

1 **Proteomic associations with forced expiratory volume – a Mendelian randomisation study**

2 Axelsson GT^{1,2}, Jonmundsson T¹, Woo YJ³, Frick EA¹, Aspelund T^{1,5}, Loureiro JJ³, Orth AP⁴, Jennings LL³

3 Gudmundsson G^{5,6}, Emilsson V^{1,5}, Gudmundsdottir V^{1,5}, and Gudnason V^{1,5}

4 1: Icelandic Heart Association, Holtasmari 1, IS-201 Kopavogur, Iceland.

5 2: Landspítali University Hospital, Department of Internal Medicine, 101 Reykjavik, Iceland

6 3: Novartis Institutes for Biomedical Research, 22 Windsor Street, Cambridge, MA 02139, USA

7 4: Novartis Institutes for Biomedical Research, 10675 John Jay Hopkins Drive, San Diego, CA 92121,

8 USA

9 5: Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

10 6: Landspítali University Hospital, Department of Respiratory Medicine and Sleep, 108 Reykjavik,

11 Iceland

12 Correspondence:

13 Keywords: Lung function tests, Forced expiratory volume, Proteomics, Mendelian randomisation

14 **Word count:**

15 Main text: 3528

16 Abstract: 251

17 Author contributions:

18 Conceptualization, methodology, and design of the study: GTA, TJ, VaG, GG, VE and ViG.

19 Data acquisition: VaG, TA, LLJ, JLL, APO, VE, ViG

20 Analysis of the data: GTA, TJ, VaG, EAF, TA, YJW, JLL

21 Administration and supervision: VaG, LLJ, GG, VE, ViG

22 Drafting of the initial manuscript: GTA, TJ, VaG and ViG.

23 All authors revised and/or edited the manuscript for scientific content.

24 Funding. The AGES-Reykjavik study was funded by Icelandic Heart Association, National Institute on
25 Aging contract N01-AG-12100 and HHSN271201200022C, and Althingi (the Icelandic Parliament).

26 GTA was supported by the Eimskip University Fund from the University of Iceland. GG was supported
27 by the University of Iceland and by Landspítali University Hospital. TJ and VaG were supported by the
28 University of Iceland, VaG by its postdoctoral fund. All protein measurements were funded by
29 Novartis.

30 Competing interests: The study was supported by the Novartis Institute for Biomedical Research,
31 and protein measurements for the AGES-Reykjavik cohort were performed at SomaLogic. JLL, YJW.
32 and LLJ are employees and stockholders of Novartis. GTA and GG report travel support from

- 33 Boehringer-Ingelheim for work not related to the study. No other potential conflicts of interest
- 34 relevant to this article were reported.

35 Abstract

36 A decline in forced expiratory volume (FEV1) is a hallmark of obstructive respiratory diseases, an
37 important cause of morbidity among the elderly. While some data exist on biomarkers that are
38 related to FEV1, we sought to do a systematic analysis of causal relations of biomarkers with FEV1.

39 Data from the general population-based AGES-Reykjavik study were used. Proteomic measurements
40 were done using 4,782 DNA aptamers (SOMAmers). Data from 1,648 participants with spirometric
41 data were used to assess the association of SOMAmer measurements with FEV1 using linear
42 regression. Bi-directional Mendelian randomisation (MR) analyses were done to assess causal
43 relations of observationally associated SOMAmers with FEV1, using genotype and SOMAmer data
44 from 5,368 AGES-Reykjavik participants and genetic associations with FEV1 from a publicly available
45 GWAS (n = 400,102).

46 In observational analyses, 473 SOMAmers were associated with FEV1 after multiple testing
47 adjustment. The most significant were R-Spondin 4, Alkaline Phosphatase, Placental Like 2 and
48 Retinoic Acid Receptor Responder 2. Of the 235 SOMAmers with genetic data, eight were associated
49 with FEV1 in MR analyses. Three were directionally consistent with the observational estimate,
50 Thrombospondin 2 (THBS2), Endoplasmic Reticulum Oxidoreductase 1 Beta and Apolipoprotein M.
51 THBS2 was further supported by a colocalization analysis. Analyses in the reverse direction, testing
52 whether changes in SOMAmer levels were caused by changes in FEV1, were performed but no
53 significant associations were found after multiple testing adjustments.

54 In summary, this large scale proteogenomic analyses of FEV1 reveals protein markers of FEV1, as
55 well as several proteins with potential causality to lung function.

56

57 Introduction

58 Chronic respiratory diseases such as chronic obstructive pulmonary disease (COPD) are a leading
59 global cause of mortality and morbidity, with their relative importance increasing in the last decades
60 (1). Diagnosis of COPD is based on obstruction on pulmonary function testing, by a low forced
61 expiratory volume in one second (FEV_1) relative to the forced vital capacity (FVC), and a progressive
62 decline in pulmonary function is a hallmark of the disease (2). While it is undisputed that exposure to
63 external harmful stimuli such as cigarette smoke and biomass fumes are substantial risk factors for
64 pulmonary function decline, intrinsic factors such as genetics and gene-environment interactions
65 play a significant part as well (2-4). Genome wide association studies (GWAS) have found several
66 genetic polymorphisms that are associated with COPD and lung function decline, including
67 polymorphisms in or near genes encoding matrix metalloproteinase 12, nicotinic acetylcholine
68 receptor, hedgehog interacting protein, glutathione S-transferase, C-terminal domain-containing
69 protein (5-9) and the antiprotease alpha-1-antitrypsin (10). In addition to genetic polymorphisms,
70 biomarkers that predict COPD or lung function have been discovered. Among them are inflammatory
71 markers, soluble receptor for advanced glycoprotein end products (AGER), club cell secretory
72 protein 16 (SCGB1A1) and surfactant protein D (SFTPD) (11-14). For several of these biomarkers,
73 estimations of potential causality have been made. AGER, SCGB1A1 and SFTPD have been suggested
74 to be causally associated with COPD or with lung function, while analyses have pointed against such
75 an association for the inflammatory markers CRP and IL-6 (15-19). Still, while substantial
76 epidemiologic data exist regarding single genetic markers and biomarkers that predict COPD or lung
77 function, a systematic large-scale analysis of observational and causal associations of protein
78 markers with lung function has not been undertaken to our knowledge.

