Genetic evidence supports the development of SLC26A9 targeting therapies for the treatment of lung disease

Jiafen Gong[1], Gengming He[1,2], Cheng Wang[1], Claire Bartlett[3] , Naim Panjwani[1], Fan Lin[1], Katherine Keenan[1,3], Julie Avolio[3], Anat Halevy[1], Michelle Shaw[3], Mohsen Esmaeili[1] , Guillaume Côté-Maurais[4], Damien Adam[4,5], Stéphanie Bégin[4], Candice Bjornson[6], Mark Chilvers[7], Joe Reisman[8], April Price[9], Michael Parkins[10], Richard Van Wylick[11], Yves Berthiaume [5], Lara Bilodeau[12], Dimas Mateos-Corral[13], Daniel Hughes[13], Mary J. Smith[14], Nancy Morrison[15], Janna Brusky[16], Elizabeth Tullis[17], Anne L. Stephenson[17], Bradley S. Quon[18], Pearce Wilcox[18], Winnie M. Leung[19], Melinda Solomon[20], Lei Sun [2,21], Emmanuelle Brochiero[4,5], Theo J. Moraes[3,20], Tanja Gonska[3,22], Felix Ratjen[3,23], Johanna M. Rommens[1,24], Lisa J. Strug[1,2,21,25, 26,*]

Affiliations:

[1]Program in Genetics and Genome Biology, The Hospital for Sick Children; Toronto, ON, Canada.

[2]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto; Toronto, ON, Canada.

[3]Translational Medicine, The Hospital for Sick Children; Toronto, ON, Canada.

[4]CRCHUM; Montréal, QC, Canada.

[5]Department of Medicine, Faculty of Medicine, Université de Montréal; Montréal, QC, Canada.

[6]Alberta Children's Hospital; Calgary, AB, Canada.

[7] British Columbia Children's Hospital; Vancouver, BC, Canada.

22 [8]The Children's Hospital of Eastern Ontario; Ottawa, ON, Canada.

23 [9] The Children's Hospital, London Health Science Centre; London, ON, Canada.

24 [10] Foothills Medical Centre; Calgary, AB, Canada.

25 [11] Kingston Health Sciences Centre; Kingston, ON, Canada.

26 [12]Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec-
27 Université Laval; Québec City, QC, Canada.

28 [13]IWK Health Centre; Halifax, NS, Canada.

29 [14] Faculty of Medicine, Memorial University of Newfoundland; St. John's, NL, Canada.

30 [15]Queen Elizabeth II Health Sciences Centre; Halifax, NS, Canada.

31 [16]Jim Pattison Children's Hospital; Saskatoon, SK, Canada.

32 [17]St. Michael's Hospital; Toronto, ON, Canada.

33 [18]St. Paul's Hospital; Vancouver, BC, Canada.

34 [19] University of Alberta Hospital; Edmonton, AB, Canada.

35 [20]Respiratory Medicine, Hospital for Sick Children; Toronto, ON, Canada.

36 [21]Department of Statistical Sciences, University of Toronto; Toronto, ON, Canada.

37 [22]Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children;
38 Toronto, ON, Canada.

39 [23]Department of Paediatrics, University of Toronto; Toronto, ON, Canada.

40 [24]Department of Molecular Genetics, University of Toronto; Toronto, ON, Canada.

41 [25]The Centre for Applied Genomics, Hospital for Sick Children; Toronto, ON, Canada.

42 [26]Department of Computer Science, University of Toronto; Toronto, ON, Canada.

43 [*]Corresponding author: Lisa J. Strug, PhD

44 Program in Genetics and Genome Biology

45 SickKids Research Institute

46 Room 12.9705, PGCRL

47 686 Bay Street

48 Toronto, ON

49 M5G 0A4

50 Email: lisa.strug@utoronto.ca

51 Tel: +1416-813-7654 X301762

52

53

54 **Abstract**

55 **Background:** Over 400 variants in the cystic fibrosis (CF) transmembrane conductance

56 regulator (*CFTR*) are CF-causing. CFTR modulators target different variants to improve lung

57 function, but large variability in response exists and current therapies do not address all CF-

58 causing variants highlighting an unmet need. Alternative epithelial ion channels such as

59 SLC26A9 could compensate for CFTR dysfunction, providing a therapeutic target that benefits

60 all individuals with CF.

Method: We investigate the relationship between *SLC26A9* and lung function pre- and post-treatment with CFTR modulators in Canadian and US CF cohorts, in the general population, and in those with chronic obstructive pulmonary disease (COPD).

Results: *SLC26A9* rs7512462 CC genotype is associated with greater lung function in individuals with minimal function variants (for which there are currently no approved therapies; p=0.002); and gating (p=0.03) and p.Phe508del/ p.Phe508del (p=0.009) genotypes upon treatment with CFTR modulators. Analogously, p.Phe508del/p.Phe508del human nasal epithelia with CC after triple combination modulator treatment show greatest CFTR function ($p=8x10^{-4}$). Beyond CF, rs7512462 is associated with lung function in the UK Biobank (meta-$p=2.74x10^{-44}$) and in COPD (min p=0.007).

Conclusion: These findings support SLC26A9 as a therapeutic target to improve lung function for all people with CF and in individuals with other obstructive lung diseases such as COPD.

Key words: SLC26A9, Cystic Fibrosis, CFTR modulator treatment, precision medicine, COPD

Background

Cystic Fibrosis [CF (MIM:219700)] is a common life-limiting autosomal recessive genetic disease caused by pathogenic variants in the cystic fibrosis transmembrane conductance regulator (CFTR; MIM:602421). CF affects multiple organs; morbidity in the pancreas and intestine are seen at birth (1-3), while progressive lung disease is experienced throughout the course of disease and is the major cause of morbidity and mortality. Variability in disease severity across the affected organs is due, in part, to variation in other genes, referred to as

83 modifier genes. Modifier genes impact phenotypic expressivity in the presence of a
84 dysfunctional major causal gene, for example, *CFTR* (1, 4-7).
85
86 CFTR is localized to the cell membrane of epithelial cells and functions as an ion channel. Over
87 400 variants have been reported to cause CF through variable effects on CFTR protein function
88 (8, 9). CF-causing mutations impact apical chloride and bicarbonate transport, altering hydration
89 and pH of airway surface liquid resulting in viscous mucus. Accumulation of this viscous mucus
90 leads to cycles of infection and inflammation, obstructing and damaging the airways, resulting in
91 progressive lung damage and end-stage lung disease (10).
92
93 Developments in CF therapeutics over the past decade have been transformative, altering the
94 management paradigm from treating the downstream consequences of dysfunctional CFTR and
95 delaying the progression of the disease, to treating the basic defect in an individual's CFTR
96 itself: *precision medicine*. New drugs are enhancing CFTR function by targeting the different
97 defects in the protein. For example, individuals with gating mutations have access to a highly
98 effective modulator, ivacaftor (IVA), which is a potentiator that increases the opening probability
99 of CFTR to aid chloride and bicarbonate ion transport in CF epithelia (11). The most common
100 CF causing allele Phe508del (c.1521_1523delCTT; p.Phe508del) (12) displays minimal CFTR
101 at the apical membrane due to processing defects (5, 6) and once at the cell surface the
102 Phe508del protein exhibits reduced opening probability and stability. Pharmaceuticals targeting
103 this defect include a combination therapy of IVA and a CFTR corrector lumacaftor (LUM) or
104 tezacaftor (TEZ) that improves the Phe508del processing to increase cell surface localized
105 protein as well as channel gating. More recently, a triple combination therapy of another

106 corrector, elezacaftor, combined with tezacaftor and ivacaftor (ETI) has been approved in the
107 United States (US) for individuals with at least one Phe508del aged 6 and above or other
108 mutations responsive to Trikafta (13). Elezacaftor stabilizes the protein within the cell
109 membrane, resulting in greater improvements in lung function over LUM/IVA or TEZ/IVA
110 alone (14, 15). ETI is now indicated by the FDA for 90% of individuals with CF (16), and thus
111 the vast majority of individuals with CF have approved pharmaceuticals that address the basic
112 defect of their CF disease.
113
114 Although significant progress has been made in the development of pharmaceuticals for
115 precision medicine in CF, several challenges remain. First, not all individuals with *CFTR*
116 genotypes for which eligible pharmaceuticals are available respond to those treatments. For
117 those who do show a positive improvement in lung function, there is significant variability in the
118 response (14, 15, 17, 18), which could be augmented with additional therapies. Second, there are
119 mutations that do not result in a functional protein that can be addressed using the current
120 paradigm, and therefore an alternative approach to therapy beyond potentiators and correctors of
121 CFTR is necessary.
122
123 Several alternative approaches are being actively pursued, such as gene correction or alternative
124 targets (10). Alternative targets to CFTR can compensate for the abnormal dehydrated airway
125 surface liquid that results from dysfunctional CFTR by modulating other ion channels,
126 transporters and pumps (19-23) and could address the variation in response to existing CFTR
127 modulators while providing a therapeutic option for those individuals with genotypes that do not
128 produce a CFTR protein.

129

130 The clinical success rate of drugs in development is appreciably higher when there is human
131 genetic evidence that supports a drug target (24). Genome-wide modifier gene studies in CF have
132 aimed to identify genetic loci that impact disease severity in the presence of CFTR dysfunction
133 in a hypothesis-free approach. First identified in a genome-wide association study (GWAS) of
134 intestinal obstruction in CF (7), the C allele of rs7512462 in intron 5 of Solute Carrier Family 26
135 member 9 (*SLC26A9)* (chr1:205899595, hg19) has consistently demonstrated a beneficial effect
136 for several CF co-morbidities, including the exocrine (2, 3) and endocrine pancreas (4, 25).
137 These co-morbidities appear to originate from pre-natal dysfunction in the CF pancreas (1, 3,
138 26). According to the Genotype Tissue Expression project (GTEx; (27, 28)), rs7512462 is an
139 expression quantitative locus (eQTL) for *SLC26A9* where the C allele is associated with greater
140 expression in the adult pancreas (p=1.2E-05, effect size=0.27). The *SLC26A9* eQTLs in the
141 region of chr1: 205806897- 206006897(hg19) colocalize with the CF co-morbidity GWAS
142 summary statistics (1), suggesting gene expression variation is responsible for the observed CF
143 GWAS results.

