Single-nucleotide polymorphisms in the sulfatase-modifying factor 1 gene are associated with lung function and COPD

Linnea Jarenbäck [1], Sophia Frantz[2], Julie Weidner [1], Jaro Ankerst[1], Ulf Nihlén [1], Leif Bjermer[1], Per Wollmer[2] and Ellen Tufvesson [1]

[1]Dept of Clinical Sciences Lund, Respiratory Medicine and Allergology, Lund University, Lund, Sweden. [2]Dept of Translational Science, Clinical Physiology, Lund University, Skåne University Hospital, Malmö, Sweden.

Corresponding author: Ellen Tufvesson (ellen.tufvesson@med.lu.se)

Shareable abstract (@ERSpublications)
This study shows that several SNPs in the *SUMF1* gene are associated with COPD and impaired lung function. These SNPs appear in two blocks: block 2 is associated with COPD and airway obstruction, and block 1 is associated with impaired lung volumes. https://bit.ly/3IiZXcC

Cite this article as: Jarenbäck L, Frantz S, Weidner J, *et al*. Single-nucleotide polymorphisms in the sulfatase-modifying factor 1 gene are associated with lung function and COPD. *ERJ Open Res* 2022; 8: 00668-2021 [DOI: 10.1183/23120541.00668-2021].

Received: 4 Dec 2021
Accepted: 17 Feb 2022

Abstract

Single nucleotide polymorphisms (SNPs) in various genes have been shown to associate with COPD, suggesting a role in disease pathogenesis. Sulfatase modifying factor (SUMF1) is a key modifier in connective tissue remodelling, and we have shown previously that several SNPs in *SUMF1* are associated with COPD. The aim of this study was to investigate the association between *SUMF1* SNPs and advanced lung function characteristics.

Never-, former and current smokers with (n=154) or without (n=405) COPD were genotyped for 21 SNPs in *SUMF1* and underwent spirometry, body plethysmography, diffusing capacity of the lung for carbon monoxide (D_{LCO}) measurement and impulse oscillometry.

Four SNPs (rs793391, rs12634248, rs2819590 and rs304092) showed a significantly decreased odds ratio of having COPD when heterozygous for the variance allele, together with a lower forced expiratory volume in 1 s (FEV_1) and FEV_1/forced vital capacity (FVC) ratio and an impaired peripheral resistance and reactance. Moreover, individuals homozygous for the variance allele of rs3864051 exhibited a strong association to COPD, a lower FEV_1/FVC, FEV_1 and D_{LCO}, and an impaired peripheral resistance and reactance. Other SNPs (rs4685744, rs2819562, rs2819561 and rs11915920) were instead associated with impaired lung volumes and exhibited a lower FVC, total lung capacity and alveolar volume, in individuals having the variance allele.

Several SNPs in the *SUMF1* gene are shown to be associated with COPD and impaired lung function. These genetic variants of *SUMF1* may cause a deficient sulfation balance in the extracellular matrix of the lung tissue, thereby contributing to the development of COPD.

Introduction

COPD is mainly caused by long-term cigarette smoking [1], but multiple genetic factors may influence predisposition to lung damage and thereby the susceptibility for developing COPD [2]. Given the heterogeneous nature of COPD, it is most likely that COPD patients have different genetic patterns. The decreased lung function seen in COPD patients is mainly due to airway inflammation caused by oxidative stress, which subsequently leads to airways remodelling and tissue destruction [3–6]. Different gene polymorphisms related to these processes are important to investigate to better understand their role in disease development. The most well-known genetic factor associated with COPD is alpha-1 antitrypsin deficiency, which is associated with polymorphism in the *SERPINA1* gene [7]. In addition, mainly single nucleotide polymorphisms (SNPs) associated with inflammatory processes [8] and biological stress pathways [9] have been identified to be linked to COPD. Some polymorphisms related to connective tissue remodelling have also been identified, such as matrix metalloproteinase (MMP)-7, which is suggested to influence early development of COPD [10], and MMP-12, which was found to be associated with severe/very severe COPD [11].

In a prior study, we have found several SNPs in sulfatase modifying factor 1 (*SUMF1*) to be associated with COPD [12]. SUMF1 is the main regulator of all known sulfatases in the body [13], and modifies them into their active state. The role of the different sulfatases is then to remove sulphate from specific sulfated carbohydrate chains, and they thereby have an important role in the delicate balance of connective tissue remodelling. SUMF1 is well known for its implication in multiple sulfatase deficiency [14–18], but ARTEAGA-SOLIS *et al.* [19] observed that *Sumf1*$^{-/-}$ mice exhibited an emphysema-like pattern in their lungs, due to post-natal alveolarisation arrest. In our previous clinical study, we showed that several SNPs in *SUMF1* were associated with COPD, of which rs793391 was the most significant. 12 *SUMF1* SNPs were found to be significant by expression quantitative trait loci (eQTL) analysis, and certain splice variants of *SUMF1* exhibited decreased expression levels in sputum cells from COPD patients compared to control subjects. In association with the *SUMF1* SNP rs11915920, which was a top hit in the eQTL analyses, we found decreased mRNA expression levels in sputum cells and lung fibroblasts in subjects with the variance allele, confirming the results of the lung tissue eQTL analysis [12]. A previous genome-wide association study has also identified *SUMF1* to be associated with prominent emphysema, but was not studied further [20].

The main objective of the present study was to investigate the association between *SUMF1* SNPs and advanced lung function characteristics. A secondary aim was to verify findings from our previous study in a larger, well-defined cohort and to investigate if additional SNPs in *SUMF1* showed an association to lung function and COPD.

Material and methods

Study population

598 patients were recruited from two sites within Skåne University Hospital (Sweden): the lung clinic in Lund and at the clinical physiology department in Malmö. In Lund, smokers and former smokers with or without COPD were recruited [21]. In the Malmö cohort, never-smokers without any reported lung disease or lower respiratory symptoms, smokers and former smokers with and without COPD and subjects with a self-reported diagnosis of chronic bronchitis/emphysema/COPD were included [22].

All subjects signed written informed consent and the regional ethical review board in Lund approved the studies (431/2008 and 786/2003+amendment 101/2015).

Lung function measurements

Spirometry (MasterScreen; Erich Jaeger, Würzburg, Germany), body plethysmography (MasterScreen Body; Erich Jaeger), diffusing capacity of the lung for carbon monoxide (D_{LCO}) (MasterScreen Diffusion; Erich Jaeger) and impulse oscillometry (MasterScreen) were performed after inhalation of β-agonist (Lund: 400 mg salbutamol, HandiHaler using a spacer device; Malmö: 1 mg terbutaline, Turbohaler). All measurements were performed according to manufacturers' protocols and according to European Respiratory Society/American Thoracic Society recommendations [23] when applicable. Reference values from the Global Lung Function Initiative [24–26] were used to calculate percentage of predicted normal.

