



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

DEPARTMENT OF BIOTECHNOLOGY

Ph.D. Thesis

**Tumorspheres as an *in vitro*
model for cancer stem-like cells
characterization in non-small cell
lung cancer. Prognostic
implications**

Alejandro Herreros Pomares

Supervisors:

Dr. Carlos Camps Herrero

Dr. Eloísa Jantus Lewintre

November 2019



VNIVERSITAT
DE VALÈNCIA



H.R.G. CONSORCI
HOSPITAL GENERAL
UNIVERSITARI
VALÈNCIA

CARLOS CAMPS HERRERO, Dr. por la Universitat de València, Catedrático del Departamento de Medicina de la Universitat de València y Jefe del Servicio de Oncología del Hospital General Universitari de València.

CERTIFICA:

Que D. Alejandro Herreros Pomares, Graduado en Biotecnología, ha realizado bajo mi dirección la Tesis Doctoral que lleva por título **“Tumorspheres as an *in vitro* model for Cancer Stem-like Cells characterization in Non-Small Cell Lung Cancer. Prognostic implications”**. Dicha tesis reúne todos los requisitos necesarios para su juicio y calificación.

En Valencia, 15 de noviembre de 2019



Fdo.: Carlos Camps
Profesor Titular de Medicina
Facultad de Medicina
Universitat de València

Prof. Carlos Camps Herrero

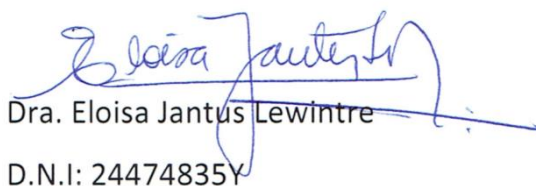
D.N.I.: 22517012N

ELOISA JANTUS LEWINTRE, Dra. por la Universitat de València, Jefa del Laboratorio de Oncología Molecular de la Fundación del Hospital General Universitari de València y Profesora Asociada del Departamento de Biotecnología de la Universitat Politècnica de València.

CERTIFICA:

Que D. Alejandro Herreros Pomares, Graduado en Biotecnología, ha realizado bajo mi dirección la Tesis Doctoral que lleva por título **“Tumorspheres as an *in vitro* model for Cancer Stem-like Cells characterization in Non-Small Cell Lung Cancer. Prognostic implications”**. Dicha tesis reúne todos los requisitos necesarios para su juicio y calificación.

En Valencia, 15 de noviembre de 2019


Dra. Eloisa Jantus Lewintre
D.N.I: 24474835Y

**“If we knew what it was we were doing,
it would not be called *research*, would it?”**

Albert Einstein

AGRADECIMIENTOS

Me gustaría comenzar esta sección dando las gracias a todas las personas que durante estos años han contribuido a que esta tesis doctoral sea una realidad.

En primer lugar, me gustaría agradecer al Dr. Carlos Camps la confianza que depositó en mí, abriéndome las puertas de su laboratorio y dándome la oportunidad de formar parte de un equipo excepcional. Muchas gracias por tu apoyo, por tus lecciones y también por tu exigencia, que me ha llevado a ser cada día un poco mejor.

Del mismo modo, me gustaría dar las gracias a la Dra. Eloísa Jantus, la que ha supuesto para mí, no sólo una directora de tesis inmejorable, sino una referente, un ejemplo a seguir de disciplina, trabajo duro y rigurosidad, una madre científica. Gracias por tu ayuda inestimable en este viaje, por tus correcciones, consejos y palabras siempre acertadas en el momento preciso.

También me gustaría agradecer a la Dra. Silvia Calabuig todo lo que ha hecho por mí durante estos años. Crec que ja saps tot el que he de dir-te, però el deixaré escrit perquè no ho dubtes mai. Eres collonuda. Un exemple de sacrifici, dedicació, altruisme i treball dur. He trobada una germana major en tu a qui contar-li qualsevol cosa, des dels meus problemes personals fins a laborals i sé que sempre trobaré un recolze en tu, un ànim i la cara positiva de la vida. Som molt diferents i això ens ha dut a no estar sempre d'acord, però també són eixes relacions les que més ens fan créixer i aprendre. Gràcies, no canvies mai.

Expresar también mi más sincero agradecimiento al Dr. Rafael Sirera por apoyarme, ayudarme y animarme a convertirme en un gran científico. Durante mi formación has sido un gran referente y sé que lo seguirás siendo a lo largo de mi carrera investigadora. Muchas gracias por todo.

Me gustaría también dar las gracias a todos los compañeros del Hospital General de Valencia, tanto médicos adjuntos, como residentes, enfermeras, data managers, etc. por su importante labor asistencial, su ayuda, dedicación y colaboración. Por supuesto,

agradecer a todos los pacientes sus donaciones desinteresadas para la investigación contra el cáncer y a sus familiares, por ser su gran apoyo y consuelo.

Igualmente, dar las gracias a todos los colaboradores externos que han hecho posible esta tesis doctoral. En especial, a la Dra. Rosa Farràs por su contribución a los análisis proteicos e *in vivo*, al Dr. Sergio Alonso por su ayuda en los análisis bioinformáticos y a la Dra. María Dolores Chiara por su colaboración en los análisis de invasión celular.

I would also like to thank Prof. Lijie Grace Zhang, who opened her laboratory at the George Washington University for me, not only for her scientific support during my internship, but also for her concern about my well-being. I would like to extend my gratitude to all the members of the lab, especially to Xuan, Tim and Se-Jun for their kindness and for providing me with help whenever I needed it.

Y como no, agradecer a todos los compañeros de la Fundación para la Investigación del Hospital General de Valencia, por su contribución a este trabajo, por crear un ambiente de trabajo increíble y por hacer que me lleve muy buen recuerdo de mi paso por aquí. Gracias a Román y Sisco por su buen humor, a Miriam, Ángela, Carla y Joana y, cómo no, a mis Corti, con las que tan buenos ratos he pasado. Muchas gracias a Bea por ser mi bastón, mi amiga y asesora y, por supuesto, a las Sonias, Adela y Anselm.

Dar las gracias a todas las personas que durante estos años han pasado por el laboratorio y que, de una manera u otra, han aportado su granito de arena para que este trabajo saliera adelante. Ester y Borja, 'mis pollitos', porque entramos juntos y fuisteis mi gran apoyo en los inicios, seguid así, porque vais a llegar muy lejos. Andrea y Adrián por guiarme y enseñarme cómo hacer una buena PCR, que ha sido núcleo central de esta tesis doctoral y también a todos los estudiantes que tanto me han enseñado. Gracias a Lidia, Víctor, Dani, Judith, Elena, María y Marta.

A Eva, sabes que eres un gran ejemplo para mí y te deseo la mejor de las suertes. Ojalá ser como tú de mayor. Muchas gracias por tu apoyo, por tu ayuda y por consentirme. Este agradecimiento no es menos para Marais. Entramos juntos y te convertiste en la mamá del laboratorio. Muchas gracias por apoyarnos, estar siempre pendiente y dispuesta a todo, eres una grande y te mereces todo lo bueno que te pase, lucha por

ello. A Sandra, que durant anys has sigut la meua amiga, consellera i confident. Valora't, perquè no eres conscient de la teua vàlua i és moltíssima. Els teus límits només els estableixes tu, Dra. Gallach. Per al que necessites, ací tens un amic.

Dar las gracias también a Eva Aguilera, Cristóbal y Álvaro por su apoyo en este tiempo.

A todos mis compañeros predocs, con los que durante estos años he sufrido, reído y llorado, haciendo que nuestros trabajos salieran adelante. Muchas gracias a Franklin, Héctor, Bruno, Clara y Elena. Sin vosotros este proceso no habría sido igual. Gracias a Andrea. Aún recuerdo cuando llegaste aquí y nos podíamos pasar horas en cultivos con nuestros secretos, confesiones y preocupaciones y ahora, mírate, pronto serás la Dra. Moreno. Llegarás muy alto, pequeña. Algo similar me pasa con Susana. Sabía que lo conseguirías y aquí estás. Sigue adelante pese a la adversidad, eres una curranta y llegarás donde quieras, gracias por las risas y el apoyo.

A mi familia biotec, Alba, Cristina, Ainhoa, Raquel e Isa, porque sois un regalo, y a mi familia vasca, Ane y María, porque no se encuentran 'top, top friends' todos los días y sé que a pesar del tiempo y la distancia puedo contar con vosotras siempre.

Pero como no podía ser de otra manera, esta tesis tiene que ir dedicada a mi familia, la que ha creído en mí y me ha dado todo para que hoy pueda estar aquí. Gracias a mis abuelos, tíos y primos por su apoyo, en especial a mi tía Julia y a mi prima Cristina por estar siempre ahí, pero sobre todo a mi madre, por su amor incondicional y por sacrificar su vida para darnos todo a sus hijos. Eres única y te adoro.

Va para ti, José Daniel.

ABSTRACT

Lung cancer is the most commonly diagnosed type of cancer and the leading cause of cancer-related death worldwide, with approximately 15% of patients surviving 5 years after diagnosis. Curative surgery is the standard of care for early-stage patients with a good performance status, but 75% are diagnosed at advanced stages, when surgery is not possible, and 35-50% of the resected patients relapse after an apparently successful surgical treatment. Significant advances in the development of therapies against driver mutations and immune-based treatments for these patients have been achieved in recent years, but many patients still develop treatment resistance, progress, and die. The high resistance against these therapies has been associated to cancer stem-like cells (CSCs), a population with stem properties which is able to survive after conventional treatments and regenerate tumor even when are undetectable.

In this thesis, primary cultures from early-stage non-small cell lung cancer (NSCLC) patients were established, using sphere-forming assays for CSCs enrichment and adherent conditions for the control counterparts. Patient-derived tumorspheres showed self-renewal and unlimited exponential growth potentials, resistance against chemotherapeutic agents, invasion and differentiation capacities *in vitro*, and superior tumorigenic potential *in vivo*. Using RTqPCR, gene expression profiles were analyzed, and *NANOG*, *NOTCH3*, *CD44*, *CDKN1A*, *SNAI1*, and *ITGA6* were selected as the best contributors to distinguish tumorspheres from adherent cells. Immunoblot and immunofluorescence analyses confirmed that proteins encoded by these genes were consistently increased in tumorspheres from adenocarcinoma patients and showed differential localization and expression patterns. The prognostic role of genes significantly overexpressed in tumorspheres was evaluated *in silico* in a cohort of 661 NSCLC patients from TCGA. Based on a Cox regression analysis, *CDKN1A*, *SNAI1* and *ITGA6* were found to be associated with prognosis and used to calculate a gene expression score, named CSCs score. Kaplan-Meier survival analysis showed that patients with high CSCs score have shorter overall survival (OS) in the entire cohort [37.7 vs. 60.4 months, $p = 0.001$] and in the adenocarcinoma (ADC) subcohort [36.6 vs. 53.5 months, $p = 0.003$], but not in the squamous cell carcinoma one. Multivariate analysis

indicated that this gene expression score is an independent biomarker of prognosis for OS in both, the entire cohort [hazard ratio (HR): 1.498; 95% confidence interval (CI), 1.167-1.922; $p = 0.001$], and the ADC subcohort [HR: 1.869; 95% CI, 1.275-2.738; $p = 0.001$]. This score was also analyzed in an independent group of 245 patients from *Consorti Hospital General Universitari de València*, confirming its prognostic value in the ADC subtype [42.90 vs. not reached (NR) months, $p = 0.020$]. In conclusion, our findings provide relevant prognostic information for lung ADC patients and the basis for developing novel therapies.

RESUMEN

El cáncer de pulmón es el tipo de cáncer más frecuentemente diagnosticado y la principal causa de muerte debida a cáncer en el mundo, con sólo un 15% de pacientes con una supervivencia mayor a 5 años tras el diagnóstico. La resección quirúrgica es el tratamiento estándar para los pacientes en estadios tempranos con un buen ECOG, pero el 75% de los pacientes son diagnosticados en estadios avanzados, cuando la intervención quirúrgica no es posible y entre un 35% y un 50% de los pacientes operados recaen tras una cirugía aparentemente exitosa. En los últimos años, se han logrado importantes avances en el desarrollo de la inmunoterapia y de tratamientos contra mutaciones conductoras, pero muchos pacientes todavía desarrollan resistencia, progresan y mueren. Esta resistencia terapéutica ha sido asociada a las células madre tumorales (CMTs), una población tumoral con propiedades de célula madre capaz de sobrevivir a las terapias convencionales y regenerar el tumor incluso cuando son indetectables.

En esta tesis doctoral, se establecieron cultivos primarios de pacientes de cáncer de pulmón no microcítico (CPNM) resecados, usando ensayos de formación de tumoresferas para el enriquecimiento en CMTs y condiciones de adherencia para los controles. Las tumoresferas derivadas de pacientes mostraron capacidad de autorenovación y crecimiento exponencial ilimitado, alta resistencia a agentes quimioterápicos, capacidad de invasión y diferenciación *in vitro* y un elevado potencial tumorigénico *in vivo*. Usando PCR cuantitativa, se analizaron los perfiles de expresión de los cultivos y se determinó que *NANOG*, *NOTCH3*, *CD44*, *CDKN1A*, *SNAI1* e *ITGA6* eran los genes más diferencialmente expresados entre tumoresferas y células adherentes. Los análisis de inmunoblot e inmunofluorescencia confirmaron que las proteínas codificadas por estos genes se encuentran aumentadas en tumoresferas de los pacientes con adenocarcinoma y mostraron patrones de expresión y localización diferencial entre éstas y los controles en adherencia. El valor pronóstico de los genes significativamente sobreexpresados en tumoresferas fue evaluado *in silico* en una cohorte de 661 pacientes con CPNM procedente del TCGA. De todos ellos, *CDKN1A*, *SNAI1* y *ITGA6* mostraron estar relacionados con el pronóstico de los pacientes de

acuerdo a un análisis de regresión de Cox y fueron seleccionados para construir una firma de expresión génica, denominada firma de CMTs. Los análisis de supervivencia por Kaplan-Meier mostraron que los pacientes con valores elevados de la firma tienen una supervivencia global (SG) menor para la cohorte completa de CPNM [37,7 vs. 60,40 meses, $p = 0,001$] y para la subcohorte de adenocarcinoma (ADC) [36,6 vs. 53,5 meses, $p = 0,003$], pero no para la de los epidermoides. Además, el análisis multivariante mostró que la firma de CMTs es un marcador pronóstico independiente para la SG de los pacientes en la cohorte completa [*hazard ratio* (HR): 1,498; intervalo de confianza (IC) 95%, 1,167-1,922; $p = 0,001$] y la subcohorte de ADC [HR: 1,869; IC 95%, 1,275-2,738; $p = 0,001$]. Esta firma fue también analizada en un grupo independiente de 245 pacientes procedentes del *Consorci Hospital General Universitari de València*, confirmando su valor pronóstico en los pacientes con ADC [42,90 vs. no alcanzado (NA) meses, $p = 0,020$]. En resumen, nuestros hallazgos aportan información pronóstica relevante para los pacientes con ADC de pulmón y establecen las bases para el desarrollo de nuevos tratamientos.

RESUM

El càncer de pulmó és el tipus de càncer més diagnosticat i la principal causa de mort deguda a càncer en el món, amb només un 15% de pacients amb una supervivència major a 5 anys després del diagnòstic. La resecció quirúrgica és el tractament estàndard per als pacients en estadis primaris amb un bon ECOG, però el 75% dels pacients són diagnosticats en estadis avançats, quan la intervenció quirúrgica no és possible i entre un 35% i un 50% dels pacients operats recauen després d'una cirurgia aparentment satisfactòria. En els últims anys, s'han aconseguit importants avanços en el desenvolupament de la immunoteràpia i de tractaments contra mutacions conductores, però molts pacients encara desenvolupen resistència, progressen i moren. Aquesta resistència a les teràpies ha estat relacionada amb les cèl·lules mare tumorals (CMTs), una població tumoral amb propietats de cèl·lula mare capaç de sobreviure a les teràpies convencionals i regenerar el tumor fins i tot quan són indetectables.

En aquesta tesi doctoral, es van establir cultius primaris de pacients de càncer de pulmó no microcític (CPNM) ressecats, usant assajos de formació de tumoresferes per a l'enriquiment en CMTs i condicions d'adherència per als controls. Les tumoresferes derivades de pacients van mostrar capacitat d'autorenovació, creixement exponencial il·limitat, alta resistència a agents quimioteràpics, capacitat d'invasió i diferenciació *in vitro* i un elevat potencial tumorigènic *in vivo*. Usant PCR quantitativa, es van analitzar els perfils d'expressió dels cultius i es va determinar que *NANOG*, *NOTCH3*, *CD44*, *CDKN1A*, *SNAI1* i *ITGA6* eren els gens més diferencialment expressats entre tumoresferes i cèl·lules adherents. Les anàlisis de immunoblot i immunofluorescència van confirmar que les proteïnes codificades per aquests gens es troben augmentades en tumoresferes dels pacients amb adenocarcinoma i van mostrar patrons d'expressió i localització diferencial entre aquestes i els controls en adherència. El valor pronòstic dels gens significativament sobreexpressats en tumoresferes va ser avaluat *in silico* en una cohort de 661 pacients amb CPNM procedent del TCGA. De tots ells, *CDKN1A*, *SNAI1* i *ITGA6* van mostrar estar relacionats amb el pronòstic dels pacients d'acord a una anàlisi de regressió de Cox i van ser seleccionats per a construir una signatura d'expressió gènica, denominada signatura de CMTs. Les anàlisis de supervivència per Kaplan-Meier

van mostrar que els pacients amb valors elevats de la signatura tenen una supervivència global (SG) menor per a la cohort completa de CPNM [37,7 vs. 60,40 mesos, $p = 0,001$] i per a la subcohort d'adenocarcinoma (ADC) [36,6 vs. 53,5 mesos, $p = 0,003$], però no per a la dels escamosos. A més, l'anàlisi multivariant va mostrar que la signatura de CMTs és un marcador pronòstic independent per a la SG dels pacients en la cohort completa [*hazard ratio*, (HR): 1,498; interval de confiança (IC) 95%, 1,167-1,922; $p = 0,001$] i la subcohort d'ADC [HR: 1,869; IC 95%, 1,275-2,738; $p = 0,001$]. Aquesta signatura va ser també analitzada en un grup independent de 245 pacients procedents del Consorci Hospital General Universitari de València, confirmant el seu valor pronòstic en els pacients amb ADC [42,90 vs. no arribat (NA) mesos, $p = 0,020$]. En resum, els nostres resultats aporten informació pronòstica rellevant per als pacients amb ADC de pulmó i estableixen les bases per al desenvolupament de nous tractaments.

INDEX

LIST OF ABBREVIATIONS.....XXIII

LIST OF FIGURES.....XXVII

LIST OF TABLES.....XXXI

I. INTRODUCTION..... 1

1. CANCER..... 3

1.1. CONCEPT 3

1.2. MOLECULAR BIOLOGY..... 3

2. LUNG CANCER 6

2.1. EPIDEMIOLOGY 6

2.2. ETIOLOGY 7

2.3. DIAGNOSIS AND PROGNOSIS 8

2.4. HISTOLOGICAL CLASSIFICATION..... 11

2.5. MOLECULAR CLASSIFICATION 13

2.6. TREATMENT..... 16

3. CANCER STEM CELLS 18

3.1. TUMOR HETEROGENEITY 18

3.2. PROPERTIES 21

3.3. *IN VITRO* CULTURE 24

3.4. MARKERS..... 26

3.5. SIGNALING PATHWAYS 30

3.5.1. Notch signaling pathway..... 31

3.5.2. Wnt signaling pathway..... 35

3.5.3. Hedgehog signaling pathway 39

3.5.4. Other signaling pathways..... 42

II. OBJECTIVES.....45

III. MATERIALS & METHODS.....49

1. MATERIALS.....	51
1.1. PATIENTS AND SAMPLES.....	51
1.1.1. PRIMARY CULTURE ESTABLISHMENT.....	53
1.2. CELL LINES	54
1.3. ANIMALS.....	55
2. METHODS.....	56
2.1. CELL PROLIFERATION AND INVASION ANALYSIS.....	56
2.1.1. GROWTH CURVES	56
2.1.2. CITOTOXICITY ASSAYS	56
2.1.3. CELL INVASION ASSAYS AND TIME-LAPSE VIDEO RECORDING	57
2.2. NUCLEIC ACID ANALYSIS	57
2.2.1. RNA AND DNA ISOLATION	57
2.2.2. DETERMINATION OF THE MUTATIONAL STATUS	58
2.2.2.1. DETECTION OF <i>EGFR</i> MUTATIONS	58
2.2.2.2. DETECTION OF <i>KRAS</i> MUTATIONS	59
2.2.3. QUANTIFICATION OF GENE EXPRESSION.....	61
2.2.3.1. REVERSE TRANSCRIPTION	61
2.2.3.2. QUANTITATIVE REAL TIME PCR.....	61
2.3. PROTEIN ANALYSIS	64
2.3.1. IMMUNOBLOTTING	64
2.3.2. IMMUNOFLUORESCENCE.....	66
2.3.3. IMMUNOHISTOCHEMISTRY	68
2.3. BIOINFORMATIC AND STASTITICAL ANALYSIS	70

IV. RESULTS & DISCUSSION.....73

1. ESTABLISHMENT OF ADHERENT AND NON-ADHERENT CULTURES ...	75
1.1. PHENOTYPIC CHARACTERIZATION OF PRIMARY CULTURES.....	81
1.1.1. SELF-RENEWAL, PROLIFERATION AND DIFFERENTIATION CAPACITIES.....	81
1.1.2. INVASION AND TUMOR INITIATION CAPACITIES.....	83
1.1.3. RESISTANCE TO CLASSICAL CHEMOTHERAPEUTIC AGENTS	85
2. GENE EXPRESSION ANALYSIS	87
2.1. RNA QUANTIFICATION AND QUALITY ASSESSEMENT.....	87
2.2. RELATIVE mRNAs EXPRESSION.....	88
2.3. UNSUPERVISED PRINCIPALCOMPONENTS ANALYSIS	93
3. PROTEIN EXPRESSION ANALYSIS	99
3.1. IMMUNOBLOTTING.....	99
3.2. IMMUNOFLUORESCENCE.....	102
3.1. IMMUNOHISTOCHEMISTRY	106
4. <i>IN SILICO</i> SURVIVAL ANALYSIS.....	108
4.1. CLINICOPATHOLOGICAL VARIABLES.....	108
4.2. INDIVIDUAL BIOMARKERS.....	113
4.3. GENE EXPRESSION SCORE	120
4.4. MULTIVARIATE ANALYSIS.....	122
5. VALIDATION OF THE PROGNOSTIC GENE EXPRESSION SCORE	124
5.1. CLINICOPATHOLOGICAL VARIABLES.....	124
5.2. INDIVIDUAL BIOMARKERS.....	130
5.3. GENE EXPRESSION SCORE	131
6. PHARMACOLOGICAL INHIBITION OF CSCs POTENTIAL TARGETS	133
7. INTEGRATION OF RESULTS.....	135

V. CONCLUSIONS	139
VI. REFERENCES	143
VII. APPENDICES	179
1. SUPPLEMENTARY MATERIAL.....	181
2. APPROVAL FROM THE INSTITUTIONAL ETHICAL AND SCIENTIFIC REVIEW BOARD	183
3. FUNDING	185
4. NATIONAL AND INTERNATIONAL CONGRESS COMMUNICATIONS .	185
5. PUBLICATIONS.....	190
6. AWARDS	190

LIST OF ABBREVIATIONS

ABC: ATP binding cassette
ADC: Adenocarcinoma
AIC: Akaike's information criterion
ALDH: Aldehyde dehydrogenases
ALK: Anaplastic lymphoma kinase
AMOTL2: Angiomotin-like protein 2
AP-1: Activator protein 1
APC: Adenomatous polyposis coli
AREG: Amphiregulin
AXIN: Axis inhibition protein
BCRP: Breast cancer resistance protein
bFGF: Basic fibroblast growth factor
BRAF: B-Raf proto-oncogene
BSA: Bovine serum albumin
CEA: Carcinoembryonic antigen
CI: Confidence interval
CK1 α : Casein kinase 1 α
CK7: Cytokeratin 7
COS: Atypical kinesin-like protein Costa
CREB: cAMP response element-binding protein
CSCs: Cancer stem cells
CT: Computed tomography
CTGF: Connective tissue growth factor
CYR61: Cysteine-rich angiogenic inducer 61
DDR2: Discoidin domain receptor tyrosine kinase 2
DHH: Desert hedgehog
DTT: Dithiothreitol
EGF: Epidermal growth factor

EGFR: Epidermal growth factor receptor
EML4: Echinoderm microtubule-associated protein-like 4
EMT: Epithelial-mesenchymal transition
EUS: Endoscopic ultrasound
EBUS: Endobronchial ultrasound
FAP: Familial adenomatous polyposis
FBS: Fetal bovine serum
FDA: Food and Drug Administration
FGFR: Fibroblast growth factor receptor 1
FFPE: Formalin-fixed paraffin-embedded
FZD: Frizzled
GSK3 β : Glycogen synthase kinase 3 β
HDPR: Human homolog of Dapper
HE: Hematoxylin and eosin staining
HER2: human epidermal growth factor receptor 2
HES: Hairy enhancer of split
HR: Hazard ratio
HRP: Horseradish peroxidase
IAP: Inhibitors of apoptosis
ICAM: Intercellular cell adhesion molecule
ICGC: International cancer genome consortium
IHC: Immunohistochemistry
IHH: Indian hedgehog
ITS : Insulin-transferrin-selenium
JNK: cJun N-terminal kinase
KIF5B: kinesin family member 5b
KLC-1: kinesin light chain 1
KRAS: Kirsten rat sarcoma viral oncogene homolog
LATS: Large tumor suppressor kinase
LEF1: Lymphoid enhancer-binding factor

LCC: Large cell carcinoma
LN: Lymph nodes
LRP: Lipoprotein receptor-related protein 5/6
MDR: Multidrug resistance transporters
MDSC: Myeloid-derived suppressor cell
MEK1: Mitogen-activated protein kinase kinase 1
MET: Mesenchymal epithelial transition factor proto-oncogene
MOB1: MOB kinase activator 1
MST: Mammalian Ste20-like protein kinase
NCI: National cancer institute
NGS: Next generation sequencing
NICD: Notch intracellular domain
NSCLC: Non-small cell lung carcinoma
OS: Overall survival
PC1: Principal component 1
PCA: Principal component analysis
PCP: Planar cell polarity
PD1: Programmed cell death 1
PD-L1: Programmed cell death 1 ligand
PET: Positron emission tomography
PES: Phenazine ethosulfate
PI3K: Phosphatidylinositol-4,5 bisphosphate 3-kinase
PLS-DA: Partial least square-discriminant analysis
PTCH: Patched homolog
PS: Performance status
PTEN: Phosphatase and tensin homolog
RB: Retinoblastoma-associated
RECIST: Response evaluation criteria for solid tumors
RET: Ret proto-oncogene
RFA: Radiofrequency ablation

RFS: Relapse-free survival

ROS1: Reactive oxygen species proto-oncogene 1

RT: Room temperature

RTqPCR: Quantitative real time PCR

SBRT: Stereotactic body radiotherapy

SAV1: Salvador family WW domain containing protein 1

SCC: Squamous cell carcinoma

SCLC: Small cell lung carcinoma

SDS: Sodium dodecyl sulfate

SHH: Sonic hedgehog

SP: Side population

SUFU: Serine-threonine protein kinase suppressor of fused

TACE: Tumor necrosis factor- α -converting enzyme

TAZ: Transcriptional co-activator with PDZ-binding motif

TBX5: T-box transcription factor 5

TCGA: The cancer genome atlas

TICs: Tumor initiating cells

TGF: TRK-fused gene

TKI: Tyrosine-kinase inhibitors

TNF: Tumor necrosis factor

TNM: Tumor/Lymph Nodes/Metastasis

TRAIL: TNF-related apoptosis-inducing ligand

Treg: Regulatory T cells

TV: Tumor volume

UV: Unit variance

VCAM: Vascular cell adhesion molecule

VEGF: Vascular endothelial growth factor

WHO: World Health Organization

WIF-1: Wnt inhibitory factor 1

WNT: Wingless-type protein

LIST OF FIGURES

Figure 1. The ten hallmarks of cancer	5
Figure 2. Top cancer site associated with number of deaths estimated for both sexes and all ages per country in 2018	7
Figure 3. Prognostic information regarding TNM classification.....	11
Figure 4. Evolution of NSCLC subtyping from histological to molecular classification ..	13
Figure 5. Therapeutic management of advanced NSCLC patients	17
Figure 6. Inter-tumor and intra-tumor heterogeneity	19
Figure 7. Unified model of Clonal Evolution and Cancer Stem Cells.....	21
Figure 8. The canonical Notch signaling pathway	34
Figure 9. The canonical and non-canonical Wnt signaling pathway	36
Figure 10. The canonical HH-signaling pathway.....	41
Figure 11. Representation of the histograms for the codons and nucleotides according to the sequences for analysis	60
Figure 12. Immunofluorescence images of fibroblast controls and tumor cells	77
Figure 13. Representative images of the primary patient-derived cancer cells cultured under adherent (A) and non-adherent (B) conditions	78
Figure 14. Representative images of adherent FIS299 and FIS301 primary cultures	78
Figure 15. Representative images of the human non-small cell lung cancer cell lines included in the study cultured under (A) adherent and (B) suspension conditions	80
Figure 16. Self-renewal and unlimited exponential growth potential of tumor cells grown under suspension conditions.....	81
Figure 17. Cell growth curves from the primary adherent-cell cultures established	82
Figure 18. Differentiation capacity of lung tumorspheres	83
Figure 19. Tumor development capacity of adherent cells and tumorspheres.....	85

Figure 20. Cell viability of FIS302, FIS320, and FIS343 patients and NCI-H1650, PC9 and HCC827 cell lines after 48 hours of exposure to cisplatin 50 μ M, docetaxel 10 μ M, paclitaxel 10 μ M, vinorelbine 10 μ M, pemetrexed 50 μ M, and salinomycin 1 μ M.....	86
Figure 21. Analysis of the transcriptomic distribution of the samples included in the gene expression part of study	89
Figure 22. Transcription levels of the CSCs-related genes in tumorspheres versus adherent- cultured cells.....	90
Figure 23. A. PCA score plot of primary cultures and cell lines based on the gene expression of 44 CSCs-related genes. B. Loading plot of PCA analysis showing the contributing of each gene to culture distribution	94
Figure 24. A. PLS-DA score plot of primary cultures and cell lines based on the gene expression of 44 CSCs-related genes. B. Loading plot of PLS-DA analysis showing the contributing of each gene to culture distribution	95
Figure 25. A. PCA score plot of cell lines based on the gene expression of 44 CSCs-related genes. B. Loading plot of PCA analysis showing the contributing of each gene to culture distribution	97
Figure 26. A. PCA score plot of primary cultures based on the gene expression of 44 CSCs-related genes. Green dots represent adherent-cultured cells whereas blue dots represent tumorspheres. B. Loading plot of PCA analysis showing the contributing of each gene to culture distribution	98
Figure 27. A. Representative immunoblotting images of p21, Notch3, CD44, Integrin α 6, Nanog and Snail in adherent-cultured cells and tumorspheres from ADC patients. B. Relative protein expression of the same proteins in adherent-cultured cells and tumorspheres from ADC patients.	99
Figure 28. A. Representative immunoblotting images of p21, Notch3, CD44, Integrin α 6, Nanog and Snail in adherent-cultured cells and tumorspheres from SCC patients. B. Relative protein expression of the same proteins in adherent-cultured cells and tumorspheres from SCC patients	101
Figure 29. Immunoblotting images of E-cadherin, β -catenin, Vimentin, ALDH1A1, EpCAM and Sox2 in adherent-cultured cells and tumorspheres from primary cultures.....	101
Figure 30. Immunofluorescence images of p21, Notch3, CD44, Integrin α 6, Nanog and Snail in adherent-cultured cells and tumorspheres from ADC patients	103

Figure 31. Immunofluorescence images of CD133, CD166, ALDH1A1, B-catenin, E-cadherin and Vimentin in adherent-cultured cells and tumorspheres from FIS302, FIS303, FIS315, FIS317, and FIS320 patients	105
Figure 32. Representative immunohistochemical staining for CEA, CD56, CK5-6, CK7 and p63 in tumorspheres and adherent cells from FIS320 and FIS343 primary cultures...	107
Figure 33. Hematoxylin and eosin staining of a tumorsphere from FIS343 patient	108
Figure 34. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the entire cohort.....	110
Figure 35. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the ADC subcohort.....	112
Figure 36. Kaplan-Meier plots for RFS according to gene expression levels for the entire cohort.....	114
Figure 37. Kaplan-Meier plots for OS according to gene expression levels for the entire cohort from TCGA.....	115
Figure 38. Kaplan-Meier plots for OS according to gene expression levels for the ADC subcohort.....	117
Figure 39. Kaplan-Meier plots for RFS according to gene expression levels for the ADC subcohort.....	118
Figure 40 Kaplan-Meier plots for RFS and OS according to gene expression levels for the SCC subcohort.....	120
Figure 41. Univariate analysis of the expression of the 17 genes overexpressed in lung tumorspheres for OS	121
Figure 42. Kaplan-Meier plots for OS according to the gene expression score levels.	122
Figure 43. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the validation cohort	127
Figure 44. Kaplan-Meier plots for RFS according to clinicopathological variables for the ADC validation cohort.....	128
Figure 45. Kaplan-Meier plots for OS according to the gene expression levels of the ADC validation cohort	131

Figure 46. Kaplan-Meier plot for OS according to the gene expression score levels of ADC validation subcohort..... 132

Figure 47. Cell viability of FIS302, FIS320, and FIS343 patients and NCI-H1650, PC9 and HCC827 cell lines after 48 hours of exposure to curcumin, apigenin, and UC2288 134

LIST OF TABLES

Table 1. The T, N, and M descriptors in the eighth edition of the TNM classification for lung cancer	9
Table 2. Staging criteria based on the eighth edition of the TNM classification for lung cancer	10
Table 3. Compilation of proposed CSC markers for distinct solid tumor types.....	26
Table 4. Clinicopathological characteristics of the patients from <i>Consortio Hospital General Universitario de Valencia</i> included in the study	52
Table 5. Clinicopathological characteristics of TCGA patients included in the study	52
Table 6. Clinicopathological characteristics and lung tumorspheres formation capacity of the patients included in the study.....	54
Table 7. Clinicopathological characteristics of the cell lines included in the study.	55
Table 8. <i>EGFR</i> mutations detected using TheraScreen® kit.....	58
Table 9. TheraScreen® <i>EGFR</i> RGQ PCR cycling parameters	59
Table 10. TheraScreen® <i>KRAS</i> Pyro® PCR cycling parameters.....	59
Table 11. Cycling program for reverse transcription reaction	61
Table 12. Genes analyzed in this study, their description and the ID of TaqMan® assays used for RTqPCR	62
Table 13. Endogenous gene TaqMan® assays used for the normalization of results....	63
Table 14. Cycling program for RTqPCR.....	64
Table 15. Characteristics of the primary antibodies used for protein detection by immunoblotting.....	66
Table 16. Characteristics of the primary antibodies used for protein detection by immunofluorescence.....	67
Table 17. Characteristics of the primary antibodies and conditions used to detect each protein by immunohistochemistry.....	69

Table 18. Results from survival analysis based on clinicopathological variables for the entire cohort from TCGA	110
Table 19. Results from survival analysis based on clinicopathological variables for the ADC cohort from TCGA.....	111
Table 20. Results from survival analysis based on clinicopathological variables for the SCC cohort from TCGA.....	111
Table 21. Results from survival analysis based on gene expression biomarkers for the entire cohort from TCGA	113
Table 22. Results from survival analysis based on gene expression biomarkers for the ADC subcohort from TCGA	116
Table 23. Results from survival analysis based on gene expression biomarkers for the SCC subcohort from TCGA	119
Table 24. Results from the multivariate model for OS with genes included in the gene expression score	121
Table 25. Multivariate Cox regression model results including all the significant variables from TCGA	123
Table 26. Results for survival analysis based on clinicopathological variables for the validation cohort	125
Table 27. Results for survival analysis based on clinicopathological variables of the ADC validation subcohort.....	128
Table 28. Results for survival analysis based on clinicopathological variables of the SCC validation subcohort.....	129
Table 29. Results from survival analysis based on gene expression biomarkers of the validation cohort	130
Table 30. Results from survival analysis based on gene expression biomarkers of the ADC validation subcohort.....	130
Table 31. Results from survival analysis based on gene expression biomarkers of the SCC validation subcohort.....	131

I. INTRODUCTION

1. CANCER

1.1. CONCEPT

According to the World Health Organization (WHO) and the National Cancer Institute (NCI), cancer is a generic term that defines a large group of diseases that can affect almost any part of the body. Indeed, there are more than 100 types of cancer, and many subtypes of tumors can be found in each specific organ. In cancer, abnormal cells have defects in the regulatory mechanisms that control them, making cells able to grow uncontrolledly and spread into surrounding tissues. In addition to this reductionist view of cancer, there are two other dimensions of complexity. Firstly, tumors are complex tissues in which cancer cells can recruit normal cell types to serve as active collaborators, creating tumor microenvironments in which proliferation and invasion are favored (1). Secondly, the genetic diversity in populations of tumor cells is an unavoidable consequence of the genome instability, which is caused by defects affecting components of the DNA-maintenance machinery in combination with the large number of cell divisions required for the formation of macroscopic tumors (2).

1.2. MOLECULAR BIOLOGY

In the past decade, Hanahan and Weinberg suggested that all cancers have in common six characteristics which are acquired during multistep tumorigenesis (1). However, the manner of acquisition of these features as well as the order in which they are gained varies significantly across cancers. Several years later, in 2011, these authors proposed two enabling characteristics that make possible the acquisition of these six features of cancer and two emerging hallmarks consequence of new research facts (Figure 1) (3). Hence, the ten hallmarks of cancer are the following:

1. **Self-sufficiency in growth signals.** In normal conditions, cells require mitogenic growth signals to move from a quiescent state into a proliferative state. By the union of signaling molecules, such as soluble growth factors or extracellular matrix components, to transmembrane receptors, growth

signals are transmitted into cells. In the absence of these signals, normal cells are not capable of growing. However, tumor cells grow even when these interactions do not take place because of the autocrine and paracrine growth signals they generate.

2. **Evading growth suppressors.** In normal conditions, cells receive anti-proliferative signals in order to maintain their quiescence, many of which depend on the actions of tumor suppressor genes. For instance, RB (retinoblastoma-associated) protein integrates signals from diverse extracellular and intracellular sources, making it a cell-cycle progression gatekeeper. Therefore, when the RB pathway is disrupted, cells become insensitive to inhibitory growth signals and inappropriate replication continues.
3. **Apoptosis evasion.** Programed cell death by apoptosis is a natural barrier to cancer development. However, different strategies can be used by tumor cells to escape from these mechanisms. One of the most common strategies is the loss of the proapoptotic regulator *TP53*, which induces apoptosis when DNA damage is detected.
4. **Enabling replicative immortality.** The number of cell divisions that can occur during a lifetime in mammals is limited by an intrinsic cell program, which is known as the Hayflick limit. Once cells have achieved this limit, they stop growing and start senescing (4), which results from the loss of the protective function of telomeres. It has been shown that in neoplastic cells, telomeres maintain their length due to higher activity of the telomerase enzyme (5).
5. **Inducing angiogenesis.** Like normal tissues, tumors require nutrients and oxygen, and need to evacuate metabolic wastes. As a consequence, the generation of new vasculature from the pre-existing one is essential for tumor growth (6). This process is known as angiogenesis and it is regulated by the equilibrium between inducer and inhibitor factors. Tumors have the capacity to activate angiogenesis by producing inducer factors.
6. **Invasion and metastasis.** Neoplastic cells can escape from primary tumor masses and invade adjacent tissues or distant sites. The success of this

process depends on the other five characteristics and on complex changes in the physiological relationship between cells and their microenvironment.

7. **Genome instability and mutations.** Defects affecting components of the DNA-maintenance machinery lead to the accumulation of a large number of alterations in neoplastic cells, which are related to the aforementioned characteristics. The different steps involved in tumor progression are a succession of clonal expansions produced by the accumulation of mutations that generate selectively advantageous neoplastic cells.
8. **Inflammation.** Tumors are densely infiltrated by immune cells which were initially thought to be acting against them. However, it has now become clear that inflammation can contribute to tumorigenesis and tumor progression by supplying different molecules to the tumor microenvironment such as growth factors, angiogenic factors, and extracellular matrix-modifying enzymes. Inflammation seems to play a pivotal role in the earliest stages of cancer because inflammatory cells release potentially mutagenic chemicals that induce genetic mutations in the neoplastic cells.

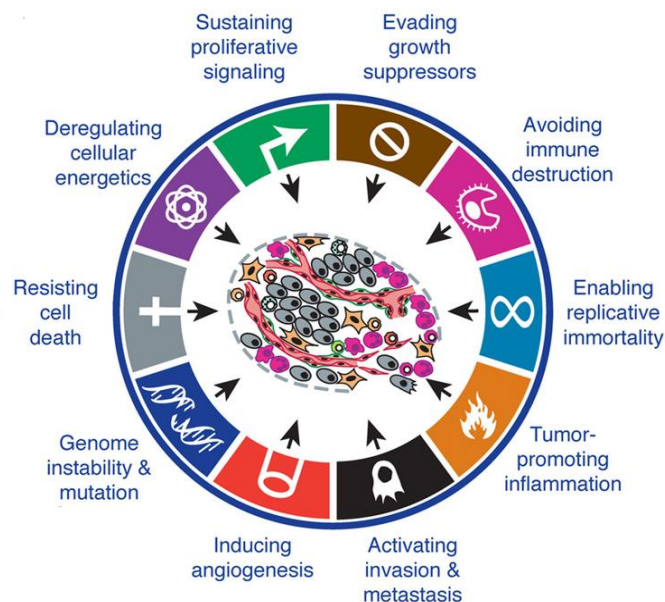


Figure 1. The ten hallmarks of cancer. This illustration represents the capabilities of cancer cells as proposed by Hanahan and Weinberg. Adapted from (3).

9. **Reprogramming energy metabolism.** During the neoplastic process, changes in energy metabolism are produced in order to avoid apoptosis and to maintain and stimulate the growth and division of neoplastic cells.
10. **Avoiding immune destruction.** It has been proven that neoplastic cells have developed different strategies in order to avoid being detected by the immune system. For instance, inflammatory cells, such as regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), which are actively immunosuppressive, are recruited to the tumor environment, where they suppress the action of cytotoxic lymphocytes against tumor cells.

2. LUNG CANCER

2.1. EPIDEMIOLOGY

In 2018, 18.1 million new cases of cancer were diagnosed and there were 9.6 million deaths worldwide (7). Among the different cancer sites, lung is the most commonly diagnosed and the leading cause of cancer death, with 2.1 million new lung cancer cases and 1.8 million deaths predicted in 2018, representing close to 1 in 5 cancer deaths (Figure 2). Trends in lung cancer mortality in Spain are similar to those observed in most developed countries. In 2017, 230,000 new cases of cancer were diagnosed in Spain, being 20,000 of them lung cancer (information obtained from Sociedad Española de Oncología Médica website). Furthermore, although cancer mortality has moderately declined across Europe in the last decade, female lung cancer is an exception in most European countries, including Spain (8). This persistent increase in lung cancer mortality reflects the women's pattern of tobacco consumption in countries of western Europe (9).

Despite the advances in biomedical research and the improvements in diagnosis and therapies of the past decades, five-year survival for lung cancer remains below 15% (10). The main reason for this poor prognosis relay on the late diagnosis of the disease when curative surgery is no longer possible.

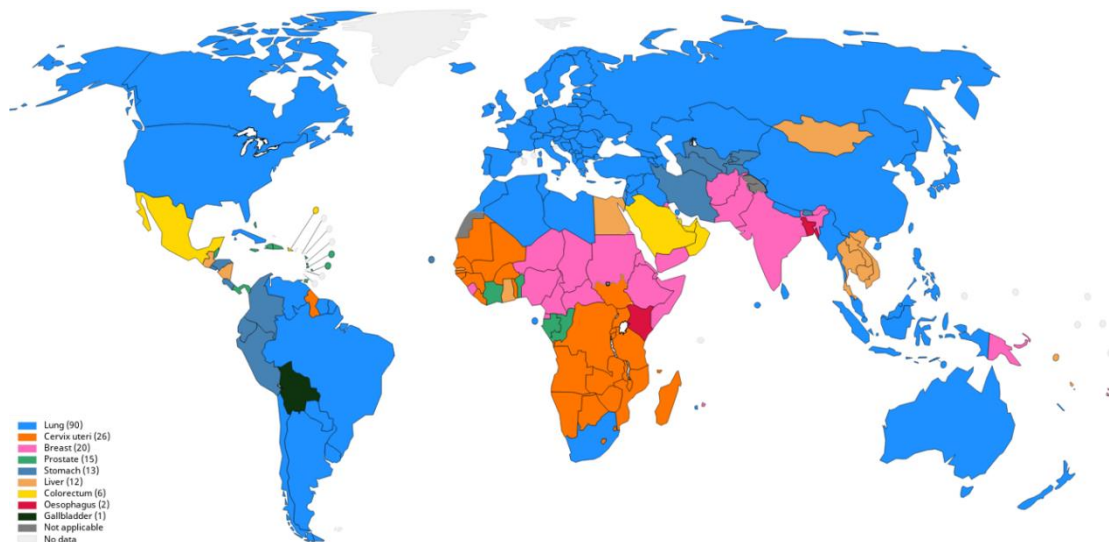


Figure 2. Top cancer site associated with number of deaths estimated for both sexes and all ages per country in 2018 (Data Source GLOBOCAN 2018, graph production: IARC).

2.2. ETIOLOGY

Tobacco smoking is the principal cause of lung cancer, being responsible for 85-90% of these tumors (11). About 4000 chemical substances, 60 of them carcinogenic, are present in cigarette smoke, triggering the accumulation of a large number of genetic mutations. As a result, it has been observed that in lung cancer, an average of 200 mutations are accumulated per tumor, while in other frequent tumors, such as breast or prostate cancer, the number of mutations ranges from 25 to 50 on average (12). Interestingly, the number of somatic mutations that appeared in smokers is ten times higher than the number of mutations found in non-smoker lung cancer patients (13). In addition, the risk of developing lung cancer increases depending on the intensity and duration of smoke exposure (14), starting to decrease after two or three years of quitting smoking (15). An increased risk of developing lung cancer (around 30%) has also been observed in passive smokers (16).

Other factors associated with lung cancer development are environmental or occupational exposures to pollutants such as radon gas, arsenic, asbestos, and polycyclic hydrocarbons (17). Additionally, diet type, alcohol consumption, some infectious agents and other diseases such as diffuse cystic fibrosis are also considered risk factors (18).

2.3. DIAGNOSIS AND PROGNOSIS

Nowadays, lung cancer is commonly detected because of symptomatic manifestations such as pain, hemoptysis, dyspnea or weight loss are developed by patients. As a result, 70-75% of patients with lung cancer present advanced disease at the time of diagnosis, with no curative surgery possible and having distant metastases in the 50% of cases (19). Routine cancer screening allows earlier diagnosis, but it presents several risks and limitations in this tumor type, including a high-rate of false-positive results, cumulative radiation exposure from multiple computed tomography scans, and unnecessary lung biopsies and surgeries. These potential harms together with economic difficulties makes the development of risk models and biomarkers a necessity (10).

Diverse technologies are now available for locating the primary tumor and further staging, including chest radiograph, low-dose chest computed tomography (CT) or fluorodeoxyglucosa (18F-FDG)-positron emission tomography (PET) (20). These imaging methods are frequently complemented with tumor biopsies using bronchoscopic techniques and endoscopic ultrasound (EUS) or endobronchial ultrasound (EBUS) for better diagnosis (21). In addition, prognostic assessment is an important factor to take into consideration when selecting an appropriate treatment regimen. The anatomical extent of disease, as described by the TNM criteria (T: size of primary tumor, N: number and position of lymph nodes, and M: presence of local or distant metastasis) is one of the most important prognostic factors in lung cancer (22). The 8th edition of *Cancer Staging Manual* was published in 2017 by the American Joint Commission on Cancer and the Union for International Cancer Control (23). The subcategories of each component of this classification are shown in Table 1.

Table 1. The T, N, and M descriptors in the eighth edition of the TNM classification for lung cancer.
Adapted from (24).

T: Primary tumor	
Tx	Primary tumor cannot be assessed or tumor proven by presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor ≤ 3 cm in greatest dimension surrounded by lung or visceral pleura without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus) ^a
T1a(mi)	Minimally invasive adenocarcinoma
T1a	Tumor ≤ 1 cm in greatest dimension ^a
T1b	Tumor >1 cm but ≤ 2 cm in greatest dimension ^a
T1c	Tumor >2 cm but ≤ 3 cm in greatest dimension ^a
T2	Tumor >3 cm but ≤ 5 cm or tumor with any of the following features ^c : - Involves main bronchus regardless of distance from the carina but without involvement of the carina - Invades visceral pleura - Associated with atelectasis or obstructive pneumonitis that extends to the hilar region, involving part or all of the lung
T2a	Tumor >3 cm but ≤ 4 cm in greatest dimension
T2b	Tumor >4 cm but ≤ 5 cm in greatest dimension
T3	Tumor >5 cm but ≤ 7 cm in greatest dimension or associated with separate tumor nodule(s) in the same lobe as the primary tumor or directly invades any of the following structures: chest wall (including the parietal pleura and superior sulcus tumors), phrenic nerve, parietal pericardium
T4	Tumor >7 cm in greatest dimension or associated with separate tumor nodule(s) in a different ipsilateral lobe than that of the primary tumor or invades any of the following structures: diaphragm, mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, and carina
N: Regional lymph node involvement	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
M: Distant metastasis	
M0	No distant metastasis
M1	Distant metastasis present
M1a	Separate tumor nodule(s) in a contralateral lobe; tumor with pleural or pericardial nodule(s) or malignant pleural or pericardial effusion ^d
M1b	Single extrathoracic metastasis ^e
M1c	Multiple extrathoracic metastases in one or more organs

^aThe uncommon superficial spreading tumor of any size with its invasive component limited to the bronchial wall, which may extend proximal to the main bronchus, is also classified as T1a.

^bSolitary adenocarcinoma, ≤ 3 cm with lepidic pattern and ≤ 5 mm invasion in any one focus.

^cT2 tumors with these features are classified as T2a if ≤ 4 cm in greatest dimension or if size cannot be determined, and T2b if >4 cm but ≤ 5 cm in greatest dimension.

^dMost pleural (pericardial) effusions with lung cancer are due to tumor. In a few patients, however, multiple microscopic examinations of pleural (pericardial) fluid are negative for tumor and the fluid is nonbloody and not an exudate. When these elements and clinical judgment dictate that the effusion is not related to the tumor, the effusion should be excluded as a staging descriptor.

^eThis includes involvement of a single distant (nonregional) lymph node.

The combination of these parameters defines the staging of lung tumors from IA1 to IVB (Table 2) and although it provides pivotal prognostic information (Figure 3), histological confirmation (including molecular characterization of the tumor) is also needed in order to make adequate decisions regarding the best treatment option for patients (24).

Table 2. Staging criteria based on the eighth edition of the TNM classification for lung cancer.
Adapted from (24).

TNM staging		N categories Overall stage		
T/M	N0	N1	N2	N3
T1a	IA1	IIB	IIIA	IIIB
T1b	IA2	IIB	IIIA	IIIB
T1c	IA3	IIB	IIIA	IIIB
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T3	IIB	IIIA	IIIB	IIIC
T3	IIB	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T4	IIIA	IIIA	IIIB	IIIC
M1a	IVA	IVA	IVA	IVA
M1b	IVA	IVA	IVA	IVA
M1c	IVB	IVB	IVB	IVB

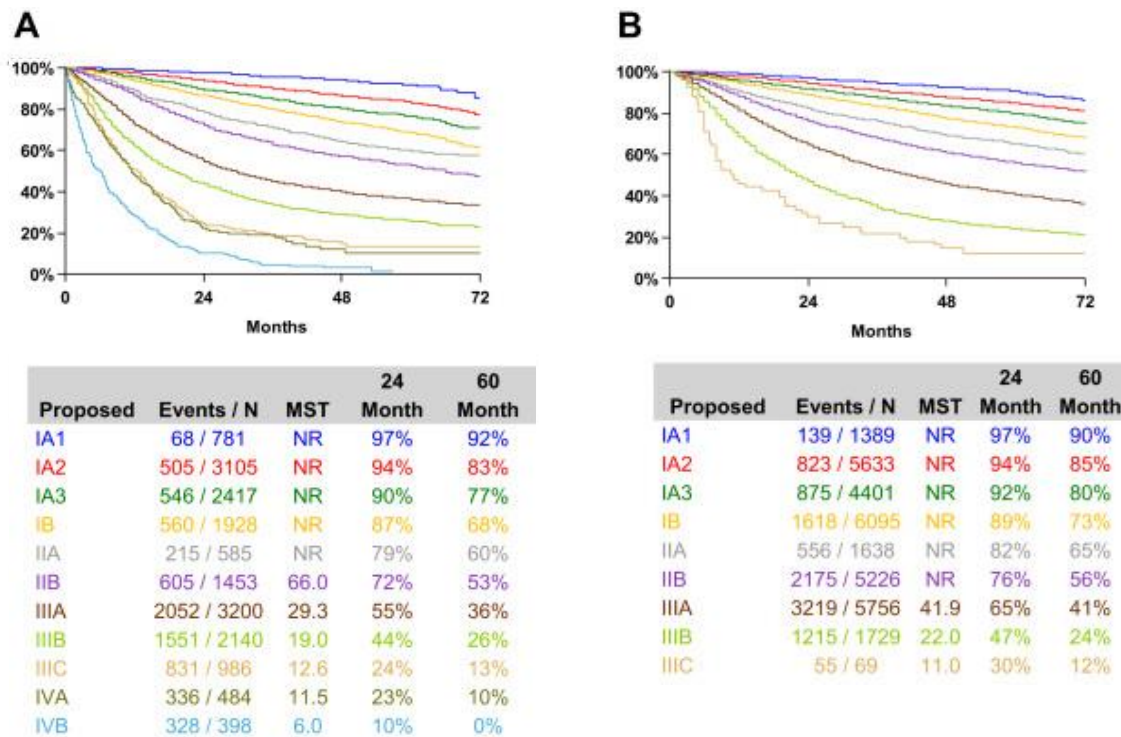


Figure 3. Prognostic information regarding TNM classification. Overall survival expressed as median survival time of NSCLC patients, classified by a) clinical stage and b) pathological state, according to the eighth edition of the TNM classification. Adapted from (24).

2.4. HISTOLOGICAL CLASSIFICATION

To date, the most commonly used classification for lung cancer is based on histological and pathological techniques. Approximately a 15% of tumors present a small cell component and are classified as small cell lung carcinomas (SCLC), whereas the other 85% of pulmonary tumors are known as non-small cell lung carcinoma (NSCLC). In addition, NSCLC is divided into three main histological subtypes: adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC).

ADC is the most common of the histologic subtypes of lung cancer, accounting for more than 40% of all lung cancer cases. It originates in broncho-alveolar cells and is usually found in the peripheral areas of the lung, being more likely to be surgically resected. Tumor cells tend to present translucent cytoplasm often vacuolated and peripheral nucleus with a prominent nucleolus (25). If the tumor is well differentiated, commonly named low grade, it resembles the normal glandular structure with acinar, papillary, micropapillary, lepidic or solid patterns. In contrast, poorly differentiated or

high grade ADC have lost the glandular morphology and immunohistochemical (IHC) or mucin staining is required for diagnosis (26).

SCC accounts for approximately 30% of all lung carcinomas and is closely linked to tobacco smoking and more common in men than in women. It usually occurs in the central portion of the lung or in one of the main airway branches and can form cavities in the lung if they grow to a large size as a consequence of central necrosis. Keratin pearls and intercellular bridges are typical histological features in well-differentiated SCC. In these tumors, cells usually have a dense cytoplasm with sharply defined cell borders and a central hyperchromatic nucleus without prominent nucleolus (25). In contrast, minimal residual squamous-cell features are found in undifferentiated ones.

LCC accounts for around 10% of lung cancers. This subtype is formed by tumors that do not present the cytologic, architectural and immunohistochemical features to be included in the other subtypes. However, the increasingly improved classification of lung carcinomas has triggered a huge decrease in the classification of pulmonary tumors as LCC over the past three decades. In general, this subtype is originated by undifferentiated lung epithelial cells and is usually found in the periphery of the lung. Tumors are usually more necrotic and cells exhibit a large cytoplasm, a big vesicular nucleus and clearly visible nucleolus. In addition, they are often organized in sheets or nests and the clinical evolution is aggressive due to their rapid growth and their capacity to generate metastasis (26).

Regarding SCLC, the last edition of the WHO classification grouped this tumor type with neuroendocrine tumors together with large cell neuroendocrine carcinoma and carcinoids. SCLC cells are small in size with scant cytoplasm, small nucleus with round, oval or spindle shape, granular chromatin, and absent or small nucleolus. Cells borders are rarely seen and nuclear molding is usual (25). It is highly associated to tobacco smoking and is one of the fastest growing and spreading tumors with a very high mitotic rate and frequent extensive necrosis. Most patients relapse within a year of their initial treatment, making SCLC one of the most aggressive type of tumors with very poor prognosis (27). However, due to the low frequency of SCLC, this study will focus on NSCLC.

2.5. MOLECULAR CLASSIFICATION

Over last decades, though, substantial advances have been made in the understanding of the molecular and cellular mechanisms driving cancer initiation, maintenance, and progression (28). NSCLC is one of the most genomically diverse tumors, and therefore, there are a variety of molecularly defined subsets of patients characterized by specific sets of driver mutations, such as *EGFR*, *KRAS* or *ALK* in ADC, among others (Figure 4).

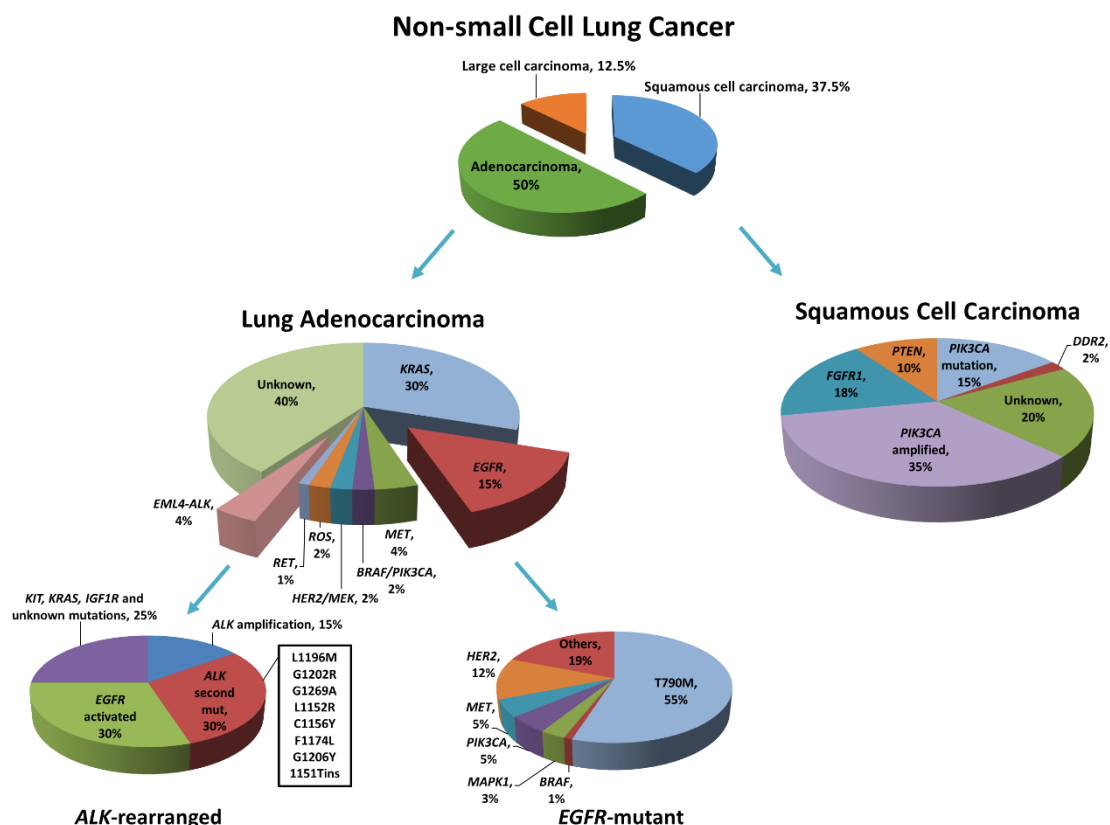


Figure 4. Evolution of NSCLC subtyping from histological to molecular classification. *ALK*, anaplastic lymphoma kinase; *BRAF*, B-Raf proto-oncogene; *DDR2*, discoidin domain receptor tyrosine kinase 2; *EGFR*, epidermal growth factor receptor; *EML4*, echinoderm microtubule-associated protein-like 4; *FGFR1*, fibroblast growth factor receptor 1; *HER2*, human epidermal growth factor receptor 2; *IGF1R*, insulin-like growth factor 1 receptor; *KIT*, proto-oncogene c-Kit; *KRAS*, Kirsten rat sarcoma viral oncogene homolog; *MAPK1*, mitogen-activated protein kinase 1; *MEK*, mitogen-activated protein kinase kinase; *MET*, mesenchymal epithelial transition factor proto-oncogene, receptor tyrosine kinase; *PIK3CA*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *PTEN*, phosphatase and tensin homolog; *RET*, ret proto-oncogene; *ROS1*, reactive oxygen species proto-oncogene 1, receptor tyrosine kinase.

EGFR (epidermal growth factor receptor) alterations are implicated in many types of cancer, including lung cancer. Overexpression or aberrant activation of this gene is more common in never-smokers and female populations and also in ADC patients, where is found in about 15% of patients (29,30). This gene encodes a transmembrane tyrosine kinase with an extracellular binding domain and an intracellular component including a tyrosine-kinase (TK) domain. Binding to its ligand leads to receptor homo- or heterodimerization with other members of the *EGFR* family and activation of the tyrosine kinase domain (31). In NSCLC, *EGFR* mutations have been observed in the first four exons of the intracellular tyrosine kinase domain, mainly in frame deletions in exon 19. The next commonest *EGFR* mutations are missense mutations, particularly L858R, a single nucleotide point mutation in exon 21 that leads to a single amino acid change from leucine to arginine at codon 858. Less common mutations, including in-frame duplications or insertions in exon 20 (representing 5-10% of cases), have also been identified (32). All these mutations, as will be discussed below, are known to confer sensitivity or resistance to *EGFR* specific inhibitors. For instance, patients harboring tumors with T790M mutation in exon 20 are known to be resistant to first and second-generation *EGFR* tyrosine-kinase inhibitors (TKIs), although they respond to the third-generation inhibitor Osimertinib. In contrast, insertions in exon 20 have been found to confer resistance to currently approved *EGFR* TKIs (33).

KRAS (Kirsten rat sarcoma viral oncogene homolog) is part of the *RAS* family of proto-oncogenes, which in humans comprises *KRAS*, *NRAS*, and *HRAS* and encodes a G-protein with a critical role in controlling signal transduction pathways. These pathways regulate cell proliferation, differentiation, and survival, playing a critical role in downstream signal transduction induced by several growth factor receptors, including *EGFR*. The most common mutations are found in codons 12, 13 and 61, being G12C and G12V the most frequent ones (34). *KRAS*-activating mutations are the most recurrent oncogenic alterations identified in lung ADCs and are considered non-druggable targets, being associated with poor response to conventional and targeted therapies. As a result, therapeutic strategies focused on the inhibition of *KRAS* downstream pathways are being studied for the treatment of patients with mutations in this gene (34).

ALK (anaplastic lymphoma kinase) rearrangements, which result in fusions of the intracellular kinase domain with the amino terminal end of *EML4* (echinoderm microtubule-associated protein-like 4), are also very important in NSCLC. The rearrangement occurs in chromosome 2p (intron 13 of *EML4* is fused to intron 19 of *ALK*) and causes a constitutive oligomerization that produces a mitogenic signal and therefore a malignant transformation (35). More recently, different partner genes have been identified in a small subset of *ALK* rearrangements (less than 1% of cases) including *KIF5B* (kinesin family member 5b), *TFG* (TRK-fused gene) and *KLC-1* (kinesin light chain 1). *ALK* rearrangements have been observed in around 4% of ADC NSCLC patients, predominantly in never-smokers (36). *ALK* inhibition with the TKI crizotinib produces profound responses. However, drug resistance eventually develops, and there is evidence for the formation of secondary *ALK* point mutations and *EGFR* signaling activation implicated in some cases (37).

In addition to the most frequent alterations described above, less common alterations have also been identified in ADC tumors (28). Among them, *ROS1* (reactive oxygen species proto-oncogene 1, receptor tyrosine kinase), *RET* (*ret proto-oncogene*) and *NTRK1* translocations, *BRAF* (B-Raf proto-oncogene) alterations, *HER2* (human epidermal growth factor receptor 2) and *MEK1* (mitogen-activated protein kinase kinase 1) mutations, as well as *MET* (mesenchymal epithelial transition factor proto-oncogene, receptor tyrosine kinase) mutations and amplifications are potential driver mutations present in lower percentages in patients (38–40).

On the other side, SCC tumors are characterized by a high overall mutational load and genomic complexity, probably due to its high association with tobacco smoking. To date, only a few clinical trials are ongoing, and diagnostic methods still need to be standardized (41). In fact, 80% of ADC patients exhibit a known driver mutation, whereas only 45% of patients with SCC present identified mutations, being *PI3K* the most frequent one (42). Interestingly, mutations found in *EGFR* in SCC tumors differ from those seen in ADC patients. Deletions in exon 19 and L858R mutation in exon 21 are absent, whereas amplifications and L861Q mutation have been identified (43). Other important targets include fibroblast growth factor receptor 1 (*FGFR1*), which is amplified

in 18% of the cases, phosphatase and tensin homolog (*PTEN*) which is mutated in 10% of SCC (all of them smokers), discoidin domain receptor tyrosine kinase 2 (*DDR2*), present in 2% of SCC patients and *AKT* and *MET* mutations.

2.6. TREATMENT

Lung cancer treatment essentially depends on pathological classification, tumor stage, and performance status (PS). Surgery is still the standard treatment for 20-30% of diagnosed NSCLC patients who are in early-stages (stages I, II and some IIIA) and have a good PS. Alternatively, radiofrequency ablation (RFA) or stereotactic body radiotherapy (SBRT) are options for patients in stage I or II and a borderline medical indication for surgery (44). Adjuvant chemotherapy or radiotherapy can be recommended in resected patients, since it seems to improve local control (45).

Regarding unresectable NSCLC, patients' management has evolved substantially over the last 15 years. Specific anti-target therapies have emerged, increasing patients' survival and decreasing the toxicity that can be produced by conventional chemotherapy (Figure 5). Good examples are the TKIs gefitinib, erlotinib, afatinib, dacomitinib and osimertinib, which are specific treatments for metastatic NSCLC patients bearing mutations or deletions in *EGFR* (46). However, acquired resistance to EGFR TKIs has become a great issue. The main mechanisms leading to resistance to these therapies are *EGFR*-T790M mutation in exon 20, as well as *MET* or *HER2* amplifications, or *PI3KCA* mutations, among others(47–50). Osimertinib is the only *EGFR* TKI approved for the treatment of patients with acquired T790M mutations previously treated with other EGFR inhibitor. Unfortunately, resistance mutations to this third-generation TKI have also been identified, including *EGFR* C797S mutation, loss of *EGFR* T790M, amplification of *MET* or *HER2* and mutations in *KRAS*, *PIK3CA*, and *HER2* (51–53).

Another example of personalized treatment that has improved objective responses and survival in combination with chemotherapy in patients with a non-SCC histology is bevacizumab, a recombinant, humanized, monoclonal antibody that selectively inhibits vascular endothelial growth factor (*VEGF*), preventing tumor angiogenesis, growth and metastasis (54).

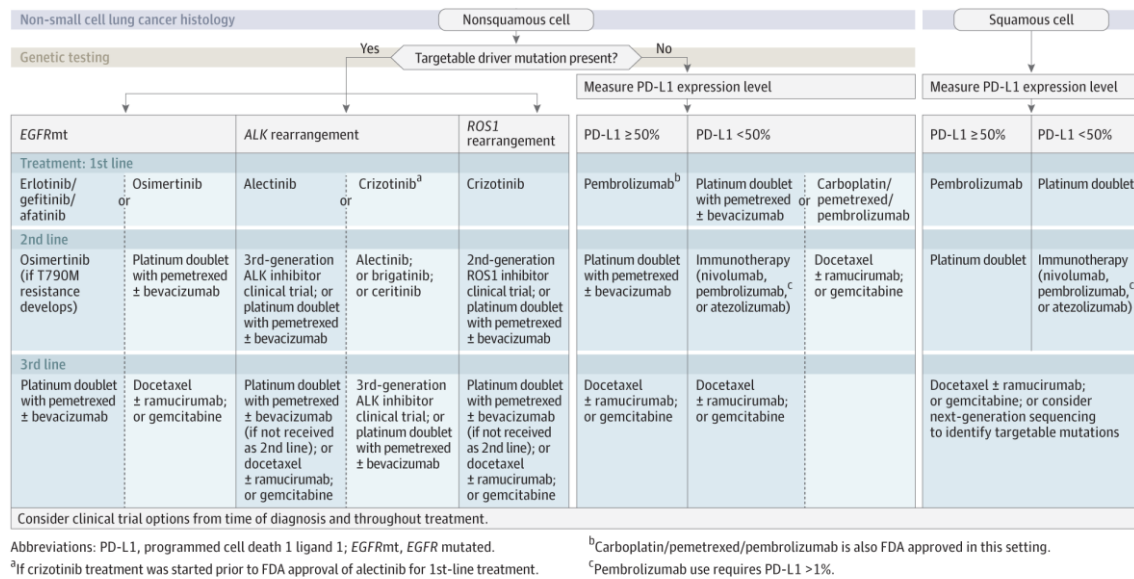


Figure 5. Therapeutic management of advanced NSCLC patients (55).

For patients harboring metastatic NSCLC with *ALK* or *ROS1* rearrangements, crizotinib, has been approved as first-line targeted therapy. Like happens with *EGFR* inhibitors, resistance to crizotinib are developed through different secondary mutations in *ALK* or via alternative tyrosine kinase activation (56–59). To solve this problem, lorlatinib has been recently designed for the treatment of patient with metastatic NSCLC *ALK* fusions previously treat with one or more *ALK* inhibitors and many trials are undergoing combining this antibody with other agents (60). For *ROS1* rearranged patients, the most common mechanism of resistance is *ROS1* G2032R mutation (61). As for patients with *ALK* rearrangements, lorlatinib has shown potential activity against *ROS1* G2032R mutation in preclinical models and the second-generation ceritinib, brigatinib and entrectinib have also shown to be effective (62–64). Moreover, the combination of a BRAF inhibitor, dabrafenib, and a MEK inhibitor, trametinib, has been recently approved for the treatment of advanced NSCLC patients harboring *BRAF* V600E mutation (46).

Besides the major advances in the discovery of driver molecular alterations that can be targeted with specific drugs, cancer immunotherapy has set a new standard for the treatment of NSCLC as well. It refers to a number of approaches intended to activate the immune system in order to induce objective responses and disease stabilization. Among the different possible options, blocking immune checkpoint molecules with monoclonal antibodies has emerged as a promising strategy. Pembrolizumab, a

monoclonal antibody that blocks the interaction between the immune checkpoint molecule programmed cell death 1 (PD1) and its ligand (PD-L1), has already got approval as a single agent for the first-line treatment of patients with metastatic NSCLC harboring high expression of PD-L1 (more than 50%) in the absence of *EGFR* mutations or *ALK* or *ROS1* fusions (65). In addition to its first-line indication, pembrolizumab was also approved, along with nivolumab and atezolizumab, for the treatment of metastatic NSCLC after progression to platinum-based chemotherapy or targeted therapies (65). Like pembrolizumab, nivolumab targets PD-1, whereas atezolizumab binds to PD-L1. However, while nivolumab and atezolizumab are used irrespective of PD-L1 tumor expression, pembrolizumab is used if PD-L1 expression is at least 1% (66).

Despite these advances, treatment resistance is still the main cause of death in lung cancer. In that sense, it is known that cancer cells are heterogeneous, causing that not all the subclones in the tumor are affected by treatments in the same way. Certainly, there is strong evidence pointing out that treatment resistance is highly associated to populations of tumor cells with stem-like properties, named cancer stem-like cells (CSCs), which are able to survive using different mechanisms that will be addressed in the next section.

3. CANCER STEM CELLS

3.1. TUMOR HETEROGENEITY

Genetic and phenotypic variations are observed between tumors from different tissues and cell types, as well as between individuals with the same tumor type, which is known as inter-tumor heterogeneity. In addition, cancers evolve over time in every particular patient in terms of clonal structure, genotype and phenotype (67). As a result, genetic and phenotypic diversity exists not only between tumors, but also within populations of cells in a single tumor (intra-tumor heterogeneity), making difficult definitive diagnostics and obstructing therapeutic decision-making (Figure 6).

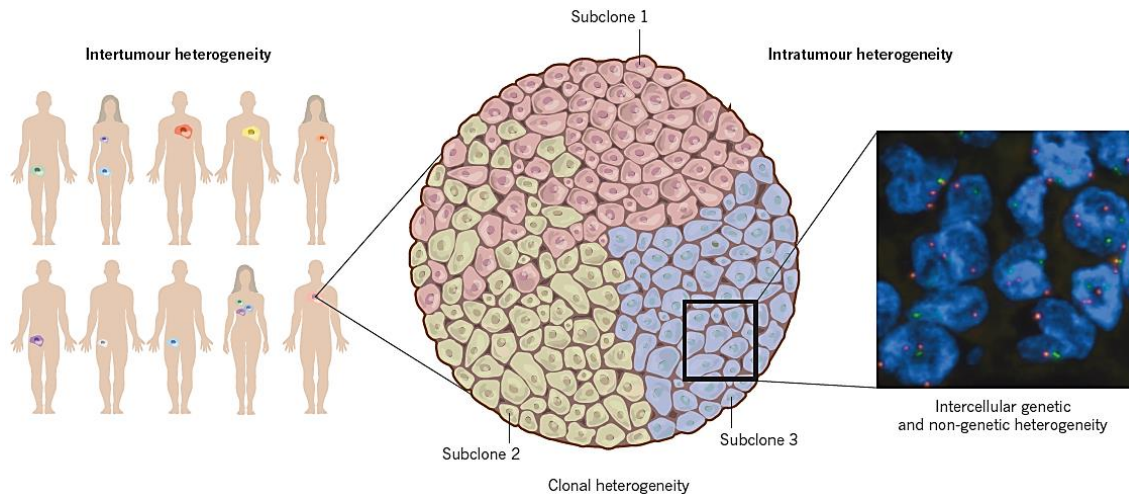


Figure 6. Inter-tumor and intra-tumor heterogeneity. Genetic and phenotypic variation are observed between tumors of different patients, tissues and cell types, but also between clones in a single tumor. Subclones may intermingle (as shown by subclones 1 and 2) or be spatially separated (as shown by subclone 3). Separation between subclones could reflect physical barriers such as blood vessels or micro-environmental changes. Tumor subclones may show differential gene expression due to both genetic and epigenetic heterogeneity. Within a subclonal population of tumor cells — shown here as a tumor section, hybridized to two fluorescent probes for the centromeres of two chromosomes (chromosome 2, red; chromosome 18, green) with DNA (blue) — there is intercellular genetic and non-genetic variation of, for example, chromosome copy number, somatic point mutations or epigenetic modifications that results in phenotypic diversity. Reproduced from (68).

First of all, spatial phenotypic heterogeneity could trigger that a biopsy did not supply an adequate picture of the whole tumor. Secondly, decision-making based on scoring the dominant phenotype in a given specimen might be biased if they do not account for minor subpopulations with clinically and biologically important distinguishing features (68). For these reasons, the study of the tumor evolution is important as well as finding good markers to properly define tumor populations. Currently, there are two major frameworks to explain cancer cell heterogeneity: the clonal evolution model and the cancer stem cell model (69).

The clonal evolution model, also known as the stochastic model, was proposed by Nowell in 1976 and relies on the concept that neoplasms arise from a single cell of origin, and tumor progression results from acquired genetic variability within the original clone (70). As a result, the genetic and epigenetic changes that occur over time in individual cancer cells can confer a selective advantage in a Darwinian-like way, allowing individual clones to generate other clones and leading to genetic heterogeneity and phenotypic and functional differences among the cancer cells within a single patient. In this model, the frequency of cancer cells with tumorigenic potential is high,

the tumor organization is not necessarily hierarchical and the rational approach to therapy has been to target most or all cells (71,72).

The cancer stem cell model, also known as the deterministic model, proposes that the growth and progression of many cancers are driven by small subpopulations of stem-like cells with self-renewal and differentiation properties, named cancer stem cells (CSCs). It was developed in the late 1990s, when researchers began to address the possible relationship between hematopoietic stem cells and human leukemias via transplantation experiments of hematological tumor cells into immunodeficient mice (73,74). The CSC concept was coined in 1997, when Bonnet and Dick demonstrated that human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell (75). Since then, CSCs have been identified in several solid tumors, including brain (76), breast (77), lung (78), colon (79) and pancreas (80). In this model, the frequency of cancer cells with tumorigenic potential varies from rare to moderate, the tumor organization is always hierarchical and the therapy approach enables to target only tumorigenic cells.

Even though these two models were considered mutually exclusive at first (81), nowadays clonal evolution and CSC models are proposed as a unified model by some authors (82). In the integrated model, the gaining of favorable mutations can result in clonal expansion of a founder cell (Figure 7). At some point, another cell may acquire new mutations that permit it to produce a new subclone. Over time, genetic mutations accumulate and subclones evolve concurrently. Here, CSCs are not considered static entities, since they can evolve over lifetime and genetic changes can influence their frequency in each subclone. Some subclones may present a marked hierarchical development, where few self-renewing CSCs exist among a larger number of bulk non-CSCs. Other subclones may contain an intermediate hierarchy, where the number of CSCs is relatively high and some other subclones may have genetic alterations that confer high-renewal potential, where most cells are tumorigenic (83).

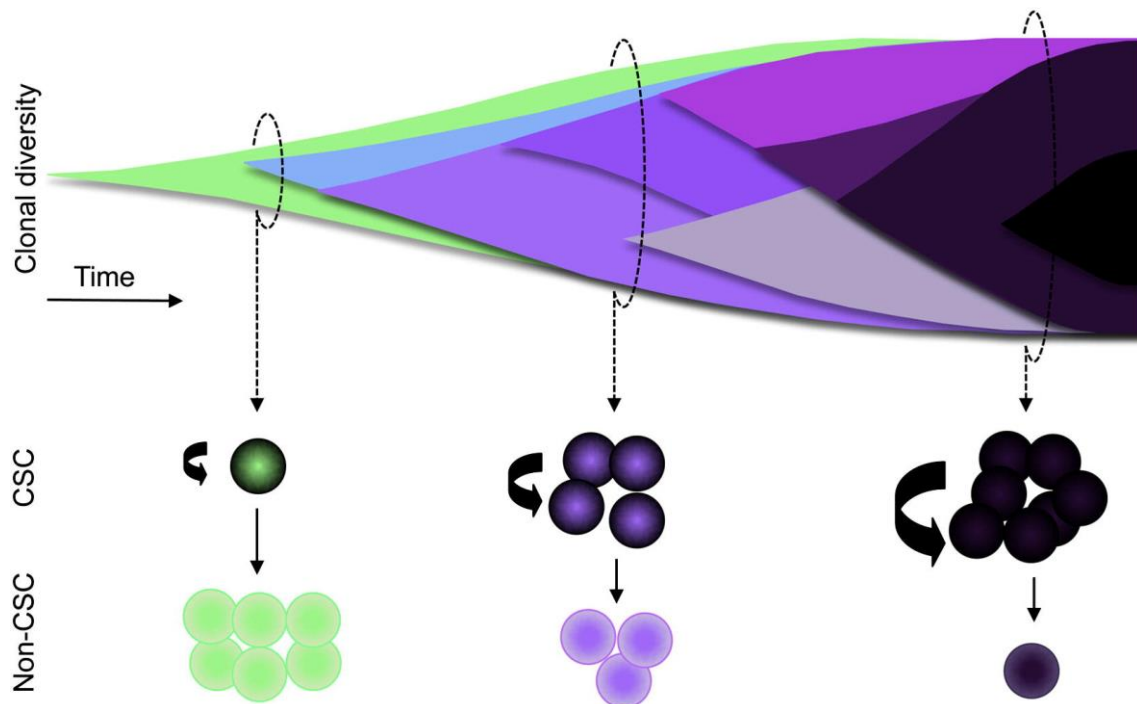


Figure 7. Unified model of Clonal Evolution and Cancer Stem Cells (CSCs). Top panel shows how clonal evolution resulted in the parallel expansion of tumor cells that have acquired favorable mutations creating different subclones in the tumor. Bottom panel shows CSCs evolving over time with different frequencies in different subclones. Reproduced from (83).

Even if the origin of CSCs remains unknown, there are some theories about their cellular precursors (84). It was hypothesized that they arise from non-tumorigenic stem cells when a mutation that disturbs their ability to control cell division is generated. Stem cells are pluripotent and show self-renewal ability, so that CSCs could just use aberrantly stem cell pathways to support their self-renewal. Alternatively, it was suggested that partially differentiated precursor cells, which are more frequent in adult tissue, could experience a few mutations that trigger their transformation to CSCs. CSCs could also arise from differentiated cells that have undergone a de-differentiation process because of oncogenic mutations (85).

3.2. PROPERTIES

Although aberrantly regulated, CSCs share most of the mechanisms governing other stem cell populations. The main difference between them relies on their tumorigenic activity, given that CSCs are able to form tumors when transplanted into animals, acting like tumor initiating cells (TICs), but the rest of stem cells are not. The

main features which are characteristic of CSCs include:

- A) Self-renewal and asymmetric division. CSCs are able to generate daughter CSCs and high proliferative bulk cancer populations by asymmetric cell division, which is defined as any division that gives rise to two sister cells that have different fates—a feature that can be recognized by differences in size, morphology, gene expression pattern, or the number of subsequent cell divisions undergone by the two daughter cells (86). CSCs have been described to undergo this type of division, maintaining their pluripotency by asymmetrically segregating their gene products into differentiating daughter cells (87). Since CSCs have lost balance in networks regulating proto-oncogenes (promoting self-renewal), gate-keeping tumor suppressors (limiting self-renewal), and care-taking tumor suppressors (maintaining genomic integrity), they have the potential to proliferate indefinitely with minimal niche support. Soft agar formation and limiting dilution assays are commonly used to determine the frequency of self-renewing cells in a cell culture since CSCs can be serially transplanted through multiple generations (88).
- B) Cell quiescence. CSCs are known for an aberrant regulation of cell cycling, which permits them to display a quiescent slow-cycling phenotype tightly associated to treatment resistance and tumor dormancy. Traditional chemotherapy regimens target proliferating cells, potentially missing their effect on slower dividing CSCs. This tumor cell population is able to enter into a reversible G₀ phase from which cells may escape to reenter the cell cycle in response to physiological cell stimuli, like cell death after treatment (89). In this way, cell quiescence is not just a passive state but rather a condition actively maintained and regulated by signaling pathways allowing rapid activation of quiescent cells and reentry in the cell cycle (90).
- C) Overexpression of multidrug efflux transporters. Protection of the stem cell population from damage or death is critical because these cells need to remain intact throughout the life of an organism. Ironically, it has been shown that efflux pumps also afford higher protection to CSCs, shielding them from the adverse effects of therapeutic agents (91). ATP binding cassette (ABC) transporters, including ABCG2, ABCB1 or ABCC1, to name but a few, are known to be associated with drug resistance

in CSCs. A subset of CSCs that have a high capability for effluxing antimitotic drugs can be isolated by their capacity to efflux fluorescent dye Hoechst 33342 or rhodamine 123 with the help of a flow cytometer. This particular population is called side population (SP) because during flow-cytometry analysis these cells can be visualized as a negatively stained population off to “the side” of the main cell population. The current understanding is that the drug transporting capability of these cells is likely conferred by certain ABC transporters including ABCB1 (rhodamine 123) and ABCG2 (Hoechst 33342) (92).

