



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

Multiplex analysis of circRNAs in lung cancer using the nCounter technology



Carlos Pedraz Valdunciel

Biochemistry, Molecular Biology
and Biomedicine PhD Program
Universitat Autònoma de
Barcelona

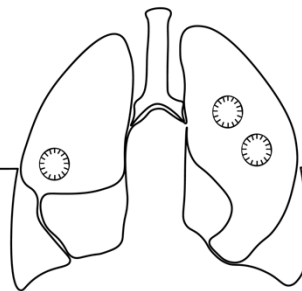
UAB

Universitat Autònoma de Barcelona

Multiplex analysis of circRNAs in lung cancer using the nCounter technology

"The science of today is the technology of tomorrow."
– **Edward Teller**

Carlos Pedraz Valdunciel



Multiplex analysis of circRNAs in lung cancer using the nCounter technology

By

Carlos Pedraz Valdunciel

A dissertation submitted in fulfillment of the requirements for the PhD degree
in Biochemistry, Molecular Biology and Biomedicine of the Universitat
Autònoma de Barcelona



Barcelona, September 2022



Universitat Autònoma
de Barcelona

The present doctoral thesis entitled “Multiplex analysis of circRNAs in lung cancer using the nCounter technology”, by Carlos Pedraz Valdunciel, was carried out at the German Trias i Pujol Health Institute (IGTP), under the supervision of Dr. Rafael Rosell from the Cancer Biology and Precision Medicine Program at the IGTP, and the tutoring of Dr. Anna Maria Bassols Teixidó, from the Department of Biochemistry and Molecular Biology at the Universitat Autònoma de Barcelona (UAB).

Rafael Rosell

Anna Maria Bassols Teixidó

Barcelona, September 2022

“What I love about science is that as you learn, you don’t really get answers. You just get better questions.” – **John Green**

Acknowledgements

La presente tesis doctoral es la expresión máxima de años de arduo esfuerzo dedicados al estudio de los circRNAs como biomarcadores de cáncer de pulmón, un trabajo complejo donde sin la ayuda y apoyo de cada una de las personas involucradas en este trabajo, esta tesis no hubiera podido llevarse a cabo. Es por eso, que me gustaría dedicarle unas palabras a cada una de esas personas que hemos hecho de éste, nuestro proyecto.

En primer lugar, quería mostrar mi gratitud y admiración al Dr. Rafael Rosell, mi mentor y director de tesis. Gracias por creer en mí, por darme la libertad necesaria para desarrollar y llevar a cabo mis proyectos, guiándome y dándome soporte en cada paso. Tu conocimiento, experiencia y consejos han servido de gran apoyo durante la duración de esta tesis doctoral. Por esto, te estaré siempre agradecido.

A mi tutora de tesis, Dr. Anna María Bassols, muchas gracias por tu apoyo, por ser tan cercana y por dar respuesta a todas las dudas y preguntas surgidas durante estos años.

A los miembros del comité de seguimiento anual, Dr. Jordi Codony, Dr. Jaume Ferrer y Dr. Manel Monreal: muchísimas gracias por vuestra ayuda y buenos consejos.

Gracias al Instituto Germans Trias i Pujol (IGTP) por acogerme con esta beca doctoral, especialmente al laboratorio del Dr. Rosell.

Una mención especial se merecen todos y cada uno de mis compañeros de Pangaea Oncology. Gracias por ayudarme y acogerme como un compañero más cuando más lo necesitaba. En tiempos difíciles de pandemia, hicisteis un huequito para que pudiera seguir con mi tesis, y eso es algo que os agradezco de corazón. Especialmente, me gustaría agradecer al Dr. Miguel Ángel Molina, no sólo su acogedora recepción en el laboratorio, sino su implicación personal en este proyecto, revisando todos los artículos y capítulos que componen esta tesis. A Ana Giménez, gracias por ser tan buena compañera y amiga, y por estar siempre dispuesta a echar una mano. En lo personal, gracias a Beatriz García, Jordi Bertrán y Cristina Aguado. Gracias por vuestros buenos consejos, por las risas y por brindarme vuestra amistad.

Del mismo modo, me gustaría agradecer a todos los oncólogos del Instituto Oncológico Dr. Rosell (IOR). Muchas gracias Dr. Rosell, Dr. Andrés Aguilar, Dr. María González, Dr. Alejandro Martínez por acercarme al mundo de la oncología clínica, por enseñarme tanto y por vuestra activa implicación en este proyecto. Siempre llevare conmigo las ponencias y los comités de pulmón con el Dr. Rosell y el Dr. Andrés Aguilar.

Gracias al resto de oncólogos colaboradores, Dr. Santiago Viteri y Dr. Carlos Cabrera por vuestro apoyo y contribución mediante la inclusión de muestras de pacientes en este estudio. Del mismo modo, me gustaría agradecer a todos los pacientes oncológicos y voluntarios que donaron material biológico para que este proyecto fuese posible.

No me gustaría olvidarme de la Union Europea, la cual ha financiado este proyecto mediante el Programa Horizon 2020, sin el cual el consorcio ELBA que dio cobijo a este programa de doctorado, no existiría.

Additionally, I would like to thank all the ELBA partners, particularly the University of Granada and the Champalimaud Foundation for hosting me during my two placements. Thanks to Dr. Tom Würdinger for coordinating such an amazing consortium, and Dr. Danijela Koppers-Lalic for been such an extraordinary human being. I know no one as empathetic and

helpful as you are. We will be forever grateful for all the great work you have done during all these years, in addition to your continuous support.

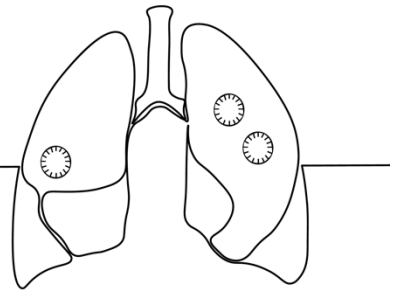
A big thank you also to all my ELBA colleagues, with whom I have been lucky enough to establish amazing collaborations. Special thanks to Jill, Martyna, and Stavros. Jill, thank you very much for being who you are, for our endless conversations and for being there every time I needed it. Thank you for all the moments of joy, I do really miss you so much. However, even far apart, our friendship grows stronger every day. Thank you very much, and don't forget we have a trip to California pending! Martyna, my little sister, I would have not been able to finish this PhD without your unconditional support. I felt so lucky I could count with you at the IGTP to support each other whenever needed. Thank you, you were, without a doubt, an important pillar of this PhD. Stavros, I need to give you all the credit for the bioinformatic analyses of this thesis. Not only you are an amazing professional, but also you have proved to be one of the kindest persons I know. It is such an honor to call you my friend. I wish all the very best for you.

A todos mis amigos de Salamanca, Estela, María José, Hugo, Helena, Diego, Silvia, Bea, Carmen, Vicky y Edu. Gracias por estar siempre ahí. Aunque en la distancia, siempre os he sentido cerca apoyándome y escuchándome. Del mismo modo, quiero incluir en este agradecimiento a mis amigos de Barcelona, especialmente a mi familia catalana Sergio, Patricia, Ash y Gregor, y a la internacional, Neil, Mark and Shanon. Thank you all for being such amazing and supportive friends.

A big thank you also to my Oxonian friends, specially Romisa and Amandine. I love you both, thank you for everything, and really hope we can continue with our adventures for many more years.

Finalmente, quería dedicar esta tesis a mi familia, mi madre María Ángeles, mis hermanos Anselmo y Laura, mi cuñada Emma, y mis sobrinos Diego y Paula. Hemos pasado por mucho en los últimos años, pero hemos demostrado ser una unidad fuerte y sólida. Gracias por apoyarme siempre en cada una de mis decisiones, por estar a mi lado en mis mejores y peores momentos. Sin vosotros no hubiera encontrado la inspiración ni la fuerza necesaria para poder concluir esta tesis. Os quiero inmensamente y os llevo conmigo siempre. Ahora sí, empieza un nuevo capítulo para todos.

Acronyms and abbreviations



5-CV	5-fold cross validation
AASDH	aminoadipate-semialdehyde dehydrogenase
ACACA	acetyl-CoA carboxylase alpha
ACP6	acid phosphatase 6, lysophosphatidic
ADAM22	ADAM metallopeptidase domain 22
ADAR1	adenosine deaminase acting on RNA 1
ADC	adenocarcinoma
AI	artificial intelligence
AKT	serine/threonine kinase
AKT	protein kinase B
ALK	anaplastic lymphoma kinase
AUC	area under the curve
B ₄ GALT2	beta-1,4-galactosyltransferase 2
BACH2	BTB domain and CNC homolog 2
BANP	BTG3 associated nuclear protein
BBX8	B-box domain protein 8
BNC2	basonuclin 2
C ₁ GALT1	core 1 synthase, glycoprotein-N-acetylgalactosamine -3-beta-galactosyltransferase 1
C ₁ ORF116	chromosome 1 open reading frame 116
CCNB1	cyclin B1
cDNA	complementary DNA
cfDNA	circulating cell-free DNA
cfRNA	cell-free RNA
CHD2	chromodomain helicase DNA binding protein 2
CHD9	chromodomain helicase DNA binding protein 9
CHST15	carbohydrate sulfotransferase 15
CI	confidence interval
circRNA	circular RNA
CLK1	CDC like kinase 1
COL11A1	collagen type XI alpha 1 chain
CORO1C	coronin 1C
Cq	quantification cycle
CSPP1	centrosome and spindle pole associated protein 1
Ct	cycle threshold
CT	computed tomography
CTC	circulating tumor cell
ctDNA	circulating tumor DNA
DE	differential expression
DENN1B	DENN domain containing 1B
DHCR24	24-dehydrocholesterol reductase

Acronyms and abbreviations

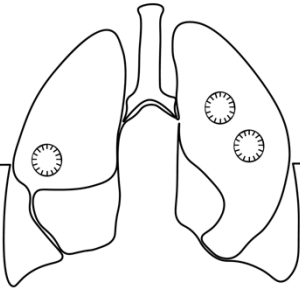
DHX9	DExH-box helicase 9
DNA	deoxyribonucleic acid
DUS2L	dihydrouridine synthase 2
EcirRNA	exonic-circRNA
ElcircRNA	exonic-intronic-circRNA
EGFR	epidermal growth factor receptor
EM	electron microscopy
EML4	echinoderm microtubule associated protein-like 4
EPB41L2	circular erythrocyte membrane protein band 4.1 Like 2
ERK	extracellular signal-regulated kinase
ETC	extra tree classifier
EV	extracellular vesicle
FAM13B	family with sequence similarity 13 member B
FARSA	phenylalanyl-tRNA synthetase subunit alpha
FC	fold change
FDA	US food and drug administration
FFPE	formalin-fixed paraffin embedded
FOXP1	forkhead box P1
FUT8	fucosyltransferase 8
G3BP1	G3BP stress granule assembly factor 1
GAPDH	gliceraldehído-3-fosfato deshidrogenasa
GBM	gradient-boosting classifier
(geo)mean	geometrical mean
GNA14	G protein subunit alpha 14
h	hour
HIPK3	homeodomain interacting protein kinase 3
IARC	international agency for research on cancer
IcircRNA	intronic circRNA
IQR plot	interquartil range
ITGAX	integrin subunit alpha X
KNN	k-nearest neighbors
LC	lung cancer
lncRNA	long non-coding RNA
LDCT	low dose computed tomography
LOOCV	leave-one-out cross validation
LPCAT1	lysophosphatidylcholine acyltransferase 1
LYPLAL1	lysophospholipase like 1
M-MLV	moloney murine leukemia virus
m6A	N6-methyladenosine
MAPK	mitogen-activated protein kinase
MBL	muscleblind transcription factor
MDCT	multidetector computed tomography

MGA	MAX gene-associated protein
miRNA	microRNA
ML	machine learning
mRNA	messenger RNA
MRPL19	mitochondrial ribosomal protein L19
n	number
nanoFCM	nanoparticle flow cytometry
NC	normalized counts
NCs	negative probe counts
NEDD4L	NEDD4 like E3 ubiquitin protein ligase
NF90	interleukin enhancer binding factor 3 protein
NF110	
NGS	next-generation sequencing
NLST	national lung screening trial
NPV	negative predictive value
ns	not significant
NSCLC	non-small cell lung cancer
nt	nucleotide
NUP98	nucleoporin 98 and 96 precursor
NUPL2	nucleoporin 42
OS	overall survival
PBS	phosphate-buffered saline
PC	positive control
PCA	principal component analysis
PCR	polymerase chain reaction
PET	positron emission tomography
PFS	progression-free survival
PI3K	phosphoinositide 3-kinase
PIK3R1	phosphoinositide-3-Kinase regulatory subunit 1
PPV	positive predictive value
PSMC4	proteasome 26S subunit, ATPase 4
PTEN	phosphatase and tensin homolog
PTPRM	protein tyrosine phosphatase receptor type M
QBP	quaking binding protein
RANGAP1	ran GTPase activating protein 1
RBP	RNA binding protein
RDH11	retinol dehydrogenase 11
RF	random forest
RFE	recursive feature elimination
RHD	Rh blood group D antigen
RHOQ	Ras Homolog Family Member Q
RLE	relative log expression

Acronyms and abbreviations

RNA	ribonucleic acid
RNAseq	RNA sequencing
ROC	receiver operating characteristic
RPLPo	ribosomal protein lateral stalk subunit Po
RSR	relative survival rate
RT	room temperature
RT-qPCR	reverse transcription quantitative PCR
RUNX1	RUNX family transcription factor 1
SCLC	small-cell lung cancer
SD	standard deviation
SEC	size-exclusion chromatography
SEMA5A	semaphorin 5A
SF3	splicing factor 3
SLC8A1	solute carrier family 8 member A1
SMAD	suppressor of mothers against decapentaplegic
SMARCA5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5
SND1	staphylococcal nuclease and tudor domain containing 1
SNX25	sorting nexin 25
SOX13	SRY-box transcription factor 13
TASP1	taspase 1
TEM	transmission electron microscopy
TEP	tumor educated platelet
TMEM39B	transmembrane protein 39B
UBB	ubiquitin B
UBXN7	UBX domain protein 7
UC	ultracentrifugation
USP3	ubiquitin specific peptidase 3
UV	ultra violet
VRK1	VRK serine/threonine kinase 1
WNT	wingless-related integration site
YES1	YES proto-oncogene 1
ZCCHC6	zinc finger CCHC domain-containing protein 6
ZFR	zinc finger RNA binding protein

Table of contents



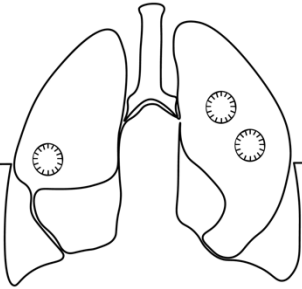
Summary / Resumen / Resum	1
<i>Summary</i>	3
<i>Resumen</i>	5
<i>Resum</i>	7
Chapter I: Introduction	9
1. <i>Lung cancer</i>	11
1.1. Lung cancer definition and classification	11
1.2. Global lung cancer trends, etiology, and detection of the disease	12
1.3. Current landscape of lung cancer early detection and research	13
2. <i>Circular RNAs</i>	14
2.1. CircRNA expression in humans	14
2.2. Biosynthesis and regulation of circRNAs.....	15
2.3. CircRNAs in NSCLC.....	16
2.4. Current landscape of circRNAs in liquid biopsies as NSCLC biomarkers	16
3. <i>The NanoString nCounter platform</i>	17
3.1. The nCounter technology	18
3.2. nCounter in the clinical setting	19
4. <i>Central motivation and objectives</i>	20
5. <i>Accepted manuscripts</i>	20
Chapter II: nCounter circRNA assay testing in lung cancer cell lines and FFPE specimens	21
<i>Abstract</i>	23
<i>Graphical Abstract</i>	23
1. <i>Introduction</i>	23
2. <i>Materials and methods</i>	25
2.1. Patient samples and cell lines	25
2.2. RNA extraction	25
2.3. Rnase R treatment	25
2.4. RT-qPCR and Sanger sequencing analysis	27
2.5. miRNA prediction and circRNA-miRNA network construction	28
2.6. NanoString nCounter panel design and sample processing.....	28
2.7. Data normalization and differential expression analysis	29
2.8. Machine Learning classification	29
3. <i>Results</i>	29
3.1. nCounter for circRNA detection in fresh NSCLC samples.....	29
3.2. nCounter for circRNA detection in FFPE NSCLC samples	31
3.3. circRNA expression in NSCLC fresh cell lines.....	31
3.4. circRNA expression in FFPE NSCLC vs. non-tumor tissue	31

Table of contents

3.5. CircRNA expression in early-stage NSCLC tissues	34
3.6. Univariate analysis related to lung cancer risk	36
3.7. Validation by RT-qPCR and Sanger Sequencing of circRNA junction sites	36
4. Discussion	38
5. Conclusions	41
6-Supporting information.....	41
Chapter III: Plasma-derived EV-circRNA analysis using the nCounter platform	51
<i>Abstract</i>	53
<i>Keywords</i>	53
1. Introduction.....	54
2. Materials and Methods	55
2.1. Patient samples.....	55
2.2. Plasma processing.....	55
2.3. Enrichment of EVs.....	56
2.4. Transmission electron microscopy (TEM)	56
2.5. Nano-flow cytometry measurements	56
2.6. RNA isolation and DNase treatment	56
2.7. RT-qPCR and Sanger sequencing analysis	57
2.8. nCounter processing	58
2.9. Differential expression analysis	58
2.10. Data pre-processing and normalization for signature development	58
2.11. Machine Learning (ML) for signature development	59
2.12. Univariate and multivariate analyses	59
3. Results	59
3.1. Enrichment of plasma EVs and workflow development for nCounter circRNA analysis.....	59
3.2. CircRNA expression in plasma EV samples.....	62
3.3. Development of a circRNA-signature associated with early-stage NSCLC	63
4. Discussion	68
5. Conclusions	70
6-Supporting information.....	70
Chapter IV: Plasma circRNA analysis for the early detection of lung cancer using nCounter	75
<i>Abstract</i>	77
<i>Keywords</i>	77
1. Introduction.....	78
2. Methods.....	79
2.1. Patient samples.....	79
2.2. Plasma processing and RNA isolation	80

2.3. nCounter processing	80
2.4. Differential expression analysis	80
2.5. Machine Learning (ML) for signature development	81
3. <i>Results</i>	81
3.1. CircRNA expression in plasma samples.....	81
3.2. Comparison of circRNA expression in EVs and total plasma.....	82
3.3. Differential expression analysis of circRNAs in NSCLC plasma samples.....	83
4. <i>Discussion</i>	86
5. <i>Conclusions</i>	88
6. <i>Supporting Information</i>	89
Chapter V: Discussion	91
Chapter VI: Future perspectives.....	99
Chapter VII: General conclusions.....	103
Bibliography	107
Publications	117
Annexes	121

Summary / Resumen / Resum



Summary

Lung cancer stands as one of the two most commonly diagnosed types of cancer, and the foremost cause of cancer-related death worldwide, primarily due to a late-stage diagnosis and lack of effective treatments for a significant number of patients. Although innovative single-agent therapies and their combination are constantly being tested in clinical trials, the five-year survival rate of late-stage lung cancer remains only at 5% (Cancer Research, UK). Consequently, advances in the early diagnosis of lung cancer are critical for improving the overall survival. Circular RNAs (circRNAs) are a re-discovered type of RNA showing a stable structure and tissue-specific expression. Accumulating evidence indicates how abnormal expression of some circRNAs can play a role in carcinogenesis and tumor progression, positioning these molecules as putative lung cancer biomarkers. In this context, many laboratories are currently investigating the clinical value of circRNAs found in liquid biopsies. Unfortunately, the lack of standardized methodologies for the study of circRNAs, together with some limitations inherent to liquid biopsies such as the low quantity and poor quality of tumor-derived material, is holding back their clinical implementation.

The NanoString nCounter FLEX system can quantify the expression levels of up to 800 RNAs extracted from formalin-fixed paraffin-embedded (FFPE) material, cell lysates or liquid biopsies. Among the different advantages of this platform, we can highlight its short turnaround time and the ability to work with very low quantity of highly degraded samples. However, to the best of our knowledge, there is no published evidence on the use of this platform for the study of circRNAs in lung cancer specimens. Therefore, the main objective of this doctoral thesis was to test the nCounter technology for the study of circRNA expression in lung cancer samples, including cell lines, FFPE samples and liquid biopsies. The results of this thesis have generated two manuscripts accepted for publication (Chapter 2 and 3) and a third manuscript in preparation (Chapter 4).

In Chapter 2, we present a proof-of-concept study of nCounter for the detection of circRNAs in lung cancer samples. First, we designed and validated a customized nCounter circRNA panel in lung cancer cell lines. Then, the panel was used to study circRNA expression in tissue biopsies. As a result, we found a cluster of differentially expressed circRNAs in early and late-stage lung cancer ($n=53$) vs. non-cancer controls ($n=16$). Also, machine learning (ML) analysis allowed us to develop two circRNA signatures to discriminate early and late-stage lung cancer biopsies from non-tumor samples with high accuracy.

In Chapter 3, we transitioned from solid specimens to liquid biopsies. In this second proof-of-concept study, we established and validated a protocol for the analysis of extracellular vesicle (EV)-circRNAs by nCounter. To this end, we tested different key points such as initial volume of plasma, EV purification method, number of cycles of pre-amplification, data normalization procedures or ML approaches. Then, we used the protocol to analyze the EV-circRNAs in plasma samples of lung cancer patients ($n=36$) and non-cancer controls ($n=30$). Again, we found a cluster of differentially expressed EV-circRNAs that could potentially be used as biomarkers of lung cancer. Also, ML allowed us to find a 10-circRNA signature that discriminated lung cancer samples from controls.

Purification of EVs may result challenging to implement in the clinical setting. Consequently, in Chapter 4, we tested nCounter in whole plasma and detected a higher number of circRNAs

compared to EV-enriched samples. Then, we analyzed plasma samples of early-stage lung cancer patients (n=49) and non-cancer controls (n=49), which included individuals with benign nodules (n=19/49). Subsequently, ML techniques generated a signature to discriminate early-stage lung cancer with an area under the ROC curve (AUC ROC) of 0.9.

In conclusion, this thesis has proved that nCounter can be used for the study of circRNAs in lung cancer samples, setting the ground for the development of clinically relevant circRNA assays. In particular, we have demonstrated the potential of some circRNAs as lung cancer biomarkers and developed preliminary circRNA signatures that discriminate early-stage lung cancer patients. Validation of such signatures in bigger cohorts is warranted. If successful, circRNA signatures, particularly in plasma, could be incorporated in the clinical setting for early detection of lung cancer.

Resumen

El cáncer de pulmón es uno de los dos tipos de cáncer más comúnmente diagnosticados y la principal causa de muerte relacionada con el cáncer en todo el mundo, principalmente debido a su diagnóstico en estadíos avanzados y a la falta de tratamientos efectivos para un número significativo de pacientes. Aunque numerosas terapias innovadoras basadas en un solo agente y su combinación son testadas constantemente en ensayos clínicos, la tasa de supervivencia a cinco años del cáncer de pulmón en estadíos avanzados sigue siendo tan solo del 5% (Cancer Research, Reino Unido). En consecuencia, avances en el diagnóstico precoz del cáncer de pulmón son muy necesarios para poder mejorar la supervivencia global de estos pacientes. Los ARN circulares (circRNA) son un tipo de ARN redescubierto recientemente que muestra una estructura estable y una expresión específica a cada tejido. Numerosos trabajos científicos muestran cómo la expresión anormal de algunos circRNAs puede desempeñar un papel importante en la carcinogénesis y la progresión del tumor, posicionando a estas moléculas como posibles biomarcadores del cáncer de pulmón. En este contexto, muchos laboratorios están investigando actualmente el valor clínico de los circRNAs que se encuentran presentes en las biopsias líquidas. Desafortunadamente, la falta de un método estandarizado para el estudio de los circRNAs, además de las limitaciones inherentes a las biopsias líquidas, como la baja cantidad y mala calidad del material tumoral, está frenando su implementación clínica.

El sistema nCounter FLEX de NanoString puede cuantificar los niveles de expresión de hasta 800 moléculas diferentes de ARN extraídas de material fijado en formalina e incluido en parafina (FFPE), lisados celulares o biopsias líquidas. Entre las diferentes ventajas de esta plataforma, podemos destacar el breve periodo de tiempo de espera para la obtención de resultados, así como su capacidad para trabajar con muy poca cantidad de muestras altamente degradadas.

Sin embargo, hasta donde nosotros sabemos, no hay evidencia publicada sobre el uso de esta plataforma para el estudio de los circRNAs en muestras de cáncer de pulmón. Por ello, el principal objetivo de esta tesis doctoral ha sido testar la tecnología nCounter para el estudio de la expresión de los circRNAs en muestras de cáncer de pulmón, incluyendo líneas celulares, muestras FFPE y biopsias líquidas. Como resultado de esta tesis, dos manuscritos han sido generados y aceptados para publicación (capítulo 2 y 3), así como un tercer manuscrito que se encuentra en preparación (capítulo 4).

En el capítulo 2, presentamos un estudio de prueba de concepto donde hacemos uso del nCounter para la detección de circRNAs en muestras de cáncer de pulmón. En primer lugar, diseñamos y validamos en líneas celulares de cáncer de pulmón un panel personalizado de circRNAs para nCounter. A continuación, utilizamos nuestro panel para estudiar la expresión de circRNAs en biopsias de tejido. Como resultado, encontramos un grupo de circRNAs diferencialmente expresado en cáncer de pulmón, tanto en etapas tempranas como tardías de la enfermedad ($n = 53$), frente a controles sin cáncer ($n = 16$). Además, análisis de aprendizaje automático (en inglés, *Machine Learning*, ML) nos permitió desarrollar dos firmas de circRNAs para discriminar con alta precisión, biopsias de cáncer de pulmón (etapas temprana y tardía) de muestras no tumorales.

En el capítulo 3, hicimos la transición de muestras sólidas a biopsias líquidas. En este segundo estudio piloto, establecimos y validamos un protocolo para el análisis de los circRNAs

contenidos en las vesículas extracelulares (EVs) por nCounter. Para ello, probamos diferentes puntos clave como el volumen inicial de plasma, el método de purificación EVs, el número de ciclos de pre-amplificación, los procedimientos de normalización de datos o diferentes métodos de ML. A continuación, usamos nuestro protocolo para analizar los EV-circRNAs en muestras de plasma de pacientes con cáncer de pulmón (n = 36) y donantes sin cáncer (n = 30). Nuevamente, encontramos un grupo de EV-circRNAs expresados diferencialmente en muestras de pacientes con cáncer que podrían usarse potencialmente como biomarcadores de cáncer de pulmón. Además, ML nos permitió encontrar una firma de 10 circRNAs que permitió diferenciar las muestras de cáncer de pulmón de los controles.

La purificación de los EVs es un proceso que puede resultar difícil de implementar en el entorno clínico. En consecuencia, en el capítulo 4, probamos nCounter en muestras de RNA extraídas directamente de plasma, detectando una mayor cantidad de circRNAs en estas muestras en comparación con aquellas procesadas para la obtención de EVs. A continuación, analizamos muestras de plasma de pacientes con cáncer de pulmón en estadio temprano (n=49) y de controles sin cáncer (n=49), los cuales incluían también individuos con nódulos benignos (n=19/49). Posteriormente, a través de ML, pudimos generar una firma para discriminar el cáncer de pulmón en etapa temprana con un área bajo la curva (AUC) ROC de 0,9.

En conclusión, esta tesis ha demostrado que la plataforma nCounter puede ser utilizada para el estudio de circRNAs en muestras de cáncer de pulmón, sentando las bases para el desarrollo de ensayos de circRNAs clínicamente relevantes. En particular, hemos demostrado el potencial de algunos circRNAs como biomarcadores de cáncer de pulmón, desarrollando varias firmas preliminares de circRNA que discriminan a los pacientes con cáncer de pulmón en etapa temprana. La validación de tales firmas en cohortes más grandes está garantizada. En caso de éxito, las firmas de circRNA, particularmente en plasma, podrían incorporarse en el entorno clínico para la detección temprana del cáncer de pulmón.

Resum

El càncer de pulmó és un dels dos tipus de càncer més comunament diagnosticats i la principal causa de mort relacionada amb el càncer a tot el món, principalment degut al seu diagnòstic en estadis tardans i a la manca de tractaments efectius per a un nombre significatiu de pacients. Encara que es porten a terme constantment assajos amb nombroses teràpies innovadores basades amb un sol agent o en combinacions, la taxa de supervivència a cinc anys en estadis avançats continua sent sol del 5% (Cancer Research, el Regne Unit). En conseqüència, avanços en el diagnòstic precoç del càncer de pulmó són essencials per a poder millorar la supervivència global d'aquests pacients. Els ARN circulars (circRNA) són un tipus d'ARN re-descobert recentment que mostra una estructura estable i una expressió teixit específica. L'evidència acumulada indica com l'expressió anormal d'alguns circRNA pot exercir un paper en la carcinogènesis i la progressió del tumor, posicionant a aquestes molècules com a possibles biomarcadors del càncer de pulmó. En aquest context, molts laboratoris estan investigant actualment el valor clínic dels circRNAs que es troben en les biòpsies líquides. Desafortunadament, la falta d'una metodologia estandarditzada per a l'estudi de circRNAs, a més de les limitacions inherents a les biòpsies líquides, com la baixa quantitat i mala qualitat del material tumoral, està frenant la seva implementació clínica.

El sistema nCounter FLEX de NanoString pot quantificar els nivells d'expressió de fins a 800 molècules de diferents ARNs extretes de material fixat en formalina i inclòs en parafina (FFPE), lisats cel·lulars o biòpsies líquides. Entre els diferents avantatges d'aquesta plataforma, podem destacar l'obtenció de resultats en un curt període de temps, així com la seva capacitat per a treballar amb molt poca quantitat de mostra altament degradada. No obstant això, fins on nosaltres sabem, no hi ha evidència publicada sobre l'ús d'aquesta plataforma per a l'estudi dels circRNAs en mostres de càncer de pulmó. Per això, el principal objectiu d'aquesta tesi doctoral ha estat testar la tecnologia nCounter per a l'estudi de l'expressió dels circRNAs en mostres de càncer de pulmó, incloent-hi línies cel·lulars, mostres FFPE i biòpsies líquides. Els resultats d'aquesta tesi han generat dos manuscrits acceptats per a publicació (capítol 2 i 3), així com un tercer que es troba en preparació (capítol 4).

En el capítol 2, presentem un estudi de prova de concepte de nCounter per a la detecció de circRNA en mostres de càncer de pulmó. En primer lloc, vam dissenyar i validar en línies cel·lulars de càncer de pulmó un panell personalitzat de circRNAs per a nCounter. A continuació, el panell es va utilitzar per a estudiar l'expressió de circRNAs en biòpsies de teixit. Com a resultat, hem trobat un grup de circRNAs diferencialment expressats en càncer de pulmó en etapes primerenques i tardanes de la malaltia ($n = 53$), enfront de controls sense càncer ($n = 16$). A més, l'anàlisi d'aprenentatge automàtic (Machine Learning, ML) ens va permetre desenvolupar dues signatures de circRNA per a discriminar biòpsies de càncer de pulmó, d'estadis primerencs i tardans, de mostres no tumorals amb una alta precisió.

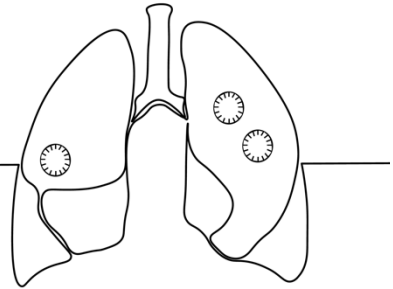
En el capítol 3, vam fer la transició de mostres sòlides a biòpsies líquides. En aquest segon estudi pilot, vam establir i validar un protocol per a l'anàlisi dels circRNAs continguts en les vesícules extracel·lulars (EVs) per nCounter. Amb aquest propòsit, vàrem provar diferents punts clau com el volum inicial de plasma, el mètode de purificació d'EVs, el nombre de cicles de pre-amplificació, els procediments de normalització de dades o diferents mètodes de ML. A continuació, vam utilitzar el nostre protocol per analitzar els EV-circRNAs en

mostres de plasma de pacients amb càncer de pulmó (n = 36) i donants sense càncer (n = 30). Novament, vam trobar un grup d'EV-circRNAs expressats diferencialment en mostres de pacients amb càncer que podrien usar-se potencialment com biomarcadors de càncer de pulmó. A més, ML ens va permetre trobar una signatura de 10 circRNAs que va permetre diferenciar les mostres de càncer de pulmó dels controls.

La purificació dels EVs és un procés que pot resultar difícil d'implementar en l'entorn clínic. En conseqüència, en el capítol 4, vam analitzar nCounter en mostres d'RNA extretes directament de plasma, detectant una major quantitat de circRNAs en aquestes mostres en comparació amb aquelles enriquides en EV. A continuació, vam analitzar mostres de plasma de pacients amb càncer de pulmó en estadi primerenc (n=49) i de controls sense càncer (n=49), els quals incloïen també individus amb nòduls benignes (n=19/49). Posteriorment, a través de ML, vam poder generar una signatura per discriminar el càncer de pulmó en etapa primerenca amb un àrea sota la corba (AUC) ROC de 0,9.

En conclusió, aquesta tesi ha demostrat que nCounter es pot utilitzar per a l'estudi de circRNAs en mostres de càncer de pulmó, establint les bases per al desenvolupament d'assajos de circRNA clínicament rellevants. En particular, hem demostrat el potencial d'alguns circRNAs com biomarcadors de càncer de pulmó, hem desenvolupat diverses signatures preliminars de circRNA que discriminen els pacients amb càncer de pulmó en estadis primerencs. Es necessari la validació d'aquestes signatures en cohorts més grans. En cas d'èxit, les signatures de circRNA, particularment en plasma, podrien incorporar-se en l'entorn clínic per a la detecció precoç del càncer de pulmó.

Chapter I: Introduction



Some parts of this chapter have been adapted from the following publication:

- 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 Nucleic Acids* 2021;2:179-201. <http://dx.doi.org/10.20517/evcna.2020.07>

1. Lung cancer

1.1. Lung cancer definition and classification

The process of tumorigenesis starts with the accumulation of genetic mutations within cells affecting different types of cell-cycle control genes, including tumor suppressor genes and oncogenes. This causes dysregulation of mitosis and inhibition of apoptosis, triggering the abnormal growth of the human cells, and leading to the formation of a lump [1]. When these tumor cells have the ability to spread to other parts of the human body, we then denote it malignant tumor or cancer. During the respiratory process, air goes into the lungs through the trachea. The trachea divides into bronchi, which enter the lungs and subsequently divide into smaller bronchioles that lead to the alveoli at the tip end. Lung cancers typically start in those cells lining the bronchi and parts of the lung such as the bronchioles or alveoli, where they crowd out normal cells thus, impeding the lungs to function correctly.

Depending on the type of cell the tumor originates from, we can classify lung cancer into small-cell lung cancer (SCLC) or non-small-cell lung cancer (NSCLC) (**Figure 1**). SCLC cells are flatter and smaller in appearance than those from NSCLC. In addition, SCLC tumors tend to grow and spread faster than NSCLC [2]. As a result, patients with this type of cancer tend to benefit from chemo- and radiotherapy. However, these patients show a higher recurrence than those with NSCLC, probably attributed to this elevated growing rate.

In the case of NSCLC, we can subcategorize the disease in adenocarcinoma, when the tumor starts in the goblet cells; squamous cell carcinoma, when the cancer commences in the squamous cells lining the large airways in the lungs; and large-cell carcinoma, when the tumor develops in other outer regions of the lungs. This subtype of NSCLC tends to grow and spread quickly, which can make it harder to treat. A few other subtypes of NSCLC, such as adenosquamous carcinoma and sarcomatoid carcinoma, are much less common [3].

Other types of lung tumors may include lung carcinoid tumors (accounting for less than 5% of lung tumors), or adenoid cystic carcinomas, lymphomas, and sarcomas, which are rarely found (**Figure 1**).

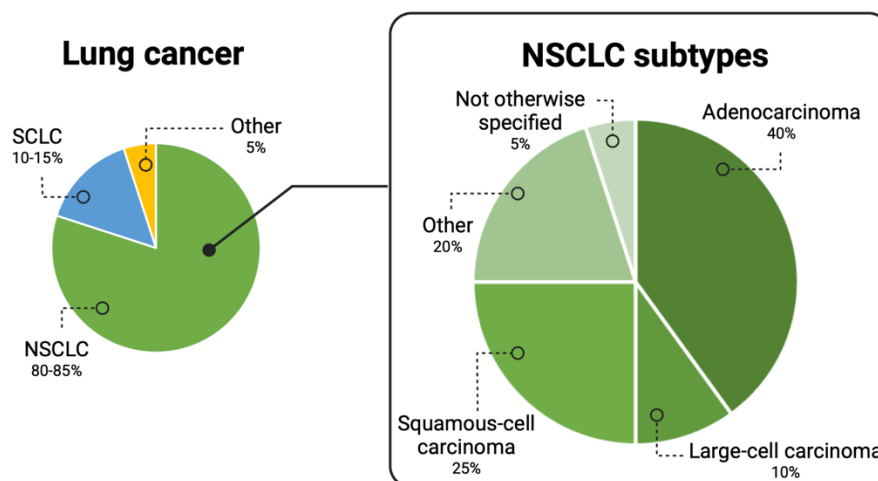


Figure 1. Pie charts indicating the categories of lung cancer and NSCLC subtypes. The area of the pie chart reflects the proportion of each type of lung cancer / NSCLC subtype. Percentages are indicated.

1.2. Global lung cancer trends, etiology, and detection of the disease.

With 11.4% of new diagnosed cases for 2020 – considering both sexes, lung cancer is the leading cause of cancer mortality in the world, surpassing breast cancer with 18% of the almost 10 million demises estimated for this year [4] (Figure 2).

From all new lung cancer diagnoses, 85% corresponds to NSCLC, while 10-15% is associated to SCLC [5] (Figure 1). The development of this disease is attributed to multileveled and elusive complex interactions between genetic liabilities, sex (incidence and mortality rates are roughly two times higher in men than in women [4]), environmental toxins, and imbalanced signaling processes. In this context, tobacco smoking remains the leading risk factor for lung cancer development. As a matter of fact, an escalation both in number of lung cancer cases and deaths is being observed in those developing regions where the smoking habit is increasing, such as China, Indonesia, Eastern Europe, and the Northern and Southern parts of Africa [6, 7]. In contrast, lung cancer incidence is declining in those countries that established smoking cessation and avoidance campaigns [6].

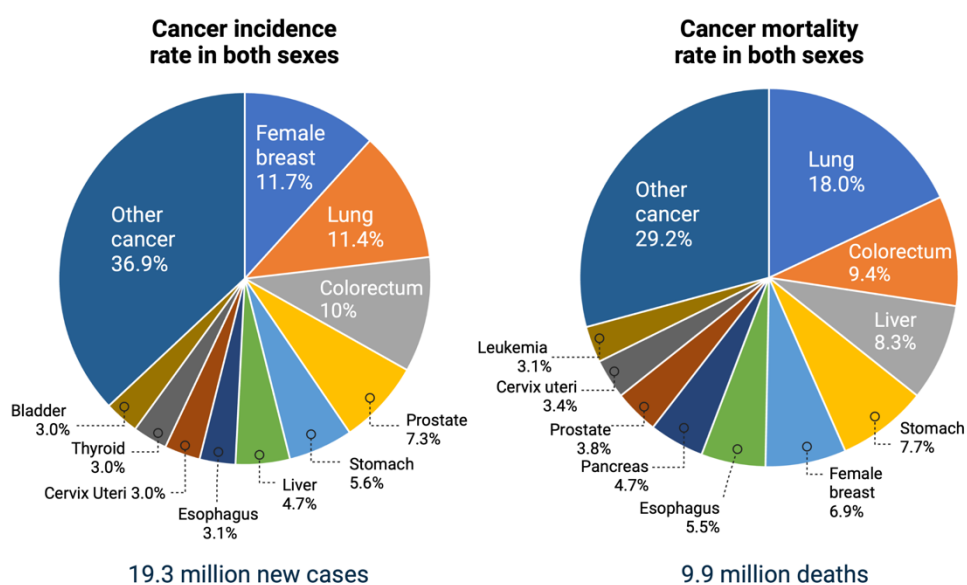


Figure 2. Pie charts representing cancer incidence and mortality rates for the Top 10 Most Common Cancers in 2020 for both sexes. The area of the chart reflects the proportion of the number of cases or deaths. Percentages are indicated. Source: GLOBOCAN 2020.

Detecting lung cancer at an early stage has the potential to drastically improve patient outcomes. At early stages (stages I and II), more options for treatment with curative intent are available, causing a positive shift in patient survival and quality of life. Instead, at later stages of the disease (stages IIIB and IV), fewer, if any, curative options are available. Despite all efforts for early detection, lung cancer is usually diagnosed at an advanced stage owing to inadequate screening programs, which translates into a poor 5-year relative survival rate (RSR) of 3–8% [8, 9].

Invasive tumor biopsy remains the primary method for the early diagnosis of lung cancer. Difficulty accessing the tumor, changes in the genetic composition or intra-tumor

heterogeneity are just some of the limitations related to this procedure that may hamper subsequent tumor characterization [10].

In that respect, liquid biopsies can be used to overcome, or complement, invasive tissue biopsies. Different body fluids can be utilized as liquid biopsies, including blood, urine, and saliva. Circulating molecules, such as DNA, RNA or proteins, can either be freely present within these liquids, or can be extracted from circulating tumor cells (CTCs) [11], extracellular vesicles (EVs) [12] or tumor-educated platelets (TEPs) [13]. They can be exploited to monitor therapy resistance and provide a more heterogeneous readout of the tumor burden [14]. This allows for identification of resistance mechanisms and can guide second-line therapy selection. Besides more effective personalized treatment regimens, a diagnostic stage-shift would significantly enhance patient survival. New tools for improved early lung cancer detection are therefore an unmet clinical need. In this regard, many liquid biopsy biomarkers are actively being investigated, holding potential for future clinical implementation [14].

1.3. Current landscape of lung cancer early detection and research

A longstanding challenge in public health is to develop and expand innovative solutions for earlier detection, screening, and diagnosis of cancers. One major approach in early cancer detection is screening of asymptomatic individuals, often at higher risk for cancer. For lung cancer, early detection screening by low dose computed tomography (LDCT) scan, has demonstrated to be effective among high-risk individuals (current and former heavy smokers), reducing lung cancer mortality in many randomized clinical trials [15-18]. However, the translation of this benefit to the general population has proven challenging, likely impeding the implementation of lung cancer screening programs as part of a global strategy to reduce the disease burden, at least in the near future.

The synergistic alliance between life science researchers and clinical investigators have thrived the development of new analytics and diagnostic methods and tools for the early identification of lung cancer, mainly through the analysis of liquid biopsies.

The continuous development in liquid biopsy biomarkers and genetic signatures has spurred greater capacity for identifying the malignancy, offering the promise of real-time detection and minimal invasiveness. In this regard, circulating tumor DNA (ctDNA) is one of the most investigated biomarkers, counting with many studies evaluating its potential clinical value as diagnostic tool for early lung cancer detection. Its presence has been validated in the plasma of lung cancer patients [19], enabling discrimination of these patients from non-cancer controls, including individuals with chronic respiratory inflammation [20]. However, the presence of ctDNA in the plasma of early stage (stage I) lung cancer patients is not as frequent, being represented by only 50% of the individuals from this cohort [19].

In addition to ctDNA, the screening of CTCs in blood plasma is being tested as a tool for early detection [21]. The value of these biomarkers has already been demonstrated as indicators of tumor progression in lung cancer patients [22]. However, CTCs are rather scarce in the blood of cancer patients, especially at early stage, hence, challenging the implementation of these molecules as a tool for early detection.

Lung cancer involves massive changes in RNA metabolism, both in the tumor and tumor microenvironment. Expression analysis of the RNA present in plasma can offer a valuable snapshot of the disease, providing information about the development and progression of the cancer [23], and guidance to better tailor a treatment plan [24]. RNA can be found either freely circulating in the blood stream (cfRNA) or inside EVs, CTCs or TEPs. EVs are multisized

vesicles (ranging from 30 nm to 100 μ m) that can be released by many types of cells, including those with tumor origin, being therefore an attractive source of lung cancer biomarkers [12]. As a result, many laboratories are exploiting their potential, investigating their clinical value for early detection [25]. Efforts on signature development for either lung cancer detection or monitoring of treatment response have mainly focused on expression of mRNA and miRNA. However, the number of studies focusing on other types of RNA (such as circular RNAs or long non-coding RNAs, lncRNAs) in lung cancer is very limited.

2. Circular RNAs

Circular RNAs (circRNAs) are a recently re-discovered type of RNA generated by coupling of the 5' and 3' ends in a non-canonical process known as backsplicing [26]. These circular structures lack a poly(A) tail, which confers them resistance to the exonuclease RNase R and turns most of them into robustly stable molecules when compared to linear mRNA.

While thousands of circRNAs have been described thanks to the technological burst of deep sequencing [27], only the function of only a fraction of them has been elucidated. Nevertheless, recent investigations have shown the role of some circRNAs as important players in lung cancer, positioning as plausible biomarkers for the early detection, and promising candidates for seeking therapeutic and prevention strategies towards this disease [28].

2.1. CircRNA expression in humans

Although circRNAs have been acknowledged for many years as abnormally spliced "scrambled" transcripts [29], only recently have they been re-defined as biologically active molecules with a significant role in human homeostasis, having a tissue-specific expression profile during the different stages of development [30].

More than 60% of human genes can express circRNAs [31]. However, their expression levels in tissue remain rather low, accounting for only 5%-10% of the canonical (linear) mRNA expression [32, 33].

CircRNAs are originated by an alternative process called "backsplicing", where the 5' splice donor can stick to the 3' splice acceptor of an upstream exon. This process results in forming a circular structure that can include one or multiple exonic/intronic regions, depending on the specific mechanism that was inferred during this non-canonical process [34].

They have arisen as key post-transcriptional regulators through different functions (**Figure 3**), with micro-RNA (miRNA) sponging being the most studied. During this process, the circRNA binds to the argonaute-miRNA complex, and either via miRNA degradation or inhibition of the miRNA-mRNA interaction, it triggers further mRNA expression at post-transcriptional level [35].

Recent studies have also revealed that circRNAs could associate with ribosomes and be translated into functional short peptides, in a cap-independent manner [36]. Alternatively, they can also associate with proteins acting as scaffolds for enzymatic reactions. In addition, the process of circRNA synthesis generates an imbalance of the canonical splicing; hence, the backsplicing process itself stands as a direct regulator of the circRNA precursor gene at the transcriptional level.

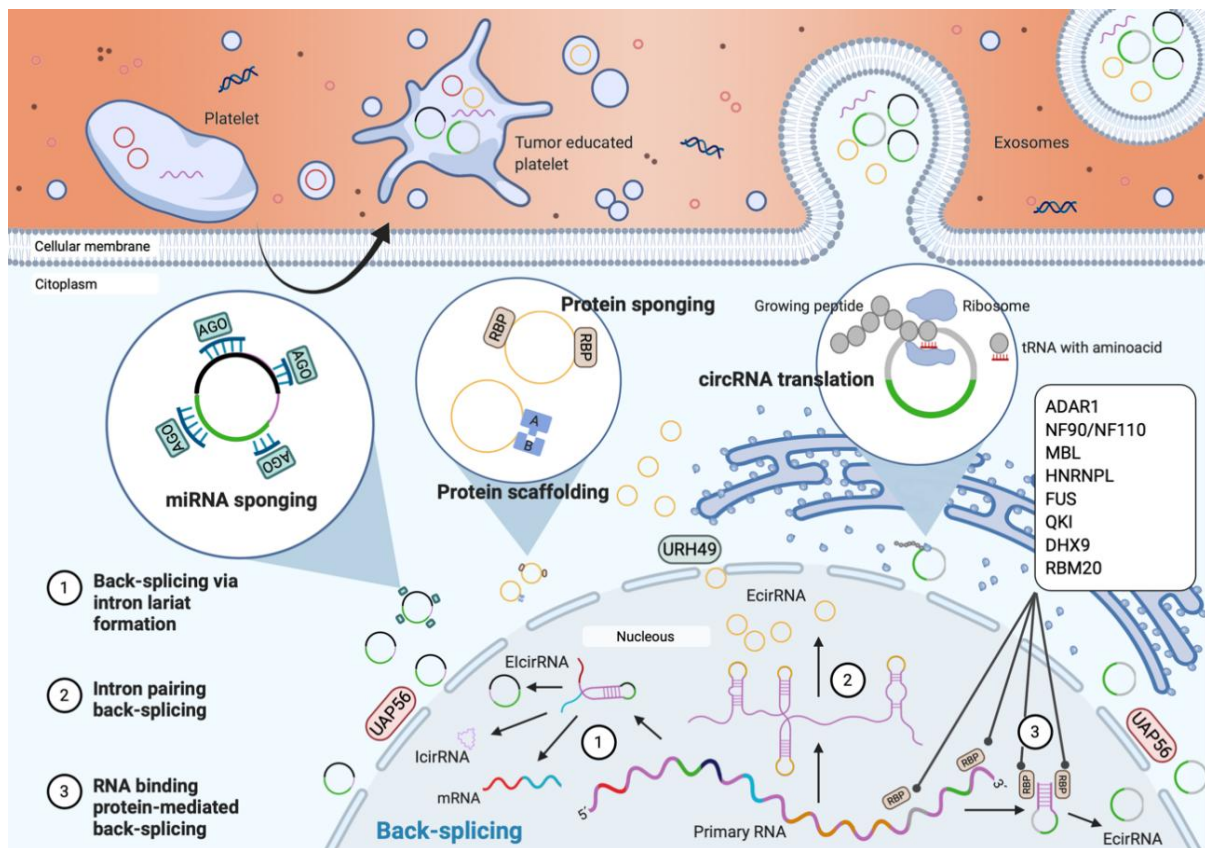


Figure 3. Biosynthesis and molecular functions of circRNAs. CircRNAs are generated by three different mechanisms of backsplicing (via lariat formation, intron pairing or RNA binding proteins). Resultant circRNAs can be formed by only exonic regions (EcircRNAs), intronic regions (IcircRNAs) or both (ElicirRNAs). circRNAs are exported into the cytoplasm in a size-mediated manner by URH49 and UAP56. Once in the cytoplasm, circRNAs will perform their functions including miRNA and protein sponging, protein scaffolding, or even translate into small functional peptides. CircRNAs are released into the blood stream inside exosomes mediating cellular communication. Most cellular types, including tumor cells, will secrete circRNA-containing EVs. Platelets can modify its content when in contact with the tumor, including their circRNA expression profile.

2.2. Biosynthesis and regulation of circRNAs

Different backsplicing mechanisms have been reported in the nucleus, including RNA binding protein (RBP)-mediated circularization, circRNA synthesis by intron pairing, or circularization by intron-lariat formation [34] (**Figure 3**). The first mechanism is normally executed by associating two adjacent exons and skipping the intronic region during an RBP-assisted circularization, resulting in an exonic-circRNA (EcircRNA). Numerous RBPs have been described in literature to regulate this mechanism, such is the case of the adenosine deaminase RNA specific-1 protein (ADAR1) [37], NF90/NF110 immune factors [38], muscleblind transcription factor (MBL) [39], heterogeneous nuclear ribonucleoprotein L [40], FUS protein [41], Quaking binding protein (QKI) [42], RNA helicase DHX9 [43], and the RNA-binding motif protein 20 [44].

Exon-intron circRNAs are the result of two or more exons circularized along with their corresponding introns via intron-lariat formation. Intron pairing backsplicing is usually the common process in conserved RNAs with high frequency of Alu repeats in flanking sequences. These Alu elements complement each other, promoting the hairpin formation and further backsplicing, creating mono-EcircRNAs as a result [45]. Intronic circRNAs are another type of such a class; however, the mechanism of generation of these molecules remains yet unclear.

After synthesis in the nucleus, circRNAs are exported to the cytoplasm. Recent studies have shown the active role of the UAP56/URH49 helicases in this size-mediated process. UAP56 is required to transfer molecules longer than 1300 nucleotides, while URH49 intervenes only in short transcript exporting [46]. Once in the cytoplasm, circRNAs accumulate and exert their function by regulating transcription, normally via sponging targeted miRNAs.

How circRNA gets degraded still remains unclear; however, recent investigation has shed light on this conundrum, unveiling some intriguing mechanisms that underpin circRNA decay. Hansen *et al.* [47] describe an Ago2-miR-671-mediated degradation of the circRNA CDR1as (aka ciRS-7). In another study by Park *et al.* [47, 48], a cleavage mechanism induced by RNase P/MRP was elucidated in N6-methyladenosine (m6A)-enriched circRNAs. More recently, a study by Liu *et al.* [49] demonstrated that some circRNAs tend to form intricate duplexes which makes them susceptible to degradation by RNase L upon viral infection.

A different mechanism was described by Fischer *et al.* [50] revealing an alternative structure-mediated circRNA regulation process that selectively degrades circRNAs based on 3'-UTR structure complexity via the UPF1/G3BP1 protein complex.

2.3. *CircRNAs in NSCLC*

The implication of circRNAs in cancer metabolism has been studied in recent years. Their contribution to mutant glycolysis (via transporter, enzyme, and/or transcription factor regulation), lipogenesis and lipolysis, glutaminolysis, and oxidative respiration has been widely demonstrated [51].

CircRNAs are becoming a new area of interest within cancer research, including NSCLC, where several authors are contributing by investigating the effect that dysregulated circRNA expression can have on the different cancer stages. Although their implication in NSCLC has not been as intensively investigated as other types of non-coding RNAs, circRNAs have been shown to play a significant role in tumorigenesis, tumor development, proliferation, migration, invasion, and sensitivity to NSCLC therapy [52]. In light of these findings, recent publications highlight the potential of some of these circular transcripts as plausible biomarkers to assess the disease status.

2.4. *Current landscape of circRNAs in liquid biopsies as NSCLC biomarkers*

Non-coding RNA-enriched exosomes are strategic players in different cancer stages, especially regarding malignant tumor metastasis [53]. The assessment of circRNA expression by RNAseq analysis in EVs was first reported by Li *et al.* [54], finding circRNAs enriched at least 2-fold in exosomes compared to producer cells. Although some authors

defend the theory that exosomal circRNA enrichment may be a mechanism of cellular circRNA clearance [55], few investigators have shown that these circRNA are directly involved in cellular communication, henceforth, acting as direct readouts of several human malignancies, including NSCLC [56].

As a result, circRNAs stand as important liquid biopsy-derived biomarkers, holding potential for NSCLC diagnosis and prediction of treatment response [57].

In a recent study, Chen *et al.* [58] performed high throughput sequencing of plasma-EV RNA cargo of lung adenocarcinoma patients, finding 182 circRNA dysregulated when compared to cancer-free donors, including 105 up-regulated and 78 downregulated. Four upregulated circRNAs were successfully validated by RT-qPCR (hsa_circ_0001492, hsa_circ_0001346, hsa_circ_0000690, and hsa_circ_0001439) [58]. Although authors elucidated the specific circRNA-miRNA-mRNA interaction, not much information about their biological impact was provided.

Fei *et al.* [59] also presented in a recent study a novel circRNA, hsa_circRNA_005661, that could be found enriched in plasma EVs from lung adenocarcinoma patients with lymph node metastasis, presenting it as a biomarker of this stage [59].

Not only plasma-EVs, but serum and whole plasma can serve as a good source of circRNAs. Xian *et al.* [60] studied the circRNA differential expression profile in serum EVs from NSCLC patients. As a result, 3 circRNAs stood out showing suitable biomarker potential (hsa_circ_0047921, hsa_circ_0007761, and hsa_circ_0056285) with the latter correlating with clinical stages and lymph node metastasis in all Chinese patients included in the study [60]. Exploring circRNA expression in plasma samples, Liu *et al.* found a two circRNA-based signature that could potentially be used to classify lung adenocarcinoma patients [61]. Hsa_circ_0005962 was found upregulated in lung cancer while hsa_circ_0086414 was found barely expressed when compared to controls. In addition, they observed that overexpression of hsa_circ_0005962 correlated to mutant EGFR expression. In vitro experiments suggested that this circRNA could be involved in cancer proliferation of EGFR+ lung cancer patients.

Alhasan *et al.* showed for the first time that platelets are enriched in circRNAs when compared to nucleated cells [62]. In addition, Preußner *et al.* demonstrated that platelets are not only a good source of circRNA, but also platelet-derived EVs are enriched in these biomolecules, representing yet another source of potential biomarkers that may be involved in different signaling pathways [63]. Platelets change their RNA profile when in contact with the tumor, enabling them to contribute to the systemic and local responses to tumor growth. As a result, TEP-RNA could be used as a potential biomarker for cancer diagnostics [64].

In this context, a recent publication demonstrates the diagnostic potential of platelet circRNA cargo as biomarker for lung cancer [65].

3. The NanoString nCounter platform

The NanoString nCounter technology has lately grown popularity among translational investigators, both in solid and liquid biopsies, to validate previously identified gene expression profiles as well as to discover novel signatures. In addition to transcriptional research, this multiplex, fluorescence-based hybridization method allows for other type of

assessments, namely analysis of copy number variation or genomic mutations, among others[66].

Advantages of this platform include high accuracy, short turnaround time, and the generation of reliable user-friendly data, especially when compared with RNAseq or next generation sequencing (NGS). For solid biopsies, an additional benefit of the nCounter platform is the fact that it can work with little amount of starting material, and does not require amplification, cDNA or library preparation, in comparison to the other abovementioned methods. In case of liquid biopsies, a pre-amplification step is required prior hybridization due to the minimal amount of material in this type of biosources. Although this process could entail additional bias to the protocol as previously reported [67], the low input kit developed by NanoString can overcome this issue as demonstrated by a recent study [68].

3.1. The nCounter technology

The technology of the nCounter platform is based on fluorescent molecular barcode chemistry developed at the Institute for Systems Biology (ISB) in Seattle, USA [69]. A combination of short, sequence-specific reporter and biotin-labeled capture probes configures a CodeSet representing the target RNA/DNA molecules for the assay. This CodeSet is mixed with the sample for overnight hybridization, followed by a purification step on a preparation station machine (nCounter Prep Station) where unbound nucleic acids and excess probes are washed out. Following that, the resulting immobilized molecules on the streptavidin coated cartridge are analyzed by digital counting of each of the distinctively color-coded barcodes in the nCounter Digital Analyzer, to determine relative expression of each target. Different normalization methods can then be applied to determine the expression of each of the analyzed RNA/DNA targets (**Figure 4**).

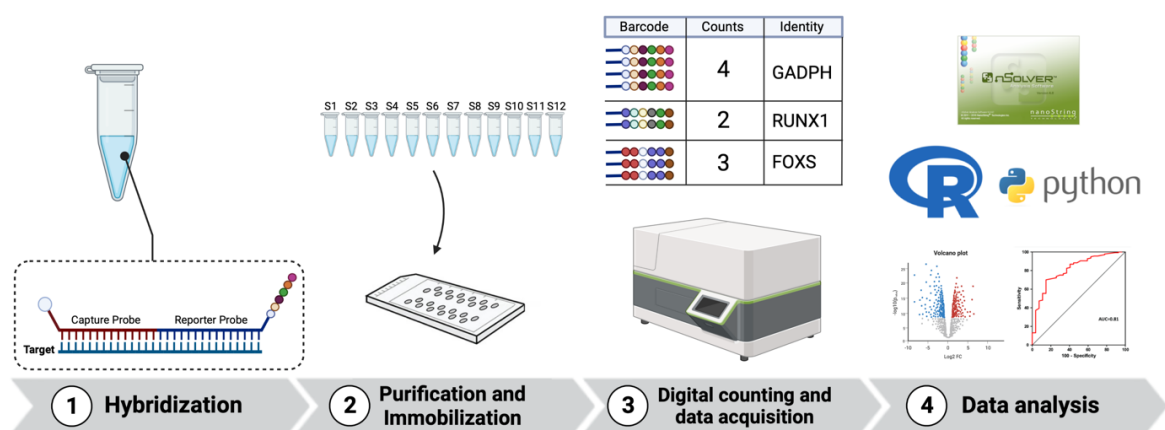


Figure 4. NanoString nCounter workflow. In a first step, capture and reporter probes configuring the CodeSet are hybridized overnight with the target of interest. Next, in a second phase, up to 12 samples are handled by the nCounter Prep Station where the nucleic acid-probe complexes are immobilized in the cartridge and purified from the excess of reagent. Then, the cartridge is placed into the nCounter Digital Analyzer to determine relative expression by digital counting of color-coded molecules. Finally, data are exported in form of RCC files and further analyzed by different software.

Taking this technology as a foundation, NanoString has developed two different types of assays: standard chemistry and TagSet chemistry. The former allows identification of more molecules (up to 800) through the direct binding of the pre-designed color-coded probes to

the sequence of interest (**Figure 5a**); the latter can analyze up to 228 molecules simultaneously and allows more flexibility to the user by giving the option to change the targets during the development of the assay (**Figure 5b**). In this case, the probes are unlabeled oligonucleotides designed by NanoString, but obtained from an external oligonucleotide synthesis provider. These probes (called probe A and probe B) bind both to the molecular target and the NanoString TagSet (which includes both capture and reporting tags) during the process of hybridization. In this case, probe A and B oligonucleotides are designed with 35-50 nt long target-specific regions, as well as a region complementary to the specific reporter and capture probes of the TagSet (**Figure 5b**).

Nowadays, we can find many commercially available panels for the study of immunology and oncology profiles using standard chemistry [66]. In addition, fully customized panels can be designed, using both standard and TagSet chemistry assays, for the identification of not only mRNA or DNA, but other forms of RNA such as miRNAs, lncRNAs or circRNAs. For the latter, a special divergent configuration of the probes is designed to target regions overlaying the backsplicing junctions, allowing the uniquely recognition of these circular transcripts (**Figure 5c**).

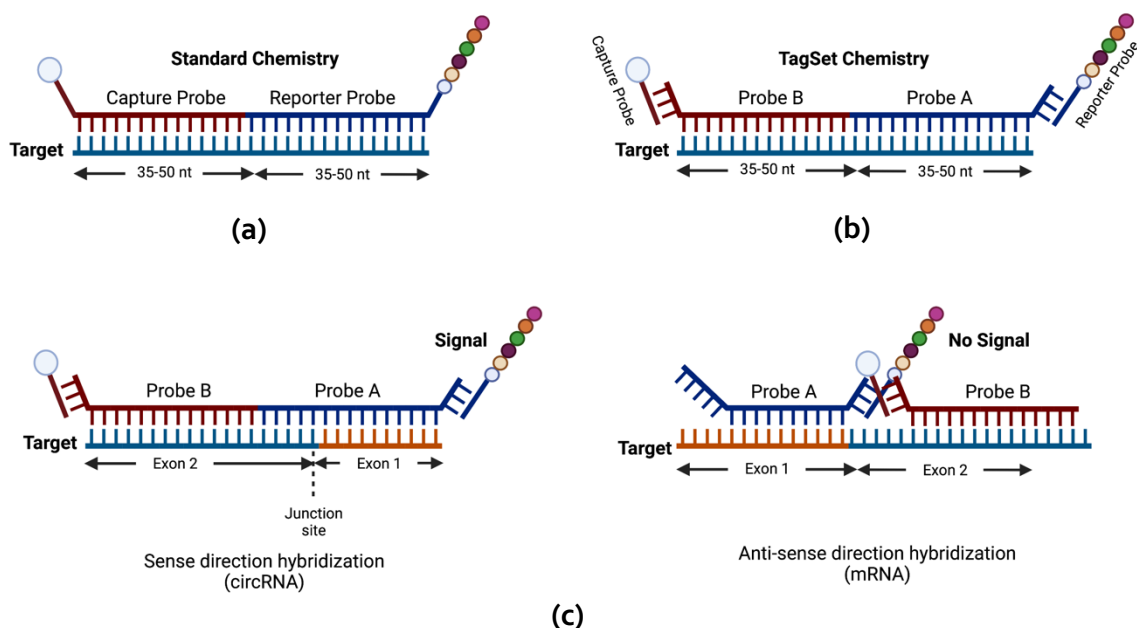


Figure 5. Standard and TagSet chemistry comparison including probe design for circRNA detection. **(a)** A color-coded reporter probe and biotin-labelled capture probe bind directly to target nucleic acid adjacently to each other in standard chemistry. **(b)** In TagSet chemistry, probes A and B bind to a 35-50 nt long regions of target nucleic acid as well as regions complementarity to reported and capture probes. **(c)** All probes are designed divergently spanning circRNA junction sites, so they can bind adjacently to one another and produce target-specific signal (left). In case of binding to a linear counterpart, probes will bind in an opposite orientation impeding the emission of the target-specific signal (right).

