Alemany-Cosme *et al. Clinical Epigenetics* (2026) 18:8
https://doi.org/10.1186/s13148-025-02026-4

Clinical Epigenetics

RESEARCH Open Access

Genome-wide methylation analysis of circulating cell-free DNA identifies an episignature in lung cancer: validation of EMP2 as a potential biomarker

Ester Alemany-Cosme[1], Dario Cruz-Chamorro[2], David Hervas[3], Diana Garcia[4], Raquel Martinez-Tomas[2], Andres Briones-Gomez[2], Jose Galbis-Caravajal[5], Alfonso Calvo[6,7], Oscar Juan[1], Agustin Lahoz[1], Enrique Cases[2], Cora Palanca-Ballester[1] and Juan Sandoval[4*]

Abstract

Background Lung cancer (LC) remains the leading cause of cancer-related mortality worldwide. Its poor prognosis is largely attributed to late-stage diagnosis, highlighting the need for sensitive, minimally invasive biomarkers for early detection. In this study, we aimed to identify circulating cell-free DNA (cfDNA) methylation-based biomarkers for LC through genome-wide profiling of plasma-derived cfDNA from patients with LC (stages I–IV) using the Infinium DNA MethylationEPIC array (> 850,000 CpGs). A relaxed elastic net penalized logistic regression model was applied to identify differentially methylated CpGs between LC patients and non-neoplastic controls, and selected candidates were validated in an independent cohort.

Results A 21-CpG episignature was identified, predominantly hypomethylated in LC, which distinguished cases ($n=25$) from controls ($n=8$) with a cross-validated area under the curve (AUC) of 0.70. Two CpGs with the largest effect sizes (associated with *EMP2* and *IKZF2*) were selected for validation. Digital droplet PCR (ddPCR) analysis in an independent cohort of 47 individuals (23 LC patients and 24 controls) confirmed significant hypomethylation at the *EMP2* locus (median % methylation: 73.60% in tumors vs. 86.83% in controls; $p=0.0034$), achieving an AUC of 0.75.

Conclusions Our findings support the utility of methylome-wide cfDNA profiling for LC biomarker discovery. *EMP2* promoter methylation represents a promising candidate for the minimally invasive detection of LC across heterogeneous clinical presentations.

Keywords Lung cancer, Cell-free DNA, DNA methylation, Biomarker, Digital droplet PCR (ddPCR), EMP2

Ester Alemany-Cosme and Dario Cruz-Chamorro should be considered as first authors

Cora Palanca-Ballester and Juan Sandoval should be considered as last authors

*Correspondence:
Juan Sandoval
epigenomica@iislafe.es
[1]Biomarkers and precision medicine Unit, Health research institute La Fe, Valencia, Spain
[2]Pneumology Service, University Hospital La Fe, Valencia, Spain
[3]Department of Applied Statistics and Operational Research and Quality, Universitat Politècnica de València, Valencia, Spain
[4]Epigenomics Core Facility and Translational Epigenomics in disease and longevity (EPITELONG). Health research institute La Fe, Valencia, Spain
[5]Thoracic Surgery Service, University Hospital La Ribera, Alzira, Spain
[6]CIBERONC, ISCIII, 28029 Madrid, IDISNA, Spain
[7]Center for Applied Medical Research (CIMA), Department of Pathology, Anatomy and Physiology, School of Medicine, University of Navarra, Pamplona, Spain

Background

Lung cancer (LC) is one of the most common cancers and the leading cause of cancer-related death worldwide (representing around 18.7% of cancer deaths). The five-year survival rate is generally below 20% in most countries [1, 2]. The poor prognosis of LC is primarily attributed to its late diagnosis, since patients typically exhibit nonspecific symptoms that become clinically apparent only at advanced stages of the disease [3].

Although several clinical trials have demonstrated that LC screening programs utilizing low-dose computed tomography (LDCT) can effectively increase survival rates, routine LC screening is not widely adopted in clinical practice [4, 5]. This is mainly due to the false-positive results and the risk of overdiagnosis in high-risk individuals, radiation concerns, and the associated medical costs to insurance coverage or national healthcare systems. Consequently, there is an urgent need to develop novel, sensitive, and specific biomarkers to improve early diagnosis of LC. In this context, blood-based liquid biopsies have emerged as a minimally invasive approach for cancer detection. In particular, plasma-derived circulating cell-free (cfDNA) has garnered considerable interest due to its ability to capture tumor heterogeneity while offering high stability and ease of isolation [6]. Recent advances suggest that integrating molecular biomarkers such as DNA methylation profiles with LDCT could further improve screening accuracy and cost-effectiveness [7].

Epigenetic alterations, including DNA methylation, are considered to play a crucial role in the acquisition of hallmark traits during tumor development and malignant progression [8]. In LC, the hypermethylation of tumor suppressor gene promoters causes their silencing, whereas global genome hypomethylation is associated with genomic instability and oncogene activation [9]. Over two decades ago, the discovery of the hypermethylated promoter region of *CDKN2A* within precursor lesions of LC led to the proposal of aberrant DNA methylation as a biomarker for early LC detection [10]. Since then, considerable efforts have focused on detecting abnormal DNA methylation patterns in non-invasive (e.g. sputum), semi-invasive samples (such as bronchoalveolar lavages or bronchial aspirates); and predominantly, in peripheral blood, for a minimally invasive approach and early detection potential. For instance, our group previously validated a four-gene epigenetic signature in bronchial fluids [11], and demonstrated the potential of *BCAT1* methylation in cfDNA as a LC biomarker [12]. Notably, *SHOX2* DNA methylation in blood plasma has been demonstrated to reliably distinguish patients with LC [13]. Yet, most studies to date have focused on targeted validation of previously identified, tissue-based methylation markers [14–16], rather than leveraging the full potential of cfDNA for genome-wide discovery-based approaches.

