

1 A transcription regulatory network within the ACE2 locus may promote a pro-viral
2 environment for SARS-CoV-2 by modulating expression of host factors.

3 Tayaza Fadason^{1,2¶}, Sreemol Gokuladhas^{1¶}, Evgeniia Golovina^{1¶}, Daniel Ho¹, Sophie Farrow¹,
4 Denis Nyaga¹, Hong Pan³, Neerja Karnani^{3,4}, Conroy Wong⁵, Antony Cooper^{6,7}, William
5 Schierding^{1,2}, Justin M. O’Sullivan^{1,2,8†}.

6 ¹ The Liggins Institute, The University of Auckland, Auckland, New Zealand

7 ² Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand

8 ³ Singapore Institute for Clinical Sciences (SICS), Agency for Science, Technology and Research (A*STAR),
9 Singapore, Singapore

10 ⁴ Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore (NUS),
11 Singapore

12 ⁵ Respiratory Medicine, Middlemore Hospital and University of Auckland, Auckland, New Zealand

13 ⁶ Australian Parkinsons Mission, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

14 ⁷ St Vincent’s Clinical School, UNSW Sydney, Sydney, New South Wales, Australia

15 ⁸ Brain Research New Zealand, The University of Auckland, Auckland, New Zealand.

16 ¶ Equal contributors

17 † Corresponding author: Justin M. O’Sullivan, The Liggins Institute, The University of Auckland, Phone:+64 9
18 9239868, justin.osullivan@auckland.ac.nz

19

20 Abstract

21 *Introduction:* A novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was
22 recently identified as the pathogen responsible for the COVID-19 outbreak. SARS-CoV-2
23 triggers severe pneumonia, which leads to acute respiratory distress syndrome and death in
24 severe cases. As reported, SARS-CoV-2 is 80% genetically identical to the 2003 SARS-CoV
25 virus. Angiotensin-converting enzyme 2 (ACE2) has been identified as the main receptor for
26 entry of both SARS-CoV and SARS-CoV-2 into human cells. ACE2 is normally expressed in
27 cardiovascular and lung type II alveolar epithelial cells, where it positively modulates the RAS
28 system that regulates blood flow, pressure, and fluid homeostasis. Thus, virus-induced
29 reduction of ACE2 gene expression is considered to make a significant contribution to severe
30 acute respiratory failure. Chromatin remodelling plays a significant role in the regulation of
31 ACE2 gene expression and the activity of regulatory elements within the genome.

32 *Methods:* Here, we integrated data on physical chromatin interactions within the genome
33 organization (captured by Hi-C) with tissue-specific gene expression data to identify spatial
34 expression quantitative trait loci (eQTLs) and thus regulatory elements located within the
35 ACE2 gene.

36 *Results:* We identified regulatory elements within *ACE2* that control the expression of *PIR*,
37 *CA5B*, and *VSP13C* in the lung. The gene products of these genes are involved in inflammatory
38 responses, *de novo* pyrimidine and polyamine synthesis, and the endoplasmic reticulum,
39 respectively.

40 *Conclusion:* Our study, although limited by the fact that the identification of the regulatory
41 interactions is putative until proven by targeted experiments, supports the hypothesis that viral
42 silencing of *ACE2* alters the activity of gene regulatory regions and promotes an intra-cellular
43 environment suitable for viral replication.

44

45 Keywords: COVID-19, SARS-CoV-2, ARDS, ACE2

46 **Introduction**

47 Within months of the first reports [1], the COVID-19 outbreak has become a pandemic
48 infecting and killing thousands of people worldwide [2]. COVID-19 is an infectious disease
49 associated with acute respiratory distress syndrome (ARDS) that is caused by SARS-CoV-2, a
50 Betacoronavirus that is 80% identical to the SARS-CoV virus [3]. Betacoronaviruses,
51 including SARS-CoV, Murine Hepatic Virus (MHV), and SARS-CoV-2, utilize the ACE2
52 protein for cell entry [4,5]. The Spike protein on SARS-CoV-2 has a 10 to 20 fold higher
53 affinity for the ACE2 protein than its SARS-CoV homologue [3,6].

54 The ACE2 protein is highly expressed in cardiovascular and lung type II alveolar epithelial
55 cells [3,7,8], where ACE2 is a primary modulator of the renin–angiotensin (RAS) system that
56 regulates blood flow, pressure and fluid homeostasis [9]. The ACE2 protein and the products
57 of the reactions it catalyses have also been implicated in immune responses and anti-
58 inflammatory pathways [10–12].

59 SARS-CoV infection reduces *ACE2* gene expression and this is thought to contribute to severe
60 acute respiratory failure [4] by triggering an imbalance in the RAS system that causes a loss of
61 fluid homeostasis, induces inflammatory responses [10,13,14], and results in severe acute
62 injury in heart and lung [3,15,16]. As mentioned above, both SARS-CoV and SARS-CoV-2
63 utilize the ACE2 protein for cell entry. Poor prognoses in elderly SARS-CoV-2 patients (≥ 65
64 years old) are frequently associated with a pre-existing reduction in *ACE2* expression and
65 imbalance in ACE2-related host derived pathways [17,18]. *ACE2* is an X-linked gene whose
66 expression is regulated by chromatin structure. Brg1, a chromatin remodeler, and the FoxM1
67 transcription factor recognize the *ACE2* promoter and reduce expression through a mechanism
68 involving structural chromatin changes [19]. This control is complex, as illustrated by the

69 finding that *ACE2* gene escapes X chromosome-inactivation and shows a heterogeneous sex
70 bias that is tissue dependent [20].

71 Chromatin structure in the nucleus involves non-random folding of DNA on different scales
72 [21]. This folding and the resulting contacts that form are dynamic, and can be disrupted (*e.g.*
73 by genetic variation) leading to altered enhancer-promoter interactions that result in changes
74 in gene expression [22]. Changing chromatin structure rewires interactions between regulatory
75 elements and the genes they control. Theoretically and practically, each component of this
76 change contributes to the observed pathogenesis [23,24], and can lead to developmental
77 disorders[25] and cancer [26–28].

78 Virus-induced chromatin changes at the *ACE2* locus could induce expression changes in
79 additional genes regulated by elements located within this locus and thus may alter/modulate
80 host factors important for SARS-CoV-2 replication. How can you identify the elements within
81 a gene regulatory network like this? One approach to identify the networks that form between
82 regulatory elements and the genes they control is to use information on the physical interactions
83 that are captured occurring between the elements. Physical interactions between two sites can
84 be captured and identified using Hi-C [29,30]. We have used this insight to develop a
85 discovery-based pipeline (CoDeS3D; S1 Fig) [23]. Our approach uses genetic variation (*e.g.*
86 single nucleotide polymorphisms) to identify changes in gene expression and thus determine if
87 a region that physically contacts a gene contains a regulatory element. This enables the rapid
88 identification of the regulatory networks that form in cells and tissues (*e.g.* [23,31]).

