Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens

Carlos Pedraz-Valdunciel[1,2] (iD), Stavros Giannoukakos[3], Nicolas Potie[4], Ana Giménez-Capitán[5], Chung-Ying Huang[6], Michael Hackenberg[3], Alberto Fernandez-Hilario[4], Jill Bracht[2,5], Martyna Filipska[1,2] (iD), Erika Aldeguer[5], Sonia Rodríguez[5], Trever G. Bivona[7], Sarah Warren[6], Cristina Aguado[5], Masaoki Ito[8], Andrés Aguilar-Hernández[9], Miguel Angel Molina-Vila[5] (iD) and Rafael Rosell[1,9,10] (iD)

1 Germans Trias I Pujol Research Institute, Badalona, Spain
2 Department of Biochemistry, Molecular Biology and Biomedicine, Autonomous University of Barcelona, Spain
3 Department of Genetics, University of Granada, Spain
4 Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Spain
5 Laboratory of Oncology, Pangaea Oncology, Barcelona, Spain
6 NanoString Technologies, Seattle, WA, USA
7 UCSF Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA, USA
8 Department of Surgical Oncology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Japan
9 Oncology Institute Dr. Rosell, IOR, Quirón-Dexeus University Institute, Barcelona, Spain
10 Autonomous University of Barcelona, Spain

Although many studies highlight the implication of circular RNAs (circRNAs) in carcinogenesis and tumor progression, their potential as cancer biomarkers has not yet been fully explored in the clinic due to the limitations of current quantification methods. Here, we report the use of the nCounter platform as a valid technology for the analysis of circRNA expression patterns in non-small cell lung cancer (NSCLC) specimens. Under this context, our custom-made circRNA panel was able to detect circRNA expression both in NSCLC cells and formalin-fixed paraffin-embedded (FFPE) tissues. CircFUT8 was overexpressed in NSCLC, contrasting with circEPB41L2, circBNC2, and circSOX13 downregulation even at the early stages of the disease. Machine learning (ML) approaches from different paradigms allowed discrimination of NSCLC from nontumor controls (NTCs) with an 8-circRNA signature. An additional 4-circRNA signature was able to classify early-stage NSCLC samples from NTC, reaching a maximum area under the ROC curve (AUC) of 0.981. Our results not only present two circRNA signatures with diagnosis potential but also introduce nCounter processing following ML as a feasible protocol for the study and development of circRNA signatures for NSCLC.

1. Introduction

circRNAs are a newly re-defined type of endogenous RNA molecules originated by a noncanonical process called 'back-splicing'. Through this mechanism, the 5′ splice donor covalently links to the 3′ end of an upstream exon, resulting in a single-stranded circular structure which can include one or different exonic/intronic regions [1]. This particular assembly lacking a poly(A) tail makes them very stable and resistant to exonuclease-mediated degradation when compared to their linear counterparts [2].

The existence of circRNAs has been acknowledged for more than 45 years. First evidence was reported in 1976 with the first description of viroids as 'single-stranded and covalently closed circular RNA molecules' [3], and their discovery in humans followed almost two decades later [4]. However, it is not until recently that their role has been clarified, evolving from abnormally spliced unfunctional 'scrambled' transcripts to circular RNA molecules with a marked role in homeostasis [5,6].

CircRNAs have been classified as noncoding RNA for many years, due to the lack of a 5′ cap structure and their inability to bind to ribosomes. However, recent studies reported that some circRNAs can be translated into small functional peptides in a cap-independent manner [7]. Other functions may include serving as protein decoys, scaffolds, and/or recruiters [8], or regulating the canonical transcription by competing with the formation of linear cognates via back-splicing [9,10]. Nonetheless, the most well-studied function is their interaction with miRNAs. A single circRNA can have several miRNA-binding sites through which targeted miRNAs get 'sponged', thereby blocking their activity [11]. It is throughout this mechanism how they predominantly exert their role as cell proliferation regulators targeting mediators of classical signaling pathways, such as MAPK/ERK, PI3K/AKT, and WNT/β-catenin, or cell cycle checkpoint regulators [11]. Because of their implication in the above-mentioned processes, dysregulation of circRNA expression can be associated to the development of different malignancies, including lung cancer. CircRNAs are significantly associated with tumorigenesis, proliferation, migration, and sensitivity to lung cancer therapies [12] and, as a result, have been presented in many recent studies as novel biomarkers to assess disease status.

However, the number of studies focusing on the development of circRNA signatures with either diagnostic or prognostic value in human malignancies is rather small, probably due to the lack of standardized circRNA quantification methods, which in turn is hampering the development of clinically applicable assays. RT-qPCR is widely used as a quantification tool for circRNA expression studies. While its sensitivity and short turnaround time proves beneficial for circRNA research, several events such as template switching, rolling circle amplification, or the bias attached to this technique may hinder the results [13]. In addition, it does not allow high-throughput analysis, which is necessary for biomarker discovery. Microarrays or RNAseq may overcome these limitations; however, the first have a limited range of detection disregarding those targets with either very low or high expression, while the latter not only results rather expensive but also includes other restrictions such as the use of long time-consuming protocols, or complex data analysis [14,15].

The nCounter technology allows multiplex analysis of up to 800 transcripts by direct capture and counting of individual targets [16]. With a short turnaround time and minimal hands-on work, it provides results in < 48 h with the use of an intelligible software. However, despite the growing number of laboratories using this platform, it still gets mostly restricted to mRNA analysis.

In this proof-of-concept study, we retrospectively analyzed the circRNA expression profiles in NSCLC cell lines and FFPE tissues by using a custom-designed 78 circRNA nCounter panel. Our data demonstrate that nCounter can be employed not only for basic circRNA research but also for the development of clinically useful circRNA signatures.

2. Materials and methods

2.1. Patient samples and cell lines

This study was carried out in accordance with the principles of the Declaration of Helsinki, under an approved protocol of the institutional review boards of Quirón Hospitals, and the IGTP-HUGTP Biobank. FFPE lung cancer tissues were retrospectively collected from 27 early-stage and 26 late-stage cancer patients from the different Quirón hospitals (Table 1). FFPE tissue samples from 16 donors were collected as controls from the IGTP-HUGTP Biobank. Most controls did not present any type of cancer, except for four samples which were extracted from the nontumorigenic region of the lung from a cancer patient. Individuals with different pathologies were also included to ensure

Table 1. Clinicopathologic characteristics of enrolled patients ($n = 69$). NSCLC, non-small cell lung cancer.

Clinicopathological characteristics	Lung cancer patients ($n = 53$)	Noncancer controls ($n = 16$)
Gender—no. (%)		
Male	28 (52.8)	10 (62.5)
Female	25 (47.2)	6 (37.5)
Age—years		
Median	66	59
Range	32–85	29–76
Smoking status—no. (%)		
Ex- or current smoker	40 (75.5)	9 (56.25)
Never smoker	11 (20.8)	5 (31.25)
Not information	2 (3.7)	2 (12.5)
Histological type		
Adenocarcinoma	43	–
Squamous carcinoma	1	–
Other NSCLC	9	–
Driver mutation		
EGFR	6	–
Exon19	3	–
Exon21	1	–
Exon20-21	1	–
Exon21 and amplification	1	–
KRAS	12	–
G12A	2	–
G12C	3	–
G12V	4	–
G12R	1	–
Other	2	–
BRAF	1	–
ROS	1	–
RET	2	–
ALK	1	–
MET (exon14 mutation)	1	–
Other alterations	5	
Not information	24	–
Tumor stage—no. (%)		
I	16 (30.2)	–
II	4 (7.5)	–
IIIA	7 (13.2)	–
IIIB	3 (5.6)	–
IV	23 (43.4)	–

Table 2. Characteristics of the lung cell lines included in the study. AD, adenocarcinoma; ATCC, American Type Culture Collection; NE, normal epithelial; UCSF, University California San Francisco; UTSW, University of Texas Southwestern.

Cell line	Histology	Gene	Mutation	Origin
A549	AD	*KRAS*	G12S	ATCC
HOP-62			G12C	ATCC
PC9		*EGFR*	E746_A750 DL	Hoffmann-La Roche, with the authorization of Dr. Mayumi Ono
HCC-827			E746_A750 DL	ATCC
NCI-H1666		*BRAF*	G466V	ATCC
NCI-H2228		*ALK*	*EML4-ALK*, variant 1	ATCC
NCI-H3122			*EML4-ALK*, variant 3	ATCC
AALE	NE	–	wt	Dr. Trever Bivona Lab, UCSF
HBEC30KT				Dr. Minna Lab, UTSW

A panel of seven human lung cancer cell lines harboring different mutations was selected along with two normal epithelial cell lines (Table 2). Cell lines were maintained following standard culture conditions [17] in RPMI-1640 or DMEM (Gibco, Life Technologies, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (Gibco). All cell lines were tested for mycoplasma infection.

2.2. RNA extraction

RNA extraction was performed following previously published methods [18,19]. RNA from fresh cell lines was isolated using the Allprep DNA/RNA/miRNA universal kit (Qiagen, Hilden, Germany). FFPE cells and tissues were deparaffined with xylene. After the removal of xylene using ethanol, RNA was extracted using the High Pure FFPET RNA isolation Kit (Roche, Rotkreuz, Switzerland). RNA quantification was performed using the Qubit 4 Fluorometer (Invitrogen, Carlsbad, CA, USA) with the Qubit RNA HS Assay Kit (Invitrogen). RNA integrity was assessed with the 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA) using the RNA 6000 Nano kit (Agilent Technologies).

2.3. Rnase-R treatment

5 μg of total RNA was either treated or mock-treated with RNase-R (Lucigen, Madison, WI, USA). RNA

the development of signatures specific of lung cancer (Table S1).

All collected samples were assessed for tumor and lymphocyte infiltration by a pathologist (Table S2).