79 Proteomics have emerged as a way of exploring biologic predictors of disease, especially with the
80 advent of methods that allow for measurement and evaluation of thousands of proteins in biological
81 samples from participants of large cohort studies (20). In addition to predicting disease and disease-
82 related outcomes, integration of genetic data allows one to make assumptions regarding causality
83 with proteomic markers. Mendelian randomization (MR) is such a method and utilizes genetic
84 polymorphisms as instrumental variables to assess the relationship of an exposure with an outcome.
85 As chromosomal alleles are randomly allocated during gamete formation, this methodology allows
86 one to avoid the effect of confounders and to infer causality from epidemiologic data (21).

87 The aim of the study was to systematically explore the associations of a multitude of protein markers
88 with lung function and then to assess the potential causal relationships of these protein markers
89 with lung function by use of bi-directional Mendelian randomization.

90 Methods

91 Study phenotyping

92 The Age/Gene Environment Susceptibility (AGES)-Reykjavik study is a population-based cohort study
93 of 5,764 elderly Icelanders that was carried out between 2002 and 2006. The participants, aged from
94 66 to 96 years (mean 76 years), were all prior participants of the Reykjavik Study done decades
95 earlier. As part of AGES-Reykjavik, the participants underwent extensive phenotyping by
96 questionnaires, physiological measurements, imaging studies and laboratory measurements, during
97 a single clinic visit. The study was approved by the Icelandic National Bioethics Committee (VSN-00-
98 063) and the Institutional Review Board of the Intramural Program of the National Institute for
99 Aging, with informed consent obtained from all participants. Further details of the study design are
100 previously published (22).

101 A subset of the study participants underwent lung function testing in a standardised manner. The
102 device used was a Vitalograph Gold Standard Plus (Vitalograph Ltd., UK). Each participant completed
103 three attempts. Participants with spirometry of acceptable quality were included ($n = 1,660$).
104 Spirometry measures from participants that completed at least two attempts with a no more than
105 300 ml difference between the attempts and exhalation for at least 6 seconds were deemed
106 acceptable (23). Smoking history was ascertained from questionnaires while anthropometric
107 measurements were done during the clinic visit. Protein measurements were done in serum samples
108 from participants using a high throughput proteomics technology, the SOMAscan (SomaLogic,
109 Boulder, CO) in which DNA aptamers (Slow-Off Rate Modified Aptamers (SOMAmers)) bind to target
110 protein epitopes and are then quantified with the help of fluorescence after wash-out of unbound
111 aptamers and proteins. Measurements in AGES-Reykjavik were done with a 5,034 SOMAmer
112 platform in serum from 5,457 AGES-Reykjavik participants, targeting 4,135 individual human
113 proteins. For the analyses, 4,782 SOMAmers targeting human proteins were used. Measurement
114 data were transformed using Box-Cox transformation and extreme outliers were excluded, as
115 previously described (24).

116 Statistical analyses

117 A flow chart of study design is shown in Figure 1. Descriptive statistics were compiled for
118 participants with acceptable pulmonary function tests, demographic covariate data and protein
119 measurements available ($n = 1,648$). Using data from these participants, the association of all
120 measured human SOMAmers ($n = 4,782$) with FEV₁ was assessed using linear regression modelling.
121 These analyses were adjusted for variables that are commonly used to predict FEV₁ in clinical
122 practice, age, sex, height, age squared, and height squared (25). Adjustment for multiple testing was

123 done using the Benjamini-Hochberg False Discovery Rate (FDR). To understand how tobacco smoking
124 affected these associations, the analyses were repeated with participants stratified by smoking
125 history (ever-smokers versus never-smokers). Genes encoding proteins significantly associated with
126 the outcome after adjustment for multiple testing were subjected to over-representation analysis of
127 Gene Ontology terms (26). In secondary analyses we explored the protein associations with
128 additional outcomes, including FVC, FEV₁/FVC ratio and COPD. Additional covariates included
129 smoking status, pack years of smoking, weight and eGFR.

130 For SOMAmers that were associated with FEV₁ after adjustment for multiple testing, a bi-
131 directional two-sample MR analysis was performed for the association of SOMAmers with FEV₁. MR
132 analyses were done for SOMAmers that were observationally associated with FEV₁ and had genetic
133 instruments available. Genetic instruments for SOMAmers were defined as follows. The associations
134 of single-nucleotide polymorphisms (SNPs) with SOMAmers were calculated using data from 5,368
135 participants of AGES-Reykjavik, as previously published and described in detail (24). The possible
136 instrumental variables for each SOMAmer were defined as SNPs located within a cis-window near
137 the gene coding for the protein measured by the SOMAmer, defined as within 500 kb up- and
138 downstream of the gene. SNPs with a window-wide significant association ($p < 0.05/\text{number of SNPs}$
139 in cis-window) with a given SOMAmer were considered as potential instruments. For each gene,
140 SNPs were filtered based on linkage disequilibrium (LD; $r^2 < 0.2$) or distance (>1 mb) using Plink v1.9.
141 Instruments were considered valid if $F > 10$ for the association of instruments with SOMAmers. The
142 associations of the instrumental variables with FEV₁ were obtained from a GWAS of lung function in
143 which data from two cohorts, UKBiobank and SpiroMeta Consortium, were meta-analysed. The total
144 number of participants in that analysis was 400,102 (27). A MR analysis in the reverse direction was
145 performed using all observationally associated SOMAmers as outcomes. Genetic instruments for
146 FEV₁ were defined based on the same GWAS on lung function as described above (27). For these
147 analyses, SNPs associated with FEV₁ ($p < 5 \times 10^{-8}$) were used as instruments after clumping within a
148 cis-window 5 mb up- and downstream of the gene, by an LD threshold of 0.01. The associations of
149 the instrumental variables with SOMAmer levels were obtained from AGES-Reykjavik data (24).

150 For the MR analyses, SNPs were harmonised using the TwoSampleMR R package. The MR
151 estimates for FEV₁ were obtained using the generalized weighted least squares (GWLS) method,
152 accounting for correlation between instruments (28), except for instances where only one
153 instrument was available, in which Wald ratios were calculated. For each MR analysis with more
154 than two instruments, sensitivity analyses using the weighted median and Egger estimators were
155 done to assess the validity of instruments and limit the effect of pleiotropic associations,
156 respectively. Results were considered to pass these sensitivity analyses when the following

157 conditions were met. For the weighted median estimator, the weighted median estimate had to be
158 significant and directionally consistent with the GWLS estimate and for Egger, the Egger estimate
159 had to be directionally consistent with the GWLS estimate and the intercept not significant. In
160 analyses in the reverse direction, estimates were obtained using the inverse variance weighted
161 method. Adjustment for multiple testing was done using the Benjamini-Hochberg False Discovery
162 Rate (FDR).

