UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

II. Medizinische Klinik und Poliklinik
(Onkologie, Hämatologie, Knochenmarktransplantation mit Abteilung für Pneumologie)

Direktor:
Prof. Dr. med. Carsten Bokemeyer

Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Espen Elias Groth
aus Eutin

Hamburg 2020

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: **06.05.2021**

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: **Prof. Dr. Klaus Pantel**

Prüfungsausschuss, zweite/r Gutachter/in: **Prof. Dr. Carsten Bokemeyer**

Meinen Eltern

Inhaltsverzeichnis

1 Artikel

Groth *et al. Respir Res (2020) 21:274*
https://doi.org/10.1186/s12931-020-01544-4

Respiratory Research

RESEARCH Open Access

Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0

Espen E. Groth[1,2,3,4*], Melanie Weber[5], Thomas Bahmer[1,2,3], Frauke Pedersen[1,2,6], Anne Kirsten[2,6], Daniela Börnigen[7], Klaus F. Rabe[1,2], Henrik Watz[2,6], Ole Ammerpohl[2,8†] and Torsten Goldmann[2,9†]

Abstract

Background: To date, most studies involving high-throughput analyses of sputum in asthma and COPD have focused on identifying transcriptomic signatures of disease. No whole-genome methylation analysis of sputum cells has been performed yet. In this context, the highly variable cellular composition of sputum has potential to confound the molecular analyses.

Methods: Whole-genome transcription (Agilent Human 4 × 44 k array) and methylation (Illumina 450 k BeadChip) analyses were performed on sputum samples of 9 asthmatics, 10 healthy and 10 COPD subjects. RNA integrity was checked by capillary electrophoresis and used to correct in silico for bias conferred by RNA degradation during biobank sample storage. Estimates of cell type-specific molecular profiles were derived via regression by quadratic programming based on sputum differential cell counts. All analyses were conducted using the open-source R/Bioconductor software framework.

Results: A linear regression step was found to perform well in removing RNA degradation-related bias among the main principal components of the gene expression data, increasing the number of genes detectable as differentially expressed in asthma and COPD sputa (compared to controls). We observed a strong influence of the cellular composition on the results of mixed-cell sputum analyses. Exemplarily, upregulated genes derived from mixed-cell data in asthma were dominated by genes predominantly expressed in eosinophils after deconvolution. The deconvolution, however, allowed to perform differential expression and methylation analyses on the level of individual cell types and, though we only analyzed a limited number of biological replicates, was found to provide good estimates compared to previously published data about gene expression in lung eosinophils in asthma. Analysis of the sputum methylome indicated presence of differential methylation in genomic regions of interest, e.g. mapping to a number of human leukocyte antigen (HLA) genes related to both major histocompatibility complex (MHC) class I and II molecules in asthma and COPD macrophages. Furthermore, we found the SMAD3 (SMAD family member 3) gene, among others, to lie within differentially methylated regions which has been previously reported in the context of asthma.

*Correspondence: e.groth@lungenclinic.de
†Ole Ammerpohl and Torsten Goldmann contributed equally to this work
[1] LungenClinic Grosshansdorf, Großhansdorf, Germany
Full list of author information is available at the end of the article

Conclusions: In this methodology-oriented study, we show that methylation profiling can be easily integrated into sputum analysis workflows and exhibits a strong potential to contribute to the profiling and understanding of pulmonary inflammation. Wherever RNA degradation is of concern, in silico correction can be effective in improving both sensitivity and specificity of downstream analyses. We suggest that deconvolution methods should be integrated in sputum omics analysis workflows whenever possible in order to facilitate the unbiased discovery and interpretation of molecular patterns of inflammation.

Keywords: Sputum, Omics, Transcriptome, Methylome, Deconvolution, RNA, Degradation, Biobanking, Asthma, COPD

Background

Respiratory research has greatly benefited from the application of molecular high-throughput ("omics") technologies [1, 2]. Significant contributions could be made to the understanding of chronic-inflammatory respiratory disease, ranging from phenotyping and classifying disease to modeling therapy responses [3–7]. Amongst other materials, induced sputum has proven to be very valuable for the molecular profiling of both bronchial asthma and chronic obstructive pulmonary disease (COPD) [3, 6, 8–11]. With growing availability of computational infrastructure and analysis platforms, multi-omics approaches gained attractivity and have been applied successfully [2, 12–14]. In this context, epigenetic analyses, such as DNA methylation profiling, have contributed to the molecular characterization of inflammation in asthma and COPD [15–21]. So far, however, methylation analyses of sputum samples have been limited to subsets of loci, e.g. by the means of methylation-sensitive polymerase chain reaction (PCR), in cancer research [22–24]. The use of whole-genome methylation analyses of sputum for the molecular profiling of asthma or COPD has not been evaluated yet.

Sputum samples contain a mixture of immune cells (mainly alveolar macrophages, neutrophils, eosinophils and lymphocytes), but also contaminating cells (such as ciliated epithelium from the airways and squamous epithelium from the pharyngeal and oral region). The relative abundancy of cell types varies substantially and can be used to distinguish disease subgroups such as "T2 high" from "T2 low" types in asthma by eosinophil counts [25, 26]. This variability, in turn, has potential to confound whole-sputum omics analyses [27]. Previously, methods such as fluorescence activated cell sorting (FACS) [28], gradient centrifugation [29, 30] or selection by cellular adherence [31] have been used to purify certain cell types from blood and bronchoalveolar lavage (BAL) samples in asthma and COPD. Due to the corresponding procedural and financial effort, however, the implementation of such methods becomes complicated in large-scale settings. Apart from physical cell separation, attempts to correct for cellular composition in silico have been made in omics analyses of BAL [32] and blood samples [33]. Recently, a reference-based transcriptomic method thought to be less sensitive to sputum composition bias has been suggested for use in asthma research [10].

However, the aforementioned approaches do not allow to infer cell-type specific molecular profiles from mixed-cell data and, so far, the high-throughput molecular analysis of mixed-cell sputum samples has generally been limited to be used as a molecular "fingerprint" to describe inflammatory processes.

Over the last years, tailored in silico methods, so-called deconvolution algorithms, have been designed to solve the problem of inferring cell type-specific omics profiles from mixed-cell data [34–41]. These methods have been primarily developed and evaluated on blood, brain and cancer data sets but exhibit a strong potential to be of avail for omics analyses of sputum and other respiratory mixed-cell materials.

In this exploratory and methodology-oriented study, we examine the applicability of sputum whole-genome methylation analysis in molecular profiling of asthma and COPD. Furthermore, we provide insight into how in silico RNA quality correction can benefit the transcriptome analysis of sputum samples from long-term storage and, for the first time, apply a deconvolution method based on linear regression by quadratic programming to infer cell type-specific omics profiles from mixed-cell sputum data.

Materials and methods

Sputum samples

Sputum samples were obtained from biomaterial depositories of the German prospective cohort studies ALLIANCE [42] (asthma and controls) and COSYCONET [43] (COPD) at the LungenClinic Grosshansdorf, Germany. To evaluate the applicability of methylation profiling to sputum samples from (long-term) biobank storage, we focused on samples collected during early recruiting periods (11/2013–05/2015).

Asthma samples were selected from 9 subjects representing a phenotype with eosinophilic (type 2)

inflammation and overall good disease control at the time of sample collection based on sputum and blood eosinophil count (≥ 2% eosinophils in sputum differential cell count and/or ≥ 400 eosinophils/µL in blood differential) as well as total asthma control test (ACT) score (≥ 20 points). In addition, subjects had to be non-smokers (smoking cessation > 12 months before sample collection) with a neglectable smoking history (2/9 subjects, maximum of 5 packyears). No subject was treated with biologics, antihistamines or oral corticoids at the time of sputum collection. Out of the 9 subjects, 7 had a proven allergy (pollinosis, food allergy and/or atopic dermatitis). A total of 5 subjects presented with severe asthma (defined as requiring high doses of inhaled corticosteroids with > 500 µg fluticasone equivalent per day) and 4 presented with a mild-to-moderate type (≤ 500 µg per day).

A total of 10 COPD samples was selected from subjects that had not experienced any moderate or severe exacerbation (defined as requiring use of oral corticoids or inpatient hospital treatment) for ≥ 12 months as well as had successfully accomplished smoking cessation for ≥ 12 months. Of the 10 selected subjects, 7 had moderate (GOLD 2) and 3 had severe COPD (GOLD 3).

Healthy controls (10 samples) were defined as subjects without any history of pulmonary or systemic-inflammatory disease, allergies or respiratory tract infection within the last 12 months. None of the selected control subjects had any smoking history. Descriptive statistics can be found in Table 1.

For a graphical representation of the overall workflow, see Fig. 1.

Details about sputum induction and processing are provided in Additional file 1. Differential sputum cell counts (alveolar macrophages, neutrophils, eosinophils, lymphocytes, monocytes, ciliated epithelium and squamous cells) were performed on Diff-Quick-stained slides by two independent evaluators, each of whom evaluated a total of 400 cells per slide.

Samples were either stored in RLT Plus extraction buffer (proprietary buffer by Qiagen, Hilden, Germany) at − 80 °C [9], or preserved via the HOPE technique (Hepes-glutamic acid buffer-mediated organic solvent protection effect) by incubation with HOPE medium (DCS innovative diagnostic systems, Hamburg, Germany), followed by embedding in low-melting paraffin and subsequent storage at 4 °C [44]. The HOPE technique is a preservation technique originally developed for tissue samples in pathology diagnostics and research to allow for a variety of processing and analysis protocols without the constraints imposed by conventional formalin fixation [45]. Previous studies have demonstrated that HOPE-preserved material can successfully be processed to retrieve nucleic acids suitable for omics analysis [45, 46] and that the technique can be transferred to sputum samples [44] as well as bronchoalveolar lavage fluid [47].

Extraction of nucleic acids

From HOPE-preserved, paraffin embedded samples, sputum cells were extracted by cutting slices on a microtome (using alcohol- and heat-sterilized, RNase-free blades) which were deparaffinized subsequently by incubation with xylene (2 × 10 min) and ethanol (2 × 10 min), followed by a drying step using a vacuum centrifuge before addition of RLT Plus lysis buffer [44, 47]. Sputum samples stored in RLT buffer were thawed on ice. DNA and RNA were simultaneously extracted using the AllPrep Micro Kit (Qiagen) following the manufacturer's instructions. Total DNA and RNA yield were measured on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity was determined on a BioAnalyzer system (Agilent Technologies, Waldbronn, Germany). For optimal electrophoresis resolution, the RNA 6000 Pico Kit (Agilent) was used after adjusting aliquots of RNA extracts to the maximum input RNA concentration. Samples with RIN < 3 (RNA Integrity Number) were excluded from further processing for microarray analysis which applied to a total of 4 HOPE-preserved samples, limiting the total number of

Table 1 Descriptive statistics of study subjects

	Age (years) Mean ± SD (min/max)	Gender (male/female)	Smoking history (PY) Mean ± SD (min/max)	Daily ICS (µg FE) Mean ± SD (min/max)
Asthma n = 9	59 ± 14 (35/76)	6/3	1 ± 2 (0/5)	511 ± 388 (0/1000)
COPD n = 10	68 ± 10 (44/77)	9/1	38 ± 21 (15/80)	235 ± 206 (0/500)
Controls n = 10	44 ± 20 (19/76)	7/3	0	0

PY pack years, *ICS* inhaled corticosteroid, *FE* fluticasone equivalent

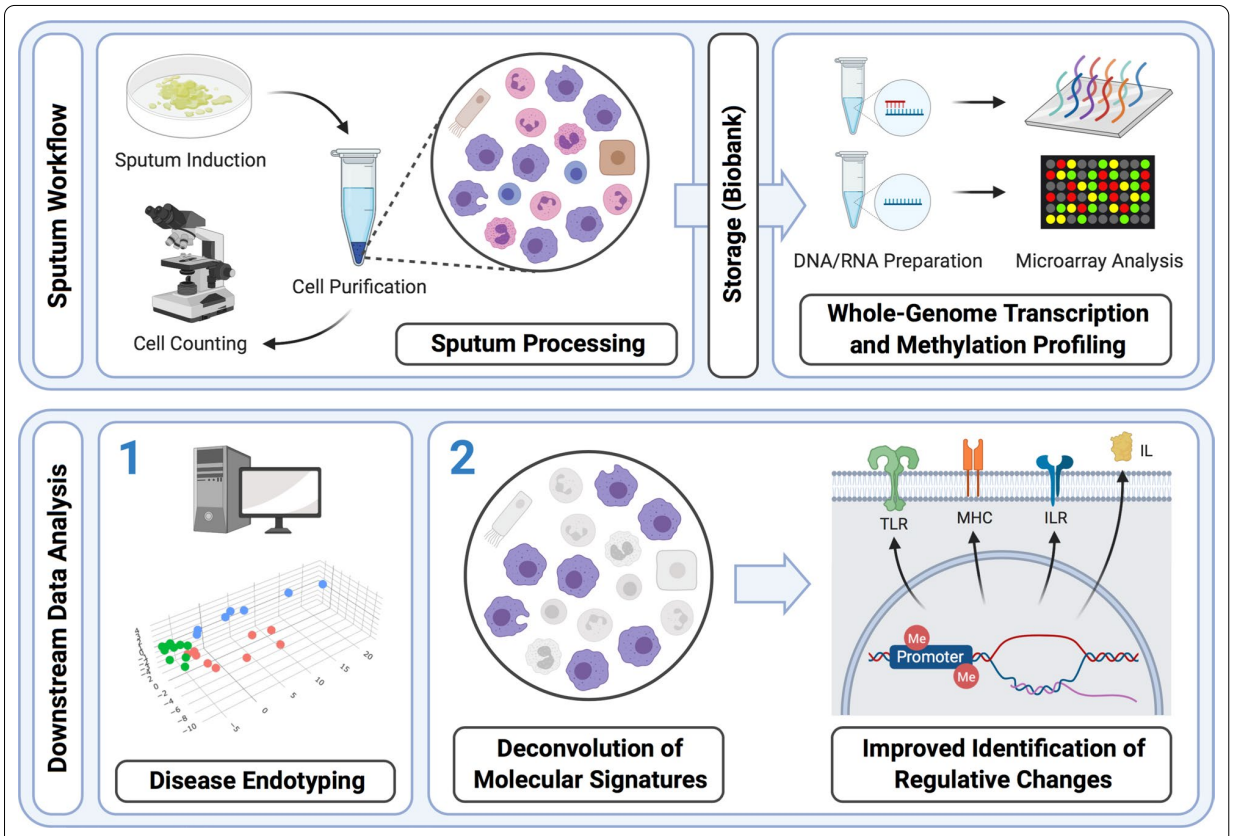

Fig. 1 Graphical abstract. Induced sputum, containing a variety of inflammatory cells, exhibits potential to directly reflect inflammatory processes in the lower airways. Progress in understanding the underlying mechanisms has been made by supplying sputum samples to high-throughput molecular analyses, primarily transcriptomics. To date, these have provided valuable insights to disease mechanisms and have led to differentiation of molecular endotypes (1) that are associated with distinct clinical presentations. However, most high-throughput analyses of sputum samples are prone to substantial bias by variation in cellular composition. Here, we introduce an unbiased deconvolution approach to sputum omics analysis in order to improve the identification of molecular patterns and dysregulation (2). Furthermore, were provide an example that sputum analysis can be extended by whole-genome methylation profiling to broaden the view on molecular mechanisms of pulmonary inflammation. Created with BioRender.com

successfully hybridized expression arrays to 25 out of 29 (9 asthma, 7 COPD, 9 controls).

Transcription microarray analysis

Extracted total RNA was processed with Agilent's Low Input Quick Amp Labeling Kit. Labeled complementary RNA (cRNA) was purified using the RNeasy Mini Kit (Qiagen) and 1650 ng of labeled cRNA per sample was hybridized to Agilent Human GE 4 × 44 K v2 arrays. All steps were performed according to the manufacturers' standard instructions. Hybridized arrays were scanned with an Agilent SureScan microarray scanner (5 µm resolution, default settings) and scan images were analyzed with Agilent's Feature Extraction Software (version 11.5.1, default parameters, protocol GE1_1105_Oct12). All hybridized arrays passed the manufacturer's standard quality controls.

Methylation microarray analysis

Genomic DNA was bisulfite converted utilizing the EZ DNA Methylation kit (ZymoResearch, Irvine, CA, USA) following the manufacturer's instructions. Converted DNA was further processed and hybridized to Infinium HumanMethylation 450 k BeadChips (Illumina Inc., San Diego, CA, USA) following the standard Illumina workflow. Hybridized chips were scanned with an Illumina iScan system on default settings. All chips passed the manufacturer's standard quality controls as well as further quality controls applied within the downstream in silico analysis. Due to a technical error, three samples were lost during processing, limiting the total number of samples from which methylation data was available to 26 (9 asthma, 10 COPD, 7 controls).

Data analysis

Downstream data analysis was entirely performed using the open-source R/Bioconductor software framework [48] (https://www.r-project.org, https://www.bioconductor.org). Supplementary methodological information as well as a comprehensive list of utilized software packages is provided in Additional file 1: Table S3). All annotation data used throughout this study was entirely based on the human genome version hg19 in accordance with the utilized array platforms. Transcriptome and methylome data have been made publicly available via NCBI's Gene Expression Omnibus [49] (see "Availability of data and materials").

Methylation array data was imported, annotated and processed by stratified quantile normalization utilizing the *minfi* package [50]. Individual CpG loci were filtered before further analysis by detection p values (threshold $p = 0.01$), mapping to sex chromosomes (X and Y chromosomes were excluded), affection by single nucleotide polymorphisms (SNPs) as well as potential for cross hybridization based on data published by Chen et al. [51]. Hereafter, a total of 429,236 CpGs (of initial 485,512) was further analyzed.

For gene expression data, median foreground signals were background corrected by subtracting the mean background signals ("minimum" method) via the *limma* package [52]. Quantile normalization was applied and control probes were filtered out. We used flagging information generated by the Feature Extraction Software (Agilent) to further exclude probes that were classified as non-uniform, saturated or feature population outlier in any of the arrays. Furthermore, at least 50% of features per array probe in any of the sample groups had to be classified as being found as well as being positive and significant to be kept in the data set. Replicate probes were averaged and further analysis steps were carried out on probe level (27,380 array probes of 44,495 initial feature reads retained).

Differential expression/methylation analysis

Differentially expressed genes (DEGs) as well as differentially methylated CpGs (differentially methylated positions, DMPs) were determined via *limma* comparing disease entities to healthy controls (asthma vs. controls and COPD vs. controls).

For DEGs, statistics were calculated on $\log_2$-transformed expression values with Benjamini–Hochberg (BH)-adjusted p value < 0.05 and absolute $\log_2$-fold change ($\log_2 FC$) ≥ 1.5 as statistical significance cutoffs. To remove redundancy from the data set and to simplify the biological interpretation of results, DEGs were filtered for well-annotated transcripts (based on available ENSEMBL and RefSeq annotation) and the most significantly differentially expressed transcript per gene was reported.

DMPs were determined on the beta value scale and considered statistically significant at a BH-adjusted p value < 0.05 and delta beta ≥ 0.1.

Deconvolution of cell type-specific expression and methylation

Generally speaking, the deconvolution we applied is based on the idea that estimates for cell type-specific expression/methylation can be derived by finding (optimizing) estimates that, given the relative cell counts for each sample, best match the observed (measured) mixed-cell expression/methylation. This poses a classical regression problem which gets complicated by the circumstance that both expression and methylation have biological limits (e.g., expression cannot be negative) that must not be violated by the mathematical optimization process in order to get biologically possible and meaningful results. In technical detail, we performed regression-based deconvolution (by quadratic programming) using the differential sputum cell counts as predictor variables and the measured mixed-cell omics profiles (expression/methylation) as response variables in the underlying linear models. To allow for linear combinability of the input data, expression values had to be analyzed on the linear (instead of $\log_2$-transformed) and methylation values on the beta value scale for the purpose of deconvolution. The general performance of a multiple linear regression approach was evaluated by fitting models with built-in functions of R (*stats* R core package). The estimation was carried out by quadratic programming (QP), allowing us to specify biological constraints under which the regression parameters were estimated. This approach had previously been successfully applied to methylation and gene expression data [36, 38]. A detailed mathematical description is provided in Additional file 1. In short, we utilized the *quadprog* R package (https://cran.r-project.org/web/packages/quadprog/index.html) which implements the dual method of Goldfarb and Idnani to solve quadratic programs [53]. Estimation was performed for each sample group separately, methylation estimates were constrained to the interval between 0 and 1, expression estimates to the dynamic range of the array. We estimated the standard errors of the estimates following a standard approach in regression analysis as previously applied [38]. Comparisons of methylation and expression estimates across disease groups was followingly carried out with a Welch modified two-sample (unequal variance) t-test. Taking into account that, due to the distribution of the analyzed methylation and expression values, one of the core assumptions of parametric testing (normality) was likely violated, we applied more stringent p

cutoffs to assign statistical significance: DMPs were considered statistically significant at a BH-adjusted $p < 0.001$ and delta beta ≥ 0.1. For DEGs, the BH-adjusted p cutoff was set to 0.005 at a $\log_2 FC \geq 1.5$.

Identification of DMRs

Differentially methylated regions (DMRs) were identified via *DMRcate* [54]. For the mixed-cell methylation data, the overall false discovery rate (FDR) was set to 0.05. For the analysis of the deconvolved estimates of cell type-specific methylation, the overall FDR was set to 0.001 (as for the identification of individual DMPs). DMRs had to contain at least one CpG with delta beta > 0.1 to be considered for further analysis. Mapping of DMRs to genomic regions of interest was performed to promoter (defined as up to 1500 base pairs upstream of the transcription start site) and gene body regions with a minimum required overlap of 200 base pairs.

GO and KEGG enrichment

Analyses for enrichment in Gene Ontology (GO) terms [55] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [56] were carried out via the *clusterProfiler* package [57]. Hypergeometric overrepresentation tests (cutoff $p < 0.1$, cutoff $q < 0.2$) were performed with custom backgrounds based on the array designs after probe filtering.

Results

Sample and data evaluation

Asthma samples were largely composed of eosinophils, alveolar macrophages and neutrophils. Whilst eosinophils were nearly absent in COPD and control samples, COPD sputum was greatly composed of neutrophils and that of healthy controls mainly of alveolar macrophages (see Table 2 and Fig. 2). Apart from the differences observed across conditions, substantial interindividual variation could be observed within the respective groups (for further information see Additional file 1: Tables S1, S2 and Figure S1.

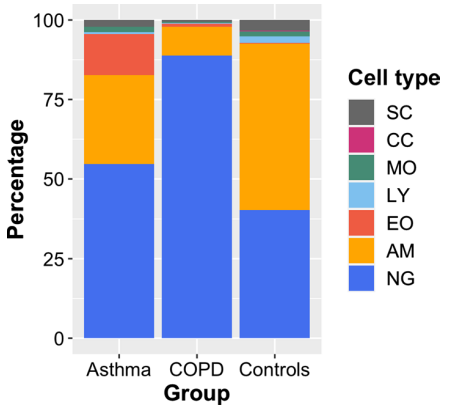

Fig. 2 Mean cellular composition of sputum samples. *AM* alveolar macrophages, *NG* neutrophil granulocytes, *EO* eosinophils, *LY* lymphocytes, *MO* monocytes, *CC* ciliated cells (respiratory epithelium), *SC* squamous cells

Those samples that had been specifically preserved for nucleic acid preparation by storage in RLT extraction buffer were found to provide RNA of overall good quality (RIN ranging from 7.6 to 9.1, see Table 3). Concomitantly, HOPE-preserved samples were subject to a higher amount of RNA degradation during biobank storage (maximum RIN 5.1).