Written informed consent was obtained from all patients and further documented; samples were de-identified for patient confidentiality. Clinical information collected from each patient was limited to gender, age, smoking status, tumor histology, driver mutation, and stage.

samples were denatured at 95 °C for 30 s following addition of a master mix containing RNase-R (or molecular grade water in the case of mock-treated samples), 10× RNase-R buffer adjusted to the final volume, and molecular grade water. Samples were incubated 160 min at 40 °C and kept at 4 °C prior RNA quantification and subsequent nCounter hybridization.

2.4. RT-qPCR and Sanger sequencing analysis

RT-qPCR and Sanger sequencing of circRNA junction sites were performed as previously described [18]. 10 μL of total RNA was converted into cDNA using the M-MLV reverse transcriptase enzyme and random hexamers (Invitrogen).

A 1 : 3 dilution of cDNA was performed, and 2.5 μL were added to the Taqman Universal Master Mix (Applied Biosystems) in a 12.5 μL reaction containing a specific pair of primers and probe for each gene. Three replicas of each sample were run for the quantification of the expression of each assessed circRNA. Three replicas of each sample were run for the quantification of the expression of each assessed circRNA. Divergent primers and probe sets were designed using Primer Express 3.0 Software (version 3.0.1, Applied Biosystems) with the latter spanning the circRNA junction site (Table 3). Quantification of gene expression was performed using the QuantStudioTM 6 Flex System (Applied Biosystems) and calculated according to the comparative Ct method.

In all quantitative experiments, a sample was considered not evaluable when the standard deviation of the Cq values was > 0.30 in two of the three independent analyses ($n = 3$).

For Sanger sequencing, 10 μL of each PCR product was loaded on a Precast Agarose HT-1gel and visualized under UV light (E-Gel™ Safe Imager™ Real-Time Transilluminator, Invitrogen) after electrophoresis (E-Gel™ iBase™ Power System, Invitrogen).

Five microliters of each cDNA sample were purified using the PCR ExoSAP-IT Product Clean up Reagent (Applied Biosystems). Sequencing PCRs were set up using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), forward primer, cDNA and water in a final volume of 20 μL. Sequencing PCR was performed using a Verity 96 well thermal cycler (Applied Biosystems).

After sequencing amplification, samples were loaded into a 96-well plate and subjected to Sanger sequencing using the 3130 Genetic Analyzer (Applied Biosystems).

2.5. miRNA prediction and circRNA-miRNA network construction

MiRNAs targeted by the differentially expressed circRNAs found in early-stage FFPE lung cancer tissues were predicted using the CIRCINTERACTOME tool (https://circinteractome.nia.nih.gov). circRNA-miRNA interaction network was built using CYTOSCAPE (v3.8.2; https://cytoscape.org). Association of miRNAs with cancer-associated downstream signaling pathways was investigated using the miRCancer database (https://mircancer.ecu.edu).

2.6. NanoString nCounter panel design and sample processing

A custom-made panel of 78 circRNAs was produced, including both highly and lowly expressed circRNAs that could be related to lung cancer (Table S3). Each probe was designed to target a flanking exonic sequence between 35–55 nucleotides of the circRNA junction site. They also contain a complementary region to capture and reporter probes, conforming a precise configuration that allows specific recognition of circular transcripts (Fig. S1). In addition, six linear reference genes (GAPDH, MRPL19, PSMC4, RPLP0, SF3, and UBB) and four mRNAs of FAM13B,

Table 3. Primer and probe design for circRNA validation by RT-qPCR. In blue marked the junction site.

circRNA		
circEPB41L2 (hsa_circRNA_0001640)	Forward	GAAGACCAAAACTGTCCAGTGTAAAG
	Reverse	CACTTCAGACACAGAGCCTACTTCA
	Probe	TGACCTGGAGCATAAG
circSOX13 (hsa_circRNA_0004777)	Forward	CAGTGACTGGAAGGAGAGGTTTC
	Reverse	CTGGGCAGAGATGGGGCT
	Probe	AAAGATGTCAAAGGATGTCCATGA
circBNC2 (hsa_circ_0086414)	Forward	GTCTGCACAGTGGCTGGTTG
	Reverse	GGTGATGATTTCCTCTTCTCGAG
	Probe	AGACAGGATGCTGCTG

HIPK3, MGA, and UBXN7 genes were included (Table S2).

Sample processing in the nCounter was performed as previously described [18] following NanoString's guidelines (Fig. S2).

2.7. Data normalization and differential expression analysis

Raw count values were exported to Microsoft Excel (version 16.40, Microsoft, Redmond, WA, USA) using nSolver Analysis Software (version 4.0.70, NanoString Technologies, Seattle, WA, USA). For each of the circRNAs included in the panel, raw counts lower than the cut-off value established as background were automatically excluded from further analysis. Background was calculated for each sample by using the mean of the negative probe counts plus two times the standard deviation. Only circRNAs with a value > 10 counts after background subtraction were considered as expressed. Subsequent circRNA-specific counts were normalized by dividing this number by the total number of counts for this sample. Resulting number was multiplied by 10 000 (units expressed in counts per 10 000).

Further differential expression analysis of raw nCounter data was carried out with R (version 4.0.2; R Core Team and the R Foundation for Statistical Computing, Vienna, Austria) and R studio (version 1.3.1056; RStudio PBC, Boston, MA, USA). Technical variability correction, normalization, and differential expression analysis was performed using the RUVSEQ (version e1.24.0; Bioconductor Core Team, Buffalo, NY, USA) and DESEQ2 (version 1.30.0; Bioconductor Core Team) packages (RUVSEQ-DESEQ2, Bioconductor Core Team). Firstly, the RUVg function was used to estimate the unwanted variation among samples based on the positive controls. The positive controls used in the NanoString panel are Spike-In control sequences; therefore, analogous constant expression of positive controls is expected across all samples. Secondly. DESeq2 was used to perform the normalization of the data, while accommodating the estimated factors provided by the RUVg function. Finally, DESeq2 was used to perform hypothesis testing in order to identify differentially expressed circRNAs. Shrunken $\log_2$ fold-change ($\log_2$FC) was then reported by DESeq2 along with adjusted *P*-values. Batch effect was considered during normalization using RUVSeq-DESeq2. The normalized data were employed for ML techniques.

Volcano plots were used to visualize $\log_2$FC on the *x*-axis and $-\log_{10}$ adjusted *P*-values on the *y*-axis.

2.8. Machine learning classification

Recursive feature elimination (RFE) was used to perform feature selection and the LOOCV algorithm was applied on the full panel of circRNA transcripts. The number of features to select was set by default at 4, 8, 16, and 78. The number of features that yielded best performance after cross-validation was automatically selected. To test whether generated data had enough discriminative information to build a robust model for the classification of cancer samples from controls, different paradigms of classification models were tested to provide the most accurate results. Under this context, three classification approaches were performed with the selected features: an 'instance-based' model (KNN). This model uses the distances among samples to obtain a predictive label; and two different ensemble mechanisms with decision trees—bagging (RF) and boosting (GBM).

For the analysis of early-stage lung cancer samples versus control samples, GBM was excluded due to the high volume of samples is required for this model.

The model with the highest ROC AUC value was then selected as the final model. A confidence threshold of 0.5 was considered for the calculation of PPV and NPV. Additional statistical indicators such as accuracy, sensitivity, and specificity were also calculated.

3. Results

3.1. nCounter for circRNA detection in fresh NSCLC samples

Based on a literature review, 78 circRNAs were selected according to their differential expression in lung cancer specimens for the development of an nCounter panel (Table S3). To test the reproducibility of this panel for circRNA detection, RNA from fresh PC9 cells was subjected to nCounter analysis in three independent reactions. As a result, a strong correlation was found between the normalized counts for each individual circRNA, represented by a Spearman's *r* of 0.82–0.88, *P* < 0.01 (Fig. S3).

Then, RNA from the same cell line was used in an experiment with RNAase-R, an enzyme that degrades linear RNA, to elucidate if the nCounter probes bind specifically to the circRNA of the genes included in the panel (Fig. 1A). As a result, 18 new transcripts that could not be detected in mock-treated samples were observed after RNase-R treatment (Fig. 1B). In addition, among the 34 transcripts identified in both
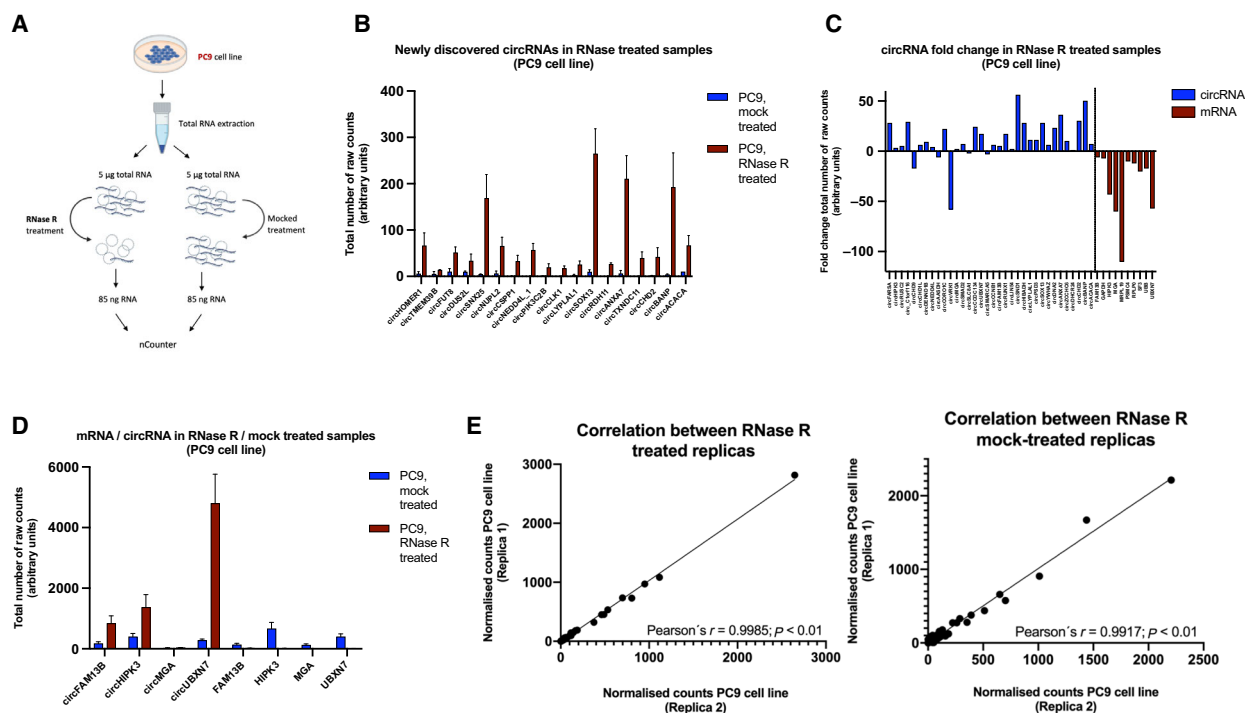
Fig. 1. Analysis of circRNA from RNase-R-treated samples. (A) Workflow for circRNA enrichment with RNase-R. (B) Representation of the newly discovered circRNAs after RNase-R treatment. Bars indicate the mean of the replicas ($n = 2$). Error bars indicate SD. (C) CircRNA/linear HK fold-change after RNase-R treatment ($n = 2$). (D) Comparison of circRNAs/mRNA cognates in RNase-R/mock-treated samples. Bars indicate the mean of the replicas ($n = 2$). Error bars indicates SD. (E) Correlation of the nCounter replicas ($n = 2$) for each treatment. Pearson's coefficient is indicated.