163 A colocalization analysis was performed to provide additional causal support for analytes
164 associated with FEV1 in MR analyses. In that aim, AGES-Reykjavik serum pQTLs (24), summary
165 statistics for FEV1 (29) and plasma pQTLs from deCODE (30) were harmonized to account for strand
166 orientation and differences in genome builds. All studies were lifted over to build GRCh38 for
167 colocalization when needed. For the QTLs, only putative cis-regulatory variants were examined
168 defined by 500kb away from the gene body. All regions with variants that had associations of $P <$
169 1×10^{-5} were fine-mapped using Sum of Single Effects (SuSiE) (31) based on 1000 genome population
170 reference (32). 95% credible-sets were filtered and colocalization between GWAS and QTLs was
171 performed using fastENLOC using the posterior inclusion probabilities estimated in SuSiE (31, 33).
172 Regional colocalization probability (RCP) was calculated by summing the colocalization probability
173 within each 95% credible-set and filtered colocalization results at $RCP > 80\%$ for further examination.
174 RCP represents the probability that a given genomic region contains a single colocalized variant (34).
175 Visualization was done using Julia 1.7 and libraries within (35).

176

177 Results

178 The protein profile of pulmonary function

179 Descriptive statistics for participants that had pulmonary function testing data are shown in Table 1.
180 A majority of participants were female (56%) and participants were on average 76 years old. Most
181 participants had a history of smoking (59%).

182 Of the 4,782 SOMAmers tested, 473 were observationally associated with FEV₁ after adjustment for
183 multiple testing (Table 2, Table S1, Figure 2). The most significantly associated SOMAmers measured
184 were R-Spondin 4 (RSPO4, $\beta = -0.099$, $p = 1.13 \times 10^{-13}$), Alkaline Phosphatase, Placental Like 2 (ALPPL2,
185 $\beta = -0.084$, $p = 4.76 \times 10^{-11}$), Retinoic Acid Receptor Responder 2 (RARRES2, $\beta = -0.093$, $p = 7.03 \times 10^{-11}$),
186 Sushi, Von Willebrand Factor Type A, EGF And Pentraxin Domain Containing 1 (SVEP1, $\beta = -0.085$, $p =$
187 1.67×10^{-10}) and Polymeric Immunoglobulin Receptor (PIGR, $\beta = -0.080$, 1.80×10^{-10}). Results for
188 proteins that have been previously suggested as biomarkers of FEV₁ (11) are shown in Table S2. Prior
189 associations of SFTPD ($\beta = -0.05$, $p = 0.0002$), CRP ($\beta = -0.06$, $p = 1.2 \times 10^{-7}$), fibrinogen ($\beta = -$
190 0.04 , $p = 0.0008$ for the stronger associated SOMAmer), IL6 ($\beta = -0.05$, $p = 9.6 \times 10^{-5}$ for the
191 stronger associated SOMAmer) and eotaxin ($\beta = -0.05$, $p = 0.0002$) were reproduced in the AGES-
192 Reykjavik data while associations for other suggested biomarkers of FEV₁ were not.

193 The proteins most significantly associated with FEV1 overall also had strong associations with FEV1
194 among ever-smokers, such as RSPO4 ($\beta = -0.106$, $p = 4.83 \times 10^{-9}$), RARRES2 ($\beta = -0.11$, $p = 1.19 \times 10^{-8}$)
195 and ALPPL2 ($\beta = -0.092$, $p = 5.73 \times 10^{-8}$). Trefoil Factor 2 (TFF2, $\beta = -0.095$, $p = 3.84 \times 10^{-8}$), Complement
196 9 (C9, $\beta = -0.095$, $p = 5.65 \times 10^{-8}$) and TIMP Metallopeptidase Inhibitor 1 (TIMP1, $\beta = -0.122$, $p =$
197 1.18×10^{-7}) also had notably strong associations with the outcome. However, associations of these
198 proteins were much weaker among never-smokers, with the strongest associations observed for two
199 SOMAmers measuring SVEP1 ($\beta = -0.081$, $p = 4.52 \times 10^{-6}$ for the stronger associated SOMAmer) (Table
200 S3) in this subgroup. The proteins associated with FEV1 were most strongly enriched for Gene
201 Ontology (GO) terms related to humoral immune response, regulation of neurogenesis, compliment
202 activation and extracellular matrix organization (Figure S1, Table S4).

203 Secondary analyses showed a large overlap between the protein profile of FEV₁ compared to that of
204 FVC, FEV₁/FVC ratio and COPD, with some additional proteins detected (204 significant SOMAmers in
205 primary analysis were among 497 significant SOMAmers in secondary analyses; Tables S8-S15).

206 Mendelian randomization analysis

207 Of 473 SOMAmers associated with FEV₁, 235 (50%) had genetic instruments available and were
208 included in MR analyses of the association of SOMAmers with FEV₁. The instruments are shown in

209 Table S5 for significant SOMAmers. Of the 235 SOMAmers tested, eight were significantly associated
210 with FEV₁ in MR analyses (Table 3), suggesting they may have a causal effect on lung function. Of
211 those, three, Thrombospondin 2 (THBS2, $\beta = -0.037$, $p = 9.53 \times 10^{-5}$), Endoplasmic Reticulum
212 Oxidoreductase 1 Beta (ERO1B, $\beta = -0.025$, $p = 8.05 \times 10^{-4}$) and Apolipoprotein M (APOM, $\beta = 0.053$, p
213 $= 9.72 \times 10^{-4}$) were directionally consistent with the observational analyses. The other five SOMAmers
214 with significant causal estimates, R-Spondin-2 (RSPO2), TIMP Metalloproteinase Inhibitor 4 (TIMP4),
215 CD14, Heparin Binding Growth Factor (HDGF) and Parkinsonism Associated Deglycase (PARK7) were
216 directionally inconsistent. Table 3 and Figure 4 show the 29 SOMAmers that were nominally
217 associated with FEV₁ (unadjusted $p < 0.05$) using MR and passed weighted median sensitivity testing.
218 Among nominally associated SOMAmers, the directional consistency between causal and
219 observational estimates was low, or 38%, suggesting that the potentially causal effects of the protein
220 are generally not reflected in the observational estimates. Of the five previously suggested
221 biomarkers of FEV₁ listed in Table S2 that were observationally associated with FEV₁, only three
222 proteins had genetic instruments available, the acute phase reactants CRP and fibrinogen as well as
223 eotaxin. None of these proteins were significantly associated with FEV₁ in MR analyses (Table S6).

