Heritability and genome-wide association study of diffusing capacity of the lung

Natalie Terzikhan[1,2,8], Fangui Sun[3,8], Fien M. Verhamme[1,8], Hieab H.H. Adams[2,4], Daan Loth[2], Ken R. Bracke[1], Bruno H.C. Stricker[2], Lies Lahousse[2,5,9], Josée Dupuis[3,9], Guy G. Brusselle[1,2,6,9] and George T. O'Connor[7,9]

Affiliations: [1]Dept of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [2]Dept of Epidemiology, Erasmus MC – University Medical Center Rotterdam, Rotterdam, The Netherlands. [3]Dept of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [4]Dept of Radiology and Nuclear Medicine, Erasmus MC – University Medical Center Rotterdam, Rotterdam, the Netherlands. [5]Dept of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium. [6]Dept of Respiratory Medicine, Erasmus MC – University Medical Centre Rotterdam, Rotterdam, The Netherlands. [7]Pulmonary Center, Boston University Schools of Medicine and Public Health, Boston, MA, USA. [8]These authors contributed equally. [9]These authors contributed equally.

Correspondence: Bruno H.C. Stricker, Dept of Epidemiology, Erasmus MC – University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands. E-mail: b.stricker@erasmusmc.nl

@ERSpublications
This is the first population-based heritability and GWAS study of gas exchange. We identified a functional variant in ADGRG6, and demonstrated differential expression in lung tissue of patients with COPD and decreased diffusing capacity. http://ow.ly/Rvy430kHIT4

Cite this article as: Terzikhan N, Sun F, Verhamme FM, *et al*. Heritability and genome-wide association study of diffusing capacity of the lung. *Eur Respir J* 2018; 52: 1800647 [https://doi.org/10.1183/13993003.00647-2018].

ABSTRACT Although several genome-wide association studies (GWAS) have investigated the genetics of pulmonary ventilatory function, little is known about the genetic factors that influence gas exchange. The aim of the study was to investigate the heritability of, and genetic variants associated with the diffusing capacity of the lung.

GWAS was performed on diffusing capacity of the lung measured by carbon monoxide uptake (D_{LCO}) and per alveolar volume (V_A) using the single-breath technique, in 8372 individuals from two population-based cohort studies, the Rotterdam Study and the Framingham Heart Study. Heritability was estimated in related (n=6246) and unrelated (n=3286) individuals.

Heritability of D_{LCO} and D_{LCO}/V_A ranged between 23% and 28% in unrelated individuals and between 45% and 49% in related individuals. Meta-analysis identified a genetic variant in *ADGRG6* that is significantly associated with D_{LCO}/V_A. Gene expression analysis of *ADGRG6* in human lung tissue revealed a decreased expression in patients with chronic obstructive pulmonary disease (COPD) and subjects with decreased D_{LCO}/V_A.

D_{LCO} and D_{LCO}/V_A are heritable traits, with a considerable proportion of variance explained by genetics. A functional variant in *ADGRG6* gene region was significantly associated with D_{LCO}/V_A. Pulmonary *ADGRG6* expression was decreased in patients with COPD.

This article has supplementary material available from erj.ersjournals.com

Received: April 04 2018 | Accepted after revision: June 22 2018

Introduction

The respiratory system can be separated functionally into two zones. The first is the conducting zone, which includes the trachea, bronchi, bronchioles and terminal bronchioles and which is functional in ventilation, *i.e.* conducting the air in and out of the lungs. The second zone is the respiratory zone, which consists of the respiratory bronchioles, alveolar ducts and alveoli, the site where oxygen and carbon dioxide are exchanged between the lungs and the blood.

Different pulmonary function tests are available to measure these distinct functions of ventilation and gas exchange. These tests help to evaluate and manage patients with respiratory symptoms and diseases, and include spirometry, measurements of lung volumes and the diffusing capacity of the lung for carbon monoxide (D_{LCO}). The latter, also known as transfer factor of the lung for carbon monoxide, provides a quantitative measure of gas transfer in the lung [1, 2] and reflects processes in the alveolar compartment and pulmonary microcirculation.

The D_{LCO} provides clinical insights complimentary to those obtained by spirometry and lung volume measurements, for example, in discriminating asthma from chronic obstructive pulmonary disease (COPD), to identify causes of hypoxaemia or dyspnoea, and to monitor patients with interstitial lung disease [3]. D_{LCO} is decreased in patients with emphysema due to a decrease in the total surface area of the lung and the loss of capillary beds [1, 4]. In contrast to the abundance of genome-wide association studies (GWAS) investigating genetic variation of spirometry measures [5–8], the heritability of, and genetic influences on D_{LCO} are largely unknown.

Therefore, we first investigated the heritability of D_{LCO} to understand which proportion of the variance in D_{LCO} can be explained by genetics. Next, we performed a GWAS, to identify genetic variants affecting the variability in D_{LCO}, using data from two prospective population-based cohorts, the Rotterdam Study and the Framingham Heart Study. Finally, we investigated the expression of the lead GWAS association in lung tissue of individuals with COPD and (nonsmoking and smoking) controls.

Methods

The methods are described briefly here; refer to the supplementary material for more detailed information.

Setting

The present meta-analysis combined results from two population-based studies, *i.e.* the Rotterdam Study and the Framingham Heart Study. In both cohorts, only individuals of European ancestry were included in the analyses. The Rotterdam Study [9] is an ongoing prospective population-based cohort study that includes three cohorts encompassing 14 926 participants aged ⩾45 years, living in the Netherlands. D_{LCO} was measured between 2009 and 2013. The Rotterdam Study has been approved by the medical ethics committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study).

The Framingham Heart Study is a population-based family study that recruited residents of Framingham (MA, USA), starting in 1948. D_{LCO} was measured at the eighth and ninth examinations of the offspring cohort (2005–2008 and 2011–2014) and the first and second examinations of the third-generation cohort (2002–2005 and 2008–2011). For participants with measurements at both time points, we analysed the later measurement. The Framingham Heart Study has been approved by the institutional review board of the Boston University Medical Campus. All participants provided written informed consent to participate in the study and to obtain information from their treating physician.

Lung function

D_{LCO} ($mmol \cdot min^{-1} \cdot kPa^{-1}$) and alveolar volume ($V_A$) were measured using the single-breath technique in accordance with European Respiratory Society/American Thoracic Society guidelines [2]. The D_{LCO} per alveolar volume (D_{LCO}/V_A; $mmol \cdot min^{-1} \cdot kPa^{-1} \cdot L^{-1}$) was calculated by dividing the D_{LCO} by V_A. Analyses were restricted to participants with two interpretable and reproducible measurements of D_{LCO} and D_{LCO}/V_A.