types of samples, the counts of 28 (82.3%) increased at least 2-fold after RNAse-R treatment. CircSND1 and circBANP were found with the highest enrichment, with a 56- and 33-fold change, respectively. CircCHD9, circAASDH, circVRK1, circSLC8A1, and circSMARCA5 were the only circular transcripts affected by the exonuclease activity of RNase-R, showing a lower number of counts after incubation with the enzyme (Fig. 1C). All mRNA controls, including the linear forms of FAM13B, HIPK3, MGA, and UBXN7, were found with reduced or null expression after treatment (Fig. 1D).

A high correlation was found between the two replicas included for each of the conditions (Pearson's $r = 0.99917$; $P < 0.01$ and $r = 0.9985$; $P < 0.01$ for mock-treated and RNase-R-treated samples, respectively) demonstrating the specificity of the assay (Fig. 1E).

3.2. nCounter for circRNA detection in FFPE NSCLC samples

To assess the performance of our panel in FFPE samples, RNA from paired FFPE and fresh PC9 cell line was extracted and processed in the nCounter. The

number of total raw counts in PC9 FFPE samples was significantly lower compared to fresh PC9 samples (771.870 versus 1.353.811). However, despite the suboptimal quality observed in the RNA extracted from the FFPE cells (Fig. S4), a statistically significant correlation was found when comparing both types of input (Fig. S5A).

Next, we assessed the feasibility of RNase-R treatment in FFPE samples. As a result, overall circRNA enrichment was not achieved, in contrast to what was previously observed in RNA extracted from fresh cells. Most circRNAs were found to be degraded to different extents in RNase-R-treated replicas when compared with the controls, indicating that such treatment should be avoided when working with FFPE samples (Fig. S5B).

Then, different concentrations of FFPE-derived RNA (between 250 and 2000 ng of total RNA) were tested assessing the effect on downstream nCounter analysis. As a result, saturation was not achieved with the highest concentration, suggesting that a greater RNA input could be applied. Analysis of normalized counts across all samples indicated similar performance of 250 ng compared to the rest of tested concentrations, with a Pearson's correlation between

0.99–1.00 (Fig. S6). As a result, 250 ng of total RNA was selected for the rest of the study.

3.3. circRNA expression in NSCLC fresh cell lines

A set of seven lung cancer cell lines were selected according to their driver mutation, along with two normal epithelial cell lines (Table 2). Duplicates of equal RNA concentrations were run in all cases.

Out of the 78 circRNAs included in the panel, 33 were expressed in all cell lines. Nineteen were expressed in epithelial cells and not in all lung cancer cells, while only one, circFUT8 was only expressed in all lung cancer cell lines (Fig. S7). Nineteen circRNAs included in the panel were not found in any of the assessed cell lines. Fifty-one was the highest number of circular transcripts displayed by any cell line (AALE). The NCI-H2228 cell line showed the lowest number, with only 40 circRNAs detected (Fig. 2A). Overall, total raw counts were significantly higher in normal epithelial lung cell lines compared to cancer cell lines (Fig. S8). Hierarchical clustering led to a separation of the KRAS cell lines and normal epithelial cell
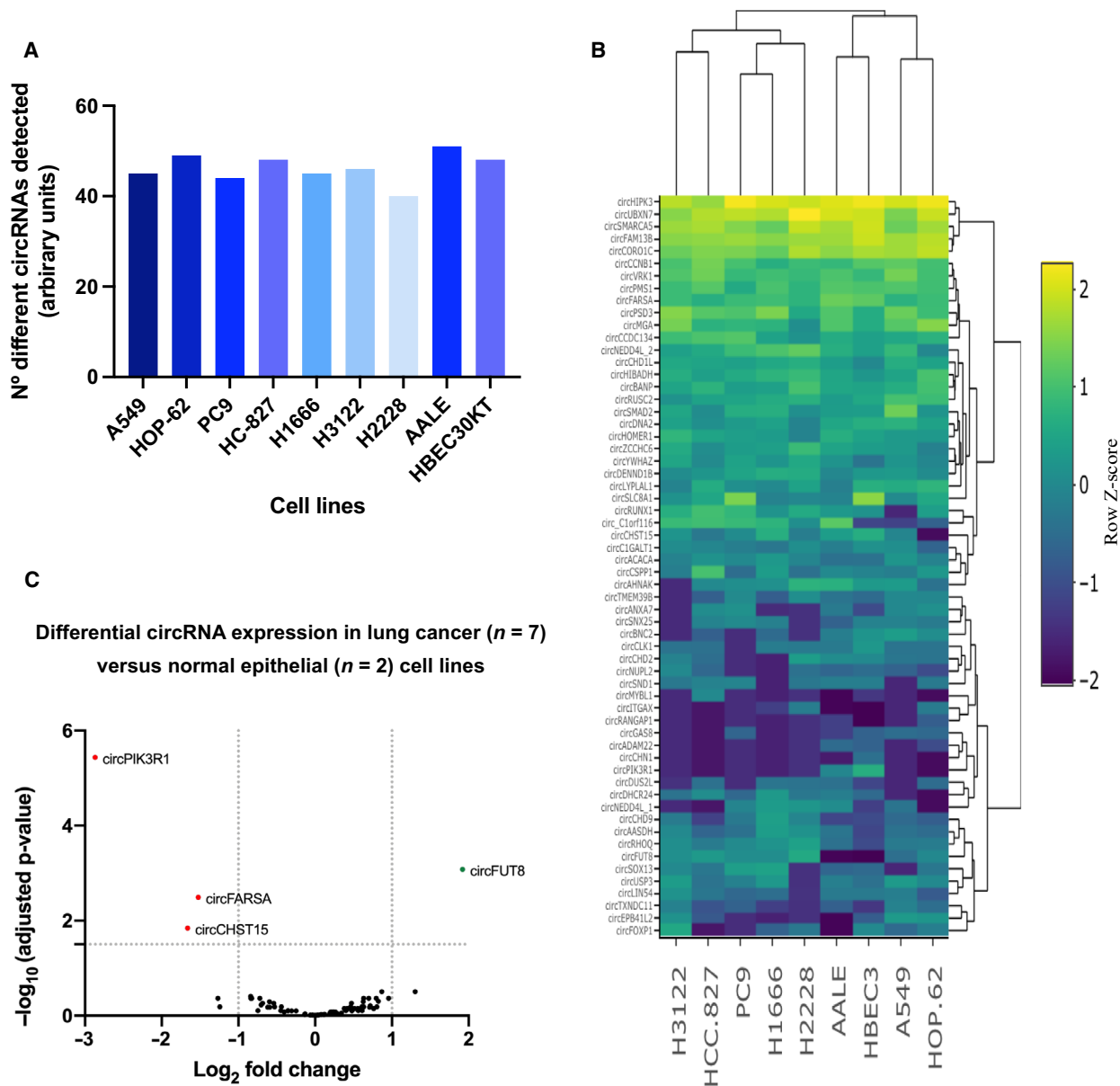
Fig. 2. CircRNA analysis in lung cancer (A549, HOP-62, PC9, HCC-827, H1666, H3122, and H2228) and epithelial cells (AALE and HBEC30KT). (A) Bar plot representing total circRNAs detected in each of the cell lines. (B) Hierarchical clustering of cell lines based on circRNA expression. (C) Differential circRNA expression analysis of $\log_2$-normalized counts between lung cancer and normal lung cells.

lines from the rest (Fig. 2B). The two EGFR mutant cell lines positioned together, showing a distinctive group of downregulated circRNAs (circBNC2, circCLK1, circCHD2, and circNUPL2) compared to the rest of the cell lines.

Finally, differential expression analysis revealed four circRNAs that allowed for differentiation between the seven-lung cancer cells and normal epithelial cells. CircPIK3R1, circFARSA, and circCHST15 were found downregulated in the cancer cell lines, while circFUT8 was upregulated (Fig. 2C).