224 The 29 proteins identified from the MR analysis (Table 3) were further examined for
225 colocalization evidence. Among the 8 proteins with significant MR association, a 3' UTR variant
226 (rs3253) from the FEV1 GWAS colocalized with THBS2 protein expression in AGES-Reykjavik ($P =$
227 1.4×10^{-19} , $\beta = -0.18$, RCP=0.9976). The same variant was associated with THBS2 protein expression in
228 the deCODE cohort ($P = 2.1 \times 10^{-186}$, $\beta = -0.26$) (Table 4, Figure 4). Among 22 proteins with nominal
229 association with MR analysis, an upstream variant (rs4968200) from the FEV1 GWAS colocalized with
230 TNFSF12 protein levels in AGES-Reykjavik ($P = 5.6 \times 10^{-133}$, $\beta = 0.63$, RCP = 0.9995).

231 Finally, all 473 SOMAmers observationally associated with FEV₁ were included in a MR
232 analysis in the reverse direction, i.e., to evaluate if the changes in SOMAmer levels are downstream
233 of changes in lung function. FEV₁ was not causally associated with levels of any SOMAmers after
234 adjustment for multiple testing. Data for the 18 SOMAmers that FEV₁ was nominally associated with
235 ($p < 0.05$) in the reverse-MR analysis are shown in Table S7.

236

237

238 Discussion

239 We present findings from a proteomic analysis of pulmonary function with more candidate protein
240 analytes than previously published to our knowledge (11), highlighting several proteins as strong
241 markers of FEV₁. Stratification by smoking shows that most associations are driven by ever-smokers.
242 Mendelian randomisation was systematically applied to the candidate markers, revealing proteins
243 whose levels may have a causal effect on lung function. Reverse causation analyses failed to
244 demonstrate that protein level changes associated with lung function occur downstream of the
245 phenotype change, although this could partly be due to insufficient power in the FEV1 GWAS.
246 Among the proteins identified through MR analysis, probabilistic colocalization pointed towards
247 *THBS2* and *TNFSF12* as putative upstream targets for COPD.

248 Eight proteins were suggested to be causally implicated in lung function based on the MR analysis.
249 However, only three (*THBS2*, *ERO1B* and *APOM*) had consistent direction of effect for the
250 observational and causal estimates. Such discrepancy has been observed when comparing causal
251 and observational estimates (36, 37) for serum proteins. Based on probabilistic quantification, a 3'
252 UTR variant within *THBS2* was identified as a putative causal variant for FEV1 changes and this lead
253 variant colocalized with *THBS2* protein expression in AGES-Reykjavik (Table 4, Figure 4). The effect of
254 this variant was replicated in an independent cohort for *THBS2* protein levels (Table 4, Figure 4).
255 Because directions of effects were consistent across the datasets and *THBS2* had non-reversible
256 causal association to FEV1, supported by colocalization, *THBS2* may be a candidate target for COPD.
257 *THBS2* is an extracellular matrix protein that has been implicated in various cardiovascular disorders
258 and is also a candidate biomarker for non-small cell lung cancer (38, 39). *THBS2* is involved in tissue
259 repair and interacts with many different ligands in the extracellular matrix, among them matrix
260 metalloproteases and elastase (40). Although mechanism of *THBS2* needs further experimental
261 validation, our work suggests that protein levels of *THBS2* may be a target for extracellular matrix
262 and regenerative properties associated with COPD. Meanwhile, *ERO1B* is a disulfide oxidase in the
263 endoplasmic reticulum that is shown to predict survival in pancreatic and pulmonary cancers (41-43)
264 and *APOM*, an apolipoprotein that is mainly a component of high density lipoproteins and has been
265 associated with COPD severity (44). Genetic variants flanking this gene have been associated with
266 obstructive spirometry measurements (45). However, it must be kept in mind that the association of
267 *APOM* is based on a single SNP, rs2736163, which is intronic to *PRRC2A*.

268 Despite not reaching the study threshold for statistical significance, some proteins that were
269 nominally associated in MR analyses are of interest. First, alpha-1-antitrypsin (*SERPINA1*) is the best-
270 known protein known to cause COPD as severe deficiency of alpha-1-antitrypsin results in

271 obstructive lung disease (46). However, in our data, serum levels of alpha-1-antitrypsin are inversely
272 associated with FEV₁ in both observational and MR analyses, contrary to what would possibly be
273 expected, although this directionality is known from previous observational analyses of FEV₁ and
274 explained by alpha-1-antitrypsin's role as an acute phase reactant (47). Also, polymorphisms that
275 cause mild or intermediate alpha-1-antitrypsin deficiency are not consistently associated with
276 decreased lung function, suggesting that levels of the protein may only affect lung function below a
277 threshold level (48). Second, TNFSF12 was a protein with a colocalizing pQTL variant with the FEV₁
278 GWAS (Table 4, Figure 4). TNFSF12 is a member of the Tumor Necrosis Factor (TNF) superfamily, of
279 which one key cytokine, TNF-alpha, is a well-known protein target disrupted in COPD patients (49).
280 Despite a concrete association between TNF-alpha levels and COPD in both human and murine
281 models (49-51), TNF-alpha monoclonal antibody was a failure as a COPD therapy (52). This is likely
282 due to COPD being a highly complex inflammatory disease in which single cytokine blockage may not
283 be sufficient for treating the disease. These additional TNF family proteins with MR based causal
284 evidence and probabilistic colocalization may be novel targets for COPD patients impacted by TNF-
285 alpha associated pathobiology. Notable other nominally significant or directionally inconsistent
286 proteins in the findings are matrix metalloproteinase 8 (MMP-8), one of the proteases
287 observationally associated with lung function and implicated in the pathogenesis of COPD (53, 54),
288 TIMP4, an inhibitor of matrix metalloproteinases that has been shown to be upregulated in COPD
289 patients (55) and CD14, levels of which have been shown to be elevated in lungs of smokers (56).
290 While only 8 proteins were significantly associated with FEV₁ in MR analyses, FEV₁ was not
291 associated with any proteins in analyses in the reverse direction. The data were therefore unable to
292 support prior causal analyses involving inflammatory markers that have suggested reverse causality
293 of FEV₁ with inflammatory markers (18) (Table S6).

294 The study reveals novel markers with strong observational relations to lung function such as RSPO4,
295 a signalling molecule that is part of the Wnt signalling pathway (57), the tumor marker ALPL2 (58),
296 the adipokine chemerin (RARRES2), and SVEP1, a protein that is thought to play a role in
297 inflammation in atherosclerosis (59). In addition, the findings validate the observational associations
298 of some of the previously suggested protein markers of FEV₁, such as SFTPD, fibrinogen, eotaxin and
299 CRP (Table S2) (13), while the associations of other previously suggested biomarkers of FEV₁ such as
300 AGER were not corroborated in this study (13, 60). Finally, the findings show that proteins that take
301 part in immune responses and extracellular matrix modulation are over-represented among the
302 proteins related to FEV₁.