3.2. nCounter in the clinical setting

The robustness of the nCounter technology has facilitated the translation from bench to bedside of many molecular signatures, leading to the development of different assays for the prediction of tumor response [70], breast cancer profiling [71], identification of lung

cancer gene fusions [72], diagnosis of leukemia [73] and lymphoma [74], or the identification of genetic alterations and translocations associated with sarcoma gene fusions [75]. However, to this date, the Prosigna breast cancer profiling assay [71, 76], a test for the identification and stratification of breast cancer samples for the prediction of recurrence -or risk of recurrence, stands as the only nCounter panel granted with approval from the US Food and Drug Administration (FDA).

When it comes to liquid biopsies, most clinical research is restricted to the analysis of cfDNA [77], cfRNA [11, 78], CTCs [11, 67] and EV cargo [79, 80]. In addition, the amount of published studies based on nCounter for the early detection of lung cancer is scarce, and the signatures found are far from clinical implementation [81].

To the best of our knowledge, there is no scientific publication assessing nCounter for the detection of circRNAs in lung cancer, neither in solid tumors nor liquid biopsies.

4. Central motivation and objectives

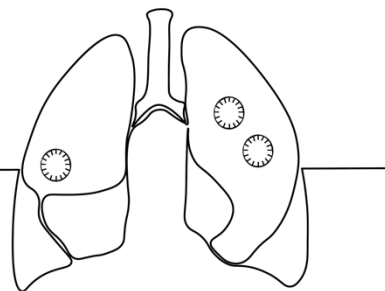
Considering the lack of published data on the use of the nCounter platform for the study of circRNAs as lung cancer biomarkers, both in solid and liquid biopsies, this PhD thesis aimed to validate this platform for such purpose, in particular:

- Chapter 2 contains a proof-of-concept study where we test a custom-made circRNA panel in lung cancer cell lines and FFPE tissues, using the nCounter FLEX platform. We also compared lung cancer and non-cancer tissues generating a circRNA signature of lung cancer via machine learning (ML).
- In Chapter 3, we adapted our procedures to generate a protocol for the study of EV-circRNAs from plasma of lung cancer patients. We also performed differential expression and ML analyses to discriminate lung cancer patients from non-cancer controls.
- We transition from EVs to whole plasma in Chapter 4, to analyze the circRNAs of this bio-source, generating a preliminary circRNA signature able to discriminate lung cancer patients from non-cancer controls, including individuals with benign nodules.
- Regarding Chapter 5, we go through a comprehensive discussion of presented results, providing an overview of the future perspectives.
- Finally, the conclusions of this PhD thesis can be found in Chapter 6.

5. Accepted manuscripts

- Pedraz-Valdunciel C, Giannoukakos S, Potie N, *et al.* Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens. *Mol Oncol.* 2022. doi: 10.1002/1878-0261.13182.
- Pedraz-Valdunciel C, Giannoukakos S.; Giménez-Capitán A, *et al.* Multiplex analysis of circRNAs from plasma extracellular vesicle-enriched samples for the detection of early-stage non-small cell lung cancer. *Pharmaceutics* 2022; 14, 2034. <https://doi.org/10.3390/pharmaceutics14102034>.

Chapter II: nCounter circRNA assay testing in lung cancer cell lines and FFPE specimens



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

Carlos Pedraz-Valdunciel, Stavros Giannoukakos, Nicolas Potie, Ana Giménez-Capitán, Chung-Ying Huang, Michael Hackenberg, Alberto Fernández-Hilario, Jill Bracht, Martyna Filipaska, Erika Aldeguer, Sonia Rodríguez, Trever G Bivona, Sarah Warren, Cristina Aguado, Masaoki Ito, Andrés Aguilar-Hernández, Miguel Ángel Molina-Vila, Rafael Rosell.

Molecular Oncology (*First published: 21 January 2022*)

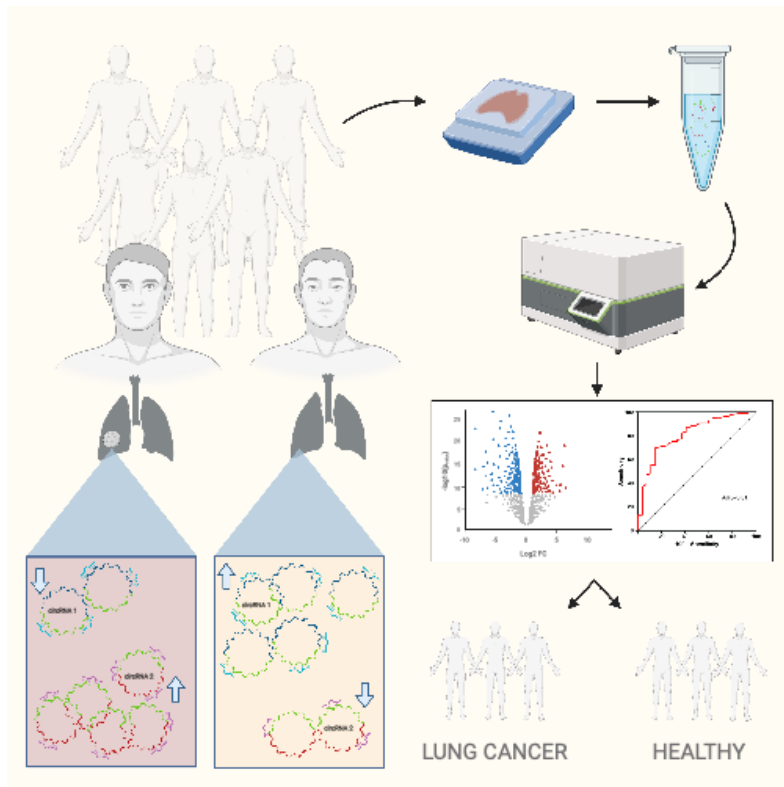
DOI: 10.1002/1878-0261.13182

Abstract

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 non-tumor 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.

Graphical Abstract

Aberrant circular RNA (circRNA) expression is present in lung cancer. Using nCounter with machine learning, we discovered two signatures able to discriminate FFPE lung cancer samples from controls even at early stage. Our results not only highlight the potential of circRNAs as lung cancer biomarkers but also introduce nCounter as a suitable platform for circRNA expression studies in these samples.



1. Introduction

circRNAs are a newly re-defined type of endogenous RNA molecules originated by a non-canonical process called “backsplicing”. 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 [82]. This particular assembly lacking a poly(A) tail makes them very stable and resistant to exonuclease-mediated degradation when compared to their linear counterparts [83].

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” [84], and their discovery in humans followed almost two decades later [85]. 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 [86, 87].

CircRNAs have been classified as non-coding 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 [88]. Other functions may include serving as protein decoys, scaffolds and/or recruiters [89], or regulating the canonical transcription by competing with the formation of linear cognates via backsplicing [39, 90]. 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 [91]. 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 [91]. 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 [92] 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. Whilst 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 [93]. 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 [94, 95].

The nCounter technology allows multiplex analysis of up to 800 transcripts by direct capture and counting of individual targets [96]. With a short turnaround time and minimal hands-on work, it provides results in less than 48h 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 non-tumorigenic region of the lung from a cancer patient. Individuals with different pathologies were also included to ensure 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.

A panel of 7 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[97] in RPMI-1640 or DMEM medium (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 [98, 99]. 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 FFPE 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 samples were denatured at 95°C for 30 seconds following addition of a master mix containing RNase R (or molecular grade water in the case of mock-treated samples), 10x RNase R buffer adjusted to the final volume, and molecular grade water. Samples were incubated 160 minutes at 40°C and kept at 4°C prior RNA quantification and subsequent nCounter hybridization.

Table 1. Clinicopathologic characteristics of enrolled patients (n = 69).

Clinicopathological characteristics	Lung cancer patients (n= 53)	Non-cancer controls (n = 16)
Gender – no. (%)		
Male	28 (52.8)	10 (62.5)
Female	25 (47.2)	6 (37.5)
Age – yr.		
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)	-

2.4. RT-qPCR and Sanger sequencing analysis

RT-qPCR and Sanger sequencing of circRNA junction sites were performed as previously described [98]. 10 µl of total RNA was converted into cDNA using the M-MLV reverse transcriptase enzyme and random hexamers (Invitrogen).

Table 2. Characteristics of the lung cell lines included in the study.

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

AD, adenocarcinoma; NE, normal epithelial; ATCC, American Type Culture Collection; UCSF, University California San Francisco; UTSW, University of Texas Southwestern

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 QuantStudio™ 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 PCR reactions 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 μ l. 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).

Table 3. Primer and probe design for circRNA validation by RT-qPCR.

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

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://mirccancer.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 (**Figure S1**). In addition, six linear reference genes (GAPDH, MRPL19, PSMC4, RPLP0, SF3 and UBB) and four mRNAs of FAM13B, HIPK3, MGA, and UBXN7 genes were included (**Table S2**).

Sample processing in the nCounter was performed as previously described [98] following NanoString's guidelines (**Figure S2**).

2.7. Data normalization and differential expression analysis

Raw count values were exported to Microsoft Excel (version 16.40, Microsoft) using nSolver Analysis Software (version 4.0.70, NanoString Technologies). 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) and R studio (version 1.3.1056). Technical variability correction, normalization and differential expression analysis was performed using the RUVSeq (version e1.24.0) and DESeq2 (version 1.30.0) packages (RUVseq-DESeq2). 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_2FC) 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_2FC on the x-axis and $-\log_{10}$ adjusted p-values on the y-axis.

2.8. Machine Learning classification

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 were 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 3 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, $p < 0.01$ (**Figure S3**).

Then, RNA from the same cell line was used in an experiment with RNase 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 (**Figure 1a**). As a result, 18 new transcripts that could not be detected in mock-treated samples were observed after RNase R treatment (**Figure 1b**). In addition, among the 34 transcripts identified in both 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 (**Figure 1c**). All mRNA controls, including the linear forms of FAM13B, HIPK3, MGA, and UBXN7, were found with reduced or null expression after treatment (**Figure 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 (**Figure 1e**).

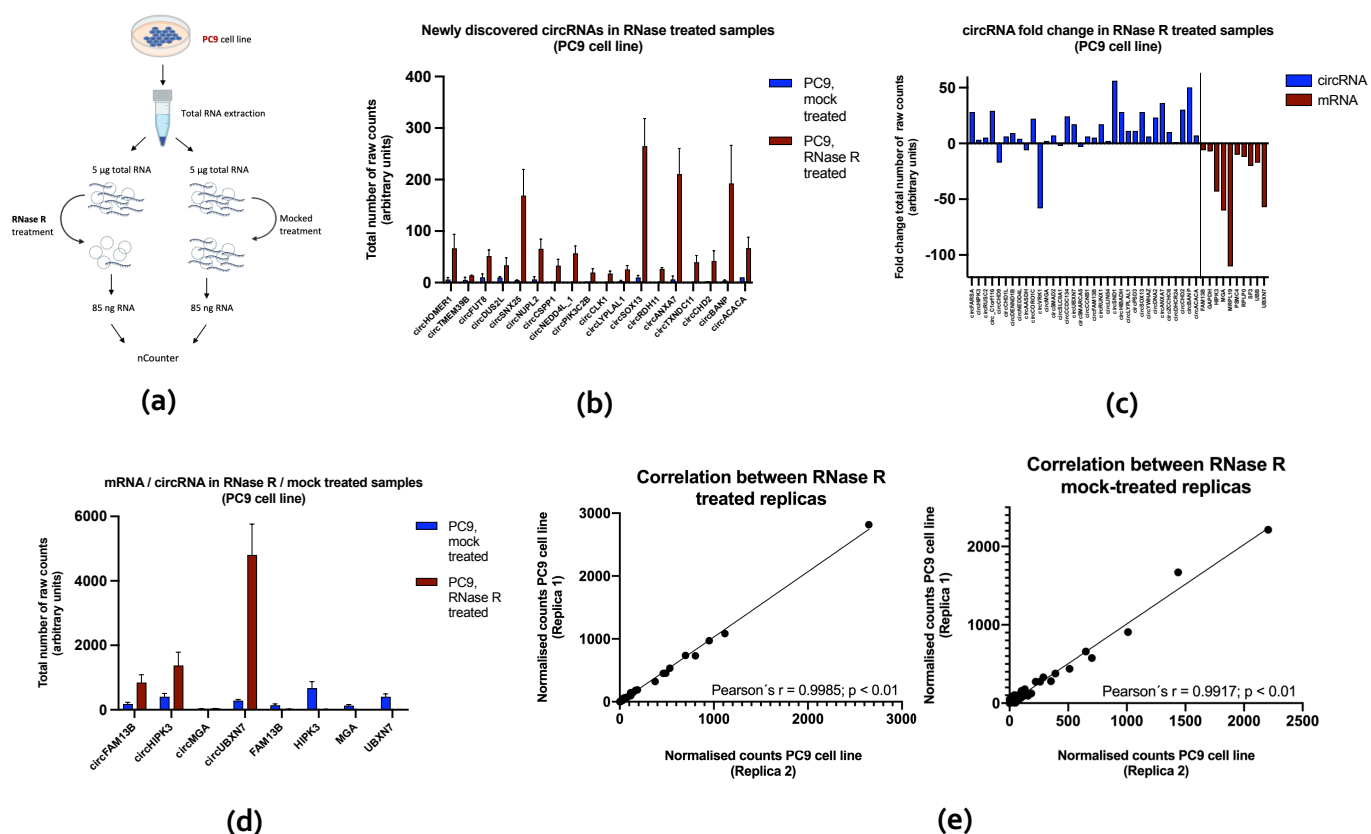


Figure 1. Analysis of circRNA from RNase R treated samples. **(a)** Workflow for circRNA enrichment with RNase R. **(b)** Representation of the newly discovered circRNAs after RNase R treatment. Bars indicate the mean of the replicas ($n=2$). Error bars indicate SD. **(c)** CircRNA/linear HK fold-change 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.

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 (**Figure S4**), a statistically significant correlation was found when comparing both types of input (**Figure 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 (**Figure S5b**).

Then, different concentrations of FFPE-derived RNA (between 250 ng 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 (**Figure 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 (**Figure 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 (**Figure 2a**). Overall, total raw counts were significantly higher in normal epithelial lung cell lines compared to cancer cell lines (**Figure S8**). Hierarchical clustering led to a separation of the KRAS cell lines and normal epithelial cell lines from the rest (**Figure 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 4 circRNAs that allowed for differentiation between the 7-lung cancer cells and normal epithelial cells. CircPIK3R1, circFARSA, and circCHST15 were found downregulated in the cancer cell lines, while circFUT8 was upregulated (**Figure 2c**).

3.4. circRNA expression in FFPE NSCLC vs. non-tumor 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 (**Figure 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 (Figure 3b).

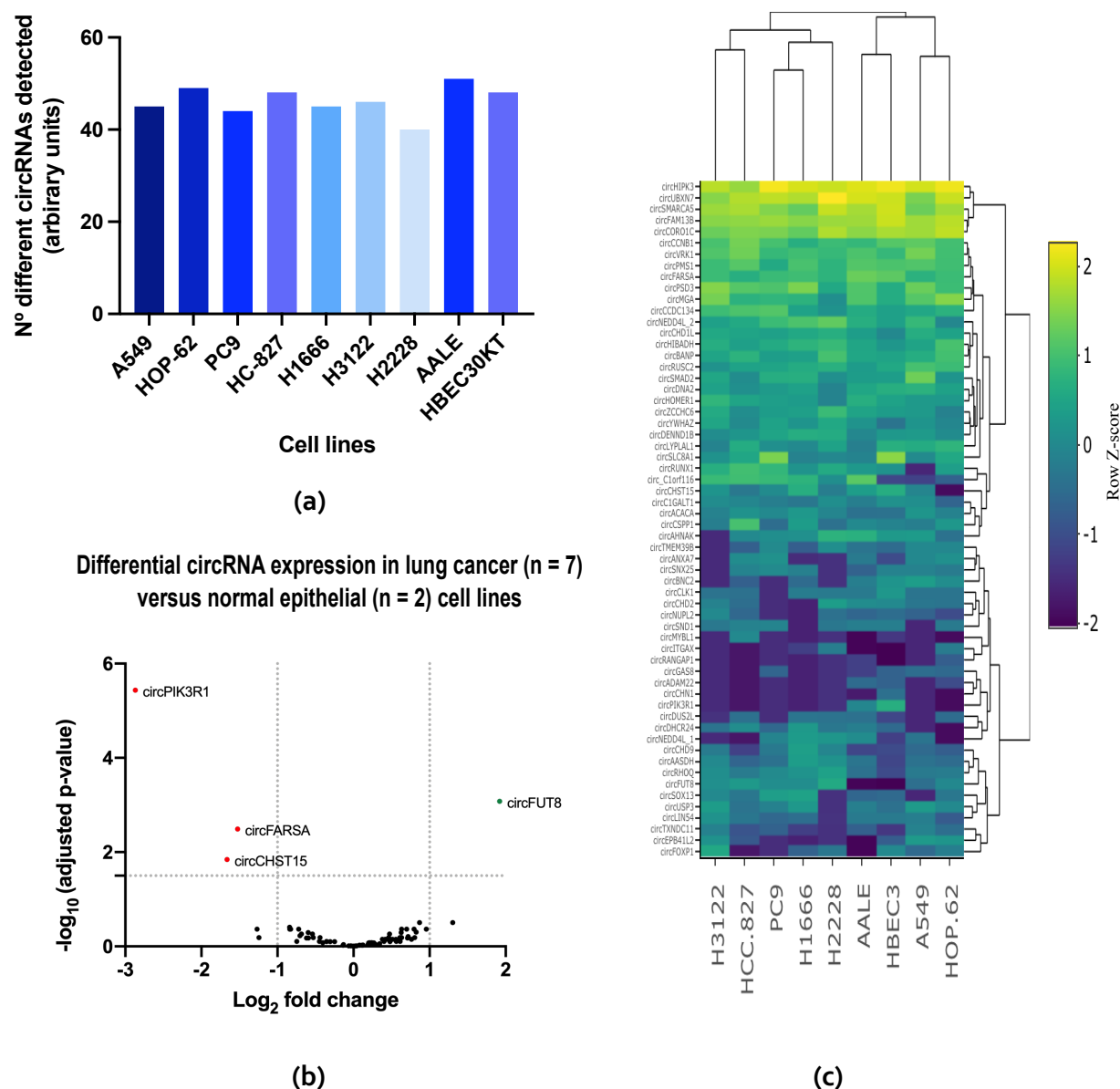
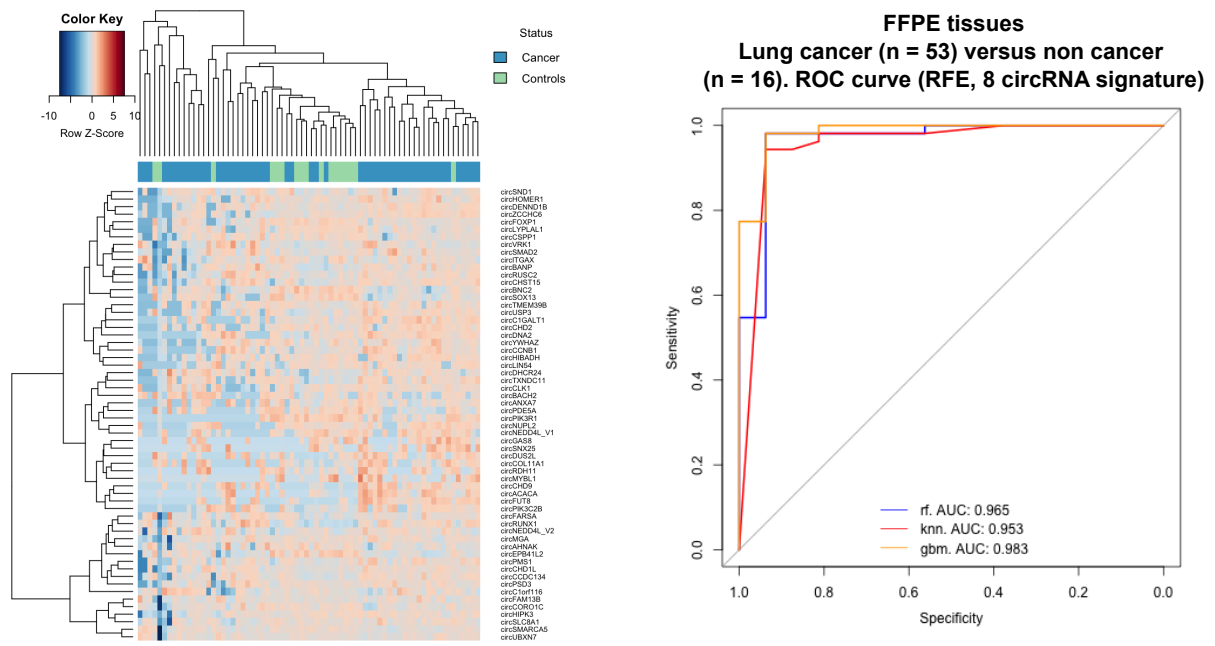


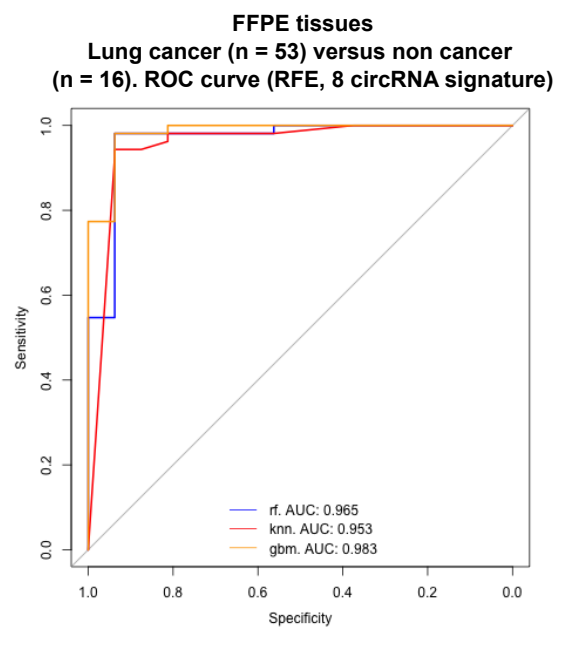
Figure 2. CircRNA analysis in lung cancer (A549, HOP-62, PC9, HCC-827, H1666, H3122 and H2228) and epithelial cells (AALE and HBEC30KT). **(a)** Bar plot representing total circRNAs detected in each of the cell lines. **(b)** Hierarchical clustering of cell lines based on circRNA expression. **(c)** Differential circRNA expression analysis of \log_2 -normalized counts between lung cancer and normal lung cells.

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).

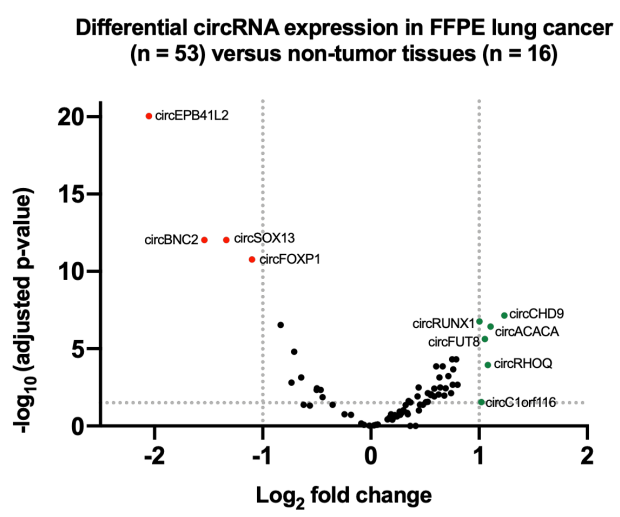


(a)



	Healthy	Cancer	Class.Err.
Healthy	13	3	0.1875
Cancer	1	52	0.0188

(c)



(b)

Figure 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 log₂ 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.

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 (**Figure 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.

3.5. 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 (**Figure 4a**). Only 6 out of these 39 transcripts were differentially expressed when compared with the control specimens (**Figure 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 6 circRNAs, a circRNA-miRNA network was built based on sequence-pairing prediction (**Figure 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 (**Figure S9**).

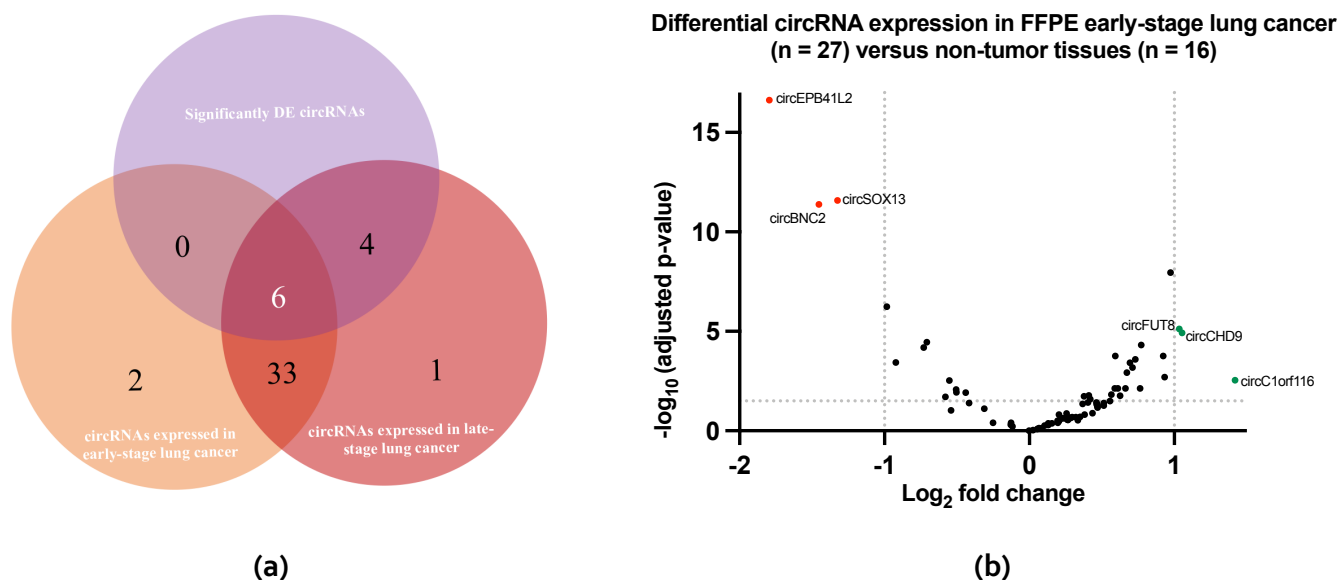


Figure 4. CircRNA expression in early-stage NSCLC samples. **(a)** Venn diagram displaying circRNAs identified in early- and late-stage samples, featuring those shared by both cohorts. DE circRNAs are indicated. **(b)** Differential expression analysis of log₂-normalized counts between the early-stage 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

Additional ML analysis was performed in early-stage lung cancer and control samples. RFE algorithm provided a signature that included 4 circRNAs (circEPB41L2, circSOX13, circBNC2,

circCORO1C) and provided a ROC AUC of 0.981, 0.918 with RF and KNN respectively (**Figure 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 4 circRNA included in the signature allowed a clear differentiation between both cohorts (**Figure 6b**).

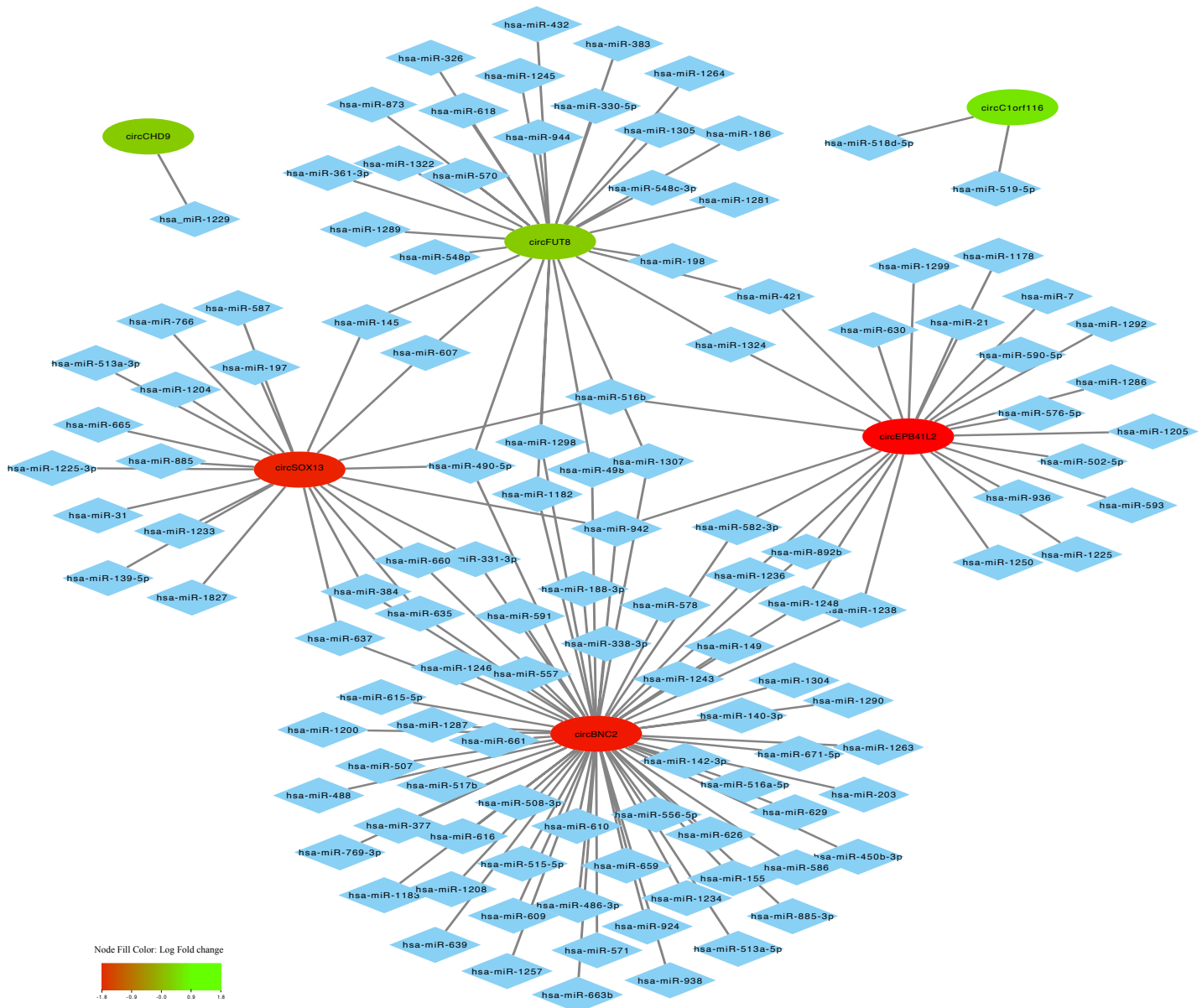


Figure 5. Mapping network showing predicted sequence-pairing circRNA-miRNA interaction of differentially expressed circRNA found in early-stage lung cancer tissues. CircRNAs are represented by elliptic nodes and colored based on their log fold change. Complementary binding miRNAs are represented by diamond shaped nodes.

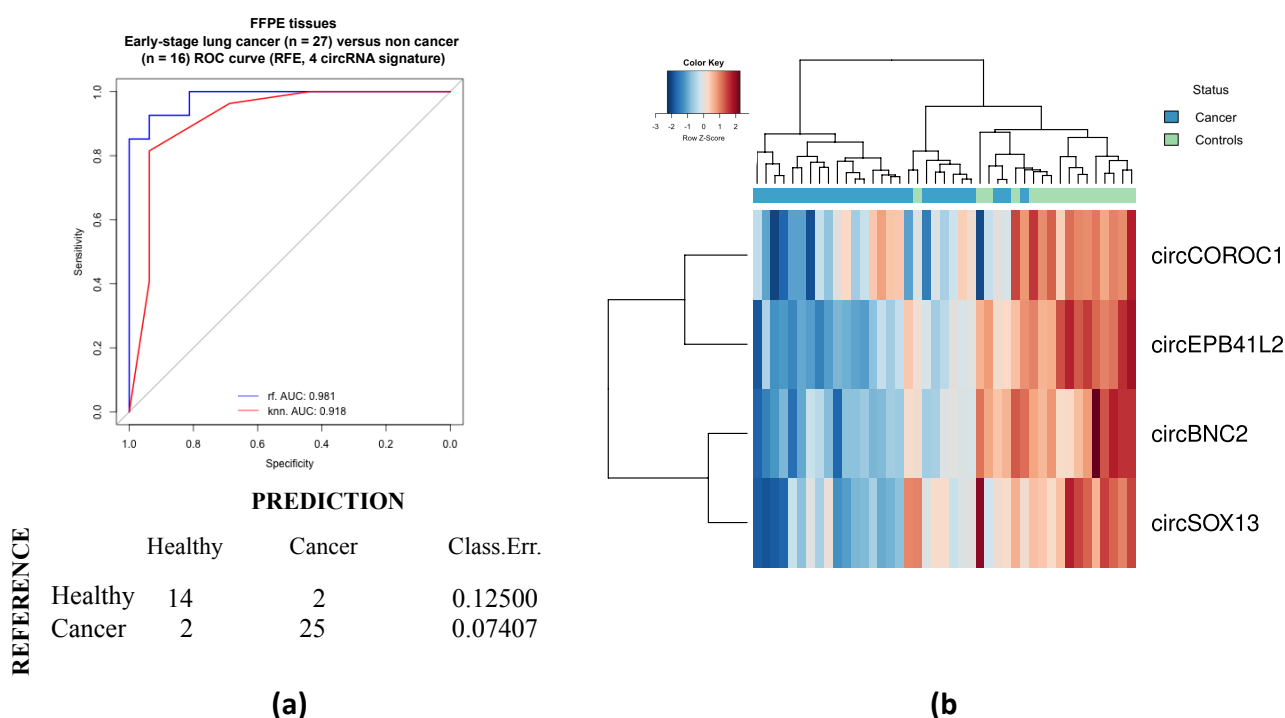


Figure 6. (a) Area under the ROC curve of the 4 circRNA-signature using recursive feature elimination (RFE) for cohort classification. Confusion matrix was generated based on the RF classification scores. Classification error scores are indicated. **(b)** Hierarchical clustering of samples based on the 4-circRNA signature.

3.6. Univariate analysis related to lung cancer risk

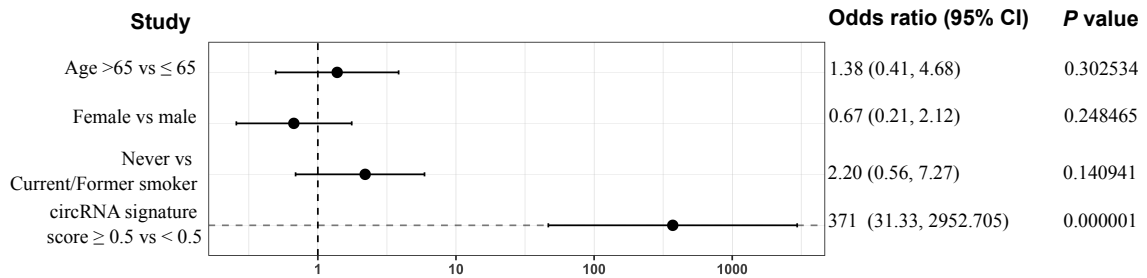
We then explored if certain patient characteristics could provide risk factors for lung cancer by performing a univariate analysis (**Figure 7**). Several characteristics that could be associated with higher risk of lung cancer such as age, gender, and smoking status, were evaluated. No significant association 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.

3.7. Validation by RT-qPCR and Sanger Sequencing of circRNA junction sites

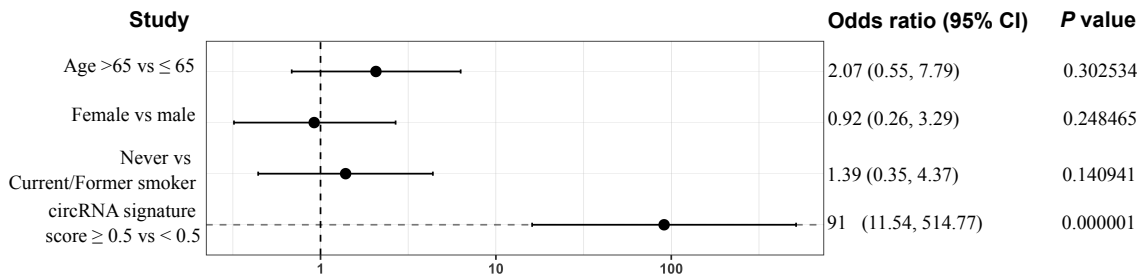
CircEPB41L2, circSOX13 and circBNC2 not only were significantly downregulated both in early and late stages showing the highest fold-change, but also were selected by the RFE algorithm as part of the two predictive signatures. As a result, these 3 targets were selected to validate the nCounter performance using RT-qPCR.

Divergent primers and probes spanning the junction sites were designed for the specific amplification of cited circular transcripts (**Figure 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 (**Figure 8b**). A gel electrophoresis of the PCR products revealed 3 bands corresponding to the size of expected amplicons (**Figure 8c**). Further Sanger sequencing validated these findings by exposing the circRNA junction sites (**Figure 8d**).

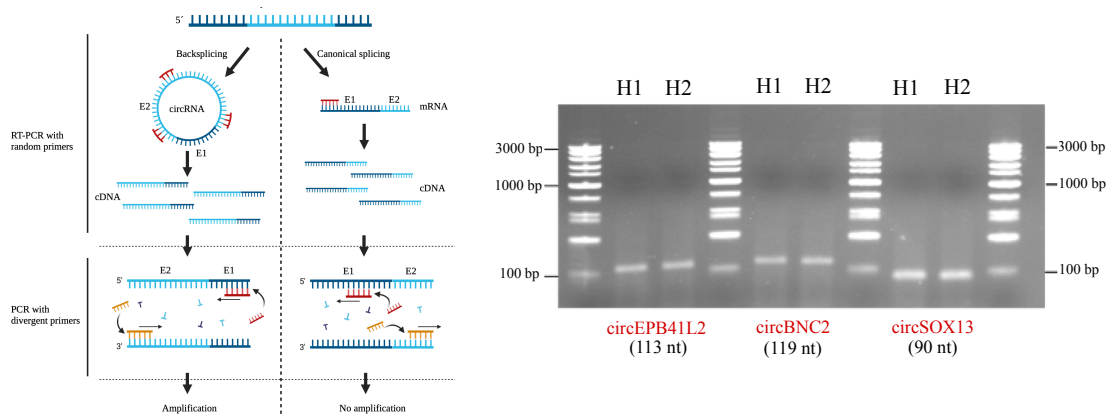


(a)



(b)

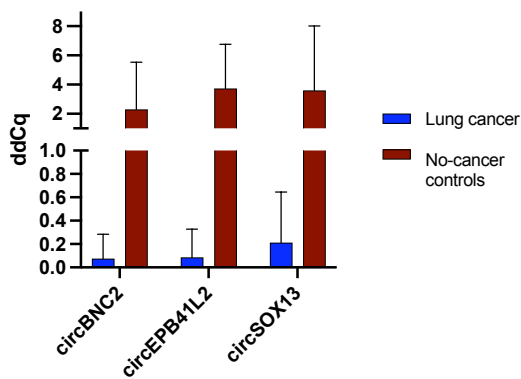
Figure 7. Univariate analysis exploring associations between patient characteristics and lung cancer to determine risk factor. Forest plot represents the odds ratios in (a) lung cancer; and (b), early-stage lung cancer cohorts with a 95% wald confidence limit. Student's t-test was used for the calculation of P-values.



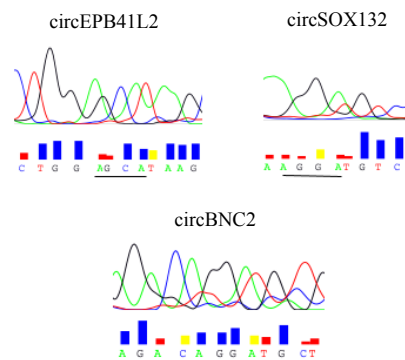
(a)

(c)

circRNA expression assessed by RT-qPCR



(b)



(d)

Figure 8 (reverse page). 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.

4. Discussion

Precision oncology currently relies on genomic, transcriptomic or proteomic-based features that serve as decision-making support, predicting treatment outcome [100].

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 [101]. 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 [99]. On this regard, some authors have proved the benefits of this technology to study circRNA subcellular distribution [102], or elucidate the potential roles in skin [103, 104] or brain diseases[105]. 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 (**Figure 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 thirteen 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 [61, 106-109]. From the detected circRNAs, seventeen were found upregulated, according to literature review [106-109], contrasting to the results achieved with nCounter; however, only

3 of those could be further validated by RT-qPCR [106, 108, 109]. 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 [107]. 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 [96]. 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 [110]. 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 nCounter processing. Consequently, most expressed circRNAs were enriched up to 56-fold when compared to controls, and only 5 circRNAs were found affected by this treatment. Sensitivity of specific circRNAs towards the endonuclease activity of the RNase R enzyme was expected since it could be found reported in other publications [111]. Full or partial degradation of all linear transcripts included in the panel was observed, hence, validating the circRNA nCounter panel. Eighteen 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 [32, 33] 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 [112]. 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 [113]; Therefore, untreated total RNA samples were used for the expression experiments in our study.

Since circRNA represents only 5 to 10% of total RNA [32, 33], 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 [114].

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 [108], 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 [115]. Similarly, circFARSA upregulation has been described in tumor cells, promoting migration and invasion [116]. 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 [117, 118]; 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.* in a recent publication, where this circRNA was shown to increase proliferation, invasion and migration of NSCLC cells via miR-944/YESI axis [119].

For circCHD9, only one miRNA, hsa-miR-1229, was predicted. This miRNA was found upregulated in breast cancer activating β -Catenin/Wnt signaling [120]; 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 [121].

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 (**Figure S9**). This miRNA was previously described to be involved in colorectal and esophageal cancer progression activating the Wnt/ β -catenin signaling pathway [122, 123]. 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.

6. Supporting information

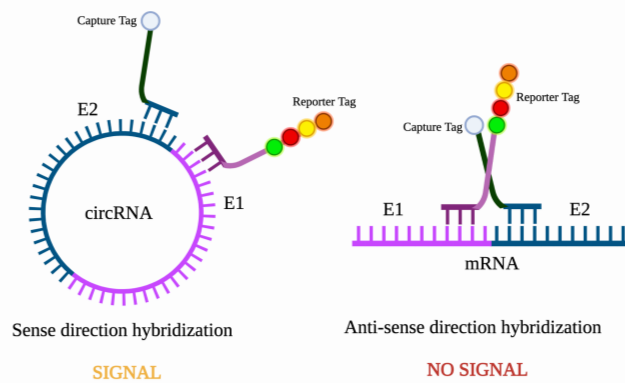
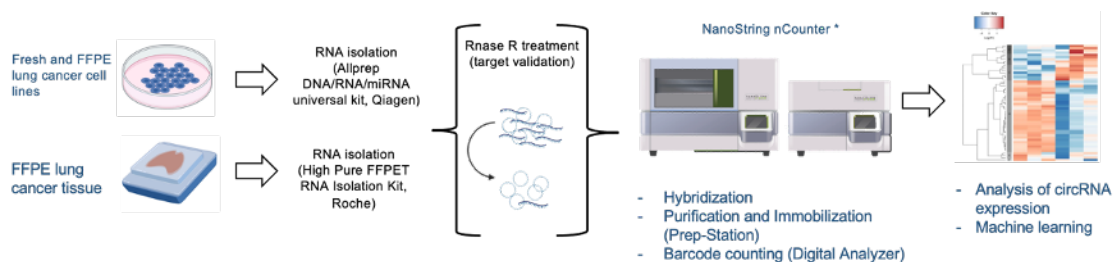


Figure S1. nCounter probe design allows specific recognition of the circRNAs included in the panel.



* Custom panel including 78 circRNA targets and 10 house keeping mRNA

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

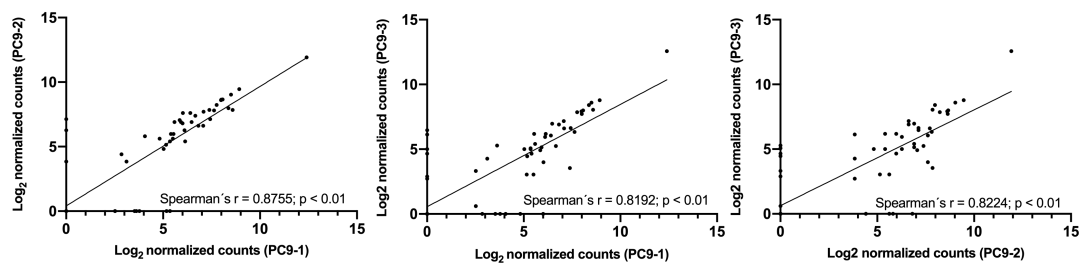


Figure S3. Reproducibility experiment comparing the log₂ of normalized counts by nCounter from three independent RNA samples derived from the PC9 cell line. Spearman's correlation coefficient is indicated.

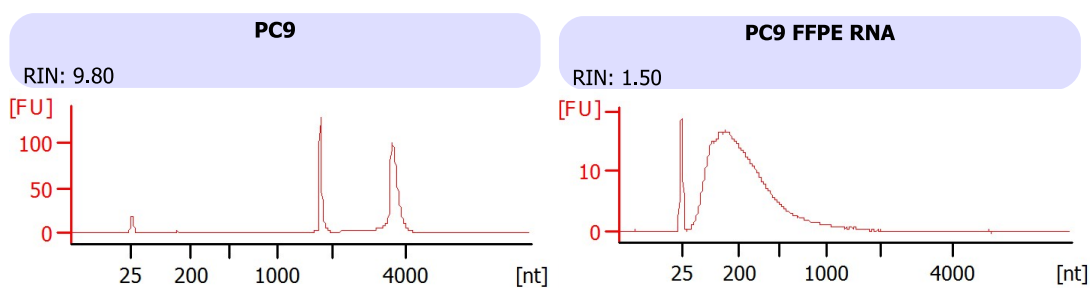


Figure S4. Bioanalyzer profiles of faired fresh (left) and FFPE (right) PC9 cell line-derived RNA. RNA integrity number (RIN) score is indicated on the upper-left corner of each figure.

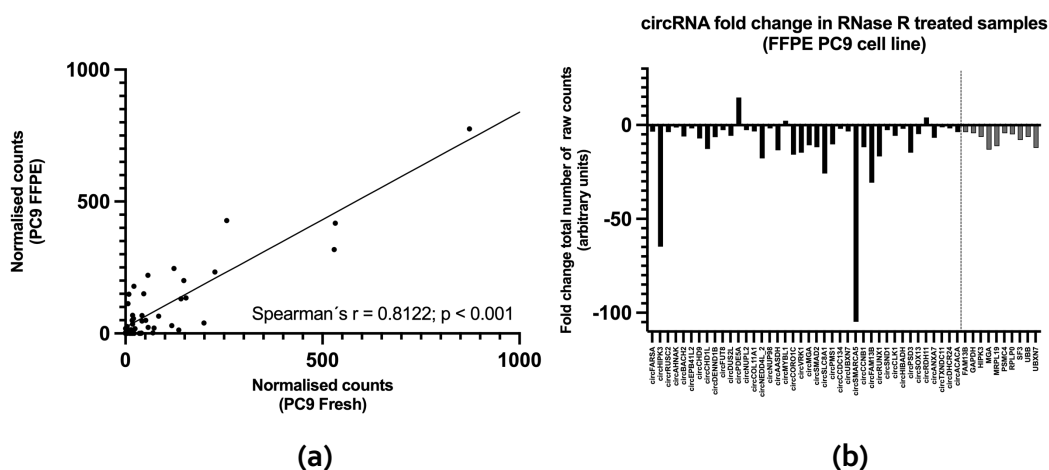


Figure S5. nCounter analysis of FFPE PC9 cell line. (a) Correlation between circRNAs from FFPE PC9 versus fresh PC9 cell lines. Spearman's correlation coefficient is indicated. (b) Bar plot showing circRNA and linear HK fold change after RNase R treatment.

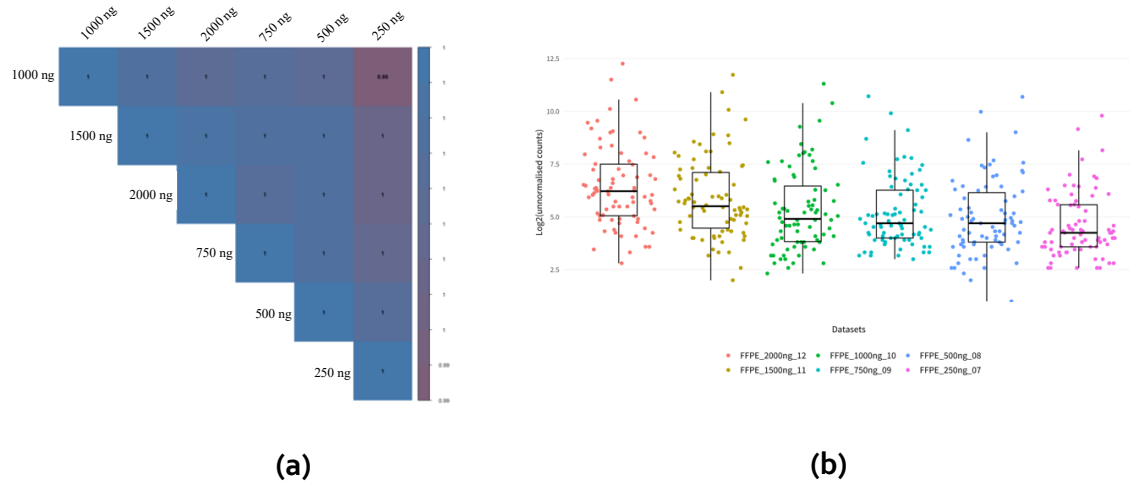


Figure S6. Total RNA concentration assessment for circRNA analysis using the nCounter platform. **(a)** Figure showing the correlation of normalized counts among the different RNA concentrations tested. **(b)** Boxplot of the log₂ unnormalized datasets showing the count distribution in each of the different RNA concentrations tested.

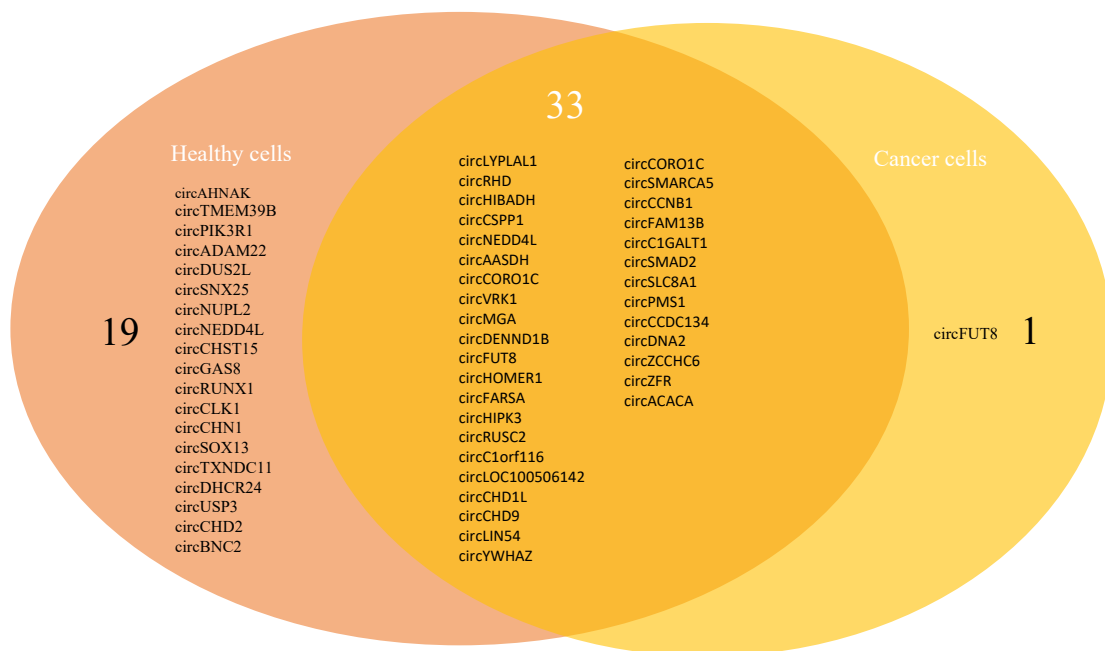


Figure S7. Venn diagram showing circRNAs identified in all healthy cells (19) versus those only expressed in all lung cancer cell lines (1). Those circRNAs expressed in all cell lines are indicated (33).

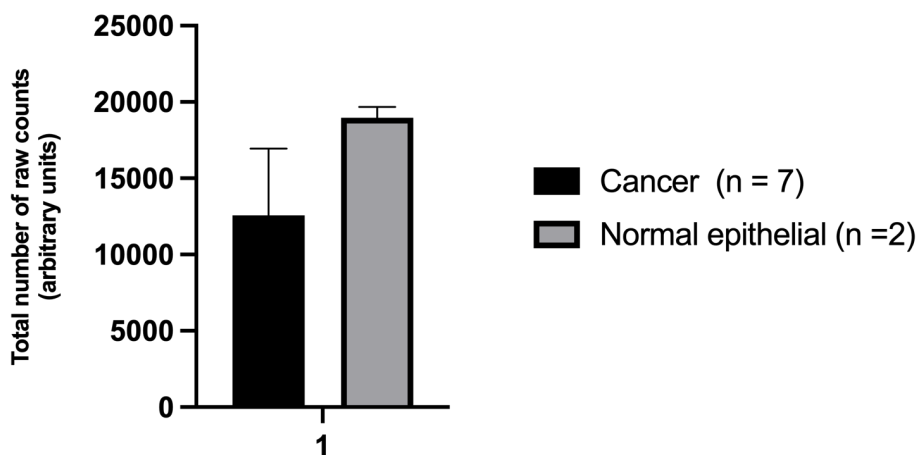


Figure S8. Overall total number of raw counts in lung cancer (A549, PC9, H2228, H3122, HOP-62, HCC-827, H1666) and normal epithelial cell lines (aale, HBEC3KT).

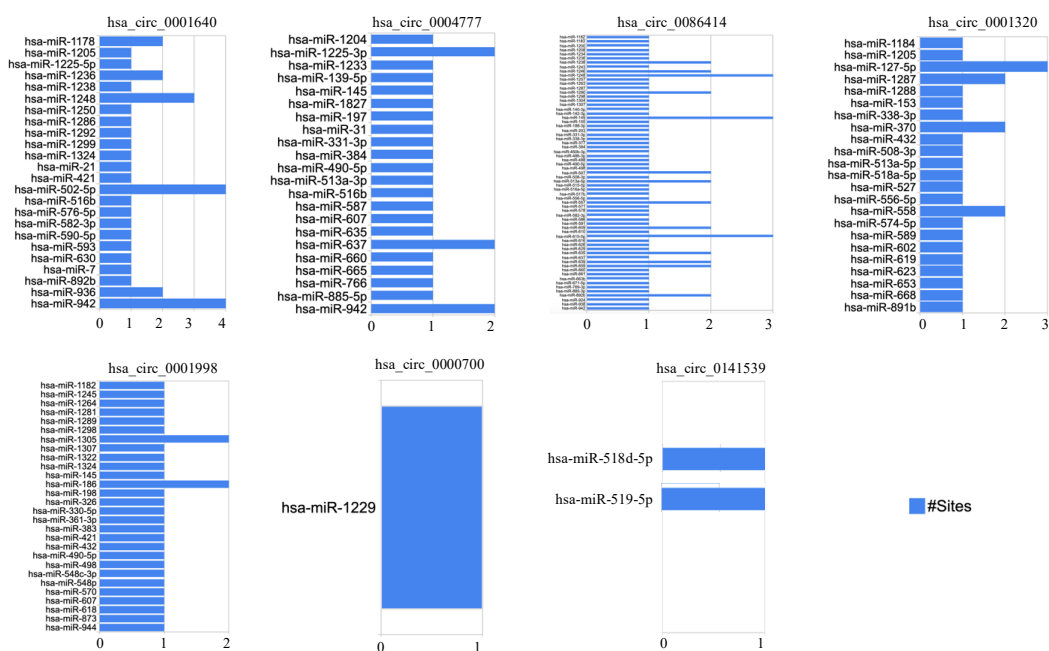


Figure S9. Different miRNA binding sites of dysregulated circRNAs in early-stage lung cancer tissues.

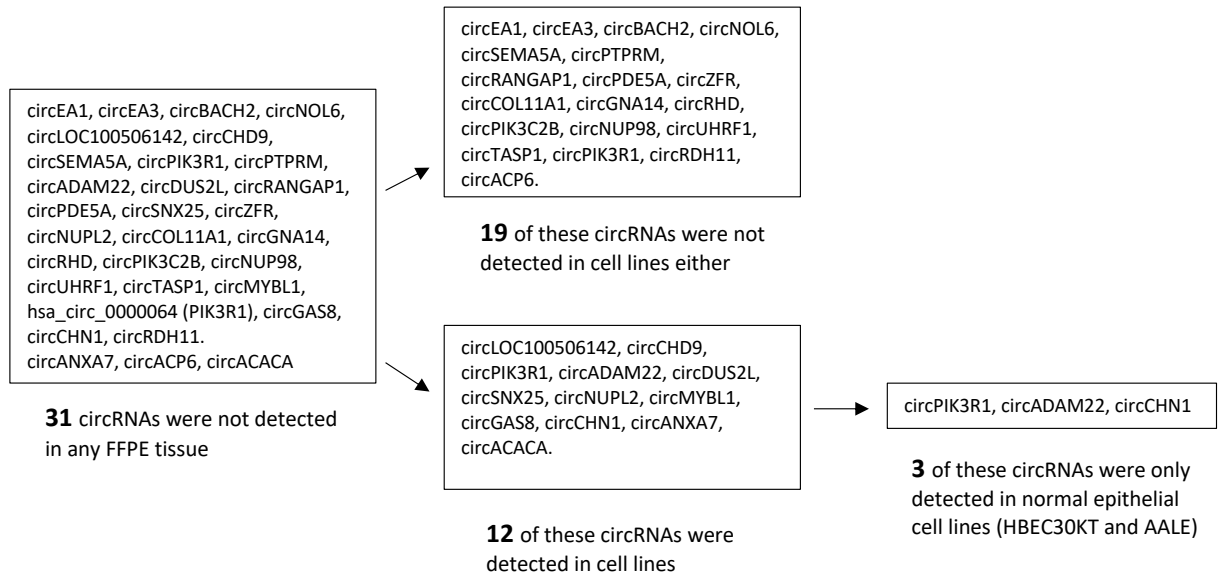


Figure S10. Diagram showing the tracking of those circRNAs of the circRNA nCounter panel not detected in assessed FFPE tissues.

Table S1. Diagnosis and associated pathologies of the control cohort. Different individuals with different characteristics were selected as controls to ensure that both, potential differentially expressed circRNAs and ML signatures, would be specific for lung cancer classification.

Diagnosis and associated pathologies of control cohort					
Patient ID	Sex	Age	Diagnosis	Smoking status	Associated pathologies
IGTP-1	M	46	Pulmonary congestion and emphysematous changes	Severe	Aortic insufficiency Polyglobulia, thrombopenia, moderate-severe COPD, hypothyroidism Not information
IGTP-2	M	50	Pneumonia	Severe	
IGTP-3	M	69	Multiple infracentimetric, fibrous and whitish lesions in the right upper lobe, subpleural in location, suggestive of an old inflammatory process	Ex-smoker	Not information
IGTP-4*	F	61	Infiltrating acinar adenocarcinoma	Moderate	Tuberculosis in 1992
IGTP-5*	F	47	Adenocarcinoma	Severe	Not information
IGTP-6*	F	76	Adenocarcinoma	N.I	Not information
IGTP-7*	M	73	Metastatic carcinoma	Ex-smoker	Gout disease
IGTP-8	M	29	Bullous emphysema	N.I	Not information

Chapter II

IGTP-9	F	62	Emphysema, edema and pulmonary congestion, diffuse and bilateral.	Smoker	Endoprosthesis from the arch to the proximal segment of the descending aorta with debranching of the right brachiocephalic trunk and the left common carotid artery. And the second in 2018 due to a new type A aortic dissection with severe AI, performing biological aortic replacement and substitution of the ascending aorta up to the first segment of the arch, with a Dacron tube, is complicated by endocarditis and vegetative growth in biological prosthesis, is reoperated for valve replacement and repair of the aortic ring. Arteriosclerosis
IGTP-10	M	75	No remarkable evidence.	N.I	
IGTP-11	M	57	Aortic rupture at the thoracic level secondary to type B aortic dissection	Severe	Obesity with a pathological history of type- B aortic dissection, in addition to severe aortic regurgitation without stenosis and Bentall-Bono surgery with ATS mechanical prosthesis and replacement of the aortic arch with implantation of a Thoraflex prosthesis.
IGTP-12	M	76	Intestinal ischemia	N.I	Ischemic heart
IGTP-13	M	34	Pulmonary embolism	N.I	disease Arterial hypertension, dyslipidemia, and type 2 diabetes mellitus
IGTP-14	M	47	Dilated cardiomyopathy.	Ex-smoker	Dilated cardiomyopathy with two-vessel coronary disease, stage 3 chronic kidney disease.
IGTP-15	F	N.I	N.I	N.I	Not associated pathologies
IGTP-16	F	N.I	N.I	N.I	Not associated pathologies

*N.I = Not information; F = Female; M = Male. *A non-tumoral region of the lung was used as control*

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

Characteristics FFPE sample	Lung cancer patients (n= 53)	Controls (n = 16)
Tumor infiltration – no. (%)		
0-20 %	1 (1.9)	-
21-40 %	6 (11.3)	-
41-60 %	13 (24.5)	-
61-80 %	16 (30.2)	-
81-100 %	4 (7.6)	-
Not information	13 (24.5)	-
Lymphocyte infiltration		
0-5 %	9 (17.0)	10 (62.5)
6-10 %	14 (26.4)	4 (25.0)
11-20 %	9 (17.0)	-
21-30 %	2 (3.8)	-
31-40 %	4 (7.5)	-
40-50 %	2 (3.8)	-
Not information	13 (24.5)	2 (12.5)

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

Accession	Gene	Type of transcript
hsa_circ_0000264	CHST15	Circular RNA
hsa_circ_0039161	ITGAX	Circular RNA
hsa_circ_0002360	RUNX1	Circular RNA
hsa_circ_0006276	ANXA7	Circular RNA
hsa_circ_0004062	CHN1	Circular RNA
hsa_circ_0004417	LYPLAL1	Circular RNA
hsa_circ_0001320	FOXP1	Circular RNA
hsa_circ_0004777	SOX13	Circular RNA
hsa_circ_0003941	CLK1	Circular RNA
hsa_circ_0003148	LIN54	Circular RNA
hsa_circ_0001675	C1GALT1	Circular RNA
hsa_circ_0003655	SND1	Circular RNA
hsa_circ_0035654	USP3	Circular RNA
hsa_circ_0037007	CHD2	Circular RNA
hsa_circ_0009128	TXNDC11	Circular RNA
hsa_circ_0001998	FUT8	Circular RNA
hsa_circRNA_404643	PIK3C2B	Circular RNA
hsa_circ_0005066	SNX25	Circular RNA
hsa_circ_0008144	PDE5A	Circular RNA
hsa_circRNA_401977	NEDD4L	Circular RNA
hsa_circ_0006296	CHD1L	Circular RNA
hsa_circ_0063526	RANGAP1	Circular RNA
hsa_circ_0001623	BACH2	Circular RNA
hsa_circ_0006324	DENND1B	Circular RNA
hsa_circ_0005096	NUPL2	Circular RNA
hsa_circRNA_407081	MYBL1	Circular RNA
hsa_circRNA_406483	AASDH	Circular RNA
hsa_circRNA_404833	NUP98	Circular RNA
hsa_circ_0072088	ZFR	Circular RNA
hsa_circ_0039908	DUS2L	Circular RNA
hsa_circRNA_404458	RHD	Circular RNA
hsa_circ_0001806	CSPP1	Circular RNA
hsa_circ_0006349	TMEM39B	Circular RNA
hsa_circ_0080968	ADAM22	Circular RNA
hsa_circ_0000700	CHD9	Circular RNA

hsa_circRNA_406083	TASP1	Circular RNA
hsa_circRNA_400294	COL11A1	Circular RNA
hsa_circ_0008584	PTPRM	Circular RNA
hsa_circ_0000317	AHNAK	Circular RNA
hsa_circ_0001640	EPB41L2	Circular RNA
hsa_circRNA_405718	UHRF1	Circular RNA
hsa_circ_0002099	SEMA5A	Circular RNA
hsa_circ_0001845	NOL6	Circular RNA
hsa_circ_0000999	LOC100506142	Circular RNA
hsa_circRNA_404185	GNA14	Circular RNA
hsa_circ_0006411	PIK3R1	Circular RNA
hsa_circRNA_401977	NEDD4L	Circular RNA
hsa_circ_0000729	GAS8	Circular RNA
hsa_circ_0005962	YWHAZ	Circular RNA
hsa_circ_0086414	BNC2	Circular RNA
hsa_circ_0040809	BANP	Circular RNA
hsa_circ_0003958	HIBADH	Circular RNA
hsa_circ_0005139	RDH11	Circular RNA
F-circEA3	EML4-ALK	Circular RNA
F-circEA1	EML4-ALK	Circular RNA
hsa_circ_0013958	ACP6	Circular RNA
hsa_circ_0043256	ACACA	Circular RNA
hsa_circ_0012673	DHCR24	Circular RNA
hsa_circ_0021592	HIPK3	Circular RNA
hsa_circ_0000064	B4GALT2	Circular RNA
circ_001569	RUSC2	Circular RNA
hsa_circ_0049627	FARSA	Circular RNA
hsa_circ_0006916	HOMER1	Circular RNA
hsa_circ_0001495	CCNB1	Circular RNA
hsa_circ_0000566	VRK1	Circular RNA
hsa_circ_0001238	CCDC134	Circular RNA
hsa_circ_0007037	ZCCHC6	Circular RNA
hsa_circ_0141539	C1ORF116	Circular RNA
hsa_circ_0001083	PMS1	Circular RNA
hsa_circ_0006151	DNA2	Circular RNA
hsa_circ_0004458	PSD3	Circular RNA
hsa_circ_0000847	SMAD2	Circular RNA
hsa_circ_0000994	SLC8A1	Circular RNA

hsa_circ_0001535	FAM13B	Circular RNA
hsa_circ_0000591	MGA	Circular RNA
hsa_circ_0001380	UBXN7	Circular RNA
hsa_circ_0000437	CORO1C	Circular RNA
hsa_circ_0001445	SMARCA5	Circular RNA
NM_018955.2	UBB	Linear RNA
NM_001256799.1	GAPDH	Linear RNA
NM_014763.3	MRPL19	Linear RNA
NM_006503.2	PSMC4	Linear RNA
NM_001002.3	RPLP0	Linear RNA
NM_005877.5	SF3A1	Linear RNA
NM_005734	HIPK3	Linear RNA
NM_001101800	FAM13B	Linear RNA
NM_001080541	MGA	Linear RNA
NM_015562	UBXN7	Linear RNA

Author contributions

C.P.V., J.W.P.B., R.R. and M.A.M.V. conceptualized and designed the experiments. M.I. and A.A.H were responsible of patient recruitment and sample collection. C.P.V, A.G.C., E.A, and S.R.M. performed the experiments. S.P.G, N.P., C.P.V, performed data analysis with contributions of A.F.H and M.H. C.P.V., R.R. and M.A.M.V. wrote the main manuscript and prepared the figures. C.Y.H. and S.W. contributed to reagents and materials. A.F.H, M.H. T.B., M.F. and C.A.E. provided editing, comments, and experimental guidance. All authors reviewed the manuscript.