There is increasing recognition of the importance of conducting genome-wide analysis directly in cfDNA, to identify specific, blood-measurable DNA methylation biomarkers [17]. In this study, we sought to identify a novel cfDNA methylation-based episignature for LC detection using a genome-wide discovery approach, and to validate top-ranked CpGs in an independent cohort using digital droplet PCR (ddPCR), a clinically scalable technique.

Methods

Patients: peripheral blood samples

This observational prospective study included two independent cohorts: a discovery cohort and a validation cohort. Participants were recruited between 2019 and 2022 from the University Hospital La Fe (Valencia, Spain) and University Hospital La Rivera (Alcira, Spain), and included patients diagnosed with LC (stages I–IV) as well as non-cancer control individuals. Control participants in the discovery cohort had non-malignant respiratory diseases, while those in the validation cohort included individuals with both pulmonary and non-pulmonary conditions. The absence of malignancy in control individuals was confirmed by clinical evaluation and a five-year follow-up.

Exclusion criteria included being a minor, significant deterioration in general condition (Karnofsky score < 30%), uncontrolled coagulopathy, or unwillingness to participate in the study. All participants provided written informed consent before enrollment, in accordance with protocols approved by the Clinical Research Ethics Committee of Hospital La Fe (ref: 2019/0114). Available resources determined the sample size for the discovery cohort. For the validation cohort, sample size was estimated to achieve a target 95% confidence interval half-width for the AUC. Assuming an anticipated AUC of approximately 0.75 and a balanced case-control design, a sample size of $n = 24$ per group provides a precision of about ± 0.12 for the AUC estimate [18]. An AUC in the 0.70–0.80 range is generally regarded as acceptable discrimination, including in thoracic oncology [19].

CfDNA isolation and bisulfite conversion

All peripheral blood samples were collected in PAXgene® Blood cfDNA Tubes (Qiagen) and were centrally processed at Hospital La Fe. A total of 30 mL of blood was obtained from each participant in the discovery cohort, whereas 10 mL were collected from participants in the validation cohort. The reduced sample volume in the validation phase reflects the high sensitivity of bisulfite-based ddPCR methylation assays, which yield reliable results with minimal cfDNA input, making larger draws

unnecessary [12, 20]. Blood plasma was isolated within 1 to 24 h of collection using double centrifugation at 4 °C (1,900 x g for 15 min; 1.900 x g for 10 min) and stored at − 80 °C until further processing.

cfDNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) following the manufacturer's instructions. Quantification of cfDNA was performed using the Qubit 1× dsDNA High-Sensitivity Assay Kit and a Qubit 4.0 Fluorometer (Thermo Fisher Scientific). Bisulfite-conversion of cfDNA was carried out using the EZ DNA Methylation Lightning Kit (Zymo Research), following the manufacturer's protocol. For the discovery cohort, up to 200 ng of cfDNA per sample (minimum 100 ng) were processed, while 50 ng of cfDNA were used for bisulfite conversion in samples from the validation cohort. Bisulfite-treated DNA was eluted in 10 µL of elution buffer.

Epigenomic studies

Bisulfite-converted cfDNA was analyzed using the Infinium DNA MethylationEPIC (850 K) BeadChip arrays (Illumina) targeting 853,307 CpG sites (CpGs) across gene promoters, enhancers, and intergenic regions. Hybridization was performed with 10 µL of converted DNA as described by our former lab [21]. Arrays were scanned on the iScan SQ System (Illumina) and the raw data (IDAT files) were normalized using functional normalization, implemented in the R-package minfi (version 1.54.0). The methylation level of each cytosine was expressed as a beta value calculated as the fluorescence intensity ratio of the methylated to the unmethylated versions of the probes ($\beta = M/[M + U + 100]$) (according to a combination of the Cy3 and Cy5 fluorescence intensities). For probe filtering, unreliable beta-values with high detection p-values (>0.01) (85,805 CpGs) were removed. Subsequently, probes targeting CpGs associated with single-nucleotide polymorphisms (SNPs; 2,932 CpGs), as well as those located on sex chromosomes (19,681 CpGs), were excluded. After the filtering, the remaining 761,457 CpGs were considered valid for downstream analysis.

Cell lines

A549, NCI-H358, NCI-H1975, and HCC827 cell lines obtained from the American Type Culture Collection (ATCC) were used for primer optimization. These cell lines were selected for their methylation status at the *EMP2* and *IKFG2* gene loci, as determined from an internal laboratory database of DNA methylation profiles generated using the Illumina Infinium HumanMethylation450 BeadChip array (Illumina). All cell lines were routinely tested and confirmed to be free of mycoplasma contamination.

Digital droplet PCR methylation analysis

For the validation of specific differentially methylated CpGs (DMCs), digital droplet PCR (ddPCR; Bio-Rad Laboratories) analysis was conducted on bisulfite-converted cfDNA from the validation cohort. Methylation-specific primers and hydrolysis probes were designed to amplify bisulfite-converted DNA (bs-DNA), with FAM-labeled probes targeting methylated CpGs and HEX-labeled probes for unmethylated sequences. Primers and probe sequences were designed according to Bio-Rad recommendations using Primer3Plus [22] and are listed in Supplementary Table S1 (Additional file 3). The QX200 Droplet Generator (Bio-Rad Laboratories) was used before DNA amplification with the following conditions: 95 °C for 10 min; 40 cycles of 94 °C for 30 s and 55 °C for 1 min; 98 °C for 10 min. The optimal annealing temperature (55 °C) was chosen after performing a temperature gradient assay for the designed primer sets in bs-DNA isolated from the cell lines. DNA amplification was carried out with the C1000 Touch Thermal Cycler (Bio-Rad Laboratories). After PCR, QuantaSoft™ software Analysis Pro v2.0 (Bio-Rad Laboratories) was used for the analysis [with the RED (Rare Event Detection) option]. Methylation percentages were calculated with QuantaSoft™ .