- D) Anchorage-independent growth ability. Under anchorage-free conditions, the majority of cell types undergo anoikis, a specific type of cell death provoked by loss of cell adhesion. Nevertheless, CSCs can grow independently of a solid surface (88). It appears that cells that escape anoikis, represented mainly by stem cells and possibly early progenitor cells, synthesize higher levels of growth factors and extracellular matrix receptors, creating an *in vitro* niche that supports their survival and proliferation in suspension being less dependent on cell-matrix and cell-cell interactions for survival (93). In addition, CSC have been reported to overexpress antiapoptotic molecules, including BCL2 and BCLXL, which act as negative regulators of mitochondrial membrane permeabilization and cytochrome C release, survivin, belonging to the inhibitors of apoptosis (IAP) family members and lower levels of caspase 8 associated with tumor necrosis factor- α (TNF- α) related apoptosis-inducing ligand (TRAIL) resistance (90).
- E) Expression of signaling pathways and molecules governing stem-cell fate. The expression of signaling pathways related to self-renewal, proliferation and differentiation by stem cell populations is essential for development and tissue homeostasis (94). However, dysregulation and aberrant activation of these highly conserved pathways result in the formation of tumorigenic CSCs (95). The most studied and characterized pathways responsible for CSCs formation an maintenance will be reviewed later.

3.3. *IN VITRO* CULTURE

Since stem cells display anchorage-independent growth ability, sphere-forming assays have become the gold standard method for their isolation and enrichment. These assays consist in culturing cells under non-adherent conditions using serum-free medium supplemented with minimum growing requirements using fresh tumor tissue or commercial cell lines as starting material. Cells directly isolated from surgical resections specimens have been proven to be better models for tumor characterization than cell lines, but establishment of primary cultures is problematic and time-consuming, mainly because of the frequent lack of cell viability, the excessive necrosis of some tumor samples and the proliferation of non-tumorigenic cell types in cultures. As a result, most studies to date have been performed in commercial cell lines.

Sphere-forming assays were first used to culture cells from the adult brain, obtaining stem-like cells as free-floating spheres, called neurospheres (96). Since then, these culture conditions have been widely used to evaluate self-renewal and differentiation at the single-cell level *in vitro*, assuming a relevant role to isolate CSCs due to the lack of unique CSCs markers and the absence of distinctive morphological phenotypes (97). Several studies have reported the identification of cells with stem cell properties in cell lines and primary cultures established from resectable NSCLC and SCLC patients (98–105). However, most publications display substantial variations in different aspects of sphere-forming assays, including changes in medium composition and volume, cell density, and surface area of the culture dish (106,107). This diversity in procedures has promoted differing and even opposing results, ranging from those who argue that formation of tumorspheres with increased stemness can be done without the addition of any external mitogenic stimulation (108) to those who reported that tumorspheres cultivation requires exogenous mitogens supplementation, including epidermal growth factor (EGF), basic fibroblast growth factor (bFGF), insulin-transferrin-sodium selenite, B27 or hydrocortisone (109,110). Moreover, high variability in kinetics of formation of tumorspheres between cell lines and primary cultures has also been detected, making even more difficult to know if reported variations are due to changes in culture conditions or due to the source of the CSC population. Another key parameter

of sphere-forming assays is cell density, given that it plays a critical role on clonality. Initially, these assays were conceived to obtain clonal spheres formed from one single cell. However, tumorspheres are highly dynamic entities and have been observed to frequently aggregate and fuse, even at low densities, so that true clonality can only be assured by plating single cells per well. Nonetheless, single cell plating has also a negative impact on cell viability due to the lack of autocrine/paracrine signals released by cells into the medium and can be extremely challenging in some cases (111–113). In addition, sphere-forming assays allow determining *ex vivo* proliferation, but they cannot evaluate the ability of CSCs to initiate nor propagate tumors.

To overcome these limitations, the usage of alternative approaches, like transplantation assays in animal models, is widely extended. However, animal models are costly, lack immune system, and contain limitations regarding controllability, reproducibility and flexibility of design, which have motivated extensive research on developing 3D cell culture models using scaffolds and scaffold-free techniques that better mimicked the *in vivo* tumor microenvironment (114–116). There are several processing techniques to design and fabricate 3D porous scaffolds able to provide large internal volume and 3D space where cells can aggregate and form tissue-like structures. Electrospinning is a versatile polymer processing method commonly used to fabricate microfiber structures. These fibers form a non-woven fabric that highly mimics the fibrous network of native extracellular matrix with fiber diameters that is close to extracellular matrix fibrils (117). In addition, microfluidic devices have brought a great degree of versatility to the field of cancer research. They consist on a chamber with manageable submillimeter channels where cell location, fluid flow, mechanical cues and gradients can be controlled. As a result, complex tumor microenvironment can be created, recreating metastatic microenvironments, immune-cancer cell interactions, and specific behavior of cancer cells. However, although microfluidic systems seem promising, few studies have been conducted using them to date and further investigation to standardize parameters are still required (118,119).

3.4. MARKERS

Over the last decade, the evaluation of the expression profiles of cancer cells with stem properties in different solid tumors has allowed the identification of several biomarkers, pathways and therapeutic targets against CSCs (120,121). The main markers used for identifying and isolating CSCs include surface cell-adhesion molecules, such as CD133 or CD44, cytoprotective enzymes, like aldehyde dehydrogenases (ALDH), and transcription factors and drug-efflux pumps, such as ABC drug transporters or multidrug resistance transporters (MDR) (122). However, CSCs markers in a determined organ or tissue are not completely shared with those markers that work in other, being few of them common between organs and tissues (see Table 3). Regarding lung cancer, several markers for CSCs have been proposed, but there are conflicting data reported between studies in cell lines. Molecules suggested as potential lung CSCs markers include CD133, CD90, EpCAM, CD44, ALCAM, BMI1, ALDH1A1, and ABCG2, and are further discussed below (123–125).

Table 3. Compilation of proposed CSC markers for distinct solid tumor types.

Tumor type	Putative markers	References
Breast	ALDH1, CD24, CD44, CD90, CD133, ITGA6	(120,126)
Colon	ABCB5, ALDH1, CD24, CD26, CD29, CD44, CD133, CD166, CXCR4	(127–129)
Glioma	CD15, CD36, CD90, CD133, ITGA6	(130,131)
Liver	CD13, CD24, CD44, CD90, CD133	(132,133)
Lung	ABCG2, ALDH1, BMI, CD90, CD133, CD166, CD326, CD44	(134–137)
Melanoma	ABCB5, ABCG2, ALDH1, CD20, CD133	(138–140)
Ovarian	CD24, CD44, CD117, CD133	(141,142)
Pancreatic	ABCG2, ALDH1, CD24, CD44, CD133, CXCR4	(143–145)
Prostate	ALDH1, CD44, CD133, CD166, ITGA2, ITGB1, ITGA6, TROP2	(146,147)

CD133 (also known as PROM1) is a pentaspan transmembrane glycoprotein overexpressed in both humans and mice tumors. Some evidence has suggested that CD133⁺ CSCs display strong resistance to chemo-, radio- and immunotherapy. Several studies showed that CD133⁺ stem-like cells survived standard chemotherapeutic treatment with oxaliplatin and 5-fluorouracil (5-FU), and the downregulation of CD133 using short hairpin RNAs has been associated with slower cell growth, reduced cell motility and decreased ability to form spheroids and metastasize (148,149). Consequently, CD133 has been considered a potential CSC marker and an important

therapeutic target for many CD133-expressing cancer types. In lung cancer, there are a number of publications that focused on CD133 as a potential CSC marker (134,150–152). However, some authors reported that the use of CD133 expression to discriminate lung CSC is overstated (153). For example, some CD133⁻ lung cancer cells also possess the ability to self-renew and generate the formation of xenograft when transplanted into recipient mice (154).

CD90 (or THY1) is a highly conserved glycoposphatidylinositol (GPI)-anchored cell surface glycoprotein that is used as a marker for a variety of stem cells and axonal processes of mature neurons. CD90 is involved in cell-cell and cell-matrix interaction and its overexpression has been identified in CSCs from several tumors, including glioblastoma, breast and liver malignances (155). Cell lines experiments suggest CD90 as a promising CSC marker for lung cancer. *In vitro* and *in vivo* data showed that the sphere formation ability and the tumorigenic capacity of CD90⁺ cells from A549 and NCI-H446 cell lines is superior in number and volume than that for CD90⁻ cells (135). Previous investigation associated CD90 overexpression to CSC-like phenotype in A549, NCI-H2122, NCI-H226 and NCI-H1437 lung cancer cell lines (156). However, few data are available about CD90 expression in lung cancer patients' samples. Flow cytometry analysis suggested the co-expression of CD90, CD326 and CD133 as a signature with prognostic value in patients with NSCLC (157).

EpCAM (also known as CD326) is a transmembrane glycoprotein that plays a role in balancing cell proliferation and differentiation. In healthy tissue, high EpCAM levels are associated with proliferation during morphogenesis, tissue regeneration and stem cell maintenance (158). Elevated levels of EpCAM expression have been found to promote tumor progression and therefore, it has been explored as a prognostic and diagnostic marker and as a therapeutic target. EpCAM overexpression has been found in CSCs from different solid tumors and was associated with poor prognosis and aggressiveness of tumorus (159,160). Karimi-Busheri et al. proposed the overexpression of *EPCAM1* and *ALDH1A1* as a signature of enriched CSCs in NCI-H460 non-small cell lung cancer (NSCLC) cell line (136) and *in vivo* and *in vitro* experiments suggested that CD133⁺/ EpCAM⁺ and CD34⁺/ EpCAM⁺ subpopulations represent CSCs in lung primary

tumors and cell lines (161).

CD44 is a cell-surface glycoprotein that acts as a receptor for many extracellular matrix components, including acid hyaluronic, collagen, integrins as well as matrix metalloproteinases (MMPs). CD44 molecule integrates signals from microenvironment, growth factors and cytokines, transducing them to cytoskeletal proteins and nucleus, and regulating the expression of genes involved in cell-matrix adhesion, cell migration, proliferation, differentiation, and survival (162,163). CD44 was proposed as a CSC marker with prognosis value in various types of tumors, including colorectal (164), breast (165) and gastric (166) cancers. Regarding NSCLC, disparate data has been published. CD44 gene overexpression was significantly correlated to tumor differentiation, histological type, clinical stage, lymph node metastasis and poor prognosis (167,168). On the contrary, no significant association between CD44 and clinical outcome was found by an immunohistochemical staining analysis of 159 patients with resected NSCLC (169), and flow cytometry characterization of A549 concluded that CD44 cannot be considered a marker for isolating lung CSCs in that cell line (150).

ALCAM (or CD166) is a highly preserved transmembrane protein that belongs to the immunoglobulin superfamily. CD166 is expressed in several tissues, including neural, epithelial and hematopoietic stem cells and has been proposed as a CSC marker in colorectal (170), head and neck (171) and prostate (172) cancers. Regarding NSCLC cancer, contradictory information about CD166 involvement in CSCs biology has been reported. Some authors have pointed out that CD166⁺ cells from NSCLC cell lines and patients display stem cell-like features (137,173). However, others observed no growth change when knocking-down CD166 in NSCLC cell lines *in vitro* nor *in vivo* (174). In patients' tissue, Ishiguro and colleagues found no significant association with clinical or pathologic factors in an immunohistochemical analysis of 147 NSCLC patients, but described CD166 expression as an independent prognostic factor for overall survival (OS) (174). In contrast, an inverse association between CD166 expression and tumor size and lymph node status without prognostic effect on patients' survival was observed in a large cohort of NSCLC patients (n=1910) using immunohistochemistry on a tissue microarray basis (175).

BMI1 is a catalytic member of epigenetic repressor polycomb group proteins who regulates several cellular processes such as cell cycle progression, senescence, aging, DNA damage repair, apoptosis, angiogenesis, and self-renewal of stem cells (176). BMI1 was identified as a CSC-marker in head and neck (177), breast (178) and bladder cancers (179). In lung cancer, few data have been published to date. RTqPCR and immunohistochemistry analysis of 57 samples from resected NSCLC patients revealed a strong correlation between BMI1 and ALDH1A1 and CD133 (125). Moreover, recent studies have associated *BMI1* expression to chemotherapeutic resistance via direct transcriptional targeting of ABC drug transporters in lung (137) and breast cancers (178), whereas analysis of 60 tissue samples from NSCLC patients using RTqPCR associated higher *BMI1* expression with advanced tumor stage and more extent of lymph nodes metastasis, suggesting that BMI1 plays a relevant role in cancer cells invasion capability (180).

Aldehyde dehydrogenases (ALDH) are a group of NAD(P)⁺-dependent enzymes that catalyze the oxidization of aldehydes into carboxylic acids. It is thought that ALDH can act as drug-detoxifying enzymes and be responsible for therapeutic resistance (181). ALDH1A1 is an isozyme of ALDH superfamily that is believed not only to act as a marker for stem cells, including CSCs, but also to play important functional roles in self-protection, differentiation and expansion. The evaluation of ALDH1A1 activity determined that its expression is useful for isolating CSCs from various malignancies, including breast (182) and colorectal cancers (183), but its capacity to act as a potential CSC marker in some other malignancies, such as ovarian cancer, is questionable (184). Studies on NSCLC patient-derived cells and cell lines reported that isolated ALDH⁺ cells were highly tumorigenic, clonogenic as well as capable of self-renewal compared with their ALDH⁻ counterparts (185,186). On the prognostic impact of this marker, interesting data have been reported. A recent meta-analysis including 1836 patients concluded that higher *ALDH1* expression is connected with decreased OS and disease free survival (DFS) and a worse prognosis (187). Complementarily, You and colleagues investigated the prognostic value of each ALDH1 isozyme in 1926 NSCLC patients through the Kaplan-Meier plotter database, finding that high expression of ALDH1A1 correlates to better OS whereas high expression of ALDH1A2 and ALDH1B1 correlates to worse OS (188).

ABC drug transporters are overexpressed in both stem cells and CSCs, acting as efflux pumps to protect them from xenobiotic toxins. ABCG2, also known as breast cancer resistance protein (BCRP), is one of the most important members of ABC transporter family in cancer because it can be used to determine the SP phenotype, an important population in CSCs studies which shows many features of CSCs with regard to self-renewal, lineage capacity and tumorigenicity (95). ABCG2 is considered a potential marker of CSCs in many cancer types, including breast (189) and liver (190), acting as a mechanism in multidrug resistance. However, there are controversial data about ABCG2 in lung cancer. Elevated expression of ABCG2 was observed in lung cancer SP cells, which were resistant to several chemotherapeutic drugs, more tumorigenic than non-SP cells, and had self-renewal capacity (191,192). Tang et al. established drug-resistant H460 and A549 cell lines by repeated exposure to cisplatin and found a significant increase in ABCG2 expression. They established stable ABCG2-overexpressing and shRNA-ABCG2-knockdown cell lines, observing that overexpression of ABCG2 significantly increased the proportion of symmetric division, whereas knockdown of ABCG2 increased the proportion of asymmetric division (193). However, previous immunohistochemistry analysis in tissue from 133 early-stage NSCLC patients showed no association of CD133 and ABCG2 with prognosis (194). Furthermore, Liang et al. detected in NSCLC patients using immunohistochemistry that, in addition to the membranous form, ABCG2 proteins are also found inside the nucleus, increasing *CDH1* (E-Cadherin) expression. Additional mice xenografts of A549 cells showed that those cells with lower expression levels of ABCG2 were more likely to metastasize, suggesting that ABCG2 may play a dual role in cancer metastasis (195).

3.5. SIGNALING PATHWAYS

As discussed above, the major difference between stem cells and CSCs relies on their ability to regulate self-renewal and differentiation pathways. In non-tumorigenic stem cell populations, stem-related pathways, such as Notch, Wnt/ β -catenin, Hedgehog, JAK/STAT, TGF- β or Hippo, are tightly controlled with intact genetics or epigenetics. However, in CSCs, deregulation of these pathways along with improper interactions between them represent key events for cell propagation and pathogenesis (196). Among

these signaling cascades, abnormal activity of Notch, Wnt and Hedgehog pathways are the most studied and correlated to the tumorigenicity of CSCs, making these developmental pathways important therapeutic targets for blockade of CSC self-renewal and proliferation, and tumor progression (94).

3.5.1. Notch signaling pathway

Notch pathway is an evolutionarily conserved signaling cascade that constitutes a critical component in the molecular circuits that regulate a broad range of events during embryonic and post-natal development, including border formation, cell fate decisions, differentiation, migration, proliferation and apoptosis (197,198). The role of Notch in human malignancies has been highlighted recently by the presence of gain-of-function mutations and amplifications of Notch genes in different types of cancer, and by the evidence that genes in the Notch cascade are potential therapeutic targets (199,200).

The core components of the Notch pathway comprises four transmembrane receptors—from Notch1 to Notch4—and five structurally similar ligands: Delta-Like (DLL) -1, -3, -4, and Jagged-1, -2, although there is very little evidence that Delta-like3 physically binds to the Notch receptors or that it truly functions as a Notch ligand (Figure 8) (201). The Notch signaling cascade is initiated by ligand-receptor interaction between two neighboring cells resulting in two successive proteolytic events as part of the activation mechanism (202). The first cleavage is mediated by a metalloprotease of the ADAM family (TACE, tumor necrosis factor- α -converting enzyme) in close proximity to the extracellular side of the plasma membrane. The released extracellular domain is then transendocytosed by the ligand-expressing cell. The second one occurs within the transmembrane domain, mediated by a multi-protein complex, called γ -secretase, consisting of presenilin, nicastrin, APH1 and PEN2, which leads to the release of the Notch intracellular domain (NICD). Upon cleavage, NICD translocates to the nucleus, forming a complex with the ubiquitously expressed transcription factor CBF1. The translocation of NICD is counteracted by Numb, through a mechanism that is not completely understood (203). In the absence of NICD, CBF1 is a transcriptional repressor due to its association with co-repressors. When NICD associates with CBF1, a number of

co-activators are recruited, including mastermind-like (MAML) -1, -2 and -3, resulting in a multiprotein complex which acts as a potent transcriptional activator. The most well-defined targets of the NICD-CBF1 complex are the hairy enhancer of split (HES) family, the Hes-related repressor protein (HERP, also called HEY) family, cell cycle regulators, such as *CDKN1A* and *CCND1* and apoptosis regulators (204,205).

The first link between Notch and human tumors was made in the late 1980s in a small number of patients suffering from T cell acute lymphoblastic leukemia (206). More recently, deregulated expression of members of Notch signaling pathway has also been reported in many solid tumors, including breast (207) and lung cancers (208). Notch pathway function has been seen to be context dependent, since different Notch receptors or ligands could induce diverse gene expression programs, explaining the heterogeneous, and even opposite, outcomes observed in this signaling pathway for different solid tumors. There are four major pleiotropic effects that Notch pathway plays and are relevant during tumorigenesis. A) Gate-keeper function: Notch maintains stem cells in an undifferentiated state. In the intestine for example, Notch prevents crypt progenitor cells from differentiating. B) Binary cell fate decisions: in the lymphoid system, it specifies the T cell lineage at the expense of the B cell lineage from a bi-potent early thymocyte progenitor. C) Induction of differentiation: in the skin, Notch induces terminal differentiation events, and during thymocyte differentiation, *NOTCH1* promotes differentiation of pro-T-cells into pre-T-cells. D) Tumorigenesis: overexpression of Notch within hematopoietic bone marrow cells or in T cell progenitors results in T cell leukemias and as such, Notch functions as an oncogene. However, in the skin Notch functions as a tumor repressor since loss of Notch signaling results in the development of basal cell carcinoma-like tumors (209).

Notch pathway expression is not only tissue-dependent, but cell-dependent as well. While in SCLC Notch signaling is not active, in NSCLC it is, possibly due to loss of Numb inhibitor expression or to the presence of gain-of-function mutations in Notch receptors (203), leading to high expression levels of Notch target genes and making this tumor type susceptible to therapies based on Notch inhibition (210). The involvement of Notch on lung cancer was experimentally proved in a transgenic mouse model by

inducing the specific expression of activated Notch. The mice developed alveolar hyperplasia as early as 7 days after *NOTCH1* induction and, when crossed with mice conditionally overexpressing *MYC* in the alveolar epithelium, mice developed adenocarcinomas (211).

Furthermore, using a model of lung adenocarcinoma with *KRAS* mutation and *NOTCH1* deletion, it was found that *NOTCH1* function was required for tumor initiation via suppression of p53-mediated apoptosis (212). Other studies have shown that under hypoxic conditions, *NOTCH1* stimulated NSCLC tumor growth through direct upregulation of IGF1-R and survivin, both of which enhanced cell proliferation and survival (213,214). Additionally, the overexpression of this gene has been significantly correlated with disease progression, metastasis and poorer prognosis of NSCLC patients (215). These data support a strong and direct role of Notch signaling in NSCLC initiation and proliferation. In CSCs from NSCLC, Hassan et al. reported the decisive role of Notch signaling for sphere formation and self-renewal *in vitro* and for tumor initiation and tumor heterogeneity formation *in vivo* (216). Interestingly, molecular analyses defined a subpopulation of CD24⁺ ITGB4⁺ Notch^{hi} cells that were capable of propagating tumor growth in both clonogenic assays and in serial orthotopic transplantation assays (217). In addition, upregulation of *NOTCH2* has been linked to progression of early-stage lung adenocarcinoma and aggressive phenotype at advanced stages (218) and targeting of Notch signaling pathway with Tarextumab (a *NOTCH2/NOTCH3* antagonist) inhibited tumor growth and decreases TICs frequency in patient-derived xenografts (219). Similarly, Notch target *HES1* has also been seen to enhance CSCs phenotype, promoting cell proliferation and migration by activating *BMI1* (220).

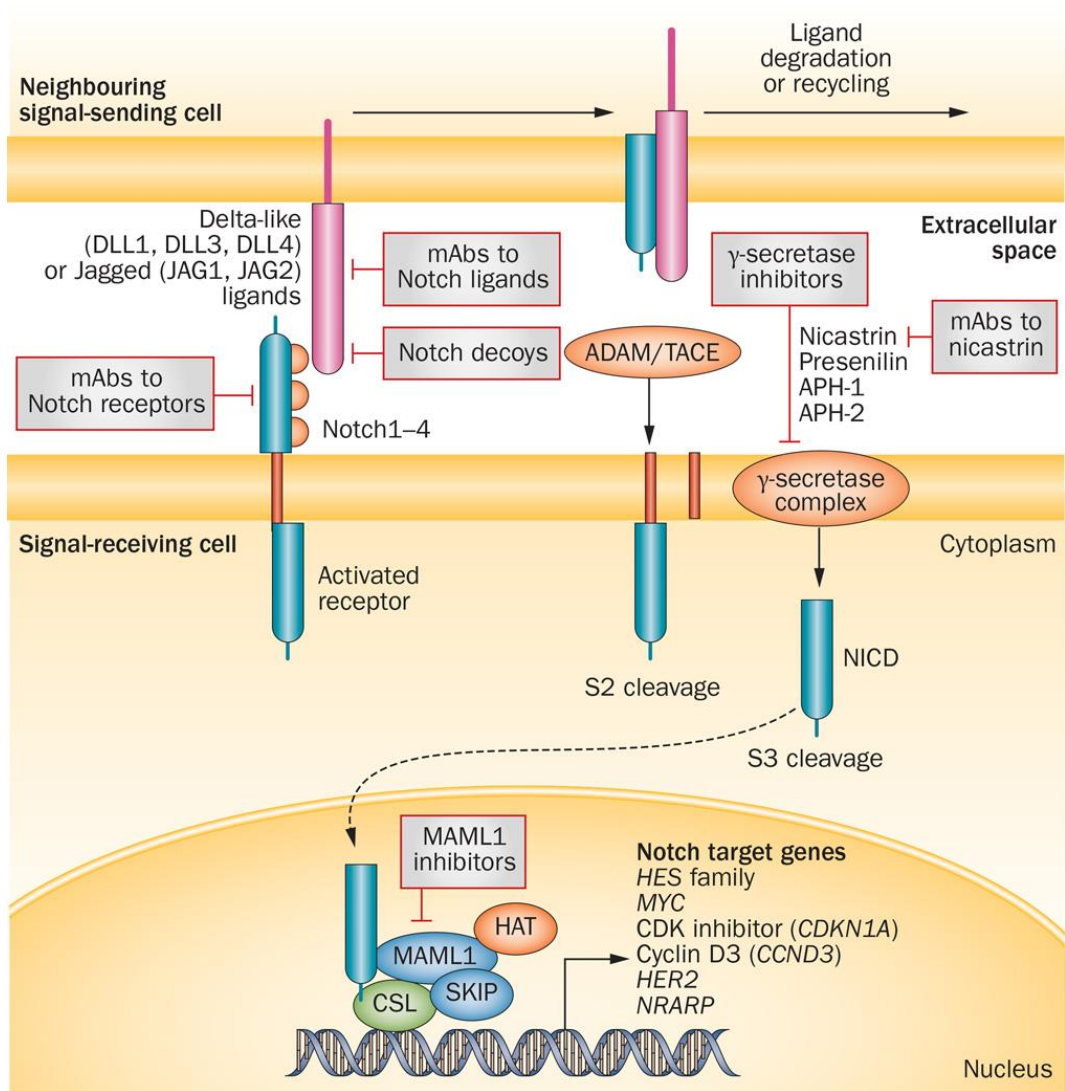


Figure 8. The canonical Notch signaling pathway. DLL1, DLL3 and DLL4, and Jagged ligands (JAG1 and JAG2) expressed on the cell surface can induce signaling in adjacent cells expressing their cognate receptors Notch1–4. Ligand binding promotes sequential cleavage of the Notch receptors by ADAM/TACE enzymes (S2 cleavage) and then γ -secretase (S3 cleavage), resulting in release the NICD, which interacts with transcriptional regulators in the nucleus to instigate a Notch gene-expression profile. Notch target genes, in turn, regulate pivotal cell-fate choices, including differentiation, cell-cycle progression and survival. The final phenotypic effect is dependent on the specific signaling context, paralogue, ligand and dosage. Under many conditions, and in several types of cancer stem-like cells, Notch signaling can delay differentiation, and maintain proliferative and survival potential. Abbreviations: ADAM, a disintegrin and metalloproteinase; APH-1/2, anterior pharynx-defective-1/2; CSL, CBF1/Su(H)/Lag-1; DLL, delta-like ligand; HAT, histone acetyltransferase; HES, hairy and enhancer of split-1; JAG1, Jagged-1; JAG2, Jagged-2; mAb, monoclonal antibody; MAML1, Mastermind-like 1; NICD, Notch intracellular domain; NRARP, Notch-regulated ankyrin-repeated protein; SKIP, ski-interacting protein; TACE, TNF- α -converting enzyme (also known as ADAM17). Reproduced from (94).

3.5.2. Wnt signaling pathway

The evolutionarily conserved Wingless-type protein (Wnt) signaling pathway plays an important role in controlling a number of embryonic development processes and the maintenance of tissue homeostasis in adults by regulating proliferation, differentiation, migration and polarity, survival, genomic stability and self-renewal of stem cells (221). Not surprisingly, aberrant Wnt signaling underlies a wide range of diseases, including cancer (222), fibrosis (223) and neurodegenerative disorders (224).

The Wnt signaling network is extremely complex. Firstly, there are 19 Wnt ligands, which are glycoproteins of 40 kDA in size that contain lipid modifications with many conserved cysteines, and more than 15 receptors and co-receptors distributed over seven protein families in mammals (225). Moreover, Wnt proteins can trigger a variety of responses, often gathered at two groups: the canonical Wnt signaling pathway, for the classical Wnt-induced activation of β -catenin-TCF (T-cell factor) transcriptional complexes, and the non-canonical Wnt signaling pathway, which includes the planar cell polarity (PCP) signaling pathway (226), the Wnt/ Ca^{++} flux pathway and the protein kinase A pathway (227,228) and cJun N-terminal kinase (JNK) and small GTPase Rho, Rac and Cdc 42 signaling networks (229,230). Moreover, crosstalk from various non-Wnt factors has also been reported to modulate nuclear β -catenin accumulation (231).

In the absence of Wnt proteins, canonical Wnt signaling is inhibited due to a β -catenin degradation complex consisting of Axis inhibition protein (AXIN), adenomatous polyposis coli (APC), casein kinase 1 α (CK1 α) and glycogen synthase kinase 3 β (GSK-3 β) (Figure 9). This complex phosphorylates β -catenin, resulting in its ubiquitination and proteosomal destruction, making β -catenin unavailable to interact with other factors (232). If Wnt -1, -2 or -3 are present, porcupine contributes to its secretion, making them available to interact with members of the Frizzled (FZD) family of receptors. Binding of Wnt to FZD results in the formation of a stable receptor complex between Wnt, FZD, lipoprotein receptor-related protein 5/6 (LRP), Disheveled (Dvl) and the degradation complex. This new complex phosphorylates Dvl, inactivating GSK-3 β as a result, and reducing the proteolytic destruction of β -catenin. Consequently, cytoplasmic levels of β -

catenin increase, from where it can move to the nucleus to bind to transcription factors of the TCF/lymphoid enhancer-binding factor (LEF1) family. The basic transcription machinery as well as transcriptional coactivators are then recruited, including the cAMP response element-binding protein (CREB)-binding protein/E1A binding protein p300 (CBP/p300) and Pygopus 2, which triggers the transcription of a number of factors that promote cell proliferation and treatment resistance, including cyclin D1 and c-Myc. In addition, β -catenin has been seen to interact with E-cadherin at the cytoplasmic membrane, promoting cellular adhesion. Therefore, even though the influence of risen β -catenin on transcription may enhance cell growing and chemoresistance, the interaction of β -catenin with E-cadherin has been proposed to reduce malignant characteristics, such as metastasis, by decreasing cell migration in favor of cell adhesion (233).

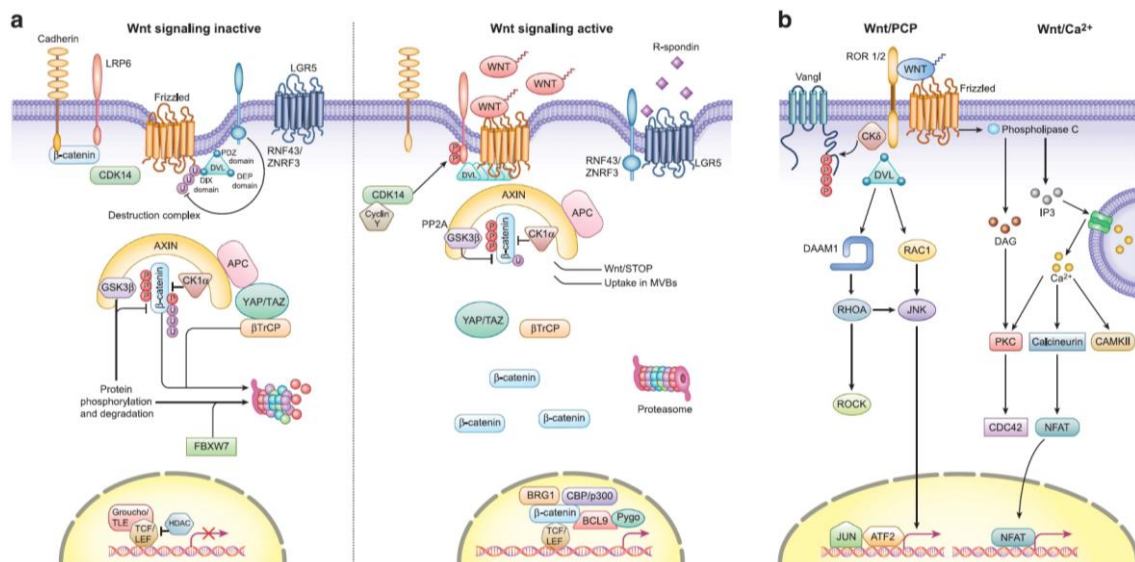


Figure 9. The canonical and non-canonical Wnt signaling pathway. (a) In canonical Wnt signaling, absence of Wnt ligands (left) leads to phosphorylation of β -catenin by the destruction complex. β -catenin is phosphorylated, ubiquitinated and targeted for proteosomal degradation. In the absence of nuclear β -catenin, a repressive complex recruits HDACs to repress target genes. The canonical pathway is activated upon binding of secreted Wnt ligands to Fzd receptors and LRP co-receptors (right). LRP receptors are then phosphorylated which recruit Dvl proteins. Dvl polymers inactivate the destruction complex and this results in accumulation of β -catenin which then translocates into the nucleus. There, β -catenin forms an active complex with LEF and TCF by displacing TLE/Groucho complexes and recruitment of histone modifying co-activators. (b) Non-canonical Wnt signaling is defined by β -catenin-independent mechanisms. During Wnt/PCP signaling, Wnt ligands bind to the ROR-Frizzled receptor complex to recruit and activate Dvl. Dvl binds to the small GTPase Rho, leading to rearrangements of the cytoskeleton and/or transcriptional responses. Next to Dvl, Vangl is activated by phosphorylation in a Wnt5a-dependent manner. Wnt/Ca²⁺ signaling is initiated by G-protein triggered phospholipase C activity leading to intracellular calcium fluxes and downstream calcium dependent cytoskeletal and/or transcriptional responses. Reproduced from (234).

In addition to the inhibition of the canonical Wnt pathway by the β -catenin degradation complex, this cascade can be inhibited or enhanced in many different ways. On the one hand, frizzled-related proteins (sFRPs) compete with Wnt for binding to FZD, Wnt inhibitory factor 1 (WIF-1) and Cerberus arrest secreted Wnt, Disabled 2 (Dab2) and the Dickkopf (Dkk) family inhibit Wnt signaling by binding to the component of the Wnt receptor complex, LRP5/6, and the human homolog of Dapper (HDPR1) and Idax antagonist Dvl. On the other hand, tankyrases-1 and -2 stimulate Wnt signaling by destabilizing AXIN and EGFR increases signaling by phosphorylating and inactivating GSK-3 β (235,236). Given the importance of Wnt signaling for adult stem cell biology, it is not surprising that Wnt pathway mutations are frequently observed in cancer. A role for Wnt pathway in cancer was first described in the 1980s and 1990s in mouse models of mammary cancer and in human and mouse colon cancer. Researchers found that the induction of an aberrant overexpression of *WNT1* using a proviral insertion at its locus or by transgenesis generated breast tumors in mice (237). Other studies suggested a critical role of *WNT-CTNNB1* (gene encoding for β -catenin), signaling in colorectal cancer (238,239). Germline inactivating mutations in *APC* gene, which is a negative regulator of *CTNNB1*, were found in patients with a hereditary cancer syndrome termed familial adenomatous polyposis (FAP), which can progress to colorectal carcinomas following concomitant activating mutations in *KRAS* and inactivating mutations in *TP53* (240). Both *APC* gene and *CTNNB1* are often mutated in colorectal cancers of non-FAP patients, and overexpression of constitutively active *CTNNB1* or loss of APC function can result in colorectal tumorigenesis (222).

Regarding lung cancer, there is strong evidence pointing to a decisive role of Wnt pathway in the development of NSCLC (241). Most of NSCLC cell lines have active Wnt signaling and, in murine models, activation of Wnt signaling is associated with increasing tumor initiation potential (242,243). Downregulation of Wnt signaling by anti-Wnt-1 monoclonal antibody or small interfering RNA (siRNA) induced apoptosis, inhibited tumor cells proliferation (244), blocked xenografts and reduced cell motility and invasion (245). Additionally, Li et al. showed that *in vitro* overexpression of *WNT3* by NSCLC cells correlated to metastasis, cell invasion, anchorage-independent growth, EMT-like morphological changes and F-actin reorganization (246). Wnt pathway activation and

overexpression of Wnt-1, -2 or -3 are also associated with poor prognosis. In a cohort of 262 resected NSCLCs patients, 37% stained positively for Wnt ligands by immunohistochemistry, which was associated with a low apoptotic index, aberrant β -catenin expression, increased expression of c-Myc, cyclin D1, VEGF-A, MMP-7, Ki-67, survivin and bigger intratumoral microvessel density (247). In addition, Nakashima et al. found that WNT3 promotes tumor progression in a study including 128 resected NSCLC patients (248). Moreover, it has been reported that the Wnt signaling pathway helps to maintain CSCs population in lung cancer. Increased expression of *CTNNB1* has been associated with the overexpression of putative stem cells markers, such as *CD44*, *EPCAM*, *OCT4* and *CCND1*, and resistance to a number of chemotherapeutic drugs in sorted lung CSCs (249). Furthermore, the inhibition of β -catenin enhanced the anticancer effect of irreversible EGFR-TKI in *EGFR* T790M-mutated NSCLC cell lines (250).

The role of GSK-3 β in β -catenin degradation is clear in the absence of Wnt proteins. Nevertheless, when Wnt proteins are present, GSK-3 β functions remain unclear. Recent results suggest a role of GSK-3 β in cell survival and migration. Tivantinib has been found to target GSK-3 β and that pharmacological inhibition caused apoptosis in NSCLC cells (251). Additionally, in a lung xenograft model, astrocyte elevated gene-1 (AEG-1) behaved as a critical protein in the activation of EMT by directly targeting GSK-3 β (252). Some other inhibitors of the canonical Wnt pathway seem to have a dual function in cancer. Recombinants *DKK1* and *WNT5A* inhibited and increased, respectively, mice mammospheres formation (253). Differential expression of *DKK1* has been found among cancer cells from 98 NSCLC patients and cell lines, and it was the overexpression of *DKK1* that promoted migratory and invasive activity, suggesting an oncogenic role of *DKK1* in lung cancer (254,255). Then again, loss of *WNT5A* in hepatocellular carcinoma has been associated with poor prognosis (256), but overexpression has also been correlated with unfavorable prognosis and angiogenesis promotion in NSCLC patients (257,258). Therefore, the Wnt pathway seems to be a promising target in NSCLC tumorigenesis, but it is still necessary to explore some components of this complex network in order to consider targeting Wnt pathway as a therapeutical option for NSCLC patients.

3.5.3. Hedgehog signaling pathway

The hedgehog (Hh) signaling pathway is an important component on the regulation of stem cells properties during embryonic development and in adult tissues. During embryogenesis, it controls proliferation and differentiation on a time and position-dependent manner and it plays a central role in tissue repair and regeneration in adult tissues. Mutations and deregulations of genes related to Hh pathway have been reported in some solid tumors, contributing to the onset of cancer and accelerating the rate of tumor growth (259).

Mammalian Hh signaling pathway is mainly constituted by three Hh ligands homologues with different spatial and temporal distribution patterns: sonic hedgehog (SHH), indian hedgehog (IHH) and desert hedgehog (DHH), two transmembrane receptors: patched homolog 1 and 2 (PTCH1, -2), a G protein-coupled receptor called smoothened (SMO), and a cytoplasmic complex that regulates the glioma-associated oncogene homolog (GLI) family (Figure 10). GLI1 is a transcription activator, and GLI2 and GLI3 are both activators and repressors of transcription. The Hh signaling cascade is initiated by Hh ligands binding to the PTCH1 protein on the target cell. In the absence of Hh ligands, PTCH1 represses the activity of SMO, preventing its localization to the cell surface from intracellular endosomes, where SMO is predominantly located. Under these circumstances, different kinases phosphorylate and activate repressor forms of GLI transcription factors. The active form of GLI is prevented from transactivating Hh-responsive genes by the serine-threonine protein kinase suppressor of fused (SUFU) and the atypical kinesin-like protein Costa (COS) in a manner that is still not completely understood. Upon binding of the Hh ligand, PTCH1 is internalized, and apparently destabilized, so that it can no longer transport the endogenous agonist molecules outwards. This allows them to accumulate intracellularly and activate SMO, which sequester COS and SUFU, releasing the GLI transcription factors to exert their effects in nucleus. KIF3A and β -arrestin are required for SMO activation (260,261).

The first connection between aberrant Hh signaling and cancer was the discovery of a mutation in the transmembrane receptor *PTCH1* that causes a rare condition, named Gorlin syndrome (262). Gorlin syndrome patients suffer from various basal cell

carcinomas throughout their lifetimes and are predisposed towards other types of cancer. Additionally, increased Hh signaling has been reported to be involved in a third of all human medulloblastoma cases, frequently due to *PTCH1* and *SUFU* mutations. In all these cases, it is believed that deregulated Hh signaling leads to increased cell proliferation and tumor formation (263,264). Many malignancies have been linked to aberrant Hh signaling in CSCs, including oral and esophageal cancers (265). Regarding lung tissue, Hh signaling pathway is probably inactive in all cells of the human adult lung epithelium except for the progenitor cells. The persistence of Hh signaling in the epithelial progenitor cells may perhaps facilitate these cells maintenance and play a decisive role in tissue response to injuries in the airway epithelia. Studies carried out on mice models of lung airway epithelial regeneration showed that persistent injuries in the airway are a potent stimulus for the activation of the Hh signaling, which promotes epithelial progenitor cells expansion and attenuates the proliferative expansion of the pulmonary mesenchyme, maintaining tissue homeostasis (266,267). As happened for Notch pathway, Hh signaling expression seems to differ between SCLC and NSCLC. Activation of Hh pathway has been clearly reported on SCLC cell lines and tumors (268,269), in contrast to NSCLC, where the role of Hh pathways remains unclear (270). Recent research on NSCLC lines concluded that the blockade of Hh signaling increases sensitivity to EGFR-TKIs (271). On the contrary, immunohistochemistry analysis of clinical samples of human lung cancer samples demonstrated that 50% of SCLC expressed both *SHH* and *GLI1* compared to only the 10% found in NSCLC (272). In the present context of developing targeted therapies, the Hh signaling modulators offer a potential new opportunity in the management of lung cancer, but further investigation of this pathway is still required.

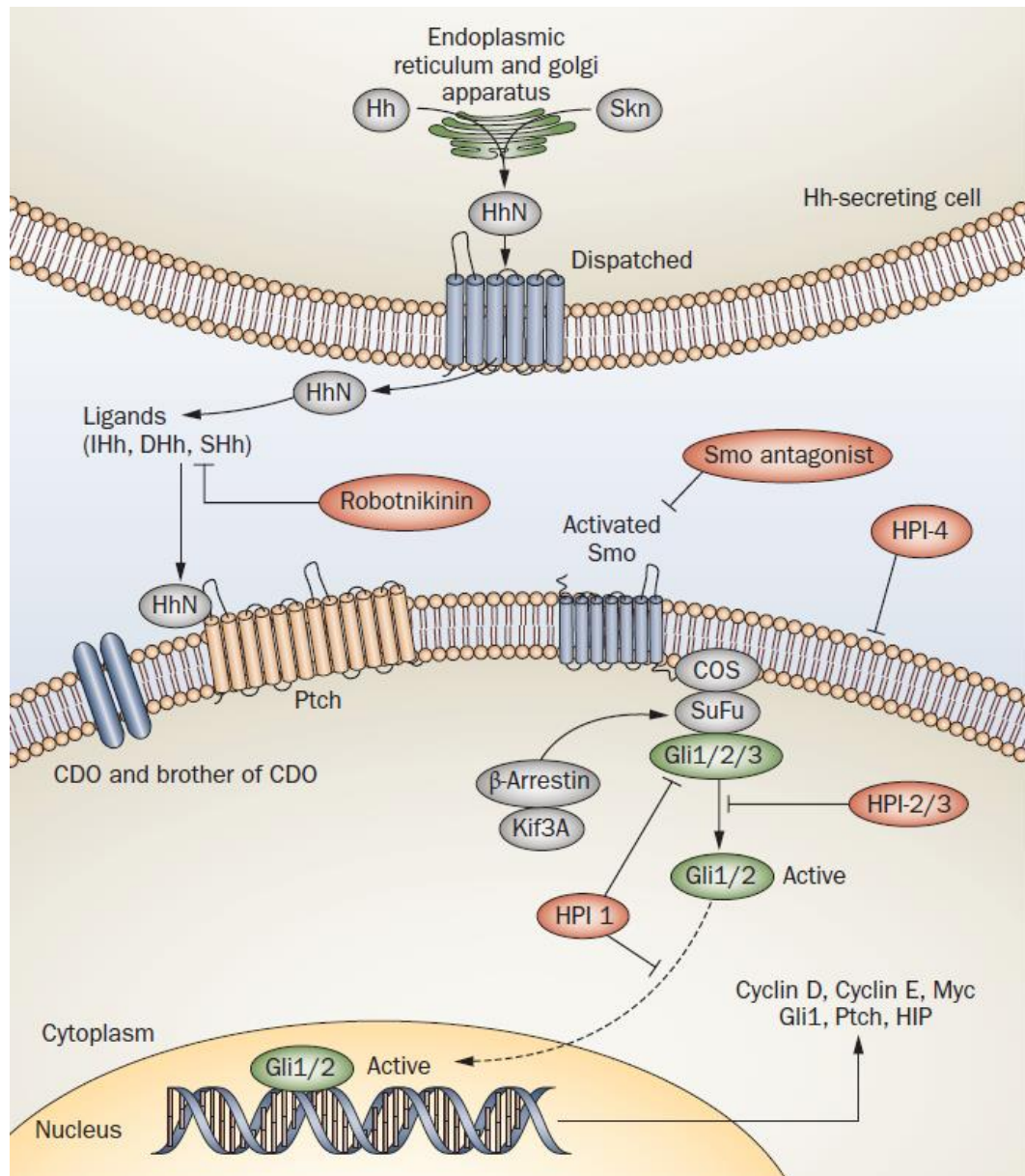


Figure 10. The canonical HH-signaling pathway. The HH-processing pathway involves HHC autocatalysis, and SKN and Dispatched proteins, which mediate the release of HHN ligands (IHH, DHH and SHH). In the absence of HHN binding, PTCH interacts with and inhibits the activity of SMO; HHN binding to PTCH releases its inhibitory effects on SMO, resulting in SMO accumulation and sequestration of COS and SUFU proteins in cilia, which releases the GLI transcription factors to exert their effects in the nucleus. KIF3A and β -arrestin are required for localization of SMO to cilia. GLI1/2 promote a gene-expression pattern relevant to tumorigenesis. Abbreviations: COS, Costal; DHH, Desert hedgehog; HH, Hedgehog; HHC, Hedgehog C-terminal domain; HHN, Hedgehog N-terminal domain; HIP, Hedgehog interacting protein; IHH, Indian hedgehog; Ptch, patched; SHH, Sonic hedgehog; SKN, Skinny hedgehog; SMO, Smoothened; SUFU, suppressor of fused. Reproduced from (273).

3.5.4. Other signaling pathways

In addition to Notch, Wnt and Hh pathways, a number of signaling pathways with different levels of evidence have been related to CSCs in lung cancer and other solid tumors, showing the complex signaling network that involves CSCs. Proteins from Hippo, JAK/STAT, NF- κ B and others signaling pathways that have been frequently related to CSC biology in lung cancer are reviewed below.

The Hippo signaling pathway has emerged as an evolutionarily conserved regulator of diverse cellular processes, including cell survival, proliferation and differentiation. Like, Notch, Wnt and Hh, Hippo plays a fundamental role in tissue homeostasis, organ size, regeneration and its deregulation has been associated with tumorigenesis in several malignancies, including breast (274) or oral (275) cancers. The central components of the Hippo pathway in mammals are well defined, but most of the upstream regulators of this pathway remain unknown. Essentially, it consists of an inhibitory serine/threonine kinase module and a transcriptional module. Silencing of components from the inhibitory module and overexpression of those from the transcriptional module of Hippo pathway has been associated with CSCs phenotype. In lung cancer, transcriptome meta-analysis of 153 samples, including lung ADC, SCC, LCC patients and cell lines, revealed recurrent aberrations in Hippo pathway genes (276). Likewise, different studies have shown the relevant role that Hippo pathway deregulation plays in lung cancer (277,278). On the other hand, JAK/STAT pathway activation has been associated with an increase in the expression of adhesion molecules, which has been related in turn to growth and survival in cancer cells, tumor progression and metastasis. For instance, human prostate CD133^{high}/CD44^{high} CSCs exhibited ICAM-1 and VCAM-1 immunoreaction compared with non-CSCs (279). In addition, STAT3-NF κ B signaling is known to be activated in breast CSCs and inhibition of STAT3 in two breast cancer cell lines reduced CD44⁺ CSC marker cell population (280). In lung cancer, STAT3 is known to function as a transcription factor with target genes that are important for cell proliferation, induction of angiogenesis, prevention of apoptosis, evasion of host immune surveillance and CSCs renewal (281). It was reported that CD133⁺ and ALDH⁺ cells from NSCLC patient samples contain higher levels of the activated form of STAT3

and its inhibition suppressed CSC-properties (282,283). Moreover, knockdown of eIF4E/eIF4GI in NCI-H1299 and NCI-H460 cell lines decreased expression levels of different targets, including NFkB, attenuating cell migration, EMT promoters and enhancing EMT suppressors (284). In the same way, overexpression of CD164 in normal lung epithelial cells (BEAS2B cells) increased CXCR4 expression and activated AKT/mTOR signaling, leading to a malignant transformation of cells *in vitro* with the acquisition of CSC-like properties and tumorigenicity in xenografted mice (285).

Altogether, data obtained so far indicate that CSCs are tightly linked to patient outcome, being an important tumor population to be addressed. Identifying molecular alterations that could act as potential targets and biomarkers against this tumor population could have major implications in NSCLC management. Therefore, it is important to continue characterizing CSCs to better understand their role in treatment resistance, relapse and metastasis associated to the high mortality that is typical of lung cancer

II. OBJECTIVES

Non-small cell lung cancer (NSCLC) is a heterogeneous and etiopathologically complex disease with poor prognosis. In the current context of lung cancer research, the working hypothesis of this thesis is that a minority group of cells, called Cancer Stem Cells (CSCs), are responsible for treatment resistance and recurrences accounting in cancer patients thanks to their abilities and properties. However, few data are available from NSCLC patient samples and, despite their interest, many key questions on CSCs properties, markers and druggability remain elusive. Therefore, the objective of this study is to identify CSCs-associated markers and targets in NSCLC patient specimens and to correlate them with clinicopathological and prognostic variables.

The specific aims of this study are:

1. To establish NSCLC primary cultures from surgical resections under adherent (2D) and non-adherent (3D tumorspheres) conditions.
2. To characterize adherent cultures and tumorspheres.
3. To analyze biomarkers related to CSCs by performing RTqPCR on adherent cultures and tumorspheres from primary cultures and cell lines.
4. To assess expression patterns of proteins encoded by genes that better distinguish both culture conditions.
5. To evaluate the prognostic impact of genes significantly over- and underexpressed in tumorspheres in TCGA.
6. To validate the prognostic role of these genes in an independent cohort of early-stage NSCLC patients.
7. To find new CSCs-related profiles and signatures and to evaluate their possible roles as biomarkers for better NSCLC patient stratification.

III. MATERIALS & METHODS

1. MATERIALS

1.1. PATIENTS AND SAMPLES

This study included a first cohort of 245 patients from *Consorcio Hospital General Universitario de Valencia* who underwent lobectomy or pneumonectomy between 2004 and 2016 and met the eligibility criteria: resected, non-pre-treated stage I to IIIA (according to the 7th edition of the American Joint Committee on Cancer staging manual) with a histological diagnosis of NSCLC. The study was conducted in accordance with the Declaration of Helsinki, and the institutional ethical review board approved the protocol. Lung tumor tissue specimens were obtained from patients at the time of surgery and immediately separated by a pathologist. Tissue samples were preserved in RNALater® (Applied Biosystems) to avoid RNA degradation, and were fresh-frozen at -80° C until further analysis. A pathology report was available for all the samples, enabling their characterization. We included the different NSCLC histology subtypes in this study. The most relevant demographic and clinicopathological characteristics of the cohort are shown in Table 4. The median patient age was 65 years [range: 26-85], 82.4% were male, and 46.5% had ADCs. Moreover, 54.3% of the patients were diagnosed at stage I of the disease and 62.9% presented a PS = 0.

A second cohort of patients from The Cancer Genome Atlas (TCGA) was included as well. This cohort included all patients with histological diagnosis of ADC or SCC. However, patients with post-surgical complication were excluded from analyses and only those patients who had at least 1 month of follow-up were included. The most relevant clinicopathological characteristics of these patients are summarized in Table 5. The median patient age was 68 years, ranging from 38 to 88, 59.8% were male, and 52.2% had SCC. Moreover, 56.7% of the patients were diagnosed at stage I of the disease and only 17.2 never smoked.

Table 4. Clinicopathological characteristics of the patients from *Consortio Hospital General Universitario de Valencia* included in the study.

Characteristics	N	%
Age at surgery (median, range):	65 [26-85]	
Gender		
Male	202	82.4
Female	43	17.6
Stage		
I	133	54.3
II	70	28.6
IIIA	42	17.1
Histology		
SCC	111	45.3
ADC	114	46.5
Others	20	8.2
Performance Status		
0	154	62.9
1-2	91	37.1
Differentiation grade		
Poor	57	23.3
Moderate	96	39.2
Well	46	18.8
NS	46	18.8
Smoking Status		
Current	116	47.3
Former	101	41.2
Never	28	11.4

ADC, adenocarcinoma; SCC, squamous cell carcinoma; NS, not specified.

Table 5. Clinicopathological characteristics of the TCGA patients included in the study.

Characteristics	N	%
Age at surgery (median, range):	68 [38-88]	
Gender		
Male	395	59.8
Female	266	40.2
Stage		
I	375	56.7
II	179	27.1
IIIA	107	16.2
Histology		
SCC	345	52.2
ADC	316	47.8
Exitus		
No	400	60.5
Yes	261	39.5
Relapse		
No	394	59.6
Yes	208	31.5
NS	59	8.9
Smoking Status		
Current	165	25.0
Former	382	57.8
Never	114	17.2

ADC, adenocarcinoma; SCC, squamous cell carcinoma; NS, not specified.

1.1.1. PRIMARY CULTURE ESTABLISHMENT

20 surgical tumor samples from patients from the first cohort were selected for primary culture establishment. Detailed clinicopathological information from each of the 20 patients included in this part of the study is summarized in Table 6. Fresh tissue specimens were washed with cold PBS containing 2% penicillin-streptomycin (Gibco) three times before dissociation using mechanical and enzymatic digestion. Each sample was minced into small pieces using a sterile scalpel and incubated in approximately 7 mL of Advanced DMEM/F12 with 0,001% DNase (Sigma-Aldrich), 1 mg/mL collagenase (Gibco), 1 mg/mL dispase (Gibco), 200 U/mL penicillin and 200 µg/mL streptomycin at 37°C for 3 hours in a water bath with intermittent shaking. After incubation, suspensions were repeatedly pipetted and passed through 70 µm and 40 µm cell strainers (BD Falcon) and centrifuged at 200 g for 5 minutes. Next, cells were resuspended in red blood cell lysis buffer 10x (eBioscience) and incubated at 37°C with intermittent shaking for 20 minutes. After erythrocytes lysis, cells were centrifuged at 130 g for 5 minutes and cell viability was evaluated by trypan blue dye exclusion (Gibco).

Half of cells were transferred to standard collagen-coated flasks and cultured in DMEM/F12 (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco), 200 µg/mL penicillin-streptomycin, and 2 mM of L-glutamine (Gibco). Sphere-forming assays were used for primary CSCs enrichment. For spheres formation, the rest of the disaggregated cells were grown in serum-free DMEM/F12 medium supplemented with 50 µg/mL epidermal growth factor (EGF, Gibco), 20 µg/mL basic fibroblast growth factor (bFGF, Gibco), insulin-transferrin-selenium (ITS PREMIX, Corning), 0.4% Bovine Serum Albumin (BSA, Gibco), 2% B-27 (Gibco), 200 µg/mL penicillin-streptomycin, and 2 mM of L-glutamine. Cells were plated at low density in ultra-low attachment plates (Corning) for 7 to 10 days and expanded by mechanical dissociation of spheroids, followed by replating of both single cells and residual small aggregates in complete fresh medium. In all cases, cells were maintained at 37°C in 5% CO₂ atmosphere and medium was replaced twice a week.

Table 6. Clinicopathological characteristics and lung tumorspheres formation capacity of the patients included in the study.

Patient code	Sex	Age (years)	TNM Stage	Hist.	Smoking status	Relapse or Exitus	DFS (months)	Mutational status	SPH Form.
FIS291	M	66	IIA	SCC	Current	Yes	25.77	KRAS G12C	No
FIS299	M	69	IIIA	SCC	Former	Yes	6.10	TP53 K132E	Yes
FIS301	M	71	IIB	SCC	Former	No	30.77	TP53 D259fs*84	Yes
FIS302	F	74	IIA	ADC	Never	No	33.23	KRAS G12D TP53 E285K	Yes
FIS303	F	57	IB	ADC	Former	Yes	11.13	TP53 R175H	Yes
FIS308	M	72	IIB	SCC	Current	No	27.80	No detected	No
FIS310	M	68	IIIA	ADC	Former	Yes	6.43	TP53 V157F R213SNP	No
FIS312	M	62	IA	ADC	Current	No	24.80	No detected	No
FIS315	F	65	IA	ADC	Never	No	20.93	No detected	Yes
FIS317	M	76	IIB	SCC	Current	Yes	18.40	No detected	Yes
FIS320	M	65	IB	ADC	Current	No	23.60	TP53 P153fs*26	Yes
FIS321	M	83	IB	SCC	Current	No	22.50	TP53 Q156*	No
FIS325	F	67	IB	ADC	Never	No	16.97	EGFR L858R	No
FIS326	F	64	IB	ADC	Former	Yes	6.97	EGFR L858R TP53 G244C	No
FIS330	M	54	IA	ADC	Current	No	5.10	TP53 R283P	No
FIS331	M	75	IIA	ADC	Current	Yes	6.20	TP53 R175H	No
FIS337	M	73	IB	ADC	Former	No	8.27	KRAS G12S	No
FIS343	F	60	IB	ADC	Former	Yes	7.00	TP53 R158L	Yes
FIS345	M	74	IIIA	SCC	Current	No	7.80	No detected	No
FIS347	F	68	IB	ADC	Never	No	17.67	No detected	No

F, Female; M, Male; Hist., Histology; DFS, Disease free survival; ADC, adenocarcinoma; SCC, squamous cell carcinoma. SPH Form: Tumorspheres formation.

1.2. CELL LINES

A549, NCI-H1395, NCI-H1650, NCI-H1975, NCI-H1993, NCI-H2228, NCI-H23, NCI-H358, NCI-H460, HCC827, PC9 and SW900 commercial cell lines were purchased from American Type Culture Collection. The main characteristics of these lung cancer cell lines are described in Table 7. Cell lines were cultured in RPMI-1640 medium containing 10% FBS, 200 µg/mL penicillin/streptomycin, and 0.001% non-essential amino acids. To obtain tumorspheres, monolayer cells were trypsinized using 0.05% trypsin-EDTA when they reached 80% confluence. Trypsinized cells were seeded at low density in ultra-low attachment plates with serum-free RPMI-1640 medium supplemented with 0.4% BSA, 50 µg/mL EGF, 20 µg/mL bFGF, 5 µg/mL ITS PREMIX, 2% B-27, 200 µg/mL penicillin/streptomycin, and 0.001% non-essential amino acids. Again, cultures were

expanded by enzymatic dissociation of spheroids, followed by re-plating of both single cells and residual small aggregates in complete fresh medium replaced twice a week.

Table 7. Clinicopathological characteristics of the cell lines included in the study.

Cell line	Gender	Age (years)	Smoking status	Histology	Mutational status
A549	Male	58	NS	ADC	<i>KRAS</i> G12S
NCI-H1395	Female	55	Current	ADC	<i>BRAF</i> G469A
NCI-H1650	Male	27	Current	ADC	<i>EGFR</i> E746-A750 Del
NCI-H1975	Female	NS	Never	ADC	<i>EGFR</i> L858R, T790M <i>TP53</i> R273H; <i>PIK3CA</i> G118D
NCI-H1993	Female	47	Current	ADC	c-MET amplification
NCI-H2228	Female	NS	Never	ADC	<i>EML4-ALK</i> gene fusion <i>TP53</i> Q331*
NCI-H23	Male	51	NS	ADC	<i>KRAS</i> G12C; <i>TP53</i> M246I
NCI-H358	Male	NS	NS	ADC	<i>KRAS</i> G12C
NCI-H460	Male	NS	NS	LCC	<i>KRAS</i> Q61H; <i>PI3KCA</i> E545K
HCC-827	Female	39	NS	ADC	<i>EGFR</i> E746-A750 Del <i>TP53</i> V218delV
PC9	Male	NS	NS	ADC	<i>EGFR</i> E746-A750 Del <i>TP53</i> R248Q
SW900	Male	53	NS	SCC	<i>KRAS</i> G12V; <i>TP53</i> Q167*

ADC, adenocarcinoma; LCC, large cell carcinoma; NS, not specified; SCC, squamous cell carcinoma.

1.3. ANIMALS

6-8-week-old NOD.CB17-Prkdc^{scid}/NcrCrI mice were purchased from Jackson Laboratories. All procedures and experimental protocols involving the use of animals in this study were in accordance with ethical standards of and approved by the Animal Care and Use Committee, Dirección General de Producción agraria y ganadería, Generalitat Valenciana (2013/027/CIPF/0284). Mice were anesthetized with 40 μ L ketamine (Imalgene 50 mg/mL, Merial) plus xylazine (Rompun 2%, Bayer HealthCare) (9:1) and subcutaneously transplanted with adherent cells and tumorspheres suspensions in serum-free medium and Matrigel (BD) (1:1). Tumor volume (TV) measurements were recorded once a week using the formula: $TV (mm^3) = d^2 * D/2$, where d and D are the shortest and the longest diameters, respectively (286). Animals were euthanized when xenografts reached 1000 mm^3 or when they appeared to be in distress according to the protocol, including weight loss, ulcerated tumor or tumor interfering with normal behavior, and animal showing respiratory difficulties, hypothermia, or pain.

2. METHODS

2.1. CELL PROLIFERATION AND INVASION ANALYSIS

Cell cultures were phenotypically characterized, including analyses of morphology, growth, differentiation and invasion capacities, as well as cell viability after treatment with several drugs.

2.1.1. GROWTH CURVES

Cultures were trypsinized at 80% confluence and counted in a Neubauer camera with trypan blue dye exclusion. Tumor cells were plated at a density of 1,000 cells per well using standard 96-well plates (Corning) for adherent cells and ultra-low attachment 96-well plates (Corning) for tumorspheres and 200 μ L of their corresponding culture medium previously described. Cell viability was evaluated 24, 48 and 72h after seeding using the CellTiter 96[®] AQueous One Solution Cell Proliferation Assay (Promega). This assay contains a tetrazolium compound, [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt; MTS], and an electron coupling reagent (phenazine ethosulfate; PES). The MTS tetrazolium compound (Owen's reagent) is bio-reduced by cells into a colored formazan product that is soluble in culture medium. This conversion is presumably accomplished by NADPH or NADH produced by dehydrogenase enzymes in metabolically active cells. Assays were performed by adding 30 μ L of the CellTiter 96[®] AQueous One Solution Reagent (Promega) directly to culture wells, incubating for 3 hours and then recording the absorbance at 490nm with a Victor 3 96-well plate reader (Wallac). Data represented are the mean of three replicates in three independent experiments.

2.1.2. CYTOTOXICITY ASSAYS

Adherent cells and lung tumorspheres were cultured at desired density according to their growth curves into standard 96-well plates in the case of adherent cells and ultra-low attachment 96-well plates in the case of tumorspheres with 200 μ L of their corresponding culture medium previously described. Drugs were added 24 hours after at the following final concentrations: docetaxel, paclitaxel, vinorelbine,

curcumin, UC-2288, and apigenin were added at 10 μ M, cisplatin and pemetrexed were added at 50 μ M, and salinomycin was added at 1 μ M. All drugs were purchased from Selleckchem, except for UC-2288, which was purchased from Sigma-Aldrich. Cells treated with DMSO (Gibco) served as controls (vehicle control). Cell viability was evaluated after 48 h with the CellTiter 96[®] Aqueous One Solution Cell Proliferation Assay and analyzed with a Victor 3 plate reader. Data represented are the mean of three replicates in three independent experiments.

2.1.3. CELL INVASION ASSAYS AND TIME-LAPSE VIDEO RECORDING

For cell invasion assays, cells were cultured in suspension conditions supplemented with 0.2% methylcellulose in a non-adhesive convex environment for 12 h at 37°C and 5% CO₂. Tumorspheres were mixed with collagen matrix (2.5 mg/ml) and incubated for 30 min at 37°C prior to microscopic analysis. Time-lapse microscopy imaging was performed on a Zeiss AxioObserver Z1 microscope with a Plan-Apochromat 40X/1.3 (NA=1.3, working distance = 0.21 mm), a camera, and an Apotome attachment (Carl Zeiss). Mosaic images were collected using AxioVision software over a period of 20 h with a time resolution of 30 minutes.

2.2. NUCLEIC ACID ANALYSIS

Genomic DNA was extracted from fresh-frozen tissue and cell cultures to assess the most common mutations in lung cancer patients. Additionally, RNA was also isolated in order to perform gene expression analyses.

2.2.1. RNA AND DNA ISOLATION

RNA and DNA isolation was carried out using Tri Reagent[®] (Invitrogen) according to the manufacturer's instructions. For tumor samples, a piece of 10-20 mg of tissue was dissected and 1 mL of Tri Reagent[®] was added. Samples were homogenized using TissueLyser (Qiagen) and 200 μ L of chloroform was added in order to separate the aqueous phase containing the RNA. Isopropanol was used to precipitate the nucleic acids and ethanol was used for washing. Total RNA was dissolved in nuclease-free water and stored at -80° C until further analysis. DNA interphase was collected in absolute

ethanol and washed, first with 10% ethanol/0.1 M sodium citrate buffer, and then with 75% ethanol. DNA was dissolved in nuclease-free water and stored at -80° C until further analysis. For cell cultures, tumor cells were trypsinized using 0.05% trypsin-EDTA when they reached 80% confluence and centrifuged at 290 g for monolayer cells and 200 g for tumorspheres during 5 minutes. Cell pellets were washed twice with PBS and the same procedure followed for RNA and DNA extraction from fresh-frozen tissue specimens was used. RNA and DNA quantity and quality was assessed using a nanospectrophotometer (Nano Drop 2000C, Thermo Fisher Scientific).

2.2.2. DETERMINATION OF THE MUTATIONAL STATUS

2.2.2.1. DETECTION OF EGFR MUTATIONS

The theraScreen® EGFR RGQ PCR (Qiagen) kit was used to analyze the *EGFR* mutations. This kit allows detecting 29 somatic mutations in the *EGFR* gene in exons 18-21 (Table 8) by RTqPCR combined with ARMS® and Scorpions® technologies. This method is highly selective and, depending on the total amount of DNA present, enables detection of a low percentage of the mutant gene in a background of wild-type genomic DNA. Each reaction was performed in 96-well plates with a final reaction volume of 12.5 µL, comprising 9.75 µL of reaction mix (primers, probes, dNTPs and reaction buffer containing Cl₂Mg), 0.25 µL of Taq DNA Polymerase, and 2.5 µL of DNA (2-10 ng/µL) isolated from fresh-frozen lung cancer specimens. A positive and negative control (provided by the manufacturer) was included in each run. The reaction was performed using a LightCycler 480 II thermocycler (Roche) following the conditions detailed in Table 9.

Table 8. *EGFR* mutations detected using theraScreen® kit.

Exon	Mutations detected
19	19 deletions*
20	T790M
21	L858R
21	L861Q
18	G719X (G719S, G719A and G719C)*
20	S768I
20	3 insertions *

* The kit does not distinguish between them.

Table 9. TheraScreen® EGFR RGQ PCR cycling parameters.

Cycles	Time	Temperature
1	15 minutes	95° C
40	30 seconds	95° C
	60 seconds	60° C

2.2.2.2. DETECTION OF KRAS MUTATIONS

The analysis of *KRAS* mutations was carried out using a TheraScreen® *KRAS* Pyro® kit (Qiagen). This kit is used for quantitative detection of mutations in codons 12, 13, and 61 of the human *KRAS* gene by pyrosequencing. Codons 12/13 and codon 61 were amplified by PCR using 5 µL of template DNA (10 ng of genomic DNA), 12.5 µL of PyroMark® PCR Master Mix 2x, 2.5 µL of Coral Load Concentrate 10x, 4 µL of nuclease-free water, and 1 µL of *KRAS* 12/13 or 61 PCR primers. The reactions took place in a MasterCycler® thermocycler (Eppendorf) following the conditions described in Table 10.

The amplicons were immobilized on Streptavidin Sepharose® High Performance beads (GE Healthcare) to prepare the single-stranded DNA and anneal the sequencing primers to it using a PyroMark Q24 plate and a vacuum workstation. PyroMark Gold Q24 reagents (enzyme mixture, substrate mixture, and nucleotides; Qiagen) were then prepared and loaded into a cartridge so they could be dispensed during the sequencing process. Finally, the plate and the cartridge were loaded into the PyroMark Q24 System and the sequencing process was started. The sequences were analyzed using software provided by the manufacturer. In each run, two controls were included: unmethylated control DNA which worked as a positive control for PCR and sequencing reactions, and a negative control (without template DNA).

Table 10. TheraScreen® *KRAS* Pyro® PCR cycling parameters.

Phases	Time	Temperature
Initial activation step	15 minutes	95° C
3-step cycling:		
Denaturation	20 seconds	95° C
Annealing	30 seconds	
Extension	20 seconds	72° C
Number of cycles	42	
Final extension	5 minutes	72° C

This kit allows the detection of the most frequent mutations at codons 12, 13, and 61. The nucleotide dispensation order is *TACGACTCAGATCGTAG* for sequencing codons 12/13, and *GCTCAGTCAGACT* for codon 61. The analyzed sequence for codons 12/13 is *GNTGRCGTAGGC*, which allows detecting the most frequent mutation in codon 12, nucleotide 35 (second position). To analyse if the mutation is present in nucleotide 34 (first position), the analysis sequence is changed to *NGTGRCGTAGGC*. In the case of exon 61, the analysis sequence is *CTCDTGACCTG*, which represents the most frequent mutation in this codon, detected in nucleotide 183 (third position). To analyse if the mutation is present in nucleotide 182 (second position) the analysis sequence is changed to *CTCTHGACCTG*, and to analyze if it is present in nucleotide 181 (first position), the sequence is *CTCTTSACCTG*. The expected histograms for each sequence are represented in Figure 11.

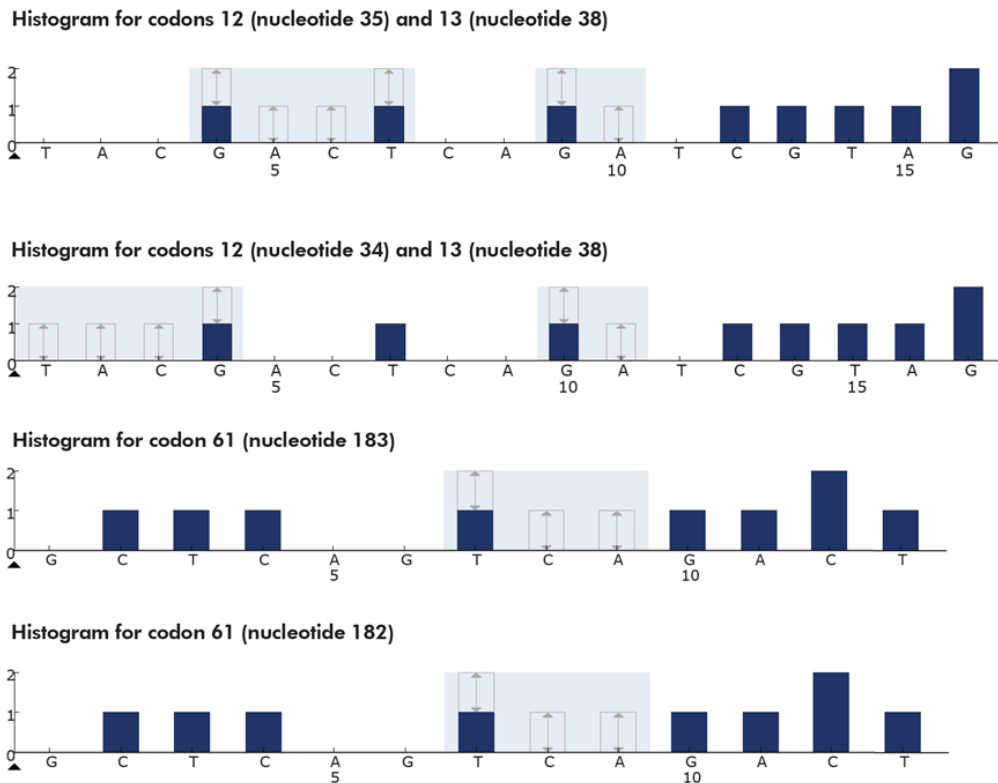


Figure 11. Representation of the histograms for the codons and nucleotides according to the sequences for analysis.

2.2.3. QUANTIFICATION OF GENE EXPRESSION

2.2.3.1. REVERSE TRANSCRIPTION

Reverse transcription was performed in order to transform RNA into complementary DNA (cDNA), which was required for the subsequent analyses, by using a High Capacity cDNA Reverse Transcription Kit® (Applied Biosystems). Each reaction comprised 2 µL of reverse transcription (RT) buffer, 0.8 µL of dNTP mix, 2 µL of RT random primers, 1µL MultiScribe™ Reverse Transcriptase, 1µL RNase inhibitor, and a variable volume of RNA depending on the sample concentration (500 ng RNA per reaction), made up to a final volume of 20 µL with nuclease-free water. The reactions took place in a MasterCycler® thermocycler (Eppendorf) following the conditions described in Table 11 and the resulting cDNA was stored at -80° C until further analysis.

Table 11. Cycling program for reverse transcription reaction.

Phase	Time	Temperature
1	10 minutes	25° C
2	2 hours	37° C
3	5 seconds	85° C

2.2.4. QUANTITATIVE REAL TIME PCR

Target gene quantification was performed by RTqPCR using hydrolysis probes labelled with a reporter dye linked to the 5' end of the probe (TaqMan®, Applied Biosystems). In this study, we analyzed a total of 52 genes that were selected according to their relevance to the biology of CSCs. The relevance of these genes was established from a PubMed database search, which revealed published information demonstrating or suggesting a role for these genes in CSCs maintenance, self-renewal, pluripotency, invasion and proliferation capacities. Gene expression levels were assessed using TaqMan® Gene Expression Assays (Applied Biosystems) which are listed in Table 12.

Table 12. Genes analyzed in this study, their description and TaqMan® assays used for RTqPCR.

Gene	Description	Assay	Amplicon length
<i>ABCG2</i>	ATP-binding cassette, sub-family G	Hs01053790_m1	83
<i>ALDH1A1</i>	Aldehyde dehydrogenase 1 family, member A1	Hs00946916_m1	61
<i>BMI1</i>	BMI1 proto-oncogene, polycomb ring finger	Hs00180411_m1	105
<i>CCND1</i>	Cyclin D1	Hs00765553_m1	57
<i>CD133</i>	CD133 molecule	Hs01009257_m1	80
		Hs01009250_m1	75
		Hs00195682_m1	107
<i>CD166</i>	CD166 molecule	Hs00233455_m1	70
<i>CD44</i>	CD44 molecule	Hs01075861_m1	70
<i>CD45</i>	Protein Tyrosine Phosphatase, Receptor Type C	Hs00898488_m1	61
<i>CDH1</i>	Cadherin 1, type 1	Hs01023894_m1	61
<i>CDKN1A</i>	Cyclin-dependent kinase inhibitor 1A	Hs999999142_m1	99
<i>CDKN2A</i>	Cyclin-dependent kinase inhibitor 2A	Hs00923894_m1	115
<i>CEACAM5</i>	Carcinoembryonic antigen-related cell adhesion molecule 5	Hs00944025_m1	71
<i>CTNNB1</i>	Catenin beta 1	Hs00355049_m1	67
<i>DKK1</i>	Dickkopf WNT signaling pathway inhibitor 1	Hs00183740_m1	68
<i>DLL1</i>	Delta-like 1	Hs00194509_m1	74
<i>DLL4</i>	Delta-like 4	Hs00184092_m1	78
<i>EPCAM1</i>	Epithelial cell adhesion molecule	Hs00158980_m1	64
<i>FZD7</i>	Frizzled class receptor 7	Hs00275833_s1	70
<i>GLI1</i>	GLI family zinc finger 1	Hs01110766_m1	83
<i>GSK3B</i>	Glycogen synthase kinase 3 beta	Hs01047719_m1	65
<i>HES1</i>	Hes family bHLH transcription factor 1	Hs00172878_m1	78
<i>HEY1</i>	Hairy ears, Y-linked	Hs01114113_m1	82
<i>ITGA2</i>	Integrin α 2	Hs00158127_m1	67
<i>ITGA6</i>	Integrin α 6	Hs01041011_m1	64
<i>ITGB1</i>	Integrin β 1	Hs00559595_m1	75
<i>JAG1</i>	Jagged 1	Hs00164982_m1	65
<i>KLF4</i>	Kruppel-like factor 4 (gut)	Hs00358836_m1	110
<i>LGALS2</i>	Lectin, galactoside-binding, soluble, 2	Hs00197810_m1	73
<i>LIN28B</i>	Lin-28 homolog B	Hs01013729_m1	130
<i>MDM2</i>	MDM2 proto-oncogene, E3 ubiquitin protein ligase	Hs01066930_m1	99
<i>MMP2</i>	Matrix metalloproteinase 2	Hs01548727_m1	65
<i>MMP9</i>	Matrix metalloproteinase 9	Hs00234579_m1	54
<i>MUC1</i>	Mucin 1, cell surface associated	Hs00159357_m1	84
<i>MYC</i>	V-Myc avian myelocytomatosis viral oncogene homolog	Hs00153408_m1	107
<i>NANOG</i>	Nanog homeobox	Hs02387400_g1	109
<i>NOTCH1</i>	<i>NOTCH1</i>	Hs01062014_m1	80
<i>NOTCH2</i>	<i>NOTCH2</i>	Hs01050702_m1	60
<i>NOTCH3</i>	<i>NOTCH3</i>	Hs01128541_m1	81
<i>NUMB</i>	NUMB, Endocytic Adaptor Protein	Hs01105433_m1	71
<i>OCT4</i>	POU class 5 homeobox 1	Hs01895061_u1	130
<i>PTCH1</i>	Patched 1	Hs00181117_m1	72
<i>SHH</i>	Sonic hedgehog	Hs00179843_m1	70
<i>SMO</i>	Smoothed, frizzled class receptor	Hs01090242_m1	54

<i>SNAI1</i>	Snail family zinc finger 1	Hs00195591_m1	66
<i>SOX2</i>	SRY (sex determining region Y)-box 2	Hs01053049_s1	91
<i>THY1</i>	Thy-1 Cell Surface Antigen	Hs00264235_s1	99
<i>VIM</i>	Vimentin	Hs00185584_m1	73
<i>WEE1</i>	WEE1 G2 checkpoint kinase	Hs00268721_m1	66
<i>WNT1</i>	Wingless-type MMTV integration site family, member 1	Hs01011247_m1 Hs00180529_m1	108 77
<i>WNT2</i>	Wingless-type MMTV integration site family, member 2	Hs00608224_m1	119
<i>WNT3</i>	Wingless-type MMTV integration site family, member 3	Hs00902257_m1	76
<i>WNT5A</i>	Wingless-type MMTV integration site family, member 5A	Hs00998437_m1	61

Different endogenous gene controls were tested in fresh-frozen and cell cultures specimens in order to evaluate the best internal control for each case using GeNorm software (Table 13). This software automatically calculates the gene-stability measurement 'M' for all control genes and allows the worst-scoring housekeeping genes to be eliminated (287). Each reaction was performed twice in 384-well plates with a final volume of 5 μL comprising 2.5 μL of TaqMan[®] Gene Expression Master Mix (Applied Biosystems), 1.25 μL of nuclease-free water, 0.25 μL of TaqMan[®] Gene Expression Assay mix (Applied Biosystems), and 1 μL of cDNA. Non-template controls (NTCs) were included in each run, as well as a commercially available reference cDNA (Clontech) as positive reference control. The reactions took place in a Light Cycler 480 thermocycler system following the cycling conditions described in Table 14. The efficiency of each TaqMan[®] assay was evaluated by carrying out serial dilutions (50 ng/ μL , 5 ng/ μL , 0.5 ng/ μL , 0.05 ng/ μL , 0.005 ng/ μL , and 0.0005 ng/ μL) using the cDNA as a template. The efficiency was calculated by using the following equation: $E = 10^{-1/\text{slope}}$ and the results indicated that almost all the assays used were adequately efficient (Supplementary Table 1).

Table 13. Endogenous gene TaqMan[®] assays used for the normalization of the results.

Gen	Description	Assay	Amplicon length
<i>ACTB</i>	Actin, Beta	Hs99999903_m1	171
<i>CDKN1B</i>	Cyclin-dependent kinase inhibitor	Hs00153277_m1	71
<i>GUSB</i>	Glucuronidase, beta	Hs01558067_m1	71
<i>HPRT1</i>	Hypoxanthine phosphoribosyltransferase 1	Hs01003267_m1	72
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase	Hs99999905_m1	122

Relative gene expression levels were expressed as the ratio of target gene expression to reference gene expression by using the Pfaffl formula (288). Herein, relative quantification determines the changes in steady-state mRNA levels of a target gene across multiple samples and expresses it relative to the levels of control RNA. The expression is normalized against a reference gene, which is often a housekeeping gene.

Table 14. Cycling program for RTqPCR.

	Step	Time	Temperature
Pre-PCR	UNG incubation	2 min	50° C
	Taq activation	10 min	95° C
PCR (40 cycles)	Denature	15 sec	95° C
	Anneal/Extend	1min	60° C

The step at 50° C is required for optimal UNG enzyme activity. The step at 95° C is required to activate the AmpliTaq Gold enzyme.

2.3. PROTEIN ANALYSIS

In addition to the gene expression studies, protein analyses were carried out to determine their expression and location patterns using immunoblotting and immunofluorescence.

2.3.1. IMMUNOBLOTTING

For protein isolation, culture medium was removed and tumorspheres were washed with cold PBS and lysed on ice with lysis buffer containing 50 mM Tris-HCl pH 8, 150 mM NaCl, 0.02% NaN₃, 0.1% sodium dodecyl sulfate (SDS), 1% Igepal CA-630 (NP40), 0.5% sodium deoxycholate, 2 mg/ml leupeptin, 2 mg/ml aprotinin, and 1 mM phenyl-methyl-sulphonyl fluoride. Adherent cultures were washed with cold PBS and scraped out of the dishes with lysis buffer. All lysates were incubated for 30 minutes on ice and centrifuge at 10000 g for 10 min at 4° C. Supernatants were collected and stored at -80° C until further analysis. Protein quantification was performed by spectrophotometry using the Bradford protein assay (Bradford reagent, Sigma-Aldrich). Absorbance was measured at 595 nm using a Victor3™ -1420 Multilabel Plate Counter (Perkin Elmer, Boston, MA, USA) and protein concentration was calculated by interpolating absorbance in a standard curve prepared with standard solutions of BSA.

Proteins were then separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). 20 µg of each protein extract were mixed with Laemmli buffer containing 50mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 0.1M Dithiothreitol (DTT) and 0.1% bromophenol blue and denatured for 5 min at 95° C. Electrophoresis was carried out at 150 V for approximately 1 hour and molecular weights were determined using a protein ladder (Rainbown Molecular Weight Markers, Amershan, Ge Healthcare). Separated proteins were next transferred to polyvinylidene difluoride (PVDF) membranes (Immobilion®-P, Millipore) at 100 V and 4 °C for 60 min. Transference was confirmed by staining membranes with 0.5% Ponceau S solution and 1% glacial acetic acid. This staining was removed using a washing solution, including 0.05% Tween 20 (Panreac) in PBS, and blocked with 5% defatted milk in washing solution for an hour at room temperature (RT) to prevent non-specific binding of antibodies. Afterwards, membranes were incubated with a dilution of the primary antibodies listed in Table 15 in blocking solution overnight at 4° C, washed three times for 10 min each at RT with washing solution and incubated with anti-Mouse IgG (whole molecule)–horseradish peroxidase (HRP) conjugated secondary antibody (A9044, Sigma-Aldrich) or anti-Rabbit IgG-HRP conjugated secondary antibody (sc-2357, Santa Cruz Biotechnology) for 1 hour at RT. Membranes were washed three times again and peroxidase activity was detected by incubating membranes with the chemiluminescent immunodetection system Amersham ECL Western Blotting Analysis System (GE Healthcare) and the photographic films Amersham Hyperfilm ECL (GE Healthcare). Densitometric analysis was performed using ImageJ (NIH) and all results were normalized over β-actin.

