

## The TERT-CLPTM1L locus for lung cancer predisposes to bronchial obstruction and emphysema

Els Wauters<sup>1,2,3\*</sup>, Dominiek Smeets<sup>1,2\*</sup>, Johan Coolen<sup>4</sup>, Johny Verschakelen<sup>4</sup>, Paul De Leyn<sup>5</sup>, Marc Decramer<sup>3</sup>, Johan Vansteenkiste<sup>3</sup>, Wim Janssens<sup>3</sup>, Diether Lambrechts<sup>1,2</sup>

1. Vesalius Research Center (VRC), VIB, 3000 Leuven, Belgium.
2. Vesalius Research Center (VRC), K.U.Leuven, 3000 Leuven, Belgium.
3. Respiratory Division, University Hospital Gasthuisberg, K.U.Leuven, Belgium.
4. Department of Radiology, University Hospital Gasthuisberg, K.U.Leuven, Belgium.
5. Department of Thoracic surgery, University Hospital Gasthuisberg, K.U.Leuven, Belgium.

\* These authors contributed equally to this work

### **Corresponding author:**

Diether Lambrechts, MSc, PhD.  
Vesalius Research Center  
VIB, KULeuven  
Campus Gasthuisberg, Herestraat 49, box 912  
B-3000, Leuven, Belgium  
Tel: +32-16-34.61.31; Fax: +32-16-34.59.90  
E-mail: [diether.lambrechts@vib-kuleuven.be](mailto:diether.lambrechts@vib-kuleuven.be)

**Funding:** E.W. is supported by the Vlaamse Liga tegen Kanker (VLK), W.J. by the Fund for Scientific Research Flanders (FWO).

**Keywords:** chronic obstructive pulmonary disease, emphysema, lung cancer, susceptibility, telomerase reverse transcriptase gene

**Running title:** The 5p15.33 locus in bronchial obstruction and emphysema

## ABSTRACT

Clinical studies suggest that bronchial obstruction and emphysema increase susceptibility to lung cancer. We assess the possibility of a common genetic origin and investigate whether the lung cancer susceptibility locus on chromosome 5p15.33 increases the risk for bronchial obstruction and emphysema.

Three variants in the 5p15.33 locus encompassing the TERT and CLPTM1L genes were genotyped in 777 heavy smokers and 212 lung cancer patients. Participants underwent pulmonary function tests and computed tomography (CT) of the chest, and took questionnaires assessing smoking behaviour.

The rs31489 C-allele correlated with reduced forced expiratory volume in 1 second (FEV1;  $P=0.006$ ). Homozygous carriers of the rs31489 C-allele exhibited increased susceptibility to bronchial obstruction with an odds ratio (OR) of 1.82 (95% confidence interval [CI]=1.24-2.69;  $P=0.002$ ). A similar association was noticed for lung diffusing capacity (DLCO;  $P=0.004$ ). Consistent herewith, CC-carriers had an increased risk of emphysema (OR=2.04; CI=1.41-2.94;  $P=1.73 \times 10^{-4}$ ) and displayed more alveolar destruction. Finally, CC-carriers also had an increased risk for lung cancer (OR=1.90; CI=1.21-2.99;  $P=0.005$ ) and were more susceptible to develop both lung cancer and bronchial obstruction than lung cancer alone (OR=2.11; CI=1.04-4.26;  $P=0.038$ ).

The rs31489 variant on 5p15.33 is associated with bronchial obstruction, presence and severity of emphysema and lung cancer.

## INTRODUCTION

Around 50-70% of patients diagnosed with lung cancer exhibit some evidence of chronic obstructive pulmonary disease (COPD). Although cigarette smoking is generally considered to be responsible for this co-occurrence, substantial epidemiological data show that the prevalence of COPD in lung cancer patients is two- to six-fold higher compared to the prevalence of COPD in a cohort of heavy smokers, and this independently of age and pack-years smoked [1]. Moreover, reduced forced expiratory volume in one second (FEV1) has been shown to increase the risk for incident lung cancer independently of smoking history [2]. COPD is also associated with lung cancer in never smokers [3], thereby suggesting that additional risk factors might explain the link between COPD and lung cancer.

Emphysema, which is characterized by alveolar destruction, is an important clinical phenotype of COPD. Indeed, COPD patients with emphysema have a poorer health-related quality of life and exhibit an increased risk of non-cancer related respiratory death [4]. Epidemiological data suggest that emphysema and bronchial obstruction independently increase susceptibility to lung cancer. For instance, a large study revealed a >2-fold increase of lung cancer risk in patients with bronchial obstruction, after adjusting for smoking history, and a >3-fold increase in patients with emphysema, after adjusting for both smoking history and COPD severity [2]. Although these data were not confirmed in a second smaller study [5], a 1.85-fold increased risk for lung cancer was subsequently reported in women with a prior history of bronchial obstruction, but as much as a 6.36-fold increased risk when emphysema was diagnosed within 9 years prior to the lung cancer [6]. Overall, this suggests that emphysema and lung cancer are also intertwined and that common genetic or other unidentified susceptibility factors may predispose to both diseases.

Variants in the genes encoding the nicotinic acetylcholine receptor subunits (nAChR) were the first to be implicated in lung cancer, COPD and emphysema [7-9]. Intriguingly, the same variants were also discovered as genetic predictors of smoking addiction and were subsequently shown to be associated with a more intense smoking behavior [10, 11]. Although studies assessing the risk effect of nAChR variants on lung cancer, COPD or emphysema were corrected for smoking history, it remains difficult to interpret these associations independently from the patient's smoking behaviour. In contrast herewith, the TERT-CLPTM1L locus, which is the second locus that was identified using genome-wide association (GWA) in lung cancer [7], is thought to associate directly with lung cancer and

entirely independent of smoking behaviour. Indeed, the association between this locus and lung cancer occurs independently of tobacco-usage since it has been replicated in never-smokers [12]. Furthermore, the TERT-CLPTM1L locus has also been identified in non-smoking related cancers, such as cervix and prostate cancer [13], but was not identified in GWA studies involving smoking addiction [11]. The locus maps to two potential cancer susceptibility genes: the telomerase reverse transcriptase (TERT) gene, which encodes the telomerase protein that sustains telomere length [14], and the cleft lip palate transmembrane 1-like (CLPTM1L) gene, which encodes a protein linked to cisplatin resistance [15].