The amount of RNA degradation could be shown to have a substantial effect on the overall variation in the expression data set (see principal component analysis in Fig. 3a). Gene expression values were both positively and negatively correlated with RNA integrity which follows from the rank-based process of quantile normalization applied to the data (see Additional file 1: Figure S2). Excluding array probes from further analysis by the extent of correlation (correlation filtering—reducing the data set to 15,550 transcripts out of 27,380; further details are provided in Additional file 1) efficiently removed major degradation effects (Fig. 3b) but was observed to be biased towards medium-to-highly expressed transcripts (Additional file 1: Figures S4 and

Table 2 Differential cell count of sputum samples

	AM	NG	EO	LY	MO	CC	SC
Asthma n = 9	27.9 ± 21.9 (6.3/60.4)	54.7 ± 24.4 (14.1/84.8)	12.9 ± 24.5 (1.5/77.0)	0.7 ± 0.5 (0.1/1.6)	0.1 ± 0.1 (0.0/0.3)	1.6 ± 1.0 (0.5/3.3)	2.1 ± 3.4 (0.3/10.8)
COPD n = 10	9.0 ± 5.9 (1.1/21.1)	88.9 ± 6.6 (76.6/98.1)	1.0 ± 1.2 (0.0/4.0)	0.2 ± 0.3 (0.0/0.8)	0.0	0.4 ± 0.4 (0.0/1.4)	0.6 ± 0.5 (0.0/1.8)
Controls n = 10	52.3 ± 24.9 (16.3/81.3)	40.3 ± 25.1 (6.5/76.1)	0.2 ± 0.4 (0.0/1.1)	2.0 ± 2.1 (0.0/7.6)	0.2 ± 0.3 (0.0/0.9)	1.6 ± 0.8 (0.4/2.6)	3.4 ± 3.3 (0.4/10.4)

Cell proportions are reported as mean percentage ± SD (min/max)

AM alveolar macrophages, *NG* neutrophil granulocytes, *EO* eosinophils, *LY* lymphocytes. *MO* monocytes, *CC* ciliated cells (respiratory epithelium), *SC* squamous cells

Table 3 RNA integrity of sputum samples supplied to gene expression analysis

	RIN Mean ± SD (min/max)	Preservation (RLT/HOPE)
Asthma n = 9	6.4 ± 2.5 (3.2/8.7)	5/4
COPD n = 7	5.6 ± 1.7 (4.2/8.5)	2/7
Controls n = 9	7.4 ± 2.1 (4.2/9.1)	6/3
RLT n = 13	8.6 ± 0.4 (7.6/9.1)	13/0
HOPE n = 12	4.3 ± 0.6 (3.2/5.1)	0/12

RIN RNA integrity number, *RLT* preservation by storage in RLT buffer, *HOPE* preservation via HOPE-fixation technique

S5). This approach was outperformed by a correction via a linear model (for details see Additional file 1) which ensured removal of degradation bias (Fig. 3c) without reducing the number of transcripts eligible for analysis (see also Additional file 1: Figure S6).

In case RNA degradation is equally distributed among compared sample groups, it can be assumed to primarily affect analysis sensitivity (more false negatives) whilst not necessarily leading to a higher proportion of false positives. Should the extent of degradation be distributed unequally among samples (the COPD samples exhibited a higher extent of degradation than the asthma or control samples), however, RNA degradation-biased expression data can in fact be expected to lead to falsely identified DEGs (false positives). Concordantly, only a minor number of DEGs in asthma identified in the uncorrected data was discarded by correlation filtering whilst RNA integrity correction by linear regression led to identification of additional DEGs (Fig. 4a). In COPD, reduction of the data set by correlation filtering primarily resulted in a concomitant reduction of the number of DEGs, whereas the linear model performed better in discarding and identifying new DEGs (Fig. 4b). Accordingly, all further analyses were performed on expression data corrected for RNA degradation by a linear model.

The methylation data, in contrast, exhibited no major influence by the respective sample preservation method (Fig. 5). Whereas asthma and control samples did not form separate clusters among the major principal components in an unsupervised analysis of the expression data (Fig. 4c), the methylation data interestingly allowed for a better separation of control and asthma samples (Fig. 5).

Deconvolution of cell type-specific gene expression and methylation

Since monocytes were only present in an overall very low quantity (and were not determinant in the COPD samples, see Table 2 and Additional file 1: Table S2), they were excluded from the deconvolution model ab initio. For both the expression and methylation data, estimates for macrophages, neutrophils and eosinophils (the latter in the asthma samples only) were found to be the most reliable, as inferred from the distribution of the p values associated with the respective fits of linear models (see Additional file 1: Figure S8). Consistently, the initial expression and beta value distributions were best retained in estimates for these cell types (see Additional file 1: Figures S6, S7, S9 and S10). This essentially follows from the mathematical nature of a linear model—estimation performs best for cell types that are prevalent whilst exhibiting variance across samples within a group. Strictly speaking, the estimation of macrophage profiles did not perform as well in the COPD samples as in asthma or controls. However, as the p value

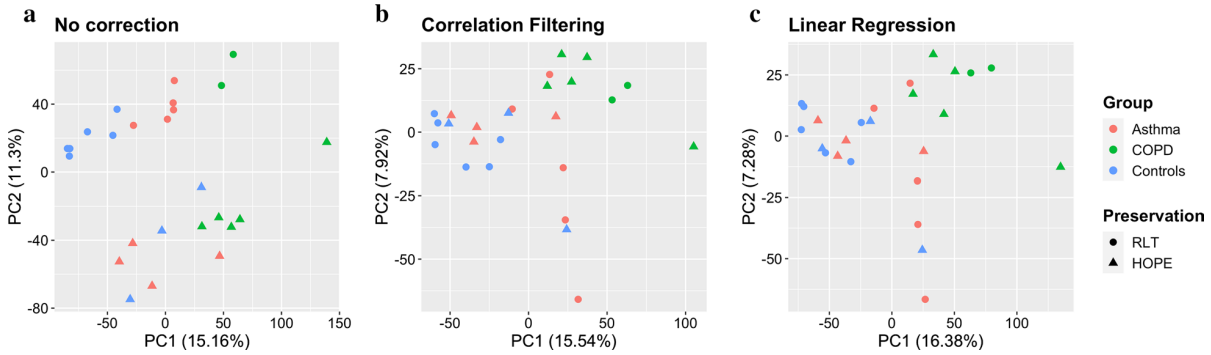

Fig. 3 Principal component analysis of the gene expression data. Before correction for RNA degradation (**a**), after correlation filtering (**b**) and after correction by linear regression (**c**)

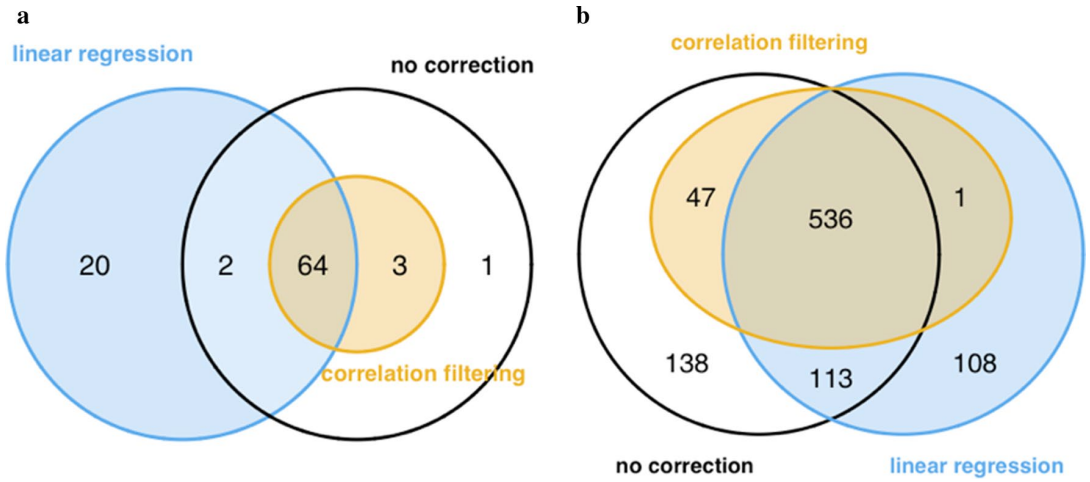

Fig. 4 Venn diagram visualizations of differentially expressed genes (DEGs). Asthma vs. controls (**a**) and COPD vs. controls (**b**). Analyses were performed on the uncorrected, mixed-cell transcriptome dataset (white/black circle), after correction for RNA degradation by correlation filtering (yellow) and after correction by linear regression (blue)

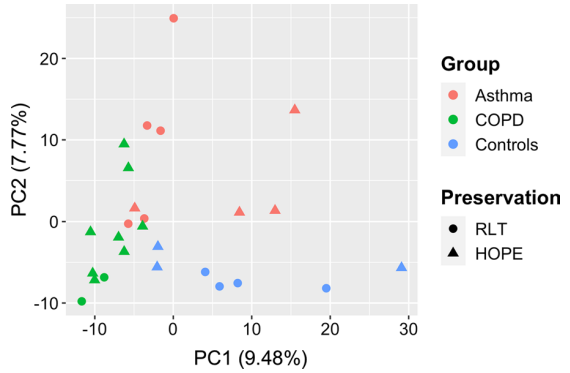

Fig. 5 Principal component analysis of the whole-genome methylation data

and expression/methylation value distributions still suggested a better performance than for the remaining cell types that could not be reliably estimated whatsoever, we decided to keep alveolar macrophages as predictor in the model fitted on the COPD data. We subsequently summarized the quantities of all other cell types and kept their sums as weighed intercepts in the models, thereby increasing the degrees of freedom of our analysis (see Additional file 1: Figures S11 to S13 and Tables S4 to S7).

Differential expression analysis of the cell type-specific estimates after deconvolution resulted in DEGs that only shared a minor proportion with those determined by analysis of the mixed-cell whole-sputum data (see Additional file 1: Figure S14). Similar observations could be made for DMPs.

DEGs identified as being upregulated in asthma in the mixed-cell analysis only (and discarded after deconvolution) clearly showed a pattern of estimated predominant expression in eosinophils (and partly neutrophils) whilst being lowly expressed in macrophages. In the background of higher sputum eosinophil counts in asthma this exemplifies how mixed-cell sputum analyses can be biased by disease-specific variation of cellular composition. For downregulated DEGs, the picture was the opposite around (see Additional file 1: Figure S15). In COPD, upregulated DEGs were estimated to be highly expressed in neutrophils and showed lower expression in macrophages in both the COPD and control samples. Though the overall expression in neutrophils seemed to be actually higher in COPD than in controls, the higher proportion of neutrophils in the COPD sputa is likely to still have had a major skewing influence. Downregulated DEGs showed a clear trend towards high estimated expression in macrophages. Similar patterns were observed for DMPs derived from the mixed-cell methylation analysis (see Additional file 1: Figure S16).

Differential expression analysis

A comprehensive compilation of results is provided as Additional files 2, 3, 4, 5, 6, 7, 8 and 9. In total, 86 genes were found to be differentially expressed in the asthma samples via the mixed-cell analysis (84 upregulated and 2 downregulated). After deconvolution by quadratic programming, 155 DEGs were identified for alveolar macrophages (13 up, 142 down) and 552 DEGs for neutrophils (145 up, 407 down).

DEGs identified by mixed-cell analysis were enriched in Gene Ontology (GO) terms highly related to immune response and regulation (see Additional file 5), e.g. by including CXCR1 and CXCR2 (chemokine CXC-motif receptors 1 and 2) as well as IL5RA (interleukin 5 receptor alpha). After deconvolution, DEGs in macrophages continued to be highly related to immune regulation but presented a greatly different picture of involved genes (such as TLR6, Toll-like receptor 6, and CD8A, cluster of differentiation 8a) and processes. In neutrophils, though immune-related genes could be identified, such as IL4R (interleukin 4 receptor) and CXCL2 (chemokine CXC-motif ligand 2), they did not significantly enrich in GO terms at the specified significance cutoffs. Results of KEGG pathway enrichment were related to immune regulation in the mixed-cell results and in macrophages (Additional file 6). As with GO enrichment, no KEGG pathways were significantly enriched in the neutrophil gene set.

The analysis of COPD samples resulted in 758 DEGs (612 up, 146 down) in the mixed-cell data and in 39 (10 up, 29 down) and 2161 (168 up, 1993 down) DEGs in macrophages and neutrophils after deconvolution, respectively. Whereas enriched GO terms were highly related to inflammatory processes (foremost neutrophil immunity) before deconvolution (with e.g. IL6R, interleukin-6 receptor, amongst them), no enriched GO terms could be identified in macrophages after deconvolution. CXCL9, chemokine (CXC motif) ligand 9, however, was found to be among the DEGs along with MMP13 (matrix metallopeptidase 13). GO enrichment in the neutrophils' DEGs resulted in a predominant picture of metabolic processes and regulation. Whilst enriched KEGG pathways were immunity-related in the mixed-cell analysis (e.g. containing the tumor necrosis factor, TNF signaling pathway), enrichment in macrophages, similar to the GO analysis, did not produce statistically significant results. Pathway terms significantly enriched in neutrophils were again related to metabolism, including the peroxisome and lysosome.

Differential methylation analysis

Genes that could be associated with differentially methylated regions in asthma (mixed-cell analysis) included a small quantity of immunity-related members such as IL27RA (interleukin 27 receptor alpha), IL20 (interleukin 20) and TNF but were overall dominated by small nucleolar RNA (snoRNA) as well as small Cajal body RNA (scaRNA) genes and thereby enriched in the GO term "Cajal body" (Additional file 7). After deconvolution, DMRs found in macrophages were still largely associated with small nucleolar RNAs, but also IL23A (interleukin 23 alpha), and CCL24 (chemokine C–C motif ligand 24, previously known as eotaxin-2). GO enrichment resulted in terms largely related to regulation of development and differentiation as well as cellular interaction by adhesion. In neutrophils, amongst a number of snoRNA genes, IL5RA (interleukin receptor 5 alpha) was found to be DMR-associated. Here, GO enrichment again resulted in terms primarily associated with developmental regulation and cell adhesion. Enriched KEGG pathways could be strongly related to inflammatory processes and regulation in the mixed-cell analysis (with TNF and the HLA, human leukocyte antigen, loci HLA-DRA and HLA-DOB largely contributing to this finding), including the KEGG pathway "Asthma" (Additional file 8). After deconvolution, KEGG enrichment was limited to "Antigen processing and presentation" in macrophages but with a higher number of gene hits, comprising genes associated with both MHC (major histocompatibility complex) class I (HLA-E, HLA-F) and class II (HLA-DMA, HLA-DMB, HLA-DOA, HLA-DPA1, HLA-DPB1) as well as heat shock protein genes, amongst others. In neutrophils, no KEGG pathways were significantly enriched.

In COPD, DMR-associated genes were related to immunity by being involved in neutrophil activation as well as antigen presentation. After deconvolution, enriched GO terms mostly related to developmental regulation in both macrophages and neutrophils. However, in macrophages, the immunity-related genes IL1RN (interleukin 1 receptor antagonist) and IL20RA (interleukin receptor 20 alpha) were found to be DMR-associated. Enriched KEGG pathways were related to inflammation and immunological regulation in the mixed-cell analysis and continued to be in macrophages after deconvolution. Again, HLA loci associated both with MHC class I and II (HLA-B, HLA-E, HLA-F as well as HLA-DMA, HLA-DMB, HLA-DQA2, HLA-DRA) contributed to this finding. No pathways were significantly enriched in COPD neutrophils (see also Additional file 1: Figure S17).

Integrative analysis

The performed deconvolution limits the applicability of some approaches for the integrative analysis of methylation and gene expression after deconvolution. E.g., a conventional correlation analysis of promoter-CpG beta values with gene expression values is not applicable to the estimates of cell-specific expression/methylation and their respective variances derived from the deconvolution. As a straightforward workaround, we decided to find overlaps between DMR-associated genes and DEGs: in the mixed-cell analysis, 1 gene was found to be both differentially expressed and methylated in asthma (see Additional file 1: Figure S18). After deconvolution, the respective quantities of DMR-associated DEGs were 23 for macrophages and 15 for neutrophils. In COPD, 74

DMRs mapped to DEGs that were identified without deconvolution. In the deconvolved data, 3 DEGs were differentially methylated in macrophages and 39 in neutrophils (see Additional file 9 for more details).

Data comparison

We did not encounter publicly available data sets that we could subject to the deconvolution approach described here for a validation of our results, largely because detailed information about the cellular composition of samples was not reported (further details are provided in Additional file 1).

Instead, we compared our data to a transcriptome study by Esnault et al. [29] who defined a core gene set predominantly expressed by lung eosinophils in asthma by expression profiling of BAL and sputum in the context of allergen challenges that was subsequently validated in purified lung eosinophils. In good agreement, this gene set was estimated to be predominantly expressed in eosinophils in our deconvolved data (see Additional file 1: Figure S19 and accompanying information). In accordance, we found the eosinophil marker genes RNASE2 (ribonuclease A family member 2), RNASE3, SIGLEC8 (sialic acid binding Ig-like lectin 8) and IL5RA [29] as well as PRSS33 (serine protease 33) [58] being exclusively expressed in eosinophils in the deconvolved asthma data (see Additional file 10). We extended this comparison to cell type-specific genes previously defined on transcriptomic reference sets derived from blood and bone marrow samples [10]. We observed the genes discriminating eosinophils, macrophages and neutrophils in our deconvolved asthma profiles predominantly overlapping with the respective genes defined on blood and bone-marrow data (see Additional file 1: Figure S20). However, the proportionate overlaps were smaller than with the gene set defined on lung eosinophils and in fact, a small number of discordances could be observed. Exemplarily, this will be further discussed in the following section.

Discussion

Here, we present first whole-genome methylation data from sputum indicating that the methylation profile of sputum cells can be used to further the molecular characterization of chronic pulmonary inflammation in asthma and COPD. By performing omics deconvolution based on quadratic programming, taking sputum differential cell counts as input, we show that the analysis of mixed-cell sputum samples is strongly biased by the interindividual variation of cellular composition. In this context, we present data indicating a high potential of omics deconvolution to deliver results that are ultimately more closely relatable to pathophysiological regulation by making differential expression and methylation attributable to individual cell types.

Genomic methylation and gene expression represent distinct entities of cellular regulation [59]. Whilst promoter methylation has traditionally been connected to gene repression, recent advances have brought up a much more complex picture of epigenetic regulation and its influence on gene expression [60, 61]. Because epigenetic changes are thought to also reflect long-term alterations, they exhibit a strong potential for the profiling of chronic diseases linked to the environment, such as asthma and COPD [21, 62].

We set the scope of this study to be primarily methodology-oriented and used a limited number of biological replicates in our explorative analysis which imposes a limitation. As statistical analyses and particularly regression models rely on an adequate number of degrees of freedom, the presented results should be interpreted carefully and not be considered conclusive. However, though our study is underpowered to draw a detailed picture of the interaction of gene methylation and expression, we found methylation changes in genetic regions of interest.

In asthma macrophages, we observed differential methylation of the IL23A gene which could be seen in the context of macrophage polarization [63]. Furthermore, we found differential methylation related to the CCL24 (previously known as eotaxin-2) gene, which can potentially be attributed to macrophage differentiation, microbiota interaction [64] and eosinophil stimulation [65]. In asthma neutrophils, the IL5RA (interleukin 5 receptor alpha) gene was identified to be differentially methylated. Interestingly, though gene expression was estimated to be present only in eosinophils in our study (and IL5RA expression has traditionally been seen as eosinophil-specific), this had recently been described to be expressed by airway neutrophils in the context of treatment-refractory asthma in children [66].

In both asthma and COPD macrophages, our data indicated differential methylation of HLA genes related predominantly to MHC class II, but also class I molecules. We did not observe concordant expression changes; however, changes in methylation of HLA loci have been described in a variety of autoimmune diseases and states of immune dysregulation [67–70] and were linked to atopic asthma [71] in whole-blood profiling of children with atopy after rhinovirus-induced wheezing. In the latter study, differential methylation of SMAD3 (SMAD family member 3) was found to be particularly associated with asthma [71]. Congruently, in our analysis, we also observed the SMAD3 gene to lie within DMRs both in asthma and, interestingly, COPD macrophages (see Additional file 4).

A particular interesting case is the differential methylation and expression of CD8A in asthma that was counterintuitively attributed to macrophages. Since we were not able to reliably estimate lymphocyte profiles in our small data set and we did not use lymphocyte counts as independent predictors in the deconvolution, the possibility of this being a "contamination" by lymphocyte-specific expression and methylation arises. However, upon inspection of the differential cell counts for each sample, we did not observe collinearity between the macrophage and lymphocyte cell counts. In fact, several studies have described CD8A expression in (alveolar) macrophages before [72–75]. A more detailed discussion of potential sources of error in the deconvolution process is provided as supplementary information (Additional file 1).

Deconvolution of transcriptomic signatures further allowed to identify interesting candidates regarding cellular regulation: Our data indicated an upregulation of IL4R (interleukin 4 receptor) in asthma neutrophils which had been shown to play an important role in the regulation of neutrophil apoptosis [76]. Furthermore, CXCL2 was observed to be upregulated for which autocrine regulation of neutrophils had been demonstrated previously [77], offering potential to contribute to inflammation in asthma. In COPD, macrophages were found to upregulate CXCL9 expression which is known to be a macrophage-derived inflammatory cytokine and indicator of M1 differentiation [63], whereas upregulated expression of MMP13 could be attributed to imbalanced protease homeostasis [78, 79].

For now, the experimental gold standard to retrieve cell type-specific molecular data remains to be cellular separation by techniques such as gradient centrifugation [29, 30]. However, these methods do not always allow to purify more than one cell type simultaneously and their applicability on a large scale (e.g. in biobanking studies) may not be given due to infrastructural or financial limitations. In contrast, in silico deconvolution is suitable for application to data from conventionally processed whole-sputum samples. The required differential cell counts are frequently performed in cohort and clinical studies. Unfortunately, findings derived from deconvolved data, unless cellular separation had been performed in parallel, will often not be able to be validated in the same, mixed-cell sputum samples from which the omics data was generated (like in this study). However, quadratic programming has previously been found to deliver accurate deconvolution estimates [36]. Accordingly, we found an overall good agreement with the data published by Esnault et al. [29] for eosinophils. With higher biological replication in larger studies, the partial lack of estimation performance for some cell types (foremost lymphocytes) observed here is likely to resolve, allowing for an even more comprehensive gain of information by applying a deconvolution. In this context, the applicability of omics deconvolution to sputum data is not limited to methods based on manually performed sputum differential cell counts. Some approaches use cell type-specific transcriptomic reference profiles to infer the respective cellular quantities in mixed-cell samples and use these quantities for the deconvolution process subsequently [41]. An important pitfall that investigators should be aware of before employing such reference-based methods in sputum analyses becomes apparent from the comparison of our deconvolved data to cell type-specific gene sets defined on blood and bone marrow-derived data by Peters et al. [10]. Exemplarily, the gene GPR97 (G-protein coupled receptor 97, also known as adhesion G-protein coupled receptor G3, ADGRG3) was found to be selectively expressed in neutrophils in their analysis, whilst both the data by Esnault et al. and our deconvolved expression profiles indicated a predominant expression in eosinophils in the asthmatic lung environment. In fact, expression of GPR97 has been described for all granulocytes [80]. This does not contest the analysis of Peters et al. since they followed a completely different approach in analyzing the sputum transcriptome but is rather intended to illustrate the potential bias that can be introduced to estimating the cellular composition of sputum based on reference sets derived from other sources. If reference sets are used for the purpose of deconvolution, we recommend they should be created based on cells derived from the lung environment (sputum or BAL) in the respective disease state [81] which becomes particularly important since alveolar macrophages are considered to be developmentally distinct from monocyte (blood)-derived macrophages [82]. Otherwise, performing detailed differential cell counts as shown here is a viable alternative.

From a phenotypical perspective, a large variability of cellular composition within a given set of samples, whilst benefiting the fit of a regression model, potentially indicates that several disease entities are comprised (such as T2-low, T2-high, T2-ultra high etc. in asthma). The validity of estimates derived from deconvolution steps thereby directly depends on the accuracy of the preceding definition of sample groups. Therefore, the application of a regression-based deconvolution approach has to be critically evaluated in any experiment and might find complimentary use to deepen the molecular understanding after distinct pheno-/endotypes were separated (e.g. via the method established by Peters and colleagues [10]).

