1 2	Alternat	ive splicing in the lung influences COVID-19 severity and respiratory diseases.			
3	Tomoko Nakanishi <sup>1–4</sup> ; Yossi Farjoun <sup>2</sup> ; Julian Willett <sup>2,5,6</sup> ; Richard J. Allen <sup>7</sup> , Beatriz Guillen-Guio <sup>7</sup> , Sirui				
4	Zhou <sup>1,5,6</sup>	Zhou <sup>1,5,6</sup> ; J. Brent Richards <sup>1,2,8–10</sup>			
5					
6	Affiliatio	ons:			
7	1.	Department of Human Genetics, McGill University, Montréal, Québec, Canada.			
8	2.	Centre for Clinical Epidemiology, Department of Medicine, Lady Davis Institute, Jewish General			
9	Hospital,	McGill University, Montréal, Québec, Canada.			
10	3.	Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate School of			
11	Medicine	e, Kyoto University, Kyoto, Japan.			
12	4.	Research Fellow, Japan Society for the Promotion of Science, Tokyo, Japan.			
13	5.	Quantitative Life Sciences Program, McGill University, Montréal, Canada.			
14	6.	McGill Genome Centre, McGill University, Montréal, Québec, Canada.			
15	7.	Department of Health Sciences, University of Leicester, Leicester, United Kingdom.			
16	8.	Department of Epidemiology, Biostatistics and Occupational Health, McGill University,			
17	Montréal, Québec, Canada.				
18	9.	Department of Twin Research, King's College London, London, United Kingdom.			
19	10.	5 Prime Sciences Inc, Montréal, Québec, Canada.			
20					
21	Corresp	onding author:			
22	J. Brent Richards				
23	Professo	or of Medicine, McGill University			
24	Senior Lecturer, King's College London (Honorary)				
25	Pavilion H-413, Jewish General Hospital				
26	3755 Côte-Ste-Catherine Montréal, Québec, Canada, H3T 1E2				
27	T: +1 514 340 8222 x24362 F: +1 514 340 7529				
28 20	E: brent.	richards@mcgill.ca www.mcgill.ca/genepi			
29 30 31	Key wor	ds: Genome-wide association studies, COVID-19, Mendelian randomization, alternative splicing			

#### 33 Abstract

- 34 Hospital admission for COVID-19 remains common despite the successful development of vaccines
- 35 and treatments. Thus, there is an ongoing need to identify targets for new COVID-19 therapies.
- 36 Alternative splicing is an essential mechanism for generating functional diversity in protein isoforms
- 37 and influences immune response to infection. However, the causal role of alternative splicing in
- 38 COVID-19 severity and its potential therapeutic relevance is not fully understood. In this study, we
- 39 evaluated the causal role of alternative splicing in COVID-19 severity and susceptibility using
- 40 Mendelian randomization (MR). To do so, we performed two-sample MR to assess whether cis-sQTLs
- 41 spanning 8,172 gene splicing in 5,295 genes were associated with COVID-19 outcomes in the
- 42 COVID-19 Host Genetics Initiative, including up to 158,840 COVID-19 cases and 2,782,977
- 43 population controls. We identified that alternative splicing in lungs, rather than total RNA expression of
- 44 OAS1, ATP11A, DPP9 and NPNT, was associated with COVID-19 severity. MUC1 splicing was
- 45 associated with COVID-19 susceptibility. Further colocalization analyses supported a shared genetic
- 46 mechanism between COVID-19 severity with idiopathic pulmonary fibrosis at ATP11A and DPP9 loci,
- 47 and with chronic obstructive lung diseases at NPNT. We lastly showed that ATP11A, DPP9, NPNT,
- 48 and MUC1 were highly expressed in lung alveolar epithelial cells, both in COVID-19 uninfected and
- 49 infected samples. Taken together, these findings clarify the importance of alternative splicing of
- 50 proteins in the lung for COVID-19 and other respiratory diseases, providing isoform-based targets for
- 51 drug discovery.
- 52

## 54 Introduction

55 Despite the successful development of vaccines<sup>1,2</sup> and treatments<sup>3,4</sup>, hospital admission for severe 56 COVID-19 and long-term sequela for COVID-19 remain common<sup>5,6</sup>. COVID-19 is now a leading cause 57 of death, accounting for 350,000 deaths<sup>7</sup> and the estimated COVID-19 mortality rate is generally higher 58 than that of influenza by 2 to 100-fold, which vary by the study design<sup>7,8</sup>. Thus, there is an ongoing need 59 to identify mechanistic targets for therapeutic development to reduce the risk of severe COVID-19. 60 61 Using human genetics methods, several host factors have been identified to influence COVID-19 62 severity, including OAS1, type I interferon and chemokine genes<sup>9–14</sup>. Human genetics can provide new 63 biological insights into disease pathogenesis and the therapeutic targets and evidence from human 64 genetics increases the probability of drug development success<sup>15,16</sup>. While prior studies have evaluated 65 causal roles of both RNA expression and circulating proteins in COVID-19 outcomes<sup>9,12,13</sup>, the causal 66 role of alternative splicing has not been fully investigated. 67 68 Alternative splicing is an essential mechanism for generating functional diversity in the isoforms, through 69 which multiple mRNA isoforms are produced from a single gene, often in tissue-specific patterns<sup>17</sup>. 70 Alternative splicing has been implicated to play an important role in immune response to infections in 71 humans<sup>17</sup>, and this might be the case for SARS-CoV-2 infection. The genetic determinants of alternative 72 splicing have been identified using splicing quantitative trait loci (sQTL) studies, which indicate that 73 splicing is under strong genetic control in humans and often has direct effects on protein isoforms. 74 Indeed, we have recently identified that a Neanderthal-interrogated isoform of OAS1, which is strongly 75 regulated by an sQTL, affords protection against COVID-19 severity<sup>9</sup>. Given this evidence, we 76 hypothesized that alternative splicing could partially explain the variability in COVID-19 outcomes in 77 humans.

78

79	In this study, to test our hypothesis, we undertook two-sample MR and colocalization analyses to
80	combine results from genome-wide association studies (GWASs) of gene splicing and COVID-19
81	outcomes <sup>18</sup> . To do so, we first obtained the <i>cis</i> -regulatory genetic determinants of gene splicing ( <i>cis</i> -
82	sQTLs) in lungs and whole blood, two relevant tissues that influence acute SARS-CoV-2 infection, from
83	the GTEx Consortium <sup>19</sup> . We then used MR to assess whether these gene splicing events have effects on
84	COVID-19 outcomes using the GWASs from the COVID-19 Host Genetics Initiative <sup>11</sup> . Next, all findings
85	were assessed for colocalization to ensure that the gene splicing and COVID-19 outcomes shared a
86	common etiological genetic signal and that the MR results were not biased by linkage disequilibrium
87	(LD). We also compared effects of alternative splicing to total RNA expression and evaluated the
88	expression in the lung transcriptome datasets. Finally, we evaluated whether such alternative splicing is
89	a shared pathophysiological mechanism with other respiratory diseases.

- 90
- 91

## 92 Results

## 93 MR using cis-sQTLs, and colocalization analyses

94 The study design is illustrated in Fig. 1. We obtained the genetic determinants of gene splicing 95 (splicing quantitative trait locus [sQTL]) that influence alternative splicing events, which were 96 quantified as normalized intron excision ratios, *i.e.* the proportion of reads supporting each 97 alternatively excised intron, by LeafCutter<sup>20</sup> (using a multiple-testing correction threshold of a false 98 discovery rate of 5% per tissue) in GTEx v.8<sup>19</sup>. We chose to examine only *cis*-sQTLs, which act *cis* to 99 the coding genes, as it is less likely that these act through horizontal pleiotropy, where the SNPs have 100 effects on diseases (here, COVID-19 outcomes) independently of the gene splicing. We also focused 101 on two COVID-19-relevant tissues; lungs (N=515, of which 452 were of European American ancestry) 102 and whole blood (N=670, of which 570 were of European American ancestry) since pulmonary 103 symptoms are the major determinant of hospital admission and the immune system plays a major role 104 in host response to SARS-CoV-2.

