EBioMedicine 2 (2015) 929–936

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Original Article

Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD

Matthias Wielscher [a], Klemens Vierlinger [a], Ulrike Kegler [a], Rolf Ziesche [b], Andrea Gsur [c], Andreas Weinhäusel [a,*]

[a] AIT — Austrian Institute of Technology, Health & Environment Department, Molecular Diagnostics Unit, Muthgasse 11/2, 1190 Vienna, Austria
[b] Medical University of Vienna, Clinical Department for Pulmonology, Spitalgasse 23, 1090 Vienna, Austria
[c] Medical University of Vienna, Institute of Cancer Research, Borschkegasse 8A, 1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form 23 June 2015
Accepted 26 June 2015
Available online 2 July 2015

Keywords:
Liquid biopsy
methyl-sensitive restriction enzyme
multiplex PCR
Biomarker
PAX9
HOXD10

ABSTRACT

Disease-specific alterations of the cell-free DNA methylation status are frequently found in serum samples and are currently considered to be suitable biomarkers.
Candidate markers were identified by bisulfite conversion-based genome-wide methylation screening of lung tissue from lung cancer, fibrotic ILD, and COPD. cfDNA from 400 μl serum (n = 204) served to test the diagnostic performance of these markers. Following methylation-sensitive restriction enzyme digestion and enrichment of methylated DNA via targeted amplification (multiplexed MSRE enrichment), a total of 96 markers were addressed by highly parallel qPCR.
Lung cancer was efficiently separated from non-cancer and controls with a sensitivity of 87.8%, (95%CI: 0.67–0.97) and specificity 90.2%, (95%CI: 0.65–0.98). Cancer was distinguished from ILD with a specificity of 88%, (95%CI: 0.57–1), and COPD from cancer with a specificity of 88% (95%CI: 0.64–0.97). Separation of ILD from COPD and controls was possible with a sensitivity of 63.1% (95%CI: 0.4–0.78) and a specificity of 70% (95%CI: 0.54–0.81). The results were confirmed using an independent sample set (n = 46) by use of the four top markers discovered in the study (HOXD10, PAX9, PTPRN2, and STAG3) yielding an AUC of 0.85 (95%CI: 0.72–0.95).
This technique was capable of distinguishing interrelated complex pulmonary diseases suggesting that multiplexed MSRE enrichment might be useful for simple and reliable diagnosis of diverse multifactorial disease states.

1. Introduction

DNA methylation analysis of cell-free blood samples has a substantial potential to serve as a minimally invasive tool for early diagnosis and clinical monitoring of diseases with considerable heterogeneity Mouliere and Rosenfeld, 2015, in particular cancer and cancer-related diseases. Important lung pathologies, such as chronic obstructive lung disease (COPD) and fibrotic interstitial lung diseases (ILD) are associated with a significantly increased risk of cancer development Tomassetti et al., 2015 rendering diagnosis and monitoring of pulmonary lesions in these states is a highly challenging task. Usually, both diagnosis and monitoring of these diseases require repeated sampling of pulmonary tissue. This procedure harbors considerable risks for the patients Fibla et al., 2012. As a result, identification of pulmonary lesions frequently relies on advanced pulmonary imaging, such as computed tomography scans Lederlin et al., 2013. This approach, however, yields increasing numbers of pulmonary lesions of questionable clinical value Fang et al., 2014 adding to diagnostic insecurity.

A non-invasive approach allowing for a reliable differentiation between lung cancer and lung cancer-associated diseases like ILD and COPD is an unmet diagnostic issue. In this proof of principle study, we thus asked whether these three associated disease states could be discerned by DNA methylation analysis in cell free plasma samples. Analysis of DNA methylation was chosen for three reasons: (a) the specific features of cancerous tissue are equally well represented by changes of DNA methylation and DNA point mutation Li et al., 2009; Thierry et al., 2010; Diehl et al., 2005; (b) while point mutations are randomly distributed across a genetic locus, DNA methylation is clustered in specific regulatory regions Brock et al., 2008; Esteller, 2008; and (c) changes of methylation have been identified for all three clinical states Bruse et al., 2014; Sanders et al., 2012; Qiu et al., 2012.

Most detection methods analyzing DNA methylation rely on bisulfite conversion of DNA allowing for the detection of single CpGs while at the same time impeding further PCR amplification and thus, analysis of

Abbreviations: AUC, area under curve; cfDNA, cell-free DNA; COPD, chronic obstructive pulmonary disease; Ct-value, cycle threshold; HP, hypersensitivity pneumonitis; ILD, interstitial lung disease, IPF, idiopathic pulmonary fibrosis; MSP, methyl specific priming; MSRE, methyl sensitive restriction enzyme; NSIP, non-specific interstitial pneumonitis; ROC, receiver operating characteristics; UIP, usual interstitial pneumonia.

* Corresponding author at: AIT Austrian Institute of Technology GmbH, Muthgasse 11/2, 1190 Vienna, Austria.
E-mail address: Andreas.Weinhaeusel@ait.ac.at (A. Weinhäusel).

http://dx.doi.org/10.1016/j.ebiom.2015.06.025

additional regions of interest. This limits detection of methyl-specific priming (MSP) to 10 simultaneous performed reactions Sanders et al., 2012 and less than 5 for MethyLight™, respectively Fackler et al., 2014; Olkhov-Mitsel et al., 2014. In order to address as many methylation changes as possible in limited clinical samples, we chose high multiplexing methyl-sensitive restriction enzyme (MSRE) qPCR for stabilization of prediction and diagnostic accuracy Weinhaeusel et al., 2008; Melnikov et al., 2005 which allowed for the simultaneous performance of 96 qPCRs in cfDNA from 400 µl plasma.

Our multiplexed MSRE enrichment approach enabled us to effectively differentiate between lung cancer, pulmonary fibrosis, and healthy subjects allowing for the formulation of a four-marker model capable of distinguishing lung cancer from controls, fibrotic ILD, and COPD.

