and KAT2A constituted the
 189 predominant contribution to PC2 (Fig. 4D). Serum levels of anti-BCORP1 and
 190 anti-KAT2A Abs were prominently elevated in COVID-19, along with a part of cases
 191 with SLE (Fig. 4E and 4F). This trend was consistent among both sex and age groups
 192 (Supplementary Fig. 4). We also explored the link between COVID-19 clinical
 193 outcomes and the presence of anti-BCORP1 or anti-KAT2A Abs. Applying a threshold
 194 set at the mean plus two standard deviations above healthy controls (indicated by the
 195 red dashed lines in Fig. 4B and 4C), we identified 61 of the 73 total COVID-19 cases
 196 as having elevated levels of anti-BCORP1 Abs and 34 cases with elevated levels of
 197 anti-KAT2A Abs. While no significant association was found between elevated
 198 anti-BCORP1 Abs and clinical features (Supplementary Table 4), elevation of
 199 anti-KAT2A Abs was significantly linked to a reduced need for intensive care,
 200 including intubation and mechanical ventilation (odds ratio = 0.19, $P = 0.02$;
 201 Supplementary Table 5).

202

203 *Time course of autoantibody levels during COVID-19*

204 Finally, we conducted longitudinal analysis of the humoral immune response in
 205 COVID-19 patients targeting paired serum samples from an "early" timepoint (within
 206 10 days after symptoms began) and a "late" timepoint (11-20 days after symptom onset)
 207 from 41 individuals. These samples were used for PWAS and for assessing IgG levels
 208 against SARS-CoV-2 particles: the nucleocapsid protein (N), spike protein (S), and the
 209 receptor binding domain (RBD) of S. Consistent with our prior findings, early timepoint
 210 samples showed IgG against N, S, and RBD in only a small part of the patients, with a
 211 marked increase in most patients by the late timepoint (**Fig. 5A**). To the contrary, SAL
 212 remained unchanged over time (**Fig. 5B**). Further exploration revealed a significant
 213 decrease in 293 autoantibodies and increase in 116 autoantibodies over the course of the
 214 infection, including those targeting BCORP1 and KAT2A (**Fig. 5C and 5D**). There was
 215 no observed correlation between these autoantibodies and IgG levels against N, S, and
 216 RBD (**Fig. 5E**). Notably, both anti-BCORP1 and anti-KAT2A Ab levels rose over time
 217 independent of disease severity (**Fig. 5F**). Additional analysis did not find any
 218 correlation between these two autoantibodies and IgG targeting N, S, and RBD at either
 219 timepoint or their progression over time (**Fig. 5G and 5H**).
 220

221 Discussion

222 Herein we conducted PWAS subjecting serum samples derived from HCs and
 223 patients with COVID-19, AD, AAV, SLE, and SSc (**Fig. 1**). We demonstrated that our
 224 PWAS methodology enables us to identify disease-specific autoantibodies (**Fig. 2A, 2B,**
 225 **and 2C**), which demonstrate the distinct distribution of autoantibodies and common
 226 biological processes among different conditions (**Fig. 2D, 2E, and 2F**). The contrast in
 227 autoantibody profiles was accentuated through the application of a machine learning
 228 approach, particularly leveraging the XGBoost framework (**Fig. 3 and 4**). We also
 229 investigated the longitudinal change of autoantibody profiles along with the time course
 230 of COVID-19 and its correlation with emergence of antibodies targeting COVID-19
 231 particles (**Fig. 5**). Collectively, these results supported our hypothesis that the
 232 combination of PWAS and omics-based bioinformatic methodologies is adaptable to
 233 human disorders including COVID-19. Additionally, our findings in this study as well
 234 as our previous works provide a comprehensive catalog of autoantibody profiles in
 235 various diseases and open the door to creating innovative diagnostic methods that can
 236 differentiate between various disease mechanisms affecting multiple organs, utilizing
 237 distinct autoantibody patterns measured by WPAs.^{15,17}

238 The machine learning-based approach has discovered that the presence of
239 anti-BCORP1 antibodies is highly specific to COVID-19 (**Fig. 4E**). Despite *BCORP1*
240 being categorized as a pseudogene, it appears to undergo transcription into mRNA, as
241 several transcriptomic studies have reported the presence of *BCORP1*-derived
242 sequences.^{24,25} Especially, Deng MC's research has shown that the transcription levels
243 of *BCORP1* in peripheral blood mononuclear cells of COVID-19 patients correlate with
244 early functional recovery and 1-year survival.²⁵ However, it remains unclear whether
245 *BCORP1* mRNA is translated into functional proteins, and further investigation is
246 needed in this regard. Moreover, our discovery of anti-BCORP1 Abs in both males and
247 females raises questions, as BCORP1 is located on the Y chromosome. This leads us to
248 consider the possibility of cross-reactivity of these antibodies against foreign antigens,
249 such as proteins comprising SARS-CoV-2 virions. However, we could not find any
250 correlation between serum levels of anti-BCORP1 Abs and antibodies targeting
251 SARS-CoV-2 particles at any of the timepoints examined, nor in their changes over
252 time (**Fig. 5G**). Another hypothesis is cross-reaction between the antigen we produced
253 from *BCORP1* cDNA and other human proteins. The *BCL2 co-repressor (BCOR)* gene,
254 the counterpart of *BCORP1* found on the X chromosome, has a nucleotide sequence that
255 is over 99% identical to *BCORP1*. It is conceivable that the BCOR protein is the actual

256 target of the detected anti-BCORP1 Abs in our study. However, this theory remains
257 unconfirmed as our cDNA library did not include *BCOR* cDNA to validate this
258 hypothesis.

