

18 **Abstract**

19 **Background:** Large genome-wide association studies (GWAS) and other genetic studies have
20 revealed genetic loci that are associated with chronic obstructive pulmonary disease (COPD).
21 However, the proteins responsible for COPD pathogenesis remain elusive. We used integrative-
22 omics by combining genetics of lung function and COPD with genetics of proteome to identify
23 proteins underlying lung function variation and COPD risk.

24 **Methods:** We used summary statistics from the GWAS of human plasma proteome from the
25 INTERVAL cohort (n=3,301) and integrated these data with lung function GWAS results from
26 the UK Biobank cohorts (n=400,102) and COPD GWAS results from the ICGC cohort (35,735
27 cases and 222,076 controls). We performed in parallel: a proteome-wide Bayesian colocalization,
28 and a proteome-wide Mendelian Randomization (MR) analyses. Next, we selected proteins that
29 colocalized with lung function and/or COPD risk and explored their causal association with lung
30 function and/or COPD using MR analysis ($P < 0.05$).

31 **Results:** We found 537, 607, and 250 proteins that colocalized with force expiratory volume in
32 one second (FEV₁), FEV₁/forced vital capacity (FVC), or COPD risk, respectively. Of these,
33 1,051 were unique proteins. The sRAGE protein demonstrated the strongest colocalization with
34 FEV₁/FVC and COPD risk, while QSOX2, FAM3D and F177A proteins had the strongest
35 associations with FEV₁. Of these, 37 proteins that colocalized with lung function and/or COPD,
36 also had a significant causal association. These included proteins such as PDE4D, QSOX2 and
37 RGAP1, amongst others.

38 **Conclusion:** Integrative-omics reveals new proteins related to lung function. These proteins may
39 play important roles in the pathogenesis of COPD.

40

41 **Introduction**

42 Chronic obstructive pulmonary disease (COPD) is a persistent respiratory condition that is
43 characterized by irreversible and progressive lung function impairment, and is responsible for
44 over three million deaths worldwide each year (Roth et al. 2018). Large genome-wide
45 association studies (GWAS) have revealed hundreds of genetic loci that are associated with
46 COPD (Cho et al. 2014; Hobbs et al. 2017; Sakornsakolpat et al. 2019) and lung function more
47 generally (Shrine et al. 2019). However, the molecular mechanisms relating these genetic
48 associations to lung function and COPD pathogenesis remain elusive, hindering the ability to
49 translate these findings into new therapeutic targets or biomarkers of disease.

50 Integrative-omic methodologies may provide insights into the biological relationships
51 between genetic variants and complex traits such as lung function and COPD (Giambartolomei
52 et al. 2014; Gusev et al. 2016). These methods aim to establish a link between gene expression or
53 protein levels and the trait by leveraging their respective associations with common genetic
54 variants, which can be determined from independent cohorts. For example, the integration of
55 COPD GWAS with transcriptomic datasets suggests that the effects of many COPD risk loci are
56 mediated through regulation of gene expression in lung tissue (Obeidat et al. 2015; Lamontagne
57 et al. 2018).

58 Determining causal associations between molecular and complex traits are critical for
59 understanding disease pathogenesis, and for translating these molecules into biomarkers or
60 therapeutic targets. One method that has recently gained momentum in assessing causality of a
61 complex trait-molecular phenotype relationship is Mendelian Randomization (MR). The MR
62 framework exploits the random allocation of alleles during meiosis and relates their effects on a
63 putative risk factor, which can be quantified by measuring biomolecules such as a blood
64 protein (Smith and Ebrahim 2003; Voight et al. 2012). This in turn can be related to a trait. MR
65 analysis measures the ‘lifetime exposure’ to this risk factor in a way that is relatively resistant to
66 confounding from environmental influences or reverse causation. This enables an unbiased
67 assessment of causality. MR analysis has established causal associations between a number of
68 candidate blood proteins and COPD (Milne et al. 2020). However, the application of MR
69 analysis at a genome-wide discovery level and by coupling it with integrative omics methods
70 will likely yield many additional novel protein associations with COPD.

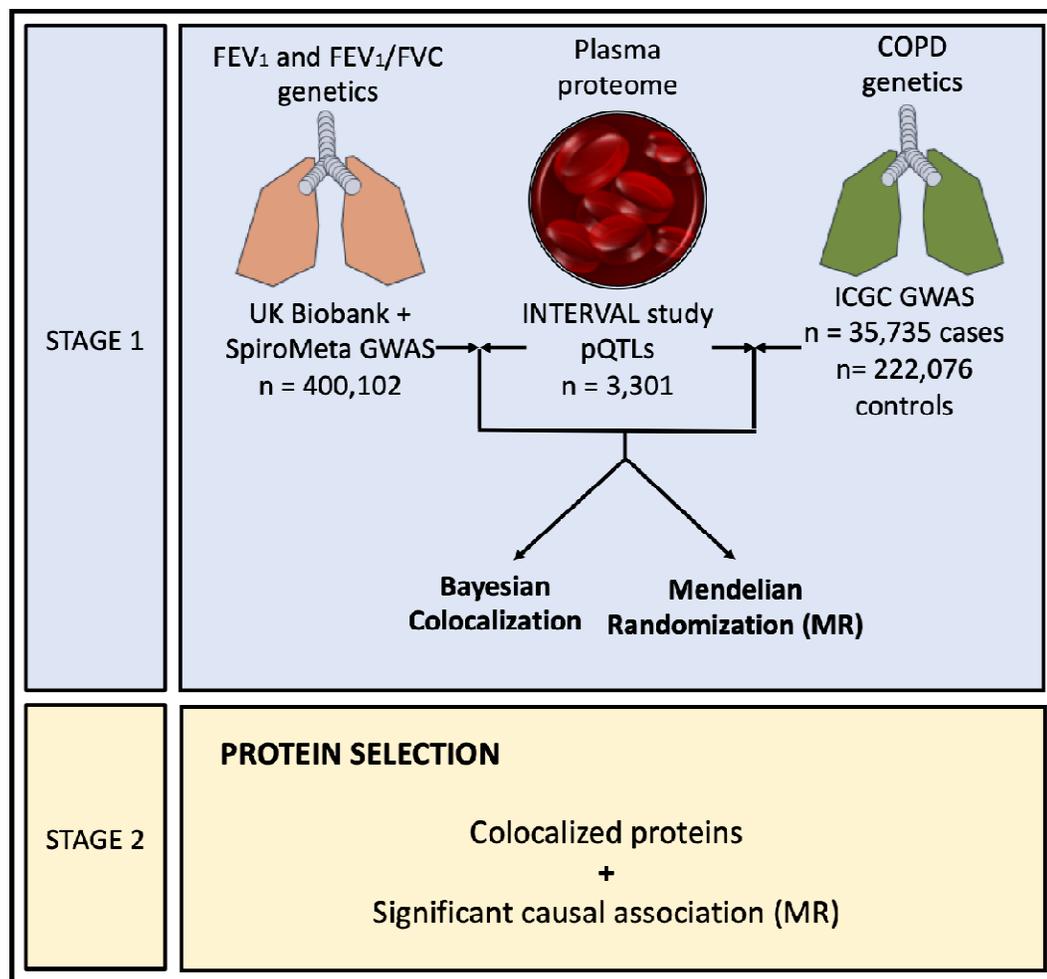
71 In this study, we combined two approaches; MR analysis of plasma proteins and COPD
72 risk and genome-wide Bayesian colocalization (COLOC). We used both of these approaches to
73 integrate a large human plasma proteome dataset (Sun et al. 2018) with GWASs for lung
74 function in a general population (Shrine et al. 2019) and for COPD risk in a large case-control
75 dataset (Sakornsakolpat et al. 2019), to nominate promising protein targets for further
76 exploration in mechanistic and biomarker studies.