We further demonstrated that, should RNA degradation be of concern, e.g. due to suboptimal or long-term biobank storage, in silico correction can remove RNA integrity-associated bias from transcriptome data,

thereby not only reducing the potential for the occurrence of false positives, but also increasing the overall sensitivity. In our data, a separate regression step performed well to remove degradation-related bias which is congruent with previous findings [83]. The cutoff we applied to select RNA samples supplied to microarray analysis was rather liberal (RIN > 3). Traditionally, cutoffs had been set to e.g. RIN > 5 [9, 84], at which it was demonstrated on cancer samples that the overall variance of gene expression is largely defined by interindividual differences and only to a much lesser extent by RNA integrity [84]. However, as overall expression differences in chronic inflammatory pulmonary disease may be more subtle than in cancer this cannot be easily assumed for sputum samples in asthma or COPD. Though several of the more strongly degraded samples in our study reached RIN values close to 5, they still clearly clustered separately from samples of higher RNA integrity. In fact, the principal component analysis of our transcriptome data showed that impaired RNA integrity has the potential of influencing the overall variance in transcription nearly as strongly as interindividual differences in asthma and COPD. Therefore, we suggest that the necessity of correction for RNA degradation should be evaluated even in data sets in which the sample quality had initially been judged suitable.

Thanks to optimized and streamlined purification workflows, parallel preparation of DNA and RNA from sputum samples is cost effective and efficient. With continuing innovation in the field of omics technologies and constantly growing affordability thereof, large-scale multi-omics analysis of sputum samples is close at hand. The application of the methods described here is by far not limited to sputum but can be expected to be successfully transferred to bronchoalveolar lavage and other respiratory samples to enhance biomarker discovery and pathophysiological understanding. Beyond microarray analysis, omics deconvolution and RNA integrity correction can further be expected to be of avail for sequencing-based methods as these can be similarly affected by RNA degradation and cell composition bias.

Conclusions

Analysis of the sputum methylome can broaden the profiling and understanding of chronic pulmonary inflammation and adds important additional information to commonly performed transcription analyses. The necessity of in silico correction for RNA degradation should be evaluated in every sputum transcriptome dataset. Finally, with suitable deconvolution approaches such as the algorithm described here, pathophysiological and regulative changes in chronic inflammatory lung diseases can be substantially better explored wherever single-cell analysis or cell separation may not be feasible. We therefore strongly recommend the application of unbiased deconvolution methods as such to all future whole-sputum omics analyses in order to complement methods that have already been established.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12931-020-01544-4.

Additional file 1. Supplementary Information, Figures and Tables.

Additional file 2. Results of the differential expression analyses.

Additional file 3. Results of the differential methylation analyses on the CpG level.

Additional file 4. Results of the differential methylation analyses on the gene region level.

Additional file 5. Results of gene ontology enrichment analyses of the gene expression data.

Additional file 6. Results of KEGG pathway enrichment analyses of the gene expression data.

Additional file 7. Results of gene ontology enrichment analyses of the methylation data.

Additional file 8. Results of KEGG pathway enrichment analyses of the methylation data.

Additional file 9. Differentially expressed genes that were found to correspond to differentially methylated genomic regions.

Additional file 10. Genes whose expression was found to discriminate eosinophils, neutrophils and macrophages in asthma sputum.

Abbreviations

ACT: Asthma control test; BAL: Bronchoalveolar lavage; BH: Benjamini–Hochberg; COPD: Chronic obstructive pulmonary disease; cRNA: Complementary ribonucleic acid; DNA: Deoxyribonucleic acid; DEG: Differentially expressed gene; DMP: Differentially methylated position; DMR: Differentially methylated region; FACS: Fluorescence-activated cell sorting; FDR: False discovery rate; GO: Gene Ontology; GOLD: Global Initiative for Chronic Obstructive Lung Disease; HLA: Human Leukocyte Antigen; HOPE: Hepes-glutamic Acid Buffer-mediated Organic Solvent Protection Effect; ICS: Inhaled corticosteroids; KEGG: Kyoto Encyclopedia of Genes and Genomes; $\log_2$FC: Logarithmized (base 2) Fold Change; MHC: Major Histocompatibility Complex; PCR: Polymerase chain reaction; QP: Quadratic programming; RIN: RNA Integrity Number; RNA: Ribonucleic acid; SNP: Single nucleotide polymorphism.

Acknowledgements

The authors would like to thank all study patients for their contribution to this study. EEG received personal funding by the German Academic Scholarship Foundation and is supported by the German Research Foundation (DFG) via the Clinician Scientist Program in Evolutionary Medicine (CSEM).

Authors' contributions

EEG, KFR, TG, OA and HW conceptualized this study. TB, FP, AK, KFR and HW were responsible for the clinical characterization of study subjects, sputum collection and processing. EEG, TG and OA generated the transcriptome and methylome data. EEG performed the data analysis with contributions by MW and DB. EEG wrote the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Funding

Analyses reported in this study were funded by the German Center for Lung Research (DZL). The ALLIANCE Cohort is funded by project grants from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) as part of DZL funding. The COSYCONET

cohort study is funded by the BMBF and by unrestricted grants from AstraZeneca GmbH, Bayer Schering Pharma AG, Boehringer Ingelheim Pharma GmbH & Co. KG, Chiesi GmbH, GlaxoSmithKline, Grifols Deutschland GmbH, MSD Sharp & Dohme GmbH, Mundipharma GmbH, Novartis Deutschland GmbH, Pfizer Pharma GmbH, Takeda Pharma Vertrieb GmbH & Co. KG for patient investigations and laboratory measurements.

Availability of data and materials
All transcriptome and methylome data have been deposited in NCBI's Gene Expression Omnibus where they are publicly accessible through the GEO series accession numbers GSE148000 and GSE148004.

Ethics approval and consent to participate
Data and sample collection as well as all analyses performed in this study were covered by the ethics approvals of the respective cohort studies. Before enrollment, all study subjects provided written informed consent to participate. The ALLIANCE cohort study is registered at clinicaltrials.gov (adult arm: NCT02419274). The COSYCONET cohort study is registered at clinicaltrials.gov (NCT01245933) and the German Clinical Trials Register (drks.de, DRKS00000284).

Consent for publication
Not applicable.

Competing interests
The authors declare they have no competing interests pertaining this work.

Author details
[1] LungenClinic Grosshansdorf, Großhansdorf, Germany. [2] Airway Research Center North (ARCN), Member of the German Center for Lung Research (DZL), Großhansdorf, Germany. [3] Department of Internal Medicine I, Pneumology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. [4] Department of Oncology, Hematology and BMT with Section Pneumology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [5] Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA. [6] Pulmonary Research Institute at LungenClinic Grosshansdorf, Großhansdorf, Germany. [7] Bioinformatics Core Unit, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [8] Institute of Human Genetics, University Medical Center Ulm, Ulm, Germany. [9] Research Center Borstel, Pathology, Borstel, Germany.

Received: 15 June 2020 Accepted: 11 October 2020
Published online: 19 October 2020

References

1. Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, Snowden S, Burg D, D'Amico A, Horvath I, et al. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. Eur Respir J. 2013;42:802–25.
2. Auffray C, Adcock IM, Chung KF, Djukanovic R, Pison C, Sterk PJ. An integrative systems biology approach to understanding pulmonary diseases. Chest. 2010;137:1410–6.
3. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, Sousa A, Corfield J, Djukanovic R, Lutter R, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. Eur Respir J. 2017;49:1602135.
4. Govoni M, Bassi M, Vezzoli S, Lucci G, Emirova A, Nandeuil MA, Petruzzelli S, Jellema GL, Afolabi EK, Colgan B, et al. Sputum and blood transcriptomics characterisation of the inhaled PDE4 inhibitor CHF6001 on top of triple therapy in patients with chronic bronchitis. Respir Res. 2020;21:72.
5. Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, Pillai SG, Hersh CP. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Med Genomics. 2015;8:1.
6. Singh D, Fox SM, Tal-Singer R, Bates S, Riley JH, Celli B. Altered gene expression in blood and sputum in COPD frequent exacerbators in the ECLIPSE cohort. PLoS ONE. 2014;9:e107381.
7. Steiling K, Lenburg ME, Spira A. Airway gene expression in chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2009;6:697–700.
8. Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, Cowan DC, Taylor DR, Cowan JO, Gibson PG. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol. 2014;133:997–1007.
9. Peters MC, Mekonnen ZK, Yuan S, Bhakta NR, Woodruff PG, Fahy JV. Measures of gene expression in sputum cells can identify T2-high and T2-low subtypes of asthma. J Allergy Clin Immunol. 2013;133:388–94.
10. Peters MC, Ringel L, Dyjack N, Herrin R, Woodruff PG, Rios C, O'Connor B, Fahy JV, Seibold MA. A transcriptomic method to determine airway immune dysfunction in T2-high and T2-low asthma. Am J Respir Crit Care Med. 2019;199:465–77.
11. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, Perez MF, Zhao H, Mane S, Martinez FD, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. Am J Respir Crit Care Med. 2015;191:1116–25.
12. Abdel-Aziz MI, Neerincx AH, Vijverberg SJ, Kraneveld AD, Maitland-van der Zee AH. Omics for the future in asthma. Semin Immunopathol. 2020;42:111–26.
13. Colas L, Hassoun D, Magnan A. Needs for systems approaches to better treat individuals with severe asthma: predicting phenotypes and responses to treatments. Front Med. 2020;7:98.
14. Sharma A, Kitsak M, Cho MH, Ameli A, Zhou X, Jiang Z, Crapo JD, Beaty TH, Menche J, Bakke PS, et al. Integration of molecular interactome and targeted interaction analysis to identify a COPD disease network module. Sci Rep. 2018;8:14439.
15. de Vries M, Nedeljkovic I, van der Plaat DA, Zhernakova A, Lahousse L, Brusselle GG, Amin N, van Duijn CM, Vonk JM, Boezen HM. DNA methylation is associated with lung function in never smokers. Respir Res. 2019;20:268.
16. Kabesch M, Tost J. Recent findings in the genetics and epigenetics of asthma and allergy. Semin Immunopathol. 2020;42:43–60.
17. Lin PI, Shu H, Mersha TB. Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. Sci Rep. 2020;10:151.
18. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, Sperling AI, Solway J, White SR, Nobrega MA, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight. 2016;1:e90151.
19. Qi C, Jiang Y, Yang IV, Forno E, Wang T, Vonk JM, Gehring U, Smit HA, Milanzi EB, Carpaij OA, et al. Nasal DNA methylation profiling of asthma and rhinitis. J Allergy Clin Immunol. 2020;145:1655–63.
20. Stadhouders R, Li BWS, de Bruijn MJW, Gomez A, Rao TN, Fehling HJ, van IJcken WF, Lim AI, Di Santo JP, Graf T, Hendriks RW. Epigenome analysis links gene regulatory elements in group 2 innate lymphocytes to asthma susceptibility. J Allergy Clin Immunol. 2018;142:1793–807.
21. Yang IV, Lozupone CA, Schwartz DA. The environment, epigenome, and asthma. J Allergy Clin Immunol. 2017;140:14–23.
22. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. 2002;62:2370–7.
23. Hubers AJ, Heideman DA, Herder GJ, Burgers SA, Sterk PJ, Kunst PW, Smit HJ, Postmus PE, Witte BI, Duin S, et al. Prolonged sampling of spontaneous sputum improves sensitivity of hypermethylation analysis for lung cancer. J Clin Pathol. 2012;65:541–5.
24. Zhang Z, Yan S, Cui H, Chen H, Liu J. Correlation between RASSF1A gene promoter hypermethylation in serum or sputum and non-small cell lung cancer (NSCLC): a meta-analysis. Med Sci Monit. 2019;25:5518–24.
25. Tiotiu A. Biomarkers in asthma: state of the art. Asthma Res Pract. 2018;4:10.
26. Medrek SK, Parulekar AD, Hanania NA. Predictive biomarkers for asthma therapy. Curr Allergy Asthma Rep. 2017;17:69.
27. Taube C, Reuter S. Transcriptome analysis of sputum cells. The modern art of assessing inflammation. Am J Respir Crit Care Med. 2019;199:402–4.
28. Zhu X, Chen Q, Liu Z, Luo D, Li L, Zhong Y. Low expression and hypermethylation of FOXP3 in regulatory T cells are associated with asthma in children. Exp Ther Med. 2020;19:2045–52.
29. Esnault S, Kelly EA, Schwantes EA, Liu LY, DeLain LP, Hauer JA, Bochkov YA, Denlinger LC, Malter JS, Mathur SK, Jarjour NN. Identification of

genes expressed by human airway eosinophils after an in vivo allergen challenge. PLoS ONE. 2013;8:e67560.
30. Poliska S, Csanky E, Szanto A, Szatmari I, Mesko B, Szeles L, Dezso B, Scholtz B, Podani J, Kilty I, et al. Chronic obstructive pulmonary disease-specific gene expression signatures of alveolar macrophages as well as peripheral blood monocytes overlap and correlate with lung function. Respiration. 2011;81:499–510.
31. Morrow JD, Chase RP, Parker MM, Glass K, Seo M, Divo M, Owen CA, Castaldi P, DeMeo DL, Silverman EK, Hersh CP. RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. Respir Res. 2019;20:65.
32. Weathington N, O'Brien ME, Radder J, Whisenant TC, Bleecker ER, Busse WW, Erzurum SC, Gaston B, Hastie AT, Jarjour NN, et al. BAL cell gene expression in severe asthma reveals mechanisms of severe disease and influences of medications. Am J Respir Crit Care Med. 2019;200:837–56.
33. Bertrams W, Griss K, Han M, Seidel K, Klemmer A, Sittka-Stark A, Hippenstiel S, Suttorp N, Finkernagel F, Wilhelm J, et al. Transcriptional analysis identifies potential biomarkers and molecular regulators in pneumonia and COPD exacerbation. Sci Rep. 2020;10:241.
34. Clarke J, Seo P, Clarke B. Statistical expression deconvolution from mixed tissue samples. Bioinformatics. 2010;26:1043–9.
35. Gaujoux R, Seoighe C. Semi-supervised nonnegative matrix factorization for gene expression deconvolution: a case study. Infect Genet Evol. 2012;12:913–21.
36. Gong T, Hartmann N, Kohane IS, Brinkmann V, Staedtler F, Letzkus M, Bongiovanni S, Szustakowski JD. Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. PLoS ONE. 2011;6:e27156.
37. Montano CM, Irizarry RA, Kaufmann WE, Talbot K, Gur RE, Feinberg AP, Taub MA. Measuring cell-type specific differential methylation in human brain tissue. Genome Biol. 2013;14:R94.
38. Onuchic V, Hartmaier RJ, Boone DN, Samuels ML, Patel RY, White WM, Garovic VD, Oesterreich S, Roth ME, Lee AV, Milosavljevic A. Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. Cell Rep. 2016;17:2075–86.
39. Perrier F, Novoloaca A, Ambatipudi S, Baglietto L, Ghantous A, Perduca V, Barrdahl M, Harlid S, Ong KK, Cardona A, et al. Identifying and correcting epigenetics measurements for systematic sources of variation. Clin Epigenet. 2018;10:38.
40. Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, Hastie T, Sarwal MM, Davis MM, Butte AJ. Cell type-specific gene expression differences in complex tissues. Nat Methods. 2010;7:287–9.
41. Avila Cobos F, Vandesompele J, Mestdagh P, De Preter K. Computational deconvolution of transcriptomics data from mixed cell populations. Bioinformatics. 2018;34:1969–79.
42. Fuchs O, Bahmer T, Weckmann M, Dittrich AM, Schaub B, Rösler B, Happle C, Brinkmann F, Ricklefs I, König IR, et al. The all age asthma cohort (ALLIANCE) - from early beginnings to chronic disease: a longitudinal cohort study. BMC Pulm Med. 2018;18:140.
43. Karch A, Vogelmeier C, Welte T, Bals R, Kauczor HU, Biederer J, Heinrich J, Schulz H, Gläser S, Holle R, et al. The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. Respir Med. 2016;114:27–37.
44. Pedersen F, Marwitz S, Seehase S, Kirsten AM, Zabel P, Vollmer E, Rabe KF, Magnussen H, Watz H, Goldmann T. HOPE-preservation of paraffin-embedded sputum samples—a new way of bioprofiling in COPD. Respir Med. 2013;107:587–95.
45. Goldmann T, Flohr AM, Murua Escobar H, Gerstmayer B, Janssen U, Bosio A, Loeschke S, Vollmer E, Bullerdiek J. The HOPE-technique permits Northern blot and microarray analyses in paraffin-embedded tissues. Pathol Res Pract. 2004;200:511–5.
46. Marwitz S, Kolarova J, Reck M, Reinmuth N, Kugler C, Schadlich I, Haake A, Zabel P, Vollmer E, Siebert R, et al. The tissue is the issue: improved methylome analysis from paraffin-embedded tissues by application of the HOPE technique. Lab Invest. 2014;94:927–33.
47. Marwitz S, Abdullah M, Vock C, Fine JS, Visvanathan S, Gaede KI, Hauber HP, Zabel P, Goldmann T. HOPE-BAL: improved molecular diagnostics by application of a novel technique for fixation and paraffin embedding. J Histochem Cytochem. 2011;59:601–14.
48. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12:115–21.
49. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2012;41:D991–5.
50. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
51. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.
52. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
53. Goldfarb D, Idnani A. A numerically stable dual method for solving strictly convex quadratic programs. Math Program. 1983;27:1–33.
54. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RV, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. Epigenet Chromatin. 2015;8:6.
55. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2018;47:D330–8.
56. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947–51.
57. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284–7.
58. Toyama S, Okada N, Matsuda A, Morita H, Saito H, Fujisawa T, Nakae S, Karasuyama H, Matsumoto K. Human eosinophils constitutively express a unique serine protease, PRSS33. Allergol Int. 2017;66:463–71.
59. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18:83.
60. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. Science. 2018;361:1336.
61. Kribelbauer JF, Lu X-J, Rohs R, Mann RS, Bussemaker HJ. Toward a mechanistic understanding of DNA methylation readout by transcription factors. J Mol Biol. 2020;432:1801–15.
62. Bae DJ, Jun JA, Chang HS, Park JS, Park CS. Epigenetic changes in asthma: role of DNA CpG methylation. Tuberc Respir Dis. 2020;83:1–13.
63. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Macrophage Polarization: different gene signatures in M1(LPS+) vs. classically and M2(LPS–) vs. alternatively activated macrophages. Front Immunol. 2019;10:1084.
64. Cheng M, Chen Y, Wang L, Chen W, Yang L, Shen G, Xu T, Shen G, Tian Z, Hu S. Commensal microbiota maintains alveolar macrophages with a low level of CCL24 production to generate anti-metastatic tumor activity. Sci Rep. 2017;7:7471–7471.
65. Palikhe NS, Kim SH, Cho BY, Ye YM, Choi GS, Park HS. Genetic variability in CRTH2 polymorphism increases eotaxin-2 levels in patients with aspirin exacerbated respiratory disease. Allergy. 2010;65:338–46.
66. Gorski SA, Lawrence MG, Hinkelman A, Spano MM, Steinke JW, Borish L, Teague WG, Braciale TJ. Expression of IL-5 receptor alpha by murine and human lung neutrophils. PLoS ONE. 2019;14:e0221113–e0221113.
67. Ramsuran V, Kulkarni S, O'Huigin C, Yuki Y, Augusto DG, Gao X, Carrington M. Epigenetic regulation of differential HLA-A allelic expression levels. Hum Mol Genet. 2015;24:4268–75.
68. Maltby VE, Lea RA, Sanders KA, White N, Benton MC, Scott RJ, Lechner-Scott J. Differential methylation at MHC in CD4+ T cells is associated with multiple sclerosis independently of HLA-DRB1. Clin Epigenet. 2017;9:71.
69. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, James T, Ewing E, Lindén M, Górnikiewicz B, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat Commun. 2018;9:2397.
70. Guo S, Zhu Q, Jiang T, Wang R, Shen Y, Zhu X, Wang Y, Bai F, Ding Q, Zhou X, et al. Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis. Mod Rheumatol. 2017;27:441–7.
71. Lund RJ, Osmala M, Malonzo M, Lukkarinen M, Leino A, Salmi J, Vuorikoski S, Turunen R, Vuorinen T, Akdis C, et al. Atopic asthma after

rhinovirus-induced wheezing is associated with DNA methylation change in the SMAD3 gene promoter. Allergy. 2018;73:1735–40.
72. Hirji N, Lin TJ, Befus AD. A novel CD8 molecule expressed by alveolar and peritoneal macrophages stimulates nitric oxide production. J Immunol. 1833;1997:158.
73. Lin TJ, Hirji N, Stenton GR, Gilchrist M, Grill BJ, Schreiber AD, Befus AD. Activation of macrophage CD8: pharmacological studies of TNF and IL-1 beta production. J Immunol. 2000;164:1783–92.
74. Gibbings DJ, Marcet-Palacios M, Sekar Y, Ng MC, Befus AD. CD8 alpha is expressed by human monocytes and enhances Fc gamma R-dependent responses. BMC Immunol. 2007;8:12.
75. Gibbings D, Befus AD. CD4 and CD8: an inside-out coreceptor model for innate immune cells. J Leukoc Biol. 2009;86:251–9.
76. Harris AJ, Mirchandani AS, Lynch RW, Murphy F, Delaney L, Small D, Coelho P, Watts ER, Sadiku P, Griffith D, et al. IL4Ralpha signaling abrogates hypoxic neutrophil survival and limits acute lung injury responses in vivo. Am J Respir Crit Care Med. 2019;200:235–46.
77. Li JL, Lim CH, Tay FW, Goh CC, Devi S, Malleret B, Lee B, Bakocevic N, Chong SZ, Evrard M, et al. Neutrophils self-regulate immune complex-mediated cutaneous inflammation through CXCL2. J Invest Dermatol. 2016;136:416–24.
78. Craig VJ, Zhang L, Hagood JS, Owen CA. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol. 2015;53:585–600.
79. Woode D, Shiomi T, D'Armiento J. Collagenolytic matrix metalloproteinases in chronic obstructive lung disease and cancer. Cancers. 2015;7:329–41.
80. Hsiao CC, Chu TY, Wu CJ, van den Biggelaar M, Pabst C, Hébert J, Kuijpers TW, Scicluna BP, Chen TC, et al. The adhesion G Protein-coupled receptor GPR97/ADGRG3 is expressed in human granulocytes and triggers antimicrobial effector functions. Front Immunol. 2018;9:2830.
81. Vallania F, Tam A, Lofgren S, Schaffert S, Azad TD, Bongen E, Haynes W, Alsup M, Alonso M, Davis M, et al. Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. Nat Commun. 2018;9:4735.
82. McQuattie-Pimentel AC, Budinger GRS, Ballinger MN. Monocyte-derived alveolar macrophages: the dark side of lung repair? Am J Respir Cell Mol Biol. 2018;58:5–6.
83. Gallego Romero I, Pai AA, Tung J, Gilad Y. RNA-seq: impact of RNA degradation on transcript quantification. BMC Biol. 2014;12:42.
84. Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, Ghadimi BM, Beissbarth T, Gaedcke J. Impact of RNA degradation on gene expression profiling. BMC Med Genomics. 2010;3:36.

Publisher's Note

2 Supplement

Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0

Espen E. Groth, Melanie Weber, Thomas Bahmer, Frauke Pedersen, Anne Kirsten, Daniela Börnigen, Klaus F. Rabe, Henrik Watz, Ole Ammerpohl, Torsten Goldmann

Additional File 1

Supplementary Information, Tables and Figures

Sputum induction and processing

Subjects underwent sputum induction by inhalation of nebulized saline in increasing concentrations from 0.9 to 3 % in repetitive inhalation/expectoration cycles (up to 4 per induction). Concomitant spirometry testing was performed to ensure subject safety during induction and saline concentration was not increased or induction of sputum was aborted if subjects experienced a decline in basal FEV1 of ≥ 10 or ≥ 20 %, respectively [1, 2].
Expectorated sputum was collected in petri dishes and assessed by microscopic evaluation. Dense plugs of sputum cells were manually separated from saliva and contaminants and incubated with dithiothreitol (Sputolysin®, Calbiochem/Merck, Darmstadt, Germany). After addition of phosphate-buffered saline (PBS), sputum cell suspensions were filtered through nylon mesh (53 µm pore size) and pelleted by centrifugation. Following removal of supernatants, cell pellets were resuspended in PBS and aliquots for subsequent cell counting with a hemocytometer were taken, followed by preparation of cell slides. Before further preservation and storage, remaining sputum cells were pelleted by centrifugation.

Sputum differential cell counts

No substantial differences in the distribution of cellular proportions could be observed between the samples submitted to methylation and transcription profiling (see Supplementary Tables S1 and S2 as well as Supplementary Figure S1).