Blood samples and SUMF1 genotyping

Whole blood was drawn from all subjects and stored in EDTA tubes at −80°C until analysis. DNA was extracted and genotyped for 21 SNPs in *SUMF1*: rs11915920, rs12634249, rs1356229, rs137852846, rs137852848, rs137852849, rs137852854, rs2322683, rs2633852, rs2819561, rs2819562, rs2819590, rs304092, rs308739, rs3864051, rs4685744, rs748169616, rs793391, rs794185, rs794187 and rs807785. This was analysed using Agena iPLEX genotyping at the mutation analysis facility at Karolinska University Hospital (Huddinge, Sweden) using iPLEX Gold chemistry and MassARRAY mass spectrometry system [27] (Agena Bioscience, San Diego, CA, USA), as described previously [12].

Statistical analysis

Baseline difference in demographic data and lung function between COPD and controls were analysed using t-test or Chi-squared test. Associations of SNPs to COPD were analysed with multivariable logistic regression models including the SNP in an additive genetic model and were adjusted for confounding factors (age, sex and smoking status). Associations of SNPs to lung function parameters were done with multivariable linear regression, corrected for confounding factors (age, sex, height and smoking status). Figures 1–3 are presented as percentage predicted when applicable, to increase visualisation, and data are analysed with one-way ANOVA for overall comparison among the groups, followed by Bonferroni's multiple comparison test between separate groups.

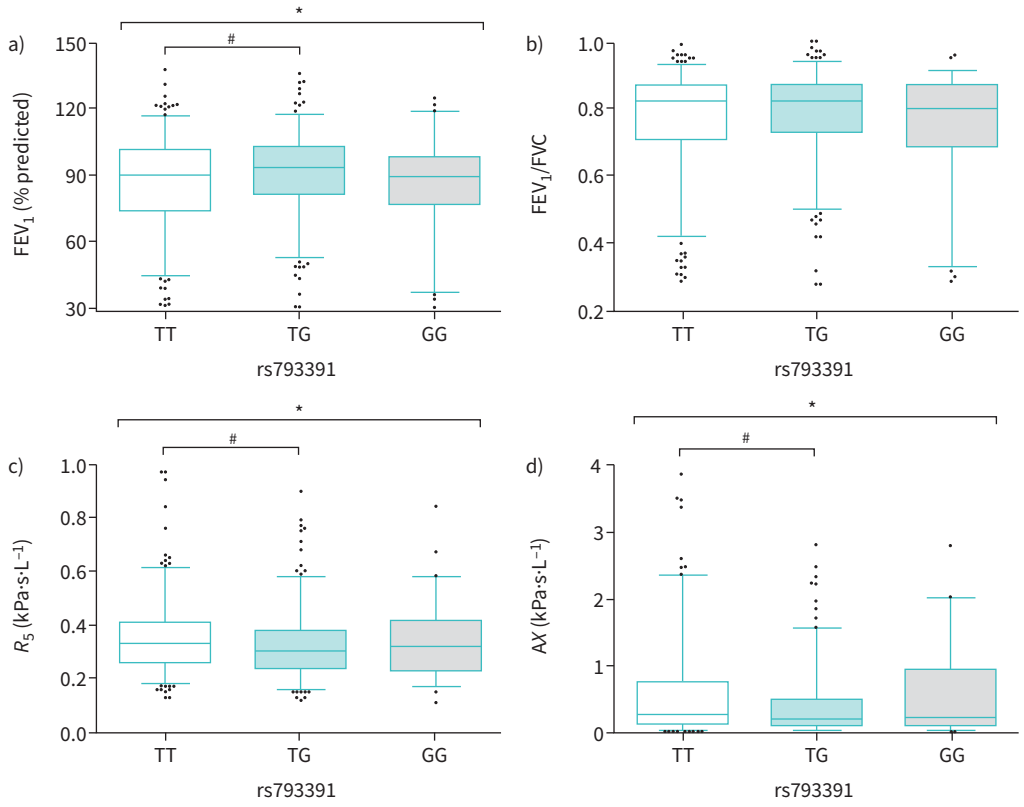

FIGURE 1 Lung function in COPD patients and controls divided by single nucleotide polymorphism rs793391 genotypes (as representative for block 2). a) Forced expiratory volume in 1 s (FEV_1), b) FEV_1/forced vital capacity (FVC) ratio, c) resistance at 5 Hz (R_5) and d) reactance area (*AX*) are divided according to the genotype of rs793391. One-way ANOVA was used for overall comparison among the groups (*: p<0.05), followed by Bonferroni's multiple comparison test between separate groups (#: p<0.05). Genotypes are presented with the reference/reference genotype to the left. Data are presented as box-plots showing the median within the box of 25th–75th percentile and whiskers of 5th–95th percentile.

Results

Subject characteristics

DNA from whole blood from a total of 559 subjects (405 controls and 154 COPD patients) were successfully genotyped and included in the final analysis (39 subjects were excluded due to other lung diseases, unsuccessful lung function performance or unsuccessful SNP analysis). Subjects were divided into two groups, COPD or healthy, according to the Global Initiative for Chronic Obstructive Lung Disease definition [28]. Subjects with COPD were significantly older and had more pack-years of smoking compared to the control group (table 1). In addition, the COPD group consisted of a larger proportion of males and former smokers, and fewer never-smokers and current smokers. As expected, all lung function variables were impaired in COPD compared to controls (table 1).

Genotype frequencies

16 out of the 21 analysed SNPs had a variation among our study subjects, while five SNP genotypes (rs137852846, rs137852848, rs137852849, rs137852854 and rs748169616) were monomorphic, *i.e.* presenting as homozygous for the reference allele in all subjects, and therefore not included in the analyses.

Genotype frequencies for COPD and controls are presented separately in table 2, and variant allele frequencies of all SNPs are presented in supplementary table S1.