105

106	A total of 4,477 genes in lungs and 2,779 genes in whole blood (5,295 unique genes in total) contained
107	conditionally independent cis-sQTLs that were also present the GWAS meta-analyses of COVID-19 Host
108	Genetics Initiative <sup>11</sup> release 7 ( <u>https://www.covid19hg.org/results/r7/</u> ), which included results from the
109	GenOMICC <sup>12</sup> study. We then undertook two-sample MR analyses using 6,110 <i>cis</i> -sQTLs in lungs and
110	3,883 <i>cis</i> -sQTLs in whole blood as genetic instruments for gene splicing for three COVID-19 outcomes:
111	1) critical illness (defined as individuals experiencing death, mechanical ventilation, non-invasive
112	ventilation, high-flow oxygen, or use of extracorporeal membrane oxygenation) owing to symptoms
113	associated with laboratory-confirmed SARS-CoV-2 infection (18,152 cases and 1,145,546 controls, of
114	which 13,769 and 1,072,442 were of European ancestry); 2) hospitalization owing to symptoms
115	associated with laboratory-confirmed SARS-CoV-2 infection (44,986 cases and 2,356,386 controls, of
116	which 32,519 and 2,062,805 were of European ancestry); and 3) reported SARS-CoV-2 infection defined
117	as laboratory-confirmed SARS-CoV-2 infection, electronic health record or clinically confirmed COVID-
118	19, or self-reported COVID-19, with or without symptoms of severity (159,840 cases and 2,782,977
119	controls, of which 122,616 and 2,475,240 are of European ancestry). Our primary MR analyses used
120	data from individuals of all ancestries, so as not to discard data for minority populations <sup>21</sup> , albeit the vast
121	majority of the individuals were of European ancestry; 86.2% in GTEx and 88.8% in COVID-19 HGI). We
122	also performed MR analyses using data derived only from European population as described below.
123	
124	MR analyses identified 29 genes, whose alternative splicing in lung influenced COVID-19 outcomes.
125	Using whole blood-derived alternative splicing we identified 20 genes influencing the same outcomes.
126	All associations survived a p-value threshold of 5.1x10 <sup>-6</sup> , which was based on Bonferroni correction
127	using the number of splicing events tested (in total 8,172 [5,999 in lung and 3,769 in whole blood],
128	Methods, Supplementary Table 1,2). We first replicated the association of COVID-19 outcomes with
129	OAS1 gene splicing which increases the excision of the intron junction at chr12:112,917,700-
130	112,919,389 [GRCh38] both in lung and whole blood, that corresponds to an increased level of the

131 p46 isoform, a prenylated form of OAS1 with higher anti-viral activity than the p42 isoform<sup>22-24</sup> (**Fig. 2**,

# 132 Extended Data Fig. 1A, Supplementary Table 1,2).

133

134 We additionally identified that ATP11A sQTL, rs12585036:T>C, which increases the excision of the

135 intron junction at chr13:112,875,941-112,880,546 in lung, was associated with protection against all

136 three adverse COVID-19 outcomes with odds ratio [OR] per normalized intron excision ratio [the

137 proportion of reads supporting each alternatively excised intron identified by LeafCutter<sup>20</sup>] of 0.78

138 (95%CI: 0.73-0.82, p=9.3x10<sup>-19</sup>) for critical illness, OR of 0.84 (95%CI: 0.81-0.87, p=3.1x10<sup>-21</sup>) for

hospitalization, and OR of 0.96 (0.94-0.97, p=2.9x10<sup>-6</sup>) for reported infection (**Fig. 2, Extended Data** 

# 140 Fig. 1B, Supplementary Table 1).

141

142 We also found novel associations of DPP9 sQTL, rs12610495:G>A, which increases the excision of

143 the intron junction at chr19:4,714,337-4,717,615 in lung, with protection against all three adverse

144 COVID-19 outcomes (OR: 0.49, 0.45-0.54, p=5.5x10<sup>-59</sup> for critical illness, OR: 0.64, 0.61-0.68,

145  $p=4.0x10^{-52}$  for hospitalization, and OR: 0.85, 0.83-0.88,  $p=3.2x10^{-29}$  for reported infection, Fig. 2).

146 DPP9 has two major isoforms, DPP9L (longer isoform, 863 aa, ENST00000262960) and DPP9S

147 (shorter isoform, 863 aa, ENST00000598800)<sup>25</sup>, and they differ in the presence of the skipping of exon

148 9 (chr19:4,715,529-4,716,152), which overlaps the intron junction at chr19:4,714,337-4,717,615

149 (Extended Data Fig. 1C, Supplementary Table 1).

150

151 NPNT rs34712979:A>G, which increases the intron junction of chr4:105,898,001-105,927,336

152 (Extended Data Fig. 1D), was associated with increased risk of severe COVID-19 outcomes (critical

153 illness and hospitalization) but with marginal evidence of increased risk of SARS-CoV-2 infection (OR:

154 1.18, 1.12-1.24, p=5.5x10<sup>-10</sup> for critical illness, OR: 1.10, 1.07-1.14, p=3.8x10<sup>-8</sup> for hospitalization, and

155 OR: 1.02, 1.00-1.03, p=4.2x10<sup>-2</sup> for reported infection, Fig. 2, Supplementary Table 1). The NPNT

- 156 sQTL, rs34712979-A allele, creates a NAGNAG splice acceptor site, which results in additional in-
- 157 frame AGT codon, coding for serine, at the 5' splice site of exon  $2^{26}$ .
- 158
- 159 Gene splicing in the two genes in the gene cluster located in chr1 q21.3-q22, namely *MUC1* and
- 160 THBS3, in lung were also associated with reported SARS-CoV-2 infection, but not with COVID-19
- 161 severity phenotypes (Fig. 2, Supplementary Table 1). Given that the *MUC1* and *THBS3* sQTL SNPs
- are in high LD with each other (rs4072037:C>T and rs2066981:A>G; r2=0.98 in European population
- 163 of 1000G), it was challenging to distinguish which genes were more likely to be causal between *MUC1*
- and *THBS3*. Nevertheless, *MUC1* sQTL SNP, rs4072037:C>T, is a recognized splice variant which
- 165 influences the 3' splice site selection of exon2<sup>27</sup>. In GTEx, the rs4072037-T allele was associated with
- decreased levels of the intron junction of chr1:155,192,310:155,192,786 (Extended Data Fig. 1E),
- 167 which leads to transcripts that have an alternative 27 bp intron retention event at the start of exon  $2^{27}$ .
- 168 We demonstrated the rs4072037-C allele, which increases the levels of the intron junction of
- 169 chr1:155,192,310:155,192,786, was associated with increased risk of reported SARS-CoV-2 infection.
- 170 Given the well-known evidence of the splicing event of *MUC1*, it is more likely that *MUC1* gene
- 171 splicing has an impact of SARS-CoV-2 infection.
- 172
- 173 All of the above associations of gene splicing, except for *MUC1*, were more pronounced with more
- severe outcomes (**Fig. 2**). We also applied MR for another hospitalization phenotype compared within
- 175 COVID-19 cases ("hospitalization vs non-hospitalization amongst individuals with laboratory-confirmed
- 176 SARS-CoV-2 infection", which corresponds to B1 phenotype in COVID-19 HGI<sup>11</sup>). We confirmed the
- 177 associations with gene splicing in ATP11A and DPP9 (ATP11A: OR 0.87, 0.81-0.94, p=4.6x10<sup>-4</sup>, and
- 178 DPP9: OR 0.83, 0.73-0.94, p=4.9x10<sup>-3</sup>) but not with gene splicing in NPNT and OAS1
- 179 (Supplementary Table 3).
- 180
- 181