89 We hypothesized that *ACE2* and its flanking region contained regulatory elements that
90 coordinate the expression of other genes, and that virus induced chromatin changes at *ACE2*
91 inadvertently modulate host factors that promote viral replication. Here we undertook an in-

92 depth characterization of the regulatory control regions within *ACE2* and their activity in lung
93 tissue. Regulatory elements located within *ACE2* affect the expression levels of the *PIR* and
94 *CA5B* genes. *PIR* and *CA5B* are involved in NF- κ B regulation and pyrimidine synthesis,
95 respectively. *VSP13C*, encoding a factor required for late stage endosome maturation, is also
96 controlled by a putative enhancer located in intron 11 of *BMX*, adjacent to *ACE2*. We propose
97 that *ACE2* repression by SARS-CoV-2 trips a chromatin-based switch that coordinates the
98 activity of these regulatory elements and thus the genes they control. Collectively, these
99 changes inadvertently lead to the development of a pro-viral replication environment.

100

101

102

103 **Methods**

104 **Identification of SNPs in the *ACE* locus**

105 We selected all common single nucleotide polymorphisms (SNPs) from dbSNP (build153)
106 with a minor allele frequency (MAF) > 1% that were located within chrX:15,519,996-
107 15,643,106, which included the *ACE2* gene and its flanking region (hereafter *ACE2* locus).
108 SNP positions are as reported for the human genome build hg38 release 75 (GRCh38).

109 **Identification of tissue-specific SNP-gene spatial relationships in the *ACE* locus**

110 We used the CoDeS3D algorithm [23] to identify putative spatial regulatory interactions for all
111 SNPs at the *ACE2* locus (S1 Fig). CoDeS3D integrates data on physical chromatin interactions
112 within the genome organization (captured by Hi-C) with tissue-specific gene expression data
113 to identify spatial expression quantitative trait loci (eQTLs). To get lung-specific spatial
114 connections, we identified SNP-gene pairs across lung-specific Hi-C libraries using published
115 data for IMR90, A549, and NCI-H460 cell lines and lung tissue (GEO accession numbers
116 GSE35156, GSE43070, GSE63525, GSE105600, GSE105725, GSE92819, GSE87112, S1
117 Table). We then queried GTEx for eQTL associations with lung tissue (dbGaP Accession
118 phs000424.v8.p2, <https://gtexportal.org/home/>). The age of the GTEx lung sample donors
119 peaks between 50-60 years (S2 Fig). SNPs were assigned to the appropriate Hi-C restriction
120 fragments by digital digestion of the hg38 reference genome (matching the restriction enzyme
121 from the Hi-C libraries: *Mbo*I or *Hind*III). All SNP-fragments were queried against the Hi-C
122 databases to identify the distal DNA fragments with which they physically interact. For each
123 distal fragment, which overlapped a gene coding region, a SNP-gene spatial connection was
124 confirmed. There was no binning or padding around restriction fragments to obtain gene
125 overlap. Spatial tissue-specific SNP-gene pairs with significant eQTLs (both cis-acting [$<1\text{Mb}$

5

126 between the SNP and gene] and trans-acting eQTLs [$>1\text{Mb}$ between the SNP and gene or on
127 different chromosomes]; FDR adjusted $p < 0.05$) within the lung were subsequently identified
128 by querying GTEx v8 lung tissue (UBERON:0008952).

129 **URLs**

130 GEO database: <https://www.ncbi.nlm.nih.gov/geo/>

131 CoDeS3D pipeline: <https://github.com/Genome3d/codes3d-v2>

132 GTEx Portal: <https://gtexportal.org/home/>

133 GUSTO study: <http://www.gusto.sg/>

134 **Data and code availability**

135 All python and R scripts used for data analysis and visualization are available at
136 <https://github.com/Genome3d/ACE2-regulatory-network>. R version 3.5.2 and RStudio version
137 1.2.5033 were used for all R scripts. All python scripts used Python 3.7.6.

138

139

140

141

142

143

144

145 **Results**

146 **The *ACE2* locus harbors regulatory variants that control SARS-CoV-2 relevant cellular**
147 **functions:**

148 We tested 367 common SNPs located across the *ACE2* locus (chrX: 15,519,996-15,643,106)
149 for their potential to act as spatial eQTLs. None of the common SNPs we tested affected *ACE2*
150 expression levels in lung tissue (S2 Table).

151 The wider *ACE2* locus (chrX: 15,519,996-15,643,106; GRCh38/hg38) sits within a
152 topologically associating domain (TAD) that is conserved across some tissues, *e.g.* IMR90 (Fig
153 1A). Therefore, it was not surprising that we identified control elements within this *ACE2* locus
154 (Fig 1A). The distribution of targets for the putative control elements we identified is consistent
155 with previous studies that show that while the majority of significant eQTLs fall within 100 kb
156 of the transcription start site of a gene, only 60% of all eQTLs are upstream of the gene they
157 regulate[32]. Notable amongst the elements we identified are long distance *trans*-regulatory
158 interactions involving: 1) rs1399200:*VPSI3C* (chr15:61,852,389-62,060,473; encodes
159 vacuolar protein sorting-associated protein 13C); and 2) rs6632680:*PHKA2* (chrX:18,892,300-
160 18,984,598; encodes phosphorylase kinase regulatory subunit alpha 2) (S2 Table).

161 We identified eighty genetic variants within the *ACE2* locus as *cis*-acting spatial eQTLs that
162 physically modulate the expression of genes *PIR* (encodes Pirin), *CA5BP1* (a pseudogene of
163 *CA5B*), and *CA5B* (encodes mitochondrial carbonic anhydrase) in lung tissues (S2 Table).
164 Fifty-eight SNPs located across the region are associated with increased expression of *PIR*
165 ($\log_2[\text{aFC, allelic fold change}] 0.462 \pm 0.07$) consistent with the elements they mark repressing
166 *PIR* transcription. Eighteen SNPs located across the region are associated with decreased
167 expression of *CA5B* ($\log_2[\text{aFC}] -0.257 \pm 0.005$) consistent with the elements they mark

168 enhancing *CA5B* transcription. These variants occurred in two clusters: 1) within the *ACE2*
169 gene; and 2) within the *CLTRN* (*TMEM27*) gene – a known homologue of *ACE2*. Expression
170 of *CA5BP1*, a pseudogene of *CA5B*, was also repressed ($\log_2[\text{aFC}] -0.21 \pm 0.01$) by 6 SNPs
171 within the *ACE2* locus. Within *ACE2* itself there were only control regions for the *PIR* and
172 *CA5B* genes (Fig 1B).