303 This work is subject to a number of limitations. First, the study is based on SOMAmer technology, a
304 relatively novel aptamer-based technology for protein measurements. While many of these

305 SOMAmers have been validated with encouraging results, it has been pointed out that a minority of
306 SOMAmers could be subject to cross-reactivity with related or homologous proteins (20). Second,
307 both the AGES-Reykjavik cohort and the UKBiobank and SpiroMeta Consortium are of European
308 ancestry (27). Therefore, the observed associations could not be generalizable to other populations.
309 Additionally, the AGES-Reykjavik cohort is older than the UKBiobank and SpiroMeta Consortium
310 which could distort comparisons between observational and genetic findings. Third, the measure of
311 lung function used in this paper, FEV₁, is disproportionately impaired in obstructive lung disease and is
312 a key diagnostic parameter for such diseases. While COPD is the archetypal obstructive lung disease,
313 especially among the elderly, the presence of other respiratory diseases such as asthma could
314 influence the findings, especially in younger cohorts. Fourth, subtle differences in genetic structure
315 between AGES, UKBiobank, and SpiroMeta cohorts is present and may be contributing to the lack of
316 colocalization for the other 6 MR identified proteins. For instance, very few MR instruments used
317 overlapped with 95% credible-sets identified (Supplementary Table 16). In comparisons to
318 UKBioBank, missing variants in AGES-Reykjavik cohort may contribute to colocalization false
319 negatives. Lastly, many causal estimates are directionally inconsistent with observational estimates.
320 While this phenomenon is known from previous proteogenomic studies, its reasons are unclear.

321 In conclusion, this proteogenomic analysis reveals several proteins that are potentially causally
322 related to lung function, most notably THBS2, ERO1B and APOB.

323

324 Figures

325 Figure 1 – A flow chart of the study design

326 Figure 2 – A volcano plot showing the observational associations of all 4782 SOMAmers with FEV1

327 Figure 3 – A forest plot showing the observational and mendelian randomisation estimates for FEV1
328 for causally associated SOMAmers

329 Figure 4. Colocalization plot for *THBS2*. FEV1 GWAS (29), AGES pQTL (24), and DECODE pQTL (30)
330 data are shown from top to bottom. Grey circles are individual SNPs from each study. Both purple
331 diamond and red vertical line represent the lead variant (rs3253, 3' UTR variant). X-axis is genomic
332 position within chromosome 6 and y-axis is $-\log_{10}$ transformed P-values. Purple horizontal line
333 delineates GWAS threshold at 5×10^{-8} and yellow vertical lines represent gene boundary for *THBS2*.
334 Visualization is restricted to 150,000 bp upstream and downstream of *THBS2*. Linkage disequilibrium
335 (r^2) within the region is plotted at the bottom in green.

336

337 **Tables**

338 Table 1 – Overview of study participants

n	1648
Sex = Female (%)	917 (55.6)
Age (mean (SD))	76.28 (5.70)
Height (mean (SD))	167.22 (9.48)
Weight (mean (SD))	75.61 (14.67)
BMI (mean (SD))	26.97 (4.38)
Eversmoker (%) [†]	969 (59.4)
FEV ₁ (mean (SD))	2.12 (0.69)
FEV ₁ % of predicted (mean (SD))*	0.89 (0.21)
FVC (mean (SD))	2.97 (0.83)
FVC % of predicted (mean (SD))*	0.92 (0.17)
FEV ₁ /FVC < 70% (%)	578 (35.1)

339 [†] 17 participants had missing smoking data.

340 * Predicted values based on Hankinson et al. (25)

341 Table 2 – Observational associations of highly-associated proteins with FEV₁

SOMA	EGS	β	95% CI	P	FDR P
8464_31_3	RSPO4	-0.099	-0.125 - -0.073	1.13×10^{-13}	5.39×10^{-10}
7813_6_3	ALPPL2	-0.084	-0.109 - -0.059	4.76×10^{-11}	1.12×10^{-7}
3079_62_2	RARRES2	-0.093	-0.121 - -0.065	7.03×10^{-11}	1.12×10^{-7}
11178_21_3	SVEP1	-0.085	-0.111 - -0.059	1.67×10^{-10}	1.72×10^{-7}
3216_2_2	PIGR	-0.08	-0.104 - -0.056	1.80×10^{-10}	1.72×10^{-7}
11109_56_3	SVEP1	-0.084	-0.11 - -0.058	3.39×10^{-10}	2.70×10^{-7}
2292_17_4	C9	-0.079	-0.103 - -0.054	4.55×10^{-10}	3.11×10^{-7}
12549_33_3	HPGDS	0.079	0.054 - 0.103	6.02×10^{-10}	3.60×10^{-7}
6379_62_3	ADAMTSL2	-0.082	-0.109 - -0.055	2.40×10^{-9}	1.27×10^{-6}
9191_8_3	TFF2	-0.076	-0.101 - -0.051	3.47×10^{-9}	1.66×10^{-6}
8841_65_3	CILP2	0.074	0.049 - 0.099	4.47×10^{-9}	1.95×10^{-6}
8323_163_3	TFF3	-0.079	-0.106 - -0.053	4.98×10^{-9}	1.99×10^{-6}
6390_18_3	NPS	0.077	0.051 - 0.103	8.04×10^{-9}	2.96×10^{-6}
2211_9_6	TIMP1	-0.096	-0.129 - -0.062	2.00×10^{-8}	6.50×10^{-6}
13722_105_3	C9	-0.07	-0.094 - -0.046	2.06×10^{-8}	6.50×10^{-6}
6605_17_3	IGFALS	0.07	0.046 - 0.095	2.18×10^{-8}	6.50×10^{-6}
12707_26_3	DPYSL3	0.07	0.046 - 0.095	2.75×10^{-8}	7.72×10^{-6}
5728_60_3	FCRL1	0.069	0.044 - 0.093	5.29×10^{-8}	1.40×10^{-5}
6075_61_3	HEXB	0.071	0.045 - 0.097	9.08×10^{-8}	2.22×10^{-5}
4496_60_2	MMP12	-0.069	-0.094 - -0.044	9.28×10^{-8}	2.22×10^{-5}
6225_3_3	PRSS8	0.072	0.046 - 0.099	1.07×10^{-7}	2.43×10^{-5}
4337_49_2	CRP	-0.063	-0.087 - -0.04	1.18×10^{-7}	2.57×10^{-5}
10620_21_3	MSMB	-0.07	-0.096 - -0.044	1.28×10^{-7}	2.67×10^{-5}
2609_59_2	CST3	-0.075	-0.103 - -0.047	2.09×10^{-7}	4.05×10^{-5}
8885_6_3	CACNA2D3	0.067	0.042 - 0.092	2.11×10^{-7}	4.05×10^{-5}