## 182 Colocalization and European-only MR analyses

183 To test whether confounding due to linkage disequilibrium may have influenced the MR estimates, we 184 tested the probability that alternative splicing and the COVID-19 outcomes shared a single causal 185 signal using colocalization analyses, as implemented in coloc<sup>28</sup>. These colocalization analyses were 186 performed using data from individuals of European ancestry, as this method assumes that the two 187 GWASs tested were derived from the individuals with the similar LD structure. We next performed MR 188 as a sensitivity analysis by using the data only from individuals of European descent to ensure that the 189 MR results were not biased from population stratification<sup>29</sup>. 190 191 Amongst the MR-prioritized splicing events (29 genes in lung and 20 genes in whole blood,

192 **Supplementary Table 1,2**), we selected the splicing events with high (>0.8) posterior probability for

193 hypothesis 4 (PP, that there is an association for both gene expression/splicing and COVID-19

194 outcomes, which is driven by the same causal variant), and with a support from European-descent

195 only MR analysis (p<5.1x10<sup>-6</sup>, where the p-value threshold was adopted from the main analysis) (**Fig1**,

196 **3, Extended Data Fig. 1**). These analyses retained 8 genes (ABO, ATP11A, DPP9, GBAP1, MUC1,

197 NPNT, OAS1, THBS3) whose alternative splicing in lung influenced at least one COVID-19 outcome

198 and 2 genes (MOSPD3, OAS1) whose alternative splicing in whole blood influenced at least one

199 COVID-19 outcome (Supplementary Table 4,5).

200

201 A posterior probability (PP) that alternative splicing of ATP11A in lungs and COVID-19 outcomes

shared a single causal signal in the 1Mb locus around the *cis*-sQTL was 1.00 for critical illness, 1.00

203 for hospitalization due to COVID-19, and 0.98 for reported infection (**Fig. 3A**). Alternative splicing of

204 DPP9 had as high PP as 0.96 for critical illness, 0.96 for hospitalization, and 0.95 for reported infection

205 (Fig. 3B). Alternative splicing of NPNT had as high PP as 1.00 for critical illness, and 1.00 for

hospitalization, but had as low PP as 0.02 for reported infection (Fig. 3C). All of these gene splicing

207 had a support from European-descent only MR analysis with similar effect estimates as the main

analyses (p<5.1x10<sup>-6</sup>, **Supplementary Table 4,5**).

209

210

211 Colocalization with eQTL GWASs from GTEx

212 Next, to understand if total gene expression levels in lung and whole blood were associated with

213 COVID-19 outcomes, without respect to the splicing or isoforms, we performed MR and colocalization

analyses using expression quantitative trait loci (eQTLs) in lung and whole blood from GTEx. The total

215 expression of *GBAP1* had MR evidence with reported SARS-CoV-2 infection (p=5.9x10<sup>-8</sup>) and the

eQTLs of ABO, GBAP1, and MOSPD3, had high colocalization with COVID-19 outcomes (PP >0.8,

217 **Supplementary Table 6, 7**). Thus, for those genes, we could not clarify whether the associations with

218 COVID-19 outcomes were driven by either total gene expression or the spliced isoform expression or

219 both.

220

221

222 Influence of prioritized sQTLs on other diseases.

We next searched for the effects of the sQTLs, which influence COVID-19 outcomes, on other diseases. To do so, we used the Open Targets Genetics<sup>30,31</sup> (https://genetics.opentargets.org) and identified GWAS lead variants that were in LD (r<sup>2</sup>>0.80) with COVID-19 associated sQTLs. We found that the *DPP9* sQTL (rs12610495:G>A) was associated with decreased risk of idiopathic pulmonary fibrosis (IPF)<sup>32</sup> and the same allele confers the protection against COVID-19 severity. Interestingly, we found the opposite direction of effects between IPF and COVID-19 in the sQTL for *ATP11A*, where the rs12585036-C allele was associated with increasing risk of IPF and decreased risk for COVID-19

severity. The *NPNT* sQTL had similar trend that rs34712979-A allele, which was protective for COVID19 severity, was associated with increased risk of COPD (*i.e.* lower FEV1/FVC ratio, a spirometry
measurement used to diagnose COPD)<sup>33,34</sup> and with increased risk of asthma<sup>35</sup>. The *OAS1* sQTL
(rs10774671:A>G) was associated with a decreased risk of systemic lupus erythematosus (in the East
Asian population)<sup>36</sup> and with an increased risk of chronic lymphocytic leukemia<sup>37</sup>. The *MUC1* sQTL
(rs4072037:T>C) was associated with decreased risk of gastric cancer<sup>38</sup>, gout<sup>39</sup>, and inflammatory
bowel disease<sup>40</sup>.

237

238 We then performed colocalization analyses in each sQTL locus between COVID-19 outcomes and the 239 other associated diseases/phenotypes if the effects were supported in European population and there 240 was genome-wide GWAS summary data available. IPF<sup>32</sup> and critical illness due to COVID-19 were 241 highly colocalized at the ATP11A and DPP9 sQTL loci with PP of 1.00. The FEV1/FVC ratio also 242 colocalized with critical illness due to COVID-19 with a PP of 1.00. Inflammatory bowel disease<sup>40</sup> and 243 urate level (which is the cause of gout) also colocalized with reported SARS-CoV-2 infections at the 244 MUC1 locus (both PPs are 1). We found no GWAS summary statistics available for the other diseases 245 (Table 1). These lines of evidence suggest that IPF and COVID-19 severity, COPD/asthma and 246 COVID-19 severity, and IBD and COVID-19 susceptibility may share causal genetic determinants in 247 these loci, respectively. The full results and the data used are summarized in Table 1. 248 249

# 250 The tissue and cell-type specific expression of the associated genes

To assess the relevant tissues and cell-types for the associated gene splicing – COVID-19 outcome relationships, we evaluated the associated gene expression in lung and peripheral blood mononuclear cell (PBMC) of healthy controls, as well as in lung of COVID-19 patients. In the consensus transcript expression levels from Human Protein Atlas (HPA)<sup>41</sup> and GTEx, the expression of *ATP11A*, and *NPNT* were highly enriched in normal lung tissue (**Fig. 4A**). At a single-cell resolution, in normal lung, *MUC1*,

ATP11A and NPNT were specifically enriched in alveolar type 1 and 2 epithelial cells (Fig. 4B), which

256

257	are known to play important roles in regeneration of alveolar epithelium following lung injury <sup>42</sup> . In 23
258	lung COVID-19 autopsy donor tissue samples from GSE171668 <sup>43</sup> , the expression of ATP11A, DPP9,
259	MUC1, and NPNT were also enriched in alveolar type 1 and 2 epithelial cells (Fig. 4C). Taken
260	together, these expression evidence suggests that the gene splicing – COVID-19 outcome
261	relationships for ATP11A, DPP9, MUC1, and NPNT were highly likely to be relevant in lung tissue,
262	especially alveolar epithelial cells for ATP11A, MUC1, and NPNT.
263	
264	
265	Discussion
266	Despite current vaccines and therapeutic options, hospitalization for COVID-19 remains high in many
267	countries. Thus, there remains a need for additional therapies, which in turn, requires target
268	identification and validation <sup>44</sup> . In this large-scale two-sample MR study of 8,172 gene splicing events
269	assessed for their effect upon three COVID-19 outcomes in up to 158,840 COVID-19 cases and
270	2,782,977 population controls, we provide evidence that alternative splicing of OAS1, ATP11A, DPP9
271	and NPNT in lung influences COVID-19 severity, and alternative splicing of MUC1 in lung influences
272	COVID-19 susceptibility. Moreover, these genetic mechanisms are likely shared with other diseases,
273	such as ATP11A and DPP9 with IPF, suggesting the potential to be targeted for multiple diseases.
274	Our results provide rationale to target gene splicing in lung as COVID-19 treatment by means of
275	therapeutic modalities such as splice-switching oligonucleotides (SSOs).
276	
277	SSOs are a type of antisense oligonucleotides which are generally 15–30 nucleotides in length and
278	are designed as complementary to specific regions of mRNA with increased stability against
279	endogenous nucleases due to chemical modifications <sup>45</sup> . SSOs can prevent splicing promoting factors
280	from binding to the target pre-mRNA, which can modulate alternative splicing <sup>46</sup> . Some SSOs have
281	been already approved by FDA, such as eteplirsen for Duchenne's muscular dystrophy, which induces