Functional enrichment analysis

Functional enrichment analysis of the genes associated with DMCs was conducted using GeneCodis4 (https://genecodis.genyo.es) [23], based on Gene Ontology (GO) Biological Process annotations. Enrichment was assessed using a hypergeometric test and corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR).

Statistical analysis

For initial data exploration, dimensionality reduction was performed via principal component analysis (PCA) to assess global methylation patterns and potential batch effects. A relaxed elastic net penalized logistic regression model (mixing parameter $\alpha = 0.4$) was adjusted to find which CpGs were able to discriminate between control and LC (tumor) samples. The relaxed elastic net penalized logistic regression model included sex, age, and smoking history as potential confounders. The penalization parameter (λ) was selected by applying the one-standard-error rule within 500 repetitions of 10-fold-cross-validation, and the final value was defined as the mode of the resulting distribution of optimal λ values across all iterations. Internal validation of the model was performed using 5-fold cross-validation. An external validation with an independent cohort was used to validate the differences in methylation levels at the selected CpGs between tumor and control samples using ddPCR within the validation cohort. Discriminatory capacity

of the selected DMC was assessed by estimating the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) values. All statistical analyses were performed using R (version 4.5.0), with R packages glmnet (version 4.1.8) and pROC (version 1.18.5).

Results

Clinical characteristics of patients in the study: discovery and validation cohorts

A total of 33 human blood samples were included in the discovery cohort, comprising 25 patients diagnosed with LC and 8 cancer-free controls with underlying respiratory diseases. The validation cohort consisted of 47 individuals, including 23 LC patients and 24 non-cancer control donors with a broader range of clinical backgrounds, encompassing both pulmonary and non-pulmonary conditions. In both cohorts, control individuals showed no histological evidence of malignancy at the time of recruitment and remained cancer-free during a five-year follow-up period. The LC group included patients with both non-small cell lung cancer (NSCLC; adenocarcinoma and squamous cell carcinoma) and small cell lung cancer (SCLC). Detailed clinical characteristics of the study participants are presented in Table 1.

Global DNA methylation profile in the discovery cohort

After pre-filtering the MethylationEPIC array results, a total of 761,457 CpGs were valid for downstream analysis. Direct comparison of the average methylation levels of these CpGs between LC patients ($n = 25$) and non-cancer controls ($n = 8$) revealed relatively similar global patterns. The initial exploratory analysis using Principal Component Analysis (PCA) showed no outliers in the

Table 1 Clinical characteristics of LC patients and cancer-free controls in the Discovery and Validation cohorts

	Discovery cohort (n = 33)		Validation cohort (n = 47)	
	LC patients (n = 25)	Cancer-free control patients (n = 8)	LC patients (n = 23)	Cancer-free control patients (n = 24)
Average Age (years)	69.0	64.9	68.0	53.0
Sex n(%)				
Male	17 (68.0)	3 (37.5)	19 (82.6)	12
Female	8 (32.0)	5 (62.5)	4 (17.4)	10
Undetermined				2
Smoking History n(%)				
Smoker/former	19 (76.0)	7 (87.5)	17 (73.9)	13 (54.2)
Non-smoker	3 (12.0)	1 (12.5)	5 (21.7)	8 (33.3)
Undetermined	3 (12.0)	0 (0)	1 (4.4)	3 (12.5)
Histology n(%)		NA		NA
Lung adenocarcinoma	19 (76)		17 (73.9)	
Lung squamous cell carcinoma	2 (8)		6 (26.1)	
Small cell lung cancer	4 (16)		0 (0)	
Stage n(%)		NA		NA
I	4 (16.0)		3 (13.0)	
II	5 (20.0)		3 (13.0)	
III	6 (24.0)		7 (30.5)	
IV	8 (32.0)		7 (30.5)	
Undetermined	2 (8.0)		3 (13.0)	
Non-cancer diseases n(%)	NA		NA	
EPOC		2 (25)		3 (12.5)
Pneumonia		2 (25)		3 (12.5)
Hemoptisis		1 (12.5)		4 (16.7)
pulmonary fibrosis		2 (25)		0 (0)
bronchiectasis		1 (12.5		0 (0)
Benign pulmonary nodule				2 (8.3)
Acute infection				3 (12.5)
Hypercholesterolemia				4 (16.7)
Dermatologic diseases				2 (8.3)
Immunomediated diseases				3 (12.5)

Control individuals in the Discovery cohort had non-tumoral respiratory diseases, while those in the Validation cohort included cancer-free donors with both pulmonary and non-pulmonary diseases. NA, not applicable

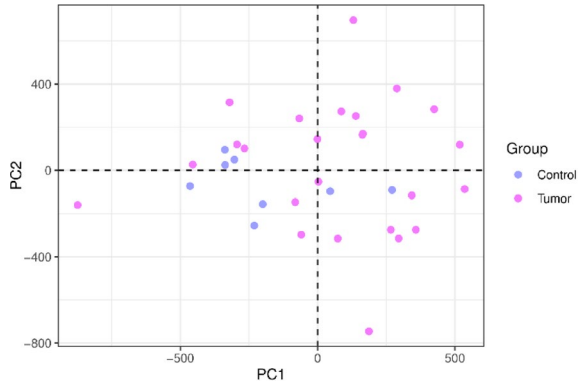

Fig. 1 Exploratory analysis of global DNA methylation. PCA illustrating global DNA methylation variation across samples