77 **Results**

78 *Study design*

79 The overall study design is shown in **Figure 1**. We aimed to identify plasma proteins
80 causally associated with the following phenotypic traits: forced expiratory volume in one sec
81 (FEV_1); FEV_1 to forced vital capacity (FEV_1/FVC) ratio; and the presence of COPD (henceforth
82 referred to as ‘COPD risk’). These associations were determined by integrating a human plasma
83 proteome GWAS (Sun et al. 2018) with the largest existing GWAS for lung function (Shrine et
84 al. 2019) and COPD (Sakornsakolpat et al. 2019). Each of these datasets (described in detail in
85 the Methods) summarizes the association of millions of genetic variants (single nucleotide
86 polymorphisms [SNPs]) with their respective traits. We performed the study in two stages. Stage
87 1 was an unbiased discovery of proteins associated with the phenotypic traits, by performing two
88 integrative omics methods in parallel: a genome-wide COLOC analysis of each associated
89 genetic locus, and a genome-wide MR analysis of all measured proteins. Stage 2 was a step-wise
90 analysis, wherein we used only proteins that were significantly colocalized with one or more of
91 the phenotypic parameters in subsequent, apply a hypothesis-driven MR threshold. From these
92 results, we generated a list of plasma proteins that showed a causal association with COPD risk
93 and/or lung function.

94



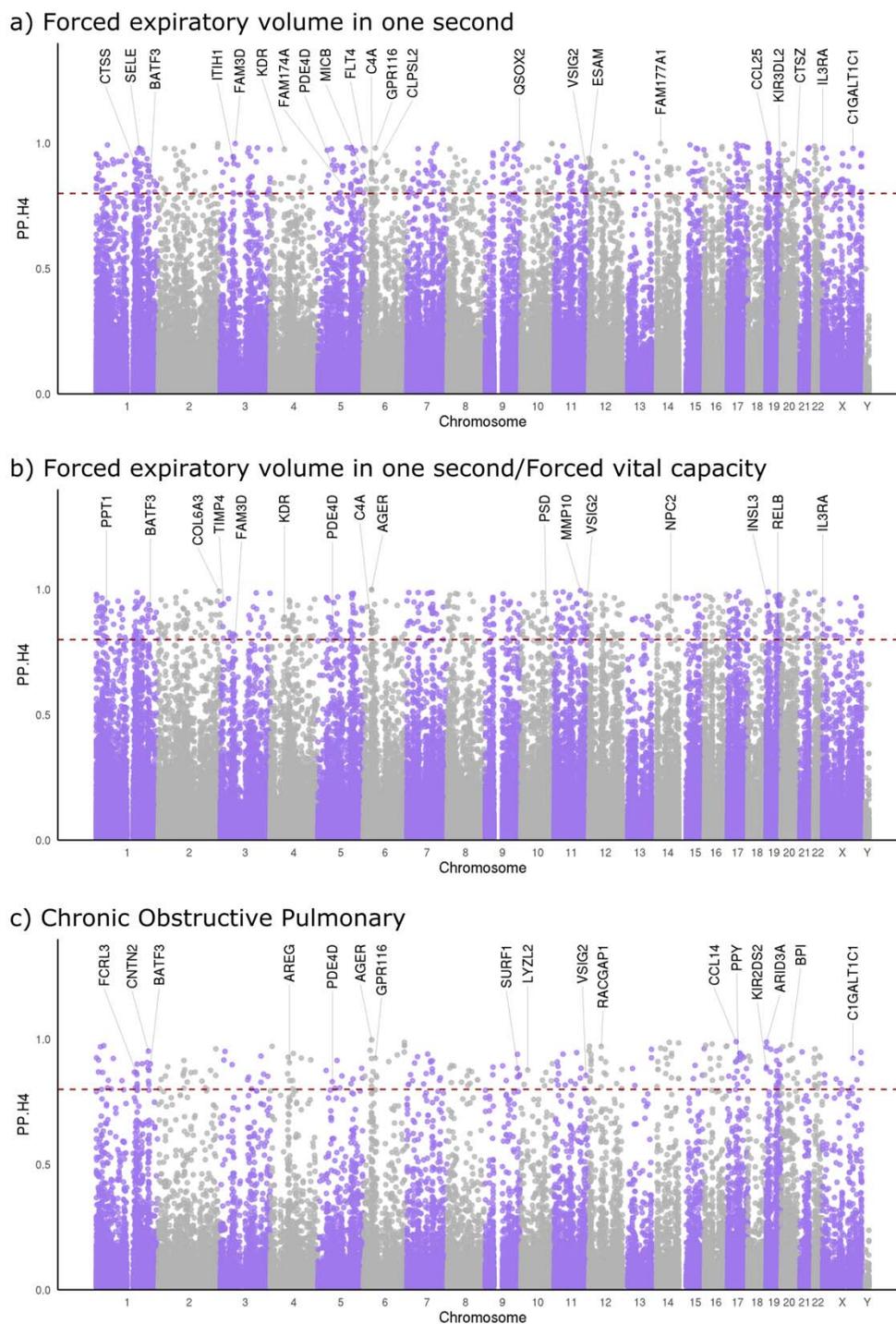
95
96 **Figure 1. Study design.** The diagram shows the workflow used to identify causal factors for lung
97 function and COPD risk. We selected as top plasma proteins those that showed significant colocalization
98 at $PP_{H4} > 0.80$ and causality at $FDR < 0.10$ or $P_{MR} < 0.05$ with lung function and/or COPD risk.

99 *Stage 1: a genome-wide discovery of plasma proteins associated with lung function and COPD*
100 *risk*

101 In Stage 1, we used two integrative omics methods. We first performed a COLOC
102 analysis across the 2,995 proteins, which were measured in the plasma proteome dataset. For this
103 analysis, we included genetic loci associated with each of the clinical parameters (lung function
104 traits and COPD) at a $P_{GWAS} < 5 \times 10^{-07}$ and set the significance of colocalization at $PP_{H4} > 0.80$.

105 In total, 1,048 unique proteins were colocalized with at least one of the COPD
106 phenotypes. Of these, 447 protein colocalized at $PP_{H4} > 0.90$. For the lung function traits, 537
107 proteins colocalized with FEV₁; proteins with the highest PP_{H4} were sulfhydryl oxidase 2
108 (QSOX2) ($PP_{H4} = 0.99$), protein FAM3D (FAM3D) ($PP_{H4} = 0.99$) and FAM177A1 (F177A)
109 ($PP_{H4} = 0.99$) (**Figure 2**). Likewise, 607 proteins colocalized with FEV₁/FVC; those with the

110 highest PP_{H4} included the advanced glycosylation end product-specific receptor (sRAGE) (PP_{H4}
111 = 0.99), stromelysin-2 (MMP-10) ($PP_{H4} = 0.99$) and collagen alpha-3(VI) ($PP_{H4} = 0.99$)
112 (**Figure 2**). We found that 200 unique proteins overlapped with both traits (**Supplemental Table**
113 **1**). For COPD risk, there were 250 colocalized proteins. Based on PP_{H4} , the top three colocalized
114 proteins were sRAGE ($PP_{H4} = 0.99$), C-C motif chemokine 14 (HCC-1) ($PP_{H4} = 0.99$), and AT-
115 rich interactive domain-containing protein 3A (ARI3A) ($PP_{H4}=0.99$) (**Figure 2**). Approximately
116 half of proteins that were colocalized with COPD risk were also colocalized with FEV_1 (94/250)
117 or FEV_1/FVC (126/250) (**Supplemental Table 1**).