282 the skipping of exon 51 of DMD gene, and nusinersen for spinal muscle atrophy, which induces the 283 expression of exon 7 of SMN2 gene<sup>45</sup>. Although delivery to target tissues after systemic administration 284 has been a key challenge in the development of SSO drugs<sup>45,47</sup>, lung might be advantageous tissue to 285 target since it is possible to deliver drugs directly to lung through inhalation and/or bronchoscopy. 286 287 There are multiple sources of biological evidence which might support the relevance of ATP11A in 288 COVID-19. ATP11A is a member of the P4-ATPase family, which codes for a phospholipid flippase at 289 the plasma membrane and translocates phosphatidylserine (PtdSer) from the outer to the inner leaflet 290 of plasma membranes. Cells that undergo apoptosis, necroptosis or pyroptosis expose PtdSer on their 291 surface through P4-ATPase, of which ATP11A is cleaved by caspase during apoptosis<sup>48</sup>. While less 292 functionally defined, given the high expression of ATP11A in the alveolar type 2 epithelial cells, 293 ATP11A sQTL (rs12585036:C>T) may impact the cell death of alveolar type 2 epithelial cells, the 294 resident stem cell population in lung<sup>49</sup>, and may impair regeneration of alveolar epithelium following 295 lung injury. 296 297 Our MR analysis showed that the excision of intron junction chr19:4,814,337-4,717,615, which may be 298 preferentially spliced in DPP9L than in DPP9S, was associated with reduced risk of COVID-19 299 severity. While both DPP9L and DPP9S can be active<sup>25</sup>, DPP9S is present in the cytosol, and DPP9L 300 localizes preferentially to the nucleus, which may be attributed to a classical monopartite nuclear 301 localization signal (RKVKKLRL) in the N-terminal extension of DPP9L<sup>25</sup>. Although it is not fully 302 understood how the DPP9 sQTL (rs12610495:A>G) regulates alternative splicing of DPP9, 303 rs12610495-G allele creates a GGGG motif (from GGAG sequence), which may have an effect on 304 alternative splicing of some genes<sup>50</sup>. DPP9 interacts with NLRP1 and represses inflammasome 305 activation<sup>51</sup> and pyroptosis, which are now recognized as important mechanisms interrupting the viral 306 replication cycle and preventing viral amplification of SARS-CoV-2<sup>52</sup>. Moreover, targeting

inflammasome-mediated hyperinflammation in COVID-19 patients may also prevent chronic phase of
 COVID-19 pathophysiology in vivo<sup>52</sup>.

309

310 Our MR analyses supported that the MUC1 gene splicing associated with susceptibility of SARS-CoV-311 2 infection. MUC1 is a mucin, which is also called KL-6 in human, and the serum level of KL-6 is used 312 as a biomarker for interstitial lung diseases<sup>53</sup>. rs4072037-T allele was associated with the decreased 313 level of the intron junction of chr1:155,192,310:155,192,786, which corresponds to transcripts with an 314 alternative 27 bp intron retention event at the start of exon 2<sup>27</sup>. MUC1 exon 2 harbors a variable 315 number of tandem repeats (VNTR) that contains 20 to 125 repeats of a 60 bp coding sequence that 316 determines the length of a heavily glycosylated extracellular domain<sup>54</sup>. Although rs4072037 may or 317 may not control the alternative splicing of VNTR region<sup>38</sup>, in addition to the alternative 27 bp intron 318 retention<sup>27</sup>, the VNTR length of *MUC1* is associated with several renal phenotypes in recent study in 319 UK Biobank<sup>55</sup>. The rs4072037-T allele has been shown to be associated with increased risk of gastric 320 cancer in both European and East Asian ancestries<sup>38,56</sup> and it is hypothesized that the MUC1 gene is 321 involved in response to *H.pylori* colonization in gastric epithelial surfaces. Taken together, it is 322 plausible that alternative splicing of MUC1 in lung alveolar cells has an impact on SARS-CoV-2 323 infection. 324 325 This study has limitations. First, the excised intron junction was quantified in an annotation-free 326 manner using LeafCutter<sup>20</sup>, without respect to the level of transcripts nor isoforms. It is thus important 327 for future work to map those disease-relevant alternative splicing events to the corresponding isoform

328 or protein product, by means of emerging technologies such as long-read sequencing<sup>57</sup> and high-

throughput protein quantification<sup>58</sup>. We anticipate our findings, and others, will motivate the ongoing

330 effort to do so. Second, we used MR to test the effect of gene splicing measured in a non-infected

- 331 state since the effect of the *cis*-sQTLs upon gene splicing was estimated in individuals who had not
- been exposed to SARS-CoV-2. Given the dynamic gene regulation of splicing during infection<sup>59</sup>, gene

333 splicing could be altered once a person experiences SARS-CoV-2 infection. Thus, the MR results 334 presented in this paper should be interpreted as an estimation of the effect of gene splicing during 335 uninfected state. Future studies may help to clarify if the same *cis*-sQTLs regulate alternative splicing 336 during infection. Third, it was not our goal to identify all alternative splicing that affects COVID-19 337 outcomes, but rather to provide strong evidence for a small set of genes with strong MR and 338 colocalization evidence. Thus, we acknowledge a high false-negative rate of our study design. 339 Although we attempted to include the data from individuals of diverse ancestries, most individuals 340 were of European descent. Therefore, we could not confirm that our findings could be robustly 341 transferrable to individuals of other populations. Fourth, while our findings demonstrate that effects of 342 alternative splicing in lung and whole blood since these are relevant to COVID-19 severity, we 343 recognize that these splicing events may not be unique to these tissues and thus the estimated effects 344 of splicing may represent the action of these same alternative transcripts in other tissues. Lastly, a 345 recent paper<sup>60</sup> reported that the shared genetic signal between IPF and COVID-19 outcomes at the 346 DPP9 and ATP11A loci are likely driven by the difference of total expression in whole blood, which 347 was supported by colocalization using eQTLGen<sup>61</sup> dataset. However, eQTLGen consists of multi-348 ancestry cohorts and could be affected by the bias due to LD. Moreover, we demonstrated that both 349 DPP9 and ATP11A expression were enriched in alveolar cells in lung, compared to whole blood. 350 Thus, we concluded that it is more likely that the alternative splicing of DPP9 and ATP11A in lung is 351 important both for IPF and COVID-19 severity.