Association to COPD

After adjusting for age, sex and smoking status, five of the SNPs were found to be significantly associated with COPD (table 2). Four of them (rs793391, rs12634248, rs2819590 and rs304092) exhibited a

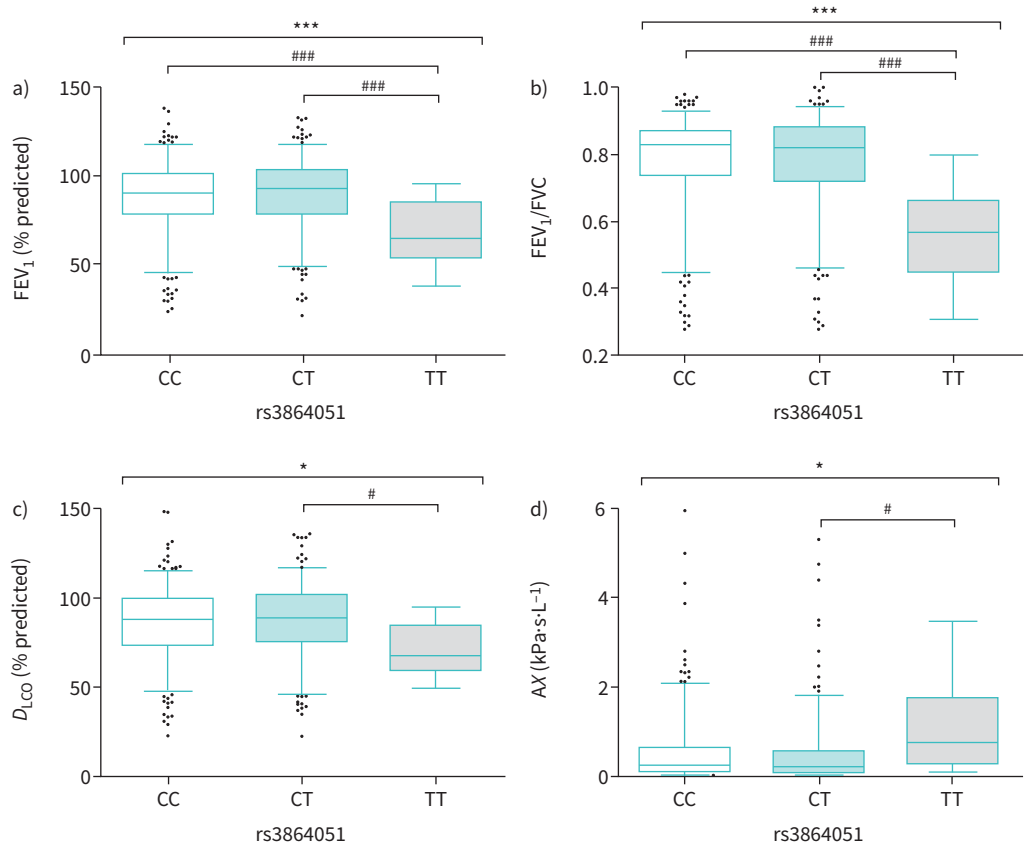

FIGURE 2 Lung function in COPD patients and controls divided by single nucleotide polymorphism rs3864051 genotypes. a) Forced expiratory volume in 1 s (FEV_1), b) FEV_1/forced vital capacity (FVC), c) diffusing capacity of the lung for carbon monoxide (D_{LCO}) and d) reactance area (AX) are divided according to the genotype of rs3864051. One-way ANOVA was used for overall comparison among the groups (*: p<0.05, ***: p<0.001), followed by Bonferroni's multiple comparison test between separate groups ($^{\#}$: p<0.05, $^{\#\#\#}$: p<0.001). Genotypes are presented with the reference/reference genotype to the left. Data are presented as box-plots showing the median within the box of 25th–75th percentile and whiskers of 5th–95th percentile.

significantly decreased odds ratio of having COPD if heterozygous for the variance allele. Furthermore, when homozygous for the variance allele for any of the aforementioned SNPs, a decreased odds ratio of having COPD was not observed.

rs3864051 displayed the most significant association with COPD, with 11 out of the 13 subjects that were homozygous for the variance allele having COPD. In addition, a tendency towards an increased risk of having COPD was seen in subjects heterozygous for the variance allele rs3864051.

Association with advanced lung function

When examining the different lung function parameters, as expected, the forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) ratio and FEV_1 were similarly associated to the same SNPs as having COPD, showing a lower FEV_1/FVC ratio and a lower FEV_1 if homozygous for the reference allele of rs793391 (figure 1a and b), rs12634248, rs2819590 and rs304092 compared to being heterozygous for the variance allele (table 3). Additionally, two other SNPs (rs794185 and rs794187) showed a similar association to FEV_1, with lower FEV_1 if homozygous for the reference allele, but were not associated with having COPD. In conjunction with the association to COPD, lower FEV_1/FVC and FEV_1 were seen if subjects were homozygous for the variance allele of rs3864051 (table 3 and figure 2a and b).

In contrast, other SNPs were associated with FVC, showing a significantly lower FVC if the variance allele (whether heterozygous or homozygous) for rs4685744, rs2819562, rs2819561 or rs11915920 (table 3