118
 119 **Figure 2. Bayesian colocalization (COLOC) analyses of the plasma proteome related to lung**
 120 **function (a and b) traits and chronic obstructive pulmonary disease (COPD) risk (c).** The horizontal
 121 axis in each plot represents the chromosomal position of the plasma protein coding genes and the vertical
 122 axis shows the posterior probability of the two phenotypes (protein level and clinical trait) sharing a
 123 common genetic variant (PP_{H4}). Significant colocalization is defined as $PP_{H4} > 0.80$ (red dashed line).

124 Each purple or grey dot represents a plasma protein. Labelled colocalized proteins are those with
125 significant causal associations with the phenotypic trait at $P < 0.05$. Note: Plasma proteins are labelled
126 based on their protein coding gene names.
127

128 We next performed MR analyses for each of the 2,995 proteins present in the plasma
129 proteome dataset (Sun et al. 2018). MR analysis uses genetic variants as instrumental variables to
130 relate their per-allele effects on a risk factor to their per-allele effects on a phenotypic trait. MR
131 therefore estimates the ‘causal’ effect of the risk factor on the phenotypic trait. For these
132 analyses, we used a two-sample multivariable inverse variance weighted MR model (IVW-MR),
133 and tested various MR assumptions (see Methods). Since this was a hypothesis-free analysis, we
134 set a false discovery rate (FDR) of < 0.1 to control for multiple comparisons.

135 For the lung function traits, three proteins – QSOX2, Rac GTPase-activating protein 1
136 (RGAP1) and NAD(P)H quinone dehydrogenase 1 (NQO1) – showed causal associations with
137 FEV₁ at FDR < 0.1 (**Table 1**). Testing of the MR assumptions showed no significant
138 heterogeneity based on a Cochran’s Q test ($P > 0.05$), or pleiotropy (Egger-MR intercept $P >$
139 0.05) for either protein. Based on the direction of the MR estimate, we inferred that increased
140 plasma levels of QSOX2 and RGAP1 were associated with decreased FEV₁ (**Table 1 and**
141 **Figure 3**), while increased plasma level of NQO1 was associated with increased FEV₁ (**Table 1**).
142 None of the analysed proteins showed a significant causal association with FEV₁/FVC at FDR $<$
143 0.1 . MHC class I polypeptide-related sequence B (MICB) showed a significant causal
144 association with COPD risk (**Table 1**) with no evidence of heterogeneity ($P > 0.05$) or
145 pleiotropy (Egger MR Intercept $P > 0.05$). The direction of effect was such that increased
146 plasma MICB level was associated with reduced COPD risk (**Table 1 and Figure 3**). None of
147 the proteins was causally associated with more than one COPD phenotype.

148 When we compared the results of the COLOC and hypothesis-free MR, we found that
149 three proteins (QSOX2, RGAP1 and MICB) were identified by both methods at FDR < 0.1 .

150

151 **Table 1. Causal association of plasma proteins with lung function and/or COPD risk from**
152 **hypothesis-free MR analysis**

Protein	Trait	Beta	SE	P-value	Het P-value
NQO1	FEV ₁	0.02	0.004	2.88×10^{-08}	0.44
RGAP1	FEV ₁	-0.04	0.005	6.49×10^{-15}	0.20
QSOX2	FEV ₁	-0.02	0.003	5.85×10^{-11}	0.44
MICB	COPD risk	-0.15	0.018	9.98×10^{-17}	0.16

153 Plasma protein with significant Mendelian randomization (MR) ($FDR < 0.10$) for lung function and
154 chronic obstructive pulmonary disease (COPD) risk. From left to right columns shows 1) the name of the
155 protein; 2) trait used for the MR, 3) inverse variance weighting (IVW)-MR estimated effect; 4) IVW-MR
156 standard error of the effect; 5) IVW-MR P -value; 6) Heterogeneity (Het) P -value of the SNPs used for
157 the MR (Cochran's Q test), 7) trait that colocalized (COLOC) with plasma protein levels (posterior
158 probability (PP_{H4}) of colocalization > 0.80). FEV₁: Forced expiratory volume in one second. ns: not
159 significant.
160

161 *Stage 2: a step-wise discovery of plasma proteins that were causally associated with lung*
162 *function and/or COPD risk*

163 In Stage 2, we extracted the MR results for proteins showing significant colocalization
164 ($PP_{H4} > 0.8$) with at least one of the phenotypes (537, 607, and 250 plasma proteins that
165 colocalized with FEV₁, FEV₁/FVC, and COPD risk, respectively). In this stage we considered
166 the MR result as a complement to the COLOC results; we therefore set the significance threshold
167 of the MR model at $P_{MR} < 0.05$. In total, 37 unique proteins showed a causal association with at
168 least one of the three phenotypic traits. These included 20 proteins associated with FEV₁, 12
169 proteins associated with FEV₁/FVC, and 9 proteins associated with COPD (**Table 2**). None of
170 the significant results showed heterogeneity ($P > 0.05$) or pleiotropy (Egger MR Intercept $P >$
171 0.05).

172 For FEV₁, the most significantly associated candidate protein was RGAP1 whose plasma
173 levels increased with decreasing FEV₁ (**Figure 3**). For FEV₁/FVC, the most significantly
174 associated protein was cAMP-specific 3', 5'-cyclic phosphodiesterase 4D (PDE4D), whose
175 plasma levels increased with increasing FEV₁/FVC (**Figure 3**). For COPD risk, the most
176 significantly associated protein was MICB, whose plasma levels increased with decreasing
177 COPD risk (**Figure 3**). RGAP1 and MICB, but not PDE4D, were also identified with the
178 hypothesis-free MR approach (stage 1) with the same direction of effect for their respective
179 phenotypic traits.

180 There was some overlap between the proteins associated with each of the COPD
181 phenotypes. For example, increased plasma QSOX2 was associated with increased COPD risk
182 and decreased FEV₁ (**Table 2**). Increased plasma contactin-2 (CNTN2) was causally associated
183 with decreased FEV₁/FVC and increased COPD risk. None of the proteins tested were found to
184 have a causal association with all three phenotypic traits.

185 *Highlighting potential candidate proteins*

186 We nominated colocalized proteins having strong support for a causal association with lung
187 function and/or COPD as candidate proteins. This list included proteins showing a significant
188 MR estimate in Stage 2; for example, **Figure 3** shows the MR results for PDE4D, MICB and
189 RGAP1; by nature of the step-wise analysis, these proteins were also significantly colocalized
190 with one or more of the COPD traits. The full MR results for these proteins are provided in
191 **Supplemental Table 2.**

192

193 **Table 2. Potential plasma protein targets associated with lung function and COPD**

Protein	Causal association (Mendelian randomization)			Colocalization		
	COPD Risk $\beta(P_{MR})$	FEV ₁ $\beta(P_{MR})$	FEV ₁ /FVC $\beta(P_{MR})$	COPD risk PP_{H4}	FEV ₁ PP_{H4}	FEV ₁ /FVC PP_{H4}
PDE4D	-0.18 (3.61×10^{-16})	-	0.08 (1.47×10^{-60})	0.83	0.91	0.84
QSOX2	0.05 (4.78×10^{-05})	-0.02 (5.85×10^{-11})	-	-	0.99	-
VEGF sR3	-0.09 (1.23×10^{-05})	-0.01 (1.45×10^{-03})	-	-	0.97	-
CNTN2	0.06 (9.14×10^{-04})	-	-0.03 (5.46×10^{-11})	0.95	-	-
MICB	-0.15 (9.98×10^{-17})	-	-	-	-	0.80
IL3-RA	-0.06 (3.21×10^{-05})	-	-	-	0.98	0.82
CF126	0.129307 (2.47×10^{-03})	-	-	-	0.88	-
NPC2	-0.03176 (8.69×10^{-03})	-	-	-	-	0.87
SURF	-0.05146 (3.85×10^{-02})	-	-	0.85	-	-
KI2S2	-	-0.05 (3.01×10^{-23})	-	0.88	-	-
RGAP1	-	-0.04 (6.49×10^{-15})	-	0.97	-	-
GP116	-	0.02 (4.42×10^{-11})	-	0.92	0.98	-
sE- Selectin	-	0.02 (2.51×10^{-10})	-	-	0.98	-
VEGF sR2	-	0.03 (2.58×10^{-10})	-	-	0.98	0.89
C4	-	-0.02 (1.82×10^{-04})	-	-	0.92	0.84
TIMP-4	-	0.02 (2.24×10^{-04})	-	-	-	0.94
Cathepsin S	-	0.01 (3.54×10^{-04})	-	-	0.93	-
LYZL2	-	-0.03 (4.82×10^{-04})	-	0.88	-	-
ITI heavy chain H1	-	-0.01 (6.58×10^{-04})	-	-	0.92	-
VSIG2	-	-0.02 (7.01×10^{-04})	-	0.86	0.91	0.90
BATF3	-	0.02 (7.78×10^{-04})	-	0.90	0.87	0.92
sRAGE	-	-0.02 (9.64×10^{-04})	-	0.99	-	0.99
INSL3	-	-0.03 (1.15×10^{-03})	-	-	-	0.90
C1GLC	-	-0.01 (1.49×10^{-03})	-	0.92	0.98	-
ESAM	-	0.02	-	-	0.81	-