Table 15. Characteristics of the primary antibodies used for protein detection by immunoblotting.

Antigen	Antibody type	Clone	Working dilution	Supplier	Catalog number
ALDH1A1	Rabbit polyclonal	Polyclonal	1:1000	Abcam	ab23375
B-catenin	Rabbit polyclonal	Polyclonal	1:500	Novus Biologicals	NBP1-89989
CD44	Mouse monoclonal	156-3C11	1:1000	Cell Signaling Technology	#3570
E-cadherin	Mouse monoclonal	4A2	1:1000	Cell Signaling Technology	#14472
EpCAM	Mouse monoclonal	VU1D9	1:1000	Cell Signaling Technology	#2929
Integrin α6	Mouse monoclonal	F-6	1:250	Santa Cruz Biotechnology	sc-374057
Nanog	Mouse monoclonal	1E6C4	1:500	Santa Cruz Biotechnology	sc-293121
Notch3	Mouse monoclonal	A-6	1:500	Santa Cruz Biotechnology	sc-515825
p21 Waf1/Cip1	Mouse monoclonal	DCS60	1:1000	Cell Signaling Technology	#2946
Snail	Mouse monoclonal	G-7	1:1000	Santa Cruz Biotechnology	sc-271977
Sox2	Rabbit polyclonal	Polyclonal	1:3333	Abcam	ab15830
Vimentin	Mouse monoclonal	V9	1:1000	Sigma-Aldrich	V6389
B-actin	Mouse monoclonal	AC-15	1:5000	Sigma-Aldrich	A5441

2.3.2. IMMUNOFLUORESCENCE

For immunofluorescence analysis, adherent cells were seeded on coverslips and cultured until they reached 80% confluence and tumorspheres were resuspended in PBS at a final concentration around 5×10^5 cell/mL. 100 μ L of cell suspensions were centrifuge at 400 g for 5 min using a cytospin3 (Thermo Shandon). Cells were fixed in 4% paraformaldehyde in PBS pH 7.4 at RT for 15 min and washed three times with PBS for 5 min. Permeabilization of cell membranes was carried out using 0.4% Triton X-100 in PBS for 10 min. Next, cells were washed three times and blocked in PBS containing 10% BSA and 0.4% Triton X-100 for 1 hour at RT. Immunodetection was performed using the antibodies listed in Table 16.

Table 16. Characteristics of the primary antibodies used for protein detection by immunofluorescence.

Antibody	Antibody type	Clone	Working dilution	Supplier	Catalog number
ALDH1A1	Rabbit polyclonal	Polyclonal	1:400	Abcam	ab23375
B-catenin	Rabbit polyclonal	Polyclonal	1:1000	Sigma-Aldrich	C2206
CD44	Mouse monoclonal	156-3C11	1:400	Cell Signaling Technology	#3570
CD133	Rabbit polyclonal	Polyclonal	1:100	Novus Biologicals	NB120-16518
CD166	Rabbit monoclonal	EPR2759(2)	1:100	Abcam	ab109215
E-cadherin	Mouse monoclonal	G-10	1:150	Santa Cruz Biotechnology	sc-8426
Integrin α6	Mouse monoclonal	F-6	1:50	Santa Cruz Biotechnology	sc-374057
Nanog	Mouse monoclonal	1E6C4	1:250	Santa Cruz Biotechnology	sc-293121
Notch3	Mouse monoclonal	A-6	1:50	Santa Cruz Biotechnology	sc-515825
p21 Waf1/Cip1	Mouse monoclonal	DCS60	1:100	Cell Signaling Technology	#2946
Snail	Mouse monoclonal	G-7	1:50	Santa Cruz Biotechnology	sc-271977
Vimentin	Mouse monoclonal	RV202	1:500	Abcam	ab8978

Cells were incubated overnight at 4°C with primary antibodies in blocking buffer and washed with PBS three times for 5 min each washing step. Thereafter, cells were incubated with Alexa Fluor 488 donkey anti-Mouse IgG (A21202, Molecular Probes) or Alexa Fluor 555 donkey anti-Rabbit IgG (A31572, Molecular Probes) in blocking solution for an hour at RT. Finally, nuclear staining was done incubating cells with 1:1000 DAPI (Thermo Fisher Scientific) in PBS for 5 min at RT and washing with PBS for 5 min. For mounting, coverslips were immersed for a few seconds in distilled water to avoid crystal precipitation from buffer salts and placed face-down on the slides, where previously a drop of mounting media (Fluoromount Aqueous Mounting Medium, Sigma-Aldrich) was added. Mounted slides were dried, sealed with nail polish and analyzed under a Leica TCS-SP2-AOBS confocal microscope (Leica Microsystems).

2.3.3. IMMUNOHISTOCHEMISTRY

The histopathological study was carried out in lung tumor tissue for the determination of ALK and ROS1 rearrangements and in primary cultures to analyze CD56, carcinoembryonic antigen, cytokeratin 5/6, cytokeratin 7 and p63. Tumor tissue was fixed in 10% formaldehyde at room temperature for 15 min, washed and embedded in paraffin, whereas cell cultures were fixed in 4% paraformaldehyde at room temperature for 15 min, washed and embedded in Richard-Allan Scientific™ HistoGel (Thermo Scientific) prior to paraffin. Each formalin-fixed paraffin-embedded (FFPE) sample was cut in 5 µm sections for hematoxylin and eosin (HE) staining and for Dako charged slides (Code K8020, Dako).

HE staining was used to assess the morphology of the tissue studied. Briefly, deparaffinization was performed by incubating slides at 60°C for 20 minutes and then submerging in xylene. Slides were hydrated with ethanol (absolute/96°/80°/70°) and rinsed with water. They were then stained with hematoxylin, decolorized with acid chloride, and immersed in lithium carbonate. Slides were counterstained with eosin, dehydrated with ethanol (96°/absolute) and cleared with xylene. Samples were mounted with a cover slip using DPX Mountant for histology (Sigma). Immunodetection was done using the antibodies listed in Table 17 from Master Diagnostica and comprised the following steps.

One charged slide was analyzed for each sample and antibody. Sections were dried at 60°C for 60 minutes prior to immunohistochemistry (IHC) and loaded into the PT Link instrument where the antigen retrieval/dewaxing process took place. This consisted in a temperature rise to 95°C over a period of 20 minutes, incubation of the samples for 20 minutes with the high pH EnVision™ FLEX Target Retrieval Solution (Dako) and cooling for an additional 10 minutes. Sections were transferred to a Dako Autostainer Link 48 instrument, which is a compact, bench-top, open system that allows decoupled pre-treatment, in which up to 48 IHC slides can be run at the same time. The immunostaining was done with Dako FLEX Ready-to-Use format. Samples were incubated with peroxidase blocking reagent for 10 minutes followed by an incubation with the primary antibody for 5 to 40 minutes. The detection was carried out by an

incubation for 20 minutes with the Dako EnVision™ FLEX detection system, which consists of a dextran backbone to which a large number of HRP and secondary antibody molecules have been coupled. A final incubation with the chromogen (3,3'-diaminobenzidine, DAB, Sigma-Aldrich) for 10 minutes was performed. Afterwards, slides were flooded with distilled water and then manually counterstained with hematoxylin, dehydrated and cover slipped. Positive and negative controls were fixed, processed in the same way than the rest of the samples and included in each run. Human normal tonsil tissue was used as a positive control, whereas a duplicate of one of the samples was included in the run without incubation with the primary antibody as a negative control.

Table 17. Characteristics of the primary antibodies and conditions used to detect each protein by immunohistochemistry.

Antigen	Antibody type	Clone	Incubation time	Dilution buffer	Catalog number
ALK/p80	Mouse monoclonal	5A4	20 minutes	Tris-EDTA buffer pH9	MAD-001720QD
ROS1	Rabbit monoclonal	D4D6	40 minutes	Tris-EDTA buffer pH9	MAD-000746QD
CD56/NCAM-1	Mouse monoclonal	123C3.D5	10 minutes	Tris-EDTA buffer pH8	MAD-000749QD
CEAm	Mouse monoclonal	COL-1	5 minutes	Tris-EDTA buffer pH8	MAD-002095QD
CEAp	Rabbit polyclonal	Polyclonal	15 minutes	Tris-EDTA buffer pH8	MAD-001115QD
Cytokeratin 5/6	Rabbit monoclonal	EP64+EP24	10 minutes	Tris-EDTA buffer pH8	MAD-000651QD
Cytokeratin 7	Mouse monoclonal	OV-TL12/30	10 minutes	Tris-EDTA buffer pH8	MAD-001004QD
p63	Mouse monoclonal	4A4	20 minutes	Tris-EDTA buffer pH8	MAD-000479QD

ALK, anaplastic lymphoma kinase; CEA, carcinoembryonic antigen.

2.4. BIOINFORMATIC AND STATISTICAL ANALYSIS

First, we evaluated if the variables followed a normal distribution by using the Kolmogorov-Smirnov test. Because the variables did not follow a normal distribution, statistical analyses were conducted by non-parametric tests. Expression of paired adherent cells and tumorspheres was analyzed using Wilcoxon signed-rank test. Continuous variables were compared using non-parametric Mann Whitney U and Kruskal Wallis tests. Spearman rank test was used to test for correlations between continuous variables, and the association between discrete variables was evaluated by the Chi-square test. In order to reduce the dimensionality and remove possible collinear expression of genes, a logistic regression model was built using a stepwise selection and minimizing Akaike's Information Criterion (AIC) to select the genes which more contributed to differentiate tumorspheres from adherent cells.

To study the correlation between genes overexpressed in tumorspheres compared to adherent cells and survival, *in silico* analyses were performed using two lung cancer data sets from TCGA consortium (28,43). Clinical and RNA-seq information was directly downloaded from The International Cancer Genome Consortium (ICGC) Data Portal (289). Overall survival (OS) and relapse-free survival (RFS) were calculated from the date of surgery to the end point of the study or to the last recorded follow-up and patient progression was assessed following the Response Evaluation Criteria in Solid Tumors (RECIST) criteria (290). The survival analysis was performed using a univariate Cox regression method using clinicopathological variables and dichotomized gene expression markers. Survival curves were created using the Kaplan-Meier method, and the statistical significance between survival curves was assessed using the log-rank test.

Furthermore, we also calculated the gene expression scores based on multi-gene signatures, which can provide more accurate predictions than a model, using single genes. For this purpose, we constructed a gene expression score using a method previously reported by Lossos et al. (291). All the genes analyzed were included, and expression values were introduced as continuous variables. First, a univariate COX regression analysis was performed to identify which genes were moderately associated with OS, which were those that had a |Z-score| higher than 1.5 ($p < 0.13$). Z-scores are

defined as the regression coefficient (b) obtained from univariate Cox regression, divided by their standard error. The selected genes were included in a multivariate model and regression coefficients from this model were multiplied by the gene expression values and summed to build the expression score. For multivariate regression models, missing values for genes were replaced with the average values (292).

Finally, to assess the independent value of the tested biomarkers, a Cox proportional hazard model for multivariate analyses was used. All significant variables (both biomarker and clinicopathological markers) from the univariate analyses were entered into the multivariate analyses in a forward stepwise Cox regression analysis. A probability of 95% ($p < 0.05$) was considered statistically significant for all analyses. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, Chicago, IL, USA) version 15.0. Principal component analysis (PCA) were performed with the SIMCA-P software (Umetrics, Umea, Sweden) version 13.0 using unit variance (UV) scaling method. RNA-seq data analyses were performed using Linux platform (Ubuntu). GraphPad Prism (GraphPad Software Inc.) version 5.0 and R version 3.5 were used in order to build some of the graphics presented here.

IV. RESULTS & DISCUSSION

1. ESTABLISHMENT OF ADHERENT AND NON-ADHERENT CULTURES

In this study, we processed 20 tumor specimens from early-stage NSCLC patients who underwent surgery (lobectomies or pneumonectomies). Primary patient-derived lung cancer cell cultures were maintained for at least four weeks before they were split for the first passage. Cultures were successfully established in 8 out of 20 cases (40%), being possible to grow tumor cells as monolayers and tumorspheres. A primary culture was considered established when cancer cells were free of stromal fibroblasts and could be cryopreserved, thawed, and re-grown, whereas it was classified as a failed culture when exhibiting no cancer cells after 6 months of culture. No significant associations were found between the establishment of primary cultures and clinicopathological variables, including sex, age, histology, progression or mutational status (see Table 5). However, a tendency in favor of never or former smokers was observed ($p = 0.06$) probably associated with healthier tissues. The success rate for the establishment of primary cultures of our study was in line with the results previously obtained for some other groups (293). For instance, Eramo et al. were the first to identify and expand lung CSCs, with a success rate of 37% (98). Intriguingly, Wang and colleagues performed another study including primary lung cancer patient samples, and reported a success rate of 75% (123). Some other studies have included lung cancer primary cultures, but no data about unviable cultures is given (294–296).

Several external causes have been described to influence the success rate of primary cell culture establishment, such as excessive necrosis of tumor samples, deficient preservation of tumor samples, fibroblast outgrowth or lack of cancer cells (293). Fibroblast outgrowth was the main reason that caused primary cultures failure in this study. Therefore, as we gained experience in the culture process, we used a number of techniques to overcome this issue. We took advantage of the fact that, in some cases, fibroblasts detached faster and adhered more slowly than cancer cells to use differential trypsinization timing and re-adhesion to separate cancer cells from fibroblasts. In addition, cancer cell colonies were sometimes specifically picked from the plate to separate them from the stromal fibroblasts. Furthermore, we tried to add 0.5M D-

valine, which has been historically used to inhibit fibroblast growth due to their lack of enzyme D-amino acid oxidase, allowing the selective growth of epithelial cells (297). However, only one fibroblast was enough to keep on their expansion. We also attempted culturing cells with Defined Keratinocyte-Serum Free Medium (Gibco, Paisley, UK). The medium is formulated to inhibit fibroblast proliferation and has a low calcium concentration (<0.1 mM) allowing for the isolation and culture of other cell populations. However, the growth of tumor cells was also affected, with cells showing altered morphology and dying in some cases. Finally, we used Anti-Fibroblast Microbeads kit (Miltenyi Biotech, Teterow, Germany), which showed better results. Here, fibroblasts are magnetically labeled and the cell suspension is loaded onto a MACS[®] Column, which is placed in the magnetic field of a MACS Separator. The magnetically labeled fibroblasts are retained within the column whereas the unlabeled cells run through depleted of fibroblasts. In all cases, established primary cultures were carefully examined using a light phase-contrast microscope and the absence of fibroblasts was confirmed as visualized by eye and using immunofluorescence staining with anti-cytokeratin 7 (anti-CK7, a lung epithelial cell marker) and anti-fibroblast surface protein (anti-FSP), in which tumor cells react with both markers whereas fibroblasts are negative for anti-CK7 (Figure 12).

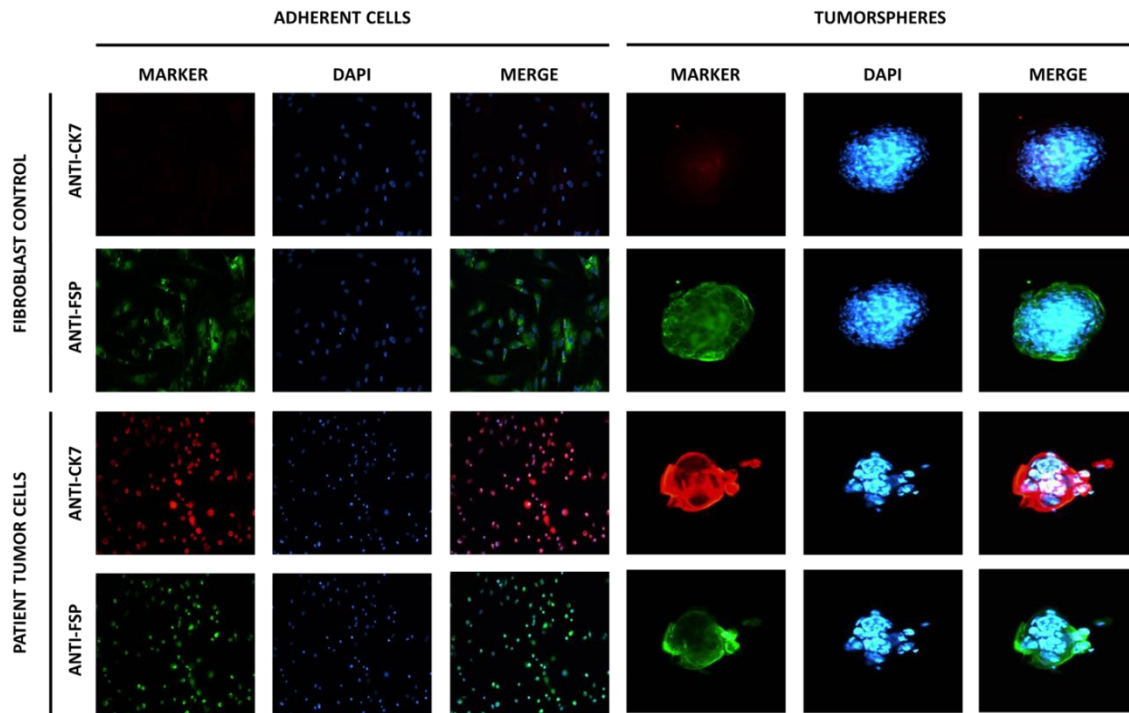


Figure 12. Representative immunofluorescence images of fibroblast controls and tumor cells. Primary tumor cultures react with both anti-fibroblast surface protein (anti-FSP) and anti-cytokeratin 7 (anti-CK7), whereas fibroblasts react only with anti-FSP.

The examination of patient-derived cultures using the light phase-contrast microscope revealed great heterogeneity between the adherent-cultured cells (Figure 13A). Cells from patients FIS299 and FIS301 grew as multilayers and formed cell colonies. These cell cultures showed abundant cell-cell interactions in the form of filopodia and lamellipodia and presence of giant cells and vesicles, which can be seen in Figure 14. FIS317 cells were cubic, grew as a monolayer, but showed tight cell-cell contact with filopodia and a high number of vesicles. In contrast, cells from patients FIS302, FIS303, and FIS315 were more elongated, with brighter nuclei, fewer interactions and a more isolated growth. FIS320 and FIS343 cells were similar to patients FIS302, FIS303, and FIS315 with minor differences in their morphology.

Regarding tumorspheres, differences in their morphologies were noticed as well. FIS299, FIS301, and FIS315 formed tight spheroids, whereas FIS302, FIS303, FIS317, and FIS320 formed more loose and irregularly-shaped spheres, and FIS343 showed a mixed behavior (Figure 13B).

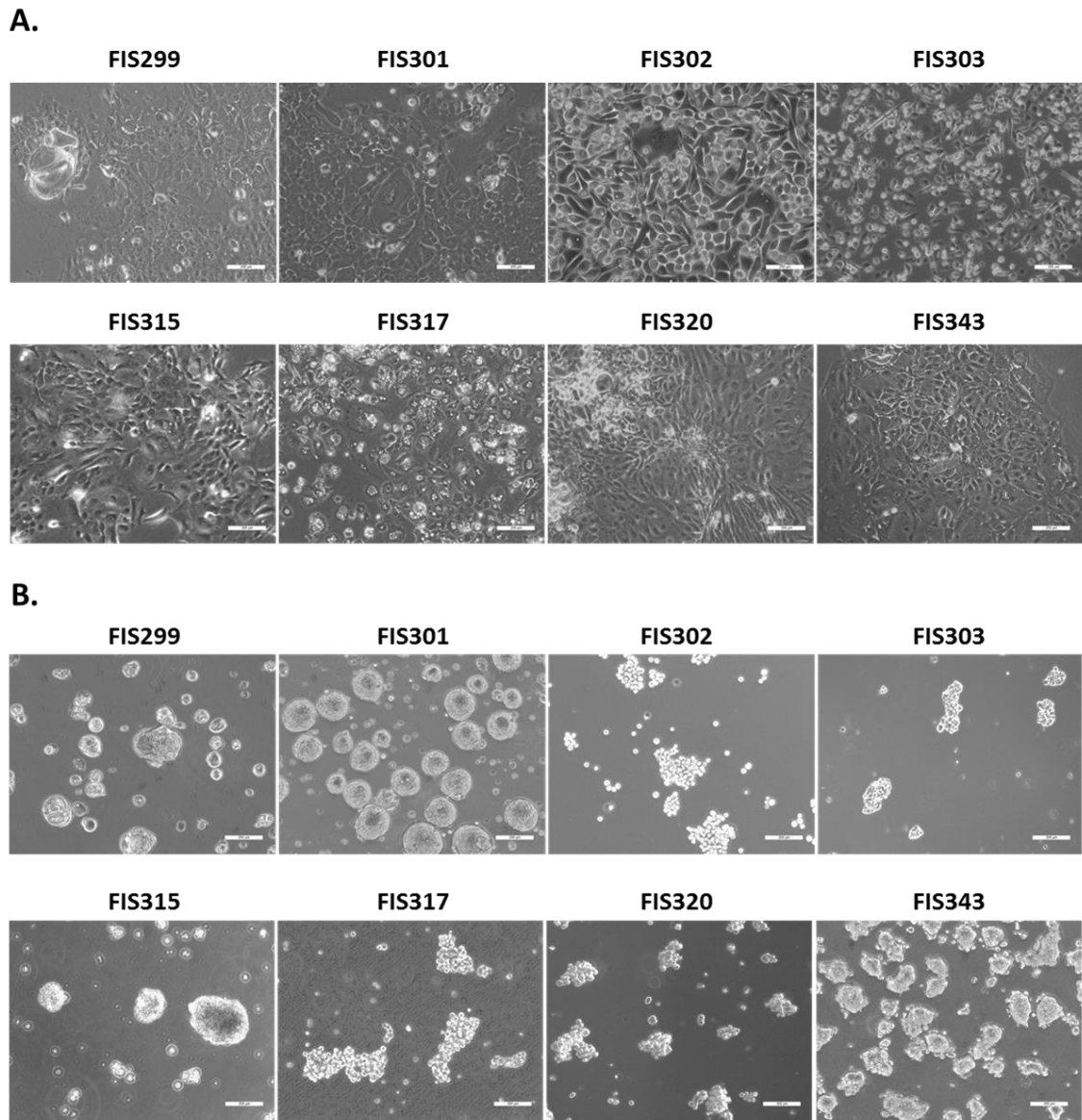


Figure 13. Representative images of the primary patient-derived cancer cells cultured under adherent (A) and non-adherent (B) conditions. All pictures were taken in well-established cultures (passage > 10).

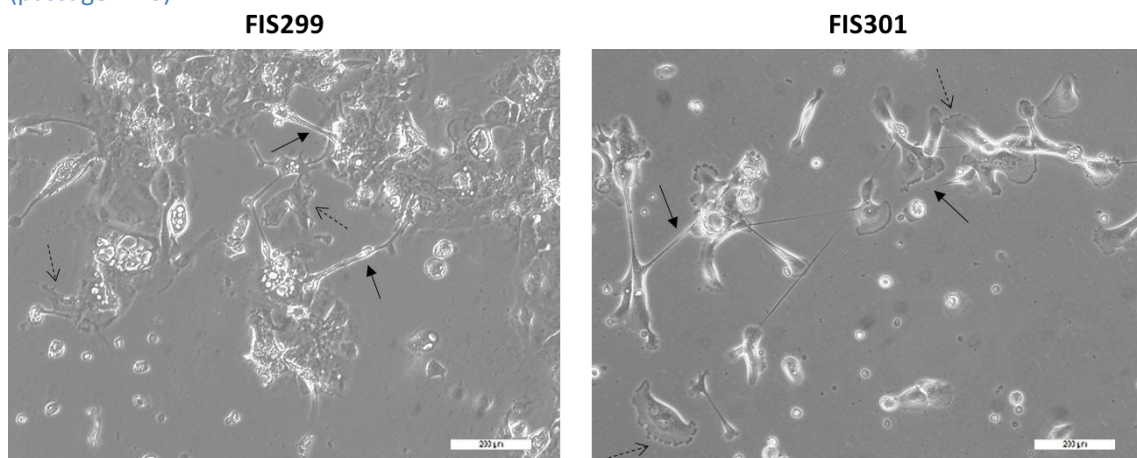


Figure 14. Representative images of adherent FIS299 and FIS301 primary cultures. These primary cultures showed abundant cell-cell interactions in the form of filopodia (solid arrows) and lamellipodia (dashed arrows) and presence of giant cells and vesicles.

Simultaneously, cell line cultures were established using both culture conditions and included in further analyses (Figure 15). Adherent-cultured cells from A549, NCI-H1975, NCI-H1993, NCI-H460, PC9 showed few interactions and isolated growth, but were notably different in terms of morphology. A549 cells were triangular with a well-defined nuclear membrane, whereas PC9 cells were elongated with less marked nuclei. H23 cells were circular, formed lamellipodia and tend to have big nucleus occupying most of the cell volume. In addition, H460 cells were cubic and small, but with presence of some giant cells. Similarly, some cells from H1993 were giant and others had marked nucleus, showed filopodia and a high number of vesicles. On their behalf, HCC827, NCI-H1395, NCI-H1975, NCI-H2228, NCI-H358, and SW900 cells formed cell colonies, whereas NCI-H1650 cells showed mixed behavior with abundant cell-cell interactions. NCI-H1395 cells were small and circular and grew as multilayers similar to cells from FIS299 and FIS301 patients. HCC827 and NCI-H1975 cells formed abundant filopodia and lamellipodia, but cells from HCC827 cell line were polygonal with marked nuclei in contrast to NCI-H1975 cells, which were more elongated and with more discreet nuclei. In terms of shape, SW900 cells resemble NCI-H1975 cells and NCI-H1650 resemble HCC827, although cells were smaller in both cases. Finally, some NCI-H2228 cells were giant, wedge-shaped and tend to form filopodia. Regarding tumorspheres, tight spheroids were formed by HCC827, NCI-H1395, NCI-H1650, NCI-H1975, NCI-H1993, NCI-H2228 and NCI-H358 cultures, whereas A549, NCI-H23, NCI-H460, PC9, and SW900 formed more loose and irregularly-shaped tumorspheres.

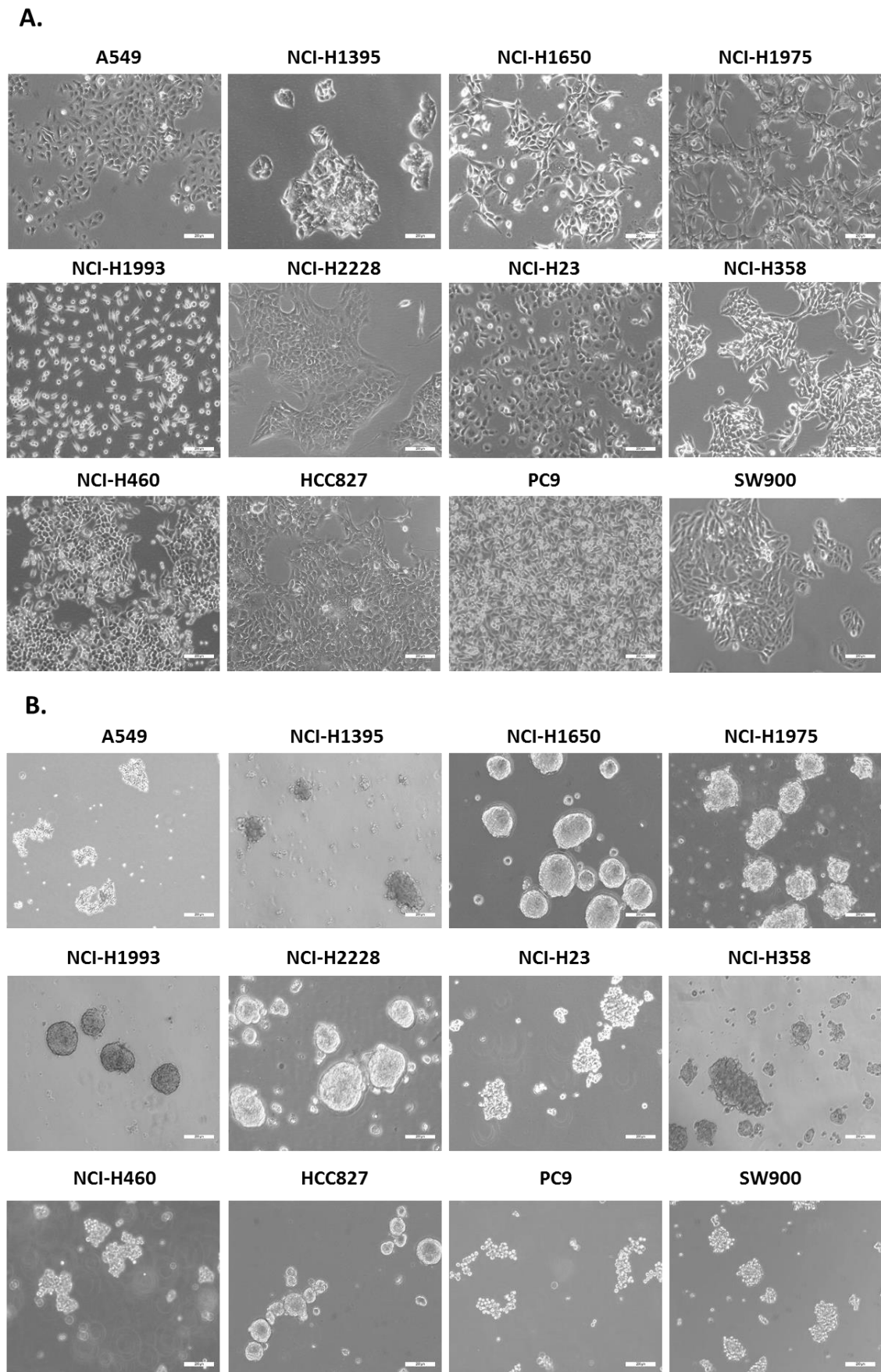


Figure 15. Representative images of the human non-small cell lung cancer cell lines included in the study cultured under (A) adherent and (B) suspension conditions.

1.1. PHENOTYPIC CHARACTERIZATION OF PRIMARY CULTURES

One of the aims of this study was to analyze if patient-derived tumorspheres display CSCs phenotype. For that purpose, multiple analyses were performed to determine their proliferation, chemoresistance, invasion and differentiation capacities *in vitro* and their tumor initiation capacity *in vivo* since these are fundamental features of this population.

1.1.1. SELF-RENEWAL, PROLIFERATION AND DIFFERENTIATION CAPACITIES

Self-renewal potential and growth characteristics of primary cultures were assessed *in vitro*. In order to do so, tumor cells were seeded at low density in suspension conditions and maintained for more than six months. Tumorspheres were formed and exhibited stable unlimited growth even in later passages (p>30), being able to continue growing limitlessly (Figure 16). As it is characteristic of CSCs, patient-derived tumorspheres had the potential to proliferate with minimal niche support indefinitely, being able to undergo cell division with maintenance of the undifferentiated state.

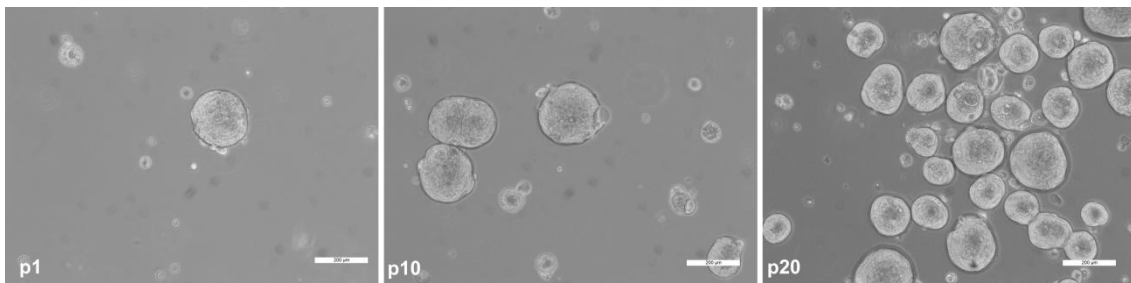


Figure 16. Self-renewal and unlimited exponential growth potential of tumor cells grown under suspension conditions. Tumorspheres displayed stable growth without declining in number. Images show the growth of FIS301 patient tumorspheres over time.

In turn, the growth of adherent-cultured cells was evaluated by calculating their growth curves. Tumor cells were plated at a density of 1,000 cells and cell viability was evaluated 24, 48 and 72h after seeding using MTS assay. FIS299 and FIS301 were slow growth cultures with low rates compared to the rest of primary cultures. In contrast, FIS303 showed the highest growth rate, whereas the growth rates of the rest of the cellular lines were as follows: FIS303 > FIS315 > FIS343 > FIS320 > FIS317 > FIS302 (Figure 17).

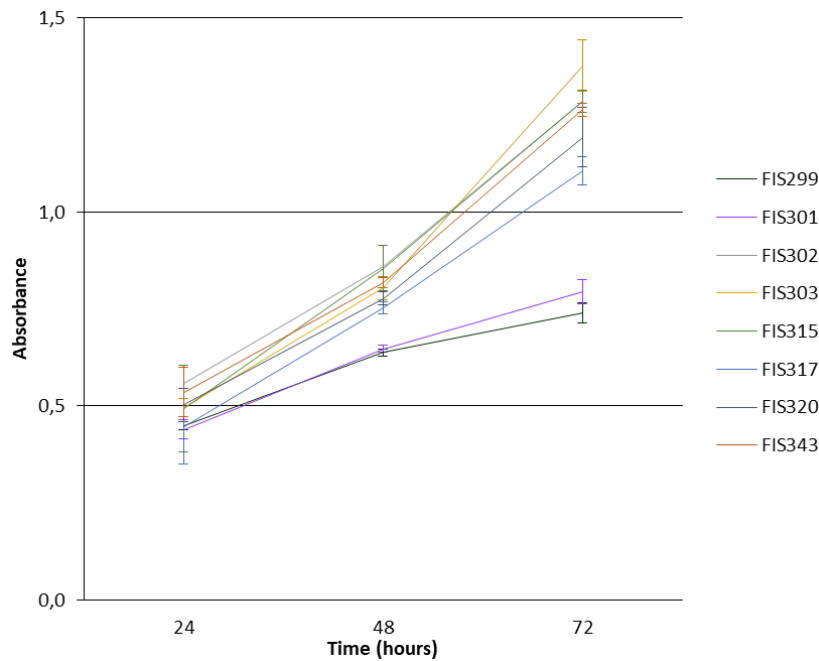


Figure 17. Cell growth curves from the primary adherent-cell cultures established in the study.

The differences observed in the growth rates of the primary cultures established were in consonance with results previously reported for lung cancer cell lines (298). For instance, NCI-H1395 was described as a slow growth cell line which do not reach a real exponential phase, whereas A549 cell line shows a remarkable exponential phase with a very fast growth (298).

Another characteristic of CSCs is their differentiation ability. CSCs cannot only form daughter CSCs by symmetric cell division, but also generate high proliferative bulk cancer populations by asymmetric cell division. To assess the differentiation potential of tumorspheres, we seeded them in conventional flask using serum-containing medium and found that tumorspheres were able to adhere and acquire the same morphology than their corresponding adherent-cultured cells (Figure 18). Moreover, further gene expression profiles showed no differences between adherent cells directly established from tissue and those established from tumorspheres, confirming that tumorspheres are able to adhere and differentiate, losing the expression of the stemness markers reported in the next section.

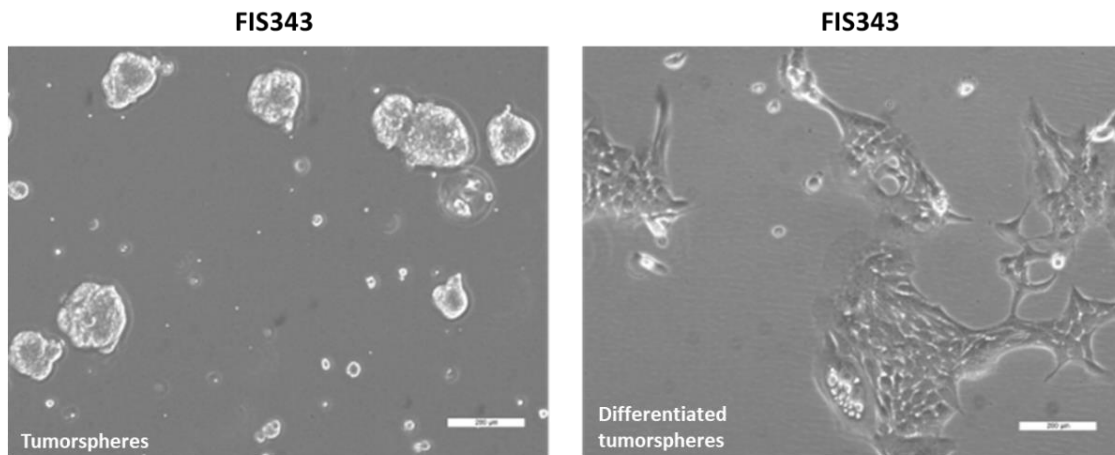
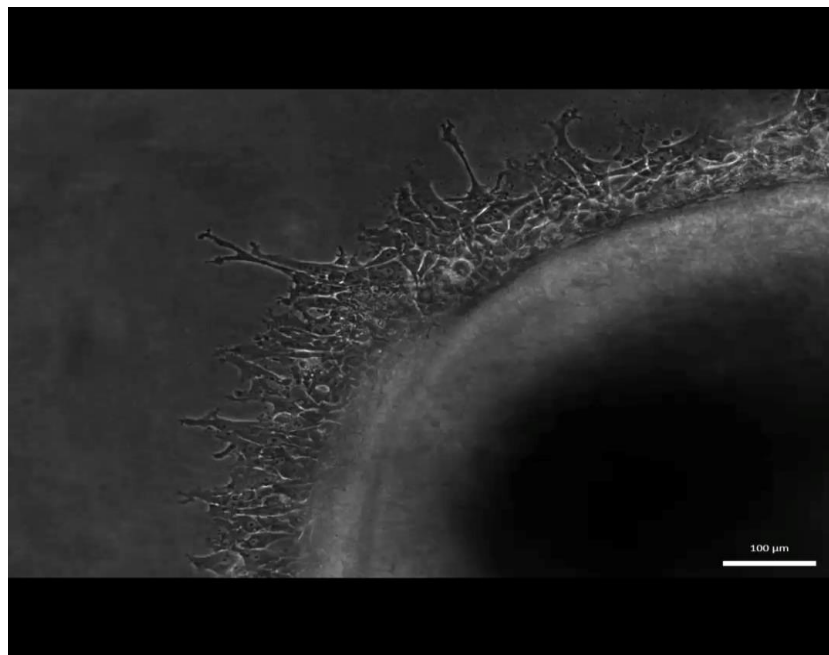


Figure 18. Differentiation capacity of lung tumorspheres. Under adherent culture conditions, tumorspheres adhere and acquire the morphologic features of the cells directly established from tumor tissue. Images show the growth of the FIS343 patient tumorspheres under suspension conditions and their differentiation capacity under adherent culture conditions.

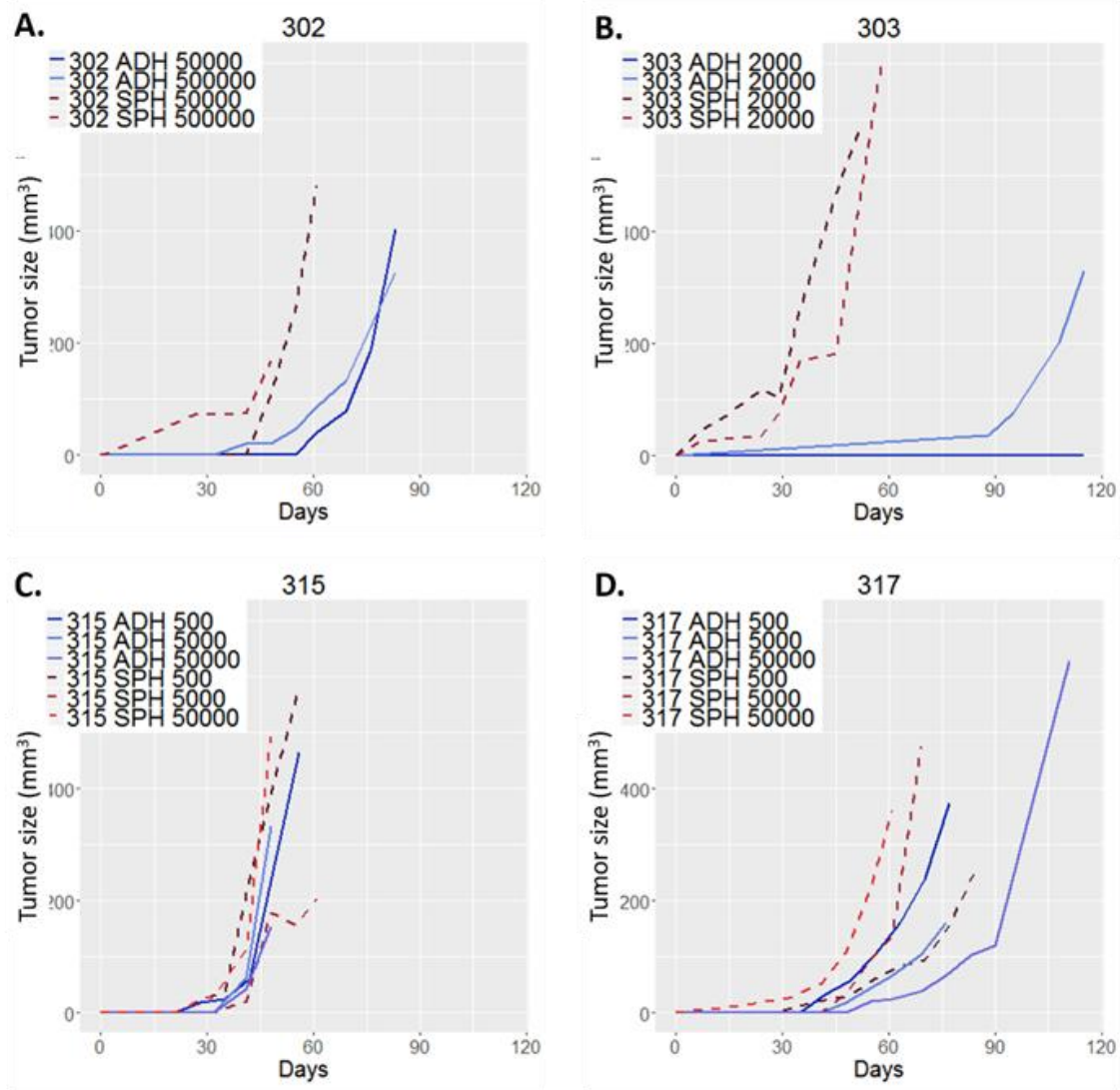
1.1.2. INVASION AND TUMOR INITIATION CAPACITIES

To determine the invasive and tumorigenic capacities of the primary cultures, tumorspheres were cultured in suspension conditions with 0.2% methylcellulose in a non-adhesive convex environment for 12 hours and then mixed with collagen matrix. Time-lapse video microscopy revealed the high invasive capacity of lung tumorspheres, being able to migrate through the collagen matrix over time (Video 1).



Video 1. Representative video of the invasion capacity of patient-derived lung tumorspheres *in vitro* through collagen matrix. Video available from: <https://www.nature.com/articles/s41419-019-1898-1>.

We also evaluated the ability of the tumorspheres and their corresponding adherent counterparts to develop tumors *in vivo* by subcutaneous transplantation of tumor cells into immunocompromised mice. Both, cells derived from tumorspheres and adherent cultures were able to initiate tumors, but a 100% success rate was only achieved by tumorspheres, and tumor latency was significantly higher in tumors induced by adherent cells (Figure 19).



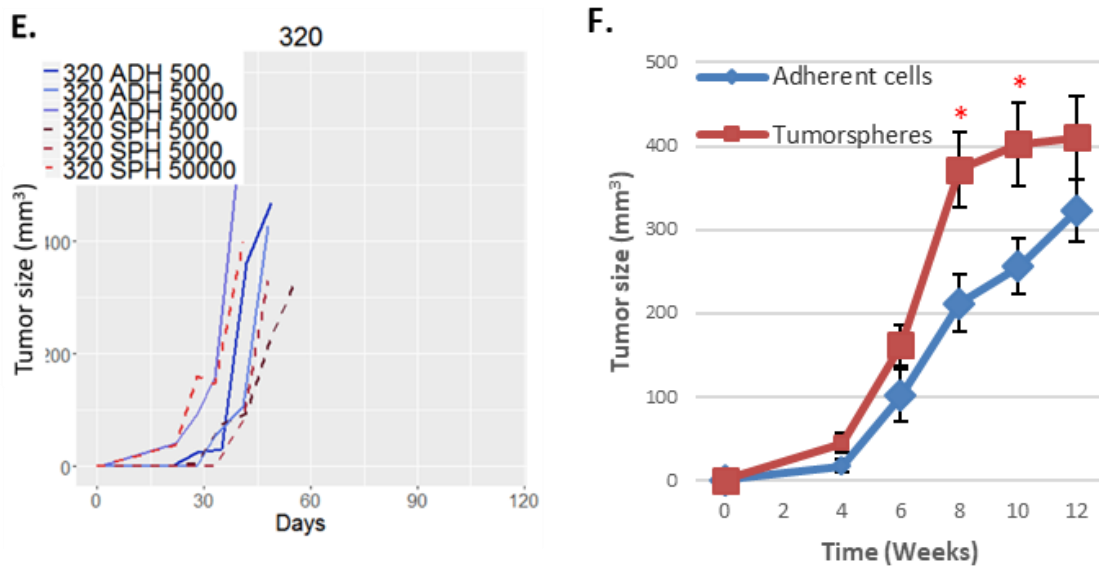


Figure 19. Tumor development capacity of adherent (ADH) cells and tumorspheres (SPH). A-E plots show the tumor growth variation in mice after injection of ADH and SPH from patients FIS302, FIS303, FIS315, FIS317, and FIS320 with different cell densities at the indicated time points. F plot shows the average tumor size after injection of ADH and SPH at the indicated time points. Error bars represent SEM. * $p < 0.01$.

1.1.3. RESISTANCE TO CLASSICAL CHEMOTHERAPEUTIC AGENTS

It is characteristic of CSCs to be highly resistant to the cancer treatments currently used in clinical practice, therefore, the resistance of adherent cells and tumorspheres to conventional chemotherapy was analyzed. Cell viability was assessed after 48 hours of exposure to cisplatin 50 μM , docetaxel 10 μM , paclitaxel 10 μM , vinorelbine 10 μM and pemetrexed 50 μM . Salinomycin, a selective agent against stem cells, was also tested at 1 μM . Figure 20 shows the results obtained for three patients (FIS302, FIS320, and FIS343) and three cell lines (HCC827, NCI-H1650, and PC9). Overall, tumorspheres were significantly more resistant to all the drugs tested, but salinomycin. Pemetrexed was the less effective drug, not reaching the IC₅₀ for any of the conditions. Tumorspheres were also more resistant to docetaxel and paclitaxel than their adherent counterparts, except for FIS343 tumorspheres, which show no significant differences between both culture conditions. The cytotoxic effect of cisplatin and vinorelbine was higher, especially against cells from HCC827 cell line. In contrast, salinomycin showed higher cytotoxic activity against tumorspheres with most of the cells dead after 48 hours of exposure to the compound.

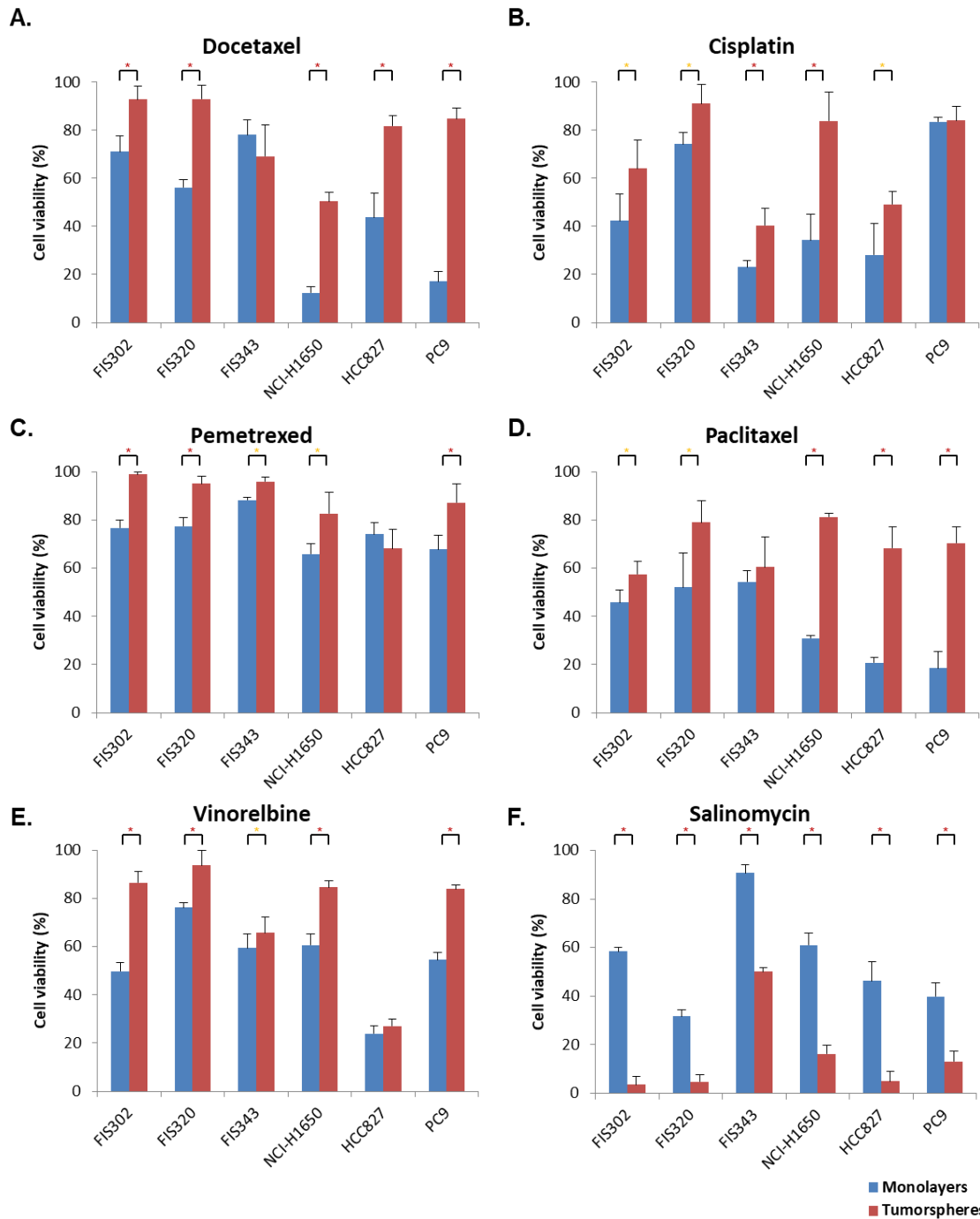


Figure 20. Cell viability of FIS302, FIS320, and FIS343 patients and NCI-H1650, PC9 and HCC827 cell lines after 48 hours of exposure to cisplatin 50 μ M, docetaxel 10 μ M, paclitaxel 10 μ M, vinorelbine 10 μ M, pemetrexed 50 μ M, and salinomycin 1 μ M. Cell viability of paired adherent cells and tumorspheres was statistically analyzed using non-parametric Wilcoxon signed-rank test. * $p < 0.05$. * $p < 0.01$.

All these results suggest that our lung tumorspheres were enriched in cells with stem-like properties. Patient-derived tumorspheres had self-renewal and unlimited exponential growth abilities, higher tumorigenic potential *in vivo* than adherent control cells, were able to differentiate and acquire the properties of their adherent counterparts, showed high invasion capacity, and were very resistant to high doses of chemotherapeutic agents. These results are in line with those previously reported on CSCs isolated from lung cancer (99,150,154) and confirm that this approach can be used for their enrichment in a simple and cost-effective way.

2. GENE EXPRESSION ANALYSIS

In addition to all the properties described, CSCs share with other stem populations the overexpression of stemness pathways, cytoprotective enzymes and efflux pumps (94,153,299). Different genes and molecules have been proposed as characteristic of lung CSCs, but most studies are focused on few cell lines and analyzed the expression of small groups of genes, being very challenging to determine which ones are characteristic of a particular cell line or patient and which ones are governing stemness in lung CSCs (103,120,135). For these reasons, we analyzed the gene expression profiles of this tumor population, trying to identify genes and molecules that could have a prognostic role or constitute the basis for developing novel therapies.

2.1. RNA QUANTIFICATION AND QUALITY ASSESSMENT

For relative gene expression analysis, RNA from cellular pellets was isolated using standard trizol extraction procedure. RNA concentration and quality was assessed using a nanospectrophotometer. Only the samples with an optimal concentration (≥ 40 ng/ μ L) and quality were included in the study. The mean RNA concentration for adherent-cultured cells was 1054.4 [171.2-2681.8] ng/ μ L and 889.5 [80.3-2131.2] ng/ μ L for tumorspheres. Regarding the quality of the RNA obtained, the ratios $A_{260/280}$ and $A_{260/230}$ showed values from 1.8 to 2.0 and from 2.0 to 2.2, respectively, meaning that all samples included in this study could be considered optimal for further analysis. Therefore, this

RNA extraction protocol from cellular pellets allows obtaining enough quantity of good quality RNA for gene expression analysis by qPCR.

2.2. RELATIVE mRNAs EXPRESSION

The relative expression of 50 genes described as potential lung CSC markers (*CD44*, *THY1/CD90*, *CD133*, *CD166/ALCAM*, *CD326/EPCAM1*, *LIN28B*), detoxifying molecules (*ABCG2*, *ALDH1A1*), pluripotency (*SOX2*, *NANOG*, *OCT4*, *KLF4*, *MYC*) and cell cycle (*BMI1*, *CCND1*, *CDKN1A*, *CDKN2A*, *MDM2*, *MUC1*, *WEE1*) regulators, invasion promoters (*CDH1*, *CEACAM5*, *ITGA2*, *ITGA6*, *ITGB1*, *JUNB*, *MMP2*, *MMP9*, *SNAI1*, *VIM*), and components of Notch (*DLL1*, *DLL4*, *NOTCH1*, *NOTCH2*, *NOTCH3*, *HES1*, *HEY1*, *JAG1*), Wnt (*CTNNB1*, *DKK1*, *FZD7*, *GSK3B*, *WNT1*, *WNT2*, *WNT3*, *WNT5A*) and Hedgehog (*GLI1*, *PTCH1*, *SHH*, *SMO*) signaling pathways was analyzed in tumorspheres and differentiated adherent cells from patient-derived cells and cell lines using RTqPCR. The relative expression levels of *LIN28B*, *CD133*, *WNT1*, *WNT2*, *SHH*, and *GLI1* were below the limit of detection of the technique in most samples and were excluded from the final analysis. Remarkably, the expression of *WNT1* and *CD133* could not be detected even when using two and three different sets of gene expression assays, respectively. As mentioned before, the efficiency of each TaqMan® was evaluated using the Cp slope method. In the Supplementary Table 1, a list of the slopes and efficiency values for each assay can be found. All the assays used in this study showed an amplification efficiency near to 100%.

The expression of five endogenous genes (*ACTB*, *GAPDH*, *GUSB*, *HPRT1*, and *CDKN1B*) was tested in all samples in order to establish the best internal control. For this purpose, we used GeNorm software (see materials and methods), which indicated that the most stable option was the *ACTB*, *GUSB*, and *CDKN1B* combination. Following the procedure proposed by Vandesompele et al., a normalization factor based on the expression of these three endogenous genes was calculated using the geometric mean (287). Finally, gene expression levels of each sample were examined, looking for abnormal distributions which could affect further analyses. As shown in Figure 21, all the samples showed similar distribution and, taking into account all the genes analyzed in the study, aberrant differences were not found in their transcriptomic levels.

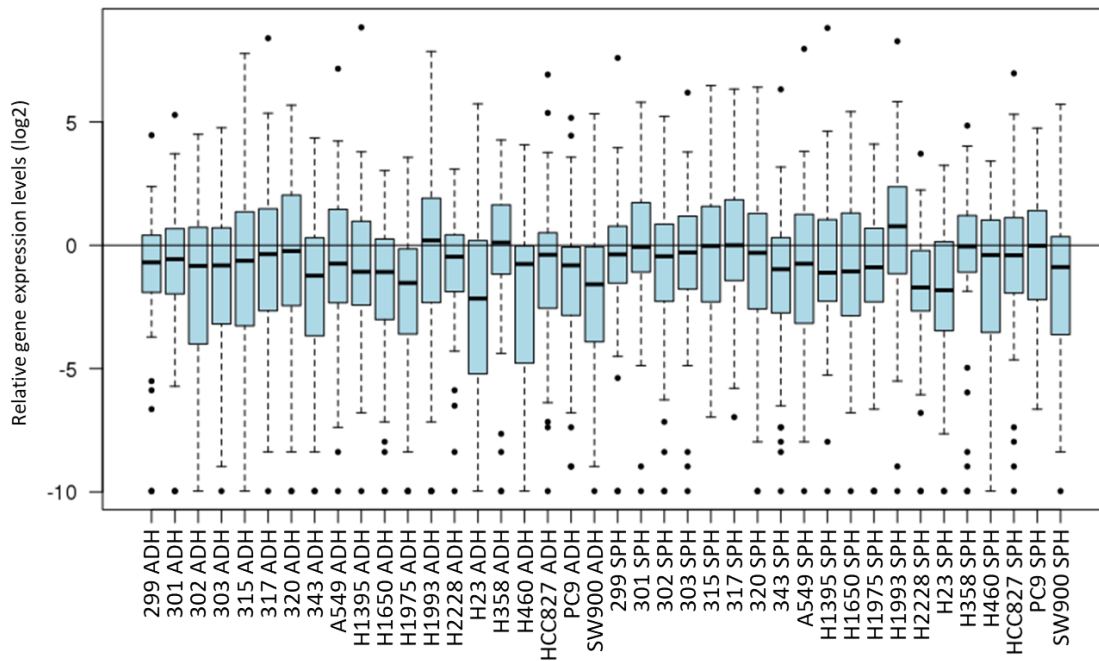


Figure 21. Analysis of the transcriptomic distribution of the samples included in this part of the study. The results shown are the log₂ of the transcription levels in 44 CSCs-related. Error bars represent SEM and dots represent outliers. mRNA was measured by RTqPCR.

We analyzed the relative gene expression pairing samples from both culture conditions in a Wilcoxon signed-rank test which included primary cultures and cell lines. Tumorspheres showed higher expression of 37 out of 44 genes compared to adherent-cultured cells, being a group of 17 genes (*ALDH1A1*, *KLF4*, *NANOG*, *CD44*, *THY1*, *CDKN1A*, *JUNB*, *MDM2*, *MMP9*, *SNAI1*, *ITGA6*, *NOTCH1*, *NOTCH3*, *DLL4*, *JAG1*, *CTNNB1*, and *GSK3B*) significantly overexpressed according to this test (Figure 22).

The gene encoding for the detoxifying molecule *ALDH1A1* was found to be the most overexpressed one. *ALDH1A1* is a NAD(P)⁺-dependent enzyme which catalyzes the oxidization of aldehydes into carboxylic acids, acting as drug-detoxifying and being responsible for therapeutic resistance (181). This enzyme has been proposed as a potential marker of CSCs in lung malignancies. For instance, Karimi-Busheri et al. proposed the overexpression of *ALDH1A1* along with *EPCAM1* as a signature of enriched CSCs in H460 NSCLC cell line (136). Moreover, isolated *ALDH*⁺ lung cancer cells have been reported to be highly tumorigenic and clonogenic as well as capable of self-renewal in contrast to their *ALDH*⁻ counterparts (185,186). We found the levels of the pluripotency genes *KLF4* and *NANOG* significantly elevated as well. The expression of these factors

are determinant for stem cell induction and maintenance, being included in cocktails to induce pluripotent stem cells (300,301) (Takahashi and Yamanaka 2006; Goldman 2008) and frequently related to tumor initiation, EMT induction, drug resistance and metastasis (302,303), which explains their overexpression in tumorspheres enriched in CSCs. Pointing the focus on the cell cycle regulators, *CDKN1A* and *MDM2* were significantly overexpressed in tumorspheres. Both molecules are described as p53 inhibitors, inducing a quiescent state when needed, preserving self-renewal potential and resisting cancer treatment against highly proliferative cells (304). *CDKN1A* encodes for p21, whose phosphorylation state changes under stressful conditions, determining its cellular location and modulating cell cycle (305–307). On the other hand, *MDM2* oncogene has been reported to enhance stemness, being required for the efficient generation of induced pluripotent stem cells (308). In that sense, its inhibition promotes cell apoptosis and differentiation (309), whereas its increased expression is associated with poor clinical outcome in NSCLC patients (310).

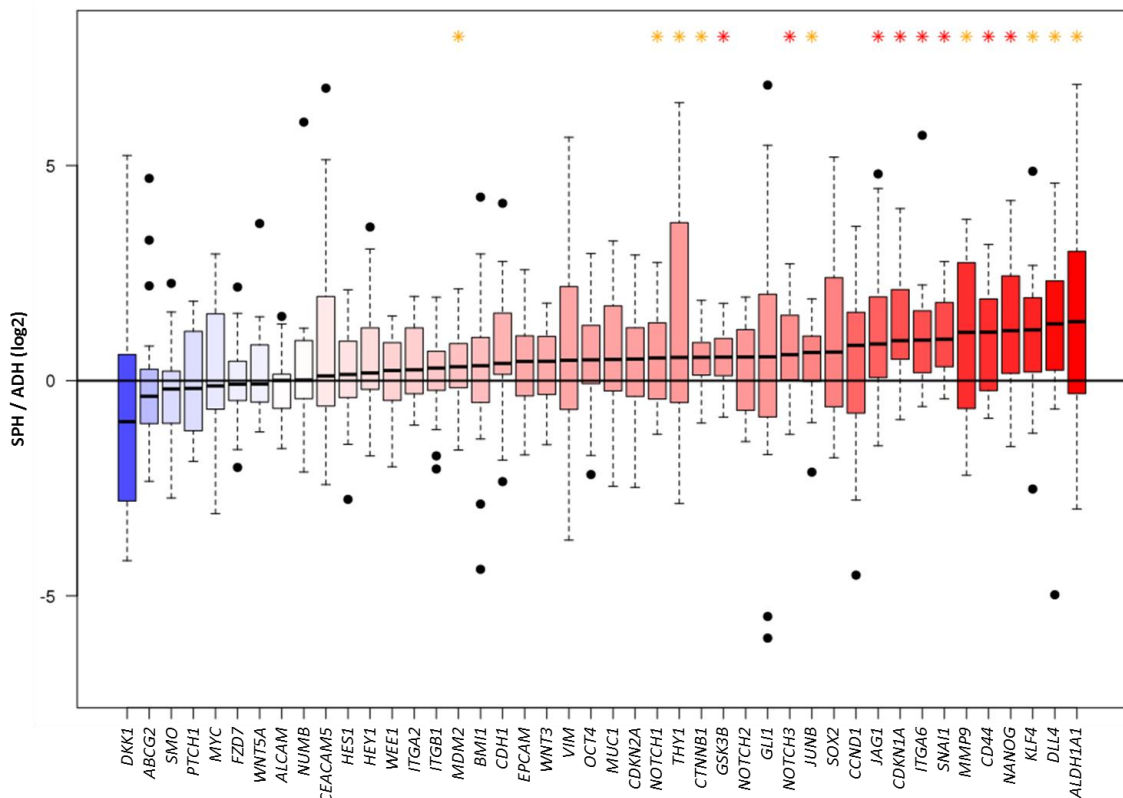


Figure 22. Transcription levels of the CSCs-related genes in tumorspheres versus adherent-cultured cells. The results shown are the log₂ of the ratio between the gene expression of tumorspheres and the gene expression of adherent-cultured cells. Error bars represent SEM and dots represent outliers. *p<0.05, *p<0.01. mRNA was measured by RTqPCR.

Regarding the CSCs markers analyzed in this study, we found the expression of *THY1* and *CD44* increased in tumorspheres. Although *THY1* (*CD90*) was suggested as a promising CSC marker for lung cancer, few data is available in patient specimens, and previous investigation had only associated its overexpression to the CSC phenotype in A549, NCI-H2122, NCI-H226 and NCI-H1437 cell lines (311). *CD44* has been also reported as a CSC marker in many solid tumors, including lung cancer, and functionally is involved in cell growth, survival, differentiation and migration through regulation of Wnt/ β -catenin pathway (312). The results shown here confirm the potential role of *CD44* in the identification of CSCs and are in line with the previous results obtained by our group, where three compounds which alter the expression of *CD44*, *ALCAM*, and *EPCAM1* triggering cell death and differentiation of ADC tumorspheres were identified (313). It is of note that we could not evaluate the expression of *CD133* even when using three different sets of assays. There are a number of publications that focused on *CD133* as a potential lung CSC marker (314–316). Nevertheless, some authors have reported that the use of *CD133* expression to discriminate lung CSCs can be overstated (101). For example, *CD133*⁻ lung cancer cells also possess self-renewal ability and can generate xenograft when transplanted into recipient mice (154). In addition, *CD133* expression in lung cancer was not associated with patient prognosis in some studies (317) and in many lung cancer samples, it could not be detected by qPCR (318–320), as it is reported in this study.

Our analyses also showed that tumorspheres have greater expression of the metastasis-related genes *ITGA6*, *SNAI1*, and *MMP9*. *ITGA6* is a member of the integrin family, which reprogram tumor cells to promote invasion and metastasis. Among these, integrin $\alpha 6$ (encoded by *ITGA6*) is the most widely observed enriching for CSCs in breast (321), prostate (322), and colorectal (323) cancers. Integrin $\alpha 6$ forms with integrin $\beta 4$ an extracellular matrix receptor implicated in carcinoma development (324). During cancer progression, the integrin $\alpha 6\beta 4$ is released from hemidesmosomes, activating RhoA, leading to membrane ruffling, lamellae formation, and enabling cell migration and invasion (325). In line with our results, integrin $\alpha 6\beta 4$ was previously proposed as a lung CSC marker associated with *CD44* expression, venous invasion, metastasis, and decreased OS (326,327). On the other hand, the overexpression of *SNAI1* has been

related with CSCs properties in a number of solid tumors such as thyroid (328), colorectal (329), head and neck (330), pancreatic (331), and lung (332) cancer. It is a known transcriptional repressor of *CDH1*, one of the key inducers of EMT and it was reported to correlate with *NANOG* expression and increased malignancy in mice models (333). In regard to *MMP9*, this gene encodes for a matrix metalloproteinase involved in extracellular matrix degradation and induction of both, tumor growth and metastasis, in lung cancer (334). As expected, increased expression of *MMP9* has been found in NSCLC patients and cell lines, especially in cells with enhanced migratory capacity like CSCs (335,336).

The genes encoding for Notch receptors (*NOTCH1* and *NOTCH3*) and ligands (*DLL4* and *JAG1*) were also overexpressed in tumorspheres, suggesting a possible activation of this pathway in lung CSCs. A possible role of Notch signaling in lung cancer was first suggested when the overexpression of Notch components was first detected in NSCLC correlating with poor prognosis (203). In CSCs, Hassan *et al.* demonstrated the decisive function of this signaling cascade for sphere formation and self-renewal *in vitro* and also for tumor initiation *in vivo* (216). *NOTCH1* overexpression has been significantly correlated with disease progression, metastasis and poorer prognosis of NSCLC patients (215), whereas targeting of Notch3 with Tarextumab inhibited tumor growth and decreased the tumor-initiating cell frequency in patient-derived xenografts (219). Concerning Notch ligands, ectopic expression of *JAG1* on lung cancer cells enhanced cell migration and invasion *in vitro* as well as metastasis *in vivo*, suggesting a potential role of this marker in lung cancer. In addition, a Phase IB trial targeting Delta-like 4 (encoded by *DLL4*) with Demcizumab in metastatic non-squamous NSCLC patients showed a 50% of objectives responses, recommending a phase II trial in combination with standard carboplatin and pemetrexed chemotherapy (337).

The overexpression detected in key genes of the Wnt pathway, *CTNNB1* and *GSK3B*, which encode for β -catenin and GSK-3 β , respectively, suggests an activation of this signaling network in lung CSCs and correlates with some findings already reported. Increased expression of *CTNNB1* has been associated with *OCT4* and *CCND1* overexpression and resistance to a number of chemotherapeutic drugs in lung CSCs

(249). Furthermore, knockdown of *CTNNB1* suppressed the metastatic potential of lung tumor xenografts (338). Regarding *GSK3B*, in the absence of Wnt proteins, GSK-3 β phosphorylates β -catenin, resulting in its ubiquitination and proteosomal degradation (242,339). However, when Wnt proteins are present, GSK-3 β functions remain unclear. Tivantinib has been found to target GSK-3 β and that pharmacological inhibition caused apoptosis in NSCLC cells (251). Additionally, in a lung xenograft model, astrocyte elevated gene-1 (AEG-1) behaved as a critical protein in the activation of EMT by directly targeting *GSK3B* (252).

2.3. UNSUPERVISED PRINCIPAL COMPONENT ANALYSIS

Next, an unsupervised principal component analysis (PCA) including adherent cells and tumorspheres from primary cultures and cell lines was performed in order to group samples according to gene expression. This analysis revealed that the adherent-cultured cell population is very homogeneous, being closely plotted, in comparison with tumorspheres, which were scattered along the score plot (Figure 23).

This suggest that tumorspheres can gain CSCs properties from the activation of multiple complementary pathways. As a result, principal component 1 (PC1) slightly separated some tumorspheres from adherent cells, so we decided to apply a supervised partial least square-discriminant analysis (PLS-DA) to better discriminate between tumorspheres and adherent-cultured cells. As shown in Figure 24A, PC1 separated CSCs from differentiated tumor cells. Tumorspheres from A549, HCC827, NCI-H2228 and NCI-H358 were the most similar to adherent cells in terms of gene expression, whereas tumorspheres from NCI-H1650, NCI-H1993 PC9 and patients 303, 315 and 317 differed the most. Loading plot (Figure 24B) revealed that the expression of *SNAI1*, *GSK3B*, *CD44*, *CDKN1A*, *NOTCH3*, *NANOG*, and *CTNNB1* genes in tumorspheres and *DKK1*, *SMO*, *ALCAM*, *PTCH1* and *FZD7* in monolayers contributed the most to this separation.

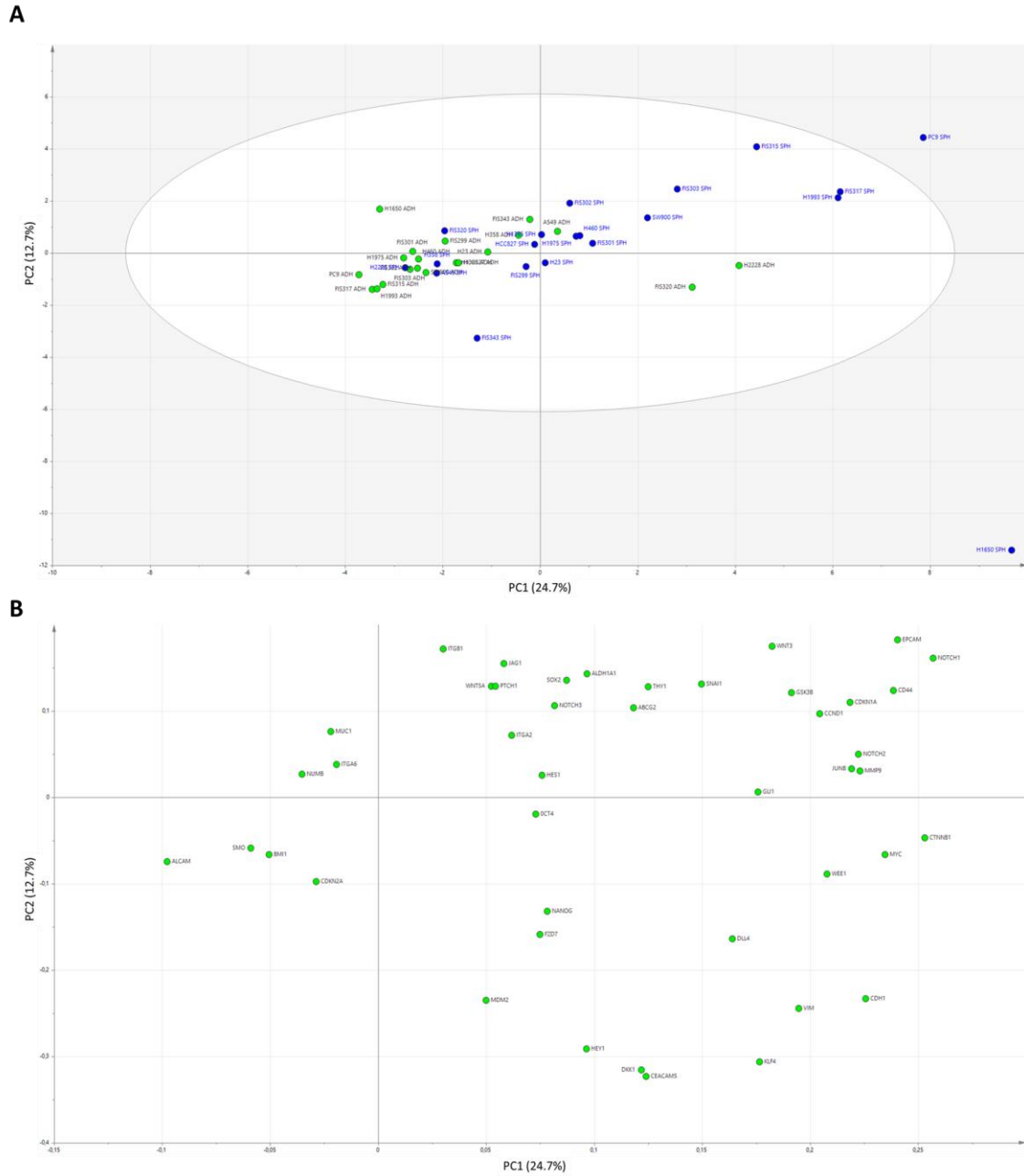


Figure 23. **A.** PCA score plot of primary cultures and cell lines based on the gene expression of 44 CSCs-related genes. Green dots represent adherent-cultured cells whereas blue dots represent tumorspheres. **B.** Loading plot of PCA analysis showing the contributing of each gene to culture distribution.

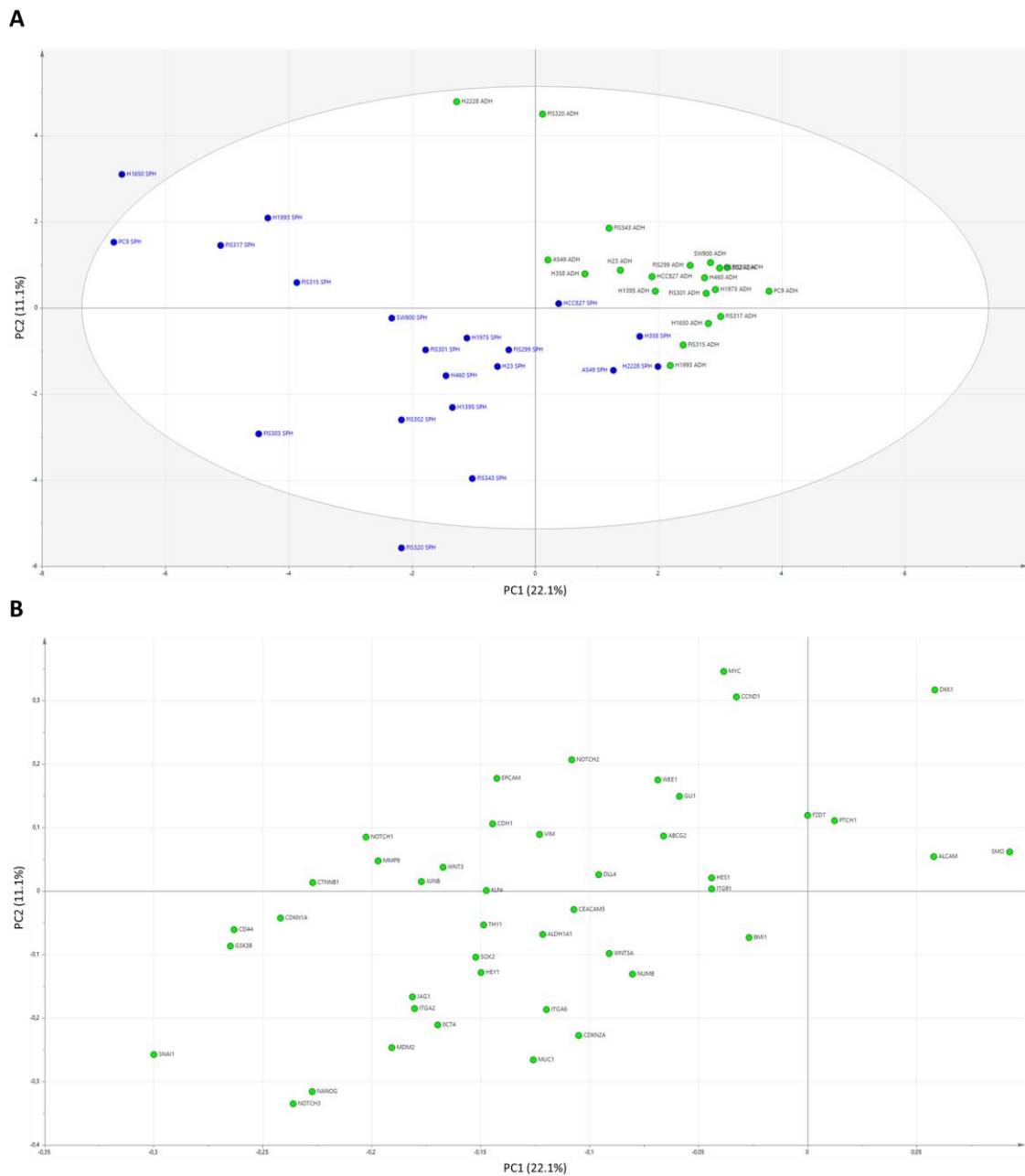


Figure 24. A. PLS-DA score plot of primary cultures and cell lines based on the gene expression of 44 CSCs-related genes. Green dots represent adherent-cultured cells whereas blue dots represent tumorspheres. **B.** Loading plot of PLS-DA analysis showing the contributing of each gene to culture distribution.

Based on the results obtained, it seems that other factors could be contributing more to sort samples than culture conditions. Thus, we postulated that most unwanted variation could be due to differences between immortalized cell lines and primary patient-derived cultures, so we decided to apply PCA analyses separately. The PCA score plot for cell lines exhibited the high variability between them (Figure 25A), being only able to separate NCI-H1650, NCI-H1975, NCI-H1993, NCI-H23, PC9, and SW900 tumorspheres from adherent cultures. In addition to *CDKN1A*, *CD44*, *GSK3B*, and *SNAI1*, these tumorspheres showed high expression levels of *EPCAM*, *NOTCH1*, *ALDH1A1*, and *WNT3* (Figure 25B). On the contrary, in the PCA score plot corresponding to the patient-derived cultures, PC2 clearly distinguished differentiated cells from lung tumorspheres, whereas PC1 separated FIS343 and FIS320 from the rest of primary cultures (Figure 26). When comparing the gene expression profiles, it was observed that the expression of *BMI1*, *CD166*, *CDKN2A*, *MDM2*, *HEY1*, *NUMB*, *ITGA6*, and *NOTCH3* was induced in tumorspheres from FIS343 and FIS320 patients, while tumorspheres from the rest of patients showed higher expression of *EPCAM*, *NOTCH1*, *NOTCH2*, *CD44*, *CTNNB1*, *MMP9*, and *CDKN1A*.

Sphere-forming assays are well-described culture methods that have been used for stem cells isolation, identification and enrichment from different tissues (76,79,80,98). Starting material for these cultures can also be commercial cell lines, but using cells directly isolated from surgical resections specimens reflects *in vivo* conditions better than they do, since immortal cell lines do not behave as primary cells and long-term manipulation alters phenotype, functions and responsiveness to stimuli. In addition, clinicopathological information from primary cultures can be correlated with each culture behavior, whereas this information is limited for most cell lines. However, in spite of not completely mimic primary cultures, cell lines are a powerful tool which offer several advantages over primary cells, triggering that most of the studies to date have been carried out on them with diverse, and sometimes conflicting, data reported. For instance, CD133 and CD44 molecules have been successfully used to identify lung cancer cells with CSCs properties in some studies (99,319), whereas other publications reported that CD133⁻ or CD44⁻ cell populations also possess the ability for self-renewal and enhanced tumor initiation capacity when transplanted into recipient mice

For these reasons, and based on PCA analyses, we decided to perform further analyses only on primary cultures. In order to reduce the dimensionality of our data and remove possible collinear expression of genes, we built a logistic regression model considering the 17 statistically significant overexpressed genes ($p < 0.05$, Wilcoxon signed-rank test). Based on this model, *CDKN1A*, *NOTCH3*, *CD44*, *ITGA6*, *NANOG*, and *SNAI1* were selected as the best contributors to separate CSCs from monolayers, so the expression of the proteins encoded by these genes was selected for further analysis.

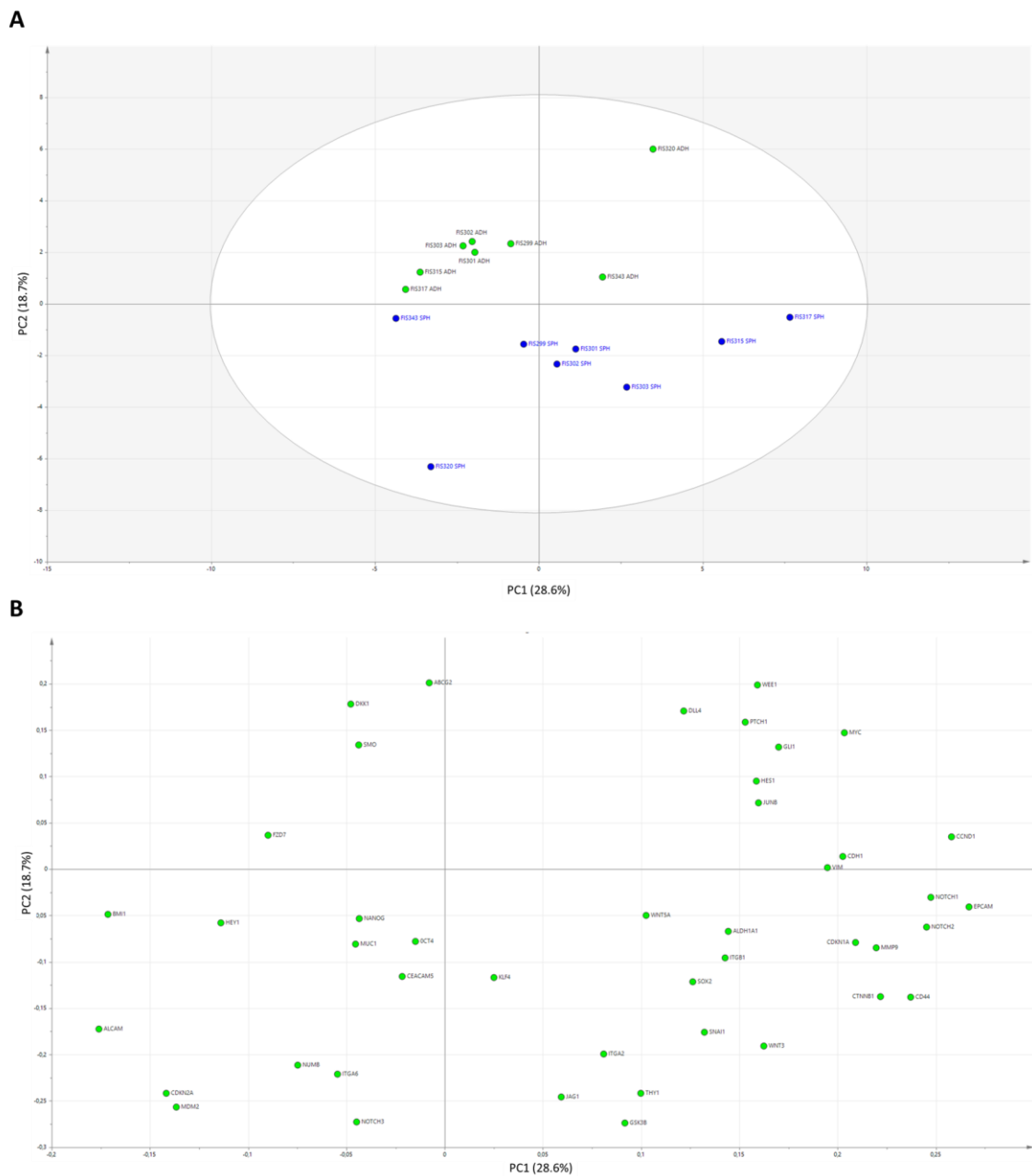


Figure 26. A. PCA score plot of primary cultures based on the gene expression of 44 CSCs-related genes. Green dots represent adherent-cultured cells whereas blue dots represent tumorspheres. B. Loading plot of PCA analysis showing the contributing of each gene to culture distribution.