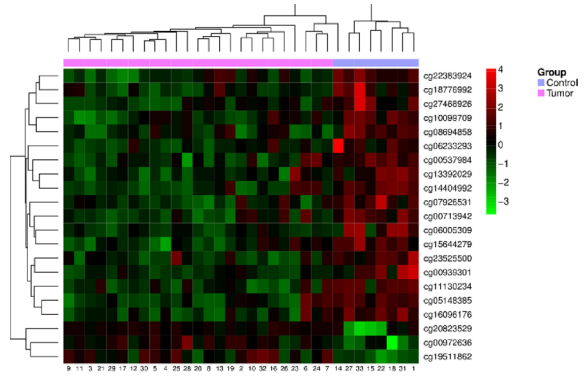

Fig. 2 Discovery of a cfDNA-methylation signature associated with lung cancer patients. Heatmap of the selected 21 CpGs resulting from the relaxed elastic net penalized logistic regression model shows how their methylation values differentiate between tumor and control groups. Rows (CpGs) and columns (patients) are ordered according to hierarchical clustering results. The Z-score color scale ranges from red (hypermethylation) to green (hypomethylation)

data, and demonstrated a mixed distribution of control and tumor samples (Fig. 1). This indicates that in our cfDNA methylation context, the discriminatory performance of unsupervised and exploratory PCA should be enhanced by incorporating supervised feature selection.

Identification of a DNA-methylation signature associated with LC patients

A relaxed elastic net penalized logistic regression model was applied to identify a set of CpGs that could distinguish the tumor and control groups from the discovery subset using the 761,457 valid CpGs. The final model retained 21 CpGs, which successfully discriminated between the groups with a cross-validated AUC of 0.7 (Fig. 2). Supplementary Figure S1 (Additional file 1) presents box plots illustrating DNA methylation beta-values across control and tumor samples at each CpG site. Genomic locations of the identified CpGs were analyzed using the MethylationEPIC array annotation to determine their distribution across functional genomic regions and CpG substructures (Table 2). Among the 21 CpGs, 5 were located at intergenic regions, while the remaining 16 were associated with one or more genes and distributed across both promoter and gene body regions. Notably, tumor samples were hypomethylated at 18 out of the 21 CpGs when compared to control samples.

The DMCs in LC samples were annotated to 17 genes, including *EMP2*, *IKZF2*, *TP73*, *AGAP1*, *DKFZp434J0226*, *ZNF638*, *ADC*, *TRAK2*, *ARTN*, *CHN1*, *STRADB*, *MAP3K10*, *GLTSCR1*, *EPCAM*, *PCGF3*, *CDC42*, and *BRI3BP*. To gain insight into potential biological processes associated with these genes, functional enrichment analysis was performed using GeneCodis4. Of the 17 genes, 14 were recognized by the GeneCodis database and included in the functional enrichment analysis. Only categories containing at least two genes and an FDR-adjusted p-value ≤ 0.025 were considered significantly enriched. The analysis highlighted biological processes related to kinase activity and the JNK signaling cascade (Table 3).

To identify the CpGs with the most pronounced methylation differences among the 21 DMCs, the standardized mean difference in methylation (SMDH) was calculated for each locus. CpGs cg14404992 and cg05148385 showed the highest SMDH values—2.28 and 2.25, respectively—indicating strong discriminatory potential between LC and control samples. Mean beta-values decreased from 0.79 to 0.63 for cg14404992 and from 0.59 to 0.43 for cg05148385 in control versus tumor groups. Complementary, both loci also exhibited absolute median methylation differences (|ΔMedian|) greater than 0.15 (-0.20 and – 0.19, respectively). Based on these metrics, cg14404992 and cg05148385 were prioritized for validation in an independent cohort. These CpGs are located in the genomic regions associated with the *Epithelial Membrane Protein 2* (*EMP2*) and the *IKAROS Family Zinc Finger 2* (*IKZF2*) genes, respectively. Interestingly, the EMP2 associated CpG is located in the promoter and CGI shore. A CpG shore is a region of DNA located within 2 kilobases upstream or downstream of the CpG island's boundaries, and when located in promoter regions, is strongly associated with transcriptional regulation and gene function in cancer [24].

Validation of EMP2 and IKZF2 methylation levels in LC patients

To assess the diagnostic performance of the two top-ranked DMCs, their methylation levels were quantified in plasma-derived cfDNA from an independent validation cohort (23 LC patients and 24 controls) using ddPCR. Clinical characteristics of this cohort are presented in Table 1. ddPCR was selected for its high sensitivity in quantifying low-abundance cfDNA methylation targets and its capacity for robust absolute quantification,

Table 2 Genomic context and differential methylation analysis of the 21 CpGs distinguishing LC patients