352

In conclusion, we have used genetic determinants of gene splicing and COVID-19 outcomes obtained from large-scale studies and found compelling evidence that splicing events in *OAS1*, *ATP11A*, *DPP9*, *NPNT*, and *MUC1* have causal effects on COVID-19 severity and susceptibility. Interestingly, the available evidence suggests shared genetic mechanisms for COVID-19 severity with IPF at *ATP11A* and *DPP9* loci, and with chronic obstructive lung diseases at the *NPNT* locus. Taken together, our

- 358 study highlights the importance of alternative splicing both in COVID-19 and other diseases, which
- 359 could be further investigated for drug discovery such as SSOs.
- 360
- 361
- 362 Methods
- 363 Splicing quantitative trait loci (sQTL) GWASs
- 364 We obtained all conditionally independent splicing quantitative trait loci (sQTLs) in lung (N=515, of which
- 365 452 were of European ancestry) and whole blood (N=670, of which 570 were of European ancestry) that
- 366 act in *cis* (in a +/- 1Mb window around the transcription start site of each gene with normalized intron
- 367 excision ratios from LeafCutter<sup>20</sup> (5% FDR per tissue). Conditional independent sQTLs were mapped
- 368 using stepwise regression in GTEx consortium<sup>19</sup>. Intron excision ratios are the proportion of reads
- 369 supporting each alternatively excised intron identified by LeafCutter<sup>20</sup>. We selected lung and whole blood
- 370 since they are the two major tissues relevant to COVID-19 pathophysiology.
- 371
- 372

### 373 COVID-19 GWASs

- 374 To assess the association of *cis*-sQTLs with COVID-19 outcomes, we used COVID-19 GWAS meta-
- 375 analyses results from the COVID-19 Host Genetics Initiative (HGI) release 7<sup>11</sup>
- 376 (https://www.covid19hg.org/results/r7/). The outcomes tested were critical illness, hospitalization, and
- 377 reported SARS-CoV-2 infection (named A2, B2, and C2, respectively by the COVID-19 HGI).
- 378
- 379 Critically ill COVID-19 cases were defined as those individuals who were hospitalized with laboratory-
- 380 confirmed SARS-CoV-2 infection and who required respiratory support (invasive ventilation, continuous
- 381 positive airway pressure, Bilevel Positive Airway Pressure, or continuous external negative pressure,
- 382 high-flow nasal or face-mask oxygen) or who died due to the disease. Simple supplementary oxygen
- 383 (e.g. 2 liters/minute via nasal cannula) did not qualify for case status. Hospitalized COVID-19 cases were

384	defined as individuals hospitalized with laboratory-confirmed SARS-CoV-2 infection (using the same
385	microbiology methods as for the critically ill phenotype), where hospitalization was due to COVID-19
386	related symptoms. Reported SARS-CoV-2 infection was defined as laboratory-confirmed SARS-CoV-2
387	infection or electronic health record, ICD coding or clinically confirmed COVID-19, or self-reported
388	COVID-19 (for example, by questionnaire), with or without symptoms of any severity. Controls were
389	defined in the same way across all three outcomes above as everybody that is not a case-for example,
390	population controls.
201	

391

392 In a sensitivity analysis, we also used another hospitalization phenotype (named B1 in the COVID-19

393 HGI), where cases were hospitalized COVID-19 cases and controls were defined as non-hospitalized

individuals with laboratory-confirmed SARS-CoV-2 infection.

395

396

# 397 Two-sample Mendelian randomization

398 We used two-sample mendelian randomization (MR) analyses using "TwoSampleMR v0.5.6" R 399 package<sup>62</sup> to screen and test the potential role of alternative splicing in lung and whole blood to influence 400 COVID-19 outcomes. In two-sample MR, the effect of SNPs on the exposure and outcome are taken 401 from separate GWASs. Two-sample MR often improves statistical power and is less biased than one-402 sample MR, with potential bias in the direction to the null<sup>63</sup>. MR is less affected by confounding and 403 reverse causality than observational epidemiology studies. The MR framework is based on three main 404 assumptions: First, the SNPs are robustly associated with the exposure (*i.e.* a lack of weak instrument 405 bias). We validated that all *cis*-sQTLs had a F-statistic >10, corresponding to T-statistics >3.16, which 406 indicate a low risk of weak instrument bias in MR analyses<sup>64</sup>. Second, the SNPs are not associated with 407 any confounding factors for the relationship between the exposure and the outcome. Third, the SNPs 408 have no effect on the outcome that is independent of the exposure (*i.e.* a lack of horizontal pleiotropy), 409 which is the most challenging assumption to assess. Nevertheless, in order to reduce the risk of

410 horizontal pleiotropy, we selected cis-sQTLs as instrumental variables, as cis-SNPs that reside close to 411 the genes are more likely to have an effect on the outcomes by directly influencing the splicing of the 412 gene. Palindromic *cis*-sQTLs with minor allele frequencies (MAF) >0.42 were removed prior to MR to 413 prevent allele-mismatches. For gene splicing with a single (sentinel) cis-sQTL, we used the Wald ratio to 414 estimate the effect of each splicing event on each of the three COVID-19 outcomes. For any gene 415 splicing event with multiple conditionally independent cis-sQTLs, an inverse variance weighted (IVW) 416 method was used to meta-analyze their combined effects. After harmonizing the cis-sQTLs with COVID-417 19 GWASs, a total of 5,999 splicing events in 4,477 genes (6,110 matched cis-sQTLs) in lung and a total 418 of 3,769 splicing events in 2,779 genes (3,883 matched cis-sQTLs) in whole blood were used for the MR 419 analyses across the three COVID-19 outcomes. We applied the Bonferroni corrected p-value (5.1x10<sup>-6</sup>) 420 to adjust for multiple testing, by using the number of splicing events tested (5,999 in lung and 3,769 in 421 whole blood, in total 8,172).

- 422
- 423

## 424 European-only MR analyses and colocalization analysis

425 Next, we assessed potential confounding by population stratification and linkage disequilibrium (LD). To 426 do so, we first repeated MR analyses using the data from individuals of European ancestry. We obtained 427 the effect estimate of each *cis*-sQTL from the data mapped in European-American subjects. If the *cis*-

- 428 sQTL was missing in the summary data mapped in European-ancestry subjects, we removed those *cis*-
- 429 sQTLs in the analyses. The three COVID-19 outcomes GWASs results were also replaced by the meta-
- 430 analyses summary results of European-ancestry subjects.

- 432 We also evaluated whether the gene splicing and COVID-19 outcomes shared a common etiological
- 433 genetic signal and that the MR results were not biased by linkage disequilibrium (LD) using colocalization
- 434 analyses with the data from individuals of European ancestry. Specifically, for each of these MR
- 435 significant splicing events, a stringent Bayesian analysis was implemented in "coloc v 5.1.0.1" R package

- 436 to analyze all variants in 1MB genomic locus centered on the *cis*-sQTL. Colocalizations with posterior
- 437 probability for high colocalization (PP >0.8), that is, that there is an association for both gene splicing and
- 438 COVID-19 outcomes, and they are driven by the same causal variant were considered to colocalize,
- 439 which means that the exposure and the outcome shared a single causal variant.
- 440
- 441
- 442 Mendelian randomization and Colocalization with eQTL GWASs from GTEx
- 443 To understand if the total gene expression levels in lung and whole blood were associated with
- 444 COVID-19 outcomes, without respect to the splicing or isoforms, we similarly performed MR and
- 445 colocalization analyses using expression quantitative trait loci (eQTLs) in lung and whole blood from
- 446 GTEx v.8<sup>19</sup> (lung: N=515, of which 452 were of European ancestry, and whole blood: N=670, of which
- 570 were of European ancestry) by restricting the regions within 1 Mb of each QTL. The genetic
- 448 instruments were conditionally independent eQTLs for the prioritized sQTL genes in lung and/or whole
- 449 blood, all of which had strong support for colocalization between sQTLs and COVID-19 outcomes.
- 450

### 451 Influence of identified sQTLs on other diseases

- 452 We assessed the effects of the sQTLs which influence COVID-19 outcomes on other diseases.
- 453 Pleiotropic search was performed using Open Targets Genetics<sup>30,31</sup> (<u>https://genetics.opentargets.org</u>)
- 454 and identified any GWAS lead variants that are in LD (r<sup>2</sup>>0.80) with those sQTLs. Colocalization
- 455 analyses were performed using "coloc v 5.1.0.1" R package when there is available GWAS summary
- 456 statistics of European ancestry.
- 457
- 458 The tissue and cell-type specific expression of the associated genes
- 459 To assess the relevant tissues and cell-types for the associated gene splicing COVID-19 outcome
- 460 relationships, we evaluated the associated gene expression in lung and peripheral blood mononuclear
- 461 cell (PBMC) of healthy controls, as well as in lung of COVID-19 patients. We first downloaded