In an effort to assess the intriguing hypothesis that lung cancer, COPD and emphysema may have a common genetic origin, we here tested whether variants in the TERT-CLPTM1L locus are also associated with bronchial obstruction or emphysema.

## **MATERIALS AND METHODS**

### **Study design**

This was an observational and prospective study carried out by the respiratory division of the University Hospital of Leuven (Belgium). The protocol was run on an outpatient basis. The protocol was approved by the Ethics Committee of the University Hospitals Leuven (Belgium), and all subjects consented prior to study participation.

### **Study subjects**

We recruited two independent study populations at the University Hospital of Leuven (Belgium). A first population consisted of 777 heavy smokers, which were prospectively included between January 2007 and August 2010. Of these, 394 were recruited at the respiratory outpatient clinic because of symptoms suggestive of chronic obstructive pulmonary disease (COPD) [9]. In addition, 383 population-based subjects, who were Belgian participants of the Dutch-Belgian randomized lung cancer screening trial (NELSON), were included [16]. Inclusion criteria were a smoking history of at least 15 pack-years, a minimum age of 50 years and the availability of a complete pulmonary function test (including lung diffusing capacity measurements). Patients with suspicion or diagnosis of asthma were excluded, as well as patients with other respiratory diseases affecting pulmonary function. All COPD patients had a stable clinical condition with no exacerbation within 6 weeks before inclusion. From all study subjects, an extensive list of demographic variables (including age, gender, body mass index in kg/m<sup>2</sup>), a computed tomography (CT) scan of the chest within one year of enrolment and questionnaires determining smoking history were also collected.

A second population consisted of 212 newly diagnosed lung cancer patients, which were prospectively recruited at the Respiratory Oncology Unit between March 2010 and January 2011. The same inclusion criteria as for the first study population were applied, i.e., >50 years of age, a smoking history of  $\geq 15$  pack-years and the availability of a complete pulmonary function test prior to medical or surgical treatment. Additionally, presence of primary lung cancer had to be confirmed by a pathologist. The following cancer subtypes were diagnosed: small cell carcinoma (10.9%), adenocarcinoma (36.3%), squamous cell carcinoma (29.2%), and other non-small cell lung carcinoma (mainly large cell and bronchoalveolar subtypes, 23.6%). No data on the radiographic extent of emphysema were obtained for this subgroup. All recruited study participants self-declared Belgian-Flemish ethnicity for three generations.

## **Pulmonary function testing**

All pulmonary function measurements were performed with standardized equipment (SensorMedics Whole Body Plethysmograph, Viasys Healthcare, Belgium) and according to American Thoracic Society and European Respiratory Society guidelines [17]. Spirometric values were post-bronchodilator measurements. Diffusing capacity of the lung was determined by the single breath carbon monoxide gas transfer method (DLCO) and corrected for alveolar ventilation (KCO) but not for hemoglobin concentration [18]. All variables are given as absolute values expressed as % predicted of reference values. To determine if the TERT-CLPTM1L region is a susceptibility locus for COPD, we performed a case-control study in which COPD cases were defined as participants with a post-bronchodilator FEV1/FVC < 0.70 and FEV1 < 80% predicted (i.e. GOLD stage 2+ according to the Global Initiative for Chronic Obstructive Lung Diseases or GOLD classification) [19]. Subjects with a FEV1/FVC > 0.70 and FEV1 > 80% predicted were selected as healthy smokers. The same definitions were used in three recently published genome-wide association studies [8, 20, 21] to limit misclassification of healthy smokers as COPD GOLD stage I patients.

## **CT Imaging Protocol**

All 777 participants of the first study population received a chest CT scan allowing the semi-quantitative assessment of alveolar destruction. The complete protocol used for CT imaging and quantification of emphysema has been described previously [22]. Briefly, a blinded radiologist specialized in thoracic imaging scored each of the CT scans for the presence and extent of emphysema at three levels in each lung. Emphysema was defined as an area of hypovascular low attenuation, graded at each level with an incremental 5% scale and averaged in a tissue score reflecting the extent of emphysema over both lungs. If emphysema was visually scored on any of the predefined fields, the patient was categorized as having emphysema. Based on National Emphysema Treatment Trial criteria [22], four categories were generated yielding an alveolar destruction score ranging from 0 to 3. These categories were defined as no emphysema, emphysema affecting < 20%, between 20-50% and > 50% of the lung, respectively.

## **Genotyping**

Peripheral blood was sampled in K2EDTA plastic vacutainer tubes, and after centrifugation germ-line DNA was extracted from the precipitated leukocyte cell fraction according to standard procedures. DNA was aliquoted into 384-well plates and genotyped at the Vesalius Research Center (Leuven, Belgium). Genotyping for the lung cancer

susceptibility variants rs31489, rs4635969 and rs2736100 (identified in [23]) was performed in a blinded manner using iPLEX technology on a MassARRAY Compact Analyser (Sequenom Inc., San Diego, CA, USA), as reported previously [9]. The rs31489, rs2736100 and rs4635969 variants are located in non-coding regions at chromosome 5, respectively, at location 1395714, 1339516 and 1361552 (NCBI dbSNP identifier, Human genome Build 36 location). Automated genotyping calls were generated using the MassARRAY RTTM software and were validated by manual review of the raw mass spectra. Quality control was performed by genotyping 13 samples in duplicate, with a duplicate concordance of 100%. To our knowledge, this is the first study assessing these variants as susceptibility factor for COPD or emphysema.