Heritability analysis

Heritability was defined as the ratio of trait variance due to additive genetic effects to the total phenotypic variance after accounting for covariates. In the Rotterdam Study, GCTA software [10] was used to estimate heritability in unrelated individuals. In the Framingham Heart Study, SOLAR (Sequential Oligogenic Linkage Analysis Routines) software [11] was used to estimate heritability based on familial relationships. Analyses were adjusted for age, sex and principal components of genetic relatedness (in GCTA only). Additional adjustments for current and former smoking were made in a subsequent analysis.

GWAS analyses

A GWAS was performed for both phenotypes D_{LCO} and D_{LCO}/V_A using ProbABEL (version 0.4.4). Variants with imputation quality (R^2) <0.3 and minor allele frequency (MAF) <0.01 were excluded from the analyses. Linear regression was conducted for each single-nucleotide polymorphism (SNP), assuming an additive model. All analyses were adjusted for age, sex and principal components (Rotterdam Study only) in model 1 and additionally adjusted for smoking, weight and height in model 2. A random effect was added to the model to account for familial relationship in the Framingham Heart Study analyses. Data were meta-analysed using METAL software (www.sph.umich.edu/csg/abecasis/metal/) and were adjusted for genomic control. Genome-wide significance threshold was set at p-value $<5\times10^{-8}$ and for suggestive associations at p-value 5×10^{-7}. Quantile–quantile plots, Manhattan plots and regional plots were generated using the R software. Analyses were repeated after 1) correction for haemoglobin in the Rotterdam Study; and 2) additional adjustment for forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) and 3) additional adjustment for quantitative emphysema (less than −950 LAA on computed tomography scan of the lungs), as measured in the Framingham Heart Study within 8 years of the lung function measurement.

Follow-up analyses

Several steps were taken in order to explore the functionality of the variants and genes of interest, and to associate those newly identified loci to clinically relevant disease outcomes, as follows. 1) Genetic correlations were investigated; 2) genetic overlap was investigated with SNPs that are significantly related to COPD [12] and emphysema [13]; 3) posterior probability of causality of the lead SNP was calculated using FINEMAP software [14]; 4) the regulatory function of the lead SNP was explored on the Haploreg server (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php); 5) the effect of the lead SNP on mRNA expression was checked (expression quantitative trait loci (eQTL) analysis), using the lung tissue dataset from the genotype tissue expression (GTEx) portal (GTEx portal: www.gtexportal.org/home/; GTEx portal eQTL data, lung tissue set obtained from location javascript:portalClient.browseDatasets.downloadFile('Lung.allpairs.txt.gz','gtex_analysis_v7/single_tissue_eqtl_data/all_snp_gene_associations/Lung.allpairs.txt.gz')); 6) tissue-specific gene expression was checked in the GTEx portal; and 7) mRNA expression of the *ADGRG6* gene was analysed in lung tissue (using real-time PCR) of 92 patients with or without COPD.

Results

Study cohorts and participant characteristics

The general characteristics of the study populations (the Rotterdam Study and the Framingham Heart Study) are shown in table 1. The mean±SD age was 67.3±8.0 years in the Rotterdam Study and 52.8±14.8 years in the Framingham Heart Study. Figure 1 shows the study flow of participants.

TABLE 1 General characteristics of the study populations

	Rotterdam Study	Framingham Heart Study
Subjects n	2574	5798
Age years	67.3±8.0	52.7±14.8
Female	51.9	53.9
Weight kg	80.5±14.9	79.7±18.5
Height cm	170.6±9.2	168.9±9.5
Former smokers	55.4	39.9
Current smokers	11.5	10.9
Never-smokers	33.1	49.2
D_{LCO} mmol·min^{-1}·kPa^{-1}	8.0±1.8	8.3±2.3
D_{LCO} corrected for Hb mmol·min^{-1}·kPa^{-1}	7.9±1.7	NA
D_{LCO}/V_A mmol·min^{-1}·kPa^{-1}· V_A^{-1}	1.5±0.2	1.5±0.2
D_{LCO}/V_A corrected for Hb mmol·min^{-1}·kPa^{-1}· V_A^{-1}	1.5±0.2	NA
FEV_1 L	2.8±0.7	3.1±0.9
FEV_1 % pred	105.3±19.8	99.2±14.9
FVC L	3.7±1.0	4.2±1.1
FVC % pred	110.1±17.6	102.7±13.5
FEV_1/FVC	76.4±7.1	75.4±6.9

Data are presented as mean±SD or %, unless otherwise stated. D_{LCO}: diffusing capacity of the lung for carbon monoxide; Hb: haemoglobin; V_A: alveolar volume; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; NA: not applicable.

Heritability

Heritability was estimated in two ways; firstly using the Rotterdam Study data with unrelated individuals, with a total number of 3286 participants with genetic data and interpretable measurements of *D*LCO; and secondly by using data from the Framingham Heart Study to estimate heritability based on familial relationships in 6246 participants with interpretable measurements of *D*LCO (figure 1). In table 2 heritability estimates for *D*LCO and *D*LCO/*V*A are presented. In unrelated individuals, we found an age- and sex- and principal components-adjusted heritability for *D*LCO of 23%, and a heritability of 28% after additional adjustment for current and past smoking. Similar heritability estimates were found for *D*LCO/*V*A with 24% after adjustment for age, sex and principal components, and 25% after additional adjustment for smoking. In the Framingham Heart Study, investigating individuals with known familial relationships, we found an age- and sex-adjusted heritability for *D*LCO of 49%, and a heritability of 47% after additional adjustment for current and past smoking. Heritability estimates for *D*LCO/*V*A were 45% after adjustment for age and sex, and 46% after additional adjustment for current and past smoking.

Genetic variants associated with diffusing capacity

We performed GWAS on *D*LCO and *D*LCO/*V*A in the Rotterdam Study (n=2574) and the Framingham Heart Study (n=5798), and subsequently meta-analysed both cohorts (n=8372). All variants with a p-value $<5\times10^{-6}$ at the meta-analysis stage are presented in table 3. The corresponding quantile–quantile plots are presented in supplementary figure E1. GWAS results of the separate cohorts with (p-value$<5\times10^{-6}$), are presented in supplementary tables E1 and E2.

Analyses were adjusted for age, sex and principal components in model 1. In model 2 analyses were adjusted for variables in model 1, in addition to weight, height, current and past smoking.