3. PROTEIN EXPRESSION ANALYSIS

According to our gene expression results, the expression of six genes can be used to identify lung CSCs. We sought to determine if the overexpression found at the transcriptomic level could be detected at protein level as well, so the expression of p21, Notch3, CD44, Integrin $\alpha 6$, Nanog and Snail proteins encoded by *CDKN1A*, *NOTCH3*, *CD44*, *ITGA6*, *NANOG*, and *SNAI1* genes, respectively, was analyzed in the primary cultures established.

3.1. IMMUNOBLOTTING

Firstly, we performed immunoblotting analyses and found notable differences between ADC and SCC cultures. All protein expressions were significantly higher in tumorspheres from lung ADC patients than in their adherent counterparts. Only one patient (FIS320) showed higher levels of Integrin $\alpha 6$ and Snail in the adherent-cultured cells than in tumorspheres (Figure 27).

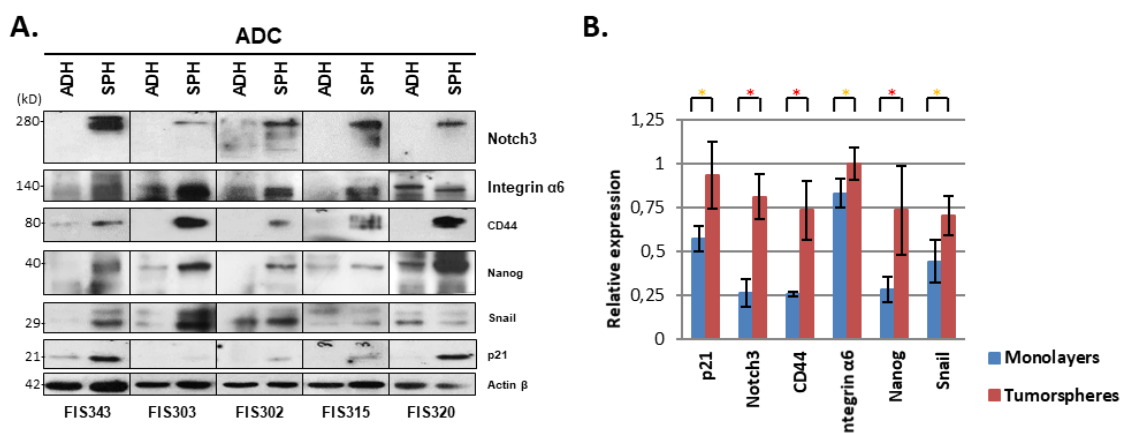


Figure 27. A. Representative immunoblotting images of p21, Notch3, CD44, Integrin $\alpha 6$, Nanog and Snail in adherent-cultured cells and tumorspheres from ADC patients. B. Relative protein expression of the same proteins in adherent-cultured cells and tumorspheres from ADC patients.

In contrast, variability was greater for SCC patients. No significant differences were found for p21, CD44 and Integrin $\alpha 6$, whereas higher expression of Notch3, Nanog and Snail were detected in adherent-cells in this histological subtype, suggesting that diverse molecular alterations govern CSCs in this tumor subtype (Figure 28).

In consonance with these notable changes we identified between CSCs from different histological subtypes, diverse and sometimes opposed roles have been described for some of these genes in ADC and SCC. A previous study performed on lung cancer cell lines reported that Notch3 expression was induced in tumorspheres from lung ADC, but not in lung SCC ones (340). In this study, Notch3 was found to regulate asymmetric cell division in lung ADC and that it was necessary for tumorigenic behavior *in vitro* and *in vivo* (340). Another study analyzed Notch3 gene expression in a cohort of 97 lung cancer patients including ADC and SCC and concluded that *NOTCH3* was associated with poor prognosis only in ADC (341). Our findings confirm these differences described in tumorspheres derived from cell lines and expand those results to primary cells. Regarding Nanog, elevated levels of this pluripotency regulator have been found in lung CSCs, showing a positive correlation with different CSCs markers, and association with poor prognosis (123,302). Interestingly, Park et al. found, in a retrospective study including 368 NSCLC patients, that Nanog expression is a negative independent factor associated with poor prognosis in lung ADC, but not in lung SCC (342). These findings suggest once again that there are significant differences between NSCLC subtypes and that markers to identify CSCs are tissue-dependent and different from lung cancer subtypes. Similarly, a study performed on 341 lung cancer specimens reported that the overexpression of Snail determined by tissue microarray analysis significantly decrease survival of lung ADC patients whereas no significant differences in prognosis were found for SCC (343). However, in our study, one ADC patient (FIS320) showed higher expression of Snail in adherent cells than in their corresponding tumorspheres. A recent study reported that Snail was upregulated in differentiated cells compared to tumorspheres from one lung ADC patient as well, concluding that this could challenge the association of CSCs with EMT (296). We found this situation only in one out of six cases, so it is difficult to estimate its frequency and implications.

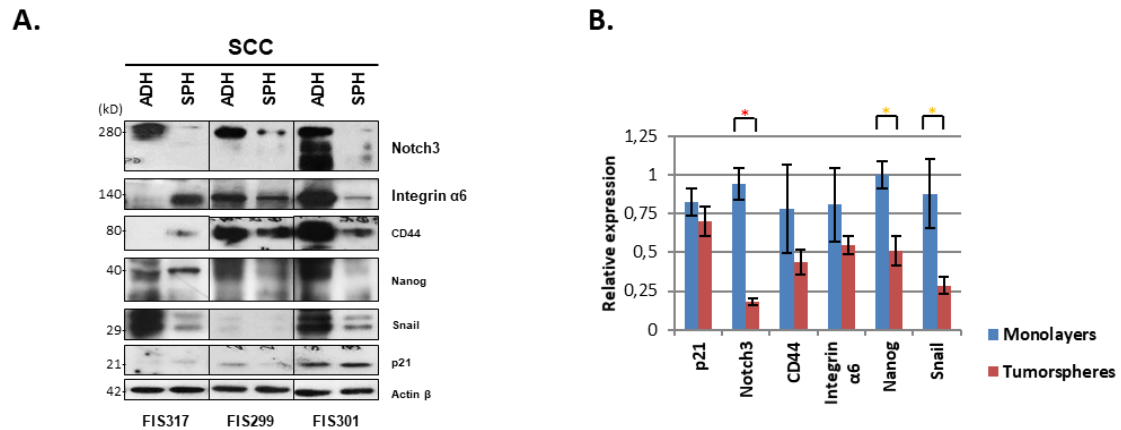


Figure 28. A. Representative immunoblotting images of p21, Notch3, CD44, Integrin α 6, Nanog and Snail in adherent-cultured cells and tumorspheres from SCC patients. **B.** Relative protein expression of the same proteins in adherent-cultured cells and tumorspheres from SCC patients.

Thus, a clearer expression pattern was obtained for ADC than for SCC, probably because of the small sample size, since only 1 out of the 12 lung cancer cell lines and 3 out of 8 patients were lung SCC in the gene expression analysis where these six molecules were selected. Complementary immunoblot analyses were carried out, analyzing the expression of E-cadherin, β -catenin, Vimentin, ALDH1A1, EpCAM and Sox2 because of their frequently reported implications in the CSCs biology (Figure 29). As expected, E-cadherin was higher expressed in adherent cells from all the primary cultures, but FIS303 and FIS299. In contrast, Vimentin was only detected in tumorspheres from FIS343. B-catenin, EpCAM and Sox2 were generally overexpressed in tumorspheres, whereas ALDH1A1 showed a more heterogeneous expression pattern. Sox2 was the only molecule analyzed to show overexpression in all SCC tumorspheres, suggesting this molecule as a potential CSC marker for this histological subtype.

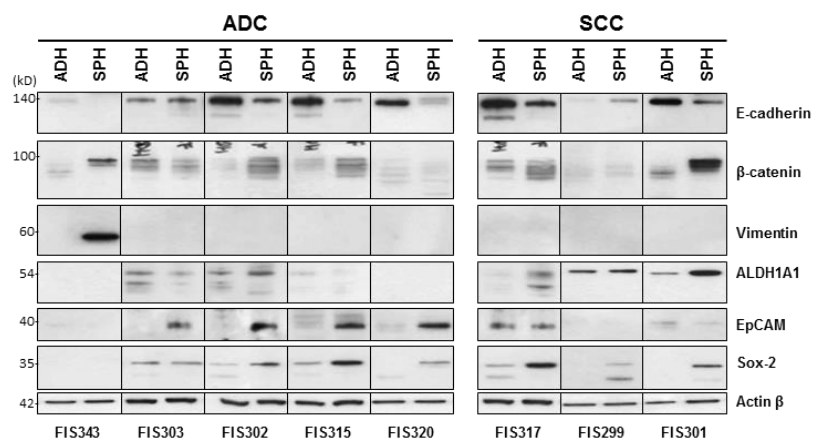


Figure 29. A. Representative immunoblotting images of E-cadherin, β -catenin, Vimentin, ALDH1A1, EpCAM and Sox2 in adherent-cultured cells and tumorspheres from primary cultures.

3.2. IMMUNOFLUORESCENCE

In addition to the western blotting, we performed immunofluorescence (IF) of these proteins in order to analyze their localization patterns in both culture conditions from ADC patients (Figure 30). The secondary antibody control can be found in Supplementary Figure 1.

Interestingly, IF showed differential subcellular localization of p21, which was nuclear and cytoplasmic in cells forming lung tumorspheres and only nuclear in adherent tumor cells. Phosphorylation of p21 by AKT1 has been reported to induce p21 cytoplasmic accumulation, where it binds to and inhibits the activity of proteins directly involved in the induction of apoptosis (305). Consequently, cytoplasmic p21 correlates positively with aggressive tumors and poor prognosis (306,307). Notch3 expression was observed in the cytoplasmic membrane and nuclei in both adherent cells and lung tumorspheres, although the localization was preferentially cytoplasmic for tumorspheres and nuclear for adherent cells. Notch3 is a cytoplasmic membrane receptor from the Notch family, which has been identified as a key driver of stemness. Particularly, overexpression of Notch3 receptor has been reported in chemoresistant tumors and related to poor prognosis in NSCLC (344). Therefore, the higher cytoplasmic detection of this component of Notch pathway maybe related to the aggressive phenotype characteristic of CSCs. In addition to Notch3, all cells forming tumorspheres expressed both CD44 and Nanog, although signals showed polarity, being higher on cell membranes of cells located in the periphery of the lung tumorspheres for CD44 and nuclear for Nanog. In contrast, the expression of Nanog and CD44 was notably lower in adherent cells. As previously discussed, CD44 is a potential stem cell marker frequently overexpressed in lung CSCs. Elevated levels of the pluripotency regulator Nanog have been found in lung CSCs, showing a positive correlation with different CSCs markers, and association with poor prognosis (123). Moreover, clear correlations between Nanog and lung CSCs features, including self-renewal, EMT induction, drug resistance, and metastasis have been reported (302). As expected, Integrin $\alpha 6$ showed similar expression pattern to CD44, with higher expression in cytoplasm, especially in the cell membrane. Lastly, Snail was overexpressed in tumorspheres and showed a non-uniform nuclear location along

them. In differentiated cells, a more homogeneous expression was detected in nuclei, which was weaker than that observed on their corresponding tumorspheres. Snail is a zinc finger protein which acts as a transcriptional repressor of E-cadherin and as a key regulator of EMT. It was reported that Snail overexpression induces CSCs properties in lung cancer as well as regulating Nanog status during EMT via the Sma1/AKT/GSK3 β signaling pathway (332). Furthermore, Snail upregulation was found on high-grade but not low-grade NSCLC tumors (333).

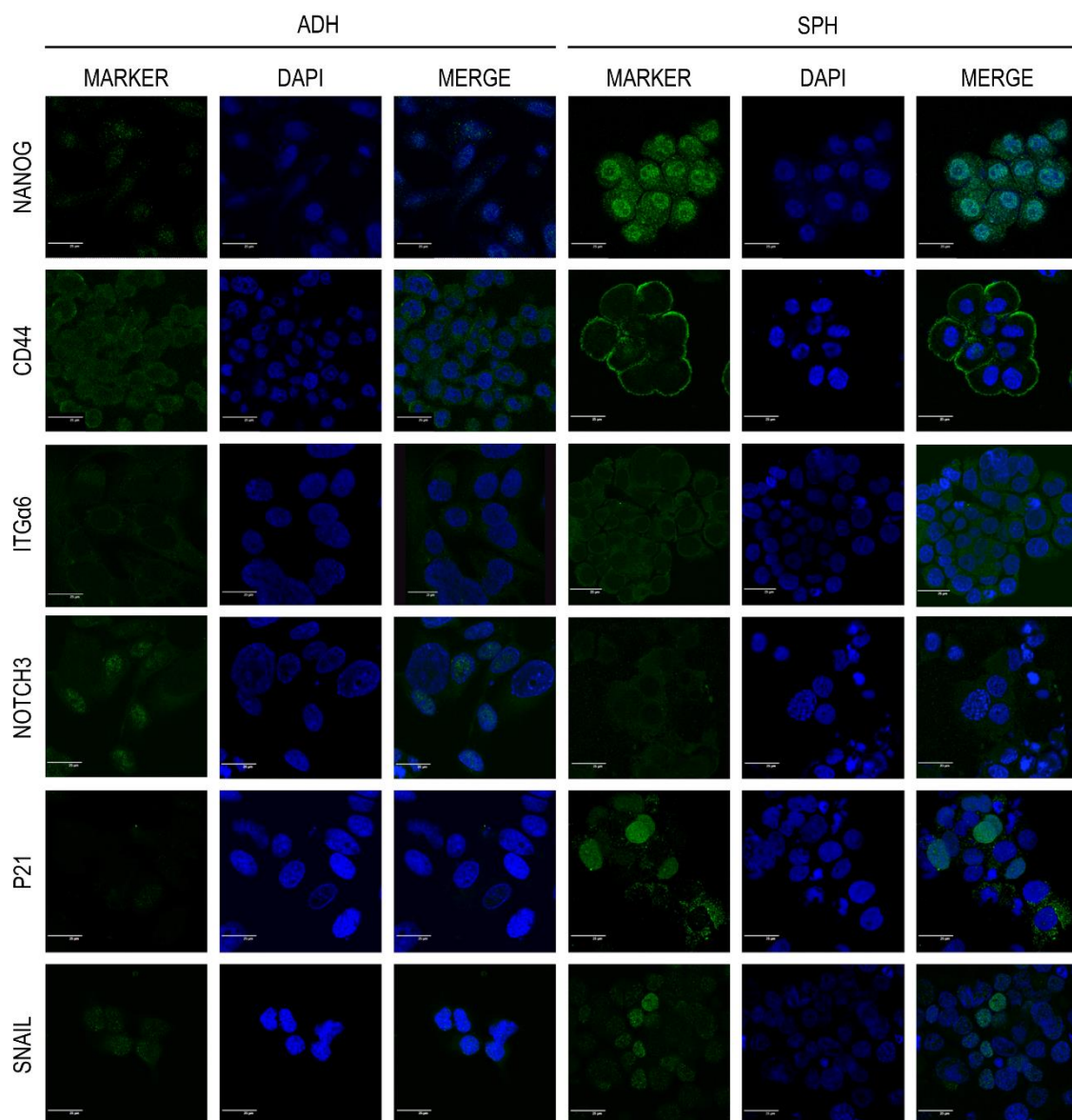


Figure 30. Representative immunofluorescence images of p21, Notch3, CD44, Integrin α 6, Nanog and Snail in adherent-cultured cells and tumorspheres from ADC patients. Green channel shows the indicated antibody staining, blue channel shows DAPI staining, and merge shows all channels merged. Scale bar represents 25 μ m.

The results obtained from the immunoblotting and IF analyses are in line with the mRNA expression and permitted to differentiate the CSCs population according to the tumor histology, so we concluded that the expression of p21, Notch3, CD44, Integrin α 6, Nanog and Snail can be used to identify CSCs populations from lung ADC. Complementary IF analyses of other potential CSCs-related proteins were also performed in adherent-cultured cells and tumorspheres (Figure 31). Interestingly, the expression of CD133 could be detected by IF and was generally overexpressed in tumorspheres. The expression of CD166 was also elevated in tumorspheres from patients FIS302, FIS303, and FIS315 compared to their adherent counterparts, but barely detectable in FIS317 and FIS320. Expression patterns were more heterogeneous for the rest of proteins analyzed. ALDH1A1 and β -catenin were overexpressed in tumorspheres from patients FIS315 and FIS317, whereas few differences between them and the adherent-cultured cells were found for the other primary cultures. Expression was generally weak for E-cadherin, while Vimentin expression was notably stronger in FIS303 and FIS320 tumorspheres compared to their corresponding adherent cells. No other remarkable differences were found between both culture conditions.

In addition, although tumorspheres from FIS320 and FIS343 shared the features characteristic of CSCs (self-renewal capacity, anchorage-independent growth ability or enhanced drug resistance, among others) mechanisms governing them seem to be different from the rest of primary cultures based on their expression profiles. We hypothesize that the molecular changes that conferred stem-like properties to CSCs from FIS343 and FIS320 could be different from those that conferred these features to CSCs from the other analyzed patients. As a result, tumorspheres from FIS320 and FIS343 showed higher expression of *NOTCH3*, *NUMB*, *HEY1*, *BMI1*, *CDKN2A* and *MDM2*. In contrast, tumorspheres derived from the rest of patient tumors showed higher expression of *NOTCH1*, *NOTCH2*, *CTNNB1*, *MMP9*, *CD44* and *CDKN1A*. In turn, this would also explain the differences noticed at the protein level. However, we found that there are significant changes between the CSCs population from different histological subtypes, so we decided to check if any histological transformation of these cultures could have taken place from ADC to SCC or a poorly differentiated neuroendocrine phenotype like SCLC.

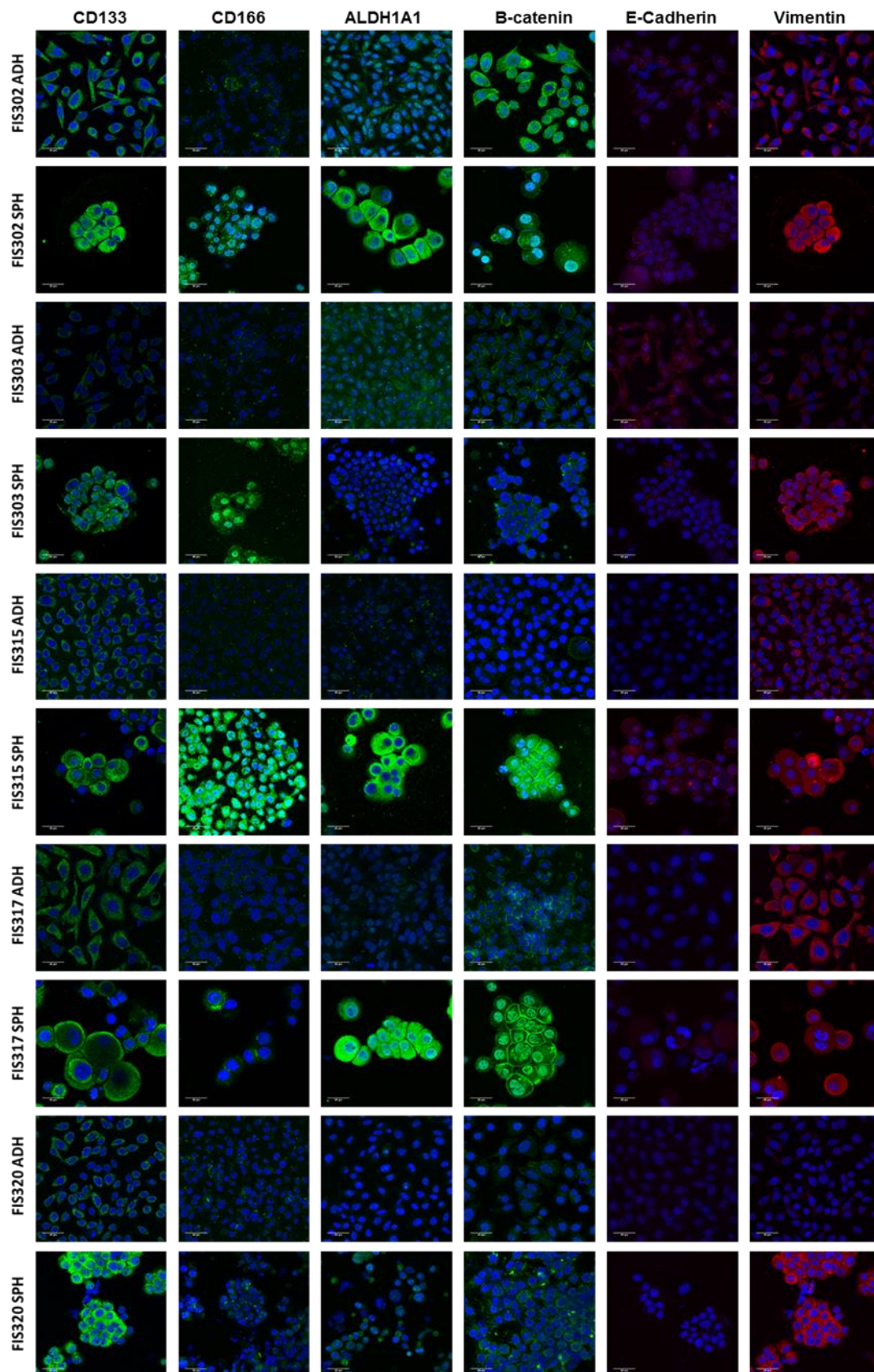


Figure 31. Representative immunofluorescence images of CD133, CD166, ALDH1A1, B-catenin, E-cadherin and Vimentin in adherent-cultured cells and tumorspheres from FIS302, FIS303, FIS315, FIS317, and FIS320 patients. Green channel shows the indicated antibody staining, blue channel shows DAPI staining, and merge shows all channels merged. Scale bar represents 25 μm.

3.3. IMMUNOHISTOCHEMISTRY

To test the histological subtype of FIS320 and FIS343 primary cultures, we analyzed the expression of common immunohistochemical markers used in the diagnosis of lung tumors: carcinoembryonic antigen (CEA), cytokeratin 7, p63, cytokeratins 5/6 and CD56. SCC samples are frequently positive for cytokeratins 5/6 and p63, whereas ADC samples are frequently positive for cytokeratin 7, and neuroendocrine samples are commonly positive for CD56. Images of the immunohistochemical staining can be found in Figure 32.

FIS320 was positive for CEA, cytokeratins 5/6 and cytokeratin 7 and negative for CD56, whereas FIS 343 was positive for cytokeratin 7 and CEA. In both cases, primary cultures showed the histological characteristics described in the pathological report from the primary tumor tissue. Patient FIS320 was diagnosed of a moderately differentiated adenocarcinoma with tumor cells positive for cytokeratin 7 in concordance with the positive staining detected in adherent cells and tumorspheres. Patient FIS343 was diagnosed of a poorly differentiated lung adenocarcinoma with signet ring cells, which can be identified in the primary culture (Figure 33). These results confirm that the differences in expression that were detected between the ADC tumorspheres are not due to changes in the histological subtype and suggest that different molecules and pathways can be activated to confer CSCs properties to tumor cells.

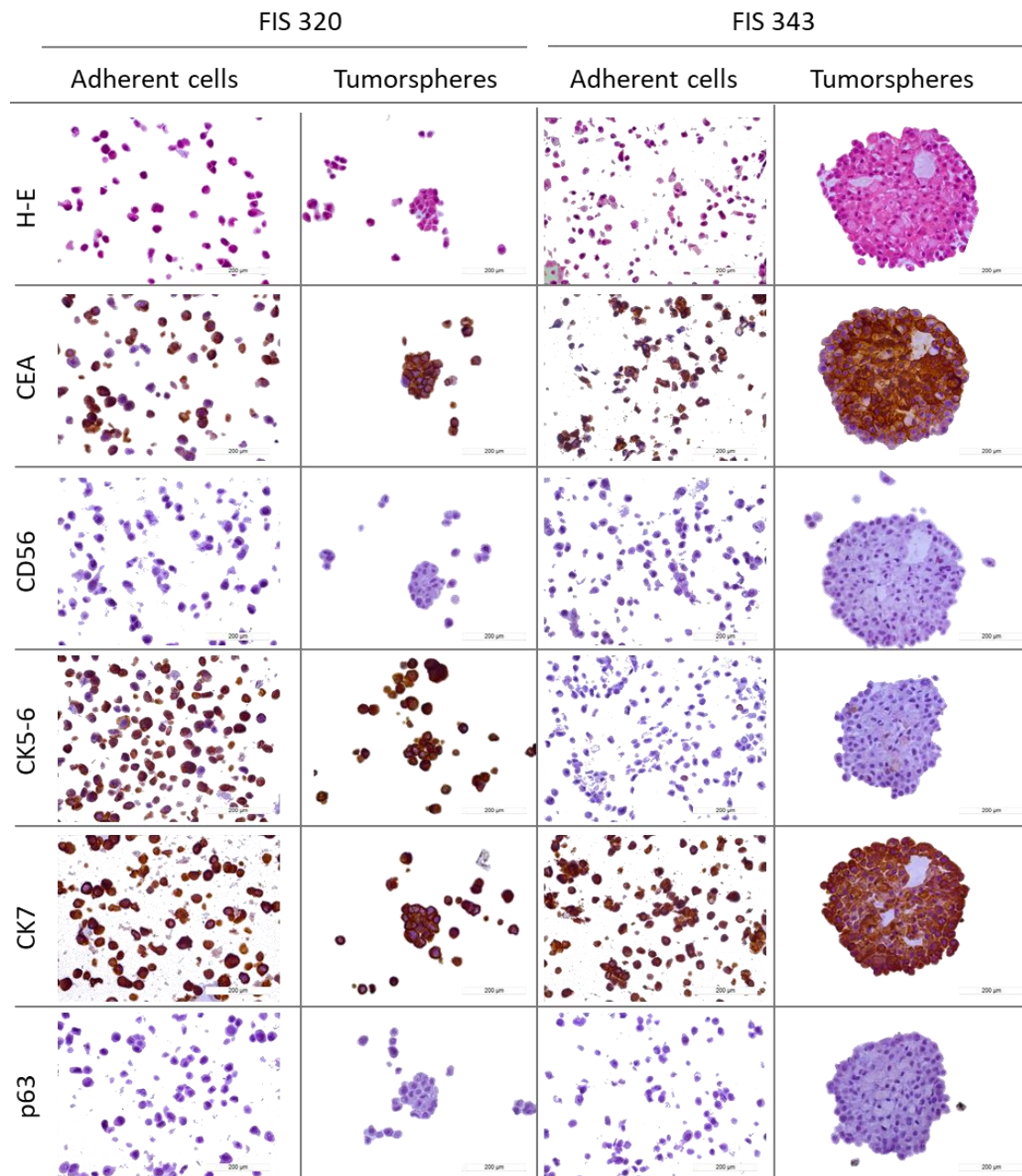


Figure 32. Representative immunohistochemical staining for CEA, CD56, CK5-6, CK7 and p63 in tumorspheres and adherent cells from FIS320 and FIS343 primary cultures. 40X magnification. H-E: show hematoxylin and eosin staining. CEA: carcinoembryonic antigen; CK: cytokeratin.

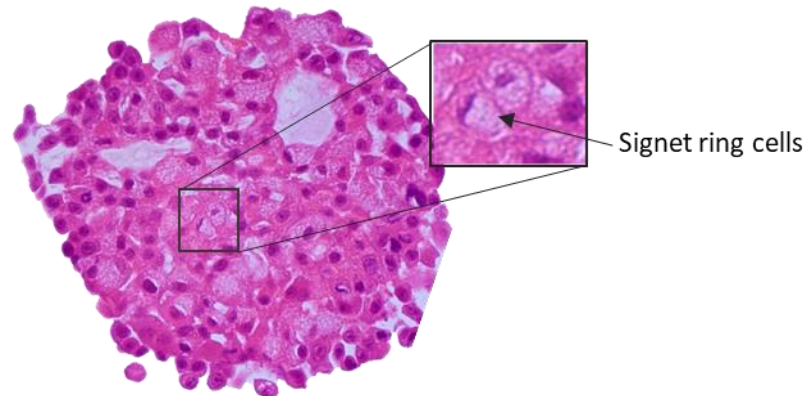


Figure 33. Hematoxylin and eosin staining of a tumorsphere from FIS343 patient. This patient was diagnosed of a poorly differentiated lung adenocarcinoma with signet ring cells in the tumor tissue that can be found in the established cultures.

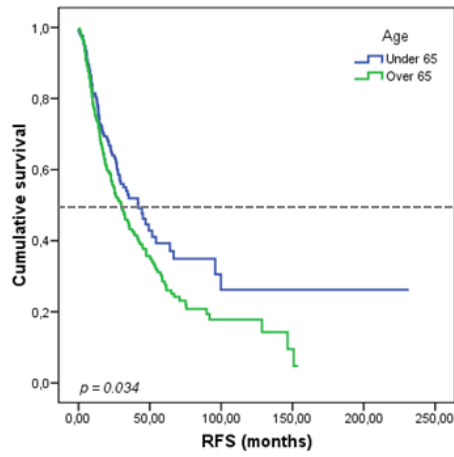
4. *IN SILICO* SURVIVAL ANALYSIS

In order to determine the prognostic implications of the CSC population and to identify potential targets against this tumor population, we correlated the expression of the genes significantly overexpressed in tumorspheres with the survival of an *in silico* cohort of resected NSCLC patients from TCGA and validated the results in a second cohort of resected patients from Hospital General Universitario de Valencia.

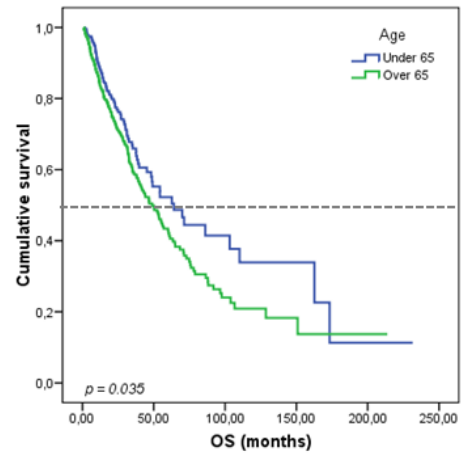
4.1. CLINICOPATHOLOGICAL VARIABLES

Of the 661 resected NSCLC patients from TCGA included in this part of the study, 208 (31.5%) relapsed and 261 (39.5%) died during the follow-up. However, relapse information was not available for 59 (8.9%) patients. The median follow-up was 23.08 months [range: 1.02-231.54]. The prognostic value of the different clinicopathological variables was assessed using the univariate Cox regression method for RFS and OS, and are shown in Table 18 along with the hazard ratios and *p*-value for each variable. Significant results obtained from the univariate Cox regression method were also analyzed using the Kaplan-Meier method (log-rank) in order to obtain the survival plots. This univariate analysis showed that patients over 65, with large tumors, lymph node (LN) involvement, or more advanced stage had shorter RFS and worse OS (Figure 34), which agrees with previously published results (345,346).

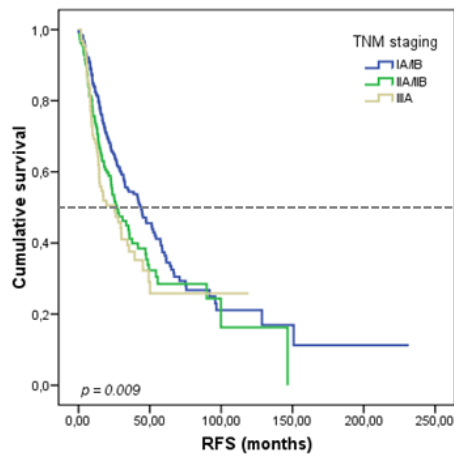
A.



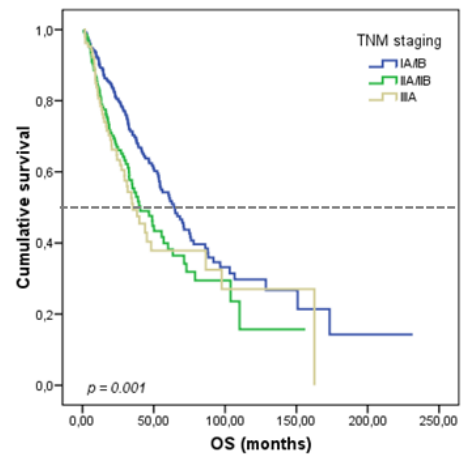
B.



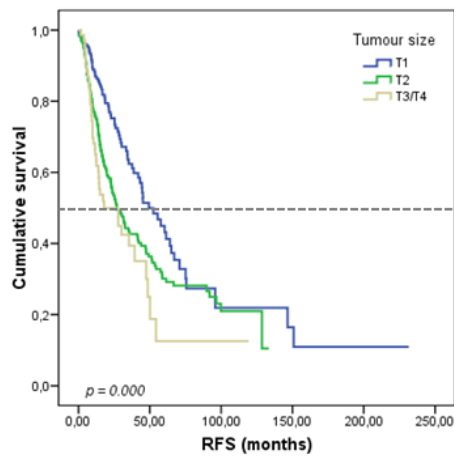
C.



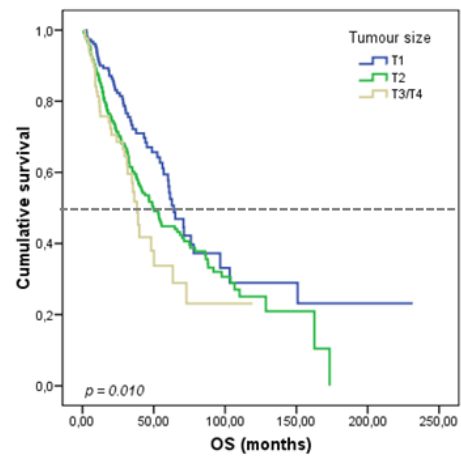
D.



E.



F.



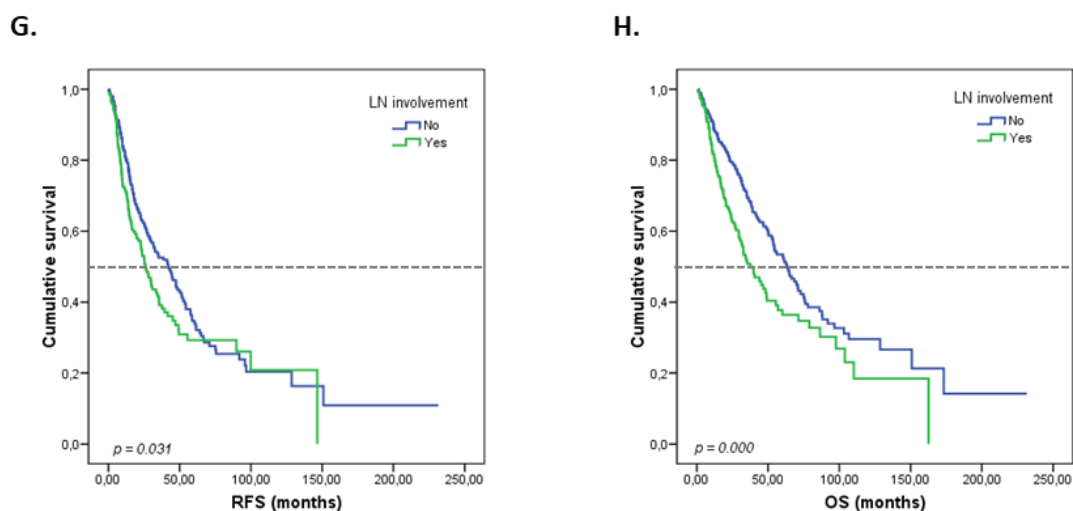


Figure 34. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the entire cohort. A-B. Age; C-D. TNM staging; E-F. Tumor size; G-H. LN involvement. *P*-values from the Kaplan-Meier test. LN: lymph node.

The prognostic value of the clinicopathological variables was also assessed according to the tumor histology. The ADC subcohort comprised 345 patients, 123 (35.7%) of whom relapsed and 114 (33.0%) who died. Again, in the univariate analysis, TNM staging, tumor size and LN involvement were associated with RFS and OS, whereas age was significantly associated with RFS ($p = 0.035$; Table 19). Survival plots from Kaplan-Meier survival analysis are represented in Figure 35.

Table 18. Results for survival analysis based on clinicopathological variables for the entire cohort.

Variable	RFS			OS		
	HR	95% CI	<i>p</i> -value	HR	95% CI	<i>p</i> -value
Gender						
Male vs. Female	0.859	0.679-1.088	0.208	0.798	0.619-1.029	0.081
Age						
> 65 vs. ≤ 65	1.305	1.020-1.671	0.034*	1.327	1.019-1.727	0.036*
TNM staging						
III vs. II vs. I	1.252	1.081-1.450	0.003*	1.312	1.125-1.529	0.001*
Histology						
ADC vs. SCC	0.893	0.708-1.126	0.338	1.204	0.941-1.541	0.141
Tumor size						
T3/T4 vs. T2 vs. T1	1.469	1.221-1.767	0.000*	1.362	1.116-1.663	0.002*
LN involvement						
Yes vs. No	1.297	1.023-1.646	0.032*	1.565	1.219-2.008	0.000*
Smoking status						
Former/Current vs. Never	0.895	0.664-1.205	0.464	0.923	0.663-1.286	0.638

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival; PS, performance status. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

Table 19. Results from survival analysis based on clinicopathological variables for the ADC cohort.

Variable	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
Gender						
<i>Male vs. Female</i>	0.909	0.663-1.247	0.555	0.899	0.622-1.299	0.570
Age						
<i>> 65 vs. ≤ 65</i>	1.431	1.026-1.995	0.035*	1.339	0.908-1.975	0.140
TNM staging						
<i>III vs. II vs. I</i>	1.407	1.158-1.709	0.001*	1.547	1.232-1.943	0.000*
Tumor size						
<i>T3/T4 vs. T2 vs. T1</i>	1.700	1.313-2.201	0.000*	1.481	1.086-2.02	0.013*
LN involvement						
<i>Yes vs. No</i>	1.587	1.153-2.185	0.005*	2.118	1.461-3.07	0.000*
Smoking status						
<i>Former/Current vs. Never</i>	0.831	0.586-1.178	0.298	0.744	0.497-1.113	0.150

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival; PS, performance status. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

The SCC subcohort comprised 316 patients, 85 (26.9%) of whom relapsed and 147 (15.5%) who died. In contrast to the findings in ADC patients, no significant associations were found between clinicopathological variables and relapse or survival in this group (Table 20).

Table 20. Results from survival analysis based on clinicopathological variables for the SCC cohort.

Variable	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
Gender						
<i>Male vs. Female</i>	0.712	0.48-1.055	0.090	0.766	0.521-1.127	0.176
Age						
<i>> 65 vs. ≤ 65</i>	1.184	0.815-1.721	0.376	1.292	0.898-1.86	0.168
TNM staging						
<i>III vs. II vs. I</i>	1.083	0.863-1.36	0.489	1.142	0.926-1.408	0.215
Tumor size						
<i>T3/T4 vs. T2 vs. T1</i>	1.263	0.966-1.651	0.088	1.269	0.976-1.65	0.076
LN involvement						
<i>Yes vs. No</i>	1.021	0.712-1.464	0.909	1.203	0.855-1.694	0.289
Smoking status						
<i>Former/Current vs. Never</i>	1.295	0.658-2.548	0.454	1.311	0.642-2.677	0.457

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival; PS, performance status. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

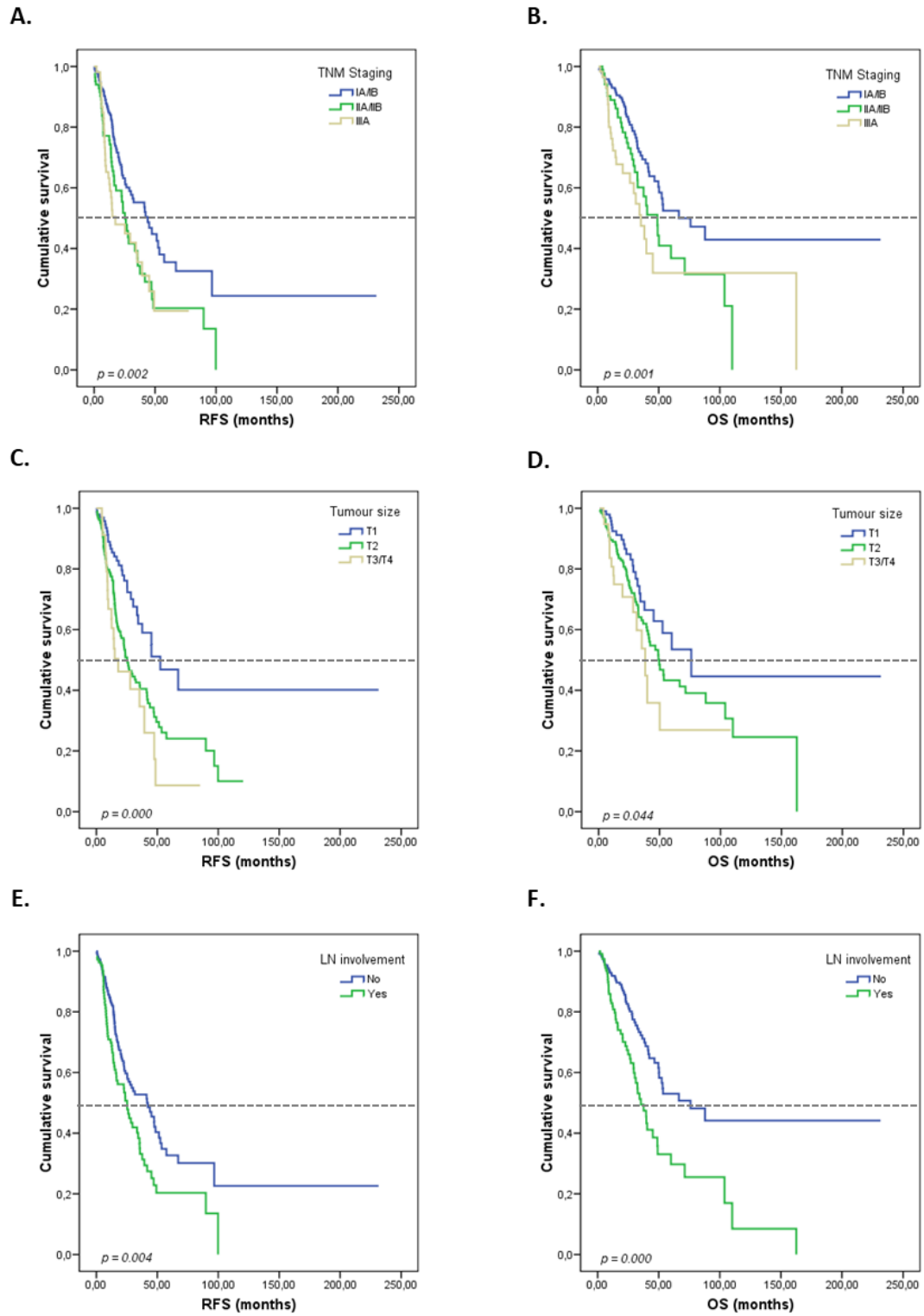


Figure 35. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the ADC subcohort. A-B. TNM staging; C-D. Tumor size; E-F. LN involvement. P-values from the Kaplan-Meier test. LN: lymph node.

4.2. INDIVIDUAL BIOMARKERS

The prognostic value of the genes significantly overexpressed in lung tumorspheres was assessed using the univariate Cox regression method for RFS and OS. Gene expression levels were dichotomized according to their median, and the results obtained are shown in Table 21.

Table 21. Results from survival analysis based on gene expression biomarkers for the entire cohort.

Gene	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
ALDH1A1						
<i>High vs. Low</i>	0.828	0.669-1.024	0.082	1.037	0.837-1.285	0.741
CD44						
<i>High vs. Low</i>	0.916	0.740-1.133	0.417	1.097	0.885-1.359	0.398
CDKN1A						
<i>High vs. Low</i>	1.136	0.918-1.405	0.241	1.251	1.010-1.551	0.041*
CTNNB1						
<i>High vs. Low</i>	1.094	0.884-1.354	0.408	1.014	0.819-1.257	0.897
DLL4						
<i>High vs. Low</i>	1.187	0.957-1.471	0.119	1.013	0.817-1.257	0.904
GSK3B						
<i>High vs. Low</i>	0.830	0.670-1.027	0.086	0.967	0.780-1.197	0.756
ITGA6						
<i>High vs. Low</i>	1.063	0.859-1.314	0.576	1.276	1.029-1.582	0.026*
JAG1						
<i>High vs. Low</i>	0.987	0.798-1.221	0.904	1.209	0.975-1.500	0.083
JUNB						
<i>High vs. Low</i>	1.233	0.997-1.527	0.054	1.035	0.836-1.283	0.752
KLF4						
<i>High vs. Low</i>	0.924	0.746-1.144	0.469	1.085	0.875-1.345	0.458
MDM2						
<i>High vs. Low</i>	0.916	0.740-1.133	0.417	0.990	0.799-1.227	0.927
MMP9						
<i>High vs. Low</i>	1.114	0.900-1.379	0.320	1.031	0.832-1.277	0.781
NANOG						
<i>High vs. Low</i>	1.137	0.882-1.466	0.321	1.062	0.821-1.374	0.648
NOTCH1						
<i>High vs. Low</i>	0.861	0.695-1.066	0.169	1.004	0.810-1.244	0.972
NOTCH3						
<i>High vs. Low</i>	0.935	0.756-1.157	0.536	0.942	0.760-1.167	0.585
SNAI1						
<i>High vs. Low</i>	1.340	1.081-1.660	0.007*	1.350	1.089-1.674	0.006*
THY1						
<i>High vs. Low</i>	1.463	1.181-1.813	<0.001*	1.420	1.144-1.762	0.001*

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; * $p < 0.05$.

Univariate Cox regression analysis revealed that high levels of *SNAI1* were associated with shorter RFS [HR, 1.340; 95% CI 1.081-1.660; $p = 0.007$] and worse OS [HR, 1.350; 95% CI 1.089-1.674; $p = 0.006$], as well as high levels of *THY1*, RFS [HR, 1.463; 95% CI 1.181-1.813; $p < 0.001$] and OS [HR, 1.420; 95% CI 1.144-1.762; $p = 0.001$]. Moreover, high expression levels of *CDKN1A* and *ITGA6* were associated with worse OS as well [HR, 1.251; 95% CI 1.010-1.551; $p = 0.041$] and [HR, 1.276; 95% CI 1.029-1.582; $p = 0.026$], respectively. Kaplan-Meier analyses were carried out in order to obtain the survival plots (Figure 36 and 37).

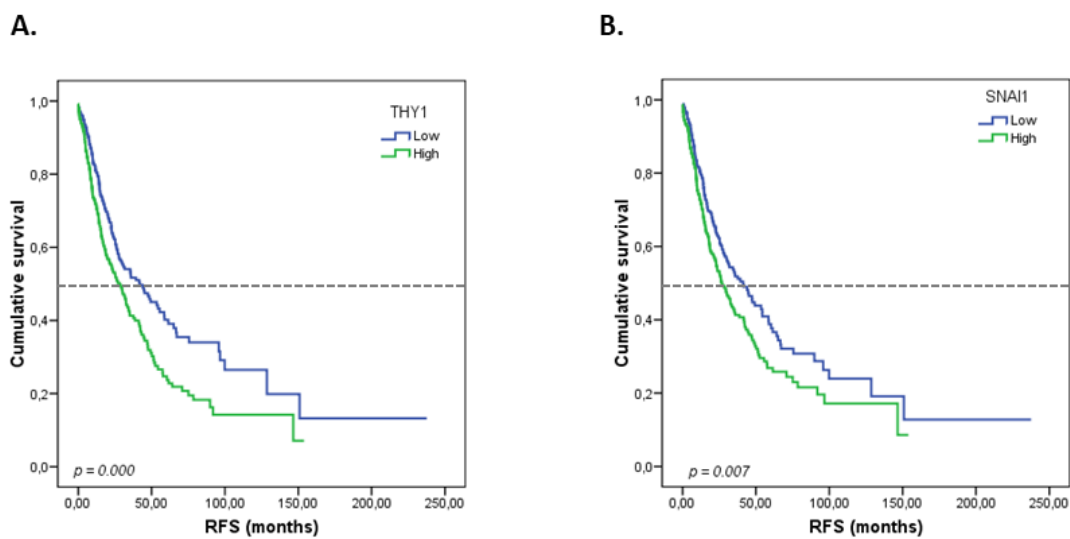


Figure 36. Kaplan-Meier plots for RFS according to gene expression levels for the entire cohort. A. *THY1*; B. *SNAI1*. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. P -values from the Kaplan-Meier test.

As expected, these four genes, which were found significantly overexpressed in lung tumorspheres compared to adherent cells, are associated with worse prognosis. The overexpression of Snail analyzed by immunohistochemistry was previously associated with poor prognosis in different cohorts of resected NSCLC patients (347,348). In contrast, few results have been published regarding the prognosis impact of *THY1* in lung cancer. Flow cytometry studies suggested the co-expression of CD90, CD326 and CD133 as a signature with prognostic value in patients with NSCLC (157), and co-expression of CD90 and CD44 was associated with significantly reduced RFS in NSCLC patients (349). Controversial data has been published about the correlation between p21 expression and survival for NSCLC patients (350,351). However, most studies did

not take into account the cellular localization of p21, and those publications where cytoplasmic and nuclear differentiation were performed, agreed in the poor prognosis of cytoplasmic p21 (306,307). In addition, previous *in silico* analysis correlated high *ITGA6* expression with poor prognosis (352), and elevated levels of the integrin $\alpha6\beta4$ heterodimer determined by immunohistochemistry were associated with decreased OS in another cohort of 216 NSCLC patients (326).

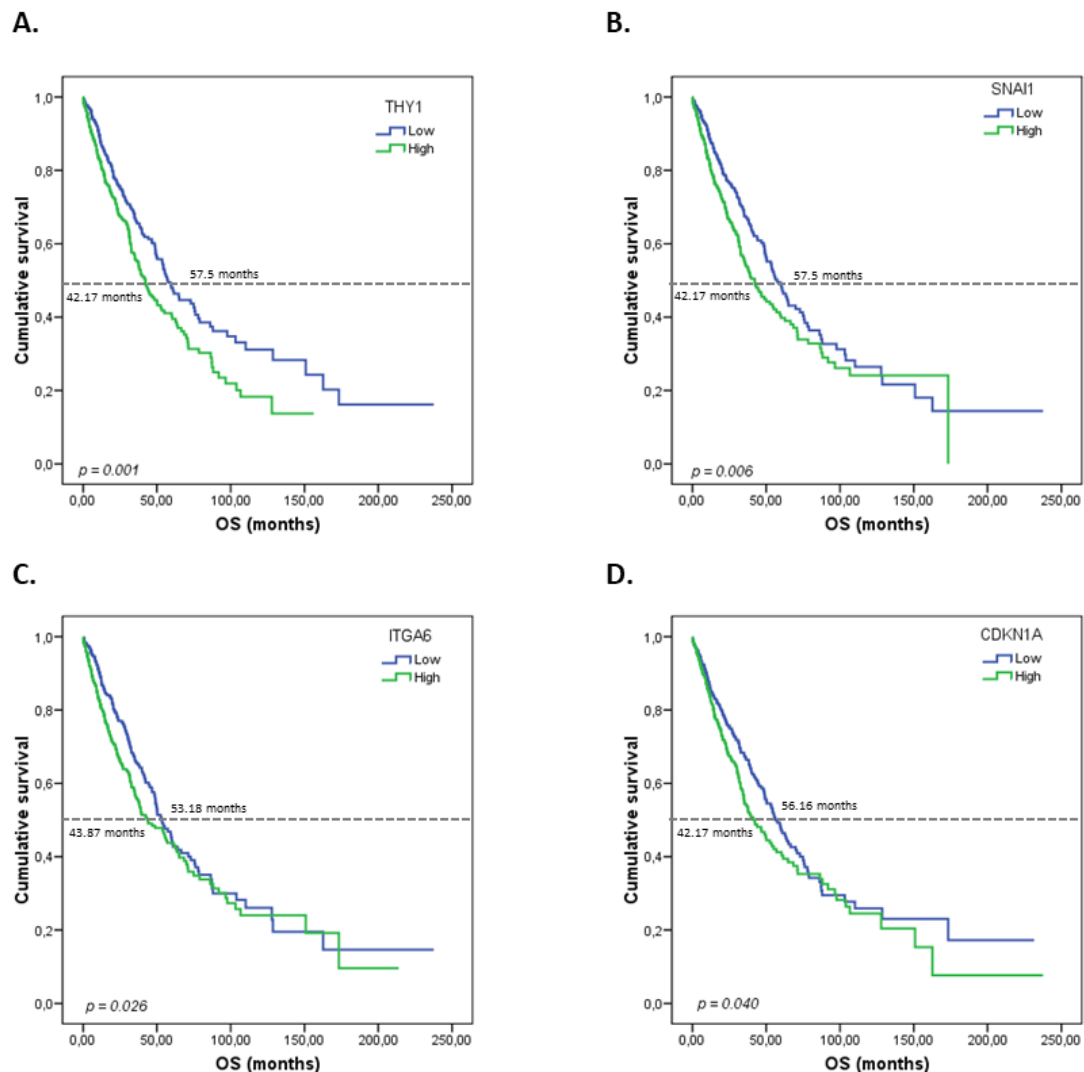


Figure 37. Kaplan-Meier plots for OS according to gene expression levels for the entire cohort. A. *THY1*; B. *SNAI1*; C. *ITGA6*; D. *CDKN1A*. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. *P*-values from the Kaplan-Meier test.

Furthermore, it is known that NSCLC is one of the most genomically diverse tumors, and there are a variety of molecularly-defined subsets of patients. Different driver mutations have been identified in SCC and ADC histologies and we found

significant differences between them *in vitro*, which led to the assumption that they are molecularly different diseases. For this reason, survival analysis was also performed according to patient’s histology. Results from univariate Cox regression analysis for ADC patients are provided in Table 22.

Table 22. Results from survival analysis based on gene expression biomarkers for the ADC subcohort.

Gene	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
ALDH1A1 <i>High vs. Low</i>	0.758	0.567-1.013	0.061	0.881	0.645-1.203	0.425
CD44 <i>High vs. Low</i>	1.031	0.773-1.374	0.837	1.047	0.768-1.428	0.772
CDKN1A <i>High vs. Low</i>	1.014	0.761-1.351	0.926	1.161	0.851-1.584	0.345
CTNNB1 <i>High vs. Low</i>	0.911	0.684-1.214	0.526	0.814	0.594-1.114	0.198
DLL4 <i>High vs. Low</i>	0.880	0.660-1.173	0.384	0.849	0.620-1.161	0.305
GSK3B <i>High vs. Low</i>	0.923	0.692-1.230	0.584	0.739	0.539-1.013	0.060
ITGA6 <i>High vs. Low</i>	1.208	0.906-1.610	0.198	1.444	1.057-1.973	0.021*
JAG1 <i>High vs. Low</i>	1.042	0.782-1.388	0.779	1.368	1.001-1.869	0.049*
JUNB <i>High vs. Low</i>	1.365	1.023-1.822	0.035*	1.116	0.818-1.523	0.488
KLF4 <i>High vs. Low</i>	0.953	0.711-1.277	0.747	1.296	0.947-1.772	0.105
MDM2 <i>High vs. Low</i>	0.893	0.669-1.191	0.440	0.941	0.689-1.284	0.701
MMP9 <i>High vs. Low</i>	1.321	0.989-1.765	0.060	1.263	0.923-1.728	0.145
NANOG <i>High vs. Low</i>	0.980	0.700-1.373	0.907	0.928	0.648-1.330	0.684
NOTCH1 <i>High vs. Low</i>	0.734	0.549-0.982	0.038*	0.805	0.586-1.105	0.179
NOTCH3 <i>High vs. Low</i>	0.957	0.717-1.277	0.764	1.121	0.821-1.530	0.472
THY1 <i>High vs. Low</i>	1.227	0.920-1.637	0.163	1.321	0.966-1.807	0.082
SNAI1 <i>High vs. Low</i>	1.277	0.957-1.703	0.096	1.250	0.915-1.707	0.162

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; **p* < 0.05.

The univariate Cox regression model performed with ADC patients showed similar association between high *ITGA6* and worse prognosis to that found in the entire cohort, OS [HR, 1.444; 95% CI 1.057-1.973; $p = 0.021$]. Remarkably, *JAG1* was also associated with worse OS [HR, 1.368; 95% CI 1.001-1.869; $p = 0.049$]. Kaplan-Meier plots represented in Figure 38 show the survival differences between the two groups of patients. As happened with the genes associated with prognosis in the entire cohort, we found *JAG1* overexpressed in tumorspheres. *JAG1* encodes for one of the five canonical ligands for Notch receptors and its increase has been correlated with metastasis induction and poor survival in NSCLC (353).

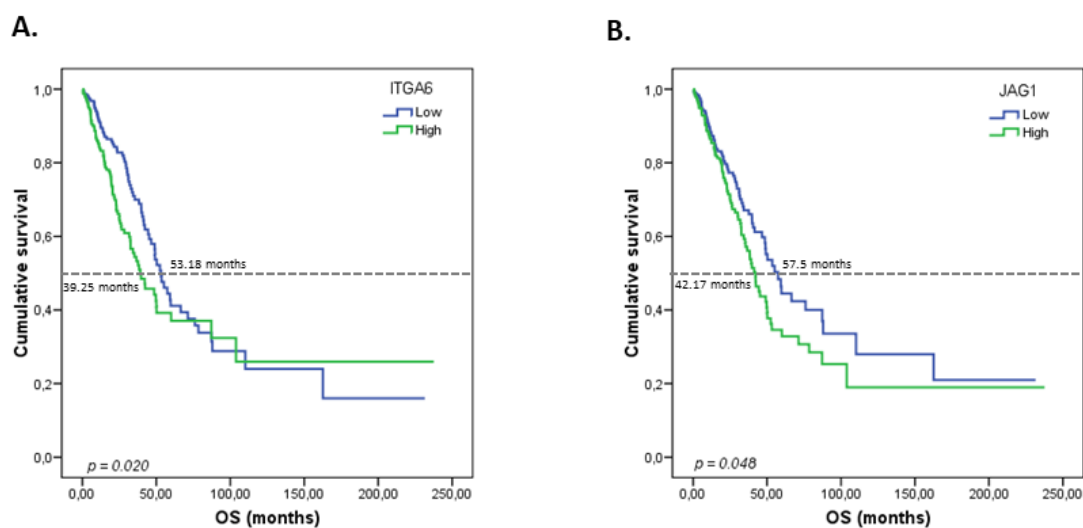


Figure 38. Kaplan-Meier plots for OS according to gene expression levels for the ADC subcohort. A. *ITGA6*; B. *JAG1*. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. P -values from the Kaplan-Meier test.

Interestingly, high *JUNB* expression correlated with shorter RFS [HR, 1.365; 95% CI 1.023-1.822; $p = 0.035$] in contrast to *NOTCH1*, whose high expression was associated with better prognosis, RFS [HR, 0.734; 95% CI 0.549-0.982; $p = 0.038$] (Figure 39). However, we found both genes significantly overexpressed in tumorspheres compared to adherent cells as well. JunB is a member of the activator protein-1 (AP-1), which plays an important role in a number of cellular processes, including proliferation, differentiation, apoptosis, cell migration, and transformation (354). However, opposite roles have been proposed for this molecule and its implication in lung cancer remains unclear (355). Many studies have been performed to analyze the prognostic value of

NOTCH1 reporting disparate results. A big meta-analysis including 2,369 NSCLC patients found no significant associations between OS and Notch1 expression in ADC or SCC patients (356), although it seems that high Notch1 expression could correlate with greater possibility of lymph node metastasis and higher tumor stages. Unexpectedly, we found that higher expression of *NOTCH1* in ADC patients from TCGA is associated with better prognosis, suggesting that more studies including this gene could be interesting to clarify the role of this molecule in NSCLC.

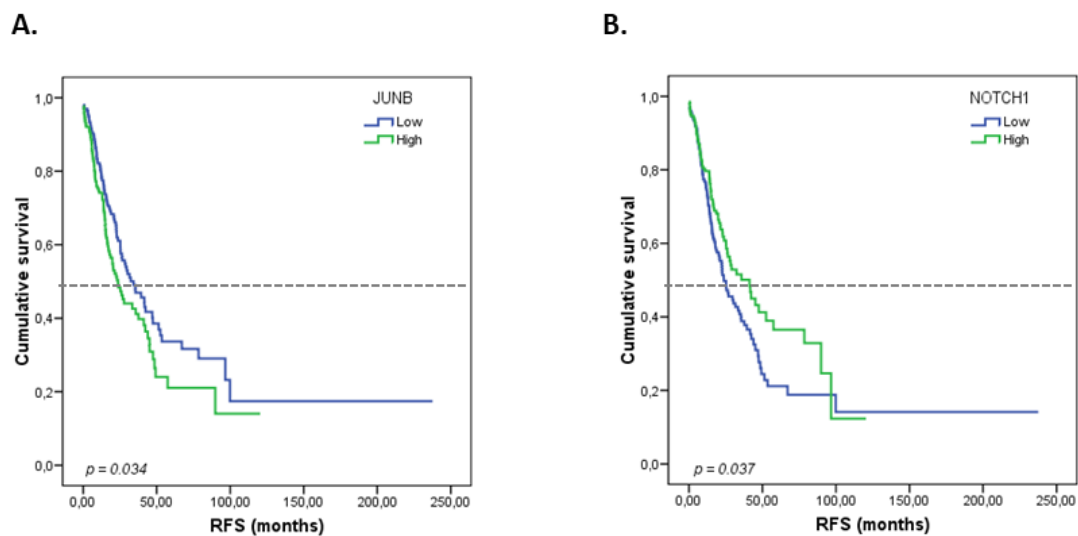


Figure 39. Kaplan-Meier plots for RFS according to gene expression levels for the ADC subcohort. A. *JUNB*; B. *NOTCH1*. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. *P*-values from the Kaplan-Meier test.

Regarding SCC patients, univariate Cox regression analysis showed that *SNAI1* and *THY1* have similar association with relapse and survival than that observed in the entire cohort (Table 23 and Figure 40) and no other significant correlations were found.

Table 23. Results from survival analysis based on gene expression biomarkers for the SCC subcohort.

Gene	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
ALDH1A1 <i>High vs. Low</i>	0.826	0.602-1.134	0.238	0.943	0.701-1.270	0.701
CD44 <i>High vs. Low</i>	0.934	0.679-1.285	0.675	1.037	0.770-1.396	0.811
CDKN1A <i>High vs. Low</i>	1.325	0.962-1.823	0.085	1.310	0.972-1.766	0.076
CTNNB1 <i>High vs. Low</i>	1.092	0.795-1.500	0.587	1.067	0.793-1.436	0.667
DLL4 <i>High vs. Low</i>	1.252	0.910-1.723	0.168	1.161	0.862-1.562	0.326
GSK3B <i>High vs. Low</i>	1.117	0.813-1.535	0.495	1.187	0.880-1.601	0.261
ITGA6 <i>High vs. Low</i>	1.275	0.928-1.752	0.134	1.166	0.866-1.569	0.312
JAG1 <i>High vs. Low</i>	1.033	0.753-1.418	0.839	0.911	0.677-1.226	0.539
JUNB <i>High vs. Low</i>	1.190	0.866-1.636	0.283	1.100	0.817-1.481	0.529
KLF4 <i>High vs. Low</i>	1.149	0.836-1.579	0.392	0.976	0.725-1.314	0.872
MDM2 <i>High vs. Low</i>	0.925	0.674-1.271	0.632	0.971	0.721-1.307	0.846
MMP9 <i>High vs. Low</i>	1.045	0.758-1.441	0.789	0.972	0.722-1.310	0.853
NANOG <i>High vs. Low</i>	1.029	0.700-1.514	0.884	1.110	0.767-1.608	0.579
NOTCH1 <i>High vs. Low</i>	1.281	0.930-1.764	0.129	1.176	0.872-1.587	0.288
NOTCH3 <i>High vs. Low</i>	0.976	0.710-1.342	0.883	0.766	0.568-1.034	0.081
THY1 <i>High vs. Low</i>	1.598	1.160-2.201	0.004*	1.532	1.137-2.065	0.005*
SNAI1 <i>High vs. Low</i>	1.718	1.242-2.377	0.001*	1.416	1.051-1.908	0.022*

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; * $p < 0.05$.

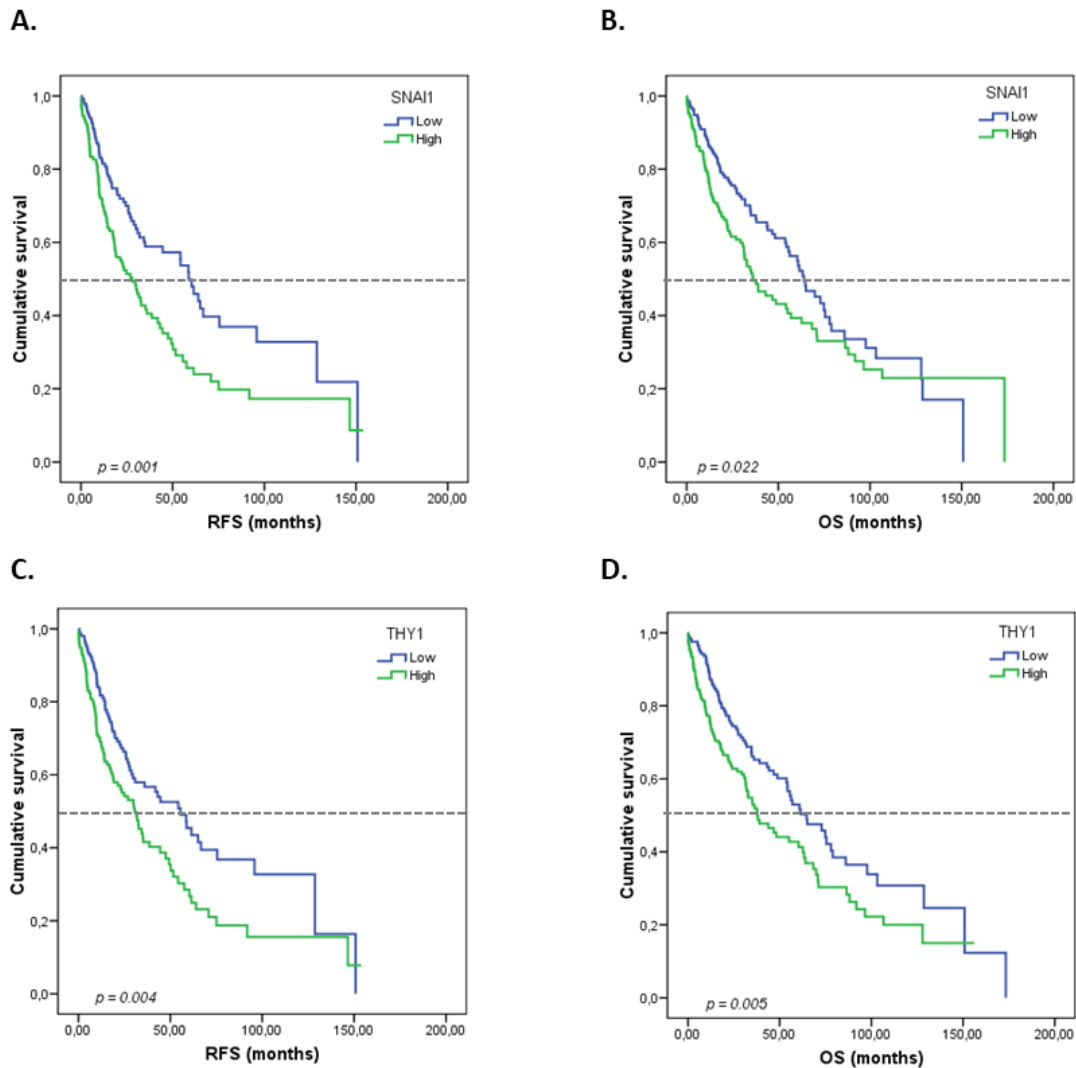


Figure 40. Kaplan-Meier plots for RFS and OS according to gene expression levels for the SCC subcohort. A-B. *SNAI1*; C-D. *THY1*. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. *P*-values from the Kaplan-Meier test.

4.3. GENE EXPRESSION SCORE

We also decided to create a gene expression score based on a multi-gene signature, which can provide more accurate predictions than a model using single genes. For this purpose, we followed the steps described in the bioinformatics and statistical analysis section of *Materials and Methods*. Based on univariate Cox regression analysis, expression of *SNAI1*, *ITGA6*, and *CDKN1A* were moderately associated with mortality ($|Z\text{-score}| > 1.5$; Figure 41).

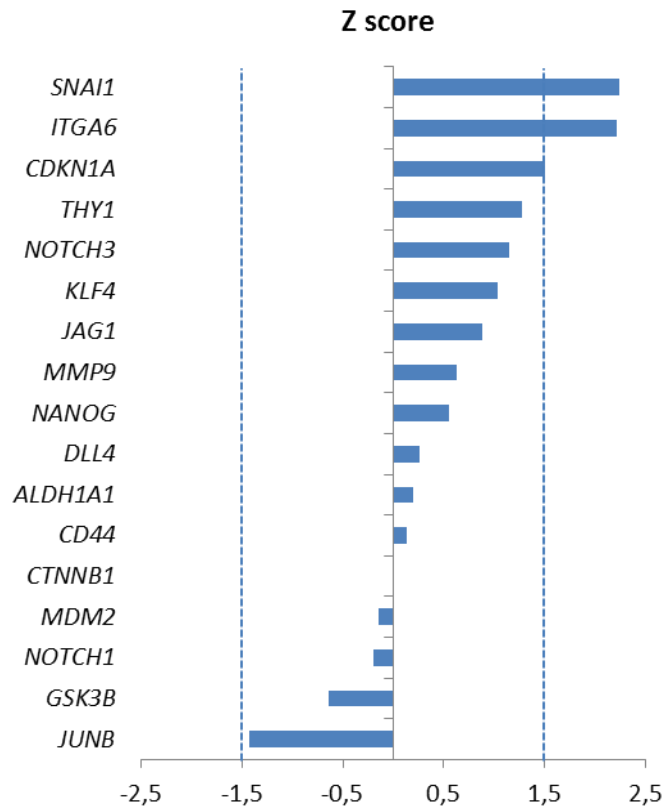


Figure 41. Univariate analysis of the expression of the 17 genes overexpressed in lung tumorspheres for OS. The genes are ranked based on their predictive power (univariate Z-score). Dashed lines indicate $|Z\text{-score}| = 1.5$. This criterion was used to select genes to include in the multivariate Cox regression model used to calculate the expression score.

These genes were selected to construct a prognostic signature by introducing them into a multivariate model, using their absolute regression coefficients from this analysis to calculate the expression prognostic score, called CSCs score. The results from the multivariate model including these 3 genes can be found in Table 24.

Table 24. Results from the multivariate model for OS with genes included in the expression score.

Variable	Regression coefficient	SE	p -value	HR	95% CI
<i>CDKN1A</i>	0.123	0.164	0.452	1.131	0.821-1.559
<i>ITGA6</i>	0.196	0.102	0.054	1.217	0.997-1.486
<i>SNAI1</i>	0.255	0.133	0.056	1.290	0.994-1.674

CI, confidence interval; HR, hazard ratio; SE, standard error.

The formula to calculate the signature was the following:

$$(CDKN1A \times 0.123) + (ITGA6 \times 0.196) + (SNAI1 \times 0.255)$$

Kaplan–Meier log-rank analysis showed that patients from the entire cohort with high CSCs score have shorter OS (37.70 vs. 60.36 months, $p = 0.001$, Figure 42A). Furthermore, to evaluate the potential use of the score as a biomarker and based on our previous results, we did a stratified analysis by histology. We observed that when we stratified patients according to histology, the association between the ADC patient score and the prognosis was similar to that found in the entire cohort for OS (36.56 vs. 53.51 months, $p = 0.003$, Figure 42B), whereas no significant association was found for SCC patients ($p = 0.100$).

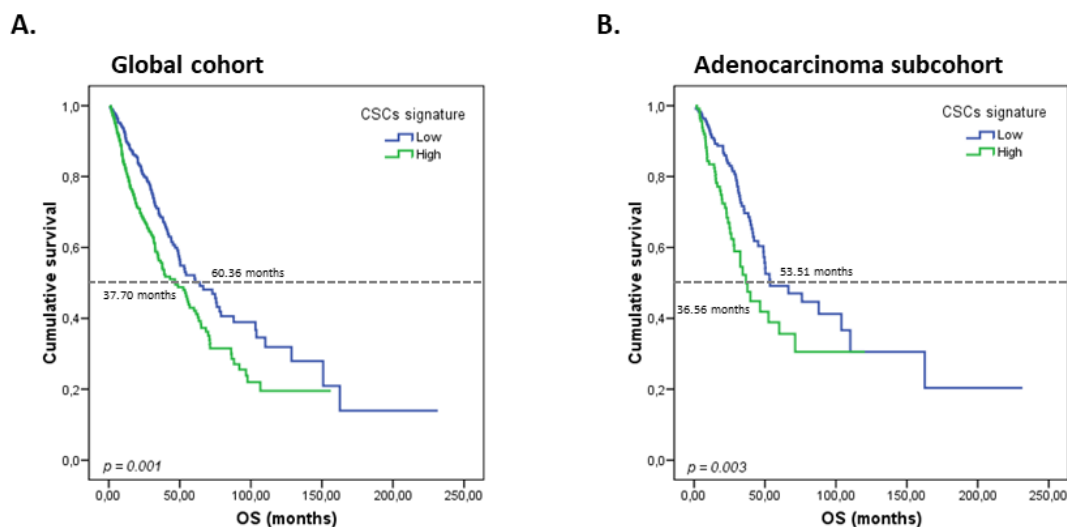


Figure 42. Kaplan-Meier plots for OS according to the gene expression score levels. A. Global cohort; B. ADC subcohort. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. P -values from the Kaplan-Meier test.

Here, we identified a prognostic score based on the expression of three CSCs-related genes which is associated with OS in NSCLC. NSCLC patients with high expression score and therefore high expression levels of these genes, which were significantly overexpressed in lung tumorspheres enriched in CSCs, had decreased survival. This score was also applicable for ADC patients, which showed a similar correlation between the expression of these three genes, *SNAI1*, *ITGA6*, and *CDKN1A*, and survival.

4.4. MULTIVARIATE ANALYSIS

In order to determine the prognostic value of the CSCs score identified, we interrogated all the biomarkers that were significantly associated with prognosis ($p <$

0.05) and included them in a multivariate model for OS. In the multivariate model the following variables were included: age, TNM staging, tumor size, LN involvement, *CDKN1A*, *ITGA6*, *SNAI1*, *THY1* and the CSCs score. Results obtained from this multivariate analysis indicated that age, TNM staging and the CSCs score were independent biomarkers for OS in the entire cohort (Table 25). Multivariate analysis was also performed with significant results for ADC patients. The following variables were included in the model: TNM staging, tumor size, LN involvement, *ITGA6*, *JAG1*, and the CSCs score. In this case, TNM staging, LN involvement and the CSCs score proved to be independent prognostic biomarkers. Our results indicate that the CSCs score is an independent biomarker for OS, and moreover, its prognostic value proved to be stronger for OS than other factors such as TNM staging in the entire cohort and in ADC patients.

Table 25. Multivariate Cox regression model results, including all the significant variables.

Variables	Global cohort			ADC subcohort		
	HR	95% CI	p-value	HR	95% CI	p-value
Age						
>65 vs. ≤65	1.398	1.069-1.827	0.014	-	-	-
TNM staging						
IIIA vs. II vs. I	1.353	1.161-1.578	0.0001	1.515	1.097-2.092	0.012
LN involvement						
Yes vs. No	-	-	-	2.108	1.453-3.059	<0.0001
CSCs signature						
High vs. low	1.498	1.167-1.922	0.001	1.869	1.275-2.738	0.001

ADC, adenocarcinoma; LN, lymph node; CI, confidence interval; HR, hazard ratio.

Many studies have tried to correlate the expression of genes associated with CSCs to patients' prognosis (353,357). Nevertheless, most of them are focused on single pathway-specific markers with limited prognostic value. Finding gene expression signatures that identify altered pathways in carcinogenesis could lead to the discovery of molecular subclasses and predict patients' outcomes better (358,359). In this study, we created a score combining the expression of *CDKN1A*, *ITGA6* and *SNAI1* which was an independent prognostic biomarker for early-stage resected lung cancer patients and even stronger than factors currently used in clinical practice like the TNM staging. These results are of great importance because current clinicopathological staging methods have limited success in predicting patient survival and today we still cannot predict which patients will be cured, and which ones will relapse after surgery. The CSCs score

proposed can help in future clinical practice, since high scores may reflect a bigger CSCs population with enhanced cell migration, invasion and tumor initiation capacity that will be able to modulate cell cycle and persist after cancer treatments. As a result, patients with high values on this score may need adjuvant treatment and should be closely followed after a successful surgery, because they have a higher risk to die. To validate this score, the prognostic value of the expression of these three genes was evaluated in an independent cohort of resected lung patients from our hospital.

5. VALIDATION OF THE PROGNOSTIC GENE EXPRESSION SCORE

To validate the prognostic value of the CSCs score proposed, the gene expression of *ITGA6*, *CDKN1A* and *SNAI1* was evaluate in an independent cohort of 245 patients diagnosed of NSCLC who underwent surgery at *Consorti Hospital General Universitari de València*.

5.1. CLINICOPATHOLOGICAL VARIABLES

Of the 245 resected NSCLC patients included in this part of the study, 101 (41.4%) relapsed and 117 (48.0%) died during the follow-up. The median follow-up was 32.27 months [range: 1.00-161.70]. The prognostic value of the different clinicopathological variables was assessed using the univariate Cox regression method for RFS and OS, and are shown in Table 26 along with the hazard ratios and *p*-value for each variable.

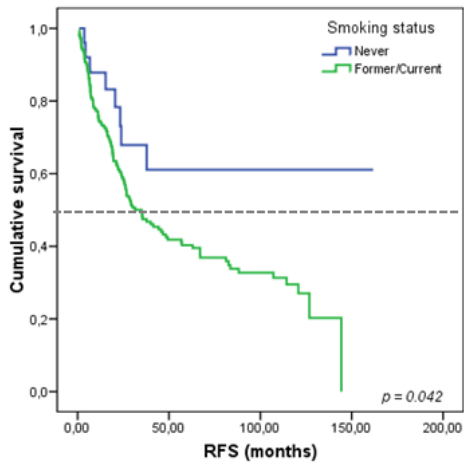
Significant results obtained from the univariate Cox regression method were also analyzed using the Kaplan-Meier method (log-rank) in order to obtain the survival plots. This univariate analysis showed that patients with large tumors, LN involvement, worse PS, more advanced stage of the disease, and smokers and former smokers had shorter RFS. In addition, those with worse PS, more advanced disease or mutated in *KRAS* had worse OS as well, which agrees with our previous finding in the entire cohort from TCGA (Figure 43).

Table 26. Results for survival analysis based on clinicopathological variables for the validation cohort.

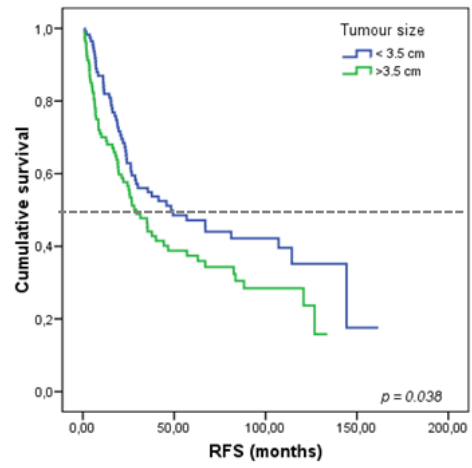
Variable	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
Gender						
<i>Male vs. Female</i>	1.471	0.856-2.528	0.162	1.504	0.822-2.752	0.185
Age						
<i>> 65 vs. ≤ 65</i>	1.198	0.830-1.730	0.335	1.358	0.912-2.023	0.132
TNM staging						
<i>III vs. II vs. I</i>	1.435	1.147-1.794	0.002*	1.436	1.129-1.826	0.003*
Histology						
<i>ADC vs. SCC</i>	1.055	0.785-1.419	0.721	1.012	0.738-1.388	0.942
Tumor size						
<i>T3/T4 vs. T2 vs. T1</i>	1.452	1.018-2.070	0.039*	1.312	0.892-1.931	0.168
LN involvement						
<i>Yes vs. No</i>	1.761	1.208-2.567	0.003*	1.501	0.992-2.271	0.055
Pleural involvement						
<i>Yes vs. No</i>	1.432	0.992-2.069	0.055	1.476	0.984-2.214	0.060
Performance status						
<i>1 vs. 0</i>	1.593	1.106-2.294	0.012*	1.678	1.121-2.513	0.012*
Differentiation grade						
<i>Poor vs. Moderate vs. Well</i>	1.199	0.923-1.558	0.175	1.088	0.819-1.446	0.559
Surgical margin						
<i>R1/R2 vs. R0</i>	1.553	0.991-2.433	0.055	1.470	0.899-2.403	0.124
Smoking status						
<i>Former/Current vs. Never</i>	2.074	1.010-4.261	0.047*	1.817	0.840-3.933	0.129
EGFR						
<i>Mutated vs. Wild type</i>	1.115	0.501-2.480	0.790	0.985	0.415-2.339	0.972
KRAS						
<i>Mutated vs. Wild type</i>	1.530	0.961-2.435	0.073	1.657	1.003-2.737	0.049*

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival; ADC, adenocarcinoma; SCC, squamous cell carcinoma. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

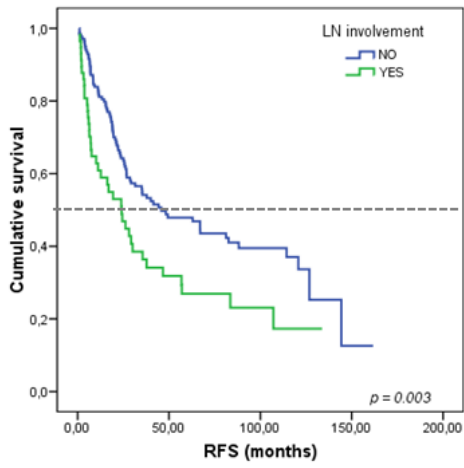
A.



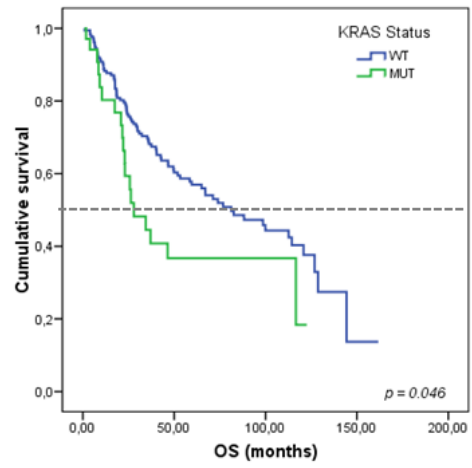
B.