173 The common variants that we tested show an unusual ancestry associated pattern of minor
174 allele frequencies (Fig 1B). Specifically, the East Asian population (1K Genomes project)
175 displays little variation across the bulk of the variants we analyzed. This observation is
176 supported by measures of genetic diversity (F_{ST}) between the Indian, Chinese and Malay
177 populations within the Growing Up in Singapore Towards healthy Outcomes (GUSTO) cohort
178 (S3 Table). However, this pattern breaks down at several positions across the *ACE2* gene
179 (including rs4646142, rs2285666, and rs2106809, which show significant selection towards
180 the reference allele) in all of the tested populations, indicating potential selective pressure at
181 these loci (Fig 1B). Notably, two of these variants alter potential transcription factor binding
182 sites (*i.e.* rs2285666 alters HNF1, and Ncx motifs, rs2106809 alters a CEBPB motif; S4 Table).
183 All three variants (rs4646142, rs2285666, and rs2106809) have previously been associated
184 with allele, sex and ethnicity specific impacts on hypertension, blood pressure, hypertrophic
185 cardiomyopathy, type 2 diabetes, myocardial infarction (reviewed in[33]). Moreover, the
186 CEBPB motif is recognized by the CCAAT enhancer binding protein- β which has been
187 implicated in inflammatory responses in lung carcinoma cells [34].

188

189 **Discussion**

190 We identified transcription regulatory elements for *CA5B* and *PIR* that are active in lung tissue
191 and are located within the *ACE2* gene. We also identified a transcription regulatory element
192 (located in the *BMX* gene, adjacent to *ACE2*) for the *PIR* and *VSP13C* genes. It is sterically
193 impossible for a single DNA sequence to simultaneously be transcribed and regulate another
194 gene through a physical connection. Therefore, we propose that SARS-CoV-2-induced
195 chromatin-dependent repression of *ACE2* expression in lung enables the regulatory sites,
196 repressing *PIR* and activating *CA5B*, to exhibit increased functionality in infected cells (Fig 2).
197 We hypothesize that this regulatory change extends to coordinate changes in the expression of
198 *VSP13C* and *PHKA2* in ways that promote viral proliferation. This host regulatory network has
199 not evolved to benefit the virus but rather, these regulatory changes inadvertently produce an
200 environment advantageous for the virus.

201 The *CA5B* gene encodes a mitochondrial carbonic anhydrase that catalyzes the reversible
202 hydration of CO₂ in the lung. This reaction is important in mitochondria as it supplies HCO₃⁻
203 ions required by pyruvate carboxylase for gluconeogenesis, and by carbamoyl phosphate
204 synthase 1 (CSP1) for pyrimidine biosynthesis (Fig 2B) [38–40]. Pyrimidines are important
205 host factors critical for viral genomic replication, mRNA synthesis for protein translation, and
206 phospholipid synthesis [41]. Inhibiting *de novo* pyrimidine biosynthesis impacts on SARS-
207 CoV-2 replication [18]. CSP1 additionally produces a precursor for the biosynthesis of
208 polyamines, small aliphatic molecules that play important roles in virus replication. Inhibition
209 of polyamine biosynthesis significantly impaired replication of the Middle East Respiratory
210 Syndrome [MERS] coronavirus [42]. Targeted inhibition of CA5B encoded carbonic
211 anhydrase might therefore decrease levels of critical host factors pyrimidines and polyamines
212 - critical host factors needed for SARS-CoV-2 replication. Intriguingly, similarities in the
213 pathologies of SARS-CoV-2 infection and high altitude pulmonary edema (HAPE) have led to
9

214 the suggestion that carbonic anhydrase inhibitors could be used to treat or prevent Covid-19
215 infection [43].

216 Interleukin expression is responsible for irreversible, pathological changes associated with
217 SARS-CoV infection in the lung (*e.g.* [44]). Human coronavirus has been shown to fine-tune
218 NF- κ B signaling [45]. *PIR* encodes a non-heme iron binding protein that is a redox switch that
219 modulates the binding of p65 (RelA) to NF- κ B responsive promoters [46]. NF- κ B regulates
220 multiple immune function aspects, including the production of pro-inflammatory
221 cytokines[47]. Therefore, it is notable that repressor regulatory sequences for *PIR* sit within
222 the *ACE2* gene (Fig 2A). We postulate that the chromatin modifications that silence *ACE2*
223 expression upon early stage infection activate the *PIR* repressor (Fig 2B). This reduces
224 responsiveness of NF- κ B, and thereby delays the expected and needed anti-viral response.
225 Reduction in *PIR* expression would also reduce the impact of any changes to intra-cellular
226 redox state caused by that SARS-CoV-2 infection, however little is known and future
227 experiments are required to clarify this.

228 The enveloped Betacoronaviruses (MHV, SARS-CoV, SARS-CoV-2) gain entry to the cell
229 through the endo/lysosomal pathway and require late endosomal maturation for fusion [48].
230 Therefore, it is interesting to speculate on the impact of coordinated changes to *VPS13C*
231 expression. The VPS13 family are endoplasmic reticulum associated lipid transporters.
232 *VPS13C* is proposed to act as a lipid transporter at organelle contact sites between (i) the
233 endoplasmic reticulum (ER) and endolysosomes, and (ii) the ER and lipid droplets, where it
234 transfers lipids, potentially bulk lipid transfer, between organelles to maintain lipid
235 homeostasis and organelle functionality [49]. Increased *VPS13C* expression is predicted to
236 increase the extent of contact and lipid transfer between these organelles. This in turn could

237 enhance the virus's replication capacity and pathogenesis, as the ER plays both a physical and
238 functional central role in the virus's capacity to replicate and form new viral progeny.
239 Moreover, SARS-CoV extensively reorganizes the host cell's membranes infrastructure to
240 produce a reticulo-vesicular network of modified ER to coordinate its replication cycle [50].
241 Alterations to ER-lipid droplet contacts mediated by VPS13C could support the virus's
242 required expansion and re-organization of ER membranes by altering lipid flow through the
243 ER [51]. Notably, cells infected with Hepatitis C virus (HCV), a positive-strand RNA virus
244 like SARS-CoV-2, contain ER-derived membraneous structures that contain significantly high
245 levels of cholesterol, despite the ER of uninfected cells possessing relatively low cholesterol
246 levels [52]. Increased VPS13C-mediated ER-endolysosomal contact sites could increase the
247 capacity of endocytosed dietary cholesterol to be delivered to the ER and enhance the virus's
248 ability to replicate. Pharmacological impairment of endolysosomal cholesterol efflux reduced
249 HCV replication, [52] suggesting another possible therapeutic approach for investigation to
250 slow SARS-CoV-2 replication. ER stress, impacted by changes to *VSP13C* expression, may
251 also contribute to late infection stage NF- κ B activation (reviewed in [53]).

252 The significance of the putative enhancer for *PHKA2*, which encodes the phosphorylase kinase
253 regulatory subunit alpha 2, is unclear. Mutations in this gene have been linked to glycogen
254 storage disorders and glucose metabolism. Thus, linkages can be drawn to the increased
255 expression of *CA5B*, which impacts on gluconeogenesis. Notably, *PHKA2* was downregulated
256 in plasma from individuals with hepatocellular-carcinoma caused by HCV infections [54].
257 Theoretically, chromatin remodeling in response to SARS-COV-2 infection could down-
258 regulate *PHKA2* expression. However, there is a paucity of information linking this gene to
259 viral infections or the lung and this conclusion requires additional experimental support.