3.4. circRNA expression in FFPE NSCLC versus nontumor tissue

A total of 53-lung cancer samples and 16 control tissue samples were selected and processed with the circRNA nCounter panel. Initial analysis included normalization of counts for each circRNA as described in the methods section, followed by unsupervised hierarchical clustering of patient samples based on total circRNA expression. A partial separation between cohorts was achieved, indicating a group of circRNAs with discriminatory potential (Fig. 3A).

A differential expression analysis revealed a cluster of 10 differentially expressed circRNAs. CircEPB41L2, circBNC2, circSOX13, and circFOXP1 were downregulated in lung cancer tissues, while circRUNX1, circCHD9, circACACA, circFUT8, circRHOQ, and circC1ORF116 were overexpressed (Fig. 3B).

Additionally, we also investigated the possible differences in circRNA expression based on the smoking habits of the lung cancer cohort. As a result, four circRNAs (circCSPP1, circNEDD4L, circSOX13, and circCORO1C) negatively correlated with smoking status with $P = 0.015$, $P = 0.043$, $P = 0.017$, and $P = 0.045$ respectively (Student's *t*-test).

Next, a ML approach was used to develop a circRNA signature predictive of lung cancer.

Due to the low number of samples to be analyzed ($n = 59$), we decided to use LOOCV as a validation model, which considers only one sample for testing in each interaction reducing the bias to the minimum when compared to other techniques such as stratified cross validation. As a result, a RFE algorithm selected an 8-circRNA signature (including circSOX13, circEPB41L2, circFOXP1, circBNC2, circCORO1C, circCHD9, circSNX25, and circPIK3R1) as the final model, providing a ROC AUC of 0.965, 0.953, 0.983 with RF, KNN, and GBM classifiers, respectively (Fig. 3C). A PPV of 98.1% and NPV of 81.2% were achieved with the final model. The accuracy, sensitivity, and specificity of the signature were of 97.1%, 94.5%, and 92.8%, respectively.

3.4.1. CircRNA expression in early-stage NSCLC tissues

Next, the 27 early-stage NSCLC samples (stages I–IIIA) of our cohort were compared to the 26 late-stage specimens (stages IIIB and IV) (Table 1) to assess those differentially expressed circRNAs emerging early in the disease.

From the 41 circRNAs expressed in early-stage samples, 39 were shared with late-stage samples (Fig. 4A). Only 6 out of these 39 transcripts were differentially expressed when compared with the control specimens (Fig. 4B). Interestingly, one of these circRNAs (circFUT8) was found upregulated in both lung cancer tissues and lung cancer cell lines. To shed some light on the potential targets of these six circRNAs, a circRNA-miRNA network was built based on sequence-pairing prediction (Fig. 5). Using circinteractome database, 64 miRNAs were found to potentially bind to differentially expressed circRNAs, with 29 of them showing more than 1 binding site (Fig. S9).

Additional ML analysis was performed in early-stage lung cancer and control samples. RFE algorithm provided a signature that included four circRNAs (circEPB41L2, circSOX13, circBNC2, circCORO1C) and provided a ROC AUC of 0.981, 0.918 with RF and KNN, respectively (Fig. 6A). PPV and NPV were of 92.6% and 87.5%, whereas accuracy, sensitivity, and specificity were of 90.6%, 92.6%, and 87.5%, respectively with the selected model. Hierarchical clustering based on the four circRNA included in the signature allowed a clear differentiation between both cohorts (Fig. 6B).

3.5. Univariate analysis related to lung cancer risk

We then explored if certain patient characteristics could provide risk factors for lung cancer by performing a univariate analysis (Fig. 7). Several characteristics that could be associated with higher risk of lung cancer such as age, gender, and smoking status were evaluated. No significant association could be found between lung cancer and any of the characteristics previously mentioned. However, presented signatures for lung cancer and early-lung cancer classification were found to be significant predictive factors for lung cancer, with an odds ratio of 371 and 91, respectively.

3.6. Validation by RT-qPCR and Sanger sequencing of circRNA junction sites

CircEPB41L2, circSOX13 and circBNC2 not only were significantly downregulated both in early and late

A

B

C

PREDICTION

|	Healthy	Cancer	Class.Err.
REFERENCE Healthy	13	3	0.1875
Cancer	1	52	0.0188
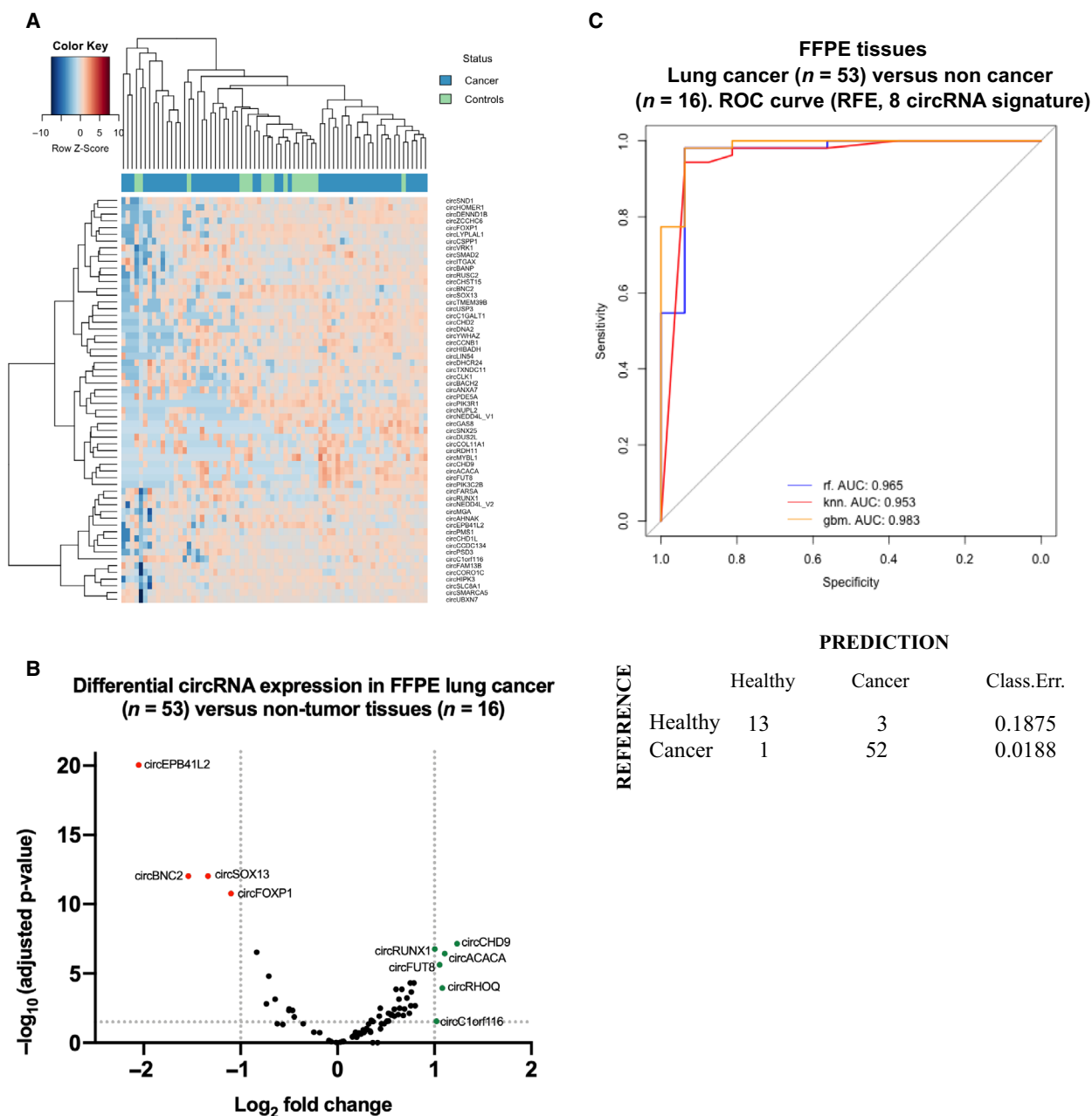
Fig. 3. (A) Heatmap showing the circRNA expression in lung cancer and control specimens. Unsupervised clustering was performed based on total circRNA expression. (B) Volcano plot showing the circRNA $\log_2$ fold-change in FFPE lung cancer ($n = 53$) versus control ($n = 16$) FFPE tissues. (C) Area under the ROC curve for the classification of lung cancer and control samples. Confusion matrix was generated based on the RF classification scores. Classification error scores are indicated.

stages showing the highest fold-change, but also were selected by the RFE algorithm as part of the two predictive signatures. As a result, these 3 targets were selected to validate the nCounter performance using RT-qPCR.

Divergent primers and probes spanning the junction sites were designed for the specific amplification of

cited circular transcripts (Fig. 8A) in 10 NSCLC and 10 control FPPE tissue samples previously assessed with the nCounter circRNA panel.

