

Both cohorts were recruited at the Center for Clinical Studies, Research Center Borstel, Germany. Cohort 1 (“Borstel Healthy Cohort”) was recruited from February 2016 to January 2017. For Cohort 2 (“LuMEEn cohort”) recruiting started in March 2017 and finished in July 2020.

The two cohorts represent a cross-section from the same population defined by identical inclusion and exclusion criteria: Caucasian adult individuals (≥ 18 years) were eligible if they had no evidence of (1) a respiratory infection or systemic glucocorticoid therapy in the month before enrolment, (2) antibiotic therapy in the last two months, (3) diabetes mellitus, (4) pregnancy or lactation, (5) active tuberculosis or history thereof, (6) immunosuppression, (7) known pulmonary disease (except for COPD stage GOLD I/II). Participants were excluded if they had a medical contraindication for bronchoscopy as judged by the investigator (e.g. allergy to sedative medication).

The demographic data of both cohorts is listed in Table S1.1. Medical, environmental and socioeconomic data were assessed by a standardised questionnaire. The study was approved by the independent Ethics Committee of the University of Luebeck (cohort 1, ref.15-194; cohort 2, ref. 16-145). Oral and written informed consent from all study subjects was obtained according to ICH/GCP standards. Cohort 1 is registered at the German Register for Clinical Trials (DRKS00016932). Cohort 2 is registered at clinicaltrials.gov (NCT03562442).

Participants were assigned either to smokers, ex-smokers or never-smokers, according to their smoking history as defined by the Centers for Disease Control and Prevention (https://www.cdc.gov/nchs/nhis/tobacco/tobacco_glossary.htm). Smokers were further classified. Long-term smokers were defined by ≥ 10 cigarettes per day and ≥ 10 packyears. Short-term smokers differed from long-term smokers by a shorter smoking history: ≥ 10 daily cigarettes, but < 10 packyears. Mild smokers consumed less than 10 daily cigarettes and had < 5 packyears.

The study group included 24 female and 34 male subjects (n=58). Gender, age and BMI were similar across smokers and never-smokers (n=58; Figure S1). Ex-smokers were significantly older compared to the other groups and had stopped smoking for 19 years (median; Table S1.2). Allocation of smokers to individual groups of smoking histories accorded with levels of urine cotinine and anabiasine (Table S1.2). Concentrations were determined from unprocessed urine samples by chromatography (LC)-tandem mass spectrometry.

SOP-guided bio-sampling and processing included deep bilateral nasal swabs (; n=46), bilateral oropharyngeal swabs (; n=54) and bronchoalveolar lavage (BAL; n=52). Oropharyngeal and nasal samples (Mast swab TM, Mast Group Ltd., UK) were frozen in cryotubes on dry ice immediately after sampling. Unused sterile swabs were included for microbiome analyses as controls per lot. Participants underwent flexible bronchoscopy according to current German guidelines.²⁹ The bronchoscope was wedged into a sub-segmental bronchus of the middle lobe. BAL was performed in 15 x 20ml fractions up to a total volume of 300ml using sterile saline solution (0.9%). Sterile saline solution was added as control for the microbiome analyses. Due to possible contamination, the recovery of the first BAL fraction was discarded. The remaining fractions were pooled and frozen at -80°C.

Table S 1.1: Summary of baseline characteristics of the two pooled cohorts.

| | Cohort 1 (Borstel healthy cohort) | Cohort 2 (LuMEEn cohort) | Pooled cohorts |
|---|--|---------------------------------|-----------------------|
| Number, n | 36 | 22 | 58 |
| Age in years, median (IQR) | 26.5 (23.3-39) | 29 (24-39) | 28 (24-39) |
| Male, n (%) | 24 (66.7) | 10 (45.5) | 34 (58.6) |
| Body mass index [kg/m²], median (IQR) | 24.2 (22.2-26.8) | 24.8 (21.8-26.5) | 24.5 (21.8-26.5) |
| Never smokers, n (%) | 4 (13.9) | 11 (50) | 15 (25.8) |
| Smokers, n (%) | 25 (69.4) | 11 (50) | 36 (62.1) |
| Smoking duration in years, median (IQR) | 8 (3-17) | 15.5 (7.3-18.3) | 10 (4.8-17.5) |
| Current cigarettes per day, median (IQR) | 12 (4-18) | 15.5 (13.3-21.5) | 14.5 (7.8-19.3) |
| Maximum number of cigarettes per day, median (IQR) | 20 (15-30) | 30 (21.3-40) | 20 (15.5-30) |
| Cumulative packyears, median (IQR) | 5 (3-12) | 15 (6.3-19.3) | 7 (4.5-16.5) |
| Ex-Smokers, n (%) | 6 (16.7) | - | 6 (10.3) |
| Smoking duration in years, median (IQR) | 19 (16-20) | - | 19 (16-20) |
| Maximum number of cigarettes per day, median (IQR) | 30 (7-30) | - | 30 (7-30) |
| Cumulative packyears, median (IQR) | 5 (3-25) | - | 5 (3-25) |

Data are presented as n, n (%) and median (interquartile range).