Table S1: Differential cell count of sputum samples supplied to methylation microarray analysis

	AM	NG	EO	LY	MO	CC	SC
Asthma n = 9	27.9 ± 21.9 (6.3/60.4)	54.7 ± 24.4 (14.1/84.8)	12.9 ± 24.5 (1.5/77.0)	0.7 ± 0.5 (0.1/1.6)	0.1 ± 0.1 (0.0/0.3)	1.6 ± 1.0 (0.5/3.3)	2.1 ± 3.4 (0.3/10.8)
COPD n = 10	9.0 ± 5.9 (1.1/21.1)	88.9 ± 6.6 (76.6/98.1)	1.0 ± 1.2 (0.0/4.0)	0.2 ± 0.3 (0.0/0.8)	0.0	0.4 ± 0.4 (0.0/1.4)	0.6 ± 0.5 (0.0/1.8)
Controls n = 7	47.8 ± 23.3 (16.3/81.3)	43.3 ± 24.4 (6.5/76.1)	0.3 ± 0.4 (0.0/1.1)	2.1 ± 2.6 (0.0/7.6)	0.3 ± 0.3 (0.0/0.9)	1.8 ± 0.7 (0.9/2.6)	4.4 ± 3.5 (0.4/10.4)

Cell proportions are reported as mean percentage ± SD (min/max).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. MO: monocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

Table S2: Differential cell count of sputum samples supplied to gene expression microarray analysis.

	AM	NG	EO	LY	MO	CC	SC
Asthma n = 9	27.9 ± 21.9 (6.3/60.4)	54.7 ± 24.4 (14.1/84.8)	12.9 ± 24.5 (1.5/77.0)	0.7 ± 0.5 (0.1/1.6)	0.1 ± 0.1 (0.0/0.3)	1.6 ± 1.0 (0.5/3.3)	2.1 ± 3.4 (0.3/10.8)
COPD n = 7	9.9 ± 6.4 (2.8/21.1)	88.4 ± 6.7 (76.6/96.1)	0.6 ± 0.6 (0.0/1.8)	0.2 ± 0.3 (0.0/0.8)	0.0	0.4 ± 0.5 (0.0/1.4)	0.6 ± 0.6 (0.0/1.8)
Controls n = 9	52.8 ± 26.3 (16.3/81.3)	40.1 ± 26.6 (6.5/76.1)	0.3 ± 0.4 (0.0/1.1)	2.1 ± 2.2 (0.0/7.6)	0.2 ± 0.3 (0.0/0.9)	1.6 ± 0.9 (0.4/2.6)	3.0 ± 3.2 (0.4/10.4)

Cell proportions are reported as mean percentage ± SD (min/max).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. MO: monocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

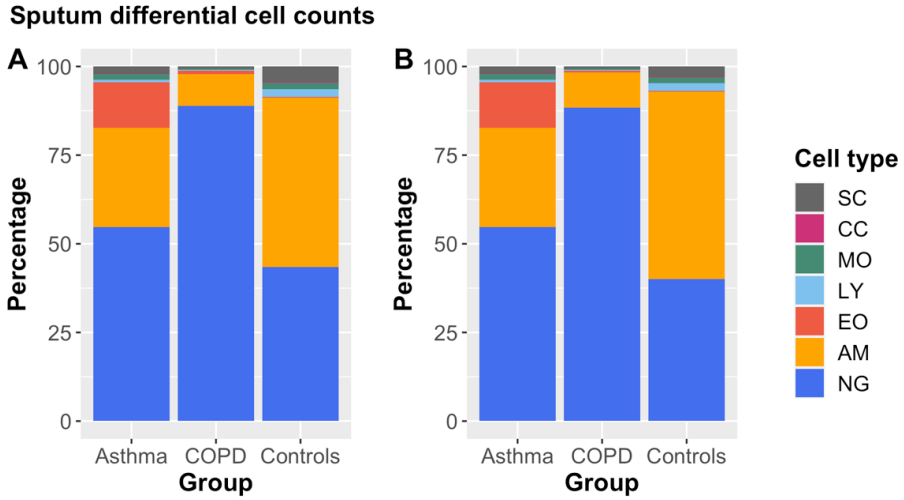

Figure S1: Mean cellular composition of sputum samples, shown separately for samples supplied to methylation profiling (A) and gene expression profiling (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. MO: monocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

R/Bioconductor software packages

Table S3: Software packages utilized for statistical analysis and data display.

Name	Version	Reference
R	3.6.1	[3]
devtools	2.2.2	[4]
Bioconductor	3.10.1	[5]
limma	3.42.2	[6]
minfi	1.32.0	[7]
IlluminaHumanMethylation450kanno.ilmn12.hg19	0.6.0	[8]
IlluminaHumanMethylation450kmanifest	0.4.0	[9]
RColorBrewer	1.1-2	[10]
matrixStats	0.56.0	[11]
DMRcate	2.0.7	[12]
DMRcatedata	2.2.1	[13]
quadprog	1.5-8	[14]
gtools	3.8.1	[15]
BSDA	1.2.1	[16]
plyr	1.8.6	[17]
clusterProfiler	3.14.3	[18]
org.Hs.eg.db	3.10.0	[19]
msigdbr	7.0.1	[20]
tidyr	1.0.2	[21]
ggplot2	3.3.0	[22]
cowplot	1.0.0	[23]
VennDiagram	1.6.20	[24]
EDec	0.9	[25]

DMR identification with *DMRcate*

All workflow steps were calculated based on methylation beta values at standard kernel settings (bandwidth lambda = 1000, scaling factor C = 2, minimum no. of consecutive CpGs = 2).
From the deconvolved data, we supplied the measures of significance of differential expression/methylation, together with the corresponding deconvolution estimates, to internal functions of the package.

Enrichment analysis with *clusterProfiler*

Gene symbols were mapped to Entrez IDs based on information of the org.Hs.eg.db annotation package.

Correction for RNA degradation

From the framework of available omics technologies, transcriptomics has by far been the one most frequently applied to sputum samples [26, 27]. However, the process of isolating high-quality nucleic acids from sputum is not trivial. Preparation of suitable samples from raw, freshly expectorated sputum is intensive and requires purifying procedures to remove saliva and break up mucin bonds by incubation with reducing chemicals. In light of the general instability of ribonucleic acid (RNA), being targeted by degrading enzymes (RNases) in the environment, challenges for handling and storage of sputum samples become obvious [28]. Methods that focus on high-throughput gene expression profiling (such as transcription microarray analysis or RNA sequencing) are optimized to work with very small amounts of input RNA and usually employ RNA amplification steps in their workflows. In the case of messenger RNA (mRNA) profiling, these are commonly based on poly-thymine (poly-T) primers that bind to poly-adenin sequences (poly-A tails), allowing for selective human mRNA amplification [29]. This has potential to introduce substantial bias to downstream analyses in light of RNA degradation. In case of the widely spread microarray platforms, transcripts may not be detected if their corresponding array probes map to distant transcript regions that had already been cleaved from the poly-A-adjacent fragment. Previous studies on this issue have shown that the exclusion of array probes from downstream analyses, based on the position in the corresponding transcript to which they map, does not necessarily remove RNA quality effects from transcriptome data. Instead, RNA degradation in biological materials is considered to be a dynamic process that not solely depends on the length of RNA molecules (as, in contrast, it can be assumed for purified RNA samples) [30, 31]. Though sputum processing and storage protocols have been optimized to ensure retrieval of high-quality RNA, degradation remains a challenge to consider in any large-scale prospective biobank study in which sample collection and analysis may take place years apart. As long as the overall variation of RNA quality can be assumed to be constant across sample groups, the effects specified above may primarily impair the sensitivity of comparative transcriptome studies. When RNA quality is found to be distributed unequally across sample groups, however, degradation may in fact lead to false-positive findings. In the context of inter-array quantile normalization, a rank-based normalization algorithm frequently applied to gene expression microarrays [32, 33], transcripts can in fact be rendered to both higher and lower expression levels upon RNA degradation [31]. Some *in silico* correction methods [31] have been suggested to overcome these challenges but have not been evaluated in the context of sputum analysis yet.

We evaluated two correction procedures for RNA degradation: First, we calculated the correlation (Pearson's product moment correlation coefficient r together with the associated two-sided p value of the correlation test) between expression and RIN (RNA Integrity Number) values across samples for each array probe. As expected in the context of rank-based normalization, gene expression was both positively (52.7 % of transcripts) and negatively (47.3 %) correlated with RIN in our data. In total, 43.2 % of transcripts were correlated at $p < 0.1$ and 36.1 % at $p < 0.05$ (see Figure S2). Probes who showed a correlation of $r \geq 0.3$ (to be seen as medium association, based on the definition of Cohen [34]) at $p < 0.1$ were removed from the data set, followingly referred to as correlation filtering.

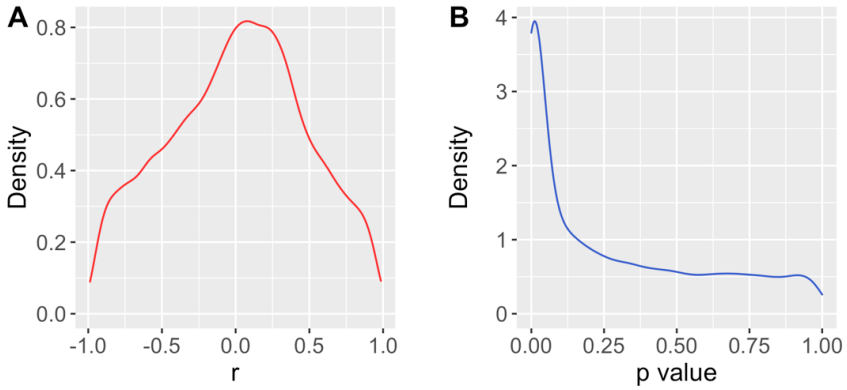

Figure S2: Correlation between measured gene expression and RNA integrity (RIN, RNA Integrity Number). Distribution of r (Pearson correlation coefficient, A) and associated p values (B).

Second, to correct for RNA degradation by linear least-squares regression, we applied a simple model of the form

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{with} \quad i = 1, \ldots, n \tag{1}$$

where n represents the number of samples supplied to the regression. For every sample i, x_i denotes the single scalar predictor variable, y_i the single scalar response variable and ε_i the stochastic component. β_0 and β_1 denote the regression parameters (β_0 can be referred to as intercept). In our case, we defined the RNA integrity (represented by the RNA integrity number, RIN) as predictor and the measured ($\log_2$) gene expression value as response variable. Consequently, β_1 represents the estimated measure of association between RNA integrity and expression, β_0 (intercept) the approximated integrity-independent measure of expression with an estimate of variation across samples retained in ε. Reversely calculating $\{y_i\}$ based on the estimates β_0, β_1 and $\{\varepsilon_i\}$ with $\{x_i\}$ set to the maximum RIN present in our study (9.1), we were able to retrieve expression values corrected for RIN-related effects (see also Figure S3). This was implemented with built-in functions of R (*stats* R core package).

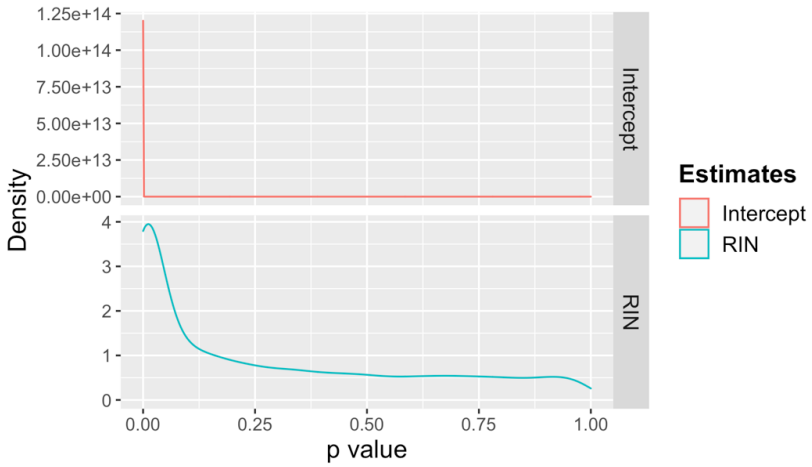

Figure S3: Applying linear regression for RNA integrity (RIN) correction: distribution of p values associated with regression parameters.

Before applying the correlation filtering or linear regression steps to the expression dataset, we took into account that the COPD group contained a higher proportion of HOPE-preserved samples with therefore more intensely degraded RNA than the asthma and control groups (see Table 3 in main manuscript). Therefore, correlation and regression characteristics were determined based on the asthma and control samples and then subsequently applied to the COPD samples in order to control for a possible skewing of correlation and regression parameters by COPD-specific expression patterns [31].

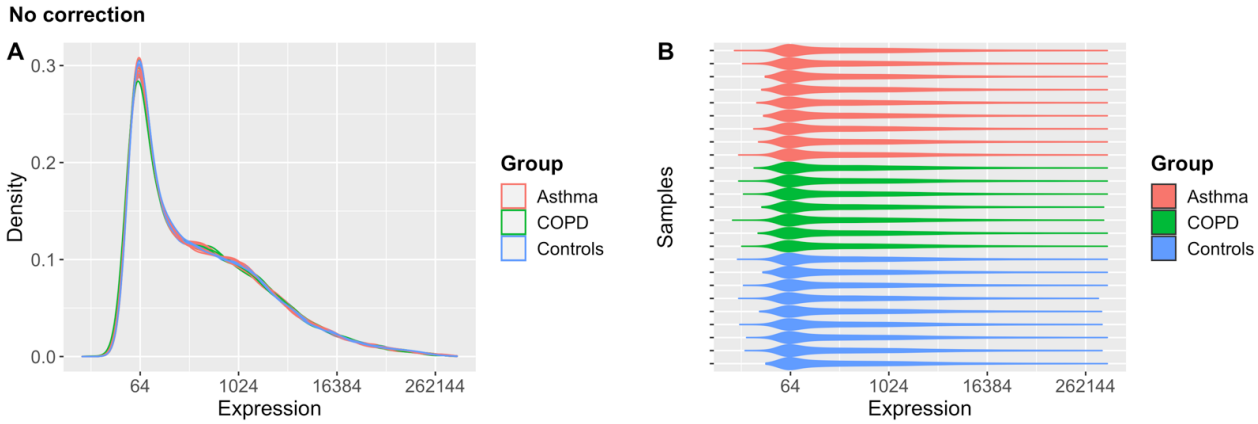

Figure S4: Distribution of gene expression values before correction for RNA integrity: density plot (A) and bean plot (B).

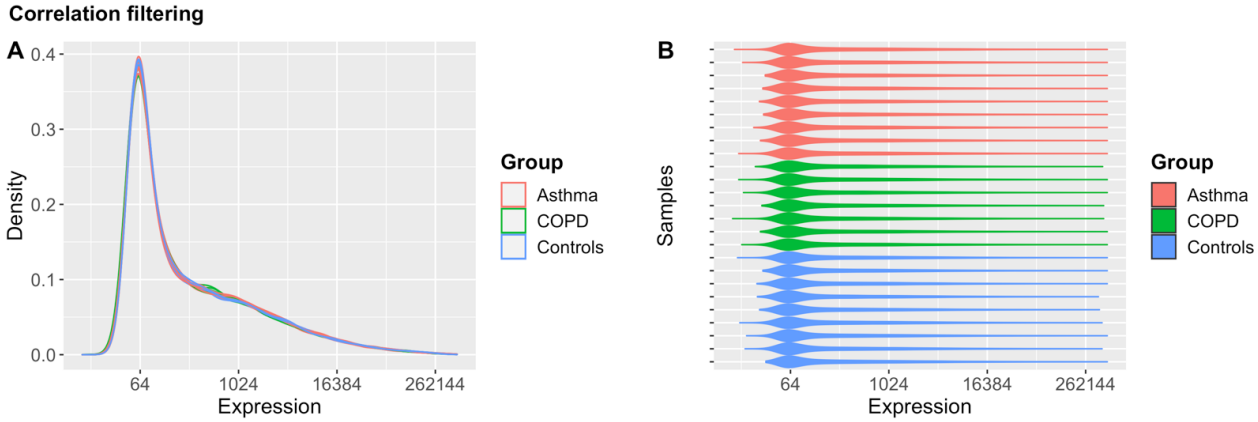

Figure S5: Distribution of gene expression values after correlation filtering: density plot (A) and bean plot (B).

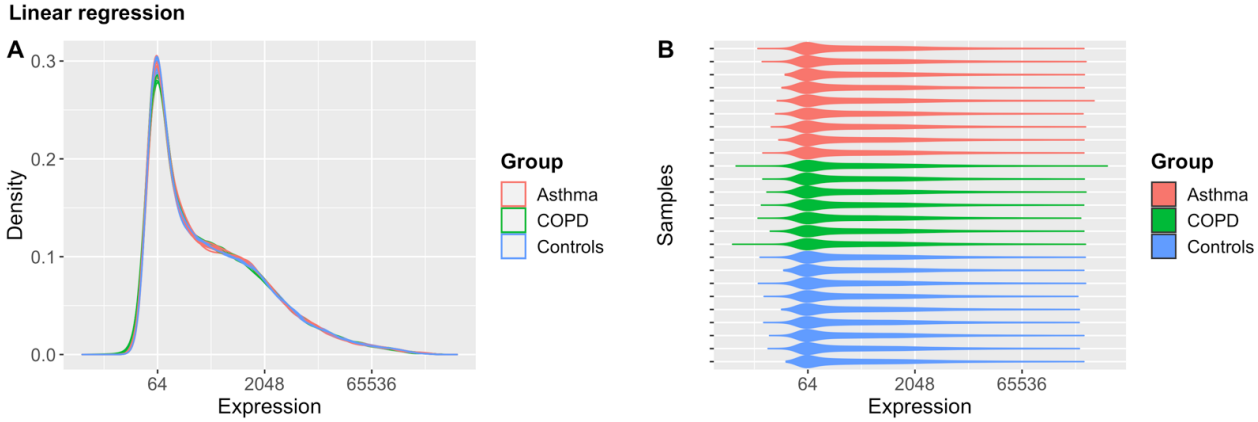

Figure S6: Distribution of gene expression values after correction for RNA integrity by linear regression: density plot (A) and bean plot (B).

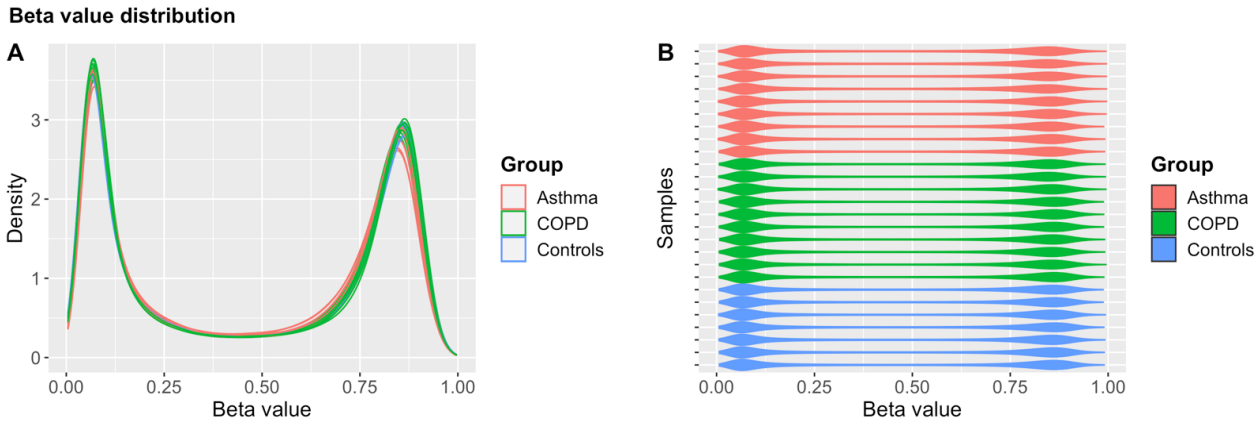

Figure S7: Distribution of beta values in the methylome data: density plot (A) and bean plot (B).

Estimation of cell type-specific gene expression and methylation (deconvolution)

Based on the general assumption that both the methylation level of any CpG and the expression level of any transcript measured in a mixed-cell sample (such as sputum) result from the linear combination of the methylation/expression levels of all analyzed cells (as it were the "mean" of expression/methylation of all cells), and further assuming that methylation/gene expression across the cells of a particular cell type can be seen as homogeneous, inferring cell type-specific omics patterns from mixed-cell measurements can be reduced to a multiple (linear) regression problem in case the cellular proportions are known:

$$y = X\beta + \varepsilon \tag{2}$$

Here, $X = \{x_{ij}\}$ represents the matrix of predictor variables where column j relates to the proportionate quantity of cell type $j = 1, \dots, c$ across samples $i = 1, \dots, n$. $y = \{y_i\}$ denotes the vector of response variables (measured methylation/expression), $\beta = \{\beta_j\}$ the vector of regression parameters (estimated cell type-specific methylation/gene expression) and $\varepsilon = \{\varepsilon_i\}$ the vector of stochastic components (residuals). Fitting the above model by the standard least-squares method, however, disregards that methylation and expression values in the biological context represent bounded variables: there exists no methylation below 0 or over 100 % and, likewise, there exists no negative expression. Therefore, during the regression process, an optimal solution has to be found which assigns to all cell types expression/methylation levels that are biologically possible. As a solution, we performed regression by quadratic programming, allowing us to specify (biological) constraints (C in (3)) under which the regression parameters were estimated. This approach had previously been successfully applied to methylation and gene expression data [35, 36]. Programmatically, we implemented the dual method of Goldfarb and Idnani [37] (via the *quadprog* package) to solve the problem

$$\arg min_{\beta} \parallel y - X\beta \parallel^2 \quad \text{with} \quad \beta \in C \tag{3}$$

The assumption of linear combinability can be seen to hold true for methylation reported on the beta value scale (where values approximately correspond to proportionate methylation). Since the commonly applied $\log_2$-transformation of expression values does not allow for linear combination, expression values had to be analyzed on the linear (instead of $\log_2$-transformed) scale.

Estimation was performed for each sample group (asthma, COPD and controls) separately. Methylation estimates were constrained to $C = [0, 1]$ and expression estimates to the dynamic range of the array (defined as respective minimum and maximum background corrected feature intensities after quantile normalization).

Following a standard approach in regression analysis as previously implemented by Onuchic et al. [35], we estimated the standard error s_j of each of the regression estimates (methylation/gene expression level in each cell type j) in the same way as for a multiple linear regression problem by

$$s_j = \sqrt{[MSE\ (X^t X)^{-1}]_{j,j}} \tag{4}$$

where *MSE* denotes the mean squared error, calculated by

$$MSE = \frac{\sum_{i=1}^{n} {\varepsilon_i}^2}{n - c} \tag{5}$$

In (5), the mean squared error (MSE) is considered to be an unbiased estimator of the true (unknown) mean squared error by dividing by the degrees of freedom [38].