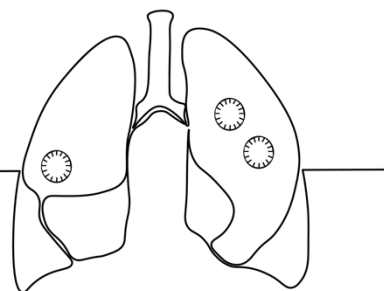
Acknowledgements

We would like to thank Stephanie Davis for her language editing assistance. The investigators also wish to thank the patients for kindly agreeing to donate samples to this study. We thank all the physicians who collaborated by providing clinical information. Graphical Abstract, Figures 1A, 8A and S1 were created with Biorender.com.

Funding sources and disclosure of conflicts 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. This project has received funding from a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No 7654

Chapter III: Plasma-derived EV-circRNA analysis using the nCounter platform



Multiplex analysis of circRNAs from plasma extracellular vesicle-enriched samples for the detection of early-stage non-small cell lung cancer

Carlos Pedraz-Valdunciel, Stavros Giannoukakos Ana Giménez-Capitán, Diogo Fortunato, Martyna Filipaska, Jordi Bertran-Alamillo, Jillian W.P. Bracht, Ana Drozdowskyj, Joselyn Valarezo, Natasa arovni, Alberto Fernández-Hilario, Michael Hackenberg, Andrés Aguilar-Hernández, Miguel Ángel Molina-Vila, Rafael Rosell.

Pharmaceutics (First published: 24 September 2022)

DOI: <https://doi.org/10.3390/pharmaceutics14102034>

Abstract

Background: The analysis of liquid biopsies brings new opportunities in the precision oncology field. Under this context, extracellular vesicle circular RNAs (EV-circRNAs) have gained interest as biomarkers for lung cancer (LC) detection. However, standardized, and robust protocols need to be developed to boost their potential in the clinical setting. Although nCounter has been used for the analysis of other liquid biopsy substrates and biomarkers, it has never been employed for EV-circRNA analysis of LC patients. Methods: EVs were isolated from early-stage LC patients (n=36) and controls (n=30). Different volumes of plasma, together with different number of pre-amplification cycles, were tested to reach the best nCounter outcome. Differential expression analysis of circRNAs was performed, along with the testing of different machine learning (ML) methods for the development of a prognostic signature for LC. Results: A combination of 500 μ L of plasma input with 10 cycles of pre-amplification was selected for the rest of the study. Eight circRNAs were found upregulated in LC. Further ML analysis selected a 10-circRNA signature able to discriminate LC from controls with AUC ROC of 0.86. Conclusions: This study validates the use of the nCounter platform for multiplexed EV-circRNA expression studies in LC patient samples, allowing the development of prognostic signatures.

Keywords

circRNAs; extracellular vesicles; nCounter; lung cancer; NSCLC; liquid biopsies.

1. Introduction

With 350 deaths per day projected for 2022, lung cancer stands as the main cause of cancer-related mortality, leading the second highest incidence in the United States and Europe [124, 125]. Treatments have proved to be more effective at the early stage of the disease, when lung cancer patients benefit from a significantly improved overall survival (OS) [126]. However, most cases are diagnosed at an advanced stage, with a 5-year survival rate dropping to only 2-8% in stage IV.

In order to achieve early detection, many challenges need first to be faced. Classical biopsy techniques for sampling and profiling of suspicious pulmonary nodules often involve invasive procedures. Limitations of such practices include restricted access to the nodules, which regularly compromise the quality and quantity of extracted biopsy specimens. Heterogeneity of resected samples also hampers the use of these methods, especially for tumor identification [127].

Liquid biopsies offer a minimally invasive procedure for sampling, providing a practical tool for continuous monitoring of lung cancer patients [10], being also actively investigated for early detection [14]. Despite the slow progression on the development of liquid biopsies in this area, many possible biomarkers have been proposed in the last few years, including circulating tumor DNA (ctDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs), proteins, extracellular vesicles (EVs) and tumor educated platelets (TEPs).

Lung cancer elicits massive changes in RNA metabolism, reflecting both in the tumor transcriptome and in the circulating EV and TEP cargo. EVs contain different RNA molecules, including mRNA and non-coding RNAs such as miRNA or circular RNAs (circRNAs) [54, 128]. The circRNA transcripts are generated by post-transcriptional circularization of the 5' and 3' ends in an alternative process called backsplicing. Their circular structure makes most of them resistant to exonucleases and, therefore, robustly stable RNA molecules, compared to the canonical (linear) mRNA. CircRNAs seem to play an important role in human homeostasis [86, 87]. Moreover, it has been reported that aberrant expression of certain circRNAs can promote cancer development and progression [92]. Also, some circRNAs have been investigated as liquid biopsy biomarkers for the early detection of lung cancer and other solid tumors [82, 129]. However, the lack of consensus on a robust and standardized protocol for circRNA quantification is holding back the development of clinically applicable assays.

RT-qPCR, microarrays and RNAseq are the three methods most commonly used in circRNA research. However, the RT-qPCR does not allow high-throughput analysis; microarrays have a limited dynamic range of RNA detection; and RNAseq is associated with high cost, long time-consuming protocols, and high grade of complexity when it comes to data analysis.

An alternative technique for multiplex analysis of circRNA is nCounter, which provides a cost-effective automated solution for analysis of more than 800 targets with minimal hands-on time, providing highly reproducible data in less than 48 hours. nCounter is based on the detection of RNA of interest using target-specific probe pairs. Each pair comprises of a reporter probe with a unique color combination at the 5' - end, allowing specific recognition of the gene of interest; and a capture probe carrying a molecule of biotin, which provides a molecular grip to the nCounter cartridge, allowing downstream digital detection [96]. The expression of a particular gene is then calculated by counting the number of times a specific color-coded probe is detected. This technology has been embraced in translational research, including the development and validation of liquid biopsies, due to its capability of working with low quantity of highly degraded samples [69, 130]. Recent studies reported the use of nCounter for the study of several categories of circulating biomarkers [11, 67, 77, 78, 131],

including EV-derived DNA [132], miRNA [79, 80], mRNA [133] and circRNA [134]. However, nCounter analysis of EV-circRNAs has not been investigated for early detection of lung cancer. Here, we report the development of a protocol for EV enrichment from plasma followed by RNA purification and circRNA analysis by nCounter.

Then, we analyzed liquid biopsies from non-cancer donors and early-stage non-small cell lung cancer (NSCLC) patients and applied machine learning (ML) to develop a prognostic signature.

2. Materials and Methods

2.1. Patient samples

The study was carried out in accordance with the principles of the Declaration of Helsinki, under an approved protocol of the institutional review board of Quirón Hospitals. We obtained and documented written informed consent from all the patients. A total of 36 samples from early-stage NSCLC (stages IA to IIIA) were selected from our institution, along with 30 samples from non-cancer controls (Table 1). Clinical information from patients and controls included age, gender, smoking status, tumor histology and stage, when applicable. All samples were de-identified before further processing for confidentiality purposes.

Table 1. Clinicopathologic characteristics of enrolled patients (n=66).

Clinicopathological characteristics	NSCLC patients (n=36)	Non-cancer controls (n=30)
Gender – no. (%)		
Male	18 (50.0)	13 (43.3)
Female	18 (50.0)	17 (56.7)
Age – yr.		
Median	71.5	38
Range	32-91	23-57
Histological type		
Adenocarcinoma	27 (75.0)	-
Squamous carcinoma	4 (11.1)	-
Not information	5 (13.9)	-
Smoking status – no. (%)		
Former-orcurrent smoker	20 (55.5)	11 (36.6)
Never smoker	13 (36.2)	17 (56.7)
Not information	3 (8.3)	2 (6.7)
Tumor stage – no. (%)		
I	19 (52.8)	-
II	2 (5.5)	-
IIIA	15 (41.7)	-

2.2. Plasma processing

Around 10 mL of whole blood was collected from the participants enrolled in the study using sterile EDTA Vacutainer tubes (BD, Plymouth, UK) and processed within the next 2 h. Blood samples were centrifuged twice at 2000 x g at room temperature (RT) in a Rotina 380 R centrifuge (Hettich, Tuttlingen, Germany) for 10 min to separate plasma from red/white

blood cells, platelets, and cell debris. Aliquoted plasma samples were then stored at -80 °C until downstream processing.

2.3. *Enrichment of EVs*

EVs were isolated from plasma using differential ultracentrifugation (UC) as described previously [135] or the miRCURY Exosome Serum/Plasma Kit (Qiagen, Hilden, Germany). In the case of UC, 500 µL plasma samples were transferred into 15 mL sterile high-speed centrifuge tubes (VWR-Avantor, PA, USA), filled up with sterile 1x phosphate-buffered saline (PBS) and centrifuged twice at 10,000 x g for 30 min at 4 °C in a Sorvall RC 6 Plus centrifuge (Thermo Fisher Scientific, Waltham, MA, USA). Supernatants were then transferred into UC tubes (Beckman Coulter, CA, USA), equilibrated with sterile 1x PBS, and spun twice at 70,000 x g for 1h at 4 °C in the Sorvall WX Ultra 100 centrifuge (Thermo Fisher Scientific). The EV enriched pellets were resuspended in 100 µL sterile PBS and stored at -80 °C until used. EV enrichment with the miRCURY Kit was performed as described [133]. Debris was cleared from 500 µL plasma samples with thrombin and subsequent centrifugation at 10,000 x g for 5 min at RT. Samples were then incubated with Precipitation Buffer overnight at 4 °C and centrifuged twice (500 x g, 5 min at RT). Supernatants were discarded, EV enriched pellets were resuspended in 270 µL of Resuspension Buffer and stored at -80 °C until used.

2.4. *Transmission electron microscopy (TEM)*

Visualization of EV samples was performed by the TEM service of the Universitat Autònoma de Barcelona (UAB). A volume of 3.9 µL of EV-enriched sample was blotted onto a Holey Carbon Film Supported Nickel Grid (Merck, Darmstadt, Germany) previously glow-discharged in a PELCO easiGlow glow cleaning system (Ted Pella Inc, CA, USA). Next, the grid containing the sample was plunged into a Leica EM GP cryo-work station (Leica, Wetzlar, Germany) comprising a liquid ethane bath cooled to -180 °C, and subsequently transferred and visualized in a JEOL 2011 TEM (Jeol Ltd, Tokyo, Japan) operating at 200 kV. Samples were kept at -180 °C during the observation and captures were obtained with a Gatan Model 895 UltraScan 4000 4k x 4k CCD camera (Gatan Inc, CA, USA). Image processing was performed using ImageJ software (version 1.8.0, National Institutes of Health, MD, USA).

2.5. *Nano-flow cytometry measurements*

The volume of EV samples was brought to 500 µL with sterile PBS. Size-exclusion chromatography (SEC) columns (qEVoriginal/35nm, Izon Science, Oxford, UK) were equilibrated with 20-30 mL of sterile PBS and eluted using the same buffer. Collection started immediately after loading the sample into the column, according to manufacturer instructions. Eluted EV-enriched samples were directly analyzed with the nanoFCM (NanoFCM Ltd., Nottingham, UK), a nanoparticle flow cytometer. Instrument calibration with standard beads enabled accurate measurements of both size and concentration of 40-200 nm particles through the detection of their side scatter [136].

2.6. *RNA isolation and DNase treatment*

EV-enriched samples were treated with 4 µg/mL of RNase A (Sigma-Aldrich, MO, USA) for 1 h at 37 °C, to eliminate any non-vesicular RNA. TRI Reagent (MRC, OH, USA) was added to a final volume of 1 mL and incubated at RT for 20 min. Then, 200 µL of a Chloroform and Isoamyl Alcohol dilution (24:1) (Panreac Química SLU, Barcelona, Spain) were added followed by vigorous shaking and centrifugation at 12,000 x g for 15 min at 4 °C. Upper

fraction was collected, and RNA was precipitated by adding 2.5 μL of glycogen (Merck) and 500 μL 2-propanol (Merck), followed by incubation at RT for 10 min and further centrifugation at 12,000 $\times g$ for 10 min at 4 °C. RNA pellet was then washed with 75% ethanol, dried at 95 °C for 3 min, and resuspended in 12 μL of nuclease-free water.

The DNA-free DNA Removal Kit (Thermo Fisher Scientific) was used to eliminate any DNA remaining in the samples. Following the manufacturer's protocol, 0.75 μL of DNase buffer and 1 μL enzyme were added to 7.5 μL RNA sample and incubated at 37 °C for 30 min. A volume of 0.75 μL of DNase inactivation reagent was then added to the reaction, incubated for 2 min at RT and centrifuged for 1.5 min at 10,000 $\times g$ and RT. The supernatant containing EV-RNA was then transferred to a fresh tube and stored at -80 °C until further use.

2.7. RT-qPCR and Sanger sequencing analysis

RT-qPCR and Sanger sequencing of circRNA junction sites were performed as previously described [98]. Divergent primers and probe sets were designed using Primer Express 3.0 Software (version 3.0.1, Applied Biosystems) with the probes spanning the circRNA junction site (Table 2). Five microliters of EV-RNA was converted into cDNA using the M-MLV reverse transcriptase enzyme and random hexamers (both from Invitrogen, MA, USA). 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 circRNA. Three replicas of each sample were run for the quantification of the expression of each assessed circular transcript. Quantification of gene expression was performed using the QuantStudio™ 6 Flex System (Applied Biosystems) and the comparative Ct method.

Table 2. Primer and probe design for circRNA validation by RT-qPCR.

CircRNA		
	Forward	5'CGGCCAGTCATGTATCAAAGAC 3'
circHIPK3	Reverse	5'AAAGGCACTTGACTGAGTTTGATAAA 3'
	Probe	FAM 5'AATCTCGGTACTACAGGTATG 3' MGB
	Forward	5'AGATGTTGTCGAATTTGTGGAAAA 3'
circZCCHC6	Reverse	5'TCTTCTACCATTGATAAAAGCCTTCAT 3'
	Probe	FAM 5'GAGGAGAAATGACAAATT 3' MGB

For Sanger sequencing, 10 μL of each PCR product was subjected to electrophoresis in a 2x agarose gel (100V, 30 minutes) 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 and 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).

2.8. nCounter processing

The nCounter Low RNA Input Amplification Kit (NanoString Technologies, WA, USA) was used to retrotranscribe and pre-amplify 4 μ L of EV-derived RNA in a Verity thermal cycler (Applied Biosystems, MA, USA) following NanoString's guidelines. Briefly, samples were denatured at 95 °C for 10 min and hybridized for 18 h at 67 °C. Our custom-made nCounter panel (including 78 circRNAs, 6 linear reference genes and 4 mRNAs [137]) was used to analyze EV-derived pre-amplified cDNA according to the manufacturer's instructions. RCC files containing data outputted by the NanoString nCounter Flex System (NanoString Technologies) from each run were exported to the nSolver Analysis Software (version 4.0.70, NanoString Technologies, Seattle, WA, USA).

2.9. Differential expression analysis

Raw count nCounter values were exported to Microsoft Excel (version 16.40, Microsoft, Redmond, WA, USA) using nSolver Analysis Software. The background was calculated for each sample as (geo)mean \pm 2SD of the negative probe counts (NCs). Raw counts lower than the background were automatically excluded from further analysis. The raw circRNA counts were normalized using the total number of counts of the sample and multiplied by 10,000. Differential expression analysis was performed comparing the means of the normalized counts for each circRNA in the early-stage NSCLC vs. non-cancer controls. The circRNAs with a fold change >1 and p-value < 0.05 were considered as differentially expressed (DE).

2.10. Data pre-processing and normalization for signature development

Raw RCC-formatted data files were exported from the nSolver Analysis Software (NanoString Technologies). R (version 4.0.3, R Core Team and the R Foundation for Statistical Computing, Vienna, Austria) and R studio (version 2021.09.0, RStudio PBC, Boston, MA, USA) were used for pre-processing and normalization analysis of the imported files. Initial evaluation of the quality and integrity of the RCC data was performed using the NanoStringQCPro (version 1.22.0) package. During this process, we looked for potential outliers based on the performance of standard control metrics provided by NanoString, such as Imaging, Binding Density, Positive Control Linearity, and Limit of Detection. After this first pre-analytical step, samples were subjected to supplementary exploratory examination, including Principal Component Analysis (PCA) and interquartile range (1.5 IQR rule) analysis. Samples found as outliers by both methods were then excluded from downstream analyses. NCs were employed to exclude lowly expressed circRNAs with excessive background noise. The arithmetic mean of the NC \pm 2SD was subtracted from each endogenous circRNA for each sample. Any transcript scoring a value below 0 in more than 75% of the analyzed samples was then excluded from further analysis. PCA plot was then used to re-assess the data after the aforementioned filtering step. Technical variability correction and normalization were performed using the RUVSeq/RUVg function (version 1.24.0) and DESeq2 (version 1.30.1) packages (RUVseq-DESeq2). First, the RUVg function was used to estimate the unwanted variation among samples based on the DE genes. DESeq2 and edgeR (version 3.32.1) performed a first pass DE analysis and the intersected least significant genes (with adjusted p-value above 0.1) were used as "in-silico empirical" negative controls. DESeq2 was then utilized with default parameters along with the RUV factors to perform the normalization of the raw filtered data. The normalization performance was assessed using the standard relative log expression (RLE) plot.

2.11. Machine Learning (ML) for signature development

The Recursive Feature Elimination (RFE) algorithm along with leave-one-out cross-validation (LOOCV) and the random forest (RF) classifier were used to perform feature selection on the normalized data previously generated by RUVseq-DESeq2. The optimal number of features was automatically selected by keeping only those yielding best performance after cross-validation. These final features were to constitute the prognostic signature. To test the predictive power of the selected signature, extra trees classifier (ETC), k-nearest neighbor (KNN) and RF models were built using default parameters. The 5-fold cross validation (5-CV) algorithm was applied for this purpose. During this process, the dataset was randomly split into k-folds ($k=5$), being $4/5$ of the data used to train the model, while the remaining $1/5$ was used to test its behavior. The classifier showing the highest area under the ROC curve (AUC ROC) value was selected as the final model. Signature scores for each sample were obtained from the final model. A confidence threshold of 0.5 was considered for the calculation of the positive and negative predictive values (PPV-NPV). Additional statistical indicators such as accuracy, sensitivity, specificity, and Cohen's κ were also calculated.

2.12. Univariate and multivariate analyses

Association between clinical characteristics and ML-generated signature was assessed with a univariate Cox proportional-hazard regression model. Odds ratios, with a Confidence Interval (CI) of 95% was calculated using the MedCalc Statistical Software (MedCalc Software Ltd. Odds ratio calculator. https://www.medcalc.org/calc/odds_ratio.php). Multivariate analysis using logistic regression was performed using SAS software (v9.4, SAS Institute, NC, USA). Significance was set at $p < 0.05$ for all statistical tests.

3. Results

3.1. Enrichment of plasma EVs and workflow development for nCounter circRNA analysis

Two replicated 500 μL plasma samples from an early-stage NSCLC patient and a non-cancer control were submitted to EV enrichment by ultracentrifugation (UC) or using the miRCURY Exosome Serum/Plasma kit. Enriched EVs were characterized by transmission electron microscopy (TEM) and nanoparticle flow cytometry via nanoFCM. TEM images revealed different clusters of diverse-sized EVs (30 to 300 nm, all within the reported EV size range [138-140]) in all samples regardless of the enrichment method used (**Figure 1a**). Samples extracted using the miRCURY kit showed a higher proportion of vesicles with an exosome-like size range (30-100 nm) by TEM, compared to the more heterogeneous UC samples (**Figure 1a**). NanoFCM analysis revealed a higher concentration of 40-100 nm particles in samples enriched using the miRCURY kit (**Figure 1b**). In addition, nanoFCM indicated a higher number of particles/mL in the NSCLC patient sample when compared to the control, both in the UC and miRCURY preparations (**Figure 1b**).

Next, different volumes of plasma (500 μL , 1000 μL and 1500 μL) from a NSCLC patient were tested in duplicates to assess the effect of initial volumes on downstream circRNA analysis by nCounter using the custom panel we previously developed [137]. Since RNA concentration from EV enriched samples has been demonstrated to be insufficient for direct nCounter analysis [133], pre-amplification steps of 14 and 20 cycles were tested. The utmost total number of counts was achieved using an input of 500 μL both with 14 and 20 cycles (14151 ± 1864 and 686525 ± 345655 , respectively; **Figure 2a**). Consequently, 500 μL of plasma was

also the volume allowing the detection of more circRNAs ($n = 27.5 \pm 4.95$ and 33 ± 7.07 for 14 and 20 cycles, respectively; **Figure 2b**), even if only those with a score above 10 counts after background removal were selected (**Table S1**, **Figure S1**).

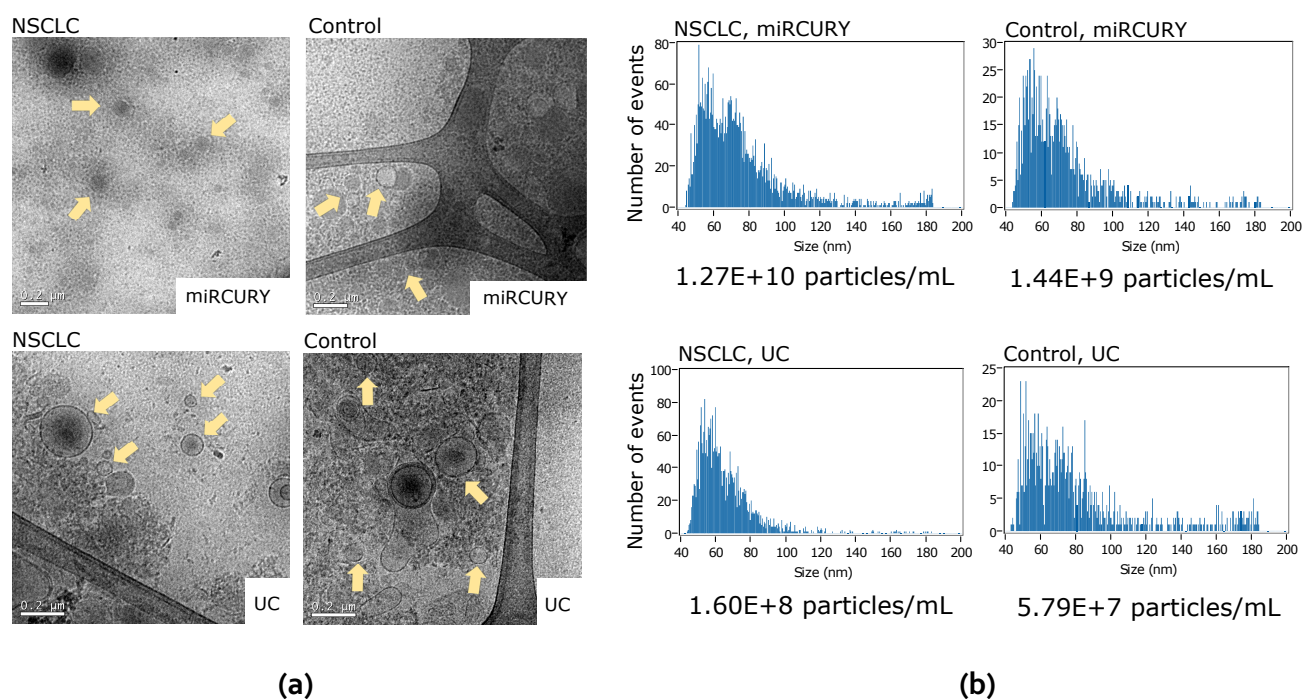


Figure 1. Characterization of extracellular vesicles (EVs) enriched either by differential ultracentrifugation (UC) or precipitation using the miRCURY Exosome Serum/Plasma kit. **(a)** Observation of EV samples on transmission electron microscopy (TEM). Yellow arrows point out EVs of different sizes. Scale bars indicate 200 nm; **(b)** Nanoflow cytometry (nanoFCM) profiles of EV samples showing size and concentration of 40-200 nm particles.

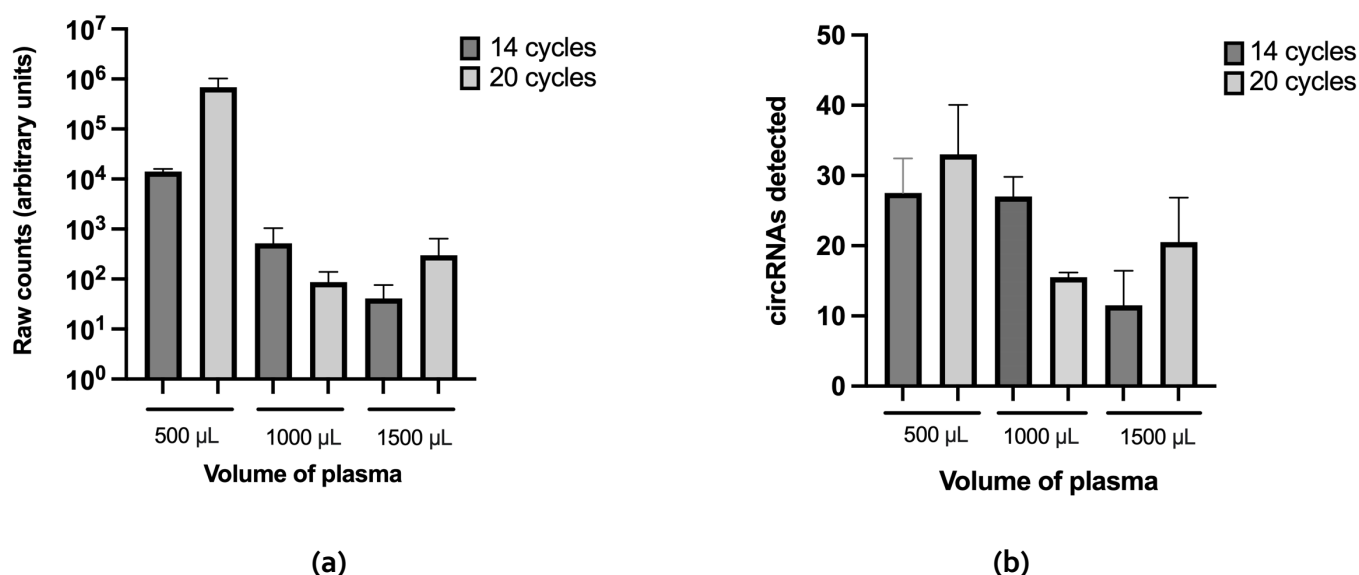


Figure 2. Plasma input testing. **(a)** Total number of counts and **(b)** number of circRNAs detected by nCounter with each of the volumes of plasma tested with 14 and 20 cycles of pre-amplification. Plasma from a NSCLC patient was used for this purpose. Error bars indicate standard deviation.

Different amplification cycles (10, 12 and 14) were subsequently tested in a 500 μ L plasma sample. The highest number of raw counts was obtained with 14 cycles (**Figure 3a**). Regarding the number of circRNAs, 10 and 12 cycles yielded similar results ($n = 51.5 \pm 9.19$ and 52.5 ± 7.78 respectively). More circRNAs were detected at 14 cycles ($n = 59 \pm 16.97$) with a high variability between replicates (**Figure 3b**, **Table S2**). In view of these results, we selected for EV-circRNA analysis a protocol that included 500 μ L of plasma input, EV enrichment with the miRCURY kit, extravesicular RNA elimination with RNase A, EV lysis and RNA extraction with TRI reagent, retrotranscription and nCounter analysis with a 10-cycle preamplification step (**Figure 4**).

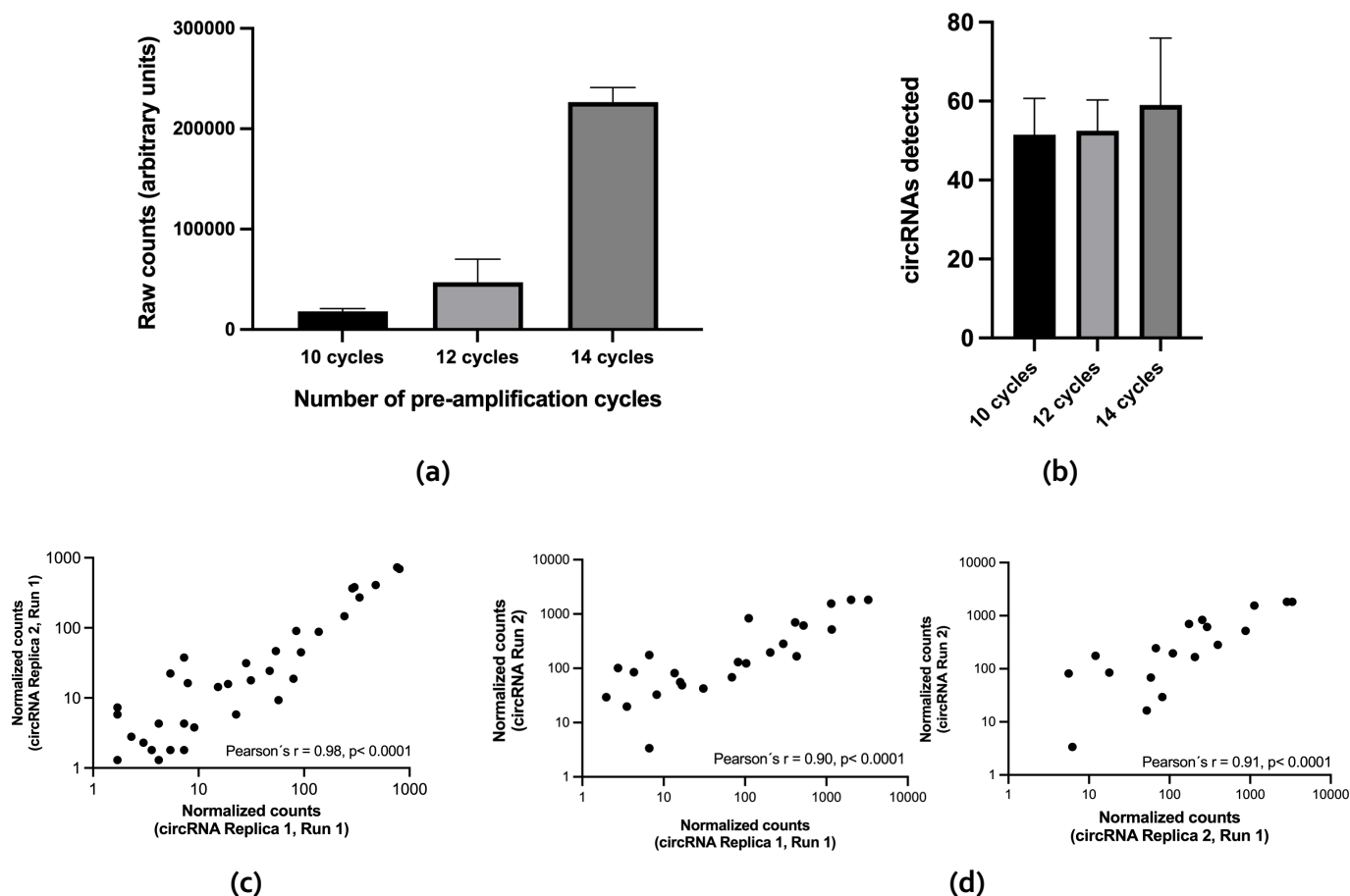


Figure 3. Testing of different number of pre-amplification cycles. Effect of the number of pre-amplification cycles (10, 12 and 14) on (a) the total number of raw counts and (b) total number of circRNAs detected. Error bars indicate standard deviation. (c) Correlation of the two technical nCounter duplicates subjected to 10 cycles of pre-amplification. Pearson's correlation coefficient is indicated. (d) Correlation of each of the technical duplicates from the same nCounter run with the results obtained in an independent nCounter assay of the same sample. Pearson's correlation coefficient is indicated

The repeatability of the protocol was first tested by submitting to nCounter duplicates of a preamplified plasma sample. A strong correlation between the normalized counts was found between the duplicates, represented by a Pearson's r of 0.98, $p < 0.0001$ (**Figure 3c**). When the same plasma sample was re-purified and re-analyzed, nCounter results also showed a strong correlation with the initial duplicates (Pearson's $r = 0.90$ - 0.91 ; $p < 0.0001$) (**Figure 3d**).

3.2. CircRNA expression in plasma EV samples

Plasma from 66 individuals, 36 early-stage NSCLC patients and 30 non-cancer donors, were analyzed using the protocol previously described in section 3.1 (**Figure 4**). An average of 40 ± 14 EV-circRNAs per sample were detected in controls vs. 47 ± 9 in the NSCLC cohort. This difference was found not significant by the Mann-Whitney U test (**Figure 5a**). Among the 78 circRNAs included in the panel, 70 were detected in at least one NSCLC sample and 68 in at least one non-cancer control. A total of 66 EV-circRNAs were shared by both cohorts, while 4 EV-circRNAs were exclusive to NSCLC patients and 2 to non-cancer donors (**Figure 5b**, **Table S3**).

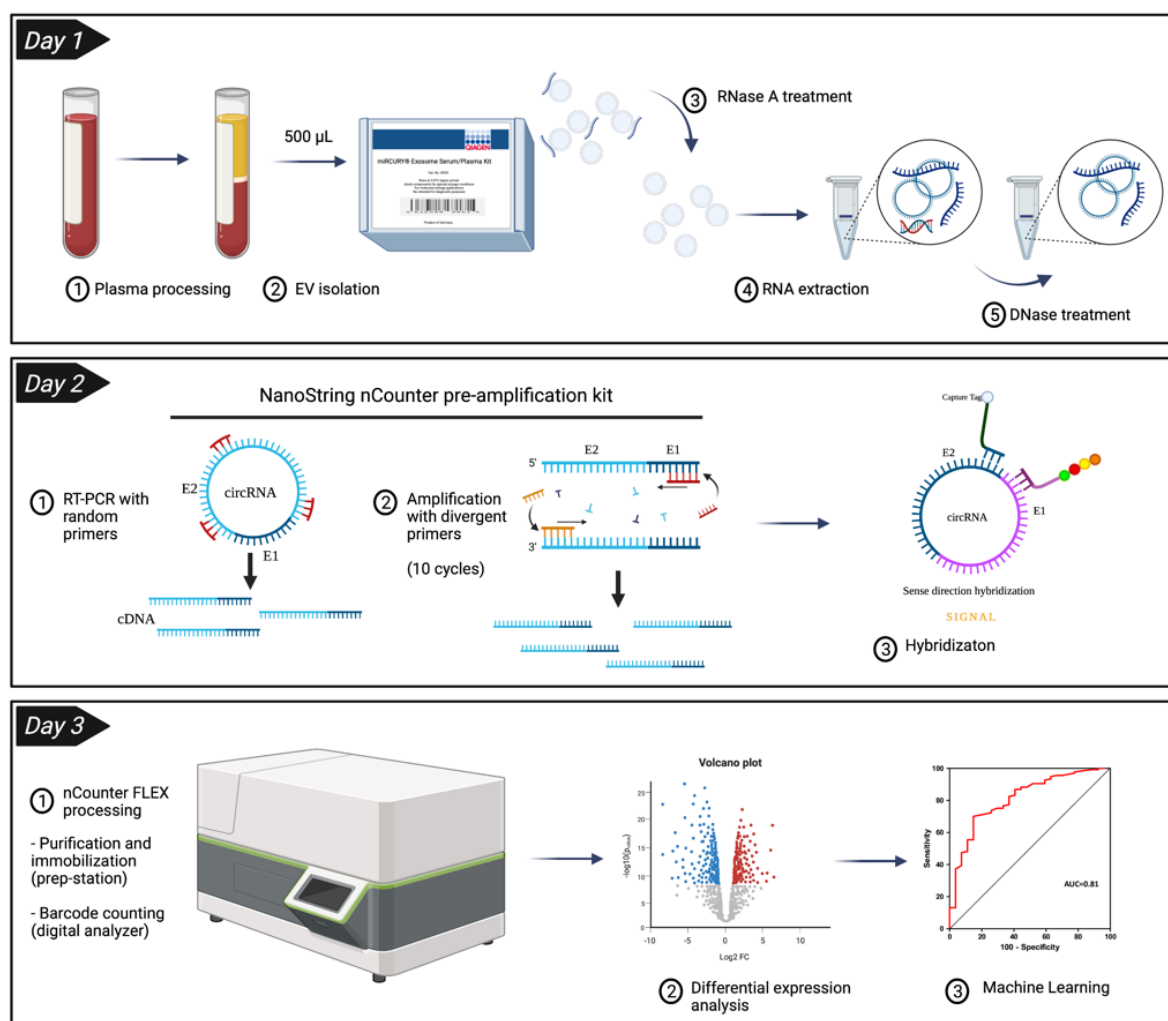


Figure 4. Final workflow established for the study of circRNAs from plasma extracellular vesicles (EVs) using the nCounter technology. A volume of 500 μ L of plasma was used in the miRCURY kit to precipitate EVs. Rnase A was used to remove any non-vesicular RNA that could be present in the sample before proceeding with manual RNA extraction with TRI reagent. RNA samples were treated with DNase to eliminate any trace of genomic DNA, followed by retro-transcription and a pre-amplification step of 10 cycles. Finally, samples were hybridized overnight before nCounter processing.

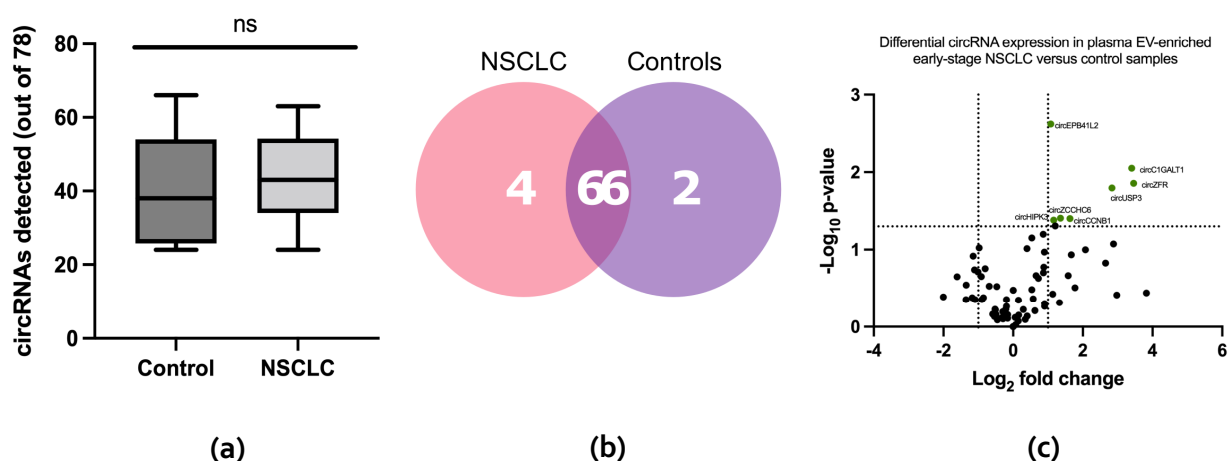


Figure 5. EV-circRNA detection and differential expression analysis. **(a)** Number of circRNAs detected in extracellular vesicle (EV) enriched samples from cancer patients and non-cancer controls using our custom circRNA nCounter panel, which targets 78 circRNA (Mann–Whitney U test, $p = 0.3807$). **(b)** Venn diagram displaying circRNAs identified in early-stage NSCLC and non-cancer controls, featuring those shared by both cohorts. **(c)** Differential expression analysis of log₂-normalized counts between the early-stage NSCLC and control EV samples. circEPB_{41L2}, circC1GALT1, circZFR, circUSP₃, circZCCHC6, circHIPK₃ and circCCNB1 were found upregulated in NSCLC samples.

DE analysis revealed eight circRNAs significantly upregulated in EV-enriched samples from NSCLC patients vs. controls; namely circular Erythrocyte Membrane protein Band 4.1 Like 2 (circEPB_{41L2}), circular Core 1 Synthase, Glycoprotein-N-Acetylgalactosamine -3-Beta-Galactosyltransferase 1 (circC1GALT1), circular Zinc Finger RNA Binding Protein (circZFR), circular Ubiquitin Specific Peptidase 3 (circUSP₃), circular Zinc Finger CCHC Domain-Containing Protein 6 (circZCCHC6), circular Cyclin B1 (circCCNB1), circular DENN Domain Containing 1B (circDENN1B) and circular Homeodomain Interacting Protein Kinase 3 (circHIPK₃) (**Figure 5c**). Of them, only circZFR and circC1GALT1 showed <10 counts in each cohort (**Table S4**). To validate these results, we tested the expression circZCCHC6 and circHIPK₃ by RT-qPCR. Divergent primers and probes spanning the junction sites were designed for the specific amplification of these two circular transcripts (**Table 2**) in samples previously assessed by nCounter with sufficient remaining material. Gel electrophoresis of the RT-qPCR products revealed bands matching the size of expected amplicons and subsequent Sanger sequencing confirmed the expected junction sites in the circRNAs (**Figure 6a-b**). Among the six samples analyzed by RT-qPCR, 4 and 6 samples produced satisfactory results for circZCCHC6 and circHIPK₃ respectively. A trend between nCounter counts and RT-qPCR $\Delta\Delta C_t$ s was observed for both circRNAs (**Figure 6c-d**), with circZCCHC6 showing a strong correlation (Pearson's $r = 0.99$; $p = 0.0076$) (**Figure 6c**).

3.3. Development of a circRNA-signature associated with early-stage NSCLC

Interquartile range analysis classified 9/66 samples as potential outliers (**Figure 7a**) and PCA revealed that they deviated from the main cluster of observations (**Figure S2**). Consequently, these 9 samples were excluded from further analysis.

Then, different R packages including DESeq2, edgeR, RUVSeq and their combination were tested in order to select the normalization approach that best adapts to our data. As a result,

RLE plots indicated a superior performance of RUVSeq-DESeq2 versus the other combinations (**Figure 7b**, **Figure S3**). Consequently, RUVSeq-DESeq2 normalization was selected for the rest of the study.

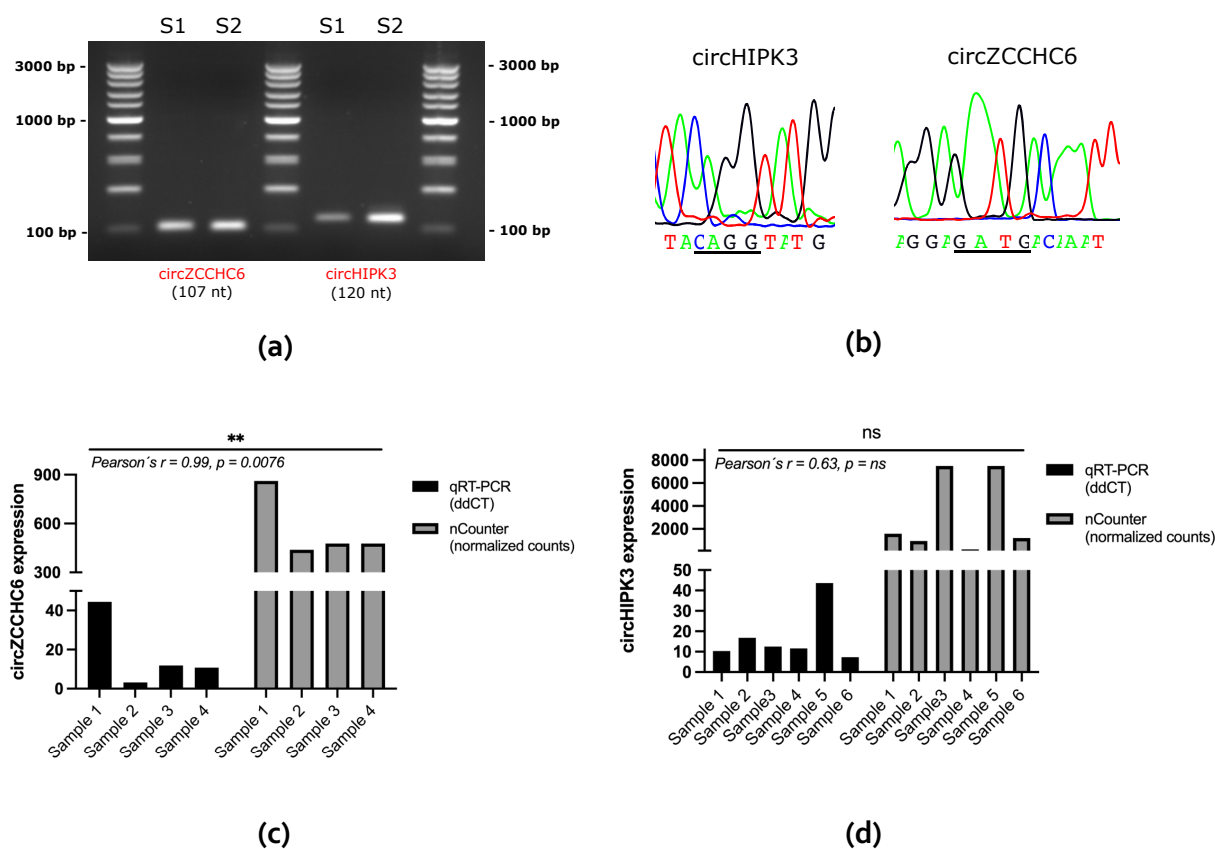
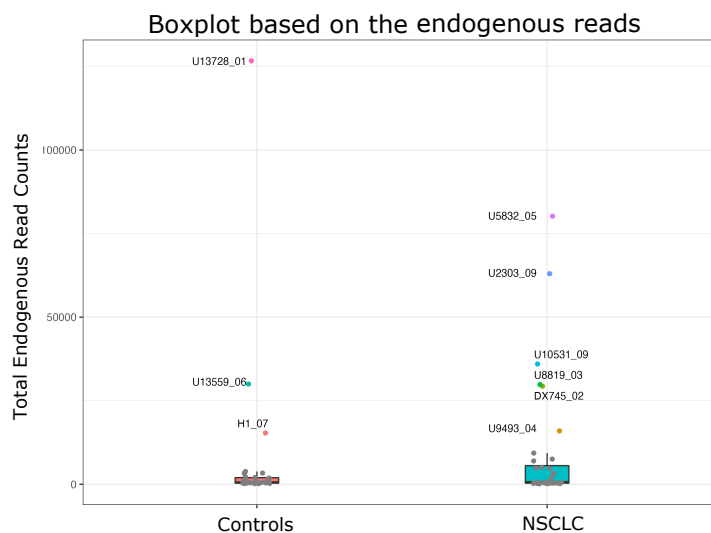


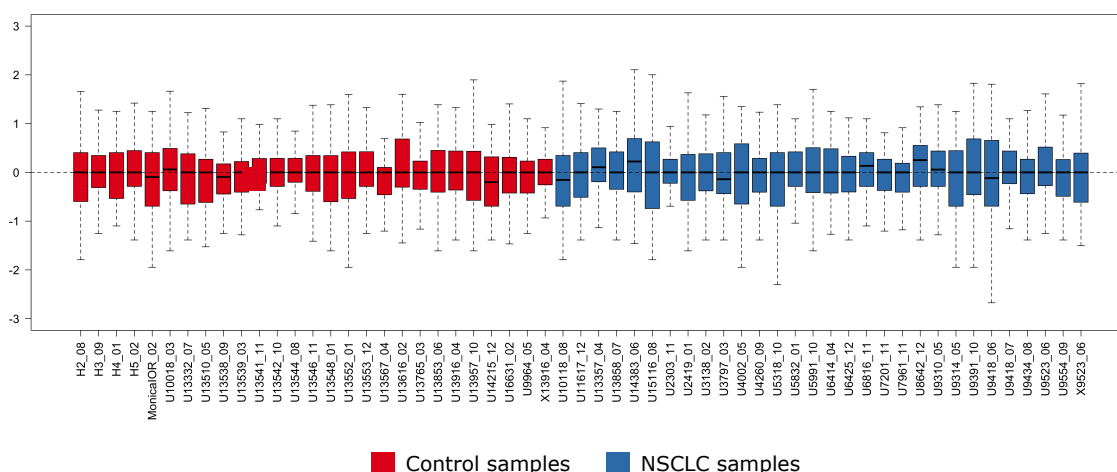
Figure 6. Validation of nCounter results by RT-qPCR and further Sanger sequencing. **(a)** Electrophoresis gel of amplified circZCCHC6 (107 nt) and circHIPK3 (120 nt). **(b)** Sanger sequencing results spanning the junction site (underlined) of cited circRNAs. Comparison of nCounter normalized counts versus $\Delta\Delta C_t$ values by RT-qPCR for circZCCHC6 **(c)** and circHIPK3 **(d)** in analyzed samples. Pearson's correlation coefficient is indicated. NS, not significant.

Next, ML was performed using RFE along with RF classifier and LOOCV, as described in Methods, in order to obtain a signature associated with NSCLC. As a result, ETC was selected as the best model, with a signature of 10 circRNAs (including circular Family With Sequence Similarity 13 Member B -circFAM13B, circular ADAM Metallopeptidase Domain 22 -circADAM22, circular UBX Domain Protein 7 -circUBXN7, circZCCHC6, circular Integrin Subunit Alpha X -circITGAX, circular Retinol Dehydrogenase 11 -circRDH11, circEPB41L2, circular CDC Like Kinase 1 -circCLK1, circular Phenylalanyl-tRNA Synthetase Subunit Alpha -circFARSA, and circular Phosphoinositide-3-Kinase Regulatory Subunit 1 -circPIK3R1) showing an AUC ROC of 0.86 (**Figure 8a**). Signature scores were found to be statistically different when comparing early-stage NSCLC and non-cancer controls (Mann-Whitney U test, $p < 0.001$; **Figure 8b**). The sensitivity and specificity of the ETC signature were of 90% (CI = 73.47% - 97.89%) and 81% (CI = 61.92% - 93.70%) respectively, outperforming the RF and KNN classifiers (**Table 3**). The accuracy achieved with ETC was 86%, resulting in 49 out of the 66 cases being correctly classified (**Figure 8c**).



(a)

DESeq2/RUVg normalization (k=1)



(b)

Figure 7. Data outlier detection and normalization for machine learning (ML) processing. **(a)** Outlier detection using the 1.5 IQR rule. **(b)** RUVSeq/DESeq2 RLE plot of normalized data ($k = 1$).

Then, a univariate analysis was performed to explore the association of the ETC circRNA signature with gender, age, smoking, cancer status and tumor stage (**Figure 9a**). A statistically significant correlation was found of the signature with age (odds ratio = 24.91, $p < 0.0001$), and particularly cancer status (odds ratio of 39.6, $p < 0.0001$).

To further evaluate the implication of age and cancer status on the ML-developed signature, we first performed an exploratory study assessing the interconnexion of both variables by performing a chi-square test. As a result, a strong association between age and cancer status was found, with a $p < 0.0001$ (**Table 4**).

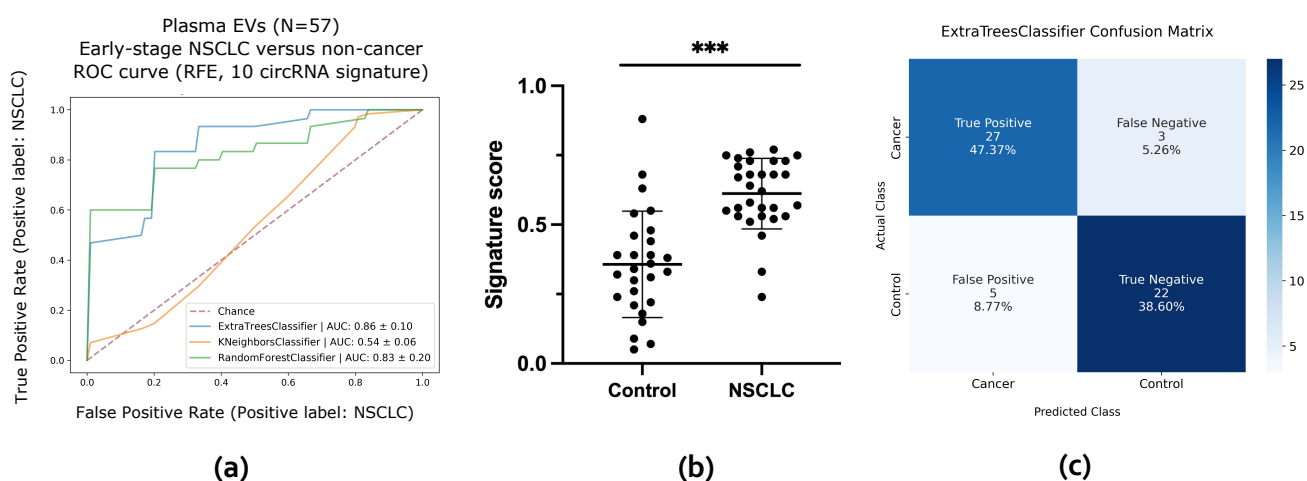
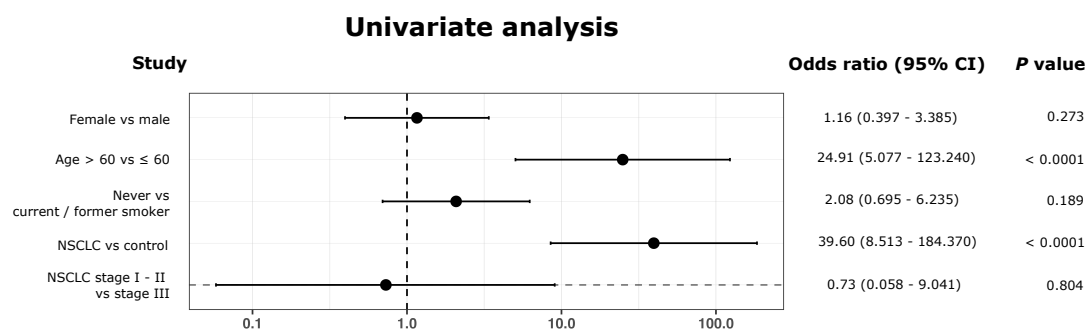


Figure 8. Machine learning (ML) analysis of extracellular vesicle (EV)-enriched samples. **(a)** Area under the ROC curve of the 10 circRNA-signature using recursive feature elimination (RFE) for cohort classification. **(b)** Scores of early-stage NSCLC versus control samples based on expression of the 10-circRNA signature ($p < 0.001$ in a two-tailed Mann–Whitney U test). **(c)** Confusion matrix based on the ETC classification scores.

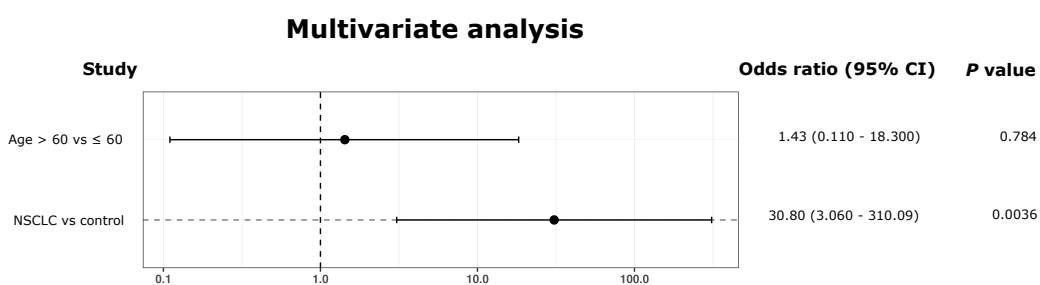
Table 3. Precision assessment of the ML generated circRNA signature with ETC, RF, and KNN. The 95% CI are indicated.

Model	ETC	RF	KNN
No. concordant samples	49	44	30
No. discordant samples	8	13	27
AUC ROC	0.86	0.83	0.54
Accuracy	86%	77%	53%
Sensitivity	90%	83%	50%
	(CI = 73.47% – 97.89%)	(CI = 65.28% – 94.36%)	CI = 31.30% – 68.70%)
Specificity	81%	70%	56%
	(CI = 61.92.1% – 93.70%)	(CI = 49.82% – 86.25%)	CI = 41.83% – 68.49%)
PPV	84%	76%	56%
	(CI = 70.81% – 92.32%)	(CI = 63.10% – 85.10%)	(CI = 41.83% – 68.49%)
NPV	88%	79%	50%
	(CI = 71.18% – 95.61%)	(CI = 62.20% – 89.77%)	(CI = 37.95% – 62.02%)
Cohen's κ	0.72	0.54	0.06
	(CI = 0.458 – 0.976)	(CI = 0.281 – 0.798)	(CI = - 0.202 – 0.313)

ML = machine learning, AUC = Area Under the Curve, ROC = Receiver Operating Characteristic, RF = Random Forest, KNN = K-Nearest Neighbor, CI = confidence interval, PPV = Positive Predictive Value, NPV = Negative Predictive Value.



(a)



(b)

Figure 9. Association between clinical characteristics and ML-generated 10-circRNA signature. (a) Univariate analysis exploring associations between presented 10-circRNA signature and patient characteristics. Forest plot represents the odds ratios with a 95% Wald confidence limit. (b) Multivariate analysis exploring associations between presented 10-circRNA signature with age and cancer status. Forest plot represents the odds ratios with a 95% Wald confidence limit.

Table 4. Association between age and cancer status

Statistic	DF	Value	p-value
Chi-Square	1	32.245	< 0.0001
Likelihood Ratio Chi-Square	1	41.232	< 0.0001

DF = Degrees of freedom

Next, a multivariate analysis was carried out. Results not only demonstrated dependency of these two variables, but also showed a statistically significant correlation between the signature and cancer status ($p = 0.0036$, **Table 5, Figure 9b**). No correlation was found between age and presented signature, in this regard ($p = 0.0784$, **Table 5, Figure 9b**)

Table 5. Analysis of maximum likelihood estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	p-value
Age	1	0.356	1.301	0.075	0.7840
Cancer status	1	3.427	1.178	8.462	0.0036

DF = Degrees of freedom

4. Discussion

EVs are released by most cell types and play an important role in cancer cell communication. Many publications have demonstrated the role of EVs as key modulators in cancer progression [141, 142], which requires intercellular communication mediated by horizontal transferring of biological information via the EV cargo of proteins, DNA, and coding/non-coding RNA, including circRNAs. Therefore, analysis of EVs can provide a snapshot of the tumor and serve as valuable tool to discover liquid biopsy biomarkers. CircRNAs are highly enriched in EVs [54] and show a relatively high stability compared to other forms of RNA [128]. Several studies have highlighted their potential as liquid biopsy biomarkers [82] but current limitations in circRNA quantification methods are limiting their implementation in the clinical setting. Consequently, new, and robust protocols for circRNA analysis are needed. The nCounter platform has gained popularity among translational investigators for transcriptional research not only for solid biopsies but also for EV samples. However, studies focusing on circRNA analysis by nCounter are limited and mostly restricted to tissue specimens [99, 102-105, 137]. In particular, to the best of our knowledge, nCounter has never been applied to the analysis of circRNA in liquid biopsies of lung cancer patients. Consequently, we developed a comprehensive protocol for nCounter-based EV-circRNA expression analysis, from EV enrichment to differential expression and subsequent ML analysis. Key points in this protocol were the initial volume of plasma, the EV purification method, and the number of cycles for the pre-amplification step prior to nCounter testing.

UC is currently still the method of choice for EV isolation in the research setting and we have previously demonstrated its utility for downstream analysis of cell line-derived EV circRNAs [143]. However, ultracentrifuges are not usually available in clinical laboratories, while precipitation-based kits such as the miRCURY Exosome Serum/Plasma Kit represent an easily implementable option with a simple, on-the-bench protocol and short hands-on time. In our study, we compared the two methodologies using plasma samples from a NSCLC patient and a healthy donor. The presence of EV-like particles in all preparations was confirmed by TEM and nanoFCM. Interestingly, a more uniform EV population with an exosomal size-range was found by TEM in both cancer and control samples processed with the miRCURY kit, along with a higher concentration of 40-200 nm particles observed by nanoFCM. A possible explanation to this event could be a size-selective enrichment attributed to this type of precipitation-based preparations, as previously reported in serum samples [144]. This finding prompted us to select miRCURY for further assay development. In addition, a higher number of EV-like particles was observed in the cancer sample compared to the control, regardless the isolation method used. Although a higher number of samples should be analyzed for further confirmation, preliminary results are in agreement with previous reports indicating a higher abundance of EVs in cancer patients [145].

Finally, adding to the evidence provided by TEM and nanoFCM, a treatment with RNase A was applied to EV-enriched samples prior to EV lysis and incorporated into our protocol to eliminate any extravesicular RNA. The resulting and subsequently analyzed RNA proved to be protected from the digestion of cited ribonuclease, indicating a vesicular origin of the transcripts.