144

145 Given robust expression in the lungs in GTEx (27, 28), and the importance of attenuating
146 progressive lung disease in CF, there was significant interest in investigating the contribution of
147 *SLC26A9* to CF lung function. There was no evidence of association with rs7512462 in the
148 largest CF GWAS of lung function to date (5), comprised predominantly of individuals
149 homozygous for Ph508del (65%). However, restricting to individuals with *CFTR* variants that
150 maintain apical membrane localization, rs7512462 was associated with CF lung function (29).
151 Moreover, upon treatment with the CFTR modulator IVA, the C allele was associated with the

152 greatest lung function improvement (5, 29). These results suggested the possibility that
153 SLC26A9 could improve CFTR function at the apical membrane in airway cells, which was
154 supported by two studies that demonstrated greater CFTR function with the C allele of
155 rs7512462 after correction of CFTR-Phe508del primary human bronchial (29) and nasal
156 epithelia (30).
157
158 SLC26A9 is an anion chloride channel in epithelial cells that contributes to constitutive apical
159 chloride conductance and enhances cAMP-regulated CFTR currents (31-33). *Slc26a9*$^{-/-}$/*Cftr*$^{-/-}$
160 rodent models show poor post weaning survival over *Cftr*$^{-/-}$ (34). That *SLC26A9* was not shown
161 to be a strong modifier of CF lung disease in individuals predominantly homozygous for
162 Phe508del (5, 10, 29) in the absence of treatment with CFTR correctors is consistent with
163 functional studies, where differential SLC26A9 interaction with wild-type CFTR and with
164 Phe508del-CFTR in human bronchial epithelia (HBE) has been reported (29, 31).
165
166 In CF population studies, the *SLC26A9* rs7512462 C allele is associated with greater *SLC26A9*
167 expression in the pancreas, reduced gastrointestinal disease severity and greater lung function
168 response to IVA, highlighting its potential as a target to improve CF outcomes across the
169 affected organs and to improve response to approved CFTR modulators. However, several
170 questions remain including whether population studies support SLC26A9 as providing
171 alternative chloride transport in individuals with genotypes for which there are no approved
172 therapies; whether enhancing SLC26A9 will improve lung function response to CFTR
173 modulators targeting the most common Phe508del variant; which patient-specific cell-based

174 models of the airway are adequate to study SLC26A9; and whether enhancing SLC26A9 could
175 benefit other obstructive lung diseases.
176
177 Through continued recruitment into the Canadian CF Gene Modifier Study (CGMS) and in
178 collaboration with the US PROSPECT observational study (35, 36), we investigate the
179 association between *SLC26A9* rs7512462 and (1) lung function in individuals with different
180 *CFTR* genotypes including variants that result in no CFTR protein product; (2) individual lung
181 function responses to CFTR modulators IVA and LUM/IVA; (3) CFTR function in primary
182 cultured human nasal epithelia (HNE) from individuals with CF in response to approved and
183 experimental CFTR modulators; and (4) other phenotypes in non-CF populations to inform the
184 clinical utility of a SLC26A9 modulator beyond applications in CF.
185

186 **Methods**

187 *Study Design*

188 The Canadian Cystic Fibrosis Gene Modifier Study (CGMS) is a registry-based observational
189 genetics study and the US Part B Cystic Fibrosis Foundation (CFF) Therapeutics PROSPECT
190 study is an observational study of modulator treatment. The current study investigates the
191 relationship between *SLC26A9* and lung function: pre- and post-CFTR modulators in Canadian
192 and US CF cohorts; in the general population; and in those with chronic obstructive pulmonary
193 disease (COPD).
194
195 Clinical data collected as part of the CGMS include forced expiratory volume in 1 second
196 (FEV_1), age, sex, and height which are obtained from the Canadian CF Patient Data Registry

197 (CCFRD), with occasional augmentation by chart review at the participating sites. Genetic data
198 is linked to clinical data through an approved data access agreement with Cystic Fibrosis
199 Canada. Part B of the CFF Therapeutics PROSPECT observational study (35, 36) in the US
200 evaluated the effectiveness of LUM/IVA and collected buffy coat and clinical data from
201 individuals with two copies of the Phe508del variant who were prescribed LUM/IVA. We
202 received the PROSPECT study buffy coat and corresponding clinical data from the US CFF.
203 These and the CGMS samples are genotyped on different Illumina platforms (please
204 Supplementary File for sample genotyping and quality control).
205
206 *Phenotypes, Sampling and Inclusion Criteria*
207 In this study we investigated the effect of *SLC26A9* on pre-treatment lung function in three
208 subgroups from CGMS: individuals with at least one gating mutation, individuals homozygous
209 for Phe508del and those with two MF alleles in trans. Lung function severity in the absence of
210 modulator treatment for participants in CGMS was measured by the Survival adjusted average
211 CF-specific Kulich FEV_1 percentiles that is normalized (Saknorm) (5, 29, 37). For participants
212 whose recruitment date was after 2008, Saknorm is calculated using Canadian CF-specific
213 reference equations from 2008–2014 (38) rather than US reference equations (39). After quality
214 control (QC), 91 participants with at least one gating mutation, 1279 who are homozygous
215 Phe508del and 62 with MF alleles in trans are included in the analysis (Fig. 1).
216
217
218
219

220 **Table 1.** Characteristics of participants included for response to CFTR modulators from the

221 CGMS and US PROSPECT.

		CGMS Participants in CFTR modulator study			US PROSPECT
Participants on CFTR modulator	Participant Characteristics	CGMS IVA (In (29), n=22[#])	CGMS IVA (new, n=23)	CGMS LUM/IVA (n=105)	LUM/IVA (n=91)
rs7512462	TT/TC/CC	11/10/1	12/9/2	41/48/16	30/50/15
age	mean (range)	26.49 (6.1-58.3)	31.3 (14.0-55.3)	25.7 (10.5-55)	20.9 (6-57)
	age <= 12	5	0	2	23
	age (12,20]	4	3	38	27
	age (20,30]	3	11	38	24
	age>30	10	9	27	17
sex	female (%)	16 (72.7%)	13 (56.3%)	57 (54.3%)	52 (57.1%)
FEV_{1pp} baseline	mean (range)	66.6 (30.7-90.4)	58.9 (30.9-89.6)	59.6 (30.8-95.5)	75.7 (31.7-95.8)
Number of post-treatment FEV_{1pp} in 400 days	median (range)	4 (1-6)	4 (1-20)	6 (1-33)	4 (2-4)

222 The lung function measure in the CFTR modulator study is forced expiratory volume in one
223 second percent predicted (FEV_{1pp}).
224 [#] updated data removed 1 sample with FEV_{1pp} baseline outside the inclusion criteria and 1 with
225 no data after lung transplant.
226

227 CGMS participants prescribed IVA or LUM/IVA were also included in the treatment response

228 study, together with the US PROSPECT participants for LUM/IVA (Table 1). Similar to (29), all

229 participants included for modulator lung response analysis had a baseline measurement between

230 30 and 96 FEV_{1pp} measured within 3 months prior to, or on, the treatment initiation date. The

231 CGMS data is obtained from the CCFRD with variable longitudinal entry depending on the
232 resources of the individual clinics across the country while PROSPECT collects regular
233 measurements at 1,3,6 and 12 months after treatment initiation. To account for this difference,
234 the LUM/IVA treatment response is defined as the difference in FEV_{1pp} between the first visit on
235 treatment within 5-7 months and that measured at baseline following convention in the literature
236 (40); for the IVA study, the difference between mean FEV_{1pp} within 15 to 400 days to FEV_{1pp}
237 baseline is used (29). We also investigated an IVA analysis with treatment response defined as
238 the difference from the 1st FEV_{1pp} within 15 to 60 days after the treatment initiation; the
239 conclusion is comparable but because the sample size is reduced to 33, we report the mean
240 treatment response in 15 to 400 days. After phenotype and genotype QC, 91 participants from
241 PROSPECT and 105 CGMS participants were included in the genetic analysis for LUM/IVA
242 and 45 for the IVA study (please see Supplementary Table 1 for sample exclusion).
243
244 We also investigated CFTR function in 46 CF Canada Sick Kids Program in CF Individualized
245 Therapy (CFIT; (41)) participants homozygous for Phe508del whose nasal epithelia were
246 brushed, cultured to passage 2 (P2) and mounted in a circulating Ussing chamber. Thirty-seven
247 individuals who underwent brushing prior to modulator initiation and for whom we had
248 *SLC26A9* genotype are included in the Ussing chamber analysis (please see Supplementary File
249 for cell culture and Ussing chamber studies).
250
251 A subset of the CGMS cohort (n=82) and 6 healthy controls had RNA from their nasal cells
252 sequenced as part of CFIT. In addition, we have sequenced nasal cells for 9 CFIT samples
253 cultured to passage 3 (P3) at two time points (14-16 days and 26-30 days). We also sequenced

254 the RNA from CF individuals who underwent lung transplantation, obtained from paired human
255 bronchial and nasal epithelia; n=17 independent pairs of uncultured primary HNE and HBE and
256 n=16 cultured HNE and HBE pairs (Supplementary Material and Methods for sampling and cell
257 culturing).
258
259 Single cell RNAseq data for lung and pancreas tissue were obtained from the Human Protein
260 Atlas (42, 43) from individuals without CF. Lung tissue samples were obtained from 4 donors
261 aged 58 to 69 years, and pancreas tissue data were from 3 donors between 35 to 51 years of age
262 (Supplementary Material for Single Cell RNA-seq Sample processing).
263
264 Summary association statistics from other GWAS phenotypes were obtained from the GWAS
265 ATLAS (44) (please see Supplementary file for PheWAS data extraction). Association analysis
266 between rs7512462 and lung function in individuals with COPD was carried out using the UK
267 Biobank data under application #40946. The summary statistics for population-level GWAS in
268 the UK Biobank (45) for COPD (physician-diagnosed COPD, UID=22130, case-control study
269 with cases=1,357, controls=90,430) were obtained from Ben Neale's lab (46).
270
271 *Statistical Analysis for association with rs7512462*
272 *i. association with lung function by CFTR genotype group*
273 To assess whether the rs7512462 genotype is associated with lung function in the different *CFTR*
274 genotype groups, we carried out a stratified analysis where the Saknorm association with
275 rs7512462 was assessed separately for the gating, homozygous Phe508del or two MF genotype
276 groups, adjusted by an indicator for which cohort was used to calculate Saknorm. We used

277 linear regression (LR) with a robust variance estimator for those homozygous for Phe508del and
278 for those with two MF variants using the R package rms (47). Previous studies of gating variants
279 reported association of Saknorm with rs7512462 in G551D only, so to mimic those analyses , we
280 implemented LRusing the newly recruited unrelated samples with at least one G551D (n=23) and
281 meta-combined the results with those from (29) and the three cohorts from (48). For this meta-
282 analysis, we used the R function metagen in the package meta (49) using a fixed effect model
283 and the R function forestplot from the package forestplot (50). To assess whether lung function is
284 different between the three *CFTR* genotype groups, we used LR with a robust variance estimator,
285 we also regressed Saknorm on two indicators for the three CFTR genotype categories.
286 Regression was used to determine whether the effect of rs7512462 genotype on lung function
287 (Saknorm) was different depending on the *CFTR* genotype group. This was achieved by
288 including an interaction term in the multiple regression model and the p-value was obtained by a
289 likelihood ratio test comparing nested models. All models include a binary indicator for the
290 cohort used to calculate Saknorm and only two-sided p-values less than 0.05, and with the
291 direction that the C alleles is beneficial, were considered significant.
292
293 *ii. association with response to modulator treatment*
294 We also used multivariable regression with robust variance estimates to assess the modulator
295 FEV_{1pp} response association with rs7512462. For the analysis of rs7512462 association with
296 treatment response to IVA, covariates included age and FEV_{1pp} at baseline. For treatment
297 response to LUM/IVA, besides rs7512462, age and FEV_{1pp} at baseline, principal components
298 (PCs) were also included to adjust for population structure, which were calculated from the
299 PROSPECT or the combined Canadian and PROSPECT studies by the R function PC-AiR in the

300 GENESIS package (51, 52) using the kinship matrix estimated by KING 2.2.4 (53). The
301 significant PCs were selected based on the Tracy-Widom test result using the function twtable in
302 POPGEN of Eigensoft (54) ; we included 7 PCs for the PROSPECT analysis and 4 PCs for the
303 combined CGMS+PROSPECT analysis. Analysis of the Ussing chamber data to assess
304 association of CFTR functional response to CFTR modulators with rs7512462 implementedLR
305 with a robust variance estimator, adjusted for a binary indicator of culture media. The boxplots
306 are implemented using the functions ggplot and geom_boxplot from the package ggplot2 in R
307 (55) and the function geom_jitter from ggplot2 is used to overlay the individual measurements
308 in a stripchart.
309
310 iii. *association analysis with COPD in the UK Biobank*
311 Genetic association analyses for spirometry measures in COPD cases in the UK Biobank were
312 conducted using imputed (v3) phenotypic data obtained from the UK Biobank (45, 56). In-house
313 scripts, R package rbgen and C++ tool bgenix (57), were used to index, subset and perform
314 association analysis using the imputed dosage (56). Genotyping QC of the microarray data,
315 recommended genomic analysis exclusions (UID=22010), inclusion of only Caucasians
316 (UID=22006), and unrelated individuals as identified by KING 2.2.4 (53) resulted in a final
317 n=1,001 COPD cases. Phenotypic analysis at the *SLC26A9* locus included the best measures for
318 peak expiratory flow (PEF), FEV_1, the forced vital capacity (FVC), FEV_{1pp} and FEV_1/FVC ratio,
319 which is the ratio of FEV_1 to FVC of the lungs. The best measure for FEV_1 (UID=20150) and
320 FVC (UID=20151) were defined as the highest measure from the array of values of up to three
321 blows (UID=3063 and 3062, respectively) that were deemed acceptable according to UID=3061
322 as defined in (58), and the best FEV_1/FVC is calculated from the selected best FEV_1 and FVC.