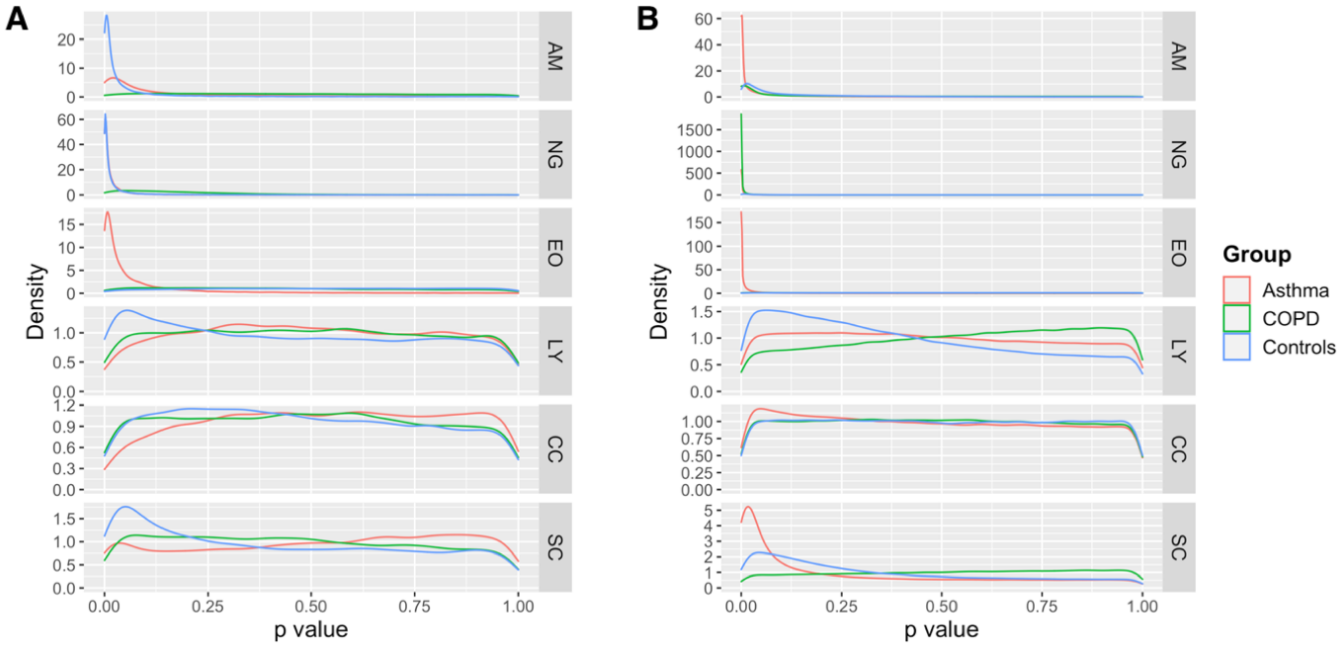

Figure S8: Distribution of the regression p values associated with fitting multiple linear models to the gene expression (A) and methylation data (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

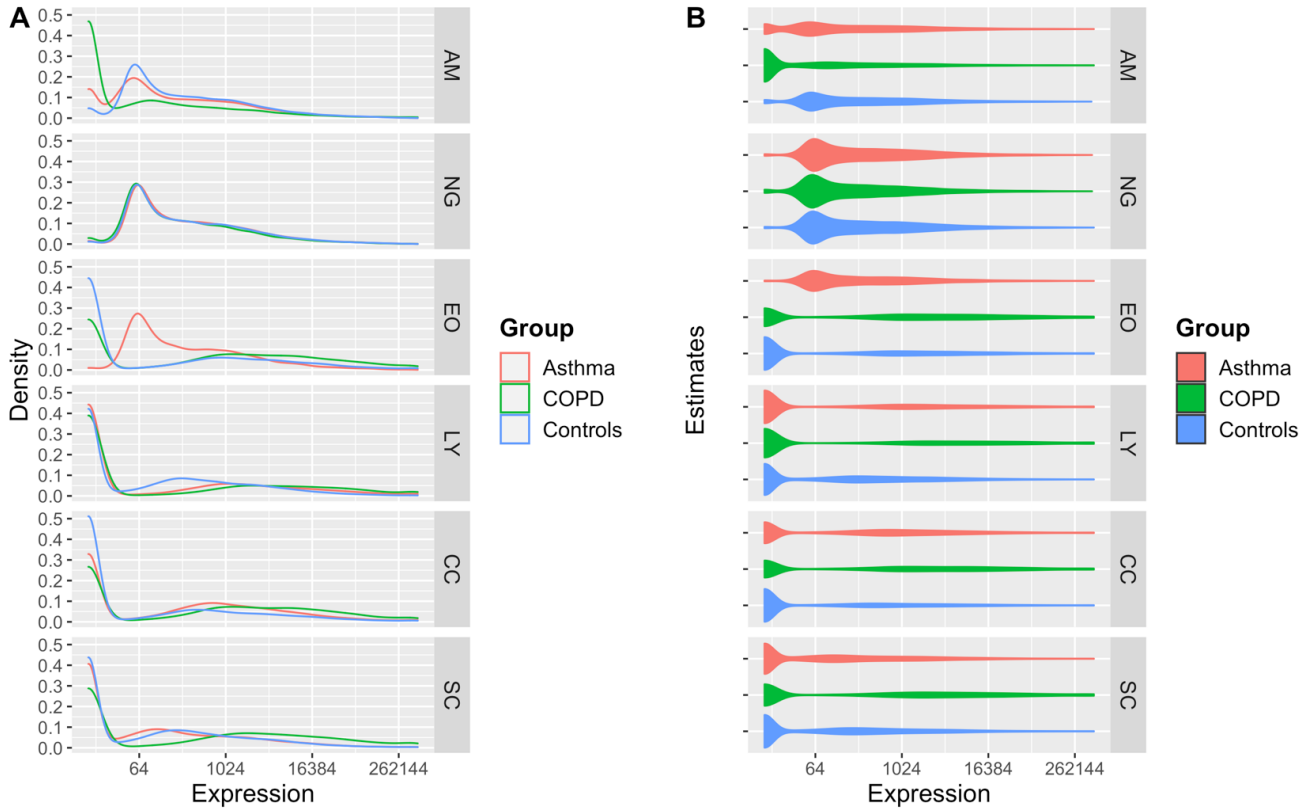

Figure S9: Distribution of the cell type-specific gene expression estimates after deconvolution: density plot (A) and bean plot (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

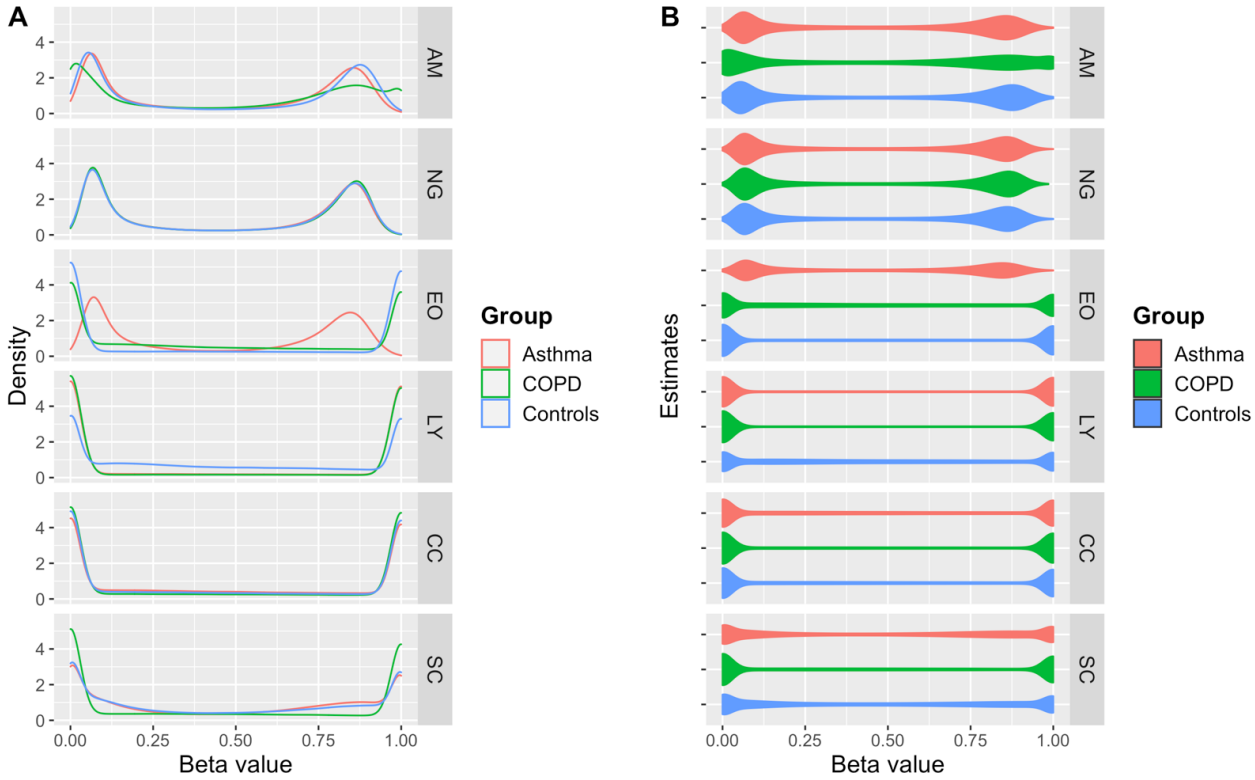

Figure S10: Distribution of the cell type-specific methylation beta value estimates after deconvolution: density plot (A) and bean plot (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

Table S4: Percentiles of the regression p value distributions associated with fitting multiple linear models to the gene expression data.

		Percentiles								% sig at	
Group	**Cell**	5%	10%	25%	50%	75%	95%	98%	99%	p < 0.05	p < 0.001
Asthma	**AM**	0.00	0.01	0.02	0.07	0.26	0.79	0.92	0.96	41.2	1.6
	NG	0.00	0.00	0.00	0.01	0.02	0.13	0.33	0.56	87.3	13.1
	EO	0.00	0.00	0.01	0.03	0.09	0.43	0.69	0.84	62.4	5.1
	LY	0.07	0.12	0.27	0.50	0.74	0.95	0.98	0.99	3.8	0.1
	CC	0.08	0.15	0.30	0.54	0.77	0.95	0.98	0.99	2.8	0.1
	SC	0.03	0.08	0.27	0.54	0.78	0.96	0.98	0.99	7.4	0.3
COPD	**AM**	0.04	0.09	0.22	0.45	0.70	0.94	0.98	0.99	5.7	0.1
	NG	0.01	0.03	0.07	0.16	0.28	0.61	0.82	0.91	17.0	0.4
	EO	0.04	0.08	0.21	0.43	0.70	0.94	0.98	0.99	6.3	0.1
	LY	0.05	0.10	0.25	0.49	0.74	0.95	0.98	0.99	4.9	0.1
	CC	0.05	0.10	0.24	0.49	0.72	0.95	0.98	0.99	5.2	0.1
	SC	0.04	0.08	0.22	0.45	0.71	0.94	0.98	0.99	6.0	0.2
Controls	**AM**	0.00	0.00	0.00	0.02	0.06	0.57	0.84	0.92	71.3	7.6
	NG	0.00	0.00	0.00	0.01	0.03	0.20	0.45	0.65	84.4	14.1
	EO	0.06	0.12	0.28	0.53	0.77	0.95	0.98	0.99	4.1	0.1
	LY	0.02	0.06	0.18	0.43	0.72	0.94	0.98	0.99	9.0	0.2
	CC	0.05	0.10	0.23	0.46	0.71	0.94	0.98	0.99	4.8	0.1
	SC	0.02	0.04	0.14	0.38	0.68	0.94	0.97	0.99	11.7	0.3

AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

Table S5: Percentiles of the regression p value distributions associated with fitting multiple linear models to the methylation data.

		Percentiles								% sig at	
Group	**Cell**	5%	10%	25%	50%	75%	95%	98%	99%	p < 0.05	p < 0.001
Asthma	**AM**	0.00	0.00	0.00	0.01	0.05	0.34	0.59	0.75	74.5	33.3
	NG	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.13	96.0	59.4
	EO	0.00	0.00	0.00	0.00	0.01	0.12	0.27	0.41	88.9	48.2
	LY	0.05	0.09	0.23	0.47	0.72	0.94	0.98	0.99	5.2	0.1
	CC	0.04	0.08	0.22	0.47	0.73	0.95	0.98	0.99	6.1	0.1
	SC	0.00	0.01	0.03	0.16	0.53	0.90	0.96	0.98	32.3	2.0
COPD	**AM**	0.00	0.00	0.00	0.06	0.39	0.87	0.95	0.97	47.7	13.1
	NG	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.12	97.2	68.5
	EO	0.03	0.07	0.20	0.45	0.72	0.94	0.98	0.99	7.9	0.2
	LY	0.07	0.13	0.31	0.56	0.79	0.96	0.98	0.99	3.6	0.1
	CC	0.05	0.10	0.25	0.49	0.74	0.95	0.98	0.99	5.2	0.1
	SC	0.06	0.12	0.29	0.54	0.78	0.96	0.98	0.99	4.2	0.1
Controls	**AM**	0.00	0.01	0.02	0.06	0.19	0.51	0.71	0.83	44.9	1.2
	NG	0.00	0.00	0.01	0.03	0.09	0.26	0.40	0.51	62.4	2.5
	EO	0.05	0.11	0.26	0.52	0.76	0.95	0.98	0.99	4.7	0.1
	LY	0.03	0.07	0.17	0.37	0.64	0.92	0.97	0.98	7.7	0.2
	CC	0.05	0.10	0.25	0.50	0.75	0.95	0.98	0.99	5.0	0.1
	SC	0.02	0.04	0.11	0.28	0.57	0.91	0.96	0.98	11.8	0.2

AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. LY: lymphocytes. CC: ciliated cells (respiratory epithelium). SC: squamous cells.

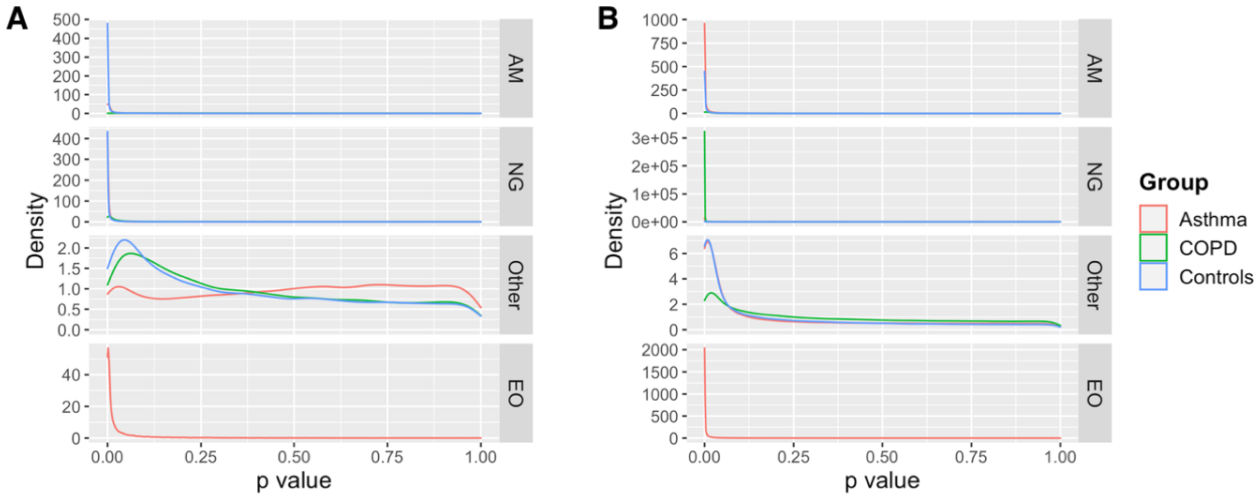

Figure S11: Distribution of the regression p values associated with fitting multiple linear models to the gene expression (A) and methylation data (B) after summarization of cell types with low prevalence.
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: sum of residual cell types (weighed intercept).

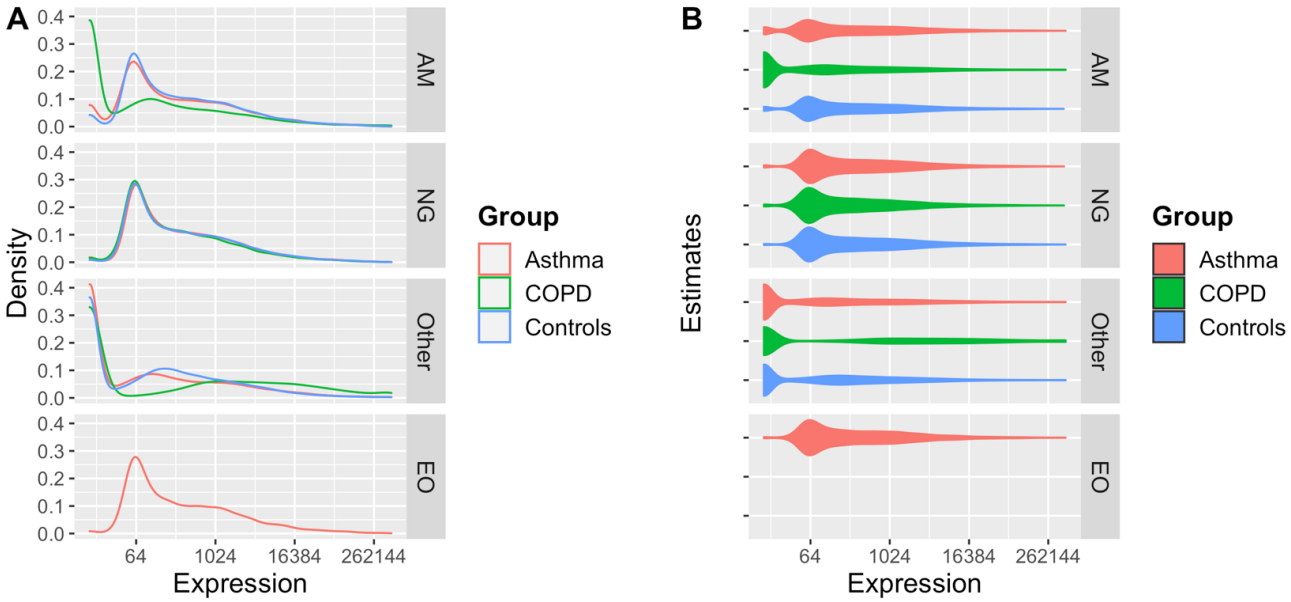

Figure S12: Distribution of the cell type-specific gene expression estimates after deconvolution with the reduced set of cell types: density plot (A) and bean plot (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: sum of residual cell types (weighed intercept).

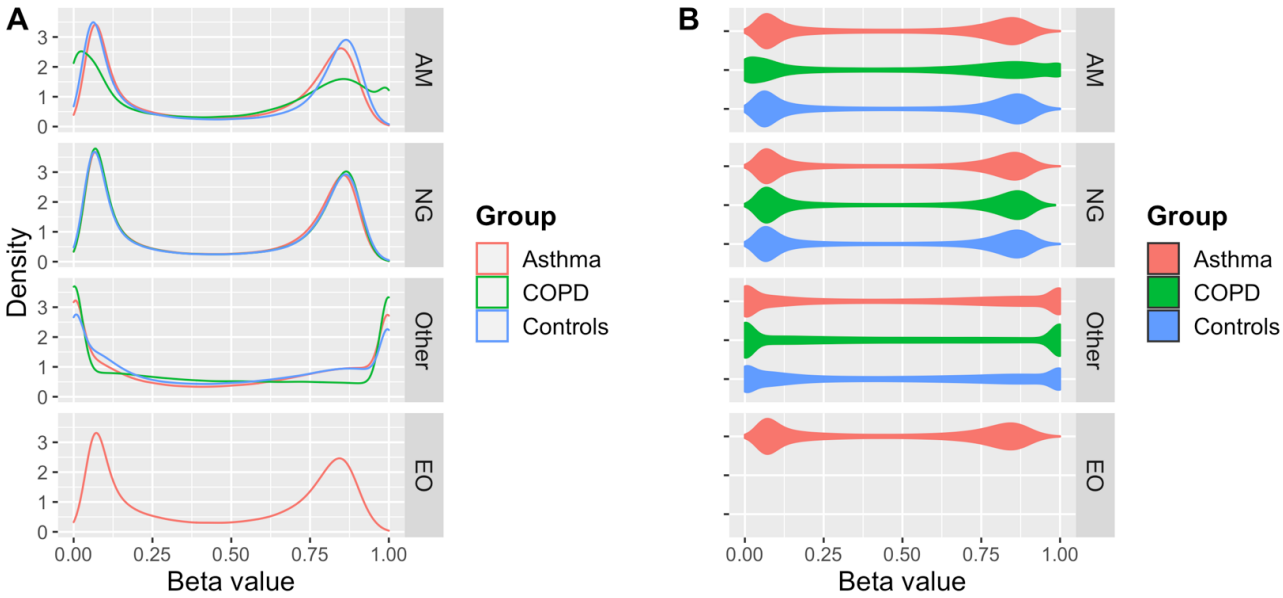

Figure S13: Distribution of the cell type-specific methylation beta value estimates after deconvolution with the reduced set of cell types: density plot (A) and bean plot (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: sum of residual cell types (weighed intercept).

Table S6: Percentiles of the regression p value distributions associated with fitting multiple linear models to the gene expression data after summarization of cell types with low prevalence.

		Percentiles								% sig at	
Group	**Cell**	5%	10%	25%	50%	75%	95%	98%	99%	p < 0.05	p < 0.001
Asthma	**AM**	0.00	0.00	0.00	0.00	0.04	0.60	0.83	0.92	76.2	27.9
	NG	0.00	0.00	0.00	0.00	0.00	0.07	0.34	0.56	93.6	52.5
	Other	0.02	0.07	0.26	0.53	0.77	0.95	0.98	0.99	8.5	0.6
	EO	0.00	0.00	0.00	0.01	0.04	0.32	0.59	0.77	79.2	24.0
COPD	**AM**	0.03	0.06	0.18	0.40	0.68	0.93	0.97	0.99	8.2	0.2
	NG	0.00	0.00	0.00	0.02	0.07	0.46	0.75	0.87	70.3	10.1
	Other	0.02	0.04	0.13	0.33	0.63	0.93	0.97	0.99	11.1	0.2
Controls	**AM**	0.00	0.00	0.00	0.00	0.00	0.23	0.59	0.79	88.6	63.5
	NG	0.00	0.00	0.00	0.00	0.01	0.15	0.41	0.57	90.6	56.2
	Other	0.01	0.03	0.10	0.31	0.62	0.93	0.97	0.98	15.3	0.4

AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: Sum of residual cell types (weighed intercept).

Table S7: Percentiles of the regression p value distributions associated with fitting multiple linear models to the methylation data after summarization of cell types with low prevalence.

		Percentiles								% sig at	
Group	**Cell**	5%	10%	25%	50%	75%	95%	98%	99%	p < 0.05	p < 0.001
Asthma	**AM**	0.00	0.00	0.00	0.00	0.00	0.07	0.18	0.31	93.8	66.0
	NG	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.06	98.7	81.4
	Other	0.00	0.00	0.01	0.10	0.49	0.90	0.96	0.98	41.9	8.2
	EO	0.00	0.00	0.00	0.00	0.00	0.05	0.15	0.27	95.0	70.1
COPD	**AM**	0.00	0.00	0.00	0.02	0.26	0.82	0.92	0.96	56.7	35.0
	NG	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	99.6	93.6
	Other	0.01	0.02	0.09	0.31	0.63	0.92	0.97	0.98	18.0	1.6
Controls	**AM**	0.00	0.00	0.00	0.00	0.01	0.11	0.29	0.49	91.3	58.8
	NG	0.00	0.00	0.00	0.00	0.00	0.05	0.13	0.25	95.4	67.0
	Other	0.00	0.00	0.01	0.10	0.44	0.88	0.95	0.98	41.5	8.5

AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: Sum of residual cell types (weighed intercept).

When implementing a regression model-based deconvolution approach several points have to be critically considered. The model's performance is not only dependent on the achieved number of degrees of freedom, but also directly dependent on the quality and accuracy of the input data. Sputum differential cell counts should therefore be prepared with the highest resolution possible. However, cell types that only represent a small proportion of the cell set or that cannot be identified via a conventional differential cell count (such as specific subsets of lymphocytes) won't be able to be estimated via the regression. The overall performance of the model in our data shows that this holds true: estimates were anticipated to be most reliable for those cell types that were most prevalent whilst exhibiting variability within the respective sample group. Consequently, in our data, this led to the estimation of expression and methylation in eosinophils only performing well in asthma samples. In larger-scale studies with well-defined phenotypes, however, it may well be possible to compare transcription/methylation profiles of eosinophils, e.g. between asthma subgroups.

