**Supplementary materials**

**GLUCOLD**

Samples were collected and processed as previously described 1,2. Methods for the processed for RNA-Seq are previously described3

**INDURAIN**

Samples were collected and processed for RNA-Seq and genome wide methylation as previously described 3.

**NORM**

The Study to Obtain Normal Values of Inflammatory Variables From Healthy Subjects (NORM; NCT00848406) included healthy individuals. Bronchial and nasal brushings were collected at the same time, using a Cellebrity bronchial brush (Boston Scientific, Marlborough, Mass) or a Cyto-Pak CytoSoft nasal brush (Medical Packaging Corporation, Camarillo, Calif),for genome wide gene expression profiling using microarrays (Affymetrix Hugene\_ST1.0 arrays). The methods for RNA extraction, labelling and microarray processing, have previously been published. To determine is COV-19 related genes were differentially expressed between nasal and bronchial brushes, we ran a paired analysis, using limma correcting for age and sex. We next investigated whether there was an association with expression and age, sex, BMI or age.

Acute smoking cohort

The study is registered at clinicaltrials.gov under the National Clinical Trial (NCT) identifier: NCT00850863. Samples were collected and processed for microarray analysis as previously described 1.

**PIAMA**

Analysis of adolescents was performed within the PIAMA (Prevention and Incidence of Asthma and Mite Allergy) birth cohort study 5. Nasal brushes were collected from the lower inferior turbinate of adolescents at 16 years old. Genome-wide DNA methylation was determined with use of Infinium HumanMethylation450 BeadChips (Illumina, San Diego, Calif). β-values were transformed to M-values as log2(β/(1-β)); M-values were used in downstream analysis.RNA-seq was performed on the Illumina HiSeq2500 platform, raw count data were transformed to log2CPM using voom in the limma package, and used in the downstream analysis. Extreme outliers in the gene expression and methylation data were identified using the Tukey method (<1st quartile−3 × IQR; >3rd quartile+3 × IQR) and set as missing.The methods for DNA and RNA extraction and quality control of DNA methylation and RNAseq have been published previously 6. Genotype imputation was performed on Michigan sever with the reference panel of HRC r1.1 2016 (GRCH37/hg19). Dosage is used for the downstream analysis. In total, 325 samples have RNAseq data, with 291 samples with all covariates available. Besides, 303 samples have matched RNAseq and genotype data that used for eQRL analysis; 245 samples have matched RNAseq and DNA methylation data that used for eQTM analysis.

**Finnish Biopsy study**

Full-term infants who had been referred to Helsinki University Central Hospital for investigation of severe recurrent respiratory symptoms, including dyspnea, cough, and wheeze, and were symptomatic for at least 4 weeks, underwent a bronchoscopy as part of their clinical work-up. Bronchoscopy with endobronchial biopsy was performed under general anesthesia, with a 3.5-mm rigid bronchoscope (Karl Storz GmbH & Co, Tuttingen, Germany), using biopsy forceps (No. 10378L; Karl Storz GmbH & Co). For RNA isolation, samples were homogenized in Eurozol (EuroClone, Siziano, Italy) or TRIsure (Bioline, London, UK) using the Ultra-Turrax T10 (IKA Labortechnik, Staufen, Germany). Glycogen (Roche Diagnostics GmbH, Mannheim, Germany) was used as a carrier molecule during RNA isolation (20 µg / sample). Total RNA extraction was performed according to the manufacturer’s instructions (Eurozol / TRIsure), and RNA was dissolved in 25 µl of diethylpyrocarbonate (DEPC) treated water and stored at -70°C.

Initial quality control (QC) and RNA quantification of the samples was performed by capillary electrophoresis using the LabChip GX (Perkin Elmer, Waltham, MA, USA). PolyA amplification was used to select RNA over DNA and account for DNA contamination that was found during initial QC.

Sequence libraries were generated using the TruSeq RNA sample preparation kits (Illumina, San Diego, CA, USA) using the Sciclone NGS Liquid Handler (Perkin Elmer). In case of contamination of adapter duplexes an extra purification of the libraries was performed with the automated agarose gel separation system Labchip XT (Perkin Elmer). The obtained cDNA fragment libraries were sequenced on an Illumina HiSeq2500 using default parameters (paired end 2x150bp) in pools of multiple samples.

The trimmed fastQ files were aligned to build b37 human reference genome using HISAT (version 0.1.5) allowing for 2 mismatches. SAMtools (version 1.2) was used to sort the aligned reads. The gene level quantification was performed by HTSeq (version 0.6.1p1).

QC metrics were calculated for the raw sequencing data with FastQC and Picard tools.

Raw count data were transformed to gene expression levels (TPM and variance stabilised) using DeSeq2, and used in the downstream analysis.

**CRUKPAP**

487 donors were recruited into the CRUKPAP cohort at Royal Papworth Hospital, Cambridge (UK), including 114 healthy volunteers (HV) and 337 patients being investigated for suspicion of lung cancer. Participants were stratified into smoking cessation categories as follows: 45 never smokers (NV), 289 former smokers (FS) and 153 current smokers (CS). Former smokers were further divided into categories: > 1 year after cessation (FS1, n=234), 1-12 months after cessation (FS2, n=45) and < 1 month after cessation (FS3, n=10). Smoking history measured in pack years was recorded and stratified into four categories: ‘none’ (PY1), < 10 years (PY2), 10-30 years (PY3) and > 30 years (PY4). For suspected lung cancer patients, both COPD status and final cancer diagnosis (lung cancer / no lung cancer) were recorded.