342 Results of linear regression analyses adjusted for sex, age, age squared, height and height squared.

343 Results are shown for the proteins with the most significant associations.

344

345 Table 3 – Results of Mendelian randomisation analysis of the association of proteins with FEV₁

SOMA	Protein	nSNP	B - MR	SE	P - MR	FDR P - MR	Egger	B - Obs	P - Obs	FDR P - Obs
8409_3_3	RSPO2	3	0.109	0.023	2.44×10⁻⁶	5.73×10⁻⁴	Yes	-0.064	1.52×10⁻⁶	1.75×10⁻⁴
3339_33_1	THBS2	24	-0.037	0.009	9.53×10⁻⁵	7.85×10⁻³	Yes	-0.062	1.96×10⁻⁶	2.09×10⁻⁴
6462_12_3	TIMP4	17	0.028	0.007	1.00×10⁻⁴	7.85×10⁻³	Yes	-0.05	1.70×10⁻⁴	5.29×10⁻³
8969_49_3	CD14	6	0.084	0.024	5.85×10⁻⁴	0.034	Yes	-0.045	2.21×10⁻³	0.029
7994_41_3	ERO1B	12	-0.025	0.007	8.05×10⁻⁴	0.038	Yes	-0.036	3.78×10⁻³	0.042
14125_5_3	APOM	1	0.053	0.016	9.72×10⁻⁴	0.038	-	0.041	1.12×10⁻³	0.018
8953_47_3	HDGF	7	0.028	0.009	1.26×10⁻³	0.041	Yes	-0.054	5.10×10⁻⁵	2.26×10⁻³
9845_33_3	PARK7	2	-0.056	0.017	1.40×10⁻³	0.041	-	0.036	4.63×10⁻³	0.048
2668_70_2	CAPN1	3	-0.051	0.016	1.94×10 ⁻³	0.051	Yes	0.04	2.93×10 ⁻³	0.035
3290_50_2	CD109	49	-0.013	0.005	4.91×10 ⁻³	0.111	Yes	0.044	4.23×10 ⁻⁴	0.01
2797_56_2	APOB	9	-0.029	0.011	5.18×10 ⁻³	0.111	Yes	0.054	2.39×10 ⁻⁵	1.28×10 ⁻³
3580_25_8	SERPINA1	9	-0.027	0.01	6.03×10 ⁻³	0.118	Yes	-0.055	3.11×10 ⁻⁵	1.61×10 ⁻³
12395_86_3	DARS2	2	-0.058	0.022	9.64×10 ⁻³	0.164	-	-0.037	3.09×10 ⁻³	0.036
5939_42_3	TNFSF12	77	0.011	0.004	0.01	0.164	Yes	0.037	3.33×10 ⁻³	0.038
8427_118_3	RSPO3	17	0.033	0.013	0.01	0.164	Yes	-0.06	1.68×10 ⁻⁵	9.92×10 ⁻⁴
2630_12_2	IL1RAP	58	0.006	0.002	0.018	0.259	No	0.037	3.38×10 ⁻³	0.038
5070_76_3	TNFRSF6B	3	0.174	0.075	0.02	0.283	Yes	-0.044	3.14×10 ⁻⁴	8.25×10 ⁻³
8992_1_3	TMEM2	1	-0.052	0.023	0.025	0.312	-	0.043	5.41×10 ⁻⁴	0.012
8288_27_3	APOH	5	-0.032	0.014	0.025	0.312	Yes	0.044	7.37×10 ⁻⁴	0.014
13130_150_3	HK2	2	-0.04	0.019	0.032	0.338	-	-0.042	8.84×10 ⁻⁴	0.015
9870_17_3	WARS	5	-0.021	0.01	0.032	0.338	Yes	0.041	1.31×10 ⁻³	0.02
4297_62_3	SPON1	8	-0.026	0.012	0.032	0.338	Yes	-0.052	1.17×10 ⁻⁴	3.93×10 ⁻³
2789_26_2	MMP7	8	0.02	0.009	0.033	0.338	Yes	-0.054	2.53×10 ⁻⁵	1.34×10 ⁻³
9172_69_3	MMP8	13	-0.014	0.007	0.036	0.343	Yes	-0.041	1.03×10 ⁻³	0.017
5646_20_3	RNASE6	33	0.008	0.004	0.037	0.343	Yes	-0.048	1.95×10 ⁻⁴	5.80×10 ⁻³
4546_27_3	ADGRE2	29	0.012	0.006	0.039	0.343	Yes	0.057	7.36×10 ⁻⁶	5.41×10 ⁻⁴
3449_58_2	SERPINA4	56	-0.007	0.004	0.039	0.343	Yes	0.04	1.42×10 ⁻³	0.021
10746_24_3	DKK3	6	-0.031	0.015	0.044	0.356	Yes	0.049	2.10×10 ⁻⁴	6.09×10 ⁻³
12387_7_3	PDLIM4	11	-0.025	0.012	0.044	0.356	Yes	0.046	5.08×10 ⁻⁴	0.011

346 Shown are data for SOMAmers that had nominally significant associations using Mendelian
 347 randomisation and passed weighted median sensitivity analyses. In the column with results from
 348 Egger analyses, “Yes” denotes passing the sensitivity test, “No” denotes failing and “-” denotes
 349 insufficient number of SNPs to perform the test. Observational data are adjusted for sex, age, age
 350 squared, height and height squared.

351 **Bold** = Significantly associated with FEV₁ in MR analyses after multiple testing correction.