323 For PEF (UID=3064), acceptable measurements were first determined using the European
324 Respiratory Society/American Thoracic Society (ERS/ATS) criteria (UID=20152) and the best
325 measure for PEF was obtained for n=842 samples which is corresponding to the blow with the
326 highest value for the sum of FEV_1 (UID=3063) and FVC (UID=3062), similar to (58). There
327 were n=537 FEV_{1pp} values provided among the COPD cases (UID=20154). The lower sample
328 size for PEF and FEV_{1pp} may be due to more stringent quality assessment using the ERS/ATS
329 criteria. PCs were calculated on the final 1,001 COPD cases or the relevant subsets for each
330 spirometry association analysis using PC-AiR (51, 52) and all association models included two
331 principal components, sex, age, age^2, sex*age and sex*age^2. All spirometry measures were
332 inverse rank normal transformed prior to association analysis using the RNOmni R package's
333 (v0.7.1) rankNorm function (59). Association statistics from each phenotype were then used for
334 colocalization analysis with the CF GWAS summary statistics using LocusFocus (v1.4.8; (60)).
335
336 *RNA Sequencing, Quality Control and analysis*
337 The HNE cells were sequenced on the Illumina HiSeq 2500 platform (Illumina Inc. San Diego,
338 California, USA) and aligned as described in (41) Expression counts were quantified using
339 RNA-SeQC (ver. 2.0.0) and normalized to transcripts per million (TPM) (61) as well as trimmed
340 mean of M values (TMM) measures (62) .
341
342 To compare the expression level for *SLC26A9* across different airway models, we calculated the
343 TPM from naïve HNE for 82 CGMS participants and 6 healthy controls, 16 pairs of cultured
344 HNE and HBE, 17 pairs of naïve HNE and HBE and 9 CFIT cultured HNE samples.
345

346 eQTLs were calculated from 79 CGMS participants for whom both genotype and RNA sequence
347 from naïve HNEs were available. Quality control required TPM ≥ 0.1 and read counts ≥ 6 in
348 greater than 20% of the sample to be analyzed. FastQTL (ver. 2.0) was used to conduct
349 differential gene expression analysis of TMM-normalized read counts on SNP genotypes (63).
350 Covariates included the top 3 genotype principal components, 15 *probabilistic estimation of*
351 *expression residual* (PEER) factors, sample study source, sex, genotyping platform, RNA
352 integrity number (RIN) and PTPRC/CD45 gene expression adjusting for immune cell
353 composition in the samples. The genotype principal components and PEER factors were
354 generated using R packages GENESIS (51, 52) and peer (64), respectively. The average
355 expression level of *SLC26A9* was compared across HBE and HNE in both cultured and naïve
356 paired tissues. Paired t-tests were conducted using the *t.test()* function in R v3.6.1 (65), based on
357 TPM.

358

359 **Results**

360 The CGMS has enrolled 3,257 participants from 9 provinces and 35 clinics across Canada.
361 These individuals have various causal *CFTR* genotypes that are reflective of the Canadian CF
362 population. Here we specifically study the subgroups of the CGMS who are: (1) homozygous
363 for Phe508del; (2) carrying at least one copy of the G551D (c.1652G>A;p.Gly551Asp) variant
364 or another gating variant approved for IVA; and (3) individuals with two minimal function
365 (MF) variants in trans (66) for which there are no CFTR modulators approved (these include
366 nonsense, splicing and small indel variants). A subset of these individuals homozygous for
367 Phe508del or having at least one gating mutation are on a CFTR modulator (Table 1) and we
368 investigate the lung function response to LUM/IVA in this CGMS subset and in an independent

369 sample of 91 participants from PROSPECT, an observational study in US. Participants from the
370 PROSPECT study were younger and healthier than the Canadian cohort (as measured by forced
371 expiratory volume in one second (FEV_1) percent predicted (FEV_{1pp})) at treatment initiation
372 (Table 1).
373
374 *SLC26A9 is associated with CF lung function in the absence of treatment*
375 Using the International CF Gene Modifier consortium lung phenotype, Saknorm ((37);
376 Methods), we compared lung function across the three genotype groups (Fig.1A). Saknorm
377 measures differ significantly between individuals with a gating variant (n=91) and those who are
378 homozygous for Phe508del (n= 1279; effect size=0.21, p=0.025), while individuals with two MF
379 variants (n=62) do not differ in Saknorm from those who are Phe508del homozygous (effect
380 size= -0.09, p=0.35).
381
382 Using a fixed effect meta-analysis we combine our newly recruited unrelated individuals from
383 the CGMS with at least one G551D variant (n=23) with our previously published G551D cohort
384 (29) and a recently published independent G551D study (48). This meta-analysis further
385 supports that better lung function is associated with the C allele (n=365, p=0.035, Fig. S1). In
386 individuals from the CGMS who are homozygous Phe508del, the CC genotype provides some
387 evidence of increased lung function albeit with a smaller effect size (effect size = 0.107,
388 p=0.065, Fig. 1a). Importantly, in individuals with two MF variants, the CC genotype is
389 associated with improved lung function (effect size=0.54, p=0.0019, Fig. 1a). This is consistent
390 with the hypothesis that SLC26A9 may be providing alternative chloride transport properties in
391 individuals with *CFTR* variants for whom no current approved therapies exist.

392 **Table 2.** The association between Saknorm and rs7512462 for additive and recessive coding in
393 combined samples.

Models	Covariate	Effect size	S.E.	t	P-value
Additive	Intercept	0.554	0.069	7.997	2.66E-15
	rs7512462_C	0.072	0.030	2.403	0.016
	CFTR:Gating	0.215	0.092	2.329	0.02
	CFTR:MF/MF	-0.081	0.101	-0.802	0.423
	Saknorm cohort	-0.245	0.05	-4.906	1.04E-06
Recessive	Intercept	0.589	0.065	9.038	<1.0E-15
	rs7512462_CC	0.131	0.056	2.341	0.019
	CFTR:Gating	0.212	0.092	2.312	0.021
	CFTR:MF/MF	-0.083	0.101	-0.830	0.407
	Saknorm cohort	-0.246	0.05	-4.918	9.74E-07

394 The combined samples include individuals with gating mutations (n=91), individuals
395 homozygous for Phe508del (n=1279) or those with two minimal function (MF) mutations
396 (n=62). Saknorm is calculated as in (5, 37) with FEV_1 measurements taken prior to modulator
397 treatment, if applicable. All analyses use the robust covariance matrix estimates by the R
398 package 'rms' (47). In addition to rs7512462, both models include a *CFTR* mutation group
399 indicator, with reference group as homozygous Phe508del and Saknorm cohort. Saknorm cohort
400 is an indicator for the reference equation used to calculate Saknorm (Methods) depending on the
401 year the lung function measurement was taken.
402

403 When combining CGMS participants from the three genotype groups in a joint multivariable
404 regression model to assess the association with rs7512462, the C allele is associated with greater
405 lung function (p=0.016 and 0.019 for additive and recessive models, respectively, Table 2). The
406 C allele does not demonstrate a statistically significant difference in effect size depending on
407 one's *CFTR* genotype (interaction p-values>0.05 for both additive and recessive models).

408 **Table 3.** Association of rs7512462 for CFTR modulator response or change in CFTR function

409 to modulator treatment.

Studies	Covariates	Effect Size	S.E.	P-value
IVA: CGMS	Intercept	10.226	5.121	0.053
	Age at baseline	-0.077	0.095	0.419
	FEV1pp at baseline	-0.025	0.073	0.731
	rs7512462_CC	9.905	4.477	0.033
	Old cohort	5.336	2.606	0.047
LUM/IVA: PROSPECT	Intercept	14.460	5.299	0.008
	Age at baseline	-0.188	0.074	0.014
	FEV1pp at baseline	-0.123	0.053	0.024
	rs7512462_CC	8.518	2.657	0.002
LUM/IVA: PROSPECT+CGMS	Intercept	8.635	3.273	0.009
	Age at baseline	-0.128	0.064	0.047
	FEV1pp at baseline	-0.071	0.037	0.057
	rs7512462_CC	4.858	1.828	0.009
	Prospect Cohort	0.965	1.401	0.492
CFTR-mediated current (VX-770+VX-809)	Intercept	-0.878	0.249	0.001
	rs7512462_CC	-0.512	0.436	0.248
	Standard media	-1.815	0.465	0.0004
CFTR-mediated current (VX-770+VX-809+ amplifier)	Intercept	-2.153	0.397	<0.0001
	rs7512462_CC	-2.064	0.546	0.00076
	Standard media	-2.011	0.69	0.0066

410 Sample sizes in each CFTR modulator treatment study is IVA:CGMS (n=45), LUM/IVA:
411 PROSPECT (n=91), LUM/IVA: PROSPECT+CGMS (n=196). LUM/IVA studies were also
412 adjusted for population structure by PCs (n=7 PCs for PROSPECT and n=5 for
413 PROSPECT+CGMS). Functional study includes HNEs from individuals homozygous for
414 Phe508del with VX-770+VX-809 (n=37) and VX-770+VX-809+ amplifier (n=31; a subset of
415 the 37 samples tested with 770+VX-809) measured as difference in CFTR-mediated current after
416 applying forskolin regressed on rs7512462 with a recessive coding. All results use robust
417 variance estimates.
418
419 *Rs7512462 C allele is associated with improved response to CFTR modulators*

420 Newly recruited participants into the CGMS on IVA (n=23) were, on average, older and had

421 worse baseline lung function than those included in (29) (Table 1). Combining the two samples