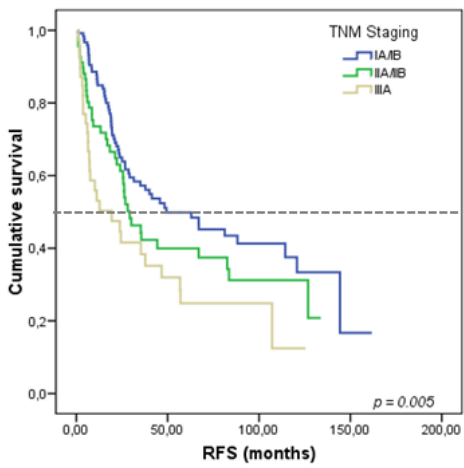
C.



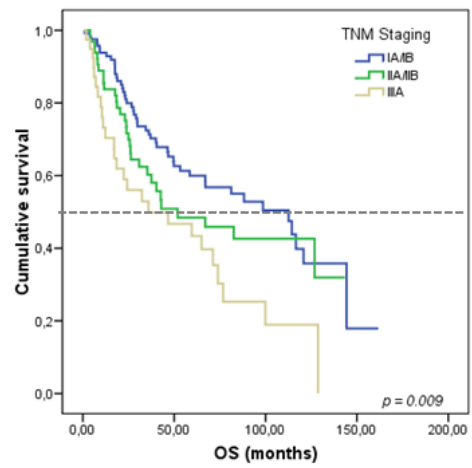
D.



E.



F.



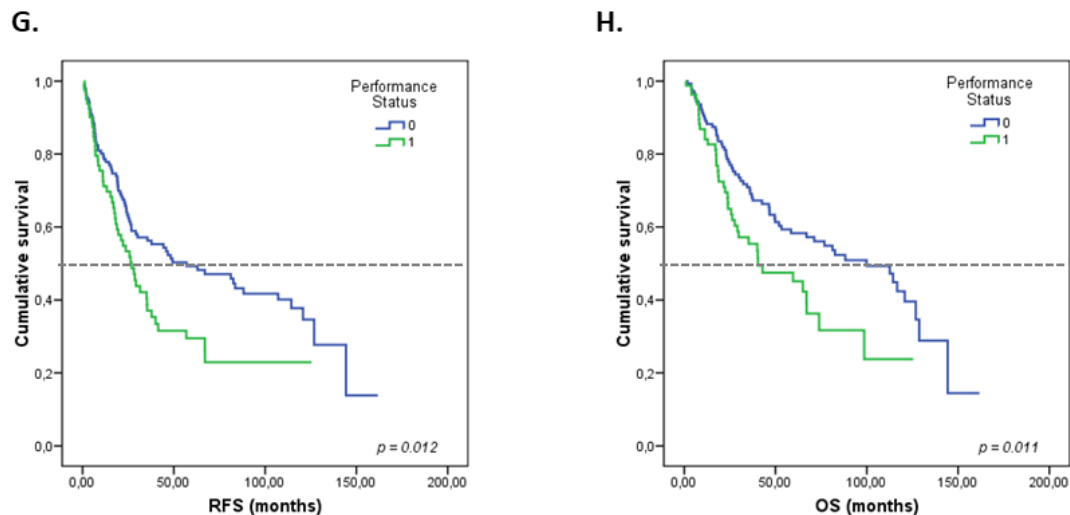


Figure 43. Kaplan-Meier plots for RFS and OS according to clinicopathological variables for the validation cohort. A. Smoking status; B. Tumor size; C. LN involvement; D. KRAS status; E-F. TNM staging; G-H. Performance status. *P*-values from the Kaplan-Meier test. LN: lymph node; MUT: mutated; WT: wild type.

The prognostic value of clinicopathological variables was also assessed according to histology. The ADC subcohort comprised 114 patients, 41 (36.6%) of whom relapsed and 48 (42.9%) who died. In the univariate analysis, LN involvement and surgical margins of non-tumorous tissue around resected tumors were the only variables to be significantly associated with prognosis (Table 27). Nevertheless, other variables which were significantly associated with prognosis in the TCGA cohort, such as tumor size, showed the same tendency in the validation cohort, suggesting that a bigger sample size is required to get significant results for those correlations. Survival plots from Kaplan-Meier survival analysis are represented in Figure 44.

Table 27. Results for survival analysis based on clinicopathological variables of the ADC validation subcohort.

Variable	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
Gender						
Male vs. Female	1.422	0.720-2.809	0.311	1.380	0.651-2.926	0.401
Age						
> 65 vs. ≤ 65	1.217	0.678-2.185	0.510	1.510	0.790-2.886	0.212
TNM staging						
III vs. II vs. I	1.301	0.889-1.905	0.176	1.245	0.812-1.911	0.315
Tumor size						
T3/T4 vs. T2 vs. T1	1.695	0.944-3.044	0.077	1.205	0.628-2.310	0.575
LN involvement						
Yes vs. No	2.273	1.201-4.302	0.012*	1.836	0.901-3.741	0.095
Pleural involvement						
Yes vs. No	1.536	0.824-2.863	0.177	1.447	0.727-2.883	0.293
Performance status						
1 vs. 0	1.463	0.752-2.846	0.262	1.217	0.572-2.589	0.610
Differentiation grade						
Poor vs. Moderate vs. Well	0.975	0.635-1.497	0.907	0.943	0.584-1.523	0.810
Surgical margin						
R1/R2 vs. R0	2.358	1.086-5.118	0.030*	1.732	0.719-4.172	0.221
Smoking status						
Former/Current vs. Never	2.097	0.887-4.959	0.092	1.900	0.739-4.887	0.183
EGFR						
Mutated vs. Wild type	1.030	0.426-2.487	0.948	0.949	0.360-2.501	0.916
KRAS						
Mutated vs. Wild type	1.588	0.815-3.094	0.174	1.360	0.638-2.896	0.426

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

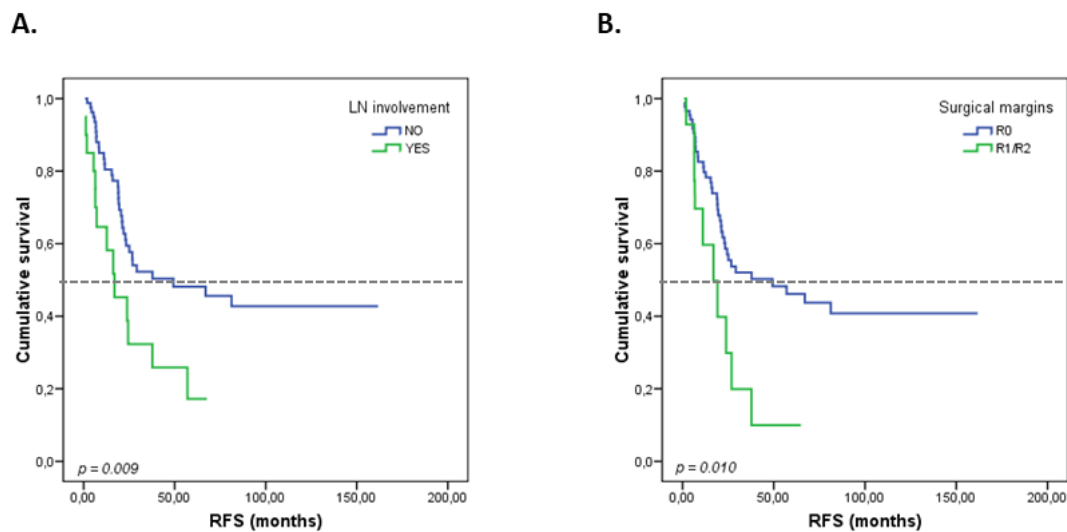


Figure 44. Kaplan-Meier plots for RFS according to clinicopathological variables for the ADC validation cohort. A. LN involvement; B. Surgical margins. P-values from the Kaplan-Meier test. LN: lymph node.

Regarding SCC, this subcohort comprised 111 patients, 50 (45.0%) of whom relapsed and 58 (52.3%) who died. In contrast to the findings in ADC patients and in concordance with our findings in TCGA analysis, no significant associations were found between clinicopathological variables and relapse or survival in this group (Table 28). Interestingly, only 3 out of the 111 SCC patients were never smokers, so statistical analysis for smoking status could not be performed.

Table 28. Results for survival analysis based on clinicopathological variables of the SCC validation subcohort.

Variable	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
Gender						
<i>Male vs. Female</i>	1.171	0.161-8.489	0.876	1.036	0.142-7.535	0.972
Age						
<i>> 65 vs. ≤ 65</i>	1.231	0.727-2.083	0.439	1.615	0.901-2.895	0.108
TNM staging						
<i>III vs. II vs. I</i>	1.338	0.970-1.846	0.076	1.295	0.918-1.826	0.141
Tumor size						
<i>T3/T4 vs. T2 vs. T1</i>	1.096	0.650-1.850	0.730	1.127	0.641-1.982	0.677
LN involvement						
<i>Yes vs. No</i>	1.358	0.794-2.321	0.263	1.128	0.629-2.025	0.686
Pleural involvement						
<i>Yes vs. No</i>	1.191	0.701-2.025	0.518	1.258	0.705-2.242	0.437
Performance status						
<i>1 vs. 0</i>	1.321	0.773-2.256	0.309	1.595	0.887-2.868	0.119
Differentiation grade						
<i>Poor vs. Moderate vs. Well</i>	1.268	0.836-1.924	0.264	1.090	0.695-1.710	0.708
Surgical margin						
<i>R1/R2 vs. R0</i>	0.930	0.469-1.845	0.837	1.134	0.566-2.274	0.723
Smoking status						
<i>Former/Current vs. Never</i>	-	-	-	-	-	-
EGFR						
<i>Mutated vs. Wild type</i>	5.076	0.565-45.572	0.147	3.277	0.379-28.349	0.281
KRAS						
<i>Mutated vs. Wild type</i>	2.068	0.639-6.700	0.225	1.749	0.420-7.279	0.442

CI, confidence interval; HR, hazard ratio; LN, lymph node; OS, overall survival; RFS, relapse-free survival. The results were obtained using the univariate Cox regression method. * $p < 0.05$.

5.2. INDIVIDUAL BIOMARKERS

The prognostic value of *CDKN1A*, *ITGA6* and *SNAI1* was assessed using the univariate Cox regression method for RFS and OS. Gene expression levels were dichotomized according to their median. The results obtained are shown in Table 29.

Table 29. Results from survival analysis based on gene expression biomarkers of the validation cohort.

Gene	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
<i>CDKN1A</i>						
<i>High vs. Low</i>	1.063	0.723-1.563	0.755	1.136	0.751-1.719	0.546
<i>ITGA6</i>						
<i>High vs. Low</i>	0.993	0.672-1.469	0.973	1.077	0.710-1.633	0.727
<i>SNAI1</i>						
<i>High vs. Low</i>	1.135	0.768-1.676	0.526	1.210	0.798-1.835	0.368

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; * $p < 0.05$.

No significant results were obtained for the entire validation cohort. However, we found significant differences between the histological subtypes for the clinicopathological variables, so we decided to perform survival analysis according to the patient histology as well. Results from univariate Cox regression analysis for ADCs are provided in Table 30.

Table 30. Results from survival analysis based on gene expression biomarkers of the ADC validation subcohort.

Gene	RFS			OS		
	HR	95% CI	p-value	HR	95% CI	p-value
<i>CDKN1A</i>						
<i>High vs. Low</i>	1.551	0.810-2.971	0.185	1.733	0.859-3.494	0.124
<i>ITGA6</i>						
<i>High vs. Low</i>	1.844	0.948-3.588	0.072	2.117	1.031-4.344	0.041*
<i>SNAI1</i>						
<i>High vs. Low</i>	1.849	0.968-3.530	0.062	2.248	1.092-4.629	0.028*

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; * $p < 0.05$.

Univariate Cox regression analysis revealed that high levels of *ITGA6* and *SNAI1* were associated with worse OS [HR, 1.350; 95% CI 1.089-1.674; $p = 0.006$] and [HR,

1.350; 95% CI 1.089-1.674; $p = 0.006$], respectively. Survival plots from Kaplan-Meier analyses are shown in Figure 45.

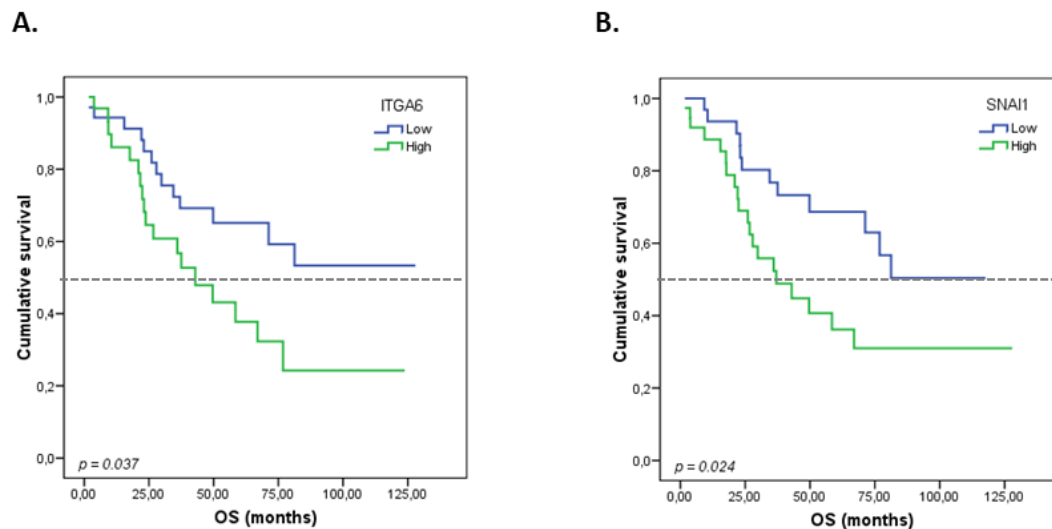


Figure 45. Kaplan-Meier plots for OS according to the gene expression levels of the ADC validation cohort. A. *ITGA6*; B. *SNAI1*. *P*-values from the Kaplan-Meier test.

Regarding SCC, univariate Cox regression showed no associations between the expression of these genes and prognosis (Table 31).

Table 31. Results from survival analysis based on gene expression biomarkers of the SCC validation subcohort.

Gene	RFS			OS		
	HR	95% CI	<i>p</i> -value	HR	95% CI	<i>p</i> -value
<i>CDKN1A</i>						
<i>High vs. Low</i>	1.074	0.637-1.812	0.789	1.281	0.726-2.258	0.393
<i>ITGA6</i>						
<i>High vs. Low</i>	1.146	0.680-1.933	0.608	1.143	0.650-2.009	0.642
<i>SNAI1</i>						
<i>High vs. Low</i>	1.167	0.696-1.954	0.558	0.982	0.560-1.722	0.950

Gene expression levels dichotomized as high and low according to their medians. The results were obtained using the univariate Cox regression method. CI, confidence interval; HR, hazard ratio; OS, overall survival; RFS, relapse-free survival; * $p < 0.05$.

5.3. GENE EXPRESSION SCORE

In order to validate the gene expression score, which we built for OS based on the expression of TCGA patients, we applied the same equation to the expression of *CDKN1A*, *ITGA6* and *SNAI1* of patients from the validation cohort. Due to the differences we found according to histology we also applied analysis separately. We found no

significant associations for the entire cohort including subcohorts and the SCC subcohort ($p = 0.116$ and $p = 0.434$, respectively). On the contrary, the CSCs score showed that ADC patients with a high CSCs expression score have shorter OS (42.90 vs. NR months, $p = 0.020$, Figure 46).

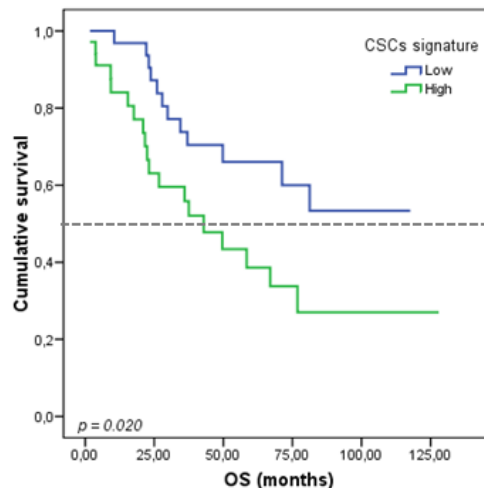


Figure 46. Kaplan-Meier plot for OS according to the gene expression score levels of ADC validation subcohort. Gene expression levels were dichotomised according to the median. Blue lines represent patients with low levels of expression, whilst green lines represent patients with high levels of expression. P -values from the Kaplan-Meier test.

The association between high CSCs score and worse prognosis was confirmed in an independent ADC cohort, validating its prognostic power for this lung cancer subtype. In order to do so, we have used RTqPCR, which is the gold standard method for gene expression quantification, because of its high sensitivity and specificity. Other advantages of this technology are that it requires a low RNA input, it is less time consuming than other methods, and it is robust and flexible. Thus, RTqPCR is clinically applicable for detecting patient subgroups with specific prognostic characteristics. Gene expression scores based on RTqPCR have demonstrated being useful for classifying tumors and predicting prognosis, being even approved as prognostic tools in clinical practice (360). Moreover, this technology is a well-implemented methodology in our group for biomarkers' research, previously reporting angiogenesis and immune checkpoint scores for NSCLC (361,362).

However, although this CSCs score seems promising, the development of target therapies against this tumor population is essential to prevent relapse of patients and improve their future outcome. We found that the genes constituting the CSCs score,

CDKN1A, *ITGA6* and *SNAI1*, were significantly overexpressed in tumorspheres with a more aggressive phenotype and higher resistant to chemotherapeutic agents than their adherent counterparts, so we decided to assess if the molecules encoded by these genes could be potential targets against CSCs.

6. PHARMACOLOGICAL INHIBITION OF CSCs POTENTIAL TARGETS

In order to determine the value of Snail, Integrin $\alpha 6$, and p21 as potential therapeutic targets against CSCs, we analyzed the effect of 10 μ M Curcumin (integrin $\alpha 6\beta 4$ inhibitor), 10 μ M Apigenin (Snail pathway inhibitor), and 10 μ M UC2288 (p21 inhibitor) in three primary cultures derived from patients (FIS302, FIS320, and FIS343) and three cell lines (NCI-H1650, HCC827, and PC9). Cell viability results after 48 hours of exposure to each inhibitor are displayed in Figure 47. Cell viability was significantly lower in tumorspheres from FIS320, FIS343, NCI-H1650, and PC9 after treatment with curcumin, although more than 50% were alive in all cases except for PC9 tumorspheres. Regarding Apigenin, cell viability was high in both conditions after 48 hours, although significantly lower in tumorspheres than in adherent cells for all the cultures tested except for FIS343, which showed no significant differences. UC2288 achieved the highest reduction in cell viability after exposure, being it significantly lower in tumorspheres than in adherent cells from all cases except for FIS320. To date, curcumin, also known as diferuloylmethane, is the only molecule which has been reported to target integrin $\alpha 6$ (363). It is a highly pleiotropic molecule that exhibits antibacterial, anti-inflammatory, hypoglycemic, antioxidant, wound-healing, and antimicrobial activities (364). Due to these properties, curcumin has been investigated for the treatment and supportive care of clinical conditions including proteinuria, breast cancer, multiple myeloma, depression, and NSCLC. However, it has been reported as pan-assay interference compound, which often gives false positive results in high-throughput screens (365), challenging its therapeutic efficacy. We also analyzed the effect of Apigenin as a Snail pathway inhibitor, since no specific molecules against the transcription factor Snail are available (366). Apigenin, known chemically as 4',5,7-trihydroxyflavone, belongs to the flavone subclass and is abundant in vegetables and

fruits. It is reported to have anti-inflammatory, antioxidant, anti-telomerase, antidepressant, and anti-cancer properties (367). Moreover, it is also known for its low toxicity, so combinatorial strategies have been proposed to enhance the anti-cancer effect of Apigenin on various cancers and its use as an adjuvant chemotherapeutic agent to overcome cancer drug resistance or to alleviate other adverse effects of chemotherapy (368). Simultaneously, a small molecule structurally related to sorafenib, UC2288, was used to target p21. This molecule selectively downregulates the expression of p21 and has no significant effect on the activities of Raf kinases, VEGFR2 kinase and ERK (369). We consider that the results obtained after the inhibition of these molecules are promising, but more experiments are needed in order to determine the depleting effect that this molecule have in their targets and trying different approaches, such as RNA interference, CRISPR or small interfering RNA, to obtain robust results and select a proper approach before trying their efficacy *in vivo*.

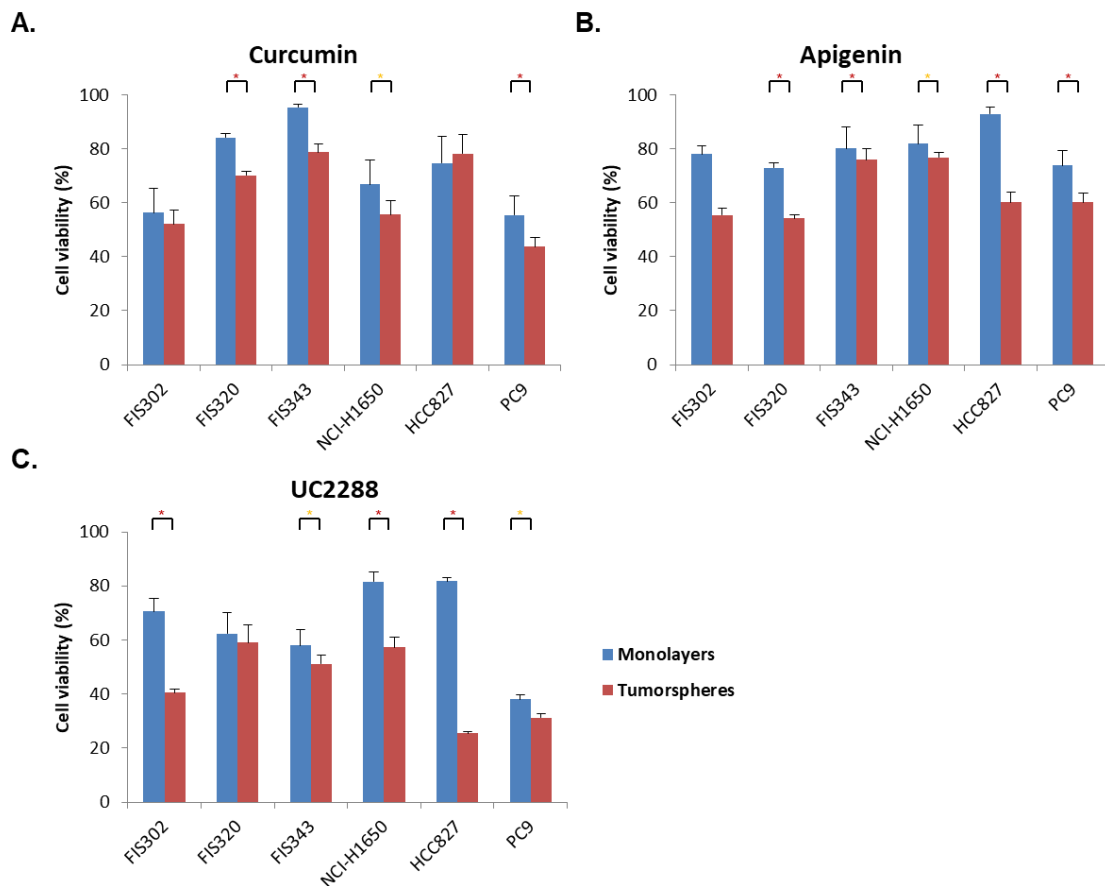


Figure 47. Cell viability of FIS302, FIS320, and FIS343 patients and NCI-H1650, PC9 and HCC827 cell lines after 48 hours of exposure to (A) curcumin, (B) apigenin and (C) UC2288 10 μ M. Cell viability of paired adherent cells and tumorspheres was statistically analyzed using non-parametric Wilcoxon signed-rank test. * $p < 0.05$. ** $p < 0.01$.

7. INTEGRATION OF RESULTS

Cancer is a generic term that defines a large group of diseases that can affect almost any part of the body. Indeed, there are more than 100 types of cancer, and subtypes of tumors can be found in each specific organ. Among them, lung cancer is the most commonly diagnosed and the leading cause of death cancer-related worldwide, with a five-year survival rate around 15% after diagnosis. There are two main reasons why lung cancer has such a poor prognosis. Firstly, it is mainly diagnosed in advanced states, when curative surgery is no longer possible, and secondly, patients frequently develop resistance against the rest of treatments currently used in clinical practice. In that sense, it is known that cancer cells are heterogeneous, causing that some subclones in tumors are more affected by certain treatments than others. Indeed, there is strong evidence pointing out that elevated resistance against cancer therapies and recurrence are linked to CSCs, a population of tumor cells with stem-like properties for which many questions remain unexplained.

In this study, we processed 20 tumor samples from early-stage resected NSCLC patients and successfully established eight primary cultures using two conditions: sphere-forming assays for CSCs enrichment and standard adherent conditions for their corresponding control counterparts. We found that external factors, such as fibroblast presence or tissue preservation, contribute more to the success in the establishment of cultures than clinicopathological variables, like tumor histology or the mutational status. Remarkably, phenotypical changes between primary cultures were found. Adherent-cultured cells show either tight growth, forming colonies or abundant cell-cell interactions in the form of filopodia and lamellipodia or isolated growth, whereas tumorspheres were spherical and tight or more irregularly-shaped and loose. Regardless of the shape, tumorspheres from all cases showed unlimited exponential growth, differentiation and invasion capacities and compared to their control counterparts, showed greater tumor development potential when injected in NOD/SCID mice and higher resistance to classical chemotherapeutic agents, including cisplatin, docetaxel, paclitaxel, vinorelbine and pemetrexed.

We also analyzed the gene expression profile of tumorspheres and adherent cells, trying to identify genes and molecules which could have prognostic role or be potential therapeutic targets against CSCs. In this part of the study, we also included commercial cell lines in order to compare their expression with that from primary cultures. The relative gene expression of potential lung CSCs markers, detoxifying enzymes, pluripotency and cell cycle regulators, invasion promoters, and components of Notch, Wnt and Hedgehog signaling pathways was determined by RTqPCR, which is considered the gold standard in gene expression quantification with major advantages like its speed, sensitivity, and the low amount of RNA required. Using Wilcoxon signed-rank test and PCA analyses, *ALDH1A1*, *KLF4*, *NANOG*, *CD44*, *THY1*, *CDKN1A*, *JUNB*, *MDM2*, *MMP9*, *SNAI1*, *ITGA6*, *NOTCH1*, *NOTCH3*, *DLL4*, *JAG1*, *CTNNB1*, and *GSK3B* were found significantly overexpressed in tumorspheres and the gene expression from immortalized cell lines was found more heterogeneous than that from primary cultures. At that point, we built a logistic regression model considering the 17 statistically significant overexpressed genes with the purpose of reducing the data dimensionality and removing possible collinear expression of genes. Based on this model, *CDKN1A*, *NOTCH3*, *CD44*, *ITGA6*, *NANOG*, and *SNAI1* were selected as the best contributors to separate CSCs from monolayers, so the expression of the proteins encoded by these genes was selected for further analysis. Immunoblotting analysis revealed that these are significantly higher in tumorspheres than in adherent-cultured cells from ADC patients, in contrast to tumorspheres from SCC patients, which showed greater variability. Immunofluorescence analysis confirmed the higher expression of these molecules in tumorspheres and showed differential localization pattern between both culture conditions, which are related to tumor cells with more aggressive phenotypes.

In the light of these results, the expression of the 17 genes significantly overexpressed in tumorspheres was correlated *in silico* with the RFS and OS of a cohort of 661 early-stage NSCLC patients from TCGA. Individual survival analysis revealed associations between elevated *CDKN1A*, *ITGA6*, *JAG1*, *JUNB*, *SNAI1*, and *THY1* and worse outcomes. Furthermore, following a mathematical model, a score with the expression of *CDKN1A*, *ITGA6*, and *SNAI1* was constructed, since these genes showed the highest association with mortality. The score was named CSCs score and was demonstrated to

be an independent prognostic biomarker for OS in NSCLC patients. Because of the significant differences reported for lung cancer histologies and based on our previous results, survival analyses were also applied according to tumor histology. We found that the CSCs score proposed is an independent prognostic biomarker for ADC patients, but not for SCC patients. Furthermore, we validated the prognostic value of the CSCs score proposed in an independent cohort of 245 patients from *Consortio Hospital General Universitario de Valencia*, finding that high levels of the CSCs score was associated with worse OS in ADC patients.

Finally, in order to determine the value of p21, Integrin $\alpha 6$ and Snail (encoded by *CDKN1A*, *ITGA6*, and *SNAI1*, respectively) as potential therapeutic targets against CSCs, we analyzed the effect of Curcumin, Apigenin and UC2288 in adherent cells and tumorspheres. We found a higher reduction of cell viability in tumorspheres for the three compounds tested which indicates that these molecules constitute promising targets that could change the current management of lung ADC patients.

In summary, our approach allowed the establishment of long-term lung CSCs cultures, being a powerful tool for identifying the molecular alterations present in this tumor-cell population and providing new insight into the field of CSCs in NSCLC. Tumorspheres can be used for CSCs enrichment and a common set of genes were found to be significantly and consistently overexpressed in them. We created a gene score based on molecules overexpressed in this tumor-cell population which predicts worse outcome in two independent cohorts of patients. This score represents an independent prognostic biomarker that can help in future clinical practice, since patients with high values can be followed up closely after a successful surgery to prevent their relapse and death. Developing new therapeutic strategies against these molecules could have major implications in patients' survival, since targeting cells with the ability to resist current cancer treatment is an unresolved matter and it will improve the future outcome of lung cancer patients.

V. CONCLUSIONS

1. Sphere-forming assays allow the establishment of long-term cultures enriched in lung CSCs from resected NSCLC tissue and cell lines in a simple and cost-effective way.
2. Lung tumorspheres have self-renewal and unlimited exponential growth abilities, differentiation and invasion capacities, higher tumorigenic potential *in vivo* and higher resistance to chemotherapeutic agents than adherent cells *in vitro*.
3. Tumorspheres also recapitulate tumor histology and allow isolating DNA, RNA and protein to perform extensive expression analysis and profiling, becoming suitable *in vitro* models for CSCs study.
4. The expression of cytoprotective enzymes (*ALDH1A1*), pluripotency inducers (*KLF4* and *NANOG*), cell cycle regulators (*CDKN1A*, *JUNB* and *MDM2*), metastasis-related genes (*CD44*, *THY1*, *MMP9*, *SNAI1* and *ITGA6*), and components of Notch (*NOTCH1*, *NOTCH3*, *DLL4* and *JAG1*) and Wnt (*CTNNB1* and *GSK3B*) was significantly increased in lung tumorspheres.
5. According to this gene expression analysis, *CDKN1A*, *ITGA6*, *NOTCH3*, *NANOG*, *SNAI1*, and *CD44* were the major contributors to distinguish tumorspheres from adherent cells. In consonance, the proteins encoded by these genes were found increased in tumorspheres from ADC patients and showed in some cases differential localization patterns compared to their adherent control counterparts.
6. We demonstrated that a CSCs score, defined by *CDKN1A*, *SNAI1* and *ITGA6* expressions, is an independent prognostic factor for early-stage NSCLC patients' survival, so that patients with high CSCs score have significantly shorter OS.
7. Pharmacological inhibition of proteins encoded by *CDKN1A*, *SNAI1* and *ITGA6* revealed that these molecules are promising therapeutic targets against lung CSCs.

VI. REFERENCES

1. Hanahan D, Weinberg RA, Yang Q, Goldstein A., Tucker M., Walker G., et al. The hallmarks of cancer. *Cell*. 2000 Jan;100(1):57–70.
2. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012 Apr;12(5):323–34.
3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011 Mar 4;144(5):646–74.
4. Hayflick L. Mortality and immortality at the cellular level. A review. *Biochemistry (Mosc)*. 1997 Nov;62(11):1180–90.
5. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. *EurJCancer*. 1997 Apr;33(5):787–91.
6. Folkman J. Fundamental concepts of the angiogenic process. *Curr Mol Med*. 2003;3:643–51.
7. 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 Nov;68(6):394–424.
8. Bosetti C, Bertuccio P, Malvezzi M, Levi F, Chatenoud L, Negri E, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. *Ann Oncol*. 2013 Oct;24(10):2657–71.
9. Ferlay J, Soerjomataram I I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015 Sep;136(5):E359-86.
10. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin*. 2015 Mar;65(2):87–108.
11. de Groot P, Munden RF, Fuller C, al. et. Lung Cancer Epidemiology, Risk Factors, and Prevention. *Radiol Clin North Am*. 2012 Mar;50(5):863–76.
12. Fry JS, Lee PN, Forey BA, Coombs KJ. How rapidly does the excess risk of lung cancer decline following quitting smoking? A quantitative review using the negative exponential model. *Regul Toxicol Pharmacol*. 2013 Nov;67(1):13–26.

References

13. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW, et al. Cancer genome landscapes. *Science*. 2013 Mar;339(6127):1546–58.
14. Jemal A, Thun MJ, Ries LA, Howe HL, Weir HK, Center MM, et al. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. *J Natl Cancer Inst*. 2008 Dec;100(23):1672–94.
15. Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR, et al. Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J Clin Oncol*. 2003 Mar;21(5):921–6.
16. Bilello KS, Murin S, Matthay RA. Epidemiology, etiology, and prevention of lung cancer. *Clin Chest Med*. 2002;23:1–25.
17. Tyczynski JE, Bray F, Parkin DM. Lung cancer in Europe in 2000: epidemiology, prevention, and early detection. *Lancet Oncol*. 2003;4(1):45–55.
18. Sawyers CL. AACR Cancer Progress Report 2013. *Clin Cancer Res*. 2013;19(1):1–98.
19. Morgensztern D, Ng SH, Gao F, Govindan R. Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. *J Thorac Oncol*. 2010;5:29–33.
20. Cuaron J, Dunphy M, Rimner A. Role of FDG-PET scans in staging, response assessment, and follow-up care for non-small cell lung cancer. *Front Oncol*. 2012;2:208.
21. Jantus-Lewintre E, Uso M, Sanmartin E, Camps C, Jantus-Lewintre M, Usó E, et al. Update on biomarkers for the detection of lung cancer. *Lung Cancer Targets Ther*. 2012 Jun;3:21–9.
22. Shepherd FA, Crowley J, Van Houtte P, Postmus PE, Carney D, Chansky K, et al. The International Association for the Study of Lung Cancer lung cancer staging project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. *J Thorac Oncol*. 2007 Dec;2(12):1067–77.
23. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. *Chest*. 2017 Jan;151(1):193–203.
24. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The

- IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2016 Jan;11(1):39–51.
25. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol.* 2015 Sep;10:1243-60.
26. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol.* 2011 Feb;6(2):244–85.
27. Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nature reviews.* 2017;17:765.
28. Cancer Genome Atlas Research Network, Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014 Jul;511(7511):543–50.
29. Gerber DE, Gandhi L, Costa DB. Management and future directions in non-small cell lung cancer with known activating mutations. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meet.* 2014;e353-65.
30. Barlesi F, Mazieres J, Merlio J-P, Debieuvre D, Mosser J, Lena H, et al. Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the French Cooperative Thoracic Intergroup (IFCT). *Lancet.* 2016 Apr;387(10026):1415–26.
31. Scagliotti G V, Selvaggi G, Novello S, Hirsch FR. The biology of epidermal growth factor receptor in lung cancer. *Clin.Cancer Res.* 2004;10:4227s-4232s.
32. Okabe T, Okamoto I, Tamura K, Terashima M, Yoshida T, Satoh T, et al. Differential constitutive activation of the epidermal growth factor receptor in non-small cell lung cancer cells bearing EGFR gene mutation and amplification. *Cancer Res.* 2007 Mar;67(5):2046–53.
33. Yasuda H, Kobayashi S, Costa DB. EGFR exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *Lancet Oncol.* 2012 Jan;13(1):e23-31.

34. Kempf E, Rousseau B, Besse B, Paz-Ares L. KRAS oncogene in lung cancer: focus on molecularly driven clinical trials. *Eur Respir Rev.* 2016 Mar;25(139):71–6.
35. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med.* 2012 Feb;18(3):378–81.
36. Peters S, Taron M, Bubendorf L, Blackhall F, Stahel R. Treatment and detection of ALK-rearranged NSCLC. *Lung Cancer.* 2013;81(2):145–54.
37. Choi YL, Soda M, Yamashita Y, Ueno T, Takashima J, Nakajima T, et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. *N Engl J Med.* 2010 Oct;363(18):1734–9.
38. Bergethon K, Shaw AT, Ou S-HI, Katayama R, Lovly CM, McDonald NT, et al. ROS1 rearrangements define a unique molecular class of lung cancers. *J Clin Oncol.* 2012 Mar;30(8):863–70.
39. Vaishnavi A, Capelletti M, Le AT, Kako S, Butaney M, Ercan D, et al. Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer. *Nat Med.* 2013 Oct;19(11):1469–72.
40. Paik PK, Drilon A, Fan P-D, Yu H, Rekhtman N, Ginsberg MS, et al. Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. *Cancer Discov.* 2015 Aug;5(8):842–9.
41. Drilon A, Rekhtman N, Ladanyi M, Paik P. Squamous-cell carcinomas of the lung: emerging biology, controversies, and the promise of targeted therapy. *Lancet Oncol.* 2012 Oct;13(10):e418-26.
42. Rosell R, Karachaliou N. Large-scale screening for somatic mutations in lung cancer. *Lancet (London, England).* 2016 Apr;387(10026):1354–6.
43. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, Getz G, Sougnez C, Imielinski M, Helman E, Hernandez B, Pho NH, Meyerson M, Chu A, Chun HJ, Mungall AJ, Pleasance E, Robertson A, Sipahimala TE, Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012 Sep;489(7417):519–25.
44. Robinson CG, DeWees TA, El Naqa IM, Creach KM, Olsen JR, Crabtree TD, et al. Patterns of failure after stereotactic body radiation therapy or lobar resection for clinical stage I

- non-small-cell lung cancer. *J Thorac Oncol*. 2013 Feb;8(2):192–201.
45. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular Determinants of Response to Anti–Programmed Cell Death (PD)-1 and Anti–Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non–Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J Clin Oncol*. 2018 Mar;36(7):633–41.
 46. Yu HA, Planchard D, Lovly CM. Sequencing Therapy for Genetically Defined Subgroups of Non-Small Cell Lung Cancer. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meet*. 2018 May;38:726–39.
 47. Chong CR, Janne PA. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med*. 2013 Nov;19(11):1389–400.
 48. Gainor JF, Shaw AT. Emerging paradigms in the development of resistance to tyrosine kinase inhibitors in lung cancer. *J Clin Oncol*. 2013 Nov;31(31):3987–96.
 49. Ohashi K, Maruvka YE, Michor F, Pao W. Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. *J Clin Oncol*. 2013 Mar;31(8):1070–80.
 50. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res*. 2013 Apr;19(8):2240–7.
 51. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med*. 2015 Jun;21(6):560–2.
 52. Planchard D, Loriot Y, Andre F, Gobert A, Auger N, Lacroix L, et al. EGFR-independent mechanisms of acquired resistance to AZD9291 in EGFR T790M-positive NSCLC patients. *Ann Oncol Off J Eur Soc Med Oncol*. 2015 Oct;26(10):2073–8.
 53. Ichihara E, Westover D, Meador CB, Yan Y, Bauer JA, Lu P, et al. SFK/FAK Signaling Attenuates Osimertinib Efficacy in Both Drug-Sensitive and Drug-Resistant Models of EGFR-Mutant Lung Cancer. *Cancer Res*. 2017 Jun;77(11):2990–3000.
 54. Soria JC, Mauguen A, Reck M, Sandler AB, Saijo N, Johnson DH, et al. Systematic review and meta-analysis of randomised, phase II/III trials adding bevacizumab to platinum-based chemotherapy as first-line treatment in patients with advanced non-small-cell lung

- cancer. *Ann Oncol*. 2013 Jan;24(1):20–30.
55. Doroshow DB, Herbst RS. Treatment of Advanced Non-Small Cell Lung Cancer in 2018. *JAMA Oncol*. 2018 Apr;4(4):569–70.
56. Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, et al. Mechanisms of Acquired Crizotinib Resistance in ALK-Rearranged Lung Cancers. *Sci Transl Med*. 2012 Feb;4(120):120ra17-120ra17.
57. Iacono D, Chiari R, Metro G, Bennati C, Bellezza G, Cenci M, et al. Future options for ALK-positive non-small cell lung cancer. *Lung Cancer*. 2015 Mar;87(3):211–9.
58. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Discov*. 2016 Oct;6(10):1118–33.
59. Doebele RC, Pilling AB, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, et al. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. *Clin Cancer Res*. 2012 Mar;18(5):1472–82.
60. Tripathy D, Bardia A, Sellers WR. Ribociclib (LEE011): Mechanism of Action and Clinical Impact of This Selective Cyclin-Dependent Kinase 4/6 Inhibitor in Various Solid Tumors. *Clin Cancer Res*. 2017 Jul;23(13):3251–62.
61. Zou HY, Friboulet L, Kodack DP, Engstrom LD, Li Q, West M, et al. PF-06463922, an ALK/ROS1 Inhibitor, Overcomes Resistance to First and Second Generation ALK Inhibitors in Preclinical Models. *Cancer Cell*. 2015 Jul;28(1):70–81.
62. Lim SM, Kim HR, Lee J-S, Lee KH, Lee Y-G, Min YJ, et al. Open-Label, Multicenter, Phase II Study of Ceritinib in Patients With Non-Small-Cell Lung Cancer Harboring ROS1 Rearrangement. *J Clin Oncol*. 2017 Aug;35(23):2613–8.
63. Davare MA, Vellore NA, Wagner JP, Eide CA, Goodman JR, Drilon A, et al. Structural insight into selectivity and resistance profiles of ROS1 tyrosine kinase inhibitors. *Proc Natl Acad Sci U S A*. 2015 Sep;112(39):E5381-90.
64. Drilon A, Siena S, Ou S-HI, Patel M, Ahn MJ, Lee J, et al. Safety and Antitumor Activity of the Multitargeted Pan-TRK, ROS1, and ALK Inhibitor Entrectinib: Combined Results from Two Phase I Trials (ALKA-372-001 and STARTRK-1). *Cancer Discov*. 2017 Apr;7(4):400–9.

65. Brown NA, Aisner DL, Oxnard GR. Precision Medicine in Non-Small Cell Lung Cancer: Current Standards in Pathology and Biomarker Interpretation. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meet.* 2018 May;38:708–15.
66. Raju S, Joseph R, Sehgal S. Review of checkpoint immunotherapy for the management of non-small cell lung cancer. *ImmunoTargets Ther.* 2018;7:63–75.
67. Greaves M, Maley CC. Clonal Evolution in Cancer. *Nature.* 2012;481(7381):306–13.
68. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature.* 2013;501:338–45.
69. Shackleton M, Quintana E, Fearon E, Morrison S. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell.* 2009;138(5):822–9.
70. Nowell PC. The clonal evolution of tumor cell populations. *Science.* 1976;194(4260):23–8.
71. Yong Wang, Jill Water, Marco L. Leung, Anna Unruh, Whijae Roh, Xiuqing Shi¹ K, Chen, Paul Scheet, Selina Vattathil, Han Liang, Asha Multani, Hong Zhang R, Zhao, Franziska Michor, Funda Meric-Bernstam and NEN. Clonal Evolution in Breast Cancer Revealed by Single Nucleus Genome Sequencing. *Nature.* 2014;512(7513):155–60.
72. Landau D-A, Carter SL, Getz G, Wu CJ. Clonal evolution in hematologic malignancies and therapeutic implications. *Leukemia.* 2014;28(1):34–43.
73. Dick JE. Stem cell concepts renew cancer research. *Blood.* 2008;112(13):4793–807.
74. Sourisseau T, Hassan KA, Wistuba I, Penault-Llorca F, Adam J, Deutsch E, et al. Lung cancer stem cell: fancy conceptual model of tumor biology or cornerstone of a forthcoming therapeutic breakthrough? *J Thorac Oncol.* 2014;9(1):7–17.
75. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med.* 1997;3(7):730–7.
76. Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. *Cancer Res.* 2003;63:5821–8.
77. Smalley M, Ashworth A. Stem cells and breast cancer: A field in transit. *Nat Rev Cancer.* 2003 Nov;3(11):832–44.

78. Kim CFB, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell*. 2005;121(6):823–35.
79. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. *Nature*. 2007 Jan;445(7123):111–5.
80. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, et al. Identification of pancreatic cancer stem cells. *Cancer Res*. 2007;67:1030–7.
81. Cheng L, Zhang S, Davidson DD, Montironi R, Lopez-beltran A. Implications of Cancer Stem Cells for Cancer Therapy. *Cancer Drug Discov Dev*. 2009;(1):255–62.
82. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013;501(7467):328–37.
83. Kreso A, Dick JE. Evolution of the Cancer Stem Cell Model. *Cell Stem Cell*. 2014;14(3):275–91.
84. Bu Y, Cao D. The origin of cancer stem cells. *Front Biosci*. 2012;4:819–30.
85. Pattabiraman DR, Weinberg RA. Tackling the cancer stem cells - what challenges do they pose? *Nat Rev Drug Discov*. 2014 Jul;13(7):497–512.
86. Knoblich JA. Mechanisms of asymmetric stem cell division. *Cell*. 2008 Feb;132(4):583–97.
87. Ragoussis J. Regulators of Asymmetric Cell Division in Breast Cancer. *Trends in cancer*. 2018 Dec;4(12):798–801.
88. Seyfrid M, Bobrowski D, Bakhshinyan D, Tatari N, Venugopal C, Singh SK. In Vitro Self-Renewal Assays for Brain Tumor Stem Cells. *Methods Mol Biol*. 2019;1869:79–84.
89. Moore N, Lyle S. Quiescent, slow-cycling stem cell populations in cancer: a review of the evidence and discussion of significance. *J Oncol*. 2011;2011.
90. Chen W, Dong J, Haiech J, Kilhoffer M-C, Zeniou M. Cancer Stem Cell Quiescence and Plasticity as Major Challenges in Cancer Therapy. *Stem Cells Int*. 2016;2016:1740936.
91. Moitra K, Lou H, Dean M. Multidrug efflux pumps and cancer stem cells: insights into multidrug resistance and therapeutic development. *Clin Pharmacol Ther*. 2011 Apr;89(4):491–502.

92. Moitra K. Overcoming Multidrug Resistance in Cancer Stem Cells. *Biomed Res Int.* 2015;2015:635745.
93. Dontu G, Liu S, Wicha MS. Stem cells in mammary development and carcinogenesis: implications for prevention and treatment. *Stem Cell Rev.* 2005;1(3):207–13.
94. Takebe N, Miele L, Harris PJ, Jeong W, Bando H, Yang SX, et al. Targeting Notch, Hedgehog, and Wnt Pathways in cancer stem cells: clinical update. *Nat Rev Clin Oncol.* 2015;12(8):445–64.
95. Han L, Shi S, Gong T, Zhang Z, Sun X. Cancer stem cells: therapeutic implications and perspectives in cancer therapy. *Acta Pharm Sin B.* 2013;3(2):65–75.
96. Reynolds BA, Weiss S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science.* 1992 Mar;255(5052):1707–10.
97. Bielecka ZF, Maliszewska-Olejniczak K, Safir IJ, Szczylik C, Czarnecka AM. Three-dimensional cell culture model utilization in cancer stem cell research. *Biol Rev Camb Philos Soc.* 2016 Aug;92(3):1505-1520.
98. Eramo A, Lotti F, Sette G, Pilozzi E, Biffoni M, Di VA, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. *Cell Death Differ.* 2008;15:504–14.
99. Qiu X, Wang Z, Li Y, Miao Y, Ren Y, Luan Y. Characterization of sphere-forming cells with stem-like properties from the small cell lung cancer cell line H446. *Cancer Lett.* 2012 Oct;323(2):161–70.
100. Ghani FI, Yamazaki H, Iwata S, Okamoto T, Aoe K, Okabe K, et al. Identification of cancer stem cell markers in human malignant mesothelioma cells. *Biochem Biophys Res Commun.* 2011 Jan;404(2):735–42.
101. Zakaria N, Yusoff NM, Zakaria Z, Lim MN, Baharuddin PJN, Fakiruddin KS, et al. Human non-small cell lung cancer expresses putative cancer stem cell markers and exhibits the transcriptomic profile of multipotent cells. *BMC Cancer.* 2015;15:84.
102. Ouadah Y, Rojas ER, Riordan DP, Capostagno S, Kuo CS, Krasnow MA. Rare Pulmonary Neuroendocrine Cells Are Stem Cells Regulated by Rb, p53, and Notch. *Cell.* 2019 Oct;179(2):403-416.e23.
103. Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, et al. Glycine

- decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell*. 2012 Jan;148(1–2):259–72.
104. Zhang D-G, Jiang A-G, Lu H-Y, Zhang L-X, Gao X-Y. Isolation, cultivation and identification of human lung adenocarcinoma stem cells. *Oncol Lett*. 2015;9(1):47–54.
105. Hashida S, Yamamoto H, Shien K, Miyoshi Y, Ohtsuka T, Suzawa K, et al. Acquisition of cancer stem cell-like properties in non-small cell lung cancer with acquired resistance to afatinib. *Cancer Sci*. 2015 Oct;106(10):1377–84.
106. Chaichana K, Zamora-Berridi G, Camara-Quintana J, Quinones-Hinojosa A. Neurosphere assays: growth factors and hormone differences in tumor and nontumor studies. *Stem Cells*. 2006 Dec;24(12):2851–7.
107. Pastrana E, Silva-Vargas V, Doetsch F. Eyes Wide Open: A Critical Review of Sphere-Formation as an Assay For Stem Cells. *Cell Stem Cell*. 2011 May 6;8(5):486–98.
108. Yakisich JS, Azad N, Venkatadri R, Kulkarni Y, Wright C, Kaushik V, et al. Formation of Tumorspheres with Increased Stemness without External Mitogens in a Lung Cancer Model. *Stem Cells Int*. 2016 Jan 6;2016:5603135.
109. Calvet CY, Andre FM, Mir LM. The culture of cancer cell lines as tumorspheres does not systematically result in cancer stem cell enrichment. *PLoS One*. 2014;9(2):e89644.
110. Weiswald L-B, Bellet D, Dangles-Marie V. Spherical cancer models in tumor biology. *Neoplasia*. 2015 Jan;17(1):1–15.
111. Singec I, Knoth R, Meyer RP, Maciaczyk J, Volk B, Nikkhah G, et al. Defining the actual sensitivity and specificity of the neurosphere assay in stem cell biology. *Nat Methods*. 2006 Oct;3(10):801–6.
112. Rota LM, Lazzarino DA, Ziegler AN, LeRoith D, Wood TL. Determining mammosphere-forming potential: application of the limiting dilution analysis. *J Mammary Gland Biol Neoplasia*. 2012 Jun;17(2):119–23.
113. Qureshi-Baig K, Ullmann P, Rodriguez F, Frاسquilho S, Nazarov P V, Haan S, et al. What Do We Learn from Spheroid Culture Systems? Insights from Tumorspheres Derived from Primary Colon Cancer Tissue. *PLoS One*. 2016;11(1):e0146052.
114. Malanchi I, Santamaria-Martinez A, Susanto E, Peng H, Lehr H-AA, Delaloye J-FF, et al.

- Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature*. 2012 Jan;481(7379):85–9.
115. Seo J, Park S-J, Kim J, Choi S-J, Moon S-H, Chung H-M. Effective method for the isolation and proliferation of primary lung cancer cells from patient lung tissues. *Biotechnol Lett*. 2013 Aug;35(8):1165–74.
116. Kurpios NA, Girgis-Gabardo A, Hallett RM, Rogers S, Gludish DW, Kockeritz L, et al. Single unpurified breast tumor-initiating cells from multiple mouse models efficiently elicit tumors in immune-competent hosts. *PLoS One*. 2013;8(3):e58151.
117. Pal M, Chen H, Lee BH, Lee JYH, Yip YS, Tan NS, et al. Epithelial-mesenchymal transition of cancer cells using bioengineered hybrid scaffold composed of hydrogel/3D-fibrous framework. *Sci Rep*. 2019 Jun;9(1):8997.
118. Sreepadmanabh M, Toley BJ. Investigations into the cancer stem cell niche using in-vitro 3-D tumor models and microfluidics. *Biotechnol Adv*. 2018 Jul;36(4):1094–110.
119. Aboulkheyr Es H, Montazeri L, Aref AR, Vosough M, Baharvand H. Personalized Cancer Medicine: An Organoid Approach. *Trends Biotechnol*. 2018 Apr;36(4):358–71.
120. Medema JP. Cancer stem cells: the challenges ahead. *Nat Cell Biol*. 2013;15(4):338–44.
121. Wang J, Li Z, White J, Zhang L. Lung cancer stem cells and implications for future therapeutics. *Cell Biochem Biophys*. 2014 Jul;69(3):389–98.
122. Mannelli G, Gallo O. Cancer stem cells hypothesis and stem cells in head and neck cancers. *Cancer Treat Rev*. 2012 Aug;38(5):515–39.
123. Wang P, Gao Q, Suo Z, Munthe E, Solberg S, Ma L, et al. Identification and Characterization of Cells with Cancer Stem Cell Properties in Human Primary Lung Cancer Cell Lines. Tang DG, editor. *PLoS One*. 2013 Mar 4;8(3):e57020.
124. Yuan X, Wu H, Han N, Xu H, Chu Q, Yu S, et al. Notch signaling and EMT in non-small cell lung cancer : biological significance and therapeutic application. *J Hematol Oncol*. 2014 Dec 5;7(87):1–10.
125. Koren A, Rijavec M, Kern I, Sodja E, Korosec P, Cufer T. BMI1, ALDH1A1, and CD133 Transcripts Connect Epithelial-Mesenchymal Transition to Cancer Stem Cells in Lung Carcinoma. *Stem Cells Int*. 2016 Dec 7;2016:1.

126. Reuben JM, Lee B-N, Gao H, Cohen EN, Mego M, Giordano A, et al. Primary breast cancer patients with high risk clinicopathologic features have high percentages of bone marrow epithelial cells with ALDH activity and CD44(+)CD24lo cancer stem cell phenotype. *Eur J Cancer*. 2011 Jul;47(10):1527–36.
127. Pang R, Law WL, Chu ACY, Poon JT, Lam CSC, Chow AKM, et al. A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. *Cell Stem Cell*. 2010 Jun;6(6):603–15.
128. Zhang S, Han Z, Jing Y, Tao S, Li T, Wang H, et al. CD133(+)CXCR4(+) colon cancer cells exhibit metastatic potential and predict poor prognosis of patients. *BMC Med*. 2012 Aug;10:85.
129. Todaro M, Gaggianesi M, Catalano V, Benfante A, Iovino F, Biffoni M, et al. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell*. 2014 Mar;14(3):342–56.
130. Kahlert UD, Bender NO, Maciaczyk D, Bogiel T, Bar EE, Eberhart CG, et al. CD133/CD15 defines distinct cell subpopulations with differential in vitro clonogenic activity and stem cell-related gene expression profile in in vitro propagated glioblastoma multiforme-derived cell line with a PNET-like component. *Folia Neuropathol*. 2012;50(4):357–68.
131. Erhart F, Blauensteiner B, Zirkovits G, Printz D, Soukup K, Klingenbrunner S, et al. Gliomasphere marker combinatorics: multidimensional flow cytometry detects CD44+/CD133+/ITGA6+/CD36+ signature. *J Cell Mol Med*. 2019 Jan;23(1):281–92.
132. Wang R, Sun Q, Wang P, Liu M, Xiong S, Luo J, et al. Notch and Wnt/beta-catenin signaling pathway play important roles in activating liver cancer stem cells. *Oncotarget*. 2016 Feb;7(5):5754–68.
133. Qiu L, Li H, Fu S, Chen X, Lu L. Surface markers of liver cancer stem cells and innovative targeted-therapy strategies for HCC. *Oncol Lett*. 2018 Feb;15(2):2039–48.
134. Sarvi S, Mackinnon AC, Avlonitis N, Bradley M, Rintoul RC, Rassl DM, et al. CD133+ cancer stem-like cells in small cell lung cancer are highly tumorigenic and chemoresistant but sensitive to a novel neuropeptide antagonist. *Cancer Res*. 2014 Mar;74(5):1554–65.
135. Yan X, Luo HU, Zhou X, Zhu B, Wang Y, Bian X. Identification of CD90 as a marker for lung cancer stem cells in A549 and H446 cell lines. *Oncol Rep*. 2013;30(6):2733–40.

136. Karimi-Busheri F, Zadorozhny V, Li T, Lin H, Shawler DL, Fakhrai H. Pivotal role of CD38 biomarker in combination with CD24, EpCAM, and ALDH for identification of H460 derived lung cancer stem cells. *J Stem Cells*. 2011;6(1):9–20.
137. Su J, Wu S, Tang W, Qian H, Zhou H, Guo T. Reduced SLC27A2 induces cisplatin resistance in lung cancer stem cells by negatively regulating Bmi1-ABCG2 signaling. *Mol Carcinog*. 2016 Nov;55(11):1822–32.
138. Boonyaratanakornkit JB, Yue L, Strachan LR, Scalapino KJ, LeBoit PE, Lu Y, et al. Selection of tumorigenic melanoma cells using ALDH. *J Invest Dermatol*. 2010 Dec;130(12):2799–808.
139. Fang D, Nguyen TK, Leishear K, Finko R, Kulp AN, Hotz S, et al. A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res*. 2005 Oct;65(20):9328–37.
140. Monzani E, Facchetti F, Galmozzi E, Corsini E, Benetti A, Cavazzin C, et al. Melanoma contains CD133 and ABCG2 positive cells with enhanced tumourigenic potential. *Eur J Cancer*. 2007 Mar;43(5):935–46.
141. Baba T, Convery PA, Matsumura N, Whitaker RS, Kondoh E, Perry T, et al. Epigenetic regulation of CD133 and tumorigenicity of CD133+ ovarian cancer cells. *Oncogene*. 2009 Jan;28(2):209–18.
142. Klemba A, Purzycka-Olewiecka JK, Wcisło G, Czarnecka AM, Lewicki S, Lesyng B, et al. Surface markers of cancer stem-like cells of ovarian cancer and their clinical relevance. *Contemp Oncol*. 2018 Mar;22(1A):48–55.
143. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell*. 2007 Sep;1(3):313–23.
144. Li C, Wu J-J, Hynes M, Dosch J, Sarkar B, Welling TH, et al. c-Met is a marker of pancreatic cancer stem cells and therapeutic target. *Gastroenterology*. 2011 Dec;141(6):2218–2227.e5.
145. Wang YH, Li F, Luo B, Wang XH, Sun HC, Liu S, et al. A side population of cells from a human pancreatic carcinoma cell line harbors cancer stem cell characteristics. *Neoplasma*. 2009;56(5):371–8.

146. Klonisch T, Wiechec E, Hombach-Klonisch S, Ande SR, Wesselborg S, Schulze-Osthoff K, et al. Cancer stem cell markers in common cancers - therapeutic implications. *Trends Mol Med*. 2008 Oct;14(10):450–60.
147. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res*;65:10946–51.
148. Baumann M, Krause M, Thames H, Trott K, Zips D. Cancer stem cells and radiotherapy. *Int J Radiat Biol*. 2009 Jan;85(5):391–402.
149. Li J, Chen J-N, Zeng T-T, He F, Chen S-P, Ma S, et al. CD133+ liver cancer stem cells resist interferon-gamma-induced autophagy. *BMC Cancer*. 2016;16(1):15.
150. Roudi Raheleh, Zahra M, Marzieh E, Fazel S, Ali S, Roudi R, et al. CD44 and CD24 cannot act as cancer stem cell markers in human lung adenocarcinoma cell line A549. *Cell Mol Biol Lett*. 2014;19(1):23.
151. Moro M, Bertolini G, Pastorino U, Roz L, Sozzi G. Combination Treatment with All-Trans Retinoic Acid Tumor-Initiating Cells and Reveals Heterogeneity of Cancer Stem Cell Compartment in Lung Cancer. *J Thorac Oncol*. 2015 Jul;10(7):1027-36.
152. Liu Q-F, Zhang Z-F, Hou G-J, Yang G-Y, He Y. Polymorphisms of the Stem Cell Marker Gene CD133 and the Risk of Lung Cancer in Chinese Population. *Lung*. 2016 Jun;194(3):393–400.
153. Sullivan JP, Spinola M, Dodge M, Raso MG, Behrens C, Gao B, et al. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on Notch signaling. *Cancer Res*. 2010 Dec;70(23):9937–48.
154. Meng X, Li M, Wang X, Wang Y, Ma D. Both CD133+ and CD133– subpopulations of A549 and H446 cells contain cancer-initiating cells. *Cancer Sci*. 2009 Jun 1;100(6):1040–6.
155. Kotiyal S, Bhattacharya S. Lung Cancer Stem Cells and Their Therapeutic Targeting. *Arch Stem Cell Res*. 2015;2(2):1009.
156. Jacob MK, Male H, Diaz EB, Huang C, Farassati F. CD90 (thy1) As A Potential Cancer Stem Cell Markers And Therapeutic Target In Non-Small Cell Lung Cancer. *Am J Respir Crit Care Med*. 2011 May 1;183:A5070.
157. Pirozzi G, Tirino V, Camerlingo R, La Rocca a, Martucci N, Scognamiglio G, et al. Prognostic

- value of cancer stem cells, epithelial-mesenchymal transition and circulating tumor cells in lung cancer. *Oncol Rep.* 2013;29:1763–8.
158. Schnell U, Kuipers J, Giepmans BNG. EpCAM proteolysis: new fragments with distinct functions? *Biosci Rep.* 2013;33(2):e00030.
 159. Patriarca C, Macchi RM, Marschner AK, Mellstedt H. Epithelial cell adhesion molecule expression (CD326) in cancer: a short review. *Cancer Treat Rev.* 2012 Feb;38(1):68–75.
 160. Ni J, Cozzi P, Hao J, Beretov J, Chang L, Duan W, et al. Epithelial cell adhesion molecule (EpCAM) is associated with prostate cancer metastasis and chemo/radioresistance via the PI3K/Akt/mTOR signaling pathway. *Int J Biochem Cell Biol.* 2013 Dec;45(12):2736–48.
 161. Lin S, Sun J, Wu J, Long H, Zhu C, Xiang T, et al. Aberrant microRNAs Expression in CD133(+)/CD326(+) Human Lung Adenocarcinoma Initiating Cells from A549. *Mol Cells.* 2012 Mar 31;33(3):277–83.
 162. Jordan AR, Racine RR, Hennig MJP, Lokeshwar VB. The Role of CD44 in Disease Pathophysiology and Targeted Treatment. *Front Immunol.* 2015 April;6:182.
 163. Yan Y, Zuo X, WEI D, Cells CS. Concise Review: Emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target. *Stem Cells Transl Med.* 2015 Sep;4(9):1033–43.
 164. Sahlberg SH, Spiegelberg D, Glimelius B, Stenerl??w B, Nestor M. Evaluation of cancer stem cell markers CD133, CD44, CD24: Association with AKT isoforms and radiation resistance in colon cancer cells. *PLoS One.* 2014 Apr;9(4):e94621.
 165. de Beca FF, Caetano P, Gerhard R, Alvarenga CA, Gomes M, Paredes J, et al. Cancer stem cells markers CD44, CD24 and ALDH1 in breast cancer special histological types. *J Clin Pathol.* 2013 Mar;66(3):187–91.
 166. Yoon C, Ph D, Park DJ, Ph D, Schmidt B, Thomas J, et al. CD44 expression denotes a subpopulation of gastric cancer cells in which Hedgehog signaling promotes chemotherapy resistance. *Clin Cancer Res.* 2014;20(15):3974–88.
 167. Luo Z, Wu R, Lv L, Li P, Zhang L, Hao Q, et al. Prognostic value of CD44 expression in non-small cell lung cancer: a systematic review. *Int J Clin Exp Pathol.* 2014;7(7):3632–46.
 168. Li L, Qi L, Liang Z, Song W, Liu Y, Wang Y, et al. Transforming growth factor- β 1 induces

- EMT by the transactivation of epidermal growth factor signaling through HA/CD44 in lung and breast cancer cells. *Int J Mol Med*. 2015;36(1):113–22.
169. Ko YH, Won HS, Jeon EK, Hong SH, Roh SY, Hong YS, et al. Prognostic significance of CD44s expression in resected non-small cell lung cancer. *BMC Cancer*. 2011;11(1):340.
170. Margaritescu C, Pirici D, Cherciu I, Barbalan A, Cartana T, Saftoiu A. CD133/CD166/Ki-67 triple immunofluorescence assessment for putative cancer stem cells in colon carcinoma. *J Gastrointest Liver Dis*. 2014 Jun;23(2):161–70.
171. Yan M, Yang X, Wang L, Clark D, Zuo H, Ye D, et al. Plasma Membrane Proteomics of Tumor Spheres Identify CD166 as a Novel Marker for Cancer Stem-like Cells in Head and Neck Squamous Cell Carcinoma. *Mol Cell Proteomics*. 2013 Nov;12(11):3271–84.
172. Jiao J, Hindoyan A, Wang S, Tran LM, Goldstein AS, Lawson D, et al. Identification of CD166 as a surface marker for enriching prostate stem/progenitor and cancer initiating cells. *PLoS One*. 2012;7(8):1–14.
173. Jiang Z, Hao Y, Ding X, Zhang Z, Liu P, Wei X, et al. The effects and mechanisms of SLC34A2 on tumorigenicity in human non-small cell lung cancer stem cells. *Tumour Biol*. 2016 Aug;37(8):10383-92.
174. Ishiguro F, Murakami H, Mizuno T, Fujii M, Kondo Y, Usami N, et al. Membranous expression of activated leukocyte cell adhesion molecule contributes to poor prognosis and malignant phenotypes of non-small-cell lung cancer. *J Surg Res*. 2013 Jan;179(1):24–32.
175. Tachezy M, Zander H, Wolters-Eisfeld G, Müller J, Wicklein D, Gebauer F, et al. Activated Leukocyte Cell Adhesion Molecule (CD166): An “Inert” Cancer Stem Cell Marker for Non-Small Cell Lung Cancer? *Stem Cells*. 2014;32(6):1429–36.
176. Sahasrabudde AA. BMI1 : A Biomarker of Hematologic Malignancies. *Biomark Cancer*. 2016;8:65–75.
177. Allegra E, Trapasso S, Pisani D, Puzzo L. The Role of BMI1 as a Biomarker of Cancer Stem Cells in Head and Neck Cancer: A Review. *Oncology*. 2014;86(4):199–205.
178. Ge G, Zhou C, Ren Y, Tang X, Wang K, Zhang W, et al. Enhanced SLC34A2 in breast cancer stem cell-like cells induces chemotherapeutic resistance to doxorubicin via SLC34A2-

- Bmi1-ABCC5 signaling. *Tumor Biol.* 2016;37(4):5049–62.
179. Zhu D, Wan X, Huang H, Chen X, Liang W, Zhao F, et al. Knockdown of Bmi1 inhibits the stemness properties and tumorigenicity of human bladder cancer stem cell-like side population cells. *Oncol Rep.* 2014 Feb;31(2):727–36.
180. Physiology C, He Z, Xia Y, Pan C, Ma T, Liu B, et al. Up-regulation of MiR-452 inhibits metastasis of non-small cell lung cancer by regulating BMI1. *Cell Physiol Biochem.* 2015;37(1):387–98.
181. Tomita H, Tanaka K, Hisamatsu K, Hara A. The role of aldehyde dehydrogenase 1A1 in stem cells and cancer stem cells. 2016;7(10):11018–32.
182. Mansour SF, Atwa MM. Clinicopathological Significance of CD133 and ALDH1 Cancer Stem Cell Marker Expression in Invasive Ductal Breast Carcinoma. *Asian Pac J Cancer Prev.* 2015;16:7491–6.
183. Fitzgerald TL, Lea CS, Brinkley J, Zervos EE. Colorectal cancer outcome inequalities : association between population density , race , and socioeconomic status. *Rural Remote Health.* 2014;14(3):2668.
184. Liu S, Liu C, Min X, Ji Y, Wang N, Liu D, et al. Prognostic Value of Cancer Stem Cell Marker Aldehyde Dehydrogenase in Ovarian Cancer : A Meta-Analysis. 2013;8(11):1–11.
185. Huang C-P, Tsai M-F, Chang T-H, Tang W-C, Chen S-Y, Lai H-H, et al. ALDH-positive lung cancer stem cells confer resistance to epidermal growth factor receptor tyrosine kinase inhibitors. *Cancer Lett.* 2013 Jan;328(1):144–51.
186. Liu J, Xiao Z, Wong SK, Tin VP, Ho KY, Wang J, et al. Lung cancer tumorigenicity and drug resistance are maintained through ALDH(hi)CD44(hi) tumor initiating cells. *Oncotarget.* 2013;4.
187. Huo W, Du M, Pan X, Zhu X, Li Z. Prognostic value of ALDH1 expression in lung cancer: a meta-analysis. *Int J Clin Exp Med.* 2015 Feb 15;8(2):2045–51.
188. You Q, Huanchen G, Dongxiang X. Distinct prognostic values and potential drug targets of ALDH1 isoenzymes in non-small-cell lung cancer. *Drug Des Devel Ther.* 2015;9:5087–97.
189. Zhou Q, Ye M, Lu Y, Zhang H, Chen Q, Huang S, et al. Curcumin Improves the Tumoricidal

- Effect of Mitomycin C by Suppressing ABCG2 Expression in Stem Cell-Like Breast Cancer Cells. Castresana JS, editor. PLoS One. 2015 Aug 25;10(8):e0136694.
190. Jia Q, Zhang X, Deng T, Gao J. Positive correlation of Oct4 and ABCG2 to chemotherapeutic resistance in CD90(+)CD133(+) liver cancer stem cells. *Cell Reprogram*. 2013 Apr;15(2):143–50.
191. To KKW, Poon DC, Wei Y, Wang F, Lin G, Fu L. Vatalanib sensitizes ABCB1 and ABCG2-overexpressing multidrug resistant colon cancer cells to chemotherapy under hypoxia. *Biochem Pharmacol*. 2015 Sep;97(1):27–37.
192. Yang B, Ma Y-F, Liu Y. Elevated Expression of Nrf-2 and ABCG2 Involved in Multi-drug Resistance of Lung Cancer SP Cells. *Drug Res (Stuttg)*. 2015 Oct;65(10):526–31.
193. Tang Y, Hou J, Li G, Song Z, Li X, Yang C, et al. ABCG2 regulates the pattern of self-renewing divisions in cisplatin-resistant non-small cell lung cancer cell lines. *Oncol Rep*. 2014 Nov;32(5):2168–74.
194. Herpel E, Jensen K, Muley T, Warth A, Schnabel PA, Meister M, et al. The cancer stem cell antigens CD133, BCRP1/ABCG2 and CD117/c-KIT are not associated with prognosis in resected early-stage non-small cell lung cancer. *Anticancer Res*. 2011 Dec;31(12):4491–500.
195. Liang S-CC, Yang C-YY, Tseng J-YY, Wang H-LL, Tung C-YY, Liu H-WW, et al. ABCG2 localizes to the nucleus and modulates CDH1 expression in lung cancer cells. *Neoplasia*. 2015 Mar;17(3):265–78.
196. Ajani JA, Song S, Hochster HS, Steinberg IB. Cancer stem cells: The promise and the potential. *Semin Oncol*. 2015 Apr;42 Suppl 1(S1):S3–17.
197. Borggrefe T, Oswald F. The Notch signaling pathway: Transcriptional regulation at Notch target genes. *Cell Mol Life Sci*. 2009 May;66(10):1631–46.
198. Takebe, Naoko Nguyen, Dat & Yang S, Takebe N, Nguyen D, Yang SX, Naoko Takebea, Dat Nguyenb and SXY. Targeting notch signaling pathway in cancer: clinical development advances and challenges. *Pharmacol Ther*. 2014;141(2):140–9.
199. Sun W, Gaykalova DA, Ochs MF, Mambo E, Liu Y, Loyo M, et al. Activation of the NOTCH pathway in head and neck cancer. *Cancer Res*. 2014 Feb;74(4):1091–104.

200. Vinson KE, George DC, Fender AW, Bertrand FE, Sigounas G. The Notch pathway in colorectal cancer. *Int J Cancer*. 2016 Apr 15;138(8):1835–42.
201. Xiao MJ, Han Z, Shao B, Jin K. Notch signaling and neurogenesis in normal and stroke brain. *Int J Physiol Pathophysiol Pharmacol*. 2009 Nov 10;1(2):192–202.
202. Penton AL, Leonard LD, Spinner NB. Notch signaling in human development and disease. *Semin Cell Dev Biol*. 2012 Jun;23(4):450–7.
203. Westhoff B, Colaluca IN, D’Ario G, Donzelli M, Tosoni D, Volorio S, et al. Alterations of the Notch pathway in lung cancer. *Proc Natl Acad Sci U S A*. 2009 Dec 29;106(52):22293–8.
204. Yu Z, Pestell T, Lisanti MP. Cancer Stem Cells. *Int J Biochem Cell Biol*. 2012;44(12):2144–51.
205. Egloff AM, Grandis JR. Molecular Pathways: Context-dependent approaches to Notch targeting as cancer therapy. *Clin Cancer Res*. 2012 Oct 1;18(19):5188–95.
206. Reynolds TC, Smith SD, Sklar J. Analysis of DNA surrounding the breakpoints of chromosomal translocations involving the beta T cell receptor gene in human lymphoblastic neoplasms. *Cell*. 1987 Jul;50(1):107–17.
207. Bhola NE, Jansen VM, Koch JP, Li H, Formisano L, Williams JA, et al. Treatment of Triple-Negative Breast Cancer with TORC1/2 Inhibitors Sustains a Drug-Resistant and Notch-Dependent Cancer Stem Cell Population. *Cancer Res*. 2016 Jan;76(2):440–52.
208. Yuan R, Ke J, Sun L, He Z, Zou Y, He X, et al. HES1 promotes metastasis and predicts poor survival in patients with colorectal cancer. *Clin Exp Metastasis*. 2015 Feb;32(2):169–79.
209. Kushwah R, Guezguez B, Lee JB, Hopkins CI, Bhatia M. Pleiotropic roles of Notch signaling in normal, malignant, and developmental hematopoiesis in the human. *EMBO Rep*. 2014 Nov;15(11):1128–38.
210. Bousquet Mur E, Bernardo S, Papon L, Mancini M, Fabbriozio E, Goussard M, et al. Notch inhibition overcomes resistance to Tyrosine Kinase Inhibitors in EGFR-driven lung adenocarcinoma. *J Clin Invest*. 2019 Oct. pii: 126896.
211. Allen TD, Rodriguez EM, Jones KD, Bishop JM. Activated NOTCH1 induces lung adenomas in mice and cooperates with MYC in the generation of lung adenocarcinoma. *Cancer Res*. 2011 Sep 15;71(18):6010–8.

212. Licciulli S, Avila JL, Hanlon L, Troutman S, Cesaroni M, Kota S, et al. Notch1 is required for Kras-induced lung adenocarcinoma and controls tumor cell survival via p53. *Cancer Res.* 2013;73(19):5974-84.
213. Elias S, Liang S, Chen Y, De Marco MA, Machek O, Skucha S, et al. Notch-1 stimulates survival of lung adenocarcinoma cells during hypoxia by activating the IGF-1R pathway. *Oncogene.* 2010;29(17):2488-98.
214. Chen Y, Li D, Liu H, Xu H, Zheng H, Qian F, et al. Notch-1 signaling facilitates survivin expression in human non-small cell lung cancer cells. *Cancer Biol Ther.* 2011;11(1):14-21.
215. Zhou L, Wu S, Yu L, Gong X, Song W, Cheng Z. Expression of CD133 and Notch1 in non-small cell lung cancer and the clinicopathological significance. *Nan Fang Yi Ke Da Xue Xue Bao.* 2015 Feb;35(2):196–201.
216. Hassan KA, Wang L, Korkaya H, Chen G, Maillard I, Beer DG, et al. Notch pathway activity identifies cells with cancer stem cell-like properties and correlates with worse survival in lung adenocarcinoma. *Clin Cancer Res.* 2013;19(8):1972–80.
217. Zheng Y, de la Cruz CC, Sayles LC, Alleyne-Chin C, Vaka D, Knaak TD, et al. A rare population of CD24(+)ITGB4(+)Notch(hi) cells drives tumor propagation in NSCLC and requires Notch3 for self-renewal. *Cancer Cell.* 2013 Jul;24(1):59-74.
218. Mimae T, Okada M, Hagiyaama M, Miyata Y, Tsutani Y, Inoue T, et al. Upregulation of notch2 and six1 is associated with progression of early-stage lung adenocarcinoma and a more aggressive phenotype at advanced stages. *Clin Cancer Res.* 2012 Feb;18(4):945–55.
219. Yen W-C, Fischer MM, Axelrod F, Bond C, Cain J, Cancilla B, et al. Targeting Notch signaling with a Notch2/Notch3 antagonist (tarextumab) inhibits tumor growth and decreases tumor-initiating cell frequency. *Clin Cancer Res.* 2015 May;21(9):2084–95.
220. Gao F, Huang W, Zhang Y, Tang S, Zheng L, Ma F, et al. Hes1 promotes cell proliferation and migration by activating Bmi-1 and PTEN/Akt/GSK3beta pathway in human colon cancer. *Oncotarget.* 2015 Nov;6(36):38667–80.
221. Clevers H, Nusse R. Wnt/B-catenin signaling and disease. *Cell.* 2012;149(6):1192–205.
222. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev*

- Cancer. 2012;13(1):11–26.
223. Enzo MV, Rastrelli M, Rossi CR, Hladnik U, Segat D. The Wnt/ β -catenin pathway in human fibrotic-like diseases and its eligibility as a therapeutic target. 2015;1–13.
224. Purro SA, Galli S, Salinas PC. Dysfunction of Wnt signaling and synaptic disassembly in neurodegenerative diseases. 2014;75–80.
225. Niehrs C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol*. 2012 Dec;13(12):767–79.
226. Sokol SY. Spatial and temporal aspects of Wnt signaling and planar cell polarity during vertebrate embryonic development. *Semin Cell Dev Biol*. 2015 Jun;42:78–85.
227. Viale-Bouroncle S, Klingelhoffer C, Ettl T, Reichert TE, Morscheck C. A protein kinase A (PKA)/beta-catenin pathway sustains the BMP2/DLX3-induced osteogenic differentiation in dental follicle cells (DFCs). *Cell Signal*. 2015 Mar;27(3):598–605.
228. Kim S, Nie H, Nesin V, Tran U, Outeda P, Bai C-X, et al. The polycystin complex mediates Wnt/Ca²⁺ signalling. *Nat Cell Biol*. 2016 Jul;18(7):752–64.
229. Mezzacappa C, Komiya Y, Habas R. Activation and Function of Small GTPases Rho, Rac, and Cdc42 During Gastrulation. *Methods Mol Biol*. 2012;(839):119–31.
230. Puvirajesinghe TM, Bertucci F, Jain A, Scerbo P, Belotti E, Restouin A, et al. Identification of p62/SQSTM1 as a component of non-canonical Wnt VANGL2–JNK signalling in breast cancer. *Nat Commun*. 2016 Jan;7:10318.
231. Kahn M. Can we safely target the WNT pathway? *Nat Rev Drug Discov*. 2014;13(2):513–32.
232. Stamos JL, Weis WI. The β -Catenin Destruction Complex. *Cold Spring Harb Perspect Biol*. 2013;5(a007898):1–16.
233. Shukla S, Sinha S, Khan S, Kumar S, Singh K, Mitra K. Cucurbitacin B inhibits the stemness and metastatic abilities of NSCLC via downregulation of canonical Wnt/ β -catenin signaling axis. *Sci Rep*. 2016 Feb;6:21860.
234. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. 2017 Mar;36(11):1461–73.

235. Sharma G, Sharma AR, Seo EM, Nam JS. Genetic polymorphism in extracellular regulators of Wnt signaling pathway. *Biomed Res Int.* 2015;2015.
236. Yang K, Wang X, Zhang H, Wang Z, Nan G, Li Y, et al. The evolving roles of canonical WNT signaling in stem cells and tumorigenesis: implications in targeted cancer therapies. *Lab Investig.* 2016;96:116–36.
237. Nusse R, Varmus HE. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell.* 1982 Nov;31(1):99–109.
238. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, et al. Constitutive Transcriptional Activation by a β -Catenin-Tcf Complex in APC^{-/-} Colon Carcinoma. *Science.* 1997 Mar;275(5307):1784–7.
239. Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, et al. Activation of β -Catenin-Tcf Signaling in Colon Cancer by Mutations in β -Catenin or APC. *Science.* 1997 Mar;275(5307):1787–90.
240. Cao X, Eu KW, Seow-Choen F, Cheah PY. Germline mutations are frequent in the APC gene but absent in the beta-catenin gene in familial adenomatous polyposis patients. *Genes Chromosomes Cancer.* 1999 Aug;25(4):396–8.
241. Tanaka K, Kumano K, Ueno H. Intracellular signals of lung cancer cells as possible therapeutic targets. *Cancer Sci.* 2015 May;106(5):489–96.
242. Teng Y, Wang X, Wang Y, Ma D. Wnt/b-catenin signaling regulates cancer stem cells in lung cancer A549 cells. *Biochem Biophys Res Commun.* 2010 Feb;392(3):373–9.
243. Vaughan AE, Halbert CL, Wootton SK. Lung Cancer in Mice Induced by the Jaagsiekte Sheep Retrovirus Envelope Protein Is Not Maintained by Rare Cancer Stem Cells , but Tumorigenicity Does Correlate with Wnt Pathway Activation Lung Cancer in Mice Induced by the Jaagsiekte Sheep. *Mol Cancer Res.* 2012;10(1):86–95.
244. Akiri G, Cherian MM, Vijayakumar S, Liu G, Bafico A, Aaronson SA. Wnt pathway aberrations including autocrine Wnt activation occur at high frequency in human non-small-cell lung carcinoma. *Oncogene.* 2009 May 28;28(21):2163–72.
245. Lee JS, Hur MW, Lee SK, Choi W Il, Kwon YG, Yun CO. A novel sLRP6E1E2 inhibits canonical

- Wnt signaling, epithelial-to-mesenchymal transition, and induces mitochondria-dependent apoptosis in lung cancer. *PLoS One*. 2012;7(5):e36520.
246. Li C, Song G, Zhang S, Wang E, Cui Z. Wnt3a increases the metastatic potential of non-small cell lung cancer cells in vitro in part via its upregulation of Notch3. *Oncol Rep*. 2015 Mar;33(3):1207–14.
247. Xu X, Sun P-L, Li J-Z, Jheon S, Lee C-T, Chung J-H. Aberrant Wnt1/ β -catenin expression is an independent poor prognostic marker of non-small cell lung cancer after surgery. *J Thorac Oncol*. 2011 Apr;6(4):716–24.
248. Nakashima N, Liu D, Huang C-L, Ueno M, Zhang X, Yokomise H. Wnt3 gene expression promotes tumor progression in non-small cell lung cancer. *Lung Cancer*. 2012 May;76(2):228–34.
249. Jiang H-L, Jiang L-M, Han W-D. Wnt/beta-catenin signaling pathway in lung cancer stem cells is a potential target for the development of novel anticancer drugs. *J BUON*. 2015 Jul-Aug;20(4):1094–100.
250. Togashi Y, Hayashi H, Terashima M, De Velasco MA, Sakai K, Fujita Y, et al. Inhibition of β -Catenin enhances the anticancer effect of irreversible EGFR-TKI in EGFR-mutated non-small-cell lung cancer with a T790M mutation. *J Thorac Oncol*. 2015 Jan;10(1):93–101.
251. Remsing Rix LL, Kuenzi BM, Luo Y, Remily-Wood E, Kinose F, Wright G, et al. GSK3 alpha and beta are new functionally relevant targets of tivantinib in lung cancer cells. *ACS Chem Biol*. 2014 Feb;9(2):353–8.
252. He W, He S, Wang Z, Shen H, Fang W, Zhang Y, et al. Astrocyte elevated gene-1(AEG-1) induces epithelial-mesenchymal transition in lung cancer through activating Wnt/beta-catenin signaling. *BMC Cancer*. 2015;15:107.
253. Many AM, Brown AMC. Both canonical and non-canonical Wnt signaling independently promote stem cell growth in mammospheres. *PLoS One*. 2014;9(7):e101800.
254. Li S, Qin X, Guo X, Cui A, He Y, Wei S, et al. Dickkopf-1 Is Oncogenic and Involved in Invasive Growth in Non Small Cell Lung Cancer . Deb S, editor. *PLoS One*. 2013 Dec;8(12):e84944.
255. Xiang XJ, Liu YW, Chen DD, Yu S. Differential expression of Dickkopf-1 among non-small

- cell lung cancer cells. *Mol Med Rep*. 2015 Aug;12(2):1935–40.
256. Geng M, Cao Y-C, Chen Y-J, Jiang H, Bi L-Q, Liu X-H. Loss of Wnt5a and Ror2 protein in hepatocellular carcinoma associated with poor prognosis. *World J Gastroenterol*. 2012 Mar;18(12):1328–38.
257. Yao L, Sun B, Zhao XX, Zhao XX, Gu Q, Dong X, et al. Overexpression of Wnt5a promotes angiogenesis in NSCLC. *Biomed Res Int*. 2014;2014:832562.
258. Lu C, Wang X, Zhu H, Feng J, Ni S, Huang J. Over-expression of ROR2 and Wnt5a cooperatively correlates with unfavorable prognosis in patients with non-small cell lung cancer. *Oncotarget*. 2015 Sep;6(28):24912–21.
259. Gonnissen A, Isebaert S, Haustermans K. Targeting the Hedgehog signaling pathway in cancer: beyond Smoothened. *Oncotarget*. 2015;6(16):13899–913.
260. Gupta S, Takebe N, Lorusso P. Targeting the Hedgehog pathway in cancer. *Ther Adv Med Oncol*. 2010;2(4):237–50.
261. Jung B, Padula D, Burtscher I, Landerer C, Lutter D, Theis F, et al. Pitchfork and Gprasp2 target smoothened to the primary cilium for hedgehog pathway activation. *PLoS One*. 2016;11(2):1–23.
262. Hahn H, Wicking C, Zaphiropoulos PG, Gailani MR, Shanley S, Chidambaram A, et al. Mutations of the Human Homolog of *Drosophila* patched in the Nevoid Basal Cell Carcinoma Syndrome. *Cell*. 1996 Jun;85(6):841–51.
263. Thalakoti S, Geller T. Basal cell nevus syndrome or Gorlin syndrome. *Handb Clin Neurol*. 2015;132:119–28.
264. Shanley S, McCormack C. Diagnosis and Management of Hereditary Basal Cell Skin Cancer. *Recent results cancer Res*. 2016;205:191–212.
265. Cochrane C, Szczepny A, Watkins D, Cain J. Hedgehog Signaling in the Maintenance of Cancer Stem Cells. *Cancers (Basel)*. 2015 Aug;7(3):1554–85.
266. Peng T, Frank DB, Kadzik RS, Morley MP, Komal S, Wang T, et al. Hedgehog actively maintains adult lung quiescence and regulates repair and regeneration. *Nature*. 2015 Oct;526(7574):578–82.