259 Anti-KAT2A antibodies were also found to be specifically elevated in the sera
260 of COVID-19 patients (**Fig. 4F**). Like anti-BCORP1 Abs, the levels of anti-KAT2A
261 Abs increased over the course of COVID-19 but were not correlated to antibodies
262 targeting SARS-CoV-2 particles (**Fig. 5H**). This observation indicates that
263 autoantibodies to KAT2A emerge because of autoantigen exposure due to tissue
264 damage triggered by COVID-19, not as a reflection of cross-reaction between
265 SARS-CoV-2 virions. KAT2A functions as a histone acetyltransferase that plays a role
266 in the epigenetic regulation of the genome by modifying chromatin structures. There is
267 notable research by John K. et al., indicating that the SARS-CoV-2 ORF8 protein
268 mimics the histone H3's ARKS motifs, which interfere with the role of KAT2A role in
269 host cell epigenetic regulation.²⁶ Intriguingly, patients who did not show an increase in
270 anti-KAT2A antibodies were more often those requiring intensive care and mechanical
271 ventilation (**Supplementary Table 5**). This observation has led to the proposition that
272 the presence of anti-KAT2A Abs could be indicative of an effective immune response
273 to the virus by KAT2A upregulation, a hypothesis that warrants further research.

274 Our study has multiple strengths. First, the wheat-germ *in vitro* protein
275 synthesis system and technique for manipulation of WPAs realized high-throughput
276 expression of various human proteins including exoproteome upon a single platform.
277 Second, as a result, our autoantibody measurement could cover a wider range of
278 antigens at an almost proteome-wide level, which enabled us to apply omics-based
279 bioinformatical approaches for interpreting the data. Third, we investigated the
280 longitudinal change of autoantibody profiles within COVID-19 patients, along with the
281 measurement of antibodies targeting SARS-CoV-2 particles. The limitation of our
282 present study includes its retrospective design and a relatively small number of the
283 subjects. Furthermore, we could not distinguish whether the autoantibodies found in our
284 measurement were predisposed before COVID-19 or newly appeared after infection. In
285 addition, functional assays for the autoantibodies such as neutralizing assays or *in vivo*
286 studies are lacking. Therefore, insights into the direct contribution of each autoantibody
287 to the pathophysiology of COVID-19 are limited. Our next challenges would include
288 collecting serum samples before and after COVID-19 by accessing to population-based
289 cohorts, evaluating the function of each autoantibody against their target molecules, and
290 testing their contribution to the pathogenesis in animal experiments.
291

292 **References**

- 293 1. Grp, C. S. & Version, P. The species Severe acute respiratory syndromerelated
294 coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.*
295 5, 5346–544 (2020).
- 296 2. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and
297 COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
- 298 3. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and
299 immunosuppression. *Lancet* **295**, 1033–1034 (2020).
- 300 4. Huang, C. *et al.* Clinical features of patients infected with 2019 novel
301 coronavirus in Wuhan , China. *Lancet* **395**, 497–506 (2020).
- 302 5. Dessie, Z. G. & Zewotir, T. Mortality□related risk factors of COVID□19: a
303 systematic review and meta□analysis of 42 studies and 423,117 patients. *BMC*
304 *Infect. Dis.* **21**, 855 (2021).
- 305 6. Puel, A., Bastard, P., Bustamante, J. & Casanova, J.-L. Human autoantibodies
306 underlying infectious diseases. *J. Exp. Med.* **219**, e20211387 (2022).
- 307 7. Cheng, A. & Holland, S. M. Anti-cytokine autoantibodies: mechanistic insights
308 and disease associations. *Nat. Rev. Immunol.* (2023)
309 doi:10.1038/s41577-023-00933-2.

- 310 8. Muri, J. *et al.* Autoantibodies against chemokines post-SARS-CoV-2 infection
311 correlate with disease course. *Nat. Immunol.* **24**, 604–611 (2023).
- 312 9. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with
313 life-threatening COVID-19. *Science* (80-.). **370**, (2020).
- 314 10. Bastard, P. *et al.* Autoantibodies neutralizing type I IFNs are present in ~4% of
315 uninfected individuals over 70 years old and account for ~20% of COVID-19
316 deaths. *Sci. Immunol.* **6**, (2021).
- 317 11. Eto, S. *et al.* Neutralizing Type I Interferon Autoantibodies in Japanese Patients
318 with Severe COVID-19. *J. Clin. Immunol.* **42**, 1360–1370 (2022).
- 319 12. Wang, E. Y. *et al.* Diverse functional autoantibodies in patients with COVID-19.
320 *Nature* **595**, 283–288 (2021).
- 321 13. Chang, S. E. *et al.* New-onset IgG autoantibodies in hospitalized patients with
322 COVID-19. *Nat. Commun.* **12**, 5417 (2021).
- 323 14. Cabral-marques, O. *et al.* Autoantibodies targeting GPCRs and RAS-related
324 molecules associate with COVID-19 severity. *Nat. Commun.* **13**, 1220 (2022).
- 325 15. Matsuda, K. M. *et al.* Autoantibody Landscape Revealed by Wet Protein
326 Array: Sum of Autoantibody Levels Reflects Disease Status. *Front. Immunol.*
327 **13**, 1–14 (2022).

- 328 16. Matsuda, K. M., Kotani, H., Yamaguchi, K., Okumura, T. & Fukuda, E.
329 Significance of anti-transcobalamin receptor antibodies in cutaneous arteritis
330 revealed by proteome-wide autoantibody screening. *J. Autoimmun.* **135**, 102995
331 (2023).
- 332 17. Kuzumi, A. *et al.* Comprehensive autoantibody profiling in systemic
333 autoimmunity by a highly-sensitive multiplex protein array. *Front. Immunol.* **14**,
334 (2023).
- 335 18. Goshima, N. *et al.* Human protein factory for converting the transcriptome into
336 an in vitro-expressed proteome. *Nat. Methods* **5**, 1011–1017 (2008).
- 337 19. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein
338 synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. U. S. A.*
339 **99**, 14652–14657 (2002).
- 340 20. Sawasaki, T. *et al.* A bilayer cell-free protein synthesis system for
341 high-throughput screening of gene products. *FEBS Lett.* **514**, 102–105 (2002).
- 342 21. Endo, Y. & Sawasaki, T. Cell-free expression systems for eukaryotic protein
343 production. *Curr. Opin. Biotechnol.* **17**, 373–380 (2006).