Figure 2 represents the Manhattan plots of *D*LCO GWAS at the meta-analysis level. For both *D*LCO analyses (models 1 and 2), no variant reached genome-wide significance threshold. In model 2, two variants at 10q22.1 (rs1665630, gene *CDH23*, MAF 0.44; p-value=2.8×10^{-7}) and at 20p12.3 (rs2423124, close to gene *GPCPD1*, MAF 0.19; p-value=4.2×10^{-7}) showed a suggestive association with *D*LCO.

Figure 3 represents the Manhattan plots of *D*LCO/*V*A GWAS at the meta-analysis level. 19 variants at the same locus at 6q24.1 (top: rs17280293, gene *ADGRG6*, MAF 0.03; p-value=1.4×10^{-10}) were significantly

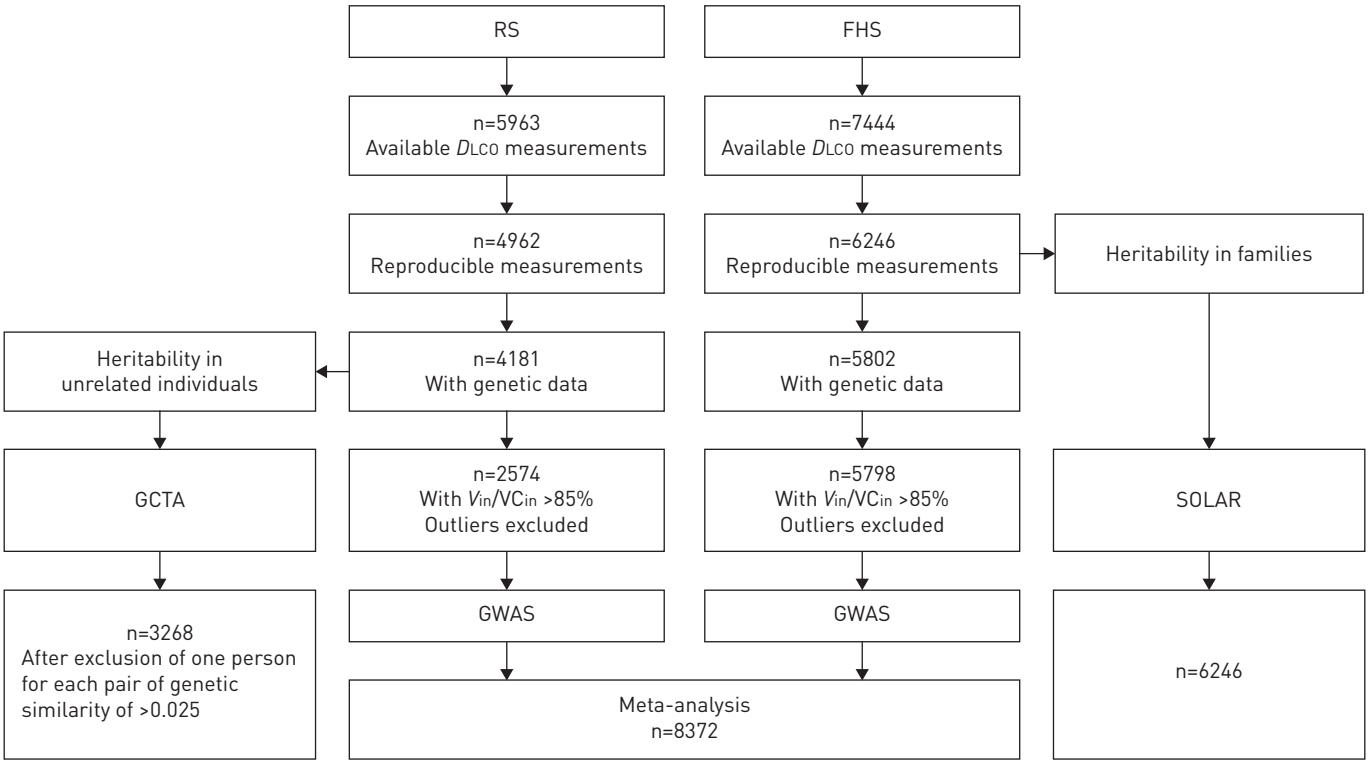

FIGURE 1 Flowchart of study participants. RS: Rotterdam Study; FHS: Framingham Heart Study; *D*LCO: diffusing capacity of the lung for carbon monoxide; *V*in: inspiratory volume; VCin: vital capacity measured during maximal inspiration; GWAS: genome-wide association study; GCTA: Genome-wide Complex Trait Analysis software; SOLAR: Sequential Oligogenic Linkage Analysis Routines package.

TABLE 2 Heritability of diffusing capacity of the lung

	Rotterdam Study				Framingham Heart Study			
	D_{LCO}		D_{LCO}/V_A		D_{LCO}		D_{LCO}/V_A	
	h^2±SE	p-value	h^2±SE	p-value	h^2±SE	p-value	h^2±SE	p-value
Subjects n	3286[+]				6246[§]			
Model 1[#]	0.23±0.10	0.01	0.24±0.10	0.009	0.49±0.03	2.3×10^{-106}	0.45±±0.03	5.0×10^{-82}
Model 2[¶]	0.28±0.10	0.002	0.25±0.10	0.0075	0.47±0.03	8.5×10^{-100}	0.46±0.03	7.6×10^{-84}

D_{LCO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; h^2: heritability estimate. [#]: adjusted for age, sex and principal components of genetic relatedness (Rotterdam Study only); [¶]: adjusted for age, sex, smoking and principal components of genetic relatedness (Rotterdam Study only); [+]: unrelated individuals; [§]: individuals with known family relationships.

associated with D_{LCO}/V_A in model 1 (see regional plot in figure 4). Of these, six variants at the same locus at 6q24.1 reached the genome-wide significance threshold in model 2. Sensitivity analysis by adjusting for FEV1/FVC did not explain the effect of the association between rs17280293 and D_{LCO}/V_A (β±SE= −0.07±0.01, p-value=1.51×10^{-10} after adjustment for FEV1/FVC *versus* β±SE=−0.07±0.01, p-value=7.9×10^{-11} before adjustment for FEV1/FVC in model 2) (supplementary figure E2). Similarly, adjusting for quantitative emphysema (<−950 LAA on computed tomography scan of the lungs) in a subset of the Framingham Heart Study (n=2176) did not alter the association between rs17280293 and D_{LCO}/V_A (β±SE= −0.06±0.02, p-value=0.003 after adjustment for emphysema *versus* β±SE=−0.06±0.02, p-value=0.002 before adjustment for emphysema). Moreover, in both models, a variant at 5q12.1 (rs918606, gene *IPO11*, MAF 0.44; p-value model 1=5.96×10^{-8}, p-value-model 2=7.49×10^{-8}) was found to be suggestively associated with D_{LCO}/V_A. Additional sensitivity analysis by adjusting for haemoglobin blood concentrations did not materially change the results of the D_{LCO}/V_A GWAS (supplementary material).