		(3.70×10^{-03})				
TM157	-	-0.03 (5.47×10^{-03})	-	-	0.83	-
PPT1	-	-0.01 (8.25×10^{-03})	-	-	-	0.88
AREG	-	-	-0.02 (1.05×10^{-02})	0.91	-	-
RELB	-	-	-0.01 (1.36×10^{-02})	-	-	0.85
PSD1	-	-	0.01 (1.59×10^{-02})	-	-	0.88
FCRL3	-	-	0.01 (1.59×10^{-02})	0.87	-	-
KI3L2	-	-	-0.02 (1.93×10^{-02})	-	0.81	-
TECK	-	-	0.01 (2.36×10^{-02})	-	0.99	-
FAM3D	-	-	-0.01 (2.37×10^{-02})	-	0.99	0.82
BPI	-	-	-0.01 (3.03×10^{-02})	0.98	-	-
CATZ	-	-	-0.01 (3.31×10^{-02})	-	0.88	-
PH	-	-	0.02 (4.08×10^{-02})	0.92	-	-

194 The table shows the proteins that were selected as potential candidates. From left to right column: 1)
 195 Protein symbol; 2) effect of plasma protein on Chronic Obstructive Pulmonary Disease (COPD) risk and
 196 IVW-MR *P-value*; 3) effect of plasma protein on and forced expiratory volume in one second (FEV₁) and
 197 IVW-MR *P-value*; 4) effect of plasma protein on FEV₁/FVC and IVW-MR *P-value*; 5) posterior
 198 probability of colocalization (PP_{H4}) between plasma protein levels and COPD risk; 6) PP_{H4} between
 199 plasma protein levels and FEV₁; 7) PP_{H4} between plasma protein levels and FEV₁/FVC.

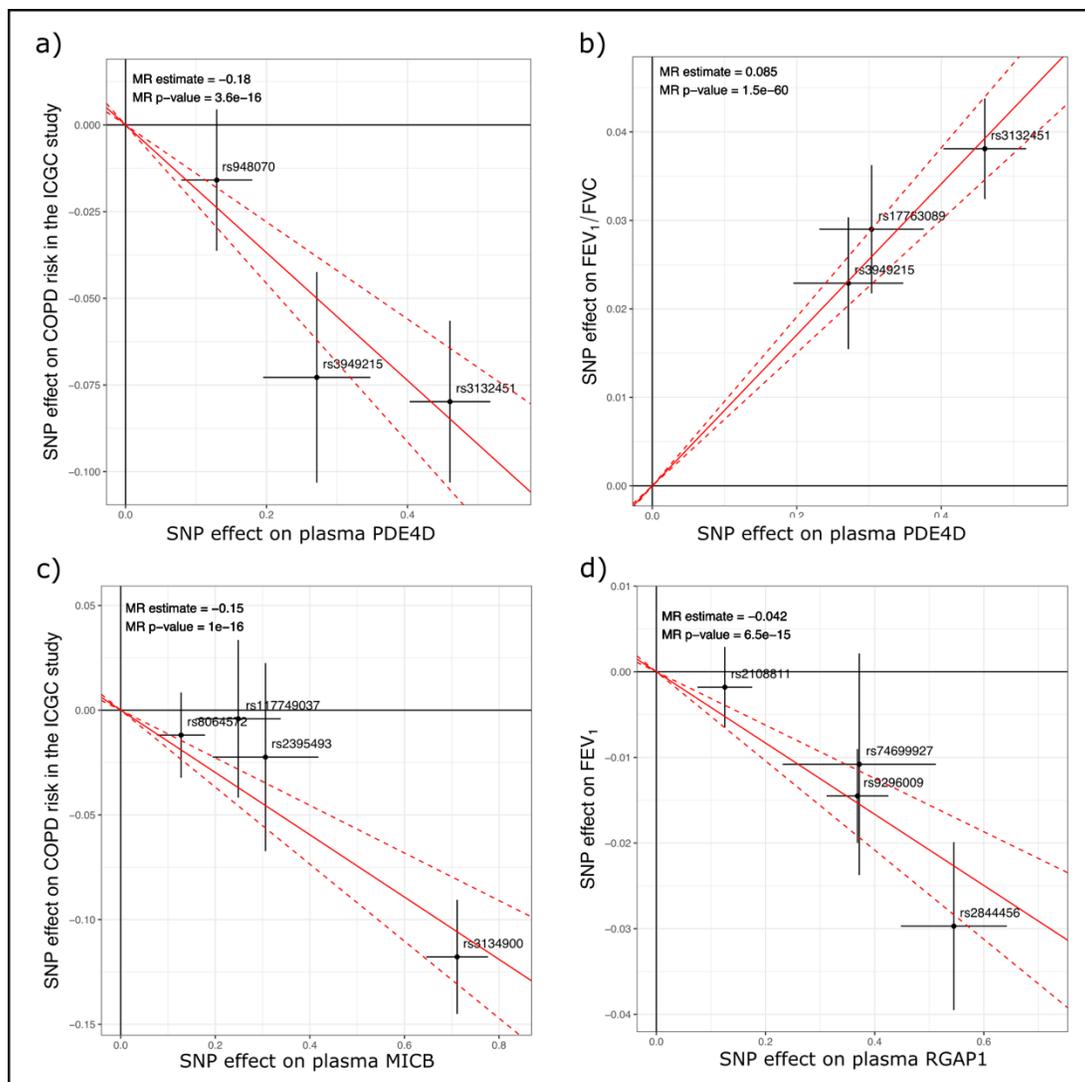
200

201

202

203

204



205

206 **Figure 3. Mendelian Randomization (MR) of protein biomarkers for lung function and Chronic**
 207 **Obstructive Pulmonary Disease (COPD) risk.** Inverse variance weighting (IVW) MR (IVW-MR) of
 208 three plasma proteins on lung function and/or COPD risk. The figure shows the IVW-MR plot for PDE4D
 209 (a and b), MICB (c) and RGAP1 (d). Dots represent the effect of the SNPs used for the IVW-MR on the
 210 plasma protein levels (horizontal axis) and lung function traits or COPD risk (vertical axis). Estimates
 211 were derived from 1) a plasma genome-wide association study (GWAS) for each protein, 2) GWAS of
 212 the International COPD Genetics Consortium (ICGC) for COPD risk and 3) a Spirometry GWAS meta-
 213 analysis (UK biobank and SpiroMeta cohorts) for the forced expiratory volume in one second (FEV₁) and
 214 FEV₁/Forced vital capacity (FVC). Error bars for each SNP represents the 95% confidence intervals. The
 215 slope of the red line is the instrumental variable regression estimate of the effect of protein on the lung
 216 function traits and/or COPD risk. IVW-MR *P-value* is shown at the top left corner of each MR plot.
 217

218 **Discussion**

219 Translating COPD genetics findings into actionable biomarkers and therapeutic targets
220 requires understanding of the genes and proteins underlying the genetic associations. To our
221 knowledge, this is the largest integrative proteomics and GWAS for lung function and COPD to
222 date. The key findings were that: 1) the genetic loci associated with plasma levels of 1,048 and
223 250 unique proteins colocalized with lung function and COPD risk, respectively; 2) using an
224 standard MR approach, we identified 4 unique plasma proteins causally associated with lung
225 function and/or COPD risk; and 3) using an step-wise approach (COLOC coupled with MR) we
226 identified 37 candidates proteins for lung function and/or COPD.