267. Metcalfe C, Siebel CW. The Hedgehog Hold on Homeostasis. *Cell Stem Cell*. 2015 Nov 5;17(5):505–6.
268. Park K-S, Martelotto LG, Peifer M, Sos ML, Karnezis AN, Mahjoub MR, et al. A crucial requirement for Hedgehog signaling in small cell lung cancer. *Nat Med*. 2011;17(11):1504–8.
269. Kaur G, Reinhart RA, Monks A, Evans D, Morris J, Polley E, et al. Bromodomain and hedgehog pathway targets in small cell lung cancer. *Cancer Lett*. 2016 Feb;371(2):225–39.
270. Giroux Leprieur E, Antoine M, Vieira T, Rozensztajn N, Ruppert A-M, Rabbe N, et al. Role of the Sonic Hedgehog pathway in thoracic cancers. *Rev Mal Respir*. 2015 Oct;32(8):800–8.
271. Bai X-Y, Zhang X-C, Yang S-Q, An S-J, Chen Z-H, Su J, et al. Blockade of Hedgehog Signaling Synergistically Increases Sensitivity to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors in Non-Small-Cell Lung Cancer Cell Lines. *PLoS One*. 2016;11(3):e0149370.
272. Li Y, Yang P. GPC5 gene and its related pathways in lung cancer. *J Thorac Oncol*. 2011;6(1):2–5.
273. Takebe N, Harris PJ, Warren RQ, Ivy SP. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nat Rev Clin Oncol*. 2011 Feb;8(2):97–106.
274. Cordenonsi M, Zanconato F, Azzolin L, Forcato M, Rosato A, Frasson C, et al. The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell*. 2011 Nov;147(4):759–72.
275. Li Z, Wang Y, Zhu Y, Yuan C, Wang D, Zhang W, et al. The Hippo transducer TAZ promotes epithelial to mesenchymal transition and cancer stem cell maintenance in oral cancer. *Mol Oncol*. 2015 Jun;9(6):1091–105.
276. Dhanasekaran SM, Balbin OA, Chen G, Nadal E, Kalyana-Sundaram S, Pan J, et al. Transcriptome meta-analysis of lung cancer reveals recurrent aberrations in NRG1 and Hippo pathway genes. *Nat Commun*. 2014 Dec;5:5893.
277. Lau AN, Curtis SJ, Fillmore CM, Rowbotham SP, Mohseni M, Wagner DE, et al. Tumor-propagating cells and Yap/Taz activity contribute to lung tumor progression and

- metastasis. *EMBO J.* 2014 Mar;33(5):468–81.
278. Yuan Y, Zhong W, Ma G, Zhang B, Tian H. Yes-associated protein regulates the growth of human non-small cell lung cancer in response to matrix stiffness. *Mol Med Rep.* 2015 Jun;11(6):4267–72.
279. Duzagac F, Inan S, Ela Simsek F, Acikgoz E, Guven U, Khan SA, et al. JAK/STAT pathway interacts with intercellular cell adhesion molecule (ICAM) and vascular cell adhesion molecule (VCAM) while prostate cancer stem cells form tumor spheroids. *J BUON.* 2015;20(5):1250–7.
280. Chung SS, Vadgama J V. Curcumin and epigallocatechin gallate inhibit the cancer stem cell phenotype via down-regulation of STAT3-NFkappaB signaling. *Anticancer Res.* 2015 Jan;35(1):39–46.
281. Dutta P, Sabri N, Li J, Li WX. Role of STAT3 in lung cancer. *JAK-STAT [Internet].* 2014 Jan 20;3(4):e999503.
282. Hsu HS, Huang PI, Chang YL, Tzao C, Chen YW, Shih HC et al. Cucurbitacin I inhibits tumorigenic ability and enhances radiochemosensitivity in nonsmall cell lung cancer-derived CD133-positive cells. *Cancer.* 2011 Jul 1;117(13):2970-85.
283. Shao C, Sullivan JP, Girard L, Augustyn A, Yenerall P, Rodriguez-Canales J, et al. Essential role of aldehyde dehydrogenase 1A3 for the maintenance of non-small cell lung cancer stem cells is associated with the STAT3 pathway. *Clin Cancer Res.* 2014 Aug;20(15):4154–66.
284. Attar-Schneider O, Drucker L, Gottfried M. Migration and epithelial-to-mesenchymal transition of lung cancer can be targeted via translation initiation factors eIF4E and eIF4GI. *Lab Invest.* 2016 Sep;96(9):1004–15.
285. Chen W-L, Huang A-F, Huang S-M, Ho C-L, Chang Y-L, Chan JY-H. CD164 promotes lung tumor-initiating cells with stem cell activity and determines tumor growth and drug resistance via Akt/mTOR signaling. *Oncotarget.* 2016 Dec;46(13): 1380-1391.
286. Bertolini G, D'Amico L, Moro M, Landoni E, Perego P, Miceli R, et al. Microenvironment-Modulated Metastatic CD133+/CXCR4+/EpCAM- Lung Cancer-Initiating Cells Sustain Tumor Dissemination and Correlate with Poor Prognosis. *Cancer Res.* 2015 Sep;75(17):3636–49.

287. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002 Jun;3(7):research0034.1-11.
288. Pfaffl MW, Duquenne M, François JM, Parrou J-L, Francois J, Gancedo C, et al. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001 May;29(9):45e – 45.
289. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford).* 2011;2011:bar026.
290. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *J Natl Cancer Inst.* 2000 Feb;92(3):205–16.
291. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, et al. Prediction of Survival in Diffuse Large-B-Cell Lymphoma Based on the Expression of Six Genes. *N Engl J Med.* 2004 Apr;350(18):1828–37.
292. Schetter AJ, Nguyen GH, Bowman ED, Mathe EA, Yuen ST, Hawkes JE, et al. Association of Inflammation-Related and microRNA Gene Expression with Cancer-Specific Mortality of Colon Adenocarcinoma. *Clin Cancer Res.* 2009 Sep;15(18):5878–87.
293. Kodack DP, Farago AF, Dastur A, Held MA, Dardaei L, Friboulet L, et al. Primary Patient-Derived Cancer Cells and Their Potential for Personalized Cancer Patient Care. *Cell Rep.* 2017 Dec;21(11):3298–309.
294. Bartucci M, Svensson S, Romania P, Dattilo R, Patrizii M, Signore M, et al. Therapeutic targeting of Chk1 in NSCLC stem cells during chemotherapy. *Cell Death Differ.* 2012 May;19(5):768–78.
295. Zeuner A, Francescangeli F, Contavalli P, Zapparelli G, Apuzzo T, Eramo A, et al. Elimination of quiescent/slow-proliferating cancer stem cells by Bcl-XL inhibition in non-small cell lung cancer. *Cell Death Differ.* 2014 Dec;21(12):1877–88.
296. Tiran V, Lindenmann J, Brcic L, Heitzer E, Stanzer S, Tabrizi-Wizsy NG, et al. Primary patient-derived lung adenocarcinoma cell culture challenges the association of cancer stem cells with epithelial-to-mesenchymal transition. *Sci Rep.* 2017 Aug;7(1):10040.

297. Cholewinski AJ, Reid JC, McDermott AM, Wilkin GP. Purification of astroglial-cell cultures from rat spinal cord: the use of d-valine to inhibit fibroblast growth. *Neurochem Int.* 1989;15(3):365–9.
298. Assanga I, Luján L. Cell growth curves for different cell lines and their relationship with biological activities. *Int J Biotechnol Mol Biol Res.* 2013;4(4):60–70.
299. Nunes T, Hamdan D, Leboeuf C, El Bouchtaoui M, Gapihan G, Nguyen TT, et al. Targeting Cancer Stem Cells to Overcome Chemoresistance. *Int J Mol Sci.* 2018 Dec;19(12).
300. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell.* 2006 Aug;126(4):663–76.
301. Goldman B. Embryonic Stem Cells 2.0. *Nat Reports Stem Cells.* 2008 May.
302. Chiou SH, Wang ML, Chou YT, Chen CJ, Hong CF, Hsieh WJ, et al. Coexpression of Oct4 and Nanog enhances malignancy in lung adenocarcinoma by inducing cancer stem cell-like properties and epithelial-mesenchymal transdifferentiation. *Cancer Res.* 2010 Dec;70:10433–44.
303. Slawek S, Szmyt K, Fularz M, Dziudzia J, Boruckowski M, Sikora J, et al. Pluripotency transcription factors in lung cancer—a review. *Tumour Biol.* 2016 Apr;37(4):4241-9.
304. Insinga A, Cicalese A, Faretta M, Gallo B, Albano L, Ronzoni S, et al. DNA damage in stem cells activates p21, inhibits p53, and induces symmetric self-renewing divisions. *Proc Natl Acad Sci U S A.* 2013 Mar;110(10):3931–6.
305. Xia W, Chen J-S, Zhou X, Sun P-R, Lee D-F, Liao Y, et al. Phosphorylation/cytoplasmic localization of p21Cip1/WAF1 is associated with HER2/neu overexpression and provides a novel combination predictor for poor prognosis in breast cancer patients. *Clin Cancer Res.* 2004 Jun;10(11):3815–24.
306. Shiraki K, Wagayama H. Cytoplasmic p21(WAF1/CIP1) expression in human hepatocellular carcinomas. *Liver.* 2006;26:1018–9.
307. Zhou BP, Liao Y, Xia W, Spohn B, Lee MH, Hung MC. Cytoplasmic localization of p21Cip1/WAF1 by Akt-induced phosphorylation in HER-2/neu-overexpressing cells. *Nat Cell Biol.* 2001 Mar;3(3):245–52.
308. Wienken M, Dickmanns A, Nemajerova A, Kramer D, Najafova Z, Weiss M, et al. MDM2

- Associates with Polycomb Repressor Complex 2 and Enhances Stemness-Promoting Chromatin Modifications Independent of p53. *Mol Cell*. 2015 Dec;61(1):68–83.
309. Daniele S, Costa B, Zappelli E, Da Pozzo E, Sestito S, Nesi G, et al. Combined inhibition of AKT/mTOR and MDM2 enhances Glioblastoma Multiforme cell apoptosis and differentiation of cancer stem cells. *Sci Rep*. 2015 Apr;5:9956.
310. Javid J, Mir R, Julka PK, Ray PC, Saxena A. Association of p53 and mdm2 in the development and progression of non-small cell lung cancer. *Tumour Biol*. 2015 Jul;36(7):5425–32.
311. Shaikh MV, Kala M, Nivsarkar M. CD90 a potential cancer stem cell marker and a therapeutic target. *Cancer Biomark*. 2016;16(3):301–7.
312. Schmitt M, Metzger M, Gradl D, Davidson G, Orian-Rousseau V. CD44 functions in Wnt signaling by regulating LRP6 localization and activation. *Cell Death Differ*. 2015 Apr;22(4):677–89.
313. Soto-Cerrato V, Manuel-Manresa P, Hernando E, Calabuig-Fariñas S, Martínez-Romero A, Fernández-Dueñas V, et al. Facilitated Anion Transport Induces Hyperpolarization of the Cell Membrane That Triggers Differentiation and Cell Death in Cancer Stem Cells. *J Am Chem Soc*. 2015 Dec;137(50):15892–8.
314. Mizrak D, Brittan M, Alison MR. CD133: molecule of the moment. *J Pathol*. 2008 Jan;214(1):3–9.
315. Rappa G, Fodstad O, Lorico A. The Stem Cell-Associated Antigen CD133 (Prominin-1) Is a Molecular Therapeutic Target for Metastatic Melanoma. *Stem Cells*. 2008 Dec;26(12):3008–17.
316. Smith LM, Nesterova A, Ryan MC, Duniho S, Jonas M, Anderson M, et al. CD133/prominin-1 is a potential therapeutic target for antibody-drug conjugates in hepatocellular and gastric cancers. *Br J Cancer*. 2008 Jul 8;99(1):100–9.
317. Howard B, Boockvar J. Stem cell marker CD133 expression predicts outcome in glioma patients. *Neurosurgery*. 2008;62(6):N6.
318. Tirino V, Camerlingo R, Franco R, Malanga D, La RA, Viglietto G, et al. The role of CD133 in the identification and characterisation of tumour-initiating cells in non-small-cell lung

- cancer. *Eur J Cardiothorac Surg*. 2009 Sep;36:446–53.
319. Bertolini G, Roz L, Perego P, Tortoreto M, Fontanella E, Gatti L, et al. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. *Proc Natl Acad Sci U.S.A.* 2009 Sep;106:16281–6.
320. Salnikov A V, Gladkich J, Moldenhauer G, Volm M, Mattern J, Herr I. CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. *Int J Cancer*. 2010 Feb 15;126(4):950–8.
321. Martin TA, Jiang WG. Evaluation of the expression of stem cell markers in human breast cancer reveals a correlation with clinical progression and metastatic disease in ductal carcinoma. *Oncol Rep*. 2014 Jan;31(1):262–72.
322. Hoogland AM, Verhoef EI, Roobol MJ, Schroder FH, Wildhagen MF, van der Kwast TH, et al. Validation of stem cell markers in clinical prostate cancer: alpha6-integrin is predictive for non-aggressive disease. *Prostate*. 2014 May;74(5):488–96.
323. Haraguchi N, Ishii H, Mimori K, Ohta K, Uemura M, Nishimura J, et al. CD49f-positive cell population efficiently enriches colon cancer-initiating cells. *Int J Oncol*. 2013 Aug;43(2):425–30.
324. Guo W, Giancotti FG. Integrin signalling during tumour progression. *Nat Rev Mol Cell Biol*. 2004 Oct;5(10):816–26.
325. O'Connor KL, Nguyen BK, Mercurio AM. RhoA function in lamellae formation and migration is regulated by the alpha6beta4 integrin and cAMP metabolism. *J Cell Biol*. 2000 Jan;148(2):253–8.
326. Stewart RL, West D, Wang C, Weiss HL, Gal T, Durbin EB, et al. Elevated integrin alpha6beta4 expression is associated with venous invasion and decreased overall survival in non-small cell lung cancer. *Hum Pathol*. 2016 Aug;54:174–83.
327. Hsu Y-L, Wu C-Y, Hung J-Y, Lin Y-S, Huang M-S, Kuo P-L. Galectin-1 promotes lung cancer tumor metastasis by potentiating integrin alpha6beta4 and Notch1/Jagged2 signaling pathway. *Carcinogenesis*. 2013 Jun;34(6):1370–81.
328. Yasui K, Shimamura M, Mitsutake N, Nagayama Y. SNAIL induces epithelial-to-mesenchymal transition and cancer stem cell-like properties in aldehyde

- dehydrogenase-negative thyroid cancer cells. *Thyroid*. 2013 Aug;23(8):989–96.
329. Fan F, Samuel S, Evans KW, Lu J, Xia L, Zhou Y, et al. Overexpression of snail induces epithelial-mesenchymal transition and a cancer stem cell-like phenotype in human colorectal cancer cells. *Cancer Med*. 2012 Aug;1(1):5–16.
330. Ota I, Masui T, Kurihara M, Yook J-I, Mikami S, Kimura T, et al. Snail-induced EMT promotes cancer stem cell-like properties in head and neck cancer cells. *Oncol Rep*. 2016 Jan;35(1):261–6.
331. Zhou W, Lv R, Qi W, Wu D, Xu Y, Liu W, et al. Snail Contributes to the Maintenance of Stem Cell-Like Phenotype Cells in Human Pancreatic Cancer. Klymkowsky M, editor. *PLoS One*. 2014 Jan 29;9(1):e87409.
332. Wang HH-SH, Zhang G, Zhang H, Zhang F, Zhou B, Ning F, et al. Acquisition of epithelial-mesenchymal transition phenotype and cancer stem cell-like properties in cisplatin-resistant lung cancer cells through AKT/ β -catenin/Snail signaling pathway. *Eur J Pharmacol*. 2014 Jan;723:156–66.
333. Liu C-W, Li C-H, Yi-Jen P, Cheng Y-W, Chen H-W, Liao P-L, et al. Snail regulates Nanog status during the epithelial–mesenchymal transition via the Smad1/Akt/GSK3 β signaling pathway in non-small-cell lung cancer. *Oncotarget*. 2014 Jun 26;5(11):3880–94.
334. Li J, Wang H, Ke H, Ni S. MiR-129 regulates MMP9 to control metastasis of non-small cell lung cancer. *Tumor Biol*. 2015 Aug;36(8):5785–90.
335. Tirino V, Camerlingo R, Bifulco K, Irollo E, Montella R, Paino F, et al. TGF-beta1 exposure induces epithelial to mesenchymal transition both in CSCs and non-CSCs of the A549 cell line, leading to an increase of migration ability in the CD133+ A549 cell fraction. *Cell Death Dis*. 2013 May;4:e620.
336. Liu J, Ping W, Zu Y, Sun W. Correlations of lysyl oxidase with MMP2/MMP9 expression and its prognostic value in non-small cell lung cancer. *Int J Clin Exp Pathol*. 2014 Aug 15;7(9):6040–7.
337. McKeage MJ, Kotasek D, Markman B, Hidalgo M, Millward MJ, Jameson MB, et al. Phase IB Trial of the Anti-Cancer Stem Cell DLL4-Binding Agent Demcizumab with Pemetrexed and Carboplatin as First-Line Treatment of Metastatic Non-Squamous NSCLC. *Target Oncol*. 2018 Feb;13(1):89–98.

338. Chen X, Song X, Yue W, Chen D, Yu J, Yao Z, et al. Fibulin-5 inhibits Wnt/beta-catenin signaling in lung cancer. *Oncotarget*. 2015 Jun;6(17):15022–34.
339. Takahashi-Yanaga F, Kahn M. Targeting Wnt signaling: Can we safely eradicate cancer stem cells? *Clin Cancer Res*. 2010 Jun;16(12):3153–62.
340. Ali SA, Justilien V, Jamieson L, Murray NR, Fields AP. Protein Kinase Ciota Drives a NOTCH3-dependent Stem-like Phenotype in Mutant KRAS Lung Adenocarcinoma. *Cancer Cell*. 2016 Mar;29(3):367–78.
341. Chen C-Y, Chen Y-Y, Hsieh M-S, Ho C-C, Chen K-Y, Shih J-Y, et al. Expression of Notch Gene and Its Impact on Survival of Patients with Resectable Non-small Cell Lung Cancer. *J Cancer*. 2017;8(7):1292–300.
342. Park E, Park SY, Sun P-L, Jin Y, Kim JE, Jheon S, et al. Prognostic significance of stem cell-related marker expression and its correlation with histologic subtypes in lung adenocarcinoma. *Oncotarget*. 2016 Jul;7(27):42502–12.
343. Yanagawa J, Walser TC, Zhu LX, Hong L, Fishbein MC, Mah V, et al. Snail promotes CXCR2 ligand-dependent tumor progression in non-small cell lung carcinoma. *Clin Cancer Res*. 2009 Nov;15(22):6820–9.
344. Ma Y, Li M, Si J, Xiong Y, Lu F, Zhang J, et al. Blockade of Notch3 inhibits the stem-like property and is associated with ALDH1A1 and CD44 via autophagy in non-small lung cancer. *Int J Oncol*. 2016 Jun;48(6):2349–58.
345. Meng D, Yuan M, Li X, Chen L, Yang J, Zhao X, et al. Prognostic value of K-RAS mutations in patients with non-small cell lung cancer: a systematic review with meta-analysis. *Lung Cancer*. 2013 Jul;81(1):1–10.
346. Suzuki K, Kadota K, Sima CS, Nitadori J, Rusch VW, Travis WD, et al. Clinical Impact of Immune Microenvironment in Stage I Lung Adenocarcinoma: Tumor Interleukin-12 Receptor β 2 (IL-12R β 2), IL-7R, and Stromal FoxP3/CD3 Ratio Are Independent Predictors of Recurrence. *J Clin Oncol*. 2013 Feb;31(4):490–8.
347. Wang G, Ma W, Li Y, Jiang Y, Ma G, Zhang X, et al. Prognostic value of Twist, Snail and E-cadherin expression in pathological N0 non-small-cell lung cancer: a retrospective cohort study. *Eur J Cardiothorac Surg*. 2018 Aug;54(2):237–45.

348. Hung J-J, Yang M-H, Hsu H-S, Hsu W-H, Liu J-S, Wu K-J. Prognostic significance of hypoxia-inducible factor-1alpha, TWIST1 and Snail expression in resectable non-small cell lung cancer. *Thorax*. 2009 Dec;64(12):1082–9.
349. Mehta K, Moravcikova E, McFall D, Luketich JD, Pennathur A, Donnenberg AD, et al. The Mesenchymal State Predicts Poor Disease-Free Survival in Resectable Non-Small Cell Lung Cancer. *Ann Thorac Surg*. 2017 Jul;104(1):321–8.
350. Shoji T, Tanaka F, Takata T, Yanagihara K, Otake Y, Hanaoka N, et al. Clinical significance of p21 expression in non-small-cell lung cancer. *J Clin Oncol*. 2002 Sep;20:3865–71.
351. Kwon MS, Lee Y Il, Lee KY. p21 as a prognostic factor in non-small cell lung carcinomas. *Pathol Res Pract*. 2006;202(12):849–56.
352. Zheng W, Jiang C, Li R. Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. *Onco Targets Ther*. 2016 Apr;9:2317–27.
353. Chang W-H, Ho B-C, Hsiao Y-J, Chen J-S, Yeh C-H, Chen H-Y, et al. JAG1 Is Associated with Poor Survival through Inducing Metastasis in Lung Cancer. *PLoS One*. 2016 Mar;11(3):e0150355.
354. Hyakusoku H, Sano D, Takahashi H, Hatano T, Isono Y, Shimada S, et al. JunB promotes cell invasion, migration and distant metastasis of head and neck squamous cell carcinoma. *J Exp Clin Cancer Res*. 2016 Jan;35:6.
355. Farras R, Baldin V, Gallach S, Acquaviva C, Bossis G, Jariel-Encontre I, et al. JunB breakdown in mid-/late G2 is required for down-regulation of cyclin A2 levels and proper mitosis. *Mol Cell Biol*. 2008 Jun;28(12):4173–87.
356. Yuan X, Wu H, Xu H, Han N, Chu Q, Yu S, et al. Meta-analysis reveals the correlation of Notch signaling with non-small cell lung cancer progression and prognosis. *Sci Rep*. 2015 May;5:10338.
357. Zhao S, Qiu Z, Jin Y, Zhang L, Li W. Prognostic value of Thy1 in non-small cell lung cancer: A RNA-Seq transcriptome analysis. *Int J Clin Exp Pathol*. 2016;9(7):6999–7009.
358. Huang E, Ishida S, Pittman J, Dressman H, Bild A, Kloos M, et al. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet*. 2003 Jun;34(2):226–30.

359. Raponi M, Zhang Y, Yu J, Chen G, Lee G, Taylor JMG, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res.* 2006 Aug;66(15):7466–72.
360. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N Engl J Med.* 2004 Dec;351(27):2817-26.
361. Sanmartín E, Sirera R, Usó M, Blasco A, Gallach S, Figueroa S, et al. A Gene Signature Combining the Tissue Expression of Three Angiogenic Factors is a Prognostic Marker in Early-stage Non-small Cell Lung Cancer. *Ann Surg Oncol.* 2014 Feb;21(2):612–20.
362. Usó M, Jantus-Lewintre E, Calabuig-Fariñas S, Blasco A, García del Olmo E, Guijarro R, et al. Analysis of the prognostic role of an immune checkpoint score in resected non-small cell lung cancer patients. *Oncoimmunology.* 2017;6(1):e1260214.
363. Kim HI, Huang H, Cheepala S, Huang S, Chung J. Curcumin inhibition of integrin ($\alpha 6 \beta 4$)-dependent breast cancer cell motility and invasion. *Cancer Prev Res (Phila).* 2008 Oct;1(5):385–91.
364. Gupta SC, Patchva S, Aggarwal BB. Therapeutic roles of curcumin: lessons learned from clinical trials. *AAPS J.* 2013 Jan;15(1):195–218.
365. Baker M. Deceptive curcumin offers cautionary tale for chemists. *Nature.* 2017 Jan;541:144–5.
366. Salehi B, Venditti A, Sharifi-Rad M, Kregiel D, Sharifi-Rad J, Durazzo A, et al. The Therapeutic Potential of Apigenin. *Int J Mol Sci.* 2019 Mar;20(6).
367. Seo H-S, Ku JM, Choi HS, Woo J-K, Lee BH, Kim DS, et al. Apigenin overcomes drug resistance by blocking the signal transducer and activator of transcription 3 signaling in breast cancer cells. *Oncol Rep.* 2017 Aug;38(2):715–24.
368. Yan X, Qi M, Li P, Zhan Y, Shao H. Apigenin in cancer therapy: anti-cancer effects and mechanisms of action. *Cell Biosci.* 2017;7:50.
369. Wettersten HI, Hee Hwang S, Li C, Shiu EY, Weckslers AT, Hammock BD, et al. A novel p21 attenuator which is structurally related to sorafenib. *Cancer Biol Ther.* 2013 Mar;14(3):278–85.

VII. APPENDICES

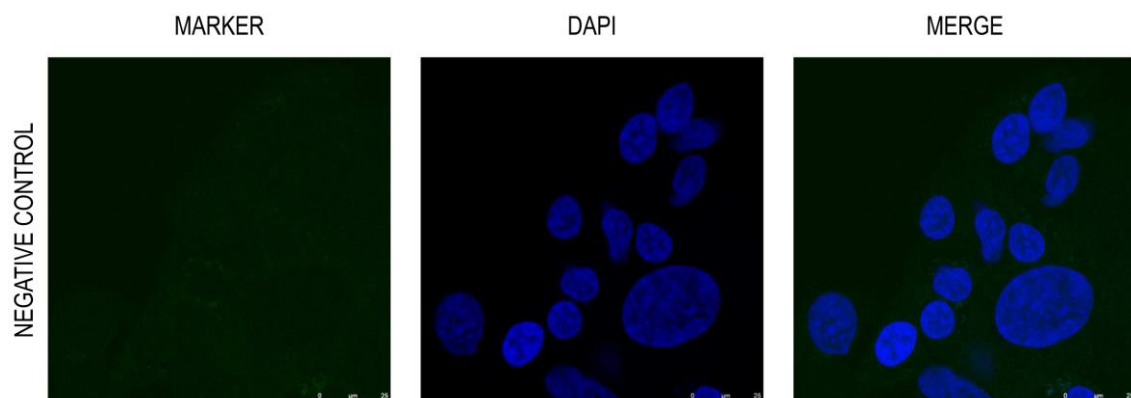
1. SUPPLEMENTARY MATERIAL

Supplementary Table 1. Efficiency results for the assays used in this study. The efficiency of each TaqMan® assay was evaluated by carrying out serial dilutions of a reference cDNA.

Gen	Slope	Efficiency	Percentage Efficiency
ABCG2	-3,396	1,970	99
ACTB	-3,322	2,000	100
ALDH1A1	-3,165	2,070	104
BMI1	-3,322	2,000	100
CCND1	-3,993	1,780	89
CD133	-3,322	2,000	100
CD166	-3,422	1,960	98
CD44	-3,644	1,881	94
CD45	-3,654	1,878	94
CDH1	-3,322	2,000	100
CDKN1A	-3,623	1,888	94
CDKN1B	-3,704	1,862	93
CDKN2A	-3,475	1,940	97
CEACAM5	-3,322	2,000	100
CTNNB1	-3,932	1,796	90
DKK1	-3,186	2,060	103
DLL1	-3,322	2,000	100
DLL4	-3,322	2,000	100
EPCAM1	-3,165	2,070	104
FZD7	-3,845	1,820	91
GAPDH	-3,322	2,000	100
GLI1	-3,805	1,831	92
GSK3B	-3,322	2,000	100
GUSB	-3,322	2,000	100
HES1	-3,322	2,000	100
HEY1	-3,743	1,850	93
HPRT1	-3,570	1,906	95
ITGA2	-3,322	2,000	100
ITGA6	-3,475	1,945	97
ITGB1	-3,644	1,881	94
JAG1	-3,551	1,929	96
JUNB	-3,810	1,830	92
KLF4	-3,541	1,916	96
LGALS2	-3,464	1,944	97

LIN28B	-3,831	1,824	91
MDM2	-3,617	1,890	95
MMP2	-3,899	1,805	90
MMP9	-3,947	1,792	90
MUC1	-3,322	2,000	100
MYC	-3,743	1,850	93
NANOG	-3,541	2,000	100
NOTCH1	-3,881	1,810	91
NOTCH2	-3,932	1,796	90
NOTCH3	-3,396	1,970	99
OCT4	-3,322	2,000	100
PTCH1	-3,711	1,860	93
SHH	-3,388	1,973	99
SMO	-2,946	2,185	109
SNAI1	-3,322	2,000	100
SOX2	-4,001	1,778	89
THY1	-3,928	1,790	90
VIM	-3,393	1,971	99
WEE1	-4,073	1,760	88
WNT1	-3,814	1,829	91
WNT2	-3,322	2,000	100
WNT3	-3,682	1,869	93
WNT5A	-3,710	1,860	93

Efficiency values were measured using the CT slope method. This method involves generating a dilution series of the target template and determining the CT value for each dilution. A plot of CT versus log cDNA concentration is constructed.



Supplementary Figure 1. Representative immunofluorescence images of the secondary antibody control. Green channel shows the secondary antibody staining, blue channel shows DAPI staining, and merge shows all channels merged. Scale bar represents 25 μ m.

2. APPROVAL FROM THE INSTITUTIONAL ETHICAL AND SCIENTIFIC REVIEW BOARD



Consorcio Hospital General Universitario de Valencia

Comité Ético de Investigación Clínica

APROBACIÓN PROYECTOS DE INVESTIGACIÓN

- ANEXO 11 -

Este CEIC tras evaluar en su reunión de 26 de Julio de 2012 el Proyecto de Investigación:

Título:	"Identificación de biomarcadores moleculares asociados a células madre tumorales en cáncer de pulmón no microcítico. Implicación en el desarrollo de nuevas estrategias terapéuticas."		
I.P.:	Dr. Carlos Camps	Servicio/Unidad:	Oncología

Acuerda respecto a esta documentación:

- Que el Proyecto de Investigación y Hoja de Información al Paciente y Consentimiento Informado presentado reúnen las condiciones exigidas por este CEIC, por tanto se decide su APROBACIÓN.

Los miembros que evaluaron esta documentación:

		Presente
Presidente	Dr. Severiano Marín	_____
Vocales	D. Ernesto Bataller	_____
	D. Alejandro Moner	_____ X _____
	D. Germán García	_____
	Dr. D. José Manuel Iranzo	_____
	Dr. D. Miguel Armengot	_____
	Dr. D. Julio Cortijo	_____ X _____
	Dra. Dña. Elena Rubio	_____ X _____
	Dr. D. Gustavo Juan	_____ X _____
	Dra. Pilar Blasco	_____
	Dña. M ^a Teresa Jareño	_____ X _____
	Dra. M ^a José Safont	_____
	Dra. Ana Blasco	_____
	Dr. Antonio Martorell	_____ X _____
	Dr. Aurelio Quesada	_____ X _____
	Dra. Begoña Peris	_____ X _____
	Dr. Fco. Javier Cervera	_____ X _____
	Dr. José Vte. Roig Vila	_____

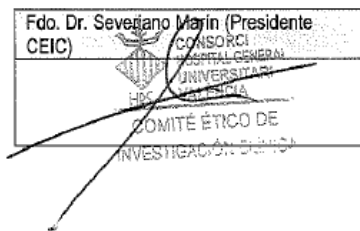


Consorcio Hospital General Universitario de Valencia

Comité Ético de Investigación Clínica

	Dr. Rafael Poveda	
	Dra. Inmaculada Sáez	X
	Dr. Alberto Berenguer	X
	Dr. Javier Milara	X
	Dña. Encarna Domingo	X
Secretario	Dra. Ana Miguez	X

Lo que comunico a efectos oportunos:



3. FUNDING

This thesis was supported by the following Spanish institutions:

- ❖ Red Temática de Investigación Cooperativa en Cáncer.
 - Project RD12/0036/0025.
- ❖ Centro de Investigación Biomédica en Red Cáncer.
 - Project CB16/12/00350.
- ❖ Fondo de Investigación Sanitaria-Fondo Europeo de Desarrollo Regional.
 - Projects PI12/00956, PI12/02838, PI15/00209 and PI15/00753.
- ❖ Fundación Arnal Planelles.
- ❖ Ministerio de Educación Cultura y Deporte.
 - Project FPU14/06911.

4. NATIONAL AND INTERNATIONAL CONGRESS COMMUNICATIONS

- ❖ Alejandro Herreros Pomares; Eloísa Jantus Lewintre; Silvia Calabuig Fariñas; Juan Diego De Maya Girones; Rut Lucas; Ana Blasco; Ricardo Guijarro; Miguel Martorell; Eva Escorihuela; María Dolores Chiara; Elena Duréndez; Carolina Gandia; Rafael Sirera; Rosa Farràs; Carlos Camps. “Lung tumorspheres characterization reveals cancer stem-like cells potential targets and prognostic biomarkers in non-small cell lung cancer”. (Oral Communication) IASLC 20th World Conference on Lung Cancer. Barcelona, Spain. 7-10 September 2019.
- ❖ Alejandro Herreros Pomares. “Analysis of cancer stem cells in human non-small cell lung cancer”. (Oral Communication) V Meeting of PhD Students from the Polytechnic University of Valencia. Valencia, Spain. 27 June 2019.
- ❖ Alejandro Herreros Pomares; Héctor Amado; Silvia Calabuig Fariñas; Eva Escorihuela; Juan Murga; Susana Torres; Elena Duréndez; Feiyu Zhang; Ana Blasco; Atilio Navarro; Cora Sampedro; Eloísa Jantus Lewintre. “Lung tumorspheres as a drug screening platform against cancer stem cells”. (Poster) European Lung Cancer Congress (ELCC) 2019. Geneva, Switzerland. 10-13 April 2019. *Annals of Oncology* (2019) 30 (suppl_2): ii7-ii13. 10.1093/annonc/mdz073.

- ❖ Alejandro Herreros Pomares; Elena Duréndez; Silvia Calabuig Fariñas; Eva Escorihuela; Cora Sampedro; Atilio Navarro; Ana Blasco; Rut Lucas; Rosa Farràs; Carlos Camps; Eloísa Jantus Lewintre. “An expression signature characterizes cancer stem cells from lung adenocarcinoma patients”. (Poster) 11as Jornadas de formación CIBERES/CIBERONC. Madrid, Spain. 15-16 November 2018.
- ❖ Alejandro Herreros Pomares; Elena Duréndez; Silvia Calabuig Fariñas; Eva Escorihuela; Cora Sampedro; Atilio Navarro; Ana Blasco; Rut Lucas; Rosa Farràs; Carlos Camps; Eloísa Jantus Lewintre. “An expression signature characterizes cancer stem cells from lung adenocarcinoma patients”. (Poster) I Encuentro de Jóvenes Investigadores de CIBERONC. Madrid, Spain. 14 November 2018.
- ❖ Alejandro Herreros Pomares; Elena Duréndez; Silvia Calabuig Fariñas; Juan Diego De Maya; Héctor Amado; Ana Blasco; Rut Lucas; Rosa Farràs; Eloísa Jantus Lewintre; Carlos Camps. “Patient-derived tumorspheres: a gene expression signatures characterizes lung adenocarcinoma stem cells”. (Poster) 16th Asociación Española de la Investigación sobre el Cáncer (ASEICA) International Congress. Valencia, Spain. 6-8 November 2018.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Juan Diego De Maya; Eva Escorihuela; Elena Duréndez; Ana Blasco; Atilio Navarro; Eva García Del Olmo; Rut Lucas; Rosa Farràs; Eloísa Jantus Lewintre; Carlos Camps. “An expression signature characterizes cancer stem cells from lung adenocarcinoma patients”. (Poster) European Society of Medical Oncology (ESMO) 2018 Congress. Munich, Germany. 19-23 October 2018. *Annals of Oncology*, Volume 29, Issue suppl_8, 1 October 2018, mdy303.051.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Juan Diego De Maya; Eva Escorihuela; Atilio Navarro; Eva García Del Olmo; Rut Lucas; Rosa Farràs; Eloísa Jantus Lewintre; Carlos Camps. “Establecimiento de una firma de genes capaces de identificar células madre tumorales en adenocarcinoma de pulmón”. (Oral Communication) Congreso de la Sociedad Española de Oncología Médica (SEOM) 2018. Madrid, Spain. 25-28 September 2018.
- ❖ Alejandro Herreros Pomares; Juan De Maya Gironés; Héctor Amado; Cristóbal Aguilar Gallardo; Eva Escorihuela; Ana Blasco; Silvia Calabuig Fariñas; Miguel Martorell; Ricardo Guijarro; Eloísa Jantus Lewintre; Rosa Farràs; Carlos Camps. “A gene expression signature to

characterize human lung adenocarcinoma cancer stem cells.” (Abstract book) 2018 American Society of Clinical Oncology (ASCO) Annual Meeting. Chicago, Illinois, USA. 1-5 June 2018. *Journal of Clinical Oncology* 36, no. 15_suppl.e20547.

- ❖ Alejandro Herreros Pomares; Cristóbal Aguilar; Héctor Amado; Eva Escorihuela; Ana Blasco; Silvia Calabuig Fariñas; Juan Murga Clausell; Eloísa Jantus Lewintre; Carlos Camps. “Lung tumor spheres as in vitro platform for testing new therapeutic strategies against cancer stem cells”. (Poster) American Association for Cancer Research (AACR) Annual Meeting 2018. Chicago, Illinois, USA. 14-18 April 2018. *Cancer Res* 2018;78(13 Suppl):Abstract nr 3056.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Susana Torres; Elena Duréndez; Rut Lucas; Cristóbal Aguilar; Eva García Del Olmo; Miguel Martorell; Eloísa Jantus Lewintre; Carlos Camps Herrero. “Stemness characterization of tumorspheres from non-small cell lung cancer. Differential expression in CSC-related markers and signaling pathways.” (Poster) II CIBERONC General Meeting. Madrid, Spain. 2 February 2018.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Héctor Amado; Critobal Aguilar; Eva Escorihuela; Rut Lucas; Miguel Martorell; Ricardo Guijarro; Eloísa Jantus Lewintre; Carlos Camps. “Characterization of tumor spheres derived from non-small cell lung cancer patients. Differential expression in stemness markers and signaling pathways.” (Oral Communication). II National Congress of Young Researchers in Biomedicine. Valencia, Spain. 23-24 November 2017.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Héctor Amado Labrador; Cristóbal Aguilar Gallardo; Eva Escorihuela; Rut Lucas; Miguel Martorell; Ricardo Guijarro; Eloísa Jantus Lewintre; Carlos Camps. “Characterization of tumor spheres derived from non-small cell lung cancer patients. Differential expression in stemness markers and signaling pathways”. (Poster) 1st ASEICA EDUCATIONAL SYMPOSIUM. Madrid, Spain. 14-15 November 2017.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Susana Torres; Elena Duréndez; Rut Lucas; Cristóbal Aguilar; Eva García Del Olmo; Miguel Martorell; Eloísa Jantus; Carlos Camps. “Análisis de marcadores y vías de señalización de célula madre tumoral en tumoresferas de

cáncer de pulmón no microcítico.” (Highlighted Poster) Congreso SEOM 2017. Madrid, Spain. 24-27 October 2017.

- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Eva Escorihuela; Héctor Amado; Ricardo Guijarro; Sandra Gallach; Atilio Navarro; Eloísa Jantus Lewintre; Carlos Camps Herrero. “Stemness characterization of tumorspheres from non-small cell lung cancer. differential expression in CSC-related markers and signaling pathways.” (Poster Discussion) ESMO 2017 Congress. Madrid, Spain. 8-12 September 2017. *Annals of Oncology* (2017) 28 (suppl_5): v453-v456".
- ❖ Alejandro Herreros Pomares; Ester Munera; Silvia Calabuig Fariñas; Rut Lucas; Rosa Farràs; Ana Blasco; Susana Torres; Jose Ferri; Ricardo Guijarro; Miguel Martorell; Eloísa Jantus Lewintre; Carlos Camps. “Gene characterization of lung-tumorspheres for their usage as an in vitro screening platform for testing new therapeutic strategies” (Poster) AACR Annual Meeting 2017. Washington, USA. 1-5 April 2017. *Cancer Res* 2017;77 (13 Suppl):Abstract nr 1918".
- ❖ Alejandro Herreros Pomares; Ester Munera; Silvia Calabuig Fariñas; Borja Lafuente; Ana Blasco; Ricardo Guijarro; Rosa Farràs; Eloísa Jantus Lewintre; Carlos Camps Herrero. “Expression analysis of tumorspheres from non-small cell lung cáncer show significant differences in CSC-markers and signaling pathways.” (Poster) II Encuentro de investigadores en cáncer ciudad de Alcoy: ‘La investigación al servicio del paciente oncológico’. Alcoy, Spain. 23 March 2017.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Ester Munera Maravilla; Ana Blasco; Aminta Isabel Martínez; Eva García del Olmo; Eloísa Jantus Lewintre; Carlos Camps. “Stemness gene expression profile of tumorspheres from non-small cell lung cancer.” (Poster) IASLC 17th World Conference on Lung Cancer. Vienna, Austria. 4-7 December 2016. *Journal of Thoracic Oncology*, Volume 12, Issue 1, S620 – S621.
- ❖ Alejandro Herreros Pomares; Eloísa Jantus Lewintre; Ester Munera Maravilla; Silvia Calabuig Fariñas; Sandra Gallach; Atilio Navarro; Eva García del Olmo; Ricardo Guijarro; Ana Blasco; Carlos Camps. “Análisis de expresión de tumoresferas derivadas de líneas celulares y

pacientes con cáncer de pulmón no microcítico.” (Poster) SEOM 2016. Madrid, Spain. 26-28 October 2016.

- ❖ Alejandro Herreros Pomares; Ester Munera Maravilla; Silvia Calabuig Fariñas; Aminta Isabel Martínez; Ana Blasco; Atilio Navarro; Carlos Camps; Eloísa Jantus Lewintre. “Gene expression analysis of cancer stem cell markers and signaling pathways from non-small cell lung cancer.” (Oral communication) VI Meeting of Young Researchers of Red Temática de Investigación Cooperativa en Cáncer (RTICC). Salamanca, Spain. 23 September 2016.
- ❖ Alejandro Herreros Pomares; Ester Munera Maravilla; Eloísa Jantus Lewintre; Silvia Calabuig Fariñas; Ana Blasco; Aminta Isabel Martínez; Atilio Navarro; Ricardo Guijarro; Miguel Martorell; Carlos Camps. “Differential expression of tumorspheres in CSC-markers and signaling pathways from non-small cell lung cancer.” (Abstract book) ASCO Congress 2016. Chicago, USA. 3-7 June 2016. *J Clin Oncol* 34, 2016 (suppl; abstr e23276).
- ❖ Alejandro Herreros Pomares; Ester Munera Maravilla; Alicia Martínez Romero; Sandra Tejedor; Silvia Calabuig Fariñas; Eloísa Jantus Lewintre; Rut Lucas; Eva Escorihuela; Rosa Farràs; Carlos Camps Herrero. “Characterization of lung-tumorspheres by gene expression and flow cytometry. Differential expression in CSC-related markers and signaling pathways.” (Poster) AACR Annual Meeting 2016. New Orleans, USA. 16-20 April 2016. *Cancer Res* 2016;76(14 Suppl):Abstract nr 3354.
- ❖ Alejandro Herreros Pomares; Ester Munera Maravilla; Silvia Calabuig Fariñas; Borja Lafuente Gutiérrez; Ana Blasco; Ricardo Guijarro; Rosa Farràs; Eloísa Jantus Lewintre; Carlos Camps Herrero. “Expression analysis of tumorspheres from non-small cell lung cancer show significant differences in CSC-markers and signaling pathways.” (Poster) European Lung Cancer Congress ELCC 2016 Geneva, Switzerland. 13-16 April 2016. *J Thorac Oncol.* 2016 Apr;11(4 Suppl):S105.
- ❖ Alejandro Herreros Pomares; Silvia Calabuig Fariñas; Ester Munera Maravilla; Andrea Palomar Ríos; Ricardo Guijarro; Eloísa Jantus Lewintre; Carlos Camps Herrero. “Comparison of gene expression profile between lung tumorspheres and adherent cells in non-small cell lung cancer (NSCLC)”. (Poster) V Meeting of Young Researchers (RTICC) Pamplona, Spain. 28-30 October 2015.

5. PUBLICATIONS

- ❖ “Lung tumorspheres reveal cancer stem cell-like properties and a score with prognostic impact in resected non-small cell lung cancer”. Herreros-Pomares A, de Maya Girones JD, Calabuig-Fariñas S, Lucas R, Martínez A, Pardo-Sánchez JM, Alonso S, Blasco A, Guijarro R, Martorell M, Eva Escorihuela E, Chiara MD, Duréndez E, Gandia C, Forteza J, Sirera R, Jantus-Lewintre E, Farràs R, Camps C. Manuscript sent to Cell Death & Disease in May 2019.
- ❖ “Passenger mutations in cancer evolution”. Aparisi F, Amado-Labrador H, Calabuig-Fariñas S, Torres S, Herreros-Pomares A, Jantus-Lewintre E, Blasco A, Iranzo V and Camps C. Cancer Rep Rev, 2019 Jun; 3:1-8.
- ❖ “EpCAM duality becomes this molecule in a new Dr. Jekyll and Mr. Hyde tale”. Herreros-Pomares A, Aguilar-Gallardo C, Calabuig-Fariñas S, Sirera R, Jantus-Lewintre E, Camps C. Crit Rev Oncol Hematol. 2018 Jun; 126:52-63.
- ❖ “Circulating tumor cells versus circulating tumor DNA in lung cancer-which one will win?”. Calabuig-Fariñas S, Jantus-Lewintre E, Herreros-Pomares A, Camps C. Transl Lung Cancer Res. 2016 Oct;5(5):466-482.

6. AWARDS

- ❖ Best oral communication for “Characterization of tumor spheres derived from non-small cell lung cancer patients. Differential expression in stemness markers and signaling pathways” presented at the II National Congress for Young Researchers in Biomedicine. Valencia, Spain. 23-24 November 2017.
- ❖ Second prize in the XII edition of the scientific and technical award <<ciutat de Algemesí>>. Algemesí, Spain. February 2017.

ARTICLE

Open Access

Lung tumorspheres reveal cancer stem cell-like properties and a score with prognostic impact in resected non-small-cell lung cancer

Alejandro Herreros-Pomares^{1,2}, Juan Diego de-Maya-Girones³, Silvia Calabuig-Fariñas^{1,2,4}, Rut Lucas⁵, Alicia Martínez⁶, José Miguel Pardo-Sánchez³, Sergio Alonso⁷, Ana Blasco⁸, Ricardo Guijarro⁹, Miguel Martorell¹⁰, Eva Escorihuela^{1,2}, María Dolores Chiara^{2,11}, Elena Duréndez^{1,2}, Carolina Gandía³, Jerónimo Forteza¹², Rafael Sirera^{2,13}, Eloísa Jantus-Lewintre^{1,2,13}, Rosa Farràs³ and Carlos Camps^{1,2,8,14}

Abstract

The high resistance against current therapies found in non-small-cell lung cancer (NSCLC) has been associated to cancer stem-like cells (CSCs), a population for which the identification of targets and biomarkers is still under development. In this study, primary cultures from early-stage NSCLC patients were established, using sphere-forming assays for CSC enrichment and adherent conditions for the control counterparts. Patient-derived tumorspheres showed self-renewal and unlimited exponential growth potentials, resistance against chemotherapeutic agents, invasion and differentiation capacities in vitro, and superior tumorigenic potential in vivo. Using quantitative PCR, gene expression profiles were analyzed and *NANOG*, *NOTCH3*, *CD44*, *CDKN1A*, *SNAI1*, and *ITGA6* were selected to distinguish tumorspheres from adherent cells. Immunoblot and immunofluorescence analyses confirmed that proteins encoded by these genes were consistently increased in tumorspheres from adenocarcinoma patients and showed differential localization and expression patterns. The prognostic role of genes significantly overexpressed in tumorspheres was evaluated in a NSCLC cohort ($N = 661$) from The Cancer Genome Atlas. Based on a Cox regression analysis, *CDKN1A*, *SNAI1*, and *ITGA6* were found to be associated with prognosis and used to calculate a gene expression score, named CSC score. Kaplan–Meier survival analysis showed that patients with high CSC score have shorter overall survival (OS) in the entire cohort [37.7 vs. 60.4 months (mo), $p = 0.001$] and the adenocarcinoma subcohort [36.6 vs. 53.5 mo, $p = 0.003$], but not in the squamous cell carcinoma one. Multivariate analysis indicated that this gene expression score is an independent biomarker of prognosis for OS in both the entire cohort [hazard ratio (HR): 1.498; 95% confidence interval (CI), 1.167–1.922; $p = 0.001$] and the adenocarcinoma subcohort [HR: 1.869; 95% CI, 1.275–2.738; $p = 0.001$]. This score was also analyzed in an independent cohort of 114 adenocarcinoma patients, confirming its prognostic value [42.90 vs. not reached (NR) mo, $p = 0.020$]. In conclusion, our findings provide relevant prognostic information for lung adenocarcinoma patients and the basis for developing novel therapies. Further studies are required to identify suitable markers and targets for lung squamous cell carcinoma patients.

Correspondence: Eloísa Jantus-Lewintre (jantus_elo@gva.es) or Rosa Farràs (rfarras@cipf.es)

¹Molecular Oncology Laboratory, Fundación Hospital General Universitario de Valencia, Valencia, Spain

²CIBERONC, Valencia, Spain

Full list of author information is available at the end of the article.

These authors contributed equally: Alejandro Herreros-Pomares, Juan Diego de-Maya-Girones

Edited by N. Barlev

© The Author(s) 2019



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, 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. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

large-cell carcinoma, and 75% are diagnosed at advanced stages, when surgery is not possible². Significant advances in the development of treatments against driver mutations and immune-based therapies for these patients were achieved in recent years^{3,4}, but many patients still develop treatment resistance, progress, and die^{5,6}. Curative surgery is the standard of care for early-stage patients with a good performance status, but the recurrence rate ranges from 35 to 50% and, after an apparently successful surgical treatment, appearance of secondary tumors often leads to the relapse of resected patients⁷. This poor prognosis greatly supports the efforts to establish prognostic biomarkers and therapeutic targets for improving the management of NSCLC.

Among these targets, cancer stem-like cells (CSCs) are proposed as a promising tumor population, since they are believed to survive after conventional cancer treatments and regenerate tumors even when are undetectable⁸. These slow-diving cells are characterized by their self-renewal potential, the capacity to undergo asymmetric division to highly proliferative cells, and a great tumorigenic activity, acting like tumor-initiating cells and producing aggressive tumors when transplanted in immune-compromised mice⁹. Additionally, CSCs share with other stem cells the overexpression of cytoprotective enzymes, enhanced ability to efflux molecules, and the anchorage-independent growth ability, which can be used for enriching cell cultures in this population. Nevertheless, specific strategies against CSCs are not approved in clinical practice and a better understanding of their impact on patients' prognosis is required.

CSCs have been analyzed in several solid tumors, including brain, lung, breast, colon, or pancreas, finding aberrant expression of different molecules and signaling pathways^{10–14}. However, most of the studies have been performed on cell lines and conflicting data can be found. For instance, CD133 and CD44 molecules have been successfully used to identify lung cancer cells with CSC properties in some studies^{15,16}, whereas other publications reported that CD133⁻ or CD44⁻ cell populations also possess the ability for self-renewal and enhanced tumor initiation capacity when transplanted into mice^{17,18}. Cell lines are a powerful tool and offer several advantages over primary cells, but they do not completely mimic them and studies including primary cultures are required when analyzing stemness properties, markers, pathways, and novel approaches. To date, few publications include primary cultures from NSCLC patients' tissue and, if included, the number of patients is small, since primary culture establishment is difficult and CSCs constitute an uncommon population^{19–21}.

The aim of this study was to characterize the population of CSCs derived from resected NSCLC patients to identify genes and molecules that could have a prognostic role or

constitute the basis for developing novel therapies focusing on this tumor population. We have confirmed that tumorspheres from NSCLC patients are enriched in cells with stem properties, identified potential targets against this aggressive population, and established a three-gene signature that is an independent prognostic marker for patients' OS.

Materials and methods

Patients and tissue samples

This study included 134 patients from the General University Hospital of Valencia who underwent surgery between 2004 and 2016 and who fit the eligibility criteria: resected, non-pretreated stage I–IIIA patients (according to the American Joint Committee on Cancer staging manual) with a histological diagnosis of NSCLC. Lung tumor specimens were obtained at the time of surgery. Tumor samples from 20 patients were immediately processed for primary culture establishment. The rest of samples were preserved in RNAlater (Applied Biosystems, USA) to avoid degradation of RNA and were frozen at -80°C until gene expression analyses. The mutation status of *KRAS*, *TP53*, *EGFR*, *ALK*, and *ROS1* was assessed for the whole cohort. *KRAS* gene mutations in codons 12, 13, and 61 were quantitatively detected by pyrosequencing using the theraScreen® KRAS Pyro® kit (Qiagen, Germany). *EGFR* mutations were analyzed by quantitative real-time PCR (RTqPCR) using the theraScreen® EGFR RGQ PCR (Qiagen, Germany), whereas *TP53* mutations were determined using standard PCR followed by Sanger sequencing. *ALK* and *ROS1* rearrangements were determined by immunohistochemistry (IHC) using ALKp80 (MAD-001720QD) and *ROS1* (MAD-000746QD). Antibodies were from Master Diagnostica (Granada, Spain), respectively.

Establishment of primary cell cultures

Unless specified, all reagents were obtained from Gibco Paisley, UK. Surgical tumor specimens were washed and minced into small pieces. Tumor dissociation was carried out by enzymatic digestion (1 mg/mL collagenase type IV, 1 mg/mL dispase, and 0.001% DNase, Sigma, St. Louis, USA) for 3 h at 37°C . Half of cells were cultured in collagen-coated flasks with Advanced DMEM-F12 supplemented with 10% fetal bovine serum (FBS), 200 $\mu\text{g}/\text{mL}$ penicillin/streptomycin, and 2 mM L-glutamine. The rest of the cells were seeded at low density in ultra-low attachment plates (Corning, Lowell, MA, USA) with serum-free Advanced DMEM-F12 medium supplemented with 0.4% bovine serum albumin (BSA), 50 $\mu\text{g}/\text{mL}$ epidermal growth factor (EGF), 20 $\mu\text{g}/\text{mL}$ basic fibroblast growth factor (bFGF), 5 $\mu\text{g}/\text{mL}$ insulin-transferrin-selenium (ITS) PREMIX (Corning, Lowell, MA, USA), 2% B-27, 200 $\mu\text{g}/\text{mL}$ penicillin/streptomycin, and 2 mM L-

glutamine to support their growth as undifferentiated tumorspheres. Cultures were expanded by mechanical dissociation of spheres, followed by re-plating of both single cells and residual small aggregates in complete fresh medium. In all cases, cells were maintained at 37 °C in 5% CO₂ atmosphere and the medium was replaced twice a week.

Cell line cultures

A549, NCI-H1395, NCI-H1650, NCI-H1975, NCI-H1993, NCI-H2228, NCI-H23, NCI-H358, NCI-H460, HCC827, PC9, and SW900 cells were purchased from American Type Culture Collection (Supplementary Table S1). Cell lines were cultured in RPMI-1640 containing 10% FBS, 200 µg/mL penicillin/streptomycin, and 0.001% non-essential amino acids. To obtain tumorspheres, the cells were trypsinized using 0.05% trypsin-EDTA when they reached 80% confluence. The cells were seeded at low density in ultra-low attachment flasks with serum-free RPMI-1640 medium supplemented with 0.4% BSA, 50 µg/mL EGF, 20 µg/mL bFGF, 5 µg/mL ITS PREMIX, 2% B-27, 200 µg/mL penicillin/streptomycin, and 2 mM L-glutamine.

Animals and xenografts

To test the tumorigenic potential of adherent cells and tumorspheres, 6-week-old NOD.CB17-Prkdc^{scid}/NcrCr1 mice (Jackson Laboratories) were subcutaneously transplanted with cell suspensions in serum-free medium and Matrigel (BD) (1:1). Tumor volume (TV) measurements were recorded once a week using the formula: TV (mm³) = $d^2 \times D/2$, where d and D are the shortest and the longest diameter, respectively²². Animals were terminated when xenografts were 1000 mm³.

Cell invasion assays and time-lapse video recording

For cell invasion assays, cells were cultured in the medium used for tumorsphere formation supplemented with 0.2% methylcellulose in a non-adhesive convex environment for 12 h at 37 °C and 5% CO₂. Tumorspheres were mixed with collagen matrix (2.5 mg/ml) and incubated for 30 min at 37 °C prior to microscopic analysis. Time-lapse microscopy imaging was performed on a Zeiss AxioObserver Z1 microscope with a Plan-Apochromat ×40 /1.3 (NA = 1.3, working distance = 0.21 mm), a camera, and an Apotome attachment (Carl Zeiss, Germany). Mosaic images were collected using AxioVision software over a period of 20 h with a time resolution of 30 min.

Cell growth curves

Cultures were trypsinized at 80% confluence and counted in a Neubauer camera with Trypan Blue dye exclusion (Sigma, USA). Tumor cells were plated at a

density of 1000 cells per well in 96-well plates and cell viability was evaluated 24, 48, and 72 h after seeding with the CellTiter 96[®] Aqueous One Solution Cell Proliferation Assay (Promega, WI, USA) according to standard protocols and analyzed with a Victor 3 plate reader (Wallac, Turku, Finland). Data represented are the mean of three replicates in three independent experiments.

Cytotoxicity assays

Adherent cells and lung tumorspheres were cultured at desired density according to their growth curves into 96-well plates. Chemotherapeutic agents were added after 24 h at the following final concentrations: cisplatin 50 µM, docetaxel 10 µM, paclitaxel 10 µM, vinorelbine 10 µM, and pemetrexed 50 µM (Selleckchem, Germany). The selective agent against CSCs, salinomycin, was added at 1 µM (Selleckchem, Germany). Cells treated with dimethyl sulfoxide (vehicle control) served as controls. Cell viability was evaluated after 48 h with the CellTiter 96[®] Aqueous One Solution Cell Proliferation Assay and analyzed with a Victor 3 plate reader. Data represented are the mean of three replicates in three independent experiments.

Gene expression analysis

RTqPCR was performed to analyze the relative expression of 51 CSC-related genes on a Roche LightCycler[®] 480 II system (Roche Ltd., Basel, Switzerland) (Supplementary Table S2). RNA from cell pellets and frozen tissue samples was extracted using standard TRIZOL (Invitrogen) method. Reverse transcription reactions were performed from 1.0 µg of total RNA using random hexanucleotides and a High-Capacity cDNA (complementary DNA) Reverse Transcription Kit (Applied Biosystems, USA) following the manufacturer's instructions. The thermal cycling conditions were as follows: 10 min at 25 °C, 120 min at 37 °C, and 5 s at 85 °C. RTqPCR was performed with assays based on hydrolysis probes using 1 µL of cDNA, TaqMan Gene Expression Master Mix, and a TaqMan Gene Expression Assay (Applied Biosystems, USA) in a 5 µL final reaction volume. The thermal cycling parameters were as follows: 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. For efficiency calculations, we used random-primed qPCR Human Reference cDNA (Clontech, USA). *ACTB*, *GUSB*, and *CDKN1B* were selected as endogenous controls using GeNorm software. Relative gene expression levels were expressed as the ratio of target gene expression to the geometric mean of the endogenous gene expressions according to Pfaffl formula²³.

Immunoblot analysis

Tumorspheres were washed with cold phosphate-buffered saline (PBS) and lysed on ice with lysis buffer

(50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.02% NaN_3 , 0.1% SDS, 1% NP40, 0.5% sodium deoxycholate, 2 mg/ml leupeptin, 2 mg/ml aprotinin, and 1 mM phenylmethylsulfonyl fluoride), whereas adherent cells were washed with cold PBS and scraped out of the dishes before lysis. Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, transferred to polyvinylidene difluoride membranes, probed with the indicated antibodies (Supplementary Table S3), and detected by chemiluminescence. Densitometric analysis was performed using ImageJ (NIH, USA) and all results were normalized over β -actin.

Immunofluorescence analysis

Cells were fixed in 4% paraformaldehyde in PBS at room temperature for 15 min, washed and permeabilized with 0.4% Triton X-100 in PBS for 10 min, and washed again. Fixed cells were blocked in PBS containing 10% BSA and 0.4% Triton X-100 for 1 h. Immunodetection was carried out using the same antibodies described above (Supplementary Table S3). Cells were incubated overnight at 4 °C with primary antibodies in blocking buffer. Thereafter, secondary antibodies contained in the blocking buffer were incubated for 1 h. Slides were incubated with 4',6-diamidino-2-phenylindole for 3 min, mounted with Fluoromount Aqueous Mounting Medium (Sigma, USA), and analyzed using a Leica confocal microscope (Leica Microsystems, IL, USA).

IHC analysis

Adherent cells and tumorspheres were fixed in 4% paraformaldehyde in PBS at room temperature for 15 min, washed, and embedded in Richard-Allan Scientific™ His-to-Gel (Thermo Scientific, UK) prior to paraffin. Immunodetection was done using CD56 (MAD-000749QD), CEAm (MAD-002095QD), CEAp (MAD-001115QD), cytokeratin 5/6 (MAD-000651QD), cytokeratin 7 (MAD-001004QD), and p63 (MAD-000479QD) antibodies from Master Diagnostica (Granada, Spain).

Bioinformatic analysis

In silico analysis was performed using two lung cancer data sets from the The Cancer Genome Atlas (TCGA) consortium^{24,25}. Clinical and RNA-sequencing (Illumina HiSeq platform) information was directly downloaded from the ICGC Data Portal²⁶, <https://dcc.icgc.org/releases/current/projects/LUAD-US>, and <https://dcc.icgc.org/releases/current/projects/LUSC-US>.

Statistical analysis

Expression of paired adherent cells and tumorspheres were analyzed using non-parametric Wilcoxon's signed-rank test. In order to reduce the dimensionality and remove possible collinear expression of genes, a logistic

regression model was built using a stepwise selection and minimizing Akaike's information criterion to select the genes, which contributed more to differentiate tumorspheres from adherent cells. Continuous variables were compared by non-parametric Mann–Whitney *U* and Kruskal–Wallis tests. A Spearman's rank test was used to test for correlations between continuous variables and the association between discrete variables was evaluated by the χ^2 test. Survival analyses were performed using univariate Cox regression analysis and Kaplan–Meier (log-rank) test method with clinicopathological variables and dichotomized gene expression levels. To assess the independent value of the tested biomarkers, a Cox proportional hazard model for multivariate analyses was used. All significant variables from the univariate were entered into the multivariate analyses in a forward stepwise Cox regression analysis. Furthermore, we also calculated gene expression score based on multi-gene signature using a method previously reported^{27,28}. Univariate Cox regression analysis was used to select genes associated with mortality (*Z*-score >1.5), which were afterwards included in a multivariate risk model. All genes were included for these purposes, and expression values for all analyses were continuous variables. A probability of 95% ($p < 0.05$) was considered statistically significant for all analyses. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, Chicago, IL, USA) version 15.0. Principal component analysis (PCA) for gene expression analyses were performed with the SIMCA-P software (version 13.0, Umetrics, Umea, Sweden) using unit variance scaling method.

Results

Generation of adherent and non-adherent primary NSCLC cultures

Clinicopathological information from each of the 20 patients included in this part of the study is summarized in Table 1. The median patient age was 68 years [range: 54–83], 65% were males, 65% had ADC, and 55% of patients were diagnosed at stage I of the disease. The median follow-up was 17.32 months [range: 5.1–33.23], and eight (40%) relapsed or died during the follow-up period.

Primary patient-derived lung cancer cell cultures were maintained for 4 weeks before they were split for the first passage. Patient-derived cultures were successfully established in 8 out of 20 cases (40%), being able to grow tumor cells as monolayers and tumorspheres. No significant associations were found between the analyzed clinicopathological variables and the establishment of primary cultures. The morphology of cells from patient-derived cultures was examined and heterogeneity was observed on the adherent-cultured cells between samples (Fig. 1a). Cells from patients FIS299 and FIS301 grew as

Table 1 Clinicopathological characteristics and lung tumorspheres formation of the patients included in the study^a

Patient code	Gender	Age (years)	TNM stage	Histology	Smoking status	Progression/exitus	DFS (months)	Mutational status	Tumorspheres formation
FIS291	Male	66	IIA (T2aN0M0)	SCC	Current	Yes	25.77	<i>KRAS</i> G12C	No
FIS299	Male	69	IIIA (T4N1M0)	SCC	Former	Yes	6.10	<i>TP53</i> K132E	Yes
FIS301	Male	71	IIB (T3N0M0)	SCC	Former	No	30.77	<i>TP53</i> D259fs*84	Yes
FIS302	Female	74	IIA (T2aN0M0)	ADC	Never	No	33.23	<i>KRAS</i> G12D, <i>TP53</i> E285K	Yes
FIS303	Female	57	IB (T2aN0M0)	ADC	Former	Yes	11.13	<i>TP53</i> R175H	Yes
FIS308	Male	72	IIB (T3N0M0)	SCC	Current	No	27.80	No mutation detected	No
FIS310	Male	68	IIIA (T3N2M0)	ADC	Former	Yes	6.43	<i>TP53</i> V157F R213SNP	No
FIS312	Male	62	IA (T1bN0M0)	ADC	Current	No	24.80	No mutation detected	No
FIS315	Female	65	IA (T1aN0M0)	ADC	Never	No	20.93	No mutation detected	Yes
FIS317	Male	76	IIB (T3N0M0)	SCC	Current	Yes	18.40	No mutation detected	Yes
FIS320	Male	65	IB (T2aN0M0)	ADC	Current	No	23.60	<i>TP53</i> P153fs*26	Yes
FIS321	Male	83	IB (T2aN0M0)	SCC	Current	No	22.50	<i>TP53</i> Q156*	No
FIS325	Female	67	IB (T2aN0M0)	ADC	Never	No	16.97	<i>EGFR</i> L858R	No
FIS326	Female	64	IB (T2aN0M0)	ADC	Former	Yes	6.97	<i>EGFR</i> L858R, <i>TP53</i> G244C	No
FIS330	Male	54	IA (T1aN0M0)	ADC	Current	No	5.10	<i>TP53</i> R283P	No
FIS331	Male	75	IIA (T2aN1M0)	ADC	Current	Yes	6.20	<i>TP53</i> R175H	No
FIS337	Male	73	IB (T2aN0M0)	ADC	Former	No	8.27	<i>KRAS</i> G12S	No
FIS343	Female	60	IB (T2aN0M0)	ADC	Former	Yes	7.00	<i>TP53</i> R158L	Yes
FIS345	Male	74	IIIA (T1aN2M0)	SCC	Current	No	7.80	No mutation detected	No
FIS347	Female	68	IB (T2aN0M0)	ADC	Never	No	17.67	No mutation detected	No

DFS disease-free survival, ADC adenocarcinoma, SCC squamous cell carcinoma, WT wild type

^aNo significant associations were found between clinicopathological characteristics and tumorspheres formation

multilayers and formed cell colonies. These cell cultures showed abundant cell–cell interactions in the form of filopodia and lamellipodia and presence of giant cells and vesicles (Supplementary Fig. S1). FIS317 cells were cubic, grew as a monolayer, but showed tight cell–cell contact with filopodia and a high number of vesicles. In contrast, cells from patients FIS302, FIS303, and FIS315 were more elongated, with brighter nuclei, fewer interactions, and a more isolated growth. FIS320 and FIS343 cells were similar to patients FIS302, FIS303, and FIS315 in terms of growth rate (Supplementary Fig. S2), but had a different morphology. Regarding tumorspheres, tight spheroids

were formed by FIS299, FIS301, and FIS315 cultures, whereas FIS302, FIS303, FIS317, and FIS320 formed more loose and irregularly shaped spheres, and FIS343 showed a mixed behavior (Fig. 1b). Simultaneously, cell line cultures were established in both conditions, adherence and suspension, and were included in further gene expression analyses (Supplementary Fig. S3).

Lung tumorspheres exhibit stemness features

One of our aims was to analyze if lung tumorspheres displayed stemness properties. To determine the self-renewal and growth potentials of lung tumorspheres, we

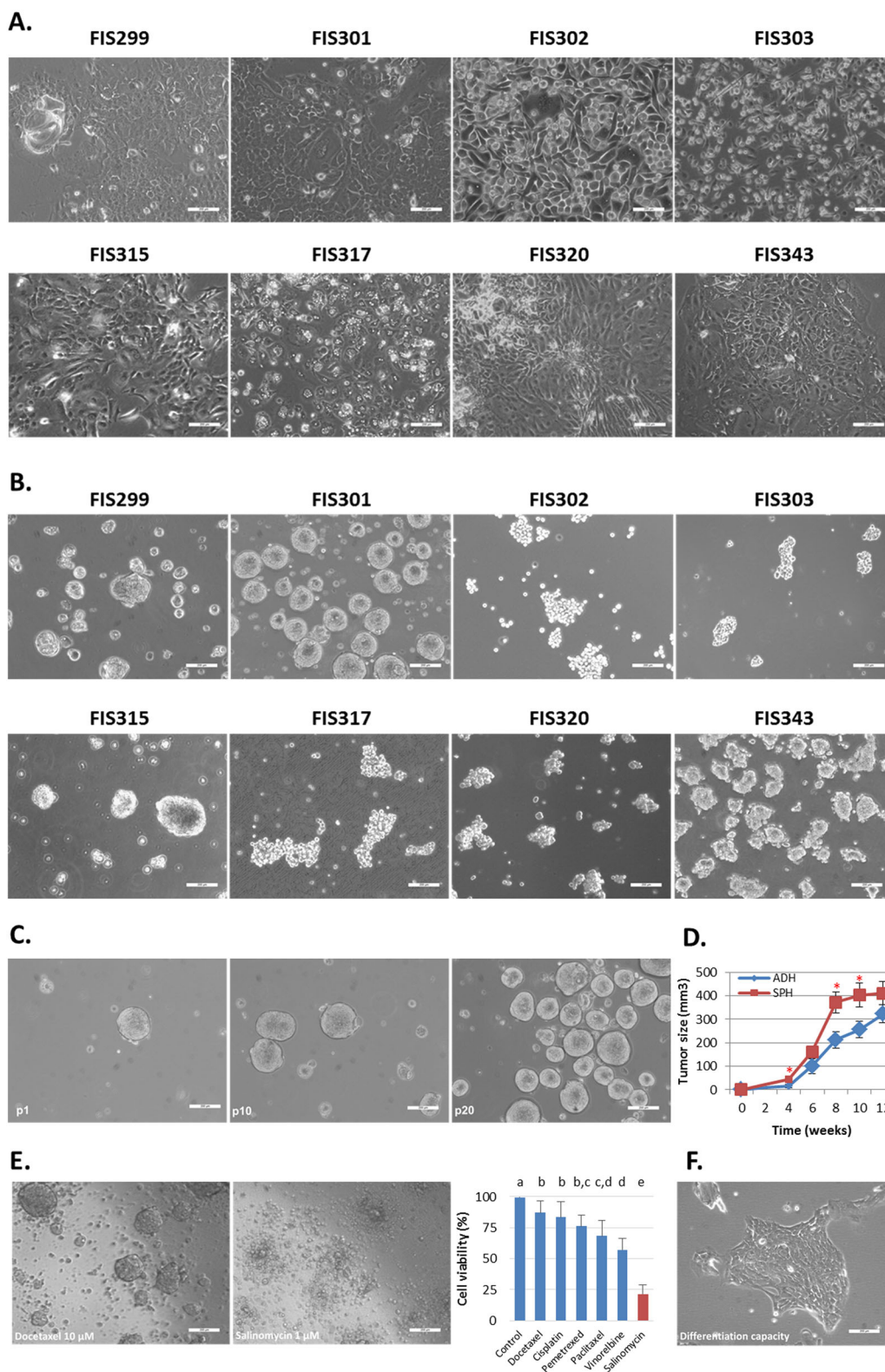


Fig. 1 (See legend on next page.)

(see figure on previous page)

Fig. 1 Results for the primary cultures establishment and characterization. **a** Representative images of the primary patient-derived cancer cells grown under adherent conditions. **b** Representative images of cells from the same patients under suspension conditions. **c** Self-renewal and unlimited exponential growth potentials of suspension cultures. Tumorspheres displayed stable growth without declining in number. **d** Tumor development capacity of adherent cells (blue line) and tumorspheres (red line) in vivo. The graph shows tumor growth in mice after injection of tumorspheres and their adherent counterparts at the indicated time points. Error bars represent SEM. * $P < 0.01$. **e** Representative images of the cytotoxic effects of the exposure of FIS301 patient tumorspheres to chemotherapeutic agents and salinomycin. Bar chart represents the cell viability of tumorspheres from primary cultures after 48 h exposure to cisplatin 50 μM , docetaxel 10 μM , paclitaxel 10 μM , vinorelbine 10 μM , pemetrexed 50 μM , and salinomycin 1 μM . ^{a,b,c,d} Bars with different superscripts are statistically different ($p < 0.05$). **f** Differentiation capacity of lung tumorspheres. Under adherent conditions, tumorspheres adhere and acquire the morphologic features of cells directly established from tumor tissue

maintained the suspension cultures for more than 6 months. In all cases, lung tumorspheres exhibited stable unlimited exponential growth even in later passages (>30 passages) (Fig. 1c). To determine the invasive and tumorigenic capacities of tumorspheres, time-lapse video microscopy was performed, revealing that tumorspheres possess a high invasive capacity, being able to migrate through collagen matrix (Supplementary Video 1). We also evaluated the ability of tumorspheres and its corresponding adherent counterparts to develop tumors in vivo by subcutaneous transplantation of cells into immunocompromised mice. Both cells derived from tumorspheres and adherent cultures were able to initiate tumors in vivo, being the tumor latency higher in tumors induced by adherent cells (Fig. 1d). It is characteristic of CSCs to be highly resistant to conventional therapies as well. We investigated the cytotoxic activity of cisplatin, paclitaxel, vinorelbine, and pemetrexed at high doses. A selective agent against stem cells, salinomycin, was also tested at lower concentration. All chemotherapeutic drugs had a mild effect on tumorspheres from primary cultures after 48 h of exposure to antineoplastic agents, not reaching the half-maximal inhibitory concentration (Fig. 1e). Cisplatin, docetaxel, and pemetrexed displayed a modest cytotoxic effect with 83.5%, 86.9%, and 76.2% of cells alive after treatment, respectively. Paclitaxel and vinorelbine were more effective with 68.2% and 56.9% of cells alive after exposure, respectively. In contrast, salinomycin showed higher cytotoxic activity against tumorspheres with 21.7% of cells alive after 48 h exposure. Moreover, these drugs were more effective in cells cultured under adherent conditions (Supplementary Fig. S4). Finally, we assessed the differentiation potential of tumorspheres. Using serum-containing medium and conventional flask, we seeded them and found that tumorspheres were able to adhere and acquire the same morphology than their corresponding adherent-cultured cells (Fig. 1f). Moreover, gene expression profiles showed no differences between adherent cells directly established from tissue and those established from tumorspheres, confirming that tumorspheres are able to adhere and differentiate, losing the expression of stemness markers reported in the next section.

Lung tumorspheres overexpress genes related to stemness and invasion

The expression at mRNA level of 51 genes described as potential lung CSC markers, pluripotency and cell cycle regulators, invasion promoters, and components of Notch, Wnt, and Hedgehog signaling pathways was analyzed in tumorspheres and adherent cells from patient-derived cells and cell lines using RTqPCR. The relative expression levels of *LIN28B*, *CD133*, *WNT1*, *WNT2*, *SHH*, and *GLI1* were below the limit of detection of the technique in most samples and were excluded from the final analysis. Remarkably, the expression of *CD133* could not be detected using three different sets of gene expression assays. Tumorspheres showed higher expression of 37 out of 44 genes compared to adherent-cultured cells, being a group of 17 genes: *ALDH1A1*, *KLF4*, *NANOG*, *CD44*, *THY1*, *CDKN1A*, *JUNB*, *MDM2*, *MMP9*, *SNAI1*, *ITGA6*, *NOTCH1*, *NOTCH3*, *DLL4*, *JAG1*, *CTNNB1*, and *GSK3B*, significantly overexpressed according to Wilcoxon's signed-rank test (Fig. 2).

Unsupervised PCA including patient samples and cell lines was performed in order to group samples according to gene expression. PCA score plot revealed that the adherent-cultured cells population is more homogeneous than tumorspheres in terms of gene expression (Supplementary Fig. S5). Next, a supervised partial least square-discriminant analysis was applied to discriminate tumorspheres and adherent-cultured cells. As shown in Supplementary Fig. S6, principal component 1 (PC1) was able to separate most CSCs from adherent tumor cells. Loading plot revealed that the expression of *SNAI1*, *GSK3B*, *CD44*, *CDKN1A*, *NOTCH3*, *NANOG*, and *CTNNB1* genes in tumorspheres contributed the most to this separation. To analyze the differences between cell lines and patient-derived cultures in their gene expression profile, PCA analyses were applied separately to cell lines and patient-derived cultures. The PCA score plot from cell lines exhibited the high variability between them. PC1 separated 7 out of 12 tumorspheres cultures from adherent cells (Supplementary Fig. S7). On the contrary, in the PCA score plot corresponding to the patient-derived cultures, PC1 separated FIS343 and FIS320 from the rest of primary cultures, whereas PC2 clearly

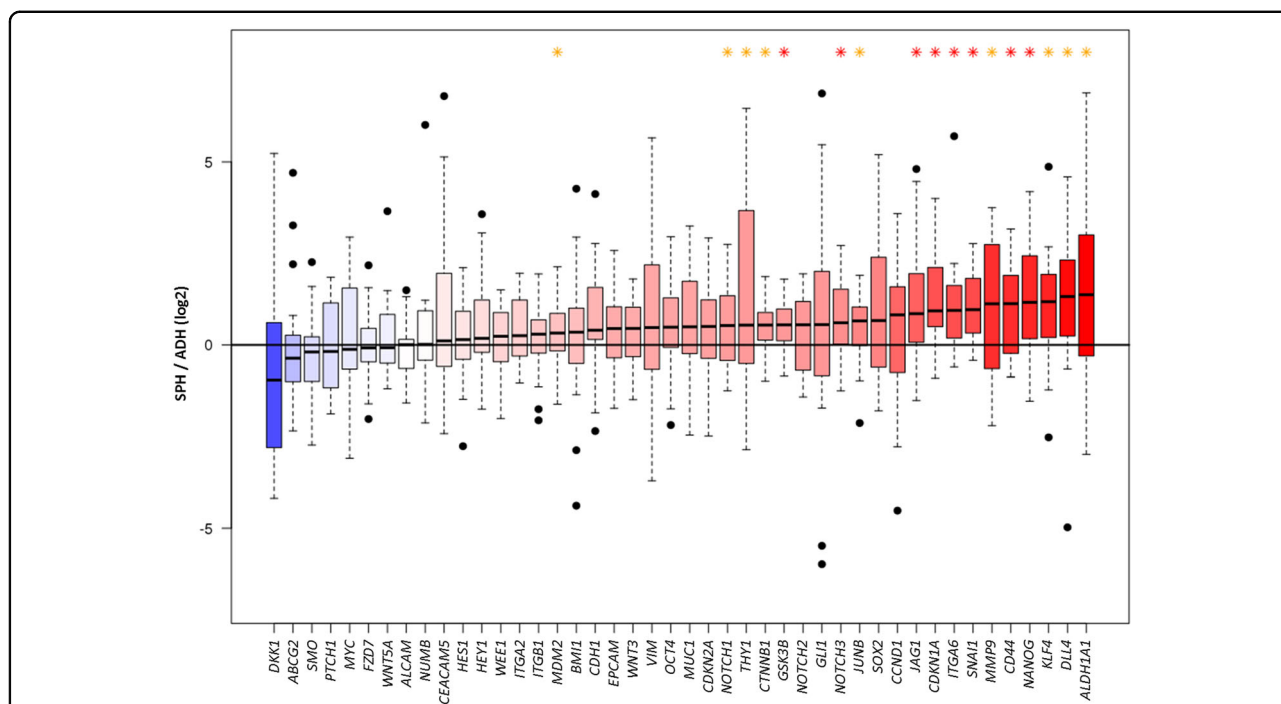


Fig. 2 Transcription levels of CSC-related genes in tumorspheres vs. adherent-cultured cells. mRNA was measured by RTqPCR. The results shown are the log₂ of the ratio between the gene expression of tumorspheres and the gene expression of adherent-cultured cells. Error bars represent SEM. Asterisks indicate $p < 0.05$ (yellow) and $p < 0.01$ (red)

distinguished adherent tumor cells from lung tumorspheres (Supplementary Fig. S8). When comparing the gene expression profiles, it was observed that expression of *BMI1*, *CD166*, *CDKN2A*, *MDM2*, *HEY1*, *NUMB*, *ITGA6*, and *NOTCH3* was induced in tumorspheres from FIS343 and FIS320 patients, while tumorspheres from the rest of patients showed higher expression of *EPCAM*, *NOTCH1*, *NOTCH2*, *CD44*, *CTNNB1*, *MMP9*, and *CDKN1A*. Thereafter, a logistic regression model was used to reduce dimensionality and remove collinear expression of genes. The model was built considering the 17 statistically significant overexpressed genes ($p < 0.05$, Wilcoxon’s signed-rank test), being *CDKN1A*, *NOTCH3*, *CD44*, *ITGA6*, *NANOG*, and *SNAIL* the genes selected for further analyses.