2. Methods

2.1. Study Design and Patients

A total of 250 serum or plasma samples derived from COPD, ILD, and lung cancer patients were analyzed (Table 1). Cancer Cases were recruited between 1996 and 2006 from the Pulmological Centers, Baumgartner Hoehe, Vienna and Grimmenstein-Hochegg, Lower Austria. Additional, population-based controls were selected from the ongoing colorectal cancer screening project "Burgenland PREvention trial of colorectal cancer Disease with ImmunologiCal Testing" (B-PREDICT). Interstitial lung disease (ILD) included of usual interstitial pneumonia (UIP) of early, limited from, and of progressive, advanced form, as well as of non-specific interstitial pneumonitis (NSIP) patients. Further hypersensitivity pneumonitis, which is due to its durance categorized into of an acute and chronic from, was also included in the group of ILD patients. All ILD and COPD sera were collected from 2008 to 2012 at the Medical University of Vienna. The study was approved by the concordant responsible ethics committee of Lower Austria (No: GS4-EK-1/121-2005), the Burgenland (No: 33/2010) and the Medical University of Vienna (No: 533/2004; No: 013/2009) and was carried out in compliance with the Helsinki Declaration. Patients gave a written informed consent.

2.2. Experimental Procedures

The 204 serum samples were randomized and split into three parts each one underwent separate multiplexed MSRE enrichment and Biomark qPCR read out. 46 samples from proof of principal (POP) set were processed separately. Technical replicates of samples as well as negative controls were introduced by the time of cell free DNA (cfDNA) isolation and randomly distributed to the three experimental blocks. Standard curves, which also served as amplification control, as well as additional negative controls, were introduced before multiplexed enrichment. cfDNA was isolated out of 400 µl patients' serum using Roche High pure template preparation kit (Roche) with a modified isolation protocol published previously Wielscher et al., 2011. In a volume of 12 µl cfDNA was digested with four methylsensitive restriction enzymes (MSRE) namely HpaII (ThermoFisher), Hin6I (ThermoFisher), AciI (NewEnglandBiolabs) and HpyCHIV4 (NewEnglandBiolabs). 10 µl of the MSRE digestion reaction was subjected to a multiplexed preamplification. Preamplification reaction was performed in a volume of 20 µl using 96 primer pairs per sample. 19 cycles of preamplification were performed. qPCR detection, consisting of 96 single qPCR reactions per sample, was performed on a BioMark™ HD Reader (Fluidigm). For detailed information on all experimental procedures and patients, see study flow figure (Fig. 1) and in supplemental material pp 1–3.

2.3. Statistical Analysis

All computations and statistical analyses were performed using R 3.0.2 Team RC, 2005 and Bioconductor 2.13 (http://www.bioconductor.org) Gentleman et al., 2004. Binominal classification comparing each disease group to healthy controls was performed on basis of linear models using lasso penalized logistic regression taken from the R package glmnet Anon, 2008. Differential multinomial classifications were performed using boosted trees method Friedman, 2001. Detailed information on Ct value preprocessing and data normalization, outlier identification and handling, prediction and resampling approach, the determination of optimal cutoffs, and simulation of automated assignment of clinical samples are given in supplemental material pp 3–7.

Table 1
Clinical Characteristics of patient serum/plasma samples.
250 samples were analyzed. 204 within the original set and 46 within the PoP (proof of principle) set. Type indicates whether a serum or plasma was available from the patients. Age indicates the mean age of each patient group. S indicates serum; P Plasma; female reflects the portion of female patients in percent. Information on the characteristic histopathologic pattern of ILD was not available for 12 fibrotic ILD patients and cancer staging information was not available for 7 sera of the original set and 3 sera of the PoP set. AdenoCa, lung adenocarcinoma; SqCC: squamous cell carcinoma, SCLC: small cell lung cancer, LCLC: large cell lung cancer.

	Original set (n = 204)				PoP set (n = 46)			
	Patients	Type	Female	Age	Patients	Type	Female	Age
Healthy	**N = 61 (29.9%)**				**N = 23 (50%)**			
Healthy	n = 27 (13.2%)	S&P	27.5%	57.5	n = 23 (50%)	P	17.4%	63.3
COPD 0	n = 34 (16.6%)	S	26.5%	48	–			
Lung cancer	**N = 33 (16.1%)**				**N = 23 (50%)**			
TNM I&II	n = 9 (4.4%)	P	44.4%	65.6	n = 8 (17.4%)	P	37.5%	60.4
TNM III&IV	n = 15 (7.3%)	P	40%	65	n = 12 (26%)	P	58.3%	63.7
Lung cancer	**N = 33 (16.1%)**				**N = 23 (50%)**			
AdenoCa	n = 11 (5.4%)	P	54.5%	63.9	n = 9 (19.5%)	P	77.7%	58.3
SqCC	n = 8 (3.9%)	P	12.5%	60.1	n = 4 (8.6%)	P	0%	65.3
SCLC	n = 7 (3.4%)	P	57.1%	70.1	n = 6 (13%)	P	66.6%	69.3
LCLC	n = 7 (3.4%)	P	0%	70.7	n = 4 (8.7%)	P	0%	58.2
COPD	**N = 42 (20.5%)**				–			
GOLD I–II	n = 31 (15.1%)	S	19.4%	52.6	–			
GOLD III–IV	n = 11 (5.4%)	S	18.2%	62.4	–			
ILD	**N = 68 (33.3%)**				–			
IPF, limited UIP	n = 15 (7.3%)	S	20%	65.3	–			
IPF, advanced UIP	n = 10 (4.9%)	S	50%	56.6	–			
NSIP	n = 11 (5.4%)	S	27.3%	68	–			
HP	n = 22 (10.7%)	S	27.3%	51.7	–			

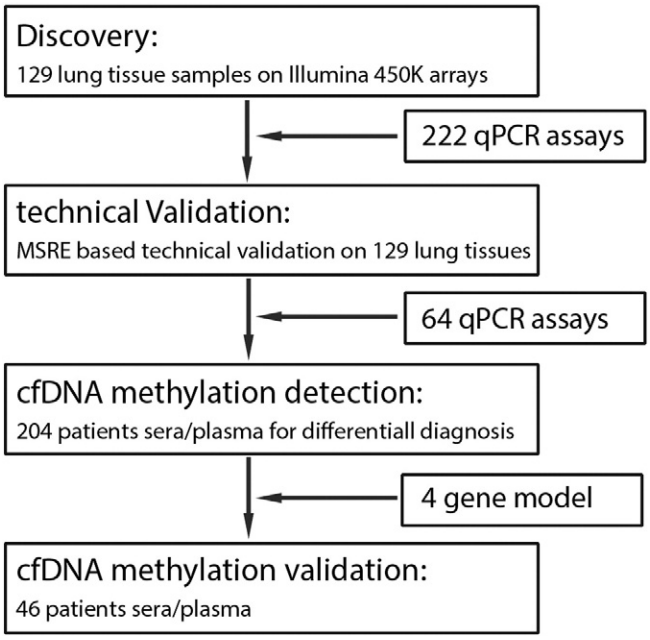

Fig. 1. Study flow diagram. The discovery was performed on bisulfite converted DNA derived from lung tissue of lung cancer patients, fibrotic ILD, COPD patients and healthy controls. Candidate markers revealed by Illumina 450K arrays were validated on the same sample material via MSRE digestion based qPCR. The best performing assays in this procedure were then used for minimal invasive cfDNA methylation detection. Thus a 4-gene model could be established, which was validated on an independent sample set of 46 patients.