227 By integrating lung function and COPD genetics and blood proteomics, we identified
228 537, 607, and 250 unique proteins, whose plasma levels significantly associated with FEV₁,
229 FEV₁/FVC, or COPD risk, respectively. Of these, 74 were common to all three traits. This is a
230 substantial increase in the number of peripheral proteins associated with COPD phenotypes when
231 compared to previous reports, which had focused on candidate genes (Obeidat et al. 2017; Milne
232 et al. 2020). Our findings suggest that the expression levels of plasma proteins are associated
233 with lung function and COPD risk, supporting the notion that GWAS variants (even when
234 located within non-coding regions) have potential biological consequences that ultimately
235 contribute to the phenotypic variation of a complex trait or disease. Our findings could be used
236 as a starting point to support future biomarker and/or drug developmental studies. This is
237 particular true since targets with genetic support are more likely to be successful in drug
238 development (Nelson et al. 2015).

239 Of the proteins we identified in this study, there are several with known, biologically-
240 plausible links to lung physiology and lung disease. For example, PDE4D is an isoenzyme that is
241 part of the phosphodiesterase subfamily 4 (PDE4 A-D), a group of proteins that are involved in
242 the pathogenesis of multiple inflammatory diseases, including but not limited to asthma and
243 COPD. Inhibition of PDE4 increases cAMP, which appears to have a positive effect in asthma
244 and COPD by decreasing lung inflammation (Huang and Mancini 2006) and inducing airway
245 smooth muscle relaxation (Méhats et al. 2003). PDE4 isoforms have different physiological roles
246 (Manning et al. 1999), and the specific role of the PDE4D isoform in lung pathology is not fully
247 understood. Importantly, pathological contribution may be tissue specific. For example, cigarette

248 exposure has opposite effects on PDE4D expression in alveolar compared to airway epithelial
249 tissue (Zuo et al. 2019). In our study, PDE4D was measured in the plasma compartment, and the
250 direction of the MR estimate suggested that increased PDE4D plasma levels were causally
251 associated with increased lung function and reduced the risk of COPD. It is therefore plausible
252 that a soluble form of PDE4D has different effects than the intracellular form or ones that are
253 expressed in the local lung or immune cells. Further research is needed to understand the role of
254 the soluble PDE4D in the pathophysiology of COPD.

255 Another interesting candidate we found was IL3-RA. Our results suggested that this
256 protein shares causal loci with lung function, and that increased plasma levels of IL3-RA are
257 associated with decreased risk of COPD. IL3-RA is one of the two subunits of the interleukin-3
258 (IL3) receptor, a cytokine produced mostly by T-cells and involved in several
259 immunopathologies (Hercus et al. 2013). IL3 signalling may be important for surfactant
260 homeostasis (Campo et al. 2012), alveolar macrophage function (Notarangelo and Pessach
261 2008), and activation and recruitment of eosinophils (George and Brightling 2016) (Davoine and
262 Lacy 2014). Surfactant homeostasis appears to be a contributing factor in obstructive lung
263 diseases, including asthma, COPD and cystic fibrosis (Devendra and Spragg 2002). Although
264 this concept has not been fully validated by experimental studies, it is plausible that soluble IL3-
265 RA in plasma may disrupt IL3 signalling via competitive binding inhibition. This potential
266 function of IL3-RA in plasma may warrant further investigation especially in the context of
267 COPD.

268 We found that increased sRAGE plasma protein levels were causally associated with
269 decreased lung function. sRAGE is the soluble isoform of the receptor for advanced glycation
270 end-products (RAGE). Under normal conditions this receptor is mainly expressed in lung tissue,
271 particularly in type I alveolar epithelial cells, and mediates proinflammatory responses (Buckley
272 and Ehrhardt 2010). sRAGE acts as a decoy for RAGE since it is capable of binding to RAGE
273 ligands, therefore inhibiting RAGE (Demling et al. 2006). Dysregulation of sRAGE has been
274 linked to COPD and lung function. Healthy individuals have been reported to have higher levels
275 of sRAGE compared to COPD patients (Gopal et al. 2012). In contrast, our results suggest that in
276 a general population increased sRAGE is linked to decreased lung function. This is in agreement
277 with a study of current smokers, where a missense variant in the gene encoding for RAGE was
278 associated with lower serum sRAGE levels and increased lung function (Miller et al. 2016).

279 Differential effects of sRAGE may therefore depend on an individual's genotype. To fully
280 understand these conflicting results it is crucial to study RAGE regulation in COPD and in
281 normal healthy individuals.

282 We also found that increased MICB levels were associated with decreased COPD risk.
283 Variants within the *MICB* gene have previously been associated with lung function (Soler
284 Artigas et al. 2011). *MICB* belongs to a major histocompatibility complex class I chain-related
285 (MIC) gene family (Bahram 2000). MICB is a cellular stress-induced molecule that contributes
286 to the innate and adaptive immune responses (Jamieson et al. 2002; Carapito and Bahram 2015).
287 Soluble MICB is decreased or undetectable in bronchial washes of smokers with normal lung
288 function and COPD patients compared to those of never-smokers (Roos-Engstrand et al. 2010).
289 This is in agreement with the direction of the MICB effect suggested by our study, and may
290 therefore be an interesting plasma protein candidate for further studies.

291 Our study has a number of limitations. First, the cohorts used for our analysis only
292 included white-European individuals; therefore the conclusions based on our results may not be
293 generalisable to populations of different ancestry. Second, our study was limited to the blood
294 plasma proteome. Since COPD is a systemic disease, studying the proteome of other tissues and
295 cell types may further elucidate the mechanisms underlying the GWAS associations. Third, the
296 2,995 proteins found in the INTERVAL Study data set represent ~15% of the whole human
297 proteome. It is probable that proteins not measured on this platform may contribute to the
298 phenotypic variability of the traits; these proteins remain undiscovered. Furthermore, it is likely
299 that other mechanisms that we did not explore (e.g. epigenetics and gene expression) could
300 contribute to COPD risk. Fourth, our integrative-omics approach only explored cis regions,
301 therefore future research should also evaluate the trans or distal regions associated with the traits.
302 Lastly, although our approach allowed us to prioritize a manageable set of proteins, further
303 efforts are necessary to validate the role of these peripheral proteins and their relationship to lung
304 function and COPD.

305 In summary, our integrative-omics approach revealed several novel plasma proteins that
306 were significantly linked with COPD risk and/or lung function. Using a MR framework, we
307 provide evidence suggesting that the plasma levels of multiple proteins have causal effects on
308 these phenotypic traits. These proteins represent promising candidates for future development of

309 biomarkers and/or therapeutic targets in COPD and other lung pathologies associated with
310 reduced lung function.