RT-qPCR results correlated with the data previously obtained from nCounter, indicating downregulation of cited circRNAs in NSCLC samples (Fig. 8B). A gel electrophoresis of the PCR products revealed three
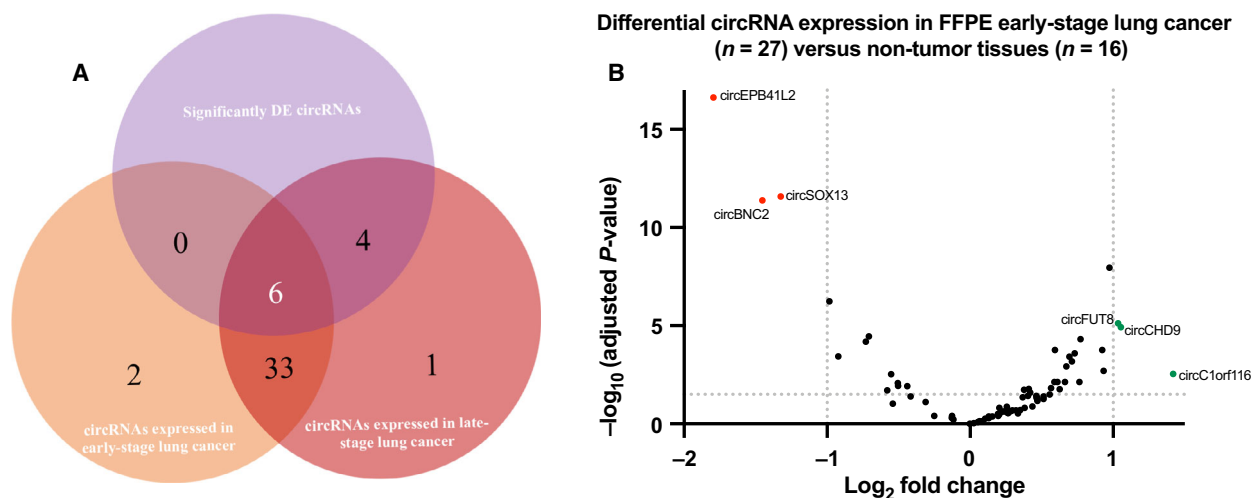
Fig. 4. CircRNA expression in early-stage NSCLC samples. (A) Venn diagram displaying circRNAs identified in early- and late-stage samples, featuring those shared by both cohorts. DE circRNAs are indicated. (B) Differential expression analysis of log2-normalized counts between the early-stage lung cancer cohort (*n* = 27) and control (*n* = 16) FFPE tissues. circEPB41L2, circSOX13, and circBNC2 were found downregulated and circFUT8, circCHD9, and circ_C1orf116 were found upregulated as previously described with all stages of lung cancer.

bands corresponding to the size of expected amplicons (Fig. 8C). Further Sanger sequencing validated these findings by exposing the circRNA junction sites (Fig. 8D).

4. Discussion

Precision oncology currently relies on genomic, transcriptomic, or proteomic-based features that serve as decision-making support, predicting treatment outcome [20].

The re-discovered role of circRNAs as regulatory entities of miRNAs, affecting the occurrence and development of different malignancies, has been supported by the growing number of studies that highlight their potential as cancer biomarkers and therapeutic targets in future personalized medicine [21]. Investigation of novel signatures based on these biomolecules could therefore be of interest to achieve earlier diagnosis by developing new tests or complimenting existing ones. However, the lack of standardized methods for their study is preventing their clinical validation and further implementation in the clinical practice.

The nCounter platform allows multiplexed digital gene expression analysis by direct counting of RNA molecules. With a wide use for transcriptomic studies, nCounter has been recently adapted for the detection of circRNAs using a specific probe design where sequences span the circRNAs junction site [19]. On this regard, some authors have proved the benefits of this technology to study circRNA subcellular

distribution [22], or elucidate the potential roles in skin [23,24] or brain diseases [25]. However, to date, no one has explored this platform in FFPE samples for the development of lung cancer signatures.

Here, we prove the use of nCounter for circRNA studies in FFPE lung cancer tissues and cell lines, developing a protocol for their study.

Due to the lack of any commercially available nCounter circRNA panel, we first performed an extensive literature research, looking for circular RNA candidates described to be differentially expressed in lung cancer cells, tissues, or liquid biopsies. Out of the 78 circRNAs conforming the panel (Table S3), 40–51 circRNAs were detected in assessed cell lines, whereas 41 and 44 were detected in early- and late-stage NSCLC tissues, respectively. From the resulting circRNAs that could not be detected by nCounter, 19 could not be found in any control tissue nor in any assessed cell line (Fig. S10). Additional experiments using liquid biopsies would be of interest to address if cited circRNAs are present in such material according to our nCounter protocol, or if on the contrary, observed discrepancies may be due to the technical differences (including normalization) among the diverse platforms used in previous studies such as RNA-seq, microarrays, or RT-qPCR when compared to our nCounter workflow.

The resulting 13 circRNAs were detected in the cell lines with three of them, circPIK3R1, circADAM22, circCHN1 being only present in the normal epithelial cell lines (AALE and HBEC30KT). These three were described in literature to be downregulated in NSCLC
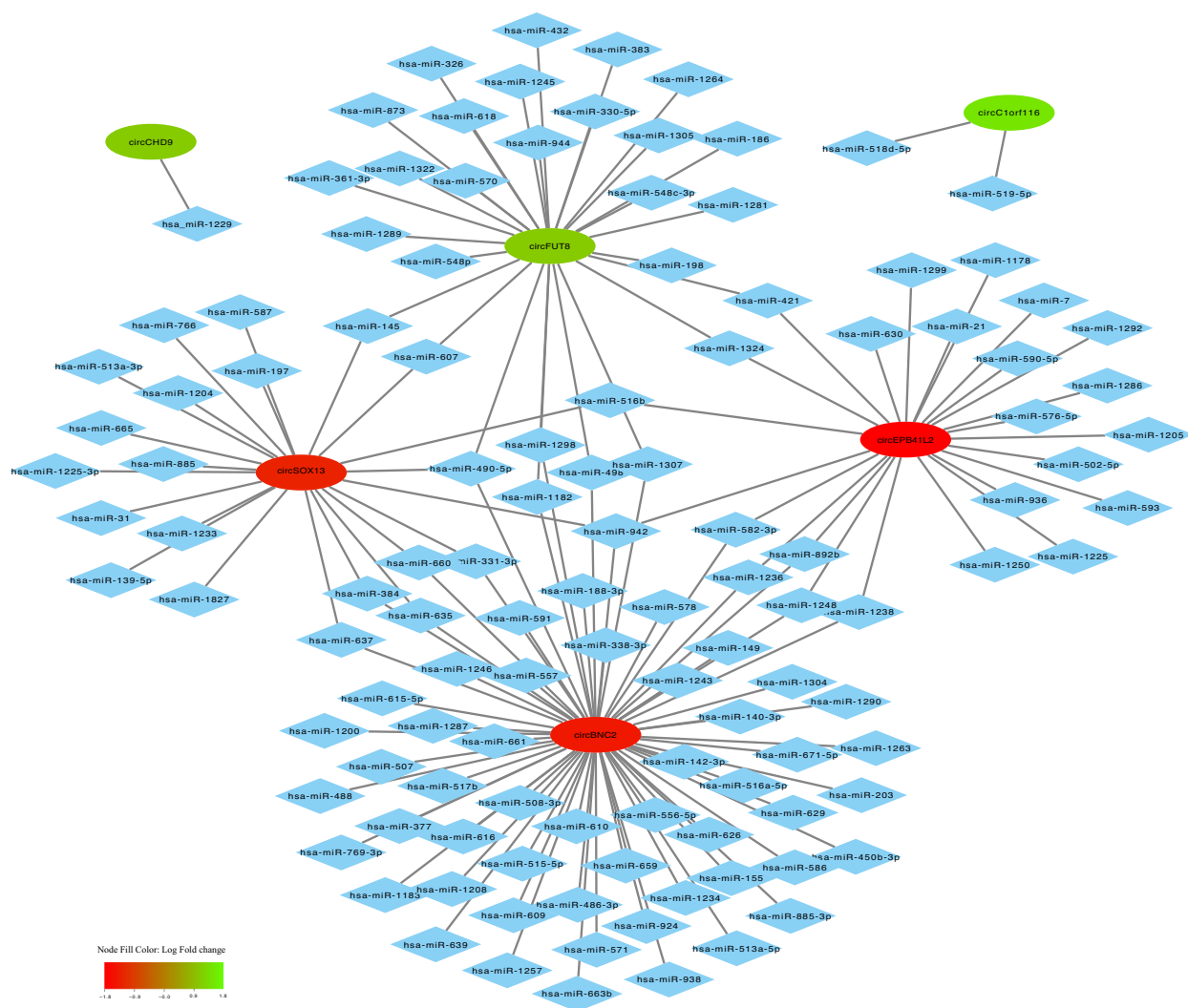
Fig. 5. Mapping network showing predicted sequence-pairing circRNA-miRNA interaction of differentially expressed circRNA found in early-stage lung cancer tissues. CircRNAs are represented by elliptic nodes and colored based on their log fold-change. Complementary binding miRNAs are represented by diamond shaped nodes.

[26–30]. From the detected circRNAs, 17 were found upregulated, according to literature review [26,28–30], contrasting to the results achieved with nCounter; however, only three of those could be further validated by RT-qPCR [26,28,30]. Although we believe that direct comparison with another circRNA panel for lung cancer detection is instrumental to fully assess the clinical utility of our panel, this was not performed due to the absence of the latter; however, this comparison will be warranted at the time other panels become available.

Most genetic analyses performed in the clinic come from paraffined specimens with either very little material or compromised quality. CircRNAs are very stable, even in this type of samples, due to their circular configuration [29]. Also, the nCounter technology performs quite well with highly degraded samples compared to other techniques since it only requires a short fragment of RNA (100 nt) for the capture and reporter probes to hybridize and emit a signal [16]. In consequence, we tested and compared results of circRNA expression from FFPE and fresh PC9 cells after nCounter analysis even if the RNA did not pass the quality control, as observed in the case of the FFPE PC9 samples, and we did obtain comparable results.

