Supplementary Data

Inclusion / Exclusion Criteria for the Study Population

Smokers with normal spirometry and normal diffusion capcity (DLCO)

Inclusion criteria

- Males and females, at least 18 years old
- Capable of providing informed consent
- Willingness to participate in the study
- Good health without history of chronic lung disease, including asthma, and without recurrent or recent (within 3 months) acute pulmonary disease
- Normal physical examination
- Normal routine laboratory evaluation, including general hematologic studies, general serologic/immunologic studies, general biochemical analyses, and urine analysis
- Normal PA and lateral chest X-ray
- Normal electrocardiogram (sinus bradycardia, premature atrial contractions are permissible)
- Females - not pregnant
- No history of allergies to medications used in the bronchoscopy procedure
- Not taking any medications relevant to lung disease or having an effect on the airway epithelium
- Normal serum α1-antitrypsin level
- HIV1 negative
- Self-reported current daily smokers with >5 pack-yr, validated by urine nicotine >30 ng/ml and/or cotinine >50 ng/ml
- Normal FEV1 (≥80% predicted), FVC (≥80% predicted), FEV1/FVC (≥0.7) based on post-bronchodilator spirometry, TLC (≥80% predicted)
- DLCO ≥80% predicted

Exclusion criteria

- Unable to meet the inclusion criteria
- Alcohol or drug abuse within the past 6 months
- Evidence of malignancy within the past 5 years
- Current active infection or acute illness of any kind

Smokers with normal spirometry but low DLCO

Inclusion criteria

- Males and females, at least 18 years old
- Capable of providing informed consent
- Willingness to participate in the study
- Good health without history of chronic lung disease, including asthma, and without recurrent or recent (within 3 months) acute pulmonary disease
- Normal physical examination

- Normal routine laboratory evaluation, including general hematologic studies, general serologic/immunologic studies, general biochemical analyses, and urine analysis
- Normal PA and lateral chest X-ray
- Normal electrocardiogram (sinus bradycardia, premature atrial contractions are permissible)
- Females - not pregnant
- No history of allergies to medications used in the bronchoscopy procedure
- Not taking any medications relevant to lung disease or having an effect on the airway epithelium
- Normal serum α1-antitrypsin level
- HIV1 negative
- Self-reported current daily smokers with >5 pack-yr, validated by urine nicotine >30 ng/ml and/or cotinine >50 ng/ml
- Normal FEV1 (≥80% predicted), FVC (≥80% predicted), FEV1/FVC (≥0.7) based on post-bronchodilator spirometry, TLC (≥80% predicted)
- DLCO <80% predicted and below the 95% range of normal DLCO calculated for each individual separately based on gender, age and height

Exclusion criteria

- Unable to meet the inclusion criteria
- Alcohol or drug abuse within the past 6 months
- Evidence of malignancy within the past 5 years
- Current active infection or acute illness of any kind

Inclusion / Exclusion Criteria for the Nonsmoker Dataset Population

Inclusion criteria

- Males and females, at least 18 years old
- Capable of providing informed consent
- Willingness to participate in the study
- Good health without history of chronic lung disease, including asthma, and without recurrent or recent (within 3 months) acute pulmonary disease
- Normal physical examination
- Normal routine laboratory evaluation, including general hematologic studies, general serologic/immunologic studies, general biochemical analyses, and urine analysis
- Normal PA and lateral chest X-ray
- Normal electrocardiogram (sinus bradycardia, premature atrial contractions are permissible)
- Females - not pregnant
- No history of allergies to medications used in the bronchoscopy procedure
- Not taking any medications relevant to lung disease or having an effect on the airway epithelium
- Normal serum α1-antitrypsin level
- HIV1 negative
- Self-reported never-smokers, validated by urine nicotine <20 ng/ml and cotinine <30 ng/ml

- Normal FEV1 (≥80% predicted), FVC (≥80% predicted), FEV1/FVC (≥0.7) based on post-bronchodilator spirometry, TLC (≥80% predicted)

Exclusion criteria

- Unable to meet the inclusion criteria
- Alcohol or drug abuse within the past 6 months
- Evidence of malignancy within the past 5 years
- Current active infection or acute illness of any kind