One of the key assumptions on which the here described models rely is linearity (linear combinability) of the data. For methylation measured by beta values from 0 to 1 this can be seen as holding true, though, needless to say, beta values themselves represent an estimate of the overall methylation level in a given set of cells (there exists no "50 % CpG methylation" in a single DNA strand as methylation is either present or absent for a single CpG site). In case of expression data, specifically from microarray experiments, however, a restricted dynamic range and the effect of fold change compression [39], are complications to the linearity assumption. Moreover, not every cell or cell type necessarily contributes the overall same amount of RNA molecules to the extract from mixed-cell samples whereas the amount of contributed DNA (copies of the genome) is constant. As RNA is generally more unstable than DNA, the RNA amounts contributed by cells that are seen as contaminants in sputum (such as respiratory epithelium or squamous cells), which, due to their loss of original cellular integrity, might exhibit higher RNA degradation rates than viable immune cell fractions, might furthermore be unreliably approximated by their respective cellular proportions. However, linearity-based deconvolution on transcriptome data has proven before to provide valuable information [36, 40], so we expect the assumption to approximately hold true.
Wherever linearity is assumed, collinearity between predictors can cause problems. Given a large enough number of biological replicates, this is unlikely to occur for sputum differential cell counts. However, the biological interconnectedness of transcription and cell proportions (imagine a cytokine being strongly upregulated in one cell type and thereby enhancing attraction, proliferation or diapedesis of a second cell type with consequently increased presence) exhibits a further potential for transcript levels correlating with "wrong" cell types. Since the measured outcomes of transcriptional regulation (expression level) and chemotaxis (cell count) vary by different rates and relative, not absolute, cell counts are used as predictor in the linear regression, the risk of this resulting in false positives or negatives in deconvolved data can be seen as minor. Here, biological replication again benefits the accuracy of the deconvolution process.

Simultaneously, fulfilling the requirement of linearity for deconvolution comes with statistical compromises that investigators should be aware of: Performing methylation analysis on beta values instead of the logit-transformed counterpart, M values, causes heteroscedasticity [41] which has to be considered when applying parametric statistical testing procedures. This extends to expression data that are not $\log_2$-transformed. In addition, the assumption of normality is likely violated in both cases. A consequent workaround for these statistical limitations can be permutative testing, as it had been implemented before in the *csSAM* package [40] for expression deconvolution (the package has not been maintained lately and is outdated by now). However, as permutative testing comes with additional limitations itself, we decided to stay with parametric testing due to computational simplicity, speed and reproducibility whilst setting strict significance cutoffs. In any case, significance cutoffs should be chosen carefully and evaluated regarding their respective performance in the analyzed data. The relatively low number of genes that we identified to be simultaneously regulated on both the methylation and transcription level could therefore be caused by applying overly strict cutoffs. Nevertheless, future evaluation of deconvolution performance and determination of a best practice has to take place in larger data sets. Some limitations in this context, such as the limited dynamic range of microarrays or fold change compression, are meanwhile likely to resolve with application of sequencing-based methods.

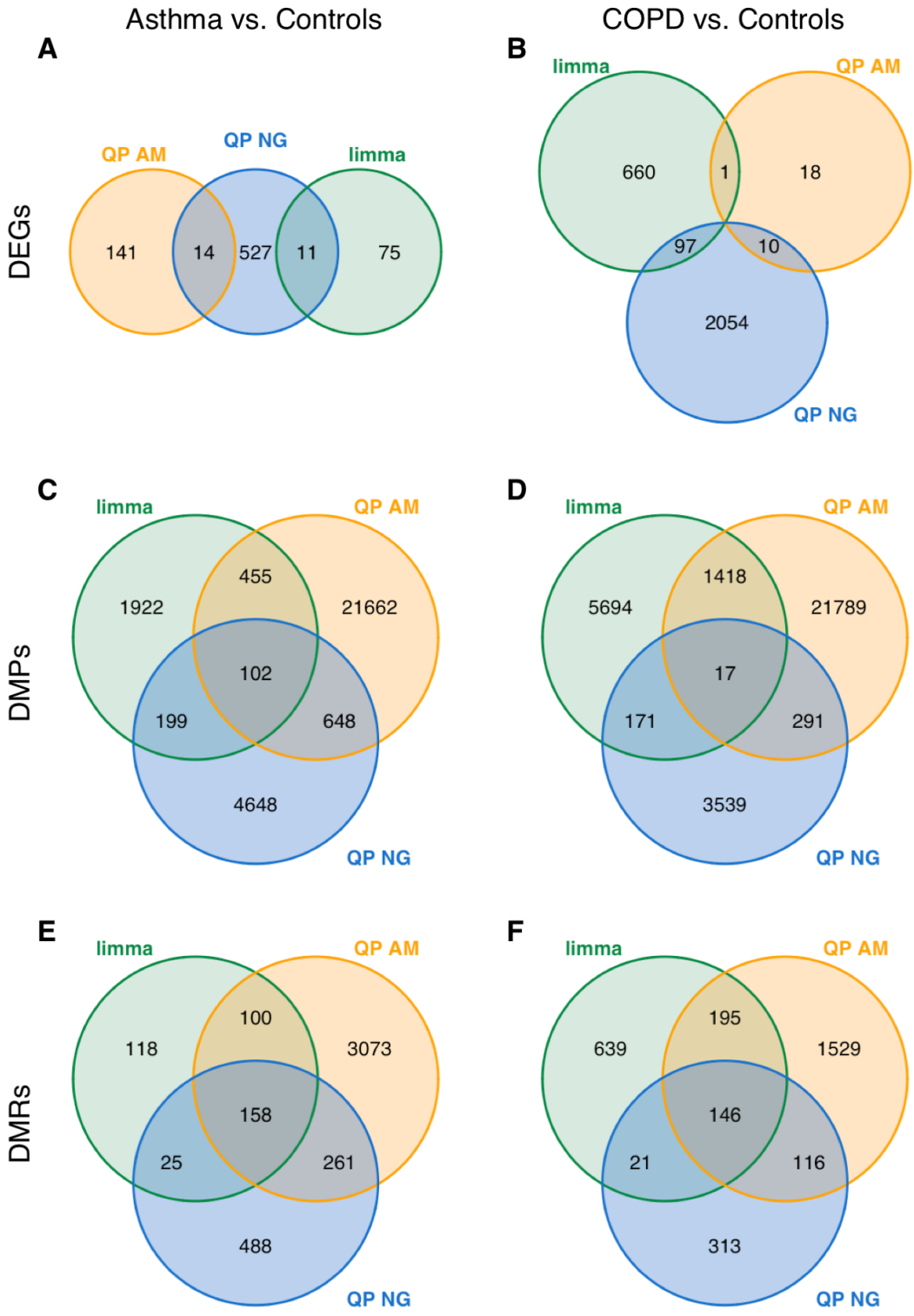

Figure S14: Venn diagram visualizations of differential expression and methylation analysis results before and after deconvolution: DEGs (A, B), DMPs (C, D) and genes associated with DMRs (E, F). Results of both comparisons, asthma vs. controls (A, C, E) and COPD vs. controls (B, D, F), are shown.
limma: conventional group comparison on mixed cell data (via the *limma* package). QP AM/NG: differential expression/methylation analysis on estimates after deconvolution by quadratic programming (QP), for alveolar macrophages (AM) and neutrophils (NG), respectively.

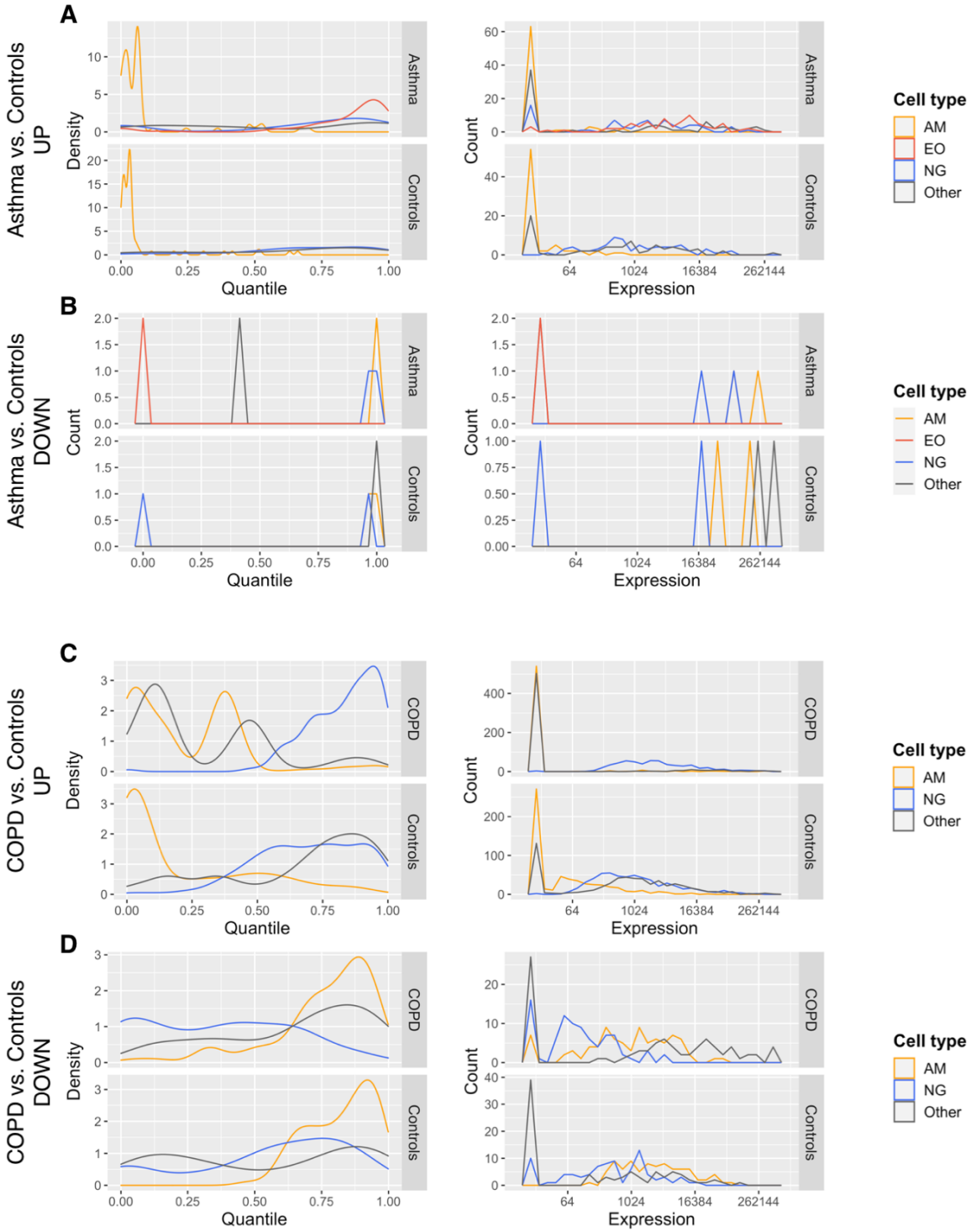

Figure S15: Deconvolved expression values and quantile ranks among estimated expression values of those transcripts that had been identified as DEGs in the mixed-cell analysis only (and not after QP deconvolution), upregulated (A) or downregulated (B) in asthma or upregulated (C) or downregulated (D) in COPD.
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: Sum of residual cell types (weighed intercept).
Of those transcripts identified to be differentially expressed in the asthma group, 12.8 % were still found after deconvolution by quadratic programming (QP). For the COPD group, the rate was 12.9 %. However, only 1.6 and 4.5 % of DEGs found after QP deconvolution were identified in the mixed-cell analyses, respectively (see also Figure S13).

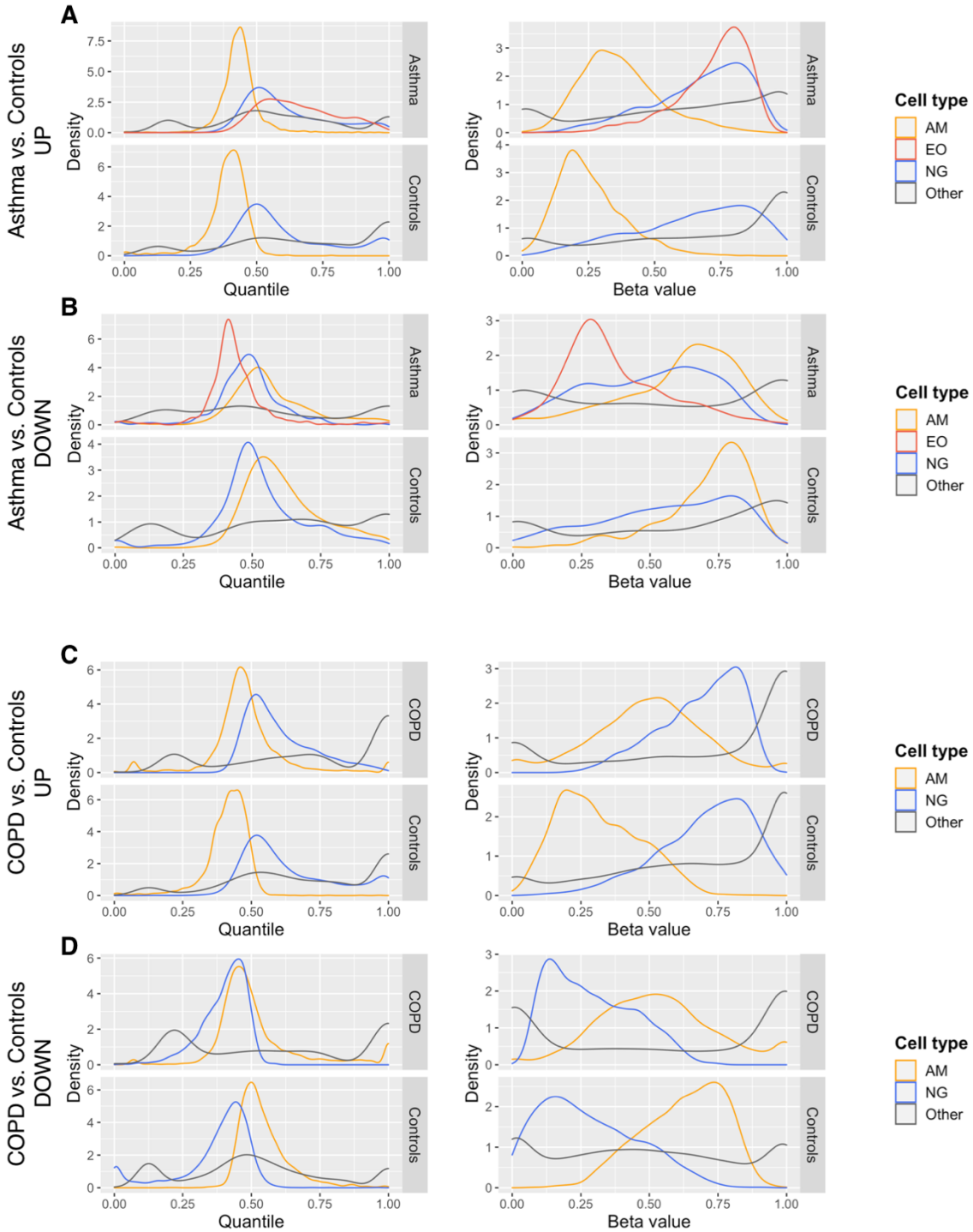

Figure S16: Deconvolved methylation beta values and quantile ranks among estimated beta values of those CpGs that had been identified as DMPs in the mixed-cell analysis only (and not after QP deconvolution), hypermethylated (A) or hypomethylated (B) in asthma or hypermethylated (C) or hypomethylated (D) in COPD.
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: Sum of residual cell types (weighed intercept).
As the overall distribution of beta values follows a bipolar pattern (see also Figure S6), the reader might find beta values easier to interpret than quantile ranks in this context.
Of those DMPs found by mixed-cell analysis in the asthma group, 28.2 % were still identified after deconvolution by quadratic programming (QP). For the COPD group, the rate was 22.2 %. However, only 2.7 and 5.9 % of DMPs found after QP deconvolution were identified in the mixed-cell analyses, respectively (see also Figure S13).

Enrichment results

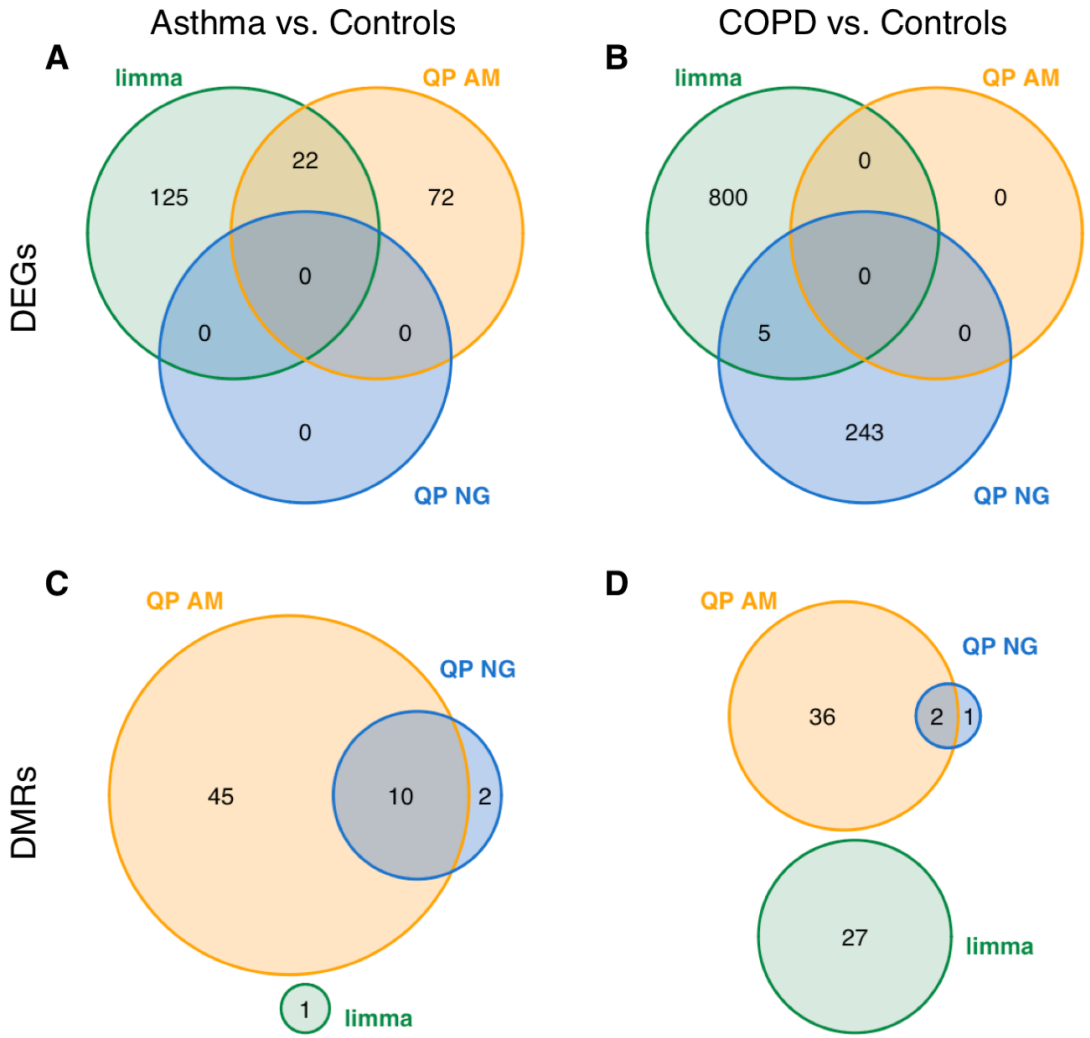

Figure S17: Venn diagram visualizations of Gene Ontology terms enriched in DEGs (A, B) and DMRs (C, D) in asthma (A, C) and COPD (B, D).
limma: conventional group comparison on mixed cell data (via the *limma* package). QP AM/NG: differential expression/methylation analysis on estimates after deconvolution by quadratic programming (QP), for alveolar macrophages (AM) and neutrophils (NG), respectively.

Integrative analysis

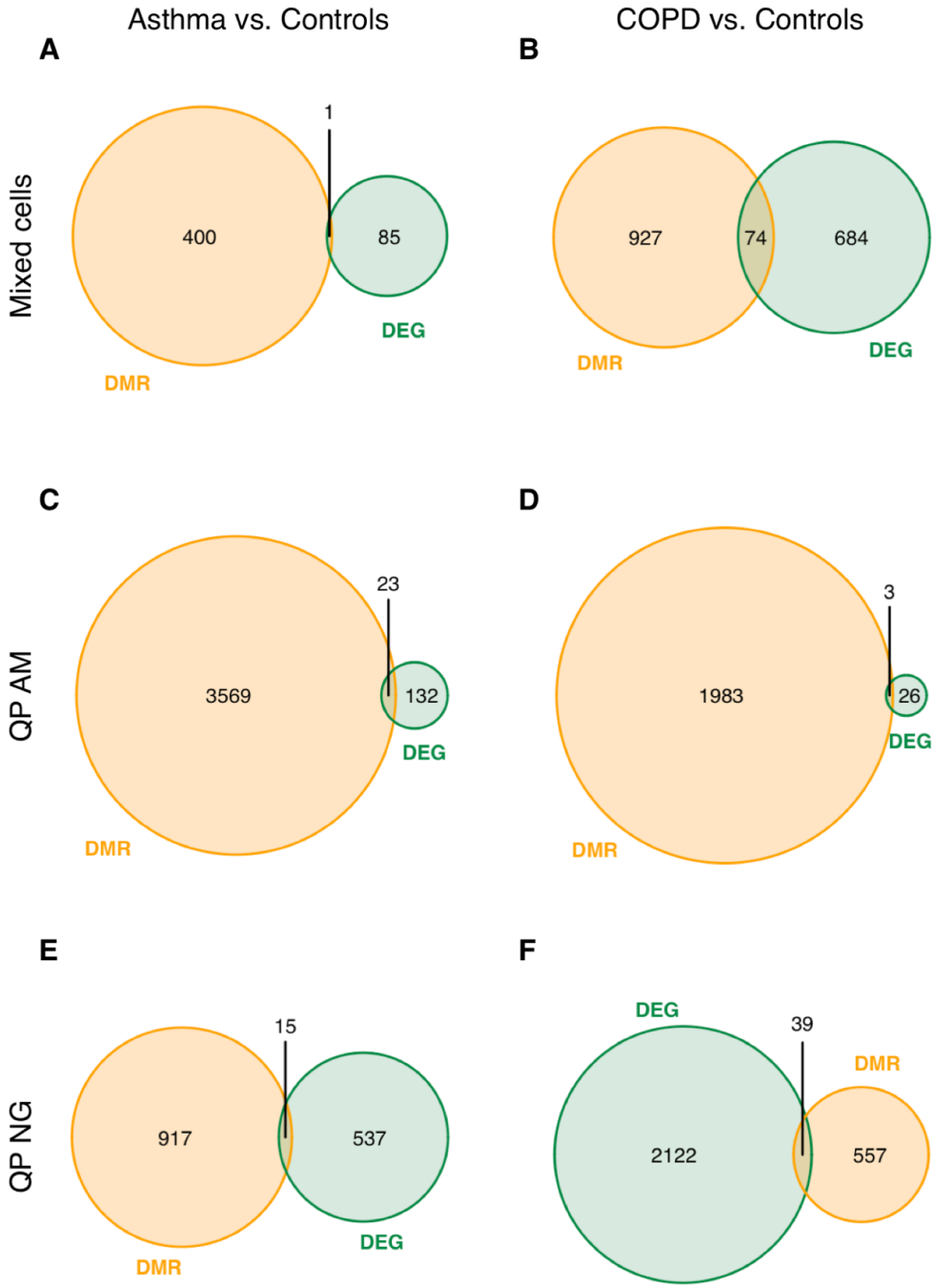

Figure S18: Venn diagram visualizations of DEGs and DMR-associated genes in asthma (A, C, E) and COPD (B, D, F) as identified via the analysis of mixed-cell data (A, B) and deconvolved cell type-specific data for macrophages (C, D) and neutrophils (E, F).
QP AM/NG: differential expression/methylation analysis on estimates after deconvolution by quadratic programming (QP), for alveolar macrophages (AM) and neutrophils (NG), respectively.