3. Results

3.1. From Illumina 450 K Probe to MSRE-qPCR Assay

We first employed Illumina's 450K arrays for the detection of methylation signatures specific for lung cancer, fibrotic ILD and COPD in pulmonary tissue (GSE 63704). For the biomarker selection we pursued an exploratory data analysis approach. First a moderated t-test was performed for every disease group, followed by inspection of P-values, fold changes and boxplots as well as ROC-curve analysis indicating the ability of each marker to distinguish between inspected disease class and healthy or all remaining classes. In sum, 251 qPCR assays were designed to meet the genomic vicinity of the selected Illumina probes (Supplemental Fig. S2, Supplemental Table S1 & S2). Samples were validated in lung tissue achieving a validation rate of 65.7% (Supplemental Table S3). The methylation signature of the selected probes was rated as valid if qPCR delta Ct values and micro array beta-values showed a significant correlation ($P < 0.05$) (Supplemental Table S3).

The best selection criteria of probes for qPCR validation were (in order of efficacy): fold change of methylation beta-values from microarray data, number of MSRE cut sites, ROC curve and P values from microarray data (Supplemental Figs. S2 & S3). A strong enrichment of PCR assays located on a CpG island was observed (Supplemental Fig. S4).

As described in detail in the Method section and in Fig. 1, 92 markers out of 222 data producing qPCR assays after multiplex PCR (Supplemental Fig. S2) were selected for cfDNA methylation assessment, and 63 assays passed the quality control (Fig. 1). Detailed information on all qPCR assays is given in Supplemental Table S1–S3.

3.2. cfDNA Amounts in Patient Sera

All samples representing cancer, idiopathic pulmonary fibrosis (IPF; limited and advanced UIP), fibrotic non-specific interstitial pneumonia (NSIP), and COPD (GOLD grade 3) showed a significant difference regarding cfDNA concentrations compared to healthy controls (Fig. 2A and Supplemental Table S4). No difference was detected between serum and plasma samples (Supplemental Fig. S5). As previously shown, the highest amount of cfDNA was observed in lung cancer patients. cfDNA amounts in cancer increased by factor 2.8 to 35.6 ng/ml (95%CI: 29.05–43.23) for TNM stages I and II and 47.4 ng/ml (95%CI: 30.79–63.28), respectively, for TNM stages III and IV when compared to normal.

3.3. cfDNA Reveals Disease Specific Methylation Signature

To increase efficacy of cfDNA methylation analysis in patient plasma, we used multiplexed methyl sensitive restriction enzyme (MSRE) enrichment leaving the methylated DNA fraction intact (Fig. 2B), while fragmenting unmethylated cfDNA. With this approach, it was possible to analyze a total of 204 serum and plasma samples (lung cancer, $n = 33$; fibrotic ILD, $n = 68$; COPD, $n = 42$; healthy, $n = 61$; Table 1). Reproducibility was further addressed by analysis of 16 randomly chosen replicates demonstrating a mean Spearman's rank correlation coefficient of 0.69 (95%CI: 0.63–0.74) (Supplemental Fig. S8).

On the basis of differential methylation observed in patient samples, a penalized logistic regression was performed. qPCR Ct values of the 63 candidate loci and the cfDNA amounts were used to create a prediction model. Based on the randomly assigned training set (75% of samples), coefficients were build. Prediction was then performed on the remaining 25% of the samples (test set) (Fig. 2B.). This process was repeated 200 times and the resulting probabilities of sample class membership were averaged. To estimate the diagnostic value of the methylation markers tested, each disease group was compared to healthy controls (Fig. 2C). ROC analysis revealed an area under curve (AUC) of 0.91 (95%CI: 0.84–0.96) for lung cancer, 0.815 (95%CI: 0.73–0.88) for ILD, 0.73 (95%CI: 0.62–0.83) for COPD, and 0.828 (95%CI: 0.76–0.89) for all diseases versus healthy controls.

3.4. Differential Diagnosis of Lung Cancer, ILD and COPD

In order to achieve a stringent analytical approach capable of complying with the diagnostic needs, we performed simultaneous analysis of cfDNA methylation for all disease groups based on highly multiplexed MSRE enrichment. L1 regression, support vector machines or random forest was compared with regression tree-based gradient boosting for all 63 candidate loci and cfDNA amounts. The latter one turned out to be the most effective one for sample classification.

For differential diagnosis, the same resampling strategy was used (Fig. 2B). As depicted in Fig. 3A, three prediction rounds were performed to classify each sample of the entire sample pool. The first prediction round separated cancer from non-cancer cases in 90.6% (Fig. 3A and B). In the second round, cancer cases were subdivided into TNM classes I and II or TNM classes III and IV (Supplemental Fig. S9). The third prediction round distinguished between healthy controls, COPD, and ILD cases (Fig. 3A, 3C).

Introduction of cutoff values for each diagnostic group was determined by group membership probabilities obtained from gradient boosting classification (Fig. 3B, C). Significance of differentially methylated loci is given as relative variable importance (Fig. 3B, C) reflecting the number of decisions from gradient boosting classification made on basis of each candidate loci. Nonetheless, correlating markers, such as *CACNA1B*, *ZIC1*, *DLX1*, or *SIM1*, being potentially equally discriminative as the top markers *PAX9*, *HOXD10*, *PTPRN2*, and *STAG3* given in Fig. 3B and C, do not appear in the figure as a result of the model building process. Detailed information of all assays including P-values und fold changes is given in Supplemental Table S5 & S6. The four top markers found by multiplexed MSRE enrichment strategy were *HOXD10*, *PAX9*, *PTPRN2*, and *STAG 3*. *HOXD10* and *STAG3* were capable of discriminating lung cancer, ILD and COPD from healthy (Fig. 4B), while *PAX9* and *PTPRN2* demonstrated a strong specificity for lung cancer (Fig. 4A).