311 **Materials and methods**

312 *Datasets analysed for the current study*

313 *INTERVAL study*

314 We analysed plasma protein quantitative trait loci (pQTL) obtained from the INTERVAL
315 study. The INTERVAL study was a randomized trial of blood donation intervals that comprises
316 around 45,000 participants that were recruited between June 11, 2012, and June 15, 2014 (Di
317 Angelantonio et al. 2017; Sun et al. 2018). The full details on the criteria for participants'
318 recruitment, informed consent, description of the cohort, sample collection, INTERVAL study
319 design, and objectives have been previously published (Di Angelantonio et al. 2017; Sun et al.
320 2018). Briefly, participants from the INTERVAL study were aged 18 years and older, in general
321 good health (based on blood donation criteria), and were recruited at 25 static donor centres of
322 NHS Blood and Transplant (NHSBT). Blood collection was performed using standard
323 venepuncture. Participants were genotyped for about 830,000 genetic variants using the
324 Affymetrix Axiom UK Biobank genotyping array and imputed to the 1000 Genome phase 3
325 UK10K reference panel. Genetic variants with imputation score (r^2) >0.7 , Hardy-Weinberg
326 Equilibrium (HWE) $P > 5 \times 10^{-6}$ and minor allele count of > 8 were retained (10,572,788 genetic
327 variants). A full description of the genotyping protocol and quality control has been previously
328 described (Astle et al. 2016).

329 After filtering out participants who failed to pass the genetic quality controls (HWE,
330 minor allele count, r^2), a randomly-selected subset of 3,301 participants was used for the plasma
331 pQTL analyses of 2,995 plasma proteins levels. The protein levels were log-transformed and
332 adjusted for age, sex, waiting period between blood collection and processing and the first three
333 genetic principal components. The protein residuals then were extracted and rank-inverse
334 normalized. Later these residuals were used for genome-wide associations study (GWAS). The
335 genetic associations were tested with linear regression using an additive genetic model and the
336 results from each cohort were combined using a fixed-effect inverse-variance meta-analysis.

337 Significant associations were defined at $P < 1.5 \times 10^{-11}$. We used the totality of the pQTL results
338 (summary statistics) for the subsequent analyses described later in this section.

339

340 *International COPD Genetics Consortium (ICGC) study*

341 We determined the relationship between plasma proteins and COPD risk using the ICGC
342 dataset. The ICGC study is one of the largest COPD GWASs that has been conducted to date.
343 Briefly, this cohort included over 200,000 participants (35,735 cases and 222,076 controls) from
344 individual COPD GWASs (Sakornsakolpat et al. 2019). Each individual study obtained informed
345 consent from individual participants, and approval from the local human research
346 ethics/regulatory bodies (Cho et al. 2014; Hobbs et al. 2017; Sakornsakolpat et al. 2019). COPD
347 cases were defined based on pre-bronchodilator spirometry, with $FEV_1 < 80\%$ predicted and
348 FEV_1 to FVC ratio of < 0.70 . Controls were defined as $FEV_1 > 80\%$ predicted and $FEV_1/FVC >$
349 0.70 . Each COPD GWAS was evaluated using logistic regression, which adjusted for age, sex,
350 pack-year of smoking, ever smoking status, current smoking status, and genetic ancestry
351 (principal component, as required for each study). The GWAS results were combined using a
352 fixed-effects meta-analysis. A full description of the cohort and study methods have been
353 previously published (Cho et al. 2014; Hobbs et al. 2017; Sakornsakolpat et al. 2019).

354

355 *UK biobank and SpiroMeta lung function meta-analysis*

356 We used genome-wide associations with lung function (FEV_1 and FEV_1/FVC) from a
357 previously-published meta-analysis of the UK Biobank and SpiroMeta datasets (Shrine et al.
358 2019). This meta-analysis included 321,047 and 79,055 white European participants from the
359 UK Biobank project (Bycroft et al. 2018) and the SpiroMeta consortium, respectively.

360 In the UK Biobank, phenotypic information was collected in 22 recruitment centres
361 across the United Kingdom (Bycroft et al. 2018). Human research ethics approval for the UK
362 Biobank project was granted by the North West Multi-centre Research Ethics Committee
363 (MREC). Participants were genotyped with the Affymetrix Axiom UK BiLEVE and UK biobank
364 array (Wain et al. 2015) and imputed to the Haplotype Reference Consortium panel; genotypes
365 were retained if the minor allele count was ≥ 3 and imputation $r^2 > 0.5$ (Bycroft et al. 2018).

366 Association testing was conducted between genome-wide-significant genotypes and lung
367 function traits (FEV_1 and FEV_1/FVC) using a linear mixed model (LMM), assuming additive
368 genetic effects (Loh et al. 2015). The LMM was adjusted for confounders (age, age², sex, height,
369 smoking status, and genotyping array) (Shrine et al. 2019).

370 The SpiroMeta consortium includes participants from 22 individual studies, which have
371 been previously described (Shrine et al. 2019). Approval was granted for each individual study
372 from their respective local human research ethics committees. Participants were genotyped
373 according to the protocol described in each of the studies; 13 studies were imputed to the 1000
374 Genomes Project Phase 1 panel (1000 Genomes Project Consortium et al. 2010) and nine to the
375 Haplotype Reference Consortium panel (McCarthy et al. 2016). Genotypes with imputation
376 $r^2 < 0.30$ were excluded from further analysis (Shrine et al. 2019). For each individual study,
377 genetic associations with FEV_1 and FEV_1/FVC were determined using a linear regression model
378 adjusted for age, age², sex and height. Genetic principal components were included as covariates
379 in studies of unrelated participants, while LMM were used in studies of related participants to
380 account for kinship and population structure. The results across all studies in the SpiroMeta
381 consortium were combined using an inverse-variance weighted meta-analysis. Shrine and
382 colleagues (Shrine et al. 2019) then combined the UK Biobank and SpiroMeta GWAS results
383 using an inverse-variance-weighted fixed-effects meta-analysis. In total, 19,819,130 genetic
384 variants (imputed or genotyped) in both cohorts (UK biobank and SpiroMeta consortium) were
385 used for the meta-analysis. The LD regression score intercept was close to 1 (Shrine et al. 2019),
386 therefore no genomic control was applied.

387

388 *Integrative -omics methods*

389 *Bayesian colocalization (COLOC).*

390 We used COLOC to determine whether the associations between lung function traits
391 (FEV_1 , FEV_1/FVC , and COPD risk) and plasma protein levels were consistent with a causal
392 variant (colocalization). We performed the analysis using the coloc package (Giambartolomei et
393 al. 2014) implemented in R (R Core Team 2018). We evaluated the summary statistics for 3,248
394 plasma pQTLs (Sun et al. 2018) from the INTERVAL study. We included all SNPs associated
395 with lung function (FEV_1 or FEV_1/FVC) in the UK Biobank/SpiroMeta GWAS meta-analysis

396 (19,819,130) (Shrine et al. 2019) and all SNPs associated with COPD risk in the ICGC dataset
397 (6,224,355). In order to maximise the number of variants tested per genomic locus, which
398 overlapped across these studies, we did not set an *a priori* p value threshold for inclusion.

399 The COLOC method used for our research calculates the posterior probability (*PP*) of
400 colocalization of lung function (or COPD risk) and plasma protein-associated variants within a
401 defined genomic region. In summary, a large *PP* is supportive evidence of a single shared causal
402 variant for the two traits. For this test we applied the following parameters: 1) a prior probability
403 of a SNP being associated with the lung function trait or plasma protein level to 1×10^{-04} ; and 2) a
404 prior probability of a SNP being associated with the lung function trait and plasma protein level
405 to 1×10^{-06} . All SNPs across the genome were assumed to have equal prior probabilities (as set
406 before). In addition, we defined loci to be tested as genomic regions ± 0.5 Mb windows
407 surrounding the top lung function-associated (or COPD risk-associated) variants. As a result, we
408 applied COLOC to 988,183 loci for FEV₁, 1,109,654 for FEV₁/FVC and 282,338 for COPD risk.
409 We placed the significance threshold of the COLOC analysis to be $PP_{H4} > 0.80$.