260 SARS-CoV is known to repress *ACE2* expression [4]. *ACE2* regulation involves chromatin
261 remodeling and structural chromatin changes [19]. Several of the regions that we identified
262 overlapped or were adjacent to CTCF binding sites (*e.g.* rs1399200, which regulates *PIR* and
263 *VPS13C*; rs6629111, which regulates *CA5B*; and sites [rs714205, rs1514280, rs4240157 and
264 rs4646131] within *ACE2*[33]). We also note that the regulatory sites we identified included
265 transcription factor binding sites for Ap-1, RXRA (a DNA-binding receptor involved in host-
266 virus interactions) [55], GR or NR3C1 (a regulator of inflammation in asthma and COPD) [56],
267 Pou2f2 (trans-activator of NR3C1)[57], and P300 (a chromatin modifier; S4 Table) [58].
268 Expression data within the search-based exploration of expression compendium (SEEK)
269 supports a strong co-expression relationship between *ACE2* and *PIR* (lung cancer, ovarian
270 tumour) and a weaker association with *CA5B* (ovarian cancer) (<http://seek.princeton.edu/> [59]).
271 However, the possible mechanism(s) that link *ACE2* silencing to alterations associated with
272 these regulatory regions remains unknown until empirically determined in lung cells in the
273 presence/absence of real or simulated viral infection.

274 Whilst this study is novel and uses empirically derived data in the analyses, the observations
275 are limited by the fact that the identification of the regulatory interactions is only putative until
276 proven by targeted experiments. Ideally, the Hi-C datasets should be derived from matched
277 tissues prior and post SARS-CoV-2 infection. Finally, the GTEx database has recognized
278 limitations, including the ethnic diversity of the samples. These limitations will form the basis
279 of future studies.

280

281 **Conclusion**

282 We identified putative regulatory regions in and surrounding *ACE2* that regulate the expression
283 of *PIR*, *CA5B*, and *VSP13C* in the lung. We contend that viral induced chromatin-dependent
284 repression of the *ACE2* gene increases the activity of these regulatory sites and promotes an
285 intra-cellular environment suitable for viral replication. The altered gene products represent
286 new targets for anti-SARS-CoV-2 therapeutics.

287

288

289 **Acknowledgements**

290 The authors would like to thank the Genomics and Systems Biology Group (Liggins Institute),
291 Mark Hampton, and Elizabeth Ledgerwood for useful discussions. We would like to thank the
292 funders of GTEEx Project - common Fund of the Office of the Director of the National Institutes
293 of Health, and by National Cancer Institute, National Human Genome Research Institute,
294 National Heart, Lung, and Blood Institute, National Institute on Drug Abuse, National Institute
295 of Mental Health, National Institute of Neurological Disorders and Stroke. The authors thank
296 the GUSTO study group for comments on the manuscript. This work has been released as a
297 pre-print.

298 **Conflict of interest statement**

299 The authors declare that the research was conducted in the absence of any commercial or
300 financial relationships that could be construed as a potential conflict of interest.

301 **Author contributions**

302 TF, EG, SG and WS performed analyses and co-wrote the manuscript. DH, DN, SF, and AC
303 performed literature searches and co-wrote the manuscript. HP and NK provide FST data from
304 GUSTO and commented on the manuscript. CW reviewed the findings and commented on the
305 manuscript. JOS led the study and co-wrote the manuscript.

306 **Funding**

307 SG and DH are recipients of scholarships funded by a Ministry of Business, Innovation and
308 Employment Catalyst grant (New Zealand- Australia LifeCourse Collaboration on Genes,
309 Environment, Nutrition and Obesity; UOAX1611) to JOS. JOS and WS are funded by a Royal
310 Society of New Zealand Marsden Fund [Grant 16-UOO-072]. JOS and TF are funded by an
311 HRC explorer grant (HRC19/774) to JOS. DN was supported by the Sir Colin Giltrap Liggins
312 Institute Scholarship fund. AC received grant funding from the Australian government.
313 GUSTO study is supported by Singapore National Research Foundation under its Translational
314 and Clinical Research (TCR) Flagship Program administered by the Singapore Ministry of
315 Health's National Medical Research Council (NMRC/TCR/004-NUS/2008; NMRC/TCR/012-
316 NUHS/2014). SNP variant analysis in GUSTO cohort was supported by Industry Alignment
317 Fund – Pre-positioning Programme (IAF-PP H17/01/a0/005, available to NK. The funders had
318 no role in study design, data collection and analysis, decision to publish, or preparation of the
319 manuscript.

320

321

322

323

324

325 References

- 326 1. Chen J, Jiang Q, Xia X, Liu K, Yu Z, Tao W, et al. Individual Variation of the SARS-
327 CoV2 Receptor ACE2 Gene Expression and Regulation. 2020.
- 328 2. WHO. Coronavirus disease 2019. World Heal Organ. 2020;2019: 2633.
329 doi:10.1001/jama.2020.2633
- 330 3. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin,
331 transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak
332 - an update on the status. *Mil Med Res.* 2020;7: 11. doi:10.1186/s40779-020-00240-0
- 333 4. Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, et al. A crucial role of angiotensin
334 converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nat Med.*
335 2005;11: 875–879. doi:10.1038/nm1267
- 336 5. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al.
337 SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a
338 Clinically Proven Protease Inhibitor. *Cell.* 2020; 1–10. doi:10.1016/j.cell.2020.02.052
- 339 6. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM
340 structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020;1263:
341 1260–1263. doi:10.1126/science.abb2507
- 342 7. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling
343 of ACE2, the putative receptor of Wuhan 2019-nCov. *bioRxiv.* 2020;
344 2020.01.26.919985. doi:10.1101/2020.01.26.919985
- 345 8. Harmer D, Gilbert M, Borman R, Clark KL. Quantitative mRNA expression profiling
346 of ACE 2, a novel homologue of angiotensin converting enzyme. *FEBS Lett.* 2002;532:
347 107–110. doi:10.1016/S0014-5793(02)03640-2
- 348 9. Tikellis C, Thomas MC. Angiotensin-converting enzyme 2 (ACE2) is a key modulator
349 of the renin angiotensin system in health and disease. *Int J Pept.* 2012;2012.
350 doi:10.1155/2012/256294
- 351 10. Rodrigues Prestes TR, Rocha NP, Miranda AS, Teixeira AL, Simoes-e-Silva AC. The
352 Anti-Inflammatory Potential of ACE2/Angiotensin-(1-7)/Mas Receptor Axis: Evidence
353 from Basic and Clinical Research. *Curr Drug Targets.* 2017;18: 1301–1313.
354 doi:10.2174/1389450117666160727142401
- 355 11. Sodhi CP, Nguyen J, Yamaguchi Y, Werts AD, Lu P, Ladd MR, et al. A Dynamic
356 Variation of Pulmonary ACE2 Is Required to Modulate Neutrophilic Inflammation in
357 Response to *Pseudomonas aeruginosa* Lung Infection in Mice . *J Immunol.* 2019;203:
358 3000–3012. doi:10.4049/jimmunol.1900579
- 359 12. Steck AK, Dong F, Wong R, Fouts A, Liu E, Romanos J, et al. Improving prediction of
360 type 1 diabetes by testing non-HLA genetic variants in addition to HLA markers. *Pediatr*
361 *Diabetes.* 2014;15: 355–362. doi:10.1111/pedi.12092
- 362 13. Hashimoto T, Perlot T, Rehman A, Trichereau J, Ishiguro H, Paolino M, et al. ACE2
363 links amino acid malnutrition to microbial ecology and intestinal inflammation. *Nature.*
364 2012;487: 477–481. doi:10.1038/nature11228