RNase-R treatment can efficiently degrade highly structured RNA in 3′ end-dependent manner [31]. Since most circRNAs are resistant to this exoribonuclease activity, we tested the specificity of our panel by treating cell line-derived RNA with this enzyme prior
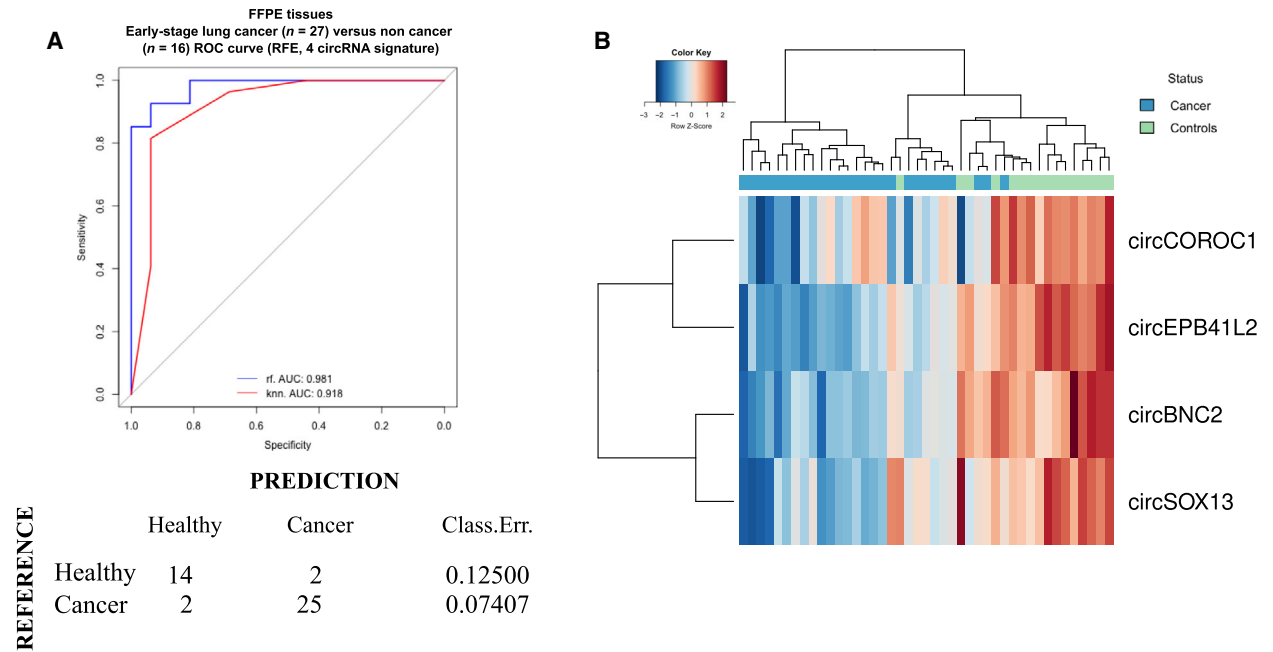
Fig. 6. (A) Area under the ROC curve of the 4 circRNA-signature using recursive feature elimination (RFE) for cohort classification. Confusion matrix was generated based on the RF classification scores. Classification error scores are indicated. (B) Hierarchical clustering of samples based on the 4-circRNA signature.
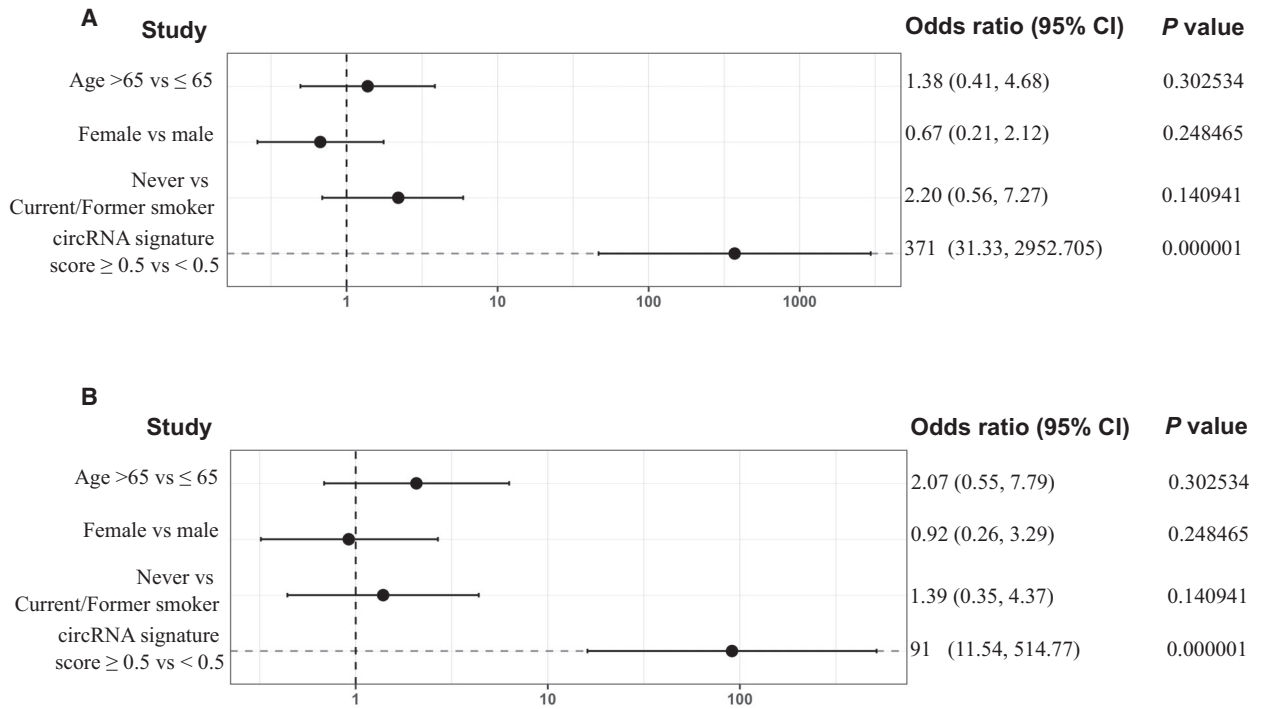
Fig. 7. Univariate analysis exploring associations between patient characteristics and lung cancer to determine risk factor. Forest plot represents the odds ratios in (A) lung cancer; and (B), early-stage lung cancer cohorts with a 95% Wald confidence limit. Student's t-test was used for the calculation of *P*-values.
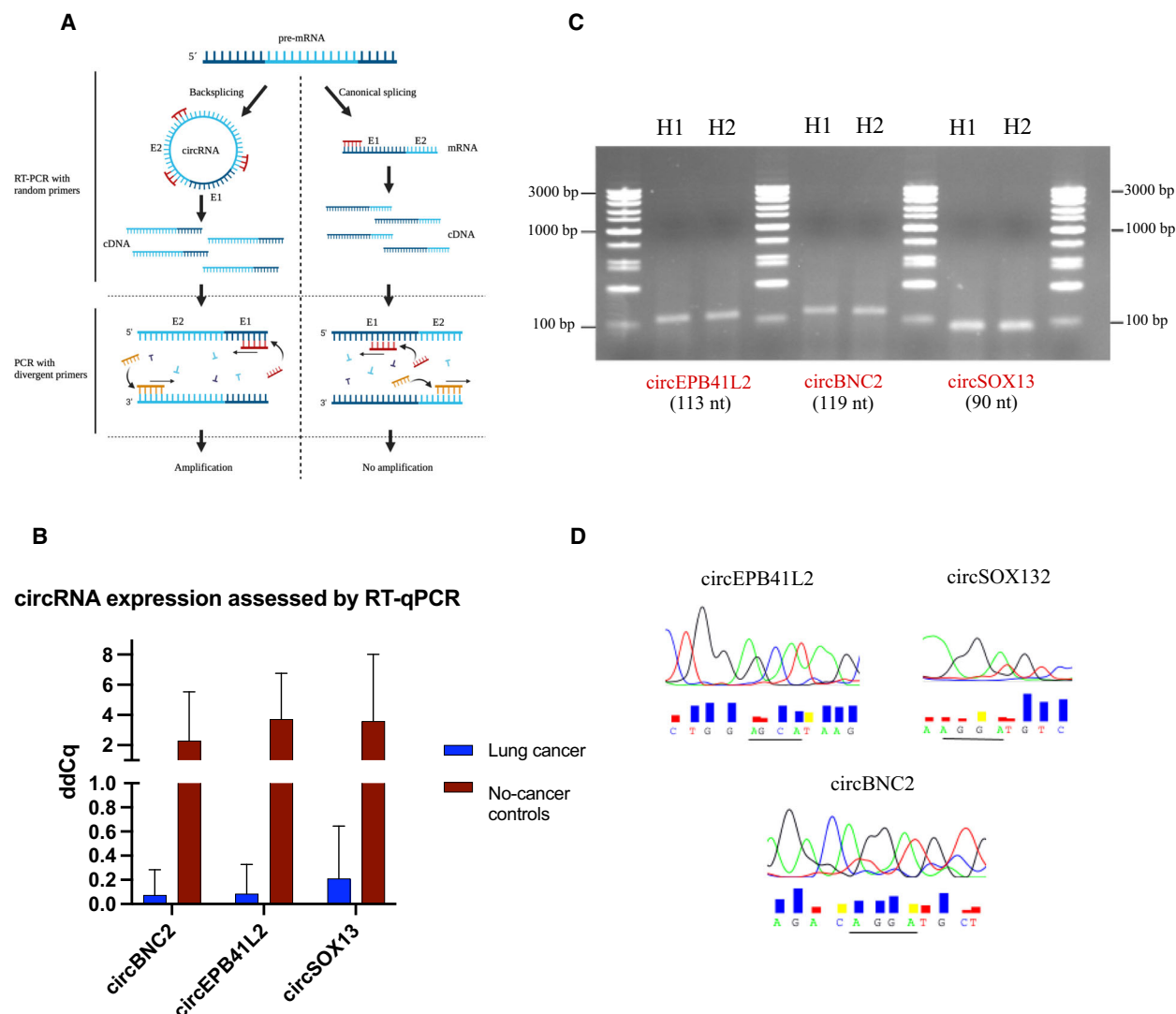
Fig. 8. Validation of nCounter results by RT-qPCR and further Sanger sequencing. (A) Representation of circRNA amplification using divergent primers. (B) Bar plot of RT-qPCR results depicting downregulation of circEPB41L2, circSOX13, and circBNC2 in lung cancer versus control tissues validating previous nCounter results. Bars indicate the mean of the 10-lung cancer (n = 3) and 10 control samples (n = 3). Error bars indicate SD. (C) Electrophoresis gel of amplified circEPB41L2 (113 nt), circBNC2 (119 nt) and circSOX13 (90 nt). (D) Sanger sequencing results spanning the junction site (underlined) of cited circRNAs.