- 344 22. Fukuda, E. *et al.* Identification and characterization of the antigen recognized by
345 the germ cell mAb TRA98 using a human comprehensive wet protein array.
346 *Genes to Cells* **26**, 180–189 (2021).
- 347 23. Moritz, C. P. *et al.* Autoantigenomics: Holistic characterization of autoantigen
348 repertoires for a better understanding of autoimmune diseases. *Autoimmun. Rev.*
349 **19**, 102450 (2020).
- 350 24. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by
351 genome-wide integration of transcriptomics and antibody-based proteomics. *Mol.*
352 *Cell. Proteomics* **13**, 397–406 (2014).
- 353 25. Deng, M. C. Multi-dimensional COVID-19 short- and long-term outcome
354 prediction algorithm. *Expert Rev. Precis. Med. drug Dev.* **5**, 239–242 (2020).
- 355 26. Kee, J. *et al.* SARS-CoV-2 disrupts host epigenetic regulation via histone
356 mimicry. *Nature* **610**, 381–388 (2022).
- 357 27. Yamakawa, K. *et al.* Japanese rapid/living recommendations on drug
358 management for COVID □ 19: updated guidelines (July 2022). *Acute Med. Surg.*
359 **9**, 1–21 (2022).

- 360 28. Nakano, Y. *et al.* Time course of the sensitivity and specificity of
361 anti-SARS-CoV-2 IgM and IgG antibodies for symptomatic COVID-19 in Japan.
362 *Sci. Rep.* **11**, 1–10 (2021).
- 363 29. Qian, C. *et al.* Development and multicenter performance evaluation of fully
364 automated SARS-CoV-2 IgM and IgG immunoassays. *Clin. Chem. Lab. Med.* **58**,
365 1601–1607 (2020).
- 366 30. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis
367 of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
- 368
- 369

370 **Materials and Methods**

371 *Human subjects*

372 We consecutively enrolled patients administered to our institution for
373 COVID-19 from April 2020 to April 2021. Inclusion criteria were a SARS-CoV-2
374 positive nasopharyngeal swab test by real-time reverse transcription-polymerase chain
375 reaction (RT-PCR) and age ≥ 18 years. Clinical data were collected by retrospective
376 review of electric medical records. We gathered basic patient information, symptoms,
377 medications, histopathologic features, and laboratory findings from the closest time
378 point from the date of serum collection. The disease severity was assessed following the
379 Japanese guideline for managing COVID-19 patients.²⁷ In brief, individuals requiring
380 intensive care or mechanical ventilation were categorized as severe COVID-19, those
381 exhibiting hypoxemia among the remaining cases were classified as moderate to severe
382 COVID-19, and all other patients were considered to have mild COVID-19. We also
383 gathered serum samples from HCs and patients with AD, AAV, SLE, and SLE. This
384 study has been approved by The University of Tokyo Ethical Committee (Approval
385 number 0695). Written informed consent has been obtained from all the participants.

386

387 *Measurement of IgG targeting SARS-CoV-2 particles*

388 The process of quantifying IgG antibodies that target specific SARS-CoV-2
389 proteins, namely the nucleocapsid protein, spike protein, and the spike protein's
390 receptor binding domain, was conducted as outlined previously using a commercial
391 SARS-CoV-2 IgG kit (YHLO Biotechnology Company, Ltd., Shenzhen, China).^{28,29}
392 This involved an assay where serum samples were combined with magnetic beads
393 coated with the viral proteins and a substance to prepare the samples. This mix was then
394 washed, combined with an acridinium-conjugated anti-human IgG, and washed again.
395 The subsequent steps included adding solutions to induce a chemiluminescent reaction,
396 the intensity of which was measured by the iFlash3000 CLIA analyzer (YHLO
397 Biotechnology Company, Ltd.) A threshold of 10 AU/mL was used for the detection,
398 following the guidelines provided by the manufacturer.

399

400 *Autoantibody measurement*

401 WPAs were arranged as previously described.¹⁵ First, proteins were
402 synthesized *in vitro* utilizing a wheat germ cell-free system from 13,350 clones of the
403 HuPEX.¹⁸ Second, synthesized proteins were plotted onto glass plates (Matsunami
404 Glass, Osaka, Japan) in an array format by the affinity between the GST-tag added to
405 the N-terminus of each protein and glutathione modified on the plates. The WPAs were

406 treated with human serum diluted by 3:1000 in the reaction buffer containing 1x
 407 Synthetic block (Invitrogen), phosphate-buffered saline (PBS), and 0.1% Tween 20.
 408 Next, the WPAs were washed, and goat anti-Human IgG (H+L) Alexa Flour 647
 409 conjugate (Thermo Fisher Scientific, San Jose, CA, USA) diluted 1000-fold was added
 410 to the WPAs and reacted for 1 hour at room temperature. Finally, the WPAs were
 411 washed, air-dried, and fluorescent images were acquired using a fluorescence imager
 412 (Amersham Typhoon, Cytiva, Marlborough, MA, USA). Fluorescence images were
 413 analyzed to quantify serum levels of autoantibodies targeting each antigen, following
 414 the formula shown below:

415

$$Autoantibody\ level\ [AU] = \frac{F_{autoantigen} - F_{negative\ control}}{F_{positive\ control} - F_{negative\ control}} \times 100$$

416 AU : arbitrary unit

417 $F_{autoantigen}$: fluorescent intensity of autoantigen spot

418 $F_{negative\ control}$: fluorescent intensity of negative control spot

419 $F_{positive\ control}$: fluorescent intensity of positive control spot

420

421 *Machine learning*

422 We applied supervised machine learning techniques using the Python code
423 with the scikit-learn library to analyze the measurement data for autoantibodies from
424 284 patients. At random forest, decision trees were built and trained in parallel on
425 subsets of sampled instances and features. Meanwhile, at XGBoost decision trees were
426 built sequentially to improve each other. The final prediction of the random forest was
427 the majority of its decision trees, while that of XGBoost was from their weighted
428 average. The performance of the classifiers was evaluated in area under the
429 operator-receiver characteristics curve (AUC), accuracy, precision, recall, and F1-score,
430 calculated by 5-fold cross validation. The accuracy is the ratio of the correct positive
431 and negative prediction, the precision is the ratio of the correct positive prediction, the
432 recall (or, sensitivity) is the ratio of the correct positive prediction among all true
433 positive instances, and F1-score is the harmonic mean of precision and sensitivity.

434

435 *Statistical analysis*

436 Differentially elevated autoantibodies were defined as more than 2-fold
437 changes in the serum levels with a P value < 0.01 . Gene Ontology Analysis using
438 web-based tools targeted the list of the entry clones coding the differentially highlighted
439 autoantigens was performed for gene-list enrichment analysis, gene-disease association

440 analysis, and transcriptional regulatory network analysis with Metascape.³⁰ Other data

441 analyses and presentations were conducted using Stata IC/15.0 (StataCorp, TX, USA).