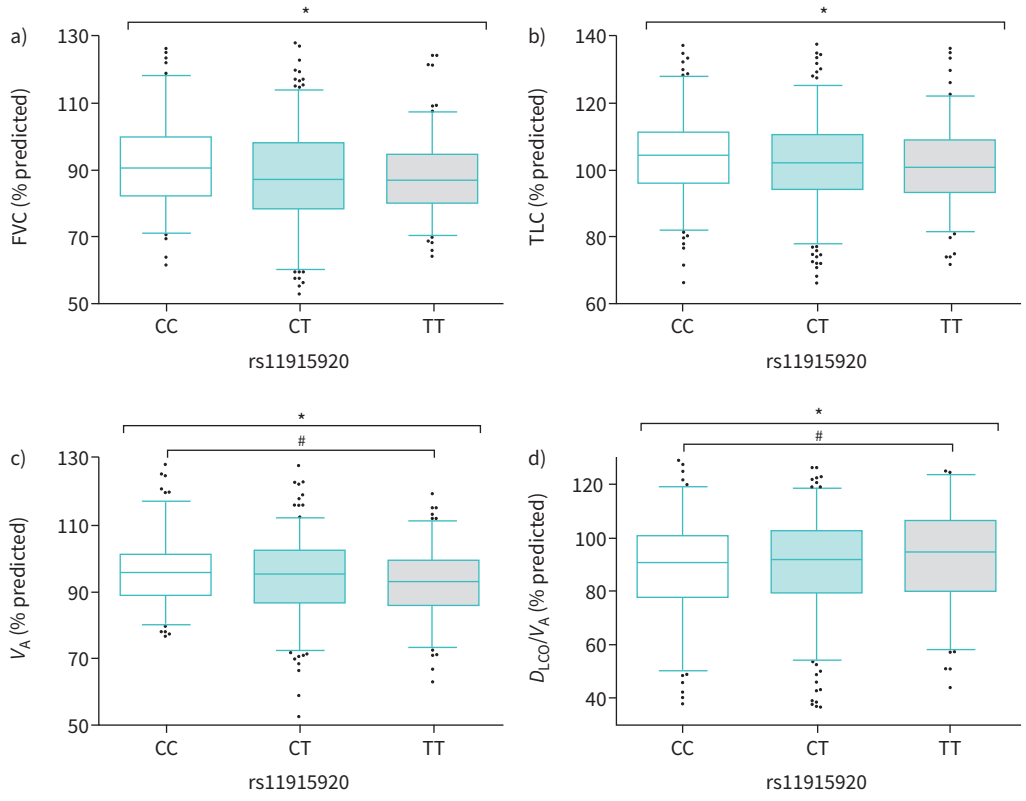

FIGURE 3 Lung function in COPD patients and controls divided by single nucleotide polymorphism rs11915920 genotypes (as representative for block 1). a) Forced vital capacity (FVC), b) total lung capacity (TLC), c) alveolar volume (V_A) and d) diffusing capacity of the lung for carbon monoxide (D_{LCO})/V_A are divided according to the genotype of 11915920. One-way ANOVA was used for overall comparison among the groups (*: $p<0.05$), followed by Bonferroni's multiple comparison test between separate groups (#: $p<0.05$). Genotypes are presented with the reference/reference genotype to the left. Data are presented as box-plots showing the median within the box of 25th–75th percentile and whiskers of 5th–95th percentile.

and figure 3a) was present, and a tendency of association also observed for rs2633852. In contrast, subjects heterozygous for the variance allele of rs794185 and rs794187 had a higher FVC if heterozygous for the variance allele (table 3), which followed the pattern of FEV_1 for these SNPs.

Besides being associated to FVC, the same SNPs (rs4685744, rs2819562, rs2819561, rs2633852 and rs11915920) were also significantly associated with total lung capacity (TLC), exhibiting a lower TLC if the variance allele was present, and most significant if homozygous for the variance allele (table 4 and figure 3b). These SNPs also showed a similar association to residual volume (RV), but this was not significant, with a lower RV if homozygous for the variance allele. In addition, these SNPs show a higher D_{LCO}/alveolar volume (V_A) ratio if homozygous for the variance allele, which is mostly due to a low V_A (figure 3c and d), but also a higher D_{LCO}. Consistent with the association between rs3864051 and COPD, a lower D_{LCO} was observed if this SNP was homozygous for the variance allele (figure 2c).

The resistance and reactance results were consistent with the findings of the SNPs associated to COPD and FEV_1 (table 5). Total resistance (R_5) (possibly here reflecting the peripheral airways as it is in concordance with R_5–R_{20}, which is a variable that is commonly taken to represent resistance of peripheral airways), is lower if heterozygous for the variance allele of rs793391 (figure 1c), rs12634248, rs2819590, rs304092, rs794185 and rs794187. A similar pattern was seen in airway reactance (reactance at 5 Hz and reactance area), which was also less impaired if heterozygous for the variance allele of these SNPs (figure 1d). Consistent with its association to COPD and FEV_1, rs3864051 exhibited impaired resistance and reactance if homozygous for the variance allele (figure 2d).

There were no associations between rs1356229 or rs308739 and lung function (data not shown).

TABLE 1 Patient characteristics and lung function

	Controls	COPD	p-value
Patients	405	154	
Female/male	239/166 (59/41)	71/83 (46/54)	**0.006**
Age, years	62±8	66±6	**<0.001**
Weight, kg	77±16	77±15	0.99
Height, cm	170±9	171±9	**0.025**
Smoking status, never-/former/current smoker	76/145/184	5/92/57	**<0.001**
Smoking, pack-years	22±18	36±20	**<0.001**
GOLD stage 1/2/3/4	NA	23/96/31/4	NA
Lung function			
FEV_1, L	2.85±0.71	1.87±0.64	**<0.001**
FEV_1, % pred	97±13	63±17	**<0.001**
FVC, L	3.40±0.85	3.19±1.02	**0.023**
FVC, % pred	91±12	83±19	**<0.001**
FEV_1/FVC	0.84±0.06	0.60±0.17	**<0.001**
RV, % pred	116±21	150±45	**<0.001**
TLC, % pred	102±11	106±17	**0.003**
D_{LCO}, % pred	93±16	66±20	**<0.001**
V_A, % pred	96±16	91±13	**<0.001**
D_{LCO}/V_A, % pred	97±15	72±20	**<0.001**
R_5, kPa·s·L^{-1}	0.31±0.11	0.41±0.17	**<0.001**
R_{20}, kPa·s·L^{-1}	0.24±0.08	0.28±0.09	**<0.001**
R_5–R_{20}, kPa·s·L^{-1}	0.07±0.05	0.13±0.10	**<0.001**
X_5, kPa·s·L^{-1}	−0.09±0.05	−0.18±0.13	**<0.001**
A*X*, kPa·s·L^{-1}	0.34±0.39	1.13±1.33	**<0.001**

Data are presented as n, n (%) or mean±SD, unless otherwise stated. Significant differences are depicted in bold. Categorical data were analysed using the Chi-squared test and numeric data were analysed with the t-test. GOLD: Global Initiative for Chronic Obstructive Lung Disease; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; RV: residual volume; TLC: total lung capacity; D_{LCO}: diffusing lung capacity of the lung for carbon monoxide; V_A: alveolar volume; R_5: resistance at 5 Hz; R_{20}: resistance at 20 Hz; X_5: reactance at 5 Hz; A*X*: reactance area; NA: not applicable.

Discussion

In the present study, several SNPs in the *SUMF1* gene were shown to be associated with COPD and with impaired lung function. These SNPs appear in two blocks (supplementary figure S1A) where one block was associated with COPD and airway obstruction, and the other block was associated with impaired lung volumes. One block (block 2) included rs7933191, rs12634249, rs2819590 and rs304092, which were associated with having COPD, affecting FEV_1 and FEV_1/FVC, and airway resistance and reactance. This was also partly the case for rs794187 and rs794185. The other block (block 1) included rs4685744, rs2819562, rs2819561, rs2633852 and rs11915920, and was associated with impaired FVC, TLC, RV and V_A. Furthermore, the most evident SNP associated with COPD was rs3864051, which also exhibited associations with impaired FEV_1, FEV_1/FVC, D_{LCO} and airway resistance and reactance.