In a previous study, a volume of 500 μ L of plasma was found to be sufficient for the analysis of EV-derived mRNA by nCounter [133]. Here, we compared several plasma volumes and

found that 500 μ L outperformed 1000 and 1500 μ L for circRNAs analysis, both in terms of the number of circRNA molecules detected and total counts. A possible explanation for these results may rely on saturation issues with the circRNAs/reporter-probe complexes when a higher plasma input is applied, which impede a correct molecule identification by the digital analyzer. Regarding the number of cycles for the preamplification step, we investigated a range from 10 to 20 in an effort to reduce amplification-related background noise to a minimum, and we found that a 10-cycle pre-amplification step yielded adequate results.

Then, we applied our protocol to assess circular transcripts in early-stage NSCLC samples (n=36) and to non-tumor controls (n=30). We found that 8 circRNAs were found differentially expressed between the two cohorts. Among them, circEPB41L2, circZCCHC6 and circHIPK3 showed the highest number of counts in early-stage cancer patients (**Table S4**). Interestingly, we previously found circEPB41L2 differentially expressed in FFPE tissues of early-stage lung cancer patients [137] and found that it displayed 4 binding sites with hsa-miR-942, which has been described as an activator of the Wnt/ β -catenin signaling pathway [122, 123] in colorectal and esophageal cancers. Our results warrant further investigation in the biology of this circRNA to characterize its role in lung cancer. Regarding circHIPK3, it has been extensively investigated in lung cancer and found to exert a dual activity over miR-149 [146] and miR-124 [147, 148], inducing cell proliferation and inhibiting apoptosis. Our results are in agreement with these findings, since circHIPK3 was upregulated in EV samples from early-stage NSCLC patients. Finally, circZCCHC6 has been recently described to regulate lysophosphatidylcholine acyltransferase 1 (LPCAT1) levels via miR-433-3p [149] in lung cancer. We used circinteractome (www.circinteractome.nia.nih.gov) to investigate possible additional miRNA binding sites, finding matches for 7 additional transcripts (miR-579-3p, miR-623, miR-1197, miR-1304, miR-548l, miR-605, and miR-935). All these miRNAs have been reported as downregulated in lung tumors and have been related with poor prognosis, tumor growth and metastases [150-156].

ML and other computational methods based on artificial intelligence (AI) have emerged in the last decade for multileveled analysis of different datasets. In particular, ML enables computers to make predictions by finding patterns within analyzed data [157], offering a novel approach for the development of predictive signatures that often reach a higher predictive value than biomarkers found by differential expression analyses. Consequently, we decided to use ML in our study. To this end, we developed a pipeline with several steps. First, using IQR and PCA plots, we identified nine outliers, which were excluded from downstream analyses. An RLE plot from each different normalization procedure was generated, showing a higher performance of the RUVSeq-DESeq2 function when compared to the other combinations (**Figure 7B**, **Figure S3A-C**). Finally, we used RFE along with LOOCV and the RF classifier as the feature selection algorithm to automatically determine the most significant circRNAs which are best suited for the construction of the prognostic signature. The final 10-circRNA signature included 2 of the eight circular transcripts previously found by differential expression analysis and eight additional transcripts, including circFARSA. Interestingly, circFARSA has been described as a plasma biomarker of NSCLC [116], promoting tumor invasion and metastases via the PTEN/PI3K/AKT axis [158].

Since we did not sort EV populations, we could not verify the vesicular cell or tissue origin of the circRNAs included in the ML signature nor the origin of the circular transcripts, either

cancer cells or tumor microenvironment. Also, we did not investigate the biological role of the circRNAs, being out of the scope of our work.

In addition, while multivariate analysis could demonstrate that classification accuracy of presented signature is based on cancer status and no other clinicopathological characteristics (**Figure 9**), the lack of > 60-year-old individuals was a limitation in the study. The inclusion of equivalent cohorts in terms of age should be taking into consideration for the design of forthcoming validation studies.

Finally, all 36 cancer samples included in this study were lung adenocarcinomas, with the exception of 4 squamous carcinoma and 5 NSCLC samples with unknown histological subtype. A uniform inclusion of the different lung cancer histologies is suggested for future validation studies to assess the predictive power of the signature for other subtypes of NSCLC.

5. Conclusions

We have demonstrated the feasibility of using nCounter for the multiplex study of plasma-EV circRNAs in liquid biopsies of lung cancer patients, including differential expression analysis and development of predictive ML signatures. Further studies of larger cohorts are warranted in order to determine the clinical applicability of such signatures.

6. Supporting information

Table S1. CircRNAs detected in the different plasma volumes of the same patient with 14 and 20 cycles of pre-amplification.

	500 μ L				1000 μ L				1500 μ L			
	14 cycles		20 cycles		14 cycles		20 cycles		14 cycles		20 cycles	
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
Detected circRNA (counts > 0)	31	24	38	28	29	25	15	16	8	15	16	25
Detected circRNA (counts > 10)	20	18	33	24	9	5	4	0	0	1	1	5

R= replica

Table S2. CircRNAs detected in the plasma of the same individual subjected to 10, 12 and 14 pre-amplification cycles.

	500 μ L					
	10 cycles		12 cycles		14 cycles	
	R1	R2	R1	R2	R1	R2
Detected circRNA (counts > 0)	58	45	58	47	71	47
Detected circRNA (counts > 10)	38	32	39	41	48	40

R= replica

Table S3. circRNAs identified in early-state NSCLC and non-cancer control cohorts.

circRNAs expressed only in the NSCLC cohort	circRNAs expressed in NSCLC and control cohorts	circRNAs expressed only in the control cohort
<i>circTMEM39B, circZFR, circRHD, circPIK3C2B</i>	<i>circHOMER1, circFARSA, circHIPK3, circRUSC2, circC1orf116, circAHNAK, circBACH2, circEPB41L2, circSEMA5A, circCHD1L, circPIK3R1, circADAM22, circDENND1B, circFUT8, circDUS2L, circRANGAP1, circPDE5A, circSNX25, circNUPL2, circCSPP1, circCOL11A1, circNEDD4L, circNEDD4L-2, circUHRF1, circTASP1, circAASDH, circMYBL1, circB4GALT2, circCHST15, circCORO1C, circVRK1, circMGA, circGAS8, circSMAD2, circSLC8A1, circPMS1, circCCDC134, circFOXP1, circUBXN7, circSMARCA5, circCCNB1, circFAM13B, circC1GALT1, circRUNX1, circLIN54, circSND1, circCLK1, circHIBADH, circCHN1, circLYPLAL1, circPSD3, circSOX13, circRDH11, circYWHAZ, circDNA2, circANXA7, circZCCHC6, circTXNDC11, circDHCR24, circACP6, circUSP3, circCHD2, circITGAX, circBANP, circACACA, circBNC2</i>	<i>circNOL6, circNUP98</i>

Table S4. Normalized counts of differentially expressed circRNAs found in the early-stage NSCLC cohort.

circRNA	Gene	Mean non-cancer cohort	Mean early-stage NSCLC cohort (controls)	t-test	Fold change
circRNA HIPK3	HIPK3	161	361	0.042	2.24
hsa_circRNA_001640	EPB41L2	109	230	0.002	2.11
hsa_circRNA_100421	DENN1B	27	62	0.049	2.31
hsa_circRNA_103809	ZFR	0.41	4	0.014	10.98
hsa_circ_0001495	CCNB1	7	21	0.040	3.10
hsa_circ_0001675	C1GALT1	1	7	0.009	10.58
hsa_circ_0007037	ZCCHC6	54	138	0.039	2.57
hsa_circ_0035654	USP3	3	22	0.016	7.13

NSCLC= non-small cell lung cancer

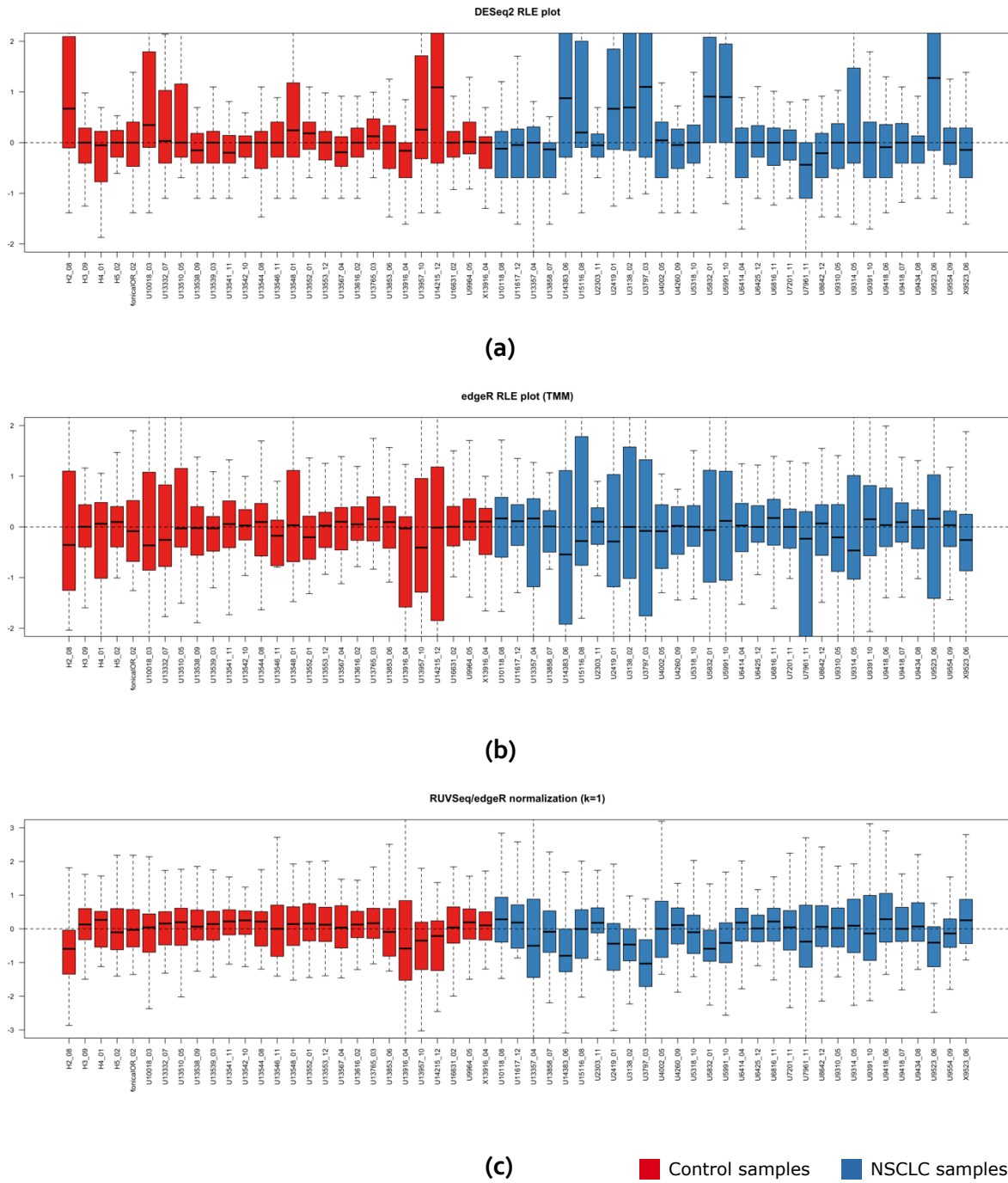


Figure S3. Assessment of the different normalization processes by RLE plot analysis. **(a)** RLE plot of normalized counts by DESeq2. **(b)** RLE plot of normalized counts by edgeR. **(c)** RLE plot of normalized counts using a combination of the edgeR-RUVg methods ($k = 1$).

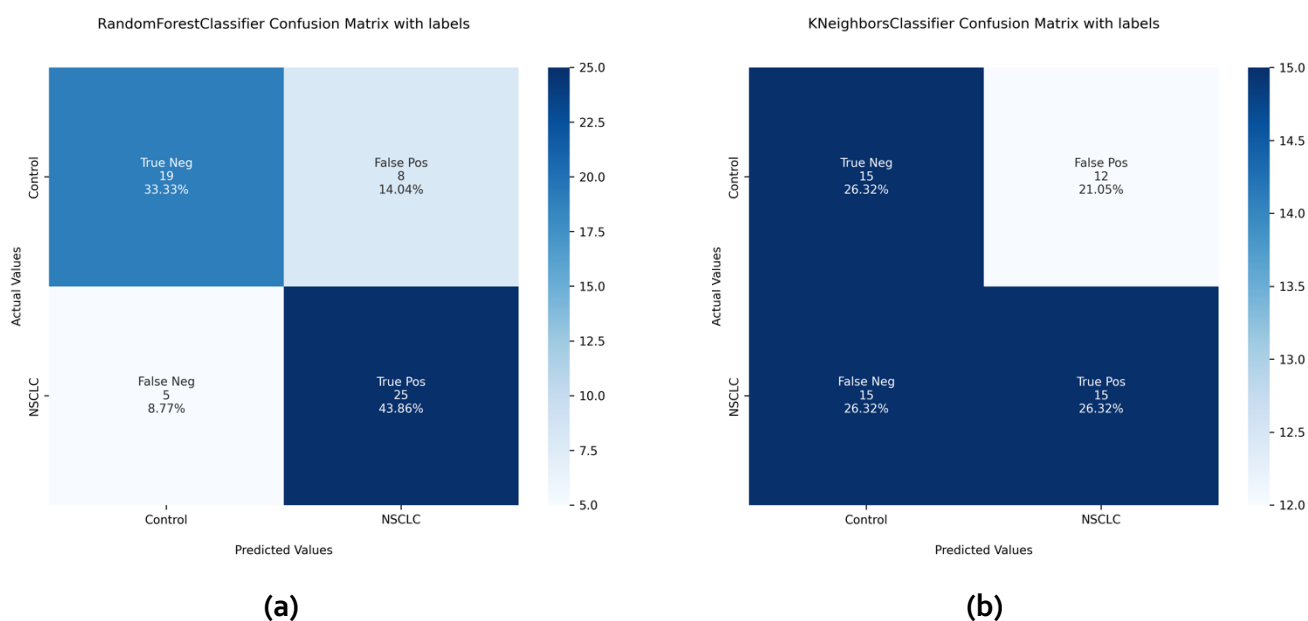


Figure S4. Confusion matrices summarizing the performance of the different classification algorithms. (a) Confusion matrix of the of the RFF classifier. (b) Confusion matrix of the KNN classifier. In both cases, 5-Fold CV was used.

RF = Random Forest, KNN = K-Nearest Neighbor, 5-CV = 5-Cross Validation, RFE = Recursive Feature Elimination.

Author Contributions

Conceptualization, C.P.V. and R.R.; methodology, C.P.V., S.G and J.W.P.B.; software, S.G.; formal analysis, C.P.V., S.G. and A.D.; investigation, C.P.V., S.G., A.G.C., D.F., M.F., J.B.A. and J.V.; resources, A.A.H, R.R. and M.M.V; data curation, C.P.V.; writing—original draft preparation, C.P.V.; writing—review and editing, M.M.V.; visualization, C.P.V and S.G.; supervision, R.R., M.M.V, A.F.H, M.H and N.Z; validation, C.P.V; project administration, C.P.V.; funding acquisition, R.R. All authors have read and agreed to the published version of the manuscript.

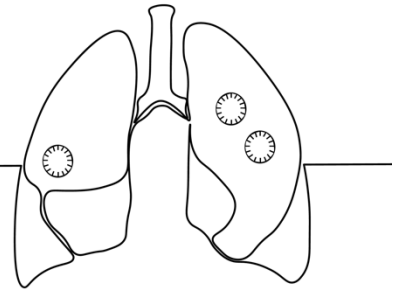
Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No 765492.

Acknowledgments

We would like to thank Neil Bertram for his language editing assistance. The investigators also wish to thank the patients for kindly agreeing to donate samples to this study. Figure 4 was created with Biorender.com.

Chapter IV: Plasma circRNA analysis for the early detection of lung cancer using nCounter



Development of a plasma circRNA signature for early-stage NSCLC detection using nCounter technology

Carlos Pedraz-Valdunciel, Giovanna Casagrande, Stavros Giannoukakos, Elizabeth Martinez, Ana Giménez-Capitán, Joselyn Valarezo, Michael Hackenberg, Andrés Aguilar-Hernández, Rui Reis, Miguel Ángel Molina-Vila and Rafael Rosell.

Manuscript in preparation.

Abstract

Background: Lung cancer is the most lethal form of cancer worldwide, showing the highest prevalence across all types of tumors. The high mortality rate attached to the malignancy is attributed to a late diagnosis, with treatment proving ineffective in most cases. For this reason, different screening programs have been established worldwide as an effort to reduce these elevated lung cancer mortality rates. Under this context, many strategies based on the analysis of the different body fluids (liquid biopsies) are currently being tested; however, the implementation of this type of analysis in screening programs remains still incipient. Circular RNAs (circRNAs) are a type of regulatory RNAs with a stable structure which can be found circulating in blood plasma. Numerous studies suggest that some circRNAs could potentially be used as lung cancer biomarkers, improving the diagnosis and treatment of the disease. Throughout this project, we will make use of the nCounter technology for the analysis of plasma circRNA expression. Methods: circRNA was purified from early-stage NSCLC patients (n=49), patients with benign nodules (n=19) and healthy controls (n=30). Differential expression and machine learning (ML) methods were performed for the development of a NSCLC signature. Results: 6 circRNAs were found dysregulated in NSCLC samples. A 64-circRNA signature selected by ML was able to differentiate NSCLC patients from controls, including those with benign nodules, with an AUC ROC of 0.90. Conclusions: Our study not only demonstrates the feasibility of using the nCounter for the study of plasma circRNAs, but also highlights the potential of a circRNA signature for the diagnosis of early-stage NSCLC.

Keywords

circRNAs; plasma; early-stage; nCounter; lung cancer; NSCLC; liquid biopsies.

1. Introduction

According to the International Agency for Research on Cancer (IARC), lung cancer still ranks as the most common type of cancer in the world, with 2.2 million new diagnosed cases just for 2020. This type of tumor is also at the foremost position in terms of lethality, with an estimated 1.8 million annual deaths worldwide [4, 159]. This high number of demises is mainly attributed to a late diagnosis, which usually accounts for 85-90% of the cases, when curative treatments are less effective [160]. Consequently, the overall 5-year survival rate of approximately 14-17% can drastically drop to less than 2-8% in advanced metastatic stages [9]. Currently, some emerging screening programs, such as the National Lung Screening Trial (NLST) in the United States, have proved effective for the early detection of lung cancer by using low dose computed tomography in high-risk populations, resulting in a decrease of 6.7% of the mortality rate [161, 162].

Lung cancer is divided into two histological types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [163], with the latter accounting for 85% of the cases. These histological types differ not only in terms of morphological features, but also in terms of clinical and molecular characteristics [164]. Consequently, investigation of cited subtypes can aid the identification of novel biomarkers with diagnostic, prognostic, and predictive value, contributing to an early detection of the disease. However, evaluation of such molecular alterations is currently being carried out on tumor tissue fragments obtained by classical invasive methods, such as tumor surgical biopsy [165]. Yet, tumor sampling may sometimes be challenging, compromising both quality and quantity of extracted biopsy specimens. Heterogeneity of resected samples may also hamper the use of this method, especially for genetic assessment [162]. Nevertheless, surgical biopsy still stands as the gold standard procedure for tumor diagnosis and characterization.

In contrast, liquid biopsies offer a minimally invasive and reproducible tool for diagnosis and monitoring of lung cancer patients, detecting tumor-related biomarkers present in the different body fluids [166]. The use of liquid biopsies was initially restricted to circulating tumor DNA (ctDNA) with the FDA approval of Cobas EGFR Mutation test v.2 CE-IVD (Roche, Basel, Switzerland) and Therascreen mutation kits (Qiagen, Hilden, Germany) to detect EGFR mutations in advanced NSCLC patients [167]. Nowadays, circulating tumor cells (CTCs) and other subcellular components such as extracellular vesicles (EVs), proteins and cell-free RNA (cfRNA), can also be used as potential diagnostic biomarkers [168-171].

Within this context, some circRNAs have recently been investigated as liquid biopsy biomarkers for the early detection of lung cancer and other solid tumors.

Although many scientific reports have demonstrated circRNA enrichment in extracellular vesicles (EVs) [58], implementation of an EV-purification step may be challenging in the clinical setting. In addition, isolation of these molecules often leads to RNA loss during this process, which may sometimes affect downstream analysis.

During this project, we investigated the expression of whole-plasma circRNAs by using the nCounter FLEX technology. Next, we analyzed plasma samples from non-cancer controls, patients with benign nodules, and early-stage non-small cell lung cancer (NSCLC) patients and performed machine learning (ML) to develop a prognostic signature of the disease.

2. Methods

2.1. Patient samples

The study was carried out in accordance with the principles of the Declaration of Helsinki, under an approved protocol of the institutional review board of Quirón Hospitals. We obtained and documented written informed consent from all the patients. A total of 49 samples from early-stage NSCLC (stages IA to IIIA) were selected from our institution, along with 49 samples from non-cancer controls (**Table 1**). Clinical information from patients and controls included age, gender, smoking status, tumor histology and stage, when applicable. All samples were de-identified before further processing for confidentiality purposes.

Table 1. Clinicopathological characteristics of enrolled patients (n=98)

Clinicopathological characteristics	NSCLC Patients (n=49)	Non-cancer controls (n=49)
Gender - no. (%)		
Male	25 (51.0)	19 (38.77)
Female	24 (49.0)	30 (61.22)
Age - yr.		
Median	65.83	51.85
Range	32-91	23-86
Smoking Status - no. (%)		
Former- or current smokers	33 (67.34)	21 (42.85)
Never smoker	13 (26.53)	20 (40.82)
Not information	3 (6.13)	8 (16.32)
Nodules- no. (%)		
Malignant	49 (50.0)	-
Benign	-	19 (19.38)
None	-	30 (30.62)
Histological type - no. (%)		
Adenocarcinoma	39 (44.82)	-
Squamous carcinoma	8 (9.19)	-
Not information	2 (2.29)	-
Tumor Stage - no. (%)		
I	21 (42.85)	-
II	5 (10.20)	-
III	20 (40.81)	-
Not information	3 (6.12)	-

2.2. Plasma processing and RNA isolation

Around 10 mL of whole blood was collected from the participants enrolled in the study using sterile EDTA Vacutainer tubes (BD, Plymouth, UK) and processed within the next two hours. Blood samples were centrifuged twice at 2000 x g at room temperature (RT) in a Rotina 380 R centrifuge (Hettich, Tuttlingen, Germany) for 10 minutes to separate plasma from red/white blood cells, platelets, and cell debris. Aliquoted plasma samples were then stored at -80°C until downstream processing.

For cfRNA purification, 1.2 mL of plasma was processed using the QIAasymphony® DSP Virus/Pathogen Midi Kit in a QIAasymphony robot (Qiagen), following the manufacturer's instructions. Final elution volume was set at 50 µL. RNA concentration was estimated the Qubit 3.0 kit (Thermo Fisher Scientific, MA, USA).

2.3. nCounter processing

The nCounter Low RNA Input Amplification Kit (NanoString Technologies, WA, USA) was used to retrotranscribe and pre-amplify 4 µL of plasma RNA in a Verity thermal cycler (Applied Biosystems, MA, USA) following NanoString's guidelines. Briefly, samples were denatured at 95°C for 10 minutes and hybridized for 18 h at 67°C. Our custom-made nCounter panel (including 78 circRNAs, 6 linear reference genes and 4 mRNAs) was used to analyze plasma-derived pre-amplified cDNA according to the manufacturer's instructions. RCC files containing data outputted by the NanoString nCounter Flex System (NanoString Technologies) from each run were exported to the nSolver Analysis Software (version 4.0.70, NanoString Technologies, Seattle, WA, USA).

2.4. Differential expression analysis

Raw count nCounter values were exported to Microsoft Excel (version 16.40, Microsoft, Redmond, WA, USA) using nSolver Analysis Software. The background was calculated for each sample as (geo)mean \pm 2SD of the negative probe counts (NCs) Raw counts lower than the background were automatically excluded from further analysis. The raw circRNA counts were normalized using the total number of counts of the sample and multiplied by 10,000. Differential expression analysis was performed comparing the mean of the normalized counts for each circRNA in the early-stage NSCLC vs. non-cancer controls. The circRNAs with a fold change >1 and p-value < 0.05 were considered as differentially expressed.

Data pre-processing and normalization for signature development Raw RCC-formatted data files were exported from the nSolver Analysis Software (NanoString Technologies). R (version 4.0.3, R Core Team and the R Foundation for Statistical Computing, Vienna, Austria) and R studio (version 2021.09.0, RStudio PBC, Boston, MA, USA) were used for pre-processing and normalization analysis of the imported files. Initial evaluation of the quality and integrity of the RCC data was performed using the NanoStringQCPro (version 1.22.0) package. During this process, we looked for potential outliers based on the performance of standard control metrics provided by NanoString, such as Imaging, Binding Density, Positive Control Linearity, and Limit of Detection. After this first pre-analytical step, samples were subjected to supplementary exploratory examination, including Principal Component Analysis (PCA) and interquartile range (1.5 IQR rule) analysis. Samples found as outliers by both methods were then excluded from downstream analyses. NCs were employed to exclude lowly expressed circRNAs with excessive background noise. The arithmetic mean of the NC \pm 2SD was subtracted from each endogenous circRNA for each sample. Any transcript scoring a value below 0 in more than 75% of the analyzed samples was then excluded from

further analysis. PCA plot was then used to reassess the data after the aforementioned filtering step. Technical variability correction and normalization were performed using the RUVseq/RUVg function (version 1.24.0) and DESeq2 (version 1.30.1) packages (RUVseq-DESeq2). First, the RUVg function was used to estimate the unwanted variation among samples based on the DE genes. DESeq2 and edgeR (version 3.32.1) performed a first pass DE analysis and the intersected least significant genes (with adjusted p-value above 0.1) were used as “in-silico empirical” negative controls. DESeq2 was then utilized with default parameters along with the RUV factors to perform the normalization of the raw filtered data. The normalization performance was assessed using the standard relative log expression (RLE) plot.

2.5. Machine Learning (ML) for signature development

Recursive Feature Elimination (RFE) was used to perform feature selection, and the 3-cross validation (3-CV) algorithm was applied across the normalized data previously generated by RUVseq-DESeq2. During this process, the dataset was randomly split into k-folds ($k = 3$), being $2/3$ of the data used to train the model, while the remaining $1/3$ was used to test its behavior. The number of features yielding best performance after cross-validation was automatically selected to shape the prognostic signature. To test the predictive power of the selected signature, k-nearest neighbor (KNN) and random forest (RF) models were built using different parameters. The model showing the highest ROC AUC value was selected as the final model. Signature scores for each sample were obtained from 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 performed.

3. Results

3.1. CircRNA expression in plasma samples

Plasma samples from 98 individuals were analyzed, including 49 early-stage NSCLC and 49 controls (19 of which were presenting benign lung nodules). Different levels of circRNA expression could be observed in all samples included in the study (**Figure S1**).

An overall statistically significant increased number of raw counts was detected in NSCLC samples versus controls (Mann-Whitney U test, p -value = 0.0285, **Figure 1a**). In addition, an average of 42 ± 12 circRNAs were detected in controls and 41 ± 13 in the NSCLC cohort (Mann-Whitney U test, p -value = 0.3993, **Figure 1b**). Individual assessment of each of the circRNAs included in the panel indicated that 64 of them were detected in at least one NSCLC sample, and 66 in at least one non-cancer control. Out of all the circRNAs detected, only 3 were exclusive to NSCLC and 5 to the control cohort (**Figure 1c**).

Additionally, we assessed the repeatability of the nCounter assay by submitting to nCounter duplicates of two preamplified plasma samples (one NSCLC and one control samples). As a result, a strong correlation between the normalized counts was found between the duplicates, represented by a Pearson's $r = 0.99$, $p < 0.001$ for the NSCLC, and 0.96 , $p < 0.001$ for the control respectively (**Figure 2a**).

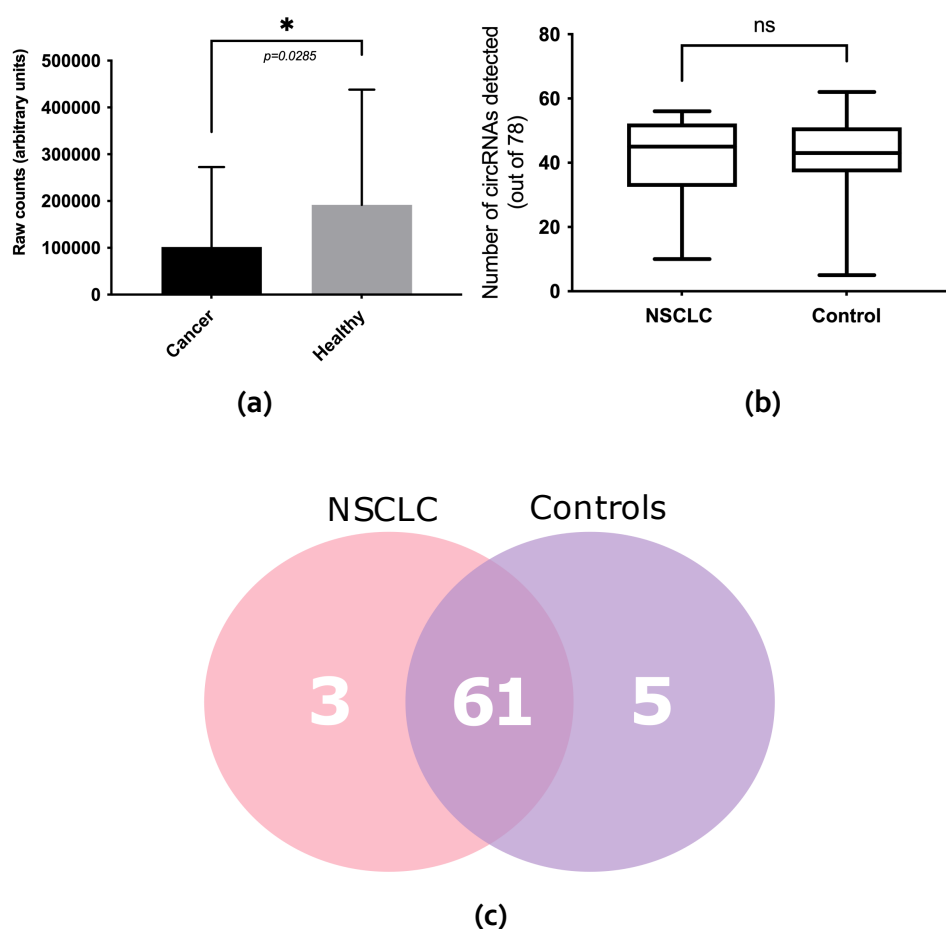


Figure 1. Plasma circRNA detection by nCounter. **(a)** Total number of raw counts in NSCLC patients and controls. **(b)** Total number of circRNAs detected in each cohort, out of the 78 circRNAs included in the nCounter panel. **(c)** Venn diagram displaying those circRNAs identified in at least one of the NSCLC patients and non-cancer controls, along with those shared by both cohorts.

Next, we assessed reproducibility of our nCounter assay. To this end, we retrotranscribed and pre-amplified duplicates of the same circRNA samples (one NSCLC and one control samples) on two independent nCounter reactions performed on different days and compared the results. Consequently, a strong correlation was found between normalized counts for each expressed circRNA obtained in the different assay, represented by a Spearman's $r = 0.96$ and 0.93 for NSCLC and control samples respectively ($p < 0.001$) (**Figure 2B**).

3.2. Comparison of circRNA expression in EVs and total plasma

Since we previously reported circRNA detection in EV preparations using our circRNA nCounter panel [172], we tested 71 matched plasma samples (40 lung cancer and 31 non-cancer controls) from the previous study and further assessed the number of circRNAs detected in each of the two biosources for each cohort. In total, 69 out of the 78 circRNA included in the nCounter panel could be detected in at least one individual of any of the analyzed cohorts (**Figure 3**). A total of 56 different circRNAs could be found in at least one

EV sample, compared to the 67 that could be detected in plasma. In this regard, 50 circRNAs were shared by both, EV and plasma samples, while 13 were uniquely detected

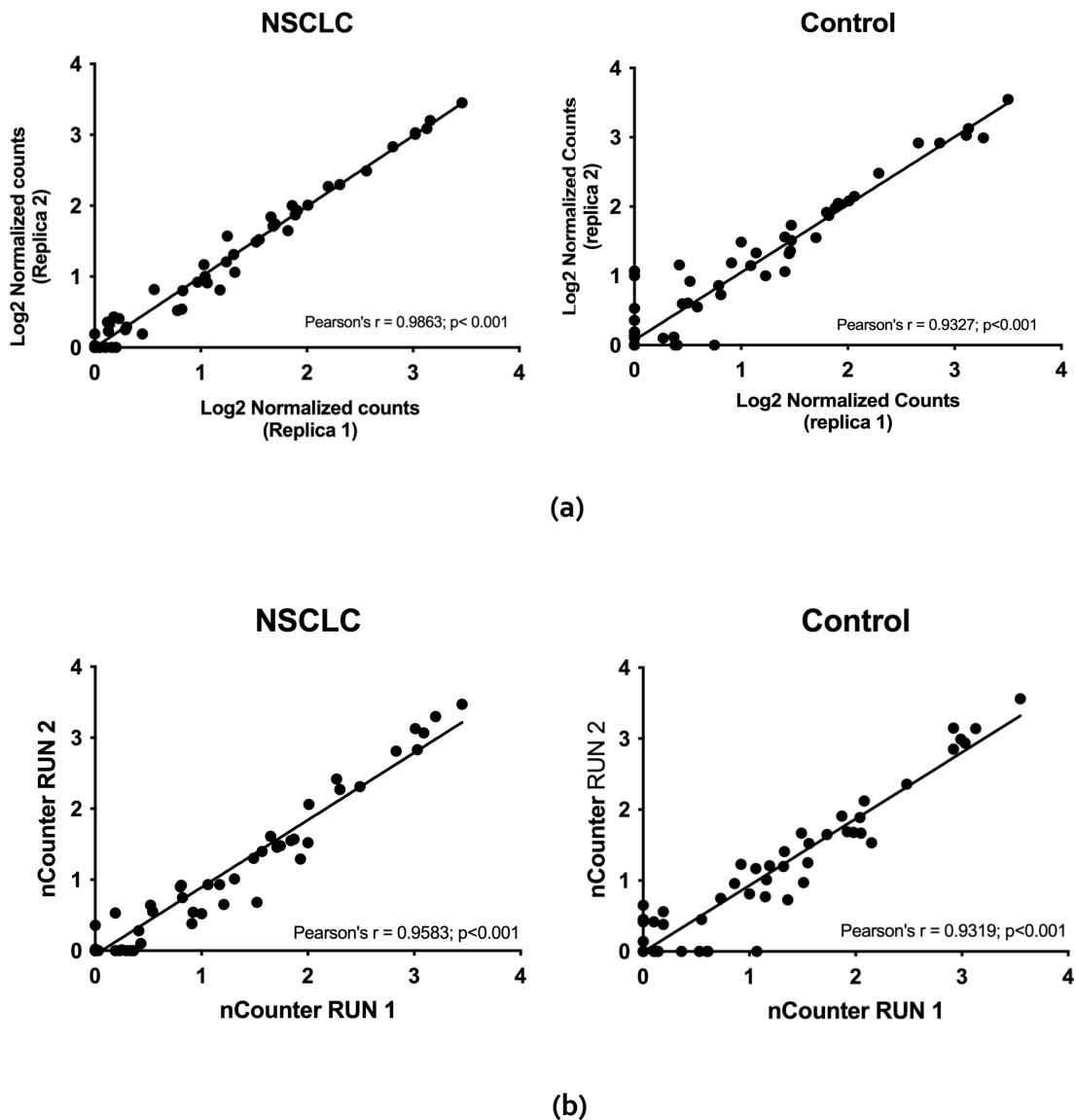


Figure 2. Reproducibility and repeatability assays. **(a)** Correlation of the two NSCLC and control technical nCounter duplicates. Pearson's correlation coefficient is indicated. **(b)** Correlation of the results obtained in two independent nCounter assays of the same NSCLC and control samples. Pearson's correlation coefficient is indicated

in at least one plasma sample. Interestingly, 3 circRNAs (circPTPRM, circRANGAP1, and circGNA14) were only observed in plasma samples from some NSCLC patients (**Figure 3**). Overall, a greater number of circRNAs could be detected compared to EV samples (67 vs 56).

3.3. Differential expression analysis of circRNAs in NSCLC plasma samples

After assessing the robustness of our protocol, differential expression analysis was performed to study the circRNA expression levels in the samples included in the study. As a result, 4 circRNAs (namely circSMAD, circCOL11A1, circCHST15 and circFUT8) were found upregulated in NSCLC, whereas only two circRNAs (circACP6, and circLYPLAL1) were

significantly downregulated in plasma of these patients when compared to the control cohort (Figure 4).

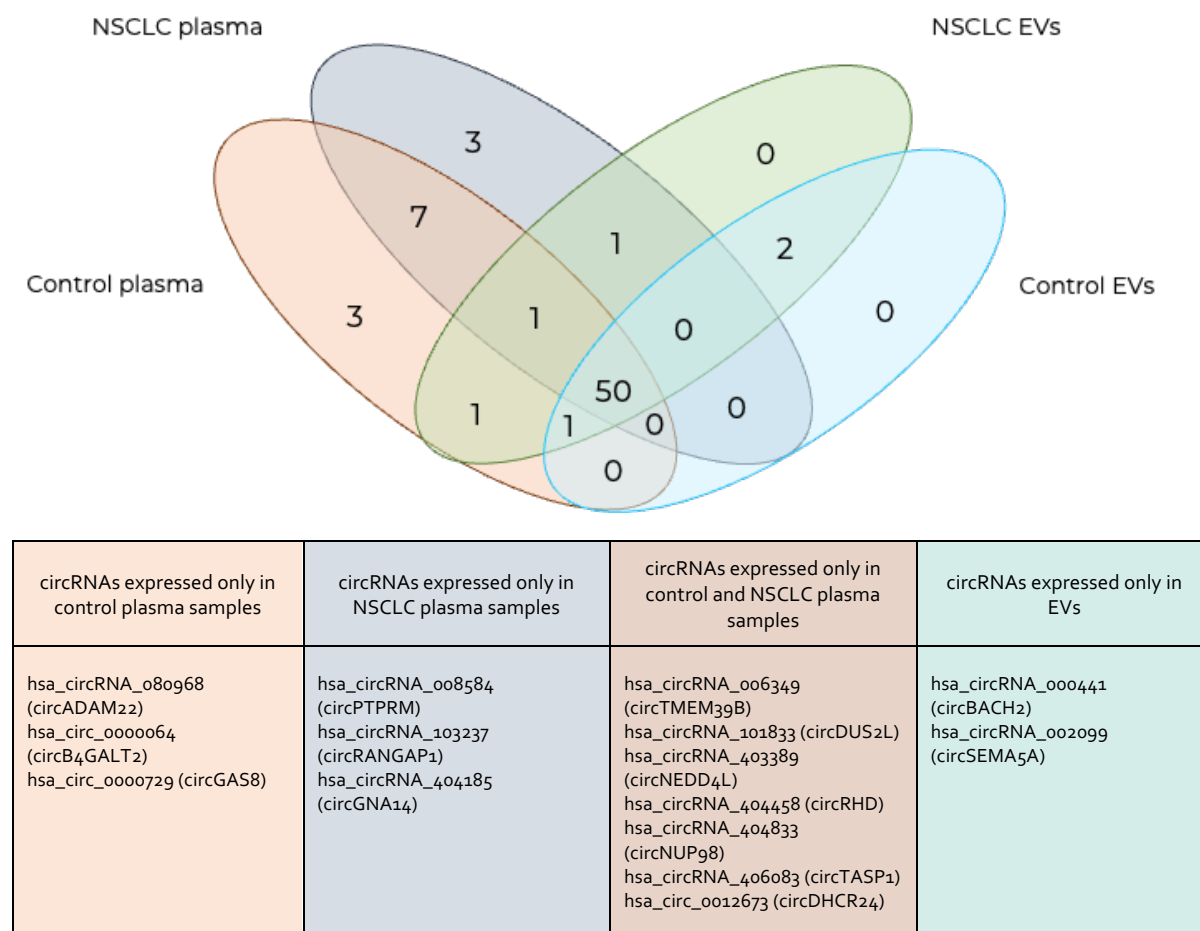


Figure 3. Venn diagram highlighting those circRNAs identified in at least one of the NSCLC patients and/or non-cancer controls, both in EVs and matched plasma samples. Overlapping of expressed circRNAs by the different groups are also indicated.

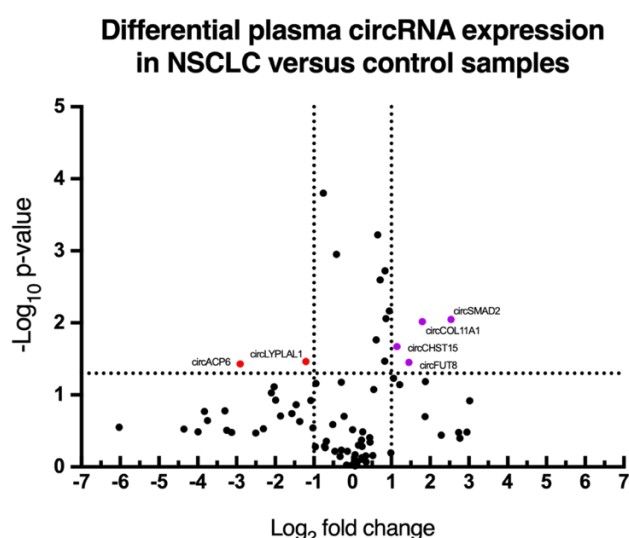


Figure 4. Differential expression analysis in plasma circRNAs from early-stage NSCLC (n=49) patients versus control (n=49).

3.4. Development of a circRNA signature for the detection of early-stage NSCLC

As previously described in the methods section, different normalization methods including DESeq2, RUVSeq, edgeR and their combination were tested to select the method that best adjusts to our circRNA raw data. As a result, RLE plots indicated a superior performance of RUVSeq-DESeq2 over the rest of combinations (**Figure 5a**, **Figure S1**). In addition, PCA plot did not show any significant batch effect or cluster separation (**Figure 5b**). Lastly, 1.5 IQR analysis identified 10 out of the 98 samples as possible outliers. Consequently, these samples were excluded from further analysis.

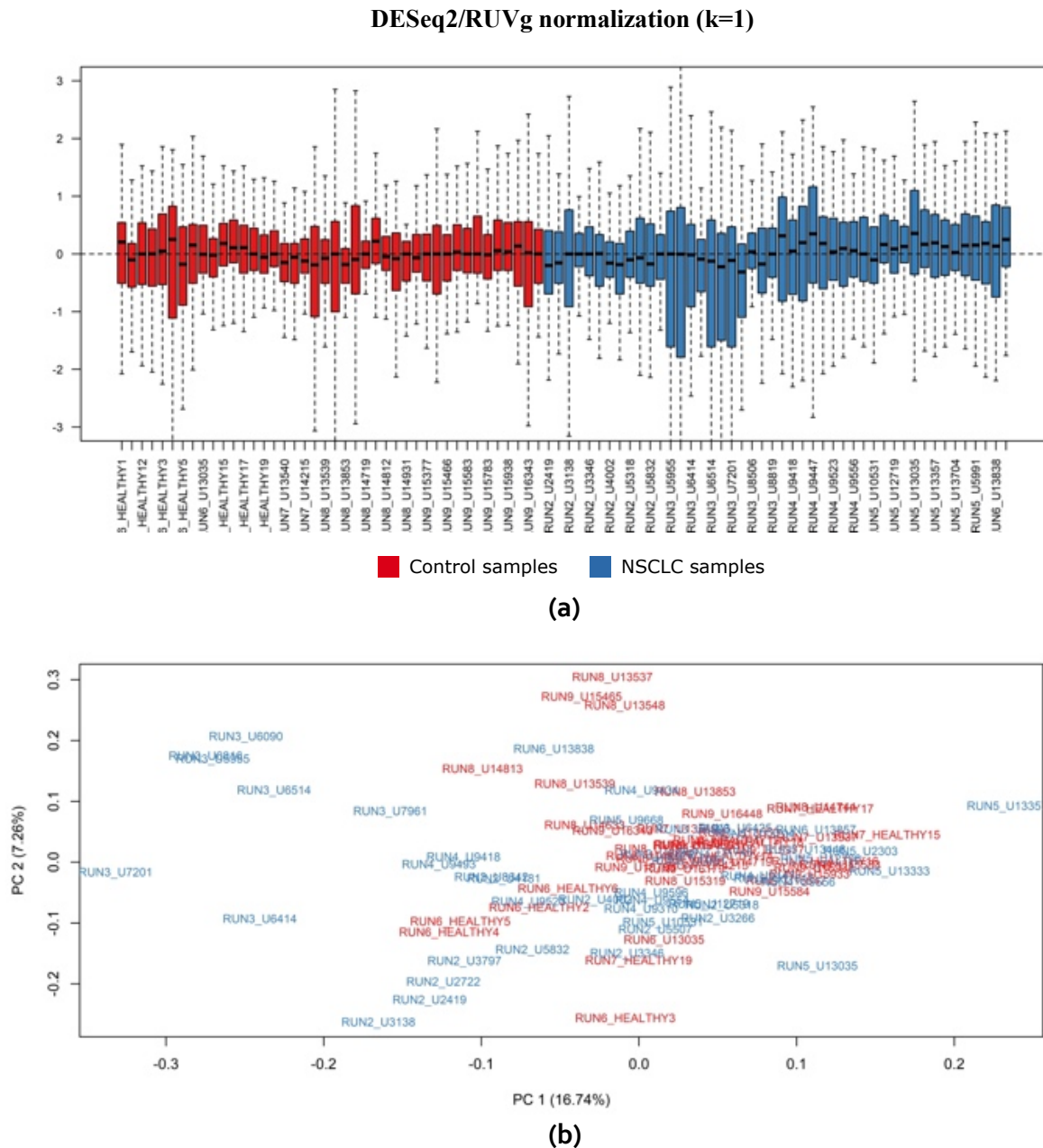


Figure 5. Data normalization and sample analysis. **(a)** RUVSeq/DESeq2 RLE plot of normalized data ($k = 1$). **(b)** PCA plot of the transformed raw data.

Next, ML analysis was performed in early-stage NSCLC and control samples. Hence, RFE algorithm provided a signature that included 64 circRNAs (**Table S1**), delivering a ROC AUC of 0.90 and 0.782 with RF and KNN respectively (**Figure 6a**). As a result, RF was selected as the final model. In addition, assessment of the signature scores allowed a statistically significant separation of the cohorts (Mann-Whitney U test, $p < 0.0001$; **Figure 6b**). The sensitivity and specificity of the RF signature were of 82.22% (CI = 67.95% - 92.00%) and 74.42% (CI = 58.83% - 86.48%) respectively, outperforming the KNN classifier (**Table 2**). Based on the algorithm determined cutoff value of 0.5, 69 out of the 88 cases were correctly classified, resulting in an accuracy of 78.41% (CI = 68.35% - 86.47%) with the final model (**Figure 6c**). Amongst the 19 wrongly classified samples, 8 corresponded to NSCLC patients and 11 to controls. Further analysis on the clinicopathological characteristics of these individuals showed that only 1 of the misclassified controls presented a benign nodule. In addition, false negative samples were found to be of varied NSCLC stages (4 IIIA, 2 IB, and 2IA) possibly indicating an independency of the ML-generated signature from tumor stage.

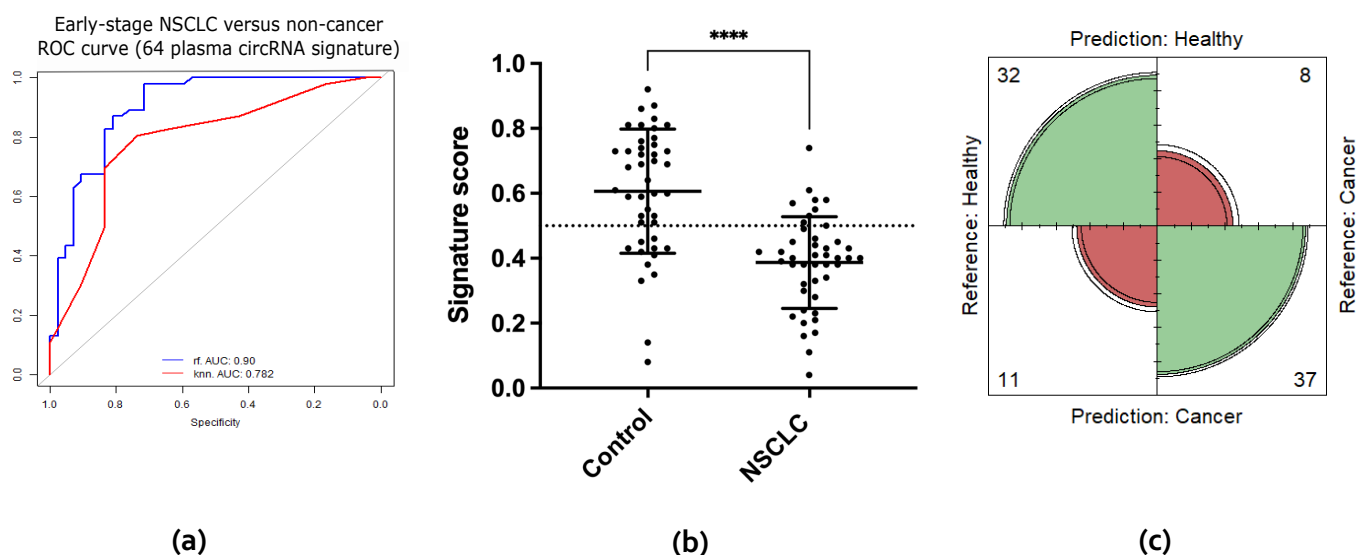


Figure 6. Machine learning analysis of plasma circRNA samples. **(a)** Area under the ROC curve of the 64 circRNA signature using RFE with RF and KNN models. **(b)** Signature scores of control samples versus NSCLC samples ($p < 0.0001$ in a two-tailed Mann-Whitney U test). **(c)** Confusion matrix based on the RF classification scores.

4. Discussion

Lung cancer is frequently diagnosed at a late stage when treatment options are not as effective anymore, which in turn leads to a high mortality rate. The development of a liquid biopsy-based assay for the detection of early-stage lung cancer could benefit these patients, reducing this elevated mortality rate drastically [173].

In addition, pulmonary nodules can routinely be found on imaging studies, particularly during multidetector computed tomography (MDCT). Advances in this technology and positron emission tomography (PET) have improved nodule characterization. However, besides continuous advancements in this field, many nodules remain uncharacterized and require further assessment for a definitive diagnosis, which often involves either invasive and limited procedures, or time-consuming follow-ups [16].

In the current study, we used the nCounter technology to analyze the expression of plasma circRNAs to further develop a circRNA signature that could discriminate between early-stage NSCLC patients and control individuals, including benign nodules.

Table 2. Precision assessment of the ML generated circRNA signature with RF, and KNN. The 95% CIs are indicated.

Model	RF	KNN
No. concordant samples	69	63
No. discordant samples	19	25
AUC ROC	0.90	0.782
Accuracy	78.41% (CI = 68.35% – 86.47%)	73.68% (CI = 63.65% – 82.19%)
Sensitivity	82.22% (CI = 67.95% – 92.00%)	72.92% CI = 58.15% – 84.72%)
Specificity	74.42% (CI = 58.83% – 86.48%)	74.47% CI = 59.65% – 86.06%)
PPV	77.08% (CI = 66.49% – 85.08%)	74.47% (CI = 63.47% – 83.04%)
NPV	80.00% (CI = 67.57% – 88.48%)	72.92% (CI = 62.17% – 81.52%)
Cohen's κ	0.567 (CI = 0.358 – 0.775)	0.428 (CI = 0.219 – 0.637)

As it has been widely described that circRNAs are enriched in plasma EVs [174], we previously described a full protocol for the analysis of EV-derived circRNAs for the generation of NSCLC signatures [172]. However, isolation of EV-RNA may sometimes be associated with some hurdles such as RNA loss, extra cost, or additional hands-on time. To assess the feasibility of circRNA nCounter analysis directly from plasma samples, we adapted our protocol to this biological liquid, and compared results to those obtained with matched EV-RNA samples. As a result, more circRNAs could be detected when using whole plasma instead of EV preparations as initial input, with the exception of 2 circRNAs (circBACH2 and circSEMA5A). Yet, these 2 circRNAs were neither significantly expressed nor confirming part of previously described EV-circRNA signature [172]. Therefore, we concluded that the overall circRNA expression detected through nCounter analysis of plasma samples was sufficient for downstream differential expression analysis and subsequent ML-based circRNA signature development.

Next, we performed differential expression analysis and found 4 circRNAs upregulated and 2 downregulated in NSCLC when compared to the control samples (**Figure 4**). Interestingly, circFUT8 was previously found upregulated by our group both in FFPE tissues and cell lines [137]. Although further investigation is crucial to unveil the biological significance of these findings, and the specific origin of these transcripts within the plasma, our preliminary data highlight the biomarker potential held by these circRNAs.

Then, ML was performed in order to find a signature for the early detection of NSCLC. As a result, the final model selected a 64-circRNA signature that achieved an AUC ROC of 0.90 and accuracy of 78.41% (**Table 2**). In addition, our signature was able to correctly classify patients with benign nodules as non-cancer, with only one false positive in the presented study. These results lay the grounds for the development of state-of-the-art circRNA-based companion diagnostics assays that, together with MDCT, may aid medical decisions in prospective lung screening programs.

This is a proof-of-concept study. Therefore, several limitations need first to be addressed before further validation of our plasma-based circRNA signature. Primary, recruitment of bigger cohorts is necessary not only to reassess cited signature, but to feed subsequent training and validation phases. Also, the inclusion of symptomatic controls including other inflammatory diseases such as COVID-19, pneumonia, or pneumonitis is recommended in order to ensure the cancer specificity of the signature. Finally, only lung cancer patients were included in this study; therefore, we do not know if the current signature could also detect other types of cancer. Inclusion of other malignancies should be considered in prospective investigations to shed light on this matter.

5. Conclusions

Presented work demonstrates the use of the nCounter FLEX system for circRNA expression analysis in blood plasma. ML provided a 64-circRNA signature predictive of early-stage NSCLC. To the best of our knowledge, this is the first nCounter-based plasma circRNA assay for the detection of early-stage NSCLC. Further studies of larger cohorts are warranted in order to determine the clinical applicability of the signature

6. Supporting Information

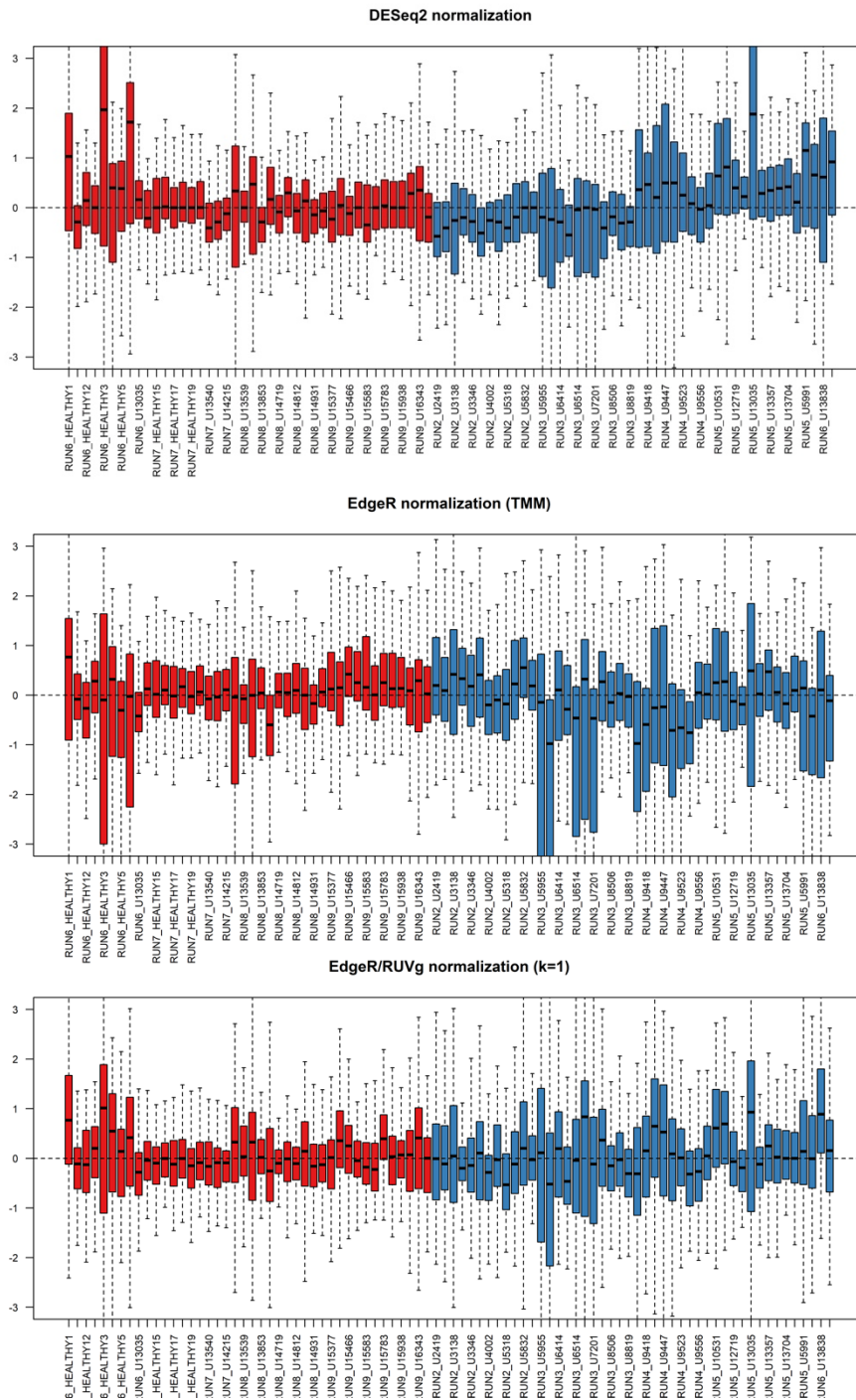


Figure S1. Assessment of the different normalization processes by RLE plot analysis, including DESeq2, edgeR, and edgeR-RUVg methods ($k = 1$).

Table S1. CircRNAs confirming the ML-generated signature of early-stage NSCLC

hsa_circ_0003148	hsa_circ_0000437	hsa_circ_0004458
hsa_circ_0001380	hsa_circRNA_103237	circ_001569
hsa_circ_000153	hsa_circ_0039161	hsa_circRNA_404185
hsa_circ_0004417	hsa_circRNA_404833	hsa_circRNA_406483
hsa_circ_0006276	hsa_circ_0001238	hsa_circRNA_006296
circ_C1orf116	hsa_circRNA_001937	hsa_circRNA_002099
hsa_circ_0001445	hsa_circ_0003958	hsa_circ_0043256
hsa_circRNA_001640	hsa_circ_0007037	hsa_circ_0001083
hsa_circRNA_080968	hsa_circ_0004062	hsa_circRNA_101367
hsa_circRNA_404643	circFARSA	hsa_circRNA_404458
hsa_circRNA_000997	hsa_circRNA_000317	hsa_circ_0001495
hsa_circRNA_101833	circRNA_HIPK3	hsa_circRNA_401977
hsa_circRNA_001288	F-circEA1 (Variant 1)	hsa_circRNA_006411
hsa_circ_0037007	hsa_circRNA_103775	hsa_circRNA_000441
hsa_circ_0005139	hsa_circRNA_103809	hsa_circ_0001320
hsa_circ_0005962	hsa_circRNA_403389	hsa_circRNA_400294
hsa_circRNA_405718	hsa_circ_0000847	hsa_circRNA_104640
hsa_circ_0003941	hsa_circ_0000064	hsa_circ_0002360
F-circEA3 (Variant 3)	hsa_circRNA_406083	hsa_circ_0000566
hsa_circ_0012673	hsa_circRNA_006349	hsa_circ_0009128
hsa_circ_0001675	hsa_circRNA_104329	hsa_circ_0000264
hsa_circRNA_103727		

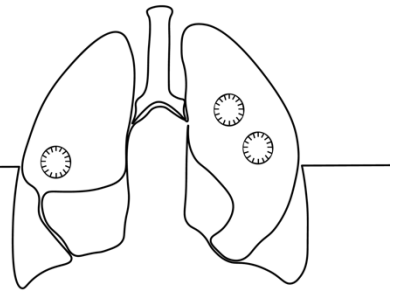
Acknowledgements

We would like to thank the patients and donors for kindly providing samples to this study. We also thank all the physicians who collaborated by providing clinical information.

Financial support

This project has received funding from a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No. 765492.

Chapter V: Discussion



While former investigations considered circRNAs as background noise derived from defective splicing, accumulating evidence demonstrates they are functional players, acting as regulators in many biological processes. Additionally, many studies highlight their implication in carcinogenesis and tumor progression, reporting also their contribution to mutant glycolysis, lipogenesis and lipolysis, glutaminolysis, and oxidative respiration [51]. Whilst many scientific reports position these circular transcripts as plausible lung cancer biomarkers and promising candidates for seeking therapeutic and prevention strategies, limitations of current high-throughput methods hamper the development of clinically applicable assays.

The Nanostring nCounter may stand as an alternative to current quantification methods, facilitating the transition of circRNA biomarkers from bench to bedside. Besides its capacity to work with very low amount of highly degraded samples, other advantages of this technology include the ability to perform multiplex analysis, minimal hands-on time, and subsequent user-friendly data analysis. As a result, many translational investigations have embraced this technology performing transcriptional research in many biomaterials, including liquid biopsies [77, 133]. However, to our knowledge, no other group has reported the use of nCounter for the study of circRNAs in lung cancer as yet.

Based on these premises, this thesis focused on testing the use of the nCounter FLEX platform for the study of circRNA expression in lung cancer specimens. As a second objective, we aimed to explore the biomarker potential of some of these circular transcripts by generating ML-based signatures for the detection of lung cancer.

Since there is no commercially available nCounter circRNA panel, we designed a customized panel targeting the backsplicing junction of 78 different circRNAs. In addition, we also included 10 mRNAs to be used as internal controls [137]. Selection of the circRNAs conforming the assay was based on scientific reports showing dysregulation of cited transcripts in either lung cancer biopsies or cell lines. Next, we tested our circRNA assay by performing different experiments with RNase R in cell line-derived RNA samples to assess the specificity of designed probes towards the circular form of the different genes included in the panel. As a result, by assessing the number of counts for each circRNA in RNase R treated samples, we observed an overall enrichment for most circRNAs, whereas the linear mRNA targets that we used as control in our study were degraded to various extents as expected.

As some circRNAs have proved to be sensitive towards the effect of RNase R [111] and the efficiency of this enzyme to remove linear RNA may range depending on different factors [45, 113], we cannot claim the solely recognition of circRNA based on this data. However, the circRNA enrichment achieved after RNase R treatment agrees to what we were expecting, strongly supporting the specificity of our nCounter probes towards these circular transcripts. Also, the specific configuration of designed probes flanking the circRNA junction sites allows the correct binding of capture and reporter probes. Conversely, in case of binding to a linear counterpart, the probes would bind divergently therefore preventing the production of a target-specific signal [69]. Lastly, we assessed expression levels of 3 circRNAs (circEPB41L2, circSOX13, and circBNC2) in different FFPE lung cancer and control samples by nCounter and RT-qPCR. As a result, similar expression patterns could be observed for both quantification methods, with subsequent Sanger sequencing unveiling the corresponding junction sites. Although RNase R treatment can result useful for the screening or validation of circRNAs, its use should be avoided in circRNA expression studies since, as previously mentioned, variability in the digestion efficiency of this enzyme may bias circRNA expression

quantification. In addition, purification of RNA from FFPE material may involve both mechanical and chemical RNA breakage, making any circRNA susceptible to the effect of the RNase R. Therefore, enrichment of circRNA by RNase R treatment should be restricted only to RNA extracted from fresh or frozen specimens.