### **Statistical analysis**

Data are summarized as frequencies and percentages for categorical variables. Continuous variables are presented as means with standard deviation or medians with 25th and 75th percentiles. Differences in baseline characteristics between patients with bronchial obstruction, emphysema or lung cancer and healthy participants were compared by the student's t-test, Mann-Whitney or Chi-square test. Hardy-Weinberg disequilibrium was tested using standard Chi-square analysis (1 degree of freedom), and all 3 single nucleotide polymorphisms (SNPs) fulfilled the criteria ( $P > 0.05$ ). Differences in baseline characteristics between rs31489 genotypes were compared by univariate analysis, Kruskal-Wallis or Chi-square test. The lambda statistic was used to calculate the genetic model of inheritance [24]. According to this method, values equal to 0, 0.5 and 1 correspond to a recessive, additive, and dominant genetic model, respectively. For the identification of disease-associated genotypes (COPD, emphysema and lung cancer versus healthy smokers), we found a lambda coefficient of  $< 0.001$ , 0.089 and 0.055, respectively, suggesting a recessive inheritance risk model. Chi-square analysis (1 degree of freedom) and logistic regression was used to assess the association of rs31489 with presence of bronchial obstruction, emphysema or lung cancer under a recessive model, respectively, without adjustment or after adjusting for other co-variables. The P value threshold for significance was adjusted for 6 multiple comparisons using the Bonferroni correction method, resulting in a significance threshold of  $P < 0.008$ . In particular, we corrected for testing the association between 3 variants and 2 variables (COPD and emphysema). We did not correct for additional variables such as FEV1 and DLCO, as these are closely linked to COPD or emphysema status. We did also not correct for testing the association with lung cancer, as all three variants were identified through a GWA

study, in which they were already extensively replicated. All statistical analyses were performed using SPSS software version 17.0 (SPSS Inc., Chicago, Ill. USA).



## RESULTS

### Population characteristics

In total, 777 heavy smokers free of lung cancer were included in this study. Details on demographics, smoking behavior and pulmonary function status are shown in Table 1. Briefly, the mean age of study participants was 64.9 years, 77.7% was male smoker with an average pack-years history of 48.8 years. Genotyping for the rs31489, rs4635969 and rs2736100 variants succeeded in respectively 739 (95.1%), 776 (99.6%) and 764 (98.7%) participants. These variants were identified in lung cancer GWA studies and each represent tagging SNPs of different linkage disequilibrium (LD) blocks: their pairwise  $r^2$ -values were 0.415, 0 and 0.003, respectively for rs4635969 and rs31489, rs4635969 and rs2736100, and rs31489 and rs2736100 (according to HapMap for CEU-based European ancestry [25]).

### The rs31489 variant correlates with reduced pulmonary function

First, we assessed whether any of the 3 SNPs were associated with demographic parameters, smoking exposure or pulmonary function. Significant differences between the rs31489 genotypes were noticed at the pulmonary function level (Table 1). Homozygous carriers of the rs31489 C-allele had significantly lower FEV1 values compared to AA carriers ( $P=0.006$  and  $P=0.014$ , respectively for mean absolute values and % predicted values, Table 1). Interestingly, a similar association was seen for the lung diffusing capacity parameter DLCO, with mean DLCO values being the lowest for carriers of the CC-genotype ( $P=0.004$ ; Table 1). Adjustments for age, gender, height, pack-years and years-quit confirmed that the associations of rs31489 with FEV1 and DLCO were independent from other known risk factors ( $P=0.008$  and  $P=0.007$ , respectively). No significant associations were observed for rs4635969 and rs2736100 ( $P=0.650$  and  $P=0.724$  for FEV1 and  $P=0.827$  and  $P=0.374$  for DLCO values, respectively for both variants).

### Association of rs31489 with bronchial obstruction

Since rs31489 correlated with reduced FEV1, we also assessed the association between rs31489 and COPD. A case-control association analysis, in which cases were defined as COPD patients diagnosed with moderate to severe bronchial obstruction (GOLD stage 2+) and controls as participants with a  $FEV1/FVC > 0.70$  and  $FEV1 > 80\%$  was performed. Baseline characteristics for the 653 participants, of which 245 were healthy smokers and 408 were GOLD stage 2+ COPD patients, are shown in Table 2. In a crude

analysis, the rs31489 C-allele was more common in patients with bronchial obstruction (60.8% in obstructive versus 54.8% in healthy smokers,  $P=0.037$ , Table 3). Using the lambda statistic to calculate the genetic model of inheritance for rs31489, the C-allele was identified as a recessive susceptibility allele [24]. At the genotypic level, 39.1% of COPD patients were CC-carriers compared to 28.6% of the healthy smokers (OR=1.60; CI=1.13-2.27;  $P=0.010$  according to a recessive risk model). Logistic regression confirmed that CC-carriers significantly increased the risk for bronchial obstruction independently of age, gender, height, pack-years and years-quit (OR=1.82; CI=1.24-2.69;  $P=0.002$ ; Table 3). The rs31489 variant did, however, not correlate with severity of COPD since genotype distributions between the different GOLD stages were similar ( $P=0.268$ ; online supplementary material Table S1). We did not find a significant association between rs4635969 or rs2736100 and bronchial obstruction ( $P=0.636$  and  $P=0.598$  for genotypes, respectively).