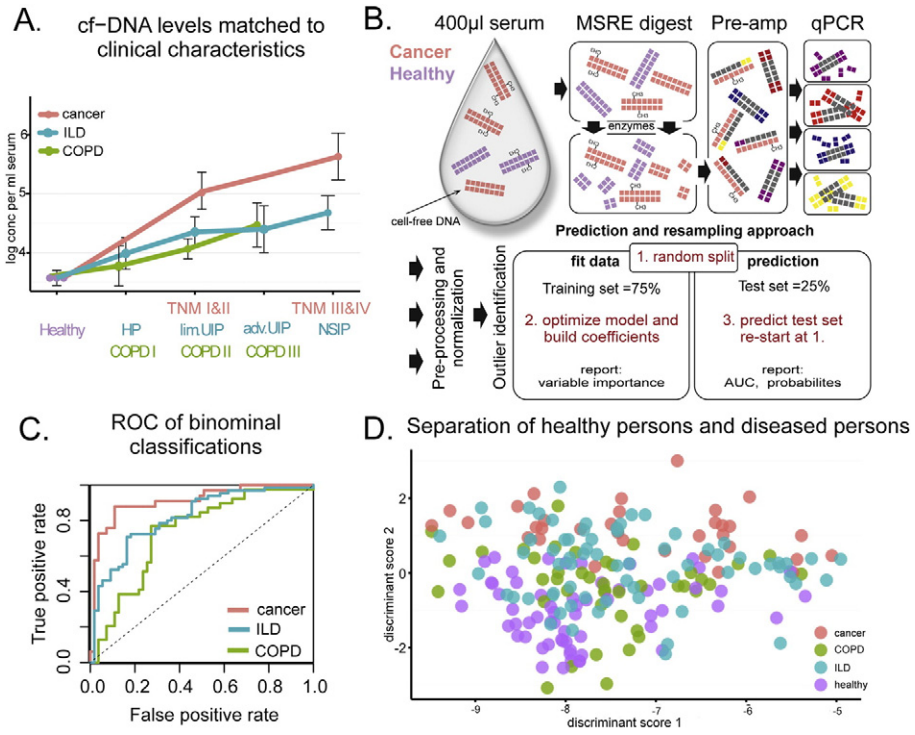

Fig. 2. Overview of cfDNA analysis. (A) Log2-transformed amount of cfDNA per ml serum/plasma is shown. Error bars indicate standard deviation of the mean. (B) The colored strings represent the cfDNAs. Those with CH_3 groups represent methylated DNA, the purple strings demonstrate cfDNA from healthy tissue (purple) and lung cancer (red), respectively. cfDNA processing workflow: Each reaction was based on one serum sample (400 µl). During enzymatic digestion, methylation protected the methylated cfDNA strings which then served as templates for targeted amplification (Pre-amp). One multiplexed preamplification was performed per sample using 96 primer pairs as indicated by different colors. Amplification results are shown in black color. Specific methylation markers were detected by individual qPCR reactions. The lower panel shows the prediction and resampling approach (see Method section). (C) Fisher discriminant analysis was performed using the top 30 markers subsequently, the data were projected to 2 most informative projection directions (discriminant scores). The plot shows separation of patient samples based on the transformed data, which may be interpreted similar to a Principal Component Analysis. (D) ROC-curve analysis shows quality of separation of each analyzed disease versus healthy controls.

3.5. Simulation of Prospective Sample Classification

The ultimate goal of our approach was an automated assignment of clinical samples to predefined diagnostic entities. Using all methylation markers detected in our analysis, we addressed their predictive power by an adjusted resampling strategy dividing all 204 plasma samples into 10 partitions. Each partition served as an unknown test sample during 10 rounds of automated clinical assignment (Supplemental Fig. S1). The synopsis of the classification is given in Fig. 5A demonstrating (a) the effectiveness of highly multiplexed MSRE enrichment for discrimination of the disease states tested (lung cancer, ILD, COPD and healthy), and (b) the overlaps between these clinical entities. Using cutoff-values derived from the corresponding training sets, it was possible to identify samples from cancer patients in 84.8% (28 of 33 cases). Patient samples derived from ILD patients were detected in 48.5% (33 of 68 cases), whereas COPD patients were discovered in 45.2% (19 of 42 cases). Healthy controls were identified in 50.8% (31 of 61 controls). Specificity was highest for diagnosis of lung cancer as depicted in Fig. 5A and B. A typical example for lung cancer (red) is shown in Fig. 5B demonstrating both the inter- and intra-individual discriminative power of multiplexed MSRE enrichment for lung cancer diagnosis. In comparison to cancer, specificity was lower for both ILD (blue) and COPD (green) samples, probably due to the considerable overlap between both diseases (Fig. 5B, patient 2 and 3). This is confirmed by the number of double positive predictions (n = 48). Discrimination of healthy samples from both cancer and ILD samples was very effective, whereas samples representing healthy and COPD demonstrated a large overlap, possibly due to the fact that in our group of COPD patients, early stage COPD (GOLD grade 1 and 2) was overrepresented (73.8%).

3.6. Independent Validation of Multiplexed MSRE Enrichment for Cancer Classification

Based on the predictive power of our approach for lung cancer, we then analyzed 46 new samples (healthy: n = 23; lung cancer: n = 23) comparing the full prediction model based on all methylation markers with a prediction model using only the 4 top markers by that addressing quality and stability of our automated prediction procedure (proof of principle; PoP-set, Table 1). ROC curve analysis showed that the 4-marker model (Fig. 6A, solid line) outperformed the full 64-marker model (Fig. 6A, dotted line) yielding an AUC of 0.85 (95%CI: 0.72–0.95) with a sensitivity of 0.97 (95%CI: 0.61–1) and a specificity of 0.73 (95%CI: 0.61–0.83). Using this approach, 22 of the 23 cancer samples were correctly identified, whereas two healthy cases were rated as COPD and 8 healthy controls as cancer (Supplemental Fig. S10).