Screening Assessment

Individuals were recruited using advertisements in newspapers and websites. After written informed consent, individuals were evaluated at the Weill Cornell NIH Clinical and Translational Science Center and at the Department of Genetic Medicine Clinical Research Facility under IRB-approved clinical protocols. All individuals had their medical history taken and had a physical exam, complete blood count, biochemical profile, serum α1-antitrypsin levels, HIV test, urine analysis, chest X-ray, EKG, and pulmonary function tests (PFTs). We excluded HIV positive individuals and those with α1-antitrypsin below normal levels. Smoking status was confirmed by history and urine nicotine and cotinine. A total of 2302 active smokers were screened. After screening, 732 of 2302 (32%) were excluded due to abnormal spirometry or other lung function abnormalities other than low diffusing capacity of the lung for carbon monoxide (DLCO), such as chronic obstructive pulmonary disease (COPD), asthma, restrictive lung disease or lung cancer. Of the remaining 1570 active smokers passing this filter, 397 (17% of the original total individuals) had normal spirometry and normal total lung capcity (TLC) but low DLCO (referred to as the "normal spirometry/low DLCO" group), and 1173 (51% of the original total individuals) had normal spirometry, normal TLC and normal DLCO (referred to as the "normal spirometry/normal DLCO" group).

In addition, 405 healthy nonsmokers, with a similar distribution of age, gender and ethnicity to the study population, were recruited from the general NY area. Their lung function results were used to calculate the 95% normal range of forced expiratory volume in 1 s (FEV1/)/forced vital capcity (FVC) and DLCO % predicted.

Pulmonary Function Tests

Individuals were instructed to refrain from smoking as of the night before the testing. PFTs included spirometry before and after the administration of salbutamol (100 μg, 4 doses)

[1], lung volumes and DLCO (Viasys Healthcare, Yorba Linda, CA). The DLCO test was performed with the individual in the sitting position. After tidal breathing, a non-forced expiratory maneuver to residual volume was performed, followed by rapid inhalation to TLC. After breath holding for ~10 sec, the individual was asked to exhale (non-forced), not exceeding 4 sec. The DLCO maneuver was carried out 2 to 4 times; the average of the best 2 trials was used. As an additional quality control measure, PFTs were performed serially in several volunteers during the course of the study. The 95% confidence interval (±2 standard deviations) for the DLCO was similar to that reported by Hathaway et al [2]. The spirometry and DLCO curves of all PFTs for all individuals were validated based on ATS/ERS guidelines [3]. For DLCO, these included: a stable calculated breath hold for 10±2 sec; no evidence of leaks or Valsalva or Mueller maneuvers in the curves; both inspiration and expiration completed in <4 sec (and sample collection time <3 sec), with appropriate clearance of deadspace volume and proper sampling/analysis of alveolar gas as assessed graphically; inspiratory vital capacity >85% of the largest expiratory forced vital capacity (from spirometry) in 96% of individuals and >80% in 98% of individuals. The DLCO % predicted value was calculated using the Gaensler et al equation [4], and corrected for hemoglobin and carboxyhemoglobin levels using ATS/ETS guidelines [3].

Study Groups and Assessment

Individuals were divided into “normal spirometry/normal DLCO” and “normal spirometry/low DLCO” groups based on their corrected DLCO prediction values. Because the study populations of both the normal and low DLCO groups had similar, but mixed ethnicities (Table I), and because of the lack of definitive, universally accepted correction criteria for DLCO for African-American and other non-European ethnicities [4-6], no correction was made for ethnicity. Instead, in addition to a predicted DLCO of <80%, a criterion of DLCO level below the 95% range of normal DLCO calculated per individual based on sex, age and height [3,7,8] was

required to place a individual in the "normal spirometry/low DLCO" group. Individuals from both groups were randomly contacted by staff not associated with the study with a goal of recruiting approximately 100 individuals total, equally divided between the 2 groups, to return for subsequent PFT assessment. The final group that returned one or more times included 59 with normal spirometry and normal DLCO and 46 with normal spirometry but low DLCO (Table I). On the average, there were more PFTs performed in the low DLCO group (3±2, *vs* normal DLCO 2±1, $p<10^{-3}$) with shorter intervals between PFTs (18±20 months *vs* normal DLCO 33±18 months , $p<10^{-6}$), but there was no difference in the time of follow-up (normal DLCO group 46±21 months *vs* low DLCO group, 41±31 months, $p>0.4$, Table I). The number of PFTs performed, the intervals between them and the follow-up for each individual was dependant of the individual's.