Lung adenocarcinoma tumorspheres overexpress p21, Notch3, CD44, integrin α6, Nanog, and Snail

Gene expression analyses were complemented with immunoblotting (IB) and immunofluorescence (IF) analyses of the proteins encoded by these six genes in primary cultures. All protein expressions were significantly higher in tumorspheres than in adherent cells in lung ADC patients according to IB and only one patient (FIS320) showed higher levels of integrin α6 and Snail in the adherent-cultured cells than in tumorspheres (Fig. 3a). Interestingly, IF showed differential subcellular

localization of p21, which was nuclear and cytoplasmic in cells forming lung tumorspheres and only nuclear in adherent tumor cells (Fig. 3b, secondary antibody control Supplementary Fig. S9). For Notch3, membrane localization in addition to cytoplasmic and nuclear was detected in both adherent cells and lung tumorspheres. In addition, all cells forming tumorspheres expressed both CD44 and Nanog, although signals showed polarity, being higher on cell membranes of cells located in the periphery of the lung tumorspheres for CD44 and nuclear for Nanog. In contrast, the expression of Nanog and CD44 was notably lower in adherent cells. Integrin α6 showed similar expression pattern to CD44 with higher expression in cytoplasm, especially in the cell membrane. Finally, Snail was overexpressed in tumorspheres and showed a non-uniform nuclear location along them. In adherent cells, a more homogeneous expression was detected in nuclei, which was weaker than that observed on their corresponding tumorspheres.

In contrast to ADC patients, variability was greater for SCC patients (Supplementary Fig. S10). No significant differences were found for p21, CD44 and integrin α6, whereas higher expression of Nanog, Snail, and Notch3 were detected in adherent cells in this histology, suggesting that different molecular changes govern CSCs in this tumor subtype. Complementary analyses of other potential CSC-related proteins (*CD133*, *CD166*,

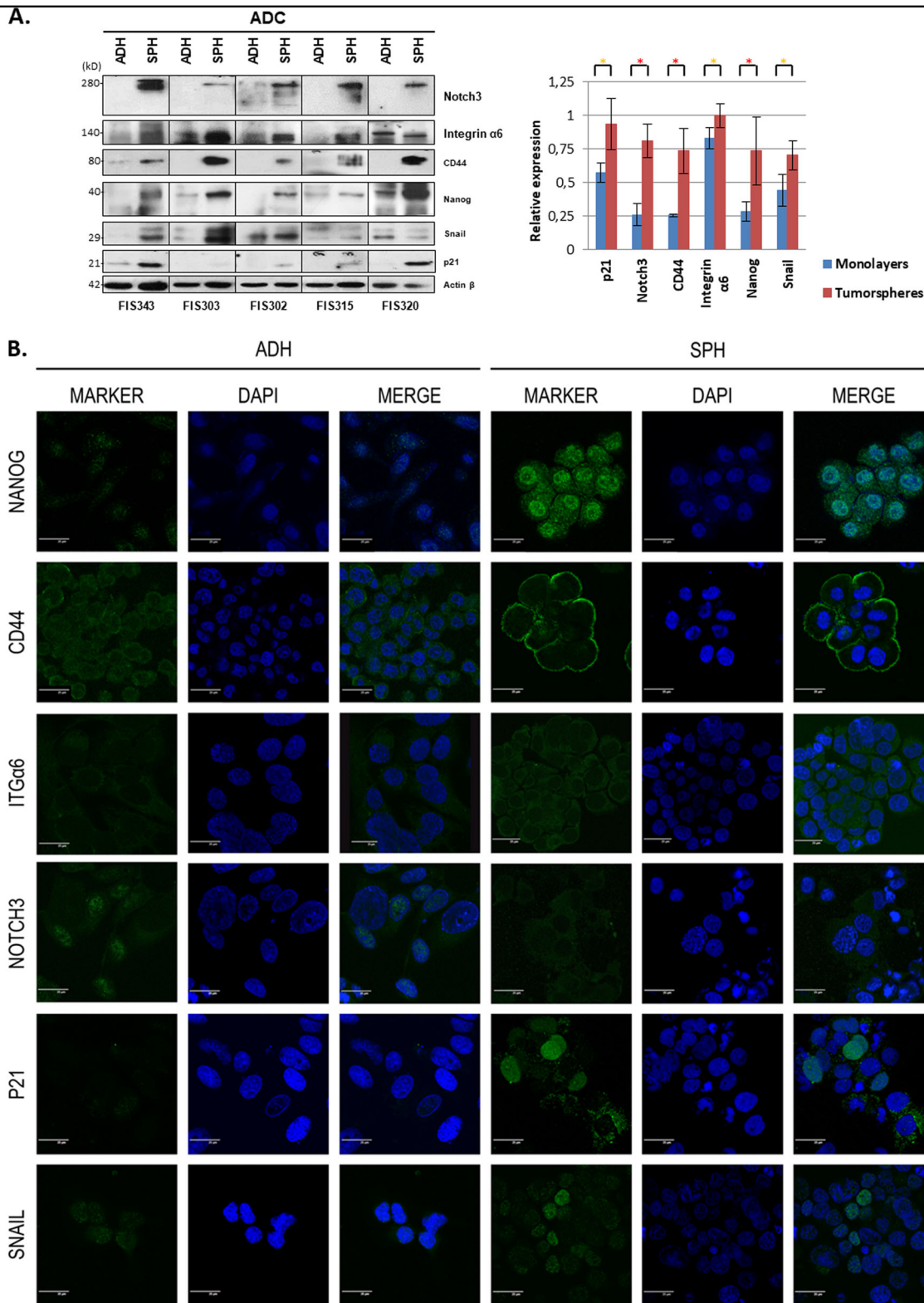


Fig. 3 Analysis of CSC-related proteins in primary cultures. Representative immunoblots (a) and immunofluorescence (b) images of Nanog, CD44, integrin α6, Notch3, p21, and Snail in adherent-cultured cells and tumorspheres from ADC patients. Bar chart represents the relative expression of each protein according to immunoblots. Asterisks indicate $p < 0.05$ (yellow) and $p < 0.01$ (red). β -Actin was used to assess equal loading in immunoblots. Green channel in immunofluorescence shows the indicated antibody staining, blue channel shows DAPI staining, and merge shows all channels merged. Scale bar represents 25 μ m

ALDH1A1, β -catenin, E-cadherin, and vimentin) and histological markers (carcinoembryonic antigen, cytokeratin 7, p63, cytokeratins 5/6, CD56) were performed and can be found in the Supplementary Figs. S11 and S12.

A CSCs score is a prognostic biomarker for OS in NSCLC

Data from TCGA for ADC and SCC patients were used to associate genes significantly overexpressed in tumorspheres with survival. Clinicopathological characteristics of these patients are summarized in Table 2 (in silico set). Patients with post-surgical complications were excluded from the survival analysis, and only those patients who had at least 1 month of follow-up were included ($N=661$). Cox regression and Kaplan–Meier analyses indicated that patients with high levels of *THY1*, *SNAI1*, *ITGA6*, and *CDKN1A* presented worse OS (Supplementary Table S4 and Supplementary Fig. S13). Survival analyses were also performed according to patient histology, associating high *ITGA6* and *JAG1* with worse prognosis in ADC patients. No other significant associations were found between survival and clinicopathological variables or gene expression.

Thereafter, we intended to create a gene expression score that can provide more accurate predictions for patients' prognostic^{27,28}. Univariate Cox regression analysis was performed considering OS as a dependent variable. Genes were ordered on the basis of their prognostic power (univariate Z-score, Supplementary Fig. S14), and according to this ranking, the expression of *CDKN1A*, *SNAI1*, and *ITGA6* were found to be associated with survival (Z-score >1.5), and therefore were selected to create a risk signature. We constructed a model based on the relative contribution of these three genes in the multivariate analysis (considering absolute regression coefficients, see Supplementary Table S5), and the resulting score was named CSCs score, with the following equation: $(CDKN1A \times 0.123) + (ITGA6 \times 0.196) + (SNAI1 \times 0.255)$. Kaplan–Meier analysis showed that patients with high CSC score (>median) had shorter OS (37.7 vs. 60.4 mo, $p = 0.001$; Fig. 4a). We also performed a stratified analysis by histology and found a similar association between high CSC score and prognosis for ADC patients (36.6 vs. 53.5 mo, $p = 0.003$; Fig. 4b). To evaluate the potential use of the CSC score as an independent prognostic biomarker, a multivariate analysis was performed including all the significant variables from the univariate analyses (age, tumor node metastasis (TNM) staging, tumor size, lymph node involvement, *CDKN1A*, *ITGA6*, *SNAI1*, and the CSC score). Results obtained from this multivariate analysis indicated that age, TNM staging, and the CSC score in the entire cohort and TNM staging, lymph node involvement, and the CSC score in the ADC cohort were independently associated with survival (see Table 3).

Table 2 Clinicopathological characteristics of the patients included in the study

	In silico set		Validation set			
	Global cohort	ADC cohort	ADC cohort			
	<i>N</i> = 661	%	<i>N</i> = 345	%	<i>N</i> = 114	%
Age at surgery (median, range)	68 [38–88]		67 [38–88]		65 [37–84]	
Gender						
Male	395	59.8	165	47.8	77	67.5
Female	266	40.2	180	52.2	37	32.5
Stage						
I	375	56.7	197	57.1	73	64.0
II	179	27.1	90	26.1	26	22.8
IIIA	107	16.2	58	16.8	15	13.2
Histology						
ADC	345	52.2	345	100.0	114	100.0
SCC	316	47.8	0	0.0	0	0.0
Others	0	0	0	0	0	0.0
Performance status						
0	NS		NS		79	69.3
1					35	30.7
Differentiation grade						
Poor	NS		NS		17	14.9
Moderate					38	33.3
Well					27	23.7
NS					32	28.1
Smoking status						
Current	165	25.0	83	24.1	52	45.6
Former	382	57.8	179	51.9	39	34.2
Never	114	17.2	83	24.1	23	20.2
Exitus						
No	400	60.5	231	67.0	65	57.0
Yes	261	39.5	114	33.0	49	43.0

ADC adenocarcinoma, NS non-specified

An independent cohort of patients with resected lung ADC was used for validation of the CSC signature. Clinicopathological characteristics of these patients are summarized in Table 2 (validation set). Cox regression and Kaplan–Meier analyses of individual genes indicated that patients with high expression levels of *SNAI1* and *ITGA6* presented worse OS (Supplementary Table S4 and Supplementary Fig. S13). In addition, the association between high CSC score and worse prognosis was confirmed in

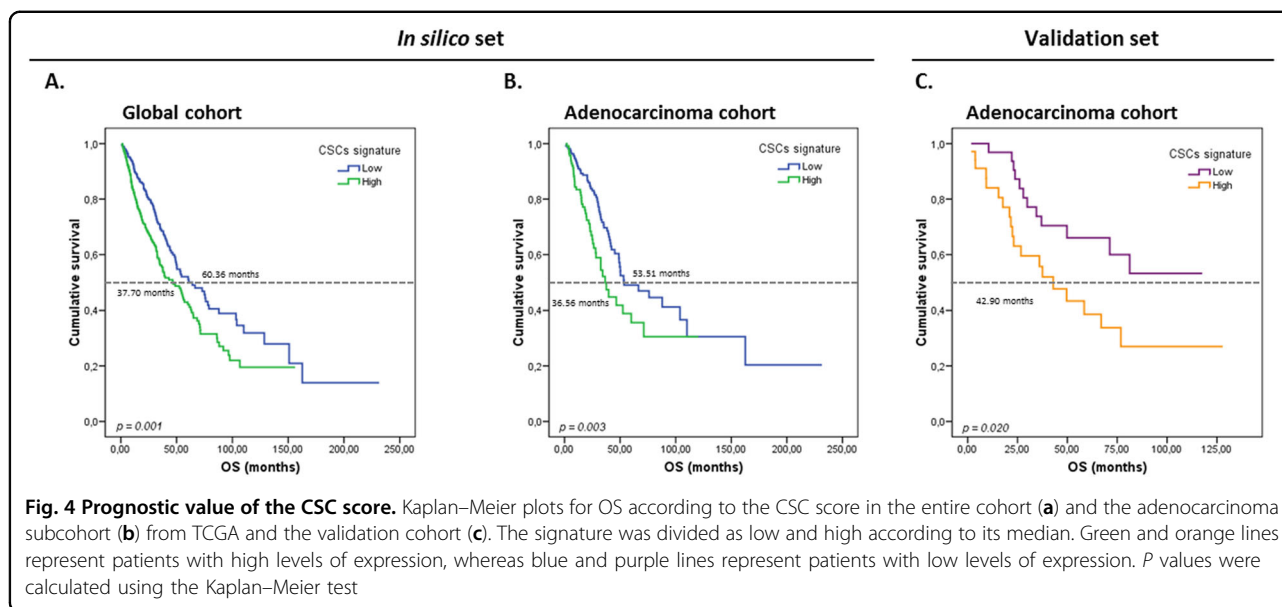


Table 3 Results from the multivariate Cox regression model for OS

	Global cohort (N = 661)			ADC cohort (N = 345)		
	HR	95% CI	P value	HR	95% CI	P value
Age (>65 vs. ≤65)	1.398	1.069–1.827	0.014	–	–	–
TNM staging (IIIa vs. II vs. I)	1.353	1.161–1.578	0.0001	1.515	1.097–2.092	0.012
LN involvement (yes vs. no)	–	–	–	2.108	1.453–3.059	<0.0001
CSC signature (high vs. low)	1.498	1.167–1.922	0.001	1.869	1.275–2.738	0.001

ADC adenocarcinoma, LN involvement, lymph node involvement, OS overall survival, HR hazard ratio, CI confidence interval

this cohort (42.90 vs. not reached (NR) mo, *p* = 0.020; Fig. 4c).

Discussion

Cancers exist in an extraordinary variety of types and subtypes, making each cancer individually unique. Tumors are heterogeneous and many cancer cell populations with different features are present. Among these tumor populations, cells with stemness properties, commonly called CSCs, have been described and associated to more aggressive phenotypes. There is strong evidence suggesting that cancer cells with stem properties selectively resist current cancer therapies, indicating the important role that CSCs play in tumor evolution, relapse, and metastasis^{8,9}. In this context, evaluation of their gene expression profiles could provide the basis for identifying biomarkers and therapeutic targets, which could improve patients’ outcomes. Sphere-forming assays are well-described culture methods that have been used for stem cells isolation, identification, and enrichment from different tissues^{10–14}. Starting material for these cultures can

also be commercial cell lines, but using cells directly isolated from surgical resections specimens reflect in vivo conditions better than they do, since immortalized cell lines do not behave as primary cultures and long-term manipulation alters phenotype, functions, and responsiveness to stimuli. In addition, clinicopathological information from primary cultures can be correlated with each culture behavior, whereas this information is limited for most cell lines. As a result, although the establishment of primary cultures can be problematic and time-consuming, experiments with this type of approach are required to strengthen the findings, especially when studying stemness properties. In addition, the expression of some surface markers has been proposed as characteristic of lung CSCs, although contradictory data can be found between different studies, even when using the same cell lines^{15–18}.

For these reasons, in this study we used unsorted primary tumor cells to form tumorspheres without selecting them according to the expression of a single marker or combination that could misrepresent the cell population with stem properties. Under these conditions, the success

rate in the establishment of primary cultures from lung cancer patients was 40%, which is in consonance with results previously reported^{12,29}. Although *EGFR*, *ALK*, and *ROS1* are common driver mutations in lung adenocarcinoma with effective targeted therapies approved for use^{30,31}, we found no correlations between their mutational status, *KRAS* or *TP53* mutational status, or other clinicopathological variables and the establishment of primary cultures.

To ensure that our lung tumorspheres were enriched in cells with stem-like properties, multiple analyses were performed to determine if they have the fundamental features of this population. Patient-derived tumorspheres had self-renewal and unlimited exponential growth abilities, higher tumorigenic potential *in vivo* than adherent control cells, were able to differentiate and acquire the properties of their adherent counterparts, showed high invasion capacity, and were very resistant to high doses of chemotherapeutic agents. These results are in line with those previously reported on CSCs isolated from lung cancer^{16–18} and confirm that this approach can be used for their enrichment in a simple and cost-effective way. In addition to all these properties, CSCs share with other stem populations the overexpression of stemness pathways, cytoprotective enzymes, and efflux pumps^{32–34}. Different genes and molecules have been proposed as characteristic of lung CSCs, but most studies are focused on few cell lines and analyzed the expression of small groups of genes, being very challenging to determine which ones are characteristic of a particular cell line or patient and which ones are governing stemness in lung CSCs^{33–37}.

In this study, lung tumorspheres exhibited increased expression of genes encoding for cytoprotective enzymes (*ALDH1A1*), pluripotency inducers (*KLF4*, *NANOG*), cell cycle regulators (*CDKN1A*, *JUNB*, *MDM2*), metastasis-related genes (*CD44*, *THY1*, *MMP9*, *SNAI1*, *ITGA6*), and components of Notch (*NOTCH1*, *NOTCH3*, *DLL4*, *JAG1*), and Wnt (*CTNNB1*, *GSK3B*) pathways. To reduce the dimensionality of our data, a mathematical algorithm was used, selecting *CDKN1A*, *NOTCH3*, *CD44*, *ITGA6*, *NANOG*, and *SNAI1* to distinguish tumorspheres from adherent tumor cells. Protein analyses of the selected genes confirmed gene expression results and showed that proteins encoded by these genes are overexpressed in tumorspheres from ADC patients and in some cases differentially located along the cells. These molecules are widely related to features observed in CSCs and could constitute potential targets. For instance, cytoplasmic p21 (encoded by *CDKN1A*) and Notch3 were associated with self-renewal, tumorigenic behavior, and aggressive tumors^{38–41}. Additionally, Snail has been reported to regulate *Nanog*, inducing stemness properties in lung cancer and, along with integrin $\alpha 6$, Snail, and *CD44*, has

been associated with cell migration, invasion, and metastasis^{42–46}. In contrast, variability was greater for SCC patients, for whom no significant results were found for p21, *CD44*, and integrin $\alpha 6$, whereas the expression of Notch3, *Nanog*, and Snail seemed higher in adherent cells than in tumorspheres. Differences in the role of these molecules have been previously reported according to tumor histology^{47,48}, highlighting that there are significant changes between NSCLC subtypes and that markers to identify CSCs are not only tissue-dependent but also histological. In this study, only one of the 12 cell lines and 3 out of the 8 primary cultures were lung SCC, so a greater number of cases could be required to properly identify markers and targets for this tumor subtype. Complementary, the protein expression of other molecules was analyzed, finding differential expression of *CD133*, *CD166*, *ALDH1A1*, β -catenin, E-cadherin, and vimentin in patients' tumorspheres. Our group previously reported the association of some of these molecules to CSCs, being their expression associated to the response to compounds targeting cell stemness⁴⁹.

Many studies have tried to correlate the expression of genes associated with CSCs to patients' prognosis^{50,51}. Nevertheless, most of them are focused on single pathway-specific markers with limited prognostic value. Finding gene expression signatures that identify altered pathways in carcinogenesis could lead to the discovery of molecular subclasses and predict patients' outcomes better^{52,53}. TCGA data of the genes significantly overexpressed in tumorspheres was ranked according to their prognostic power, trying to find a gene signature, which could provide valuable prognostic information^{24,25}. We created a score combining the expression of *CDKN1A*, *ITGA6*, and *SNAI1*, which was an independent prognostic biomarker for lung cancer patients. To validate it, the expression of these genes was evaluated in an independent cohort of resected lung ADC patients, finding that patients with elevated CSC score had shorter OS. These results are of great importance because current clinicopathological staging methods have limited success in predicting patient survival and today we still cannot predict which patients will be cured, and which ones will relapse after surgery. Gene expression scores based on RTqPCR have demonstrated being useful for classifying tumors and predicting prognosis, being even approved as prognostic tools in clinical practice⁵⁴. This technology is a well-implemented methodology in our group for biomarkers' research, previously reporting angiogenesis and immune checkpoint scores for NSCLC^{55,56}. The CSCs score proposed can help in future clinical practice, since high scores may reflect a bigger CSC population with enhanced cell migration, invasion, and tumor initiation capacity that will be able to modulate cell cycle and persist after cancer treatments. As a result, patients with high

values on this score may need adjuvant treatment and should be closely followed after a successful surgery, because they have a higher risk to die. The development of targeted therapies against this tumor population is essential to prevent relapse of patients and improve their future outcome.

Our approach allows the establishment of long-term lung CSCs cultures, being a powerful tool for identifying the molecular alterations present in this tumor cell population and providing new insight into the field of CSCs in NSCLC. Tumorspheres can be used for CSC enrichment and a common set of genes were found to be significantly and consistently overexpressed in them. Developing new therapeutic strategies against these molecules could have major implications in patients' survival. In addition, a gene score based on molecules overexpressed in this tumor cell population predicts worse outcome in two independent cohorts of patients, representing an independent prognostic biomarker that can be used to determine the outcome in resectable lung ADC patients.

Acknowledgements

This work was supported by Fondo de Investigación Sanitaria, ISCIII (PI12-02838, PI12-00956, PI15-00753, PI15-00209), and Fundación Arnal Planelles. A. H.-P. and J.M.P.-S. are supported by Ministerio de Educación, Cultura y Deporte (FPU14/06911 and FPU13/02755, respectively).

Author details

¹Molecular Oncology Laboratory, Fundación Hospital General Universitario de Valencia, Valencia, Spain. ²CIBERONC, Valencia, Spain. ³Oncogenic Signalling Laboratory, Centro de Investigación Príncipe Felipe, Valencia, Spain.

⁴Department of Pathology, Universitat de València, Valencia, Spain.

⁵Department of History of Science and Documentation, Universitat de València, Valencia, Spain.

⁶Cytomic Core Facility, Centro de Investigación Príncipe Felipe, Valencia, Spain.

⁷Program of Predictive and Personalized Medicine of Cancer, Institut de Reserca Germans Trias i Pujol (PMPPC-IGTP), Badalona, Spain.

⁸Department of Medical Oncology, Hospital General Universitario de Valencia, Valencia, Spain.

⁹Department of Thoracic Surgery, Hospital General Universitario de Valencia, Valencia, Spain.

¹⁰Department of Pathology, Hospital General Universitario de Valencia, Valencia, Spain.

¹¹Institute of Sanitary Research of Asturias, Hospital Central de Asturias, Universidad de Oviedo, Oviedo, Spain.

¹²Instituto Valenciano de Patología, Unidad Mixta de Patología Molecular, Centro Investigación Príncipe Felipe/ Universidad Católica de Valencia, Valencia, Spain.

¹³Department of Biotechnology, Universidad Politécnica de Valencia, Valencia, Spain.

¹⁴Department of Medicine, Hospital General Universitario de Valencia, Valencia, Spain

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41419-019-1898-1>).

Received: 15 May 2019 Revised: 22 July 2019 Accepted: 26 August 2019

Published online: 10 September 2019

References

- Torre, L. A. et al. Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108 (2015).
- Reck, M. & Rabe, K. F. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N. Engl. J. Med.* **377**, 849–861 (2017).
- Hirsch, F. R., Suda, K., Wiens, J. & Bunn, P. A. J. New and emerging targeted treatments in advanced non-small-cell lung cancer. *Lancet* **388**, 1012–1024 (2016).
- Rizvi, N. A. & Peters, S. Immunotherapy for unresectable stage III non-small-cell lung cancer. *N. Engl. J. Med.* **377**, 1986–1988 (2017).
- Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell* **168**, 707–723 (2017).
- Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).
- Zakaria, N. et al. Human non-small cell lung cancer expresses putative cancer stem cell markers and exhibits the transcriptomic profile of multipotent cells. *BMC Cancer* **15**, 84 (2015).
- Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nat. Rev. Cancer* **5**, 275–284 (2005).
- Jordan, C. T., Guzman, M. L. & Noble, M. Cancer stem cells. *N. Engl. J. Med.* **355**, 1253–1261 (2006).
- Li, C. et al. Identification of pancreatic cancer stem cells. *Cancer Res.* **67**, 1030–1037 (2007).
- Singh, S. K. et al. Identification of a cancer stem cell in human brain tumors. *Cancer Res.* **63**, 5821–5828 (2003).
- Eramo, A. et al. Identification and expansion of the tumorigenic lung cancer stem cell population. *Cell Death Differ.* **15**, 504–514 (2008).
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. USA* **100**, 3983–3988 (2003).
- Ricci-Vitiani, L. et al. Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
- Bertolini, G. et al. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. *Proc. Natl. Acad. Sci. USA* **106**, 16281–16286 (2009).
- Qiu, X. et al. Characterization of sphere-forming cells with stem-like properties from the small cell lung cancer cell line H446. *Cancer Lett.* **323**, 161–170 (2012).
- Meng, X., Li, M., Wang, X., Wang, Y. & Ma, D. Both CD133+ and CD133– subpopulations of A549 and H446 cells contain cancer-initiating cells. *Cancer Sci.* **100**, 1040–1046 (2009).
- Roudi, R., Zahra, M., Marzieh, E., Fazel, S. & Ali, S. CD44 and CD24 cannot act as cancer stem cell markers in human lung adenocarcinoma cell line A549. *Cell Mol. Biol. Lett.* **19**, 23–36 (2014).
- Bartucci, M. et al. Therapeutic targeting of Chk1 in NSCLC stem cells during chemotherapy. *Cell Death Differ.* **19**, 768–778 (2012).
- Zeuner, A. et al. Elimination of quiescent/slow-proliferating cancer stem cells by Bcl-XL inhibition in non-small cell lung cancer. *Cell Death Differ.* **21**, 1877–1888 (2014).
- Tiran, V. et al. Primary patient-derived lung adenocarcinoma cell culture challenges the association of cancer stem cells with epithelial-to-mesenchymal transition. *Sci. Rep.* **7**, 10040 (2017).
- Bertolini, G. et al. Microenvironment-modulated metastatic CD133+/CXCR4+/EPCAM– lung cancer-initiating cells sustain tumor dissemination and correlate with poor prognosis. *Cancer Res.* **75**, 3636–3649 (2015).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45 (2001).
- Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
- Zhang, J. et al. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford)* <https://doi.org/10.1093/database/bar026> (2011).
- Lossos, I. S. et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* **350**, 1828–1837 (2004).
- Schetter, A. J. et al. Association of inflammation-related and microRNA gene expression with cancer-specific mortality of colon adenocarcinoma. *Clin. Cancer Res.* **15**, 5878–5887 (2009).
- Wang, P. et al. Identification and characterization of cells with cancer stem cell properties in human primary lung cancer cell lines. *PLoS ONE* **8**, e57020 (2013).

30. Dogan, S. et al. Molecular epidemiology of *EGFR* and *KRAS* mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related *KRAS*-mutant cancers. *Clin. Cancer Res.* **18**, 6169–6177 (2012).
31. Soda, M. et al. A prospective PCR-based screening for the *EML4-ALK* oncogene in non-small cell lung cancer. *Clin. Cancer Res.* **18**, 5682–5689 (2012).
32. Nunes, T. et al. Targeting cancer stem cells to overcome chemoresistance. *Int. J. Mol. Sci.* **19**, 4036 (2018).
33. Takebe, N. et al. Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat. Rev. Clin. Oncol.* **12**, 445–464 (2015).
34. Sullivan, J. et al. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on Notch signaling. *Cancer Res.* **70**, 9937–9948 (2010).
35. Zhang, W. C. et al. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* **148**, 259–272 (2012).
36. Medema, J. P. Cancer stem cells: the challenges ahead. *Nat. Cell Biol.* **15**, 338–344 (2013).
37. Yan, X. et al. Identification of CD90 as a marker for lung cancer stem cells in A549 and H446 cell lines. *Oncol. Rep.* **30**, 2733–2740 (2013).
38. Insinga, A. et al. DNA damage in stem cells activates p21, inhibits p53, and induces symmetric self-renewing divisions. *Proc. Natl. Acad. Sci. USA* **110**, 3931–3936 (2013).
39. Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer* **9**, 400–414 (2009).
40. Ali, S. A., Justilien, V., Jamieson, L., Murray, N. R. & Fields, A. P. Protein kinase Ciota Drives a NOTCH3-dependent stem-like phenotype in mutant *KRAS* lung adenocarcinoma. *Cancer Cell* **29**, 367–378 (2016).
41. Chen, C. Y. et al. Expression of Notch gene and its impact on survival of patients with resectable non-small cell lung cancer. *J. Cancer* **8**, 1292–1300 (2017).
42. Stewart, R. L. et al. Elevated integrin $\alpha 6 \beta 4$ expression is associated with venous invasion and decreased overall survival in non-small cell lung cancer. *Hum. Pathol.* **54**, 174–183 (2016).
43. Wang, H. et al. Acquisition of epithelial-mesenchymal transition phenotype and cancer stem cell-like properties in cisplatin-resistant lung cancer cells through *AKT*/ β -catenin/Snail signaling pathway. *Eur. J. Pharmacol.* **723**, 156–166 (2014).
44. Liu, C. W. et al. Snail regulates Nanog status during the epithelial-mesenchymal transition via the *Smad1/Akt/GSK3 β* signaling pathway in non-small-cell lung cancer. *Oncotarget* **5**, 3880–3894 (2014).
45. Schmitt, M., Metzger, M., Gradl, D., Davidson, G. & Orian-Rousseau, V. CD44 functions in Wnt signaling by regulating LRP6 localization and activation. *Cell Death Differ.* **22**, 677–689 (2015).
46. Yanagawa, J. et al. Snail promotes CXCR2 ligand-dependent tumor progression in non-small cell lung carcinoma. *Clin. Cancer Res.* **15**, 6820–6829 (2009).
47. Chiou, S. H. et al. Coexpression of Oct4 and Nanog enhances malignancy in lung adenocarcinoma by inducing cancer stem cell-like properties and epithelial-mesenchymal transdifferentiation. *Cancer Res.* **70**, 10433–10444 (2010).
48. Park, E. et al. Prognostic significance of stem cell-related marker expression and its correlation with histologic subtypes in lung adenocarcinoma. *Oncotarget* **7**, 42502–42512 (2016).
49. Soto-Cerrato, V. et al. Facilitated anion transport induces hyperpolarization of the cell membrane that triggers differentiation and cell death in cancer stem cells. *J. Am. Chem. Soc.* **137**, 15892–15898 (2015).
50. Zhao, S., Qiu, Z., Jin, Y., Zhang, L. & Li, W. Prognostic value of Thy1 in non-small cell lung cancer: a RNA-Seq transcriptome analysis. *Int. J. Clin. Exp. Pathol.* **9**, 6999–7009 (2016).
51. Chang, W. H. et al. *JAG1* is associated with poor survival through inducing metastasis in lung cancer. *PLoS ONE* **11**, e0150355 (2016).
52. Huang, E. et al. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**, 226–230 (2003).
53. Raponi, M. et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res.* **66**, 7466–7472 (2006).
54. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
55. Sanmartin, E. et al. A gene signature combining the tissue expression of three angiogenic factors is a prognostic marker in early-stage non-small cell lung cancer. *Ann. Surg. Oncol.* **21**, 612–620 (2014).
56. Usó, M. et al. Analysis of the prognostic role of an immune checkpoint score in resected non-small cell lung cancer patients. *Oncoimmunology* **6**, e1260214 (2017).

Passenger mutations in cancer evolution

Aparisi F^{1#}, Amado-Labrador H^{2#}, Calabuig-Fariñas S^{2,3,5,7*}, Torres S^{2,5,7}, Herreros-Pomares A^{2,5}, Jantus-Lewintre E^{2,4,5,7}, Blasco A^{1,5,7}, Iranzo V^{1,5,7} and Camps C^{1,5,6,7}

¹Medical Oncology Department, Hospital General Universitario de Valencia, Valencia, Spain

²Molecular Oncology Laboratory, Fundación Investigación Hospital General Universitario de Valencia, Valencia, Spain

³Department of Pathology, Universitat de València, Valencia, Spain

⁴Biotechnology Department, Universitat Politècnica de València, Valencia, Spain

⁵CIBERONC, Madrid, Spain

⁶Medicine Department, Universidad de Valencia, Valencia, Spain

⁷Unidad Mixta TRIAL CIPF- FIHGUV, Valencia, Spain

[#]Equal contribution

Abstract

A driver mutation is an alteration that gives a cancer cell a fundamental growth advantage for its neoplastic transformation. It differs from passenger mutations in that these do not necessarily determine the development of the cancer. Genomic instability and high mutation rates cause cancer to acquire numerous mutations and chromosomal alterations during its somatic evolution; most are termed passengers because they do not confer cancer phenotypes. Studies suggest that mildly deleterious passengers accumulate and can collectively slow cancer progression. Clinical data also suggest an association between passenger load and response to therapeutics, yet no causal link between the effects of passengers and cancer progression has been established. Although in the biology of cancer, driver mutations have been given more importance, the new evidence shows that passenger mutations are more important because they impact areas such as epigenetics, in mitochondrial DNA, immunogenicity or in the response to chemotherapy. We present an extensive review of the scientific literature on the role of passenger mutations in the evolution of cancer.

Introduction

Tumorigenesis is the result of the accumulation of genomic alterations and is driven by somatic evolution: alterations that occur because of defects in the regulatory circuits governing normal cell proliferation and homeostasis [1,2]. There are many kinds of cancer and tumor subtypes, in every location in the body and this complexity means that many questions about tumorigenic processes remain to be answered. For instance, many distinct regulatory circuits within each type of target cell must be disrupted for them to become cancerous [3]. However, even though carcinogenesis is very complex, genomic instability (i.e., a high frequency of genetic, epigenetic, and chromosomal alterations, collectively referred to as ‘mutations’) is a hallmark of this process [4].

Over the past decade, next generation sequencing (NGS) has allowed the integration of cancer genomics into clinical care. This has facilitated several major mass-sequencing genome projects—such as the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA)—for almost every key type of cancer and has identified tens of thousands of tumor mutations (including lung, kidney, and breast cancer mutations) [5]. In addition, studies conceptualizing the clonal hierarchy and phylogeny of cancer have demonstrated that intra-tumoral mutational and chromosomal heterogeneity is high [6-8]. Within this context, NGS technology has allowed the scientific community to characterize the molecular classification of cancer to define mutations as ‘drivers’ or ‘passengers’, depending on their proliferative and invasive capacity [9], also providing evidence that genomic instability is the cornerstone of cancer initiation.

Driver mutations are usually defined as mutations that induce cell proliferation and tumour growth, while passenger or ‘hitchhiker’ mutations, which represent approximately 97% of all cancerous mutations do not [10]. However, the role of passenger mutations has recently become more controversial, with some authors describing them as ‘mini drivers’ [11], also referred to as latent drivers of neutral mutations [12]. This contrasts with the prevailing hypothesis that the accumulation of passenger mutations is detrimental to cancer by slowing tumor growth and reducing metastatic progression [13,14]. In this review we discuss recent evidence in support of this latter hypothesis, the main arguments against the importance of passenger mutations, and the clinical consequences of tumor evolution based on their mutational rate. We argue that current approaches should be applied in new targeted cancer therapies.

Genomic instability as a cornerstone of cancer

The ability to detect genomic variations in cancer by genome, exome, and transcriptome sequencing analysis has led to the increased use of these technologies in large-scale molecular characterization projects such as the ICGC [5] and TCGA [15], facilitating the discovery

*Correspondence to: Silvia Calabuig Fariñas, 2Molecular Oncology Laboratory, Fundación Investigación Hospital General Universitario de Valencia, Valencia, Spain, E-mail: calabuix_sil@gva.es

Key words: passenger, mutation, driver, cancer

Received: May 07, 2019; **Accepted:** May 30, 2019; **Published:** June 10, 2019

of oncogenic drivers and candidate drug targets. The genomic characterization of cancer, progress in the understanding of cancer biology, and new ways of identifying its etiology of cancer, stratifying patients, managing the disease, and monitoring its responses have been especially beneficial.

Genomic instability is considered the cornerstone of the molecular classification of cancer; the acquisition of higher mutational rates caused by inducing genomic instability leads to the accelerated accumulation of 'adaptive drivers'. In turn, this causes an increased passenger load that can cancel out the effects of these drivers, thus modulating tumorigenesis and tumor progression [16]. Chromosomal instability, the occurrence of a high rate of chromosome structural alterations in tumor cells, is the most common type of genomic instability. Another form is characterized by an increased nucleotide mutation rate; microsatellite instability is a special case of this type of genomic instability and is characterized by the expansion or contraction of the oligonucleotide repetitions present in microsatellite sequences [17,18].

Current molecular cancer classifications divides detected mutations into driver and passenger mutations. Within this paradigm, driver mutations confer a growth advantage to cancer cells and are positively selected for in the cancer-tissue microenvironment and are therefore causally involved in oncogenesis. Conversely, passenger mutations do not confer the advantage of clonal growth and therefore, do not contribute to the development of cancer. A driver mutation is not required for the maintenance of a cancer, but must have been present at some point during the cancer's evolution. Passenger mutations are present in cancer genomes because they often occur during somatic cell division and have no functional consequences. Therefore, any cell that acquires a cancerous driver mutation already contains biologically-inert somatic passenger mutations in its genome which, through clonal expansion, will be duplicated in every subsequent daughter cancer cell [19]. Genomics-driven discovery of novel driver mutations and the molecular classification of cancer have accelerated the design of rational strategies for cancer prevention,

patient stratification, the development of new drugs, and treatment options in clinical settings, thereby establishing the concept of precision medicine in cancer.

Passenger versus driver mutations

Only a small fraction of the total mutations present in a tumor are thought to be driver mutations. Tumors typically contain 40-100 gene-coding alterations, including 5-15 driver mutations [20-22], some of which may be important for tumor initiation (e.g., APC in colon cancer) [23], while others could play a role in tumor growth (e.g., VEGF) or metastasis (e.g., TWIST1) [24]. In addition, it is important to note that there is a fundamental difference between a driver gene and a driver gene mutation. A driver gene produces driver mutations but may also produce passenger mutations. For example, APC is a large driver gene, but only mutations that truncate the protein encoded in the 1,600 amino acids in its N-terminal are driver-gene mutations; missense mutations throughout the gene, as well as protein-truncating mutations in the 1,200 amino acids in its C-terminal, are normally passenger mutations [13] (Figure 1).

Because driver events are so important for cancer progression, the primary goal of cancer sequencing is usually their discovery throughout the genome [22]. Therefore, most research aims to isolate and analyse driver mutations, although for most types of cancer, these alterations in the early stages of tumorigenesis are poorly understood. Conversely, little attention has been paid to passenger mutations, which constitute the vast majority of the somatic alterations present in cancer. Evolutionary and genomic simulation studies in cancer suggest that passenger mutations accumulate and, collectively, can decrease cancer progression [13,14,16]. Clinical data also imply an association between therapeutic responses and passenger mutation loads. Furthermore, the antitumor effects of chemotherapy may be because this treatment induces genomic instability and increases the passenger mutation load.

The number of somatic passenger mutations accumulated in a tumor can provide valuable information about its evolutionary

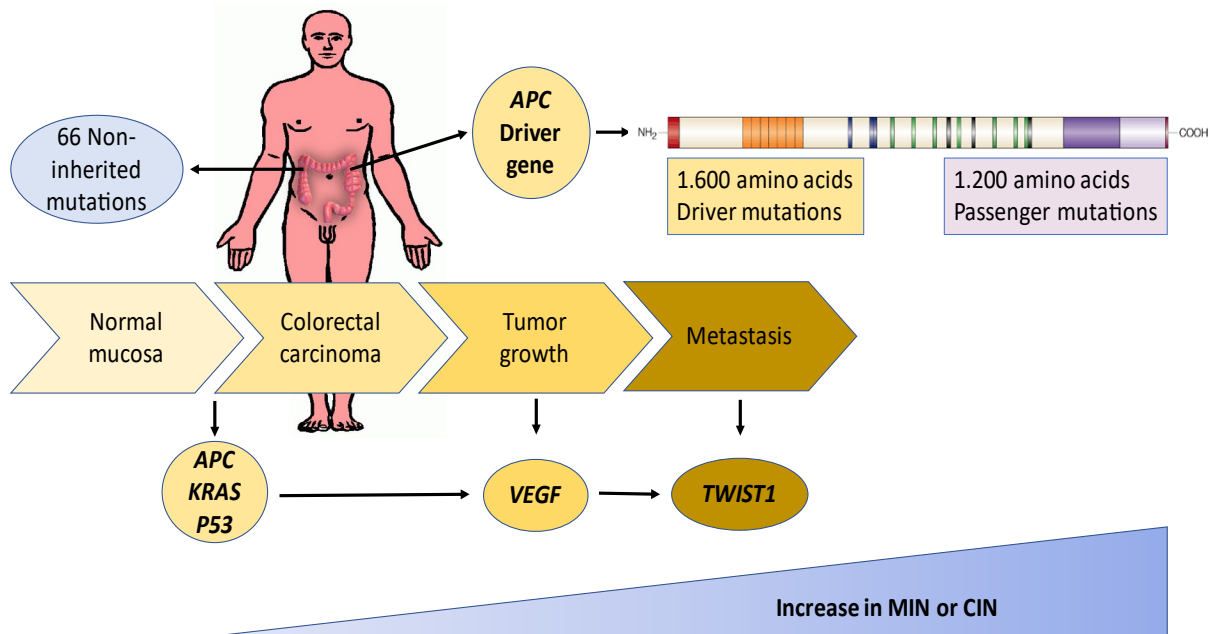


Figure 1. Mutational evolution associated with genomic instability in colorectal cancer

history because they can be used as a molecular clock to calculate the approximate age of the tumor lineage. In other words, the number of cell divisions that have occurred in the lineage of the dominant clone from the patient's birth to the time of biopsy. Passengers can also become immunotherapy targets or cause treatment resistance. Likewise, their accumulation can explain several clinical phenomena such as slow progression, long latency periods, the prevalence of small subclinical cancers, spontaneous regression, and the observed range in cancer growth rates. However, these events are not easily explained without considering deleterious passengers.

The incorporation of passengers into cancer evolution models

In 1976, genetic instability started to be considered a mechanism of tumorigenesis, even in the presence of very few chromosomal changes (such as in acute diploid leukemia and chronic granulocytic leukemia) resulting in a 'clonal evolution model'. In this paradigm, carcinogen-induced changes in normal progenitor cells produced diploid tumor cells with a growth advantage that allowed their initial clonal expansion [25]. Over the past four decades, this classic model of cancer evolution has focused exclusively on driver mutations to describe how these sequentially-acquired alterations provide an advantage to the growth of tumor cells. However, the integration of mathematical frameworks into the analysis of cancer genome sequencing data has resulted in new models: (a) in the 'Big Bang model', clonal and subclonal mutations arise early and the tumor grows as a single intermixed population; (b) in the 'neutral model' there is no significant difference between clonal populations; and (c) in the 'punctuated evolution model' clones rapidly arise between periods of relative mutational equilibrium [26] (Figure 2).

There is also a commonly used 'two-hit model' in which the first driver mutation produces no fitness benefit but the combination of two driver mutations can have a strong cumulative effect. Furthermore, the recently proposed evolutionary 'stochastic model' can explain the dynamics of cancer progression and describes how individual cells can divide and potentially acquire driver or passenger alterations, which may be involved in cell death. The size of tumors changes with the production and death of individual cells, which usually depends on the effect of accumulated drivers and passengers and the cellular

environment. Thus, assuming that all driver and passenger mutations have equal fitness advantages and disadvantages, these rates are changed by the number of drivers and passengers and the total hyperplasia, or cell population size [13].

The stochastic evolutionary model of cancer progression indicates that deleterious passengers can accumulate in cancer and that these usually have a negative selection effect on tumor cells; genomic analysis shows that the passengers present in sequenced cancers have harmful phenotypes (to the cancer). Importantly, the origin of many properties of cancer may be explained if their phenotype is considered as a balance between drivers and the deleterious effect of passenger mutations—phenomena not considered in the original model. These include (a) slow initial and rapid late growth; (b) a critical cancer size for dormancy or spontaneous regression; and (c) short-term responses to mutagenic therapies.

In addition, cancers that accumulate alterations in passengers, better respond to treatments are observed. This detrimental effect on cancer is currently an untapped therapeutic target which could be leveraged by increasing either the overall mutation rate (to augment the rate of passenger mutation accumulation) or the deleterious effect of these passengers [13,14]. This evolutionary model led to the discovery of useful treatments via molecular oncology and has facilitated the proposal of new therapeutic targets for the field of precision medicine. In the following sections, we discuss how the effects of passenger mutations could be enhanced in current and future therapies.

Passenger mutations as molecular clocks

The number of passenger mutations accumulated in each tumor lineage can provide information about its approximate age which also corresponds to the age of the tumor itself. The mean age of ovarian and lung tumor lineages is 1,113 and 749 cell divisions, respectively [26]. This is because the ovarian epithelium self-renews regularly [27] while the pulmonary epithelium renews slowly but this process is stimulated after injury [28]. Given that the cell division time for lung tumor cells is about 8 days [29], by multiplying this rate by the average lung tumor lineage age, we can say that most cancers of this type are detected 16.4 years after they first started.

Using computational methods and cell division rates for different tumor types (based on mutational signatures from genome sequencing

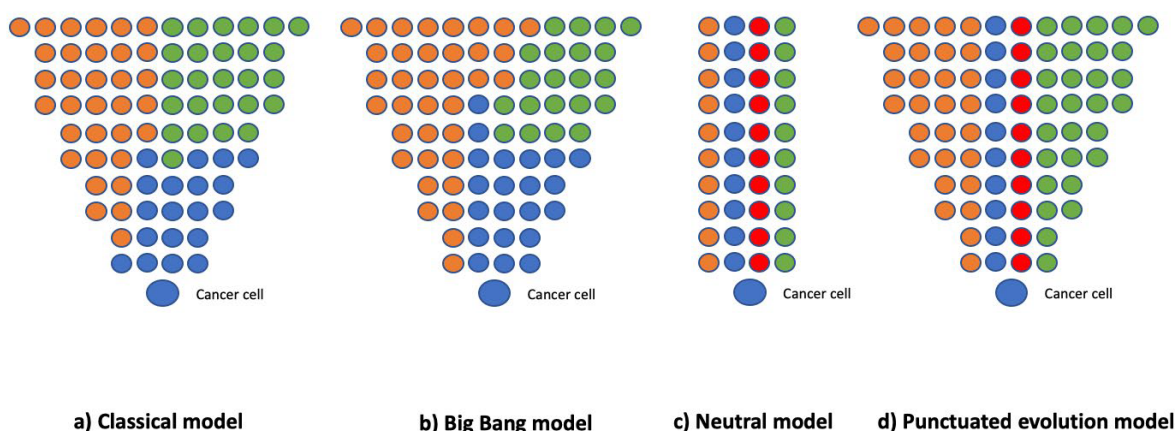


Figure 2. a) In the classical model, sequentially acquired driver mutations offer growth advantage; (b) in the Big bang model, clonal and subclonal mutations arise early, and tumor grows as a single intermixed population; (c) in neutral model, there is no major difference in fitness between the clonal populations, and (d) in the punctuated evolution model, clones arise rapidly between periods of relative mutational equilibrium

data), approximations of tumor chronological age could offer important clues into their biology. These include information about the intrinsic and extrinsic mutagenic influences present during the period of tumor initiation and the clonal evolution of the tumor. This data may also help our understanding of intratumoral heterogeneity and differences in patient prognosis and responsiveness to treatment. Together with clinical data, this can help elucidate the time required for benign tumors to become invasive and metastatic and help design methods for their early detection and treatment [26].

Using passengers to identify mutator phenotypes

The mutator-phenotype hypothesis suggests that the mutation rate of normal cells is too slow to produce the large number of alterations found in human tumors and that the elevated mutation rate of human tumor cells increases their likelihood of acquiring advantageous mutations. The hypothesis predicts that tumors contain cells harboring hundreds of thousands of mutations rather than only a few specific driver mutations [30], and that malignant cells within a tumor therefore constitute a highly heterogeneous population [31,32]. Probabilistic mathematical models can identify mutator genes that cause point mutations or increased chromosomal alteration rates and can estimate their effect during carcinogenesis.

Data for ovarian cancer reveals that alterations in the genes or regions that result in a mutator phenotype tend to occur early. For non-mutator genes and regions, those containing *MYC*, *KRAS*, *CCNE1*, and *RB1* tend to be altered early, while *CSMD3*, *USH2A*, and the region containing *MECOM* and *WWOX* tend to be altered late. In lung cancer, *TP53* and *PRKDC* increase the rate of mutations; *EGFR*, *KRAS*, *STK11*, and *TP53* tend to mutate early while *LRP1B* and *PTPRD* tends to become mutated late. However, this method of identifying altered driver genes by correlating them to passenger mutations generates many false positives because mutator genes cannot be distinguished from genes that alter later [26].

Genomic instability can be caused by dysfunction of DNA repair and cell-cycle checkpoint control genes. DNA repair genes altered in cancers include *BRCA1/2*, *MSH2/6*, *MLH1/2*, *BLM*, *RAD50*, *MRE11*, *NBS1*, *PRKDC*, *NBS1*, *BLM*, *RECQL4*, *BAP1*, *WRN*, *RAD51L3*, *RAD52*, *FANCA*, and *PALB2* [18,33]. In lung and ovarian cancer, *BRCA1/2*, *PRKDC*, and *PPP2R2A* in the 8p21.2 region are mutator genes, and *PPP2R2A* plays a role in inducing chromosomal instability in ovarian cancer. Cell-cycle checkpoint pathway genes that are altered in cancers include *TP53*, *ATM*, *MDM2/4*, *BUB1*, and *STK12*, of which, *TP53* is a mutator gene.

Epigenetic modifications: Driver methylation

NGS has helped to promote knowledge of epigenetics and its standing as a field. Epigenetics is the study of transmissible changes that does not involve DNA sequences. The three major types of epigenetic regulators are post-translational modification of histone tails, DNA methylation by covalent modification of cytosine-5', and the regulation of microRNA gene expression [34]. In particular, DNA methylation has been extensively assessed in breast, colon, esophageal, lung, pancreatic, ovarian, prostate, and other cancers [35]. Because epigenetic changes affect genomic stability and gene expression, they influence every stage of carcinogenesis over a person's whole life, and sometimes even across generations [36]. The challenge is now to identify methylation changes that are crucial to the processes of tumor initiation, progression, or metastasis and to distinguish these from non-carcinogenic passengers accompanying the transformation process [37].

Global DNA hypomethylation and gene-specific hypermethylation are among the prominent hallmarks of cancer genomes [38]. Some hypermethylated genes in cancer may be tumor suppressor genes, but it is unlikely that all of these numerous methylation changes play a causative role in tumorigenesis; rather, the majority of promoter CpG islands are probably methylated as a consequence of, or in association with, carcinogenesis (passenger methylation). Thus, key genes that are susceptible to methylation-associated gene silencing and that are functionally important in preventing tumorigenesis (driver methylation) must be pinpointed. This is perhaps an analogous situation to that of mutational changes in cancer: genome-wide DNA sequencing of either a large number of coding sequences or of entire cancer genomes have revealed the presence of a staggering number of mutational changes [39].

Driver methylation can be considered a methylation event which promotes tumorigenesis. If the tumor-driving or initiating event is a methylation change, it is more likely to occur during the early stages of tumorigenesis. In mouse models and in early-stage human tumor specimens and premalignant lesions, methylation changes can be observed from preneoplastic tissues up until late malignant disease [40]. Thus, early changes in methylation most probably drive the cancer phenotype, and later changes may simply reflect the transformed phenotype. Driver methylation can both directly and indirectly inactivate suppressor genes or activate oncogenes and this methylation-based gene silencing mechanism can be considered one of the hallmarks of cancer [4].

Mitochondrial DNA mutations

Mitochondrial DNA (mtDNA) is a small, circular, double-stranded DNA molecule approximately 16.6 Kb long that encodes 2 ribosomal RNAs (12S and 16S), 22 transfer RNAs required for protein synthesis, and 13 protein subunits that are essential for oxidative phosphorylation [41]. mtDNA is more susceptible to mutations than nuclear DNA because it lacks histones and chromatin-protective structures, has very few introns, its mtDNA repair mechanisms are inefficient, and because it is exposed to high levels of deleterious reactive oxygen species generated during ATP synthesis [42]. In 1956, Otto Warburg defined mitochondrial dysfunction as a hallmark of cancer progression; this led to the proposal of the Warburg effect—that cancer cells favor aerobic glycolic metabolism over oxidative phosphorylation. Genetic and pharmacological studies have conclusively shown that this effect is required for tumor growth, although the reason for this remains controversial [43,44]. Nonetheless, targeting damaged mtDNA could represent a promising anticancer therapy target [45].

Another characteristic cancer marker is the ability of tumor cells to reprogram their own metabolism to cope both with the abnormal protein-building requirements of their uncontrolled cell proliferation and to adapt to their everchanging microenvironments [4]. Demonstration that mitochondrial metabolism can be triggered by activation of oncogenes such as BRAF and c-Myc [46,47], loss of tumor suppressors such as p53 [48], and activation of the mTORC1 pathway [49] resulted in the redefinition of the Warburg effect. Even so, it has become evident that accelerated mitochondrial function, including ATP production, is required for cell proliferation and tumor progression [50], at least in specific phases such as adaptation to nutrient and oxygen deprivation [51].

To maintain physiological energy levels, cells have developed a sensitive molecular system that integrates multiple upstream inputs and regulates enzyme activity and transcriptional responses. The core

enzyme in this system is the AMP-activated protein kinase, AMPK. This enzyme restores energy levels when intracellular ATP drops, for instance, in response to mitochondrial dysfunction or stress. AMPK has been widely implicated in tumor initiation, progression, and metastasis [52]. However, genetic ablation resulting in AMPK loss is not sufficient to induce cell transformation in vitro or in vivo models [53]. To further complicate this scenario, AMPK is differentially expressed and activated in different cancer types and disease stages and is associated with varying outcomes and prognoses [51].

It is commonly believed that mtDNA variants arise due to positive selection of those “driver” variants conferring clonal growth advantage. Accordingly, we observed that likely non-pathogenic mtDNA variants (“passengers”) reverted to the wild-type homoplasmic status during tumor progression in colorectal cancer patients [54]. On the contrary, the mtDNA variants that are positively selected during the tumor progression might be considered the most tolerable alterations for neoplastic cells. However, a deleterious impact of mtDNA passenger variants on cancer progression may not be completely excluded, as it has previously been evidenced in nuclear DNA passenger alterations [55].

Passenger mutations in different tumor types

The association between immunotherapy response the passenger mutation load has been widely studied. Furthermore, numerous studies have estimated the mutational load in different tumor types: melanoma, lung squamous carcinoma, lung adenocarcinoma, and bladder cancer are usually associated with a higher mutational load, while pilocytic astrocytoma, acute lymphoblastic leukemia, medulloblastoma, and acute myeloid leukemia are commonly associated with lower mutational loads [56]. In addition to the overall frequency of mutations, there is also a different spectrum of mutations in each tumor type. For instance, C→A mutations related to exposure to polycyclic aromatic hydrocarbons in tobacco smoke are associated with lung cancers [57] melanomas often show C→T mutations caused by the misrepair of ultraviolet-induced DNA breaks [58] gastrointestinal tumors, show high frequencies of CpG dinucleotide transition mutations, which may reflect higher methylation levels in these tumor types [59] and finally, tumors with a mutator phenotype caused by microsatellite instability bear a mutational load that far exceeds even that of melanomas [60,61].

Overall survival and tumoral progression

Another hallmark of cancer is genomic instability, which causes many chromosomal disorders and cellular lineage mutations [62] and both driver and passenger mutations. Passengers account for an estimated 97% of tumor cell mutations [11]; they have always been presumed to be neutral and have largely been ignored in cancer research. Passengers are potential biomarkers for patient responses to mutagenic therapies, and increasing evidence now suggests that they might be deleterious to cancer cells, making them important both in clinical and cancer progression outcomes. Because of the limitations in whole-genome analyses, the properties of passenger mutations remain unclear although varying arguments suggest that they are ‘mini-drivers’ [12], ‘latent drivers’ [63], neutral [13], or potentially deleterious to cancer [14,15].

The ‘tug-of-war’ resulting both from the cumulative effect and presence of high numbers of passengers and drivers in tumor cells may explain some paradoxical cancer treatment outcomes. Furthermore, their accumulation can even cause tumor extinction by mutational meltdown [64], although this is not yet fully understood.

Some hypotheses that can explain better prognoses resulting from the accumulation of passengers include: (a) increased tumor immunogenicity [65]; (b) the correlation of high genomic instability with improved clinical outcomes [66]; and (c) reduced cell proliferation [67]. More research is still required, but preliminary studies indicate that clinical phenomena such as long periods of dormancy, slow progression, growth-rate heterogeneity, spontaneous regression, and the prevalence of small subclinical cancers, could be the result of high deleterious passenger accumulation altering the dynamics of cancer progression. For example, budding yeast [68], primary mouse cells [63], and human aneuploid cells [69] with high passenger burdens all show evidence of a proliferative tumor cell growth disadvantage.

Another indirect example is Lynch syndrome which results from a germline mutation in the *MLH1*, *MSH2*, *MSH6*, or *PMS2* DNA mismatch repair (MMR) genes or in *EPCAM*. The DNA MMR system maintains genomic integrity by correcting base substitutions and small insertion-deletion mismatches generated by base-pairing errors during DNA replication. Inactivation of both alleles of an MMR gene leads to defective MMR and may result in high mutability and target-gene inactivation. These mutations better survive in Lynch syndrome patients than in patients without MMR defects [70]. Simultaneously, many studies suggest that single-agent adjuvant fluoropyrimidine-based chemotherapy is less beneficial, or is even potentially harmful, to patients with microsatellite instable or MMR tumors [71,72]. However, the prognostic influence of microsatellite instable is less clear in patients with metastatic colorectal cancer, a population in which the prevalence of MSI-H disease is low (approximately 3.5 percent). One hypothesis that can explain it, is the adverse influence of the higher frequency of *BRAF* mutations in this population [73].

Passenger mutations, chemotherapy and immunotherapy

Tumor cell growth and survival depends on several mechanisms, including angiogenesis and immune-system avoidance; limitless replication potential is a major factor because constant biosynthesis requires continued genetic material. Thus, classical chemotherapy and radiotherapy aims to impede tumor growth by damaging DNA [74] and this may also allow the accumulation of mutations that can become neoantigens—new targets for immune-system detection. In support of this hypothesis, new data suggests that targeted tumor irradiation combined with dual CTLA-4/PD-1 blockade in melanoma is a promising treatment option. Similarly, cisplatin and checkpoint blockade has proven a successful first-line treatment for non-small cell lung cancer (NSCLC). Another strategy is the development of drug combinations such as olaparib with immunotherapy [75]. Colorectal cancer patients with germline loss-of-function mutations in DNA mismatch repair genes have a 4 to 7-fold greater response rate to pembrolizumab than patients without them. Furthermore, DNA damage response (DDR)-deficient tumors harbor 10 to 100 times more somatic mutations than DDR-proficient patient tumors.

Nitrogen mustard, an analogue of the sulfur mustard gas used as a weapon during the First World War, was introduced in 1942 as the first clinically useful alkylating agent [76] its discovery was the first step towards cancer chemotherapy (17). As a general rule, the cell cycle is always affected by drugs (such as chemotherapy agents) that interact with DNA, although final outcomes depend on the extent of the interaction and the speed at which DNA reparation can overcome any negative effects. Indeed, one of the mechanisms of tumor cell resistance to alkylating and platinum agents is attributed to enhanced DNA

cross-link repair. Therefore, a new interpretation for the effectiveness of traditional genotoxic chemotherapy is that, aside from directly inhibiting tumor cell growth, these agents increase passenger load which may at least temporarily decrease the malignant potential of cancer development [77].

The accumulation of genetic alterations in cancer cells during tumorigenesis results in tumor neoantigens; these are increasingly considered to be immuno-determinants and in the context of immunotherapy treatment, there is abundant proof that they are related to early tumor recognition and destruction by antigen-specific T cells [78,79]. The expression of neoantigens in cancer cells highlights the 'foreignness' of cancer within the human body. Specifically, mutational load (a surrogate marker for tumor neoantigen load) correlates with the expected result of experiments to block T-cell checkpoint inhibitors in melanoma and NSCLCs [80].

Two classes of antigens provide cancer-rejection epitopes: the first are created by nonmutated proteins that are not completely tolerated by T-cells (partly because of their restricted tissue expression patterns). The second type—neoantigens—are created by proteins that are missing from normal human genome, usually in tumors without a viral etiology in which tumor-specific DNA alterations cause the formation of new protein sequences. In virus-associated tumors such as cervical cancer and a subgroup of head and neck cancers, antigens derived from viruses also add to the pool of neoantigens [81]. Compared with non-mutated self-antigens, neoantigens may be especially important in tumor control because the quality of the T cell pool available for these antigens is not affected by central T-cell tolerance [82].

Deep-sequencing technologies make it relatively easy to identify mutations present that are potential neoantigens and they can be confidently predicted [83]. Two studies in mouse models provided proof that this approach can identify neoantigens recognizable by T cells [84,85]. To leverage this phenomena, neoantigens would ideally be derived from essential oncogenes common to most cancers, thus reducing the probability of their escape from the immune system. For instance, in MHC class I- and class II-restricted neoantigens [86] in validated oncogenes shared between patient subgroups [87] which are known to produce T-cell responses. Cancers with substantial exogenous mutagenic exposure, such as ultraviolet light in the case of melanoma or exposure to tobacco smoke carcinogens in lung cancers, have very high mutation rates. A single case report of a melanoma tumor found 187 non-synonymous mutations [57] and an average of 201 mutations across 14 other melanoma cases [88], and another study identified 94 non-synonymous mutations in a lung cancer cell line [89] and 4,300 mutations in one primary tumor [78]. In addition, tumors with mismatch repair deficiencies also carry large numbers of mutations [90].

Most of these mutations are 'neutral passengers', implying that T-cell reactivity towards neoantigens is usually directed against mutated gene products dispensable for tumor growth. Indeed, selective attack by the immune system may cause the loss of mutated genes; in line with this, intriguing work by Schreiber et al. in a murine model demonstrated the loss of expression of a passenger mutation after T-cell exposure [47,75]. It is unknown whether T-cell pressure during human cancer development is sufficient to lead to a similar immune selection, and so this important question should be addressed in future research. There are also 'essential passenger' mutations which occur in essential (housekeeping) genes in cases where the wild-type copy is lost, then T-cell reactivity against the neoepitope can only lead to immune escape by mutation reversal. Coulie et al. were the first to describe an essential

passenger, by identifying a mutant malate dehydrogenase enzyme epitope recognized by autologous T cells [91].

The formation of any one neoantigen by a given mutation is a probabilistic 'lottery'. This means that although tumor foreignness can likely be guaranteed for tumors with very high mutational loads, the odds of tumor foreignness can only likely be inferred for tumors with an intermediate or low mutational load. Nevertheless, mutational load represents an imperfect biomarker, even specific cases where neoantigen reactivity is the only tumor-specific T cell reactivity relevant to tumor control. Additionally, the success of the immune system attacks on cancer cells depends on several factors, including the formation of tumor-specific antigens. This concept was well described by the myriad of inhibitory and stimulatory factors involved in the cancer-immunity cycle introduced by Chen and Mellman [92].

Conclusions

Traditionally, cancer research has paid less attention to passengers, even though these represent the overwhelming majority of mutations. However, these may now take the spotlight in early diagnosis and in the improvement of future treatments because they appear to slow tumor growth and reduce metastatic progression. Cancers with the highest burden of chromosomal alterations have the best prognoses which is thought to be because passenger mutations interfere with the acquisition of new drivers and thus, reduce genetic diversity. This information will be very useful for devising new anti-cancer therapies, and therefore, we argue that current approaches in targeted cancer therapy should be carefully reconsidered. For instance, the induction of genomic instability and increased passenger-mutation loads is now considered alternative therapeutic possibilities for cancer treatment in immunotherapy. However, a greater understanding of the complexity of tumors with mutations, both driver and passenger mutations, can help us to better manage the treatment of patients with cancer, increasing their survival. The introduction of the Next Generation Sequencing is allowing this greater integration of all players present in the tumor.

Authorship and contributorship

Aparisi F and Amado-Labrador H are the main authors of the study in the search for information and development of the review. Calabuig S, Jantus-Lewintre E, Blasco A, Iranzo V, Herreros-Pomares A and Camps C have contributed in the process to obtain articles, revision of the manuscript and the supervision of the process.

Funding information

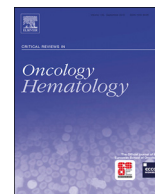
This work was supported by Fondo de Investigación Sanitaria, Instituto de Salud Carlos III and Fundación Arnal Planelles. Hector Amado is supported by a Fundación Carolina/ BBVA grant.

References

- Loeb LA, Springgate CF, Battula N (1974) Errors in DNA replication as a basis of malignant changes. *Cancer Res* 34: 2311-2321. [[Crossref](#)]
- Pepper JW, Findlay CS, Kassen R, Spencer SL, Maley CC (2009) Cancer research meets evolutionary biology. *Evol Appl* 2: 62-70. [[Crossref](#)]
- Hanahan D, Weinberg RA (2000) The Hallmarks of cancer review evolve progressively from normalcy via a series of pre. *Cell* 100: 57-70.
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674. [[Crossref](#)]
- Hudson TJ, Anderson W, Aretz A, Barker AD, Bell C, et al. (2010) International network of cancer genome projects. *Nature* 464: 993-998. [[Crossref](#)]

6. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, et al. (2014) Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* 12: e1001906. [[Crossref](#)]
7. Gerlinger M, Rowan AJ, Sc B, Horswell S, Math M (2016) Europe PMC funders group intratumor heterogeneity and branched evolution revealed by multiregion sequencing. 366: 883-892.
8. Shah Sohrah P, Roth Andrew, Goya Rodrigo, Oloumi Arusha HG, Zhao Y, et al. (2013) NIH Public Access. *Nature* 486: 1-13.
9. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* 107: 18545-18550. [[Crossref](#)]
10. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, et al. (2013) Cancer genome landscapes. *Science* 339: 1546-1558. [[Crossref](#)]
11. Castro-Giner F, Ratcliffe P, Tomlinson I (2015) The mini-driver model of polygenic cancer evolution. *Nat Rev Cancer* 15: 680-685. [[Crossref](#)]
12. Bozic I, Gerold JM, Nowak MA (2016) Quantifying clonal and subclonal passenger mutations in cancer evolution. *PLoS Comput Biol* 12: 1-19.
13. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA (2013) Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A* 110: 2910-2915. [[Crossref](#)]
14. McFarland CD, Mirny LA, Korolev KS (2014) Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proc Natl Acad Sci* 111: 15138-15143.
15. Cancer Genome Atlas Research Network JN, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, et al. (2013) The cancer genome atlas pan-cancer analysis project. *Nat Genet* 45: 1113-1120.
16. McFarland CD, Yaglom JA, Wojtkowiak JW, Scott JG, Morse DL, et al. (2017) The damaging effect of passenger mutations on cancer progression. *Cancer Res* 77: 4763-4772. [[Crossref](#)]
17. Yao Y1, Dai W2 (2014) Genomic instability and cancer. *J Carcinog Mutagen* 5. [[Crossref](#)]
18. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396: 643-649. [[Crossref](#)]
19. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458: 719-724. [[Crossref](#)]
20. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505: 495-501. [[Crossref](#)]
21. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321: 1807-1812.
22. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177-183. [[Crossref](#)]
23. Li Q, Ishikawa TO, Oshima M, Taketo MM (2005) The threshold level of adenomatous polyposis coli protein for mouse intestinal tumorigenesis. *Cancer Res* 65: 8622-8627.
24. Benjamin LE, Keshet E (1997) Conditional switching of vascular endothelial growth factor (VEGF) expression in tumors: induction of endothelial cell shedding and regression of hemangioblastoma-like vessels by VEGF withdrawal. *Proc Natl Acad Sci U S A* 94: 8761-8766.
25. Nowell PC (1976) The clonal evolution of tumor cell populations. *Science* 194: 23-28. [[Crossref](#)]
26. De S, Ganesan S (2017) Looking beyond drivers and passengers in cancer genome sequencing data. *Ann Oncol* 28: 938-945. [[Crossref](#)]
27. Tomasetti C, Vogelstein B, Parmigiani G (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* 110: 1999-2004. [[Crossref](#)]
28. Stripp BR, Reynolds SD (2008) Maintenance and repair of the bronchiolar epithelium. *Proc Am Thorac Soc* 5: 328-333. [[Crossref](#)]
29. Tinnemans MM, Schutte B, Lenders MH, Ten Velde GP, Ramaekers FC, et al. (1993) Cytokinetic analysis of lung cancer by in vivo bromodeoxyuridine labelling. *Br J Cancer* 67: 1217-1222. [[Crossref](#)]
30. Kinzler KW, Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell* 87: 159-170. [[Crossref](#)]
31. Salk JJ, Fox EJ, Loeb LA (2012) NIH Public Access. 51-75.
32. Fox EJ, Salk JJ, Loeb LA (2009) Cancer genome sequencing - An interim analysis. *Cancer Res* 69: 4948-4950. [[Crossref](#)]
33. Loeb LA (2001) A mutator phenotype in cancer. *Cancer Res* 61: 3230-3239. [[Crossref](#)]
34. Callinan PA, Feinberg AP (2006) The emerging science of epigenomics. *Hum Mol Genet* 15 Spec No 1: R95-R101. [[Crossref](#)]
35. Verma M, Srivastava S (2002) Epigenetics in cancer: Implications for early detection and prevention. *Lancet Oncol* 3: 755-763.
36. Brasset E, Chambeyron S (2013) Epigenetics and transgenerational inheritance. *Genome Biol* 14: 306. [[Crossref](#)]
37. Takeshima H, Ushijima T (2019) Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *NPJ Precis Oncol* 3: 7. [[Crossref](#)]
38. Ushijima T (2005) Detection and interpretation of altered methylation patterns in cancer cells. *Nat Rev Cancer* 5: 223-231.
39. Pfeifer GP, Besaratinia A (2009) Mutational spectra of human cancer. *Hum Genet* 125: 493-506. [[Crossref](#)]
40. Chen SS, Raval A, Johnson AJ, Hertlein E, Liu T-H, et al. (2009) Epigenetic changes during disease progression in a murine model of human chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 106: 13433-13438. [[Crossref](#)]
41. Taylor RW, Turnbull DM (2005) Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6: 389-402. [[Crossref](#)]
42. Hanes JW, Thal DM, Johnson KA (2006) Incorporation and replication of 8-oxo-deoxyguanosine by the human mitochondrial DNA polymerase. *J Biol Chem* 281: 36241-36248. [[Crossref](#)]
43. Fantin VR, St-Pierre J, Leder P (2006) Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. *Cancer Cell* 9: 425-434.
44. Shim H, Chun YS, Lewis BC, Dang C V (1998) A unique glucose-dependent apoptotic pathway induced by c-Myc. *Proc Natl Acad Sci* 95: 1511-1516.
45. Warburg O (1956) On the origin of cancer cells. *Science* 123: 309-314. [[Crossref](#)]
46. Kang HB, Fan J, Lin R, Elf S, Ji Q, et al. (2015) Metabolic rewiring by oncogenic BRAF V600E links ketogenesis pathway to BRAF-MEK1 signaling. *Mol Cell* 59: 345-358. [[Crossref](#)]
47. Fan Y, Dickman KG, Zong WX (2010) Akt and c-Myc differentially activate cellular metabolic programs and prime cells to bioenergetic inhibition. *J Biol Chem* 285: 7324-7333.
48. Ma W, Sung HJ, Park JY, Matoba S, Hwang PM (2007) A pivotal role for p53: Balancing aerobic respiration and glycolysis. *J Bioenerg Biomembr* 39: 243-246.
49. Sun Q, Chen X, Ma J, Peng H, Wang F, et al. (2011) Mammalian target of rapamycin up-regulation of pyruvate kinase isoenzyme type M2 is critical for aerobic glycolysis and tumor growth. *Proc Natl Acad Sci* 108: 4129-4134.
50. Weinberg F, Hamanaka R, Wheaton WW, Weinberg S, Joseph J, et al. (2010) Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc Natl Acad Sci* 107: 8788-8793.
51. Smolkova K, Plecitan-Hlavata L, Bellance N, Benard G, Rossignol R, et al. (2011) Waves of gene regulation suppress and then restore oxidative phosphorylation in cancer cells. *Int J Biochem Cell Biol* 43: 950-968.
52. Hardie DG (2015) Molecular pathways: Is AMPK a friend or a foe in cancer? *Clin Cancer Res* 21: 3836-3840. [[Crossref](#)]
53. Laderoute K (2006) 5'-AMP-activated protein kinase (AMPK) is induced by low-oxygen and glucose deprivation conditions found in solid-tumor microenvironments. *Mol Cell Biol* 26: 5336-5347.
54. Errichiello E, Balsamo A, Cerni M, Venesio T (2015) Mitochondrial variants in MT-CO2 and D-loop instability are involved in MUTYH-associated polyposis. *J Mol Med* 93: 1271-1281.
55. Errichiello E, Venesio T (2017) Mitochondrial DNA variants in colorectal carcinogenesis: Drivers or passengers? *J Cancer Res Clin Oncol* 143: 1905-1914. [[Crossref](#)]
56. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, et al. (2013) Signatures of mutational processes in human cancer. *Nature* 500: 415-424.

57. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, et al. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463: 184-190.
58. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463: 191-196.
59. Cancer T, Atlas G (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487: 330-337.
60. Dung T, Le J, Jennifer N, Uram H, Hao Wang H, Bjarne Bartlett B, Holly Kemberling (2015) PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 372: 2509-2520.
61. Adjiri A (2017) DNA mutations may not be the cause of cancer. *Oncol Ther* 5: 85-101. [Crossref]
62. Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501: 338-345.
63. Nussinov R, Tsai CJ (2015) 'Latent drivers' expand the cancer mutational landscape. *Curr Opin Struct Biol* 32: 25-32. [Crossref]
64. Neher RA, Shraiman BI (2012) Fluctuations of fitness distributions and the rate of Muller's ratchet. *Genetics* 191: 1283-1293. [Crossref]
65. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N (2015) Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160: 48-61. [Crossref]
66. Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, et al. (2011) Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res* 71: 3447-3452.
67. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, et al. (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* 322: 703-709.
68. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, et al. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317: 916-924. [Crossref]
69. Segal DJ, McCoy EE (1974) Studies on down's syndrome in tissue culture. I. Growth rates and protein contents of fibroblast cultures. *J Cell Physiol* 83: 85-90. [Crossref]
70. Phipps AI, Limburg PJ, Baron JA, Burnett-Hartman AN, Weisenberger DJ, et al. (2015) Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology* 148: 77-87. [Crossref]
71. Elsaleh H, Joseph D, Grieu F, Zeps N, Spry N, et al. (2000) Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet* 355: 1745-1750.
72. Hutchins G, Southward K, Handley K, Magill L, Beaumont C, et al. (2011) Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol* 29: 1261-1270.
73. Seymour MT, Richman SD, Cheadle JP, Ylstra B, Nagtegaal ID, et al. (2014) Mismatch repair status and braf mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. *Clin Cancer Res* 20: 5322-5330.
74. Kim D, Li R, Dudek SM, Wallace JR, Ritchie MD (2015) Binning somatic mutations based on biological knowledge for predicting survival: an application in renal cell carcinoma. *Pacific Symp Biocomput* 0: 96-107.
75. Gilman A, Philips FS (1946) The biological actions and therapeutic applications of beta-chloroethyl amines and sulfides. *Science* 103: 0-415.
76. Lowenstein EB (2011) A history of sulfur mustard. *Skinmed* 9: 310-311. [Crossref]
77. [No authors listed] (2016) Cancer Mutational Load: The more the merrier for immune checkpoint blockade therapies. *EBioMedicine* 13: 1-2. [Crossref]
78. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, et al. (2014) Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* 344: 641-645. [Crossref]
79. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, et al. (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 371: 2189-2199. [Crossref]
80. Blank CU, Haanen JB, Ribas A, Schumacher TN (2016) CANCER IMMUNOLOGY. The "cancer immunogram". *Science* 352: 658-660. [Crossref]
81. Schumacher TN, Schreiber RD (2015) Neoantigens in cancer immunotherapy. *Science* 348: 69-74. [Crossref]
82. Gilboa E (1999) The makings of a tumor rejection antigen. *Immunity* 11: 263-270. [Crossref]
83. Segal NH, Parsons DW, Peggs KS, Velculescu V, Kinzler KW, et al. (2008) Epitope landscape in breast and colorectal cancer. *Cancer Res* 68: 889-892.
84. Castle JC, Kreiter S, Diekmann J, Löwer M, van de Roemer N, et al. (2012) Exploiting the mutanome for tumor vaccination. *Cancer Res* 72: 1081-1091. [Crossref]
85. Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, et al. (2013) Cancer exome analysis reveals a t cell dependent mechanism of cancer immunoeediting. *Nature* 482: 400-404. [Crossref]
86. Wölfel T, Hauer M, Schneider J, Serrano M, Wölfel C, et al. (1995) A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* 269: 1281-1284.
87. Schumacher T, Bunse L, Pusch S, Sahn F, Wiestler B, et al. (2014) A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* 512: 324-327.
88. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, et al. (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465: 473-477.
89. Xiaomu Wei, Vijay Walia, Jimmy C Lin, Jamie K Teer, Todd D Prickett, et al. (2011) Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* 43: 442-446.
90. Greenman C, Stephens PJ, Smith R, Dalgliesh GL, Hunter C, et al. (2007) UKPMC funders group patterns of somatic mutation in human cancer genomes. *Nature* 446: 153-158.
91. Karanikas V, Colau D, Baurain JF, Chiari R, Thonnard J, et al. (2001) High frequency of cytolytic T lymphocytes directed against a tumor-specific mutated antigen detectable with HLA tetramers in the blood of a lung carcinoma patient with long survival. *Cancer Res* 61: 3718-3724.
92. Chen DS, Mellman I (2013) Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39: 1-10. [Crossref]



EpCAM duality becomes this molecule in a new Dr. Jekyll and Mr. Hyde tale

Alejandro Herreros-Pomares^{a,b,1}, Cristóbal Aguilar-Gallardo^{a,b,1}, Silvia Calabuig-Fariñas^{a,b,c}, Rafael Sirera^{b,d}, Eloísa Jantus-Lewintre^{a,b,d,*}, Carlos Camps^{a,b,e,f,*}

^a Molecular Oncology Laboratory, Fundació Investigación Hospital General Universitario de Valencia, Valencia, Spain

^b CIBERONC, Valencia, Spain

^c Department of Pathology, Universitat de València, Valencia, Spain

^d Department of Biotechnology, Universitat Politècnica de València, Valencia, Spain

^e Department of Medical Oncology, Consorcio Hospital General Universitario de Valencia, Valencia, Spain

^f Department of Medicine, Universitat de València, Valencia, Spain

ARTICLE INFO

Keywords:

Epithelial cell adhesion molecule (EpCAM)

Epithelial-mesenchymal transition (EMT)

Cancer stem cells

Cancer therapy

Clinical trials

ABSTRACT

EpCAM, known as an epithelial cell adhesion molecule, plays an essential role in cell adhesion, migration, metastasis and cell signalling. Rather than acting as an apoptosis antagonist, it induces cellular proliferation that impacts the cell cycle, and as a signalling transducer it uses and enhances the Wnt pathway, which is significantly relevant in cell renewal and cancer. EpCAM has become a marker of circulating tumour cells (CTCs) in lung cancer due to its specificity, and its high and stable expression level. Recent findings have allowed us to relearn and discover EpCAM again as a CSCs marker by demonstrating its role in human epithelial cancer progression. In line with this, the focus of attention on EpCAM has become an appealing therapeutic target, although the literature shows a clear controversy in information about its clinical significance. Despite this contradictory fact, solid evidence has demonstrated its dual role as a molecule with oncogenic and tumour suppressor properties, in which the microenvironment is influential. Therefore, its dual role appears to be both tissue- and tumour- dependent. In this review, we summarised the novel and updated insights in the EpCAM field by simplifying the understanding of the biological role of this fascinating molecule, and by showing the promising therapeutic tools that have been developed by various approaches which use antibodies and vaccines for different cancer types for the clear purpose of improving patient outcome.

1. EpCAM: a critical player in physiological and pathological conditions

The epithelial cell adhesion molecule (EpCAM), also known as the cluster of differentiation 326 (CD326), is a pleiotropic type I transmembrane glycoprotein that was first described as an epithelial-specific intercellular cell-adhesion molecule. However, many data suggest that its role is not limited to cell adhesion. We know that the EpCAM protein is not exclusively expressed in epithelial cells, but is also found in various tissue stem cells, precursors, and in embryonic stem cells (ESC) (Ng et al., 2010). EpCAM expression is required for the maintenance of self-renewal and the pluripotent phenotype in murine and human ESC (Lu et al., 2010). Its high expression in healthy tissue is involved in cellular signalling and migration by playing a crucial role in the balance of cell proliferation and differentiation during morphogenesis and tissue regeneration (Kuechlin et al., 2017). Alterations in the expression levels of EpCAM are associated with pathological conditions. EpCAM is

frequently overexpressed in epithelial tumours, in contrast to its low expression in normal simple epithelia. EpCAM was discovered as a cell adhesion molecule that mediates homophilic adhesive interactions; therefore, it acts as a tumour suppressor (Litvinov et al., 1994). In line with this, loss of EpCAM has been associated with increased migratory potential (Gosens et al., 2007), and EpCAM expression in metastases seems lower compared to primary tumours (Went et al., 2005). Accordingly, EpCAM overexpression has been associated with better prognosis in some tumour types (see Table 1). However, several oncogenic functions of EpCAM have been discovered over the years, including abrogation of E-cadherin-mediated cell-cell adhesion and association with claudin-7 (Philip et al., 2015), which interfere with homotypic cell-cell adhesion, and the promotion of cell motility, proliferation, survival, carcinogenesis and metastasis through the Wnt pathway (Chaves-Perez et al., 2013; Yu et al., 2017). These findings suggest a dual role of EpCAM and explain the conflicting published clinical data. Given its tumour-specific overexpression, EpCAM has also

* Corresponding authors at: Consorcio Hospital General Universitario de Valencia. Avenida Tres Cruces 2. 46014. Valencia. Spain.

E-mail addresses: jantus_elo@gva.es (E. Jantus-Lewintre), camps_car@gva.es (C. Camps).

¹ Equal contribution.

Table 1
Prognostic impact of EpCAM expression in epithelial cancer patients.