CpG ID	Gene name	†Gene region	††CpG context	CONTROL β-values mean	CONTROL β-values median	CONTROL SD	TUMOR β-values mean	TUMOR β-values median	TUMOR SD	ΔMean	Δmedian	SMDH
cg14404992	*EMP2*	5'UTR	N_Shore	0,7939	0,8051	0,0609	0,6262	0,6044	0,0813	−0,1677	−0,2006	2,28
cg05148385	*IKZF2;IKZF2*	Body;Body	Open sea	0,5930	0,5994	0,0447	0,4258	0,4080	0,0925	−0,1672	−0,1914	2,25
cg22383924	*TP73*	5'UTR	Open sea	0,6602	0,6497	0,0494	0,5382	0,5346	0,0644	−0,1220	−0,1151	2,07
cg16096176	*AGAP1;AGAP1*	Body;Body	Open sea	0,7222	0,6933	0,0637	0,5781	0,5859	0,0756	−0,1442	−0,1074	2,01
cg00537984	*DKFZp434J0226*	TSS1500	Open sea	0,5148	0,5075	0,0531	0,3528	0,3460	0,0978	−0,1620	−0,1615	2,01
cg15644279	*Intergenic*	*Intergenic*	Open sea	0,3705	0,3772	0,0501	0,2619	0,2723	0,0574	−0,1086	−0,1049	1,97
cg00713942	*Intergenic*	*Intergenic*	Open sea	0,5054	0,4953	0,0628	0,3776	0,3651	0,0652	−0,1278	−0,1303	1,95
cg08694858	*ZNF638;ZNF638*	Body;Body	Open sea	0,5537	0,5628	0,0412	0,4693	0,4783	0,0446	−0,0844	−0,0846	1,92
cg10099709	*Intergenic*	*Intergenic*	Open sea	0,6486	0,6613	0,0459	0,5533	0,5539	0,0521	−0,0953	−0,1074	1,89
cg11130234	*Intergenic*	*Intergenic*	Open sea	0,4556	0,4788	0,0825	0,3024	0,3009	0,0766	−0,1532	−0,1779	1,88
cg07926531	*ADC*	TSS200	Island	0,0385	0,0374	0,0052	0,0303	0,0300	0,0039	−0,0082	−0,0074	1,74
cg23525500	*Intergenic*	*Intergenic*	Open sea	0,5816	0,5416	0,1533	0,3250	0,3150	0,1418	−0,2566	−0,2266	1,70
cg13392029	*ARTN;ARTN;ARTN;ARTN;ARTN*	TSS1500;5'UTR;5'UTR;5'UTR;TSS1500	N_Shore	0,5570	0,5627	0,0894	0,4141	0,4174	0,0763	−0,1429	−0,1453	1,68
cg00939301	*CHN1;CHN1*	TSS200;TSS200	Island	0,0439	0,0427	0,0133	0,0285	0,0283	0,0041	−0,0154	−0,0143	1,53
cg06005309	*STRADB;TRAK2;TRAK2*	TSS200;5'UTR;1stExon	Island	0,0210	0,0232	0,0072	0,0129	0,0130	0,0023	−0,0081	−0,0103	1,49
cg27468926	*MAP3K10;MAP3K10*	1stExon;5'UTR	Island	0,0158	0,0148	0,0038	0,0113	0,0116	0,0019	−0,0045	−0,0032	1,46
cg18776992	*GLTSCR1*	5'UTR	S_Shore	0,0583	0,0543	0,0170	0,0381	0,0367	0,0088	−0,0202	−0,0176	1,45
cg06233293	*EPCAM*	TSS200	N_Shore	0,0363	0,0335	0,0117	0,0242	0,0241	0,0051	−0,0122	−0,0095	1,32
cg20823529	*PCGF3*	Body	Island	0,6010	0,6054	0,2587	0,8398	0,8501	0,0433	0,2388	0,2448	−1,26
cg00972636	*CDC42;CDC42;CDC42*	Body;Body;Body	Open sea	0,1416	0,1522	0,0632	0,2186	0,2182	0,0310	0,0770	0,0661	−1,51
cg19511862	*BRI3BP*	1stExon	Island	0,0094	0,0095	0,0007	0,0113	0,0114	0,0011	0,0020	0,0019	−2,02

†Gene Region column categorizes CpGs as "Promoter" (including TSS1500, TSS200, 5′UTR, and 1st exon), "Body" (gene body and 3′UTR), or "Intergenic" if not associated with a known gene. ††CpG context indicates whether each site is located within a CpG island, shore or open sea. Methylation metrics include mean and median beta-values and standard deviation in both tumor and control groups, along with group differences in median (ΔMedian), mean (ΔMean), and standardized mean difference in methylation (SMDH)

Table 3 Functional enrichment analysis of genes associated with DMCs in lung cancer

Annotation ID	Description	Genes found	Relative enrichment	Genes	Adjusted P-values
GO:0099159	Regulation of modification of postsynaptic structure	2	237	*CDC42, AGAP1*	0,0051
GO:0032147	Activation of protein kinase activity	2	104	*STRADB, EMP2*	0,0139
GO:0007254	JNK cascade	2	43	*STRADB, MAP3K10*	0,0243
GO:0046330	Positive regulation of JNK cascade	2	26	*MAP3K10, CDC42*	0,0243

Only categories with an FDR-adjusted p-value ≤ 0.025 and at least two genes are shown

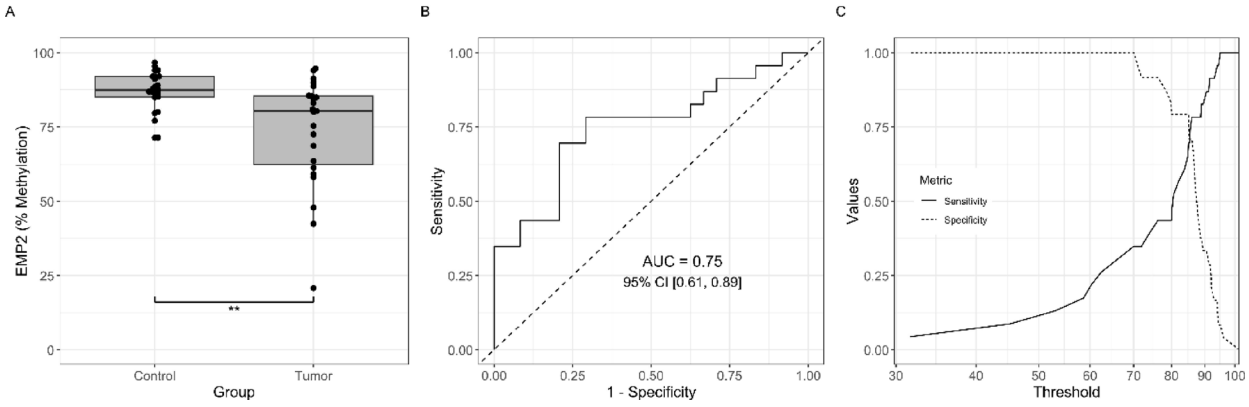

Fig. 3 *EMP2* methylation status in blood-derived cfDNA as a potential biomarker for lung cancer detection. A. DNA methylation levels of *EMP2* are significantly lower in lung cancer patients (Tumor) and non-neoplastic controls (Control) from the validation cohort. P-value (0.0034) was calculated using the two-sided Mann–Whitney U test. B. ROC curve and AUC with 95% confidence interval for *EMP2*. C. Sensitivity and specificity profile for different possible *EMP2* methylation cut-off values

enhancing precision and facilitating clinical translatability. Primer and probe sequences are listed in Table S1.