During this part of the project, we did not only validate the use of our nCounter assay for the study of circRNAs in solid lung cancer specimens, but also, we tested our protocol to perform expression analysis in FFPE lung tissue samples from 53 lung cancer patients and 16 non-cancer controls. As a result, a cluster of differentially expressed circRNAs, namely circFOXP1, circRUNX1, circRHOQ, circACACA, circC1orf116, circCHD9, circFUT8, circEPB41L2, circSOX13 and circBNC2 were found dysregulated in lung cancer specimens, even at early stages of the disease as it is the case of the last six. In addition, we examined different ML methods to explore the diagnostic potential of a plausible circRNA signature of lung cancer. As a result, a 4-circRNA signature (including circCORO1C, circEPB41L2, circSOX13 and circBNC2) was able to discriminate early-stage lung cancer specimens from controls with an AUC ROC of 0.98, resulting in only 4 out of the total 69 samples being incorrectly classified. Interestingly, 3 out of the 4 circRNA conforming the signature were previously found dysregulated, indicating the high potential that cited circRNAs hold as biomarkers. This first part of the research gave us the opportunity to develop and test our circRNA nCounter assay. In addition, we unveiled the biomarker potential of a cluster of circRNAs, which may open new lines of investigation for seeking new therapeutic strategies. This was a proof-of-concept study, therefore, validation of above presented ML-signature in solid biopsies was not pursued, due to the hurdles associated to this invasive procedure.

Majority of life-threatening cancers, including lung cancer, are nowadays detected too late, when outcomes are often lethal. A shift in these mortality rates is observed when the tumor is detected at early stages, when more therapeutic options are available [175]. Different strategies are currently being investigated for the development and subsequent implementation of screening programs. The use of low-dose CT in different annual screening campaigns has proved to be beneficial for the early detection of lung cancer, as reported by several randomized clinical trials [15-18]. In particular, the Dutch-Belgian lung cancer screening trial confirmed a reduction in lung cancer mortality of 24% and 33% in men and women respectively, attributed to the implementation of volume CT screening [18]. Nevertheless, some limitations endorsed to this practice, including the number of false positive cases that lead to unnecessary follow-up tests and invasive procedures, need first to be resolved before seeing this approach globally implemented.

Shedding light on this matter, the analysis of protein and nucleic acids circulating in blood (or other body fluids) has risen in the last years as an option to complement current strategies, providing a minimally invasive tool for the early detection of cancer.

The burst of NGS and ML advances has contributed to the development of few detection tests, based on the analysis of liquid biopsies. In this regard, the first-of-kind multi-cancer early detection blood test, the Galleri® test developed by Grail (Illumina, CA, USA), can be found commercially available since 2021, allowing the detection of 50 different types of cancer. This technology is based on the identification of anomalous methylation patterns present on the cfDNA released by tumor cells, which allows classification of the cancer type with an 88.7% of accuracy, and a specificity and sensitivity of 99.5% and 67.6% respectively for stage I-III lung cancer patients [176] (**Table 1**). Besides this test, many groups are currently studying the biomarker potential of cfDNA for the detection of lung cancer [14, 77]. However,

the proportion of tumor-derived DNA present in the plasma of early-stage lung cancer patients is rather low, which slows down the development of further clinically implementable tests for this purpose. Consequently, investigation of other liquid biosources is necessary to allow the detection of lung cancer at early stages. To this end, Würdinger and colleagues have just published the validation of a pan-cancer signature (ThromboSeq) able to identify 18 different types of cancer, including NSCLC [177]. The ML-developed signature is based on blood platelet-derived mRNA expression and demonstrates classification of early-stage NSCLC with decent accuracy (**Table 1**).

After successfully validating the nCounter technology as a useful platform for the multiplex analysis of circRNAs in solid lung cancer specimens and cell lines, we explored the use of this platform for the study of circRNAs in liquid biopsy samples from lung cancer patients. Based on many publications reporting circRNA enrichment in human EVs [54, 82, 178], we initially designed a pilot study whereby not only we established but also validated a protocol for the analysis of circRNAs from plasma EVs. Then, we applied it for the development of a ML learning signature able to discriminate stage I-IIIa lung cancer patients (n=36) from controls (n=30). Resulting 10-circRNA signature was able to discriminate lung cancer patients from non-cancer individuals, with an accuracy of 86% and sensitivity and specificity of 90% (CI = 73.47% – 97.89%) and 81% (CI = 61.92.1% – 93.70%) with the final model.

While presented results were encouraging to further validate our signature, purification of EVs, even with the final method of choice, may result difficult to implement in the clinical setting due to different aspects such as the additional cost associated to this extra step, the time dedicated to it, or the RNA-loss related to these types of procedures.

Consequently, in Chapter four, we tested nCounter in whole plasma in order to assess if first, we would pick up circRNA expression levels comparable to those attained to EV assessment; and second, develop a circRNA-signature predictive of early-stage lung cancer. Plasma samples of early-stage lung cancer patients (n=49) and non-cancer controls (n=49), including individuals with benign nodules (n=19/49) were analyzed for this purpose.

As a result, an equivalent number of circRNAs was observed in plasma compared to EVs, finding non statistically differences between the cancer and control cohorts (42±12 circRNAs in controls vs. 41±13 in the NSCLC cohort). In fact, 50 out of the 68 circRNAs were found in at least one individual of the EV or plasma samples. In addition, 7 circRNAs were present in at least one of the plasma samples, but could not be detected in matched EV samples, indicating a possible advantage of plasma samples over EVs when it comes to circRNA detection using nCounter.

Next, differential expression analysis of plasma samples revealed a cluster of 6 differentially expressed circRNAs, including circFUT8. This circRNA was recently reported to promote cancer proliferation, invasion and migration via regulation of the miR-944/YES1 pathway [119]. Additional *in-silico* investigation unveiled two miRNAs, hsa-miR-1305 and hsa-mir-186, that could potentially bind to circFUT8 in more than one binding site. While no data has been reported regarding hsa-miR-1305, hsa-mir-186 was previously described as an inhibitor of cancer proliferation and metastasis [117, 118].

Interestingly, we also found circFUT8 upregulated in lung cancer cell lines and FFPE tissues (including early stages), but not EV samples. This suggests that analysis of plasma may be a more accurate representation of the tumor carrying a higher load of tumor-derived nucleic acids. In contrast, in the EV preparations, the number of vesicles released by the normal cells may be higher than tumor cells, masking any signal coming from the tumor. Further investigation on the vesicular origin would be interesting to decipher this conundrum.

Finally, assessment of different ML techniques generated a signature to discriminate early-stage lung cancer with an accuracy of 78.41 (CI = 68.35% – 86.47%), sensitivity of 82.22% (CI = 67.95% – 92.00%) and specificity of 74.42% (CI = 58.83% – 86.48%). In addition, most samples from control individuals expressing benign nodules were correctly classified, with the exception of one that was categorized as cancer. From those lung cancer samples misclassified as controls, no pattern in tumor stage was found indicating that the power of the presented signature does not rely on tumor stage.

Although the inclusion of much bigger cohorts is necessary for both training and validation of presented plasma circRNA signature, precision assessment elements such as accuracy, sensitivity or specificity highlight its potential clinical utility (**Table 1**).

Table 1. Precision assessment of the ML generated circRNA signatures, thromboSeq and Galleri® tests.

Model	circRNA EV-5CV-ETC	circRNA Plasma-3CV-RF	Platelet signature (ThromboSeq)	Galleri®
AUC ROC	0.86	0.90	0.94 (Including all stages)	Not information published
Accuracy	86%	78.41%	50% (Stage I) 70 (Stage II) 63% (Stage III)	88.7%
Sensitivity	90% (Stage I-III A)	82.22% (Stage I-III A)	Not information published	21.9% (Stage I) 79.5% (Stage II) 90.1% (Stage III)
Specificity	81%	74.42%	99% (Asymptomatic controls) 78% (Symptomatic controls)	99.5%

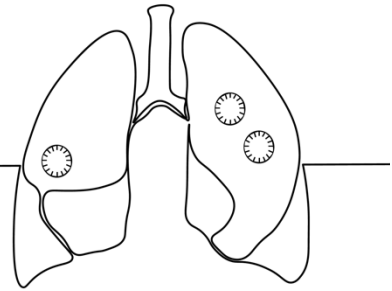
The right selection of a cohort is essential for the correct scientific design of a study. One of the limitations described in Chapter two was the hurdle of collecting non-tumor control samples due to the invasiveness associated to the acquisition of lung tissue. However, we tried to compensate this issue including individuals representative of other inflammatory diseases to ensure that the selected signature was specific of lung cancer.

Transitioning to liquid biopsies, the number of control samples was not a restraint in the study of EV-circRNAs. Yet, the disparity on the median ages of the cohorts introduced some challenges in the statistical analysis. Although multivariate analysis could finally demonstrate that classification accuracy of the EV-circRNA signature was based solely on cancer status, the lack of > 60-year-old donors in the control cohort was certainly a limitation; therefore, the inclusion of comparable cohorts in terms of age should be taken into consideration in future studies.

Lastly, in the plasma circRNA study, the clinicopathological characteristics of enrolled patients and non-cancer controls were in balance. We also included a sub-cohort of patients presenting benign nodules to ensure discrimination of cancer from this type of lumps. This would allow combination of our signature with low-dose CT scan to be implemented in

prospective screening programs, which could help in reducing the amount of false positive cases. The inclusion of symptomatic patients with other inflammatory diseases as part of the control cohort was not considered in this initial phase. However, it is already considered for the training and validation phases of this study.

Chapter VI: Future perspectives



This PhD thesis has validated the use of the nCounter FLEX platform for the multiplexed study of circRNAs in lung cancer using a custom circRNA nCounter assay. Due to the lack of previous experience, limitation in time and budget constraints, we performed an exhaustive scientific literature research in order to include the 78 circRNAs conforming the actual nCounter panel, which proved enough for the purposes of this project. However, the evidence provided in this thesis may create new horizons for the development of new circRNA-based assays including many other circRNAs that could be relevant in lung cancer.

In addition, several clusters of varied circRNAs have been found dysregulated in the different lung cancer biosources explored in this project. Besides investigating the value of nCounter for the study of circRNAs in lung cancer, our second aim was to investigate their potential biomarker value. Therefore, a thorough investigation on the biological meaning of these discoveries cannot be found in this work, being out of the scope of this project. Yet, our investigation sets the grounds for future studies where to investigate these molecules as plausible therapeutic targets.

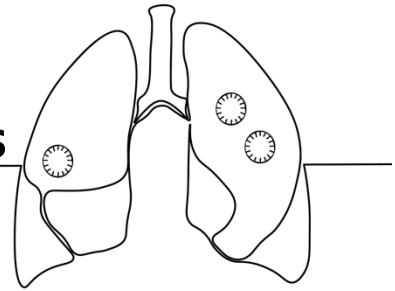
Furthermore, in this work we only focused on the diagnostic value of cited transcripts for the early detection of lung cancer. However, several reports have demonstrated the relationship between circRNAs and therapy resistance [179]. Therefore, future lines of investigation could include the development of prognostic signatures for lung cancer to predict treatment response using the nCounter technology.

EVs have been introduced in the last years as key players mediating cellular communication [56]. In Chapter three, we studied differential expression of circRNAs coming from EV preparations of lung cancer patients and controls. However, our findings cannot be attributed to a direct comparison of tumor released EVs versus normal EVs, since we did not perform any preliminary vesicular sorting. Although our results showed a circRNA signature with potential clinical value for lung cancer detection, investigation on the type of EVs present in our samples would provide more clarity to better understand the biological meaning of our findings. A collaboration with an Italian partner has been established in order to shed light on this matter.

Finally, we presented different circRNA signatures, coming out with a final protocol for the analysis of circRNAs from whole plasma samples. The absence of an EV-purification step makes our assay easier to implement in the medical setting than previous developed signatures. Since our preliminary results are promising, the continuation of this project is warranted, and a new prospective study has just been opened for the inclusion of bigger cohorts to further assess and validate our presented signature.

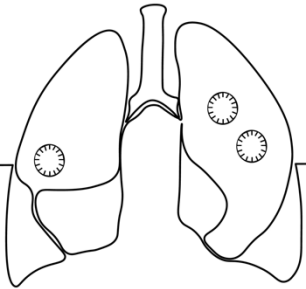
Additionally, combination of our signature with of other signatures based on mRNA o miRNA expression will be considered in a multi-omics approach, in order to achieve higher classification accuracy. We believe that our assays could act as a companion diagnostics test, and along with LDCT could help to guide medical decisions in prospective lung cancer screening campaigns.

Chapter VII: General conclusions



- The nCounter FLEX platform together with our custom-developed circRNA assay including 78 circRNAs and 10 mRNAs allows for the detection of circRNAs in lung cancer materials, including cell lines, FFPE tissues, plasma, and plasma EVs.
- RNase R treatment of RNA purified from fresh or frozen lung specimens is compatible with downstream nCounter processing, aiding in the detection of circRNAs. However, this treatment should be avoided for the detection of circRNAs from FFPE samples since circRNAs purified from this material become susceptible to the activity of this exonuclease.
- The protocols included in this thesis allowed differential expression analysis of circRNAs in lung cancer cells, tissues, and liquid biopsies, founding different clusters of dysregulated circRNAs in lung cancer patients.
- ML analysis uncovered the biomarker potential of some circRNAs developing different circRNA signatures that allowed discrimination of lung cancer from controls with promising accuracy rates.
- All circRNA-signatures presented in this thesis are proof-of-concept; therefore, the number of recruited patients and controls are small. The inclusion of larger cohorts will be necessary for the reassessment, training, and validation of cited ML-based signatures.
- Clinically relevant multiplex circRNA assays in liquid biopsies can be developed and validated in the nCounter FLEX platform, helping translational researchers with the transition from bench to bedside.

Bibliography



- [1] A. Puisieux, R. M. Pommier, A.-P. Morel, and F. Laval, "Cellular Pliancy and the Multistep Process of Tumorigenesis," *Cancer Cell*, vol. 33, no. 2, pp. 164-172, 2018/02/12/ 2018, doi: <https://doi.org/10.1016/j.ccell.2018.01.007>.
- [2] J. K. Sabari, B. H. Lok, J. H. Laird, J. T. Poirier, and C. M. Rudin, "Unravelling the biology of SCLC: implications for therapy," *Nature Reviews Clinical Oncology*, vol. 14, no. 9, pp. 549-561, 2017/09/01 2017, doi: 10.1038/nrclinonc.2017.71.
- [3] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, "Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship," (in eng), *Mayo Clinic proceedings*, vol. 83, no. 5, pp. 584-594, 2008, doi: 10.4065/83.5.584.
- [4] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, <https://doi.org/10.3322/caac.21660> vol. 71, no. 3, pp. 209-249, 2021/05/01 2021, doi: <https://doi.org/10.3322/caac.21660>.
- [5] M. G. Oser, M. J. Niederst, L. V. Sequist, and J. A. Engelman, "Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin," *The Lancet Oncology*, vol. 16, no. 4, pp. e165-e172, 2015, doi: 10.1016/S1470-2045(14)71180-5.
- [6] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, <https://doi.org/10.3322/caac.21492> vol. 68, no. 6, pp. 394-424, 2018/11/01 2018, doi: <https://doi.org/10.3322/caac.21492>.
- [7] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians*, <https://doi.org/10.3322/caac.21262> vol. 65, no. 2, pp. 87-108, 2015/03/01 2015, doi: <https://doi.org/10.3322/caac.21262>.
- [8] E. Surveillance, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat, ed.
- [9] S. E. A. i. w. f. S. c. statistics. "Surveillance Research Program, National Cancer Institute." <https://seer.cancer.gov/explorer/> (accessed 20.09.2022, 2022).
- [10] S. N. Lone *et al.*, "Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments," *Molecular Cancer*, vol. 21, no. 1, p. 79, 2022/03/18 2022, doi: 10.1186/s12943-022-01543-7.
- [11] T. N. Beck *et al.*, "Circulating tumor cell and cell-free RNA capture and expression analysis identify platelet-associated genes in metastatic lung cancer," *BMC Cancer*, vol. 19, no. 1, p. 603, 2019/06/19 2019, doi: 10.1186/s12885-019-5795-x.
- [12] L. Han, J. Xu, Q. Xu, B. Zhang, E. W. F. Lam, and Y. Sun, "Extracellular vesicles in the tumor microenvironment: Therapeutic resistance, clinical biomarkers, and targeting strategies," *Medicinal Research Reviews*, <https://doi.org/10.1002/med.21453> vol. 37, no. 6, pp. 1318-1349, 2017/11/01 2017, doi: <https://doi.org/10.1002/med.21453>.
- [13] Simon A. Joosse and K. Pantel, "Tumor-Educated Platelets as Liquid Biopsy in Cancer Patients," *Cancer Cell*, vol. 28, no. 5, pp. 552-554, 2015, doi: 10.1016/j.ccell.2015.10.007.
- [14] J. W. P. Bracht, C. Mayo-de-Las-Casas, J. Berenguer, N. Karachaliou, and R. Rosell, "The Present and Future of Liquid Biopsies in Non-Small Cell Lung Cancer: Combining Four Biosources for Diagnosis, Prognosis, Prediction, and Disease Monitoring," (in eng), no. 1534-6269 (Electronic).
- [15] "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395-409, 2011/08/04 2011, doi: 10.1056/NEJMoA1102873.
- [16] "Lung Cancer Incidence and Mortality with Extended Follow-up in the National Lung Screening Trial," *Journal of Thoracic Oncology*, vol. 14, no. 10, pp. 1732-1742, 2019, doi: 10.1016/j.jtho.2019.05.044.
- [17] U. Pastorino *et al.*, "Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy," *Annals of Oncology*, vol. 30, no. 7, pp. 1162-1169, 2019, doi: 10.1093/annonc/mdz117.
- [18] H. J. de Koning *et al.*, "Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial," *New England Journal of Medicine*, vol. 382, no. 6, pp. 503-513, 2020/02/06 2020, doi: 10.1056/NEJMoA1911793.
- [19] A. Tivey, M. Church, D. Rothwell, C. Dive, and N. Cook, "Circulating tumour DNA — looking beyond the blood," *Nature Reviews Clinical Oncology*, vol. 19, no. 9, pp. 600-612, 2022/09/01 2022, doi: 10.1038/s41571-022-00660-y.
- [20] A. Szepechinski *et al.*, "Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease," (in eng), *British journal of cancer*, vol. 113, no. 3, pp. 476-483, 2015, doi: 10.1038/bjc.2015.225.
- [21] M. Zhong *et al.*, "Clinical Utility of Circulating Tumor Cells in the Early Detection of Lung Cancer in Patients with a Solitary Pulmonary Nodule," (in eng), *Technology in cancer research & treatment*, vol. 20, pp. 15330338211041465-15330338211041465, Jan-Dec 2021, doi: 10.1177/15330338211041465.

- [22] M. Gallo *et al.*, "Clinical utility of circulating tumor cells in patients with non-small-cell lung cancer," *Translational Lung Cancer Research*, vol. 6, no. 4, pp. 486-498, 2017.
- [23] E. N. M. Nolte-^t Hoen, H. P. J. Buermans, M. Waasdorp, W. Stoorvogel, M. H. M. Wauben, and P. A. C. ^t Hoen, "Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions," *Nucleic Acids Research*, vol. 40, no. 18, pp. 9272-9285, 2012, doi: 10.1093/nar/gks658.
- [24] N. Hasegawa *et al.*, "Highly sensitive fusion detection using plasma cell-free RNA in non-small-cell lung cancers," *Cancer Science*, <https://doi.org/10.1111/cas.15084> vol. 112, no. 10, pp. 4393-4403, 2021/10/01 2021, doi: <https://doi.org/10.1111/cas.15084>.
- [25] P. Reclusa *et al.*, "Exosomes as diagnostic and predictive biomarkers in lung cancer," (in eng), *Journal of thoracic disease*, vol. 9, no. Suppl 13, pp. S1373-S1382, 2017, doi: 10.21037/jtd.2017.10.67.
- [26] X. Tang, H. Ren, M. Guo, J. Qian, Y. Yang, and C. Gu, "Review on circular RNAs and new insights into their roles in cancer," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 910-928, 2021/01/01/ 2021, doi: <https://doi.org/10.1016/j.csbj.2021.01.018>.
- [27] J. Salzman, C. Gawad, P. L. Wang, N. Lacayo, and P. O. Brown, "Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types," *PLOS ONE*, vol. 7, no. 2, p. e30733, 2012, doi: 10.1371/journal.pone.0030733.
- [28] X. Hua *et al.*, "Circular RNAs in drug resistant tumors," *Biomedicine & Pharmacotherapy*, vol. 118, p. 109233, 2019/10/01/ 2019, doi: <https://doi.org/10.1016/j.biopha.2019.109233>.
- [29] J. M. Nigro *et al.*, "Scrambled exons," (in eng), no. 0092-8674 (Print).
- [30] E. C. S. Lee *et al.*, "The roles of circular RNAs in human development and diseases," *Biomedicine & Pharmacotherapy*, vol. 111, pp. 198-208, 2019/03/01/ 2019, doi: <https://doi.org/10.1016/j.biopha.2018.12.052>.
- [31] P. Ji *et al.*, "Expanded Expression Landscape and Prioritization of Circular RNAs in Mammals," *Cell Reports*, vol. 26, no. 12, pp. 3444-3460.e5, 2020/04/27 2019, doi: 10.1016/j.celrep.2019.02.078.
- [32] J. U. Guo, V. Agarwal, H. Guo, and D. P. Bartel, "Expanded identification and characterization of mammalian circular RNAs," *Genome Biology*, vol. 15, no. 7, p. 409, 2014/07/29 2014, doi: 10.1186/s13059-014-0409-z.
- [33] J. Salzman, R. E. Chen, M. N. Olsen, P. L. Wang, and P. O. Brown, "Cell-Type Specific Features of Circular RNA Expression," *PLOS Genetics*, vol. 9, no. 9, p. e1003777, 2013, doi: 10.1371/journal.pgen.1003777.
- [34] M. Dragomir and G. A. Calin, "Circular RNAs in Cancer – Lessons Learned From microRNAs," (in English), *Frontiers in Oncology*, Review vol. 8, no. 179, 2018-May-28 2018, doi: 10.3389/fonc.2018.00179.
- [35] M.-S. Xiao, Y. Ai, and J. E. Wilusz, "Biogenesis and Functions of Circular RNAs Come into Focus," *Trends in Cell Biology*, vol. 30, no. 3, pp. 226-240, 2020/04/27 2020, doi: 10.1016/j.tcb.2019.12.004.
- [36] C. Y. Chen and P. Sarnow, "Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs," *Science*, vol. 268, no. 5209, p. 415, 1995, doi: 10.1126/science.7536344.
- [37] A. Rybak-Wolf *et al.*, "Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed," (in eng), no. 1097-4164 (Electronic).
- [38] X. Li *et al.*, "Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection," (in eng), no. 1097-4164 (Electronic).
- [39] R. Ashwal-Fluss *et al.*, "circRNA Biogenesis Competes with Pre-mRNA Splicing," *Molecular Cell*, vol. 56, no. 1, pp. 55-66, 2014, doi: 10.1016/j.molcel.2014.08.019.
- [40] T. Fei *et al.*, "Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing," *Proceedings of the National Academy of Sciences*, vol. 114, no. 26, p. E5207, 2017, doi: 10.1073/pnas.1617467114.
- [41] L. Errichelli *et al.*, "FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons," *Nature Communications*, vol. 8, no. 1, p. 14741, 2017, doi: 10.1038/ncomms14741.
- [42] Simon J. Conn *et al.*, "The RNA Binding Protein Quaking Regulates Formation of circRNAs," *Cell (Cambridge)*, vol. 160, no. 6, pp. 1125-1134, 2015, doi: 10.1016/j.cell.2015.02.014.
- [43] T. Aktaş *et al.*, "DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome," (in eng), no. 1476-4687 (Electronic).
- [44] M. A. Khan *et al.*, "RBM20 Regulates Circular RNA Production From the Titin Gene," (in eng), no. 1524-4571 (Electronic).
- [45] W. R. Jeck *et al.*, "Circular RNAs are abundant, conserved, and associated with ALU repeats," (in eng), *RNA (New York, N.Y.)*, vol. 19, no. 2, pp. 141-157, 2013, doi: 10.1261/rna.035667.112.
- [46] C. Huang, D. Liang, D. C. Tatomer, and J. A.-O. Wilusz, "A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs," (in eng), no. 1549-5477 (Electronic).

- [47] T. B. Hansen *et al.*, "miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA," (in eng), no. 1460-2075 (Electronic).
- [48] O. H. Park *et al.*, "Endoribonucleolytic Cleavage of m(6)A-Containing RNAs by RNase P/MRP Complex," (in eng), no. 1097-4164 (Electronic).
- [49] C.-X. Liu *et al.*, "Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity," *Cell*, vol. 177, no. 4, pp. 865-880.e21, 2020/04/27 2019, doi: 10.1016/j.cell.2019.03.046.
- [50] J. W. Fischer, V. F. Busa, Y. Shao, and A. K. L. Leung, "Structure-Mediated RNA Decay by UPF1 and G3BP1," (in eng), no. 1097-4164 (Electronic).
- [51] T. Yu *et al.*, "CircRNAs in cancer metabolism: a review," *Journal of Hematology & Oncology*, vol. 12, no. 1, p. 90, 2019/09/04 2019, doi: 10.1186/s13045-019-0776-8.
- [52] R. Zhou *et al.*, "Circular RNAs (circRNAs) in cancer," (in eng), no. 1872-7980 (Electronic).
- [53] H. Yang *et al.*, "Hypoxia induced exosomal circRNA promotes metastasis of Colorectal Cancer via targeting GEF-H1/RhoA axis," (in eng), *Theranostics*, vol. 10, no. 18, pp. 8211-8226, 2020, doi: 10.7150/thno.44419.
- [54] Y. Li *et al.*, "Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis," *Cell Research*, vol. 25, no. 8, pp. 981-984, 2015/08/01 2015, doi: 10.1038/cr.2015.82.
- [55] E. Lasda and R. Parker, "Circular RNAs Co-Precipitate with Extracellular Vesicles: A Possible Mechanism for circRNA Clearance," *PLOS ONE*, vol. 11, no. 2, p. e0148407, 2016, doi: 10.1371/journal.pone.0148407.
- [56] M. Wang, F. Yu, P. Li, and K. Wang, "Emerging Function and Clinical Significance of Exosomal circRNAs in Cancer," *Molecular Therapy - Nucleic Acids*, vol. 21, pp. 367-383, 2020/09/04/ 2020, doi: <https://doi.org/10.1016/j.omtn.2020.06.008>.
- [57] B.-Q. Shang *et al.*, "Functional roles of circular RNAs during epithelial-to-mesenchymal transition," *Molecular Cancer*, vol. 18, no. 1, p. 138, 2019, doi: 10.1186/s12943-019-1071-6.
- [58] F. Chen, C. Huang, Q. Wu, L. Jiang, S. Chen, and L. Chen, "Circular RNAs expression profiles in plasma exosomes from early-stage lung adenocarcinoma and the potential biomarkers," *Journal of Cellular Biochemistry*, vol. 121, no. 3, pp. 2525-2533, 2020/04/27 2020, doi: 10.1002/jcb.29475.
- [59] F. He *et al.*, "Plasma exo-hsa_circRNA_0056616: A potential biomarker for lymph node metastasis in lung adenocarcinoma," *Journal of Cancer*, Research Paper vol. 11, no. 14, pp. 4037-4046, 2020, doi: 10.7150/jca.30360.
- [60] J. Xian *et al.*, "Identification of Three Circular RNA Cargoes in Serum Exosomes as Diagnostic Biomarkers of Non-Small-Cell Lung Cancer in the Chinese Population," *The Journal of Molecular Diagnostics*, vol. 22, no. 8, pp. 1096-1108, 2020, doi: 10.1016/j.jmoldx.2020.05.011.
- [61] X.-X. Liu *et al.*, "A two-circular RNA signature as a noninvasive diagnostic biomarker for lung adenocarcinoma," *Journal of Translational Medicine*, vol. 17, no. 1, p. 50, 2019, doi: 10.1186/s12967-019-1800-z.
- [62] A. A. Alhasan *et al.*, "Circular RNA enrichment in platelets is a signature of transcriptome degradation," (in eng), *Blood*, vol. 127, no. 9, pp. e1-e11, 2016, doi: 10.1182/blood-2015-06-649434.
- [63] C. Preußner *et al.*, "Selective release of circRNAs in platelet-derived extracellular vesicles," (in eng), *Journal of extracellular vesicles*, vol. 7, no. 1, pp. 1424473-1424473, 2018, doi: 10.1080/20013078.2018.1424473.
- [64] M. G. Best *et al.*, "Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets," (in eng), *Cancer cell*, vol. 32, no. 2, pp. 238-252.e9, 2017, doi: 10.1016/j.ccell.2017.07.004.
- [65] S. D'Ambrosi *et al.*, "The Analysis of Platelet-Derived circRNA Repertoire as Potential Diagnostic Biomarker for Non-Small Cell Lung Cancer," *Cancers*, vol. 13, no. 18, doi: 10.3390/cancers13184644.
- [66] NanoString. <https://nanosttring.com> (accessed 20.09.2022, 2022).
- [67] T. B. Porras, P. Kaur, A. Ring, N. Schechter, and J. E. Lang, "Challenges in using liquid biopsies for gene expression profiling," *Oncotarget; Vol 9, No 6*, 2018.
- [68] W. S. Hanson D, Meredith G, Eagan M, Dennis L, Krouse M. et al. "Digital Gene Expression from Low Sample Input: Highly multiplexed and robust profiling of formalni-fixed paraffin-embedded (FFPE) and fresh frozen samples from as little as 1 ng of RNA using the nCounter platform." (accessed 15.12.2021, 2021).
- [69] G. K. Geiss *et al.*, "Direct multiplexed measurement of gene expression with color-coded probe pairs," *Nature Biotechnology*, vol. 26, no. 3, pp. 317-325, 2008/03/01 2008, doi: 10.1038/nbt1385.
- [70] M. Ayers *et al.*, "IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade," *The Journal of Clinical Investigation*, vol. 127, no. 8, pp. 2930-2940, 08/01/ 2017, doi: 10.1172/JCI91190.
- [71] B. Wallden *et al.*, "Development and verification of the PAM50-based Prosigna breast cancer gene signature assay," *BMC Medical Genomics*, vol. 8, no. 1, p. 54, 2015/08/22 2015, doi: 10.1186/s12920-015-0129-6.

- [72] M. E. Lira *et al.*, "A Single-Tube Multiplexed Assay for Detecting *ALK*, *ROS1*, and *RET* Fusions in Lung Cancer," *The Journal of Molecular Diagnostics*, vol. 16, no. 2, pp. 229-243, 2014, doi: 10.1016/j.jmoldx.2013.11.007.
- [73] D. Hu *et al.*, "Development of a NanoString assay to detect leukemogenic fusion transcripts in acute myeloid leukemia," *International Journal of Laboratory Hematology*, <https://doi.org/10.1111/ijlh.12555> vol. 38, no. 6, pp. 663-673, 2016/12/01 2016, doi: <https://doi.org/10.1111/ijlh.12555>.
- [74] D. W. Scott *et al.*, "New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies," *Journal of Clinical Oncology*, vol. 35, no. 15, pp. 1668-1677, 2017/05/20 2017, doi: 10.1200/JCO.2016.70.7901.
- [75] K. T. E. Chang *et al.*, "Development and Evaluation of a Pan-Sarcoma Fusion Gene Detection Assay Using the NanoString nCounter Platform," *The Journal of Molecular Diagnostics*, vol. 20, no. 1, pp. 63-77, 2018/01/01/ 2018, doi: <https://doi.org/10.1016/j.jmoldx.2017.09.007>.
- [76] J. S. Parker *et al.*, "Supervised risk predictor of breast cancer based on intrinsic subtypes," (in eng), *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 27, no. 8, pp. 1160-1167, 2009, doi: 10.1200/JCO.2008.18.1370.
- [77] A. Giménez-Capitán *et al.*, "Multiplex Detection of Clinically Relevant Mutations in Liquid Biopsies of Cancer Patients Using a Hybridization-Based Platform," *Clinical Chemistry*, vol. 67, no. 3, pp. 554-563, 2021, doi: 10.1093/clinchem/hvaa248.
- [78] A. V. Kossenkov *et al.*, "A Gene Expression Classifier from Whole Blood Distinguishes Benign from Malignant Lung Nodules Detected by Low-Dose CT," *Cancer Research*, vol. 79, no. 1, pp. 263-273, 2019, doi: 10.1158/0008-5472.CAN-18-2032.
- [79] M. Garcia-Contreras *et al.*, "Plasma-derived exosome characterization reveals a distinct microRNA signature in long duration Type 1 diabetes," *Scientific Reports*, vol. 7, no. 1, p. 5998, 2017/07/20 2017, doi: 10.1038/s41598-017-05787-y.
- [80] C. Vicentini *et al.*, "Exosomal miRNA signatures of pancreatic lesions," *BMC Gastroenterology*, vol. 20, no. 1, p. 137, 2020/05/06 2020, doi: 10.1186/s12876-020-01287-y.
- [81] A. G. Capitán *et al.*, "P22.04 Prospective Validation of an Eight Gene mRNA Signature in Plasma for the Diagnosis of Early Stage Lung Cancer," *Journal of Thoracic Oncology*, vol. 16, no. 10, p. S1026, 2021, doi: 10.1016/j.jtho.2021.08.358.
- [82] C. Pedraz-Valdunciel and R. Rosell, "Defining the landscape of circRNAs in non-small cell lung cancer and their potential as liquid biopsy biomarkers: a complete review including current methods," *Extracellular Vesicles and Circulating Nucleic Acids*, vol. 2, no. 2, pp. 179-201, 2021, doi: 10.20517/evcna.2020.07.
- [83] W. R. Jeck and N. E. Sharpless, "Detecting and characterizing circular RNAs," (in eng), *Nature biotechnology*, vol. 32, no. 5, pp. 453-461, 2014, doi: 10.1038/nbt.2890.
- [84] H. L. Sanger, G. Klotz, D. Riesner, H. J. Gross, and A. K. Kleinschmidt, "Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures," (in eng), *Proceedings of the National Academy of Sciences of the United States of America*, vol. 73, no. 11, pp. 3852-3856, 1976, doi: 10.1073/pnas.73.11.3852.
- [85] J. M. Nigro *et al.*, "Scrambled exons," *Cell*, vol. 64, no. 3, pp. 607-613, 1991, doi: 10.1016/0092-8674(91)90244-S.
- [86] T. B. Hansen *et al.*, "Natural RNA circles function as efficient microRNA sponges," *Nature*, vol. 495, no. 7441, pp. 384-388, 2013/03/01 2013, doi: 10.1038/nature11993.
- [87] S. Memczak *et al.*, "Circular RNAs are a large class of animal RNAs with regulatory potency," *Nature*, vol. 495, no. 7441, pp. 333-338, 2013/03/01 2013, doi: 10.1038/nature11928.
- [88] S. Qu *et al.*, "The emerging functions and roles of circular RNAs in cancer," *Cancer Letters*, vol. 414, pp. 301-309, 2018/02/01/ 2018, doi: <https://doi.org/10.1016/j.canlet.2017.11.022>.
- [89] W.-Y. Zhou, Z.-R. Cai, J. Liu, D.-S. Wang, H.-Q. Ju, and R.-H. Xu, "Circular RNA: metabolism, functions and interactions with proteins," *Molecular Cancer*, vol. 19, no. 1, p. 172, 2020/12/14 2020, doi: 10.1186/s12943-020-01286-3.
- [90] S. Starke *et al.*, "Exon Circularization Requires Canonical Splice Signals," *Cell Reports*, vol. 10, no. 1, pp. 103-111, 2015, doi: 10.1016/j.celrep.2014.12.002.
- [91] C.-Y. Yu and H.-C. Kuo, "The emerging roles and functions of circular RNAs and their generation," *Journal of Biomedical Science*, vol. 26, no. 1, p. 29, 2019/04/25 2019, doi: 10.1186/s12929-019-0523-z.
- [92] R. Zhou *et al.*, "Circular RNAs (circRNAs) in cancer," *Cancer Letters*, vol. 425, pp. 134-142, 2018/07/01/ 2018, doi: <https://doi.org/10.1016/j.canlet.2018.03.035>.
- [93] C. Pedraz-Valdunciel and R. Rosell, "Defining the landscape of circRNAs in non-small cell lung cancer and their potential as liquid biopsy biomarkers: a complete review including current methods.," ed. *Extracell Vesicles Circ Nucleic Acids*, 2021.

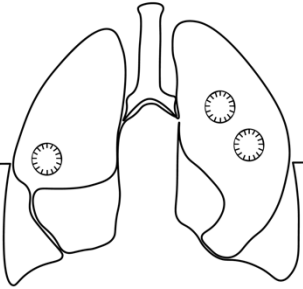
- [94] I. San Segundo-Val and C. S. Sanz-Lozano, "Introduction to the Gene Expression Analysis," in *Molecular Genetics of Asthma*, M. Isidoro García Ed. New York, NY: Springer New York, 2016, pp. 29-43.
- [95] S. D. Prokopec *et al.*, "Systematic evaluation of medium-throughput mRNA abundance platforms," (in eng), *RNA (New York, N.Y.)*, vol. 19, no. 1, pp. 51-62, 2013, doi: 10.1261/rna.034710.112.
- [96] M. M. Kulkarni, "Digital Multiplexed Gene Expression Analysis Using the NanoString nCounter System," *Current Protocols in Molecular Biology*, <https://doi.org/10.1002/0471142727.mb25b10s94> vol. 94, no. 1, pp. 25B.10.1-25B.10.17, 2011/04/01 2011, doi: <https://doi.org/10.1002/0471142727.mb25b10s94>.
- [97] C. Philippeos, R. D. Hughes, A. Dhawan, and R. R. Mitry, "Introduction to Cell Culture," in *Human Cell Culture Protocols*, R. R. Mitry and R. D. Hughes Eds. Totowa, NJ: Humana Press, 2012, pp. 1-13.
- [98] C. Aguado *et al.*, "RNA-Based Multiplexing Assay for Routine Testing of Fusion and Splicing Variants in Cytological Samples of NSCLC Patients," (in eng), *Diagnostics (Basel, Switzerland)*, vol. 11, no. 1, p. 15, 2020, doi: 10.3390/diagnostics11010015.
- [99] M. Dahl *et al.*, "Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies," *Laboratory Investigation*, vol. 98, no. 12, pp. 1657-1669, 2018/12/01 2018, doi: 10.1038/s41374-018-0108-6.
- [100] M. Olivier, R. Asmis, G. A. Hawkins, T. D. Howard, and L. A. Cox, "The Need for Multi-Omics Biomarker Signatures in Precision Medicine," (in eng), *International journal of molecular sciences*, vol. 20, no. 19, p. 4781, 2019, doi: 10.3390/ijms20194781.
- [101] H.-d. Zhang, L.-h. Jiang, D.-w. Sun, J.-c. Hou, and Z.-l. Ji, "CircRNA: a novel type of biomarker for cancer," *Breast Cancer*, vol. 25, no. 1, pp. 1-7, 2018/01/01 2018, doi: 10.1007/s12282-017-0793-9.
- [102] J. Zhang *et al.*, "Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells," (in eng), *RNA biology*, vol. 16, no. 2, pp. 220-232, 2019, doi: 10.1080/15476286.2019.1565284.
- [103] L.-I. Moldovan *et al.*, "High-throughput RNA sequencing from paired lesional- and non-lesional skin reveals major alterations in the psoriasis circRNAome," (in eng), *BMC medical genomics*, vol. 12, no. 1, pp. 174-174, 2019, doi: 10.1186/s12920-019-0616-2.
- [104] K. Das Mahapatra *et al.*, "A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma," (in eng), *Scientific reports*, vol. 10, no. 1, pp. 3637-3637, 2020, doi: 10.1038/s41598-020-59660-6.
- [105] U. Ahmadov *et al.*, "Distinct circular RNA expression profiles in pediatric ependymomas," *Brain Pathology*, <https://doi.org/10.1111/bpa.12922> vol. 31, no. 2, pp. 387-392, 2021/03/01 2021, doi: <https://doi.org/10.1111/bpa.12922>.
- [106] Y. Yan, R. Zhang, X. Zhang, A. Zhang, Y. Zhang, and X. Bu, "RNA-Seq profiling of circular RNAs and potential function of hsa_circ_0002360 in human lung adenocarcinoma," (in eng), *American journal of translational research*, vol. 11, no. 1, pp. 160-175, 2019.
- [107] F. Zhang, X. Zhao, H. Dong, and J. Xu, "circRNA expression analysis in lung adenocarcinoma: comparison of paired fresh frozen and formalin-fixed paraffin-embedded specimens," *Biochemical and Biophysical Research Communications*, vol. 500, no. 3, pp. 738-743, 2018/06/07/ 2018, doi: <https://doi.org/10.1016/j.bbrc.2018.04.145>.
- [108] J. Zhao, L. Li, Q. Wang, H. Han, Q. Zhan, and M. Xu, "CircRNA Expression Profile in Early-Stage Lung Adenocarcinoma Patients," *Cellular Physiology and Biochemistry*, vol. 44, no. 6, pp. 2138-2146, 2017, doi: 10.1159/000485953.
- [109] S. Li *et al.*, "hsa_circ_0000729, a potential prognostic biomarker in lung adenocarcinoma," *Thoracic Cancer*, <https://doi.org/10.1111/1759-7714.12761> vol. 9, no. 8, pp. 924-930, 2018/08/01 2018, doi: <https://doi.org/10.1111/1759-7714.12761>.
- [110] S. T. Hossain, A. Malhotra, and M. P. Deutscher, "How RNase R Degrades Structured RNA: ROLE OF THE HELICASE ACTIVITY AND THE S1 DOMAIN," (in eng), *The Journal of biological chemistry*, vol. 291, no. 15, pp. 7877-7887, 2016, doi: 10.1074/jbc.M116.717991.
- [111] Jakub O. Westholm *et al.*, "Genome-wide Analysis of *Drosophila* Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation," *Cell Reports*, vol. 9, no. 5, pp. 1966-1980, 2014, doi: 10.1016/j.celrep.2014.10.062.
- [112] L. Szabo and J. Salzman, "Detecting circular RNAs: bioinformatic and experimental challenges," (in eng), *Nature reviews. Genetics*, vol. 17, no. 11, pp. 679-692, 2016, doi: 10.1038/nrg.2016.114.
- [113] J. Zhang, S. Chen, J. Yang, and F. Zhao, "Accurate quantification of circular RNAs identifies extensive circular isoform switching events," *Nature Communications*, vol. 11, no. 1, p. 90, 2020, doi: 10.1038/s41467-019-13840-9.
- [114] A. Bachmayr-Heyda *et al.*, "Correlation of circular RNA abundance with proliferation – exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis and normal human tissues," *Scientific Reports*, vol. 5, no. 1, p. 8057, 2015/01/27 2015, doi: 10.1038/srep08057.

- [115] J. Yang *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," (in eng), *Frontiers in oncology*, vol. 11, pp. 595609-595609, 2021, doi: 10.3389/fonc.2021.595609.
- [116] D. Hang *et al.*, "A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer," (in eng), *Cancer medicine*, vol. 7, no. 6, pp. 2783-2791, 2018, doi: 10.1002/cam4.1514.
- [117] L. Ruan, J. Chen, L. Ruan, T. Yang, and P. Wang, "MicroRNA-186 suppresses lung cancer progression by targeting SIRT6," *Cancer Biomarkers*, vol. 21, pp. 415-423, 2018, doi: 10.3233/CBM-170650.
- [118] T. Huang *et al.*, "MicroRNA-186 suppresses cell proliferation and metastasis through targeting MAP3K2 in non-small cell lung cancer," *Int J Oncol*, vol. 49, no. 4, pp. 1437-1444, 2016/10/01 2016, doi: 10.3892/ijo.2016.3637.
- [119] H. Zhu, Q. Lu, Q. Lu, X. Shen, and L. Yu, "Matrine Regulates Proliferation, Apoptosis, Cell Cycle, Migration, and Invasion of Non-Small Cell Lung Cancer Cells Through the circFUT8/miR-944/YES1 Axis," (in eng), *Cancer management and research*, vol. 13, pp. 3429-3442, 2021, doi: 10.2147/CMAR.S290966.
- [120] Z. Tan *et al.*, "MicroRNA-1229 overexpression promotes cell proliferation and tumorigenicity and activates Wnt/ β -catenin signaling in breast cancer," (in eng), *Oncotarget*, vol. 7, no. 17, pp. 24076-24087, 2016, doi: 10.18632/oncotarget.8119.
- [121] J. Liu, D. Wang, Z. Long, and W. Li, "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," *Cellular Physiology and Biochemistry*, vol. 48, no. 1, pp. 173-184, 2018, doi: 10.1159/000491716.
- [122] A. Fasihi, B. M. Soltani, Z. S. Ranjbaran, S. Bahonar, R. Norouzi, and S. Nasiri, "Hsa-miR-942 fingerprint in colorectal cancer through Wnt signaling pathway," *Gene*, vol. 712, p. 143958, 2019/09/05/ 2019, doi: <https://doi.org/10.1016/j.gene.2019.143958>.
- [123] C. Ge *et al.*, "miR-942 promotes cancer stem cell-like traits in esophageal squamous cell carcinoma through activation of Wnt/ β -catenin signalling pathway," *Oncotarget; Vol 6, No 13*, 2015.
- [124] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, <https://doi.org/10.3322/caac.21708> vol. 72, no. 1, pp. 7-33, 2022/01/01 2022, doi: <https://doi.org/10.3322/caac.21708>.
- [125] E. Union. "Source: ECIS - European Cancer Information System." <https://ecis.jrc.ec.europa.eu>, accessed on day/month/year (accessed 2022).
- [126] D. Crosby *et al.*, "Early detection of cancer," *Science*, vol. 375, no. 6586, p. eaay9040, doi: 10.1126/science.aay9040.
- [127] S. Perakis and M. R. Speicher, "Emerging concepts in liquid biopsies," *BMC Medicine*, vol. 15, no. 1, p. 75, 2017/04/06 2017, doi: 10.1186/s12916-017-0840-6.
- [128] M.-S. Xiao and J. E. Wilusz, "An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends," (in eng), *Nucleic acids research*, vol. 47, no. 16, pp. 8755-8769, 2019, doi: 10.1093/nar/gkz576.
- [129] Q. Zhou, L.-L. Ju, X. Ji, Y.-L. Cao, J.-G. Shao, and L. Chen, "Plasma circRNAs as Biomarkers in Cancer," (in eng), *Cancer management and research*, vol. 13, pp. 7325-7337, 2021, doi: 10.2147/CMAR.S330228.
- [130] S. Warren, "Simultaneous, Multiplexed Detection of RNA and Protein on the NanoString® nCounter® Platform," in *Gene Expression Analysis: Methods and Protocols*, N. Raghavachari and N. Garcia-Reyero Eds. New York, NY: Springer New York, 2018, pp. 105-120.
- [131] T.-C. Wu *et al.*, "IL1 Receptor Antagonist Controls Transcriptional Signature of Inflammation in Patients with Metastatic Breast Cancer," *Cancer Research*, vol. 78, no. 18, pp. 5243-5258, 2018, doi: 10.1158/0008-5472.CAN-18-0413.
- [132] N. Kamyabi, R. Abbasgholizadeh, A. Maitra, A. Ardekani, S. L. Biswal, and K. J. Grande-Allen, "Isolation and mutational assessment of pancreatic cancer extracellular vesicles using a microfluidic platform," *Biomedical Microdevices*, vol. 22, no. 2, p. 23, 2020/03/11 2020, doi: 10.1007/s10544-020-00483-7.
- [133] J. W. P. Bracht *et al.*, "Analysis of extracellular vesicle mRNA derived from plasma using the nCounter platform," *Scientific Reports*, vol. 11, no. 1, p. 3712, 2021/02/12 2021, doi: 10.1038/s41598-021-83132-0.
- [134] E. B. Hansen *et al.*, "The transcriptional landscape and biomarker potential of circular RNAs in prostate cancer," *Genome Medicine*, vol. 14, no. 1, p. 8, 2022/01/25 2022, doi: 10.1186/s13073-021-01009-3.
- [135] J. Berenguer *et al.*, "Glycosylated extracellular vesicles released by glioblastoma cells are decorated by CCL18 allowing for cellular uptake via chemokine receptor CCR8," *Journal of Extracellular Vesicles*, vol. 7, no. 1, p. 1446660, 2018/12/01 2018, doi: 10.1080/20013078.2018.1446660.
- [136] Y. Tian *et al.*, "Quality and efficiency assessment of six extracellular vesicle isolation methods by nano-flow cytometry," *Journal of Extracellular Vesicles*, <https://doi.org/10.1080/20013078.2019.1697028> vol. 9, no. 1, p. 1697028, 2020/09/01 2020, doi: <https://doi.org/10.1080/20013078.2019.1697028>.

- [137] C. Pedraz-Valdunciel *et al.*, "Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens," *Molecular Oncology*, <https://doi.org/10.1002/1878-0261.13182> vol. 16, no. 12, pp. 2367-2383, 2022/06/01 2022, doi: <https://doi.org/10.1002/1878-0261.13182>.
- [138] L. Margolis and Y. Sadovsky, "The biology of extracellular vesicles: The known unknowns," *PLoS Biology*, vol. 17, no. 7, p. e3000363, 2019, doi: 10.1371/journal.pbio.3000363.
- [139] P. Reclusa *et al.*, "Exosomes genetic cargo in lung cancer: a truly Pandora's box," *Translational Lung Cancer Research*, vol. 5, no. 5, pp. 483-491, 2016.
- [140] D. K. Jeppesen *et al.*, "Reassessment of Exosome Composition," *Cell*, vol. 177, no. 2, pp. 428-445.e18, 2019, doi: 10.1016/j.cell.2019.02.029.
- [141] X. Yang *et al.*, "The Key Role of Exosomes on the Pre-metastatic Niche Formation in Tumors," (in English), *Frontiers in Molecular Biosciences*, Review vol. 8, 2021-September-14 2021, doi: 10.3389/fmolb.2021.703640.
- [142] B. Costa-Silva *et al.*, "Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver," *Nature Cell Biology*, vol. 17, no. 6, pp. 816-826, 2015/06/01 2015, doi: 10.1038/ncb3169.
- [143] C. Pedraz-Valdunciel *et al.*, *Journal of Thoracic Oncology*, vol. 16, no. 3, pp. S555-S556, 2021, doi: 10.1016/j.jtho.2021.01.998.
- [144] I. Helwa *et al.*, "A Comparative Study of Serum Exosome Isolation Using Differential Ultracentrifugation and Three Commercial Reagents," (in eng), *PloS one*, vol. 12, no. 1, pp. e0170628-e0170628, 2017, doi: 10.1371/journal.pone.0170628.
- [145] S. A. Melo *et al.*, "Glypican-1 identifies cancer exosomes and detects early pancreatic cancer," *Nature*, vol. 523, no. 7559, pp. 177-182, 2015/07/01 2015, doi: 10.1038/nature14581.
- [146] H. Lu, X. Han, J. Ren, K. Ren, Z. Li, and Z. Sun, "Circular RNA HIPK3 induces cell proliferation and inhibits apoptosis in non-small cell lung cancer through sponging miR-149," *Cancer Biology & Therapy*, vol. 21, no. 2, pp. 113-121, 2020, doi: 10.1080/15384047.2019.1669995.
- [147] X. Chen *et al.*, "Circular RNA circHIPK3 modulates autophagy via MIR124-3p-STAT3-PRKAA/AMPK α signaling in STK11 mutant lung cancer," *Autophagy*, vol. 16, no. 4, pp. 659-671, 2020, doi: 10.1080/15548627.2019.1634945.
- [148] H. Yu, Y. Chen, and P. Jiang, "Circular RNA HIPK3 exerts oncogenic properties through suppression of miR-124 in lung cancer," *Biochemical and Biophysical Research Communications*, vol. 506, no. 3, pp. 455-462, 2018/11/30/ 2018, doi: <https://doi.org/10.1016/j.bbrc.2018.10.087>.
- [149] Y. Guo, W. Xue, S. Sun, X. Chen, H. Li, and C. Yan, "Circular RNA circZCCHC6 contributes to tumorigenesis by regulating LPCAT1 via miR-433-3p in non-small cell lung cancer," *Clinical and Experimental Medicine*, 2022/01/28 2022, doi: 10.1007/s10238-021-00780-2.
- [150] Q. Z. R.-R. Wu, H.-F. Liu, S.-B. Liu, "Role of miR-579-3p in the development of squamous cell lung carcinoma and the regulatory mechanisms," vol. 23 - N. 21 ed, 2019, pp. 9464-9470.
- [151] S. Wei *et al.*, "Erratum: Hsa-miR-623 suppresses tumor progression in human lung adenocarcinoma," *Cell Death & Disease*, vol. 8, no. 5, pp. e2829-e2829, 2017/05/01 2017, doi: 10.1038/cddis.2017.254.
- [152] B. Sun, J. Hua, H. Cui, H. Liu, K. Zhang, and H. Zhou, "MicroRNA-1197 downregulation inhibits proliferation and migration in human non-small cell lung cancer cells by upregulating HOXC11," *Biomedicine & Pharmacotherapy*, vol. 117, p. 109041, 2019/09/01/ 2019, doi: <https://doi.org/10.1016/j.biopha.2019.109041>.
- [153] C.-g. Li *et al.*, "MicroRNA-1304 suppresses human non-small cell lung cancer cell growth in vitro by targeting heme oxygenase-1," *Acta Pharmacologica Sinica*, vol. 38, no. 1, pp. 110-119, 2017/01/01 2017, doi: 10.1038/aps.2016.92.
- [154] C. Liu *et al.*, "microRNA-5481 is involved in the migration and invasion of non-small cell lung cancer by targeting the AKT1 signaling pathway," *Journal of Cancer Research and Clinical Oncology*, vol. 141, no. 3, pp. 431-441, 2015/03/01 2015, doi: 10.1007/s00432-014-1836-7.
- [155] Y. Liao, L. Cao, F. Wang, and R. Pang, "miR-605-5p promotes invasion and proliferation by targeting TNFAIP3 in non-small-cell lung cancer," *Journal of Cellular Biochemistry*, <https://doi.org/10.1002/jcb.29323> vol. 121, no. 1, pp. 779-787, 2020/01/01 2020, doi: <https://doi.org/10.1002/jcb.29323>.
- [156] C. Wang, S. Li, J. Xu, and W. Niu, "microRNA-935 is reduced in non-small cell lung cancer tissue, is linked to poor outcome, and acts on signal transduction mediator E2F7 and the AKT pathway," *British Journal of Biomedical Science*, vol. 76, no. 1, pp. 17-23, 2019/01/02 2019, doi: 10.1080/09674845.2018.1520066.
- [157] S. Badillo *et al.*, "An Introduction to Machine Learning," *Clinical Pharmacology & Therapeutics*, <https://doi.org/10.1002/cpt.1796> vol. 107, no. 4, pp. 871-885, 2020/04/01 2020, doi: <https://doi.org/10.1002/cpt.1796>.
- [158] T. Chen *et al.*, "Tumor-derived exosomal circFARSA mediates M2 macrophage polarization via the PTEN/PI3K/AKT pathway to promote non-small cell lung cancer metastasis," *Cancer Treatment and*

- Research Communications*, vol. 28, p. 100412, 2021/01/01/ 2021, doi: <https://doi.org/10.1016/j.ctarc.2021.100412>.
- [159] R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014," *CA: A Cancer Journal for Clinicians*, <https://doi.org/10.3322/caac.21208> vol. 64, no. 1, pp. 9-29, 2014/01/01 2014, doi: <https://doi.org/10.3322/caac.21208>.
- [160] L. Lerner, R. Winn, and A. Hulbert, "Lung cancer early detection and health disparities: the intersection of epigenetics and ethnicity," *Journal of Thoracic Disease*, vol. 10, no. 4, pp. 2498-2507, 2018.
- [161] C. Gérard and C. Debruyne, "Immunotherapy in the landscape of new targeted treatments for non-small cell lung cancer," *Molecular Oncology*, <https://doi.org/10.1016/j.molonc.2009.09.001> vol. 3, no. 5-6, pp. 409-424, 2009/10/01 2009, doi: <https://doi.org/10.1016/j.molonc.2009.09.001>.
- [162] N. Duma, R. Santana-Davila, and J. R. Molina, "Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment," *Mayo Clinic Proceedings*, vol. 94, no. 8, pp. 1623-1640, 2019/08/01/ 2019, doi: <https://doi.org/10.1016/j.mayocp.2019.01.013>.
- [163] M. J. Thun, S. J. Henley, and E. E. Calle, "Tobacco use and cancer: an epidemiologic perspective for geneticists," *Oncogene*, vol. 21, no. 48, pp. 7307-7325, 2002/10/01 2002, doi: 10.1038/sj.onc.1205807.
- [164] A. R. Dalby, I. Emam, and R. Franke, "Analysis of Gene Expression Data from Non-Small Cell Lung Carcinoma Cell Lines Reveals Distinct Sub-Classes from Those Identified at the Phenotype Level," *PLOS ONE*, vol. 7, no. 11, p. e50253, 2012, doi: 10.1371/journal.pone.0050253.
- [165] L. Liu *et al.*, "The Combination of the Tumor Markers Suggests the Histological Diagnosis of Lung Cancer," *BioMed Research International*, vol. 2017, p. 2013989, 2017/05/18 2017, doi: 10.1155/2017/2013989.
- [166] E. Rijavec, S. Coco, C. Genova, G. Rossi, L. Longo, and F. Grossi, "Liquid Biopsy in Non-Small Cell Lung Cancer: Highlights and Challenges," *Cancers*, vol. 12, no. 1, 2020, doi: 10.3390/cancers12010017.
- [167] G. R. Oxnard *et al.*, "Noninvasive Detection of Response and Resistance in EGFR-Mutant Lung Cancer Using Quantitative Next-Generation Genotyping of Cell-Free Plasma DNA," *Clinical Cancer Research*, vol. 20, no. 6, pp. 1698-1705, 2014, doi: 10.1158/1078-0432.CCR-13-2482.
- [168] I. Koturbash, F. J. Zemp, I. Pogribny, and O. Kovalchuk, "Small molecules with big effects: The role of the microRNAome in cancer and carcinogenesis," *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, vol. 722, no. 2, pp. 94-105, 2011/06/17/ 2011, doi: <https://doi.org/10.1016/j.mrgentox.2010.05.006>.
- [169] C. Alix-Panabières and K. Pantel, "Real-time liquid biopsy: circulating tumor cells versus circulating tumor DNA," *Annals of Translational Medicine*, vol. 1, no. 2, p. 8, 2013.
- [170] G. Siravegna, S. Marsoni, S. Siena, and A. Bardelli, "Integrating liquid biopsies into the management of cancer," *Nature Reviews Clinical Oncology*, vol. 14, no. 9, pp. 531-548, 2017/09/01 2017, doi: 10.1038/nrclinonc.2017.14.
- [171] E. Heitzer, P. Ulz, and J. B. Geigl, "Circulating Tumor DNA as a Liquid Biopsy for Cancer," *Clinical Chemistry*, vol. 61, no. 1, pp. 112-123, 2015, doi: 10.1373/clinchem.2014.222679.
- [172] C. Pedraz-Valdunciel *et al.*, "Multiplex Analysis of CircRNAs from Plasma Extracellular Vesicle-Enriched Samples for the Detection of Early-Stage Non-Small Cell Lung Cancer," *Pharmaceutics*, vol. 14, no. 10, doi: 10.3390/pharmaceutics14102034.
- [173] D. Akhoundova *et al.*, "The Role of the Liquid Biopsy in Decision-Making for Patients with Non-Small Cell Lung Cancer," (in eng), *Journal of clinical medicine*, vol. 9, no. 11, p. 3674, 2020, doi: 10.3390/jcm9113674.
- [174] Y. Li *et al.*, "Extracellular Vesicles Long RNA Sequencing Reveals Abundant mRNA, circRNA, and lncRNA in Human Blood as Potential Biomarkers for Cancer Diagnosis," *Clinical Chemistry*, vol. 65, no. 6, pp. 798-808, 4/27/2020 2019, doi: 10.1373/clinchem.2018.301291.
- [175] H. Cho, A. B. Mariotto, L. M. Schwartz, J. Luo, and S. Woloshin, "When Do Changes in Cancer Survival Mean Progress? The Insight From Population Incidence and Mortality," *JNCI Monographs*, vol. 2014, no. 49, pp. 187-197, 2014, doi: 10.1093/jncimonographs/igu014.
- [176] E. A. Klein *et al.*, "Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set," *Annals of Oncology*, vol. 32, no. 9, pp. 1167-1177, 2021, doi: 10.1016/j.annonc.2021.05.806.
- [177] S. G. J. G. In 't Veld *et al.*, "Detection and localization of early- and late-stage cancers using platelet RNA," *Cancer Cell*, vol. 40, no. 9, pp. 999-1009.e6, 2022/09/12/ 2022, doi: <https://doi.org/10.1016/j.ccell.2022.08.006>.
- [178] G. Wen, T. Zhou, and W. Gu, "The potential of using blood circular RNA as liquid biopsy biomarker for human diseases," *Protein & Cell*, 2020/11/01 2020, doi: 10.1007/s13238-020-00799-3.
- [179] T. Xu *et al.*, "CircRNAs in anticancer drug resistance: recent advances and future potential," *Molecular Cancer*, vol. 19, no. 1, p. 127, 2020/08/17 2020, doi: 10.1186/s12943-020-01240-3.

Publications



-
- **Pedraz-Valdunciel C**, Giannoukakos S, Giménez-Capitán A, *et al.* Multiplex analysis of circRNAs from plasma extracellular vesicle-enriched samples for the detection of early-stage non-small cell lung cancer. *Pharmaceutics* 2022; 14, 2034.

 - **Pedraz-Valdunciel C**, Giannoukakos S, Potie N, *et al.* Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens. *Mol Oncol.* 2022;16(12):2367-2383.

 - Rosell R, Aguilar A, **Pedraz-Valdunciel C**, *et al.* KRAS inhibitors, approved. *Nat Cancer* 2, 1254–1256 (2021).

 - Rosell R, Cardona AF, Arrieta O, *et al.* Coregulation of pathways in lung cancer patients with EGFR mutation: therapeutic opportunities. *Br J Cancer* **125**, 1602–1611 (2021).

 - **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 Nucleic Acids* 2021;2:179-201.

 - Bracht JWP, Gimenez-Capitan A, Huang CY, *et al.* Analysis of extracellular vesicle mRNA derived from plasma using the nCounter platform. *Sci Rep* **11**, 3712 (2021).

 - Chaib I, Cai X, Llige D, *et al.* Osimertinib and dihydroartemisin: a novel drug combination targeting head and neck squamous cell carcinoma. *Ann Transl Med* 2019;7(22):651.

 - Rosell R, **Pedraz- Valdunciel C**. Are neutralizing anti-VEGF or VEGFR2 antibodies necessary in the treatment of EGFR-mutated non-small cell lung cancer? *The Lancet Oncology*, 2019; 12, 1617-1618.

 - Bracht JWP, Karachaliou N, Berenguer J, *et al.* Osimertinib and pterostilbene in EGFR-mutation-positive non-small cell lung cancer (NSCLC). *Int J Biol Sci* 2019; 15(12):2607-2614.

 - Karachaliou N, Codony-Servat J, Bracht JWP, *et al.* Characterizing acquired resistance to erlotinib in non-small cell lung cancer patients, *Expert Rev. Respir. Med.* 2019, 13:10,1019-1028.

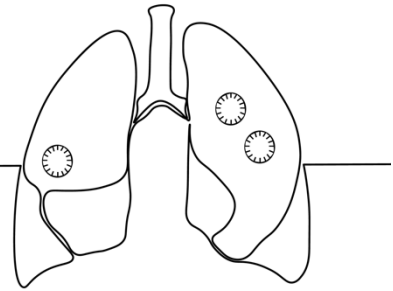
- Rosell R, Chaib I, Cai X, *et al.* Osimertinib and dihydroartemisinin: A novel drug combination targeting head and neck squamous cell carcinoma. *J. Clin. Oncol* 2019 37:15_suppl, e17526-e17526.

- Bracht JWP, Karachaliou N, Berenguer J, *et al.* PIM-1 inhibition with AZD1208 to prevent osimertinib-induced resistance in EGFR-mutation positive non-small cell lung cancer. *J Cancer Metastasis Treat* 2019;5:22.

- Filipaska M, **Pedraz-Valdunciel C**, Chaib I, Rosell R. Biological therapies in lung cancer treatment: using our immune system as an ally to defeat the malignancy, *Expert Opin. Biol. Ther.*, 19:5,457-467.

- **Pedraz-Valdunciel C**, Filipaska M, Chaib I, Rosell R. A novel miR-205-mediated ERFF1/EGFR regulatory pathway in MET-addicted cancer cells: emerging biomarkers for secondary resistance. *Non-coding RNA Investig* 2018;2:61.