4. Discussion

The goal of our study was to identify methylation biomarkers for cfDNA that permit an effective discrimination between overlapping and multifactorial lung diseases, such as cancer, ILD, and COPD (Table 1). Given the complex and intertwined pathology of these diseases, it is likely that epigenetic mechanisms, such as CpG dinucleotide methylation, contribute to their clinical phenotypes. Aberrant methylation has already been demonstrated for diseases like lung cancer,

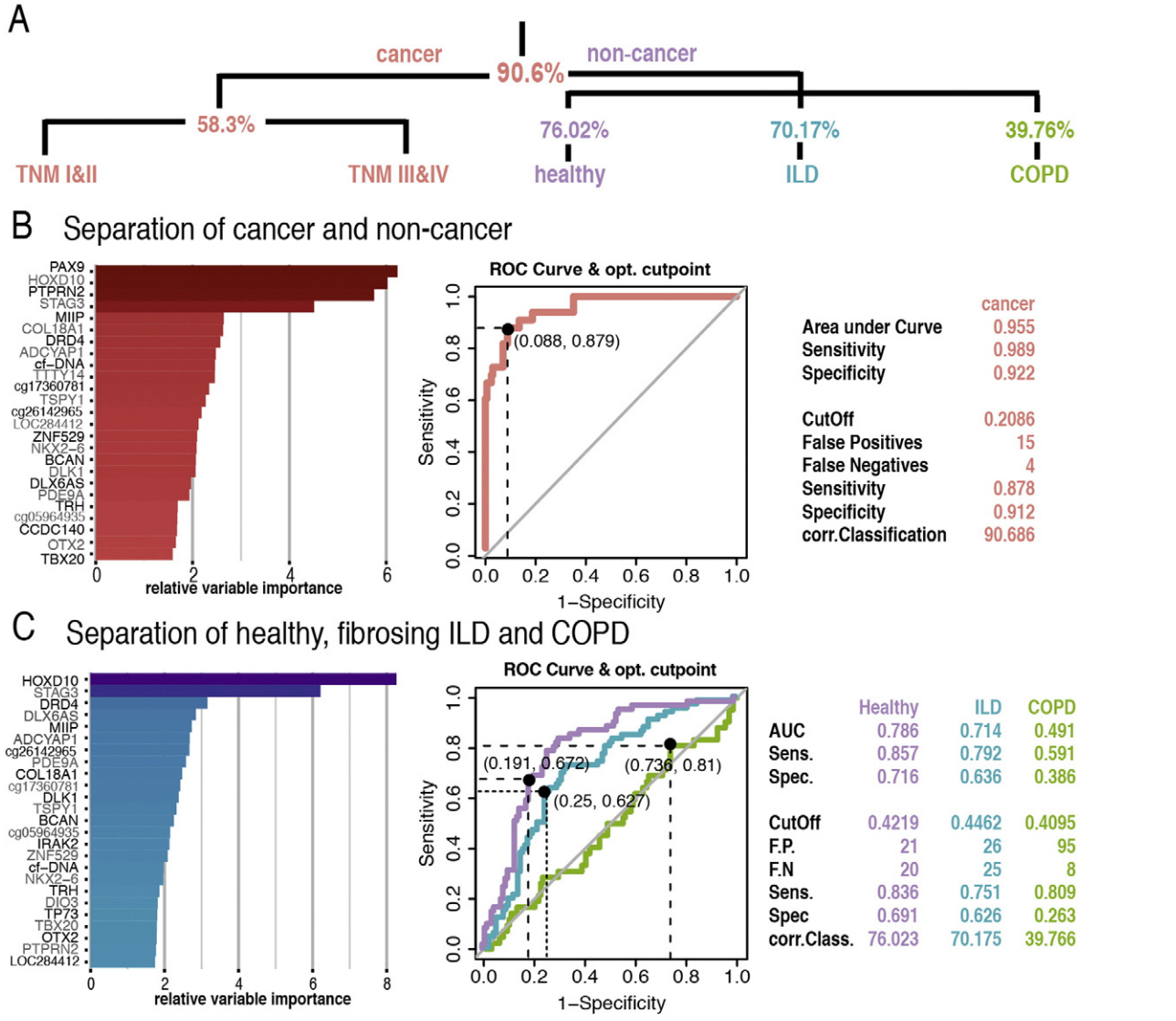

Fig. 3. Differential diagnosis approach. (A) The classification scheme starts with the separation of cancer samples, which are subsequently subdivided into TNMI&II and TNMIII&IV. The non-cancer samples are classified into healthy, fibrosis and COPD. This scheme implicates three prediction rounds, with results given in (B) and (C). The result for the separation of cancer TNMI&II and cancer TNMIII&IV is given in Supplemental Fig. S9. Percent values indicate the correct classification rate applying the determined cut off value given in sections B and C for each disease. (B and C) Bar plots indicate the relative variable importance for each model. The relative variable importance reflects the contribution of each variable to the prediction success. ROC curves indicate the quality of group separation including the chosen cut off value given as black dot. The rightmost panel summarizes the ROC curve analysis starting with the values (Area under curve, Sensitivity and Specificity) derived from conventional ROC curve analysis. The section below, starting with "CutOff", lists the values gained by the application of the specific cut off values. (B) Distinguishes between cancer and residual samples(C) classify residual samples into healthy, fibrotic ILD and COPD.

autoimmunity, immunodeficiency and neuronal regeneration defects Robertson, 2005. Yet, both detection and monitoring of such complex clinical states remain a considerable challenge Crowley et al., 2013. This is also true for the group of fibrotic ILD and for COPD, which currently accounts for 260 million patients worldwide and an annual death rate comparable to that of cancer Organization WH, 2011. Given the fact that both diseases show an increased risk of cancer development Tomassetti et al., 2015; Koshiol et al., 2009, it is conceivable that comparable epigenetic regulations will contribute to their pathologies in spite of existing phenotypical differences.

It is unlikely that a molecular discrimination between these diseases can be achieved using a single marker. Therefore, there is a need for a directed modeling of marker panels Brock et al., 2008; Bruse et al., 2014; Nikolaidis et al., 2012; Begum et al., 2011 as shown in our approach. Analysis of DNA methylation in plasma samples is well suited for such a goal Crowley et al., 2013; Fleischhacker et al., 2013. DNA methylation analysis largely relies on bisulfite conversion of cfDNA. This technique is characterized by an unavoidable degradation of DNA and a substantial loss of sequence complexity resulting in decreased sensitivity and specificity during PCR-based biomarker detection Egger et al., 2012. To overcome these technical limitations, we used a multiplexed MSRE enrichment strategy allowing for the reduction of unmethylated background DNA followed by simultaneous amplification of 96 targets (Supplemental Fig. S11).