### **Association of rs31489 with emphysema and severity of emphysema**

Since rs31489 correlated with lung diffusing capacity (DLCO), we also investigated the association of rs31489 with emphysema. Of the 408 COPD patients, 318 (78%) patients exhibited emphysema, as assessed by CT (see Table 2 for baseline characteristics). A case-control analysis between these 318 emphysema patients and the 245 healthy smokers subsequently revealed that the C-allele was more common in emphysema patients than in healthy smokers (62.5% versus 54.8%, respectively;  $P=0.010$ ; Table 3). At the genotypic level, 42.1% of the emphysema patients were CC-carriers compared to 28.6% of the healthy smokers (OR=1.80; CI=1.29-2.56;  $P=0.001$  according to a recessive risk model). Logistic regression confirmed that the association between rs31489 and emphysema was independent of age, gender, height, pack-years and years-quit (OR=2.04; CI=1.41-2.94;  $P=1.73 \times 10^{-4}$  in the recessive model; Table 3). The association remained significant after including FEV1 as a covariate in the analysis (OR=2.07; CI=1.23-3.46;  $P=0.006$ ). We did not find a significant association between rs4635969 or rs2736100 and emphysema ( $P=0.620$  and  $P=0.687$  for genotypes, respectively).

To assess severity of emphysema, emphysema was scored at three levels in each lung and the tissue score, which is the average of the scores in the individual six fields and is expressed in percentage (%) of the lung affected, was calculated. CC-carriers exhibited a significantly higher score in each of the six fields compared to the other rs31489 genotypes ( $P<0.05$  for every field; data not shown). Additionally, CC-carriers also exhibited

a higher tissue score ( $P=0.003$ , online supplementary material Table S2). Next, based on the National Emphysema Treatment Trial criteria [22], the alveolar destruction score was generated by categorizing the tissue score into four classes of increasing emphysema severity. These categories, ranging from 0 to 3, were defined as no emphysema, emphysema affecting <20%, between 20-50% and >50% of the lung, respectively. CC-carriers more frequently exhibited emphysema in >50% of the lung than AA-carriers (18.8% versus 7.6%;  $P=0.010$ ; supplementary Table S2). Overall, these data suggest that rs31489 correlates with the severity of emphysema in COPD patients.

### **Association of rs31489 with COPD in lung cancer patients**

Next, we assessed the rs31489 variant in an independent cohort of 212 newly diagnosed lung cancer patients and explored whether rs31489 contributes to a shared genetic predisposition of bronchial obstruction and lung cancer. Details on demographics, smoking behavior and pulmonary function status of the 212 patients are shown in Table 4. Briefly, all recruited lung cancer patients were heavy smokers with an average pack-years history of 43.6 years, and 72.6% was male with an average age of 66.2 years. Genotyping for rs31489, rs4635969 and rs2736100 was successful in 203 (95.7%), 208 (98.1%) and 210 (99.1%) subjects, respectively.

In a first analysis, we compared rs31489 allele frequencies between 212 lung cancer patients and 245 healthy smokers (Table 4). As expected, the at-risk C-allele was more common in patients with lung cancer (61.3% versus 54.8%,  $P=0.049$ , Table 5). At the genotypic level, 39.4% of the lung cancer patients were CC-carriers compared to 28.6% of the healthy smokers ( $OR=1.65$ ;  $CI=1.11-2.45$ ;  $P=0.018$  according to a recessive risk model; Table 5). Regression analysis revealed that the association between rs31489 and lung cancer was independent of age, gender, height, pack-years smoked, and years-quit ( $OR=1.90$ ;  $CI=1.21-2.99$ ;  $P=0.005$ ). Since rs31489 was directly correlated with FEV1 in the cohort of heavy smokers (Table 1) and was significantly associated with COPD (Table 3), we also assessed whether the at-risk rs31489 C-allele was associated with lung cancer independently of FEV1. A similar regression analysis was performed, while correcting for FEV1 and allowing for the interaction of FEV1 with rs31489. CC-carriers of rs31489 displayed a significantly higher risk to develop lung cancer ( $P=0.006$ ), suggesting a direct effect of rs31489 on lung cancer risk independently of FEV1. Remarkably, the interaction between rs31489 and FEV1 was also significant, indicating that smokers carrying the CC-genotype are more susceptible to have a reduced FEV1 and to develop lung cancer ( $P=0.002$  for interaction term). Consequently, when stratifying lung cancer patients into two categories, based on the presence or absence of COPD, we observed

that rs31489 CC-carriers were almost twice as frequent in lung cancer patients with COPD than in lung cancer patients without COPD (65.0% versus 35.0%, respectively). Binary logistic regression analysis, while correcting for age, gender, height, pack-years smoked and years-quit, confirmed that CC-carriers were more susceptible to develop both lung cancer and bronchial obstruction than lung cancer alone (OR=2.11; CI=1.04-4.26; P=0.038).

## DISCUSSION

The current study identified rs31489 on chromosome 5p15.33 as a susceptibility variant for bronchial obstruction and emphysema. Indeed, rs31489 correlated significantly with reduced lung function. In particular, CC-carriers exhibited significantly lower FEV1 values, as well as a reduced lung diffusing capacity. In addition, rs31489 conferred an increased risk to develop bronchial obstruction, whereas CT scans obtained from the same patients provided evidence for an association between rs31489 and the presence and severity of emphysema. Each of these associations remained significant after adjustment for smoking behaviour and other clinical variables, including age and height. Secondly, our data confirm that rs31489 also acts as a shared susceptibility factor for lung cancer and bronchial obstruction. Indeed, carriers of the rs31489 CC-genotype were significantly more common in lung cancer patients than in cancer-free healthy smokers. This association remained significant after incorporating FEV1 as a covariate in the analysis. Intriguingly, rs31489 interacted significantly with FEV1, indicating that rs31489 also correlates with FEV1 in lung cancer patients. Stratification of lung cancer patients into those with and without bronchial obstruction confirmed that CC-carriers were two times more susceptible to develop both bronchial obstruction and lung cancer than lung cancer alone. These findings are in agreement with previous observations from Young et al., reporting that variants in the hedgehog-interacting protein (HHIP) and nAChR subunits are more strongly associated with lung cancer and COPD than lung cancer alone [26, 27]. On a more general level, the more pronounced association of these loci with both COPD and lung cancer suggests that the pathogenetic pathways for smoking-related pulmonary disorders are shared.