Moreover, we here confirm in a separate, larger number of subjects our previous finding that SNP rs793391 was associated with COPD [12]. Furthermore, we determined additional SNPs in *SUMF1* (in block 2) associated with COPD: rs12634248, rs2819590 and rs304092. Similarly, these SNPs, together with rs794185 and rs794187, show associations to FEV_1, FEV_1/FVC and airway resistance and reactance, typical for an obstructive lung function profile. These SNPs showed weaker association to airway obstruction when the subject was heterozygous for the variance allele, suggesting a protective role for having one variance allele over two reference alleles.

In contrast, the SNPs in the other block (block 1) showed lower lung volumes (FVC, TLC, RV, V_A) when heterozygous for the variance allele, and in most cases even more impairment if homozygous for the variance allele. Two of these SNPs (rs11915920 and rs2819562) were included in a previous investigation using an eQTL analysis in lung tissue [12] and showed a very strong association between gene expression of *SUMF1* in lung tissue and the genetic polymorphisms. Lower levels of *SUMF1* gene expression were seen in subjects having the variance allele, both if being heterozygous and even more prominent in those being homozygous for the variance allele. This might cause an impaired SUMF1 function, and thereby a

TABLE 2 Genotype frequencies in healthy and COPD groups, and associations between different genotypes of sulfatase modifying factor (*SUMF1*) single nucleotide polymorphisms (SNPs)[#] and COPD[¶]

	ref/ref			ref/var var/var			p-value	OR (95% CI)[+]
		Healthy	COPD		Healthy	COPD		
rs1356229	CC	387 (96)	149 (97)	CT	16 (4)	5 (3)	0.74	1.20 (0.42–3.45)
				TT	0 (0)	0 (0)		
rs308739	CC	348 (87)	136 (88)	TT	54 (13)	18 (12)	0.98	1.01 (0.56–1.82)
				AA	0 (0)	0 (0)		
rs4685744	CC	109 (27)	43 (28)	CT	187 (46)	73 (47)	0.83	0.95 (0.60–1.51)
				TT	107 (27)	38 (25)	0.60	0.87 (0.51–1.47)
rs2819562	CC	87 (22)	36 (23)	CT	182 (45)	74 (48)	0.63	0.89 (0.54–1.45)
				TT	133 (33)	44 (29)	0.29	0.75 (0.44–1.28)
rs2819561	AA	87 (22)	36 (23)	AG	183 (46)	74 (48)	0.61	0.88 (0.54–1.44)
				GG	132 (33)	44 (29)	0.29	0.75 (0.44–1.29)
rs2633852	AA	85 (21)	36 (23)	AG	181 (45)	74 (48)	0.61	0.88 (0.54–1.44)
				GG	134 (34)	44 (29)	0.26	0.73 (0.43–1.26)
rs11915920	CC	111 (28)	43 (28)	CT	186 (46)	75 (49)	0.99	1.00 (0.63–1.58)
				TT	105 (26)	36 (23)	0.55	0.85 (0.50–1.45)
rs807785	CC	33 (8)	17 (11)	CT	164 (41)	70 (45)	0.88	0.95 (0.49–1.85)
				TT	206 (51)	67 (44)	0.39	0.75 (0.39–1.46)
rs3864051	CC	228 (57)	69 (45)	CT	171 (43)	74 (48)	0.13	1.36 (0.92–2.03)
				TT	2 (0.5)	11 (7)	**0.001**	**14.22 (3.04–66.49)**
rs794187	CC	163 (41)	68 (44)	CT	190 (47)	70 (46)	0.37	0.83 (0.55–1.25)
				TT	49 (12)	16 (10)	0.46	0.78 (0.41–1.50)
rs794185	TT	114 (28)	52 (34)	TC	211 (52)	76 (49)	0.15	0.73 (0.47–1.12)
				CC	78 (19)	26 (17)	0.33	0.75 (0.43–1.34)
rs2322683	CC	53 (13)	20 (13)	CT	181 (45)	56 (37)	0.51	0.81 (0.44–1.50)
				TT	169 (42)	77 (50)	0.46	1.25 (0.69–2.29)
rs793391	TT	174 (43)	81 (53)	TG	185 (46)	54 (35)	**0.013**	**0.59 (0.39–0.89)**
				GG	44 (11)	19 (12)	0.70	0.88 (0.48–1.64)
rs12634249	CC	225 (56)	96 (62)	CA	157 (39)	46 (30)	**0.034**	**0.63 (0.42–0.97)**
				AA	21 (5)	12 (8)	0.56	1.26 (0.58–2.73)
rs2819590	CC	192 (48)	91 (59)	CT	175 (44)	47 (31)	**0.006**	**0.55 (0.36–0.84)**
				TT	33 (8)	16 (10)	0.97	1.01 (0.52–1.98)
rs304092	GG	191 (47)	86 (56)	GA	176 (44)	53 (34)	**0.031**	**0.63 (0.42–0.96)**
				AA	35 (9)	15 (10)	0.88	0.95 (0.48–1.86)

Data are presented as n (%), unless otherwise stated. Significant associations are depicted in bold. [#]: SNPs are sorted according to chromosome localisation (chromosome localisation is from GRCh38.p12); [¶]: logistic regression between reference allele (ref)/ref and ref/variance allele (var) or var/var as indicated, and adjusted for age, sex and smoking status; [+]: odds ratio of having COPD depending on allele on different SNP in *SUMF1*.

dysfunctional downstream sulfatase imbalance in the lungs [29], thus resulting in potential extracellular matrix issues. Our findings of lower lung volumes associated to *SUMF1* SNPs in this study are in conjunction with the finding of deficient alveolarisation in *Sumf*$^{-/-}$ mice [19], which might be an explanation for the impaired lung function in these subjects.

Another SNP located between the two defined linkage blocks and strongly associated (OR 14.22) with COPD when homozygous for the variance allele was rs3864051. It has been implicated in longevity [30], and the frequency of homozygosity for the variance allele is rare. In our study, it was found in only 13 out of 555 subjects, but 11 out of these subjects had COPD and only two were healthy. However, this could be a significant risk factor for getting COPD, but since the frequency of being homozygote for this allele is low, more studies are needed to verify our findings.