462	consensus transcript expression levels summarized per gene in 54 tissues in Human Protein Atlas
463	(HPA) <sup>41</sup> , which was calculated as the maximum transcripts per million value (TPM) value for each
464	gene in all sub-tissues categories of each tissue, based on transcriptomics data from HPA and GTEx.
465	We also downloaded the single-cell type transcriptomic analyses, where we used all cell types in lung
466	(GSE13014865) and peripheral blood mononuclear cell (PBMC) (GSE11284566). We visualized RNA
467	single cell type tissue cluster data (transcript expression levels summarized per gene and cluster),
468	using $log_{10}$ (protein-transcripts per million [pTPM]) values with "corrplot v 0.92" R package. Lastly, we
469	obtained single-cell RNA expression profile of 23 lung COVID-19 autopsy donor tissue samples from
470	GSE171668 <sup>43</sup> . We calculated "pseudo-bulk" RNA expression per each cell type by taking the mean
471	value of RNA expression of the cells annotated in the same subcategory.
472	
473	
474	Data availability
475	Data from GTEx consortium (GTEx project v8 <sup>19</sup> ) are available from the referenced peer-reviewed
476	studies or their corresponding authors, as applicable. Summary statistics for eQTLs and sQTLs from
477	GTEx v8 are publicly available for download on the GTEx website (gtexportal.org/home/datasets).
478	Summary statistics for the COVID-19 outcomes are publicly available for download on the COVID-19
479	Host Genetics Initiative website <sup>67</sup> (www.covid19hg.org). The consensus transcript expression levels
480	and RNA single cell type tissue cluster data were downloadable from the Human Protein Atlas <sup>41</sup>
481	website (www.proteinatlas.org/about/download). The processed single-cell RNA expression profile of
482	23 lung COVID-19 autopsy donor tissue sample are freely available in the Gene Expression Omnibus
483	(GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE171668 <sup>43</sup> .
484	
485	Code availability
486	All code for data management and analysis is archived online at github.com/richardslab/COVID19-

487 <u>SQTLMR</u> for review and reuse.

sQTL locus (rsID)	chr	pos (b38)	EA	NEA	intron junction	COVID-19 outcome	Other outcome	PP*
ATP11A (rs12585036)	13	112881427	С	Т	chr13:112,875,941-112,880,546 ↑ (lung, WBC)	critical illness ↓	idiopathic pulmonary fibrosis ↑ [PMID: 31710517]	1
DPP9 (rs12610495)	19	4717660	A	G	chr19:4,814,337-4,717,615 ↑ (lung)	critical illness ↓	idiopathic pulmonary fibrosis ↓ [PMID: 31710517]	1
NPNT (rs34712979)	4	105897896	A	G	chr4:105,898,001-105,927,336 ↑ (lung)	critical illness ↓	FEV1/FVC ratio ↓ [PMID: 30804560]	1
							COPD ↑ [PMID: 30804561]	-
							Asthma ↑ [PMID: 31959851]	-
OAS1 (rs10774671)	11	112919388	G	A	chr12:112,917,700-112,919,389 ↑ (lung)	critical illness ↓	systemic lupus erythematosus ↓ [PMID: 33272962]	-
							chronic lymphocytic leukemia↑ [PMID: 28165464]	-
<i>MUC1</i> (rs4072037)	1	155192276	С	Т	chr1:155,192,310:155,192,786 ↑ (lung)	reported infection ↑	inflammatory bowel disease ↓ [PMID: 28067908]	1
							gastric cancer ↓ [PMID: 26098866]	-
							gout ↓ [PMID: 33832965]	-
							Urate level ↓ [PMID:33462484]	1

# 488 Table 1. Colocalization analyses of COVID-19 outcomes and other diseases at the identified sQTL loci.

489 EA: effect allele, NEA: non-effect allele, PP: a posterior probability that there is an association for both gene expression/splicing and COVID-19 outcomes,

490 which is driven by the same causal variant. PP was only estimated when there is an available GWAS summary statistics of European ancestry.

#### 491 Fig. 1. Flow Diagram of Study Design



#### 517 Fig. 2. MR estimates of the effect of gene splicing at ATP11A, DPP9, NPNT, OAS1, and MUC1

#### 518 with COVID-19 outcomes

	ATP11A (chr13:112875941-112880546)	Odds ratio (95%Cl)	p-value
Critical illness		0.78 (0.73-0.82)	9.3x10 <sup>-19</sup>
Hospitalization		0.84 (0.81-0.87)	3.1x10 <sup>-21</sup>
Reported infection	•	0.96 (0.94-0.97)	2.9x10 <sup>-6</sup>
	DPP9 (chr19:4714337-4717615)		
Critical illness		0.49 (0.45-0.54)	5.5x10 <sup>-59</sup>
Hospitalization		0.64 (0.61-0.68)	4.0x10 <sup>-52</sup>
Reported infection	•	0.85 (0.83-0.88)	3.2x10 <sup>-29</sup>
	<i>NPNT</i> (chr4:105898001-105927336)		
Critical illness		1.18 (1.12-1.24)	5.5x10 <sup>-10</sup>
Hospitalization		1.10 (1.07-1.14)	3.8x10 <sup>-8</sup>
Reported infection	•	1.02 (1.00-1.03)	4.2x10 <sup>-2</sup>
	OAS1 (chr12:112917700-112919389)		
Critical illness —•		0.23 (0.15-0.36)	5.0x10 <sup>-11</sup>
Hospitalization		0.34 (0.25-0.45)	8.3x10 <sup>-14</sup>
Reported infection		0.65 (0.57-0.74)	3.6x10 <sup>-10</sup>
	<i>MUC1</i> (chr1:155192310-155192786)		
Critical illness		1.01 (0.68-1.49)	0.98
Hospitalization		1.39 (1.07-1.81)	0.013
Reported infection		1.51 (1.33-1.71)	4.1x10 <sup>-10</sup>
	0.5 1.0 1.5	_	
(	Joos ralio (95%CI) per normalized intron excision ratios	5	

519 Forest plot showing odds ratio and 95% confidence interval from two sample Mendelian

520 Randomization analyses (two-sided). P values are unadjusted. Unit: standard deviation of intron

521 excision ratios as quantified by LeafCutter.

Fig. 3. LocusZoom plots demonstrating the colocalization of the genetic determinants of mRNA
expression, gene splicing levels and COVID-19 outcomes LocusZoom plots of genetic signal of
1MB region around each *cis*-sQTL and COVID-19 outcomes, color shows SNPs in the region in LD
(r<sup>2</sup>) to the *cis*-sQTL. (A) *ATP11A* (B) *DPP9* (C) *NPNT* (D) *MUC1*







## 532 Figure 4. The gene expression in lung and peripheral blood mononuclear cell (PBMC).

- 533 A. The consensus transcript expression levels summarized per gene in 54 tissues in Human Protein Atlas (HPA), which was calculated as the maximum
- 534 transcripts per million value (TPM) value for each gene in all sub-tissues categories of each tissue, based on transcriptomics data from HPA and
- 535 GTEx.
- 536 B. RNA single cell type tissue cluster data (transcript expression levels summarized per gene and cluster) of lung (GSE130148) and peripheral blood
- 537 mononuclear cell (PBMC) (GSE112845) were visualized using log<sub>10</sub>(protein-transcripts per million [pTPM]) values. Each c-X annotation is taken from
- 538 the clustering results performed in Human Protein Atlas.
- 539 C. Single-cell RNA expression profile of 23 lung COVID-19 autopsy donor tissue samples from GSE171668<sup>43</sup>. The mean value of RNA expression of the
- 540 cells annotated in the same subcategory was represented as a dot per each sample. The cell type annotation was manually performed in the original
- 541 publication<sup>43</sup>. AT1: alveolar type 1 epithelial cells, AT2: alveolar type 2 epithelial cells, EC: endothelial cells, *KRT8*<sup>+</sup> PATS/ADI/DATPs: *KRT8*<sup>+</sup> pre-
- 542 alveolar type 1 transitional cell state, MAST: mast cells, XXhi: XX (gene expression) high cells, RBC: red blood cells.