410 *Mendelian Randomization*

411 We used a MR approach to identify causal relationships between plasma proteins
412 (“exposure”) and the three traits (FEV₁, FEV₁/FVC, and COPD risk; “traits”). We first identified
413 pQTLs for each of the selected proteins in the INTERNAL study by extracting the effect size
414 (Beta) and standard error (SE) of each variant that at least reached the arbitrary threshold of $P <$
415 5×10^{-06} and excluded the variants that were found within a 2Mb window (± 1 Mb) to avoid
416 linkage. Next, we examined the complex trait associations in the UK Biobank/SpiroMeta meta-
417 analysis (FEV₁, FEV₁/FVC) and ICGC (COPD risk) for these variants and, if present, extracted
418 the Beta value and SE for each. We then used these genetic variants as instrumental variables
419 (IVs) in a MR analysis. The fundamental assumptions of MR analysis are: 1) that the IVs are
420 associated with the exposure; 2) that the selected IVs only affect an outcome *via* the exposure;
421 and 3) that the IVs are independent of confounders. Using two MR methods, inverse variance
422 weighting (IVW) MR (IVW-MR) and Egger-MR we aimed to identify causal risk factors. For
423 each exposure-outcome pairing, we used an inverse variance weighted linear regression model
424 (IVW-MR) to relate the per-allele SNP association with the exposure to its association with the
425 trait. IVW-MR assumes no directional pleiotropy (i.e. genetic variant associated with multiple

426 unrelated phenotypes) by constraining the intercept to zero, and accounts for linkage
427 disequilibrium (LD) between genetic variants, heterogeneity was assessed based on the
428 Cochran's Q test ($P < 0.05$) that is part of the IVW-MR outcome (Burgess et al. 2016). We also
429 performed Egger-MR, which accounts for directional pleiotropy by un-constraining the intercept
430 (Bowden et al. 2015).

431 Based on this workflow, we identified a protein as having a causal association with the
432 complex trait if the IVW-MR estimate was significant $FDR < 0.1$ and the Egger-MR intercept
433 was not different from zero ($P > 0.05$). In addition, using an *a priori* hypothesis that the
434 colocalized proteins were causally related to lung function traits and/or COPD, we determined
435 their significant causal associations based on the IVW-MR threshold of $P_{MR} < 0.05$ and Egger-
436 MR intercept $P > 0.05$.

437

438 *Potential Biomarkers*

439 Using the workflow described in **Figure 1**, we aimed to develop a list of potential
440 candidate proteins that warrant further investigation for their role in lung function and/or COPD
441 pathogenesis. We selected as top plasma proteins those that showed both significant
442 colocalization at $PP_{H4} > 0.80$ and $P_{MR} < 0.05$ with lung function and/or COPD risk.

443 **Article information**

444

445 **Data availability:** All data used for these analyses are publicly available at:
446 <http://www.phpc.cam.ac.uk/ceu/proteins/>; <https://www.ebi.ac.uk/gwas/publications/30804560>;
447 and <https://www.ebi.ac.uk/gwas/publications/30804561>.

448

449 **Conflict of Interest:**

450 D.D.S. has received research funding from AstraZeneca for an investigator-initiated research
451 project and received honoraria for speaking engagements from Boehringer Ingelheim and
452 AstraZeneca over the past 36 months. S.M. reports personal fees from Novartis and Boehringer-
453 Ingelheim, outside the submitted work.

454

455 **Funding:**

456 A.I.H.C. and S.M. are supported by MITACS accelerate and Providence Airway Centre

457

458 **Acknowledgement:**

459 Compute Canada computer cluster was used to conduct the data analyses.

460

461 **References**

462 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, et al (2010) A map of human
463 genome variation from population-scale sequencing. *Nature* 467:1061–1073.
464 <https://doi.org/10.1038/nature09534>

465 Astle WJ, Elding H, Jiang T, et al (2016) The Allelic Landscape of Human Blood Cell Trait
466 Variation and Links to Common Complex Disease. *Cell* 167:1415-1429.e19.
467 <https://doi.org/10.1016/j.cell.2016.10.042>

468 Bahram S (2000) MIC genes: from genetics to biology. *Adv Immunol* 76:1–60.
469 [https://doi.org/10.1016/s0065-2776\(01\)76018-x](https://doi.org/10.1016/s0065-2776(01)76018-x)

470 Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments:
471 effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44:512–
472 525. <https://doi.org/10.1093/ije/dyv080>

473 Buckley ST, Ehrhardt C (2010) The receptor for advanced glycation end products (RAGE) and
474 the lung. *J Biomed Biotechnol* 2010:917108. <https://doi.org/10.1155/2010/917108>

475 Burgess S, Dudbridge F, Thompson SG (2016) Combining information on multiple instrumental
476 variables in Mendelian randomization: comparison of allele score and summarized data
477 methods. *Stat Med* 35:1880–1906. <https://doi.org/10.1002/sim.6835>

478 Bycroft C, Freeman C, Petkova D, et al (2018) The UK Biobank resource with deep phenotyping
479 and genomic data. *Nature* 562:203–209. <https://doi.org/10.1038/s41586-018-0579-z>

480 Campo I, Kadija Z, Mariani F, et al (2012) Pulmonary alveolar proteinosis: diagnostic and
481 therapeutic challenges. *Multidisciplinary Respiratory Medicine* 7:4.
482 <https://doi.org/10.1186/2049-6958-7-4>

483 Carapito R, Bahram S (2015) Genetics, genomics, and evolutionary biology of NKG2D ligands.
484 *Immunol Rev* 267:88–116. <https://doi.org/10.1111/imr.12328>

485 Cho MH, McDonald M-LN, Zhou X, et al (2014) Risk loci for chronic obstructive pulmonary
486 disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2:214–
487 225. [https://doi.org/10.1016/S2213-2600\(14\)70002-5](https://doi.org/10.1016/S2213-2600(14)70002-5)