Table S1.2: Participant characteristics stratified by smoking history

| Study group | Smokers | | | | Never-smokers | Ex-smokers | p-val ¹ | p-val | p-val |
|--|------------------|--------------------|------------------|------------------|------------------|------------------|--------------------|--------------------|-------------------|
| | Total | Long-term | Short-term | Mild | Total | Total | NS-S ² | NS-ES ³ | ES-S ⁴ |
| Number, n | 36 | 15 | 9 | 12 | 16 | 6 | | | |
| Male, n (%) | 23 (63.9) | 9 (60) | 5 (55.6) | 9 (75) | 6 (37.5) | 5 (83.3) | 0.77 | 0.06 | 0.634 |
| Age in years, median (IQR) | 27.5 (23-34) | 33 (28-49.5) | 21 (19-28) | 26 (22.8-26) | 26 (24-39) | 59 (49.5-63.5) | 0.46 | 0.0002 | 0.00004 |
| Body mass index [kg/m ²], median (IQR) | 24.2 (21.5-26.5) | 26.7 (24.7-29.5) | 21.5 (21.2-23.7) | 24.7 (22.5-25.4) | 24.3 (20.6-26.2) | 24.9 (23.9-26.3) | 0.409 | 0.495 | 0.444 |
| Current cigarettes per day, median (IQR) | 14.5 (8-19) | 16.3 (13.5-20) | 14.5 (9.5-17.9) | 4 (2-7) | - | - | - | - | |
| Maximum cigarettes per day, median (IQR) | 20 (16-30) | 30 (22-37.5) | 20 (17-30) | 10 (8-15) | - | 30 (7-30) | - | - | 0.441 |
| Packyears, median (IQR) | 7 (5-17) | 18.5 (14.3-31.5) | 5 (5-6.3) | 1 (0.8-4) | - | 4.5 (3-25) | - | - | 0.471 |
| Years of smoking, median (IQR) | 10 (5-18) | 17 (13-21.5) | 5 (5-8) | 8 (0.8-11.5) | - | 14.5 (4.5-24) | - | - | 0.466 |
| Years since smoking cessation, median (IQR) | | | | | | 19 (18-23) | | | |
| Nicotine serum levels [µg/l], median (IQR) | 500 (60-643) | 500 (458-719.5) | 482 (247-1381) | 24 (5-284.5) | - | - | - | - | - |
| Cotinine urine levels [µg/l], median (IQR) | 985 (341-1645) | 1333 (1102-1666) | 1041 (500-1880) | 211 (45.8-756) | - | - | - | - | - |
| 3-hydroxy-cotinine urine levels [µg/l], median (IQR) | 2960 (938-7796) | 4354 (1982-7184.5) | 7804 (6780-8218) | 534 (144-1128) | - | - | - | - | - |
| Anabasine urine levels [µg/l], median (IQR) | 8 (0-17) | 11 (6-14) | 15 (10-21) | 0 (0-7.5) | - | - | - | - | - |

Data are presented as n, n (%) and median (interquartile range). ¹p-values are based on Chi-square tests for proportions, and t-tests for independent means of numerical variables. ²Never-smokers vs. smokers (total). ³Never-smokers vs. ex-smokers. ⁴Ex-smokers vs. smokers (total)

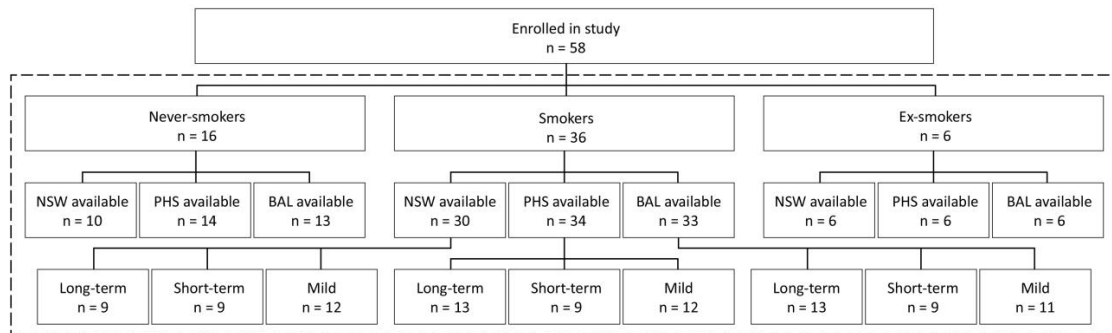


Figure S1: Schematic overview on the structure of the two cohorts and the samples analyzed