Variant allele frequencies are similar in the population of the present study and the Database of Genotypes and Phenotypes (supplementary table S1), and most SNPs in the present study follow Hardy–Weinberg equilibrium, except for rs3864051. Only 2.3% of the study population were homozygous for the variance allele (TT), instead of the expected 6% (based on variant allele frequency of 0.24 in our population compared to 0.34 in the database). This could either be due to random error in included genotypes, or a

TABLE 3 Associations between different genotypes of sulfatase modifying factor (*SUMF1*) single nucleotide polymorphisms and forced expiratory volume in 1 s (FEV_1), forced vital capacity (FVC) and FEV_1/FVC ratio in the total study population including both COPD and control subjects

	ref/ref	ref/var var/var	FEV_1	FVC	FEV_1/FVC
rs4685744	CC	CT	0.82	**0.030** **B= −0.12**	0.093 B=0.024
		TT	0.81	**0.037** **B= −0.14**	0.27
rs2819562	CC	CT	0.79	**0.048** **B= −0.12**	0.076 B=0.027
		TT	0.73	0.057 B= −0.13	0.074 B=0.030
rs2819561	AA	AG	0.76	**0.050** **B= −0.12**	0.073 B=0.0028
		GG	0.71	0.060 B= −0.12	0.076 B=0.029
rs2633852	AA	AG	0.72	0.062 B= −0.12	0.075 B=0.028
		GG	0.67	0.071 B= −0.12	0.073 B=0.030
rs11915920	CC	CT	0.98	**0.021** **B= −0.13**	0.15
		TT	0.84	**0.031** **B= −0.14**	0.23
rs807785	CC	CT	0.87	0.64	0.61
		TT	0.41	0.87	0.16
rs3864051	CC	CT	0.48	0.14	0.67
		TT	**<0.001** **B= −0.64**	0.31	**<0.001** **B= −0.20**
rs794187	CC	CT	**0.028** **B=0.12**	0.062 B=0.094	0.22
		TT	0.72	0.62	0.98
rs794185	TT	TC	**0.002** **B=0.18**	**0.004** **B=0.16**	0.11
		CC	**0.033** **B=0.16**	0.057 B=0.13	0.32
rs2322683	CC	CT	0.17	0.25	0.41
		TT	0.87	0.57	0.82
rs793391	TT	TG	**0.024** **B=0.12**	0.42	**0.039** **B=0.026**
		GG	0.66	0.65	0.79
rs12634249	CC	CA	**0.039** **B=0.11**	0.30	0.060 B=0.024
		AA	0.12	0.89	0.077 B= −0.045
rs2819590	CC	CT	**0.010** **B=0.14**	0.36	**0.016** **B=0.030**
		TT	0.76	0.74	0.42
rs304092	GG	GA	**0.025** **B=0.12**	0.44	**0.037** **B=0.026**
		AA	0.99	0.53	0.39

Data are presented as p-values from linear regression between the reference allele (ref)/ref and ref/variance allele (var) or var/var (using ref/ref as reference genotype) as indicated, and adjusted for age, sex, height and smoking status. B-value is the unstandardised β-value, and is presented when p<0.1. Significant associations are depicted in bold.

bias in sampling due to COPD patients being less prone to participate in the study. The latter case would suggest even more COPD patients in the group of homozygous for the variance allele, which would strengthen the results of rs3864051 being highly associated with having COPD.

TABLE 4 Associations between different genotypes of sulfatase modifying factor (*SUMF1*) single nucleotide polymorphisms and diffusing capacity and lung volumes in the total study population including both COPD and control subjects

	ref/ref	ref/var var/var	D_{LCO}	V_A	D_{LCO}/V_A	RV	TLC
rs4685744	CC	CT	0.26	0.23	0.080 B=0.048	0.46	0.088 B= −0.14
		TT	0.14	**0.006 B= −0.22**	**0.0 B=0.11**	0.068 B=−0,15	**0.009 B= −0.25**
rs2819562	CC	CT	0.12	0.41	**0.023 B=0.067**	0.49	0.13
		TT	**0.041 B=0.38**	0.083 B= −0.14	**<0.001 B=0.11**	0.068 B=−0.15	**0.024 B= −0.22**
rs2819561	AA	AG	0.12	0.41	**0.023 B=0.067**	0.47	0.12
		GG	**0.036 B=0.39**	0.090 B= −0.14	**<0.001 B=0.11**	0.069 B=−0.15	**0.027 B= −0.21**
rs2633852	AA	AG	0.12	0.40	**0.022 B=0.068**	0.41	0.11
		GG	**0.044 B=0.38**	0.099 B= −0.14	**0.001 B=0.11**	0.062 B= −0.15	**0.025 B= −0.22**
rs11915920	CC	CT	0.37	0.19	0.12	0.56	0.10 B= −0.14
		TT	0.11	**0.010 B= −0.21**	**<0.001 B=0.11**	0.059 B= −0.15	**0.012 B= −0.24**
rs807785	CC	CT	0.91	0.38	0.75	0.26	0.062 B= −0.24
		TT	0.72	0.87	0.67	0.18	0.23
rs3864051	CC	CT	0.12	0.057 B=0.11	0.70	0.56	0.063 B=0.13
		TT	**0.029 B= −1.01**	0.49	0.080 B= −0.14	0.16	0.29
rs794187	CC	CT	0.15	**0.030 B=0.13**	0.81	0.67	0.69
		TT	0.67	0.94	0.64	0.80	0.91
rs794185	TT	TC	0.091 B=0.26	**0.001 B=0.22**	0.86	0.32	0.36
		CC	0.14	0.29	0.33	0.37	0.76
rs2322683	CC	CT	0.96	0.061 B=0.17	0.35	0.66	0.22
		TT	0.82	0.50	0.49	0.91	0.26
rs793391	TT	TG	0.34	0.064 B=0.11	0.86	0.44	0.92
		GG	0.70	0.64	0.91	0.51	0.50
rs12634249	CC	CA	0.84	0.58	0.92	**0.022 B= −0.14**	0.29
		AA	0.18	0.73	0.25	0.27	0.81
rs2819590	CC	CT	0.32	0.078 B=0.11	0.73	0.28	0.54
		TT	0.45	0.91	0.39	0.16	0.52
rs304092	GG	GA	0.43	0.095 B=0.10	0.89	0.45	0.48
		AA	0.78	0.71	0.52	0.28	0.60

Data are presented as p-values from linear regression between the reference allele (ref)/ref and ref/variance allele (var) or var/var as indicated, and adjusted for age, sex, height and smoking status. B-value is the unstandardised β-value, and is shown when p<0.1. Significant associations are depicted in bold. D_{LCO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; RV: residual volume; TLC: total lung capacity.