422 (n=45), despite the difference in baseline characteristics and disease severity, we do observe
423 additional supportive evidence for improved lung function response (difference in FEV_{1pp} pre
424 and post treatment initiation) in individuals with the CC genotype upon treatment with IVA
425 (effect size=9.9, p=0.03, Table 3; Fig. 1b).
426 Through ongoing recruitment in the CGMS there were 105 participants prescribed LUM/IVA
427 clinically. In collaboration with the US PROSPECT observational study, we investigated lung
428 function response to LUM/IVA alone (n=91) and in a combined sample of n=196 individuals
429 homozygous for Phe508del stratified by rs7512462. Despite minimal clinical response to
430 LUM/IVA reported on average (17), we do observe a significant improvement in lung function
431 response in those with the CC genotype (p=0.002 in PROSPECT and p= 0.009 for combined,
432 respectively; Table 3), akin to observations for IVA (Fig. 1b and (29)), and in studies of
433 improved CFTR function with the rs7512462 C allele in HNE (30) and HBE (21, 29) cells
434 obtained from individuals homozygous for Phe508del. Thus, the *SLC26A9* rs7512462 genotype
435 shows improved response to the LUM/IVA combination therapy in cohorts of CF patients who
436 were monitored observationally during their real-world experience with the approved
437 modulators.
438
439 We next investigated CFTR-mediated current in 37 cultured HNE from CGMS participants
440 homozygous for Phe508del upon treatment with VX-770+VX-809 (as in (30), corresponding to
441 the IVA/LUM combination; VX-770 applied acutely) and upon treatment with VX-770+VX-809
442 + an amplifier under experimental investigation ((67); Fig. 2). We used the HNE from the
443 earliest passage available to us (P2) to align with (30). Defining the treatment response as the
444 difference in ΔIeq -forskolin from DMSO to VX-770+ VX809 (n=37) or VX-770+ VX-809 +

445 amplifier (n=31), we see a significant improvement in CFTR function in the CC group (effect
446 size=-2.064, p=0.00076, Table 3) in HNE with VX-770+VX-809+ amplifier. The increase in
447 CFTR function in the CC group in the cultured HNE with VX-809 does not reach statistical
448 significance (effect size=-0.512, p=0.248, Table 3) unlike previous reports in HNE and in HBE
449 models, although the direction of effect is consistent. This difference in effect size made us
450 question the *SLC26A9* gene expression profile across different lung tissue models, especially
451 given the low passage of the primary HNE cultured cells studied by (30).
452
453 *SLC26A9 expression differs across airway models*
454 While investigating eQTLs in various airway tissue models, we observed that there is no
455 evidence supporting rs7512462 as an eQTL in the lungs from GTEx v8 (p=0.71) or from RNA
456 obtained from naïve HNE of individuals with CF (p=0.64, n=79). *SLC26A9* expression appears
457 generally low across several different lung model systems we investigated, with average
458 transcripts per million (TPM)= 1.34 (Fig. 3a). We generated a resource that contains the
459 transcriptomes from paired cultured and fresh naïve HNE and HBE tissue on the same
460 individuals (Methods; GEO ID: GSE172232)). Of the primary lung cell models we investigated,
461 the greatest expression is in the naïve HBE cells (TPM=1.94; Fig. 3a), and this expression level
462 is 2.1-fold greater than in the naïve HNE (p=0.04, paired analysis in n=17 HNE-HBE naive
463 pairs). Unfortunately, naïve HBE cell models are not generally accessible for personalized
464 medicine approaches (41), and cultured models are the norm. Here cultured HBE show mean
465 expression with TPM=1.71 that is 2.5-fold greater than the CF cultured HNE (p=0.02, paired
466 analysis in n=16 HNE-HBE cultured pairs), although there is some indication that a reduction in
467 culturing time results in greater *SLC26A9* expression in the HNE (Fig. S2).

468

469 Using the single cell RNA-seq data catalogued in the Human Protein Atlas (42, 43), we
470 investigated the expression of *SLC26A9* and *CFTR* across cell types in the lung and pancreas.
471 *SLC26A9* is predominantly expressed in type 2 alveolar cells. Interestingly, although in general
472 *SLC26A9* is lowly expressed compared to *CFTR* across the cell types of the adult lung and
473 pancreas, when it is expressed, it is accompanied by *CFTR* (Fig. 3b).

474

475 *Phenome-wide Association Study (PheWAS) of rs7512462 and colocalization analysis*

476 We used the GWAS ATLAS database (44) that includes 4,756 GWASs from 473 unique studies
477 with 3,302 unique traits and the UK Biobank resource to investigate other non-CF traits
478 associated with rs7512462. The 10 traits with the smallest reported p-values are listed in Table
479 4, four of which are respiratory phenotypes from the UK Biobank and Spirometa consortium
480 (58): PEF and FEV_1/FVC ratio. Saknorm, the lung function measurement used in the CF
481 GWAS, is also calculated from FEV_1 (5, 37). The list of significant phenotypes also includes
482 our own CF modifier gene study where we first identified *rs7512462* as a modifier of early
483 pancreatic disease (Table 4). Interestingly, an earlier age at menarche (which is associated with
484 weight) and a higher male waist circumference and waist-hip ratio are also associated with the
485 rs7512462 C allele. The association with type 1 diabetes and the weight-related phenotypes
486 suggest that the role of SLC26A9 in these reproductive and metabolic phenotypes may likewise
487 trace back to a pancreatic origin. Significant colocalization analysis (column 'Colocalization P-
488 value', Table 4) calculated using the Simple Sum (1) implemented in LocusFocus (60) between
489 the meconium ileus CF GWAS statistics (1) and the summary statistics from the PheWAS
490 associated traits supports that the same genetic variation contributes to the traits.

Table 4. The 10 phenotypes with smallest rs7512462 association p-value obtained from the GWAS ATLAS (44).

PMID	Year	Domain	Trait	Effect Size	P-value at rs7512462	N	Colocalization P-value	Effect allele	Other allele
30804560	2019	Respiratory	PEF (UKBB)	0.038	4.33E-46	321047	3.98E-08	C	T
30804560	2019	Respiratory	PEF (Meta of UKBB and Spirometa)	0.037	2.74E-44	345265	2.69E-08	C	T
30807572	2019	Gastrointestinal	Meconium ileus in CF	-0.289	1.86E-09	6770	NA	C	T
28436984	2017	Reproduction	Age at menarche	-0.021	3.93E-06	252514	3.02E-08	C	T
30804560	2019	Respiratory	FEV_1/FVC ratio (Meta of UKBB and Spirometa)	0.01	4.11E-05	400102	3.98E-09	C	T
30804560	2019	Respiratory	FEV_1/FVC ratio (UKBB)	0.01	1.44E-04	321047	1.48E-08	C	T
23754948	2013	Metabolic	Waist circumference (male)	0.03	2.2E-04	362311	1.38E-08	C	T
25231870	2014	Reproduction	Age at menarche	-0.023	3.2E-04	132989	4.14E-09	C	T
19430480	2009	Endocrine	Type 1 Diabetes	NA	7.47E-04	7982	NA	NA	NA
23754948	2013	Metabolic	Waist-hip ratio (male)	0.028	7.7E-04	34629	1.12E-08	C	T

PheWAS significance level is $p<1.05x10^{-5}$ after adjusting for 4,756 GWASs in the database (alpha 0.05). NA in the columns effect size, effect allele and other allele denote that the effect size is not reported for the SNP in the original summary statistics from the listed publication (PMID). The column 'Colocalization P-value' represents the p-values from the colocalization analysis calculated here with the CF GWAS summary statistics in (1) for the corresponding phenotype, using LousFocus (60) calculated on the chr1: 205,895,000-205,921,000 region in hg19. NA in this column reflects a lack of information available to carry out the colocalization analysis.

502 Spirometry is a key diagnostic feature of COPD (68). Association analysis of rs7512462 with
503 lung function (PEF, FEV_1, FVC, FEV_1pp and FEV_1/FVC) in the subgroup of the UK Biobank
504 with a physician diagnosis of COPD (n=1,357 of 91,787 with a response of Yes or No)
505 demonstrates that rs7512462 is also a modifier of some lung function measures in COPD, but not
506 of COPD susceptibility itself (Supplementary Table 2). Colocalization analysis also supports a
507 common underlying mechanism between the CF and COPD lung function phenotypes
508 (Supplementary Table 2, Fig. 4).
509

510 **Discussion**

511 The availability of CFTR-modulators is altering care for many individuals with CF, although
512 variation in responses is apparent. The variability in response is due to a range of factors
513 including genes that define the genetic background of an individual. One such gene is *SLC26A9,*
514 which contributes to CF comorbidities in the organs of the digestive tract regardless of CFTR
515 genotype (1, 4). The relationship with CFTR is complex however, as an association between
516 lung function and *SLC26A9* was not immediately apparent. This could at least be partially
517 reconciled upon realization that the *SLC26A9* locus does influence lung disease in some CF
518 individuals, specifically those that carry mutations that lead to gating deficiencies in CFTR.
519 Further, experimental evidence for a relationship between *SLC26A9* genotype and CFTR
520 function in airway cell models was then observed by rescuing the traffick-defective Phe508del
521 via 'corrector' modulators. However, to assuage concerns by others (48) that the evidence for
522 *SLC26A9* remains uncertain, we now extend several of these studies to examine the complex
523 relationship further. We investigate the role of the *SLC26A9* genotype in CF patients with
524 different types of CFTR mutations and in treatment response to CFTR modulators both in patient

525 populations and in airway models. We align the findings and previously published work with
526 gene expression patterns, and now also consider the role of *SLC26A9* in the phenome and, in
527 particular, in lung function measurements in non-CF populations. Significant human genetic
528 evidence that supports a role for *SLC26A9* in CF disease severity and CFTR modulator response
529 is accumulating. Here we provide the most comprehensive investigation of *SLC26A9* in CF and
530 non-CF populations.
531
532 The relationship between the *SLC26A9* rs7512462 marker and lung function in three types of
533 *CFTR* genotypes was revealing. Individuals with CF-causing gating variants including G551D
534 do show association with rs7512462. Gating-deficient CFTR protein exhibits apical cell
535 membrane localization with reduced opening probability, resulting in reduced epithelial
536 chloride and bicarbonate secretion (69). In contrast, individuals homozygous for Phe508del,
537 comprising the majority of CF cases, show very modest evidence of association with rs7512462.
538 Phe508del-CFTR is rapidly degraded intracellularly with minimal surface membrane
539 localization. Interestingly, for the first time, we demonstrated that individuals with MF variants
540 do show rs7512462 association but would not produce protein and therefore cannot respond to
541 CFTR modulator therapy.
542
543 The studies of MF alleles argue that at least some aspect of lung function contribution by
544 SLC26A9 can be independent of CFTR. These investigations were supported by several previous
545 studies highlighting the potential for SLC26A9 to provide alternative chloride transport in the
546 absence of CFTR. Bertrand *et al*. (31, 32) demonstrated in airway cell models that SLC26A9 is
547 a constitutively active chloride channel that functions independently of CFTR prior to CFTR

548 activation, however interactions due to the poor trafficking of Phe508del-CFTR can have
549 negative consequences.
550
551 CF GWAS of early-onset pancreatic and intestinal phenotypes were CFTR genotype *independent*
552 *(1-3, 7);* that is, those homozygous for Phe508del still demonstrated benefit from the C allele of
553 rs7512462, a marker of *SLC26A9* gene expression in CF phenotypes that show pre-natal onset
554 (for example, meconium ileus). This indicates that SLC26A9 and CFTR may act independently
555 at this stage of development consistent with the observation that *Slc26a9* mRNA was high in the
556 murine pancreas in the embryonic stages of development when *Cftr* was low (29). Given the
557 small sample size, the rs7512462 association with lung function in individuals with two MF
558 alleles requires further investigation and independent replication. However, a consistent
559 beneficial effect of the CC genotype across several of our independent studies and disparate
560 outcomes provides compelling support that modulation of SLC26A9 can provide alternative
561 chloride transport and could be a therapeutic target to improve lung function in individuals with
562 any *CFTR* genotype. The modulation of alternative channels, transporters and pumps to
563 compensate for dysfunctional CFTR (22, 70), and in particular SLC26A9 (10, 23, 29), would
564 provide a mutation-agnostic approach and address the current unmet need of the CF individuals
565 with MF alleles.
566
567 Given other published reports of the rs7512462 relationship with lung function in individuals
568 with G551D variants (29, 48), we used meta-analysis to summarize the current state of evidence.
569 Together, the weight of evidence supports a relationship between the *SLC26A9* marker,
570 rs7512462, and lung function in individuals with a G551D variant where there is cell-surface