From these donors 413 nasal epithelial curettages were collected using Arlington Scientific ASI Rhino-pro nasal curettes. Briefly, the nostril is opened with a nasal speculum to identify the inferior turbinate. Under direct vision the tip of the nasal curette is gently scraped over the turbinate to obtain a 'peel or curl' of epithelial tissue. The curl of tissue is then removed by flicking the curette while the tip is submerged in RNAlater™ collection medium and presence of the curl floating in the medium is confirmed by visual inspection. This procedure is repeated twice for each nostril per donor. RNA integrity (RIN) was checked for all samples and we found >80 % of samples to have RIN 6 or better.

Additionally, bronchial brushings were collected using Olympus 2.0mm bronchoscopy cytology brushes (BC202D-2010) from 236 patients undergoing flexible bronchoscopy as part of suspected lung cancer investigations. To retrieve bronchial samples the right or left main bronchus or a segmental bronchus (upper, middle or lower lobe) on the contralateral side to the lesion under investigation was brushed gently to lift off epithelial cells while minimising blood contaminants. For 162 donors, both nasal and bronchial samples were available. All samples underwent short-read total RNA sequencing using Illumina TruSeq library generation for the Illumina HiSeq 2500 platform. Blood samples were taken from 414 donors and germline genotyped using the Illumina Infinium Oncoarray platform at 450K tagging germline variants. Phasing and imputation to 1000 Genomes phase 3 yielded a total of 8.9mio germline variants. Total gene expression levels (TPM and variance stabilised) were determined for 18,072 protein coding genes for all samples using DeSeq2.

**INCI**

79 participants were prospectively recruited into the INER-Ciencias Mexican Lung Program including 19 younger adults and 60 older adults. Nasal epithelial curettage samples were collected using a previously published method. Libraries were prepared for RNA-sequencing using poly(A) enrichment (NEB Next RNA Ultra kit) at the RNA-Seq Center . RIN scores were checked and were 7 or greater for all samples. Sequencing was performed at the NUSeq Core Facility (Northwestern University Center for Genetic Medicine) on the Illumina HiSeq 4000 platform. Reads were aligned with STAR to the human genome (GRCh38.77), and HTSeq was used to generate a table of raw counts.

**Cellular deconvolution**

Several computational methods have been proposed in the literature to infer cell-type proportions from bulk RNA samples. Their performance with noisy reference profiles highly depends on the set of genes used for deconvolution2,3. In this study, we used AutoGeneS that automatically extracts informative genes and outperforms other methods for analyzing bulk RNA samples with closely correlated cell types and noisy single-cell reference profiles4. The number of informative genes was manually set to 200, 300, and 400 genes and we selected the one (n=400) with the most stable results across cohorts. We also performed tissue-specific deconvolution for nasal and bronchial samples using nasal and bronchial reference profiles, respectively. However, we did not notice a significant difference in the inferred proportions (results not shown). Similarly, the proportions of major cell types--goblet, secretory, and ciliated cells--were consistent using different cell type resolutions.

scRNA-Seq data from bronchial biopsies and nasal brush were merged. Two samples from bronchial data were excluded because of their high batch effect. From a total of 24,363 cells, cell types with low frequency (n<100 cells) were filtered out (number of remaining cells=18,886). Due to highly similar gene expression profiles, the scRNA-Seq signatures from the club and the 2 goblet cell clusters were combined to generate a uniform scRNA-Seq signature of secretory cells. The merged scRNA-seq count data was normalized to count per million (CPM) and highly variable (HV) genes (n=5,000) were selected. We used the method implemented in single-cell analysis in Python (SCANPY)5 for selecting HV genes

in which genes are binned by their mean expression and those with the highest

variance-to-mean ratio are selected as HV genes in each bin. We then performed

AutoGeneS4 to filter 400 informative genes from the highly variable ones that differentiated the cell types. The informative genes minimized correlation and maximized distance between the clusters in the single-cell reference data. For visualisation, single-cell neighbourhood graph (kNN-graph) was computed on the first 50 principal components using 30 neighbours and low-dimensional uniform manifold approximation and projection

(UMAP) embedding was used6. Bulk deconvolution was then conducted on all bulk samples using support vector regression (SVR) method2 for samples measured by both RNA-Seq and microarray23. Resulting deconvolution predicted cell proportions were then compared within each dataset to the expression of ACE2, TMPRSS2, BSG, FURIN, NRP1 and CTSL using spearman correlations.

**Supplementary Tables**

**Table S1 Influence of sex on COVID-19 related genes in matched nasal and bronchial brushes**

**Table S2 influence of age on COVID-19 related genes in matched nasal and bronchial brushes**

**Table S3 Influence of acute smoking on COVID-19 related genes in bronchial brushes (n=63)**

**Table S4. EQTL analysis in COVID-19 related genes**

**Table S5. EQTM analysis in COVID-19 related genes**

**Table S6. Relationship with eQTM CpG sites and clinical variables in nasal brushes.**

**Table S7. Relationship with eQTM CpG sites and clinical variables in bronchial biopsies.**

**Supplementary Figures**

**Figure S1. Cellular deconvolutin of airway brushings focusing on only bronchial sectory cell signature.** Statistics for deconvolution results were conducted using Mann-Whitney test. \*=p<0.05

**Referencing**

1 Billatos E, Faiz A, Gesthalter Y, et al. Impact of acute exposure to cigarette smoke on airway gene expression. Physiol Genomics 2018; 50: 705–13.

2 Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019; 37: 773–82.

3 Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 2019; 10: 380.

4 Aliee H, Theis F. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. bioRxiv 2020. https://www.biorxiv.org/content/10.1101/2020.02.21.940650v1.abstract.

5 Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018; 19: 15.

6 McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML]. 2018; published online Feb 9. http://arxiv.org/abs/1802.03426.