The TERT-CLPTM1L locus contains numerous genetic variants that are strongly linked to each other and that are located in two genes, which could each be functionally affected by these variants [7, 12, 13, 23]. The first gene is CLPTM1L, in which the rs31489 variant is also located. Overexpression of this gene has been shown to enhance apoptosis in response to cisplatin in an ovarian tumour cell line [15]. The cytotoxic effects of cisplatin are attributed to its covalent binding to nuclear DNA, referred to as DNA adduct formation. Interestingly, there is recent evidence that a variant in the CLPTM1L gene (rs402710, pairwise  $r^2=0.67$  with rs31489) favors DNA damage through DNA adduct formation [28]. Hence, variants in the CLPTM1L gene may be functionally involved in the development of COPD or lung cancer by enhancing DNA damage in response to noxious smoke-related particles. Secondly, the region of linkage disequilibrium wherein rs31489 resides also contains the 5' end of the TERT gene. The function of this gene in human diseases,

especially in carcinogenesis, is much better understood than that of CLPTM1L [14]. Telomerase, the enzyme product of the TERT gene, ensures that the ends of human chromosomes are covered by functional telomeres. An appropriate telomere length is essential for the cell's integrity since it prevents chromosomal end-to-end fusions and subsequent chromosomal aberrations [29]. Telomerase activity is typically higher in cells where a higher proliferative potential is needed, for example in germ line stem cells. In differentiated somatic cells, however, telomerase levels are low or undetectable and telomeres get shorter each time a cell divides, until the point where telomere length becomes insufficient to prevent genomic instability and proliferative senescence is induced [30]. Intriguingly, rare germline mutations in TERT have been found in familial disorders such as dyskeratosis congenita, aplastic anemia and familial idiopathic pulmonary fibrosis [14], which are characterized by a limited cell and tissue renewal capacity and a higher rate of carcinogenesis. Lymphocytes from affected individuals are characterized by decreased telomerase activity and significantly shorter telomeres.

Based on the latter observations, it has been proposed that common variants in TERT-CLPTM1L, which were identified as susceptibility factors for cancer, are linked with reduced telomerase activity, short telomeres, limited tissue renewal capacity and chromosomal instability [31]. Indeed, a correlation between reduced telomere length and rs401681, which is strongly linked to rs31489 ( $r^2=0.87$ ), has already been observed [13], whereas Mirabello et al. also reported a direct correlation between rs31489 and shorter telomeres under a recessive genetic model [32]. Shorter telomeres have also been observed in peripheral leucocytes from COPD patients [33], and a decreased telomere length was observed in the tissue of emphysematous lungs [34]. Intriguingly, a recent GWA study also identified genetic variants in BICD1, which are known to correlate with short telomeres, as a susceptibility factor for emphysema [35]. In addition to genetic predisposition, which we observed in this study, telomere shortening in COPD and emphysema may also result from an increased number of cell divisions due to ongoing inflammation [36]. Short telomeres could possibly also predispose to more severe emphysema, as they limit tissue renewal capacity and enhance alveolar destruction in response to environmental factors such as smoking [30]. As already mentioned, critically short telomeres also lead to genomic instability and promote carcinogenesis [29], which could explain the clinically observed link between emphysema and lung cancer.

Strengths of the current study are the extensive characterization of pulmonary function, including lung diffusing capacity measurements of all participants, and the availability of CT scans for all heavy smokers. In addition, we included a second study

population to replicate the association of rs31489 with COPD. Since this population consisted of lung cancer patients with known lung function status, we were not only able to confirm the association of rs31489 with bronchial obstruction, but also to establish an independent correlation of rs31489 with lung cancer. However, since the latter study population was small, follow-up studies will be needed to confirm the independent association of the TERT-CLPTM1L locus with both COPD, emphysema and lung cancer. Another limitation is that we used subjective visual scoring to assess the severity of emphysema rather than an objective and automated quantification method. Moreover, although several studies have shown that the TERT-CLPTM1L locus has functional effects on telomere length and DNA adduct formation, the causal variants within this locus are still not known. As a consequence, we were limited to assessing the association of 3 tagging SNPs, previously identified as susceptibility variants for lung cancer in this locus, with COPD and emphysema. Finally, although our analysis may be subject to population heterogeneity we were not able to apply a genomic control method, since genome-wide data from unrelated variants in unrelated genes were not available for this study.

Overall, the conclusion of the current study is that the rs31489 variant on chromosome 5p15.33 is strongly and independently associated with bronchial obstruction, the presence and severity of emphysema and lung cancer. On a more general level these data support the hypothesis that lung cancer, bronchial obstruction and emphysema share common genetic grounds. Additionally, this study proposes an intriguing hypothesis, suggesting that genetic variability in telomere shortening could at least partially explain the link between bronchial obstruction, emphysema and lung cancer.

## **ACKNOWLEDGEMENTS**

The authors thank Kristien Debent, Claudia Carremans, Geert Celis, Bart Claes and Gilian Peuteman for their excellent assistance in the data collection, blood sampling, DNA extraction and genotyping. We also thank the lung function and radiology technicians from the UZ Leuven for their dedication and efforts. We would also like to acknowledge the reviewers for their constructive and valuable comments.