- 543 Extended Data Fig. 1. LocusZoom plots demonstrating the colocalization of the genetic
- 544 determinants of mRNA expression, gene splicing levels for OAS1 and COVID-19 outcomes.
- 545 LocusZoom plots of genetic signal of 1MB region around each cis-sQTL and COVID-19 outcomes,

OAS1 eQTL

546 color shows SNPs in the region in LD  $(r^2)$  to the *cis*-sQTL for OAS1.

10





## 548 Extended Data Fig. 2. The gene plots.

549 The major gene-coding transcripts were manually curated to visualize, in relation to the sQTLs and the excised introns quantified by LeafCutter.



## 550 (A) OAS1 (B) ATP11A (C) DPP9 (D) NPNT (E) MUC1



### 552 Contributions

- 553 Conception and design: TN, and JBR. Formal analysis: TN. Data curation: TN, and YF. Interpretation
- of data: TN, YF, RJA, BGG, and JBR. Funding acquisition: JBR. Investigation: TN, and YF.
- 555 Methodology: TN, YF, and JBR. Project administration: JBR. Resources: JBR. Supervision: YF, and
- 556 JBR. Validation: TN, YF, and JBR. Visualization: TN. Writing—original draft: TN. Writing—review and
- 557 editing: TN, YF, JW, RJA, BGG, SZ, and JBR. All authors were involved in further drafts of the
- 558 manuscript and revised it critically for content. All authors gave final approval of the version to be
- 559 published. The corresponding authors attest that all listed authors meet authorship criteria and that no
- 560 others meeting the criteria have been omitted. All authors were involved in further drafts of the
- 561 manuscript and revised it critically for content. All authors gave final approval of the version to be
- 562 published. The corresponding author attests that all listed authors meet authorship criteria and that no
- 563 others meeting the criteria have been omitted.
- 564

## 565 Ethics declarations

### 566 **Competing interests**

- 567 JBR's institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline and
- 568 Biogen for projects unrelated to this research. He is the CEO of 5 Prime Sciences
- 569 (www.5primesciences.com), which provides research services for biotech, pharma and venture capital
- 570 companies for projects unrelated to this research. No other disclosures were reported.
- 571

### 572 Funding

- 573 The Richards research group is supported by the Canadian Institutes of Health Research (CIHR)
- 574 (365825; 409511, 100558, 169303), the Lady Davis Institute of the Jewish General Hospital, the
- 575 Canada Foundation for Innovation (CFI), the NIH Foundation, Cancer Research UK, Genome
- 576 Québec, the Public Health Agency of Canada, the McGill Interdisciplinary Initiative in Infection and
- 577 Immunity and the Fonds de Recherche Québec Santé (FRQS). TN is supported by a research

- 578 fellowship of the Japan Society for the Promotion of Science for Young Scientists. Support from Calcul
- 579 Québec and Compute Canada is acknowledged.
- 580
- 581

## 582 References

- 583 1. Haas, E. J. et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-
- 584 CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a
- 585 nationwide vaccination campaign in Israel: an observational study using national
- 586 surveillance data. *The Lancet* **397**, 1819–1829 (2021).
- 587 2. Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N.
- 588 Engl. J. Med. **383**, 2603–2615 (2020).
- 589 3. Dexamethasone in Hospitalized Patients with Covid-19 Preliminary Report. N. Engl. J.
- 590 *Med.* 1–11 (2020) doi:10.1056/nejmoa2021436.
- 591 4. Hammond, J. et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-
- 592 19. N. Engl. J. Med. **386**, 1397–1408 (2022).
- 593 5. Huang, L. et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal
- 594 cohort study. *The Lancet* **398**, 747–758 (2021).
- 595 6. The Lancet. Understanding long COVID: a modern medical challenge. *The Lancet* **398**,
- 596 725 (2021).
- 597 7. Piroth, L. et al. Comparison of the characteristics, morbidity, and mortality of COVID-19
- and seasonal influenza: a nationwide, population-based retrospective cohort study. *Lancet*
- 599 *Respir. Med.* 9, 251–259 (2021).
- 600 8. Iacobucci, G. Covid and flu: what do the numbers tell us about morbidity and deaths?
- 601 *BMJ* **375**, n2514 (2021).

602	9.	Zhou, S. et al. A Neanderthal OAS1 isoform protects individuals of European ancestry
603		against COVID-19 susceptibility and severity. Nat. Med. 1-9 (2021) doi:10.1038/s41591-
604		021-01281-1.
605	10.	The Severe Covid-19 GWAS Group. Genomewide Association Study of Severe Covid-19
606		with Respiratory Failure. N. Engl. J. Med. NEJMoa2020283 (2020)
607		doi:10.1056/NEJMoa2020283.
608	11.	COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-
609		19. Nature 1-8 (2021) doi:10.1038/s41586-021-03767-x.
610	12.	Pairo-Castineira, E. et al. Genetic mechanisms of critical illness in Covid-19. Nature 1-1
611		(2020) doi:10.1038/s41586-020-03065-y.
612	13.	Kousathanas, A. et al. Whole genome sequencing reveals host factors underlying critical
613		Covid-19. Nat. 2022 1-10 (2022) doi:10.1038/s41586-022-04576-6.
614	14.	Nakanishi, T. et al. Age-dependent impact of the major common genetic risk factor for
615		COVID-19 on severity and mortality. J. Clin. Invest. 131, (2021).
616	15.	Nelson, M. R. et al. The support of human genetic evidence for approved drug

- 617 indications. Nat. Genet. 47, 856–860 (2015).
- 618 16. Cook, D. et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-
- 619 dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–31 (2014).

- 620 17. Kornblihtt, A. R. et al. Alternative splicing: a pivotal step between eukaryotic
- 621 transcription and translation. *Nat. Rev. Mol. Cell Biol. 2013* 143 14, 153–165 (2013).
- 622 18. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host
- 623 genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J.*
- 624 *Hum. Genet.* **28**, 715–718 (2020).
- 625 19. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human
- tissues. *Science* (2020) doi:10.1126/SCIENCE.AAZ1776.
- 627 20. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. Nat.
- 628 Genet. (2018) doi:10.1038/s41588-017-0004-9.
- 629 21. Ben-Eghan, C. et al. Don't ignore genetic data from minority populations Setting the
- 630 agenda in research. *ALAMY 184 Nat.* **585**, (2020).
- 631 22. Zhou, S. et al. A Neanderthal OAS1 isoform protects individuals of European ancestry
- against COVID-19 susceptibility and severity. Nat. Med. 1–9 (2021) doi:10.1038/s41591-
- 633 021-01281-1.
- 634 23. Huffman, J. E. et al. Multi-ancestry fine mapping implicates OAS1 splicing in risk of
- 635 severe COVID-19. *Nat. Genet.* 2022 1–3 (2022) doi:10.1038/s41588-021-00996-8.
- 636 24. Wickenhagen, A. et al. A prenylated dsRNA sensor protects against severe COVID-19.
- 637 *Science* **374**, (2021).