Most of the SNPs are in intron positions (table 2), and only one SNP (rs2819590 in exon 1) is defined as giving rise to missense transcript and potentially affecting the SUMF1 protein directly. In addition, one SNP (2633852) is localised in exon 9 and suggested to be a synonymous, silent mutation in a coding

TABLE 5 Associations between different genotypes of sulfatase modifying factor (*SUMF1*) single nucleotide polymorphisms and airway resistance and reactance measured by impulse oscillometry in the total study population including both COPD and control subjects

	ref/ref	ref/var var/var	R_5	R_{20}	R_5–R_{20}	X_5	AX
rs4685744	CC	CT	0.56	0.35	0.82	0.83	0.72
		TT	0.32	0.46	0.44	0.64	0.95
rs2819562	CC	CT	0.32	0.41	0.28	0.52	0.21
		TT	0.69	0.82	0.80	0.95	0.61
rs2819561	AA	AG	0.31	0.41	0.27	0.50	0.20
		GG	0.67	0.85	0.74	0.94	0.63
rs2633852	AA	AG	0.27	0.33	0.26	0.48	0.19
		GG	0.83	0.99	0.86	0.97	0.55
rs11915920	CC	CT	0.61	0.36	0.89	0.70	0.85
		TT	0.33	0.52	0.39	0.65	0.92
rs807785	CC	CT	0.45	0.85	0.14	0.13	0.22
		TT	0.49	0.76	0.15	0.095 B=0.022	0.26
rs3864051	CC	CT	0.29	0.45	0.39	0.21	0.36
		TT	**0.032** **B=0.079**	0.43	**0.003** **B=0.062**	**0.005** **B= −0.067**	**0.009** **B=0.61**
rs794187	CC	CT	0.086 B= −0.020	0.72	**0.013** **B= −0.017**	**0.018** **B=0.018**	**0.015** **B= −0.18**
		TT	0.79	0.93	0.56	0.91	0.79
rs794185	TT	TC	**0.002** **B= −0.039**	0.14	**<0.001** **B= −0.026**	<0.001 B=0.030	**<0.001** **B= −0.30**
		CC	**0.030** **B= −0.035**	0.39	**0.004** **B= −0.026**	0.031 B=0.023	**0.020** **B= −0.24**
rs2322683	CC	CT	0.21	0.33	0.27	0.11	0.11
		TT	0.62	0.89	0.48	0.89	0.81
rs793391	TT	TG	**0.014** **B= −0.029**	0.16	**0.007** **B= −0.018**	**0.012** **B=0.019**	**0.008** **B= −0.20**
		GG	0.60	0.71	0.61	0.96	1.00
rs12634249	CC	CA	0.13	0.68	**0.029** **B= −0.014**	**0.024** **B=0.017**	**0.028** **B= −0.16**
		AA	0.39	0.87	0.30	0.59	0.48
rs2819590	CC	CT	**0.010** **B= −0.030**	**0.048** **B= −0.014**	**0.021** **B= −0.015**	**0.025** **B=0.17**	**0.011** **B= −0.19**
		TT	0.85	0.92	0.84	0.65	0.71
rs304092	GG	GA	**0.010** **B= −0.030**	0.053 B= −0.014	**0.018** **B= −0.016**	**0.029** **B=0.017**	0.022 B= −0.17
		AA	0.81	0.78	0.92	0.69	0.66

Data are presented as p-values from linear regression between the reference allele (ref)/ref and ref/variance allele (var) or var/var, adjusted for age, sex, height and smoking status. B-value is the unstandardised β-value and is shown when p<0.1. Significant associations are depicted in bold. R_5: resistance at 5 Hz; R_{20}: resistance at 20 Hz; X_5: reactance at 5 Hz; AX: reactance area.

sequence (cds-syn). The role of the two aforementioned SNPs is not known, but they could be in undefined regulatory zones or act to regulate other genes. The associations between lung function and the SNPs in block 2 could also be explained by the fact that they have a neighbouring localisation on the chromosome and are in a nonrandom association of alleles (as shown in the linkage disequilibrium plot in supplementary figure S1B) and thereby a in block with rs2819590.

We have shown previously that the SNP rs793391 was associated with COPD in a selected study population of healthy current smokers and ex-smokers [12]. These findings were confirmed in the present study, but in a broader population including also many subjects who were never-smokers. In a subanalysis, we found that if the associations between the SNP and COPD were not adjusted for smoking status, the results were similar (p=0.012 and the same confidence interval of the odds ratio 0.39–0.89). Hereby we showed that rs793391 was not only associated to COPD in smoking/ex-smoking subjects, but also in never-smokers.

The association between certain SNPs and COPD may also be related to disease severity and progression. Certain SNPs located on chromosome 6 or the *HLA-DQB2* gene have been shown to be associated with susceptibility to early (mild/moderate) COPD [31], but only in comparison to never-smoking controls. The population investigated in the present study consists of mainly mild/moderate COPD, and our findings are therefore most representative for this group of patients.

Differences in lung function variables between the SNP genotypes could be read as clinically significant, which can easily be spotted when presenting the B-value, which, for example, for FEV_1 was >120 mL for most SNPs (table 3). This is within the recommended range of minimally clinical important difference [32] and suggest *SUMF1* genotypes to be a potential biomarker.

Over the past decade, the accessibility of whole-genome sequencing and the increased number of genome-wide association studies have led to a substantial increase in the understanding of genetic variants that play a role in COPD susceptibility and COPD-related phenotypes [2]. Several variants are shown to be associated to COPD, emphysema and/or spirometric values, and specifically some genetic loci have been identified as being associated to D_{LCO} [33]. In the present study, two distinct blocks of SNPs were identified, and one block (block 2) was associated to an obstructive phenotype, while the other block (block 1) was associated with a phenotype with reduced lung volumes. We have shown previously that the SNPs in the latter block were highly significant in eQTL analysis of *SUMF1* and were associated with lower levels of *SUMF1* mRNA in sputum cells and lung fibroblasts. Therefore, we believe that specific SNPs give rise to certain pathological effects, which may contribute to the heterogeneous nature of COPD.

In conclusion, we have found several SNPs in the *SUMF1* gene that are associated with COPD and with impaired lung function. These SNPs appear in two blocks: block 2 is associated to COPD and airway obstruction, and block 1 is associated to impaired lung volumes. We hereby confirm in a larger well-defined cohort, that which we have shown previously in a smaller cohort; thus, we are again proving that these SNPs may have a legitimate role in COPD pathogenesis. Additionally, we see again that the two linkage disequilibrium blocks appear to have different effects on aspects of the disease pathogenesis that we do not yet fully understand. Furthermore, a single SNP (rs3864051) is unconnected and strongly associated to COPD and impaired lung function. These genetic variants of *SUMF1* may cause a deficient sulfation balance in the extracellular matrix of the lung tissue, thereby contributing to the COPD disease.