442

443 *Data visualization*

444 Box plots, scatter plots, hierarchical clustering and correlation matrix were

445 visualized by using R (v4.2.1). Box plots were defined as follows: the middle line

446 corresponds to the median; the lower and upper hinges correspond to the first and third

447 quartiles; the upper whisker extends from the hinge to the largest value no further than

448 1.5 times the interquartile range (IQR) from the hinge; and the lower whisker extends

449 from the hinge to the smallest value at most 1.5 times the IQR of the hinge.

450

451 **Acknowledgements**

452 We thank Ms. Maiko Enomoto and her colleagues for secretary work. We
453 thank Ms. Teruko Tani and Ms. Mayumi Odagiri for their assistance in clinical data
454 collection. We appreciate Ms. Maiko Matsuda, VESPER Studio Inc., Tokyo, Japan, for
455 her contribution of illustrating skills.
456

457 **Author Contributions**

458 KM Matsuda primarily engaged in autoantibody measurement, clinical data
459 collection, data analysis, visualization, and writing the first draft of the manuscript. Y
460 Kawase primarily contributed to machine learning analysis. K Iwadoh also participated
461 in machine learning analysis. M Kurano, Y Yatomi, K Okamoto, and K Moriya
462 participated in sample collection and clinical data acquisition regarding COVID-19. H
463 Kotani, A Kuzumi, T Fukasawa, A Yoshizaki-Ogawa took part in the sample collection
464 of SSc. T Hisamoto was in charge of sample collection of AD. M Kono, T Okamura, H
465 Shoda, and K Fujio oversaw sample collection of SLE. K Yamaguchi, T Okumura, C
466 Ono, Y Kobayashi, A Sato, A Miya, and N Goshima prepared wet protein arrays,
467 provided technical assistance for autoantibody measurement, participated in data
468 analysis, setup of UT-ABCD, and revised the manuscript. R Uchino, Y Murakami and
469 H Matsunaka provided technical assistance for autoantibody measurement. H Imai and
470 R Raymond supervised the study. S Sato conceptualized and supervised the study. A
471 Yoshizaki conceptualized, launched, and supervised this study, and was involved in
472 revising the manuscript.
473

474 Conflict-of-interest statement

475 K Yamaguchi, T Okumura, C Ono, Y Kobayashi, A Miya, A Sato, and N
 476 Goshima were employed by ProteoBridge Corporation. T Fukasawa and A Yoshizaki
 477 belong to the Social Cooperation Program, Department of Clinical Cannabinoid
 478 Research, The University of Tokyo Graduate School of Medicine, Tokyo, Japan,
 479 supported by Japan Cosmetic Association and Japan Federation of Medium and Small
 480 Enterprise Organizations. T Okamura belongs to the Social Cooperation Program,
 481 Department of Functional Genomics and Immunological Diseases, The University of
 482 Tokyo Graduate School of Medicine, Tokyo, Japan, supported by Chugai
 483 Pharmaceutical Corporation. The remaining authors declare that the research was
 484 conducted in the absence of any commercial or financial relationships that could be
 485 construed as a potential conflict of interest.

486

487

Figure legends

Figure 1. Scheme of PWAS pipeline.

In the first step, proteins were synthesized *in vitro* from the proteome-wide human cDNA library (HuPEX). Promoters (P), Enhancers (E), and FLAG-GST tags were fused to open reading frames of the expression clones by Gateway LR reaction. After polymerase chain reaction amplification and *in vitro* transcription, translation was performed using the wheat germ cell-free synthesis system. In the second step, we prepared WPAs by plotting synthesized proteins onto glass slides in an array format. WPAs were treated with serum samples derived from diseased patients or HCs. Autoantibodies were detected by fluorochrome-conjugated anti-human IgG Ab. In the third step, autoantibody quantification was performed based on the fluorescent values. Analysis of acquired high-dimensional autoantibody profiles was conducted by multiple omics-based approaches. ORF: open reading frame.

Figure 2. Identification of disease-specific autoantibodies.

(A) Box plots that show SAL in each condition. NS: $P > 0.01$, *: $P < 0.01$, ***: $P < 0.0001$. **(B)** Volcano plots that illustrate differentially elevated autoantibodies within each condition. Red horizontal dash lines indicate P values = 0.01. Red vertical dash lines indicate Log_2 Fold Change (Disease/HC) = ± 1 . **(C)** Venn diagram that demonstrates the subsumptions among disease-specific autoantibodies. **(D)** UMAP plot that illustrates the distribution of disease-specific autoantibody profiles of each individual. **(E)** Heat map that shows the result of enrichment analysis targeting the genes responsible for the proteins targeted by such disease-specific autoantibodies. **(F)** Circos plot that depicts the overlap of the gene lists responsible for the proteins targeted by the disease-specific autoantibodies at the biological function level.

Figure 3. Autoantibodies highlighted in each machine learning model.

Autoantibodies that are mostly highlighted according to feature importance in two-class (A), three-class (B), and multi-class (C) classifications. Correlograms depict correlations, showcasing the connection between the highest-ranked autoantibodies in two-class (D), three-class (E), and multi-class (F) classifications. The correlation strength is denoted by Spearman's rho on the color scale. Circle sizes represent the significance of the p-values, with only those with $P < 0.01$ being displayed. Radar charts present the average normalized quantities of the most important autoantibodies in each model, with line colors distinguishing between different disease categories in two-class (G), three-class (H), and multi-class (I) classifications.

Figure 4. COVID-19 signature of autoantibody landscape. **(A)** The heatmap's columns display the serum autoantibody concentrations highlighted in the multi-class classification using XGBoost. **(B)** PCA graph plots individual participants as points, with color coding to differentiate among various disease classes. **(C)** The loading diagram illustrates the contributions to PC1 and PC2, with I, II, and III marking the clusters defined in (A). **(D)** The bar graphs show the loadings of each autoantibody on PC1 and PC2. **(E)** A box plot presents the serum levels of anti-BCORP1 Abs in the subjects. **(F)** Another box plot indicates the serum levels of anti-KAT2A Abs in the subjects. Red vertical dash lines indicate mean + 2SD in HC.