Interestingly, a more in-depth investigation of the *ADGRG6* region (figure 4) revealed the presence of two missense variants: the lead SNP rs17280293 and rs11155242 (MAF 0.19, p-value=2.1×10^{-06}). Those two SNPs showed to be in linkage disequilibrium with each other, with r^2=0.14 and D′=1.

Follow-up analyses

The most important findings of the follow-up analyses are summarised here, including genetic correlations and gene expression in lung tissue. Additional results on the genetic correlations, overlap with reported COPD and emphysema GWAS associations, posterior probability of causality, functional annotation and gene expression are presented in the supplementary material.

TABLE 3 Independent genetic variants that are significantly or suggestively associated with diffusing capacity of the lung for carbon monoxide (D_{LCO}) or D_{LCO} per alveolar volume (V_A) at meta-analysis level

	SNP	Chr:Pos	Gene[#]	A1/A2	RS		FHS		RS and FHS	
					β	p-value	β	p-value	β	p-value
Subjects n					2574		5798		8372	
D_{LCO}[¶]										
D_{LCO}[+]	rs1665630	10:73426862	CDH23	T/C	0.11	6.4×10^{-4}	0.10	9.1×10^{-5}	0.11	2.8×10^{-7}
	rs2423124	20:5636945	GPCPD1	T/C	−0.20	1.4×10^{-6}	−0.10	2.5×10^{-2}	−0.16	4.2×10^{-7}
D_{LCO}/V_A[¶]	rs17280293	6:142688969	ADGRG6	A/G	−0.06	3.0×10^{-3}	−0.08	**6.7×10^{-9}**	−0.07	**1.4×10^{-10}**
	rs918606	5:61926379	IPO11	A/G	−0.02	2.0×10^{-3}	−0.02	5.8×10^{-6}	−0.02	$6.0.x10^{-8}$
	rs75834976	4:5231710	STK32B	A/C	−0.04	2.4×10^{-3}	−0.04	5.5×10^{-5}	−0.04	6.0×10^{-7}
	rs56315120	1:165168869	LMX1A	A/G	−0.02	0.24	−0.06	1.5×10^{-7}	−0.05	7.8×10^{-7}
D_{LCO}/V_A[+]	rs17280293	6:142688969	ADGRG6	A/G	−0.06	4.3×10^{-3}	−0.07	**2.3×10^{-9}**	−0.07	**7.9×10^{-11}**
	rs918606	5:61926379	IPO11	A/G	−0.02	1.2×10^{-3}	−0.02	1.3×10^{-5}	−0.02	7.5×10^{-8}

Bold type indicates statistical significance. SNP: single-nucleotide polymorphism; Chr:Pos: chromosome and position; A1: first allele; A2: second allele; β: the effect estimate which are additive effects for each copy of A1; FHS: Framingham Heart Study; RS: Rotterdam Study (meta-analysis RSI, RSII and RSIII). [#]: the gene name is a label of the region using the closest gene, but does not necessarily pinpoint the responsible gene; [¶]: model 1 (adjusted for age, sex and principal components); [+]: model 2 (adjusted for age, sex, weight, height, smoking and principal components).

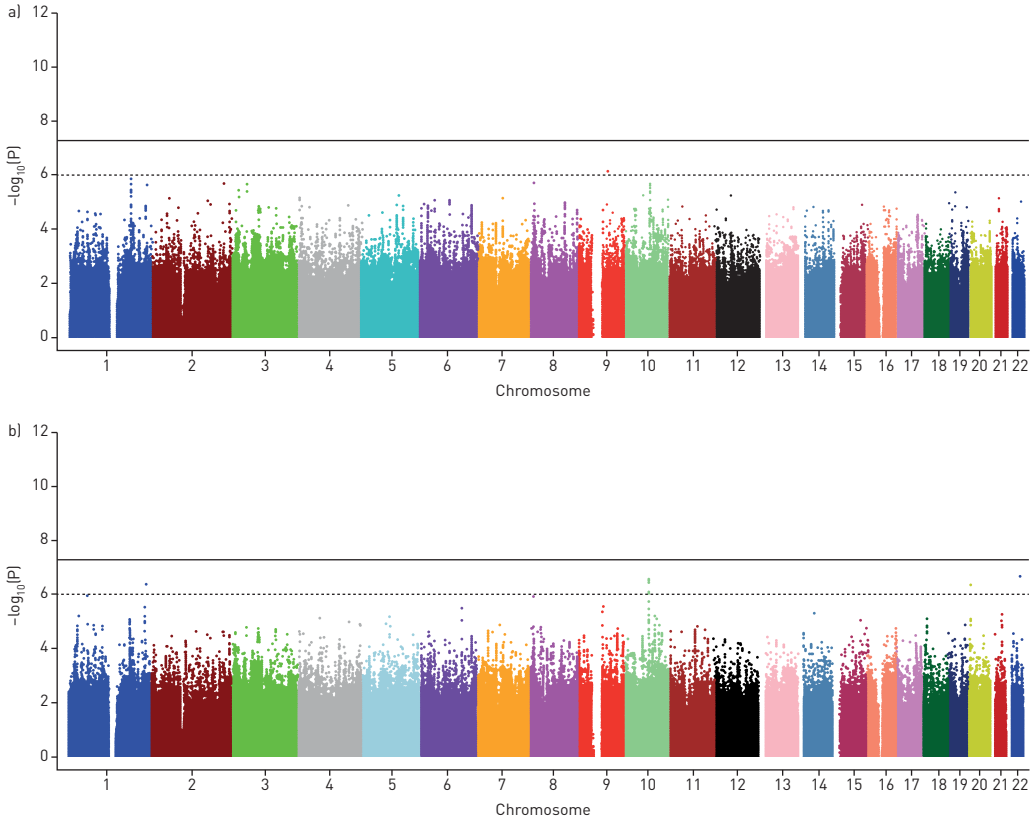

FIGURE 2 Common genetic variants associated with diffusing capacity of the lung for carbon monoxide (*D*LCO). a) Manhattan plot of the association between common genetic variants and *D*LCO, adjusted for age, sex and principal components of genetic relatedness; b) Manhattan plot of the association between common genetics variants and *D*LCO, adjusted for age, sex, weight, height, smoking and principal components of genetic relatedness.