- 365 14. Keidar S, Strizevsky A, Raz A, Gamliel-Lazarovich A. ACE2 activity is increased in
366 monocyte-derived macrophages from prehypertensive subjects. *Nephrol Dial*
367 *Transplant.* 2007;22: 597–601. doi:10.1093/ndt/gfl632
- 368 15. Imai Y, Kuba K, Rao S, Huan Y, Guo F, Guan B, et al. Angiotensin-converting enzyme
369 2 protects from severe acute lung failure. *Nature.* 2005;436: 112–116.
370 doi:10.1038/nature03712
- 371 16. Imai Y, Kuba K, Penninger JM. The discovery of angiotensin-converting enzyme 2 and
372 its role in acute lung injury in mice. *Exp Physiol.* 2008;93: 543–548.
373 doi:10.1113/expphysiol.2007.040048
- 374 17. Xudong X, Junzhu C, Xingxiang W, Furong Z, Yanrong L. Age- and gender-related
375 difference of ACE2 expression in rat lung. *Life Sci.* 2006;78: 2166–2171.
376 doi:10.1016/j.lfs.2005.09.038
- 377 18. Xiong R, Zhang L, Li S, Sun Y, Ding M, Wang Y, et al. Novel and potent inhibitors
378 targeting DHODH, a rate-limiting enzyme in de novo
379 pyrimidine biosynthesis, are broad-spectrum antiviral against RNA viruses including
380 newly emerged coronavirus SARS-CoV-2. *bioRxiv.* 2020; 2020.03.11.983056.
381 doi:10.1101/2020.03.11.983056
- 382 19. Yang J, Feng X, Zhou Q, Cheng W, Shang C, Han P, et al. Pathological Ace2-to-Ace
383 enzyme switch in the stressed heart is transcriptionally controlled by the endothelial
384 Brg1–FoxM1 complex. *Proc Natl Acad Sci.* 2016;113: E5628–E5635.
385 doi:10.1073/pnas.1525078113
- 386 20. Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, et al. Landscape of
387 X chromosome inactivation across human tissues. *Nature.* 2017;550: 244–248.
388 doi:10.1038/nature24265
- 389 21. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A
390 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin
391 Looping. *Cell.* 2014;159: 1665–1680. doi:10.1016/j.cell.2014.11.021
- 392 22. Meier M, Grant J, Dowdle A, Thomas A, Gerton J, Collas P, et al. Cohesin facilitates
393 zygotic genome activation in zebrafish. *Development.* 2018;145: dev156521.
394 doi:10.1242/dev.156521
- 395 23. Fadason T, Schierding W, Lumley T, O’Sullivan JM. Chromatin interactions and
396 expression quantitative trait loci reveal genetic drivers of multimorbidities. *Nat*
397 *Commun.* 2018;9: 5198. doi:10.1038/s41467-018-07692-y
- 398 24. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic
399 Inheritance. *Cell.* 2019;177: 1022-1034.e6. doi:10.1016/j.cell.2019.04.014
- 400 25. Paulsen J, Liyakat Ali TM, Nekrasov M, Delbarre E, Baudement M-O, Kurscheid S, et
401 al. Long-range interactions between topologically associating domains shape the four-
402 dimensional genome during differentiation. *Nat Genet.* 2019;51: 835–843.
403 doi:10.1038/s41588-019-0392-0
- 404 26. Rivas MA, Melnick AM. Role of chromosomal architecture in germinal center B cells

- 405 and lymphomagenesis. *Current Opinion in Hematology*. Lippincott Williams and
406 Wilkins; 2019. pp. 294–302. doi:10.1097/MOH.0000000000000505
- 407 27. Yang M, Vesterlund M, Siavelis I, Moura-Castro LH, Castor A, Fioretos T, et al.
408 Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid
409 childhood acute lymphoblastic leukemia. *Nat Commun*. 2019;10: 1519.
410 doi:10.1038/s41467-019-09469-3
- 411 28. Achinger-Kawecka J, Valdes-Mora F, Luu PL, Giles KA, Caldon CE, Qu W, et al.
412 Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin
413 landscape in endocrine-resistant breast cancer. *Nat Commun*. 2020;11: 320.
414 doi:10.1038/s41467-019-14098-x
- 415 29. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A,
416 et al. Comprehensive mapping of long-range interactions reveals folding principles of
417 the human genome. *Science* (80-). 2009/10/10. 2009;326: 289–293. doi:326/5950/289
418 [pii]10.1126/science.1181369
- 419 30. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A
420 3D map of the human genome at kilobase resolution reveals principles of chromatin
421 looping. *Cell*. 2014;159: 1665–1680. doi:10.1016/j.cell.2014.11.021
- 422 31. Gokuladhas S, Schierding W, Cameron-Smith D, Wake M, Scotter E, O’Sullivan JM.
423 Shared regulatory pathways reveal novel genetic correlations between grip strength and
424 neuromuscular disorders. *Front Genet*. 2020;In Press. doi:10.3389/fgene.2020.00393
- 425 32. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue
426 Expression (GTEx) project. *Nat Genet*. 2013;45: 580–585. doi:10.1038/ng.2653
- 427 33. Burrell LM, Harrap SB, Velkoska E, Patel SK. The ACE2 gene: its potential as a
428 functional candidate for cardiovascular disease. *Clin Sci*. 2013;124: 65–76.
429 doi:10.1042/CS20120269
- 430 34. Armstrong DA, Phelps LN, Vincenti MP. CCAAT enhancer binding protein- β regulates
431 matrix metalloproteinase-1 expression in interleukin-1 β -stimulated A549 lung
432 carcinoma cells. *Mol Cancer Res*. 2009;7: 1517–1524. doi:10.1158/1541-7786.MCR-
433 09-0082
- 434 35. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: a
435 web-based browser for visualizing 3D genome organization and long-range chromatin
436 interactions. *Genome Biol*. 2018;19: 151. doi:10.1186/s13059-018-1519-9
- 437 36. Hinrichs a S, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The
438 UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34: D590-8.
439 doi:10.1093/nar/gkj144
- 440 37. Ward LD, Kellis M. HaploReg v4: Systematic mining of putative causal variants, cell
441 types, regulators and target genes for human complex traits and disease. *Nucleic Acids*
442 *Res*. 2016;44: D877–D881. doi:10.1093/nar/gkv1340
- 443 38. Ghandour MS, Parkkila AK, Parkkila S, Waheed A, Sly WS. Mitochondrial carbonic
444 anhydrase in the nervous system: Expression in neuronal and glial cells. *J Neurochem*.