Figure 5. Longitudinal change of humoral immune response in COVID-19. (A)

This box plot outlines the serum concentrations of IgG antibodies against N, S, or RBD in patients with COVID-19 at two intervals: "early" signifies within 10 days of symptom onset, and "late" refers to 11-20 days after symptoms appear. A red dashed line marks the threshold for a positive test result. ***: $P < 0.0001$. **(B)** These box plots demonstrate the SAL in patients with COVID-19 during the "early" and "late" time points. NS: $P > 0.01$. **(C)** The volcano plot highlights biomarkers that are significantly increased (in red) or decreased (in blue) over time in patients with COVID-19. **(D)** A Venn diagram illustrates the overlap between autoantibodies that changed over time and those specific to COVID-19. **(E)** The correlogram visualizes the relationships between the overlapping autoantibodies that either increased over time or are specific to COVID-19, with the color scale indicating the strength of correlation according to Spearman's rho and circle sizes depicting the significance of p-values, focusing on those with $P < 0.01$. **(F)** These box plots depict the time-based evolution of serum levels of anti-BCORP1 and anti-KAT2A antibodies in COVID-19 patients, categorized by disease severity. **(G)(H)** The scatter plots illustrate the correlation between serum levels of anti-BCORP or anti-KAT2A Abs and IgG antibodies against the COVID-19 components at different time points, as well as their progression over time. The red lines

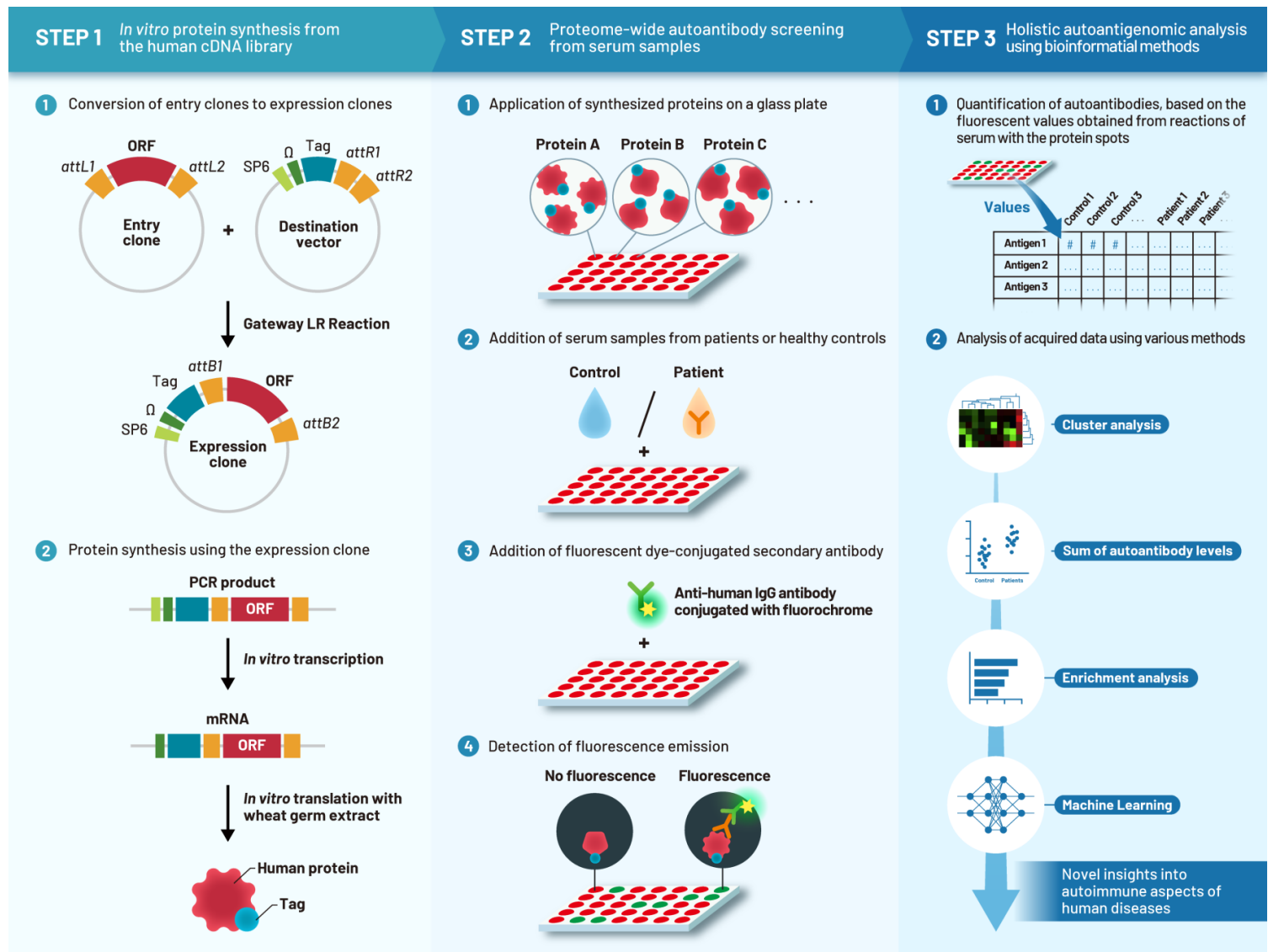
and the surrounding shaded areas indicate the regression line and the 95% confidence interval, respectively.

Supplementary Figure 1. The sum of autoantibody levels by sex and age. (A) The sum of autoantibody levels (SAL) in males. **(B)** SAL in females. **(C)** SAL for age < 50 years old. **(D)** SAL for age ≥ 50 years old.