## References

1. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble GD. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur Respir J* 2009; 34(2): 380-386.
2. Wilson DO, Weissfeld JL, Balkan A, Schragin JG, Fuhrman CR, Fisher SN, Wilson J, Leader JK, Siegfried JM, Shapiro SD, Sciurba FC. Association of radiographic emphysema and airflow obstruction with lung cancer. *Am J Respir Crit Care Med* 2008; 178(7): 738-744.
3. Turner MC, Chen Y, Krewski D, Calle EE, Thun MJ. Chronic obstructive pulmonary disease is associated with lung cancer mortality in a prospective study of never smokers. *Am J Respir Crit Care Med* 2007; 176(3): 285-290.
4. Haruna A, Muro S, Nakano Y, Ohara T, Hoshino Y, Ogawa E, Hirai T, Niimi A, Nishimura K, Chin K, Mishima M. CT scan findings of emphysema predict mortality in COPD. *Chest* 2010; 138(3): 635-640.
5. Maldonado F, Bartholmai BJ, Swensen SJ, Midthun DE, Decker PA, Jett JR. Are airflow obstruction and radiographic evidence of emphysema risk factors for lung cancer? A nested case-control study using quantitative emphysema analysis. *Chest* 2010; 138(6): 1295-1302.
6. Schwartz AG, Cote ML, Wenzlaff AS, Van Dyke A, Chen W, Ruckdeschel JC, Gadgeel S, Soubani AO. Chronic obstructive lung diseases and risk of non-small cell lung cancer in women. *J Thorac Oncol* 2009; 4(3): 291-299.
7. Brennan P, Hainaut P, Boffetta P. Genetics of lung-cancer susceptibility. *Lancet Oncol* 2010; Epub ahead of print.
8. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5(3): e1000421.
9. Lambrechts D, Buyschaert I, Zanen P, Coolen J, Lays N, Cuppens H, Groen HJ, Dewever W, van Klaveren RJ, Verschakelen J, Wijmenga C, Postma DS, Decramer M, Janssens W. The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *Am J Respir Crit Care Med* 2010; 181(5): 486-493.
10. Le Marchand L, Derby KS, Murphy SE, Hecht SS, Hatsukami D, Carmella SG, Tiirikainen M, Wang H. Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res* 2008; 68(22): 9137-9140.
11. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemenev LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010; 42(5): 448-453.
12. Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis* 2010; 31(2): 234-238.

13. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, Jakobsdottir M, Helgadóttir H, Thorlacius S, Aben KK, Blondal T, Thorgeirsson TE, Thorleifsson G, Kristjansson K, Thorisdóttir K, Ragnarsson R, Sigurgeirsson B, Skuladóttir H, Gudbjartsson T, Isaksson HJ, Einarsson GV, Benediktsdóttir KR, Agnarsson BA, Olafsson K, Salvarsdóttir A, Bjarnason H, Asgeirsdóttir M, Kristinsson KT, Matthiasdóttir S, Sveinsdóttir SG, Polidoro S, Hoiom V, Botella-Estrada R, Hemminki K, Rudnai P, Bishop DT, Campagna M, Kellen E, Zeegers MP, de Verdier P, Ferrer A, Isla D, Vidal MJ, Andres R, Saez B, Juberias P, Banzo J, Navarrete S, Tres A, Kan D, Lindblom A, Gurzau E, Koppova K, de Vegt F, Schalken JA, van der Heijden HF, Smit HJ, Termeer RA, Oosterwijk E, van Hooij O, Nagore E, Porru S, Steineck G, Hansson J, Buntinx F, Catalona WJ, Matullo G, Vineis P, Kiltie AE, Mayordomo JI, Kumar R, Kiemeny LA, Frigge ML, Jonsson T, Saemundsson H, Barkardóttir RB, Jonsson E, Jonsson S, Olafsson JH, Gulcher JR, Masson G, Gudbjartsson DF, Kong A, Thorsteinsdóttir U, Stefansson K. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* 2009; 41(2): 221-227.
14. Calado RT, Young NS. Telomere diseases. *N Engl J Med* 2009; 361(24): 2353-2365.
15. Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. *Biochem Biophys Res Commun* 2001; 280(4): 1148-1154.
16. van Iersel CA, de Koning HJ, Draisma G, Mali WP, Scholten ET, Nackaerts K, Prokop M, Habbema JD, Oudkerk M, van Klaveren RJ. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4): 868-874.
17. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J. Standardisation of spirometry. *Eur Respir J* 2005; 26(2): 319-338.
18. Rosenberg E. The 1995 update of recommendations for a standard technique for measuring the single-breath carbon monoxide diffusing capacity (transfer factor). *Am J Respir Crit Care Med* 1996; 154(1): 265-266.
19. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176(6): 532-555.
20. Cho MH, Boutaoui N, Klanderma BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, Lange C, Won S, Murphy JR, Beaty TH, Regan EA, Make BJ, Hokanson JE, Crapo JD, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Bakke P, Gulsvik A, Pillai SG, Silverman EK. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010; 42(3): 200-202.
21. Wan ES, Cho MH, Boutaoui N, Klanderma BJ, Sylvia JS, Ziniti JP, Won S, Lange C, Pillai SG, Anderson WH, Kong X, Lomas DA, Bakke PS, Gulsvik A, Regan EA, Murphy JR, Make BJ, Crapo JD, Wouters EF, Celli BR, Silverman EK, Demeo DL. Genome-Wide Association Analysis of Body Mass in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol* 2010; Epub ahead of print.
22. National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume-reduction surgery. *N Engl J Med* 2001; 345(15): 1075-1083.
23. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, Vooder T, Nelis M, Valk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeboller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF, Jr., Savage SA, Bertazzi PA, Tucker MA, Chanock S,