Chest High Resolution Computed Tomography

The percentage of the lung affected by emphsyema was evaluated at baseline in a random subset of the normal spirometry/normal DLCO (n=12) and normal spirometry/low DLCO group (n=15) at attenuation -950 Hounsfield Units (HU) using the EmphylxJ software application (EmphylxJ, Vancouver, BC, Canada) allowing automated quantitative analysis of transverse chest CT scans [9-11].

Statistical Analysis

Comparison of demographic parameters among groups was performed by two-tailed Student's t-test or Chi-square test. Progression to COPD between the 2 groups was assessed by Chi-square. A within-between ANOVA test was used to compare lung function at baseline and last visit within the normal spirometry/normal DLCO group and within the normal spirometry/low DLCO group. A 95% normal range for FEV1/FVC and DLCO % predicted was calculated based on the average ±2 standard deviations of 405 healthy nonsmokers. PFT paramteres were convert-

ed using a z-score and compared between the normal spirometry/normal DLCO and normal spirometry/low DLCO groups. To assess if DLCO level can predict the development of COPD, a binomial logistic regression model was implemented in which the response was COPD status ("1"=developing COPD, "0" = not developing COPD). In addition, Leave-one-out cross-validation was performed in order to assess the predictive accuracy. Evaluation and fit of the logistic regression model was performed using the "nnet" and "ROCR" packages in the freely available R software [12,13].

Discussion

Low DLCO in Otherwise Healthy Smokers

Several studies have reported decreased DLCO in smokers with normal spirometry. Assessment of 131 healthy Chinese male smokers with normal spirometry found that 21% had low DLCO [14]. Evaluation of 80 Caucasian cigarette smokers with normal spirometry found that 12.5% had low DLCO [15]. Assessment of 80 healthy male adolescents with normal spirometry revealed that 29 passive and 21 active smokers had a lower DLCO than the 30 neither passive nor active smokers [16]. A study of 1612 individuals found lower DLCO in smokers *vs* non-smokers [17]. A retrospective analysis of 38095 individuals showed that 179 (0.45%) had normal spirometry but low DLCO. Of these, 27 out of 179 had chest CT revealing a combination of emphysema and fibrosis [18].

Supplementary References

1. 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: 319-338

2. Hathaway EH, Tashkin DP, Simmons MS. Intraindividual variability in serial measurements of DLCO and alveolar volume over one year in eight healthy subjects using three independent measuring systems. Am Rev Respir Dis 1989; 140: 1818-1822

3. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, Gustafsson P, Hankinson J, Jensen R, McKay R, Miller MR, Navajas D, Pedersen OF, Pellegrino R, Wanger J. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J 2005; 26: 720-735

4. Gaensler EA, Smith AA. Attachment for automated single breath diffusing capacity measurement. Chest 1973; 63: 136-145

5. Neas LM, Schwartz J. The determinants of pulmonary diffusing capacity in a national sample of U.S. adults. Am J Respir Crit Care Med 1996; 153: 656-664

6. Punjabi NM, Shade D, Patel AM, Wise RA. Measurement variability in single-breath diffusing capacity of the lung. Chest 2003; 123: 1082-1089

7. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. Am Rev Respir Dis 1981; 123: 659-664

8. Knudson RJ, Burrows B, Lebowitz MD. The maximal expiratory flow-volume curve: its use in the detection of ventilatory abnormalities in a population study. Am Rev Respir Dis 1976; 114: 871-879

9. Bae KT, Slone RM, Gierada DS, Yusen RD, Cooper JD. Patients with emphysema: quantitative CT analysis before and after lung volume reduction surgery. Work in progress. Radiology 1997; 203: 705-714

10. Coxson HO, Rogers RM, Whittall KP, D'yachkova Y, Pare PD, Sciurba FC, Hogg JC. A quantification of the lung surface area in emphysema using computed tomography. Am J Respir Crit Care Med 1999; 159: 851-856

11. Madani A, Zanen J, De M, V, Gevenois PA. Pulmonary emphysema: objective quantification at multi-detector row CT-comparison with macroscopic and microscopic morphometry. Radiology 2006; 238: 1036-1043

12. W.N.Venables, B.D.Ripley, (eds). Modern Applied Statistics with S. Springer, New York, 2002

13. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics 2005; 21: 3940-3941

14. Yang, SC. Pulmonary diffusing capacity in normal smoking Chinese. *Journal of the Formosan Medical Association* 92(Suppl 2), S61-S68. 2010. Hong Kong.