Among the candidate biomarkers, *EMP2* hypomethylation in LC patients was successfully validated (Fig. 3). Methylation levels of *EMP2* were significantly lower in tumor samples (median %methylation = 73.60) compared to controls (median %methylation = 86.83). Receiver Operating Characteristic (ROC) curve analysis yielded an Area under the Curve (AUC) of 0.75 (95% CI: 0.61–0.89), indicating moderate diagnostic accuracy in distinguishing LC from control samples. In contrast, *IKZF2* did not exhibit statistically significant methylation differences between tumor (median %methylation = 92.69) and control (median %methylation = 92.95) samples in the validation cohort (Supplementary Figure S2 in Additional file 2), suggesting limited reproducibility at this locus in our validation setting.

Discussion

Over the past decade, there has been a growing interest in blood-derived cfDNA methylation signatures for the minimally invasive detection of LC. Initial studies primarily employed targeted approaches to interrogate CpGs previously identified as differentially methylated in lung tumor tissues [25, 26]. More recently, research has evolved toward blood-based, genome-wide discovery approaches. In line with this shift, an unbiased genome-wide screen of plasma-derived cfDNA was applied using the Illumina MethylationEPIC array, identifying a 21-CpG episignature that discriminates LC patients from individuals with non-malignant respiratory diseases.

Recent larger-scale efforts have already demonstrated the utility of cfDNA methylation for pan-cancer detection, including LC. The Circulating Cell-free Genome Atlas (CCGA) study [17] analyzed plasma cfDNA from over 6,000 participants using bisulfite next-generation sequencing (NGS) targeting more than 100,000 methylation regions. Their resulting classifier achieved 99.3% specificity and 67.3% sensitivity for 12 prespecified cancers (including lung) in stages I–III. Similarly, Gao et al. applied ELSA-Seq—a bisulfite-based, methylation-sensitive enrichment method with linear amplification—to profile 161,984 CpGs in cfDNA, generating two multivariate classifiers (MCDBT-1 and MCDBT-2) for the detection of six cancer types (colorectal, esophageal, hepatic, pulmonary, ovarian, and pancreatic) [27]. Other studies have explored alternative methylation enrichment strategies. Kwon et al. demonstrated that methylation-sensitive restriction enzyme digestion followed by sequencing (MRE-SEQ) was a reliable method for detecting global hypomethylation patterns in liquid biopsy and diagnosing both colorectal and lung cancers [28]. On the other hand, Xu et al. applied Methylated-DNA-Immunoprecipitation Sequencing (MeDIP-Seq) to identify promoter-region DMRs in a small LC cohort [29].

Although these studies collectively underscore the diagnostic potential of cfDNA methylation, they rely on multivariate classifiers or broad genomic panels,

and therefore entail complex workflows and substantial computational support—barriers to clinical translation. Here, we presented an alternative strategy: focusing on the identification of single-CpG markers using the MethylationEPIC array, which provides quantitative, high-throughput profiling and is unbiased toward methylated or unmethylated regions. Therefore, we have performed an epigenome-wide analysis of cfDNA using array-based profiling and validated single-CpG biomarkers for LC detection in an independent cohort using ddPCR.

The 21-CpG episignature identified to discriminate LC patients predominantly comprised hypomethylated loci (18 of 21 CpGs). While genome-wide DNA methylation alterations have been extensively characterized in tumor tissues, the methylation landscape of tumoral cfDNA remains less well defined. In cancer tissues, DNA methylation changes typically include focal hypermethylation—particularly in promoter regions of tumor suppressor genes—and widespread hypomethylation across intergenic and repetitive regions [30, 31]. Further studies are needed to better characterize the epigenomic features of circulating tumoral cfDNA.

Based on gene ontology analysis, several of the DMCs mapped to genes involved in oncogenic signaling pathways. *STRADB* and *MAP3K10*, both hypomethylated in our LC discovery set, appeared enriched in the "JNK pathway regulation". This cascade is implicated in cell proliferation, apoptosis resistance, and metastatic potential in LC [32]. Additionally, in a similar way, *EMP2* appeared enriched, along with *STRADB*, in the "activation of protein kinase activity". Interestingly, *EMP2*, our top-ranked CpG hypomethylated in the LC patient subset, has a demonstrated oncogenic role in several cancers, including the promotion of angiogenesis in endometrial cancer [33] and stemness in breast cancer cells [34]. However, it has also been reported to suppress cell growth in NSCLC through MAPK pathway inhibition [35]. Finally, other cancer-related genes from the episignature were not significantly enriched in any Gene Ontology terms.