571 localized CFTR protein, but reduced activity. In an expanded CGMS cohort, we also replicated
572 the enhancing effect of the rs7512462 C allele upon rescue of gating variants with IVA,
573 suggesting that the pre-treatment effect may be attributable to the small quantity of functional
574 CFTR that remains. Consistent with this hypothesis is the suggestive association we observed
575 between rs7512462 and lung function prior to treatment with LUM/IVA in individuals from the
576 CGMS homozygous for Phe508del, with a reduced effect size possibly reflective of reduced cell
577 membrane localization.
578
579 LUM/IVA was the first approved modulator for those homozygous for Phe508del on the basis of
580 reduced pulmonary exacerbations and lung function response. However, the average
581 improvement in lung function in clinical trials (17) and observational studies (35, 36) has been
582 modest. To investigate whether the improvements in lung function from LUM/IVA are modified
583 by the *SLC26A9* genotype, we also genotyped the DNA from participants of the US
584 observational LUM/IVA study, PROSPECT (35, 36), and demonstrated that although modest in
585 its average improvement in lung function, those with the CC rs7512462 genotype exhibited the
586 greatest benefit. Although this is the first report of *SLC26A9* impact on clinical response, and in
587 a real-world setting of treated patients, this finding is consistent with published studies of CFTR
588 function in primary HBE and HNE cells from individuals homozygous for Phe508del by us (29)
589 and others (30), respectively. Together these findings across the two *CFTR* genotype groups
590 replicate and expand previous reports (29-32) and suggest that the *SLC26A9* rs7512462 eQTL
591 correlates with CFTR residual or corrected function.
592

593 We did evaluate CFTR function in nasal brushes from 37 individuals homozygous for Phe508del
594 in Ussing chamber studies following 24h exposure to vehicle (DMSO), VX-770 with corrector
595 VX-809 (corresponding to IVA/LUM) and a combination of VX-770, VX-809 and an
596 experimental amplifier (67). We observed a similar trend to that reported in (30) with VX-809,
597 although it did not reach statistical significance. When CFTR function was, however, augmented
598 with the amplifier, the difference in CFTR function pre- and post-treatment was greatest in the
599 cells with the rs7512462 *SLC26A9* CC genotype. These functional studies further support the
600 hypothesis that SLC26A9 will likely benefit any therapeutic situation with increased apical
601 surface localized CFTR protein, such as for the latest highly effective modulator treatment, ETI.
602
603 Although the population studies provide important insights into the relationship between CFTR
604 and SLC26A9 *in vivo*, functional studies in cellular and animal models will be necessary to
605 understand the functional relationship between the two channels/transporters (71) and how they
606 may be working together. The expression studies presented here provide guidance on the use of
607 cellular models for SLC26A9 and further highlight potential limitations of cultured HNE for the
608 unique considerations of studying SLC26A9 and CFTR that are distinct from the ones for studies
609 of CFTR alone. Compared to naïve bronchial cells, we observed greater variation and lower
610 expression in naïve nasal cells. Furthermore, the culturing process of either HNE or HBE cells
611 resulted in reduced *SLC26A9* expression. Presently, the cultured HBE model appears superior to
612 HNEs for investigations of SLC26A9, although further investigation of culturing conditions
613 needs to be considered. Nevertheless, design and interpretation of functional assessment of
614 SLC26A9 in HNEs must consider this variable expression and corresponding limitations.
615

616 Interestingly, the most significant *SLC26A9* SNPs from the CF GWASs (1, 7) also associate with
617 lung function measurements in several large international studies: PEFand FEV_1/FVC ratio in
618 participants from the UK Biobank aged 40-69 (10, 58); PEF and FEV_1/FVC ratio in the
619 Spirometa consortium (58); and for the FEV_1/FVC ratio in 8-year-olds from the UK10K
620 consortium (72). These results align with our findings that, after correction of CF-causal CFTR
621 variants with modulators, *SLC26A9* SNPs are associated with improved lung function and CFTR
622 function.
623
624 It is noteworthy that decreased spirometry is diagnostic for other obstructive lung diseases such
625 as COPD (73). Several studies have reported similar pathobiology cascades between CF and
626 COPD due to dysfunctional CFTR and environmental risk factors for COPD (74-78) . For
627 example, CFTR chloride channels show reduced capacity as a result of tobacco smoke and may
628 result in the mucus obstruction characteristic of COPD (77, 79) that is akin to that seen in CF. If
629 SLC26A9 augments chloride transport, SLC26A9 agonists could also be an effective therapeutic
630 strategy for COPD. In support of this premise is the association evidence provided here
631 demonstrating rs7512462 is a modifier of lung function in individuals with physician-diagnosed
632 COPD. Moreover, the significant colocalization evidence between the CF and UK Biobank
633 summary statistics at the *SLC26A9* locus are reflective of a common underlying genetic
634 contribution.
635
636 **Conclusions**
637 A role for SLC26A9 in the CF precision medicine landscape is an exciting prospect. SLC26A9
638 shows desirable characteristics (10) as an alternative therapeutic target for CF, including the

639 urgent need for options for CF individuals with MF alleles. The association between *SLC26A9*
640 genotype and response to existing pharmaceuticals indicate the potential to refine personalized
641 combinations of modulators, and there is also support that SLC26A9 agonists may yield benefit
642 to any existing pharmacological or gene correction paradigm in CF, independent of *CFTR*
643 genotype. The observation that SLC26A9 can improve measures of lung function in non-CF
644 populations was a serendipitous finding, identified through genetic modifier studies of a
645 'monogenic' disease, and may have far-reaching impact beyond CF.

646

647

648 **List of abbreviations**

649 CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator; IVA: Ivacaftor;
650 LUM: lumacaftor; TZE: tezacaftor; ETI: elezacaftor and tezacaftor combined with ivacaftor; US:
651 United States; GWAS: genome-wide association study; SLC26A9: Solute Carrier Family 26
652 member 9; GTEx: Genotype Tissue Expression project; eQTL: expression quantitative locus;
653 HBE: human bronchial epithelia; CGMS: Canadian CF Gene Modifier Study; HNE: human nasal
654 epithelia; MF: minimal function; FEV_1: forced expiratory volume in one second; FEV_{1pp}:forced
655 expiratory volume in one second percent predicted; TPM: transcripts per million; PheWAS:
656 Phenome-wide Association Study; PEF: peak expiratory flow; FVC: the forced vital capacity;
657 FEV_1/FVC ratio: the ratio of forced expiratory volume in one second to the forced vital capacity;
658 COPD: chronic obstructive pulmonary disease; CFF: Cystic Fibrosis Foundation ; CCFRD:
659 Canadian CF Patient Data Registry; Saknorm: the Survival adjusted average CF-specific Kulich
660 FEV_1 percentiles that is normalized; QC: quality control; CFIT: CF Canada Sick Kids Program
661 in CF Individualized Therapy; P2: culture to Passage 2; P3: culture to Passage 3; LR: linear

662 regression; PCs: principal components; TMM: trimmed mean of M values; PEER: *probabilistic*
663 *estimation of expression residual* ; RIN: RNA integrity number; ERS/ATS: European
664 Respiratory Society/American Thoracic Society.

665

666

667 **References**

668 1. Gong J, Wang F, Xiao B, Panjwani N, Lin F, Keenan K, et al. Genetic association and
669 transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci.
670 PLoS Genet. 2019;15(2):e1008007.
671 2. Miller MR, Soave D, Li W, Gong J, Pace RG, Boëlle PY, et al. Variants in Solute Carrier
672 SLC26A9 Modify Prenatal Exocrine Pancreatic Damage in Cystic Fibrosis. J Pediatr.
673 2015;166(5):1152-7 e6.
674 3. Soave D, Miller MR, Keenan K, Li W, Gong J, Ip W, et al. Evidence for a causal
675 relationship between early exocrine pancreatic disease and cystic fibrosis-related diabetes: a
676 Mendelian randomization study. Diabetes. 2014;63(6):2114-9.
677 4. Blackman S, Commander C, Watson C, Arcara K, Strug L, Stonebraker J, et al. Genetic
678 modifiers of cystic fibrosis-related diabetes. Diabetes. 2013;62:3627-35.
679 5. Corvol H, Blackman SM, Boëlle PY, Gallins PJ, Pace RG, Stonebraker JR, et al.
680 Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in
681 cystic fibrosis. Nat Commun. 2015;6:8382.
682 6. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical
683 application. Nat Rev Genet. 2015;16(1):45-56.
684 7. Sun L, Rommens JM, Corvol H, Li W, Li X, Chiang TA, et al. Multiple apical plasma
685 membrane constituents are associated with susceptibility to meconium ileus in individuals with
686 cystic fibrosis. Nat Genet. 2012;44(5):562-9.
687 8. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, et al. Defining the
688 disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat
689 Genet. 2013;45(10):1160-7.
690 9. The Clinical and Functional TRanslation of CFTR (CFTR2) [Internet]. Available from:
691 http://cftr2.org.
692 10. Strug LJ, Stephenson AL, Panjwani N, Harris A. Recent advances in developing
693 therapeutics for cystic fibrosis. Hum Mol Genet. 2018;27(R2):R173-R86.
694 11. Igreja S, Clarke LA, Botelho HM, Marques L, Amaral MD. Correction of a Cystic
695 Fibrosis Splicing Mutation by Antisense Oligonucleotides. Hum Mutat. 2016;37(2):209-15.
696 12. Howard M, Frizzell RA, Bedwell DM. Aminoglycoside antibiotics restore CFTR
697 function by overcoming premature stop mutations. Nat Med. 1996;2(4):467-9.
698 13. Cystic Fibrosis Foundation. Elexacaftr-tezacaftor-ivacaftor-Tikafta 2021 [Available
699 from: https://www.cff.org/Trials/Pipeline/details/10150/Elexacaftor-tezacaftor-ivacaftor-
700 Trikafta.