- 445 2000;75: 2212–2220. doi:10.1046/j.1471-4159.2000.0752212.x
- 446 39. Adeniran C, Hamelberg D. Redox-Specific Allosteric Modulation of the
447 Conformational Dynamics of κ b DNA by Pirin in the NF- κ B Supramolecular Complex.
448 *Biochemistry*. 2017;56: 5002–5010. doi:10.1021/acs.biochem.7b00528
- 449 40. Kim J, Hu Z, Cai L, Li K, Choi E, Faubert B, et al. CPS1 maintains pyrimidine pools
450 and DNA synthesis in KRAS/LKB1-mutant lung cancer cells. *Nature*. 2017;546: 168–
451 172. doi:10.1038/nature22359
- 452 41. Löffler M, Fairbanks LD, Zameitat E, Marinaki AM, Simmonds HA. Pyrimidine
453 pathways in health and disease. *Trends Mol Med*. 2005;11: 430–7.
454 doi:10.1016/j.molmed.2005.07.003
- 455 42. Mounce BC, Cesaro T, Moratorio G, Hooikaas PJ, Yakovleva A, Werneke SW, et al.
456 Inhibition of Polyamine Biosynthesis Is a Broad-Spectrum Strategy against RNA
457 Viruses. Pfeiffer J, editor. *J Virol*. 2016;90: 9683–9692. doi:10.1128/JVI.01347-16
- 458 43. Solaimanzadeh I. Acetazolamide, Nifedipine and Phosphodiesterase Inhibitors:
459 Rationale for Their Utilization as Adjunctive Countermeasures in the Treatment of
460 Coronavirus Disease 2019 (COVID-19). *Cureus*. 2020;2019. doi:10.7759/cureus.7343
- 461 44. Yoshikawa T, Hill T, Li K, Peters CJ, Tseng C-TK. Severe Acute Respiratory Syndrome
462 (SARS) Coronavirus-Induced Lung Epithelial Cytokines Exacerbate SARS
463 Pathogenesis by Modulating Intrinsic Functions of Monocyte-Derived Macrophages and
464 Dendritic Cells. *J Virol*. 2009;83: 3039–3048. doi:10.1128/jvi.01792-08
- 465 45. Poppe M, Wittig S, Jurida L, Bartkuhn M, Wilhelm J, Müller H, et al. The NF- κ B-
466 dependent and -independent transcriptome and chromatin landscapes of human
467 coronavirus 229E-infected cells. Enjuanes L, editor. *PLOS Pathog*. 2017;13: e1006286.
468 doi:10.1371/journal.ppat.1006286
- 469 46. Liu F, Rehmani I, Esaki S, Fu R, Chen L, de Serrano V, et al. Pirin is an iron-dependent
470 redox regulator of NF- κ B. *Proc Natl Acad Sci*. 2013;110: 9722–9727.
471 doi:10.1073/pnas.1221743110
- 472 47. Liu T, Zhang L, Joo D, Sun S-C. NF- κ B signaling in inflammation. *Signal Transduct*
473 *Target Ther*. 2017;2: 17023. doi:10.1038/sigtrans.2017.23
- 474 48. Burkard C, Verheije MH, Wicht O, van Kasteren SI, van Kuppeveld FJ, Haagmans BL,
475 et al. Coronavirus Cell Entry Occurs through the Endo-/Lysosomal Pathway in a
476 Proteolysis-Dependent Manner. Perlman S, editor. *PLoS Pathog*. 2014;10: e1004502.
477 doi:10.1371/journal.ppat.1004502
- 478 49. Kumar N, Leonzino M, Hancock-Cerutti W, Horenkamp FA, Li P, Lees JA, et al.
479 VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact
480 sites. *J Cell Biol*. 2018;217: 3625–3639. doi:10.1083/jcb.201807019
- 481 50. Knoops K, Kikkert M, Worm SHE van den, Zevenhoven-Dobbe JC, van der Meer Y,
482 Koster AJ, et al. SARS-coronavirus replication is supported by a reticulovesicular
483 network of modified endoplasmic reticulum. *PLoS Biol*. 2008;6: e226.
484 doi:10.1371/journal.pbio.0060226

- 485 51. Lee J-Y, Cortese M, Haselmann U, Tabata K, Romero-Brey I, Funaya C, et al.
486 Spatiotemporal Coupling of the Hepatitis C Virus Replication Cycle by Creating a Lipid
487 Droplet- Proximal Membranous Replication Compartment. *Cell Rep.* 2019;27: 3602-
488 3617.e5. doi:10.1016/j.celrep.2019.05.063
- 489 52. Stoeck IK, Lee J-Y, Tabata K, Romero-Brey I, Paul D, Schult P, et al. Hepatitis C Virus
490 Replication Depends on Endosomal Cholesterol Homeostasis. *J Virol.* 2018;92.
491 doi:10.1128/JVI.01196-17
- 492 53. Fung TS, Liu DX. Coronavirus infection, ER stress, apoptosis and innate immunity.
493 *Front Microbiol.* 2014;5: 1–13. doi:10.3389/fmicb.2014.00296
- 494 54. Zekri ARN, Hafez MM, Bahnassy AA, Hassan ZK, Mansour T, Kamal MM, et al.
495 Genetic profile of Egyptian hepatocellular-carcinoma associated with hepatitis C virus
496 Genotype 4 by 15 K cDNA microarray: Preliminary study. *BMC Res Notes.* 2008;1.
497 doi:10.1186/1756-0500-1-106
- 498 55. Raney AK, Johnson JL, Palmer CN, McLachlan A. Members of the nuclear receptor
499 superfamily regulate transcription from the hepatitis B virus nucleocapsid promoter. *J*
500 *Virol.* 1997;71: 1058–1071. doi:10.1128/JVI.71.2.1058-1071.1997
- 501 56. Barnes PJ. Corticosteroid Resistance in Airway Disease. *Proc Am Thorac Soc.* 2004;1:
502 264–268. doi:10.1513/pats.200402-014MS
- 503 57. Préfontaine GG, Walther R, Giffin W, Lemieux ME, Pope L, Haché RJG. Selective
504 Binding of Steroid Hormone Receptors to Octamer Transcription Factors Determines
505 Transcriptional Synergism at the Mouse Mammary Tumor Virus Promoter. *J Biol*
506 *Chem.* 1999;274: 26713–26719. doi:10.1074/jbc.274.38.26713
- 507 58. Yoneyama M. Direct triggering of the type I interferon system by virus infection:
508 activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J.*
509 1998;17: 1087–1095. doi:10.1093/emboj/17.4.1087
- 510 59. Zhu Q, Wong AK, Krishnan A, Aure MR, Tadych A, Zhang R, et al. Targeted
511 exploration and analysis of large cross-platform human transcriptomic compendia. *Nat*
512 *Methods.* 2015;12: 211–214. doi:10.1038/nmeth.3249
- 513
- 514