Prioritizing single-CpG candidates through robust statistical modeling was intended to facilitate simpler and clinically implementable cfDNA methylation assays. In this way, *EMP2* hypomethylation (cg14404992) emerged as a particularly promising biomarker. Importantly, it was validated in an independent cohort from two different hospitals, within a heterogeneous population: lung cancer patients from stages I to IV and non-cancer controls with a range of pulmonary and non-pulmonary conditions. Its clinically relevant performance (AUC = 0.75) across varied clinical contexts highlights its potential translational value. Unlike most studies that validate markers using the same platform as in discovery, a different technique for validation was employed—ddPCR—a method that is more readily applicable and affordable in clinical diagnostic settings [36, 37].

cg14404992 is located in the *EMP2* promoter region and within a CpG shore, genomic contexts often linked to transcriptional regulation [38]. Although some members of the Epithelial Membrane protein family, such as EMP3 [39], have been identified to be differentially methylated in LC tissues, to our knowledge, there are no reports directly linking EMP2 CpG methylation to cancer.

While *EMP2* methylation demonstrated LC discriminatory capacity in the validation cohort, our top two-ranked DMC candidate, associated with *IKZF2*, showed no significant difference in methylation levels between groups. These results suggest limited cross-cohort reproducibility in *IKZF2* methylation levels, potentially due to the methodological differences or the greater heterogeneity of the validation cohort. This highlights the importance of validation in diverse, independent populations using standardized methods.

Despite identifying the promising cfDNA methylation-based *EMP2* candidate biomarker, several limitations merit consideration. The sample size of both cohorts was modest, which reduces statistical power. While the inclusion of two independent institutions improves the results' generalizability compared to single-center studies, larger multi-center cohorts would be necessary to validate our findings robustly. The application of Illumina MethylationEPIC arrays to cfDNA presents inherent limitations, including potential biases related to probe design and hybridization efficiency. Long-read sequencing technologies offer the potential to overcome these limitations and should be considered in future studies. Moreover, the diagnostic accuracy of *EMP2* alone was moderate, suggesting that it may be insufficient for high-confidence diagnosis and could benefit from integration into a multi-marker panel. Ultimately, both discovery and validation cohorts included patients spanning stages I through IV; therefore, stage-specific diagnostic efficacy, especially for early-stage disease, where non-invasive detection is most crucial and cfDNA can be less abundant, should be a focus of subsequent research.

Conclusions

In conclusion, the present study supports the feasibility of genome-wide cfDNA methylation profiling for the identification of clinically relevant biomarkers for LC detection. The hypomethylation of *EMP2* in plasma-derived cfDNA, validated by ddPCR, emerges as a potential non-invasive diagnostic candidate. To our knowledge, this is the first study to report *EMP2* hypomethylation in cfDNA from lung cancer patients. Larger and stage-specific cohorts are needed to confirm and extend these findings in clinically relevant settings.

Abbreviations

LC	Lung cancer
ddPCR	Digital droplet PCR
LDCT	Low-dose computed tomography
cfDNA	Cell-free DNA
DMCs	Differentially methylated CpG sites
CpGs	CpG sites

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13148-025-02026-4.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We thank Consuelo Sanz Frances, Encarna Saiz Lozoya, Maria Sanchez Samblacat and Catalina Mármol Albert for their help in sample recruitment.

Author contributions

JS, EC, and CPB conceived and designed the study. CPB, EAC, DC, MB, DG carried out the experiments. AB, EC, RMT, OJ, JGC contributed to sample acquisition, preparation and data collection. JS, DH, CPB, AC, AL contributed to the interpretation of the results. JS supervised the project and JS, EC, EAC, DC and CPB wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the ISCIII ("PI19/00572"), co-funded by ERDF/ESF, "A way to make Europe"/"Investing in your future", Proyectos de Generación de Conocimiento 2022 Proyecto PID2022-137965OB-I00 from MCIN/ AEI /https://doi.org/10.13039/501100011033/ and FEDER "A way to make Europe" and the Conselleria de innovación, universidades, ciencia y sociedad digital (CIAICO/2022/251). E. Alemany-Cosme was supported by a grant of ISCIII (FI20/00136), D. Cruz-Chamorro and C Palanca-Ballester were supported by Grants from Generalitat Valenciana (CIAPOT/2022/21 and FDEGENT/2019/014), respectively.

Data availability

The datasets generated and/or analysed during the current study are available in the ArrayExpress repository, with an assigned accession reference E-MTAB-15459.

Declarations

Ethics approval and consent to participate

All participants provided written informed consent prior to enrollment, in accordance with protocols approved by the Clinical Research Ethics Committee of Hospital La Fe (ref: 2019/0114).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 August 2025 / Accepted: 11 November 2025
Published online: 02 December 2025