701 14. Middleton PG, Mall MA, Drevinek P, Lands LC, McKone EF, Polineni D, et al.
702 Elexacaftor-Tezacaftor-Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. N Engl J
703 Med. 2019;381(19):1809-19.
704 15. Heijerman HGM, McKone EF, Downey DG, Van Braeckel E, Rowe SM, Tullis E, et al.
705 Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in
706 people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised,
707 phase 3 trial. Lancet. 2019;394(10212):1940-8.
708 16. FDA. FDA approves new breakthrough therapy for cystic fibrosis 2019 [cited 2021 Feb
709 12]. Available from: https://www.fda.gov/news-events/press-announcements/fda-approves-new-
710 breakthrough-therapy-cystic-fibrosis.
711 17. Wainwright CE, Elborn JS, Ramsey BW, Marigowda G, Huang X, Cipolli M, et al.
712 Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. N Engl
713 J Med. 2015;373(3):220-31.
714 18. Yu H, Burton B, Huang CJ, Worley J, Cao D, Johnson JP, Jr., et al. Ivacaftor potentiation
715 of multiple CFTR channels with gating mutations. J Cyst Fibros. 2012;11(3):237-45.
716 19. Haq IJ, Gray MA, Garnett JP, Ward C, Brodlie M. Airway surface liquid homeostasis in
717 cystic fibrosis: pathophysiology and therapeutic targets. Thorax. 2016;71(3):284-7.
718 20. Martin SL, Saint-Criq V, Hwang TC, Csanady L. Ion channels as targets to treat cystic
719 fibrosis lung disease. J Cyst Fibros. 2018;17(2S):S22-S7.
720 21. Balazs A, Mall MA. Role of the SLC26A9 Chloride Channel as Disease Modifier and
721 Potential Therapeutic Target in Cystic Fibrosis. Front Pharmacol. 2018;9:1112.
722 22. Enterprise Therapeutics, editor Pharmacological characterisation of TMEM16A
723 regulators. The 13th European Cystic Fibrosis Basic Science Conference; 2016; Pisa, Italy.
724 23. Mall MA, Galietta LJ. Targeting ion channels in cystic fibrosis. J Cyst Fibros.
725 2015;14(5):561-70.
726 24. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of
727 human genetic evidence for approved drug indications. Nature genetics. 2015;47(8):856-60.
728 25. Aksit MA, Pace RG, Vecchio-Pagan B, Ling H, Rommens JM, Boelle PY, et al. Genetic
729 Modifiers of Cystic Fibrosis-Related Diabetes Have Extensive Overlap With Type 2 Diabetes
730 and Related Traits. J Clin Endocrinol Metab. 2020;105(5).
731 26. Lin YC, Keenan K, Gong J, Panjwani N, Avolio J, Lin F, et al. Cystic fibrosis-related
732 diabetes onset can be predicted using biomarkers measured at birth. Genet Med. 2021;23(5):927-
733 33.
734 27. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across
735 human tissues. Science. 2020;369(6509):1318-30.
736 28. GTEx Portal [Internet]. [cited 2021]. Available from: http://www.gtexportal.org/home/.
737 29. Strug LJ, Gonska T, He G, Keenan K, Ip W, Boelle PY, et al. Cystic fibrosis gene
738 modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. Hum Mol Genet.
739 2016;25(20):4590-600.
740 30. Kmit A, Marson FAL, Pereira SV, Vinagre AM, Leite GS, Servidoni MF, et al. Extent of
741 rescue of F508del-CFTR function by VX-809 and VX-770 in human nasal epithelial cells
742 correlates with SNP rs7512462 in SLC26A9 gene in F508del/F508del Cystic Fibrosis patients.
743 Biochim Biophys Acta Mol Basis Dis. 2019;1865(6):1323-31.
744 31. Bertrand CA, Mitra S, Mishra SK, Wang X, Zhao Y, Pilewski JM, et al. The CFTR
745 trafficking mutation F508del inhibits the constitutive activity of SLC26A9. Am J Physiol Lung
746 Cell Mol Physiol. 2017;312(6):L912-L25.

747 32. Bertrand CA, Zhang R, Pilewski JM, Frizzell RA. SLC26A9 is a constitutively active,
748 CFTR-regulated anion conductance in human bronchial epithelia. J Gen Physiol.
749 2009;133(4):421-38.
750 33. Salomon JJ, Spahn S, Wang X, Fullekrug J, Bertrand CA, Mall MA. Generation and
751 functional characterization of epithelial cells with stable expression of SLC26A9 Cl- channels.
752 Am J Physiol Lung Cell Mol Physiol. 2016:ajplung 00321 2015.
753 34. Liu X, Li T, Riederer B, Lenzen H, Ludolph L, Yeruva S, et al. Loss of Slc26a9 anion
754 transporter alters intestinal electrolyte and HCO3(-) transport and reduces survival in CFTR-
755 deficient mice. Pflugers Arch. 2015;467(6):1261-75.
756 35. Rowe SM. A Two-Part Multicenter Prospective Longitudinal Study of CFTR-dependent
757 Disease Profiling in Cystic Fibrosis (PROSPECT) United States2015 [updated May 20,2020;
758 cited 2021 Jan 11]. Available from: https://clinicaltrials.gov/ct2/show/NCT02477319.
759 36. Sagel SD, Khan U, Heltshe SL, Clancy JP, Borowitz D, Gelfond D, et al. Clinical
760 Effectiveness of Lumacaftor/Ivacaftor in Patients with Cystic Fibrosis Homozygous for F508del-
761 CFTR. A Clinical Trial. Ann Am Thorac Soc. 2021;18(1):75-83.
762 37. Taylor C, Commander CW, Collaco JM, Strug LJ, Li W, Wright FA, et al. A novel lung
763 disease phenotype adjusted for mortality attrition for cystic fibrosis Genetic modifier studies.
764 Pediatr Pulmonol. 2011;46(9):857-69.
765 38. Kim SO, Corey M, Stephenson AL, Strug LJ. Reference percentiles of FEV1 for the
766 Canadian cystic fibrosis population: comparisons across time and countries. Thorax.
767 2018;73(5):446-50.
768 39. Kulich M, Rosenfeld M, Campbell J, Kronmal R, Gibson RL, Goss CH, et al. Disease-
769 specific reference equations for lung function in patients with cystic fibrosis. Am J Respir Crit
770 Care Med. 2005;172(7):885-91.
771 40. Bardin E, Pastor A, Semeraro M, Golec A, Hayes K, Chevalier B, et al. Modulators of
772 CFTR. Updates on clinical development and future directions. Eur J Med Chem.
773 2021;213:113195.
774 41. Eckford PDW, McCormack J, Munsie L, He G, Stanojevic S, Pereira SL, et al. The CF
775 Canada-Sick Kids Program in individual CF therapy: A resource for the advancement of
776 personalized medicine in CF. J Cyst Fibros. 2019;18(1):35-43.
777 42. The Human Protein Atlas Program. [Internet]. [cited 2020, Nov 19]. Available from:
778 http://www.proteinatlas.org/.
779 43. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al.
780 Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
781 44. Watanabe K, Stringer S, Frei O, Umicevic Mirkov M, de Leeuw C, Polderman TJC, et al.
782 A global overview of pleiotropy and genetic architecture in complex traits. Nature genetics.
783 2019;51(9):1339-48.
784 45. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open
785 access resource for identifying the causes of a wide range of complex diseases of middle and old
786 age. PLoS Med. 2015;12(3):e1001779.
787 46. Neale B. We're thrilled to announce an update GWAS analysis of the UK BioBank. 2021
788 [cited 2021 Mar 21]. Available from: http://www.nealelab.is/uk-biobank/.
789 47. J. Stare, H. Heinzl, Harrell F. rms: Regression Modeling Strategies 2015 [cited 2021 Mar
790 20]. Available from: https://cran.r-project.org/web/packages/rms/index.html.

48. Eastman AC, Pace RG, Dang H, Aksit MA, Vecchio-Pagan B, Lam AN, et al. SLC26A9 SNP rs7512462 is not associated with lung disease severity or lung function response to ivacaftor in cystic fibrosis patients with G551D-CFTR. J Cyst Fibros. 2021.
49. Schwarzer G. metagen: Generic inverse variance meta-analysis 2021 [cited 2021 Mar. 20]. Available from: https://rdrr.io/cran/meta/man/metagen.html.
50. Max Gordon, Lumley T. forestplot: Advanced Forest Plot Using 'grid' Graphics 2020 [cited 2021 Mar. 20]. Available from: https://rdrr.io/cran/forestplot/.
51. Conomos MP. Population Structure and Relatedness Inference using the GENESIS Package 2021 [cited 2021 Feb-06].
52. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015;39(4):276-93.
53. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-73.
54. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.
55. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
56. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
57. Band G, Marchini J. BGEN: a binary file format for imputed genotype and haplotype data. bioRxiv2018.
58. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nature genetics. 2019;51(3):481-93.
59. McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. Biometrics. 2020;76(4):1262-72.
60. Panjwani N, Wang F, Mastromatteo S, Bao A, Wang C, He G, et al. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. PLoS Comput Biol. 2020;16(10):e1008336.
61. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012;28(11):1530-2.
62. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
63. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016;32(10):1479-85.
64. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc. 2012;7(3):500-7.
65. R Core Team. R: A language and environment for statistical computing. 2015.
66. CF Foundation. Types of CFTR Mutations [cited 2021 Feb 12]. Available from: https://www.cff.org/What-is-CF/Genetics/Types-of-CFTR-Mutations/.
67. Giuliano KA, Wachi S, Drew L, Dukovski D, Green O, Bastos C, et al. Use of a High-Throughput Phenotypic Screening Strategy to Identify Amplifiers, a Novel Pharmacological

837 Class of Small Molecules That Exhibit Functional Synergy with Potentiators and Correctors.
838 SLAS Discov. 2018;23(2):111-21.
839 68. Stephens MB, Yew KS. Diagnosis of chronic obstructive pulmonary disease. Am Fam
840 Physician. 2008;78(1):87-92.
841 69. Fischer H. The G551D CFTR chloride channel spurs the development of personalized
842 medicine. J Physiol. 2014;592(9):1907-8.
843 70. Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A. Cystic fibrosis. Nature
844 Reviews Disease Primers. 2015;1:1-19.
845 71. Walter JD, Sawicka M, Dutzler R. Cryo-EM structures and functional characterization of
846 murine Slc26a9 reveal mechanism of uncoupled chloride transport. Elife. 2019;8.
847 72. Baskurt Z, Mastromatteo S, Gong J, Wintle RF, Scherer SW, Strug LJ. VikNGS: a C++
848 variant integration kit for next generation sequencing association analysis. Bioinformatics.
849 2020;36(4):1283-5.
850 73. Johns DP, Walters JA, Walters EH. Diagnosis and early detection of COPD using
851 spirometry. J Thorac Dis. 2014;6(11):1557-69.
852 74. Cantin AM, Hanrahan JW, Bilodeau G, Ellis L, Dupuis A, Liao J, et al. Cystic fibrosis
853 transmembrane conductance regulator function is suppressed in cigarette smokers. Am J Respir
854 Crit Care Med. 2006;173(10):1139-44.
855 75. Clunes LA, Davies CM, Coakley RD, Aleksandrov AA, Henderson AG, Zeman KL, et
856 al. Cigarette smoke exposure induces CFTR internalization and insolubility, leading to airway
857 surface liquid dehydration. FASEB J. 2012;26(2):533-45.
858 76. Raju SV, Jackson PL, Courville CA, McNicholas CM, Sloane PA, Sabbatini G, et al.
859 Cigarette smoke induces systemic defects in cystic fibrosis transmembrane conductance
860 regulator function. Am J Respir Crit Care Med. 2013;188(11):1321-30.
861 77. Dransfield MT, Wilhelm AM, Flanagan B, Courville C, Tidwell SL, Raju SV, et al.
862 Acquired cystic fibrosis transmembrane conductance regulator dysfunction in the lower airways
863 in COPD. Chest. 2013;144(2):498-506.
864 78. Sloane PA, Shastry S, Wilhelm A, Courville C, Tang LP, Backer K, et al. A
865 pharmacologic approach to acquired cystic fibrosis transmembrane conductance regulator
866 dysfunction in smoking related lung disease. PLoS One. 2012;7(6):e39809.
867 79. Mall MA. Unplugging Mucus in Cystic Fibrosis and Chronic Obstructive Pulmonary
868 Disease. Ann Am Thorac Soc. 2016;13 Suppl 2:S177-85.
869