15. Clark KD, Wardrobe-Wong N, Elliott JJ, Gill PT, Tait NP, Snashall PD. Cigarette smoke inhalation and lung damage in smoking volunteers. Eur Respir J 1998; 12: 395-399

16. Rizzi M, Sergi M, Andreoli A, Pecis M, Bruschi C, Fanfulla F. Environmental tobacco smoke may induce early lung damage in healthy male adolescents. Chest 2004; 125: 1387-1393

17. Viegi G, Paoletti P, Prediletto R, Di PF, Carrozzi L, Carmignani G, Mammini U, Lebowitz MD, Giuntini C. Carbon monoxide diffusing capacity, other indices of lung function, and respiratory symptoms in a general population sample. Am Rev Respir Dis 1990; 141: 1033-1039

18. Aduen JF, Zisman DA, Mobin SI, Venegas C, Alvarez F, Biewend M, Jolles HI, Keller CA. Retrospective study of pulmonary function tests in patients presenting with isolated reduction in single-breath diffusion capacity: implications for the diagnosis of combined obstructive and restrictive lung disease. Mayo Clin Proc 2007; 82: 48-54

19. Heijdra YF, Pinto-Plata VM, Kenney LA, Rassulo J, Celli BR. Cough and phlegm are important predictors of health status in smokers without COPD. Chest 2002; 121: 1427-1433

20. Fletcher M, Elmes PC, Fairbairn AS, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. Br Med J 1959; 2: 257-266

Supplementary Table I. Comparison of Smokers with Normal Spirometry and Total Lung Capacity but Low DLCO Who Developed COPD *vs* Those who Did Not[1]

Parameter	Did not develop COPD	Developed COPD	p value
Individuals	36	10	
Sex (males/females)	24/12	7/3	>0.9
Age	45 ± 9	49 ± 5	>0.1
Ethnicity (AA/E/H)[2]	29/4/3	8/1/1	>0.8
BMI (kg/m^2)	25 ± 5	23 ± 4	>0.1
Smoking history[3]			
Pack-yr	29 ± 16	31 ± 12	>0.7
Pack per day	1.1 ± 0.7	0.8 ± 0.3	>0.2
Age of smoking initiation	17 ± 4	18 ± 6	>0.9
Urine nicotine (ng/ml)	926 ± 1242	1034 ± 1490	>0.8
Urine cotinine (ng/ml)	1223 ± 950	1562 ±633	>0.2
Cough score[4]	1.9 ± 1.5	1.1 ± 1.2	>0.1
Sputum score[4]	1.4 ± 1.4	0.9 ± 1.0	>0.2
MMRC score	0.6 ± 0.7	0.3 ± 0.5	>0.2
% emphysema[5]	1.2 ± 0.01	3.7 ± 0.05	>0.2
Serology[6]			
α1-antitrypsin (mg/dl)	143 ± 18	151 ± 31	>0.3
ESR (mm/hr)	12 ± 9	12 ± 14	>0.9
IgE (IU/mL)	180 ± 283	123 ± 122	>0.5
CrP (mg/dL)	0.4 ± 0.2	0.2 ± 0.2	>0.05
Hepatitis C (negative/positive) [7]	32/3	7/3	>0.2
Lung function[8]			
VC (% predicted)	108 ± 15	109 ± 13	>0.9
FVC (% predicted)	106 ± 15	109 ±12	>0.5
FEV1 (% predicted)	105 ± 15	101 ± 10	>0.4
FEV1/FVC (% observed)	80 ± 4	75 ± 3	<0.003
TLC (% predicted)	93 ± 12	97 ± 19	<0.3
RV (% predicted)	86 ± 35	100 ± 44	>0.3
RV/TLC (% predicted)	30 ± 11	33 ± 10	>0.5
DLCO (% predicted)	69 ± 8	66 ± 11	>0.2
DLCO/VA (mL/mHg/min/L)	3.7 ± 0.6	3.3 ± 0.8	>0.05
Assessment over time (mean ± SD, range)			
Time of follow-up (months)	37 ± 30 (5-146)	54 ± 32 (17-133)	>0.1
Number of PFTs (months)	3 ± 2 (2-8)	3 ± 1 (2-6)	>0.8
Interval between PFTs (months)	17 ± 19 (1-127)	23 ± 23 (6-97)	>0.1

[1] A total of 46 active smokers with normal spirometry/low diffusion capacity (DLCO) were followed for 41±31 months with serial PFTs. Of these individiduals, 10 developed COPD by the GOLD criteria and 36 did not (Figure 3, Table II). The table compares the baseline characteristics of these 2 subgroups and the timing of their assessment.