Genetic correlations

We examined the genetic correlation between *D*LCO/*V*A and *D*LCO using the age, sex, smoking status, weight, height and principal components-adjusted model. The genetic correlation was 59% (ρgenetic=0.59, p-value=0.04). This was in line with the phenotypic correlation between *D*LCO and *D*LCO/*V*A (r^2=0.46 in the Rotterdam Study and r^2=0.57 in the Framingham Heart Study, p-value<0.01). In addition, we examined the genetic correlation with FEV1/FVC and height (supplementary material).

ADGRG6 *expression*

We extracted mRNA from lung resection specimens of 92 patients who underwent surgery for solitary pulmonary tumours or lung transplantation, including 44 patients without COPD and 48 patients with COPD (table 4). The mRNA expression of *ADGRG6* was significantly lower in lung tissue of patients with decreased *D*LCO/*V*A compared with patients with normal *D*LCO/*V*A (figure 5a) and in subjects with COPD (encompassing different categories of COPD severity according to the Global Initiative for Chronic Obstructive Lung Disease spirometric classification) compared to never-smoking controls (figure 5b). The *ADGRG6* mRNA levels were significantly associated with *D*LCO/*V*A after adjustment for age and sex in model 1 (n=67, β=0.85 (95% CI 0.06–1.64)) and after additional adjustment for weight, height and smoking in model 2 (n=66, β=0.75 (95% CI 0.03–1.47)).

Discussion

This is the first study that has investigated the heritability of, and genome-wide association with, diffusing capacity of the lung using population-based cohort studies. We found a considerable proportion of variance in diffusing capacity of the lung explained by genetics. We also identified one locus on chromosome 6, encompassing the *ADGRG6* gene, that is associated with *D*LCO/*V*A and its lead variant showed to have a high posterior probability of causality compared to other SNPs in the same region. Finally, we were able to link the pulmonary expression of *ADGRG6* directly to COPD and to low *D*LCO/*V*A (compatible with emphysema in this general population). Here, we demonstrated a differential mRNA expression of *ADGRG6* in lung tissue of COPD patients and patients with decreased *D*LCO/*V*A.

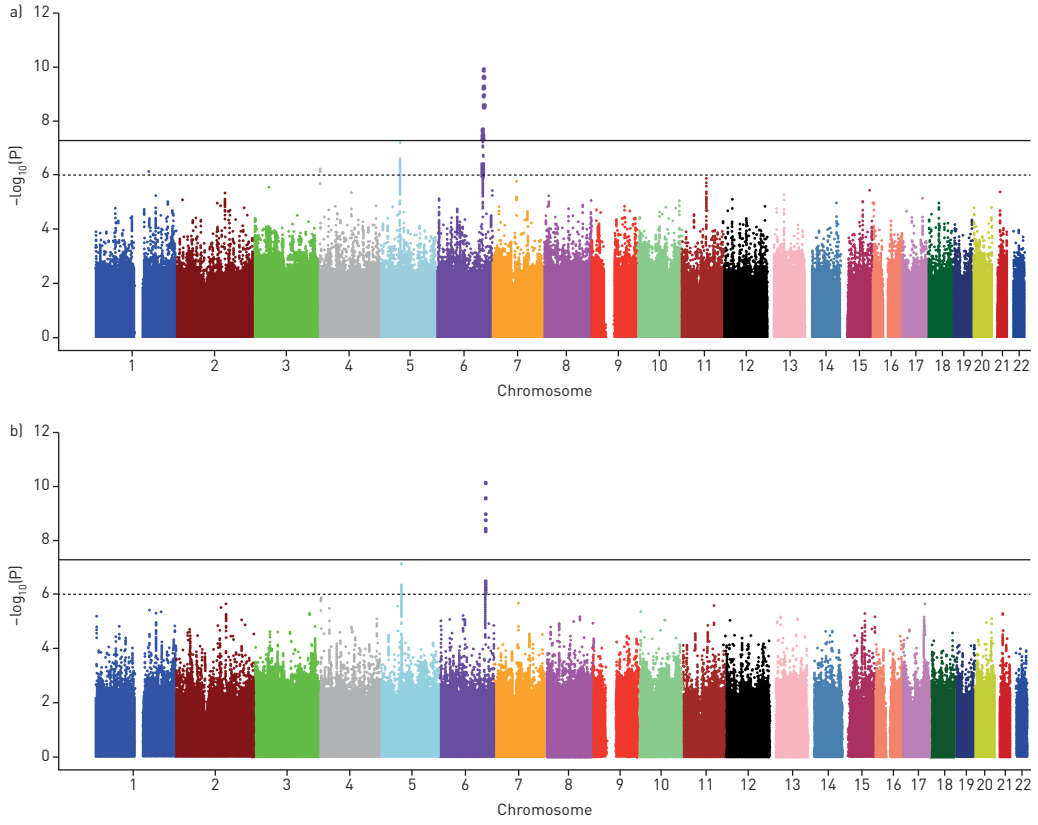

FIGURE 3 Common genetic variants associated with diffusing capacity of the lung for carbon monoxide (*D*LCO) per alveolar volume (*V*A). a) Manhattan plot of the association between common genetic variants and *D*LCO/*V*A, adjusted for age, sex and principal components of genetic relatedness; b) Manhattan plot of the association between common genetics variants and *D*LCO/*V*A, adjusted for age, sex, weight, height, smoking and principal components of genetic relatedness.

Heritability and genetic overlap

Studies on heritability of *D*LCO in the general population and unrelated individuals are lacking, and so far, *D*LCO heritability has been studied only in twins [15, 16], with a highest reported estimate of 44%. In our study, we estimated the restricted maximum likelihood-based heritability of *D*LCO using the GCTA tool in unrelated individuals of the Rotterdam Study [17], and observed an age- and sex-adjusted heritability of *D*LCO and *D*LCO/*V*A of 23% and 24%, respectively. We also investigated heritability based on known familial relationships in the Framingham Heart Study. Here we found an age- and sex-adjusted heritability of *D*LCO and *D*LCO/*V*A of 49% and 45%, respectively. The latter heritability estimates among familial related individuals are in line with the heritability estimates in twin studies and highlight the robustness of our data. Importantly, our study is the first to investigate the lower bound of heritability of *D*LCO estimated by family and twin studies [18]. The advantage of estimating heritability in unrelated individuals using GCTA in addition to the approach based upon family and twin studies is, that GCTA calculates the proportion of heritability that covers the additive effects of commons SNPs only, and does not suffer from bias due to epistatic interactions or shared environment. The latter effects might indeed be present in family and twin studies, leading to an overestimation of the heritability [18–20].

Despite their similar estimates of heritability, *D*LCO and *D*LCO/*V*A appeared to have different genetic determinants due to a genetic overlap between the two traits of 59%, explaining why we could not observe the same lead association in the two analyses.