Prior to any classification efforts based on patients DNA methylation, the contribution of the cfDNA amount was removed from the data by delta Ct normalization to be specific for disease derived methylation changes (Supplemental Figs. S6 & S7). The elevated cfDNA levels, however, were included as a separate predictor with in the study. Such an increase of cfDNA was frequently observed for diverse cancers Wielscher et al., 2011; Jung et al., 2004; Dawson et al., 2013 and late stage COPD patients Gormally et al., 2004. The presence of DNA exhibiting disease marks in serum may be due to an interplay of several mechanisms Schwarzenbach et al., 2011. Necrosis of

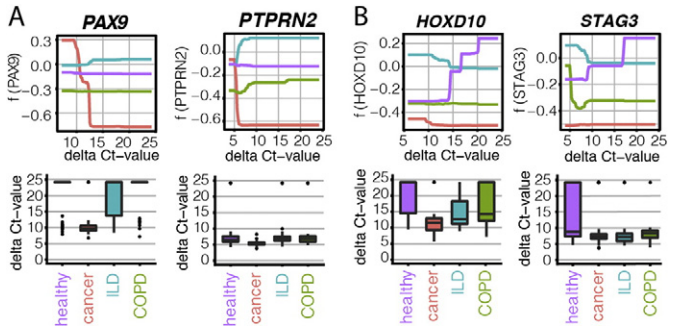

Fig. 4. Representative markers for differential diagnosis. Upper panel sections A and B demonstrate the effect of each variable on class probability. Class probability is given on the y-axis, while delta Ct-values are shown on the x-axis. Dependence of each predictor variable is averaged over the distribution of all modeled variables. The upper panel demonstrates the change of class probability (healthy, cancer, ILD, and COPD) as a function of Ct-value changes for the 4 top markers identified. The lower panels display boxplots of delta Ct-values for each marker. Due to the applied PCR methodology, lower delta Ct-values indicate increased marker methylation.

diseased cells, but also apoptotic cell death driven by inflammation or tissue repair, leads to increased DNA amounts Choi et al., 2005; Schulte-Hermann et al., 1995 in fibrosis patients and COPD patients (Fig. 2A). For cancer, additionally, the active release of DNA by cancer cells may be a reason for increased DNA amounts in patient sera Stroun and Anker, 2005.

We performed binominal classifications (case vs. control) based on logistic regression (using 63 methylation markers plus plasma-cfDNA concentration as additional predictor) to compare the diagnostic performance of our selected methylation markers with that of previously published markers. We saw the best discriminative power when comparing cancer to healthy samples yielding an AUC = 0.91 (95%CI: 0.84–0.96) with a sensitivity of 0.82 (95%CI: 0.61–0.97) and specificity of 0.89 (95%CI: 0.47–0.98). Comparable accuracy using cfDNA analysis from serum or plasma samples was previously observed in two studies addressing cancer vs. healthy (a) via measurement of global hypomethylation by massively parallel bisulfite sequencing for detection of lung, breast and nasopharyngeal cancer Chan et al., 2013, or (b) via *vimentin* hypermethylation using methyl BEAMing for diagnosis of colon cancer Li et al., 2009. The most widely used serum protein marker for lung cancer, the carcino-embryonic antigen (CEA) Okamura et al., 2013 and the best cfDNA methylation-based lung cancer assay Kneip et al., 2011, however, do not achieve a comparable discriminative power.

In view of the increased cancer risk in ILD and COPD, and given the lack of suitable blood tests for both diseases, we then applied our multiplexed minimal invasive testing for identification of ILD and COPD. For ILD we achieved classification results (Fig. 2C) comparable to the of cfDNA methylation analysis in various cancer types Egger et al., 2012. Reports on minimal invasive DNA methylation in COPD show similar classification rates as observed in this study, however they concentrate on sputum samples Bruse et al., 2014, a matrix which is not available at all times in COPD Han et al., 2010.

We then asked whether our approach was useable for differential diagnosis of all three pulmonary diseases. We identified four top markers (*HOXD10*, *PAX9*, *PTPRN2*, and *STAG3*) capable of effectively discriminating lung cancer, ILD and COPD. Of these, *PAX9* and *PTPRN2* had a particular specificity for lung cancer (Fig. 4A), while *HOXD10* and *STAG3* were well-qualified to discriminate all three diseases from healthy (Fig. 4B). To our best knowledge, changes of cfDNA methylation within these four loci have not been reported in plasma samples of patients with lung cancer, ILD or COPD. The results of our simulated prospective sample classification allow for two clinical strategies: (a) separation of cancer from non-cancer with sensitivity of 87.8% (95%CI: 0.67–0.97) and a specificity of 90.2%, (95%CI: 0.65–0.98), and (b) discrimination of cancer from ILD with a specificity 88% (95%CI: 0.57–1) and from COPD with a specificity of 88% (95%CI: 0.64–0.97) (Fig. 5). The