- 488 Davoine F, Lacy P (2014) Eosinophil cytokines, chemokines, and growth factors: emerging roles
489 in immunity. *Front Immunol* 5:570. <https://doi.org/10.3389/fimmu.2014.00570>
- 490 Demling N, Ehrhardt C, Kasper M, et al (2006) Promotion of cell adherence and spreading: a
491 novel function of RAGE, the highly selective differentiation marker of human alveolar
492 epithelial type I cells. *Cell Tissue Res* 323:475–488. <https://doi.org/10.1007/s00441-005-0069-0>
- 494 Devendra G, Spragg RG (2002) Lung surfactant in subacute pulmonary disease. *Respiratory*
495 *Research* 3:11. <https://doi.org/10.1186/rr168>
- 496 Di Angelantonio E, Thompson SG, Kaptoge S, et al (2017) Efficiency and safety of varying the
497 frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors.
498 *Lancet* 390:2360–2371. [https://doi.org/10.1016/S0140-6736\(17\)31928-1](https://doi.org/10.1016/S0140-6736(17)31928-1)
- 499 George L, Brightling CE (2016) Eosinophilic airway inflammation: role in asthma and chronic
500 obstructive pulmonary disease. *Ther Adv Chronic Dis* 7:34–51.
501 <https://doi.org/10.1177/2040622315609251>
- 502 Giambartolomei C, Vukcevic D, Schadt EE, et al (2014) Bayesian test for colocalisation between
503 pairs of genetic association studies using summary statistics. *PLoS Genet* 10:e1004383.
504 <https://doi.org/10.1371/journal.pgen.1004383>
- 505 Gopal P, Rutten EPA, Dentener MA, et al (2012) Decreased plasma sRAGE levels in COPD:
506 influence of oxygen therapy. *Eur J Clin Invest* 42:807–814.
507 <https://doi.org/10.1111/j.1365-2362.2012.02646.x>
- 508 Gusev A, Ko A, Shi H, et al (2016) Integrative approaches for large-scale transcriptome-wide
509 association studies. *Nature Genetics* 48:245–252. <https://doi.org/10.1038/ng.3506>
- 510 Hercus TR, Dhagat U, Kan WLT, et al (2013) Signalling by the βc family of cytokines. *Cytokine*
511 *Growth Factor Rev* 24:189–201. <https://doi.org/10.1016/j.cytogfr.2013.03.002>
- 512 Hobbs BD, de Jong K, Lamontagne M, et al (2017) Genetic loci associated with chronic
513 obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis.
514 *Nat Genet* 49:426–432. <https://doi.org/10.1038/ng.3752>
- 515 Huang Z, Mancini JA (2006) Phosphodiesterase 4 inhibitors for the treatment of asthma and
516 COPD. *Curr Med Chem* 13:3253–3262. <https://doi.org/10.2174/092986706778773040>
- 517 Jamieson AM, Diefenbach A, McMahon CW, et al (2002) The role of the NKG2D
518 immunoreceptor in immune cell activation and natural killing. *Immunity* 17:19–29.
519 [https://doi.org/10.1016/s1074-7613\(02\)00333-3](https://doi.org/10.1016/s1074-7613(02)00333-3)
- 520 Lamontagne M, Bérubé J-C, Obeidat M, et al (2018) Leveraging lung tissue transcriptome to
521 uncover candidate causal genes in COPD genetic associations. *Hum Mol Genet* 27:1819–
522 1829. <https://doi.org/10.1093/hmg/ddy091>

- 523 Loh P-R, Tucker G, Bulik-Sullivan BK, et al (2015) Efficient Bayesian mixed-model analysis
524 increases association power in large cohorts. *Nat Genet* 47:284–290.
525 <https://doi.org/10.1038/ng.3190>
- 526 Manning CD, Burman M, Christensen SB, et al (1999) Suppression of human inflammatory cell
527 function by subtype-selective PDE4 inhibitors correlates with inhibition of PDE4A and
528 PDE4B. *Br J Pharmacol* 128:1393–1398. <https://doi.org/10.1038/sj.bjp.0702911>
- 529 McCarthy S, Das S, Kretschmar W, et al (2016) A reference panel of 64,976 haplotypes for
530 genotype imputation. *Nat Genet* 48:1279–1283. <https://doi.org/10.1038/ng.3643>
- 531 Méhats C, Jin S-LC, Wahlstrom J, et al (2003) PDE4D plays a critical role in the control of
532 airway smooth muscle contraction. *FASEB J* 17:1831–1841.
533 <https://doi.org/10.1096/fj.03-0274com>
- 534 Miller S, Henry AP, Hodge E, et al (2016) The Ser82 RAGE Variant Affects Lung Function and
535 Serum RAGE in Smokers and sRAGE Production In Vitro. *PLoS One* 11:.
536 <https://doi.org/10.1371/journal.pone.0164041>
- 537 Milne S, Li X, Cordero AIH, et al (2020) Protective effect of club cell secretory protein (CC-16)
538 on COPD risk and progression: a Mendelian randomisation study. *Thorax* 75:934–943.
539 <https://doi.org/10.1136/thoraxjnl-2019-214487>
- 540 Nelson MR, Tipney H, Painter JL, et al (2015) The support of human genetic evidence for
541 approved drug indications. *Nature Genetics* 47:856–860. <https://doi.org/10.1038/ng.3314>
- 542 Notarangelo LD, Pessach I (2008) Out of breath: GM-CSFR α mutations disrupt surfactant
543 homeostasis. *Journal of Experimental Medicine* 205:2693–2697.
544 <https://doi.org/10.1084/jem.20082378>
- 545 Obeidat M, Hao K, Bossé Y, et al (2015) Molecular mechanisms underlying variations in lung
546 function: a systems genetics analysis. *Lancet Respir Med* 3:782–795.
547 [https://doi.org/10.1016/S2213-2600\(15\)00380-X](https://doi.org/10.1016/S2213-2600(15)00380-X)
- 548 Obeidat M, Li X, Burgess S, et al (2017) Surfactant protein D is a causal risk factor for COPD:
549 results of Mendelian randomisation. *European Respiratory Journal* 50:.
550 <https://doi.org/10.1183/13993003.00657-2017>
- 551 R Core Team (2018) R: A language and environment for statistical computing. R Foundation for
552 Statistical Computing. <https://www.r-project.org/>. Accessed 5 Sep 2019
- 553 Roos-Engstrand E, Pourazar J, Behndig AF, et al (2010) Cytotoxic T cells expressing the co-
554 stimulatory receptor NKG2 D are increased in cigarette smoking and COPD. *Respiratory*
555 *Research* 11:128. <https://doi.org/10.1186/1465-9921-11-128>
- 556 Roth GA, Abate D, Abate KH, et al (2018) Global, regional, and national age-sex-specific
557 mortality for 282 causes of death in 195 countries and territories, 1980–2017: a

- 558 systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 392:1736–
559 1788. [https://doi.org/10.1016/S0140-6736\(18\)32203-7](https://doi.org/10.1016/S0140-6736(18)32203-7)
- 560 Sakornsakolpat P, Prokopenko D, Lamontagne M, et al (2019) Genetic landscape of chronic
561 obstructive pulmonary disease identifies heterogeneous cell-type and phenotype
562 associations. *Nat Genet* 51:494–505. <https://doi.org/10.1038/s41588-018-0342-2>
- 563 Shrine N, Guyatt AL, Erzurumluoglu AM, et al (2019) New genetic signals for lung function
564 highlight pathways and chronic obstructive pulmonary disease associations across
565 multiple ancestries. *Nat Genet* 51:481–493. <https://doi.org/10.1038/s41588-018-0321-7>
- 566 Smith GD, Ebrahim S (2003) “Mendelian randomization”: can genetic epidemiology contribute
567 to understanding environmental determinants of disease? *Int J Epidemiol* 32:1–22.
568 <https://doi.org/10.1093/ije/dyg070>
- 569 Soler Artigas M, Loth DW, Wain LV, et al (2011) Genome-wide association and large-scale
570 follow up identifies 16 new loci influencing lung function. *Nat Genet* 43:1082–1090.
571 <https://doi.org/10.1038/ng.941>
- 572 Sun BB, Maranville JC, Peters JE, et al (2018) Genomic atlas of the human plasma proteome.
573 *Nature* 558:73–79. <https://doi.org/10.1038/s41586-018-0175-2>
- 574 Voight BF, Peloso GM, Orho-Melander M, et al (2012) Plasma HDL cholesterol and risk of
575 myocardial infarction: a mendelian randomisation study. *Lancet* 380:572–580.
576 [https://doi.org/10.1016/S0140-6736\(12\)60312-2](https://doi.org/10.1016/S0140-6736(12)60312-2)
- 577 Wain LV, Shrine N, Miller S, et al (2015) Novel insights into the genetics of smoking behaviour,
578 lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic
579 association study in UK Biobank. *Lancet Respir Med* 3:769–781.
580 [https://doi.org/10.1016/S2213-2600\(15\)00283-0](https://doi.org/10.1016/S2213-2600(15)00283-0)
- 581 Zuo H, Faiz A, van den Berge M, et al (2019) Cigarette smoke exposure alters
582 phosphodiesterases in human structural lung cells. *American Journal of Physiology-Lung
583 Cellular and Molecular Physiology* 318:L59–L64.
584 <https://doi.org/10.1152/ajplung.00319.2019>
- 585
- 586