Annexes



- 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 Nucleic Acids* 2021;2:179-201. <http://dx.doi.org/10.20517/evcna.2020.07>
- Pedraz-Valdunciel C, Giannoukakos S, Potie N, et al. Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens. *Mol Oncol.* 2022;16(12):2367-2383. doi:10.1002/1878-0261.13182
- Pedraz-Valdunciel C, Giannoukakos S.; Giménez-Capitán A, et al. Multiplex analysis of circRNAs from plasma extracellular vesicle-enriched samples for the detection of early-stage non-small cell lung cancer. *Pharmaceutics* 2022; 14, 2034. <https://doi.org/10.3390/pharmaceutics14102034>

Review

Open Access



Defining the landscape of circRNAs in non-small cell lung cancer and their potential as liquid biopsy biomarkers: a complete review including current methods

Carlos Pedraz-Valdunciel^{1,2}, Rafael Rosell^{1,3}

¹Cancer Biology and Precision Medicine Department, Germans Trias i Pujol Research Institute and Hospital, Badalona 08916, Spain.

²Biochemistry, Molecular Biology and Biomedicine Department, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain.

³Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain.

Correspondence to: Dr. Carlos Pedraz-Valdunciel, Cancer Biology and Precision Medicine Department, Germans Trias i Pujol Research Institute and Hospital, Camí de les Escoles, s/n, Badalon 08916, Spain. E-mail: carlospedraz@icloud.com

How to cite this article: 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 Nucleic Acids* 2021;2:179-201. <https://dx.doi.org/10.20517/evcna.2020.07>

Received: 22 Dec 2020 **First Decision:** 11 Mar 2021 **Revised:** 22 Mar 2021 **Accepted:** 2 Jun 2021 **First online:** 6 Jun 2021

Academic Editor: York Peng Loh **Copy Editor:** Yue-Yue Zhang **Production Editor:** Yue-Yue Zhang

Abstract

Despite the significant decrease in population-level mortality of lung cancer patients as reflected in the Surveillance Epidemiology and End Results program national database, lung cancer, with non-small cell lung cancer (NSCLC) in the lead, continues to be the most commonly diagnosed cancer and foremost cause of cancer-related death worldwide, primarily due to late-stage diagnosis and ineffective treatment regimens. Although innovative single therapies and their combinations are constantly being tested in clinical trials, the five-year survival rate of late-stage lung cancer remains only 5% (Cancer Research, UK). Henceforth, investigation in the early diagnosis of lung cancer and prediction of treatment response is critical for improving the overall survival of these patients. Circular RNAs (circRNAs) are a re-discovered type of RNAs featuring stable structure and high tissue-specific expression. Evidence has revealed that aberrant circRNA expression plays an important role in carcinogenesis and tumor progression. Further investigation is warranted to assess the value of EV- and platelet-derived circRNAs as liquid biopsy-based readouts for lung cancer detection. This review discusses the origin and biology of circRNAs, and analyzes their present landscape in NSCLC, focusing on liquid biopsies to illustrate the different methodological



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

trends currently available in research. The possible limitations that could be holding back the clinical implementation of circRNAs are also analyzed.

Keywords: CircRNA, extracellular vesicles lung cancer, NSCLC, liquid biopsies, biomarkers

INTRODUCTION

Lung cancer is the most commonly diagnosed cancer, contributing greatly to cancer incidence and cancer-related deaths worldwide^[1]. Of those lung cancers, non-small cell lung cancer (NSCLC) accounts for 85% of the cases; the development of the disease is attributed to multileveled and elusive complex interactions between genetic liabilities, sex, environmental toxins, and imbalanced signaling processes.

Although the mortality rate of NSCLC has decreased in previous years, presumably due to the approval and routinization of targeted therapies and immunotherapies^[2], the prognosis in late-stage lung cancer remains dismal. While the 5-year overall survival (OS) of early-stage lung cancer is 85% (stage IA), these numbers fall to only 5% in late-stage cases (stage IV). In addition to tumor tissue characterization, liquid biopsies have been introduced to overcome, or complement, invasive tissue biopsies.

Not only are they instrumental in achieving early detection of the tumor, but they can also be exploited to monitor therapy resistance and provide a more heterogeneous readout of the tumor burden^[3]. This allows the identification of resistance mechanisms and can guide second-line therapy selection.

Different body fluids can be used as liquid biopsies, including blood, urine, and saliva. Circulating molecules, such as cell-free DNA (cfDNA), RNA, or proteins, can either be freely present within these media or can be extracted and analyzed from circulating extracellular vesicles (EVs) or tumor-educated platelets (TEPs)^[4].

Lung cancer involves massive changes in RNA metabolism, both in the tumor and circulating EVs and TEPs. Traditional RNA biomarker discovery research for either lung cancer detection or monitoring of treatment response has mainly focused on the expression of mRNA and miRNA^[5-7].

Circular RNAs (circRNAs) are a recently re-discovered type of RNA generated by coupling the 5' and 3' ends in a non-canonical process known as back-splicing^[8]. This circular structure lacks a poly(A) tail, making most of them resistant to the exonuclease RNase R and, therefore, making them robustly stable molecules compared to lineal mRNA. While thousands of circRNAs have been described thanks to the technological burst of deep sequencing^[9], only the function of a fraction has been elucidated.

Recent investigations have unveiled the role of circRNAs as important players in NSCLC, positioning them as valuable biomarkers for early detection and promising candidates for seeking therapeutic and prevention strategies^[10].

This review analyzes the current state of circRNA research, starting from their biology to their different functions and implications in NSCLC, with a special focus on their not yet fully exploited potential as liquid biopsy biomarkers. We also review the most recently discovered circRNAs, both in solid and liquid specimens.

In addition, we provide a practical and complete guide on the current methodology available for their study, stressing the current limitations that may be preventing their implementation in the clinical setting.

CIRCULAR RNA EXPRESSION IN HUMANS

Although circRNAs have been acknowledged for many years as abnormally spliced “scrambled” transcripts^[11], only recently have they been re-defined as biologically active molecules with a significant role in human homeostasis, having a tissue-specific expression profile during the different stages of development^[12].

More than 60% of human genes can express circRNAs^[13]. However, their expression levels in tissue remain rather low, accounting for only 5%-10% of the canonical (linear) mRNA expression^[14,15].

CircRNAs are originated by an alternative process called “back-splicing”, where the 5' splice donor can stick to the 3' splice acceptor of an upstream exon. This process results in forming a circular structure that can include one or different exonic/intronic regions, depending on the specific mechanism that was inferred during this non-canonical process^[16].

They have arisen as key post-transcriptional regulators throughout different functions [Figure 1], with micro-RNA (miRNA) sponging being the most studied. During this process, the circRNA binds to the argonaute-miRNA complex, and either via miRNA degradation or inhibition of the miRNA-mRNA interaction, it triggers further mRNA expression^[17].

Recent studies have also revealed that circRNAs could associate with ribosomes and be translated into functional short peptides, in a cap-independent manner^[18]. Alternatively, they can also associate with proteins acting as scaffolding for enzymatic reactions. The process of circRNA synthesis generates an imbalance of the canonical splicing; hence, the back-splicing process itself stands as a direct regulator of the circRNA precursor gene at the transcriptional level.

Biosynthesis and regulation of circRNAs

Different back-splicing mechanisms have been reported in the nucleus, including RNA binding protein (RBP)-mediated circularization, circRNA synthesis by intron pairing, or circularization by intron-lariat formation^[16] [Figure 1]. The first mechanism is normally executed by associating two adjacent exons and skipping the intronic region during an RBP-assisted circularization process, resulting in an exonic-circRNA (EcircRNA). Numerous RBPs have been described to regulate this mechanism, such is the case of the adenosine deaminase RNA specific-1 protein (ADAR1)^[19], NF90/NF110 immune factors^[20], muscleblind transcription factor (MBL)^[21], heterogeneous nuclear ribonucleoprotein L^[22], FUS protein^[23], Quaking binding protein (QKI)^[24], RNA helicase DHX9^[25], and the RNA-binding motif protein 20^[26].

Exon-intron circRNAs are the result of 2 or more exons circularized along with their corresponding introns via intron-lariat formation. Intron pairing back-splicing is usually the common process in conserved RNAs with high frequency of *Alu* repeats in flanking sequences. These *Alu* elements complement each other, promoting the hairpin formation and further back-splicing, creating mono-EcircRNAs as a result^[27]. Intronic circRNAs are another type of such a class; however, the mechanism of generation of these molecules remains yet unclear.

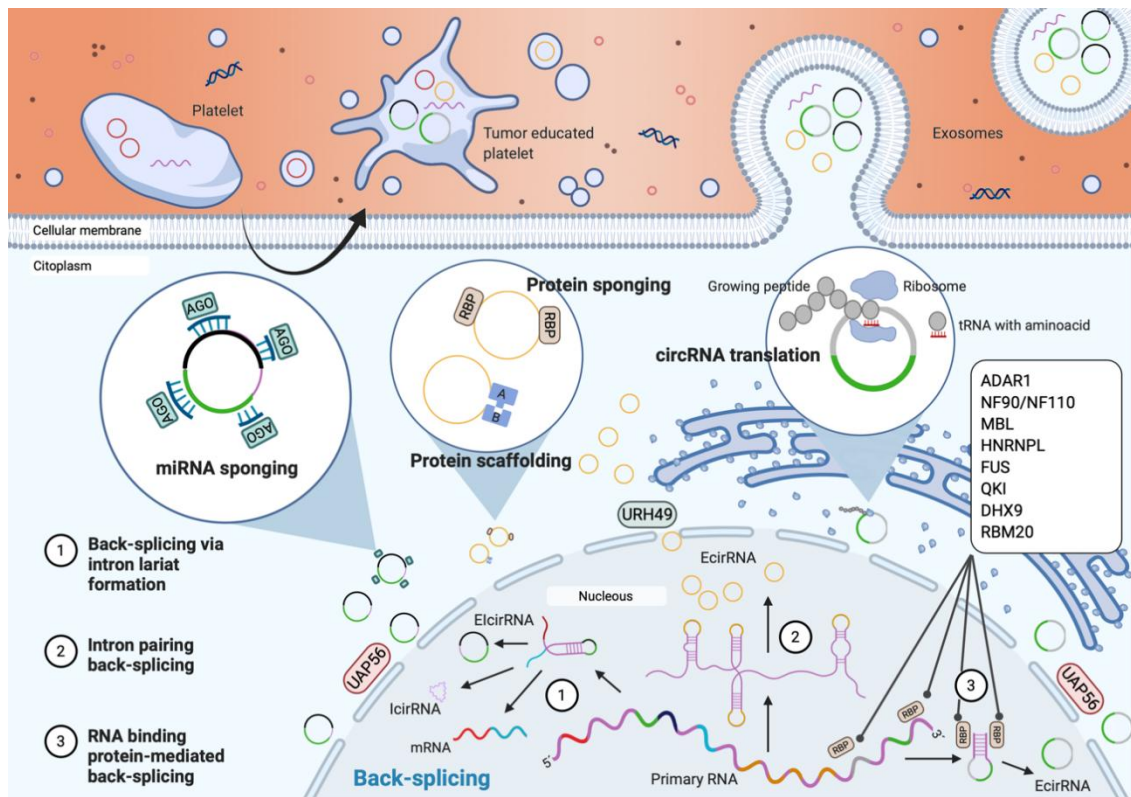


Figure 1. Biosynthesis and molecular functions of circRNAs. CircRNAs are generated by three different mechanisms of back-splicing (via lariat formation, intron pairing or RNA binding proteins). Resultant circRNAs can be formed by only exonic regions (EcircRNAs), intronic regions (lircRNAs) or both (ElcircRNAs). circRNAs are exported into the cytoplasm in a size-mediated manner by URH49 and UAP56. Once in the cytoplasm, circRNAs will perform their functions including miRNA and protein sponging, protein scaffolding, or even translate into small functional peptides. CircRNAs will be released into the blood stream inside exosomes mediating cellular communication. Most cellular types, including tumor cells, will secrete circRNA-containing EVs. Platelets can modify its content when in contact with the tumor, including their circRNA expression profile.

After synthesis in the nucleus, circRNAs are exported into the cytoplasm. Recent studies have shown the active role of the UAP56/URH49 helicases in this size-mediated process. UAP56 is required to transfer molecules longer than 1300 nucleotides, while URH49 intervenes only in short transcript exporting^[28]. Once in the cytoplasm, circRNAs accumulate and exert their function by regulating transcription, normally via sponging targeted miRNAs.

How circRNA gets degraded still remains unclear; however, recent investigation has shed light on this conundrum, unveiling some intriguing mechanisms that underpin circRNA decay. Hansen *et al.*^[29] describe an Ago2-miR-671-mediated degradation of the circRNA CDR1as (aka ciRS-7). In another study by Park *et al.*^[30], a cleavage mechanism induced by RNase P/MRP was elucidated in N6-methyladenosine (m6A)-enriched circRNAs. More recently, a study by Liu *et al.*^[31] demonstrated that some circRNAs tend to form intricate duplexes which makes them susceptible to degradation by RNase L upon viral infection.

A different mechanism was described by Fischer *et al.*^[32] revealing an alternative structure-mediated circRNA regulation process that selectively degrades circRNAs based on 3'-UTR structure complexity via the UPF1/G3BP1 protein complex.

CIRCULAR RNAS IN NSCLC

The implication of circRNAs in cancer metabolism has been studied in recent years. Their contribution to mutant glycolysis (via transporter, enzyme, and/or transcription factor regulation), lipogenesis and lipolysis, glutaminolysis, and oxidative respiration has been widely demonstrated^[33].

CircRNAs are becoming a new area of interest within cancer research, including NSCLC, where several authors are contributing by investigating the effect that dysregulated circRNA expression can have on the different cancer stages. Although their implication in NSCLC has not been as intensively investigated as other types of non-coding RNAs, circRNAs have been shown to have a significant role in tumorigenesis, tumor development, proliferation, migration, invasion, and sensitivity to NSCLC therapy^[34]. In light of these aforementioned findings, recent publications highlight the potential of these circular transcripts as plausible biomarkers to assess disease status.

CircRNAs as biomarkers of NSCLC

The number of studies on circRNA profiling in NSCLC patients has exploded exponentially in the last few years [Table 1].

ciRS-7 was the first and best characterized circRNA in cancer and served as a foundation stone for current research. Its role in carcinogenesis was first described in hepatocellular carcinoma, following breast and cervical cancer, acting as a competing endogenous RNA for miR-7^[35]. A recent study has introduced ciRS-7 as an important player in lung cancer; its expression seems to correlate with tumor size and both lymph and tumor node metastasis stages^[36].

A study by Wang *et al.*^[37] recently demonstrated the involvement of circSOX4 in lung adenocarcinoma by activating the WNT signaling pathway via sponging miR-1270 and following upregulation of PLAL2. CircSOX4 was found overexpressed in all managed lung adenocarcinoma tissue samples, and further validated across different cell-based preclinical experiments^[37].

Circular RNA HIPK3 (circHIPK3) is yet another extensively studied circRNA critical in cell proliferation of different types of cancer^[38]. Its specific role in NSCLC has been recently discovered by Xie *et al.*^[39] demonstrating impaired cell proliferation, migration, invasion and autophagy induction via the miR124-3p-STAT3-PRKAA/AMPKa axis upon silencing of the cited circular transcript. Authors also demonstrated that overexpression of circHIPK3 correlates to poor survival, especially in advanced stages.

Another well studied circRNA, circSMARCA5, plays a significant role in NSCLC via the miR-19b-3p/HOXA9 axis, setting the grounds for exploring underlying therapeutic targets^[40]. On a similar note, a circular RNA from FGFR3 was reported in NSCLC, promoting cell invasion and proliferation of tumors by sequestering miR-22-3p, thus promoting galectin-1, p-AKT, and p-ERK1/2 expression, and activating downstream pathways^[41].

The oncogenic circ-FOXM1 was first discovered overexpressed in pancreatic tissues upregulating the pancreatic progenitor cell differentiation and proliferation factor (PPDPF) and metastasis-associated in colon cancer 1 (MACC1) proteins via miR-1304-5p sponging. More recently, the same circ-FOXM1/miR-1304-5p/PPDPF/MACC1 axis was found decisive for NSCLC development and progression^[42].

Table 1. List of the most relevant recently discovered circRNAs associated with lung cancer

circRNA	Gene	CircBase ID	Source	Regulation	Target	Downstream pathway
circFGFR3	<i>FGFR3</i>	-	NSCLC tissues	Upregulated	hsa-miR 22 3p	Galectin 1 AKT/ERK1/2
ircNOL10	<i>NOL10</i>	hsa_circ_0000977	LC cells	Downregulated	hsa-miR-7	SCML1
ciRS-7	<i>CDR1</i>	-	NSCLC tissues and cell lines	Upregulated	-	-
circABCC4	<i>ABCC4</i>	hsa_circ_0030586	LUAC tissues and cell lines	Upregulated	hsa-miR 3186 3p	TNRC6B axis
circCDR1	<i>CDR1</i>	hsa_circ_0001946	LUAC tissues and cell lines	Upregulated	hsa-miR-135a-5p	SIRT1/Wnt/ β -catenin
circATXN7	<i>ATXN7</i>	hsa_circ_0007761	LC tissues and cell lines	Upregulated	-	-
circATAD3B	<i>ATAD3B</i>	hsa_circ_0000003	NSCLC tissues and cell lines	Upregulated	hsa-miR-338-3p	IRS2
circP2RX1	<i>P2RX1</i>	hsa_circ_0000735	NSCLC tissues and cell lines	Upregulated	hsa-miR-1179, miR-1182	-
circC16orf62	<i>C16orf62</i>	hsa_circ_0003645	NSCLC tissues and cell lines	Upregulated	hsa-miR-1179	TMEM14A
circPDZD8	<i>PDZD8</i>	hsa_circ_0020123	NSCLC tissues and cell lines	Upregulated	hsa-miR-488e3p	ADAM9
circTUBA1C	<i>TUBA1C</i>	hsa_circ_0026134	NSCLC tissues and cell lines	Upregulated	hsa-miR-1256, miR-12	TCTN1 and GAGE1
circCAMK2A	<i>CAMK2A</i>	hsa_circ_0128332	LUAD	Upregulated	hsa-miR-615-5p	Fibronectin 1
circFOXN1	<i>FOXN1</i>	hsa_circ_0025033	NSCLC tissues and cell lines	Upregulated	hsa-miR-1304-5p	PPDPF and MACC1
circMTO1	<i>MTO1</i>	hsa_circ_0007874	LUAD tissues and cell lines	Downregulated	hsa-miR-17	QKI-5
circPRMT5	<i>PRMT5</i>	hsa_circ_0031250	NSCLC tissues and cell lines	Upregulated	hsa-miR-377/382/498	EZH2
circRAD23B	<i>RAD23B</i>	hsa_circ_0087855	NSCLC tissues and cell lines	Upregulated	hsa-miR-593e3p, hsa-miR-653e5p	CCND2 and TIAM1
circZKSCAN1	<i>ZKSCAN1</i>	hsa_circ_0001727	NSCLC tissues and cell lines	Upregulated	hsa-miR-330-5p	FAM83A (MAP signaling)
circCRIM1	<i>CRIM1</i>	hsa_circ_0002346	LUAC cell lines	Downregulated	hsa-miR 182/miR 93	-
circHIPK3	<i>HIPK3</i>	hsa_circ_0000284	A549, H838 cell lines	Upregulated	hsa-miR-124-3p, miR-149	STAT3-PRKAA/AMPK α
circPDK1	<i>PDK1</i>	hsa_circ_0006006	LUSC tissues	Upregulated	-	-
circPIP5K1A	<i>PIP5K1A</i>	hsa_circ_0014130	NSCLC cell lines	Upregulated	hsa-miR 600	HIF-1 α
circPRKCI	<i>PRKCI</i>	hsa_circ_0067934	NSCLC cell lines	Upregulated	hsa-miR-545, hsa-miR-589	E2F7
circPTPRA	<i>PTPRA</i>	hsa_circRNA_0102984	NSCLC tissues and cell lines	Downregulated	hsa-miR-96-5p	RASSF8/E-cadherin
circPVT1	<i>PVT1</i>	Hsa_circ_0001821	NSCLC tissues and cell lines	Upregulated	hsa-miR-497	-
circTP63	<i>TP63</i>	hsa_circ_0068515	LUSC tissues and cell lines	Upregulated	hsa-miR-873-3p	FOXN1/CENPA-CENPB
circVANGL1	<i>VANGL1</i>	-	NSCLC tissues and cell lines	Upregulated	hsa-miR-195	Bcl-2
circZFR	<i>ZFR</i>	hsa_circ_0001649	NSCLC tissues and cell lines	Upregulated	hsa-miR-101-3p	CUL4B
circMras	<i>MRAS</i>	hsa_circ_0067512	LUAC samples and NSCLC cell lines	Downregulated	hsa-miR 567	PTPRG
F-circSR	<i>SLC34A2-ROS1</i>	-	HCC78 cell line	Upregulated	-	ROS
circCDK6	<i>CDK6</i>	hsa_circ_000984	NSCLC tissues and cell lines	Upregulated	-	Wnt/ β -catenin pathway
circRUNX1	<i>RUNX1</i>	hsa_circ_0002360	LUAC tissues	Upregulated	hsa-mir-3620-5p	PHF19

circZNF720	ZNF720	hsa_circ_0007059	LC tissues and cell lines	Downregulated	hsa-miR-378	Wnt/ β -catenin and ERK1/2
circRNF121	RNF121	hsa_circ_0023404	NSCLC tissues and cell lines	Upregulated	hsa-miR-217	ZEB1
circTADA2A	TADA2A	hsa_circ_0043278	NSCLC tissues and cell lines	Upregulated	hsa-miR-520f	ROCK1, CDKN1B and AKT3
circLIFR	LIFR	hsa_circ_0072309	NSCLC tissues and cell lines	Downregulated	hsa-miR-580-3p	-
circITCH	ITCH	N.A.	LC tissues and cell lines	Downregulated	hsa-miR-7 and hsa-miR-214	(PI3K)/AKT
circSMARCA5	SMARCA5	hsa_circ_0001445	NSCLC tissues and cell lines	Downregulated	hsa-miR-19b-3p	HOXA9
circRAD23B	RAD23B	hsa_circ_0087862	NSCLC tissues and cell lines	Upregulated	hsa-miR-1253	RAB3D
circPIP5K1A	PIP5K1A	hsa_circ_0014130	NSCLC tissues and cell lines	Upregulated	hsa-miR-142-5p, hsa-miR-136-5p	IGF-1 and BCL2
circABC10	ABC10	hsa_circ_0008717	NSCLC tissues and cell lines	Upregulated	-	KISS1
circIGF1R	IGF1R	hsa_circ_0005035	NSCLC tissues and cell lines	Downregulated	hsa-miR-1270	VANGL2
circSOX4	SOX4	N.A.	LUAD tissues and cell lines	Upregulated	hsa-miR 1270	PLAGL2 (WNT signaling)
circACACA	ACACA	hsa_circ_0043256	NSCLC tissues and cell lines	Upregulated	hsa-miR-1183	PI3K/PKB pathway
circBIRC6	BIRC6	hsa_circ_0003288	NSCLC tissues and cell lines	Upregulated	hsa-miR-145	FSCN1 and S6K1
circCCDC66	CCDC66	N.A.	NSCLC cell lines	Upregulated	hsa-miR-33a-5p	KPNA4/STAT3
circGFRA1	GFRA1	hsa_circ_0005239	NSCLC tissues and cell lines	Upregulated	hsa-miR-188-3p	PI3K/AKT
circLARP4	LARP4	N.A.	NSCLC tissues and cell lines	Downregulation	-	SMAD7
circTCONS	TCONS	hsa_circ_0000326	NSCLC tissues and cell lines	Upregulated	hsa-miR-338-3p	RAB14
circDHCR24	DHCR24	hsa_circ_0012673	LC tissues and cells	Upregulated	hsa-miR-320a	LIMK18521
circMACF1	MACF1	hsa_circ_0011780	NSCLC tissues and cells	Downregulated	hsa-miR-544a	FBXW7
circPANX2	PANX2	hsa_circ_0012515	NSCLC tissues and cells	Upregulated	hsa-miR-98-5p, hsa-miR-615-5p, hsa-let-7a-5p, hsa-let-7b-5p and hsa-let-7c-5p	-
circMET	MET	hsa_circ_0082003	NSCLC tissues and cells	Upregulated	miR-145-5p	CXCL3

Chromosomal translocations are cancer-associated events that may strike frequently in some genes, like *ROS* or *ALK*, leading to activation of downstream signaling pathways upon sustained expression^[43]. These events can also generate oncogenic circRNAs, as has been reported with the solute carrier family 34 member 2 (*SLC34A2*) and *ROS* proto-oncogene 1 (*ROS1*), producing two circRNAs (*F-circSR1* and *F-circSR2*) both promoting cell migration in NSCLC^[44].

Precursor mRNA of driver mutations, such as *MET*, can also lead to the generation of circRNAs. *CircMET* was first described in hepatocellular carcinoma driving immunosuppression and anti-programmed cell death 1 (PD-1) therapy resistance via the miR-30-5p/*snail*/*DPP4* axis^[45]. Its role in NSCLC was recently discovered promoting tumor proliferation via the miR-145-5p/*CXCL3* axis^[46].

Although a circRNA from epidermal growth factor receptor (EGFR) has been reported in mouse ovaries during postnatal development with a marked expression profile, the implication of this circRNA in lung cancer has not been studied yet.

There have been no circRNAs derived from the *KRAS* gene reported either; however, numerous circRNAs have been portrayed as key intermediaries of the classical pathways and may serve as a readout of these foremost altered genes.

CircRNAs as biomarkers of treatment resistance in NSCLC

Although several studies have unveiled the potential role of circRNAs in lung cancer development and progression, not much has been clarified regarding their contribution to therapeutic resistance, and only a few published studies focus on their involvement in this area [Table 2]. CircRNAs can be classified as promoters, when their high expression enhances resistance to cancer therapy; or suppressors, when their expression limits the progression of the disease during treatment, thus acting as inhibitors of resistance.

Astrocyte elevated gene-1 (AEG-1) is a key player in development, progression, and metastasis of lung cancer by regulating the Wnt/ β -catenin pathway. In a recent publication, Li *et al.*^[47] showed that circMTDH.4 regulates AEG-1 expression by sponging miR-630, leading to chemo- and radio-resistance in NSCLC cells. Sensitivity was restored via the knockdown of the cited circRNA or over expression of its target, miR-630.

Two different works have recently been published describing circRNAs that regulate the expression of STAT3. Dong *et al.*^[48] reported that upregulation of hsa_circ_0076305 confers DDP-resistance to NSCLC cells via sponging miR-296-5p, positively modulating STAT3. Xu *et al.*^[49] introduced the role of circAKT3 inhibiting cisplatin sensitivity by regulating mir-516b-5p/STAT3 axis.

Other important circRNAs described to be involved in chemotherapy resistance are hsa_circ_0071799 via miR-141 (taxol resistance)^[50], hsa_circ_0091931 via miR-34c-5p^[10], hsa_circ_0003998 via miR-326^[51], hsa_circ_0001946 via miR-7-5p, miR-671-5p, miR-1270 and miR-3156-5p (NER signaling, cisplatin resistance)^[52], circPVT1 via miR-145-5p (ABCC1, cisplatin, and pemetrexed resistance)^[53], circNFIX via miR-132 (TMZ-resistant)^[54], and cESRP1. Huang *et al.*^[55] recently discovered a suppressor circRNA that, when downregulated, allows major expression of its target miR-93-5p. This process leads to the upregulation of downstream targets, such as Smad7/p21(CDKN1A), enhancing the transforming growth factor- β (TGF- β) pathway. Furthermore, cESRP1 overexpression boosts cisplatin sensitivity by repressing miR-93-5p and TGF- β pathway in SCLC. Related to this pathway, PDPK1, intermediary of the PI3K/AKT/mTOR pathway, has been discovered to be regulated by the hsa_circ_0004015-miR-1183 axis^[56]. Overexpression of this circRNA can induce gefitinib resistance in NSCLC cells by sponging the abovementioned miRNA.

Other authors have centered their investigation on the differential expression of circRNAs that confer resistance to this and other tyrosine kinase inhibitor-based therapies. Fu *et al.*^[57] found hsa_circRNA_012515 increased in gefitinib-resistant NSCLC cell lines. Further investigation in patient tissue indicated that high expression correlated with lower OS and shorter progression free survival. Chen *et al.*^[58] found 10 differentially expressed circRNAs in different osimertinib-resistant lung cancer cell lines. Five of them were further validated and proved to correlate with resistance status (hsa_circ_0043632, hsa_circ_0048856, hsa_circ_0043634, hsa_circ_0050581, and hsa_circ_0023302)^[58]. The authors made use of specific software to predict possible targeted miRNAs; however, the axis or mechanism of action has not yet

Table 2. List of circRNAs involved in NSCLC treatment resistance

circRNA	Gene	CircBase ID	Source	Regulation	Resistance	Drug	Target	Downstream pathway	Ref.
circSEMA5A	SEMA5A	hsa_circ_0071799	NSCLC cells	Upregulated	Chemotherapy	Taxol	hsa-miR-141-5p; also, hsa-miR-1228-5p, hsa-miR-194-3p, hsa-miR-512-5p, hsa-miR-4-5p	-	Xu <i>et al.</i> ^[50] , 2018
circFLNA	FLNA	hsa_circ_0091931	NSCLC cells	Downregulated	Chemotherapy	Taxol	hsa-miR-34c-5p; also, hsa-miR-105-3p, hsa-miR-1268b, hsa-miR-1226-5p, hsa-miR-1180	-	Xu <i>et al.</i> ^[50] , 2018
circMTDH.4	SNORD115	-	NSCLC tissue and cell lines	Upregulated	Chemotherapy	5 FU, cisplatin	hsa-miR-630	AEG 1	Li <i>et al.</i> ^[47] , 2020
circESRP1	ESRP1	hsa_circ_0084927	Lung cancer cells	Downregulated	Chemotherapy	Generic chemotherapy	hsa-miR-93-5p	TGF- β pathway	Huang <i>et al.</i> ^[55] , 2020
circARFGEF2	ARFGEF2	hsa_circ_0003998	LUAC cells	Upregulated	Chemotherapy	Docetaxel	hsa-miR-326	-	Yu ^[51] , 2019
circCDR1	CDR1	hsa_circ_0001946	A549 cell line	Downregulated	Chemotherapy	Cisplatin	hsa-miR-7-5p, hsa-miR-671-5p, hsa-miR-1270, hsa-miR-3156-5p	NER signaling	Huang <i>et al.</i> ^[52] , 2019
circPGC	PGC	hsa_circ_0076305	NSCLC tissues and cell lines	Upregulated	Chemotherapy	DDP	hsa-miR-296-5p	STAT3	Dong <i>et al.</i> ^[48] , 2019
circAKT3	AKT3	hsa_circ_0017252	Lung cancer tissues and cell lines	Upregulated	Chemotherapy	DDP, cisplatin	hsa-miR-516b-5p	STAT3	Xu <i>et al.</i> ^[49] , 2020
circPVT1	PVT1	hsa_circ_0001821	LUAC tissues and cell lines	Upregulated	Chemotherapy	Cisplatin, pemetrexed	hsa-miR-145-5p	ABCC1	Zheng <i>et al.</i> ^[53] , 2020
circCDK14	CDK14	hsa_circ_0004015	NSCLC cells	Upregulated	Tyrosine Kinase Inhibitors (TKIs)	Gefitinib	hsa-miR-1183	PDPK1 gene	Zhou <i>et al.</i> ^[56] , 2019
circKRT17	KRT17	hsa_circ_0043632	AZD9291-resistant NSCLC cell lines	Upregulated	Tyrosine Kinase Inhibitors (TKIs)	Osimertinib	hsa-miR-6861-3p, hsa-miR-492, hsa-miR-4743-5p, hsa-miR-6829-3p, hsa-miR-6778-3p	-	Chen <i>et al.</i> ^[58] , 2019
circFXVD3	FXVD3	hsa_circ_0050581	AZD9291-resistant NSCLC cell lines	Downregulated	Tyrosine Kinase Inhibitors (TKIs)	Osimertinib	hsa-miR-6722-5p, hsa-miR-4641, hsa-miR-4707-3p, hsa-miR-4258, hsa-miR-652-3p	-	Chen <i>et al.</i> ^[58] , 2019
circFGFR1	FGFR1	hsa_circ_0084003	NSCLC tissues and cells	Upregulated	Immunotherapy	Anti-PD-1 therapy	hsa-miR-381-3p	PD-1	Zhang <i>et al.</i> ^[59] , 2019

been elucidated.

CircRNAs seem to also have a role mediating response to immunotherapy. CircFGFR1 has been described by Zhang *et al.*^[59] to promote progression and anti-PD-1 resistance. By sponging miR-381-3p in NSCLC cells, C-X-C motif chemokine receptor 4 would result upregulated, leading to progression and resistance

to therapy.

CURRENT LANDSCAPE OF CIRCULAR RNAs IN LIQUID BIOPSIES AS NSCLC BIOMARKERS

Non-coding RNA-enriched exosomes are strategic players in different cancer stages, especially regarding malignant tumor metastasis^[60]. The assessment of circRNA expression by RNAseq analysis in extracellular vesicles was first reported by Li *et al.*^[61], finding circRNAs enriched at least 2-fold in exosomes compared to producer cells. Although some authors defend the theory that exosomal circRNA enrichment may be a mechanism of cellular circRNA clearance^[62], few investigators have shown that these circRNAs are directly involved in cellular communication, henceforth, acting as direct readouts of several human malignancies, including NSCLC^[63].

As a result, circRNAs stand as important liquid biopsy-derived biomarkers, holding potential for NSCLC diagnosis and prediction of treatment response^[64].

In a recent study, Chen *et al.*^[65] performed high throughput sequence of plasma-EV RNA cargo of lung adenocarcinoma patients, finding 182 circRNAs dysregulated when compared to cancer-free donors, including 105 up-regulated and 78 downregulated. Four upregulated circRNAs were successfully validated by qRT-PCR (hsa_circ_0001492, hsa_circ_0001346, hsa_circ_0000690, and hsa_circ_0001439)^[65]. Although authors elucidated the specific circRNA-miRNA-mRNA interaction, not much information about their biological impact was provided.

Fei *et al.*^[66] also presented in a recent study a novel circRNA, hsa_circRNA_005661, that could be found enriched in plasma EVs from lung adenocarcinoma patients with lymph node metastasis, presenting it as a biomarker of such stage^[66].

Not only plasma-EVs, but serum and whole plasma can serve as a good source of circRNAs [Table 3]. Xian *et al.*^[67] studied the circRNA differential expression profile in serum EVs from NSCLC patients. As a result, 3 circRNAs stood out showing suitable biomarker potential (hsa_circ_0047921, hsa_circ_0007761, and hsa_circ_0056285) with the later correlating with clinical stages and lymph node metastasis in all Chinese patients included in the study^[67].

Hang *et al.*^[68] explored the use of circRNA found in total plasma of NSCLC patients in order to find some candidates that could correlate to malignancy status. Not only did they find a notorious circRNA coming from the *FARSA* gene, *circFARSA*, but they also found a set of differentially expressed circRNAs (hsa_circ_0001495, hsa_circ_0000566, hsa_circ_0001238, hsa_circ_0007037, circ_c1orf116, hsa_circ_0001083, hsa_circ_0006451, hsa_circ_0004458, and hsa_circ_0000847) based on which they were able to discriminate NSCLC patients from healthy individuals. Additionally, they performed *in silico* investigation of possible targets of *circFARSA*. Consequently, miR-330-5p and miR-326 emerged as direct target candidates. Both miR-330-5p and miR-326 may interact directly with fatty acid synthase, which has been described as a notorious oncogene in various types of cancer^[68].

Also, directly from plasma Liu *et al.*^[69] found a two circRNA-based signature that could potentially be used to classify lung adenocarcinoma patients. Hsa_circ_0005962 was found upregulated while hsa_circ_0086414 was barely expressed. In addition, they observed that overexpression of hsa_circ_0005962 was correlated to mutant *EGFR* expression. *In vitro* experiments suggested that this circRNA could be involved in cancer proliferation.

Table 3. List of the most relevant liquid biopsy-based circRNAs associated with NSCLC

circRNA	Gene	circBase ID	Source	Expression	Target	Ref.
circERBB2IP	<i>ERBB2IP</i>	hsa_circ_0001492	LUAD plasma exosomes	Upregulated	hsa-miR-130b-5p, hsa-miR-5195-3p, hsa-miR-4464, hsa-miR1236-3p, hsa-miR-106a-3p	Chen <i>et al.</i> ^[65] , 2019
circRNF13	<i>RNF13</i>	hsa_circ_0001346	LUAD plasma exosomes	Upregulated	hsa-miR-34B-5P, ha-miR-654-3p, hsa-miR-5683, hsa-miR-4452, hsa-miR-4662b	
circITGAL	<i>ITGAL</i>	hsa_circ_0000690	LUAD plasma exosomes	Upregulated	hsa-miR-7161-3p, hsa-miR-9-5p, hsa-miR-6843-3p, hsa-miR-4502, miR-372-5p	
circSCLT1	<i>SCLT1</i>	hsa_circ_0001439	LUAD plasma exosomes	Upregulated	hsa-miR-3671, hsa-miR-452-5p, hsa-miR-892c-3p, hsa-miR-223-3p, hsa-miR-4676-3p	
circCD226	<i>CD226</i>	hsa_circ_0047921	NSCLC serum exosomes	Downregulated	hsa-miR-let-7g	Xian <i>et al.</i> ^[67] , 2020
circATXN7	<i>ATXN7</i>	hsa_circ_0007761	NSCLC serum exosomes	Upregulated	-	
circRALB	<i>RALB</i>	hsa_circ_0056285	NSCLC serum exosomes	Downregulated	-	
circNPHP4	<i>NPHP4</i>	hsa_circ_0005661	LUAD plasma exosomes	Upregulated	-	He <i>et al.</i> ^[66] , 2020
circFARSA	<i>FARSA</i>	hsa_circ_0000896	NSCLC plasma	Upregulated	hsa-miR-330 5p, hsa-miR-326, hsa-miR-1270	Hang <i>et al.</i> ^[68] , 2018
circCCCNB1	<i>CCCNB1</i>	hsa_circ_0001495	NSCLC plasma	Upregulated	-	
circVRK1	<i>VRK1</i>	hsa_circ_0000566	NSCLC plasma	Upregulated	-	
circCCDC134	<i>CCDC134</i>	hsa_circ_0001238	NSCLC plasma	Upregulated	-	
circZCCCJC6	<i>ZCCCJC6</i>	hsa_circ_0007037	NSCLC plasma	Upregulated	-	
circ_c1orf116	<i>C1ORF116</i>	hsa_circ_0141539	NSCLC plasma	Upregulated	-	
circPMS1	<i>PMS1</i>	hsa_circ_0001083	NSCLC plasma	Upregulated	-	
circDNA2	<i>DNA2</i>	hsa_circ_0006451	NSCLC plasma	Upregulated	-	
PcircSD3	<i>SD3</i>	hsa_circ_0004458	NSCLC plasma	Upregulated	-	
circSMAD2	<i>SMAD2</i>	hsa_circ_0000847	NSCLC plasma	Upregulated	-	
circYWHAZ	<i>YWHAZ</i>	hsa_circ_0005962	LUAD plasma	Upregulated	hsa-miR-369-5p, hsa-miR-626, hsa-miR-326, hsa-miR-330-5p, hsa-miR-1265, and hsa-miR-622	Liu <i>et al.</i> ^[69] , 2019
circBNC2	<i>BNC2</i>	hsa_circ_0086414	LUAD plasma	Downregulated	-	
F-circEA	<i>EMLK4-ALK</i>		Lung cancer tissues, plasma and cells	Upregulated	-	Tan <i>et al.</i> ^[70] , 2018
circZNF91	<i>ZNF91</i>	hsa_circ_0109320	NSCLC plasma	Downregulated	-	Liu <i>et al.</i> ^[74] , 2019
circZNF117	<i>ZNF117</i>	hsa_circ_0134501	NSCLC plasma	Upregulated	-	

Moreover, a fusion-gene *circRNA* has been studied in liquid biopsies. Tan *et al.*^[70] started their line of research exploring the existence of a circRNA derived from the fusion gene *EML4-ALK* (*F-circEA*) in the NCI-H2228 cell line. After verification, they observed that overexpression of this circRNA could trigger cell migration and invasion, contributing to tumor development. They validated the existence of this circRNA in plasma of NSCLC patients with the *EML4-ALK* translocation, suggesting that screening of plasma *F-circEA* in this type of patients could be a valuable approach to monitor the *EML4-ALK* translocation, and provide further guidance on targeted therapy.

Alhasan *et al.*^[71] showed for the first time that platelets^[71] are enriched in circRNAs when compared to nucleated tissues, and also, that their content is superior to that on mRNA. Preußner *et al.*^[72] demonstrated that platelets are not only a good source of circRNA, but also platelet-derived extracellular vesicles are

enriched in these biomolecules, representing yet another source of potential biomarkers that may be involved in different signaling pathways.

Platelets change their RNA profile when in contact with the tumor, enabling them to contribute to the systemic and local responses to tumor growth. As a result, TEP-RNA can be used as a potential biomarker for cancer diagnostics^[73]. Although TEPs could also possibly be enriched in circRNAs, and hold potential value for NSCLC diagnosis, nothing yet has been investigated.

Little has been elucidated regarding NSCLC treatment resistance based on liquid biopsy-based circRNAs. A study of Yu-Tao *et al.*^[74] comparing gefitinib responder and non-responder NSCLC patients found that higher expression of hsa_circ_0109320 in plasma correlated with longer progression free survival in gefitinib-treated NSCLC patients^[74]; however, no information on the potentially affected signaling pathway has been provided.

Current available methods for the study of circRNAs in liquid biopsies

Although there are different methods currently available for the study of circRNAs [Table 4], no consensus has been reached on which protocol to follow for either tissue or liquid biopsy-based circRNA expression analysis.

The range of possibilities when selecting a bio-source is rather ample^[75]. Whilst plasma or serum can provide a higher yield of total RNA, tumor released EVs stand out by providing a more accurate picture of lung cancer at the transcriptional level^[76]. Procedures such as ultracentrifugation, ultrafiltration, or size-exclusion chromatography are examples of the range of methods accepted by the International Society for Extracellular Vesicles for the study and purification of these biomarkers^[77].

In the case of EV circRNA investigation, concentration levels may sometimes be the limitation factor that restricts further downstream processes. Therefore, in this case, EV isolation methods should be focused on achieving a higher EV-derived circRNA yield rather than acquiring extra pure EV samples, which are mainly attained by compromising RNA concentration^[78].

De novo discovery of circRNA

Full-length RNA sequencing emerged as the first method proving beneficial for *de novo* circRNA identification^[9]. By processing total RNA, unmatched reads are selected and assembled by remapping to custom databases containing all human intragenic exon-exon junctions. This protocol first introduced by Salzman *et al.*^[79] has since been improved with new procedures including ribosomal RNA depletion and non-polyadenylated RNA exonuclease-mediated enrichment (RNase R)^[79]. Further validation of novel identified targets requires use of specific bioinformatic tools that allow junction site identification from deep-sequencing data. The rise of newly developed bioinformatic methods have boosted the discovery and analysis of thousands of circRNA [Table 5]. However, sensitivity may be a limitation when using next-generation sequencing for circRNA discovery since library preparation is frequently associated with the loss of low-expressed molecules^[80]. Other methodologies such as microarrays or the nCounter platform have emerged to overcome this issue; however, circRNA discovery in these cases gets restricted to the candidates included either in the array or the gene panel.

Microarrays are useful tools for high-throughput analysis and expression studies of circRNAs where probes are designed to bind specifically to the junction site, getting immobilized, incubated, and further sequenced^[81]. Samples may normally be subject to RNase R to reduce background noise and enhance

Table 4. Current methods for circRNA study

Method	Application	Total RNA input	Advantages	Disadvantages	Ref.
RNAseq	circRNA discovery	Normally $\geq 1\mu\text{g}$ is needed; however, 1 ng has been used in liquid biopsies showing good results	<ul style="list-style-type: none"> - Allows whole transcriptome sequence analysis, including rare and low abundant circRNAs 	<ul style="list-style-type: none"> - Time consuming - It involves high quality RNA - Requires expertise for library preparation, sequencing, and Bioinformatics, for data normalization and analysis 	Cheng <i>et al.</i> ^[125]
Microarrays	circRNA discovery	2 μg	<ul style="list-style-type: none"> - Highly sensitive and specific for circRNA profiling - Easy technology, commercial arrays ready to use 	<ul style="list-style-type: none"> - Although it may be possible to work with less RNA, recommended input remains rather high - circRNA discovery gets restricted to the amount of circRNA included in the panel - Requires Bioinformatics expertise for data normalization and analysis 	Valladares-Ayerbes <i>et al.</i> ^[126]
nCounter	circRNA discovery and quantification	85 ng	<ul style="list-style-type: none"> - Allows multiplexed analysis of up to 800 circRNA targets - Does not require amplification (if enough RNA input) - Works well with low quality RNA samples - Very little hands-on time, with results ready within 24 h - User-friendly data analysis software reducing the need for Bioinformatics support 	<ul style="list-style-type: none"> - circRNA discovery gets restricted to the amount of circRNA included in the panel - Technology is costly, and constrained by one company 	Zhang <i>et al.</i> ^[127] , Dahl <i>et al.</i> ^[85] , 2018
qRT-PCR	circRNA quantification	250 ng (3 replicas, 1 gene)	<ul style="list-style-type: none"> - Well-established technology - Cost-effective 	<ul style="list-style-type: none"> - Does not allow analysis of a large number of genes - Susceptible to template switching and rolling circle amplification bias 	
SplintQuant	circRNA quantification	2 nM	<ul style="list-style-type: none"> - Sensitive and specific approach - Highly reproducibility rates - Eludes the template switching and rolling circle amplification bias found with qRT-PCR 	<ul style="list-style-type: none"> - Novel protocol - No tested in liquid biopsies 	Conn ^[92] , 2019
RT-PCR + end-point PCR + Sanger Sequencing	circRNA identification and validation	100 ng	<ul style="list-style-type: none"> - Well-established technology - Cost-effective - Specific - Gold standard for circRNA validation 	<ul style="list-style-type: none"> - It may require time to test divergent primers - Optimization is required for each pair of primers - Does not allow multiplexing 	Panda <i>et al.</i> ^[89] , 2018
Northern Blot	circRNA identification and validation	1-50 μg	<ul style="list-style-type: none"> - Specific circRNA detection - Allows isoform studies - Solves those problems attained to qRT-PCR such as template switching or rolling amplification biases 	<ul style="list-style-type: none"> - Low sensitivity - It requires a big amount of input which makes it incompatible with most liquid biopsy downstream processes 	Şneider <i>et al.</i> ^[128]

Table 5. Characteristics of online accessible circRNA resources

Name	Resource	Features	Website	Ref.
circBase	Database	One of the main resources with updated information discovered circRNAs. Provides a useful blat tool for circRNA alignment against the human genome	http://www.circbase.org	Garcia-Contreras <i>et al.</i> ^[84] , 2014
circBank	Database	Along with circBase, is one of the most important resources available including a database with most discovered circRNAs along with usegul information	http://www.circbank.cn	Liu <i>et al.</i> ^[93] , 2019
circInteratome	Database	Complete database with different features that allow binding site prediction and knock-down experiment designing	https://circinteractome.nia.nih.gov	Dudekula <i>et al.</i> ^[103] , 2016
CIRCpedia	Database	Database for the identification of tissue specific circRNAs	http://www.picb.ac.cn/rnomics/circpedia	Dong <i>et al.</i> ^[104] , 2018
circRNADb	Database	Searching tool for the identification of EcircRNAs.	http://reprod.njmu.edu.cn/circrnadb	Chen <i>et al.</i> ^[105] , 2016
circRNABase	Database	Allows circRNA network prediction	http://www.hzrna.com/circrn-shujuku/circrnabase	circRNABase ^[106] , 2016
circR2Disease	Database	Serves for the identification of circRNA-miRNA interactions associated to different diseases	http://bioinfo.snnu.edu.cn/CircR2Disease/	Fan <i>et al.</i> ^[107] , 2018
starBase	Database	Serves for the identification of circRNA-miRNA interactions	http://starbase.sysu.edu.cn/	Li <i>et al.</i> ^[108] , 2014
circAtlas	Database	Databased with annotation of circRNAs and with tools that allow identification of circRNA-miRNA interactions	http://circatlas.bols.ac.cn/	Wu <i>et al.</i> ^[109] , 2020
circFunBase	Database	A database for functional circRNAs	http://bis.zju.edu.cn/CircFunBase	Meng <i>et al.</i> ^[110] , 2019
circad	Database	Serves for the identification of circRNA-miRNA interactions associated to different diseases	http://clingen.igib.res.in/circad/	Rophina <i>et al.</i> ^[111] , 2020
circView	Visualization tool	Identification circRNA associated miRNAs and RBPs	http://gb.whu.edu.cn/CircView/	Feng <i>et al.</i> ^[95] , 2018
CSCD	Bioinformatic tool	Identification circRNA associated miRNAs and RBPs, with a focus on circRNA with transcription potential	http://gb.whu.edu.cn/CSCD/	Xia <i>et al.</i> ^[112] , 2018
cirRNAPL	Bioinformatic tool	Identification of circRNA based on extreme learning machine	http://server.malab.cn/CirRNAPL/index.html	Niu <i>et al.</i> ^[113] , 2020
nSolver	Program-Bioinformatic tool	Analysis of RNA expression data generated by the nCounter platform	www.nanostring.com	-
circ2Traits	Pipeline	Serves for the identification of circRNA-	http://gyanxetbeta.com/circdb/	Ghosal <i>et al.</i> ^[114] , 2013

		miRNA interactions associated to different diseases		
circMeta	Pipeline	Genomic feature annotation and differential expression analysis of circular RNAs	https://github.com/lichenlab/circMeta	Chen et al. ^[115] , 2020
circRNAwrap	Pipeline	Pipeline designed for circRNA identification, transcript prediction, and abundance estimation	https://github.com/liaoscience/circRNAwrap	Li et al. ^[116] , 2019
SpliceV	Pipeline	Analysis and publication quality printing of linear and circular RNA splicing, expression and regulation	https://github.com/flemingtonlab/SpliceV	Ungerleider et al. ^[117] , 2019
CIRCexplorer3	Pipeline	Pipeline for the direct comparison of circular and linear RNA expression	https://github.com/YangLab/CLEAR	Ma et al. ^[118] , 2019
circDeep	Pipeline	Permits circular RNA classification from other long non-coding RNA	https://github.com/UofLBioinformatics/circDeep	Chaabane et al. ^[119] , 2020
Segemehl	Pipeline	Pipeline for the identification of fusion reads	http://www.bioinf.uni-leipzig.de/Software/segemehl/segemehl_0_2_0.tar.gz	Hoffmann et al. ^[120] , 2014
MapSplice	Pipeline	Application for small segment mapping	http://www.netlab.uky.edu/p/bioinfo/MapSpliceDownload	-
DCC	Pipeline	Identification of circRNA from fusion reads	https://github.com/dieterichlab/DCC	Cheng et al. ^[121] , 2016
UROBORUS	Pipeline	Allows identification of EcircRNAs	https://github.com/WGLab/uroborus/	Song et al. ^[122] , 2016
NCLscan	Pipeline	Identification of non-coding transcripts	https://github.com/TreesLab/NCLscan	Chuang et al. ^[123] , 2016
Trcirc	High-throughput Data analysis tool	Allows the prediction of circRNA-transcription factor regulatory networks	http://www.licpathway.net/TRCirc/	Tang et al. ^[124] , 2018

detection. This systematically expression profiling process is quite sensitive and straight forward. Current methodology developed by Arraystar includes all necessary tools in order to get detailed annotation specific to circRNA biology, such as miRNA binding sites or conservation status, to reveal all possible functional roles as miRNA sponges.

The nCounter platform allows multiplex analysis of up to 800 circRNA transcripts by direct capturing and counting of individual targets^[82]. This qualitative and quantitative process is rather simple and requires minimal hands on, providing results in less than 48 h. Although nCounter is routinely used for RNA expression assessment in both FFPE and fresh tissues, only few studies have investigated its potential when it comes to liquid biopsies. EV-DNA^[83] and EV-miRNA^[84] profiles have been examined with this platform obtaining different success rates; however, investigation with circRNA remains restricted to tumor and cultured cells^[85], and in no case this platform has been explored for lung cancer research so far.

CircRNA identification and validation

For circRNA validation, end-point PCR has been established as the most extended practice using divergent primers spanning the junction site and followed by further Sanger sequencing^[63].

RNase R treatment is still a debate whether it is beneficial or not to use it in liquid biopsy samples. RNase R has been widely used for the study of circRNAs since it has the property of affecting mostly linear RNA, henceforth, enriching our samples with circRNAs^[86]. However, some circRNAs have demonstrated to be sensitive to the effect of this exonuclease^[85]. The often-long incubation periods can compromise the quality of our RNA samples. In addition, RNase treatment has been proved to not be 100% effective towards mRNA depletion which could lead to a circRNA overestimation if quantification by qPCR is the next downstream process and convergent primers are used. Xiao *et al.*^[87] proved that standard RNase R protocols result in up to 20% of highly expressed mRNAs being unaffected. Therefore, the correct design of divergent primers is instrumental for the study of circRNAs, regardless of whether RNase R treatment is applied to the samples or not. Authors also described that RNase R protocol could be enhanced by replacing K⁺ by Li⁺ in the reaction buffer so enzyme can digest complex structured linear transcripts; however, this is a convoluted process that, even though scientifically relevant, may not result practical in the laboratory routine.

Northern blot analysis has arisen as another common methodology for the study of circRNAs. Following standard protocols, once the RNA is transferred from the gel onto a blotting membrane, circRNAs are then hybridized with short probes normally designed spanning the junction site, hence, allowing circRNA identification. This method also allows studies on size, isoforms, sequence, and abundance of these circular transcripts^[88]. However, the usual high amounts of RNA required for this method is rather high, so investigations get restricted mostly to RNA from either tissue or cell lines.

Quantification of circRNA

Nowadays, different methodologies are being used for the quantification of circRNAs both in solid and liquid biopsies. qRT-PCR has been broadly established as one of the easiest and predilected mechanisms of quantification^[89]; however, different aspects may need to be taken into consideration.

Contrary to tissue, circRNAs are enriched in plasma exosomes^[61]. In this case, RNase R treatment may not be recommended due to the low overall RNA concentration that is expected in these vesicles, however, sometimes its use is necessary to validate primer specificity or due to the nature of specific experiments. In this respect, it is important to stress the need of designing divergent primers as previously cited, along with a probe spanning the junction site. Furthermore, throughout this procedure, the expression of classical reference genes, such as *beta-actin* or *GADPH*, will result altered; hence, ruling out the possibility of performing circRNA expression evaluation by using classical normalization procedures. In this case, the selection of circular RNA housekeeping genes^[90] is crucial for the correct assessment of circRNA expression.

CircRNA amplification via reverse transcription PCR (RT-PCR) often leads to extended concatemeric transcript amplification from a single priming of the reverse transcriptase. This process, triggered by the circular architecture of these molecules, is known as rolling circle amplification, and was first described by You *et al.*^[91] while studying circRNA expression in brain tissues. This event is not problematic if *de novo* circRNA discovery is intentional and direct comparison with canonical transcripts is not envisioned (in fact, it can be beneficial for the study of circRNA splice variants). However, this does not apply to transcript abundant studies, in which this mechanism can introduce biases leading to an overestimation of circRNA expression.

Conn *et al.*^[92] demonstrate this in a study with synthetic circRNAs, resulting in a five-fold increase of circRNAs compared to the expected expression upon RT-PCR and further qPCR amplification. This is a factor to take into consideration in the experimental design^[92].

The same group has developed a cutting-edge tool to avoid the bias introduced by normal qRT-PCR quantification throughout their newly designed SplintQuant method^[92]. This technology is based on the inclusion of custom DNA oligonucleotides that complement target circRNAs, and making use of the PBCV-1 DNA ligase, synthesize cDNA skipping reverse transcription. The system is sensitive, specific and reproducible, allowing the identification and quantification of canonical and non-canonical RNA transcripts including gene fusions and alternative splice variants.

nCounter technology stands out as a very effective and sensitive option for circRNA quantification. Its application for the analysis and quantification of circRNAs has been systematically studied by Dahl *et al.*^[85] in different solid biosources (including formalin fixed paraffin-embedded specimens) for the study of B-cell malignancies, becoming the first group to use this technology for the study of circRNA expression.

Bioinformatic and computational tools for the study of circRNA

Identification of circRNAs can be a straight-forward process when using microarray or nCounter data where the exploratory approach gets restricted to a specific panel of genes. However, detection of circRNA can be a much more complex in the case of deep-sequencing data analysis due to the complexity on the computational workflows. For this purpose, different pipelines and computational analysis tools have been created to facilitate this process [Table 5]. Different publicly available databases such as circBank^[93], circBase^[94], or circView^[95] have proved useful to simplify the study of circRNA throwing light on specific features such as miRNA binding sites, m6A modifications, mutations, or unveiling protein-coding potential [Table 5]. These databases also allow browsing and download of FASTA files based on specific searching criteria.

DISCUSSION

The recent impact of circRNAs in lung cancer research has become undeniable. Since ciRS-7 was introduced as the first circRNA ever described to play a role in hepatocellular carcinoma^[36], many others have followed, extending to different types of cancer, henceforth, consolidating their position as active players in cancer development and progression of malignancy. Recently, publications exploring the biomarker potential of these molecules in NSCLC have remarkably increased, with an exponential growth in the last five years. Nevertheless, despite the patent progress in this field, current research is predominantly restricted to expression analysis of circRNA in tumor samples, with very little information regarding validation in liquid specimens.

EVs, including exosomes, are released by most cells in the body and can be easily isolated from plasma^[96]. Tumor EVs can mediate intercellular communication between tumor cells and tumor microenvironment^[97]; therefore, the study of these molecules via their molecular identification can offer a valuable spatiotemporal snapshot of the state of the disease. However, while several publications have widely demonstrated that EV cargo is enriched in circRNAs^[61], not many investigators have focused on this line of research, delaying the development of novel liquid biopsy-based tools for NSCLC detection. While the potential value of liquid biopsies in the clinic has been recognized as beneficial^[98], in the research context, liquid bio-sources can be rather challenging, including plasma circRNA investigation.

With a superior relative expression and stability in EVs than the canonical mRNA, the extent of circRNA in EVs still remains very low, frequently limiting further downstream analysis. This is unlikely to be an issue in solid tumors; while circRNA overall expression is frequently low (1%-10%)^[14], RNA concentration is rarely a limitation. Furthermore, very often the study of circRNA expression relies on enzymatic amplification - qPCR. This course fueled by the circular architecture of these molecules can sometimes lead to the not-so-

well-known rolling cycle amplification events, resulting in an inaccurate yet overestimated circRNA quantification^[92], frequently leading to untruthful and irreproducible results.

On addition to the above exposed, there is not a general consensus about other fundamental matters such as EV isolation method (if we target the study of the EV circRNA cargo), potential use of RNase R, or readout assessment, among others. As a result, standardization of protocols for the study of circRNA has become instrumental for the study and implementation of these novel biomarkers into the liquid biopsy setting.

Some technologies have arisen as incipient alternatives such as the nCounter platform or the newly developed SplintQuant. Both of them rely on very low RNA input and can overcome the deviation issues that enzymatic qPCR may create.

Additionally, platelets, especially tumor educated platelets, hold a great unexplored potential as a source of circRNAs, not only due to their higher concentration in RNA when compared to EVs, but also due to the high enrichment they present towards these circular biomolecules. To elucidate whether platelet derived circRNA signatures could be of better, equal, or complementary value of the ones from EVs, additional investigation will be required.

Nowadays, most studies aim to exploit the biomarker potential of lung cancer circRNAs, frequently leaving aside any additional examination of their inherent biology. Further research elucidating the different molecular functions of these molecules is greatly needed in order to achieve a future circRNA-based liquid biopsy test.

The rediscovered role of circRNAs as lung cancer biomarkers has the potential to reshape the landscape of liquid biopsies. They count on most features needed to be considered a good biomarker: they can be measured in blood^[99], including plasma^[68], serum^[100], and urine^[101]; they are reasonably robust and very stable due to their circular architecture^[34]; and do not require special handling protocols other than those required for the rest of RNA types. Due to the diverse implications in cancer progression and development of resistance^[34], circRNAs could provide additional information improving diagnosis and treatment guidance by either generating new signatures, or complimenting existing ones.

Circulating tumor DNA is the most commonly explored liquid biopsy for NSCLC, counting with few tests already clinically implemented for the detection of classical mutations such as *EGFR Del19* and *p. L858R* mutation^[102]. However, many lung cancer cases are not linked to a specific driver mutation; therefore, research on new biomarkers, including circRNAs, and further development of multi-omic signatures of tumor microenvironment could provide additional diagnostic opportunities for these patients.

However, as mentioned above, several circRNA quantification methods have limitations, and a clear protocol needs first to be established in order to develop any clinically applicable assay. In addition, clinical utility should be demonstrated by providing convincing evidence of the new biomarker performance (in comparison to currently accepted cfDNA/mRNA liquid biopsy tests), and so far, no circRNA biomarker has achieved that status, probably due to the difficulty of recruiting large patient cohorts required to prove biomarker utility.

Further studies in biomarker discovery, molecular biology, and protocol standardization are warranted in the upcoming years to achieve the implementation of these novel biomarkers in the clinical setting.

DECLARATIONS

Acknowledgments

We would like to thank Stephanie Davis for language editing assistance, and Dr. Cristina Aguado-Esteban for her deep insights. Figure 1 created with BioRender.com.

Authors' contributions

Authors contributed equally to the article.

Availability of data and materials

Not applicable.

Financial support and sponsorship

This project has received funding from a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No 765492.