nCounter processing. Consequently, most expressed circRNAs were enriched up to 56-fold when compared to controls, and only five circRNAs were found affected by this treatment. Sensitivity of specific circRNAs toward the endonuclease activity of the RNase-R enzyme was expected since it could be found reported in other publications [32]. Full or partial degradation of all linear transcripts included in the panel was observed, hence, validating the circRNA nCounter panel. 18 new circRNAs could be seen after treatment, while they could not be detected in mock-treated samples. The degradation of the canonical mRNA which can

represent up to 95% of the total RNA expression [33,34] seems to facilitate the interaction between the circRNAs and the nCounter probes, which otherwise would be hampered by this mRNA-induced noise, making those low-expressed circRNA undetectable [35]. This, along with the enrichment of circRNA molecules upon linear RNA depletion suggests that this type of treatment may be particularly beneficial for the screening of circRNA (especially those with very low expression) derived from fresh material. Conversely, circRNA enrichment was not observed in treated FFPE-derived RNA samples. As the rest of nucleic acid present in this

type of material, circRNAs are crosslinked to the paraffin matrix. During the process of purification, these molecules are subject to both mechanical and chemical breakage; thus, any break in the circRNA would allow for RNAse-R-based degradation. As a result, we determined that this procedure can be recommended to improve circRNA detection in fresh but not paraffined specimens. However, it is imperative to mention that although RNase-R treatment is highly recommended for circRNA screening purposes, it should be avoided in circRNA expression studies since the variability of RNAase-R digestion efficiency for different samples may lead to biased circRNA expression quantification [36]. Therefore, untreated total RNA samples were used for the expression experiments in our study.

Since circRNA represents only 5–10% of total RNA [33,34], different concentrations of total RNA were tested. As a result, 250 ng of total RNA proved enough for expression studies. Technical saturation was not achieved at 2000 ng of total RNA suggesting that higher concentrations could be used if analysis of transcripts expressed at lower levels is intended.

Using the explained workflow and custom-made circRNA nCounter panel, expression analysis in lung cancer cell lines was performed. Interestingly, an overall increase in the number of circRNA raw counts was found in normal epithelial versus cancer cells. This result is in agreement with a previous study, where a global reduction of circRNA expression in cancer compared to healthy specimens was found, along with a negative correlation of overall RNA abundance and proliferation [37].

In addition, a group of differentially expressed circRNAs was discovered in the assessed cancer cell lines. Interestingly, although circPIK3R1 was downregulated in agreement with formerly published results [30], both circCHST15 and circFARSA were also downregulated. CircCHST15 was recently found highly expressed in lung cancer, correlating with PD-L1 status and promoting immune escape of lung cancer cells [38]. Similarly, circFARSA upregulation has been described in tumor cells, promoting migration and invasion [39]. Although none of the groups used AALE nor the HBEC30KT epithelial cell line for their transcriptional analyses preventing direct comparison with our study, additional experiments with other epithelial cell lines and additional transfection studies could be of interest to shed light on the biology of these circRNAs.

Furthermore, a circRNA from the FUT8 gene which was found upregulated in cancer cells and further validated in FFPE lung cancer tissues, even at the early stage of the disease. In addition, circCHD9 and circC1orf116 were found highly expressed, while circEPB41L2, circBNC2, and circSOX13 were strongly downregulated in such material. These last three circRNAs could be further seen downregulated in NSCLC samples by RT-qPCR validating previous nCounter results.

Circinteractome was used to further elucidate possible miRNA targets of aforementioned circRNAs.

Out of 28 predicted miRNAs for circFUT8, hsa-mir-186, and hsa-miR-1305 were the only ones presenting more than one potential binding site. Hsa-miR-186 was described downregulated in NSCLC, acting as an inhibitor of cancer proliferation, progression, and metastasis [40,41], whereas hsa-miR-1305 was not described in any type of cancer thus far. Another mechanism of action of circFUT8 in NSCLC has been described by Zhu et al. [42] in a recent publication, where this circRNA was shown to increase proliferation, invasion, and migration of NSCLC cells via miR-944/YESI axis.

For circCHD9, only one miRNA, hsa-miR-1229, was predicted. This miRNA was found upregulated in breast cancer activating β-Catenin/Wnt signaling [43]; however, nothing has been reported to lung cancer yet. No information regarding a possible connection between circC1orf116 and this malignancy was found either. Nonetheless, this circRNA has been described to promote cell proliferation, migration, and invasion in cervical cancer by binding to miR-518d-5p and miR-519-5p and further modulating BBX8 expression [44].

Interestingly, among the several predicted miRNAs for circEPB41L2, circBNC2, and circSOX13, hsa-miR-942 was a common target of cited circRNAs with 4, 2, and 1 binding sites, respectively (Fig. S9). This miRNA was previously described to be involved in colorectal and esophageal cancer progression activating the Wnt/β-catenin signaling pathway [45,46]. However, no evidence of its role in lung cancer has been found and would require further investigation.

Lastly, further machine learning analysis of generated data using RF, GBM, and KNN algorithms provided not only a signature able to correctly classify lung cancer samples from the control specimens, with an AUC of 0.985 (RF), 0.955 (GBM), and 0.993 (KNN) using an 8-circRNA signature, but also a 4-circRNA signature for early-stage lung cancer classification with comparable accuracy. These ML-based signatures included circEPB41L2, circSOX13, circBNC2, adding evidence of the potential of mentioned circRNAs as early-stage lung cancer biomarkers.

Since we did not perform microdissection of the tumor samples nor single cell analysis, we could not

verify whether presented signature-based circRNAs came from cancer cells or tumor microenvironment. Although this was out of our research scope since we mainly focused on the diagnostic potential of such signatures, we believe it could be of interest for future investigations. Also, most samples included in this study were lung adenocarcinomas, except for one squamous carcinoma and nine NSCLC samples with unknown histological subtype. Inclusion of different histologies in forthcoming validation studies are recommended to assess the specificity of the presented signatures. Finally, the work presented here was a proof-of-concept study and the main purpose was to demonstrate the feasibility of using nCounter for the study of circRNAs in lung cancer specimens. In consequence, the number of samples included was small and the abovementioned signature should be validated in a larger cohort.

5. Conclusions

In summary, we have developed a circRNA nCounter panel and workflow that can be used for multiplex detection of circRNA in FFPE lung cancer specimens. A cluster of differentially expressed circRNAs have been presented and further investigation is warranted to explore their potential as therapeutic targets. In addition, a 4 circRNA signature has been found through ML proving effective for early-stage lung cancer differentiation.

These findings pave the way to future biomarker investigations and validation of liquid biopsy signatures for lung cancer detection.

Acknowledgements

Conflict of interest

Chung-Ying Huang and Sarah Warren were full-time employees of NanoString Inc. at the time the study was performed. The rest of the authors declare no conflict of interests.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1002/1878-0261.13182.

Data accessibility

The data that support the findings of this study are available from the corresponding author (carlospedraz@icloud.com) upon reasonable request.

Author contributions

CPV, JWPB, RR, and MAMV conceptualized and designed the experiments. MI and AAH were responsible of patient recruitment and sample collection. CPV, AGC, EA, and SRM performed the experiments. SPG, NP, CPV, performed data analysis with contributions of AFH and MH. CPV, RR and MAMV wrote the main manuscript and prepared the figures. CYH and SW contributed to reagents and materials. AFH, MH TB, MF, and CAE provided editing, comments, and experimental guidance. All authors reviewed the manuscript.

References

1 Pedraz-Valdunciel C, Rosell R. Defining the landscape of circRNAs in non-small cell lung cancer and their potential as liquid biopsy biomarkers: a complete review including current methods. *Extracell Vesicles Circ Nucl Acids*. 2021;**2**:179–201. https://doi.org/10.20517/evcna.2020.07

2 Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;**32**:453–61. https://doi.org/10.1038/nbt.2890

3 Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci USA*. 1976;**73**:3852–6. https://doi.org/10.1073/pnas.73.11.3852

4 Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, et al. Scrambled exons. *Cell*. 1991;**64**:607–13. https://doi.org/10.1016/0092-8674(91)90244-S

5 Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;**495**:384–8. https://doi.org/10.1038/nature11993

6 Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;**495**:333–8. https://doi.org/10.1038/nature11928

7 Qu S, Liu Z, Yang X, Zhou J, Yu H, Zhang R, et al. The emerging functions and roles of circular RNAs in

cancer. *Cancer Lett.* 2018;**414**:301–9. https://doi.org/10.1016/j.canlet.2017.11.022

8 Zhou W-Y, Cai Z-R, Liu J, Wang D-S, Ju H-Q, Xu R-H. Circular RNA: metabolism, functions and interactions with proteins. *Mol Cancer.* 2020;**19**:172. https://doi.org/10.1186/s12943-020-01286-3

9 Ashwal-Fluss R, Meyer M, Pamudurti Nagarjuna R, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell.* 2014;**56**:55–66. https://doi.org/10.1016/j.molcel.2014.08.019

10 Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung L-H, et al. Exon circularization requires canonical splice signals. *Cell Rep.* 2015;**10**:103–11. https://doi.org/10.1016/j.celrep.2014.12.002

11 Yu C-Y, Kuo H-C. The emerging roles and functions of circular RNAs and their generation. *J Biomed Sci.* 2019;**26**:29. https://doi.org/10.1186/s12929-019-0523-z

12 Zhou R, Wu Y, Wang W, Su W, Liu Y, Wang Y, et al. Circular RNAs (circRNAs) in cancer. *Cancer Lett.* 2018;**425**:134–42. https://doi.org/10.1016/j.canlet.2018.03.035

13 Pedraz-Valdunciel C, Rosell R. Defining the landscape of circRNAs in non-small cell lung cancer and their potential as liquid biopsy biomarkers: a complete review including current methods. *Extracell Vesicles Circ Nucl Acids.* 2021;**2**:179–201.