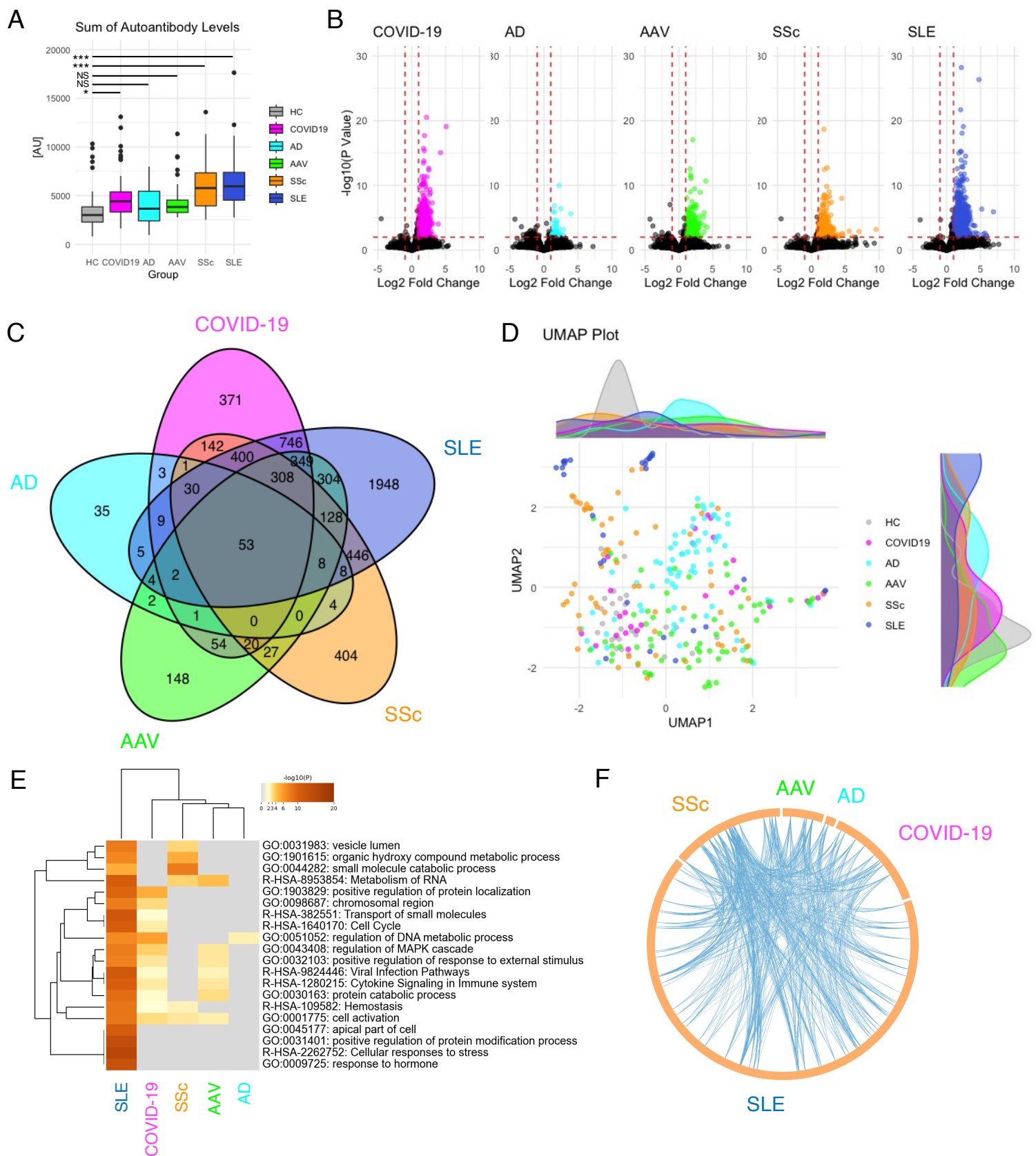
Supplementary Figure 2. Autoantibodies to cytokines or their receptors. (A) The heatmap's columns display the serum autoantibody concentrations targeting cytokines in each subject evaluated by our proteome-wide autoantibody screening. **(B)** The heatmap's columns display the serum autoantibody concentrations targeting cytokine receptors in each subject evaluated by our proteome-wide autoantibody screening. **(C)** A box plot presents the serum levels of anti-interferon alpha 2 (IFNA2) Abs in the subjects. **(D)** Another box plot indicates the serum levels of anti-interferon alpha 4 (IFN4A) Abs in the subjects.

Supplementary Figure 3. Feature importance of top highlighted autoantibodies in other candidate machine learning frameworks. (A) Simple linear regression. **(B)** Ridge regression. **(C)** Logistic regression with normalization. **(D)** Logistic regression with standardization. **(E)** SVM with normalization. **(F)** SVM with standardization. **(G)** LightBGM. **(H)** Random Forest.