870

871 **Information about Supplementary Material**

872 File name: Supplementary Material

873 File format: docx

874 Title of the data: Supplementary Methods; Supplementary Figures and Supplementary Tables.

875 Description of the Data: This file includes the supplementary materials for the paper '**Genetic**
876 **evidence supports the development of SLC26A9 targeting therapies for the treatment of**
877 **lung disease'**
878 Supplementary Methods includes the following subsections:
879 *Sample Genotyping and Quality Control* ; *HBE and HNE sampling*;*Cell Culturing;*
880 *Ussing chamber studies with primary human nasal epithelial cells*; *Single Cell RNA-seq Sample*
881 *processing* and *PheWAS data extraction and Colocalization with CF GWAS Summary Statistics.*
882 There are two figures in section **Supplementary Figures:**
883 **Fig.S1**. Forest plot of association between rs7512462 and lung function, measured as Saknorm,
884 prior to modulator treatment, if applicable, in independent samples with at least one G551D
885 variant.
886 **Fig.S2.** *SLC26A9* gene expression in cultured primary nasal cells from the same CF patients
887 (n=9) in CGMS at two different time points. Both cultures were P3.
888 There are two tables in section **Supplementary Tables:**
889 **Supplementary Table 1.** The samples removed by the exclusion criteria for the CFTR
890 modulator study.
891 **Supplementary Table 2.** The association between rs7512462 genotype and COPD lung function
892 phenotypes in the UK Biobank with colocalization p-values calculated with the CF GWAS p-
893 values reported in (1).
894

895 **Declarations**

896 **Ethics declarations**

897 The Canadian Gene Modifier Study (CGMS) was approved by the Research Ethics Board of the
898 Hospital for Sick Children (# 0020020214 from 2012-2019 and #1000065760 from 2019-
899 present) and all participating sub-sites. Written informed consent was obtained from all
900 participants or parents/guardians/substitute decision makers prior to inclusion in the study. The
901 CGMS is approved by the Research Ethics Board of the Hospital for Sick Children for the usage
902 of public and external data. The US PROSPECT study provides data from a clinical trial
903 registered at clinicaltrial.gov, identifier NCT02477319, and we obtained these data through
904 application to the US CFF at https://www.cff.org/Research/Researcher-Resources/.

905 **Consents for publication**

906 Not applicable.

907 **Availability of data and materials**

908 The datasets generated and/or analyzed in this paper are publicly available. Data from the
909 CGMS analyzed for the lung function pre- and post-modulator treatment are available from
910 Canadian CF registry at https://www.cysticfibrosis.ca/our-programs/cf-registry/requesting-
911 canadian-cf-registry-data; the functional data and RNA-seq data from CGMS is available from
912 the CFIT program at https://lab.research.sickkids.ca/cfit/cystic-fibrosis-patients-families-
913 researchers/, and the paired cultured and fresh naïve HNE and HBE is available at GEO
914 (GSE172232). The US PROSPECT data were obtained through application to the US CFF at
915 https://www.cff.org/Research/Researcher-Resources/ and the study is registered
916 on https://clinicaltrials.gov/ct2/show/NCT02477319. The single cell RNA-sequencing data are
917 downloaded from the Human Protein Atlas Program http://www.proteinatlas.org. The summary
918 statistics for the pheWAS study is available at https://atlas.ctglab.nl/PheWAS and the meconium
919 ileus association results that have been colocalized can be downloaded at

920 https://lab.research.sickkids.ca/strug/publications-software/. The data used for the COPD
921 analysis are available through application to the UKBiobank.

922 **Competing interests**

923 L.B. participated in a Vertex Virtual Advisory Board and she is a member of the CF Annual
924 Faculty, sponsored by Vertex Pharmaceuticals. D.M-C. received an honorarium for teaching
925 module development for Vertex Pharmaceuticals. N. M. is doing contract research trials for
926 Vertex Phaemaceuticals and Abbvie. A.L.S has received speaking fees for educational programs
927 sponsored by Vertex Pharmaceuticals. B.S.Q. has received speaker fees from Vertex
928 Pharmaceuticals and has served as site PI for several Vertex-sponsored clinical trials .W.M.L is a
929 study investigator for Vertex Pharmaceuticals. E.T., T.G. and F.R. act as a consultant for Vertex
930 Pharmaceuticals. M.S. participated in Vertex clinical trials and received payment for education
931 modules. J.G., G.H., C.W., C. Bartlett, N.P., F.L., K.K., J.A., A.H., M.S., M.E., G.C.M., D.A.,
932 S.B., C.Bjornson, M.C., J.R., A.P., M.P., R.V.W., Y.B., D.H., M.J.S., J.B., P.W., L.S., E.B.,
933 T.M., J.M.R and L.J.S. have no conflicts of interest.

934

935 **Funding**

936 Funding was provided by Cystic Fibrosis Foundation STRUG17PO, MORAES1610; Canadian
937 Institutes of Health Research (FRN 167282), Cystic Fibrosis Canada (2626) and the CFIT
938 Program funded by the SickKids Foundation and CF Canada; Natural Sciences and Engineering
939 Research Council of Canada (RGPIN-2015- 03742, 250053-2013). This work was also funded
940 by the Government of Canada through Genome Canada (OGI-148) and supported by a grant

941 from the Government of Ontario. The funders of the study play no role in study design, data
942 collection and analysis, decision to publish or preparation of the manuscript.

943

944 **Author contributions:**

945 Conceptualization: LJSData curation: JG, GH, CW, CBartlett, NP, FL, KK, JA, AH, MShaw,
946 ME, GCM, DA, SB, CBjornson, MC, JR,AP,MP, RVW, YB, LB, DMC, DH, MJS, NM, JB, ET,
947 ALS, BSQ, PW, WML, MSolomon, EB, TJM, TG, FR
948 Formal analysis: JG, Gh, CW, NP, LJS
949 Funding acquisition: LJS
950 Investigation: JG, LJS
951 Methodology: JG, GH, CW, NP, LJS
952 Project administration: LJS
953 Supervision: LJS
954 Visualization: JG, CW
955 Writing-original draft: JG, GH, CW, NP, JMR, LJS
956 Writing—review & editing: all authors

957

958 **Acknowledgments**

959 We thank the US CF Foundation for the use of Part B of the PROSPECT observational study.
960 We thank the Cystic Fibrosis Canada–SickKids Program for Individualized Therapy for access to
961 the RNA-sequencing of the nasal epithelia. We thank the patients, care providers and clinic
962 research assistants, collaborators, and principal investigators involved in CF Centers throughout
963 Canada for their contributions to the CF Canada Patient Registry and Canadian Gene Modifier

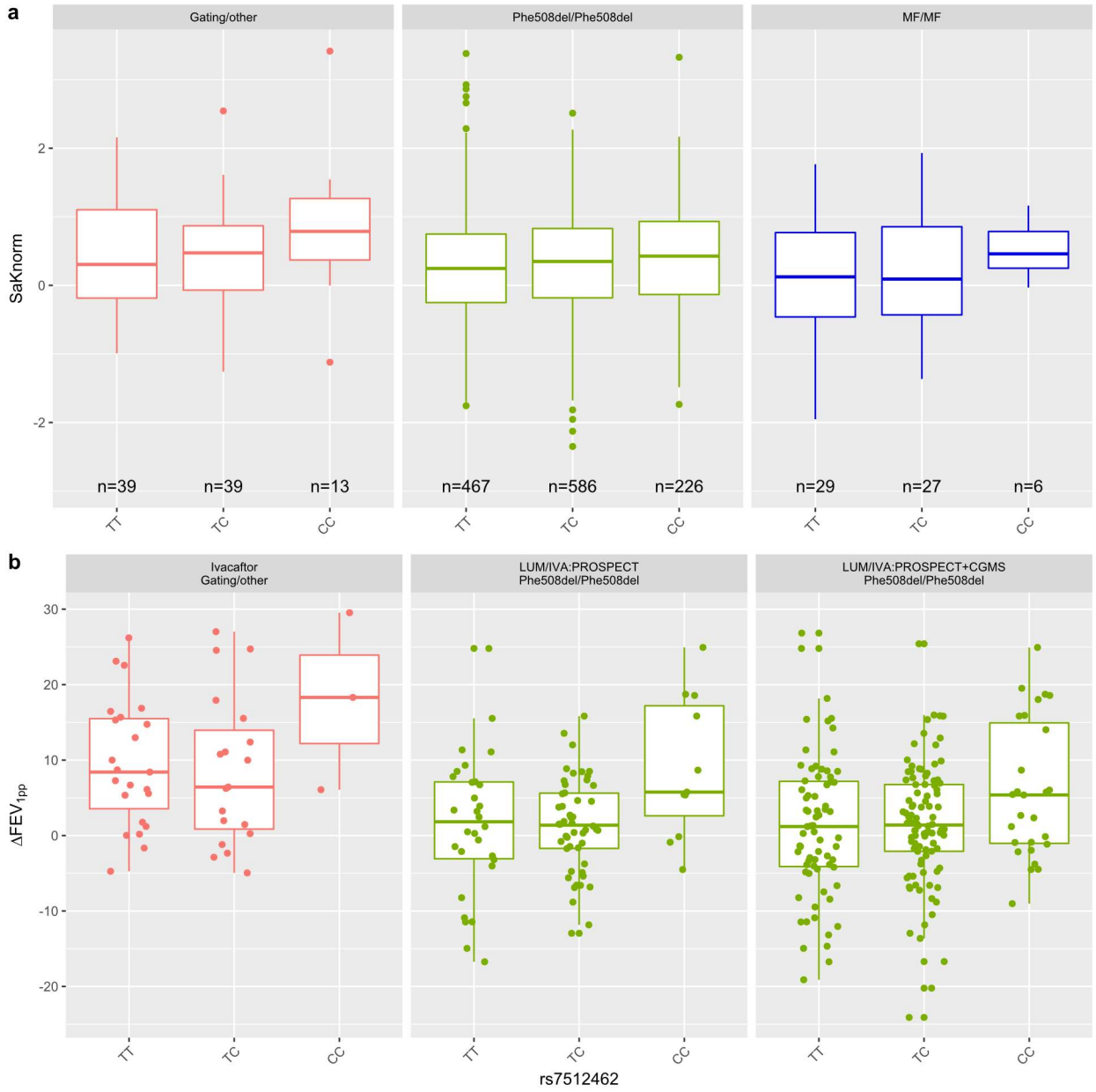
964 Study. Some of the research was conducted using the UK Biobank Resource under application
965 Number 40946. We thank Shaf Keshavjee for providing the bronchial cells for paired analysis of
966 gene expression and Hong Ouyang for culturing the epithelial cells. We thank Dr. John Miller
967 for providing the amplifier for the study. The authors wish to acknowledge the staff supporting
968 the High Performance Computing cluster and research helpdesk department and The Centre for
969 Applied Genomics at the Hospital for Sick Children, Toronto.

970

971

972

973 **Figures**



974

975

976 **Fig. 1.** Participant lung function or lung treatment response categorized by rs7512462 genotype.

977 (a) Boxplots of lung function from CGMS participants pre-modulator treatment, measured as

978 Saknorm across different CFTR mutation groups. Saknorm is calculated as in (5, 37) with FEV_1

979 measurements taken prior to modulator treatment, if applicable. Sample sizes by rs7512462 for

980 each CFTR group are displayed at the bottom of each plot. (b) Boxplot includes an overlay with
981 a stripchart for treatment response in patients on IVA (n=45), LUM/IVA in PROSPECT cohort
982 (n=91) or LUM/IVA in combined PROSPECT and CGMS samples (n=196). Each dot represents
983 an individual measurement. Following (29), the treatment response for IVA is defined as the
984 difference between the mean post-treatment FEV_{1pp} within 15 to 400 days and FEV_{1pp} baseline.
985 Response for LUM/IVA is defined as the difference in FEV_{1pp} between the first post treatment
986 within 5 to 7 months to the baseline ((40); Methods).