Conflicts of interest

Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2021.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424. [DOI](#) [PubMed](#)
2. Howlader N, Forjaz G, Mooradian MJ, et al. The effect of advances in lung-cancer treatment on population mortality. *N Engl J Med* 2020;383:640-9. [DOI](#) [PubMed](#)
3. Bracht JWP, Mayo-de-Las-Casas C, Berenguer J, Karachaliou N, Rosell R. The present and future of liquid biopsies in non-small cell lung cancer: combining four biosources for diagnosis, prognosis, prediction, and disease monitoring. *Curr Oncol Rep* 2018;20:70. [DOI](#) [PubMed](#)
4. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* 2017;14:531-48. [DOI](#) [PubMed](#)
5. He X, Park S, Chen Y, Lee H. Extracellular vesicle-associated miRNAs as a biomarker for lung cancer in liquid biopsy. *Front Mol Biosci* 2021;8:630718. [DOI](#) [PubMed](#) [PMC](#)
6. Müller Bark J, Kulasinghe A, Amenábar JM, Punyadeera C. Exosomes in cancer. Elsevier; 2021. p. 1-40. [DOI](#)
7. Pinzani P, D'Argenio V, Del Re M, et al. Updates on liquid biopsy: current trends and future perspectives for clinical application in solid tumors. *Clin Chem Lab Med* 2021. [DOI](#) [PubMed](#)
8. Tang X, Ren H, Guo M, Qian J, Yang Y, Gu C. Review on circular RNAs and new insights into their roles in cancer. *Comput Struct Biotechnol J* 2021;19:910-28. [DOI](#) [PubMed](#) [PMC](#)
9. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012;7:e30733. [DOI](#) [PubMed](#) [PMC](#)
10. Hua X, Sun Y, Chen J, et al. Circular RNAs in drug resistant tumors. *Biomed Pharmacother* 2019;118:109233. [DOI](#) [PubMed](#)
11. Nigro JM, Cho KR, Fearon ER, et al. Scrambled exons. *Cell* 1991;64:607-13. [DOI](#) [PubMed](#)
12. Lee ECS, Elhassan SAM, Lim GPL, et al. The roles of circular RNAs in human development and diseases. *Biomed Pharmacother* 2019;111:198-208. [DOI](#) [PubMed](#)
13. Ji P, Wu W, Chen S, et al. Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep* 2019;26:3444-3460.e5. [DOI](#) [PubMed](#)
14. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 2014;15:409. [DOI](#) [PubMed](#) [PMC](#)
15. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet*

- 2013;9:e1003777. DOI PubMed PMC
16. Dragomir M, Calin GA. Circular RNAs in cancer - lessons learned from microRNAs. *Front Oncol* 2018;8:179. DOI PubMed PMC
 17. Xiao MS, Ai Y, Wilusz JE. Biogenesis and functions of circular RNAs come into focus. *Trends Cell Biol* 2020;30:226-40. DOI PubMed PMC
 18. Chen CY, Sarnow P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* 1995;268:415-7. DOI PubMed
 19. Rybak-Wolf A, Stottmeister C, Glažar P, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015;58:870-85. DOI PubMed
 20. Li X, Liu CX, Xue W, et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* 2017;67:214-227.e7. DOI PubMed
 21. Ashwal-Fluss R, Meyer M, Pamudurti NR, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014;56:55-66. DOI PubMed
 22. Fei T, Chen Y, Xiao T, et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc Natl Acad Sci U S A* 2017;114:E5207-15. DOI PubMed PMC
 23. Errichelli L, Dini Modigliani S, Laneve P, et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat Commun* 2017;8:14741. DOI PubMed PMC
 24. Conn SJ, Pillman KA, Toubia J, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015;160:1125-34. DOI PubMed
 25. Aktaş T, Avşar Ilik İ, Maticzka D, et al. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* 2017;544:115-9. DOI PubMed
 26. Khan MA, Reckman YJ, Aufiero S, et al. RBM20 regulates circular RNA production from the Titin gene. *Circ Res* 2016;119:996-1003. DOI PubMed
 27. Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;19:141-57. DOI PubMed PMC
 28. Huang C, Liang D, Tatomer DC, Wilusz JE. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes Dev* 2018;32:639-44. DOI PubMed PMC
 29. Hansen TB, Wiklund ED, Bramsen JB, et al. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J* 2011;30:4414-22. DOI PubMed PMC
 30. Park OH, Ha H, Lee Y, et al. Endoribonucleolytic Cleavage of m⁶A-Containing RNAs by RNase P/MRP Complex. *Mol Cell* 2019;74:494-507.e8. DOI PubMed
 31. Liu CX, Li X, Nan F, et al. Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* 2019;177:865-880.e21. DOI PubMed
 32. Fischer JW, Busa VF, Shao Y, Leung AKL. Structure-mediated RNA decay by UPF1 and G3BP1. *Mol Cell* 2020;78:70-84.e6. DOI PubMed PMC
 33. Yu T, Wang Y, Fan Y, et al. CircRNAs in cancer metabolism: a review. *J Hematol Oncol* 2019;12:90. DOI PubMed PMC
 34. Zhou R, Wu Y, Wang W, et al. Circular RNAs (circRNAs) in cancer. *Cancer Lett* 2018;425:134-42. DOI PubMed
 35. Xu L, Zhang M, Zheng X, Yi P, Lan C, Xu M. The circular RNA ciRS-7 (Cdr1as) acts as a risk factor of hepatic microvascular invasion in hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2017;143:17-27. DOI PubMed
 36. Yan B, Zhang W, Mao XW, Jiang LY. Circular RNA ciRS-7 correlates with advance disease and poor prognosis, and its down-regulation inhibits cells proliferation while induces cells apoptosis in non-small cell lung cancer. *Eur Rev Med Pharmacol Sci* 2018;22:8712-21. DOI PubMed
 37. Wang L, Zheng C, Wu X, et al. Circ-SOX4 promotes non-small cell lung cancer progression by activating the Wnt/β-catenin pathway. *Mol Oncol* 2020;14:3253. DOI PubMed PMC
 38. Chen X, Mao R, Su W, et al. Circular RNA circHIPK3 modulates autophagy via MIR124-3p-STAT3-PRKAA/AMPKα signaling in STK11 mutant lung cancer. *Autophagy* 2020;16:659-71. DOI PubMed PMC
 39. Xie Y, Yuan X, Zhou W, et al. The circular RNA HIPK3 (circHIPK3) and its regulation in cancer progression: review. *Life Sci* 2020;254:117252. DOI PubMed
 40. Wang Y, Li H, Lu H, Qin Y. Circular RNA SMARCA5 inhibits the proliferation, migration, and invasion of non-small cell lung cancer by miR-19b-3p/HOXA9 axis. *Oncotargets Ther* 2019;12:7055-65. DOI PubMed PMC
 41. Qiu BQ, Zhang PF, Xiong D, et al. CircRNA fibroblast growth factor receptor 3 promotes tumor progression in non-small cell lung cancer by regulating Galectin-1-AKT/ERK1/2 signaling. *J Cell Physiol* 2019;234:11256-64. DOI PubMed
 42. Liu G, Shi H, Deng L, et al. Circular RNA circ-FOXM1 facilitates cell progression as ceRNA to target PDPF and MACC1 by sponging miR-1304-5p in non-small cell lung cancer. *Biochem Biophys Res Commun* 2019;513:207-12. DOI PubMed
 43. Varella-Garcia M. Chromosomal and genomic changes in lung cancer. *Cell Adh Migr* 2010;4:100-6. DOI PubMed PMC
 44. Wu K, Liao X, Gong Y, et al. Circular RNA F-circSR derived from SLC34A2-ROS1 fusion gene promotes cell migration in non-small cell lung cancer. *Mol Cancer* 2019;18:98. DOI PubMed PMC
 45. Huang XY, Zhang PF, Wei CY, et al. Circular RNA circMET drives immunosuppression and anti-PD1 therapy resistance in hepatocellular carcinoma via the miR-30-5p/snail/DPP4 axis. *Mol Cancer* 2020;19:92. DOI PubMed PMC
 46. Pei X, Chen SW, Long X, et al. circMET promotes NSCLC cell proliferation, metastasis, and immune evasion by regulating the miR-145-5p/CXCL3 axis. *Aging (Albany NY)* 2020;12:13038-58. DOI PubMed PMC
 47. Li YH, Xu CL, He CJ, Pu HH, Liu JL, Wang Y. circMTDH.4/miR-630/AEG-1 axis participates in the regulation of proliferation,





- migration, invasion, chemoresistance, and radioresistance of NSCLC. *Mol Carcinog* 2020;59:141-53. DOI PubMed
48. Dong Y, Xu T, Zhong S, et al. Circ_0076305 regulates cisplatin resistance of non-small cell lung cancer via positively modulating STAT3 by sponging miR-296-5p. *Life Sci* 2019;239:116984. DOI PubMed
 49. Xu Y, Jiang T, Wu C, Zhang Y. CircAKT3 inhibits glycolysis balance in lung cancer cells by regulating miR-516b-5p/STAT3 to inhibit cisplatin sensitivity. *Biotechnol Lett* 2020;42:1123-35. DOI PubMed
 50. Xu N, Chen S, Liu Y, et al. Profiles and bioinformatics analysis of differentially expressed circRNAs in taxol-resistant non-small cell lung cancer cells. *Cell Physiol Biochem* 2018;48:2046-60. DOI PubMed
 51. Yu W, Peng W, Sha H, Li J. Hsa_circ_0003998 promotes chemoresistance via modulation of miR-326 in lung adenocarcinoma cells. *Oncol Res* 2019;27:623-8. DOI PubMed PMC
 52. Huang MS, Liu JY, Xia XB, et al. Hsa_circ_0001946 inhibits lung cancer progression and mediates cisplatin sensitivity in non-small cell lung cancer via the nucleotide excision repair signaling pathway. *Front Oncol* 2019;9:508. DOI PubMed PMC
 53. Zheng F, Xu R. CircPVT1 contributes to chemotherapy resistance of lung adenocarcinoma through miR-145-5p/ABCC1 axis. *Biomed Pharmacother* 2020;124:109828. DOI PubMed
 54. Ding C, Yi X, Wu X, et al. Exosome-mediated transfer of circRNA CircNFIX enhances temozolomide resistance in glioma. *Cancer Lett* 2020;479:1-12. DOI PubMed
 55. Huang W, Yang Y, Wu J, et al. Circular RNA cESRP1 sensitises small cell lung cancer cells to chemotherapy by sponging miR-93-5p to inhibit TGF- β signalling. *Cell Death Differ* 2020;27:1709-27. DOI PubMed PMC
 56. Zhou Y, Zheng X, Xu B, et al. Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem Biophys Res Commun* 2019;508:527-35. DOI PubMed
 57. Fu Y, Huang L, Tang H, Huang R. hsa_circRNA_012515 is highly expressed in NSCLC patients and affects its prognosis. *Cancer Manag Res* 2020;12:1877-86. DOI PubMed PMC
 58. Chen T, Luo J, Gu Y, Huang J, Luo Q, Yang Y. Comprehensive analysis of circular RNA profiling in AZD9291-resistant non-small cell lung cancer cell lines. *Thorac Cancer* 2019;10:930-41. DOI PubMed PMC
 59. Zhang PF, Pei X, Li KS, et al. Circular RNA circFGFR1 promotes progression and anti-PD-1 resistance by sponging miR-381-3p in non-small cell lung cancer cells. *Mol Cancer* 2019;18:179. DOI PubMed PMC
 60. Yang H, Zhang H, Yang Y, et al. Hypoxia induced exosomal circRNA promotes metastasis of Colorectal Cancer via targeting GEF-H1/RhoA axis. *Theranostics* 2020;10:8211-26. DOI PubMed PMC
 61. Li Y, Zheng Q, Bao C, et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res* 2015;25:981-4. DOI PubMed PMC
 62. Lasda E, Parker R. Circular RNAs Co-precipitate with extracellular vesicles: a possible mechanism for circRNA clearance. *PLoS One* 2016;11:e0148407. DOI PubMed PMC
 63. Wang M, Yu F, Li P, Wang K. Emerging function and clinical significance of exosomal circRNAs in cancer. *Mol Ther Nucleic Acids* 2020;21:367-83. DOI PubMed PMC
 64. Shang BQ, Li ML, Quan HY, et al. Functional roles of circular RNAs during epithelial-to-mesenchymal transition. *Mol Cancer* 2019;18:138. DOI PubMed PMC
 65. Chen F, Huang C, Wu Q, Jiang L, Chen S, Chen L. Circular RNAs expression profiles in plasma exosomes from early-stage lung adenocarcinoma and the potential biomarkers. *J Cell Biochem* 2020;121:2525-33. DOI PubMed
 66. He F, Zhong X, Lin Z, et al. Plasma exo-hsa_circRNA_0056616: a potential biomarker for lymph node metastasis in lung adenocarcinoma. *J Cancer* 2020;11:4037-46. DOI PubMed PMC
 67. Xian J, Su W, Liu L, et al. Identification of three circular RNA cargoes in serum exosomes as diagnostic biomarkers of non-small-cell lung cancer in the chinese population. *J Mol Diagn* 2020;22:1096-108. DOI PubMed
 68. Hang D, Zhou J, Qin N, et al. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. *Cancer Med* 2018;7:2783-91. DOI PubMed PMC
 69. Liu XX, Yang YE, Liu X, et al. A two-circular RNA signature as a noninvasive diagnostic biomarker for lung adenocarcinoma. *J Transl Med* 2019;17:50. DOI PubMed PMC
 70. Tan S, Gou Q, Pu W, et al. Circular RNA F-circEA produced from EML4-ALK fusion gene as a novel liquid biopsy biomarker for non-small cell lung cancer. *Cell Res* 2018;28:693-5. DOI PubMed PMC
 71. Alhasan AA, Izuogu OG, Al-Balool HH, et al. Circular RNA enrichment in platelets is a signature of transcriptome degradation. *Blood* 2016;127:e1-e11. DOI PubMed PMC
 72. Preußer C, Hung LH, Schneider T, et al. Selective release of circRNAs in platelet-derived extracellular vesicles. *J Extracell Vesicles* 2018;7:1424473. DOI PubMed PMC
 73. Best MG, Sol N, In 't Veld SGJG, et al. Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* 2017;32:238-252.e9. DOI PubMed PMC
 74. Liu YT, Han XH, Xing PY, et al. Circular RNA profiling identified as a biomarker for predicting the efficacy of Gefitinib therapy for non-small cell lung cancer. *J Thorac Dis* 2019;11:1779-87. DOI PubMed PMC
 75. Konoshenko MY, Lekhnov EA, Vlassov AV, Laktionov PP. Isolation of extracellular vesicles: general methodologies and latest trends. *Biomed Res Int* 2018;2018:8545347. DOI PubMed PMC
 76. Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW, Simpson RJ. Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat Rev Clin Oncol* 2018;15:617-38. DOI PubMed
 77. Théry C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*

- 2018;7:1535750. DOI PubMed PMC
78. Tang YT, Huang YY, Zheng L, et al. Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. *Int J Mol Med* 2017;40:834-44. DOI PubMed PMC
79. Guria A, Velayudha Vimala Kumar K, Srikakulam N, et al. Circular RNA profiling by illumina sequencing via template-dependent multiple displacement amplification. *Biomed Res Int* 2019;2019:2756516. DOI PubMed PMC
80. Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 2008;29:4618-26. DOI PubMed
81. Qu S, Song W, Yang X, et al. Microarray expression profile of circular RNAs in human pancreatic ductal adenocarcinoma. *Genom Data* 2015;5:385-7. DOI PubMed PMC
82. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Curr Protoc Mol Biol* 2011;Chapter 25:Unit25B.10. DOI PubMed
83. Kamyabi N, Abbasgholizadeh R, Maitra A, Ardekani A, Biswal SL, Grande-Allen KJ. Isolation and mutational assessment of pancreatic cancer extracellular vesicles using a microfluidic platform. *Biomed Microdevices* 2020;22:23. DOI PubMed
84. Garcia-Contreras M, Shah SH, Tamayo A, et al. Plasma-derived exosome characterization reveals a distinct microRNA signature in long duration Type 1 diabetes. *Sci Rep* 2017;7:5998. DOI PubMed PMC
85. Dahl M, Daugaard I, Andersen MS, et al. Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab Invest* 2018;98:1657-69. DOI PubMed PMC
86. Vincent HA, Deutscher MP. Insights into how RNase R degrades structured RNA: analysis of the nuclease domain. *J Mol Biol* 2009;387:570-83. DOI PubMed PMC
87. Xiao MS, Wilusz JE. An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. *Nucleic Acids Res* 2019;47:8755-69. DOI PubMed PMC
88. Koch L. RNA: Translated circular RNAs. *Nat Rev Genet* 2017;18:272-3. DOI PubMed
89. Panda AC, Gorospe M. Detection and analysis of circular RNAs by RT-PCR. *Bio Protoc* 2018;8:e2775. DOI PubMed PMC
90. Zhong S, Zhou S, Yang S, et al. Identification of internal control genes for circular RNAs. *Biotechnol Lett* 2019;41:1111-9. DOI PubMed
91. You X, Vlatkovic I, Babic A, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* 2015;18:603-10. DOI PubMed PMC
92. Conn V, Conn SJ. SplintQuant: a method for accurately quantifying circular RNA transcript abundance without reverse transcription bias. *RNA* 2019;25:1202-10. DOI PubMed PMC
93. Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol* 2019;16:899-905. DOI PubMed PMC
94. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;20:1666-70. DOI PubMed PMC
95. Feng J, Xiang Y, Xia S, et al. CircView: a visualization and exploration tool for circular RNAs. *Brief Bioinform* 2018;19:1310-6. DOI PubMed
96. Liangsupree T, Multia E, Riekkola ML. Modern isolation and separation techniques for extracellular vesicles. *J Chromatogr A* 2021;1636:461773. DOI PubMed
97. Han L, Xu J, Xu Q, Zhang B, Lam EW, Sun Y. Extracellular vesicles in the tumor microenvironment: therapeutic resistance, clinical biomarkers, and targeting strategies. *Med Res Rev* 2017;37:1318-49. DOI PubMed
98. Saarenheimo J, Eigeliene N, Andersen H, Tiirola M, Jekunen A. The value of liquid biopsies for guiding therapy decisions in non-small cell lung cancer. *Front Oncol* 2019;9:129. DOI PubMed PMC
99. Wen G, Zhou T, Gu W. The potential of using blood circular RNA as liquid biopsy biomarker for human diseases. *Protein Cell* 2020. DOI PubMed
100. Fan CM, Wang JP, Tang YY, et al. circMAN1A2 could serve as a novel serum biomarker for malignant tumors. *Cancer Sci* 2019;110:2180-8. DOI PubMed PMC
101. Lam WK, Dennis Lo YM. Circular RNAs as urinary biomarkers. *Clin Chem* 2019;65:1196-8. DOI PubMed
102. Akhoundova D, Mosquera Martinez J, Musmann LE, et al. The role of the liquid biopsy in decision-making for patients with non-small cell lung cancer. *J Clin Med* 2020;9:3674. DOI PubMed PMC
103. Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* 2016;13:34-42. DOI PubMed PMC
104. Dong R, Ma XK, Li GW, Yang L. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinformatics* 2018;16:226-33. DOI PubMed PMC
105. Chen X, Han P, Zhou T, et al. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* 2016;6:34985. DOI PubMed PMC
106. circRNABase. Available from: <http://www.hzrna.com/circrn-shujuku/circrnabase> [Last accessed on 30 Jun 2021].
107. Fan C, Lei X, Fang Z, Jiang Q, Wu FX. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database (Oxford)* 2018:2018. DOI PubMed PMC
108. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;42:D92-7. DOI PubMed PMC
109. Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 2020;21:101. DOI PubMed PMC
110. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. *Database (Oxford)* 2019:2019.

[DOI](#) [PubMed](#) [PMC](#)

111. Rophina M, Sharma D, Poojary M, Scaria V. Circad: a comprehensive manually curated resource of circular RNA associated with diseases. *Database (Oxford)* 2020;2020:baaa019. [DOI](#) [PubMed](#) [PMC](#)
112. Xia S, Feng J, Chen K, et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* 2018;46:D925-9. [DOI](#) [PubMed](#) [PMC](#)
113. Niu M, Zhang J, Li Y, et al. CirRNAPL: a web server for the identification of circRNA based on extreme learning machine. *Comput Struct Biotechnol J* 2020;18:834-42. [DOI](#) [PubMed](#) [PMC](#)
114. Ghosal S, Das S, Sen R, Basak P, Chakrabarti J. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet* 2013;4:283. [DOI](#) [PubMed](#) [PMC](#)
115. Chen L, Wang F, Bruggeman EC, Li C, Yao B. circMeta: a unified computational framework for genomic feature annotation and differential expression analysis of circular RNAs. *Bioinformatics* 2020;36:539-45. [DOI](#) [PubMed](#) [PMC](#)
116. Li L, Bu D, Zhao Y. CircRNAwrap - a flexible pipeline for circRNA identification, transcript prediction, and abundance estimation. *FEBS Lett* 2019;593:1179-89. [DOI](#) [PubMed](#)
117. Ungerleider N, Flemington E. SpliceV: analysis and publication quality printing of linear and circular RNA splicing, expression and regulation. *BMC Bioinformatics* 2019;20:231. [DOI](#) [PubMed](#) [PMC](#)
118. Ma XK, Wang MR, Liu CX, et al. CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics Proteomics Bioinformatics* 2019;17:511-21. [DOI](#) [PubMed](#) [PMC](#)
119. Chaabane M, Williams RM, Stephens AT, Park JW. circDeep: deep learning approach for circular RNA classification from other long non-coding RNA. *Bioinformatics* 2020;36:73-80. [DOI](#) [PubMed](#) [PMC](#)
120. Hoffmann S, Otto C, Doose G, et al. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol* 2014;15:R34. [DOI](#) [PubMed](#) [PMC](#)
121. Cheng J, Metge F, Dieterich C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* 2016;32:1094-6. [DOI](#) [PubMed](#)
122. Song X, Zhang N, Han P, Lai RK, Wang K, Lu W. Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res* 2016;44:e87. [DOI](#) [PubMed](#) [PMC](#)
123. Chuang TJ, Wu CS, Chen CY, Hung LY, Chiang TW, Yang MY. NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. *Nucleic Acids Res* 2016;44:e29. [DOI](#) [PubMed](#) [PMC](#)
124. Tang Z, Li X, Zhao J, et al. TRCirc: a resource for transcriptional regulation information of circRNAs. *Brief Bioinform* 2019;20:2327-33. [DOI](#) [PubMed](#)
125. Cheng L, Sun X, Scicluna BJ, Coleman BM, Hill AF. Characterization and deep sequencing analysis of exosomal and non-exosomal miRNA in human urine. *Kidney Int* 2014;86:433-44. [DOI](#) [PubMed](#)
126. Valladares-Ayerbes M, Garrigos C, Taron M, et al. Circular RNAs as biomarkers in liquid biopsy in colorectal cancer. *Clinical Oncol* 2020;38:suppl. [DOI](#)
127. Zhang J, Zhang X, Li C, et al. Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells. *RNA Biol* 2019;16:220-32. [DOI](#) [PubMed](#) [PMC](#)
128. Sun L, Vella P, Schnell R, et al. Structural and Functional Characterization of the BcsG Subunit of the Cellulose Synthase in *Salmonella typhimurium*. *JMB* 2018;18:3170-89. [DOI](#) [PubMed](#) [PMC](#)

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

Carlos Pedraz-Valdunciel^{1,2} , Stavros Giannoukakos³, Nicolas Potie⁴, Ana Giménez-Capitán⁵, Chung-Ying Huang⁶, Michael Hackenberg³, Alberto Fernandez-Hilario⁴, Jill Bracht^{2,5}, Martyna Filipka^{1,2} , Erika Aldeguer⁵, Sonia Rodríguez⁵, Trevor G. Bivona⁷, Sarah Warren⁶, Cristina Aguado⁵, Masaaki Ito⁸, Andrés Aguilar-Hernández⁹, Miguel Angel Molina-Vila⁵  and Rafael Rosell^{1,9,10} 

- 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

Keywords

biomarkers; cancer; circRNA; diagnosis; nCounter; NSCLC

Correspondence

C. Pedraz-Valdunciel, Pangaea Oncology, Calle de Sabino Arana, 5, 08028 Barcelona, Spain

Tel: +34 935 46 01 19

E-mail: cpedraz@panoncology.com

and

R. Rosell, Germans Trias I Pujol Research Institute, Camí de les Escoles, s/n, 08916 Badalona, Spain

Tel: +34 930330520

E-mail: rrosell@iconcologia.net

(Received 16 August 2021, revised 22 November 2021, accepted 19 January 2022, available online 10 February 2022)

doi:10.1002/1878-0261.13182

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.

Abbreviations

AUC, area under the curve; circRNA, circular RNA; FFPE, formalin-fixed paraffin-embedded; GBM, gradient boosting machines; KNN, k-nearest neighbors; LOOCV, leave-one-out cross-validation; miRNA, micro RNA; ML, machine learning; NPV, negative predictive value; NSCLC, non-small cell lung cancer; PCR, polymerase chain reaction; PPV, positive predictive value; RF, random Forest; RFE, recursive feature elimination; RNAseq, RNA sequencing; ROC, receiver operating characteristic; RT-qPCR, quantitative reverse transcription PCR; SD, standard deviation.

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	1	–
amplification		
KRAS	12	–
G12A	2	–
G12C	3	–
G12V	4	–
G12R	1	–
Other	2	–
BRAF	1	–
ROS	1	–
RET	2	–
ALK	1	–
MET (exon14	1	–
mutation)		
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)	–

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.

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	<i>KRAS</i>	G12S	ATCC
HOP-62			G12C	ATCC
PC9		<i>EGFR</i>	E746_A750 DL	Hoffmann-La Roche, with the authorization of Dr. Mayumi Ono
HCC-827			E746_A750 DL	ATCC
NCI-H1666		<i>BRAF</i>	G466V	ATCC
NCI-H2228		<i>ALK</i>	<i>EML4-ALK</i> , variant 1	ATCC
NCI-H3122			<i>EML4-ALK</i> , 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 FFPE 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 μ g of total RNA was either treated or mock-treated with RNase-R (Lucigen, Madison, WI, USA). RNA

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 QuantStudio™ 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 *C_q* 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://mirccancer.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	GAAGACCAAACTGTCCAGTGTAAAG
	Reverse	CACTTCAGACACAGAGCCTACTTCA
	Probe	TGACCTGGAGCATAAG
circSOX13 (hsa_circRNA_0004777)	Forward	CAGTGACTGGAAGGAGAGGTTTC
	Reverse	CTGGGCAGAGATGGGGCT
	Probe	AAAGATGTCAAAGGATGTCCATGA
circBNC2 (hsa_circ_0086414)	Forward	GTCTGCACAGTGGCTGGTTG
	Reverse	GGTGATGATTTCTTCTTCGAG
	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. Shrunk \log_2 fold-change (\log_2FC) 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_2FC 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

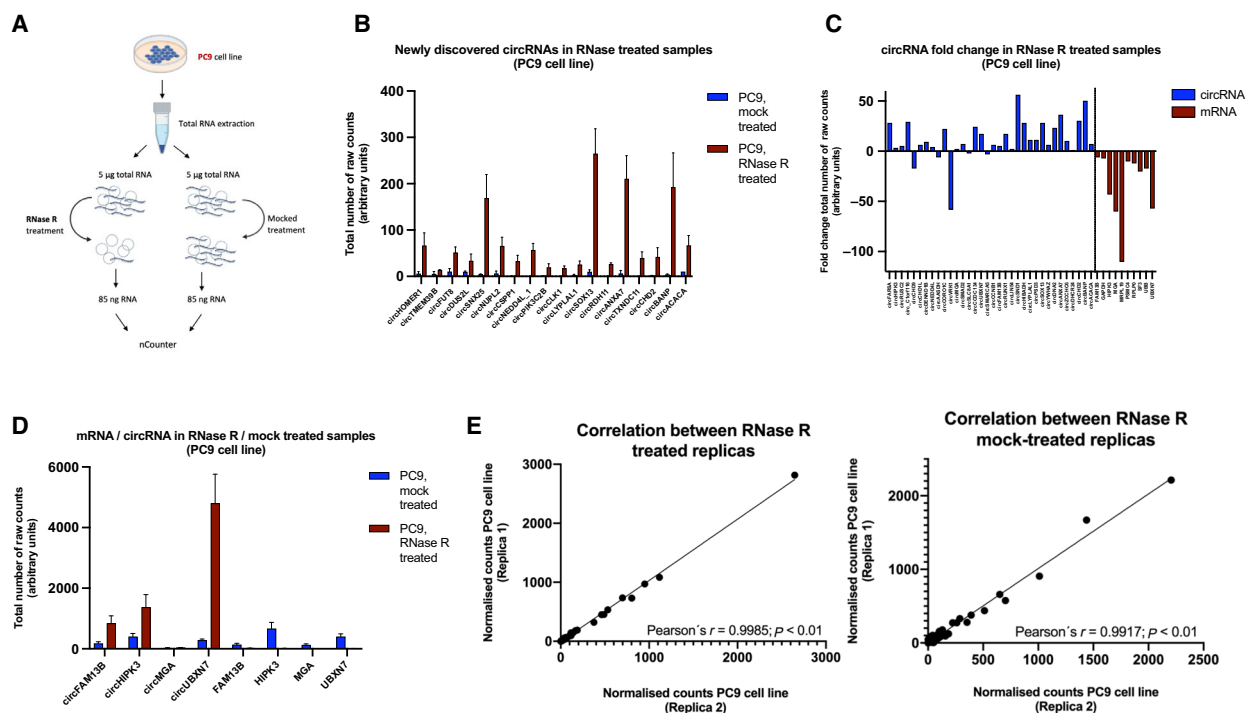


Fig. 1. Analysis of circRNA from RNase-R-treated samples. (A) Workflow for circRNA enrichment with RNase-R. (B) Representation of the newly discovered circRNAs after RNase-R treatment. Bars indicate the mean of the replicas ($n = 2$). Error bars indicate SD. (C) CircRNA/linear HK fold-change after RNase-R 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 sub-optimal 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

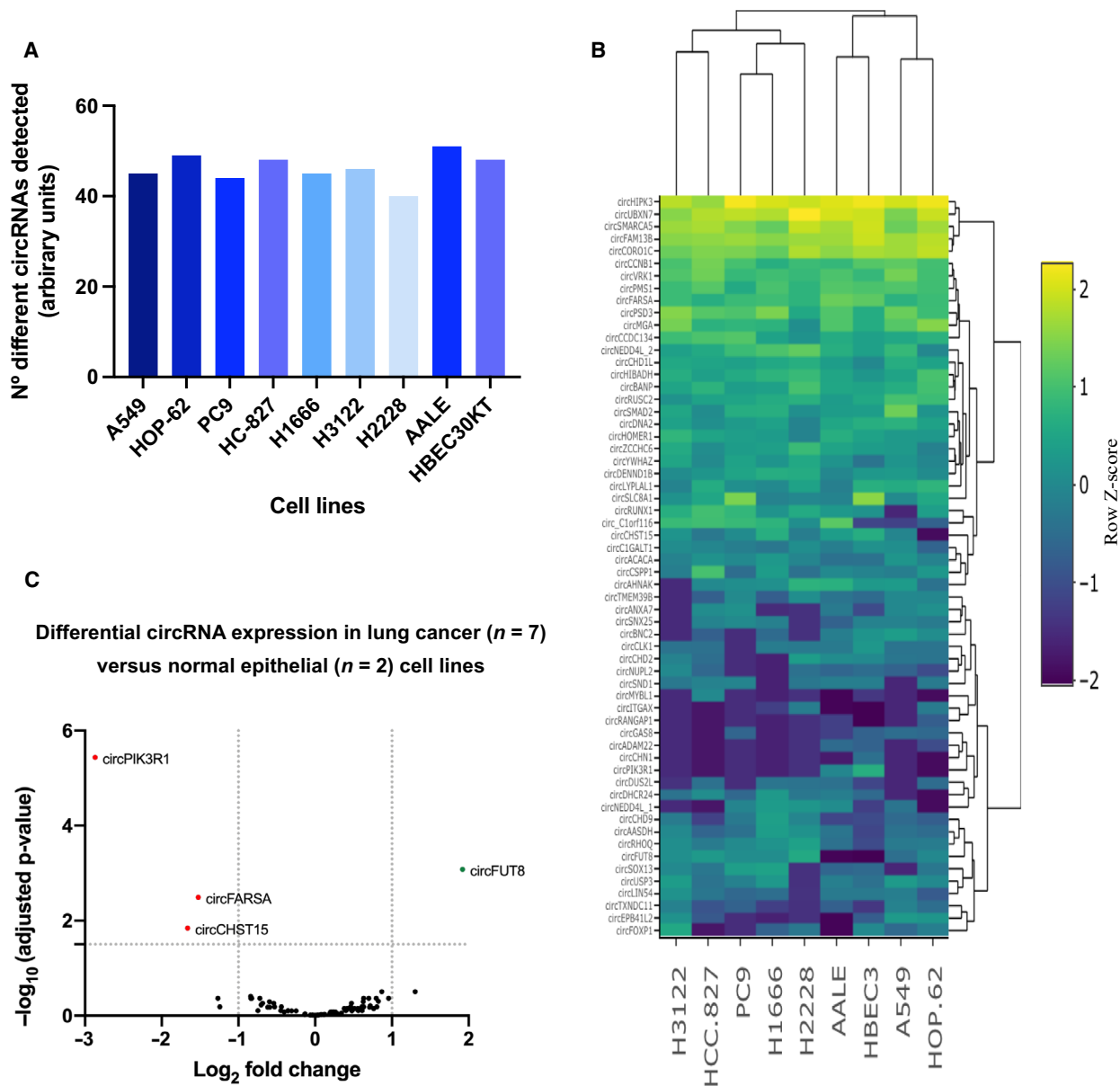


Fig. 2. CircRNA analysis in lung cancer (A549, HOP-62, PC9, HCC-827, H1666, H3122, and H2228) and epithelial cells (AALE and HBEC30KT). (A) Bar plot representing total circRNAs detected in each of the cell lines. (B) Hierarchical clustering of cell lines based on circRNA expression. (C) Differential circRNA expression analysis of log₂-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 circC10RF116 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

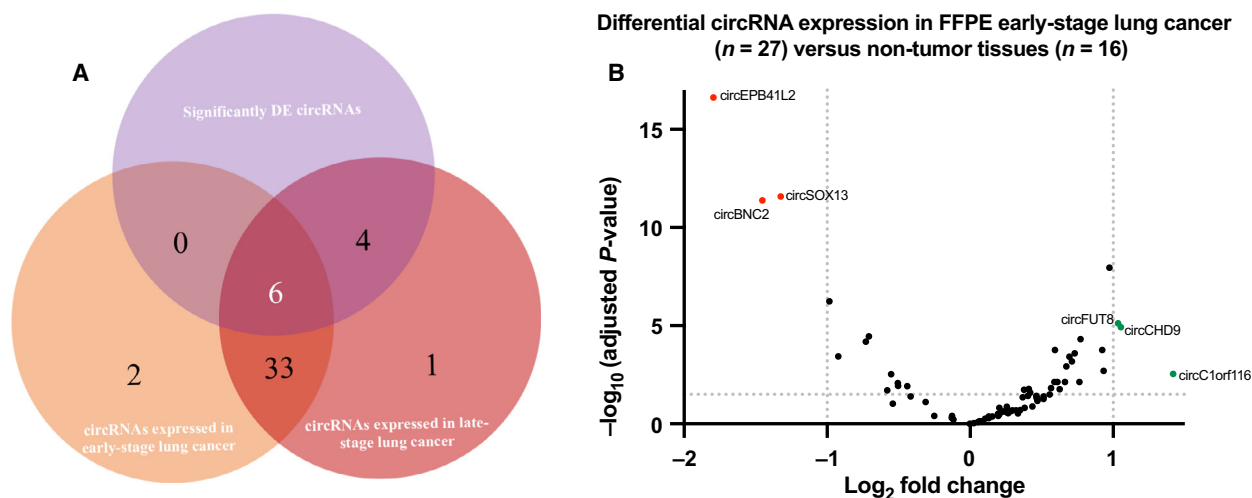


Fig. 4. CircRNA expression in early-stage NSCLC samples. (A) Venn diagram displaying circRNAs identified in early- and late-stage samples, featuring those shared by both cohorts. DE circRNAs are indicated. (B) Differential expression analysis of log₂-normalized counts between the early-stage lung cancer 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 complementing 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

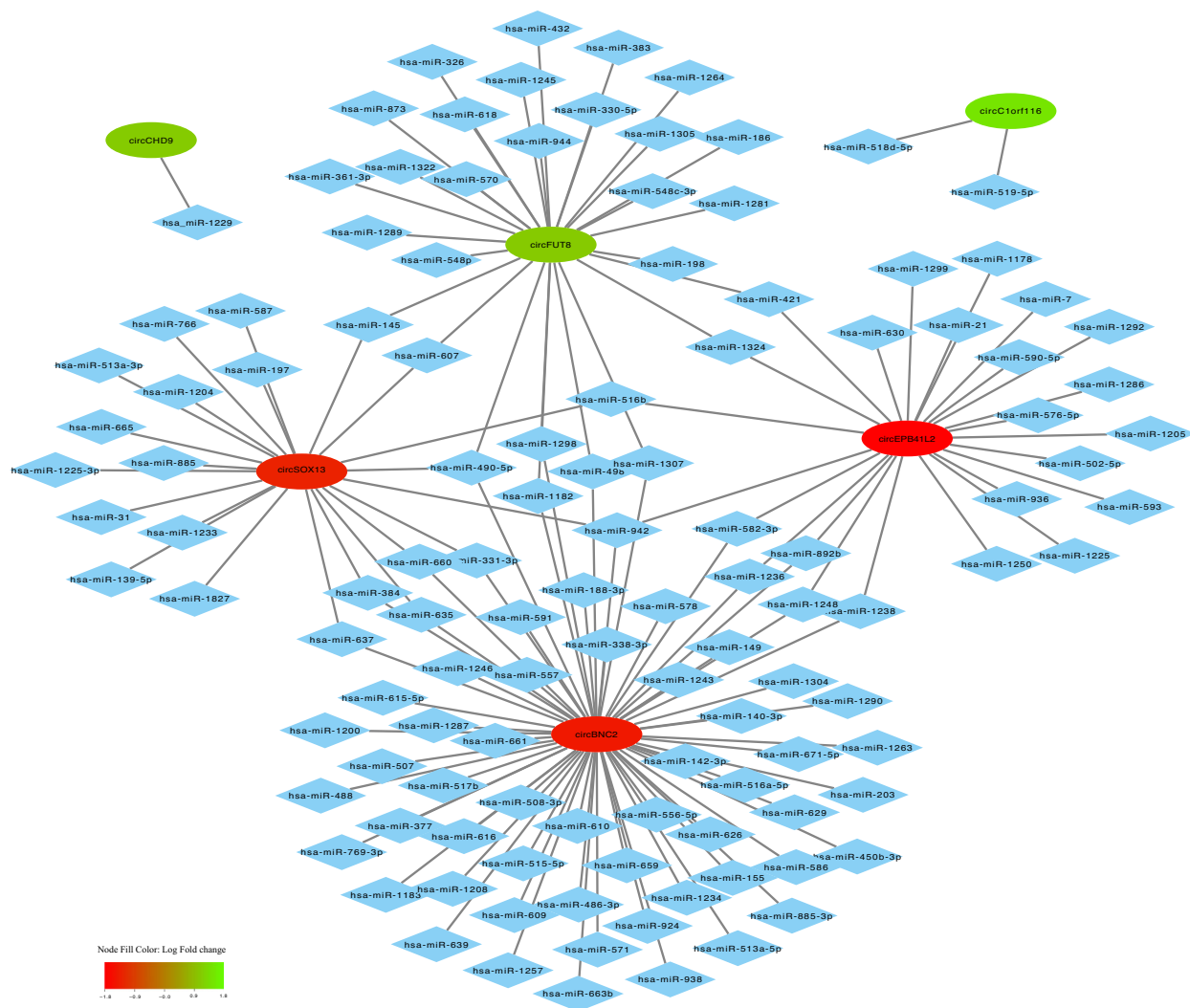


Fig. 5. Mapping network showing predicted sequence-pairing circRNA-miRNA interaction of differentially expressed circRNA found in early-stage lung cancer tissues. CircRNAs are represented by elliptic nodes and colored based on their log fold-change. Complementary binding miRNAs are represented by diamond 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

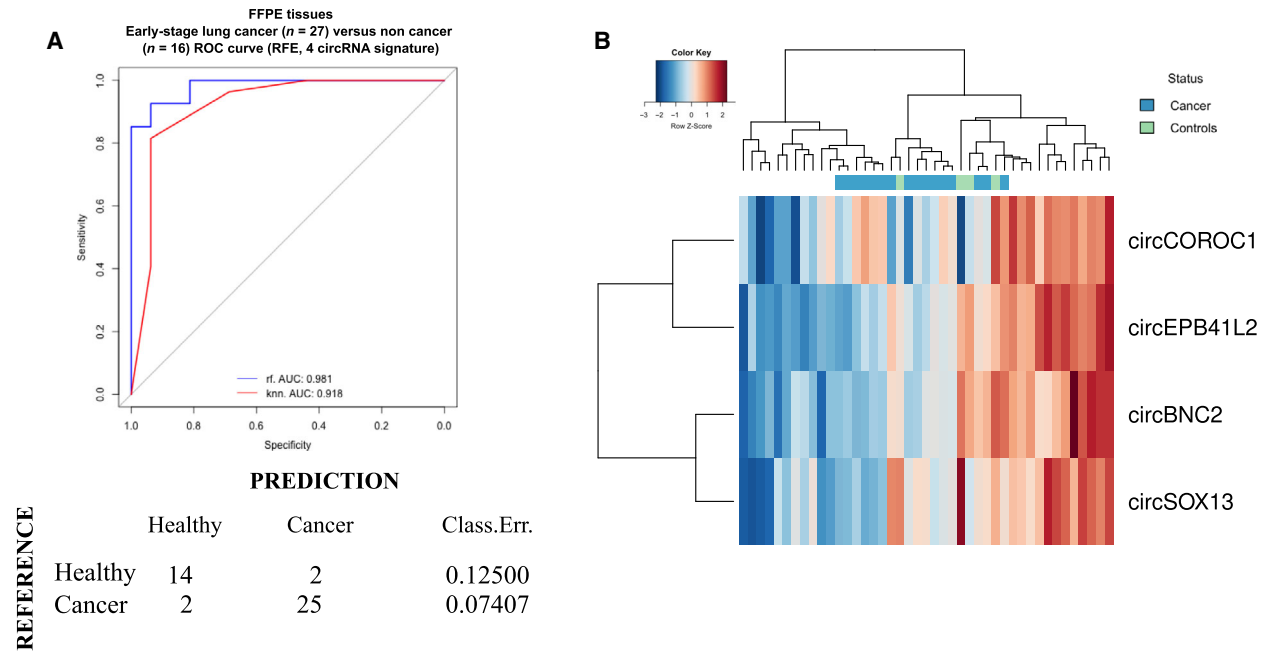


Fig. 6. (A) Area under the ROC curve of the 4 circRNA-signature using recursive feature elimination (RFE) for cohort classification. Confusion matrix was generated based on the RF classification scores. Classification error scores are indicated. (B) Hierarchical clustering of samples based on the 4-circRNA signature.

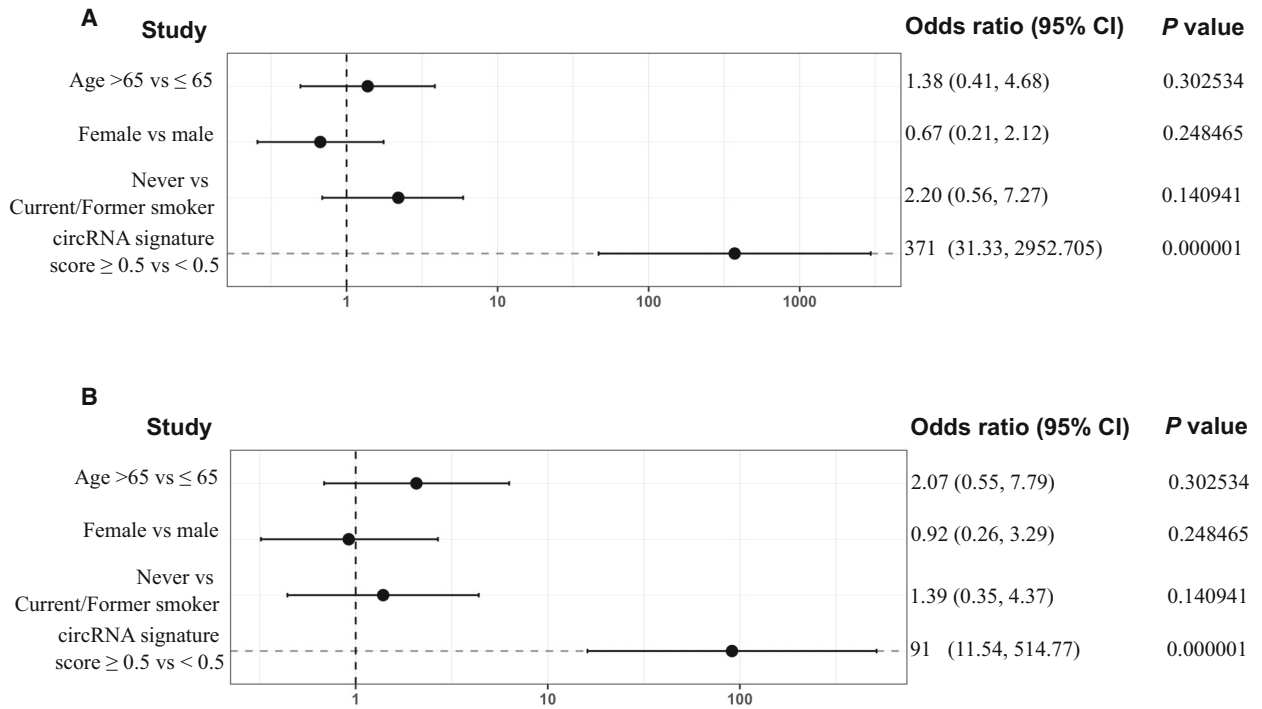


Fig. 7. Univariate analysis exploring associations between patient characteristics and lung cancer to determine risk factor. Forest plot represents the odds ratios in (A) lung cancer; and (B), early-stage lung cancer cohorts with a 95% Wald confidence limit. Student's t-test was used for the calculation of *P*-values.

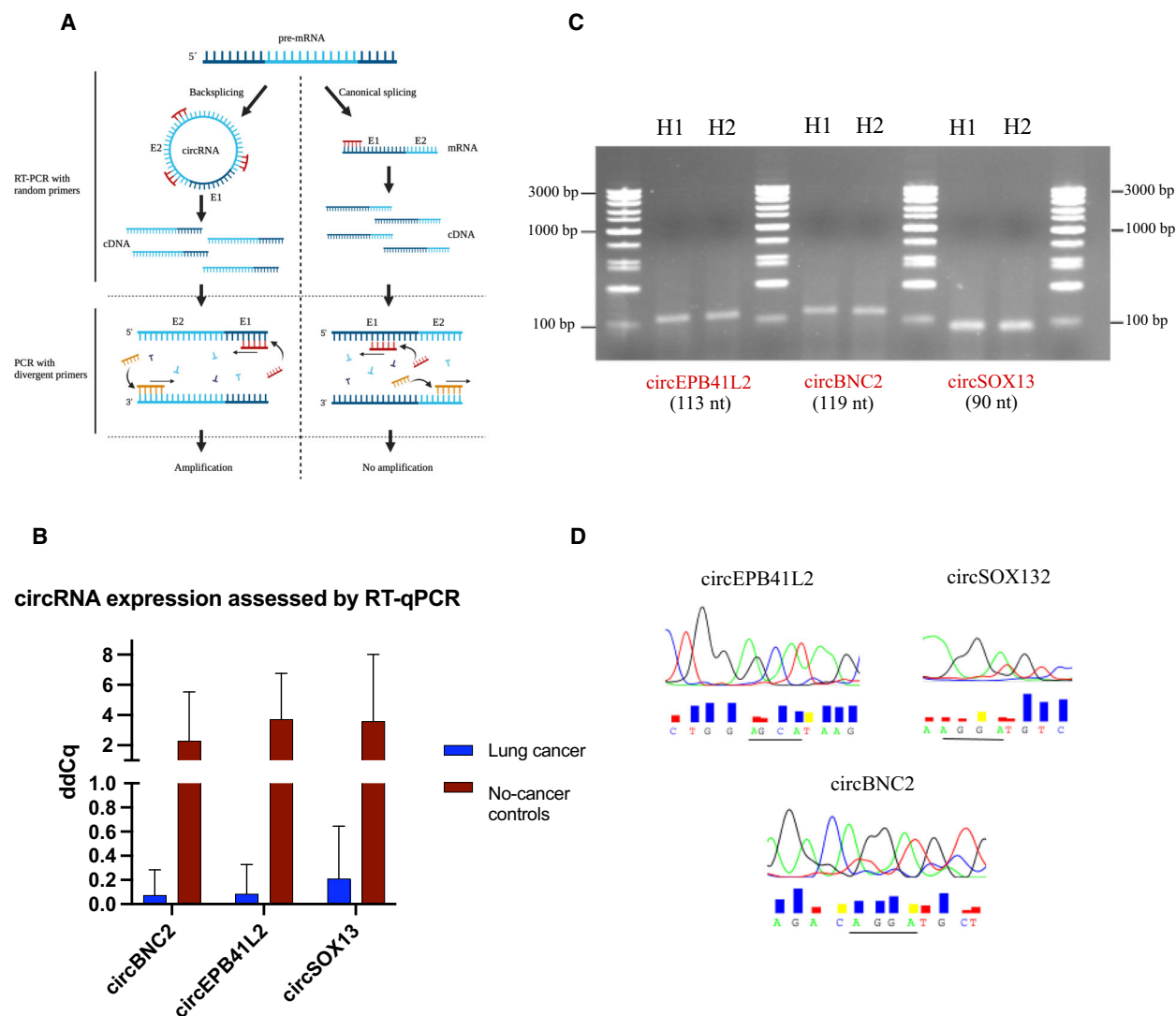


Fig. 8. Validation of nCounter results by RT-qPCR and further Sanger sequencing. (A) Representation of circRNA amplification using divergent primers. (B) Bar plot of RT-qPCR results depicting downregulation of circEPB41L2, circSOX13, and circBNC2 in lung cancer versus control tissues validating previous 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 paraformaldehyde 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 RNase-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

We would like to thank Stephanie Davis for her language editing assistance. The investigators also wish to thank the patients for kindly agreeing to donate samples to this study. We thank all the physicians who collaborated by providing clinical information. Graphical Abstract, Figs 1A, 8A and Fig. S1 were created with Biorender.com. This project has received funding from a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No 765492.

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

- 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>
- Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;**32**:453–61. <https://doi.org/10.1038/nbt.2890>
- 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>
- 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](https://doi.org/10.1016/0092-8674(91)90244-S)
- 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>
- 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>
- 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, Grønbaek 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 adenocarcinoma. *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. nCounter probe design allows specific recognition of the circRNAs included in the panel.

Fig. S2. nCounter workflow for circRNA expression studies in FFPE lung 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.

Article

Multiplex Analysis of CircRNAs from Plasma Extracellular Vesicle-Enriched Samples for the Detection of Early-Stage Non-Small Cell Lung Cancer

Carlos Pedraz-Valdunciel ^{1,2,3,*}, Stavros Giannoukacos ⁴ , Ana Giménez-Capitán ³, Diogo Fortunato ⁵,
Martyna Filipaska ^{1,6} , Jordi Bertran-Alamillo ³, Jillian W. P. Bracht ^{7,8} , Ana Drozdowskyj ⁹, Joselyn Valarezo ³,
Natasa Zarovni ⁵, Alberto Fernández-Hilario ¹⁰ , Michael Hackenberg ⁴ , Andrés Aguilar-Hernández ⁹,
Miguel Ángel Molina-Vila ³  and Rafael Rosell ^{1,9,11,*}

¹ Department of Cancer Biology and Precision Medicine, Germans Trias I Pujol Research Institute (IGTP), Campus Can Ruti, 08916 Badalona, Spain

² Department of Biochemistry, Molecular Biology and Biomedicine, Autonomous University of Barcelona, Campus de Bellaterra, 08193 Barcelona, Spain

³ Laboratory of Oncology, Pangaea Oncology, Dexeus University Hospital, 08028 Barcelona, Spain

⁴ Department of Genetics, Facultad de Ciencias, Campus Fuentenueva s/n, Universidad de Granada, 18071 Granada, Spain

⁵ Exosomics SpA, 53100 Siena, Italy

⁶ B Cell Biology Group, Hospital del Mar Biomedical Research Park (IMIM), Barcelona Biomedical Research Park (PRBB), 08003 Barcelona, Spain

⁷ Vesicle Observation Centre, Laboratory of Experimental Clinical Chemistry, Department of Clinical Chemistry, Amsterdam UMC location University of Amsterdam, 1105AZ Amsterdam, The Netherlands

⁸ Cancer Center Amsterdam, Imaging and Biomarkers, 1105AZ Amsterdam, The Netherlands

⁹ Oncology Institute Dr. Rosell (IOR), Dexeus University Institute, 08028 Barcelona, Spain

¹⁰ Department of Computer Science and Artificial Intelligence, DaSCI, University of Granada, 18071 Granada, Spain

¹¹ Catalan Institute of Oncology, Campus Can Ruti, 08916 Badalona, Spain

* Correspondence: cpedraz@panoncology.com or carlospedraz@icloud.com (C.P.-V.); rosell@iconcologia.net (R.R.)



Citation: Pedraz-Valdunciel, C.; Giannoukacos, S.; Giménez-Capitán, A.; Fortunato, D.; Filipaska, M.; Bertran-Alamillo, J.; Bracht, J.W.P.; Drozdowskyj, A.; Valarezo, J.; Zarovni, N.; et al. Multiplex Analysis of CircRNAs from Plasma Extracellular Vesicle-Enriched Samples for the Detection of Early-Stage Non-Small Cell Lung Cancer. *Pharmaceutics* **2022**, *14*, 2034. <https://doi.org/10.3390/pharmaceutics14102034>

Academic Editors: Francesco Grossi and Hye Suk Lee

Received: 8 August 2022

Accepted: 15 September 2022

Published: 24 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: The analysis of liquid biopsies brings new opportunities in the precision oncology field. Under this context, extracellular vesicle circular RNAs (EV-circRNAs) have gained interest as biomarkers for lung cancer (LC) detection. However, standardized and robust protocols need to be developed to boost their potential in the clinical setting. Although nCounter has been used for the analysis of other liquid biopsy substrates and biomarkers, it has never been employed for EV-circRNA analysis of LC patients. Methods: EVs were isolated from early-stage LC patients ($n = 36$) and controls ($n = 30$). Different volumes of plasma, together with different number of pre-amplification cycles, were tested to reach the best nCounter outcome. Differential expression analysis of circRNAs was performed, along with the testing of different machine learning (ML) methods for the development of a prognostic signature for LC. Results: A combination of 500 μ L of plasma input with 10 cycles of pre-amplification was selected for the rest of the study. Eight circRNAs were found upregulated in LC. Further ML analysis selected a 10-circRNA signature able to discriminate LC from controls with AUC ROC of 0.86. Conclusions: This study validates the use of the nCounter platform for multiplexed EV-circRNA expression studies in LC patient samples, allowing the development of prognostic signatures.

Keywords: circRNAs; extracellular vesicles; nCounter; lung cancer; NSCLC; liquid biopsies

1. Introduction

With 350 deaths per day projected for 2022, lung cancer stands as the main cause of cancer-related mortality, leading the second highest incidence in the United States and

Europe [1,2]. Treatments have proved to be more effective at the early stage of the disease, when lung cancer patients benefit from a significantly improved overall survival (OS) [3]. However, most cases are diagnosed at an advanced stage, with a 5-year survival rate dropping to only 5% in stage IV.

In order to achieve early detection, many challenges need to first be faced. Classical biopsy techniques for sampling and profiling of suspicious pulmonary nodules often involve invasive procedures. Limitations of such practices include restricted access to the nodules, which regularly compromise the quality and quantity of extracted biopsy specimens. Heterogeneity of resected samples also hampers the use of these methods, especially for tumor identification [4].

Liquid biopsies offer a minimally invasive procedure for sampling, providing a practical tool for continuous monitoring of lung cancer patients [5], being also actively investigated for early detection [6]. Despite the slow progression on the development of liquid biopsies in this area, many possible biomarkers have been proposed in the last few years, including circulating tumor DNA (ctDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs), proteins, extracellular vesicles (EVs) and tumor educated platelets (TEPs).

Lung cancer elicits massive changes in RNA metabolism, reflecting both in the tumor transcriptome and in the circulating EV and TEP cargo. EVs contain different RNA molecules, including mRNA and non-coding RNAs such as miRNA or circular RNAs (circRNAs) [7,8]. The circRNA transcripts are generated by post-transcriptional circularization of the 5' and 3' ends in an alternative process called back-splicing. Their circular structure makes most of them resistant to exonucleases and, therefore, robustly stable RNA molecules, compared to the canonical (linear) mRNA. CircRNAs seem to play an important role in human homeostasis [9,10]. Moreover, it has been reported that aberrant expression of certain circRNAs can promote cancer development and progression [11]. Additionally, some circRNAs have been investigated as liquid biopsy biomarkers for the early detection of lung cancer and other solid tumors [12,13]. However, the lack of consensus on a robust and standardized protocol for circRNA quantification is holding back the development of clinically applicable assays.

RT-qPCR, microarrays and RNAseq are the three methods most commonly used in circRNA research. However, the RT-qPCR does not allow high-throughput analysis; microarrays have a limited dynamic range of RNA detection; and RNAseq is associated with high cost, long time-consuming protocols, and high grade of complexity when it comes to data analysis.

An alternative technique for multiplex analysis of circRNA is nCounter, which provides a cost-effective automated solution for analysis of more than 800 targets with minimal hands-on time, providing highly reproducible data in less than 48 h. nCounter is based on the detection of RNA of interest using target-specific probe pairs. Each pair comprises of a reporter probe with a unique color combination at the 5'-end, allowing specific recognition of the gene of interest; and a capture probe carrying a molecule of biotin, which provides a molecular grip to the nCounter cartridge, allowing downstream digital detection [14]. The expression of a particular gene is then calculated by counting the number of times a specific color-coded probe is detected. This technology has been embraced in translational research, including the development and validation of liquid biopsies, due to its capability of working with a low quantity of highly degraded samples [15,16]. Recent studies reported the use of nCounter for the study of several categories of circulating biomarkers [17–21], including EV-derived DNA [22], miRNA [23,24], mRNA [25] and circRNA [26]. However, nCounter analysis of EV-circRNAs has not been investigated for early detection of lung cancer. Here, we report the development of a protocol for EV enrichment from plasma followed by RNA purification and circRNA analysis by nCounter.

Then, we analyzed liquid biopsies from non-cancer donors and early-stage non-small cell lung cancer (NSCLC) patients and applied machine learning (ML) to develop a prognostic signature.

2. Materials and Methods

2.1. Patient Samples

The study was carried out in accordance with the principles of the Declaration of Helsinki, under an approved protocol of the institutional review board of Quirón Hospitals. We obtained and documented written informed consent from all the patients. A total of 36 samples from early-stage NSCLC (stages IA to IIIA) were selected from our institution, along with 30 samples from non-cancer controls (Table 1). Clinical information from patients and controls included age, gender, smoking status, tumor histology and stage, when applicable. All samples were de-identified before further processing for confidentiality purposes.

Table 1. Clinicopathologic characteristics of enrolled patients ($n = 66$).

Clinicopathological Characteristics	NSCLC Patients ($n = 36$)	Non-Cancer Controls ($n = 30$)
Gender—no. (%)		
Male	18 (50.0)	13 (43.3)
Female	18 (50.0)	17 (56.7)
Age—yr.		
Median	71.5	38
Range	32–91	23–57
Histological type		
Adenocarcinoma	27 (75.0)	-
Squamous carcinoma	4 (11.1)	-
Not information	5 (13.9)	-
Smoking status—no. (%)		
Former- or current smoker	20 (55.5)	11 (36.6)
Never smoker	13 (36.2)	17 (56.7)
Not information	3 (8.3)	2 (6.7)
Tumor stage—no. (%)		
I	19 (52.8)	-
II	2 (5.5)	-
IIIA	15 (41.7)	-

2.2. Plasma Processing

Around 10 mL of whole blood was collected from the participants enrolled in the study using sterile EDTA Vacutainer tubes (BD, Plymouth, UK) and processed within the next 2 h. Blood samples were centrifuged twice at $2000 \times g$ at room temperature (RT) in a Rotina 380 R centrifuge (Hettich, Tuttlingen, Germany) for 10 min to separate plasma from red/white blood cells, platelets, and cell debris. Aliquoted plasma samples were then stored at -80°C until downstream processing.

2.3. Enrichment of EVs

EVs were isolated from plasma using differential ultracentrifugation (UC) as described previously [27] or the miRCURY Exosome Serum/Plasma Kit (Qiagen, Hilden, Germany).

In the case of UC, 500 μL plasma samples were transferred into 15 mL sterile high-speed centrifuge tubes (VWR-Avantor, Philadelphia, PA, USA), filled up with sterile $1 \times$ phosphate-buffered saline (PBS) and centrifuged twice at $10,000 \times g$ for 30 min at 4°C in a Sorvall RC 6 Plus centrifuge (Thermo Fisher Scientific, Waltham, MA, USA). Supernatants were then transferred into UC tubes (Beckman Coulter, Brea, CA, USA), equilibrated with sterile $1 \times$ PBS, and spun twice at $70,000 \times g$ for 1 h at 4°C in the Sorvall WX Ultra 100 centrifuge (Thermo Fisher Scientific). The EV enriched pellets were resuspended in 100 μL sterile PBS and stored at -80°C until used. EV enrichment with the miRCURY Kit was performed as described [25]. Debris was cleared from 500 μL plasma samples with thrombin and subsequent centrifugation at $10,000 \times g$ for 5 min at RT. Samples were then incubated with Precipitation Buffer overnight at 4°C and centrifuged twice ($500 \times g$, 5 min

at RT). Supernatants were discarded, EV enriched pellets were resuspended in 270 μL of Resuspension Buffer and stored at $-80\text{ }^{\circ}\text{C}$ until used.

2.4. Transmission Electron Microscopy (TEM)

Visualization of EV samples was performed by the TEM service of the Universitat Autònoma de Barcelona (UAB). A volume of 3.9 μL of EV-enriched sample was blotted onto a Holey Carbon Film Supported Nickel Grid (Merck, Darmstadt, Germany) previously glow-discharged in a PELCO easiGlow glow cleaning system (Ted Pella Inc, Redding, CA, USA). Next, the grid containing the sample was plunged into a Leica EM GP cryo-work station (Leica, Wetzlar, Germany) comprising a liquid ethane bath cooled to $-180\text{ }^{\circ}\text{C}$, and subsequently transferred and visualized in a JEOL 2011 TEM (Jeol Ltd., Tokyo, Japan) operating at 200 kV. Samples were kept at $-180\text{ }^{\circ}\text{C}$ during the observation and captures were obtained with a Gatan Model 895 UltraScan 4000 4k \times 4k CCD camera (Gatan Inc, Pleasanton, CA, USA). Image processing was performed using ImageJ software (version 1.8.0, National Institutes of Health, Bethesda, MD, USA).

2.5. Nano-Flow Cytometry Measurements

The volume of EV samples was brought to 500 μL with sterile PBS. Size-exclusion chromatography (SEC) columns (qEVOriginal/35 nm, Izon Science, Oxford, UK) were equilibrated with 20–30 mL of sterile PBS and eluted using the same buffer. Collection started immediately after loading the sample into the column, according to manufacturer instructions. Eluted EV-enriched samples were directly analyzed with the nanoFCM (NanoFCM Ltd., Nottingham, UK), a nanoparticle flow cytometer. Instrument calibration with standard beads enabled accurate measurements of both size and concentration of 40–200 nm particles through the detection of their side scatter [28].

2.6. RNA Isolation and DNase Treatment

EV-enriched samples were treated with 4 $\mu\text{g}/\text{mL}$ of RNase A (Sigma-Aldrich, Burlington, MA, USA) for 1 h at $37\text{ }^{\circ}\text{C}$, to eliminate any non-vesicular RNA. TRI Reagent (MRC, Cincinnati, OH, USA) was added to a final volume of 1 mL and incubated at RT for 20 min. Then, 200 μL of a Chloroform and Isoamyl Alcohol dilution (24:1) (Panreac Química SLU, Barcelona, Spain) were added, followed by vigorous shaking and centrifugation at $12,000\times g$ for 15 min at $4\text{ }^{\circ}\text{C}$. Upper fraction was collected, and RNA was precipitated by adding 2.5 μL of glycogen (Merck) and 500 μL 2-propanol (Merck), followed by incubation at RT for 10 min and further centrifugation at $12,000\times g$ for 10 min at $4\text{ }^{\circ}\text{C}$. RNA pellet was then washed with 75% ethanol, dried at $95\text{ }^{\circ}\text{C}$ for 3 min and resuspended in 12 μL of nuclease-free water.

The DNA-free DNA Removal Kit (Thermo Fisher Scientific) was used to eliminate any DNA remaining in the samples. Following the manufacturer's protocol, 0.75 μL of DNase buffer and 1 μL enzyme were added to 7.5 μL RNA sample and incubated at $37\text{ }^{\circ}\text{C}$ for 30 min. A volume of 0.75 μL of DNase inactivation reagent was then added to the reaction, incubated for 2 min at RT and centrifuged for 1.5 min at $10,000\times g$ and RT. The supernatant containing EV-RNA was then transferred to a fresh tube and stored at $-80\text{ }^{\circ}\text{C}$ until further use.

2.7. RT-qPCR and Sanger Sequencing Analysis

RT-qPCR and Sanger sequencing of circRNA junction sites were performed as previously described [29]. Divergent primers and probe sets were designed using Primer Express 3.0 Software (version 3.0.1, Applied Biosystems, Waltham, MA, USA) with the probes spanning the circRNA junction site (Table 2). Five microliters of EV-RNA was converted into cDNA using the M-MLV reverse transcriptase enzyme and random hexamers (both from Invitrogen, Waltham, MA, USA). 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 circRNA. Three replicas

of each sample were run for the quantification of the expression of each assessed circular transcript. Quantification of gene expression was performed using the QuantStudio™ 6 Flex System (Applied Biosystems) and the comparative Ct method.

Table 2. Primer and probe design for circRNA validation by RT-qPCR.

CircRNA	Primers and Probes	Sequence
circHIPK3	Forward	5'CGGCCAGTCATGTATCAAAGAC 3'
	Reverse	5'AAAGGCACTTGACTGAGTTTGATAAA 3'
	Probe	FAM 5'AATCTCGGTACTACAGGTATG 3' MGB
circZCCHC6	Forward	5'AGATGTTGTCGAATTTGTGGAAAA 3'
	Reverse	5'TCTTCTACCATTGATAAAAGCCTTCAT 3'
	Probe	FAM 5'GAGGAGAAATGACAAATT 3' MGB

For Sanger sequencing, 10 µL of each PCR product was subjected to electrophoresis in a 2× agarose gel (100 V, 30 min) 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 and 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).