Variation in ADGRG6

The meta-analysis of genetics variants of *D*LCO/*V*A yielded one genome-wide significant association, along with a number of suggestive associations that did not reach genome-wide significance. The lead variant (rs17280293) in this study is a missense SNP in *ADGRG6*, with a MAF of 0.03 which is comparable to that in public datasets (0.03 ExAC, 0.02 TOPMED and 0.03 in 1000 genomes). Mutation in this SNP causes an amino acid change (S123G), which is predicted to have a deleterious effect as indicated by both SIFT [21] and Polyphen2 [22]. It is therefore likely that this SNP is functional in *ADGRG6*. In this study,

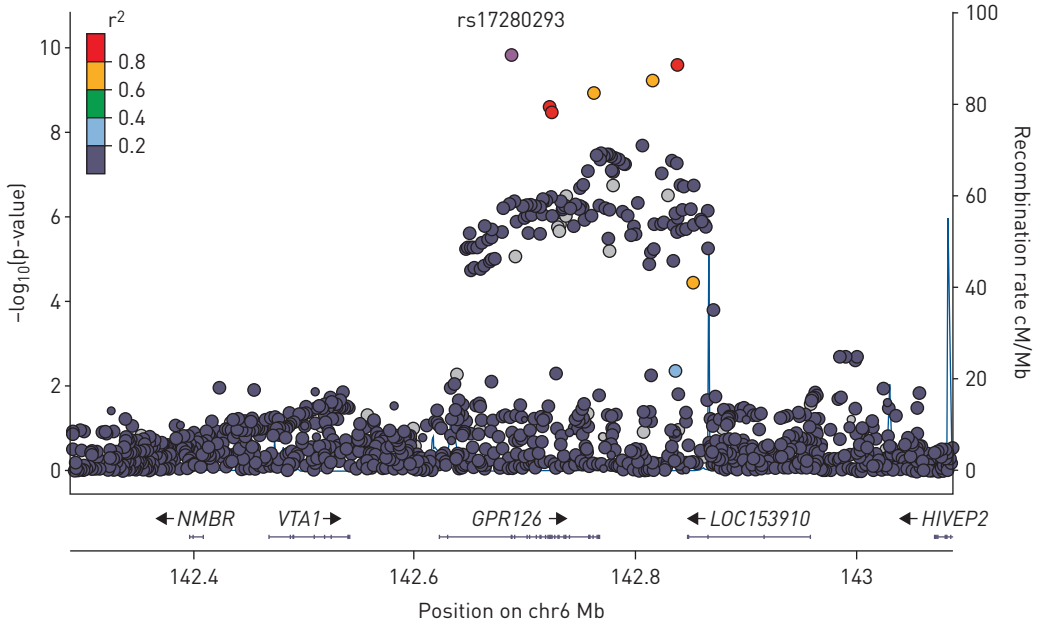

FIGURE 4 Regional association plot of the genome-wide significant locus in a genome-wide association study for diffusing capacity of the lung for carbon monoxide per alveolar volume.

we showed that this variant has a high posterior probability of causality compared to other SNPs in the same region and that this SNP is associated with different regulatory chromatin marks, promotor histone marks and enhancer histone marks in different tissue cell lines including fetal lung fibroblast cell lines and lung carcinoma cell lines. In addition, rs17280293 always co-occurs with another functional SNP in the region (rs11155242, D′=1), which is an eQTL for *ADGRG6* in human lung tissue.

Previous studies have also shown that variation at *ADGRG6* is associated with spirometric measures of lung function [5, 7]. Soler Artigas *et al.* [7] observed a strong association between spirometry, particularly FEV_1/FVC, and another SNP rs148274477, which is in strong linkage disequilibrium with rs17280293. However, since airflow limitation (*i.e.* a low FEV_1/FVC ratio) might be correlated with low diffusing capacity due to loss of elastic recoil in subjects with emphysema, we assessed the possibility that the observed association between rs17280293 and D_{LCO}/V_A might be driven by FEV_1/FVC. However, this sensitivity analysis indicated an independent association between rs17280293 and D_{LCO}/V_A because

TABLE 4 Characteristics of study individuals for lung mRNA analysis (by reverse-transcriptase PCR)

	Never-smokers	Smokers without COPD	COPD GOLD II	COPD GOLD III–IV
Subjects n	18	26	34	14
Male/female	6/12***	19/7***	31/3***	8/6***
Age years	65 (56–70)	63 (55–70)	66 (58–69)*	56 (54–60)*, #,¶
Current-smokers/ex-smokers	NA	16/10	22/12	0/14
Smoking history pack-years	NA	28 (15–45)#	45 (40–60)*,#	30 (25–30)*,#,¶
Post-bronchodilator FEV_1 L	2.7 (2.3–3.2)	2.7 (2.3–3.3)	2.0 (1.8–2.4)*,#	0.7 (0.7–0.9)*,#,¶
Post-bronchodilator FEV_1 % pred	102 (92–116)	95 (93–112)	68 (61–75)*,#	26 (20–32)*,#,¶
Post-bronchodilator FEV_1/FVC %	78 (75–83)	75 (71–79) #	56 (53–60)*,#	32 (27–35)*,#,¶
D_{LCO} % pred	90 (80–105)	80 (61–102)	67 (51–87)#	35 (33–41)*,#,¶
D_{LCO}/V_A % pred	103 (88–123)	91 (68–107)#	87 (62–108)#	59 (50–65)*,#,¶
D_{LCO} mmol·kPa^{-1}·min^{-1}	21.6 (18.1–26.8)	23.3 (17.0–27.4)	17.2 (14.2–25.0)	2.9 (2.8–3.7)###,¶¶¶,+++
D_{LCO}/V_A mmol·kPa^{-1}·min^{-1}· V_A^{-1}	4.6 (3.8–5.3)	3.9 (2.9–4.6)#	3.5 (2.7–4.2)#	0.9 (0.7–0.9)###,¶¶¶,+++

Data are presented as n or median (interquartile range). n=92. COPD: chronic obstructive pulmonary disease; GOLD: Global Initiative for Chronic Obstructive Lung Disease; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; NA: not applicable. Fisher's exact test: ***: p<0.001; Mann–Whitney U-test: *: p<0.05 *versus* smokers without COPD; #: p<0.05 *versus* never smokers; ¶: p<0.05 *versus* COPD GOLD II; ###: p<0.001 *versus* never smokers; ¶¶¶: p<0.001 *versus* COPD GOLD II; +++: p<0.001 *versus* smokers without COPD.