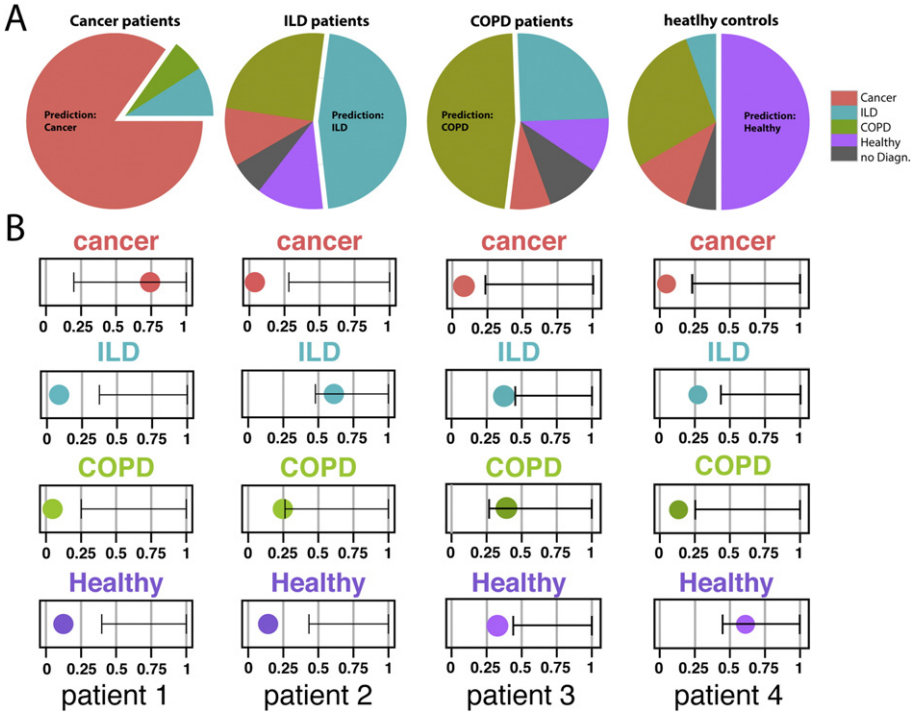

Fig. 5. Results of simulated prospective sample prediction. Simulation was achieved via an adjusted resampling strategy (Supplemental Fig. S1). (A) The upper panel shows pie diagrams of classification results derived from the simulation of prospective samples. The samples arranged according to their clinical diagnosis. Each pie represents one patient group. Each section of the pie represents the predicted sample memberships in percent. No Diagn. reflects 11 samples, which could not be classified to a specific disease or as healthy, because probabilities were below all cut off values. (B) The lower panel shows the classification of 4 representative patients. Patient 1 suffers from lung cancer; patient 2 was diagnosed with a limited UIP, patient 3 was diagnosed with COPD GOLDII and patient 4 is a healthy control. The x-axis represents the class dependence probability for each patient. The error bar indicates the range from the cut off value to a 100% probability.

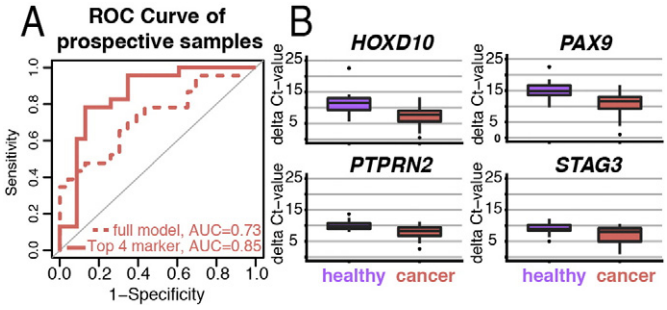

Fig. 6. Proof of Principle: Prospective sample prediction. ROC curve analysis of 46 patient samples. Prediction is based on coefficients derived from the model presented Fig. 3. The dashed line represents the separation of cancer and non-cancer patients, applying a weighted model of all 64 variables. The solid line represents the prediction based on the top 4 markers. Panel below gives boxplots of the delta Ct-values for each marker out of top four marker model. Due to the applied PCR methodology lower delta Ct-values indicate an increased methylation of the marker.

classification results further suggest that multiplexed cfDNA methylation profiling allows for the capture of possibly interconnected phenotypes in spite of different clinical diagnoses. This can be deduced on the distribution pattern across prediction probabilities from correctly classified samples (Supplemental Fig. S12). The flexibility of multiplexed MSRE enrichment is likely to allow for the introduction of further marker panels as well as the combination with SNP analysis improving the overall diagnostic capacity of both approaches.

Moreover, plasma or serum samples, often referred to as 'liquid biopsies' have several advantages over tissue sampling: (a) they are easily accessible, (b) are not subject to biopsy bias Crowley et al., 2013 and (c) can be repeatedly drawn from the same patient Dawson et al., 2013; Thierry et al., 2014; Murtaza et al., 2013. Thus, application of multiplexed MSRE enrichment to plasma or serum samples is ideally suited for clinical monitoring of disease progression, particularly as our method requires just 400 μl of plasma.

Contributors

MW contributed to the design of study, performed statistical analysis and wrote the first draft, prepared the figures, participated in execution of the experiments, performed the data analysis and interpreted the results. VK critically revised of manuscript and the design of experiments, contributed to the design of study, interpreted the results, UK executed the experiments and data analysis, RZ critically revised of manuscript and the design of experiments, contributed to the design of study, interpreted the results, obtained funding, collected serum samples and clinical data. AG critically revised of manuscript and collected serum samples and clinical data. AW, critically revised of manuscript and the design of experiments, contributed to the design of study, interpreted the results, obtained funding and gave conceptual advice on study design and data analysis.

Declaration of Interests

The authors of this manuscript have the following competing interests: Austrian Institute of Technology has filed a patent application based on the methylation markers discussed in this article, with MW, AW, KV, RZ and AG as inventors. No other authors have competing interests.

Role of the Funding Source

The funding body of this study had no role in study design, data collection, data collection, data analysis, data interpretation, or writing of the report. MW, KV, UK had access to the raw data. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Acknowledgments

This work was part of the project RESOLVE, funded by the European Commission under FP7-HEALTH-F4-2008, Contract no. 202047 (http://resolve.punkt-international.eu/).

We kindly thank all those who contributed to the sample collection, especially Karl Mach and Gernot Leeb, who collect most of our control samples and Norbert Vetter and Wolfgang Pohl both responsible of the collection of the lung cancer samples. We are grateful to John Field and Triantafillos Liloglou for primary lung tumor DNA samples.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2015.06.025.