515 **Figure legends**

516 **Fig 1. Elements located within and surrounding the *ACE2* locus regulate the lung-specific**
517 **expression of *PIR*, *CA5B*, *CA5BP1*, *VSP13C*, and *PHKA2*.** (A) Common genetic variants
518 (SNPs) located within the *ACE2* locus form spatial *cis*-acting regulatory interactions with *PIR*,
519 *CA5BP1*, and *CA5B* across sub-TAD boundaries on chrX:15,300,000-15,600,000. Inter-TAD
520 *trans*-acting interactions regulate *PHKA2* (3.2 Mb away) and *VSP13C* (located on chromosome
521 15). Visualization of TAD and chromatin interactions was performed using the 3D genome
522 browser (<http://yuelab.org/>)[35] and UCSC browser's interact tool
523 (<http://genome.ucsc.edu>)[36], respectively. (B) Within *ACE2*, MAFs for the SNPs that tag the
524 regulatory sites showed significant bias in four different populations (*i.e.* African [AFR], Ad
525 Mixed American [AMR], East Asian [ASN] and European [EUR]) at one *PIR* (rs714205) and
526 three *CA5B* regulatory sites (rs4646142, rs2285666, and rs2106809), consistent with selection
527 acting on these loci. MAFs were obtained from HaploReg v4.1
528 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) [37].

529

530 **Fig 2. Hypothesis: SARS-CoV-2 infection is associated with an *ACE2* dependent switch**
531 **that alters expression of proteins that promote an environment for viral proliferation in**
532 **the lung.** (A) In middle-aged non-infected individuals, control regions within *ACE2* are
533 capable of downregulating the expression of *PIR*, which is involved in the NF- κ B pathway.
534 Enhancer elements within *ACE2* are poised to upregulate *CA5B* expression, which encodes an
535 enzyme important for pyrimidine synthesis. In addition to this, an enhancer region within the
536 *BMX* gene (still within the same TAD) contributes to *VSP13C* regulation. (B) We hypothesize
537 that upon viral infection, SARS-COV-2 represses *ACE2* expression, which increases the
538 activity of the *PIR* repressor and *CA5B* enhancer. This results in a reduction in the production
539 of *PIR* - the redox switch necessary for NF- κ B activation, while also increasing pyrimidine
540 synthesis, which is necessary for viral replication.

541

542 **Supplementary Tables**

543 **S1 Table.** Lung-specific Hi-C libraries used in the analysis

544 **S2 Table.** Lung-specific spatial SNP-gene relationships in the *ACE* locus

545 **S3 Table.** Genetic diversity estimate (*Fst*) across *ACE2* in the Indian, Malay and Chinese
546 populations in the GUSTO cohort.

547 **S4 Table.** The common variants overlap DNA binding motifs. Data from Haploreg v4.1
548 (3/3/2020)

549

550

551 **Supplementary Figures**

552 **S1 Figure. The CoDeS3D algorithm used in this study.** Restriction fragments containing
553 SNPs located within the *ACE* locus (chrX:15,519,996-15,643,106) were identified. Lung-
20

554 specific Hi-C libraries were interrogated to identify genes in fragments that spatially interact
555 (in cis- and trans-) with SNP-containing fragments. The identified spatial SNP-gene pairs were
556 further used to query GTEx lung tissue (dbGaP Accession phs000424.v8.p2,
557 UBERON:0008952). The Benjamini-Hochberg FDR control algorithm was applied to adjust
558 the p values of the resulting eQTL associations to identify only significant (FDR < 0.05) lung-
559 specific SNP-gene spatial relationships in the *ACE* locus.

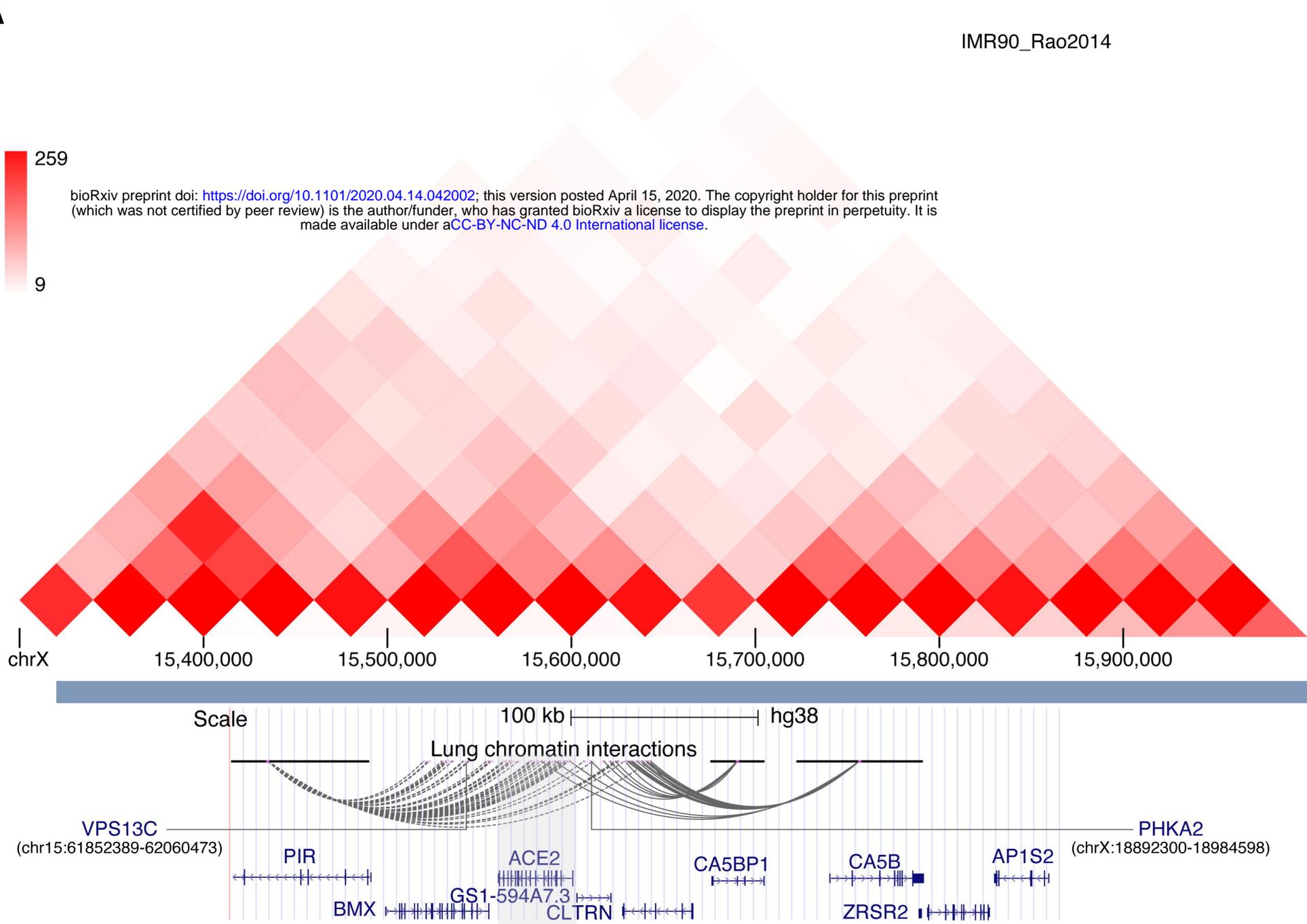
560

561 **S2 Figure. The eQTL data used in this study was obtained from lung samples taken from**
562 **middle-aged individuals.** To assess the correlation of genetic variation with the changes in
563 gene expression, the GTEx project (<https://gtexportal.org/home/>) collected and analysed lung
564 samples from donors who were densely genotyped. The age-distribution graph illustrates that
565 approximately 70% of the lung samples that were obtained were from donors aged between 50
566 and 60.

