Supplementary Materials and Methods

DNA / RNA extraction

Bronchial biopsies were taken from segmental divisions of the main bronchi and immediately frozen in Tissue-Tek (VWR, Radnor, PA) at -80°C. After thawing at room temperature, biopsies were cut from the blocks when they were semi-solid. Samples were lysed in 600 µl RLT-plus using a IKA Ultra Turrax T10 Homogenizer. Total genomic DNA was extracted using AllPrep DNA/RNA/miRNA Universal Kit, according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). DNA samples were dissolved in 30 µl elution buffer. Concentrations of DNA were measured in 1 µl of fluid using a Nanodrop-1000 and in 2 µl of fluid run on a Labchip GX (Perkin Elmer, Waltham, MA).

DNA methylation: Sample preparation and Hybridization protocol

DNA samples were purified by a precipitation step, rinsed with ethanol (70%) and dissolved in elution buffer. Then, samples were treated with sodium bisulfite to convert unmethylated cytosine bases into uracil. This was followed by a PCR-free whole genome amplification, after which the treated DNA was hybridized to the Infinium HumanMethylation450 BeadChip array (450k array). After hybridization, allele-specific single-base extension incorporated a fluorescent label for detection of methylated and unmethylated sites. The conversion and hybridization protocol was performed according to the manufacturer's instructions (Illumina, San Diego, CA).
Every beadchip was run with a control sample of blood DNA from a single female, which was used to assess the efficiency of the normalization procedure by verifying that the among sample variance was minimized. We used a randomized block design, where samples were assigned to blocks based on subject group, smoking status and gender, and randomized within blocks.

DNA methylation: Quality control, sample filtering, probe filtering and normalization

Raw intensity values were read from IDAT-files, and converted to beta values using the minfi-package [1]. We identified bad quality samples using the R-package MethylAid, based on five diagnostic filter variables with the following thresholds; MU = 10.5, OP = 11.75, BS = 12.5, HC = 13.25, and DP = 0.95 [2]. Genotypes from 65 SNP probes on the 450k array were compared to previously acquired genotypes from blood samples and discordant samples were discarded as swapped or contaminated samples and excluded from the analysis.
We excluded probes where >1% of samples had intensities indistinguishable from background levels at $P = 0.05$, probes which were crossreactive [following 3], probes which have SNPs with MAF > 10% at the interrogation or extension site, and type I probes which displayed high intensity

signals [following 4]. In addition, we excluded all sex linked probes. The probe filtering procedure resulted in the removal of 48,059 unreliable probes (10 %). After probe-and sample-filtering, raw beta-values were background-corrected and normalized using the dasen method from the wateRmelon-package [5].

DNA methylation: Correction for cell type composition

We investigated how corrections for cell type composition affected the results for a subset of differentially methylated CpG sites. We estimated surrogate variables from our DNA methylation levels as implemented in the sva package and added those to our basic model as covariates [6]. This is a reasonable alternative to deconvolution using reference cell type profiles [7]. We reran the linear modelling analysis and monitored how this impacted the significance of the differential methylation in the contrasts.

RNA sequencing: Sample preparation and sequencing

RNA samples were further processed using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina, San Diego, CA), using an automated procedure in a Caliper Sciclone NGS Workstation (PerkinElmer, Waltham, MA). In this procedure, all cytoplasmic and mitochondria rRNA was removed (RiboZero Gold kit). The obtained cDNA fragment libraries were loaded in pools of multiple samples unto an Illumina HiSeq2500 sequencer using default parameters for paired-end sequencing (2 × 100 bp).

RNA sequencing: Gene expression quantification

The trimmed fastQ files where aligned to build b37 of the human reference genome using HISAT (version 0.1.5) allowing for 2 mismatches [8]. Before gene quantification SAMtools (version 1.2) was used to sort the aligned reads [9]. The gene level quantification was performed by HTSeq (version 0.6.1p1) using Ensembl version 75 as gene annotation database [10].

RNA sequencing: Quality Control

Quality control (QC) metrics were calculated for the raw sequencing data, using the FastQC tool (version 0.11.3) [11]. Alignments of 220 subjects were obtained. QC metrics were calculated for the aligned reads using Picard-tools (version 1.130) (URL http://picard.sourceforge.net) CollectRnaSeqMetrics, MarkDuplicates, CollectInsertSize-Metrics and SAMtools flagstat. We discarded 36 samples due to poor alignment metrics. In addition, we checked for concordance between sexlinked (XIST and Y-chromosomal genes) gene expression and reported sex. All samples were concordant. This resulted in high quality RNAseq data from 184 subjects.

References

1. Aryee MJ, Jaffe AE, Corrada-bravo H, Ladd-acosta C, Feinberg AP, Hansen KD, Irizarry RA. Genome analysis Minfi : a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. 2014; 30: 1363–1369.

2. Iterson van M, Tobi EW, Slieker RC, Hollander den W, Luijk R, Slagboom PE, Heijmans BT. MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics* 2014; 30: 3435–3437.

3. Chen Y, Lemire M, Choufani S, Butcher DT, Zanke BW, Gallinger S, Hudson TJ, Lemire M, Choufani S, Butcher DT, Zanke BW, Gallinger S, Hudson TJ, Weksberg R, Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. 2013; 2294.

4. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief. Bioinform.* 2014; 15: 929–941.

5. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013; 14: 293.

6. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; 28: 882–883.

7. Teschendorff AE, Zheng SC. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* 2017; 9: 757–768.

8. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 2015; 12: 357–360.

9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078–2079.

10. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; 31: 166–169.

11. Andrews S. FastQC: A quality control tool for high throughput sequence data. Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/ 2010. p. http://www.bioinformatics.babraham.ac.uk/projects/.