352 *Italic* = Consistent in direction between MR analyses and observational analyses

353

354 Table 4. Statistics for the variants with strong colocalization (RCP > 0.9)

gene	SNP	RCP (FEV1 x AGES)	GWAS P-value [beta]	AGES P-value [beta]	deCODE P-value [beta]
THBS2	rs3253 ¹	0.9976	1.3x10 ⁻¹⁰ [0.02]	1.4x10 ⁻¹⁹ [-0.18]	2.1x10 ⁻¹⁸⁶ [-0.26]
THBS2	rs7756742 ¹	0.9775	6.3x10 ⁻¹⁰ [0.02]	2.8x10 ⁻¹⁹ [-0.19]	-- ³
THBS2	rs7771838 ¹	0.971	7.0x10 ⁻¹⁰ [0.02]	1.8x10 ⁻¹⁹ [-0.19]	-- ³
TNFSF12	rs4968200 ²	0.9995	4.5x10 ⁻¹¹ [0.02]	5.6x10 ⁻¹¹³ [0.63]	-- ³

355 RCP: Regional Colocalization Probability

356 ¹Lead variant in FEV1 GWAS is rs3253

357 ²Lead variant in FEV1 GWAS is rs4968200

358 ³Variant data was not available in DECODE

359

360

361

362 References

- 363 1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369
364 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for
365 the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-22.
- 366 2. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global Strategy for
367 the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *Am J Respir*
368 *Crit Care Med*. 2013;187(4):347-65.
- 369 3. Rennard SI, Vestbo J. COPD: the dangerous underestimate of 15%. *The Lancet*.
370 2006;367(9518):1216-9.
- 371 4. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Lam KB, et al. Respiratory risks from
372 household air pollution in low and middle income countries. *The Lancet Respiratory medicine*.
373 2014;2(10):823-60.
- 374 5. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, Lasky-Su J, et al. MMP12,
375 Lung Function, and COPD in High-Risk Populations. *N Engl J Med*. 2009;361(27):2599-608.
- 376 6. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A
377 are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42(3):200-2.
- 378 7. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A Genome-Wide Association Study
379 in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci.
380 *PLOS Genetics*. 2009;5(3):e1000421.
- 381 8. Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, et al. Effect of five
382 genetic variants associated with lung function on the risk of chronic obstructive lung disease, and
383 their joint effects on lung function. *Am J Respir Crit Care Med*. 2011;184(7):786-95.
- 384 9. Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for
385 chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The*
386 *Lancet Respiratory Medicine*. 2014;2(3):214-25.
- 387 10. Wood AM, Stockley RA. Alpha One Antitrypsin Deficiency: From Gene to Treatment.
388 *Respiration*. 2007;74(5):481-92.
- 389 11. Serban KA, Pratte KA, Bowler RP. Protein Biomarkers for COPD Outcomes. *Chest*.
390 2021;159(6):2244-53.
- 391 12. Faner R, Tal-Singer R, Riley JH, Celli B, Vestbo J, MacNee W, et al. Lessons from ECLIPSE: a
392 review of COPD biomarkers. *Thorax*. 2014;69(7):666.
- 393 13. Zemans RL, Jacobson S, Keene J, Kechris K, Miller BE, Tal-Singer R, et al. Multiple biomarkers
394 predict disease severity, progression and mortality in COPD. *Respir Res*. 2017;18(1):117.
- 395 14. Gan WQ, Man SFP, Senthilselvan A, Sin DD. Association between chronic obstructive
396 pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. *Thorax*.
397 2004;59(7):574.
- 398 15. Keefe J, Yao C, Hwang S-J, Courchesne P, Lee GY, Dupuis J, et al. An Integrative Genomic
399 Strategy Identifies sRAGE as a Causal and Protective Biomarker of Lung Function. *Chest*.
400 2022;161(1):76-84.
- 401 16. Milne S, Li X, Hernandez Cordero AI, Yang CX, Cho MH, Beaty TH, et al. Protective effect of
402 club cell secretory protein (CC-16) on COPD risk and progression: a Mendelian randomisation study.
403 *Thorax*. 2020;75(11):934.
- 404 17. Obeidat M, en, Li X, Burgess S, Zhou G, Fishbane N, et al. Surfactant protein D is a causal risk
405 factor for COPD: results of Mendelian randomisation. *Eur Respir J*. 2017;50(5):1700657.
- 406 18. Dahl M, Vestbo J, Zacho J, Lange P, Tybjaerg-Hansen A, Nordestgaard BG. C reactive protein
407 and chronic obstructive pulmonary disease: a Mendelian randomisation approach. *Thorax*.
408 2011;66(3):197.

- 409 19. van Durme YMTA, Lahousse L, Verhamme KMC, Stolk L, Eijgelsheim M, Loth DW, et al.
410 Mendelian Randomization Study of Interleukin-6 in Chronic Obstructive Pulmonary Disease.
411 *Respiration*. 2011;82(6):530-8.
- 412 20. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory
413 networks of human serum proteins link genetics to disease. *Science*. 2018;361(6404):769-73.
- 414 21. Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. *JAMA*. 2017;318(19):1925-6.
- 415 22. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, et al. Age,
416 Gene/Environment Susceptibility – Reykjavik Study: Multidisciplinary Applied Phenomics. *Am J*
417 *Epidemiol*. 2007;165(9):1076-87.
- 418 23. Gudmundsson G, Margretardottir OB, Sigurdsson MI, Harris TB, Launer LJ, Sigurdsson S, et
419 al. Airflow obstruction, atherosclerosis and cardiovascular risk factors in the AGES Reykjavik study.
420 *Atherosclerosis*. 2016;252:122-7.
- 421 24. Gudjonsson A, Gudmundsdottir V, Axelsson GT, Gudmundsson EF, Jonsson BG, Launer LJ, et
422 al. A genome-wide association study of serum proteins reveals shared loci with common diseases.
423 *Nat Commun*. 2022;13(1):480.
- 424 25. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the
425 general U.S. population. *Am J Respir Crit Care Med*. 1999;159(1):179-87.
- 426 26. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO::TermFinder--open source
427 software for accessing Gene Ontology information and finding significantly enriched Gene Ontology
428 terms associated with a list of genes. *Bioinformatics*. 2004;20(18):3710-5.
- 429 27. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New
430 genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease
431 associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-93.
- 432 28. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental
433 variables in Mendelian randomization: comparison of allele score and summarized data methods.
434 *Stat Med*. 2016;35(11):1880-906.
- 435 29. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New
436 genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease
437 associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-93.
- 438 30. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismisdottir EL, et al.
439 Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*.
440 2021;53(12):1712-21.
- 441 31. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "Sum
442 of Single Effects" model. *PLoS Genet*. 2022;18(7):e1010299.
- 443 32. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference
444 for human genetic variation. *Nature*. 2015;526(7571):68-74.
- 445 33. Wen X, Pique-Regi R, Luca F. Integrating molecular QTL data into genome-wide genetic
446 association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet*.
447 2017;13(3):e1006646.
- 448 34. Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen X. Probabilistic colocalization of
449 genetic variants from complex and molecular traits: promise and limitations. *Am J Hum Genet*.
450 2021;108(1):25-35.
- 451 35. Kim M, Vo DD, Kumagai ME, Jops CT, Gandal MJ. GeneticsMakie.jl: a versatile and scalable
452 toolkit for visualizing locus-level genetic and genomic data. *Bioinformatics*. 2023;39(1).
- 453 36. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the
454 human plasma proteome. *Nature*. 2018;558(7708):73-9.
- 455 37. Gudmundsdottir V, Zaghlool SB, Emilsson V, Aspelund T, Ilkov M, Gudmundsson EF, et al.
456 Circulating Protein Signatures and Causal Candidates for Type 2 Diabetes. *Diabetes*.
457 2020;69(8):1843-53.
- 458 38. Zhang K, Li M, Yin L, Fu G, Liu Z. Role of thrombospondin β 1 and thrombospondin β 2 in
459 cardiovascular diseases (Review). *Int J Mol Med*. 2020;45(5):1275-93.