[2] AA – African-American; E - European; H - Hispanic.

[3] Current smoking was verified at baseline by urine nicotine and its derivative cotinine; at subsequent visits for lung function testing, active smoking status was verified by questionnaire.

[4] Cough and sputum scores were each evaluated on a scale of 0-4: 0 = not at all; 1 = only with chest infections; 2 = a few days a month; 3 = several days a wk; 4 - most days a wk [19]. MMRC = Modified Medical Research Council dyspnoea scale [20].

[5] Chest high resolution computed tomography (HRCT); % emphysema at -950 Hounsfield Units (HU).

[6] All individuals tested negative for HIV and had normal levels of α1-antitrypsin; ESR - erythrocyte sedimentation rate; IgE – immunoglobin E; CrP – C-reactive protein; hepatitis C – hepatitis C serology.

[7] Data is only available for 35 of 36 low DLCO individuals who did not develop COPD.

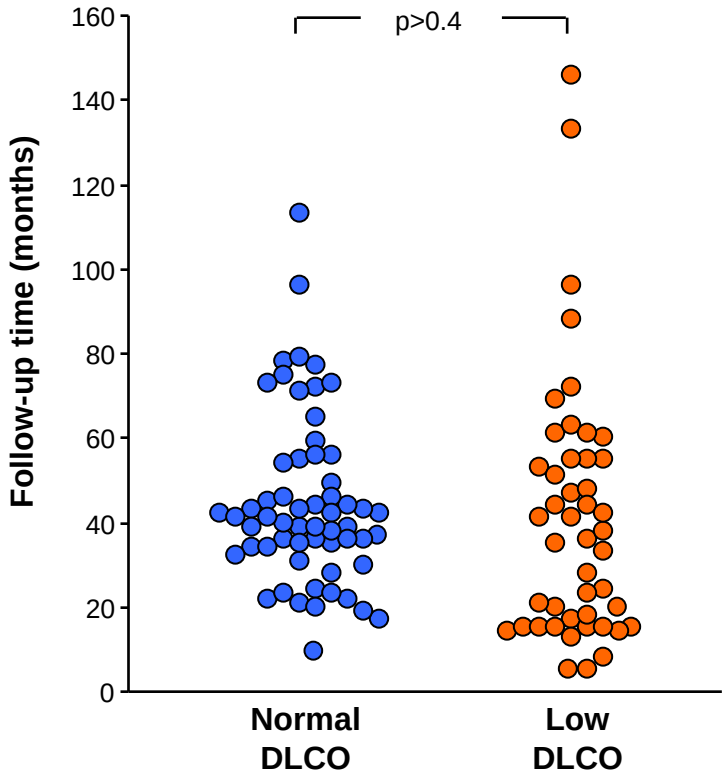
[8] Lung function parameters are presented as percent predicted except the FEV1/FVC ratio, which is presented as percent observed; VC – vital capacity; FVC - forced vital capacity; FEV1 - forced expiratory volume in 1 second; TLC - total lung capacity; RV - residual volume; DLCO - diffusion capacity; and VA – alveolar volume. The DLCO was corrected for hemoglobin and carboxyhemoglobin.[3]

Supplementary Figure Legends

Supplementary Figure 1. Total number of months each individual was followed, comparing active smokers with normal spirometry and normal diffusion capacity of the lung for carbon monoxide (DLCO) *vs* active smokers with normal spirometry but low DLCO ($p>0.4$).

Supplementary Figure 2. Percent emphysema (calculated in -950 Hu) in a subset of the active smokers with normal spirometry and normal diffusion capcity of the lung for carbon monoxide (DLCO) *vs* active smokers with normal spirometry but low DLCO ($p>0.8$).

Supplementary Figure 1



Supplementary Figure 2