Data comparison

In a search for publicly available omics data sets generated from sputum or BAL on which we could validate our deconvolution's performance, we accessed the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/, accessed on 21 August 2020). We used the search strings "sputum asthma" (19 series hits), "BAL asthma" (25 series hits), "sputum COPD" (6 series hits) and "BAL COPD" (8 series hits) and checked the series hits manually for suitability. After excluding studies performed in mice, studies that did not provide data from sputum or BAL cells and studies that were conducted on purified cell samples, none of the remaining data sets was suitable to be supplied to our deconvolution approach due to missing information about the cellular composition of individual samples (although differential cell counts had been performed as derived from the publications associated with the data sets, e.g. [42]). Only one study provided detailed sputum differential cell counts with the omics data [43], however, the transcriptomic analysis was performed on purified macrophages and not on mixed-cell samples in this case.
We further screened the encountered data and associated publications for differential transcription analyses performed on purified cell types to which we could compare our results. In [44], alveolar macrophages were purified from asthmatics and checked for differential expression against healthy controls. However, the sample size of the study was very low (n = 5 per group), some asthmatics received antihistamines whilst none of the subjects needed therapy with inhaled corticosteroids (as opposed to our study, indicating that the overall asthma phenotype was milder) and, according to the manuscript, the authors did not apply a statistical correction for multiple testing to their results. Strictly speaking, we did not observe an appreciable overlap between the reported DEGs and DEGs identified in our analysis, but considering the aforementioned points, we don't interpret this as a falsification of our results. In [43] and [45], analyses were performed on purified macrophages from COPD patients. Since we were unfortunately not able to reliably estimate COPD macrophage profiles in our small data set, we refrained from comparing the reported findings to our results. In [46], a differential expression analysis comprising purified sputum macrophages in both asthma and COPD is presented. However, the purity of macrophages in the analyzed samples (a cutoff of > 50 % was applied by the authors) is rather low so we consider the data not macrophage-specific.

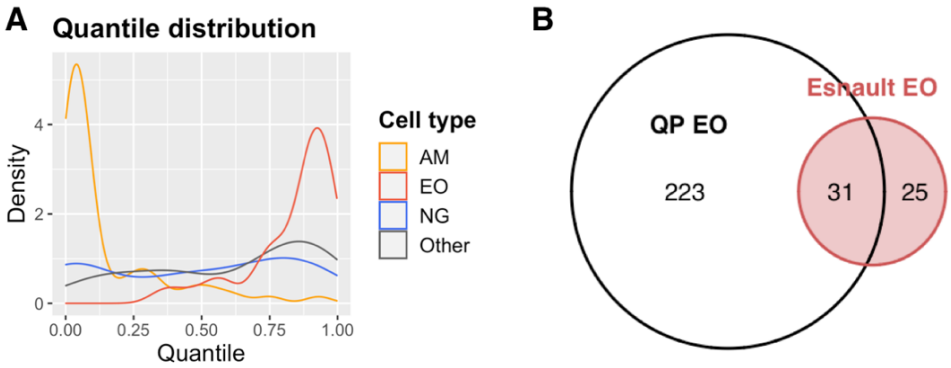

Figure S19: Quantile ranks of genes of the eosinophil set defined by Esnault et al. amongst the deconvolved expression estimates of each cell type (A) and overlap between the Esnault eosinophil gene set and genes differentiating eosinophils from macrophages and neutrophils in our data after deconvolution (B).
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. Other: Sum of residual cell types (weighed intercept). QP: deconvolution by quadratic programming.

Esnault et al. [47] defined a core set of 57 genes predominantly expressed by asthma eosinophils through transcriptome analyses in BAL and sputum in the context of allergen challenges which they validated in purified lung eosinophils subsequently. From these 57 genes, 56 were present in our expression data set after the initial data processing steps. The distribution of their respective quantile ranks among the expression estimates is shown in Figure S19A for each deconvolved cell type. Corresponding to the findings by Esnault et al., we observed the gene set to be primarily expressed in eosinophils.

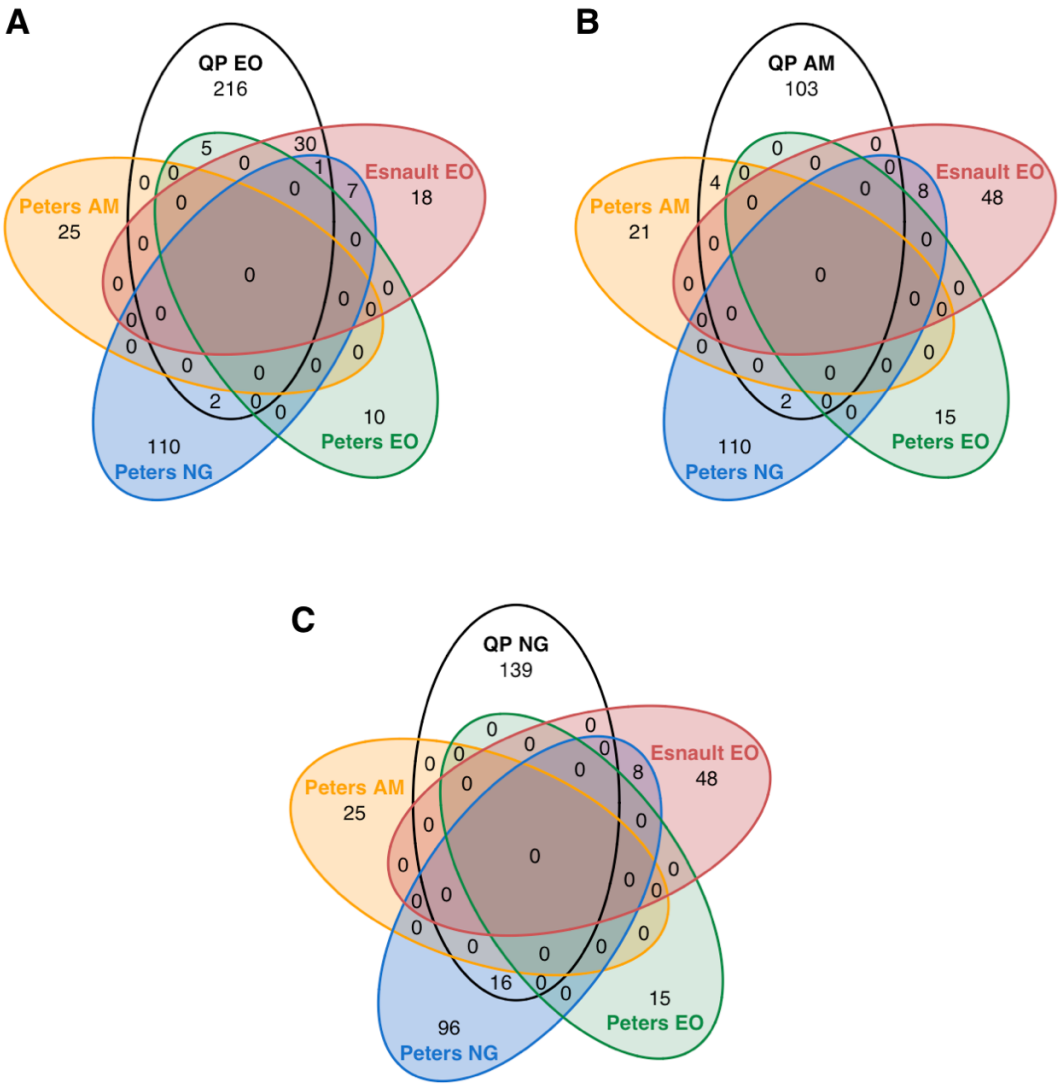

Figure S20: Overlaps between gene sets discriminating eosinophils (A), macrophages (B) and neutrophils (C) in the deconvolved asthma expression data and the eosinophil gene set defined by Esnault et al. as well as cell type-specific gene sets defined by Peters et al.
AM: alveolar macrophages. NG: neutrophil granulocytes. EO: eosinophils. QP: deconvolution by quadratic programming.

We further specified gene sets that differentiated eosinophils, neutrophils and macrophages in our deconvolved asthma expression data. To be associated with one of the cell types, genes had to exhibit a (positive) $\log_2 FC > 2$ in relation to both of the other estimated cell types at a BH-corrected (one-sided)

p value < 0.05. Note that in this constellation, statistical testing does not necessarily give meaningful results since the deconvolved estimates are derived from the same original data. Concordantly, the fold change cutoff was observed to be the predominant criterion for selection. Please note further that under these circumstances, the individual genes are not truly cell-specific (cell-exclusive) but rather estimated to be expressed at a higher level in one of the considered cell types than in both of the others. By this means, 31 out of 56 genes of the Esnault et al. gene set were found to discriminate eosinophils in our deconvolved estimates (see Figure S19B). In light of the multifaceted experimental setup employed by Esnault et al., including allergen challenges as well as application of the anti-IL5 antibody mepolizumab, we consider this to be a rather good overlap with our data of "steady state" eosinophils (no therapeutic/experimental intervention in our study).

Peters et al. previously derived cell-specific gene sets from transcriptome data that had been generated from blood and bone marrow samples and used these cell type-specific gene sets in the analysis of the sputum transcriptome in asthma [48]. Including these gene sets into our comparison, we observed predominant overlaps of our discriminating gene sets with the according cell types in the data by Peters et al. (see Figure S20). Furthermore, we observed that the eosinophil expression data provided by Esnault et al. might have been contaminated by gene expression from neutrophils in the context of the described allergen challenges (see overlap in Figure S20C). Otherwise, this implies that gene expression profiles of blood cells differ from those of cells of the same type in the lung environment. Further discussion concerning this issue is presented in the main manuscript.

References

1. Pizzichini, M.M.M., et al., *Safety of sputum induction.* European Respiratory Journal, 2002. **20**(37 suppl): p. 9s.

2. Weiszhar, Z. and I. Horvath, *Induced sputum analysis: step by step.* Breathe, 2013. **9**(4): p. 300.

3. R Core Team (2019). *R: A language and environment for statistical computing.* Available from: https://www.R-project.org/

4. Wickham, H., J. Hester, and W. Chang (2020). *devtools: Tools to Make Developing R Packages Easier. R package version 2.2.2.* Available from: https://CRAN.R-project.org/package=devtools

5. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nat Methods, 2015. **12**(2): p. 115-21.

6. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

7. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.

8. Hansen, K.D. (2016). *IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 0.6.0.*

9. Hansen, K.D. and M. Aryee (2012). *IlluminaHumanMethylation450kmanifest: Annotation for Illumina's 450k methylation arrays. R package version 0.4.0.*

10. Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.* Available from: https://CRAN.R-project.org/package=RColorBrewer

11. Bengtsson, H. (2020). *matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.56.0.* Available from: https://CRAN.R-project.org/package=matrixStats

12. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.

13. Peters, T. (2020). *DMRcatedata: Data Package for DMRcate. R package version 2.2.1.*

14. S original by Berwin A. Turlach, R port by Andreas Weingessel <Andreas.Weingessel@ci.tuwien.ac.at>, and Fortran contributions from Cleve Moler dpodi/LINPACK (2019). *quadprog: Functions to Solve Quadratic Programming Problems. R package version 1.5-8.* Available from: https://CRAN.R-project.org/package=quadprog

15. Warnes, G.R., B. Bolker, and T. Lumley (2018). *gtools: Various R Programming Tools. R package version 3.8.1.* Available from: https://CRAN.R-project.org/package=gtools

16. Arnholt, A.T. and B. Evans (2017). *BSDA: Basic Statistics and Data Analysis.* Available from: https://github.com/alanarnholt/BSDA, https://alanarnholt.github.io/BSDA/

17. Wickham, H., *The Split-Apply-Combine Strategy for Data Analysis.* Journal of Statistical Software, 2011. **40**(1): p. 29.

18. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics, 2012. **16**(5): p. 284-7.

19. Carlson, M. (2019). *org.Hs.eg.db: Genome wide annotation for Human. R package version 3.10.0.* Available from: https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html

20. Dolgalev, I. (2019). *msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.0.1.* Available from: https://CRAN.R-project.org/package=msigdbr

21. Wickham, H. and L. Henry (2020). *tidyr: Tidy Messy Data. R package version 1.0.2.* Available from: https://CRAN.R-project.org/package=tidyr

22. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York. Available from: https://ggplot2.tidyverse.org

23. Wilke, C.O. (2019). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.0.0.* Available from: https://CRAN.R-project.org/package=cowplot

24. Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20.* Available from: https://CRAN.R-project.org/package=VennDiagram

25. Onuchic, V. (2019). *EDec: Cell type specific analysis of complex tissues through Epigenomic Deconvolution. R package version 0.9.* Available from: https://github.com/BRL-BCM/EDec

26. Abdel-Aziz, M.I., et al., *Omics for the future in asthma.* Semin Immunopathol, 2020. **42**(1): p. 111-126.

27. Auffray, C., et al., *An integrative systems biology approach to understanding pulmonary diseases.* Chest, 2010. **137**(6): p. 1410-6.

28. Peters, M.C., et al., *Measures of gene expression in sputum cells can identify T2-high and T2-low subtypes of asthma.* J Allergy Clin Immunol, 2013.

29. Hrdlickova, R., M. Toloue, and B. Tian, *RNA-Seq methods for transcriptome analysis.* Wiley Interdiscip Rev RNA, 2017. **8**(1).

30. Opitz, L., et al., *Impact of RNA degradation on gene expression profiling.* BMC Med Genomics, 2010. **3**: p. 36.

31. Gallego Romero, I., et al., *RNA-seq: impact of RNA degradation on transcript quantification.* BMC Biol, 2014. **12**: p. 42.

32. Liu, X., et al., *Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review.* Front Bioeng Biotechnol, 2019. **7**: p. 358.

33. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-93.

34. Cohen, J., *A power primer.* Psychol Bull, 1992. **112**(1): p. 155-9.

35. Onuchic, V., et al., *Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types.* Cell Rep, 2016. **17**(8): p. 2075-2086.

36. Gong, T., et al., *Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples.* PLoS One, 2011. **6**(11): p. e27156.

37. Goldfarb, D. and A. Idnani, *A numerically stable dual method for solving strictly convex quadratic programs.* Mathematical Programming, 1983. **27**(1): p. 1-33.

38. Cohen, A., *Comparing Regression Coefficients Across Subsamples: A Study of the Statistical Test.* Sociological Methods & Research, 1983. **12**(1): p. 77-94.

39. Dallas, P.B., et al., *Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR -- how well do they correlate?* BMC Genomics, 2005. **6**: p. 59.

40. Shen-Orr, S.S., et al., *Cell type-specific gene expression differences in complex tissues.* Nat Methods, 2010. **7**(4): p. 287-9.

41. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC Bioinformatics, 2010. **11**: p. 587.

42. Kuo, C.S., et al., *T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED.* Eur Respir J, 2017. **49**(2).

43. Morrow, J.D., et al., *RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD.* Respir Res, 2019. **20**(1): p. 65.

44. Madore, A.M., et al., *Alveolar macrophages in allergic asthma: an expression signature characterized by heat shock protein pathways.* Hum Immunol, 2010. **71**(2): p. 144-50.

45. O'Beirne, S.L., et al., *Alveolar Macrophage Immunometabolism and Lung Function Impairment in Smoking and Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 2020. **201**(6): p. 735-739.

46. Paplińska-Goryca, M., et al., *Genetic characterization of macrophages from induced sputum of patients with asthma and chronic obstructive pulmonary disease.* Pol Arch Intern Med, 2018. **128**(9): p. 559-562.

47. Esnault, S., et al., *Identification of genes expressed by human airway eosinophils after an in vivo allergen challenge.* PLoS One, 2013. **8**(7): p. e67560.

48. Peters, M.C., et al., *A Transcriptomic Method to Determine Airway Immune Dysfunction in T2-High and T2-Low Asthma.* Am J Respir Crit Care Med, 2019. **199**(4): p. 465-477.

3 Darstellung der Publikation

Das Asthma bronchiale und die chronisch-obstruktive Lungenerkrankung (Chronic Obstructive Pulmonary Disease, COPD) zählen zu den prävalentesten chronischen Atemwegserkrankungen und zeichnen sich durch eine chronische Atemwegsinflammation aus (Soriano et al., 2017). So unterschiedlich die klinischen Präsentationen der Erkrankungen sind, so vielfältig und teils variabel stellen sich die dabei zugrundeliegenden Krankheitsprozesse dar (Barnes, 2017). Mit der Einführung von Hochdurchsatztechnologien (*engl.* „Omics“) zur Analyse zellulärer Regulation und Funktion (u.a. (Epi)genomik, Transkriptomik) wuchs das Verständnis der zugrundeliegenden Pathologien substanziell bis hin zur Identifikation molekularer Phänotypen („Endotypen“) (Kuo et al., 2017; Singh et al., 2014) und der Modellierung der Wirkung von Therapeutika (Govoni et al., 2020). Induziertes Sputum hat sich durch die Noninvasivität seiner Gewinnung und seinen immanenten Gehalt an Immunzellen, die die inflammatorischen Prozesse der Atemwege reflektieren, als für wissenschaftliche Analysen gut verfügbares und nutzbares Biomaterial hervorgetan. Bislang beschränkten sich Hochdurchsatzanalysen an Sputumzellen überwiegend auf Analysen des Transkriptoms (Abdel-Aziz et al., 2020; Wheelock et al., 2013); ergänzende epigenetische Untersuchungen waren Gewebe- oder Blutproben vorbehalten (Belinsky et al., 2002; Kabesch und Tost, 2020; Yang et al., 2017) und konnten gleichwohl mit klinischen Phänotypen und Endotypen assoziiert werden (Nicodemus-Johnson et al., 2016).

In der hier dargestellten experimentellen und explorativ ausgerichteten Arbeit führten wir erstmalig zusätzlich zu einer Transkriptom- eine genomweite Methylomanalyse an induziertem Sputum bei Asthma und COPD durch. In diesem Zusammenhang verwendeten wir bei der Datenanalyse erstmals einen angepassten „Deconvolution“-Algorithmus (*engl.* für „Entfaltung“), um aus gemischtzelligen Sputumproben auf regulative Veränderungen einzelner Zelltypen Rückschlüsse ziehen zu können.

Induziertes Sputum von Patienten mit eosinophilem Asthma bronchiale (n = 9), COPD (n = 10) sowie gesunden Kontrollen (n = 10) wurde im Rahmen der Kohortenstudien ALLIANCE (Fuchs et al., 2018) und COSYCONET (Karch et al., 2016) gewonnen und nach Lagerung in zugehörigen Biomaterialbanken für die hier beschriebene Analyse zur Verfügung gestellt. Die Auswahl der Sputumproben richtete sich hierbei nach im Manuskript näher beschriebenen Kriterien, die innerhalb der Krankheitsentitäten eine phänotypische Homogenität gewährleisteten. Aus den Proben wurden simultan DNA (Desoxyribonukleinsäure, *engl.*

deoxyribonucleic acid) und RNA (Ribonukleinsäure, *engl.* ribonucleic acid) extrahiert und zur Array-basierten Charakterisierung der genomweiten Transkription und Methylierung genutzt.
Die Extraktion hochqualitativer RNA nach mittel- bis langfristiger Probenlagerung stellt angesichts der im Sputum enthaltenen degradierenden Enzyme und reaktiver Sauerstoffspezies eine Herausforderung dar (Peters et al., 2013). In unserer Analyse zeigte sich entsprechend, dass die Biobanklagerung der Sputumproben in variablem Ausmaß zu einer Beeinträchtigung der Integrität der extrahierbaren RNA geführt hatte. Bisherige Arbeiten legten eine Verwendbarkeit von partiell degradierter RNA mit einer RIN (RNA Integrity Number) > 5 nahe (Opitz et al., 2010; Peters et al., 2013). In diesem Bereich zeigte sich in unserem Datensatz jedoch eine substanzielle, systematische Beeinflussung der Messergebnisse, sodass wir folgend zwei Ansätze zur *in silico*-Korrektur evaluierten. Unter diesen stellte sich die Anwendung eines separaten Regressionsschritts zur Korrektur der integritätsassoziierten Effekte als überlegen heraus und führte zu einer verbesserten Detektion differenziell exprimierter Gene, übereinstimmend mit einer Arbeit von Gallego Romero et al. (2014). Die genomische Methylierung zeigte sich gegenüber der mehrjährigen Materiallagerung deutlich robuster und ließ in unserem Datensatz keine systematische Beeinflussung durch die Lagerungsbedingungen erkennen. Darüber hinaus indizierte eine Hauptkomponentenanalyse unserer Datensätze eine bessere Differenzierung von asthmatischem und gesundem Sputum anhand der genomischen Methylierung als anhand des Transkriptoms.
Dies verdeutlicht zunächst, dass die Charakterisierung des Sputummethyloms methodisch in bestehende Workflows molekularer Sputumanalyse eingebunden werden kann und darüber hinaus verspricht, wesentlich zur Endotypisierung chronischer pulmonaler Inflammation beitragen zu können.

Die bereits angeführte Noninvasivität der Sputuminduktion erlaubt auch im Rahmen groß angelegter Kohortenstudien eine umfangreiche Materialgewinnung; neben den ALLIANCE- und COSYCONET-Kohorten z.B. in der U-BIOPRED-Kohorte (Shaw et al., 2015). Dabei ist bis zur Einlagerung ein umfangreiches Prozessieren der gewonnenen Sputa notwendig, um muzinöse Bestandteile zu entfernen und die enthaltenen Immunzellen aufzureinigen.
Hierbei finden sich Alveolarmakrophagen, neutrophile und eosinophile Granulozyten, Lymphozyten und darüber hinaus respiratorische und Plattenepithelien in variierender Quantität. Hochdurchsatzanalysen gemischtzelliger Sputumproben waren bislang darauf beschränkt, als molekularer „Fingerabdruck“ (bzw. molekulare Biomarker) interpretiert zu werden und erlaubten nicht ohne Weiteres Rückschlüsse auf regulative und

pathophysiologische Prozesse innerhalb der jeweiligen Zelltypen. Eine separate Analyse der Zelltypen erforderte eine umfangreichere Aufarbeitung des Materials und konnte bislang nur in kleinerem Umfang an Proben bronchoalveolärer Lavage realisiert werden (Esnault et al., 2013; Morrow et al., 2019; Poliska et al., 2011).
Während der differenzialzytologische Befund von Sputumproben genutzt werden kann, um Erkrankungen und Krankheits-Endotypen voneinander abzugrenzen (Medrek et al., 2017; Tiotiu, 2018), stellt die variable Zusammensetzung einen essenziellen Störfaktor in der Hochdurchsatzanalyse gemischtzelliger Proben dar (Taube und Reuter, 2019). Frühe Arbeiten ließen diesen Umstand in der Datenanalyse meist außer Acht (z.B. Baines et al. (2011)), während neuere Arbeiten für die variierende Zellzusammensetzung teils per kovariater Analyse (z.B. Weathington et al. (2019)) korrigierten.

Um den Informationsgewinn molekularer Sputumanalysen zu erhöhen, entwickelten wir einen auf unser experimentelles Setup zugeschnittenen Deconvolution-Algorithmus, der es in der Analyse gemischtzelliger Sputumproben erstmals erlaubt, auf die Expression und Methylierung einzelner Zelltypen schließen zu können. Eine genaue mathematische Darstellung sowie ergänzende Erläuterungen zu den statistischen Eigenschaften und der Implementation sind der Publikation sowie dem beigefügten Supplement zu entnehmen.
In Kürze dargelegt basiert die hier angewandte Deconvolution auf der Annahme, dass innerhalb einer Krankheitsgruppe (bzw. eines Endotyps) über die Individuen hinweg ein homogenes Transkriptions- bzw. Methylierungsprofil innerhalb eines jeden der im Sputum apparenten Zelltypen besteht. Dies reduziert nach vorheriger, wohldefinierter Einteilung der analysierten Proben (in unserem Fall sichergestellt durch die bei Selektion der Sputumproben angewandten Kriterien) den Vorgang der Schätzung der Zelltyp-spezifischen Expressions- und Methylierungsmuster zu einem multiplen Regressionsproblem, sofern die genaue zelluläre Zusammensetzung jeder einzelnen Probe bekannt ist. Letzteres war durch die im Rahmen von ALLIANCE und COSYCONET angefertigten Sputum-Differenzialzytologien in hinreichender Auflösung gewährleistet.
Bei der mathematischen Optimierung der Lösung dieses Regressionsproblems (der Schätzung der Zelltyp-spezifischen Expression/Methylierung) muss ferner in Betracht gezogen werden, dass Messwerte der Expression und Methylierung innerhalb biologisch (und technisch) vorgegebener Grenzen liegen: das Expressionsniveau eines Gens kann beispielsweise keinen negativen Wert annehmen. Diesem Umstand begegneten wir mit der Formulierung eines quadratischen Programms (quadratischer Programmierung, *engl.* quadratic programming; siehe

auch Goldfarb und Idnani (1983)). Dieser Ansatz hatte sich bereits zuvor in anderer Form und Kontext erfolgreich gezeigt (Onuchic et al., 2016).
In diesem Zusammenhang beobachteten wir, dass die, verglichen mit gesunden Kontrollen, in gemischtzelligem asthmatischem Sputum stärker exprimierten Gene eine prädominante (geschätzte) Expression in eosinophilen Granulozyten aufwiesen und die in COPD-Sputum hochreguliert erscheinenden Gene prädominant in neutrophilen Granulozyten exprimiert wurden. Vor dem Hintergrund der beobachteten zellulären Zusammensetzung (das asthmatische Sputum ist durch einen höheren eosinophilen, das COPD-Sputum durch einen höheren neutrophilen Anteil gekennzeichnet) verdeutlicht dies die Notwendigkeit der Anwendung einer *in silico*-Korrektur bei der Analyse gemischtzelliger Sputumproben.