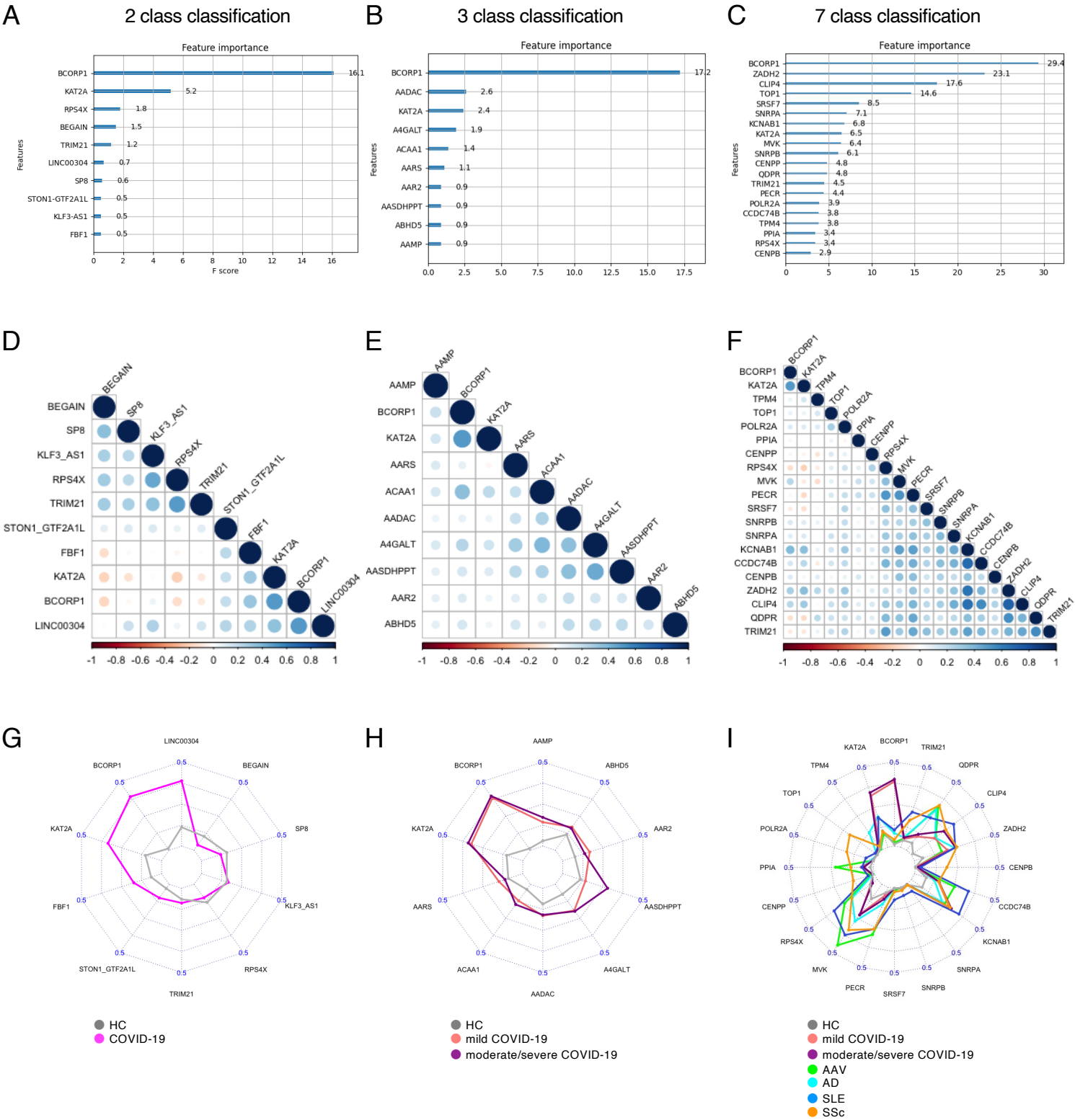
Supplementary Figure 4. Serum levels of anti-BCORP1 and anti-KAT2A Abs by sex and age. (A) Serum levels of anti-BCORP1 Abs by sex. F: female, M: male. **(B)** Serum levels of anti-BCORP1 Abs by age. **(C)** Serum levels of anti-KAT2A Abs by sex. **(D)** Serum levels of anti-KAT2A Abs by age. Red vertical dash lines indicate mean + 2SD in HC.