- Caporaso NE. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009; 85(5): 679-691.
24. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol* 2005; 34(6): 1319-1328.
25. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghorji MJ, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; 467(7311): 52-58.
26. Young RP, Hopkins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *Eur Respir J* 2008; 32(5): 1158-1164.
27. Young RP, Whittington CF, Hopkins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Chromosome 4q31 locus in COPD is also associated with lung cancer. *Eur Respir J* 2010; 36(6): 1375-1382.
28. Zienolddiny S, Skaug V, Landvik NE, Ryberg D, Phillips DH, Houlston R, Haugen A. The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis* 2009; 30(8): 1368-1371.
29. Hackett JA, Feldser DM, Greider CW. Telomere dysfunction increases mutation rate and genomic instability. *Cell* 2001; 106(3): 275-286.
30. Hao LY, Armanios M, Strong MA, Karim B, Feldser DM, Huso D, Greider CW. Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. *Cell* 2005; 123(6): 1121-1131.
31. Baird DM. Variation at the TERT locus and predisposition for cancer. *Expert Rev Mol Med* 2010; 12: e16.
32. Mirabello L, Yu K, Kraft P, De Vivo I, Hunter DJ, Prescott J, Wong JY, Chatterjee N, Hayes RB, Savage SA. The association of telomere length and genetic variation in telomere biology genes. *Hum Mutat* 2010; 31(9): 1050-1058.
33. Savale L, Chaouat A, Bastuji-Garin S, Marcos E, Boyer L, Maitre B, Sarni M, Housset B, Weitzenblum E, Matrat M, Le Corvoisier P, Rideau D, Boczkowski J, Dubois-Rande JL, Chouaid C, Adnot S. Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2009; 179(7): 566-571.
34. Tsuji T, Aoshiba K, Nagai A. Alveolar cell senescence in patients with pulmonary emphysema. *Am J Respir Crit Care Med* 2006; 174(8): 886-893.
35. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, Washko G, Hoffman EA, Bakke P, Gulsvik A, Lomas DA, Silverman EK, Pillai SG. Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. *Am J Respir Crit Care Med* 2011; 183(1): 43-49.
36. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, Coxson HO, Pare PD. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350(26): 2645-2653.

**Table 1. Baseline characteristics for all heavy smokers free of lung cancer according to rs31489 genotypes.**

	rs31489				P value
	Total	CC	CA	AA	
<b>Subjects, no (%*)</b>	777	261 (35.3)	341 (46.1)	137 (18.5)	–
<b>Demographics</b>					
Age, mean (SD), yr	64.9 (7.9)	64.2 (7.8)	65.1 (8.2)	65.1 (7.4)	0.287
Male sex, no. (%)	604 (77.7)	194 (74.3)	274 (80.4)	111 (81.0)	0.144
Height, mean (SD), cm	170.3 (8.8)	169.4 (9.0)	171.2 (8.9)	170.8 (7.8)	0.037
<b>Smoking history</b>					
Pack-years history, mean (SD), yr	48.8 (25.0)	48.7 (25.7)	49.0 (24.2)	47.9 (25.9)	0.919
Current smokers, no. (%)	370 (47.7)	127 (48.7)	159 (46.6)	66 (48.2)	0.876
Smoked years, mean (SD), yr	41.7 (9.3)	41.2 (8.8)	41.7 (9.4)	41.0 (9.5)	0.731
Years-quit former smokers, median (25th-75th percentiles)	1.0 (0.0-8.0)	1.0 (0.0-8.0)	2.0 (0.0-10.0)	2.0 (0.0-10.0)	0.237
<b>Pulmonary function tests, mean (SD)</b>					
FEV1, L, post	2.15 (1.04)	2.03 (0.97)	2.30 (1.10)	2.21 (1.00)	0.006
FEV1, % predicted, post	72.9 (30.6)	69.6 (29.1)	77.0 (32.0)	74.4 (29.5)	0.014
FVC, L, post	3.57 (1.07)	3.46 (1.00)	3.71 (1.13)	3.62 (1.04)	0.019
FVC, % predicted, post	96.8 (22.0)	95.1 (21.3)	98.7 (22.7)	97.4 (21.2)	0.146
FEV1/FVC ratio	0.58 (0.17)	0.56 (0.17)	0.60 (0.17)	0.59 (0.17)	0.058
DLCO, mmol/min/kPa	5.73 (2.37)	5.42 (2.25)	6.08 (2.42)	5.87 (2.39)	0.004
DLCO, % predicted	65.3 (23.5)	62.3 (22.6)	68.7 (24.1)	66.4 (22.8)	0.004

\*: Percentages are row percentages. Of the 777 heavy smokers (free of lung cancer) genotyping for rs31489 succeeded in 739 participants (95.1%). Due to technical limitations, 6 values (<1%) for the DLCO parameter are missing. For all other variables, there were no missing data. Abbreviations: FEV1 = forced expiratory volume in one second; FVC = forced vital capacity; DLCO = carbon monoxide diffusing capacity.

**Table 2. Baseline characteristics for healthy smokers, COPD and emphysema patients.**

	Healthy smokers	COPD	Emphysema
<b>Subjects, no.</b>	245	408	318
<b>Demographics</b>			
Age, mean (SD), yr	61.5 (5.9)	67.5 (8.6)	67.7 (8.5)
Male sex, no. (%)	183 (74.7)	324 (79.4)	253 (79.6)
Height, mean (SD), cm	171.5 (9.3)	169.1 (8.4)	169.3 (8.4)
<b>Smoking history</b>			
Pack-years history, mean (SD), yr	42.5 (20.3)	52.5 (26.3)	53.2 (27.3)
Current smokers, no. (%)	116 (47.9)	176 (43.8)	128 (40.4)
Smoked years, mean (SD), yr	38.7 (7.9)	43.5 (10.1)	43.5 (10.1)
Years-quit former smokers, median (25th-75th percentiles)	1.0 (0.0-9.0)	1.5 (0.0-8.0)	2.0 (0.0-8.2)
<b>Pulmonary function tests, mean (SD)</b>			
FEV1, L, post	3.12 (0.71)	1.36 (0.56)	1.29 (0.54)
FEV1, % predicted, post	103.6 (14.8)	47.7 (16.7)	45.4 (16.4)
FVC, L, post	4.10 (0.93)	3.01 (0.86)	3.02 (0.85)
FVC, % predicted, post	108.8 (15.1)	83.6 (18.6)	84.0 (19.0)
FEV1/FVC ratio	0.76 (0.04)	0.45 (0.12)	0.42 (0.11)
DLCO, mmol/min/kPa	7.50 (1.87)	4.30 (1.72)	3.92 (1.47)
DLCO, % predicted	83.6 (15.5)	50.3 (17.9)	45.9 (15.6)