2.8. nCounter Processing

The nCounter Low RNA Input Amplification Kit (NanoString Technologies, Seattle, WA, USA) was used to retrotranscribe and pre-amplify 4 µL of EV-derived RNA in a Verity thermal cycler (Applied Biosystems) following NanoString's guidelines. Briefly, samples were denatured at 95 °C for 10 min and hybridized for 18 h at 67 °C. Our custom-made nCounter panel (including 78 circRNAs, 6 linear reference genes and 4 mRNAs [30]) was used to analyze EV-derived pre-amplified cDNA according to the manufacturer's instructions. RCC files containing data outputted by the NanoString nCounter Flex System (NanoString Technologies) from each run were exported to the nSolver Analysis Software (Version 4.0.70, NanoString Technologies).

2.9. Differential Expression Analysis

Raw count nCounter values were exported to Microsoft Excel (Version 16.40, Microsoft, Redmond, WA, USA) using nSolver Analysis Software. The background was calculated for each sample as (geo)mean ± 2SD of the negative probe counts (NCs) Raw counts lower than the background were automatically excluded from further analysis. The raw circRNA counts were normalized using the total number of counts of the sample and multiplied by 10,000. Differential expression analysis was performed comparing the means of the normalized counts for each circRNA in the early-stage NSCLC vs. non-cancer controls. The circRNAs with a fold change >1 and *p*-value < 0.05 were considered as differentially expressed (DE).

2.10. Data Pre-Processing and Normalization for Signature Development

Raw RCC-formatted data files were exported from the nSolver Analysis Software (NanoString Technologies). R (Version 4.0.3, R Core Team and the R Foundation for Statistical Computing, Vienna, Austria) and R studio (Version 2021.09.0, RStudio PBC, Boston, MA, USA) were used for pre-processing and normalization analysis of the imported files. Initial evaluation of the quality and integrity of the RCC data was performed using the NanoStringQCPro (Version 1.22.0) package. During this process, we looked for potential

outliers based on the performance of standard control metrics provided by NanoString, such as Imaging, Binding Density, Positive Control Linearity, and Limit of Detection. After this first pre-analytical step, samples were subjected to supplementary exploratory examination, including Principal Component Analysis (PCA) and interquartile range (1.5 IQR rule) analysis. Samples found as outliers by both methods were then excluded from downstream analyses.

NCs were employed to exclude lowly expressed circRNAs with excessive background noise. The arithmetic mean of the $NC \pm 2SD$ was subtracted from each endogenous circRNA for each sample. Any transcript scoring a value below 0 in more than 75% of the analyzed samples was then excluded from further analysis. PCA plot was used to re-assess the data after the aforementioned filtering step. Technical variability correction and normalization were performed using the RUVseq/RUVg function (Version 1.24.0) and DESeq2 (Version 1.30.1) packages (RUVseq-DESeq2). First, the RUVg function was used to estimate the unwanted variation among samples based on the DE genes. DESeq2 and edgeR (Version 3.32.1) performed a first pass DE analysis and the intersected least significant genes (with adjusted p -value above 0.1) were used as “in silico empirical” negative controls. DESeq2 was then utilized with default parameters along with the RUV factors to perform the normalization of the raw filtered data. The normalization performance was assessed using the standard relative log expression (RLE) plot.

2.11. Machine Learning (ML) for Signature Development

The Recursive Feature Elimination (RFE) algorithm along with leave-one-out cross-validation (LOOCV) and the random forest (RF) classifier were used to perform feature selection on the normalized data previously generated by RUVseq-DESeq2. The optimal number of features was automatically selected by keeping only those yielding best performance after cross-validation. These final features were to constitute the prognostic signature. To test the predictive power of the selected signature, extra trees classifier (ETC), k-nearest neighbor (KNN) and RF models were built using default parameters. The 5-fold cross validation (5-CV) algorithm was applied for this purpose. During this process, the dataset was randomly split into k-folds ($k = 5$), being 4/5 of the data used to train the model, while the remaining 1/5 was used to test its behavior. The classifier showing the highest area under the ROC curve (AUC ROC) value was selected as the final model. Signature scores for each sample were obtained from the final model. A confidence threshold of 0.5 was considered for the calculation of the positive and negative predictive values (PPV–NPV). Additional statistical indicators such as accuracy, sensitivity, specificity, and Cohen’s κ were also calculated.

2.12. Univariate and Multivariate Analyses

Association between clinical characteristics and ML-generated signature was assessed with a univariate Cox proportional-hazard regression model. Odds ratios, with a Confidence Interval (CI) of 95% was calculated using the MedCal Statistical Software (MedCalc Software Ltd. Odds ratio calculator. https://www.medcalc.org/calc/odds_ratio.php. Accessed last on 5 September 2022). Multivariate analysis using logistic regression was performed using SAS software (v9.4, SAS Institute, Cary, NC, USA). Significance was set at $p < 0.05$ for all statistical tests.

3. Results

3.1. Enrichment of Plasma EVs and Workflow Development for nCounter CircRNA Analysis

Two replicated 500 μ L plasma samples from an early-stage NSCLC patient and a non-cancer control were submitted to EV enrichment by ultracentrifugation (UC) or using the miRCURY Exosome Serum/Plasma kit. Enriched EVs were characterized by transmission electron microscopy (TEM) and nanoparticle flow cytometry via nanoFCM. TEM images revealed different clusters of diverse-sized EVs (30 to 300 nm, all within the reported EV size range [31–33]) in all samples regardless of the enrichment method used (Figure 1a).

Samples extracted using the miRCURY kit showed a higher proportion of vesicles with an exosome-like size range (30–100 nm) by TEM, compared to the more heterogeneous UC samples (Figure 1a). NanoFCM analysis revealed a higher concentration of 40–100 nm particles in samples enriched using the miRCURY kit (Figure 1b). In addition, nanoFCM indicated a higher number of particles/mL in the NSCLC patient sample when compared to the control, both in the UC and miRCURY preparations (Figure 1b).

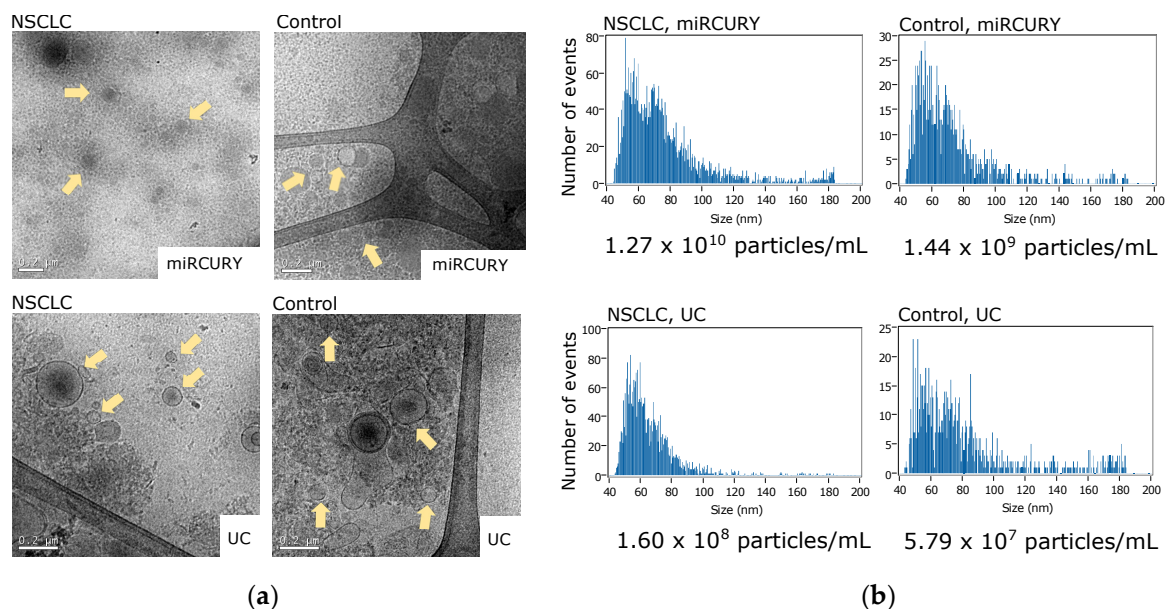


Figure 1. Characterization of extracellular vesicles (EVs) enriched either by differential ultracentrifugation (UC) or precipitation using the miRCURY Exosome Serum/Plasma kit. (a) Observation of EV samples on transmission electron microscopy (TEM). Yellow arrows point out EVs of different sizes. Scale bars indicate 200 nm.; (b) Nanoflow cytometry (nanoFCM) profiles of EV samples showing size and concentration of 40–200 nm particles.

Next, different volumes of plasma (500 μ L, 1000 μ L and 1500 μ L) from a NSCLC patient were tested in duplicates to assess the effect of initial volumes on downstream circRNA analysis by nCounter using the custom panel we previously developed [30]. Since RNA concentration from EV enriched samples has been demonstrated to be insufficient for direct nCounter analysis [25], pre-amplification steps of 14 and 20 cycles were tested. The utmost total number of counts was achieved using an input of 500 μ L both with 14 and 20 cycles (14,151 \pm 1864 and 686,525 \pm 345,655, respectively; Figure 2a). Consequently, 500 μ L of plasma was also the volume allowing the detection of more circRNAs ($n = 27.5 \pm 4.95$ and 33 ± 7.07 for 14 and 20 cycles, respectively; Figure 2b), even if only those with a score above 10 counts after background removal were selected (Table S1, Figure S1).

Different amplification cycles (10, 12 and 14) were subsequently tested in a 500 μ L plasma sample. The highest number of raw counts was obtained with 14 cycles (Figure 3a). Regarding the number of circRNAs, 10 and 12 cycles yielded similar results ($n = 51.5 \pm 9.19$ and 52.5 ± 7.78 respectively). More circRNAs were detected at 14 cycles ($n = 59 \pm 16.97$) with a high variability between replicates (Figure 3b, Table S2). In view of these results, we selected for EV-circRNA analysis a protocol that included 500 μ L of plasma input, EV enrichment with the miRCURY kit, extravesicular RNA elimination with RNase A, EV lysis and RNA extraction with TRI reagent, retrotranscription and nCounter analysis with a 10-cycle preamplification step (Figure 4).

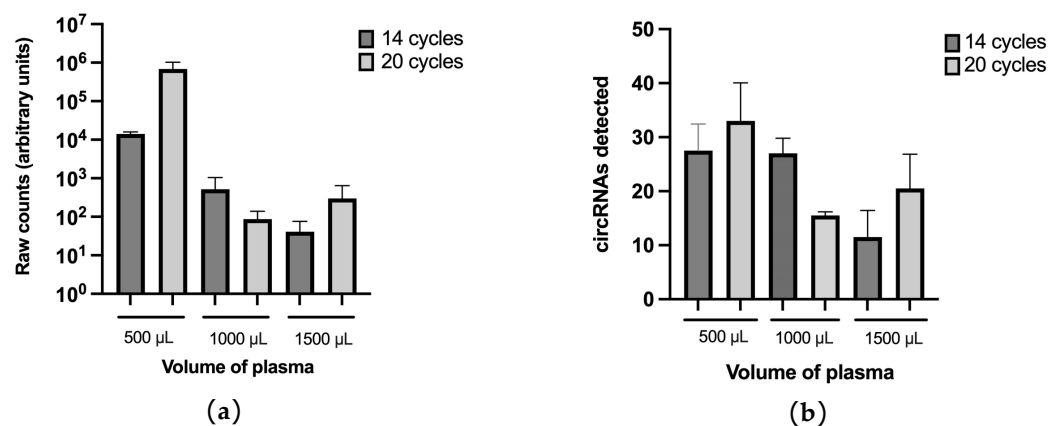


Figure 2. Plasma input testing. (a) Total number of counts and (b) number of circRNAs detected by nCounter with each of the volumes of plasma tested with 14 and 20 cycles of pre-amplification. Plasma from a NSCLC patient was used for this purpose. Error bars indicate standard deviation.

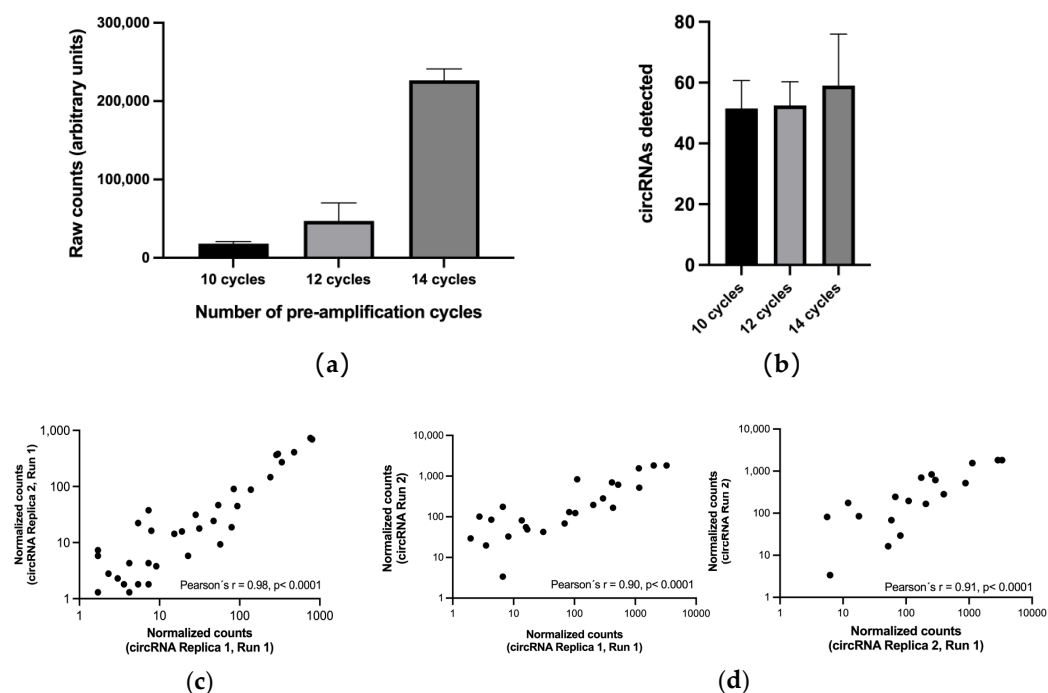


Figure 3. Testing of a different number of pre-amplification cycles. Effect of the number of pre-amplification cycles (10, 12 and 14) on (a) the total number of raw counts and (b) total number of circRNAs detected. Error bars indicate standard deviation; (c) Correlation of the two technical nCounter duplicates subjected to 10 cycles of pre-amplification. Pearson's correlation coefficient is indicated. (d) Correlation of each of the technical duplicates from the same nCounter run with the results obtained in an independent nCounter assay of the same sample. Pearson's correlation coefficient is indicated.

The repeatability of the protocol was first tested by submitting to nCounter duplicates of a preamplified plasma sample. A strong correlation between the normalized counts was found between the duplicates, represented by a Pearson's r of 0.98, $p < 0.0001$ (Figure 3c). When the same plasma sample was re-purified and re-analyzed, nCounter results also showed a strong correlation with the initial duplicates (Pearson's $r = 0.90$ – 0.91 ; $p < 0.0001$) (Figure 3d).

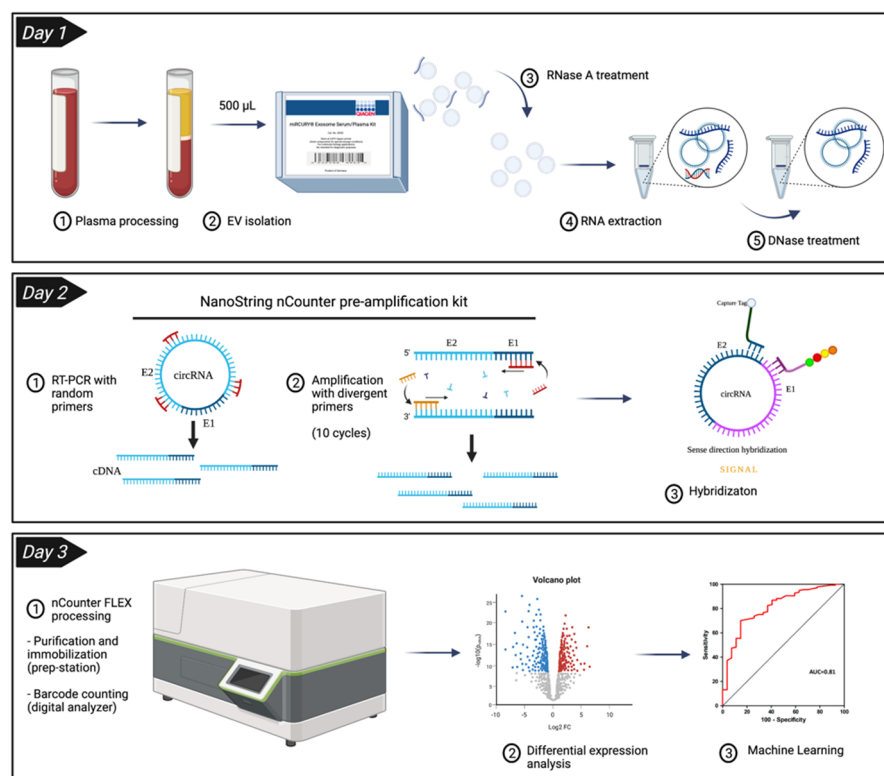


Figure 4. Final workflow established for the study of circRNAs from plasma extracellular vesicles (EVs) using the nCounter technology. A volume of 500 μ L of plasma was used in the miRCURY kit to precipitate EVs. RNase A was used to remove any non-vesicular RNA that could be present in the sample before proceeding with manual RNA extraction with TRI reagent. RNA samples were treated with DNase to eliminate any trace of genomic DNA, followed by retro-transcription and a pre-amplification step of 10 cycles. Finally, samples were hybridized overnight before nCounter processing.

3.2. CircRNA Expression in Plasma EV Samples

Plasma from 66 individuals, 36 early-stage NSCLC patients and 30 non-cancer donors, were analyzed using the protocol previously described in Section 3.1 (Figure 4). An average of 40 ± 14 EV-circRNAs per sample were detected in controls vs. 47 ± 9 in the NSCLC cohort. This difference was found to not be significant by the Mann–Whitney U test (Figure 5a). Among the 78 circRNAs included in the panel, 70 were detected in at least one NSCLC sample and 68 in at least one non-cancer control. A total of 66 EV-circRNAs were shared by both cohorts, while four EV-circRNAs were exclusive to NSCLC patients and two to non-cancer donors (Figure 5b, Table S3).

DE analysis revealed eight circRNAs significantly upregulated in EV-enriched samples from NSCLC patients vs. controls; namely circular Erythrocyte Membrane protein Band 4.1 Like 2 (circEPB41L2), circular Core 1 Synthase, Glycoprotein-N-Acetylgalactosamine -3-Beta-Galactosyltransferase 1 (circC1GALT1), circular Zinc Finger RNA Binding Protein (circZFR), circular Ubiquitin Specific Peptidase 3 (circUSP3), circular Zinc Finger CCHC Domain-Containing Protein 6 (circZCCHC6), circular Cyclin B1 (circCCNB1), circular DENN Domain Containing 1B (circDENN1B) and circular Homeodomain Interacting Protein Kinase 3 (circHIPK3) (Figure 5c). Of them, only circZFR and circC1GALT1 showed <10 counts in each cohort (Table S4). To validate these results, we tested the expression circZCCHC6 and circHIPK3 by RT-qPCR. Divergent primers and probes spanning the junction sites were designed for the specific amplification of these two circular transcripts (Table 2) in samples previously assessed by nCounter with sufficient remaining material. Gel electrophoresis of the RT-qPCR products revealed bands matching the size of expected amplicons and subsequent Sanger sequencing confirmed the expected junction sites in the circRNAs (Figure 6a,b). Among the six samples analyzed by RT-qPCR, four and six samples

produced satisfactory results for circZCCHC6 and circHIPK3 respectively. A trend between nCounter counts and RT-qPCR $\Delta\Delta C_t$ s was observed for both circRNAs (Figure 6c–d), with circZCCHC6 showing a strong correlation (Pearson’s $r = 0.99$; $p = 0.0076$) (Figure 6c).

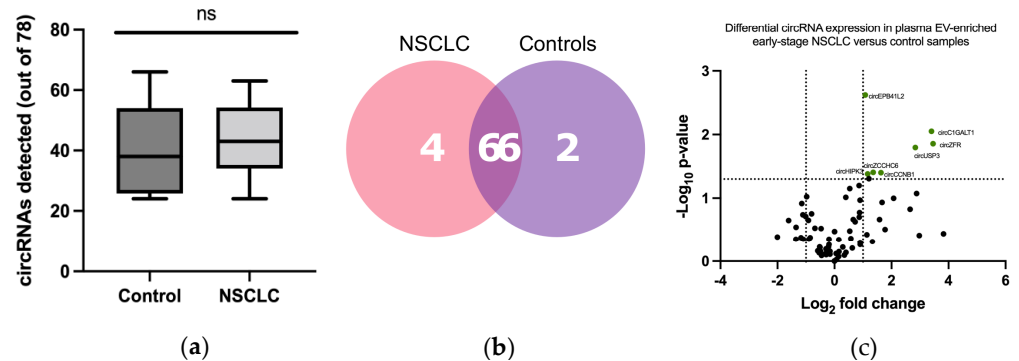


Figure 5. EV-circRNA detection and differential expression analysis. (a) Number of circRNAs detected in extracellular vesicle (EV) enriched samples from cancer patients and non-cancer controls using our custom circRNA nCounter panel, which targets 78 circRNA (Mann–Whitney U test, $p = 0.3807$); (b) Venn diagram displaying circRNAs identified in early-stage NSCLC and non-cancer controls, featuring those shared by both cohorts; (c) Differential expression analysis of \log_2 -normalized counts between the early-stage NSCLC and control EV samples. circEPB41L2, circC1GALT1, circZFR, circUSP3, circZCCHC6, circHIPK3 and circCCNB1 were found upregulated in NSCLC samples.

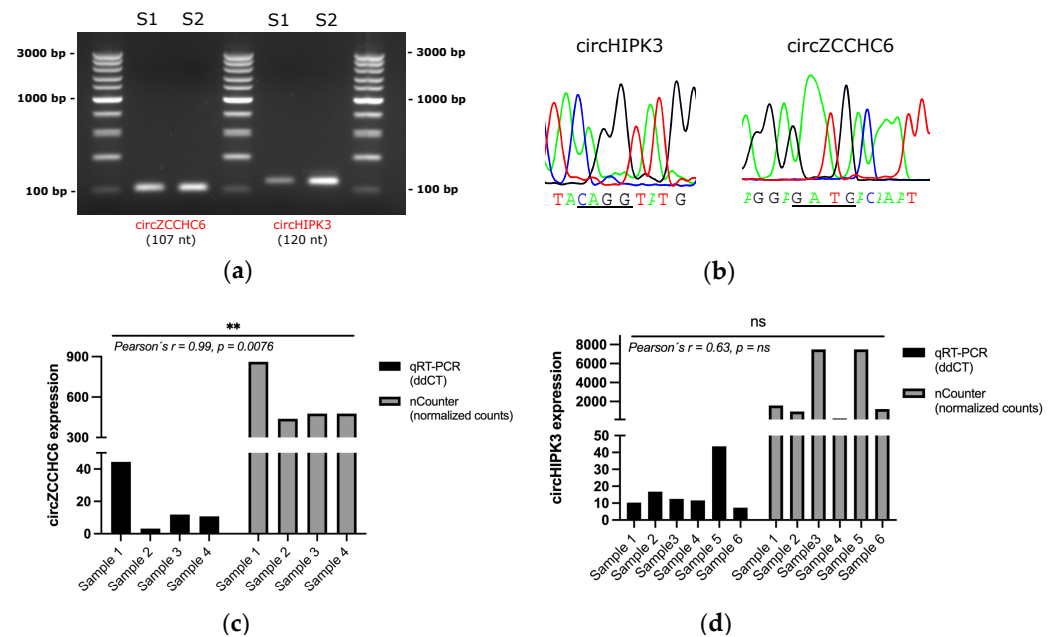


Figure 6. Validation of nCounter results by RT-qPCR and further Sanger sequencing. (a) Electrophoresis gel of amplified circZCCHC6 (107 nt) and circHIPK3 (120 nt); (b) Sanger sequencing results spanning the junction site (underlined) of cited circRNAs; Comparison of nCounter normalized counts versus $\Delta\Delta C_t$ s values by RT-qPCR for circZCCHC6 (c) and circHIPK3 (d) in analyzed samples. Pearson’s correlation coefficient is indicated. ns, not significant. ** means two grades of significant ($p < 0.01$).

3.3. Development of a CircRNA-Signature Associated with Early-Stage NSCLC

Interquartile range analysis classified 9/66 samples as potential outliers (Figure 7a) and PCA revealed that they deviated from the main cluster of observations (Figure S2). Consequently, these nine samples were excluded from further analysis.

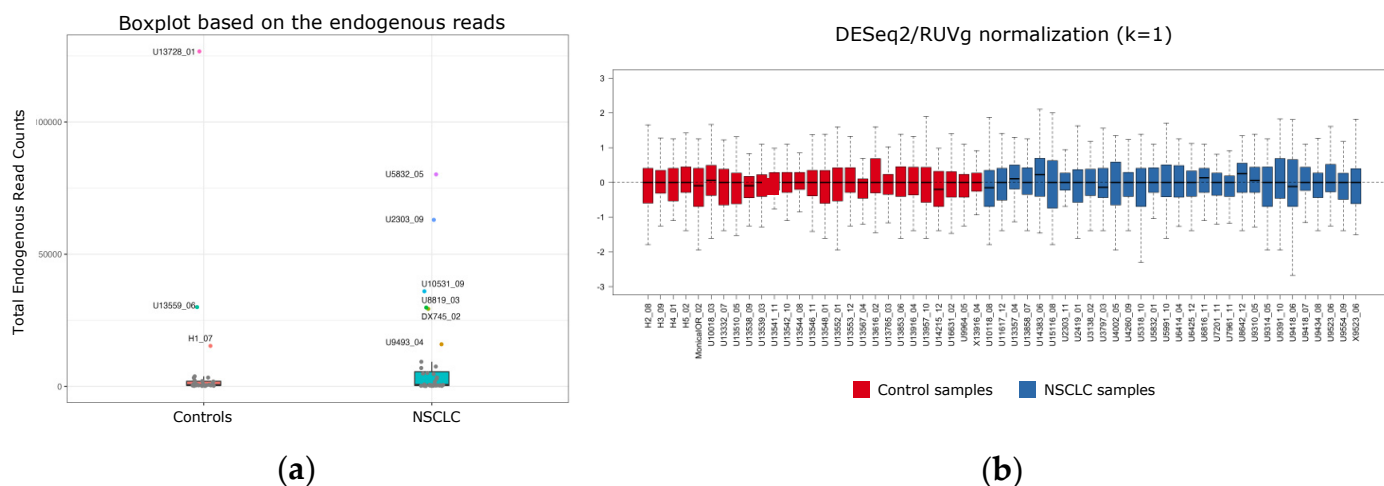


Figure 7. Data outlier detection and normalization for machine learning (ML) processing. (a) Outlier detection using the 1.5 IQR rule; (b) RUVSeq/DESeq2 RLE plot of normalized data (k = 1).

Then, different R packages including DESeq2, edgeR, RUVSeq and their combination were tested in order to select the normalization approach that best adapts to our data. As a result, RLE plots indicated a superior performance of RUVSeq-DESeq2 versus the other combinations (Figures 7b and S3). Consequently, RUVSeq-DESeq2 normalization was selected for the rest of the study.

Next, ML was performed using RFE along with RF classifier and LOOCV, as described in Methods, in order to obtain a signature associated with NSCLC. As a result, ETC was selected as the best model, with a signature of 10 circRNAs (including circular Family With Sequence Similarity 13 Member B -circFAM13B, circular ADAM Metallopeptidase Domain 22 -circADAM22, circular UBX Domain Protein 7 -circUBXN7, circZCCHC6, circular Integrin Subunit Alpha X -circITGAX, circular Retinol Dehydrogenase 11 -circRDH11, circEPB41L2, circular CDC Like Kinase 1 -circCLK1, circular Phenylalanyl-tRNA Synthetase Subunit Alpha -circFARSA, and circular Phosphoinositide-3-Kinase Regulatory Subunit 1 -circPIK3R1) showing an AUC ROC of 0.86 (Figure 8a). Signature scores were found to be statistically different when comparing early-stage NSCLC and non-cancer controls (Mann–Whitney U test, $p < 0001$; Figure 8b). The sensitivity and specificity of the ETC signature were of 90% (CI = 73.47–97.89%) and 81% (CI = 61.92%–93.70%) respectively, outperforming the RF and KNN classifiers (Table 3). The accuracy achieved with ETC was 86%, resulting in 49 out of the 66 cases being correctly classified (Figure 8c).

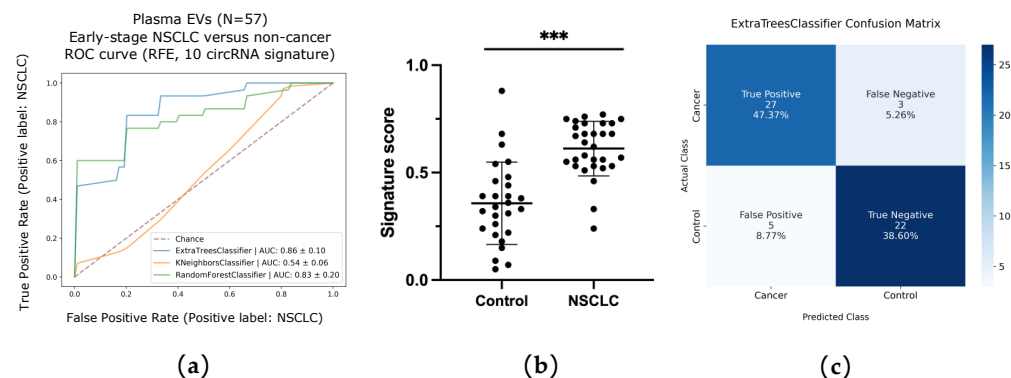


Figure 8. Machine learning (ML) analysis of extracellular vesicle (EV)-enriched samples. (a) Area under the ROC curve of the 10 circRNA-signature using recursive feature elimination (RFE) for cohort classification; (b) Scores of early-stage NSCLC versus control samples based on expression of the 10-circRNA signature ($p < 0.001$ in a two-tailed Mann–Whitney U test); (c) Confusion matrix based on the ETC classification scores. ***: $p < 0.001$.

Table 3. Precision assessment of the ML generated circRNA signature with ETC, RF and KNN. The 95% CI are indicated.

Model	ETC	RF	KNN
No. concordant samples	49	44	30
No. discordant samples	8	13	27
AUC ROC	0.86	0.83	0.54
Accuracy	86%	77%	53%
Sensitivity	90%	83%	50%
	(CI = 73.47–97.89%)	(CI = 65.28–94.36%)	CI = 31.30–68.70%)
Specificity	81%	70%	56%
	(CI = 61.92–93.70%)	(CI = 49.82–86.25%)	CI = 41.83–68.49%)
PPV	84%	76%	56%
	(CI = 70.81–92.32%)	(CI = 63.10–85.10%)	(CI = 41.83–68.49%)
NPV	88%	79%	50%
	(CI = 71.18–95.61%)	(CI = 62.20–89.77%)	(CI = 37.95–62.02%)
Cohen’s κ	0.72	0.54	0.06
	(CI = 0.458–0.976)	(CI = 0.281–0.798)	(CI = −0.202–0.313)

ML = machine learning, AUC = area under the curve, ROC = receiver operating characteristic, RF = random forest, KNN = K-nearest neighbor, CI = confidence interval, PPV = positive predictive value, NPV = negative predictive value.

Then, a univariate analysis was performed to explore the association of the ETC circRNA signature with gender, age, smoking, cancer status and tumor stage (Figure 9a). A statistically significant correlation was found for the signature with age (odds ratio = 24.91, $p < 0.0001$), and particularly cancer status (odds ratio of 39.6, $p < 0.0001$).

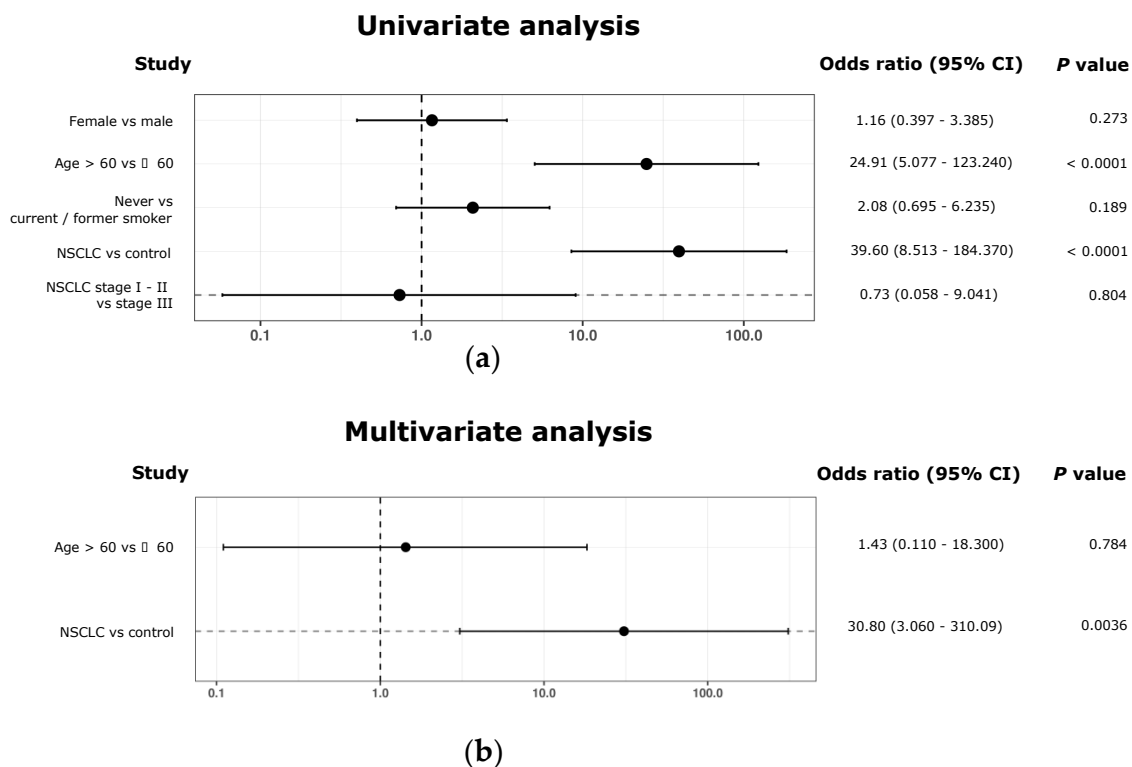


Figure 9. Association between clinical characteristics and ML-generated 10-circRNA signature. (a) Univariate analysis exploring associations between presented 10-circRNA signature and patient characteristics. Forest plot represents the odds ratios with a 95% Wald confidence limit. (b) Multivariate analysis exploring associations between presented 10-circRNA signature with age and cancer status. Forest plot represents the odds ratios with a 95% Wald confidence limit.

To further evaluate the implication of age and cancer status on the ML-developed signature, we first performed an exploratory study assessing the interconnexion of both variables by performing a chi-square test. As a result, a strong association between age and cancer status was found, with a $p < 0.0001$ (Table 4).

Table 4. Association between age and cancer status.

Statistic	DF	Value	<i>p</i> -Value
Chi-Square	1	32.245	<0.0001
Likelihood Ratio Chi-Square	1	41.232	<0.0001

DF = Degrees of freedom.

Next, a multivariate analysis was carried out. Results not only demonstrated dependency of these two variables, but also showed a statistically significant correlation between the signature and cancer status ($p = 0.0036$, Table 5, Figure 9b). No correlation was found between age and presented signature, in this regard ($p = 0.0784$, Table 5, Figure 9b)

Table 5. Analysis of maximum likelihood estimates.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	<i>p</i> -Value
Age	1	0.356	1.301	0.075	0.7840
Cancer status	1	3.427	1.178	8.462	0.0036

DF = Degrees of freedom.

4. Discussion

EVs are released by most cell types and play an important role in cancer cell communication. Many publications have demonstrated the role of EVs as key modulators in cancer progression [34,35], which requires intercellular communication mediated by the horizontal transferring of biological information via the EV cargo of proteins, DNA and coding/non-coding RNA, including circRNAs. Therefore, analysis of EVs can provide a snapshot of the tumor and serve as a valuable tool to discover liquid biopsy biomarkers. CircRNAs are highly enriched in EVs [7] and show a relatively high stability compared to other forms of RNA [8]. Several studies have highlighted their potential as liquid biopsy biomarkers [12] but current limitations in circRNA quantification methods are limiting their implementation in the clinical setting. Consequently, new, and robust protocols for circRNA analysis are needed. The nCounter platform has gained popularity among translational investigators for transcriptional research not only for solid biopsies but also for EV samples. However, studies focusing on circRNA analysis by nCounter are limited and mostly restricted to tissue specimens [30,36–40]. In particular, to the best of our knowledge, nCounter has never been applied to the analysis of circRNA in liquid biopsies of lung cancer patients. Consequently, we developed a comprehensive protocol for nCounter-based EV-circRNA expression analysis, from EV enrichment to differential expression and subsequent ML analysis. Key points in this protocol were the initial volume of plasma, the EV purification method and the number of cycles for the pre-amplification step prior to nCounter testing.

UC is currently still the method of choice for EV isolation in the research setting and we have previously demonstrated its utility for the downstream analysis of cell line-derived EV circRNAs [41]. However, ultracentrifuges are not usually available in clinical laboratories, while precipitation-based kits such as the miRCURY Exosome Serum/Plasma Kit represent an easily implementable option with a simple, on-the-bench protocol and short hands-on time. In our study, we compared the two methodologies using plasma samples from an NSCLC patient and a healthy donor. The presence of EV-like particles in all preparations was confirmed by TEM and nanoFCM. Interestingly, a more uniform EV population with an exosomal size-range was found by TEM in both cancer and control samples processed with the miRCURY kit, along with a higher concentration of 40–200 nm particles observed

by nanoFCM. A possible explanation to this event could be a size-selective enrichment attributed to this type of precipitation-based preparations, as previously reported in serum samples [42]. This finding prompted us to select miRCURY for further assay development. In addition, a higher number of EV-like particles was observed in the cancer sample compared to the control, regardless the isolation method used. Although a higher number of samples should be analyzed for further confirmation, preliminary results are in agreement with previous reports indicating a higher abundance of EVs in cancer patients [43].

Finally, adding to the evidence provided by TEM and nanoFCM, a treatment with RNase A was applied to EV-enriched samples prior to EV lysis and incorporated into our protocol to eliminate any extravesicular RNA. The resulting and subsequently analyzed RNA proved to be protected from the digestion of cited ribonuclease, indicating a vesicular origin of the transcripts.

In a previous study, a volume of 500 μ L of plasma was found to be sufficient for the analysis of EV-derived mRNA by nCounter [25]. Here, we compared several plasma volumes and found that 500 μ L outperformed 1000 and 1500 μ L for circRNAs analysis, both in terms of the number of circRNA molecules detected and total counts. A possible explanation for these results may rely on saturation issues with the circRNAs/reporter-probe complexes when a higher plasma input is applied, which impede a correct molecule identification by the digital analyzer. Regarding the number of cycles for the preamplification step, we investigated a range from 10 to 20 in an effort to reduce amplification-related background noise to a minimum, and we found that a 10-cycle pre-amplification step yielded adequate results.

Then, we applied our protocol to assess circular transcripts in early-stage NSCLC samples ($n = 36$) and to non-tumor controls ($n = 30$). We found that eight circRNAs were found differentially expressed between the two cohorts. Among them, circEPB41L2, circZCCHC6 and circHIPK3 showed the highest number of counts in early-stage cancer patients (Table S4). Interestingly, we previously found circEPB41L2 differentially expressed in FFPE tissues of early-stage lung cancer patients [30] and found that it displayed four binding sites with hsa-miR-942, which has been described as an activator of the Wnt/ β -catenin signaling pathway [44,45] in colorectal and esophageal cancers. Our results warrant further investigation in the biology of this circRNA to characterize its role in lung cancer. Regarding circHIPK3, it has been extensively investigated in lung cancer and found to exert a dual activity over miR-149 [46] and miR-124 [47,48], inducing cell proliferation and inhibiting apoptosis. Our results are in agreement with these findings, since circHIPK3 was upregulated in EV samples from early-stage NSCLC patients. Finally, circZCCHC6 has been recently described to regulate lysophosphatidylcholine acyltransferase 1 (LPCAT1) levels via miR-433-3p [49] in lung cancer. We used circinteractome (www.circinteractome.nia.nih.gov) to investigate possible additional miRNA binding sites, finding matches for 7 additional transcripts (miR-579-3p, miR-623, miR-1197, miR-1304 miR-5481, miR-605 and miR-935). All these miRNAs have been reported to be downregulated in lung tumors and have been related with poor prognosis, tumor growth and metastases [50–56].

ML and other computational methods based on artificial intelligence (AI) have emerged in the last decade for multileveled analysis of different datasets. In particular, ML enables computers to make predictions by finding patterns within analyzed data [57], offering a novel approach for the development of predictive signatures that often reach a higher predictive value than biomarkers found by differential expression analyses. Consequently, we decided to use ML in our study. To this end, we developed a pipeline with several steps. First, using IQR and PCA plots, we identified nine outliers, which were excluded from downstream analyses. An RLE plot from each different normalization procedure was generated, showing a higher performance of the RUVSeq-DESeq2 function when compared to the other combinations (Figures 7b and S3A–C). Finally, we used RFE along with LOOCV and the RF classifier as the feature selection algorithm to automatically determine the most significant circRNAs which are best suited for the construction of the prognostic signature. The final 10-circRNA signature included two of the eight circular transcripts previously

found by differential expression analysis and eight additional transcripts, including circFARSA. Interestingly, circFARSA has been described as a plasma biomarker of NSCLC [58], promoting tumor invasion and metastases via the PTEN/PI3K/AKT axis [59].

Since we did not sort EV populations, we could not verify the vesicular cell or tissue origin of the circRNAs included in the ML signature nor the origin of the circular transcripts, either cancer cells or tumor microenvironment. Also, we did not investigate the biological role of the circRNAs, being out of the scope of our work.

In addition, while multivariate analysis could demonstrate that classification accuracy of presented signature is based on cancer status and no other clinicopathological characteristics (Figure 9), the lack of > 60-year-old individuals was a limitation in the study. The inclusion of equivalent cohorts in terms of age should be taking into consideration for the design of forthcoming validation studies.

Finally, all 36 cancer samples included in this study were lung adenocarcinomas, with the exception of 4 squamous carcinoma and 5 NSCLC samples with unknown histological subtype. A uniform inclusion of the different lung cancer histologies is suggested for future validation studies to assess the predictive power of the signature for other subtypes of NSCLC.

5. Conclusions

We have demonstrated the feasibility of using nCounter for the multiplex study of plasma-EV circRNAs in liquid biopsies of lung cancer patients, including differential expression analysis and development of predictive ML signatures. Further studies of larger cohorts are warranted in order to determine the clinical applicability of such signatures.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pharmaceutics14102034/s1>, Table S1: CircRNAs detected in the different plasma volumes of the same patient with 14 and 20 cycles of pre-amplification; Table S2: CircRNAs detected in the plasma of the same individual subjected to 10, 12 and 14 pre-amplification cycles; Table S3: CircRNAs identified in early-state NSCLC and non-cancer control cohorts; Table S4: Normalized counts of differentially expressed circRNAs found in the early-stage NSCLC cohort; Figure S1: Plasma input testing; Figure S2: Principal Component Analysis of the transformed raw data; Figure S3: Assessment of the different normalization processes by RLE plot analysis; Figure S4: Confusion matrices summarizing the performance of the different classification algorithms.

Author Contributions: Conceptualization, C.P.-V. and R.R.; methodology, C.P.-V., S.G. and J.W.P.B.; software, S.G.; formal analysis, C.P.-V., S.G. and A.D.; investigation, C.P.-V., S.G., A.G.-C., D.F., M.F., J.B.-A. and J.V.; resources, A.A.-H., R.R. and M.Á.M.-V.; data curation, C.P.-V.; writing—original draft preparation, C.P.-V.; writing—review and editing, M.Á.M.-V.; visualization, C.P.-V. and S.G.; supervision, R.R., M.Á.M.-V., A.F.-H., M.H. and N.Z.; validation, C.P.-V.; project administration, C.P.-V.; funding acquisition, R.R. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement ELBA No 765492.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Quirón Hospitals (2021/11-ONC-DEX-HUSC-HUGC. 15-02-2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author (carlosedraz@icloud.com) upon reasonable request.

Acknowledgments: We would like to thank Neil Bertram for his language editing assistance. The investigators also wish to thank the patients for kindly agreeing to donate samples to this study. Figure 4 was created with Biorender.com.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [[CrossRef](#)] [[PubMed](#)]
2. European Union. ECIS—European Cancer Information System. Available online: <https://ecis.jrc.ec.europa.eu> (accessed on 30 June 2022).
3. Crosby, D.; Bhatia, S.; Brindle Kevin, M.; Coussens Lisa, M.; Dive, C.; Emberton, M.; Esener, S.; Fitzgerald Rebecca, C.; Gambhir Sanjiv, S.; Kuhn, P.; et al. Early detection of cancer. *Science* **2022**, *375*, eaay9040. [[CrossRef](#)] [[PubMed](#)]
4. Perakis, S.; Speicher, M.R. Emerging concepts in liquid biopsies. *BMC Med.* **2017**, *15*, 75. [[CrossRef](#)] [[PubMed](#)]
5. Lone, S.N.; Nisar, S.; Masoodi, T.; Singh, M.; Rizwan, A.; Hashem, S.; El-Rifai, W.; Bedognetti, D.; Batra, S.K.; Haris, M.; et al. Liquid biopsy: A step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol. Cancer* **2022**, *21*, 79. [[CrossRef](#)] [[PubMed](#)]
6. Bracht, J.W.P.; Mayo-de-Las-Casas, C.; Berenguer, J.; Karachaliou, N.; Rosell, R. The Present and Future of Liquid Biopsies in Non-Small Cell Lung Cancer: Combining Four Biosources for Diagnosis, Prognosis, Prediction, and Disease Monitoring. *Curr. Oncol. Rep.* **2018**, *20*, 70. [[CrossRef](#)]
7. Li, Y.; Zheng, Q.; Bao, C.; Li, S.; Guo, W.; Zhao, J.; Chen, D.; Gu, J.; He, X.; Huang, S. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis. *Cell Res.* **2015**, *25*, 981–984. [[CrossRef](#)]
8. Xiao, M.-S.; Wilusz, J.E. An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. *Nucleic Acids Res.* **2019**, *47*, 8755–8769. [[CrossRef](#)]
9. Hansen, T.B.; Jensen, T.I.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* **2013**, *495*, 384–388. [[CrossRef](#)]
10. Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S.D.; Gregersen, L.H.; Munschauer, M.; et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **2013**, *495*, 333–338. [[CrossRef](#)]
11. Zhou, R.; Wu, Y.; Wang, W.; Su, W.; Liu, Y.; Wang, Y.; Fan, C.; Li, X.; Li, G.; Li, Y.; et al. Circular RNAs (circRNAs) in cancer. *Cancer Lett.* **2018**, *425*, 134–142. [[CrossRef](#)]
12. 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. Nucleic Acids* **2021**, *2*, 179–201. [[CrossRef](#)]
13. Zhou, Q.; Ju, L.-L.; Ji, X.; Cao, Y.-L.; Shao, J.-G.; Chen, L. Plasma circRNAs as Biomarkers in Cancer. *Cancer Manag. Res.* **2021**, *13*, 7325–7337. [[CrossRef](#)] [[PubMed](#)]
14. Kulkarni, M.M. Digital Multiplexed Gene Expression Analysis Using the NanoString nCounter System. *Curr. Protoc. Mol. Biol.* **2011**, *94*, 25B.10.1–25B.10.17. [[CrossRef](#)]
15. Geiss, G.K.; Bumgarner, R.E.; Birditt, B.; Dahl, T.; Dowidar, N.; Dunaway, D.L.; Fell, H.P.; Ferree, S.; George, R.D.; Grogan, T.; et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **2008**, *26*, 317–325. [[CrossRef](#)]
16. Warren, S. Simultaneous, Multiplexed Detection of RNA and Protein on the NanoString® nCounter® Platform. In *Gene Expression Analysis: Methods and Protocols*; Raghavachari, N., Garcia-Reyero, N., Eds.; Springer: New York, NY, USA, 2018; pp. 105–120. [[CrossRef](#)]
17. Giménez-Capitán, A.; Bracht, J.; García, J.J.; Jordana-Ariza, N.; García, B.; Garzón, M.; Mayo-de-las-Casas, C.; Viteri-Ramirez, S.; Martínez-Bueno, A.; Aguilar, A.; et al. Multiplex Detection of Clinically Relevant Mutations in Liquid Biopsies of Cancer Patients Using a Hybridization-Based Platform. *Clin. Chem.* **2021**, *67*, 554–563. [[CrossRef](#)]
18. Porras, T.B.; Kaur, P.; Ring, A.; Schechter, N.; Lang, J.E. Challenges in using liquid biopsies for gene expression profiling. *Oncotarget* **2018**, *9*, 7036–7053. [[CrossRef](#)]
19. Beck, T.N.; Bumber, Y.A.; Aggarwal, C.; Pei, J.; Thrash-Bingham, C.; Fittipaldi, P.; Vlasenkova, R.; Rao, C.; Borghaei, H.; Cristofanilli, M.; et al. Circulating tumor cell and cell-free RNA capture and expression analysis identify platelet-associated genes in metastatic lung cancer. *BMC Cancer* **2019**, *19*, 603. [[CrossRef](#)]
20. Wu, T.-C.; Xu, K.; Martinek, J.; Young, R.R.; Banchereau, R.; George, J.; Turner, J.; Kim, K.I.; Zurawski, S.; Wang, X.; et al. IL1 Receptor Antagonist Controls Transcriptional Signature of Inflammation in Patients with Metastatic Breast Cancer. *Cancer Res.* **2018**, *78*, 5243–5258. [[CrossRef](#)]
21. Kossenkov, A.V.; Qureshi, R.; Dawany, N.B.; Wickramasinghe, J.; Liu, Q.; Majumdar, R.S.; Chang, C.; Widura, S.; Kumar, T.; Horng, W.-H.; et al. A Gene Expression Classifier from Whole Blood Distinguishes Benign from Malignant Lung Nodules Detected by Low-Dose CT. *Cancer Res.* **2019**, *79*, 263–273. [[CrossRef](#)]
22. Kamyabi, N.; Abbasgholizadeh, R.; Maitra, A.; Ardekani, A.; Biswal, S.L.; Grande-Allen, K.J. Isolation and mutational assessment of pancreatic cancer extracellular vesicles using a microfluidic platform. *Biomed. Microdevices* **2020**, *22*, 23. [[CrossRef](#)]
23. Garcia-Contreras, M.; Shah, S.H.; Tamayo, A.; Robbins, P.D.; Golberg, R.B.; Mendez, A.J.; Ricordi, C. Plasma-derived exosome characterization reveals a distinct microRNA signature in long duration Type 1 diabetes. *Sci. Rep.* **2017**, *7*, 5998. [[CrossRef](#)] [[PubMed](#)]
24. Vicentini, C.; Calore, F.; Nigita, G.; Fadda, P.; Simbolo, M.; Sperandio, N.; Luchini, C.; Lawlor, R.T.; Croce, C.M.; Corbo, V.; et al. Exosomal miRNA signatures of pancreatic lesions. *BMC Gastroenterol.* **2020**, *20*, 137. [[CrossRef](#)] [[PubMed](#)]
25. Bracht, J.W.P.; Gimenez-Capitan, A.; Huang, C.-Y.; Potie, N.; Pedraz-Valdunciel, C.; Warren, S.; Rosell, R.; Molina-Vila, M.A. Analysis of extracellular vesicle mRNA derived from plasma using the nCounter platform. *Sci. Rep.* **2021**, *11*, 3712. [[CrossRef](#)]

26. Hansen, E.B.; Fredsøe, J.; Okholm, T.L.H.; Ulhøi, B.P.; Klingenberg, S.; Jensen, J.B.; Kjems, J.; Bouchelouche, K.; Borre, M.; Damgaard, C.K.; et al. The transcriptional landscape and biomarker potential of circular RNAs in prostate cancer. *Genome Med.* **2022**, *14*, 8. [[CrossRef](#)]
27. Berenguer, J.; Lagerweij, T.; Zhao, X.W.; Dusoswa, S.; van der Stoop, P.; Westerman, B.; Gooijer, M.C.d.; Zoetemelk, M.; Zomer, A.; Crommentuijn, M.H.W.; et al. Glycosylated extracellular vesicles released by glioblastoma cells are decorated by CCL18 allowing for cellular uptake via chemokine receptor CCR8. *J. Extracell. Vesicles* **2018**, *7*, 1446660. [[CrossRef](#)] [[PubMed](#)]
28. Tian, Y.; Gong, M.; Hu, Y.; Liu, H.; Zhang, W.; Zhang, M.; Hu, X.; Aubert, D.; Zhu, S.; Wu, L.; et al. Quality and efficiency assessment of six extracellular vesicle isolation methods by nano-flow cytometry. *J. Extracell. Vesicles* **2020**, *9*, 1697028. [[CrossRef](#)] [[PubMed](#)]
29. Aguado, C.; Giménez-Capitán, A.; Román, R.; Rodríguez, S.; Jordana-Ariza, N.; Aguilar, A.; Cabrera-Gálvez, C.; Rivas-Corredor, C.; Lianes, P.; Viteri, S.; et al. RNA-Based Multiplexing Assay for Routine Testing of Fusion and Splicing Variants in Cytological Samples of NSCLC Patients. *Diagnostics* **2020**, *11*, 15. [[CrossRef](#)]
30. Pedraz-Valdunciel, C.; Giannoukagos, S.; Potie, N.; Giménez-Capitán, A.; Huang, C.-Y.; Hackenberg, M.; Fernandez-Hilario, A.; Bracht, J.; Filipiska, M.; Aldeguer, E.; et al. Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens. *Mol. Oncol.* **2022**, *16*, 2367–2383. [[CrossRef](#)]
31. Margolis, L.; Sadovsky, Y. The biology of extracellular vesicles: The known unknowns. *PLOS Biol.* **2019**, *17*, e3000363. [[CrossRef](#)]
32. Reclusa, P.; Sireira, R.; Araujo, A.; Giallombardo, M.; Valentino, A.; Sorber, L.; Bazo, I.G.; Pauwels, P.; Rolfo, C. Exosomes genetic cargo in lung cancer: A truly Pandora's box. *Transl. Lung Cancer Res.* **2016**, *5*, 483–491. [[CrossRef](#)]
33. Jeppesen, D.K.; Fenix, A.M.; Franklin, J.L.; Higginbotham, J.N.; Zhang, Q.; Zimmerman, L.J.; Liebler, D.C.; Ping, J.; Liu, Q.; Evans, R.; et al. Reassessment of Exosome Composition. *Cell* **2019**, *177*, 428–445.e418. [[CrossRef](#)] [[PubMed](#)]
34. Yang, X.; Zhang, Y.; Zhang, Y.; Zhang, S.; Qiu, L.; Zhuang, Z.; Wei, M.; Deng, X.; Wang, Z.; Han, J. The Key Role of Exosomes on the Pre-metastatic Niche Formation in Tumors. *Front. Mol. Biosci.* **2021**, *8*, 703640. [[CrossRef](#)] [[PubMed](#)]
35. Costa-Silva, B.; Aiello, N.M.; Ocean, A.J.; Singh, S.; Zhang, H.; Thakur, B.K.; Becker, A.; Hoshino, A.; Mark, M.T.; Molina, H.; et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat. Cell Biol.* **2015**, *17*, 816–826. [[CrossRef](#)]
36. Dahl, M.; Daugaard, I.; Andersen, M.S.; Hansen, T.B.; Grønbaek, K.; Kjems, J.; Kristensen, L.S. Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab. Invest.* **2018**, *98*, 1657–1669. [[CrossRef](#)]
37. Zhang, J.; Zhang, X.; Li, C.; Yue, L.; Ding, N.; Riordan, T.; Yang, L.; Li, Y.; Jen, C.; Lin, S.; et al. Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells. *RNA Biol.* **2019**, *16*, 220–232. [[CrossRef](#)] [[PubMed](#)]
38. Das Mahapatra, K.; Pasquali, L.; Søndergaard, J.N.; Lapins, J.; Nemeth, I.B.; Baltás, E.; Kemény, L.; Homey, B.; Moldovan, L.-I.; Kjems, J.; et al. A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma. *Sci. Rep.* **2020**, *10*, 3637. [[CrossRef](#)]
39. Moldovan, L.-I.; Hansen, T.B.; Venø, M.T.; Okholm, T.L.H.; Andersen, T.L.; Hager, H.; Iversen, L.; Kjems, J.; Johansen, C.; Kristensen, L.S. High-throughput RNA sequencing from paired lesional- and non-lesional skin reveals major alterations in the psoriasis circRNAome. *BMC Med. Genom.* **2019**, *12*, 174. [[CrossRef](#)]
40. Ahmadov, U.; Bendikas, M.M.; Ebbesen, K.K.; Sehested, A.M.; Kjems, J.; Broholm, H.; Kristensen, L.S. Distinct circular RNA expression profiles in pediatric ependymomas. *Brain Pathol.* **2021**, *31*, 387–392. [[CrossRef](#)]
41. Pedraz-Valdunciel, C.; Huang, C.; Ito, M.; Bracht, J.; Giménez-Capitán, A.; Aldeguer, E.; Filipiska, M.; Xu, W.; Molina-Vila, M.A.; Rosell, R. P65.04 Tracking circRNAs in Lung Adenocarcinoma Samples as Promising Biomarkers for Cancer Detection using the NanoString nCounter®. *Thorac. Oncol.* **2021**, *16*, S555–S556. [[CrossRef](#)]
42. Helwa, I.; Cai, J.; Drewry, M.D.; Zimmerman, A.; Dinkins, M.B.; Khaled, M.L.; Seremwe, M.; Dismuke, W.M.; Bieberich, E.; Stamer, W.D.; et al. A Comparative Study of Serum Exosome Isolation Using Differential Ultracentrifugation and Three Commercial Reagents. *PLoS ONE* **2017**, *12*, e0170628. [[CrossRef](#)]
43. Melo, S.A.; Luecke, L.B.; Kahlert, C.; Fernandez, A.F.; Gammon, S.T.; Kaye, J.; LeBleu, V.S.; Mittendorf, E.A.; Weitz, J.; Rahbari, N.; et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **2015**, *523*, 177–182. [[CrossRef](#)] [[PubMed](#)]
44. Fasihi, A.; Soltani, B.M.; Ranjbaran, Z.S.; Bahonar, S.; Norouzi, R.; Nasiri, S. Hsa-miR-942 fingerprint in colorectal cancer through Wnt signaling pathway. *Gene* **2019**, *712*, 143958. [[CrossRef](#)] [[PubMed](#)]
45. Ge, C.; Wu, S.; Wang, W.; Liu, Z.; Zhang, J.; Wang, Z.; Li, R.; Zhang, Z.; Li, Z.; Dong, S.; 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*, 10964–10977. [[CrossRef](#)] [[PubMed](#)]
46. Lu, H.; Han, X.; Ren, J.; Ren, K.; Li, Z.; Sun, Z. Circular RNA HIPK3 induces cell proliferation and inhibits apoptosis in non-small cell lung cancer through sponging miR-149. *Cancer Biol. Ther.* **2020**, *21*, 113–121. [[CrossRef](#)]
47. Chen, X.; Mao, R.; Su, W.; Yang, X.; Geng, Q.; Guo, C.; Wang, Z.; Wang, J.; Kresty, L.A.; Beer, D.G.; et al. Circular RNA circHIPK3 modulates autophagy via MIR124-3p-STAT3-PRKAA/AMPK α signaling in STK11 mutant lung cancer. *Autophagy* **2020**, *16*, 659–671. [[CrossRef](#)]
48. Yu, H.; Chen, Y.; Jiang, P. Circular RNA HIPK3 exerts oncogenic properties through suppression of miR-124 in lung cancer. *Biochem. Biophys. Res. Commun.* **2018**, *506*, 455–462. [[CrossRef](#)]
49. Guo, Y.; Xue, W.; Sun, S.; Chen, X.; Li, H.; Yan, C. Circular RNA circZCCHC6 contributes to tumorigenesis by regulating LPCAT1 via miR-433-3p in non-small cell lung cancer. *Clin. Exp. Med.* **2022**. [[CrossRef](#)]

50. Wu, R.-R.; Zhong, Q.; Liu, H.-F.; Liu, S.-B. Role of miR-579-3p in the development of squamous cell lung carcinoma and the regulatory mechanisms. *Eur. Rev. Med. Pharmacol. Sci.* **2019**, *23*, 9464–9470.
51. Wei, S.; Zhang, Z.-y.; Fu, S.-l.; Xie, J.-g.; Liu, X.-s.; Xu, Y.-j.; Zhao, J.-p.; Xiong, W.-n. Erratum: Hsa-miR-623 suppresses tumor progression in human lung adenocarcinoma. *Cell Death Dis.* **2017**, *8*, e2829. [[CrossRef](#)]
52. Sun, B.; Hua, J.; Cui, H.; Liu, H.; Zhang, K.; Zhou, H. MicroRNA-1197 downregulation inhibits proliferation and migration in human non-small cell lung cancer cells by upregulating HOXC11. *Biomed. Pharmacother.* **2019**, *117*, 109041. [[CrossRef](#)]
53. Li, C.-g.; Pu, M.-f.; Li, C.-z.; Gao, M.; Liu, M.-x.; Yu, C.-z.; Yan, H.; Peng, C.; Zhao, Y.; Li, Y.; et al. MicroRNA-1304 suppresses human non-small cell lung cancer cell growth in vitro by targeting heme oxygenase-1. *Acta Pharmacol. Sin.* **2017**, *38*, 110–119. [[CrossRef](#)] [[PubMed](#)]
54. Liu, C.; Yang, H.; Xu, Z.; Li, D.; Zhou, M.; Xiao, K.; Shi, Z.; Zhu, L.; Yang, L.; Zhou, R. microRNA-548l is involved in the migration and invasion of non-small cell lung cancer by targeting the AKT1 signaling pathway. *J. Cancer Res. Clin. Oncol.* **2015**, *141*, 431–441. [[CrossRef](#)] [[PubMed](#)]
55. Liao, Y.; Cao, L.; Wang, F.; Pang, R. miR-605-5p promotes invasion and proliferation by targeting TNFAIP3 in non-small-cell lung cancer. *J. Cell. Biochem.* **2020**, *121*, 779–787. [[CrossRef](#)]
56. Wang, C.; Li, S.; Xu, J.; Niu, W. microRNA-935 is reduced in non-small cell lung cancer tissue, is linked to poor outcome, and acts on signal transduction mediator E2F7 and the AKT pathway. *Br. J. Biomed. Sci.* **2019**, *76*, 17–23. [[CrossRef](#)] [[PubMed](#)]
57. Badillo, S.; Banfai, B.; Birzele, F.; Davydov, I.I.; Hutchinson, L.; Kam-Thong, T.; Siebourg-Polster, J.; Steiert, B.; Zhang, J.D. An Introduction to Machine Learning. *Clin. Pharmacol. Ther.* **2020**, *107*, 871–885. [[CrossRef](#)]
58. Hang, D.; Zhou, J.; Qin, N.; Zhou, W.; Ma, H.; Jin, G.; Hu, Z.; Dai, J.; Shen, H. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. *Cancer Med.* **2018**, *7*, 2783–2791. [[CrossRef](#)] [[PubMed](#)]
59. Chen, T.; Liu, Y.; Li, C.; Xu, C.; Ding, C.; Chen, J.; Zhao, J. Tumor-derived exosomal circFARSA mediates M2 macrophage polarization via the PTEN/PI3K/AKT pathway to promote non-small cell lung cancer metastasis. *Cancer Treat. Res. Commun.* **2021**, *28*, 100412. [[CrossRef](#)] [[PubMed](#)]