987

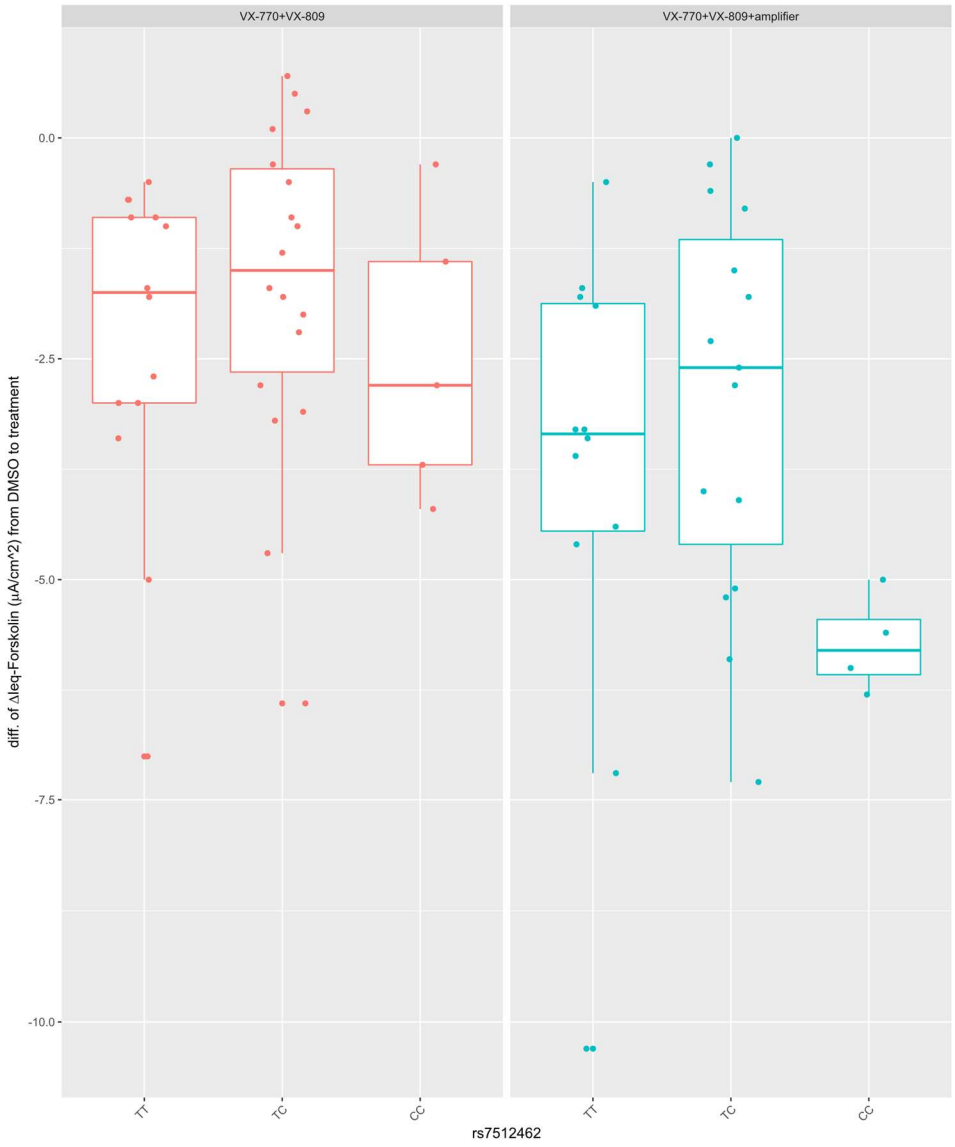


Fig.2. Boxplot overlayed with stripchart for CFTR function by rs7512462 genotype in HNE cultured to P2. Each dot represents an individual measurement. The CFTR function is defined as the difference in ΔIeq-forskolin from treatment with VX-770+VX-809 (left; n=37) and VX-770+VX-809+amplifier (right; n=31) to DMSO in CFTR-mediated current in cultured HNE from participants homozygous for Phe508del. More negative measurements reflect greater CFTR function.

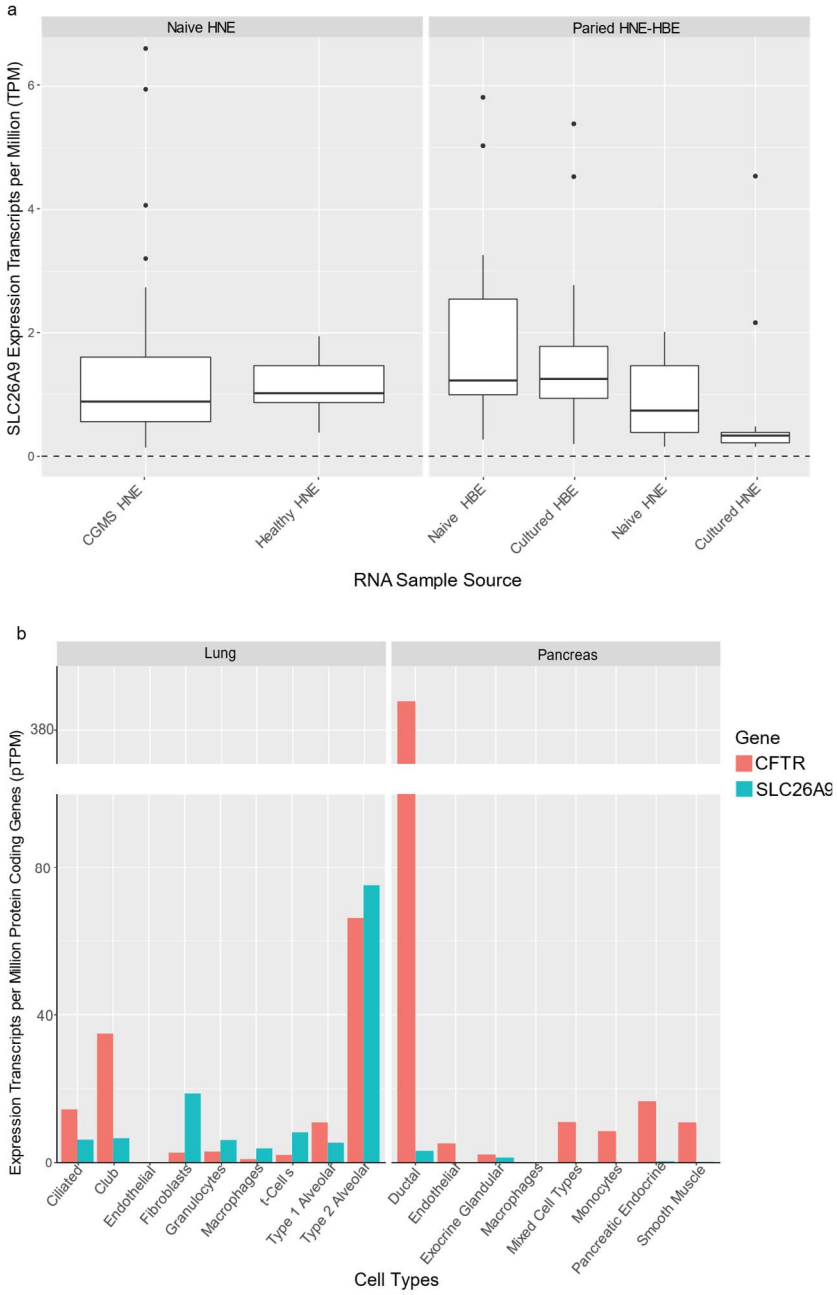


Fig. 3. *SLC26A9* gene expression in various tissue models to guide functional studies. (a) *SLC26A9* expression is low in the bulk RNA-Seq of CGMS CF participants and healthy controls in naïve HNE (left), with diminishing expression with culturing of both HNE and HBE (right). (b) Single cell RNA-Seq studies of non-CF lung and pancreas tissues from the Human Protein

1002 Atlas (42, 43). CFTR expression transcripts per million protein coding genes (pTPM) in orange

1003 and SLC26A9 pTPM in blue.

1004

1005

1006

1007

1008

1009

1010

1011

1012

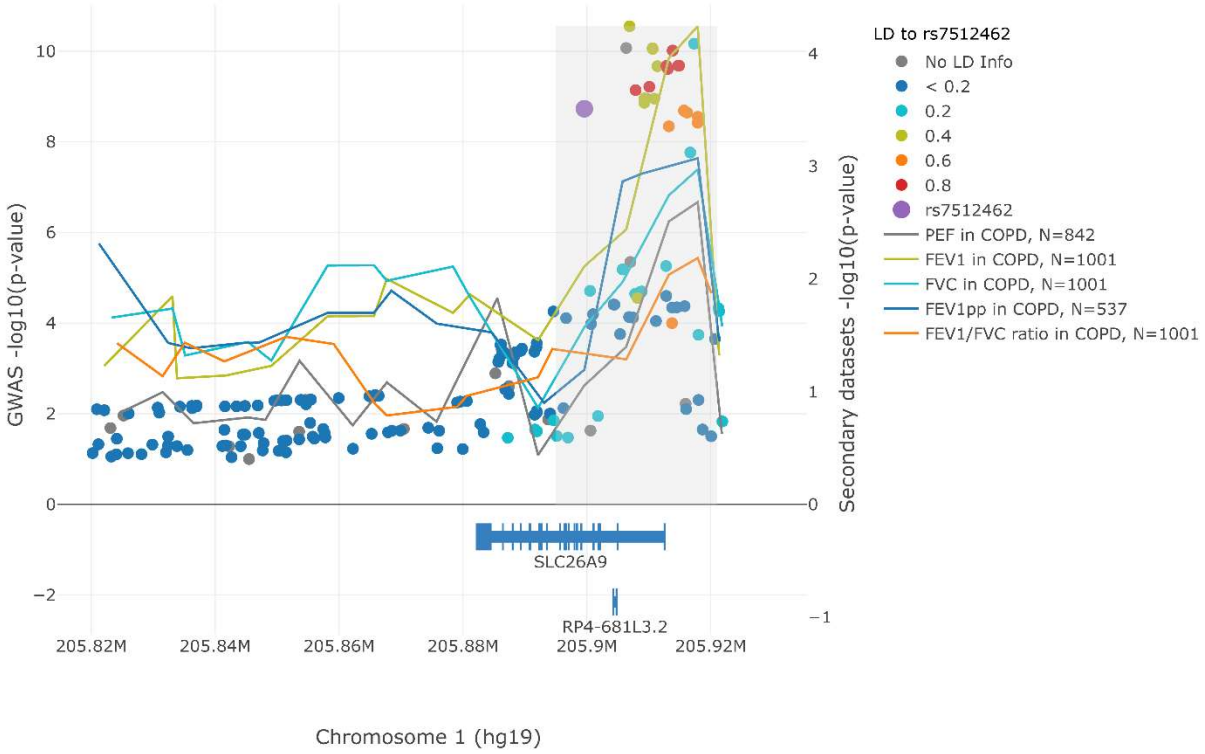


Fig. 4. Colocalization visualization of lung function phenotypes in COPD and CF GWAS p-values at SLC26A9 (1). Figure is plot by LocusFocus (60). P-value scale on the left reflects CF GWAS results plotted as dots and p-value scale on the right reflects the lung-related COPD phenotype association by a line connecting COPD lung genetic association minimum p-values in moving windows. The grey bar indicates the region for which the colocalization analysis is carried out using the Simple Sum statistic (1) with the colocalization p-values shown in Supplementary Table 2. All spirometry phenotype lines in the figure are corresponding to the best measures as defined in (58) with sample sizes reported in Supplementary Table 2.