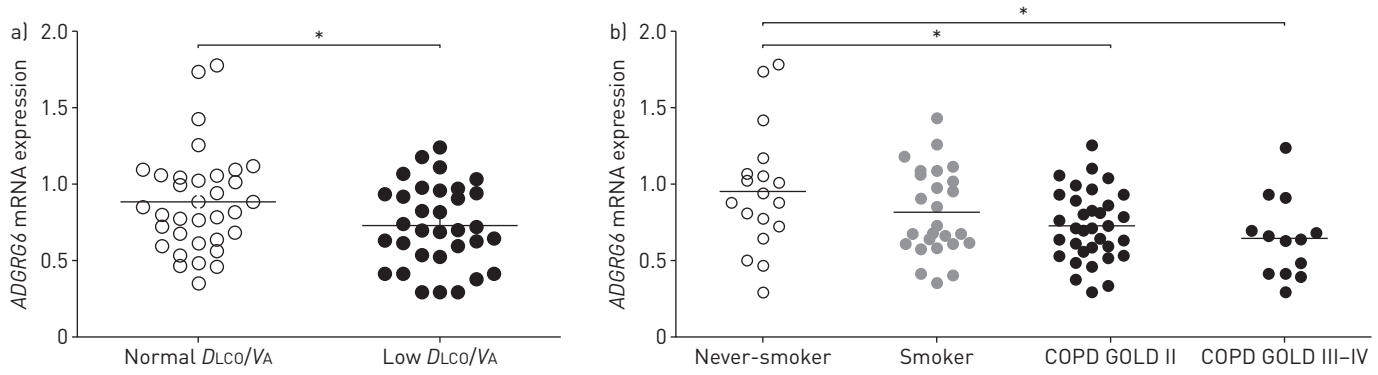

FIGURE 5 Pulmonary mRNA expression of *ADGRG6* in human subjects. a) mRNA levels of *ADGRG6* in lung tissue of individuals with normal diffusing capacity of the lung for carbon monoxide (*D*LCO) per alveolar volume (*V*A) (n=38) and low *D*LCO/*V*A (n=39). mRNA levels were corrected using a calculated normalisation factor based on mRNA expression of three reference genes (*GAPDH*, *SDHA*, *HPRT-1*); b) mRNA levels of *ADGRG6* in lung tissue of never-smokers (n=18), smokers without airflow limitation (n=26), patients with chronic obstructive pulmonary disease (COPD) Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage II (n=34) and patients with COPD GOLD stages III–IV (n=14), as measured using quantitative reverse transcriptase-PCR. For statistical analysis, Kruskal–Wallis testing followed by Mann–Whitney U-test was used for COPD and an independent sample t-test was used for *D*LCO/*V*A after rank transformation. *: p<0.05.

additional adjustment for FEV1/FVC did not affect the estimate and no genetic overlap could be proven between *D*LCO/*V*A and FEV1/FVC. Other studies have associated genetic variation in *ADGRG6* with height. In our study, adjustment for height did not affect the association between *D*LCO/*V*A and rs17280293, suggesting that the lead association in our GWAS is independent of height. In addition, genetic overlap disappeared after additional adjustment for height in the model, indicating no residual confounding by height in our analyses.

Furthermore, EICHSTAEDT *et al.* [23] recently used whole-genome sequence data from 19 Argentinean highlanders compared to 16 native American lowlanders and showed that rs17280293 might contribute to the physiological adaptations to hypobaric hypoxia.

Gene function and expression

The *ADGRG6* gene (adhesion G-protein coupled receptor G6) belongs to the G-protein coupled receptor (GPCR) superfamily, the largest known receptor family in the human genome. It has been previously shown to be essential in angiogenesis [24]. *ADGRG6*, a relatively new adhesion GPCR, has been shown to promote vascular endothelial growth factor (VEGF) signalling, by modulating the expression of endothelial growth factor receptor 2 (VEGFR2). Since *ADGRG6* is involved in angiogenesis, which is critical for the development of pulmonary capillary beds during fetal life, deletion of *ADGRG6* leads to mid-gestation embryonic lethality due to failure in cardiovascular development. GWAS of spirometric measures of airflow limitation (FEV1/FVC ratio) have indicated several genes and pathways involved in branching morphogenesis and lung development, implicating an early-life origin of complex adult respiratory diseases such as COPD. Intriguingly, this GWAS of diffusing capacity of the lung (*D*LCO and *D*LCO/*V*A) also indicates a gene (*ADGRG6*) which is implicated in cardiopulmonary development during fetal life.

The modulating effect of *ADGRG6* on *VEGFR2* expression was shown to be mediated through the transcriptional activation of *STAT5* and *GATA2* [24]. Interestingly, *GATA2* was recently linked to pulmonary alveolar proteinosis [25], a rare lung disease, characterised by an abnormal accumulation of pulmonary surfactant in the alveoli, leading to an altered gas exchange.

Moreover, knock down of *ADGRG6* in the mouse retina was shown to result in the suppression of hypoxia-induced angiogenesis [24]. This information is interesting in two ways: first it links *ADGRG6* to hypoxia, which is very much related to gas exchange; second, processes in the retina might provide a unique insight into lung microvasculature, since vascular changes in both the retina and the alveoli reflect very much the same process, *i.e.* micro-angiopathy.

Although there is a good body of evidence that *ADGRG6* is important in lung development and micro-angiopathy, mRNA expression of *ADGRG6* has not been studied in lung diseases such as COPD and decreased diffusing capacity. Therefore, we performed an expression analysis of *ADGRG6* in human lung tissue and demonstrated that mRNA expression of *ADGRG6* is decreased significantly in patients with COPD and individuals with a decreased *D*LCO/*V*A.

Strengths and limitations

We conducted our analyses using data from two population based studies; the Rotterdam Study and the Framingham Heart Study. The strength of these studies is the population-based setting including data from smokers and nonsmokers, and the standardised prospective data collection. We are not aware of other population-based cohort studies that have D_{LCO} data in genotyped individuals available. Therefore, replication in other population-based cohorts was not possible. Yet, the results of the independent analyses in the Rotterdam study and the Framingham Heart Study show that rs17280293 already reaches genome-wide significance in the Framingham Heart Study and replicates in the Rotterdam Study. Finally, a gene expression analysis on lung tissue was performed in our lab in very well-defined patient groups.