14 Prokopec SD, Watson JD, Waggott DM, Smith AB, Wu AH, Okey AB, et al. Systematic evaluation of medium-throughput mRNA abundance platforms. *RNA.* 2013;**19**:51–62. https://doi.org/10.1261/rna.034710.112

15 San Segundo-Val I, Sanz-Lozano CS. Introduction to the gene expression analysis. In: IsidoroGarcía M, editor. Molecular genetics of asthma. New York, NY: Springer New York; 2016. p. 29-43.

16 Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Curr Protoc Mol Biol.* 2011;**94**:25B.10.21–25B.10.17. https://doi.org/10.1002/0471142727.mb25b10s94

17 Philippeos C, Hughes RD, Dhawan A, Mitry RR. Introduction to cell culture. In: Mitry RR, Hughes RD, editors. Human cell culture protocols. Totowa, NJ: Humana Press; 2012. p. 1–13.

18 Aguado C, Giménez-Capitán A, Román R, Rodríguez S, Jordana-Ariza N, Aguilar A, et al. RNA-based multiplexing assay for routine testing of fusion and splicing variants in cytological samples of NSCLC patients. *Diagnostics (Basel).* 2020;**11**:15. https://doi.org/10.3390/diagnostics11010015

19 Dahl M, Daugaard I, Andersen MS, Hansen TB, Grønbæk K, Kjems J, et al. Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab Invest.* 2018;**98**:1657–69. https://doi.org/10.1038/s41374-018-0108-6

20 Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The need for multi-omics biomarker signatures in precision medicine. *Int J Mol Sci.* 2019;**20**:4781. https://doi.org/10.3390/ijms20194781

21 Zhang H-D, Jiang L-H, Sun D-W, J-c H, Ji Z-L. CircRNA: a novel type of biomarker for cancer. *Breast Cancer.* 2018;**25**:1–7. https://doi.org/10.1007/s12282-017-0793-9

22 Zhang J, Zhang X, Li C, Yue L, Ding N, Riordan T, et al. Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells. *RNA Biol.* 2019;**16**:220–32. https://doi.org/10.1080/15476286.2019.1565284

23 Das Mahapatra K, Pasquali L, Søndergaard JN, Lapins J, Nemeth IB, Baltás E, et al. A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma. *Sci Rep.* 2020;**10**:3637. https://doi.org/10.1038/s41598-020-59660-6

24 Moldovan L-I, Hansen TB, Venø MT, Okholm TLH, Andersen TL, Hager H, et al. High-throughput RNA sequencing from paired lesional- and non-lesional skin reveals major alterations in the psoriasis circRNAome. *BMC Med Genomics.* 2019;**12**:174. https://doi.org/10.1186/s12920-019-0616-2

25 Ahmadov U, Bendikas MM, Ebbesen KK, Sehested AM, Kjems J, Broholm H, et al. Distinct circular RNA expression profiles in pediatric ependymomas. *Brain Pathol.* 2021;**31**:387–92. https://doi.org/10.1111/bpa.12922

26 Li S, Sun X, Miao S, Lu T, Wang Y, Liu J, et al. hsa_circ_0000729, a potential prognostic biomarker in lung adenocarcinoma. *Thoracic Cancer.* 2018;**9**:924–30. https://doi.org/10.1111/1759-7714.12761

27 Liu X-X, Yang Y-E, Liu X, Zhang M-Y, Li R, Yin Y-H, et al. A two-circular RNA signature as a noninvasive diagnostic biomarker for lung adenocarcinoma. *J Transl Med.* 2019;**17**:50. https://doi.org/10.1186/s12967-019-1800-z

28 Yan Y, Zhang R, Zhang X, Zhang A, Zhang Y, Bu X. RNA-Seq profiling of circular RNAs and potential function of hsa_circ_0002360 in human lung adenocarcinom. *Am J Transl Res.* 2019;**11**:160–75.

29 Zhang F, Zhao X, Dong H, Xu J. circRNA expression analysis in lung adenocarcinoma: comparison of paired fresh frozen and formalin-fixed paraffin-embedded specimens. *Biochem Biophys Res Comm.* 2018;**500**:738–43. https://doi.org/10.1016/j.bbrc.2018.04.145

30 Zhao J, Li L, Wang Q, Han H, Zhan Q, Xu M. CircRNA expression profile in early-stage lung adenocarcinoma patients. *Cell Physiol Biochem.* 2017;**44**:2138–46. https://doi.org/10.1159/000485953

31 Hossain ST, Malhotra A, Deutscher MP. How RNase R degrades structured RNA: role of the helicase activity and the S1 domain. *J Biol Chem.* 2016;**291**:7877–87. https://doi.org/10.1074/jbc.M116.717991

32 Westholm Jakub O, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, et al. Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep*. 2014;**9**:1966–80. https://doi.org/10.1016/j.celrep.2014.10.062

33 Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol*. 2014;**15**:409. https://doi.org/10.1186/s13059-014-0409-z

34 Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet*. 2013;**9**:e1003777. https://doi.org/10.1371/journal.pgen.1003777

35 Szabo L, Salzman J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat Rev Genet*. 2016;**17**:679–92. https://doi.org/10.1038/nrg.2016.114

36 Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun*. 2020;**11**:90. https://doi.org/10.1038/s41467-019-13840-9

37 Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, et al. Correlation of circular RNA abundance with proliferation – exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis and normal human tissues. *Sci Rep*. 2015;**5**:8057. https://doi.org/10.1038/srep08057

38 Yang J, Jia Y, Wang B, Yang S, Du K, Luo Y, et al. Circular RNA CHST15 Sponges miR-155-5p and miR-194-5p to promote the immune escape of lung cancer cells mediated by PD-L1. *Front Oncol*. 2021;**11**:595609. https://doi.org/10.3389/fonc.2021.595609

39 Hang D, Zhou J, Qin N, Zhou W, Ma H, Jin G, et al. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. *Cancer Med*. 2018;**7**:2783–91. https://doi.org/10.1002/cam4.1514

40 Huang T, She K, Peng G, Wang W, Huang J, Li J, et al. MicroRNA-186 suppresses cell proliferation and metastasis through targeting MAP3K2 in non-small cell lung cancer. *Int J Oncol*. 2016;**49**:1437–44. https://doi.org/10.3892/ijo.2016.3637

41 Ruan L, Chen J, Ruan L, Yang T, Wang P. MicroRNA-186 suppresses lung cancer progression by targeting SIRT6. *Cancer Biomark*. 2018;**21**:415–23. https://doi.org/10.3233/CBM-170650

42 Zhu H, Lu Q, Lu Q, Shen X, Yu L. Matrine regulates proliferation, apoptosis, cell cycle, migration, and invasion of non-small cell lung cancer cells through the circFUT8/miR-944/YES1 axis. *Cancer Manag Res*. 2021;**13**:3429–42. https://doi.org/10.2147/CMAR.S290966

43 Tan Z, Zheng H, Liu X, Zhang W, Zhu J, Wu G, et al. MicroRNA-1229 overexpression promotes cell proliferation and tumorigenicity and activates Wnt/β-catenin signaling in breast cancer. *Oncotarget*. 2016;**7**:24076–87. https://doi.org/10.18632/oncotarget.8119

44 Liu J, Wang D, Long Z, Li W. CircRNA8924 promotes cervical cancer cell proliferation, migration and invasion by competitively binding to MiR-518d-5p /519-5p family and modulating the expression of CBX8. *Cell Physiol Biochem*. 2018;**48**:173–84. https://doi.org/10.1159/000491716

45 Fasihi A, Soltani BM, Ranjbaran ZS, Bahonar S, Norouzi R, Nasiri S. Hsa-miR-942 fingerprint in colorectal cancer through Wnt signaling pathway. *Gene*. 2019;**712**:143958. https://doi.org/10.1016/j.gene.2019.143958

46 Ge C, Wu S, Wang W, Liu Z, Zhang J, Wang Z, et al. miR-942 promotes cancer stem cell-like traits in esophageal squamous cell carcinoma through activation of Wnt/β-catenin signalling pathway. *Oncotarget*. 2015;**6**:10964–77.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. nCouter probe design allows specific recogition of the circRNAs included in the panel.

Fig. S2. nCounter workflow for circRNA expression studies in FFPE lun tissues.

Fig. S3. Reproducibility experiment comparing the log2 of normalized counts by nCounter from three independent RNA samples derived from the PC9 cell line.

Fig. S4. Bioanalyzer profiles of faired fresh (left) and FFPE (right) PC9 cell line-derived RNA.

Fig. S5. nCounter analysis of FFPE PC9 cell line.

Fig. S6. Total RNA concentration assessment for circRNA analysis using the nCounter platform.

Fig. S7. Venn diagram showing circRNAs identified in all healthy cells (19) versus those only expressed in all lung cancer cell lines (1).

Fig. S8. Overall total number of raw counts in lung cells.

Fig. S9. Different miRNA binding sites of dysregulated circRNAs in early-stage lung cancer tissues.

Fig. S10. Diagram showing the tracking of those circRNAs of the circRNA nCounter panel not detected in assessed FFPE tissues.

Table S1. Diagnosis and associated pathologies of the control cohort.

Table S2. Characteristics of FFPE samples included in the study. Tumor and lymphocyte infiltration is indicated.

Table S3. circRNA and mRNA candidates included in the nCounter panel.