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63.
2. Soerjomataram I, Cabasag C, Bardot A, Fidler-Benaoudia MM, Miranda-Filho A, Ferlay J, et al. Cancer survival in Africa, central and South America, and Asia (SURVCAN-3): a population-based benchmarking study in 32 countries. Lancet Oncol. 2023;24(1):22–32.
3. Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS ONE. 2011;6(3):e17479.
4. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409.
5. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer mortality with volume CT screening in a randomized trial. N Engl J Med. 2020;382(6):503–13.
6. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. Nat Rev Clin Oncol. 2017;14(9):531–48.
7. Wang Z, Xie K, Zhu G, Ma C, Cheng C, Li Y et al. Early detection and stratification of lung cancer aided by a cost-effective assay targeting circulating tumor DNA (ctDNA) methylation. Respir Res [Internet]. 2023 Jun 17 [cited 2025 Jul 10];24(1). Available from: https://respiratory-research.biomedcentral.com/articles/https://doi.org/10.1186/s12931-023-02449-8
8. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12(1):31–46.
9. Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci. 2008;105(1):252–7.
10. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of $p16^{INK4a}$ is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci. 1998;95(20):11891–6.
11. Diaz-Lagares A, Mendez-Gonzalez J, Hervas D, Saigi M, Pajares MJ, Garcia D, et al. A novel epigenetic signature for early diagnosis in lung cancer. Clin Cancer Res. 2016;22(13):3361–71.
12. Palanca-Ballester C, Hervas D, Villalba M, Valdes-Sanchez T, Garcia D, Alcoriza-Balaguer MI, et al. Translation of a tissue epigenetic signature to Circulating free DNA suggests BCAT1 as a potential noninvasive diagnostic biomarker for lung cancer. Clin Epigenetics. 2022;14(1):116.
13. Kneip C, Schmidt B, Seegebarth A, Weickmann S, Fleischhacker M, Liebenberg V, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. J Thorac Oncol. 2011;6(10):1632–8.
14. Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhäusel A. Diagnostic performance of plasma DNA methylation profiles in lung Cancer, pulmonary fibrosis and COPD. EBioMedicine. 2015;2(8):929–36.
15. Vrba L, Futscher BW. A suite of DNA methylation markers that can detect most common human cancers. Epigenetics. 2018;13(1):61–72.
16. Li P, Liu S, Du L, Mohseni G, Zhang Y, Wang C. Liquid biopsies based on DNA methylation as biomarkers for the detection and prognosis of lung cancer. Clin Epigenetics. 2022;14(1):118.
17. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol. 2020;31(6):745–59.
18. Hajian-Tilaki K. Sample size Estimation in diagnostic test studies of biomedical informatics. J Biomed Inf. 2014;48:193–204.
19. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5(9):1315–6.
20. Jensen SØ, Øgaard N, Ørntoft MBW, Rasmussen MH, Bramsen JB, Kristensen H, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. Clin Epigenetics. 2019;11(1):158.
21. Gallardo-Gómez M, Moran S, De La Páez M, Martínez-Zorzano VS, Rodríguez-Berrocal FJ, Rodríguez-Girondo M, et al. A new approach to epigenome-wide discovery of non-invasive methylation biomarkers for colorectal cancer screening in Circulating cell-free DNA using pooled samples. Clin Epigenetics. 2018;10(1):53.
22. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007;35(Web Server):W71–4.
23. Garcia-Moreno A, López-Domínguez R, Villatoro-García JA, Ramirez-Mena A, Aparicio-Puerta E, Hackenberg M, et al. Funct Enrich Anal Regul Elem Biomedicines. 2022;10(3):590.
24. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG Island Shores. Nat Genet. 2009;41(2):178–86.

25. Weiss G, Schlegel A, Kottwitz D, König T, Tetzner R. Validation of the SHOX2/PTGER4 DNA methylation marker panel for Plasma-Based discrimination between patients with malignant and nonmalignant lung disease. J Thorac Oncol. 2017;12(1):77–84.
26. Kruusmaa K, Bitenc M, Chersicola M, Weinhaeusel A, Pulverer W. P46.06 Cell-Free DNA (cfDNA) methylation assay allows for early detection and identification of lung cancer. J Thorac Oncol. 2021;16(3):S491.
27. Gao Q, Lin YP, Li BS, Wang GQ, Dong LQ, Shen BY, et al. Unintrusive multicancer detection by Circulating cell-free DNA methylation sequencing (THUNDER): development and independent validation studies. Ann Oncol Off J Eur Soc Med Oncol. 2023;34(5):486–95.
28. Kwon HJ, Shin SH, Kim HH, Min NY, Lim Y, Joo T, woon, et al. Advances in methylation analysis of liquid biopsy in early cancer detection of colorectal and lung cancer. Sci Rep. 2023;13(1):13502.
29. Xu W, Lu J, Zhao Q, Wu J, Sun J, Han B, et al. Genome-Wide plasma Cell-Free DNA methylation profiling identifies potential biomarkers for lung cancer. Dis Markers. 2019;2019:4108474.
30. Esteller M. Epigenetic gene Silencing in cancer: the DNA hypermethylome. Hum Mol Genet. 2007;16(R1):R50–9.
31. Bergman Y, Cedar H. DNA methylation dynamics in health and disease. Nat Struct Mol Biol. 2013;20(3):274–81.
32. Khatlani TS, Wislez M, Sun M, Srinivas H, Iwanaga K, Ma L, et al. c-Jun N-terminal kinase is activated in non-small-cell lung cancer and promotes neoplastic transformation in human bronchial epithelial cells. Oncogene. 2007;26(18):2658–66.
33. Gordon LK, Kiyohara M, Fu M, Braun J, Dhawan P, Chan A, et al. EMP2 regulates angiogenesis in endometrial cancer cells through induction of VEGF. Oncogene. 2013;32(46):5369–76.
34. Dillard C, Kiyohara M, Mah V, McDermott SP, Bazzoun D, Tsui J, et al. EMP2 is a novel regulator of stemness in breast cancer cells. Mol Cancer Ther. 2020;19(8):1682–95.
35. Ma Y, Schröder DC, Nenkov M, Rizwan MN, Abubrig M, Sonnemann J, et al. Epithelial membrane protein 2 suppresses Non-Small cell lung cancer cell growth by Inhibition of MAPK pathway. Int J Mol Sci. 2021;22(6):2944.
36. Wang D, O'Rourke D, Sanchez-Garcia JF, Cai T, Scheuenpflug J, Feng Z. Development of a liquid biopsy based purely quantitative digital droplet PCR assay for detection of MLH1 promoter methylation in colorectal cancer patients. BMC Cancer. 2021;21(1):797.
37. Araki K, Kurosawa A, Kumon H. Development of a quantitative methylation-specific droplet digital PCR assay for detecting Dickkopf-related protein 3. BMC Res Notes. 2022;15:169.
38. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG Island Shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet. 2009;41(12):1350–3.
39. Xue Q, Zhou Y, Wan C, Lv L, Chen B, Cao X, et al. Epithelial membrane protein 3 is frequently shown as promoter methylation and functions as a tumor suppressor gene in non-small cell lung cancer. Exp Mol Pathol. 2013;95(3):313–8.

Publisher's note