Acknowledgement: The authors would like to thank personnel at the Lung and Allergy Research Unit, Skåne University Hospital (Lund, Sweden) and Clinical Physiology, Skåne University Hospital (Malmö, Sweden) for clinical assistance.

Provenance: Submitted article, peer reviewed.

Conflict of interest: P. Wollmer reports receiving support for the present manuscript from the Swedish Heart and Lung Foundation; payment or honoraria for lectures, presentations, speaker bureaus, manuscript writing or educational events from Chiesi Pharma, outside the submitted work: and a patent issued for a device and method for pulmonary function measurement (no licence). The remaining authors have nothing to disclose.

Support statement: This work was supported by independent grants from the Swedish Heart and Lung Foundation and Region Skåne ALF grants.

References

1 Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet* 2015; 385: 899–909.
2 Ragland MF, Benway CJ, Lutz SM, *et al.* Genetic advances in chronic obstructive pulmonary disease. Insights from COPDGene. *Am J Respir Crit Care Med* 2019; 200: 677–690.
3 Barnes PJ. Mediators of chronic obstructive pulmonary disease. *Pharmacol Rev* 2004; 56: 515–548.
4 Kirkham PA, Barnes PJ. Oxidative stress in COPD. *Chest* 2013; 144: 266–273.
5 Fischer BM, Pavlisko E, Voynow JA. Pathogenic triad in COPD: oxidative stress, protease-antiprotease imbalance, and inflammation. *Int J Chron Obstruct Pulmon Dis* 2011; 6: 413–421.
6 Chung KF, Adcock IM. Multifaceted mechanisms in COPD: inflammation, immunity, and tissue repair and destruction. *Eur Respir J* 2008; 31: 1334–1356.
7 Thun GA, Imboden M, Ferrarotti I, *et al.* Causal and synthetic associations of variants in the *SERPINA* gene cluster with alpha1-antitrypsin serum levels. *PLoS Genet* 2013; 9: e1003585.

8 Reséndiz-Hernández JM, Falfán-Valencia R. Genetic polymorphisms and their involvement in the regulation of the inflammatory response in asthma and COPD. *Adv Clin Exp Med* 2018; 27: 125–133.
9 Du Y, Zhang H, Xu Y, *et al.* Association among genetic polymorphisms of GSTP1, HO-1, and SOD-3 and chronic obstructive pulmonary disease susceptibility. *Int J Chron Obstruct Pulmon Dis* 2019; 14: 2081–2088.
10 Tacheva T, Dimov D, Anastasov A, *et al.* Association of the MMP7-181A>G promoter polymorphism with early onset of chronic obstructive pulmonary disease. *Balkan J Med Genet* 2017; 20: 59–66.
11 Haq I, Chappell S, Johnson SR, *et al.* Association of MMP-2 polymorphisms with severe and very severe COPD: a case control study of MMPs-1, 9 and 12 in a European population. *BMC Med Genet* 2010; 11: 7.
12 Weidner J, Jarenback L, de Jong K, *et al.* Sulfatase modifying factor 1 (SUMF1) is associated with chronic obstructive pulmonary disease. *Respir Res* 2017; 18: 77.
13 Zito E, Fraldi A, Pepe S, *et al.* Sulphatase activities are regulated by the interaction of sulphatase-modifying factor 1 with SUMF2. *EMBO Rep* 2005; 6: 655–660.
14 Cosma MP, Pepe S, Annunziata I, *et al.* The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* 2003; 113: 445–456.
15 Dierks T, Schmidt B, Borissenko LV, *et al.* Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. *Cell* 2003; 113: 435–444.
16 Dierks T, Dickmanns A, Preusser-Kunze A, *et al.* Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme. *Cell* 2005; 121: 541–552.
17 Garavelli L, Santoro L, Iori A, *et al.* Multiple sulfatase deficiency with neonatal manifestation. *Ital J Pediatr* 2014; 40: 86.
18 Schlotawa L, Steinfeld R, von Figura K, *et al.* Molecular analysis of SUMF1 mutations: stability and residual activity of mutant formylglycine-generating enzyme determine disease severity in multiple sulfatase deficiency. *Hum Mutat* 2008; 29: 205.
19 Arteaga-Solis E, Settembre C, Ballabio A, *et al.* Sulfatases are determinants of alveolar formation. *Matrix Biol* 2012; 31: 253–260.
20 Lee JH, McDonald ML, Cho MH, *et al.* DNAH5 is associated with total lung capacity in chronic obstructive pulmonary disease. *Respir Res* 2014; 15: 97.
21 Jarenbäck L, Ankerst J, Bjermer L, *et al.* Flow-volume parameters in COPD related to extended measurements of lung volume, diffusion, and resistance. *Pulm Med* 2013; 2013: 782052.
22 Frantz S, Nihlén U, Dencker M, *et al.* Impulse oscillometry may be of value in detecting early manifestations of COPD. *Respir Med* 2012; 106: 1116–1123.
23 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 16: Suppl., 5–40.
24 Hall GL, Filipow N, Ruppel G, *et al.* Official ERS technical standard: Global Lung Function Initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J* 2021; 57: 2000289.
25 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.
26 Stanojevic S, Graham BL, Cooper BG, *et al.* Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur Respir J* 2017; 50: 1700010.
27 Jurinke C, van den Boom D, Cantor CR, *et al.* Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* 2002; 187: 179–192.
28 Fischer A, Swigris JJ, Groshong SD, *et al.* Clinically significant interstitial lung disease in limited scleroderma: histopathology, clinical features, and survival. *Chest* 2008; 134: 601–605.
29 Weidner J, Jogdand P, Jarenbäck L, *et al.* Expression, activity and localization of lysosomal sulfatases in chronic obstructive pulmonary disease. *Sci Rep* 2019; 9: 1991.
30 Malovini A, Illario M, Iaccarino G, *et al.* Association study on long-living individuals from Southern Italy identifies rs10491334 in the CAMKIV gene that regulates survival proteins. *Rejuvenation Res* 2011; 14: 283–291.
31 Lee YJ, Choi S, Kwon SY, *et al.* A genome-wide association study in early COPD: identification of one major susceptibility loci. *Int J Chron Obstruct Pulmon Dis* 2020; 15: 2967–2975.
32 Cazzola M, MacNee W, Martinez FJ, *et al.* Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008; 31: 416–469.
33 Sakornsakolpat P, McCormack M, Bakke P, *et al.* Genome-wide association analysis of single-breath DL_{CO}. *Am J Respir Cell Mol Biol* 2019; 60: 523–531.