In unserem Datensatz lieferte die angewandte Deconvolution mathematisch verlässliche Schätzungen der zellspezifischen Expression und Methylierung für Makrophagen und Neutrophile in allen drei untersuchten Gruppen (Asthma, COPD und Kontrollen) sowie für Eosinophile in der Asthma-Gruppe. Für größere Datensätzen mit höherer biologischer Replikation ist davon auszugehen, dass auch weitere Zelltypen verlässlich eingeschätzt und damit untereinander verglichen werden können (v.a. Lymphozyten).

In unserer Analyse konnten wir auf diese Weise Kandidatengene identifizieren, die ohne eine Deconvolution nicht als differenziell methyliert bzw. exprimiert detektiert worden wären:
In Makrophagen zeigten sich beispielsweise die Gene IL23A (Interleukin-23-Alpha) und CCL24 (Chemokin-C-C-Motiv-Ligand-24, auch bekannt als Eotaxin-2) innerhalb von bei Asthma differenziell methylierten Regionen liegend. Für diese konnte eine Rolle bei der Makrophagen-Polarisation (Orecchioni et al., 2019) bzw. mikrobieller Interaktion und Eosinophilen-Stimulation (Cheng et al., 2017; Palikhe et al., 2010) gezeigt werden. Neutrophile in asthmatischem Sputum zeigten interessanterweise eine differenzielle Methylierung im Bereich des IL5RA (Interleukin-5-Rezeptor-Alpha)-Gens, dessen Expression in unserem Datensatz zwar ausschließlich in Eosinophilen beobachtet wurde (und lange Zeit ausschließlich für Eosinophile beschrieben worden war), im Kontext von schwerem Asthma jedoch kürzlich in einer pädiatrischen Kohorte auch für Neutrophile demonstriert werden konnte (Gorski et al., 2019). Darüber hinaus beobachteten wir eine differenzielle Methylierung im Bereich einer Reihe von HLA (Human Leukocyte Antigen)-Genen in Makrophagen bei Asthma und COPD, die sowohl mit MHC (Major Histocompatibility Complex)-Klasse-I- als auch Klasse-II-Molekülen assoziiert werden konnten. Eine konkordante Änderung der Expression der

betreffenden Gene konnte in unserem Datensatz nicht beobachtet werden, Veränderungen der Methylierung der korrespondierenden genomischen Bereiche wurden jedoch bereits z.B. im Kontext von atopischem Asthma gezeigt (Lund et al., 2018). Ebenfalls vorbeschrieben ist eine differenzielle Methylierung des SMAD3 (SMAD family member 3)-Gens bei Asthma, welches wir in unserem Datensatz konkordant innerhalb einer differenziell methylierten Region sowohl bei Asthma, als auch interessanterweise bei COPD, verorten konnten.
In den Transkriptomsignaturen beobachteten wir in Neutrophilen im asthmatischen Sputum unter anderem eine gesteigerte Expression von IL4R (Interleukin-4-Rezeptor), welchem eine Rolle in der Regulation neutrophiler Apoptose im Rahmen einer Typ-2-Immunantwort beigemessen wird (Harris et al., 2019). Zusätzlich beobachteten wir eine höhere Expression von CXCL2 (Chemokin-C-X-C-Motiv-Ligand-2), das in die autokrine Regulation von Neutrophilen eingebunden ist (Li et al., 2016).

Wir verglichen unsere Ergebnisse mit öffentlich verfügbaren Datensätzen und konnten hierbei eine gute Übereinstimmung der nach Deconvolution extrapolierten Genexpression mit von Esnault et al. (2013) publizierten Daten zur Genexpression in Eosinophilen bei Asthma bronchiale feststellen. Im Weiteren beobachteten wir eine partielle Diskrepanz unserer Genexpressionsdaten mit zuvor anhand von Blut- und Knochenmarkdatensätzen definierten, Zelltyp-repräsentativen Expressionsprofilen (Peters et al., 2019). Letztere zeigten ebenfalls eine Diskrepanz zu den an Zellen pulmonalen Ursprungs von Esnault et al. (2013) erarbeiteten Daten.
In den letzten Jahren sind mehrere weitere, auf die jeweilige Applikation maßgeschneiderte, Omics-Deconvolution-Algorithmen entwickelt worden, die auf der Verwendung Zelltyp-spezifischer Referenzprofile basieren und keine vorherige differenzialzytologische Evaluation der Proben voraussetzen (Avila Cobos et al., 2018). Von diesen ist, soweit den Autoren der hier dargestellten Arbeit bekannt, bislang keiner an Material pulmonalen Ursprungs zum Einsatz gekommen und vor der Applikation erscheint es im Licht der o.g. Diskrepanzen daher unabdingbar, zunächst Zelltyp-Referenzprofile im pulmonalen Kontext zu definieren. Dies wird insbesondere davon unterstrichen, dass Alveolarmakrophagen nach aktuellen Erkenntnissen eine von sich aus Monozyten ableitenden Makrophagen distinkte Zellentität darstellen (McQuattie-Pimentel et al., 2018).
Unser auf der differenziellen Zytologie der Sputumproben basierender Deconvolution-Ansatz umgeht somit ein ggf. erhebliches Bias, das mit der Verwendung extrapulmonaler Referenzdaten verbunden wäre.

Die hier dargestellte Arbeit demonstriert erstmals, dass die Charakterisierung des Sputum-Methyloms in etablierte Elemente der molekularen Hochdurchsatzanalyse von Sputumproben und damit in das molekulare Phenotyping bzw. Endotyping chronisch-inflammatorischer Lungenerkrankungen wie Asthma und COPD eingebunden werden kann. Besonders vor dem Hintergrund, dass im Rahmen dieser Erkrankungen umweltbedingte Faktoren die Krankheitsentstehung und den Verlauf beeinflussen (Bae et al., 2020; Yang et al., 2017), eröffnet die Analyse epigenetischer Profile und die damit verbundene Abbildung mittel- bis langfristiger regulativer Veränderungen (Hasin et al., 2017; Luo et al., 2018) essenzielle Perspektiven in der Sputumanalyse. Die hier beschriebenen *in silico*-Methoden tragen dazu bei, im Rahmen der Analyse von Material aus Biobanken auftretende Herausforderungen zu bewältigen, indem die Qualität und Aussagekraft von Transkriptomdaten aus partiell degradiertem Material verbessert wird sowie mit dem beschriebenen Deconvolution-Algorithmus schließlich regulative Veränderungen auf zellulärer Ebene in gemischtzelligen Proben besser exploriert und von durch die variable zelluläre Zusammensetzung bedingtem Bias befreit werden können.

3.1 Literaturverzeichnis

Abdel-Aziz MI, Neerincx AH, Vijverberg SJ, Kraneveld AD, Maitland-van der Zee AH (2020). *Omics for the future in asthma*. Semin Immunopathol 42, 111-126.

Avila Cobos F, Vandesompele J, Mestdagh P, De Preter K (2018). *Computational deconvolution of transcriptomics data from mixed cell populations*. Bioinformatics 34, 1969-1979.

Bae DJ, Jun JA, Chang HS, Park JS, Park CS (2020). *Epigenetic Changes in Asthma: Role of DNA CpG Methylation*. Tuberc Respir Dis (Seoul) 83, 1-13.

Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG (2011). *Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples*. J Allergy Clin Immunol 127, 153-160, 160.e151-159.

Barnes PJ (2017). *Cellular and molecular mechanisms of asthma and COPD*. Clin Sci (Lond) 131, 1541-1558.

Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, Moots PP, Lechner JF, Stidley CA, Crowell RE (2002). *Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers*. Cancer Res 62, 2370-2377.

Cheng M, Chen Y, Wang L, Chen W, Yang L, Shen G, Xu T, Shen G, Tian Z, Hu S (2017). *Commensal microbiota maintains alveolar macrophages with a low level of CCL24 production to generate anti-metastatic tumor activity*. Sci Rep 7, 7471-7471.

Esnault S, Kelly EA, Schwantes EA, Liu LY, DeLain LP, Hauer JA, Bochkov YA, Denlinger LC, Malter JS, Mathur SK, Jarjour NN (2013). *Identification of genes expressed by human airway eosinophils after an in vivo allergen challenge*. PLoS One 8, e67560.

Fuchs O, Bahmer T, Weckmann M, Dittrich AM, Schaub B, Rösler B, Happle C, Brinkmann F, Ricklefs I, König IR, Watz H, Rabe KF, Kopp MV, Hansen G, von Mutius E (2018). *The all age asthma cohort (ALLIANCE) - from early beginnings to chronic disease: a longitudinal cohort study*. BMC Pulm Med 18, 140.

Gallego Romero I, Pai AA, Tung J, Gilad Y (2014). *RNA-seq: impact of RNA degradation on transcript quantification*. BMC Biol 12, 42.

Goldfarb D, Idnani A (1983). *A numerically stable dual method for solving strictly convex quadratic programs*. Mathematical Programming 27, 1-33.

Gorski SA, Lawrence MG, Hinkelman A, Spano MM, Steinke JW, Borish L, Teague WG, Braciale TJ (2019). *Expression of IL-5 receptor alpha by murine and human lung neutrophils*. PLoS One 14, e0221113-e0221113.

Govoni M, Bassi M, Vezzoli S, Lucci G, Emirova A, Nandeuil MA, Petruzzelli S, Jellema GL, Afolabi EK, Colgan B, Leaker B, Kornmann O, Beeh KM, Watz H, Singh D (2020). *Sputum and blood transcriptomics characterisation of the inhaled PDE4 inhibitor CHF6001 on top of triple therapy in patients with chronic bronchitis*. Respir Res 21, 72.

Harris AJ, Mirchandani AS, Lynch RW, Murphy F, Delaney L, Small D, Coelho P, Watts ER, Sadiku P, Griffith D, Dickinson RS, Clark E, Willson JA, Morrison T, Mazzone M, Carmeliet P, Ghesquiere B, O'Kane C, McAuley D, Jenkins SJ, Whyte MKB, Walmsley SR (2019). *IL4Ralpha Signaling Abrogates Hypoxic Neutrophil Survival and Limits Acute Lung Injury Responses In Vivo*. Am J Respir Crit Care Med 200, 235-246.

Hasin Y, Seldin M, Lusis A (2017). *Multi-omics approaches to disease*. Genome Biology 18, 83.

Kabesch M, Tost J (2020). *Recent findings in the genetics and epigenetics of asthma and allergy*. Seminars in Immunopathology 42, 43-60.

Karch A, Vogelmeier C, Welte T, Bals R, Kauczor HU, Biederer J, Heinrich J, Schulz H, Gläser S, Holle R, Watz H, Korn S, Adaskina N, Biertz F, Vogel C, Vestbo J, Wouters EF, Rabe KF, Söhler S, Koch A, Jörres RA (2016). *The German COPD cohort COSYCONET: Aims, methods and descriptive analysis of the study population at baseline*. Respir Med 114, 27-37.

Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, Sousa A, Corfield J, Djukanovic R, Lutter R, Sterk PJ, Auffray C, Guo Y, Adcock IM, Chung KF, Group UBS (2017). *T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED*. Eur Respir J 49.

Li JL, Lim CH, Tay FW, Goh CC, Devi S, Malleret B, Lee B, Bakocevic N, Chong SZ, Evrard M, Tanizaki H, Lim HY, Russell B, Renia L, Zolezzi F, Poidinger M, Angeli V, St John AL, Harris JE, Tey HL, Tan SM, Kabashima K, Weninger W, Larbi A, Ng LG (2016). *Neutrophils Self-Regulate Immune Complex-Mediated Cutaneous Inflammation through CXCL2*. J Invest Dermatol 136, 416-424.

Lund RJ, Osmala M, Malonzo M, Lukkarinen M, Leino A, Salmi J, Vuorikoski S, Turunen R, Vuorinen T, Akdis C, Lahdesmaki H, Lahesmaa R, Jartti T (2018). *Atopic asthma after rhinovirus-induced wheezing is associated with DNA methylation change in the SMAD3 gene promoter*. Allergy 73, 1735-1740.

Luo C, Hajkova P, Ecker JR (2018). *Dynamic DNA methylation: In the right place at the right time*. Science 361, 1336.

McQuattie-Pimentel AC, Budinger GRS, Ballinger MN (2018). *Monocyte-derived Alveolar Macrophages: The Dark Side of Lung Repair?* Am J Respir Cell Mol Biol 58, 5-6.

Medrek SK, Parulekar AD, Hanania NA (2017). *Predictive Biomarkers for Asthma Therapy*. Current Allergy and Asthma Reports 17, 69.

Morrow JD, Chase RP, Parker MM, Glass K, Seo M, Divo M, Owen CA, Castaldi P, DeMeo DL, Silverman EK, Hersh CP (2019). *RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD*. Respir Res 20, 65.

Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, Sperling AI, Solway J, White SR, Nobrega MA, Nicolae DL, Gilad Y, Ober C (2016). *DNA methylation in lung cells is associated with asthma endotypes and genetic risk*. JCI Insight 1, e90151.

Onuchic V, Hartmaier RJ, Boone DN, Samuels ML, Patel RY, White WM, Garovic VD, Oesterreich S, Roth ME, Lee AV, Milosavljevic A (2016). *Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types*. Cell Rep 17, 2075-2086.

Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, Ghadimi BM, Beissbarth T, Gaedcke J (2010). *Impact of RNA degradation on gene expression profiling*. BMC Med Genomics 3, 36.

Orecchioni M, Ghosheh Y, Pramod AB, Ley K (2019). *Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages*. Frontiers in Immunology 10.

Palikhe NS, Kim SH, Cho BY, Ye YM, Choi GS, Park HS (2010). *Genetic variability in CRTH2 polymorphism increases eotaxin-2 levels in patients with aspirin exacerbated respiratory disease*. Allergy 65, 338-346.

Peters MC, Mekonnen ZK, Yuan S, Bhakta NR, Woodruff PG, Fahy JV (2013). *Measures of gene expression in sputum cells can identify T2-high and T2-low subtypes of asthma*. J Allergy Clin Immunol.

Peters MC, Ringel L, Dyjack N, Herrin R, Woodruff PG, Rios C, O'Connor B, Fahy JV, Seibold MA (2019). *A Transcriptomic Method to Determine Airway Immune Dysfunction in T2-High and T2-Low Asthma*. Am J Respir Crit Care Med 199, 465-477.

Poliska S, Csanky E, Szanto A, Szatmari I, Mesko B, Szeles L, Dezso B, Scholtz B, Podani J, Kilty I, Takacs L, Nagy L (2011). *Chronic obstructive pulmonary disease-specific gene expression signatures of alveolar macrophages as well as peripheral blood monocytes overlap and correlate with lung function*. Respiration 81, 499-510.

Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, Pandis I, Bansal AT, Bel EH, Auffray C, Compton CH, Bisgaard H, Bucchioni E, Caruso M, Chanez P, Dahlén B, Dahlen SE, Dyson K, Frey U, Geiser T, Gerhardsson de Verdier M, Gibeon D, Guo YK, Hashimoto S, Hedlin G, Jeyasingham E, Hekking PP, Higenbottam T, Horváth I, Knox AJ, Krug N, Erpenbeck VJ, Larsson LX, Lazarinis N, Matthews JG, Middelveld R, Montuschi P, Musial J, Myles D, Pahus L, Sandström T, Seibold W, Singer F, Strandberg K, Vestbo J, Vissing N, von Garnier C, Adcock IM, Wagers S, Rowe A, Howarth P, Wagener AH, Djukanovic R, Sterk PJ, Chung KF (2015). *Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort*. Eur Respir J 46, 1308-1321.

Singh D, Fox SM, Tal-Singer R, Bates S, Riley JH, Celli B (2014). *Altered gene expression in blood and sputum in COPD frequent exacerbators in the ECLIPSE cohort*. PLoS One 9, e107381.

Soriano JB, Murray CJL, Vos T, The GBD 2015 Chronic Respiratory Disease Collaborators (2017). *Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015*. Lancet Respir Med 5, 691-706.

Taube C, Reuter S (2019). *Transcriptome Analysis of Sputum Cells. The Modern Art of Assessing Inflammation*. Am J Respir Crit Care Med 199, 402-404.

Tiotiu A (2018). *Biomarkers in asthma: state of the art*. Asthma Res Pract 4, 10.

Weathington N, O'Brien ME, Radder J, Whisenant TC, Bleecker ER, Busse WW, Erzurum SC, Gaston B, Hastie AT, Jarjour NN, Meyers DA, Milosevic J, Moore WC, Tedrow JR, Trudeau JB, Wong HP, Wu W, Kaminski N, Wenzel SE, Modena BD (2019). *BAL Cell Gene Expression in Severe Asthma Reveals Mechanisms of Severe Disease and Influences of Medications*. Am J Respir Crit Care Med 200, 837-856.

Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, Snowden S, Burg D, D'Amico A, Horvath I, Chaiboonchoe A, Ahmed H, Ballereau S, Rossios C, Chung KF, Montuschi P, Fowler SJ, Adcock IM, Postle AD, Dahlen SE, Rowe A, Sterk PJ, Auffray C, Djukanovic R (2013). *Application of 'omics technologies to biomarker discovery in inflammatory lung diseases*. Eur Respir J 42, 802-825.

Yang IV, Lozupone CA, Schwartz DA (2017). *The environment, epigenome, and asthma*. Journal of Allergy and Clinical Immunology 140, 14-23.

4 Zusammenfassung

Deutsch

Bislang beschränkten sich molekulare Hochdurchsatzanalysen von Sputumproben auf die Charakterisierung des Sputum-Transkriptoms. Hierbei hat die variable zelluläre Zusammensetzung gemischtzelliger Sputumproben ein hohes Potenzial, die Datenanalyse zu verzerren. In dieser Arbeit analysierten wir das Transkriptom sowie ergänzend das Methylom von Sputumproben von Asthmatikern (n = 9), Patienten mit COPD (n = 10) sowie gesunden Kontrollen (n = 10). Zelltypspezifische Expressions- und Methylierungsmuster wurden *in silico* mittels einer auf unser experimentelles Setup angepassten Deconvolution per quadratischer Optimierung (quadratischer Programmierung), basierend auf der Differenzialzytologie der Sputumproben, gewonnen. In dieser explorativen Analyse demonstrieren wir, dass die Analyse des Sputum-Methyloms in bestehende Sputumanalyse-Workflows integriert werden kann und zu der Charakterisierung und dem Verständnis pulmonaler Inflammation beiträgt. Für den Fall, dass durch die Probenlagerung in Biobanken die RNA-Integrität analysierter Sputumproben kompromittiert ist, zeigen wir, dass eine geeignete *in silico*-Korrektur die Sensitivität und Spezifität der Datenanalyse verbessert. Des Weiteren veranschaulichen wir, dass eine Deconvolution zelltypspezifischer molekularer Signaturen nach Möglichkeit bei Hochdurchsatzanalysen gemischtzelliger Sputumproben zur Anwendung kommen sollte, um eine verzerrungsfreie Exploration und Interpretation molekularer Signaturen pulmonaler Inflammation zu ermöglichen.

Englisch

So far, most studies involving high-throughput analyses of sputum in asthma and COPD have focused on identifying transcriptomic signatures of disease. In this context, the highly variable cellular composition of sputum has potential to confound the molecular analyses. We performed whole-genome transcription and methylation analyses on sputum samples of 9 asthmatics, 10 healthy and 10 COPD subjects. Estimates of cell type-specific molecular profiles were derived (deconvolved) via quadratic programming based on sputum differential cell counts. In this exploratory study, we show that methylation profiling can be easily integrated into sputum analysis workflows and exhibits a strong potential to contribute to the profiling and understanding of pulmonary inflammation. Wherever RNA degradation occurs in biobanking settings, *in silico* correction is recommended to improve both the sensitivity and specificity of downstream analyses. We suggest that deconvolution approaches should be integrated in

sputum omics workflows whenever possible as these facilitate the unbiased discovery and interpretation of the molecular patterns of inflammation.

5 Erklärung des Eigenanteils an der Publikation

Die Projektidee sowie das Projektkonzept wurden gemeinsam von Prof. Dr. med. Klaus F. Rabe (KFR), PD Dr. med. Henrik Watz (HW), Prof. Dr. Torsten Goldmann (TG), Prof. Dr. Ole Ammerpohl (OA) und mir entwickelt. Das verwendete Biomaterial sowie die dazugehörigen klinischen Daten stammen aus den Kohortenstudien ALLIANCE und COSYCONET, im Rahmen derer KFR, HW, PD Dr. med. Anne Kirsten (AK) und Jun-Prof. Dr. med. Thomas Bahmer (TB) für die Datenerhebung am Standort Großhansdorf verantwortlich waren. Induzierte Sputumproben wurden von Frau Dr. Frauke Pedersen prozessiert und asserviert.
Die folgende Auswahl der verfügbaren Proben für dieses Forschungsvorhaben sowie deren Weiterverarbeitung (Nukleinsäurepräparation und Transkriptomanalyse) wurden eigenständig von mir durchgeführt. Die Methylomdaten wurden von OA anhand der von mir präparierten DNA generiert. Das Design des bioinformatischen Workflows lag in meiner Verantwortung, die Datenanalyse und die statistische Auswertung wurden von mir eigenständig durchgeführt. Melanie Weber (MW) und Dr. Daniela Börnigen (DB) standen hierbei beratend zur Seite. Das Manuskript wurde von mir verfasst und von den Ko-Autoren gegengelesen.

6 Danksagung

Herzlich danken möchte ich Herrn Prof. Dr. med. Carsten Bokemeyer für die Übernahme der Doktorvaterschaft und die damit verbundenen Gespräche und Ratschläge.
Mein besonderer Dank gilt Herrn Prof. Dr. med. Klaus F. Rabe für die erhaltene Betreuung und die Begleitung meiner Ausbildung während Studium, Doktorarbeit und darüber hinaus. Herrn PD Dr. Henrik Watz, Herrn Prof. Dr. Torsten Goldmann und Herrn Prof. Dr. Ole Ammerpohl möchte ich ganz herzlich für die erhaltene Betreuung, Unterstützung und die hervorragende Zusammenarbeit danken, ohne die dieses Forschungsprojekt nicht umsetzbar gewesen wäre. Darüber hinaus möchte ich Frau Dr. Frauke Pedersen, Herrn Jun.-Prof. Dr. med. Thomas Bahmer und allen weiteren Kollegen an den Standorten Großhansdorf und Borstel meinen Dank für die außergewöhnlich gute Zusammenarbeit aussprechen. Frau Melanie Weber und Frau Dr. Daniela Börnigen gebührt mein Dank für ihre Unterstützung und die vielen Diskussionen und Gespräche zu bioinformatischen und mathematischen Problemstellungen, die zur Umsetzung dieses Projektes beigetragen haben.

Der Studienstiftung des deutschen Volkes und dem Deutschen Zentrum für Lungenforschung möchte ich meinen Dank für die während meines Studiums erhaltene Förderung aussprechen. Insbesondere die erhaltene ideelle Förderung trug wesentlich zur Realisierung dieses Forschungsprojektes bei.

Meinen Eltern danke ich ganz besonders herzlich für die unermessliche Unterstützung in allen Bereichen des Lebens. Meiner Familie und Freunden danke ich darüber hinaus für die unentwegte persönliche Anteilnahme und Bereicherung meines Lebensweges.

7 Lebenslauf

Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

8 Eidesstattliche Erklärung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: _____