This study has some limitations. First, for the measurements of diffusing capacity, the single-breath technique was used. This technique is known to underestimate measurements of V_A in individuals with obstructive disease or air trapping, since diffusing capacity cannot be measured in poorly ventilated areas of the lung. It is also known that the underestimation of V_A will be greater in more severe COPD and less in milder COPD. However, in our population-based cohorts, there are few individuals with severe COPD, thus reducing the impact of the underestimation of V_A in our study. Second, haemoglobin-corrected D_{LCO} measures were only available in the Rotterdam Study. However, the performed sensitivity analysis with or without correction for haemoglobin did not materially change the results within the Rotterdam Study. Third, the high D′ between rs17280293 and rs11155242 might suggest linked variant occurrence. However, the high D′ between those variants, estimated using data from the 1000-genomes reference panel, could result from the inflated estimation of D′ due to the low frequency of the SNPs. For this, it would be helpful to estimate the D′ in a bigger reference panel, such as the haplotype reference consortium when this information becomes available. Finally, in this study, we controlled for FEV_1/FVC in our models. In addition, we investigated the genetic correlation between gas exchange and FEV_1/FVC. While controlling for FEV_1/FVC in our analysis presents compelling evidence that rs17280293 is independently associated with D_{LCO}/V_A, lack of genome-wide genetic correlation between diffusing capacity and FEV_1/FVC does not exclude the possibility of pleiotropy at this specific locus, given that genetic correlation analyses are influenced by power, and our GWAS has a relatively small sample size. Therefore, caution is warranted in interpreting these results.

In conclusion, D_{LCO} and D_{LCO}/V_A are heritable traits with a considerable proportion of variance in diffusing capacity of the lung explained by genetics. We identified a functional variant in *ADGRG6*, a gene which is involved in gas exchange and hypoxia and differentially expressed in lung tissue of patients with COPD and subjects with decreased diffusing capacity. Therefore, experimental studies are needed to investigate the pathophysiological mechanisms and their therapeutic implications.

Acknowledgements: The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters, and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol Estrada, Yurii Aulchenko and Carolina Medina-Gomez for the creation and analysis of imputed data. Data can be obtained upon request. Requests should be directed towards the management team of the Rotterdam Study (secretariat.epi@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository.

Author contributions: N. Terzikhan, F. Sun and J. Dupuis analysed heritability and genome-wide association study data. L. Lahousse contributed to the data quality control. N. Terzikhan, H.H.H. Adams, L. Lahousse and G.G. Brusselle designed the study. L. Lahousse, J. Dupuis, G.G. Brusselle and G.T. O'Connor supervised the study. F.M. Verhamme and K.R. Bracke performed and analysed the gene expression study. K.R. Bracke supervised the gene expression study. N. Terzikhan, F. Sun and F.M. Verhamme wrote the manuscript. All authors contributed equally to revising the manuscript.

Conflict of interest: L. Lahousse reports personal fees for consultancy from Boehringer Ingelheim GmbH, accommodation support and personal fees for consultancy from Novartis, grants from AstraZeneca, travel grants and accommodation support from European Respiratory Society, grants and accommodation support from Belgian Respiratory Society, and grants from Chiesi, outside the submitted work. G.T. O'Connor reports personal fees for consulting regarding medications for asthma and COPD from AstraZeneca, and grants paid to Boston University to study genomics of COPD from Janssen Pharmaceuticals, outside the submitted work.

Support statement: This work was supported by a grant from the Fund for Scientific Research Flanders (FWO) project (G035014N). F.M. Verhamme is a post-doctoral fellow of the Fund for Scientific Research Flanders (FWO). The Rotterdam Study is funded by Erasmus MC and Erasmus University Rotterdam; the Netherlands Organisation for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (number 175.010.2005.011, 911–03-012), the Genetic Laboratory of the Dept of Internal

Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project number 050-060-810. The Framingham Heart Study was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract number N01-HC-25195 and HHSN268201500001I) and its contract with Affymetrix, Inc., for genotyping services (contract number N02-HL-6-4278). Also supported by NIH P01 AI050516. Research from the Laboratory for Translational Research in Obstructive Pulmonary Diseases in Gent is supported by the Concerted Research Action of the Ghent University (BOF/GOA, 01G02714).

References

1 Hegewald MJ. Diffusing capacity. *Clin Rev Allergy Immunol* 2009; 37: 159–166.
2 Macintyre N, Crapo RO, Viegi G, *et al.* Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 2005; 26: 720–735.
3 Crapo RO, Jensen RL, Wanger JS. Single-breath carbon monoxide diffusing capacity. *Clin Chest Med* 2001; 22: 637–649.
4 Bailey KL. The importance of the assessment of pulmonary function in COPD. *Med Clin North Am* 2012; 96: 745–752.
5 Hancock DB, Eijgelsheim M, Wilk JB, *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010; 42: 45–52.
6 Loth DW, Soler Artigas M, Gharib SA, *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* 2014; 46: 669–677.
7 Soler Artigas M, Loth DW, Wain LV, *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082–1090.
8 Wain LV, Shrine N, Artigas MS, *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017; 49: 416–425.
9 Hofman A, Brusselle GG, Darwish Murad S, *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015; 30: 661–708.
10 Yang J, Lee SH, Wray NR, *et al.* GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc Natl Acad Sci USA* 2016; 113: E4579–E4580.
11 Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; 62: 1198–1211.
12 Hobbs BD, de Jong K, Lamontagne M, *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017; 49: 426–432.
13 Cho MH, Castaldi PJ, Hersh CP, *et al.* A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am J Respir Crit Care Med* 2015; 192: 559–569.
14 Benner C, Spencer CC, Havulinna AS, *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; 32: 1493–1501.
15 Ghio AJ, Crapo RO, Elliott CG, *et al.* Heritability estimates of pulmonary function. *Chest* 1989; 96: 743–746.
16 Hallberg J, Iliadou A, Anderson M, *et al.* Genetic and environmental influence on lung function impairment in Swedish twins. *Respir Res* 2010; 11: 92.
17 Yang J, Lee SH, Goddard ME, *et al.* GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; 88: 76–82.
18 Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatry* 2015; 20: 98–108.
19 Zaitlen N, Kraft P, Patterson N, *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* 2013; 9: E1003520.
20 Yang J, Bakshi A, Zhu Z, *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* 2015; 47: 1114–1120.
21 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4: 1073–1081.
22 Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; doi: 10.1002/0471142905.hg0720s76.
23 Eichstaedt CA, Pagani L, Antao T, *et al.* Evidence of early-stage selection on EPAS1 and GPR126 genes in Andean high altitude populations. *Sci Rep* 2017; 7: 13042.
24 Cui H, Wang Y, Huang H, *et al.* GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. *J Biol Chem* 2014; 289: 34871–34885.
25 Papiris SA, Tsirigotis P, Kolilekas L, *et al.* Pulmonary alveolar proteinosis: time to shift? *Expert Rev Respir Med* 2015; 9: 337–349.