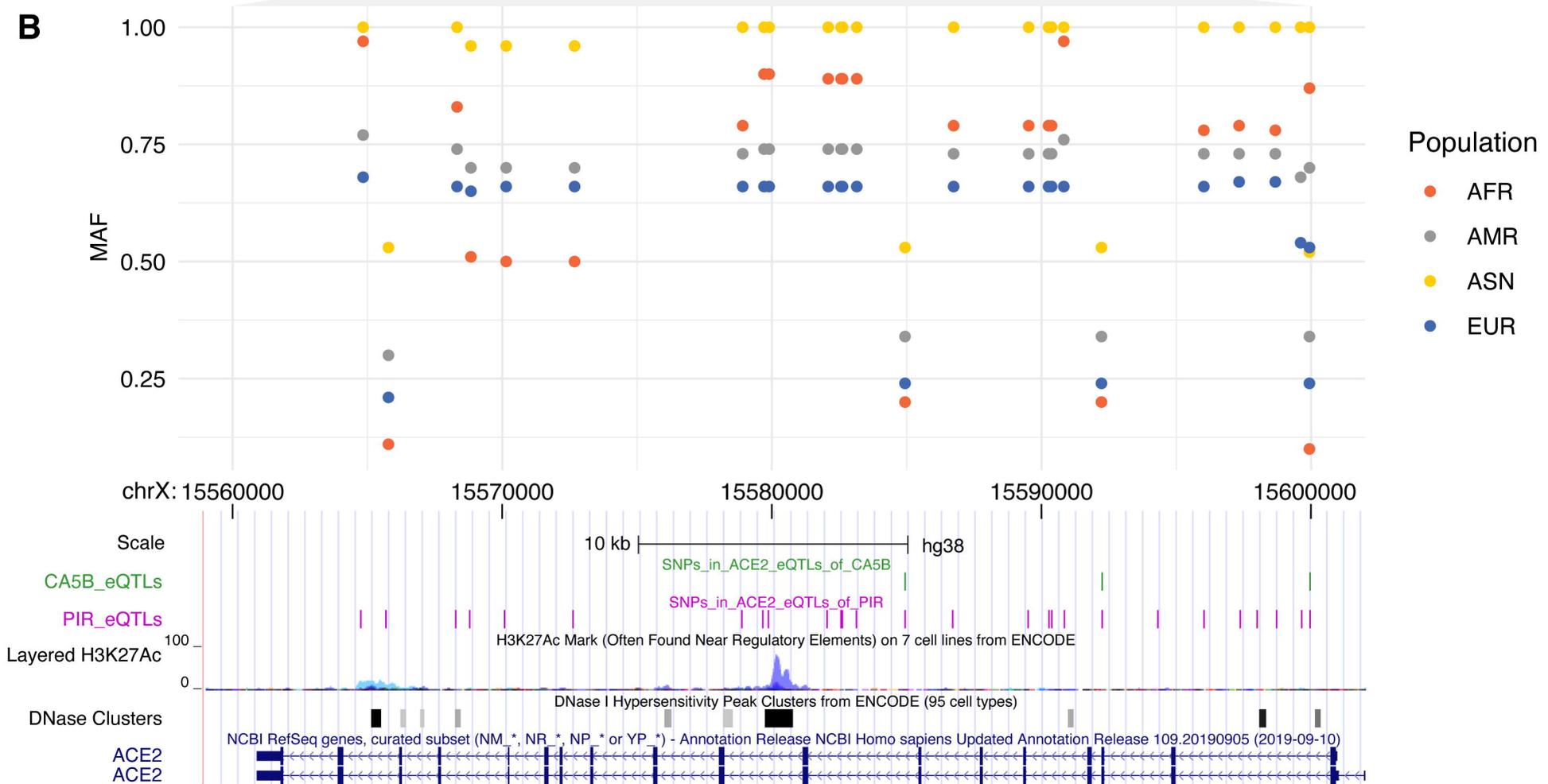
567



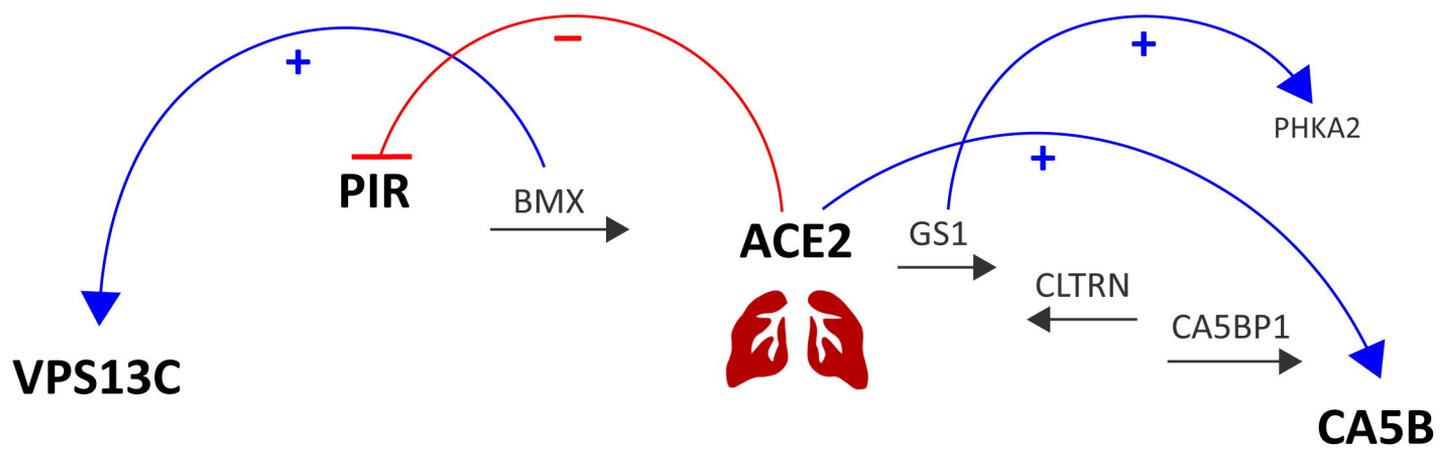
bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.14.042002>; this version posted April 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



B



a Middle-Aged Non-Infected Individual



b Early Stage Infection