Of the 408 heavy smokers with COPD (GOLD stage 2+), 52 (12.7%) subjects were derived from the NELSON study and 356 (87.3%) were recruited at the respiratory outpatient clinic. In the COPD group, 318 patients were diagnosed with emphysema. As expected, baseline characteristics, smoking history and pulmonary function tests differed significantly between COPD patients and healthy smokers ( $P < 0.001$ ). Only the number of current smokers and the years-quit smoking did not differ ( $P = 0.305$  and  $P = 0.621$ , respectively). The same observations were made when comparing baseline characteristics between emphysema patients and healthy smokers. Abbreviations: COPD = chronic obstructive pulmonary disease; GOLD = Global Initiative for Chronic Obstructive Lung Disease; FEV1 = forced expiratory volume in one second; FVC = forced vital capacity; DLCO = carbon monoxide diffusing capacity.

**Table 3. Association of rs31489 with COPD and emphysema**

	Healthy smokers	COPD			Emphysema		
	N (%)	N (%)	OR (95% CI)	P value	N (%)	OR (95% CI)	P value
C	264 (54.8)	457 (60.8)	0.78 (0.62-0.98)	0.037	365 (62.5)	1.38 (1.08-1.76)	0.010
A	218 (45.2)	295 (39.2)			219 (37.5)		
CC	69 (28.6)	147 (39.1)	1.82(1.24-2.69)	0.002*	123 (42.1)	2.04 (1.41-2.94)	1.73x10 <sup>-4</sup> *
CA	126 (52.3)	163 (43.4)			119 (40.8)		
AA	46 (19.1)	66 (17.6)			50 (17.1)		

Allele and genotype frequencies of rs31489 in COPD (GOLD stage 2+, n=408) and emphysema patients (n=318) were compared with frequencies in healthy smokers (FEV1/FVC>0.70 and FEV1>80% predicted, n=245). %: percentages are column percentages; \*P value under the assumption of a recessive genetic model as assessed by logistic regression and after correction for co-variables is given. Abbreviations: N = number; COPD = chronic obstructive pulmonary disease; GOLD = Global Initiative for Chronic Obstructive Lung Disease; OR = odds ratio; CI = confidence interval.

**Table 4. Baseline characteristics of heavy smokers diagnosed with lung cancer.**

	Healthy smokers	Lung cancer patients
<b>Subjects, no. (%)</b>	245	212
<b>Demographics</b>		
Age, mean (SD), yr	61.5 (5.9)	66.2 (9.8)
Male sex, no. (%)	183 (74.7)	154 (72.6)
Height, mean (SD), cm	171.5 (9.3)	170.0 (7.5)
<b>Smoking history</b>		
Pack-years history, mean (SD), yr	42.5 (20.3)	43.6 (31.5)
Current smokers, no. (%)	116 (47.9)	114 (54.0)
Years-quit former smokers, median (25th-75th percentiles)	1.0 (0.0-9.0)	0.0 (0.0-9.0)
<b>Pulmonary function tests, mean (SD)</b>		
FEV1, L, post	3.12 (0.71)	2.27 (0.70)
FEV1, % predicted, post	103.6 (14.8)	79.0 (21.0)
FVC, L, post	4.10 (0.93)	3.46 (0.92)
FVC, % predicted, post	108.8 (15.1)	96.1 (19.0)
FEV1/FVC ratio	0.76 (0.04)	66.3 (12.6)
DLCO, mmol/min/kPa	7.50 (1.87)	5.78 (1.76)
DLCO, % predicted	83.6 (15.5)	67.2 (17.3)

Of the 245 healthy smokers, genotyping for rs31489 succeeded in 241 participants (98%). Of the 212 lung cancer patients, genotyping for rs31489 succeeded in 203 participants (95.7%). Due to technical limitations, 6 values for the DLCO parameter, including one in the group of cancer patients, are missing. For all other variables, complete data are presented. Lung cancer patients and healthy smokers were matched with respect to their gender, height and smoking history ( $P=0.619$  and  $P=0.056$  for gender and height, respectively;  $P=0.653$ ,  $P=0.196$  and  $P=0.621$  for pack-years history, current smoking status and years-quit former smokers, respectively). All pulmonary function tests differed significantly ( $P<0.001$ ). Abbreviations: FEV1 = forced expiratory volume in one second; FVC = forced vital capacity; DLCO = carbon monoxide diffusing capacity.

**Table 5. Association of rs31489 with lung cancer versus healthy smokers**

	Healthy smokers	Lung cancer patients		
	N (%)	N (%)	OR (95% CI)	P value*
C	264 (54.8)	249 (61.3)	1.30 (1.00-1.71)	0.049
A	218 (45.2)	157 (38.7)		
CC	69 (28.6)	80 (39.4)	1.65 (1.11-2.45)	0.018
CA	126 (52.3)	89 (43.8)		
AA	46 (19.1)	34 (16.7)		

Allele and genotype frequencies of rs31489 were compared between lung cancer cases (n=212) and healthy smokers (FEV1/FVC>0.70 and FEV1>80% predicted, n=245). %: percentages are column percentages; \*P value under the assumption of a recessive model is given (CC versus CA/AA). Abbreviations: N = number; OR = odds ratio; CI = confidence interval.