References

Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455 (7216), 1061–1068.
Begum, S., Brait, M., Dasgupta, S., et al., 2011. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin. Cancer Res. 17 (13), 4494–4503.
Brock, M.V., Hooker, C.M., Ota-Machida, E., et al., 2008. DNA methylation markers and early recurrence in stage I lung cancer. N. Engl. J. Med. 358 (11), 1118–1128.
Bruse, S., Petersen, H., Weissfeld, J., et al., 2014. Increased methylation of lung cancer-associated genes in sputum DNA of former smokers with chronic mucous hypersecretion. Respir. Res. 15, 2.
Chan, K.C., Jiang, P., Chan, C.W., et al., 2013. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc. Natl. Acad. Sci. U. S. A. 110 (47), 18761–18768.
Choi, J.J., Reich III, C.F., Pisetsky, D.S., 2005. The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells. Immunology 115 (1), 55–62.
Crowley, E., Di Nicolantonio, F., Loupakis, F., Bardelli, A., 2013. Liquid biopsy: monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol. 10 (8), 472–484.
Dawson, S.J., Tsui, D.W., Murtaza, M., et al., 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N. Engl. J. Med. 368 (13), 1199–1209.
Diehl, F., Li, M., Dressman, D., et al., 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc. Natl. Acad. Sci. U. S. A. 102 (45), 16368–16373.
Egger, G., Wielscher, M., Pulverer, W., Kriegner, A., Weinhausel, A., 2012. DNA methylation testing and marker validation using PCR: diagnostic applications. Expert. Rev. Mol. Diagn. 12 (1), 75–92.
Esteller, M., 2008. Epigenetics in cancer. N. Engl. J. Med. 358 (11), 1148–1159.
Fackler, M.J., Lopez Bujanda, Z., Umbricht, C., et al., 2014. Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. Cancer Res. 74 (8), 2160–2170.
Fang, W., Xiang, Y., Zhong, C., Chen, Q., 2014. The IASLC/ATS/ERS classification of lung adenocarcinoma—a surgical point of view. J. Thorac. Dis. 6 (Suppl. 5), S552–S560.
Fibla, J.J., Brunelli, A., Cassivi, S.D., Deschamps, C., 2012. Aggregate risk score for predicting mortality after surgical biopsy for interstitial lung disease. Interact. Cardiovasc. Thorac. Surg. 15 (2), 276–279.
Fleischhacker, M., Dietrich, D., Liebenberg, V., Field, J.K., Schmidt, B., 2013. The role of DNA methylation as biomarkers in the clinical management of lung cancer. Expert Rev. Respir. Med. 7 (4), 363–383.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Gentleman, R.C., Carey, V.J., Bates, D.M., et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5 (10), R80.
Gormally, E., Hainaut, P., Caboux, E., et al., 2004. Amount of DNA in plasma and cancer risk: a prospective study. Int. J. Cancer (J. Int. Cancer) 111 (5), 746–749.
Han, M.K., Agusti, A., Calverley, P.M., et al., 2010. Chronic obstructive pulmonary disease phenotypes: the future of COPD. Am. J. Respir. Crit. Care Med. 182 (5), 598–604.
Jung, K., Stephan, C., Lewandowski, M., et al., 2004. Increased cell-free DNA in plasma of patients with metastatic spread in prostate cancer. Cancer Lett. 205 (2), 173–180.
Kneip, C., Schmidt, B., Seegebarth, A., et al., 2011. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J. Thorac. Oncol. 6 (10), 1632–1638.
Koshiol, J., Rotunno, M., Consonni, D., et al., 2009. Chronic obstructive pulmonary disease and altered risk of lung cancer in a population-based case–control study. PLoS One 4 (10), e7380.
Lederlin, M., Revel, M.P., Khalil, A., Ferretti, G., Milleron, B., Laurent, F., 2013. Management strategy of pulmonary nodule in 2013. Diagn. Interv. Imaging 94 (11), 1081–1094.
Li, M., Chen, W.D., Papadopoulos, N., et al., 2009. Sensitive digital quantification of DNA methylation in clinical samples. Nat. Biotechnol. 27 (9), 858–863.

Melnikov, A.A., Gartenhaus, R.B., Levenson, A.S., Motchoulskaia, N.A., Levenson Chernokhvostov, V.V., 2005. MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res. 33 (10), e93.
Mouliere, F., Rosenfeld, N., 2015. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. Proc. Natl. Acad. Sci. U. S. A. 112 (11), 3178–3179.
Murtaza, M., Dawson, S.J., Tsui, D.W., et al., 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497 (7447), 108–112.
Nikolaidis, G., Raji, O.Y., Markopoulou, S., et al., 2012. DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. Cancer Res. 72 (22), 5692–5701.
Okamura, K., Takayama, K., Izumi, M., Harada, T., Furuyama, K., Nakanishi, Y., 2013. Diagnostic value of CEA and CYFRA 21–1 tumor markers in primary lung cancer. Lung Cancer 80 (1), 45–49.
Olkhov-Mitsel, E., Zdravic, D., Kron, K., van der Kwast, T., Fleshner, N., Bapat, B., 2014. Novel multiplex MethyLight protocol for detection of DNA methylation in patient tissues and bodily fluids. Sci. Rep. 4, 4432.
Organization WH, 2011. World Health Statistics 2011.
Qiu, W., Baccarelli, A., Carey, V.J., et al., 2012. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am. J. Respir. Crit. Care Med. 185 (4), 373–381.
Robertson, K.D., 2005. DNA methylation and human disease. Nat. Rev. Genet. 6 (8), 597–610.
Sanders, Y.Y., Ambalavanan, N., Halloran, B., et al., 2012. Altered DNA methylation profile in idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. 186 (6), 525–535.
Schulte-Hermann, R., Bursch, W., Grasl-Kraupp, B., Torok, L., Ellinger, A., Mullauer, L., 1995. Role of active cell death (apoptosis) in multi-stage carcinogenesis. Toxicol. Lett. 82–83, 143–148.
Schwarzenbach, H., Hoon, D.S., Pantel, K., 2011. Cell-free nucleic acids as biomarkers in cancer patients. Nat. Rev. Cancer 11 (6), 426–437.
Stroun, M., Anker, P., 2005. Circulating DNA in higher organisms cancer detection brings back to life an ignored phenomenon. Cell. Mol. Biol. 51 (8), 767–774.
Team RC, 2005. R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing.
Thierry, A.R., Mouliere, F., Gongora, C., et al., 2010. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res. 38 (18), 6159–6175.
Thierry, A.R., Mouliere, F., El Messaoudi, S., et al., 2014. Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. Nat. Med. 20 (4), 430–435.
Tomassetti, S., Gurioli, C., Ryu, J., et al., 2015. The impact of lung cancer on survival of Idiopathic Pulmonary Fibrosis. Chest 147 (1), 157–164.
Weinhaeusel, A., Thiele, S., Hofner, M., Hiort, O., Noehammer, C., 2008. PCR-based analysis of differentially methylated regions of GNAS enables convenient diagnostic testing of pseudohypoparathyroidism type Ib. Clin. Chem. 54 (9), 1537–1545.
Wielscher, M., Pulverer, W., Peham, J., et al., 2011. Methyl-binding domain protein-based DNA isolation from human blood serum combines DNA analyses and serum-autoantibody testing. BMC Clin. Pathol. 11, 11.