Matsuda KM et al.
Figure 1

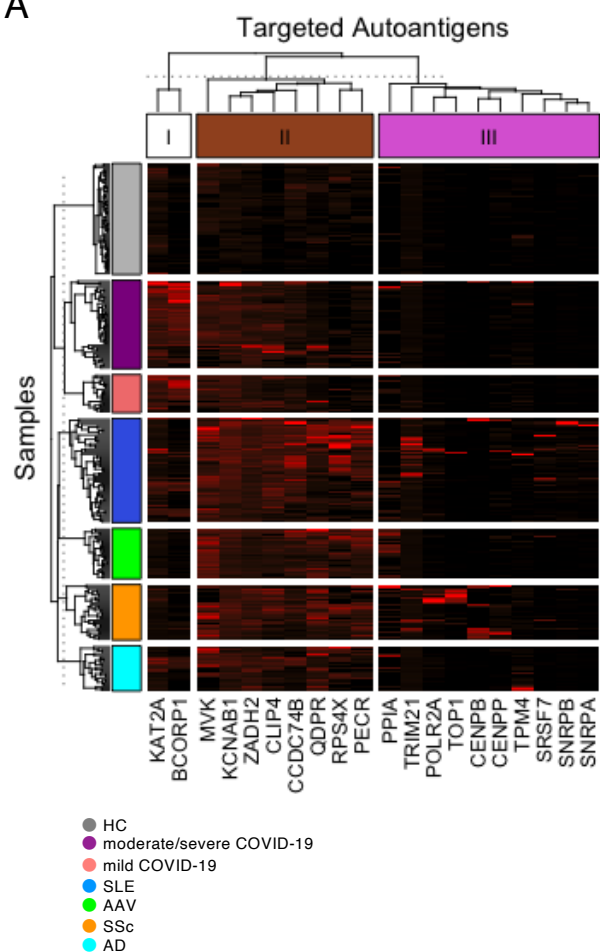


Matsuda KM et al.
Figure 2



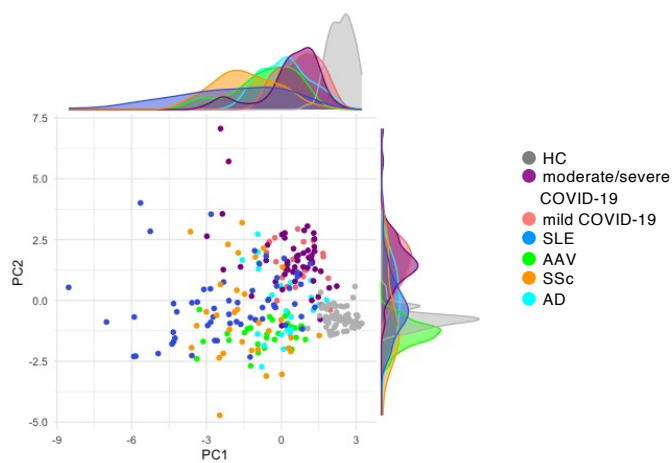
Matsuda KM et al.
Figure 3

A



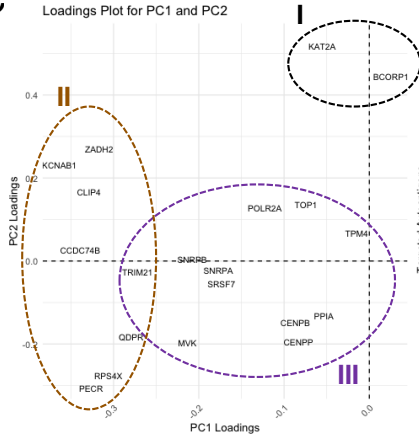
B

PCA

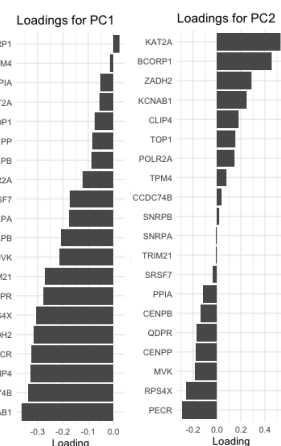


C

Loadings Plot for PC1 and PC2

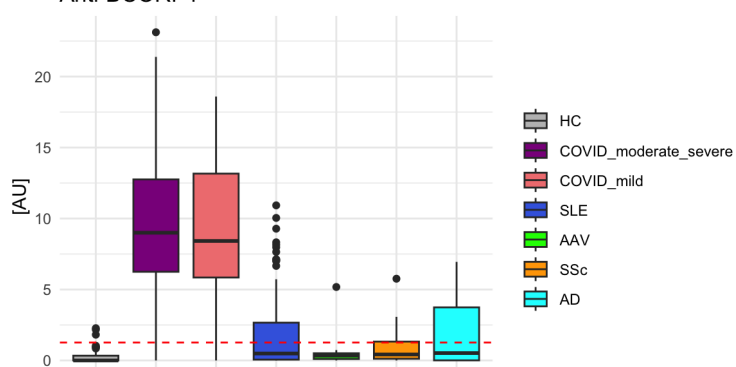


D



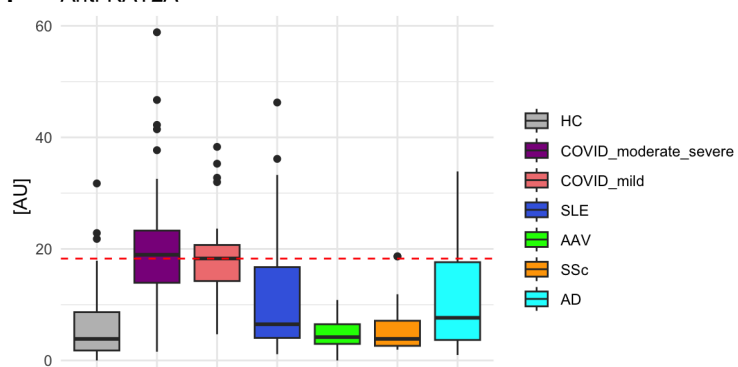
E

Anti-BCORP1

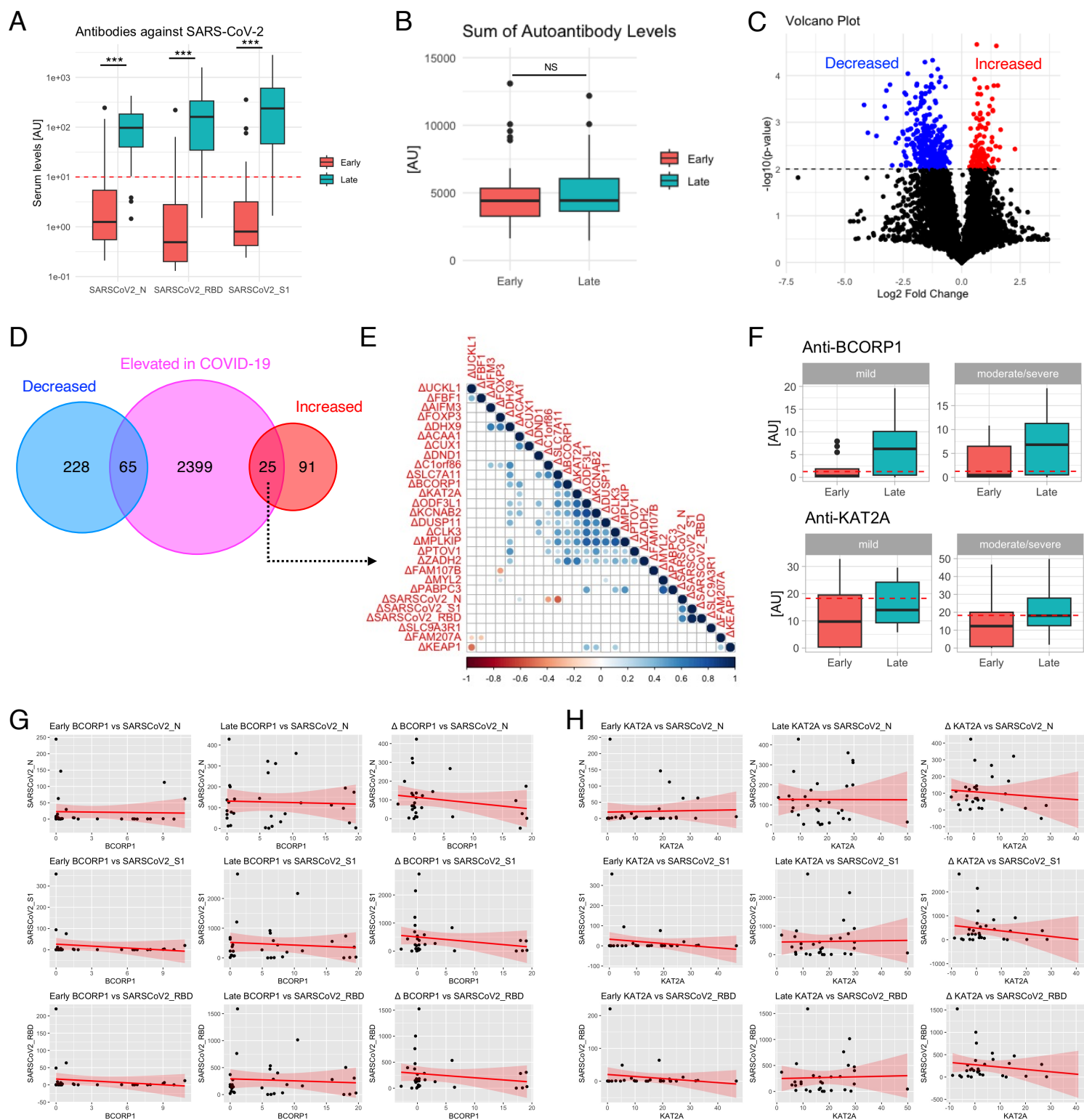


F

Anti-KAT2A



Matsuda KM et al.
Figure 4



Matsuda KM et al.
Figure 5