- 460 39. Jiang YM, Yu DL, Hou GX, Jiang JL, Zhou Q, Xu XF. Serum thrombospondin-2 is a candidate
461 diagnosis biomarker for early non-small-cell lung cancer. *Biosci Rep*. 2019;39(7).
- 462 40. Calabro NE, Kristofik NJ, Kyriakides TR. Thrombospondin-2 and extracellular matrix
463 assembly. *Biochim Biophys Acta*. 2014;1840(8):2396-402.
- 464 41. Zito E, Chin KT, Blais J, Harding HP, Ron D. ERO1-beta, a pancreas-specific disulfide oxidase,
465 promotes insulin biogenesis and glucose homeostasis. *J Cell Biol*. 2010;188(6):821-32.
- 466 42. Asada K, Kobayashi K, Joutard S, Tubaki M, Takahashi S, Takasawa K, et al. Uncovering
467 Prognosis-Related Genes and Pathways by Multi-Omics Analysis in Lung Cancer. *Biomolecules*.
468 2020;10(4).
- 469 43. Zhu T, Gao YF, Chen YX, Wang ZB, Yin JY, Mao XY, et al. Genome-scale analysis identifies
470 GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. *Oncotarget*.
471 2017;8(13):21281-9.
- 472 44. Li H, Liu Y, Wang L, Shen T, Du W, Liu Z, et al. High apolipoprotein M serum levels correlate
473 with chronic obstructive pulmonary disease. *Lipids Health Dis*. 2016;15:59.
- 474 45. Burkart KM, Manichaikul A, Wilk JB, Ahmed FS, Burke GL, Enright P, et al. APOM and high-
475 density lipoprotein cholesterol are associated with lung function and per cent emphysema. *Eur*
476 *Respir J*. 2014;43(4):1003-17.
- 477 46. Strnad P, McElvaney NG, Lomas DA. Alpha1-Antitrypsin Deficiency. *N Engl J Med*.
478 2020;382(15):1443-55.
- 479 47. Senn O, Russi EW, Schindler C, Imboden M, von Eckardstein A, Brändli O, et al. Circulating
480 alpha1-antitrypsin in the general population: Determinants and association with lung function.
481 *Respir Res*. 2008;9(1):35.
- 482 48. Thun G-A, Ferrarotti I, Imboden M, Rochat T, Gerbase M, Kronenberg F, et al. SERPINA1 PiZ
483 and PiS Heterozygotes and Lung Function Decline in the SAPALDIA Cohort. *PLoS One*.
484 2012;7(8):e42728.
- 485 49. Tanni SE, Pelegriano NR, Angeleli AY, Correa C, Godoy I. Smoking status and tumor necrosis
486 factor-alpha mediated systemic inflammation in COPD patients. *J Inflamm (Lond)*. 2010;7:29.
- 487 50. Churg A, Wang RD, Tai H, Wang X, Xie C, Wright JL. Tumor necrosis factor-alpha drives 70%
488 of cigarette smoke-induced emphysema in the mouse. *Am J Respir Crit Care Med*. 2004;170(5):492-
489 8.
- 490 51. Yao Y, Zhou J, Diao X, Wang S. Association between tumor necrosis factor- α and chronic
491 obstructive pulmonary disease: a systematic review and meta-analysis. *Ther Adv Respir Dis*.
492 2019;13:1753466619866096.
- 493 52. Rennard SI, Fogarty C, Kelsen S, Long W, Ramsdell J, Allison J, et al. The safety and efficacy of
494 infliximab in moderate to severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*.
495 2007;175(9):926-34.
- 496 53. Koo HK, Hong Y, Lim MN, Yim JJ, Kim WJ. Relationship between plasma matrix
497 metalloproteinase levels, pulmonary function, bronchodilator response, and emphysema severity.
498 *Int J Chron Obstruct Pulmon Dis*. 2016;11:1129-37.
- 499 54. Churg A, Zhou S, Wright JL. Matrix metalloproteinases in COPD. *Eur Respir J*. 2012;39(1):197.
- 500 55. Navratilova Z, Zatloukal J, Kriegova EVA, Kolek V, Petrek M. Simultaneous up-regulation of
501 matrix metalloproteinases 1, 2, 3, 7, 8, 9 and tissue inhibitors of metalloproteinases 1, 4 in serum of
502 patients with chronic obstructive pulmonary disease. *Respirology*. 2012;17(6):1006-12.
- 503 56. Regueiro V, Campos MA, Morey P, Sauleda J, Agustí AGN, Garmendia J, et al.
504 Lipopolysaccharide-binding protein and CD14 are increased in the bronchoalveolar lavage fluid of
505 smokers. *Eur Respir J*. 2009;33(2):273.
- 506 57. de Lau WB, Snel B, Clevers HC. The R-spondin protein family. *Genome Biol*. 2012;13(3):242.
- 507 58. Su Y, Zhang X, Bidlingmaier S, Behrens CR, Lee NK, Liu B. ALPPL2 Is a Highly Specific and
508 Targetable Tumor Cell Surface Antigen. *Cancer Res*. 2020;80(20):4552-64.
- 509 59. Jung IH, Elenbaas JS, Alisio A, Santana K, Young EP, Kang CJ, et al. SVEP1 is a human coronary
510 artery disease locus that promotes atherosclerosis. *Sci Transl Med*. 2021;13(586).

511 60. Merali S, Barrero CA, Bowler RP, Chen DE, Criner G, Braverman A, et al. Analysis of the
512 plasma proteome in COPD: Novel low abundance proteins reflect the severity of lung remodeling.
513 COPD: Journal of Chronic Obstructive Pulmonary Disease. 2014;11(2):177-89.

514