638	25. Justa-Schuch, D., Möller, U. & Geiss-Friedlander, R. The amino terminus extension in
639	the long dipeptidyl peptidase 9 isoform contains a nuclear localization signal targeting the
640	active peptidase to the nucleus. Cell. Mol. Life Sci. CMLS 71, 3611-3626 (2014).
641	26. Saferali, A. et al. Characterization of a COPD-Associated NPNT Functional Splicing
642	Genetic Variant in Human Lung Tissue via Long-Read Sequencing. MedRxiv Prepr. Serv.
643	Health Sci. 2020.10.20.20203927 (2020) doi:10.1101/2020.10.20.20203927.
644	27. Ng, W., Loh, A. X. W., Teixeira, A. S., Pereira, S. P. & Swallow, D. M. Genetic
645	regulation of MUC1 alternative splicing in human tissues. Br. J. Cancer 99, 978–985
646	(2008).
647	28. Hormozdiari, F. et al. Colocalization of GWAS and eQTL Signals Detects Target Genes.
648	Am. J. Hum. Genet. 99, 1245–1260 (2016).
649	29. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population
650	stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459 (2010).
651	30. Mountjoy, E. et al. Open targets genetics: An open approach to systematically prioritize
652	causal variants and genes at all published human GWAS trait-associated loci. Nat. Genet.
653	2020.09.16.299271 (2021) doi:10.1038/s41588-021-00945-5.
654	31. Ghoussaini, M. et al. Open Targets Genetics: Systematic identification of trait-associated
655	genes using large-scale genetics and functional genomics. Nucleic Acids Res. 49, D1311-
656	D1320 (2021).
	35

657	32. Allen, R. J. et al. Genome-Wide Association Study of Susceptibility to Idiopathic
658	Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med. 201, 564–574 (2020).
659	33. Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic
660	obstructive pulmonary disease associations across multiple ancestries. Nat. Genet. 51,
661	481–493 (2019).
662	34. Sakornsakolpat, P. et al. Genetic landscape of chronic obstructive pulmonary disease
663	identifies heterogeneous cell-type and phenotype associations. Nat. Genet. 51, 494–505
664	(2019).
665	35. Olafsdottir, T. A. et al. Eighty-eight variants highlight the role of T cell regulation and
666	airway remodeling in asthma pathogenesis. Nat. Commun. 11, 393 (2020).
667	36. Yin, X. et al. Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for
668	systemic lupus erythematosus. Ann. Rheum. Dis. 80, 632-640 (2021).
669	37. Law, P. J. et al. Genome-wide association analysis implicates dysregulation of immunity
670	genes in chronic lymphocytic leukaemia. Nat. Commun. 8, 14175 (2017).
671	38. Helgason, H. et al. Loss-of-function variants in ATM confer risk of gastric cancer. Nat.
672	Genet. 47, 906–910 (2015).
673	39. Sandoval-Plata, G., Morgan, K. & Abhishek, A. Variants in urate transporters, ADH1B,
674	GCKR and MEPE genes associate with transition from asymptomatic hyperuricaemia to

675	gout: results of the first gout versus asymptomatic hyperuricaemia GWAS in Caucasians
676	using data from the UK Biobank. Ann. Rheum. Dis. 80, 1220-1226 (2021).
677	40. De Lange, K. M. et al. Genome-wide association study implicates immune activation of
678	multiple integrin genes in inflammatory bowel disease. Nat. Genet. (2017)
679	doi:10.1038/ng.3760.
680	41. Karlsson, M. et al. A single-cell type transcriptomics map of human tissues. Sci. Adv. 7,
681	eabh2169 (2021).
682	42. Castranova, V., Rabovsky, J., Tucker, J. H. & Miles, P. R. The alveolar type II epithelial
683	cell: a multifunctional pneumocyte. Toxicol. Appl. Pharmacol. 93, 472–483 (1988).
684	43. Delorey, T. M. et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular
685	targets. Nature 595, 107-113 (2021).
686	44. Katzourakis, A. COVID-19: endemic doesn't mean harmless. Nature 601, 485 (2022).
687	45. Levin, A. A. Treating Disease at the RNA Level with Oligonucleotides. N. Engl. J. Med.
688	<b>380</b> , 57–70 (2019).
689	46. Kole, R., Krainer, A. R. & Altman, S. RNA therapeutics: beyond RNA interference and
690	antisense oligonucleotides. Nat. Rev. Drug Discov. 11, 125-140 (2012).
691	47. Davis, M. E. et al. Evidence of RNAi in humans from systemically administered siRNA
692	via targeted nanoparticles. Nature 464, 1067–1070 (2010).

693	48. Segawa, K.	et al. A sublethal	ATP11A mutation	associated with neurological	
-----	----------------	--------------------	-----------------	------------------------------	--

- 694 deterioration causes aberrant phosphatidylcholine flipping in plasma membranes. J. Clin.
- 695 *Invest.* **131**, 148005 (2021).
- 696 49. Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. J. Clin. Invest.
- 697 **123**, 3025–3036 (2013).
- 698 50. Královičová, J. & Vořechovský, I. Position-Dependent Repression and Promotion of
- 699 *DQB1* Intron 3 Splicing by GGGG Motifs. J. Immunol. **176**, 2381–2388 (2006).
- 51. Hollingsworth, L. R. et al. DPP9 sequesters the C terminus of NLRP1 to repress
- 701 inflammasome activation. *Nature* **592**, 778–783 (2021).
- 52. Sefik, E. et al. Inflammasome activation in infected macrophages drives COVID-19
- 703 pathology. *Nature* **606**, 585–593 (2022).
- 53. Ishikawa, N., Hattori, N., Yokoyama, A. & Kohno, N. Utility of KL-6/MUC1 in the
- clinical management of interstitial lung diseases. *Respir. Investig.* **50**, 3–13 (2012).
- 706 54. Fowler, J. C., Teixeira, A. S., Vinall, L. E. & Swallow, D. M. Hypervariability of the
- 707 membrane-associated mucin and cancer marker MUC1. *Hum. Genet.* **113**, 473–479
- 708 (2003).
- 709 55. Mukamel, R. E. et al. Protein-coding repeat polymorphisms strongly shape diverse human
- 710 phenotypes. *Science* **373**, 1499–1505 (2021).

711	56. Saeki, N. et al. A functional single nucleotide polymorphism in mucin 1, at chromosome	
712	1q22, determines susceptibility to diffuse-type gastric cancer. Gastroenterology 140, 892-	
713	902 (2011).	
714	57. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing	
715	and its applications. Nat. Rev. Genet. 21, 597-614 (2020).	
716	58. Kraemer, S. et al. From SOMAmer-based biomarker discovery to diagnostic and clinical	
717	applications: A SOMAmer-based, streamlined multiplex proteomic assay. PLoS ONE 6,	
718	(2011).	
719	59. Rotival, M., Quach, H. & Quintana-Murci, L. Defining the genetic and evolutionary	
720	architecture of alternative splicing in response to infection. Nat. Commun. 10, 1671	
721	(2019).	
722	60. Allen, R. J. et al. Genetic overlap between idiopathic pulmonary fibrosis and COVID-19.	
723	Eur. Respir. J. 60, 2103132 (2022).	
724	61. Võsa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic	
725	loci and polygenic scores that regulate blood gene expression. Nat. Genet. 53, 1300-1310	
726	(2021).	
727	62. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the	
728	human phenome. <i>eLife</i> 7, e34408 (2018).	

729	63. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-	
730	sample Mendelian randomization. Genet. Epidemiol. 40, 597-608 (2016).	
731	64. Burgess, S., Thompson, S. G., & CRP CHD Genetics Collaboration. Avoiding bias from	
732	weak instruments in Mendelian randomization studies. Int. J. Epidemiol. 40, 755–764	
733	(2011).	
734	65. Vieira Braga, F. A. et al. A cellular census of human lungs identifies novel cell states ir	
735	health and in asthma. Nat. Med. 25, 1153-1163 (2019).	
736	66. Chen, J. et al. PBMC fixation and processing for Chromium single-cell RNA sequencing.	

- 737 J. Transl. Med. 16, 198 (2018).
- 67. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-738
- 739 19. Nature 600, 472–477 (2021).