Study	Tumour Type	Patients included	Recurrence or Progression	Metastasis	Survival
Good Prognosis					
Seligson et al. (2004)	Renal cell cancer	304	DFS (HR 0.63 [0.43–0.92], p = 0.017)	NS	OS (HR 0.718 [0.541–0.953], p = 0.022)
Eichelberg et al. (2013)	Renal cell cancer	441		NS	OS (HR 0.772 [0.596–0.998], p = 0.047)
Zimpfer et al. (2014)	Renal cell cancer	948	PFS (HR 0.655 [0.414–1.037], p = 0.065)	NS	OS (HR 0.49 [0.20–1.23], p < 0.001)
Songun et al. (2005)	Gastric cancer	280		NS	OS (HR 0.54 [0.29–0.99], p = 0.05)
Dai et al. (2017)	Gastric cancer	375		NS	OS (HR 0.693 [0.291–1.653], p = 0.0185)
Kobayashi et al. (2010)	Lung cancer	130	RFS (HR 0.708 [0.524–0.956], p = 0.024)	NS	OS (HR 0.648 [0.461–0.910], p = 0.012)
Gold et al. (2014)	Lung cancer	370		NS	
Kim et al. (2016)	Colorectal cancer	725	DFS (HR 1.57 [1.04–2.39], p = 0.03)	Negative	DSS (HR 0.383 [0.201–0.727], p = 0.003)
Battista et al. (2014)	Ovarian cancer	117	PFS (HR 0.493 [0.286–0.848], p = 0.011)		
Non-significant					
Went et al. (2006)	Lung cancer	1287			
Kim et al. (2009)	Lung cancer	234			
Pak et al. (2012)	Lung cancer	164			
Goossens-Beumer et al. (2014)	Colorectal cancer	305			
Went et al. (2006)	Prostate cancer	414			
Sergeant et al. (2011)	Pancreatic cancer	48			
Fong et al. (2008)	Pancreatic cancer	153			
Laimer et al. (2008)	Head and Neck cancer	77			
Bayram et al. (2015)	Head and Neck cancer	294			
Poor prognosis					
Yanamoto et al. (2007)	Head and Neck cancer	48	NS	Positive	NS
Zhou et al. (2015)	Lung cancer	118		Positive	OS (HR 1.92 [1.02–3.64], p = 0.001)
Tayama et al. (2017)	Ovarian cancer	168	PFS (HR 1.13 [1.18–2.96], p = 0.013)	NS	OS (HR 2.17 [1.22–3.88], p = 0.008)
Schmidt et al. (2008)	Breast cancer	402	DFS (HR 1.60 [1.05–2.44], p = 0.028)	NS	
Agboola et al. (2012)	Breast cancer	726	DFI (HR 1.44 [1.09–1.92], p = 0.011)	Positive	
Soysal et al. (2013)	Breast cancer	1365			OS (HR 1.402 [1.178–1.668], p = 0.0001)
Sadeghi et al. (2017)	Breast cancer	104	PFS (HR 2.263 [1.006–2.610], p = 0.020)	Positive	
Prince et al. (2008)	Gallbladder cancer	25			OS (HR 2.9 [0.9–10.7], p = 0.005)
Bryan et al. (2014)	Urothelial bladder cancer	607			DSS (HR 1.763 [1.132–2.746], p = 0.012)

DFS: Disease free survival; PFS: Progression free survival; RFS: Relapse free survival; OS: Overall survival; DSS: Disease specific survival; HR: Hazard ratio; NS: Non-significant.

been explored for three decades as an anti-cancer target. In 2009, catumaxomab, a trispecific antibody for EpCAM, CD3 and, via the Fc γ receptor, antigen presenting cells (APC), obtained market approval in Europe to treat malignant ascites in cancer patients. Since then, different approaches that target EpCAM have been tested in clinical trials with promising results (Schmidt et al., 2010; Xiao et al., 2017; Deng et al., 2015). Thus to the best of our knowledge, this review summarises the current understanding of EpCAM's role by explaining persisting gaps and the main lines of academic debate, and by identifying new biological features that are the groundwork for new EpCAM outlooks.

2. The dual role of EpCAM in cancer: EpEX vs. EpICD

EpCAM is a transmembrane polypeptide of 314 amino acids (aa) that consists in a large extracellular domain of 242 aa, with only 26 aa facing the cytoplasm. Different proteolytic cleavages are associated with the complex regulation of EpCAM signalling (Denzel et al., 2009; Pavsic et al., 2014). The most widely studied EpCAM-mediated signalling is activated by regulated intramembrane proteolysis (RIP). The cleavage of EpCAM is initiated by a metalloprotease called TACE/ADAM17 (TNF- α converting enzyme, also known as disintegrin and metalloprotease) (Baeuerle and Gires, 2007), which cuts EpCAM at its extracellular domain and releases the EpCAM cleaved extracellular domain (EpEX). A second cleavage in the intracellular site by a multi-subunit protease complex, called γ -secretase, triggers the liberation of the cleaved intracellular domain of EpCAM (EpICD) (Denzel et al., 2009).

EpEX starts with the signal peptide, which is cleaved off between Ala-23 and Glu-24. Different models of the tertiary structure of EpEX have been developed, and three motifs have been defined: a first epidermal growth factor (EGF)-like domain, followed by a thyroglobulin (TY)-like domain and a cysteine-free domain (Schnell et al., 2013a,b). EpEX also contains three independent N-glycosylation sites in its extracellular domain at Asn74, Asn111, and Asn198 (Pavsic et al., 2014). The point mutations of the potential N-glycosylation sites have shown that all three sites are glycosylated in cancer, which results in a 3-fold increase of the EpCAM half-life (from 7 h to 21 h) (Schnell et al., 2013a,b). Since several cell surface molecules (Notch receptors, CD44, integrins, among others) are differentially glycosylated in carcinoma compared to simple epithelia, the differential glycosylation of EpCAM could be an important factor that brings about differences in the function of EpCAM in healthy versus malignant tissue. This is particularly interesting as EpCAM has been shown to be hyperglycosylated in the vast majority of head and neck carcinomas compared to healthy tissue, and the mutation of N-glycosylation has been associated with decreased adhesion capacity of breast cancer cells (Liu et al., 2017). Soon after EpCAM was discovered, a cleavage between two arginine residues at position Arg-80/Arg-81 in EpEX was identified. Following cleavage, the domains remain bound together by the disulphide bridge in the TY-like domain, which is broken under reducing conditions (Schnell et al., 2013a,b). The functional consequence of this cleavage is still under study. A protective role of EpCAM via its conserved TY motif has been proposed in some studies where EpCAM was suggested to inhibit cathepsins, matrix degradation proteases associated with greater invasiveness in some tumours (Schnell et al., 2013a,b; Baeuerle and Gires, 2007). Moreover, a possible autocrine and paracrine role of EpEX has been proposed through experiments using soluble recombinant EpEX, which concluded EpEX acts as an agonist for EpCAM⁺ cells (Liu et al., 2017). Remarkably, the analysis of serum from cancer patients revealed that detectable levels of circulating EpEX are frequent, but no significant results for progression or survival have been obtained in cohorts of patients with breast, ovarian, lung or prostate cancers (Tas et al., 2014; Karabulut et al., 2014). However, a diagnostic value has been found in breast and ovarian cancers, where patients' baseline serum EpCAM levels were significantly higher than those of controls (Tas et al., 2014; Karabulut et al., 2014). EpCAM has also been found to

be abundantly secreted by tumour cells on exosomes, but no clinical associations have yet been established (Rupp et al., 2011).

Cancer cell migration process is a concept that defines metastasis, which represents the final step in the malignant process and is the main cause of cancer patient mortality. In cancer, epithelial-mesenchymal transition (EMT) remains at the borderline between tumoural and non-tumoural tissues, which indicates that metastatic capability requires a specific microenvironment, as well as an inner aberrant cell signalling niche. Several studies have shown a clear relationship between EpCAM expression and EMT in tumours. EpCAM mediates Ca²⁺-independent cell-cell adhesion via intra- and intercellular homophilic interactions, which take place within a globular structure formed by the EGF-like domain and the TY-like domain in EpEX (Litvinov et al., 1997). As a result, EpEX is believed to act as an inhibitor of invasion. *In vitro* experiments have revealed that EpCAM expression in EpCAM⁻ cells leads to the aggregation of cells and the formation of cell-cell contacts (Litvinov et al., 1994). In addition, in a mix of EpCAM⁺ and EpCAM⁻ L-cells, cell aggregates mainly consist of EpCAM⁺ cells (90%) (Santisteban et al., 2009), and silencing EpCAM with short interfering RNA (siRNA) reduces the proliferation, migration and invasion of different prostate cancer cell lines (Ni et al., 2013). In patients, loss of EpCAM contributes to increased migratory potential (Gosens et al., 2007), and EpCAM expression in metastases of renal clear cell carcinomas is lower compared with primary tumours (Went et al., 2005). According to these results, EpCAM overexpression has been associated with improved overall survival in some epithelial tumours, including renal cell, gastric, colorectal and ovarian carcinomas (Zimpfer et al., 2014; Dai et al., 2017; Kim et al., 2016; Battista et al., 2014). However, some publications describe a role for EpCAM in promoting EMT (Chaves-Perez et al., 2013; Yu et al., 2017). Interestingly, an immunohistochemical analysis, which included 36 thyroid carcinoma patients, has found that nuclear and cytoplasmic expression and loss of membranous EpEX correlate positively with metastasis (Kunavisarut et al., 2012). These results suggest an antagonistic role of EpEX and EpICD, and could explain the conflicting results where a positive correlation between EpCAM expression and invasion are found.

3. Relationship to β -catenin/ E-cadherin/EMT

Unlike the tumour suppressive role associated with EpEX, EpICD has been described to be a promoter of cell motility, proliferation, survival and metastasis formation (Yu et al., 2017; Park et al., 2016). When soluble EpICD is released to the cytoplasm, a multi-protein nuclear complex is formed with the β -catenin and the Lymphoid Enhancer Factor 1 (Lef-1) (Fig. 1). The interaction between EpICD and various signalling proteins is achieved due to a scaffolding protein named four and one-half LIM domains protein 2 (FHL2) (Park et al., 2016). The fourth LIM domain of FHL2 is necessary for binding EpCAM, whereas LIM domains 2 and 3 bind β -catenin/p300. This complex may translocate to the nucleus, where it binds LEF-1 and DNA, and resembles the canonical Wnt signalling pathway (Yu et al., 2017; Ralhan et al., 2010). EpICD seems to be a key component of the complex by, regulating targeted genes of this evolutionarily conserved pathway involved in several developmental processes and the maintenance of the tissue homeostasis (Khosla et al., 2017). In most epithelia, E-cadherin and EpCAM are co-expressed during development and in post-natal life. We know that EpCAM overexpression is often associated with tumoural transformation, and cross-talk between EpCAM signalling and E-cadherin-mediated cell adhesion via the PI3K kinase pathway has been described. Winter et al. reported that EpCAM overexpression weakens the cadherin-mediated adhesions (Winter et al., 2007). Although the exact mechanism is unknown, it seems that EpCAM is able to increase the level of soluble β -catenin, which may be stabilised and protected from degradation due to an association with EpICD and FHL2 by enhancing its nuclear signalling via the Wnt pathway (Schnell et al., 2013a,b; Litvinov et al., 1997; Winter et al., 2007). Cadherins are

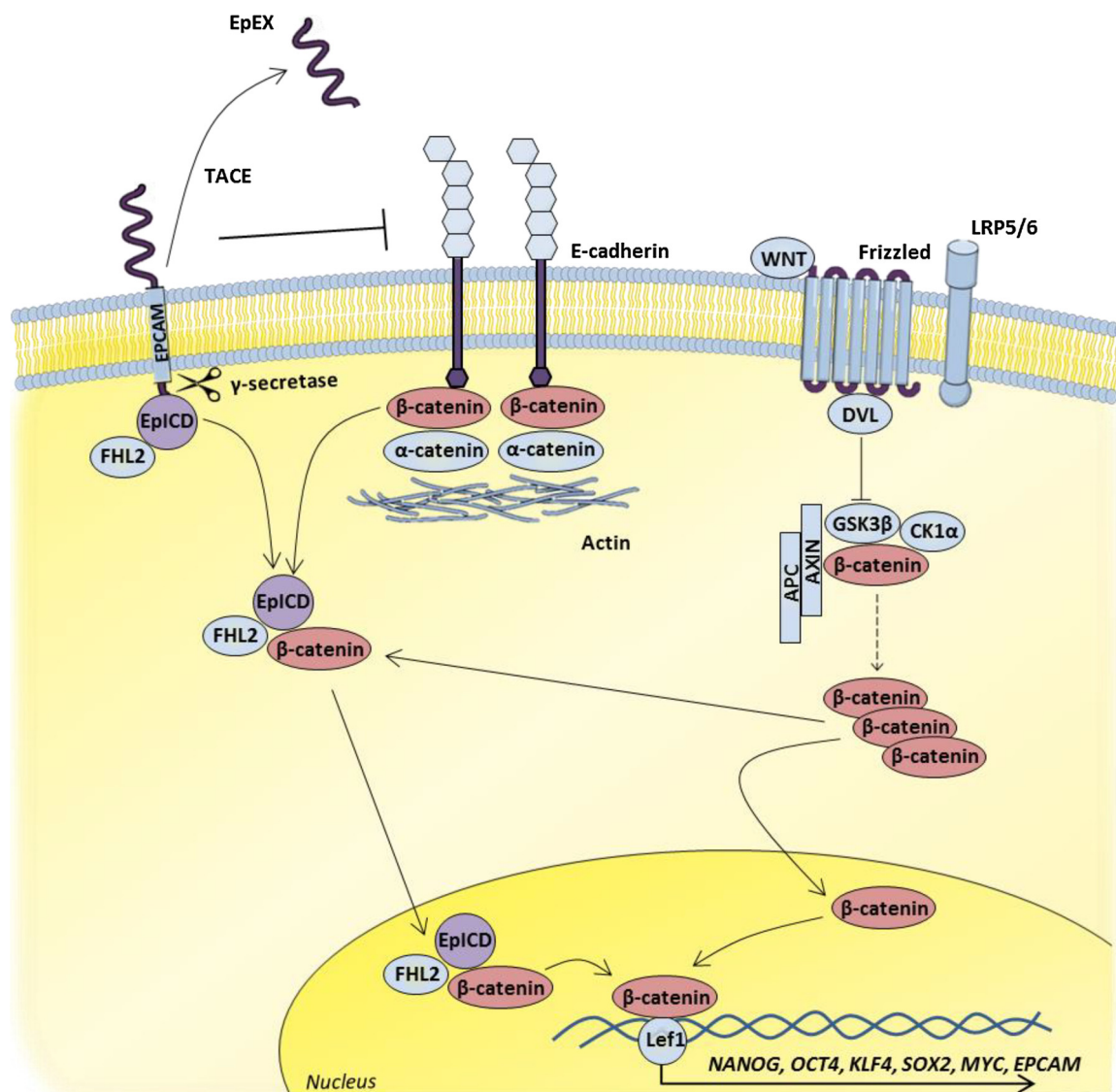


Fig. 1. Schematic representation of the cross-talk between EpCAM intracellular domain (EpICD), E-cadherin and the Wnt pathway. Following cleavage by γ -secretase, EpICD forms a multiprotein complex with β -catenin and the scaffolding protein FHL2. This complex may translocate to the nucleus where it binds Lymphoid Enhancer Factor 1 (Lef-1) and the promoters of genes related to stemness and cell cycle. It seems that EpICD is able to increase the level of soluble β -catenin inhibiting E-cadherin and protecting β -catenin from degradation, which enhances the Wnt signalling pathway.

generally crucial in both epithelial tissue morphogenesis and epithelial cell polarity maintenance; in addition, E-cadherin acts as a transcriptional repressor of Snail1, Twist, ZEB1/2 and Slug, and therefore plays a role as a suppressor of invasive phenotypes (Sanchez-Tillo et al., 2010). Elsewhere, EpCAM enhances the proliferation and invasive potential; indeed, EpCAM's effects on cadherin relate its expression level with invasion and metastasis in epithelial cancer. EpCAM mutation, which affects EpICD, has been reported to break the cadherin-adhesion complex, and suggests that EpICD negatively regulates the cell-cell interaction (Lin et al., 2012). Thus, the cells that undergo an EMT transition are subjected to a process in which epithelial cells lose their polarity, cell-cell contact and undergo cytoskeleton remodelling. Indeed, EpCAM promotes the invasion of breast cancer cells by interrupting E-cadherin, and consequently loss of cadherin could be the result of the inhibition by transcriptional repression through EMT-inducing genes, which act as repressors of E-cadherin (Lin et al., 2012). Besides, in parallel EpCAM and reprogramming factors, such as Oct4, Sox2, Nanog, and c-Myc, are frequently overexpressed (Lin et al., 2012). The nuclear translocation of EpICD regulates the expression of these stemness genes by exerting clear control in tumour cell self-renewal, hence it plays a critical role in

EpCAM signalling. EMT induction by EpCAM signalling promotes invasion and metastasis, but simultaneously induces cells with a stemness profile (Sankpal et al., 2017). Previous reports have demonstrated that EpCAM knockdown simultaneously inhibits the expression of pluripotency factors and EMT genes (Lin et al., 2012). Thus, EpCAM silencing leads to a cytoplasmic β -catenin decrease for the Wnt pathway and, consequently, the down-regulation of its target genes.

4. Relationship to cancer stem cells

Cancer Stem Cells (CSCs) have been repeatedly reported to rely on Wnt-inducing signals from their microenvironment to maintain their phenotype and, consequently, the EpCAM protein has been proposed as a strategic molecule in CSCs (Pirozzi et al., 2013). To date, a number of EpCAM-regulated target genes has been identified, including CD44, cyclin D1, and other key factors, such as c-Myc, Nanog, Klf4, Sox2 and Oct4, which are master regulators of pluripotency, and are critical for the conversion of somatic cells into induced pluripotent stem cells (iPSCs) (Chaves-Perez et al., 2013; Park et al., 2016). Thus, deregulation of EpCAM expression and signalling could trigger characteristic

traits of CSCs, such as proliferative potential and multipotency in cells. Some evidence that supports a central role of EpCAM in the CSCs phenotype has been reported. Kamrri-Busheri et al., found that the overexpression of *EPCAM1* and *ALDH1A1* can be used as a signature of enriched CSCs in the H460 cell line, and *in vivo* and *in vitro* experiments performed with lung primary cultures and cell lines identified CD133⁺/CD326⁺ and CD34⁺/CD326⁺ as the population that best represents CSCs in lung cancer (Karimi-Busheri et al., 2011). In line with these results, three compounds were identified which target CSCs by down-regulating the expression of *EPCAM1*, *CD44* and *ALCAM* (Soto-Cerrato et al., 2015).

Finally, the role of EpICD in tumour progression has been clearly described in epithelial tumours. Ralhan et al. analysed the expression of EpICD and EpEX immunohistochemically in 374 tumour samples from ten epithelial cancers, and compared them with 74 normal samples (Ralhan et al., 2010). The results showed increased nuclear and cytoplasmic EpICD accumulation and loss of membranous EpEX in all the epithelial cancers compared with normal tissues. Different studies into prostate cancer, hepatocellular carcinoma (HCC), oral squamous cell carcinoma (OSCC) and colorectal adenocarcinoma have reported a correlation between loss of membranous EpICD expression and poor prognosis in disease-free survival (DFS) and overall survival (OS) terms (Park et al., 2016; Ralhan et al., 2010). The clinical relevance of these results is discussed below (Prognostic role of EpCAM in cancer).

5. Regulation of EpCAM expression: genetics, transcription factors and epigenetics

The gene that encodes for EpCAM (*EPCAM*) is located on chromosome 2p21 with an estimated size of 14 kb. The *EPCAM* gene has nine coding exons. Exons 1–6 encode for EpCAM's extracellular domain, whereas the transmembrane region is encoded by exon 7, and the intracellular domain by exons 8 and 9. The mRNA of *EPCAM* is approximately 1.5 kb and encodes a protein of 314 amino acids with no splicing variants (van der Gun et al., 2010). Mutations in the *EPCAM* gene have been identified in patients who suffer Lynch syndrome (also known as hereditary non-polyposis colorectal cancer; HNPCC) or congenital tufting enteropathy (CTE). In the autosomal dominant disorder Lynch syndrome, heterozygous germline deletions in the 3' end of *EPCAM* lead not only to the inactivation of the adjacent *MSH2* gene through the hypermethylation of the *MSH2* promoter region, but also to loss of *MSH2* expression (Hegde et al., 2014). Nineteen different *EPCAM* deletions, all in the last two exons, have been identified in Lynch syndrome (Schnell et al., 2013a,b); Basal levels of EpCAM are remarkably high in normal colon. As a result, carriers of an *EPCAM* deletion eventually develop colorectal cancer, whereas tumour apparition in other locations is uncommon (Kempers et al., 2011). Histological studies have shown lack of EpCAM expression in a number of tumours in patients with *EPCAM* germline deletions (Schnell et al., 2013a,b). In these cases, EpCAM is only absent in the event of homozygous *EPCAM* deletion as a result of the combination of a germline and a second somatic deletion (Kloor et al., 2011). These results suggest that lack of EpCAM might not prevent tumour formation, and therapeutic strategies that target EpCAM should be considered in these situations. In autosomal recessive CTE, many different point mutations have been described, which have resulted in a single amino acid exchange, truncation or the partial deletion of the EpCAM protein (Tang et al., 2016). These mutations cause the protein from the plasma membrane to be absent, due to either the secretion or degradation of the truncated protein, or to the retention of the aberrant protein in the endoplasmic reticulum (Schnell et al., 2013a,b). The promoter region, which controls the expression of *EPCAM* as well as the sequence upstream of the transcription start site (TSS), has been completely characterised (Gires et al., 2003). It seems that promoter activity is restricted to the cells that express EpCAM, and a fragment of 3.4 kb of this regulatory sequence is capable of controlling heterologous gene expression (van der Gun et al.,

2010). Different transcription-binding sites have been described in the *EPCAM* promoter for Ets, Sp1, Lef, AP-1, nuclear factor-kappaB (NFκB) and p53. Primary evidence has been found for Ets and Sp1: the up-regulation of ESE-1 (epithelial-specific Ets-1) in metastatic lymph nodes from lung, breast and pancreas cancers are associated with the expression of EpCAM (van der Gun et al., 2010), whereas strong promoter activity has been detected in the presence of Sp1, compared with activity in the absence of Sp1 after transfection with an EpCAM promoter fragment presumably with binding sites for Sp1 (Tai et al., 2007). Regarding the Lef-1 transcription factor, it seems that EpICD exerts a positive-feedback loop on EpCAM expression at the gene transcription level since β-catenin activation induces EpCAM transcription via the binding of TCF/Lef 489 bp upstream of the EpCAM TSS. In contrast, NF-κB and p53 repress *epcam*. The repressive role of NF-κB was evidenced by cotransfection studies with an inhibitor plasmid for NF-κB and a luciferase reporter plasmid under the control of *EPCAM* promoter. In addition, it was found that treatment of EpCAM⁺ lung squamous cell carcinoma (SCCs) with TNF-α and interferon-α (INF-α) lowered *EPCAM* gene expression (Gires et al., 2003; van der Gun et al., 2010). Strong evidence exists for p53 regulation of the *EPCAM* gene. Ten possible binding sites for p53 in the *EPCAM* gene have been identified and the binding of p53 in intron 4 has been confirmed by chromatin immunoprecipitation (Sankpal et al., 2009). Moreover, induction of p53 negative correlates with *EPCAM* expression in a dose-dependent manner, whereas ablation of p53 expression correlates with significantly increased *EPCAM* levels (Sankpal et al., 2009). To demonstrate that the increase in EpCAM expression is caused by ablation of p53 expression, Sankpal et al. silenced p53 and EpCAM expression using a short hairpin RNA (shRNA), and were able to control breast cancer cell invasion *in vitro* (Sankpal et al., 2009).

Many authors have described the epigenetic regulation of EpCAM expression by DNA methylation and histone modifications. In lung, ovarian, colon and glioblastoma cancer cell lines, higher EpCAM levels correlate with the hypomethylation of a fragment of exon 1 and the proximal promoter, whereas lack of EpCAM expression correlates with hypermethylation (van der Gun et al., 2008). Remarkably, CpG at the proposed binding site for Sp1 has been found methylated in the EpCAM⁻ cell lines, in contrast to the EpCAM⁺ cell lines (Yu et al., 2008). Modulation of EpCAM by different strategies proves the relationship between the DNA methylation pattern of the EpCAM protein and mRNA expression. Treatment of the EpCAM⁻ cell lines with 5-aza-2'-deoxycytidine, a DNA demethylating agent, induces EpCAM expression and also causes increased EpCAM levels in the EpCAM⁺ cell lines (Tai et al., 2007; van der Gun et al., 2008), whereas the persistent down-regulation of EpCAM has been achieved via the methylation of its promoter (van der Gun et al., 2008). Likewise, after the delivery of DNA methyltransferase to EpCAM⁺ ovarian carcinoma cells, the methylation of the *EPCAM* promoter resulted in a silenced EpCAM expression through successive cell divisions (van der Gun et al., 2010, 2008). Gommans et al. designed trimeric and hexameric zinc finger proteins (ZFP) to target the *EPCAM* promoter, and were able to modulate the expression of *EPCAM* in colon, ovarian, astrocytoma and lung cancer cell lines (Gommans et al., 2007). Similarly, Nunna et al., used an engineered ZFP, which specifically binds the *EPCAM* promoter fused to the catalytic domain of Dnmt3a DNA methyltransferase (Nunna et al., 2014). Treatment of ovarian cancer cell line SKOV3 with this construct increased the methylation of the *EPCAM* promoter, resulting in 60–80% reduction in the EpCAM expression (Nunna et al., 2014). However, negative results have also been reported. Treatment of the hypermethylated leukaemia-derived K562 cell line and 50% of the CpGs methylated liver derived HepG2 cell line with 5-aza-2'-deoxycytidine had no effect on EpCAM expression, even though many methylated CpGs were converted into unmethylated (Yu et al., 2008). Moreover, no *de novo* re-expression of EpCAM has been detected in the lung carcinoma intermediated methylated EpCAM⁻ GLC-1 cell line after 5-aza-2'-deoxycytidine treatment (van der Gun et al., 2008). The results of in

in vitro studies correlate well with DNA methylation studies in patient samples, which suggests that the reported opposite results are due to differences between tumour types (van der Gun et al., 2010). A significant unmethylation of the CpGs islands was described in colon cancer when compared to normal tissue specimens (Yu et al., 2008). In some cases, the expression levels of EpCAM were up to 1000-fold higher in colon cancers than in normal tissue, which reflects the effect of promoter methylation on EpCAM expression. Recently, Kim et al., analysed a large cohort of 726 colorectal carcinoma patients and found that loss of EpCAM expression in tumour tissue significantly correlated with a CpG island methylation phenotype (Kim et al., 2016). Accordingly, EpCAM expression levels in the tissue from lung adenocarcinoma and oral squamous cell carcinoma correlated with the methylation status of the EpCAM promoter (Shiah et al., 2009). However, different situation has been found in other tumours like breast cancer, where no correlation was found (Spizzo et al., 2007).

Few studies have been published about the epigenetic regulation of EpCAM based on histone modifications (Lu et al., 2010; Tai et al., 2007; van der Gun et al., 2011). Tai et al. reported a clear relationship between epigenetics, EpCAM expression and invasiveness. They treated EpCAM⁻ cancer cells with 5-aza-2'-deoxycytidine and a histone deacetylase inhibitor, named trichostatin A, and found a reactivation of EpCAM expression and inhibited cancer invasiveness in both cases (Tai et al., 2007). In addition, chromatin immunoprecipitation in three lung cancer cell lines showed a correlation between low EpCAM expression and repressive histone modifications associated to the EpCAM promoter (Tai et al., 2007). Van der Gun et al. found a positive correlation between activating histone modifications (acH4, acH3 and H3K4me3) and EpCAM⁺ cells and these correlations were negative in the case of repressive histone modifications H3K9me3 and H3K27me3 (van der Gun et al., 2011). Similarly, Lu et al. reported the reduction of active histone marks and enhancement of repressive marks when silencing *EPCAM* gene in human embryonic stem cells (Lu et al., 2010).

6. Prognostic role of EpCAM in cancer

The prognostic value of EpCAM also reflects the dual role of this molecule in cancer. EpCAM expression levels increase in most epithelium-derived tumours, but the prognosis impact of high EpCAM expression levels varies between tumour types. Table 1 summarises the most relevant prognostic implications of EpCAM reported in cancer patients. Roughly speaking, high EpCAM expression levels have been reliably associated with better clinical outcome in three tumour types. In renal cell carcinomas (RCC), multiple studies have identified EpCAM expression as an independent prognostic factor for survival. EpCAM overexpression significantly correlates with a better prognosis in all the subtypes, where papillary RCC and chromophobe RCC are the histological subtypes associated with a higher EpCAM expression rate (75–90%) (Zimpfer et al., 2014; Eichelberg et al., 2013; Seligson et al., 2004). Remarkably, in a cohort of 767 clear cells RCC, Eichelberg et al., found that negative EpCAM expression is associated with high-grade disease and nodal metastases (Eichelberg et al., 2013). A positive impact of EpCAM expression on gastric cancer has also been described. A long-term study, including 280 patients, has concluded that patients with no loss of EpCAM expression have a significantly better 10-year survival rate when compared to the patients with some loss (Songun et al., 2005). The prognostic value of EpCAM seems stronger in stages I and II, with no significant correlation reported with metastases in any studies (Dai et al., 2017; Songun et al., 2005). Finally, EpCAM is highly expressed in colon tissue, and partial loss of EpCAM expression in colorectal cancer has often been associated with a poor prognosis. A cohort of 726 patients was analysed for EpCAM status, and the authors found that partial loss of EpCAM expression in 50 patients was an independent poor prognostic factor (Kim et al., 2016). Similarly, multivariate analyses in another study with 309 colorectal cancer patients identified EpCAM expression as an independent prognostic factor for

improved disease-specific survival (DSS) (Goossens-Beumer et al., 2014).

The positive correlation between EpCAM and survival is less clear in other tumour types for which studies with contrary results can be found. Akita et al., designed a study to determine the difference of the malignant potential between control and EpCAM-transfected pancreatic cancer cell lines (MiaPaCa-2, Panc1 and PSN1), and the clinical prognosis impact of tumoural EpCAM expression in a cohort of 95 pancreatic cancer patients (Akita et al., 2011). Only the PSN1 cell line showed significantly lower growth compared with the mock-transfected cells for the proliferation assay, but all the three cell lines significantly showed a reduction in invasion and migration after EpCAM-transfection. Regarding the prognosis impact, EpCAM^{high} patients showed a significantly better prognosis in terms of OS (3-year survival; 56.2% vs. 19.2%, $P = 0.0018$) and DFS (3-year survival; 40.3% versus 14.4%, $P = 0.038$) (Akita et al., 2011). However, opposing results with no statistical significance have also been reported. Fong and colleagues designed a retrospective study, including tissue samples from patients with pancreatic (N = 153) and ampullary cancer (N = 34). Although EpCAM expression failed to be an independent prognostic biomarker, subgroup analyses showed that EpCAM overexpression correlated with shorter OS among the patients with ampullary cancer and advanced stage pancreatic cancer (Fong et al., 2008). A study which used real-time PCR to detect *EPCAM* mRNA expression in the peripheral blood and peritoneal cavity of 48 patients who had undergone pancreatectomy for pancreatic ductal adenocarcinoma (PDAC) found no significant associations between EpCAM positivity (as surrogate marker for the quantification of circulating tumour cells) and the clinico-pathological variables (Sergeant et al., 2011). In lung cancer, a study that included 130 pulmonary adenocarcinoma (ADC) patients found that EpCAM expression protected patients from lymph metastasis and was associated with favourable outcome (Kobayashi et al., 2010). In line with these results, a study performed in 370 early-stage non-small cell lung cancer (NSCLC) patients found that high EpCAM expression is an independent prognostic factor for recurrence and survival (Gold et al., 2014). In contrast, two independent studies, performed in two cohorts of lung cancer patients (n = 234 ADC and n = 164 ADC and squamous cell carcinomas, respectively), found no significant correlation between EpCAM expression and survival (Kim et al., 2009; Pak et al., 2012). The results of a study published by Went et al., conducted with thousands of samples from colon, stomach, prostate and lung cancers, were also negative. This study included a cohort of 1287 lung cancer patients and only found a longer survival trend in patients with ADC, large cell and bronchioloalveolar carcinomas and a strong EpCAM expression, which was inversed in SCC (Went et al., 2006). On the contrary, a recent study done with 118 patients suffering from early-stage lung cancer has reported that EpCAM overexpression is a prognostic marker for unfavourable survival, which correlates with lymph node metastasis (Zhou et al., 2015). In epithelial ovarian cancer, contradictory results have been reported. EpCAM overexpression has been associated with a favourable prognosis in PFS and DSS terms in an unselected cohort of 117 patients (Battista et al., 2014). In fact, a recent study of 168 patients has revealed that EpCAM expression is an independent predictive biomarker of poor prognosis in terms of chemotherapeutic response (overall response rate (ORR): HR 11.12 [1.66–74.41], $p = 0.013$), PFS and OS (Tayama et al., 2017). Consistently with these clinical observations, with five ovarian cancer cell lines, researchers have found that the subpopulation of EpCAM⁺ ovarian cancer cells shows a significantly wider viability compared with EpCAM⁻ cells in response to cisplatin treatment (Tayama et al., 2017). Furthermore, platinum agents preferentially eliminate EpCAM⁻ cells compared to EpCAM⁺ cells in an *in vivo* mouse model (Tayama et al., 2017). Contradictory results have also been reported for head and neck cancers. The results reported by Hwang et al., show a significant association between a low EpCAM expression in a cohort of 84 oral SCC patients and lymph node metastasis, cancer recurrence and survival (Hwang et al., 2009). No

significant results were obtained in two studies (of 77 oral SCC patients and 294 supraglottic laryngeal carcinoma patients, respectively) associating EpCAM expression and clinico-pathologic features (Laimer et al., 2008; Bayram et al., 2015). In contrast, a study that included 48 patients with SCC of the tongue found that EpCAM expression was significantly associated with tumour size, regional lymph node metastasis, histological differentiation and invasion pattern (Yanamoto et al., 2007). These authors also reported that human tongue cancer cell lines with higher EpCAM expressions had more invasive potential. Accordingly, decreased invasion potential and proliferation were observed when EpCAM expression was inhibited using RNA interference (Yanamoto et al., 2007). In prostate cancer, a consistent association of EpCAM expression and clinico-pathological variables is lacking, and the Gleason score or survival rates in studies on large patients' cohorts have reported no significant results (Went et al., 2006).

A promoting role of EpCAM in carcinogenesis is much clearer in some other tumour types; e.g., breast, gall bladder, urothelial bladder, hepatocellular and oesophageal cancers. Many studies have established an association between EpCAM overexpression and a poor prognosis in different cohorts of breast cancer patients (Schmidt et al., 2008; Agboola et al., 2012; Soysal et al., 2013; Sadeghi et al., 2017). Schmidt et al., examined EpCAM expression in a cohort of 402 untreated node-negative breast cancer patients to find a significant association between high EpCAM levels and decreased DFS (Schmidt et al., 2008). Another study described the same prognostic role for EpCAM expression in a cohort of 726 breast cancer patients, especially for basal-like tumours, where EpCAM positivity was strongly associated with the probability of recurrence and distant metastases (Agboola et al., 2012). A larger study performed with 1365 patients confirmed the association of EpCAM expression with an unfavourable prognosis, and highlighted the poor prognosis of basal-like and luminal B HER2+ subtypes with increased EpCAM expression and the improved prognosis of the HER2 subtype (Soysal et al., 2013). Results obtained in gallbladder and urothelial bladder carcinomas have indicated that EpCAM expression predicts lower survival (Prince et al., 2008; Bryan et al., 2014). Bryan et al., analysed urinary EpCAM in 607 patients with primary bladder tumours and found that elevated urinary EpCAM is an independent indicator of poor prognosis (Bryan et al., 2014). In hepatocellular carcinoma (HCC), EpCAM expression has been associated with decreased survival, especially in advanced disease stages (Park et al., 2016), whereas Matsuda and Kimura et al., have associated EpCAM overexpression with poor survival in a study performed in cohorts of 74 and 138 patients suffering from oesophageal cancer (Matsuda et al., 2014; Kimura et al., 2007).

Remarkably, the detection methods used in most studies do not distinguish between EpEX and EpICD. Given the antagonistic role proposed for EpEX and EpICD, it is necessary to analyse them separately. First studies that differentiate EpEX and EpICD expressions have been performed and their results show that increased nuclear EpICD accumulation and loss of membranous EpEX predict poor survival in patients (Kunavisarut et al., 2012; Park et al., 2016; Kuan et al., 2017; Fong et al., 2014). We know that the main reason for cancer treatment failure is resistance to chemotherapy. This resistance of tumour cells to anticancer drugs has been associated with EpCAM overexpression. This is particularly true in epithelial tumours of breast, bladder and ovarian cancers, where the mechanism have been studied. It has been shown that EpCAM knockdown induces apoptosis, and produces cell-cycle arrest that inhibits cell proliferation and is associated with PI3K/Akt/mTOR signalling pathway inactivation (Gao et al., 2014). Thus apoptosis induction is enhanced by down-regulating the expression of anti-apoptotic protein B-cell lymphoma 2 (Bcl-2), and by the parallel up-regulation of the expression of pro-apoptotic proteins Bax and caspase3 (Gao et al., 2014). Evidence has demonstrated that an alternative to enhance chemosensitivity is possible through EpCAM knockdown. Since EpCAM overexpression clearly correlates with reduced OS and PFS, and consequently worse survival of patients suffering from these

tumour types, an EpCAM-targeting therapy to overcome chemoresistance and to achieve an effective cancer cytotoxic effect looks promising. Nevertheless, ongoing clinical trials results will be necessary to address whether the best strategy will be targeting EpCAM only in chemoresistant tumors or use this therapy alone or in combination.

7. Therapeutic role of EpCAM in cancer

EpCAM plays a role as a regulator of epithelia in normal tissue and can be described as a cancer stem cell marker since the CSCs with a higher EpCAM expression are considered more susceptible to metastasis. However, EpCAM plays a dual biological role in tumour progression, since it has been proposed to be an adhesion molecule, mediating cell-cell adhesion, and thus preventing metastasis and acting as a tumour suppressor. EpCAM has clear oncogenic properties and actively drives tumour progression through the cleavage of cell-cell adhesion mediated by E-cadherin to thus promote metastasis. EpCAM overexpression also supports proliferation and upregulates c-Myc, an oncogene that triggers the stem cells transcriptional profile in normal and cancer cells. Thus EpCAM overexpression through Wnt signalling enhancing is associated with less differentiated tumours, larger tumours, and poor patient prognosis (Gastl et al., 2000). Given its tumour-specific overexpression and its potential to target CSCs, EpCAM has been explored for three decades as an anti-cancer target (Fig. 2). Since then, EpCAM has emerged as an attractive target and many therapeutic methodologies, including immunotherapeutic strategies, vaccination approaches, RNA aptamers or engineered T-cells armed with chimeric antigen receptor (CAR-T) recognising EpCAM, have been developed, and many others are still being developed. Therapeutic strategies targeting EpCAM currently under development are summarized in Table 2.

For monoclonal antibodies, the first EpCAM targeting test in patients was a well-tolerated murine IgG2a antibody called edrecolomab (17-1A). Significant increase in the OS of colorectal cancer patients in the adjuvant was found in first clinical trials and pivotal studies (Riethmuller et al., 1994, 1998). However, subsequent larger studies could not confirm its clinical activity may be due to the short serum half-life of the antibody, and the rapid neutralisation by the human anti-mouse antibody response, which exposed the need to improve anti-EpCAM molecules (Schmoll and Arnold, 2009; Niedzwiecki et al., 2011). Subsequent studies have used very low doses of edrecolomab adsorbed on aluminium hydroxide as a subcutaneously administered vaccine (IGN101) in non-small cell lung cancer (NSCLC). Phase I studies have observed a significant reduction in circulating tumour cells, whereas Phase II trials have shown that the vaccine induces a strong immune response despite concomitant chemotherapy (Kirman, 2006; Loibner et al., 2004). However, a large double-blind placebo-controlled Phase II/III trial, with 762 NSCLC patients, found no oncological effect of this systemic adjuvant treatment type, and an overall survival of 2 years of 67.3% for placebo versus 59.7% for the IGN101-treated patients ($p = 0.07$) (Stoelben et al., 2008). Four other clinically tested anti-EpCAM monoclonal antibodies with different targets have been designed: a chimeric form of edrecolomab, humanised 3622W94, human-engineered ING-1, and the fully human adecatumumab. Remarkably, acute pancreatitis cases were detected after treatment with the highest affinities antibodies (3622W94 and ING-1), even when the concentrations used were low (1 mg/kg) (Münz et al., 2010). On the other hand, a phase I study in hormone-resistant prostate cancer patients reported minimal secondary effects, such as vomits or diarrhoea, after treatment with adecatumumab (MT201) at higher doses (2–6 mg/kg) (Oberneder et al., 2006). A few more Phase II studies with adecatumumab in monotherapy have been performed. The first randomised metastatic breast cancer patients according to high or low EpCAM expressions, and revealed that patients with high EpCAM levels treated with high-dose adecatumumab (6 mg/kg) had significantly less probability of tumour progression than patients with low EpCAM expression (Schmidt et al.,

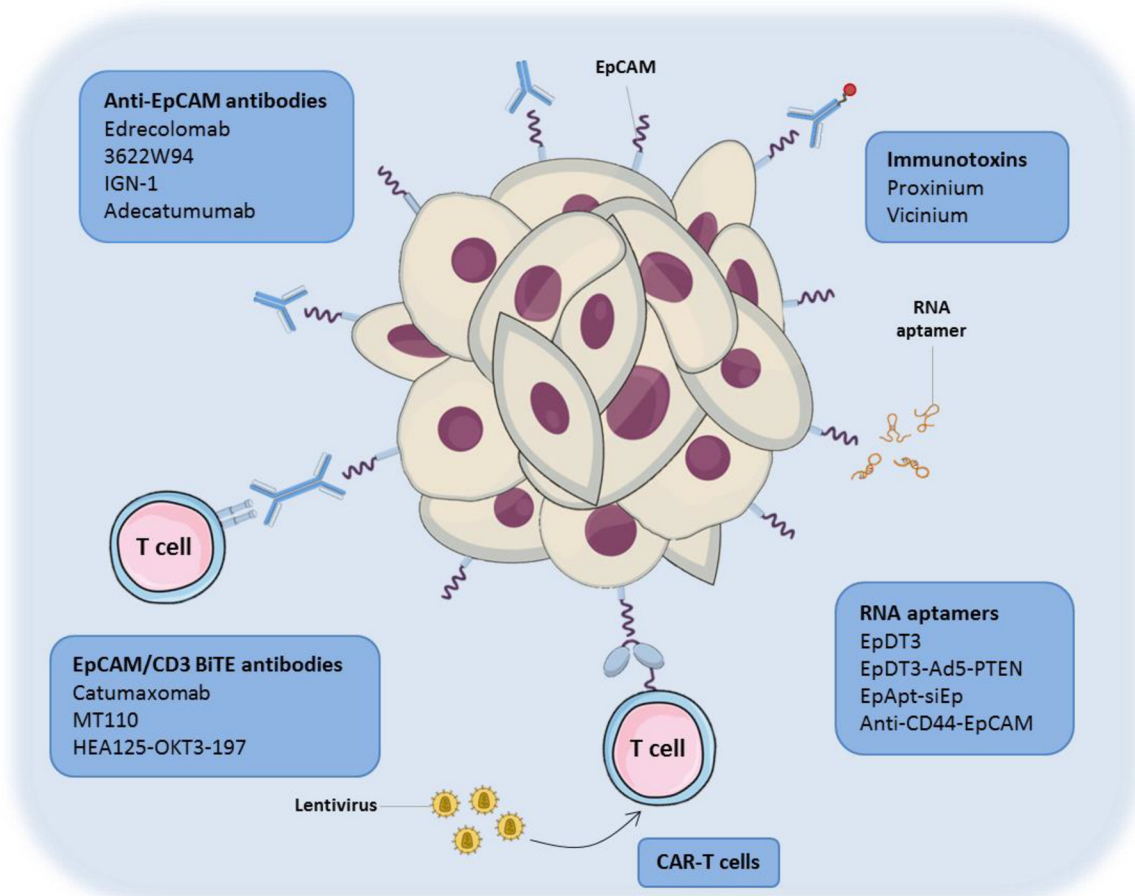


Fig. 2. Therapeutics approaches against epithelial tumours based on EpCAM expression. The therapeutic methodologies developed or under development include immunotherapeutic strategies, RNA aptamers and engineered T-cells armed with chimeric antigen receptor (CAR-T) recognising EpCAM.

2010). The second study randomised early-stage prostate cancer patients to low- (2 mg/kg) or high-dose (6 mg/kg) adecatumumab, or a placebo, and found that adecatumumab delayed disease progression in the subgroup of patients with baseline prostate-specific antigen (PSA) < 1 ng/mL and high EpCAM expression tumours (Marschner et al., 2010). Clinical trials that studied the safety and efficacy of the adecatumumab and standard chemotherapy combination have been performed. A Phase IB study that combined adecatumumab and docetaxel in patients with EpCAM⁺ relapsed or refractory advanced-stage breast cancer has determined that this combination is safe, feasible and potentially active (Schmidt et al., 2012). A randomised Phase II study (NCT00866944) to evaluate the efficacy and safety of adecatumumab alone or sequentially to FOLFOX 4 after R0 resection of colorectal liver metastases has found that the combination therapy is safe, with a 1-year DFS of 71.4% for the combination arm compared to 50% for the monotherapy with FOLFOX 4. Nevertheless, these results must be considered cautiously because of the study's small sample size (N = 35). Variants of monoclonal antibodies have been developed with promising results. Catumaxomab, a hybrid mouse IgG2a/rat IgG2b antibody that is trispecific for EpCAM, CD3 and, via the Fc γ receptor, that activates accessory cells such as macrophages, NK cells and DCs, obtained the market approval in Europe in 2009 to treat malignant ascites in cancer patients. Patients treated with catumaxomab receive an intraperitoneal infusion of 10 to 150 μ g 4–5 times over 9–13 days. Most patients develop a tolerable humoral response against catumaxomab due to the presence of the chimeric Fc γ domain, which evokes an immunogenic reaction that correlates with improved overall survival and controllable side effects (Jager et al., 2012). Clinical trials in Phases I and II are ongoing or have recently been completed in

different epithelial cancers (including ovarian, gastric, colon and breast cancers) and peritoneal carcinomatosis (Bokemeyer et al., 2015; Maus-Sorensen et al., 2015). Their results indicate a correlation between catumaxomab treatment and favourable clinical outcome. Another interesting molecule is MT110, a single-chain EpCAM/CD3-bispecific T-cell engaging (BiTE) antibody, which has been tested to target CSCs in preclinical and clinical studies of different cancer types (Cioffi and Heeschen, 2012). *In vitro* experiments in established primary cultures derived from colorectal cancer patients showed that the combination of the antibody with autologous T cells prevented tumour growth in soft agar assays (Herrmann et al., 2010). In addition, treatment with very low MT110 doses has eradicated established tumours in immunodeficient xenograft model, where mice were inoculated with colonies formed in soft agar assays (Herrmann et al., 2010). These results agree with those obtained for cell lines and primary cells of human pancreatic cancer, where the population of highly tumorigenic CSCs was efficiently targeted by MT110 *in vitro* and *in vivo* using a mouse model (Cioffi et al., 2012). Another study incubated adherent cells and spheroids from hepatic tumour cell lines with $\gamma\delta$ T-cells and MT110, and found that the antibody enhanced tumour cell lysis close to completion under both conditions (Hoh et al., 2013). A phase I clinical trial (NCT00635596) in patients with advanced solid tumours with high expression of EpCAM has already been performed, but results have not been published yet. Variations of the trispecific approach have been recently developed, among them, hybrid antibodies that combine anti-EpCAM antibody HEA125 with anti-CD3 antibody OKT3 or anti-CD64/Fc γ RI mAb 197. These have been used in small-scale clinical studies to treat ovarian cancer, and have significantly reduced malignant ascites and increased TNF- α (Souto et al., 2011). This scenario indicates strong

Table 2
EpCAM-directed therapeutic approaches in clinical development.

Therapeutic strategy	Class	Ongoing trials	Dosing	Company	Identifier
Catumaxomab	Trispecific EpCAM/CD3/Fc α antibody; mouse IgG2a/rat IgG2b hybrid	Phase II in gastric cancer with peritoneal carcinomatosis	i.p., escalating dose from 10–150 μ g	AIO-Studien-gGmbH	NCT01504256
Vicinium	single-chain antibody pseudomonas exotoxin fusion	Phase III in bladder cancer	intravesical, 30 mg once weekly	Viventia Bio	NCT02449239
EpCAM CAR-T	Autologous T cells engineered to recognize EpCAM	Phase I in nasopharyngeal carcinoma and breast cancer	i.v., escalating dose to severe toxicity	Sichuan University	NCT02915445
EpCAM CAR-T	Autologous T cells engineered to recognize EpCAM	Phase I/II in colon, esophageal, pancreatic, prostate, gastric and hepatic cancer	i.v., $1-10 \times 10^6$ EpCAM-CAR positive T cells kg^{-1}	First Affiliated Hospital of Chengdu Medical College	NCT03013712
EpCAM CAR-T	Autologous T cells engineered to recognize EpCAM	Phase II in liver cancer	i.v., five times: 0, 1, 2, 28 and 29 days	Sinobioway Cell Therapy Co., Ltd.	NCT02729493
EpCAM CAR-T	Autologous T cells engineered to recognize EpCAM	Phase II in gastric cancer	i.v., five times: 0, 1, 2, 28 and 29 days	Sinobioway Cell Therapy Co., Ltd.	NCT02725125

EpCAM = epithelial cell adhesion molecule; i.p = intraperitoneal; i.v = intravenous.

local immune stimulation. Some EpCAM-specific immunotoxins are also being used in clinical trials. Proximum is a promising humanised anti-EpCAM antibody with a Pseudomonas exotoxin (Brown et al., 2014). A first phase I/II study with proximum for the treatment of local head and neck cancers provided an 88% of stable diseases or responses, 25% of them with a complete remission of the disease (Baeuerle and Gires, 2007). However, subsequent Phase II trial (NCT00272181) for treating patients with squamous cell head and neck cancers, as well as a pivotal Phase II/III trial (NCT00412776) to compare best supportive care versus proximum plus the best supportive care in patients with advanced head and neck cancers, were discontinued due to slow accrual in North America, and also for corporate reasons unrelated to safety and efficacy. The same antibody construct, named Vicinium for use in bladder cancer, has shown satisfactory results in a Phase II study in patients with urothelial carcinoma *in situ* of the bladder, where 44% of patients achieved a complete response after treatment with 30 mg/dose for 6–12 weeks of intravesical vicinium (Kowalski et al., 2012). A Phase III study (NCT02449239) is currently ongoing for patients with non-muscle invasive bladder cancer previously treated with bacillus Calmette-Guerin.

RNA aptamers against EpCAM represent an attractive therapeutic strategy. The first aptamer against EpCAM was designed by Shigdar et al. and was called EpDT3 (Shigdar et al., 2011). It is a 19-nt RNA aptamer that interacts EpCAM with moderate affinity ($K'd = 55$ nM, 362W94 and ING-1 are approximately $K'd = 0.1$ nM, whereas adecatumumab is $K'd = 91$ nM) and is internalized after binding to cell surface. As a result, it is expected to produce manageable toxicities and to be able to channel diverse substances (Shigdar et al., 2011). Many variants of EpCAM aptamers are being developed and offer very encouraging preclinical results. Using aptamer-siRNA chimeras, the knockdown of EpCAM was able to inhibit epithelial cancer and CSCs *in vitro* and *in vivo* (Gilboa-Geffen et al., 2015). The conjugation of EpDT3 with Ad5-PTEN has been reported to enhance the antitumor effect of Ad5-PTEN with a high binding ability to hepatocellular carcinoma EpCAM⁺ cells (Xiao et al., 2017). Finally, Zheng et al. targeted a panel of ovarian cancer cell lines with a bispecific aptamer from fusing single CD44 and EpCAM aptamers and tested the results in a xenograft model. They found that the bispecific CD44-EpCAM aptamer had greater ability to inhibit cell growth and to induce apoptosis than single aptamers against CD44 or EpCAM alone or combined. *In vivo* experiments confirmed the higher ability to suppress tumours (Zheng et al., 2017).

At present, most ongoing clinical trials are focused on determining the safety and efficacy of CAR-T cells by recognising EpCAM (Table 2). In these studies, PBMC is obtained from blood samples and is activated *via* the antibodies of CD3 and CD28. Autologous T-cells are then transduced by the lentivirus carrying the EpCAM CAR gene and infused in patients. Preclinical studies using anti-EpCAM CAR in xenograft models of prostate cancer and peritoneal carcinomatosis have reported that a single intraperitoneal injection of the CAR-T cells inhibits tumour growth and prolongs mouse survival, and has even eradicated tumours in some cases (Deng et al., 2015). Based on these results, four clinical trials are currently ongoing. Two single-arm, multicentre Phase II studies are recruiting patients with relapsed or refractory liver (NCT02729493) and stomach (NCT02725125) cancers to evaluate the efficacy and safety of EpCAM-targeted CAR-T cells. Another Phase I study (NCT02915445) in EpCAM⁺ recurrent or refractory nasopharyngeal carcinoma and breast cancer patients, and a Phase I/II one (NCT03013712) for EpCAM positive cancers (colon, oesophageal, pancreatic, prostate, gastric or hepatic carcinomas) are ongoing, which indicates the great potential that CAR-T-cells therapy can have in cancer treatment.

Finally, based on recent advances on our knowledge about the role of EpICD in the tumorigenesis of cancer, new approaches have been proposed targeting different componets of the EpCAM signalling cascade. First studies targeting EpCAM were unable to distinguish between full-length and cleaved EpCAM and, in some cases, lacked patient

randomisation according to the actual EpCAM status on tumour cells can explain the disparate results sometimes obtained. Knowledge of the proteases involved in the activating proteolytic cleavage of EpCAM allows the systematic testing of combinations of the inhibitors of these proteases; *i.e.* TACE and presenilin 2, therapeutic antibodies, and conventional treatments that involve chemotherapeutics. Even more interesting could be the inhibition of the EpICD-FHL2 interaction by small molecules and their interaction with Wnt components, which is a promising strategy to specifically target the oncogenic role of EpCAM in cancer.

8. Conclusions

EpCAM molecule concept has doubtlessly evolved over time. Initially it was recognised as a cell adhesion molecule, but contrasting evidence has demonstrated that the participation of its intracellular domain is crucial in cellular malignant transformation, and in the regulation of pluripotency and self-renewal in both normal and cancer stem cells. The discovery of the antagonistic role between the extracellular and intracellular domains of EpCAM could explain the conflicting results that have been frequently reported about the prognostic role of EpCAM in cancer, which renders it necessary to analyse each domain separately. Apart from this involvement in cell cycle and stemness regulation, interest in this molecule has grown in recent years because it is expressed highly in a wide variety of solid cancers, and it is possible to use it to isolate CTCs and cancer cells with CSCs properties. Consequently, many therapeutic strategies are focus on EpCAM as a molecular target, and ongoing clinical trials employ anti-EpCAM immunotherapeutic approaches and CAR-T cells. Moreover, the development of therapeutic approaches that target the cleavage control of the EpCAM intracellular domain clearly seems promising.

Conflicts of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements

This work was supported by Fondo de Investigación Sanitaria, ISCIII (PI15-00753, PI12-02838) and a grant from the Fondo Europeo de Desarrollo Regional (FEDER).

References

Agboola, A.J., Paish, E.C., Rakha, E.A., Powe, D.G., Macmillan, R.D., Ellis, I.O., et al., 2012. EpCAM expression is an indicator of recurrence in basal-like breast cancer. *Breast Cancer Res. Treat.* 133 (June (2)), 575–582.

Akita, H., Nagano, H., Takeda, Y., Eguchi, H., Wada, H., Kobayashi, S., et al., 2011. EpCAM is a significant prognostic factor in pancreatic cancer patients by suppressing cell activity. *Oncogene* 30 (August (31)), 3468–3476.

Baeuerle, P.A., Gires, O., 2007. EpCAM (CD326) finding its role in cancer. *Br. J. Cancer* 96 (February (3)), 417–423.

Battista, M.J., Cotarello, C., Jakobi, S., Steetskamp, J., Makris, G., Sicking, I., et al., 2014. Overexpression of epithelial cell adhesion molecule protein is associated with favorable prognosis in an unselected cohort of ovarian cancer patients. *J. Cancer Res. Clin. Oncol.* 140 (July (7)), 1097–1102.

Bayram, A., Yuca, I., Cagli, S., Canoz, O., Guney, E., 2015. Predictive value of E-cadherin and Ep-CAM in cervical lymph node metastasis of supraglottic larynx carcinoma. *Am. J. Otolaryngol.* 36 (6), 736–740.

Bokemeyer, C., Stein, A., Ridwelski, K., Atanackovic, D., Arnold, D., Woll, E., et al., 2015. A phase II study of catumaxomab administered intra- and postoperatively as part of a multimodal approach in primarily resectable gastric cancer. *Gastric Cancer* 18 (October (4)), 833–842.

Brown, J., Cizeau, J., Bosc, D., Entwistle, J., Glover, N., MacDonald, G.C., 2014. A pre-clinical profile of proxiniumTM, a recombinant immunotoxin for targeting head and neck cancer. *Cancer Res.* 65 (October (9 Suppl.)), 162 LP-162.

Bryan, R.T., Shimwell, N.J., Wei, W., Devall, A.J., Pirrie, S.J., James, N.D., et al., 2014. Urinary EpCAM in urothelial bladder cancer patients: characterisation and evaluation of biomarker potential. *Br. J. Cancer* 110 (February (3)), 679–685.

Chaves-Perez, A., Mack, B., Maetzel, D., Kremling, H., Eggert, C., Harreus, U., et al., 2013. EpCAM regulates cell cycle progression via control of cyclin D1 expression. *Oncogene*

32 (January (5)), 641–650.

Cioffi, M., Heeschen, C., 2012. Immuno-targeting of pancreatic cancer stem cells: A new therapeutic strategy against a devastating disease? *Oncoimmunology* 1 (July (4)), 560–562.

Cioffi, M., Dorado, J., Baeuerle, P.A., Heeschen, C., 2012. EpCAM/CD3-bispecific T-cell engaging antibody MT110 eliminates primary human pancreatic cancer stem cells. *Clin. Cancer Res.* 18 (January (2)), 465–474.

Dai, M., Yuan, F., Fu, C., Shen, G., Hu, S., Shen, G., 2017. Relationship between epithelial cell adhesion molecule (EpCAM) overexpression and gastric cancer patients: A systematic review and meta-analysis. *PLoS One* 12 (4), e0175357.

Deng, Z., Wu, Y., Ma, W., Zhang, S., Zhang, Y.-Q., 2015. Adoptive T-cell therapy of prostate cancer targeting the cancer stem cell antigen EpCAM. *BMC Immunol.* 16 (January), 1.

Denzel, S., Maetzel, D., Mack, B., Eggert, C., Barr, G., Gires, O., 2009. Initial activation of EpCAM cleavage via cell-to-cell contact. *BMC Cancer* 9 (November), 402.

Eichelberg, C., Chun, F.K., Bedke, J., Heuer, R., Adam, M., Moch, H., et al., 2013. Epithelial cell adhesion molecule is an independent prognostic marker in clear cell renal carcinoma. *Int. J. Cancer* 132 (12), 2948–2955.

Fong, D., Steurer, M., Obrist, P., Barbieri, V., Margreiter, R., Amberger, A., et al., 2008. Ep-CAM expression in pancreatic and ampullary carcinomas: frequency and prognostic relevance. *J. Clin. Pathol.* 61 (1), 31–35.

Fong, D., Moser, P., Kasal, A., Seeber, A., Gastl, G., Martowicz, A., et al., 2014. Loss of membranous expression of the intracellular domain of EpCAM is a frequent event and predicts poor survival in patients with pancreatic cancer. *Histopathology* 64 (April (5)), 683–692.

Gao, J., Yan, Q., Liu, S., Yang, X., 2014. Knockdown of EpCAM enhances the chemosensitivity of breast cancer cells to 5-fluorouracil by downregulating the anti-apoptotic factor Bcl-2. *PLoS One.* 9 (July (7)), e102590.

Gastl, G., Spizzo, G., Obrist, P., Dunser, M., Mikuz, G., 2000. Ep-CAM overexpression in breast cancer as a predictor of survival. *Lancet* 356, 1981–1982.

Gilboa-Geffen, A., Hamar, P., Le MTN, Wheeler L.A., Trifonova, R., Petrocca, F., et al., 2015. Gene knockdown by EpCAM aptamer-siRNA chimeras suppresses epithelial breast cancers and their tumor-initiating cells. *Mol. Cancer Ther.* 14 (October (10)), 2279–2291.

Gires, O., Eskofier, S., Lang, S., Zeidler, R., Munz, M., 2003. Cloning and characterisation of a 1.1 kb fragment of the carcinoma-associated epithelial cell adhesion molecule promoter. *Anticancer Res.* 23 (4), 3255–3261.

Gold, K.A., Kim, E.S., Liu, D.D., Yuan, P., Behrens, C., Solis, L.M., et al., 2014. Prediction of survival in resected non-small cell lung cancer using a protein expression-based risk model: implications for personalized chemoprevention and therapy. *Clin. Cancer Res.* 20 (April (7)), 1946–1954.

Gommans, W.M., McLaughlin, P.M.J., Lindhout, B.I., Segal, D.J., Wiegman, D.J., Haisma, H.J., et al., 2007. Engineering zinc finger protein transcription factors to down-regulate the epithelial glycoprotein-2 promoter as a novel anti-cancer treatment. *Mol. Carcinog.* 46 (May (5)), 391–401.

Goossens-Beumer, I.J., Zeestraten, E.C.M., Benard, A., Christen, T., Reimers, M.S., Keijzer, R., et al., 2014. Clinical prognostic value of combined analysis of Aldh1, survivin, and EpCAM expression in colorectal cancer. *Br. J. Cancer* 110 (June (12)), 2935–2944.

Gosens, M.J.E.M., van Kempen, L.C.L., van de Velde, C.J.H., van Krieken, J.H.J.M., Nagtegaal, I.D., 2007. Loss of membranous Ep-CAM in budding colorectal carcinoma cells. *Mod. Pathol.* 20 (February (2)), 221–232.

Hegde, M., Ferber, M., Mao, R., Samowitz, W., Ganguly, A., a Working Group of the American College of Medical Genetics and Genomics (ACMG) Laboratory Quality Assurance Committee, 2014. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genet. Med.* 16 (January (1)), 101–116.

Herrmann, I., Baeuerle, P.A., Friedrich, M., Murr, A., Filusch, S., Ruttinger, D., et al., 2010. Highly efficient elimination of colorectal tumor-initiating cells by an EpCAM/CD3-bispecific antibody engaging human T cells. *PLoS One* 5 (October (10)), e13474.

Hoh, A., Dewerth, A., Vogt, F., Wenz, J., Baeuerle, P.A., Warmann, S.W., et al., 2013. The activity of gammadelta T cells against paediatric liver tumour cells and spheroids in cell culture. *Liver Int.* 33 (January (1)), 127–136.

Hwang, E.Y., Yu, C.-H., Cheng, S.-J., Chang, J.Y., Chen, H.-M., Chiang, C.-P., 2009. Decreased expression of Ep-CAM protein is significantly associated with the progression and prognosis of oral squamous cell carcinomas in Taiwan. *J. Oral Pathol. Med.* 38 (January (1)), 87–93.

Jager, M., Schoberth, A., Ruf, P., Hess, J., Hennig, M., Schmalfeldt, B., et al., 2012. Immunomonitoring results of a phase II/III study of malignant ascites patients treated with the trifunctional antibody catumaxomab (anti-EpCAM x anti-CD3). *Cancer Res.* 72 (January (1)), 24–32.

Karabulut, S., Tas, F., Tastekin, D., Karabulut, M., Yasasever, C.T., Ciftci, R., et al., 2014. The diagnostic, predictive, and prognostic role of serum epithelial cell adhesion molecule (EpCAM) and vascular cell adhesion molecule-1 (VCAM-1) levels in breast cancer. *Tumour Biol.* 35 (September (9)), 8849–8860.

Karimi-Busheri, F., Zadorozhny, V., Li, T., Lin, H., Shawler, D.L., Fakhrai, H., 2011. Pivotal role of CD38 biomarker in combination with CD24, EpCAM, and ALDH for identification of H460 derived lung cancer stem cells. *J. Stem Cells* 6 (1), 9–20.

Kempers, M.J.E., Kuiper, R.P., Ockeloen, C.W., Chappuis, P.O., Hutter, P., Rahner, N., et al., 2011. Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. *Lancet Oncol.* 12 (1), 49–55.

Khosla, R., Rastogi, A., Ramakrishna, G., Pamecha, V., Mukhopadhyay, A., Vasudevan, M., et al., 2017. EpCAM + liver cancer stem-like cells exhibiting autocrine Wnt signaling potentially originate in cirrhotic patients. *Stem Cells Transl. Med.* 6 (March (3)), 807–818.

Kim, Y., Kim, H.S., Cui, Z.Y., Lee, H.-S., Ahn, J.S., Park, C.K., et al., 2009. Clinicopathological implications of EpCAM expression in adenocarcinoma of the

- lung. *Anticancer Res.* 29 (May (5)), 1817–1822.
- Kim, J.H., Bae, J.M., Song, Y.S., Cho, N.-Y., Lee, H.S., Kang, G.H., 2016. Clinicopathologic, molecular, and prognostic implications of the loss of EPCAM expression in colorectal carcinoma. *Oncotarget* 7 (March (12)), 13372–13387.
- Kimura, H., Kato, H., Faried, A., Sohda, M., Nakajima, M., Fukai, Y., et al., 2007. Prognostic significance of EpCAM expression in human esophageal cancer. *Int. J. Oncol.* 30 (January (1)), 171–179.
- Kirman, I., 2006. Drug evaluation: IG101—an anti-EpCAM murine antibody vaccine for cancer. *Curr. Opin. Mol. Ther.* 8 (August (4)), 358–365.
- Kloor, M., Voigt, A.Y., Schackert, H.K., Schirmacher, P., von Knebel Doeberitz, M., Blaker, H., 2011. Analysis of EPCAM protein expression in diagnostics of Lynch syndrome. *J. Clin. Oncol.* 29 (January (2)), 223–227.
- Kobayashi, H., Minami, Y., Anami, Y., Kondou, Y., Iijima, T., Kano, J., et al., 2010. Expression of the GA733 gene family and its relationship to prognosis in pulmonary adenocarcinoma. *Virchows Arch.* 457 (July (1)), 69–76.
- Kowalski, M., Guindon, J., Brazas, L., Moore, C., Entwistle, J., Cizeau, J., et al., 2012. A phase II study of oportuzumab monatox: an immunotoxin therapy for patients with noninvasive urothelial carcinoma in situ previously treated with bacillus Calmette-Guerin. *J. Urol.* 188 (November (5)), 1712–1718.
- Kuan, I.-L., Liang, K.-H., Wang, Y.-P., Kuo, T.-W., Y.-J.J. Meir, S.-Y. Wu, et al., 2017. EpEX/EpCAM and Oct4 or Klf4 alone are sufficient to generate induced pluripotent stem cells through STAT3 and HIF2alpha. *Sci. Rep.* 7 (February), 41852.
- Kuechlin, S., Schoels, M., Slanchev, K., Lassmann, S., Walz, G., Yakulov, T.A., 2017. EpCAM controls morphogenetic programs during zebrafish pronephros development. *Biochem Biophys Res. Commun.* 487 (May (2)), 209–215.
- Kunavisarut, T., Kak, I., MacMillan, C., Ralhan, R., Walfish, P.G., 2012. Immunohistochemical analysis based Ep-ICD subcellular localization index (ESLI) is a novel marker for metastatic papillary thyroid microcarcinoma. *BMC Cancer* 12 (November), 523.
- Laimer, K., Fong, D., Gastl, G., Obrist, P., Kloss, F., Tuli, T., et al., 2008. EpCAM expression in squamous cell carcinoma of the oral cavity: frequency and relationship to clinicopathologic features. *Oral Oncol.* 44 (January (1)), 72–77.
- Lin, C.-W., Liao, M.-Y., Lin, W.-W., Wang, Y.-P., Lu, T.-Y., Wu, H.-C., 2012. Epithelial cell adhesion molecule regulates tumor initiation and tumorigenesis via activating reprogramming factors and epithelial-mesenchymal transition Gene expression in colon cancer. *J. Biol. Chem.* 287 (November (47)), 39449–39459.
- Litvinov, S.V., Velders, M.P., Bakker, H.A., Fleuren, G.J., Warnaar, S.O., 1994. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell. Biol.* 125 (April (2)), 437–446.
- Litvinov, S.V., Balzar, M., Winter, M.J., Bakker, H.A.M., Bruijn, I.H.B., Prins, F., et al., 1997. Epithelial Cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins. *J. Cell. Biol.* 139 (December (5)), 1337–1348.
- Liu, X., Gao, J., Sun, Y., Zhang, D., Liu, T., Yan, Q., et al., 2017. Mutation of N-linked glycosylation in EpCAM affected cell adhesion in breast cancer cells. *Biol. Chem.* 398 (September (10)), 1119–1126.
- Loibner, H., Eckert, H., Eller, N., Groiss, F., Himmler, G., Rosenkaimer, F., et al., 2004. A randomized placebo-controlled phase II study with the cancer vaccine IG101 in patients with epithelial solid organ tumors (IG101/2–01). *J. Clin. Oncol.* 22 (July (14, Suppl.)), 2619.
- Lu, T.-Y., Lu, R.-M., Liao, M.-Y., Yu, J., Chung, C.-H., Kao, C.-F., et al., 2010. Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. *J. Biol. Chem.* 285 (March (12)), 8719–8732.
- Marschner, N., Ruttinger, D., Zugmaier, G., Nemere, G., Lehmann, J., Obrist, P., et al., 2010. Phase II study of the human anti-epithelial cell adhesion molecule antibody adecatumumab in prostate cancer patients with increasing serum levels of prostate-specific antigen after radical prostatectomy. *Urol. Int.* 85 (4), 386–395.
- Matsuda, T., Takeuchi, H., Matsuda, S., Hiraiwa, K., Miyasho, T., Okamoto, M., et al., 2014. EpCAM, a potential therapeutic target for esophageal squamous cell carcinoma. *Ann. Surg. Oncol.* 21 (June (Suppl. 3)), S356–S364.
- Mau-Sorensen, M., Dittrich, C., Dienstmann, R., Lassen, U., Buchler, W., Martinius, H., et al., 2015. A phase I trial of intravenous catumaxomab: a bispecific monoclonal antibody targeting EpCAM and the T cell coreceptor CD3. *Cancer Chemother. Pharmacol.* 75 (May (5)), 1065–1073.
- Münz, M., Murr, A., Kvesic, M., Rau, D., Mangold, S., Pflanz, S., et al., 2010. Side-by-side analysis of five clinically tested anti-EpCAM monoclonal antibodies. *Cancer Cell Int.* 10 (November), 44.
- Ng, V.Y., Ang, S.N., Chan, J.X., Choo, A.B.H., 2010. Characterization of epithelial cell adhesion molecule as a surface marker on undifferentiated human embryonic stem cells. *Stem Cells* 28 (January (1)), 29–35.
- Ni, J., Cozzi, P., Hao, J., Beretov, J., Chang, L., Duan, W., et al., 2013. Epithelial cell adhesion molecule (EpCAM) is associated with prostate cancer metastasis and chemo/radioreistance via the PI3K/Akt/mTOR signaling pathway. *Int. J. Biochem. Cell. Biol.* 45 (December (12)), 2736–2748.
- Niedzwiecki, D., Bertagnolli, M.M., Warren, R.S., Compton, C.C., Kemeny, N.E., Benson 3rd, A.B., et al., 2011. Documenting the natural history of patients with resected stage II adenocarcinoma of the colon after random assignment to adjuvant treatment with edrecolomab or observation: results from CALGB 9581. *J. Clin. Oncol.* 29 (August (23)), 3146–3152.
- Nunna, S., Reinhardt, R., Ragozin, S., Jeltsch, A., 2014. Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells. *Tost J, editor. PLoS One* 9 (January (1)), e87703.
- Oberneder, R., Weckerhann, D., Ebner, B., Quadt, C., Kirching, P., Raum, T., et al., 2006. A phase I study with adecatumumab, a human antibody directed against epithelial cell adhesion molecule, in hormone refractory prostate cancer patients. *Eur. J. Cancer* 42 (October (15)), 2530–2538.
- Pak, M.G., Shin, D.H., Lee, C.H., Lee, M.K., 2012. Significance of EpCAM and TROP2 expression in non-small cell lung cancer. *World J. Surg. Oncol.* 10 (April), 53.
- Park, S.Y., Bae, J.S., Cha, E.J., Chu, H.H., Sohn, J.S., Moon, W.S., 2016. Nuclear EpICD expression and its role in hepatocellular carcinoma. *Oncol. Rep.* 36 (July (1)), 197–204.
- Pavsic, M., Guncar, G., Djinovic-Carugo, K., Lenarcic, B., 2014. Crystal structure and its bearing towards an understanding of key biological functions of EpCAM. *Nat. Commun.* 5 (August), 4764.
- Philip, R., Heiler, S., Mu, W., Buchler, M.W., Zoller, M., Thuma, F., 2015. Claudin-7 promotes the epithelial-mesenchymal transition in human colorectal cancer. *Oncotarget* 6 (February (4)), 2046–2063.
- Pirozzi, G., Tirino, V., Camerlingo, R., La Rocca, a, Martucci, N., Scognamiglio, G., et al., 2013. Prognostic value of cancer stem cells, epithelial-mesenchymal transition and circulating tumor cells in lung cancer. *Oncol. Rep.* 29, 1763–1768.
- Prince, S., Zeidman, A., Dekel, Y., Ram, E., Koren, R., 2008. Expression of epithelial cell adhesion molecule in gallbladder carcinoma and its correlation with clinicopathologic variables. *Am. J. Clin. Pathol.* 129 (March (3)), 424–429.
- Ralhan, R., He, H.C.-H., So, A.K.-C., Tripathi, S.C., Kumar, M., Hasan, M.R., et al., 2010. Nuclear and cytoplasmic accumulation of EpCAM is frequently detected in human epithelial cancers. *Vooijs M, editor. PLoS One* 5 (November (11)), e14130.
- Riethmuller, G., Schneider-Gadicke, E., Schlimok, G., Schmiegel, W., Raab, R., Hoffken, K., et al., 1994. Randomised trial of monoclonal antibody for adjuvant therapy of resected Dukes' C colorectal carcinoma. *Lancet* 343 (May (8907)), 1177–1183.
- Riethmuller, G., Holz, E., Schlimok, G., Schmiegel, W., Raab, R., Hoffken, K., et al., 1998. Monoclonal antibody therapy for resected Dukes' C colorectal cancer: seven-year outcome of a multicenter randomized trial. *J. Clin. Oncol.* 16 (May (5)), 1788–1794.
- Rupp, A.-K., Rupp, C., Keller, S., Brase, J.C., Ehehalt, R., Fogel, M., et al., 2011. Loss of EpCAM expression in breast cancer derived serum exosomes: role of proteolytic cleavage. *Gynecol. Oncol.* 122 (August (2)), 437–446.
- Sadeghi, S., Hojati, Z., Tabatabaieian, H., 2017. Cooverexpression of EpCAM and c-myc genes in malignant breast tumours. *J. Genet.* 96 (March (1)), 109–118.
- Sanchez-Tillo, E., Lazaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E.C., Castells, A., et al., 2010. ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene* 29 (June (24)), 3490–3500.
- Sankpal, N.V., Willman, M.W., Fleming, T.P., Mayfield, J.D., Gillanders, W.E., 2009. Transcriptional repression of epithelial cell adhesion molecule contributes to p53 control of breast cancer invasion. *Cancer Res.* 69 (January (3)), 753–757.
- Sankpal, N.V., Fleming, T.P., Sharma, P.K., Wiedner, H.J., Gillanders, W.E., 2017. A double-negative feedback loop between EpCAM and ERK contributes to the regulation of epithelial-mesenchymal transition in cancer. *Oncogene* 36, 3706–3717.
- Santesteban, M., Reiman, J.M., Asiedu, M.K., Behrens, M.D., Nassar, A., Kalli, K.R., et al., 2009. Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. *Cancer Res.* 69 (April (7)), 2887–2895.
- Schmidt, M., Hasenclever, D., Schaefer, M., Boehm, D., Cotarello, C., Steiner, E., et al., 2008. Prognostic effect of epithelial cell adhesion molecule overexpression in untreated node-negative breast cancer. *Clin. Cancer Res.* 14 (September (18)), 5849–5855.
- Schmidt, M., Scheulen, M.E., Dittrich, C., Obrist, P., Marschner, N., Dirix, L., et al., 2010. An open-label, randomized phase II study of adecatumumab, a fully human anti-EpCAM antibody, as monotherapy in patients with metastatic breast cancer. *Ann. Oncol.* 21 (February (2)), 275–282.
- Schmidt, M., Ruttinger, D., Sebastian, M., Hanusch, C.A., Marschner, N., Baeuerle, P.A., et al., 2012. Phase IB study of the EpCAM antibody adecatumumab combined with docetaxel in patients with EpCAM-positive relapsed or refractory advanced-stage breast cancer. *Ann. Oncol.* 23 (September (9)), 2306–2313.
- Schmoll, H.-J., Arnold, D., 2009. When wishful thinking leads to a misty-eyed appraisal: the story of The adjuvant colon cancer trials with edrecolomab. *J. Clin. Oncol.* 27, 1926–1929.
- Schnell, U., Kuipers, J., Mueller, J.L., Veenstra-Algra, A., Sivagnanam, M., Giepmans, B.N.G., 2013a. Absence of cell-surface EpCAM in congenital tufting enteropathy. *Hum. Mol. Genet.* 22 (July (13)), 2566–2571.
- Schnell, U., Kuipers, J., Giepmans, B.N.G., 2013b. EpCAM proteolysis: new fragments with distinct functions? *Biosci. Rep.* 33 (2), e00030.
- Seligson, D.B., Pantuck, A.J., Liu, X., Huang, Y., Horvath, S., Bui, M.H.T., et al., 2004. Epithelial cell adhesion molecule (KSA) expression. *Clin. Cancer Res.* 10 (April (8)), 2659–2669.
- Sergeant, G., Roskams, T., van Pelt, J., Houtmeyers, F., Aerts, R., Topal, B., 2011. Perioperative cancer cell dissemination detected with a real-time RT-PCR assay for EpCAM is not associated with worse prognosis in pancreatic ductal adenocarcinoma. *BMC Cancer* 11 (January), 47.
- Shiah, S.-G., Chang, L.-C., Tai, K.-Y., Lee, G.-H., Wu, C.-W., Shieh, Y.-S., 2009. The involvement of promoter methylation and DNA methyltransferase-1 in the regulation of EpCAM expression in oral squamous cell carcinoma. *Oral Oncol.* 45 (January (1)), e1–8.
- Shigdar, S., Lin, J., Yu, Y., Pastuovic, M., Wei, M., Duan, W., 2011. RNA aptamer against a cancer stem cell marker epithelial cell adhesion molecule. *Cancer Sci.* 102 (May (5)), 991–998.
- Songun, I., Litvinov, S.V., van de Velde, C.J.H., Pals, S.T., Hermans, J., van Krieken, J.H.J.M., 2005. Loss of Ep-CAM (CO17-1A) expression predicts survival in patients with gastric cancer. *Br. J. Cancer* 92 (May (9)), 1767–1772.
- Soto-Cerrato, V., Manuel-Manresa, P., Hernando, E., Calabuig-Fariñas, S., Martínez-Romero, A., Fernández-Dueñas, V., et al., 2015. Facilitated anion transport induces hyperpolarization of the cell membrane that triggers differentiation and cell death in cancer stem cells. *J. Am. Chem. Soc.* 137 (December (50)), 15892–15898.
- Souto, J.C., Vila, L., Brú, A., 2011. Polymorphonuclear neutrophils and cancer: intense and sustained neutrophilia as a treatment against solid tumors. *Med. Res. Rev.* 31

- (May (3)), 311–363.
- Soysal, S.D., Muenst, S., Barbie, T., Fleming, T., Gao, F., Spizzo, G., et al., 2013. EpCAM expression varies significantly and is differentially associated with prognosis in the luminal B HER2(+), basal-like, and HER2 intrinsic subtypes of breast cancer. *Br. J. Cancer* 108 (April (7)), 1480–1487.
- Spizzo, G., Gastl, G., Obrist, P., Fong, D., Haun, M., Grunewald, K., et al., 2007. Methylation status of the Ep-CAM promoter region in human breast cancer cell lines and breast cancer tissue. *Cancer Lett.* 246 (February (1–2)), 253–261.
- Stoelben, E., Loibner, H., Weder, W., Schmoll, C., Bijelovic, M., Hasse, J., 2008. Adjuvant active vaccination with IG101 in patients after radical lung cancer resection in stage Ib-IIIa – a prospective randomized, double-blind, placebo-controlled multicenter phase II/III study. *Chir. Forum.* 37, 329–331.
- Tai, K.-Y., Shiah, S.-G., Shieh, Y.-S., Kao, Y.-R., Chi, C.-Y., Huang, E., et al., 2007. DNA methylation and histone modification regulate silencing of epithelial cell adhesion molecule for tumor invasion and progression. *Oncogene* 26 (June (27)), 3989–3997.
- Tang, W., Huang, T., Xu, Z., Huang, Y., 2016. Novel mutations in EPCAM cause congenital tufting enteropathy. *J. Clin. Gastroenterol.*(November).
- Tas, F., Karabulut, S., Serilmez, M., Ciftci, R., Duranyildiz, D., 2014. Clinical significance of serum epithelial cell adhesion molecule (EPCAM) and vascular cell adhesion molecule-1 (VCAM-1) levels in patients with epithelial ovarian cancer. *Tumour Biol.* 35 (April (4)), 3095–3102.
- Tayama, S., Motohara, T., Narantuya, D., Li, C., Fujimoto, K., Sakaguchi, I., et al., 2017. The impact of EpCAM expression on response to chemotherapy and clinical outcomes in patients with epithelial ovarian cancer. *Oncotarget* 8 (July (27)), 44312–44325.
- van der Gun, B.T.F., Wasserkort, R., Monami, A., Jeltsch, A., Rasko, T., Slaska-Kiss, K., et al., 2008. Persistent downregulation of the pancreatic carcinoma-associated epithelial cell adhesion molecule via active intranuclear methylation. *Int. J. Cancer* 123 (July (2)), 484–489.
- van der Gun, B.T.F., Melchers, L.J., Ruiters, M.H.J., de Leij, L.F.M.H., McLaughlin, P.M.J., Rots, M.G., 2010. EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis* 31 (November (11)), 1913–1921.
- van der Gun, B.T.F., de Groote, M.L., Kazemier, H.G., Arendzen, A.J., Terpstra, P., Ruiters, M.H.J., et al., 2011. Transcription factors and molecular epigenetic marks underlying EpCAM overexpression in ovarian cancer. *Br. J. Cancer* 105 (July (2)), 312–319.
- Went, P., Dirnhofer, S., Salvisberg, T., Amin, M.B., Lim, S.D., Diener, P.-A., et al., 2005. Expression of epithelial cell adhesion molecule (EpCam) in renal epithelial tumors. *Am. J. Surg. Pathol.* 29 (January (1)), 83–88.
- Went, P., Vasei, M., Bubendorf, L., Terracciano, L., Tornillo, L., Riede, U., et al., 2006. Frequent high-level expression of the immunotherapeutic target ep-CAM in colon, stomach, prostate and lung cancers. *Br. J. Cancer* 94 (January (1)), 128–135.
- Winter, M.J., Cirulli, V., Briaire-de Bruijn, I.H., Litvinov, S.V., 2007. Cadherins are regulated by Ep-CAM via phosphatidylinositol-3 kinase. *Mol. Cell. Biochem.* 302 (August (1–2)), 19–26.
- Xiao, S., Liu, Z., Deng, R., Li, C., Fu, S., Chen, G., et al., 2017. Aptamer-mediated gene therapy enhanced antitumor activity against human hepatocellular carcinoma in vitro and in vivo. *J. Control Release* 258 (July), 130–145.
- Yanamoto, S., Kawasaki, G., Yoshitomi, I., Iwamoto, T., Hirata, K., Mizuno, A., 2007. Clinicopathologic significance of EpCAM expression in squamous cell carcinoma of the tongue and its possibility as a potential target for tongue cancer gene therapy. *Oral Oncol.* 43 (October (9)), 869–877.
- Yu, G., Zhang, X., Wang, H., Rui, D., Yin, A., Qiu, G., et al., 2008. CpG island methylation status in the EpCAM promoter region and gene expression. *Oncol. Rep.* 20 (November (5)), 1061–1067.
- Yu, T., Ma, Y., Wang, H., 2017. EpCAM intracellular domain promotes porcine cell reprogramming by upregulation of pluripotent Gene expression via beta-catenin signaling. *Sci. Rep.* 7 (April), 46315.
- Zheng, J., Zhao, S., Yu, X., Huang, S., Liu, H.Y., 2017. Simultaneous targeting of CD44 and EpCAM with a bispecific aptamer effectively inhibits intraperitoneal ovarian cancer growth. *Theranostics* 7 (5), 1373–1388.
- Zhou, N., Wang, H., Liu, H., Xue, H., Lin, F., Meng, X., et al., 2015. MTA1-upregulated EpCAM is associated with metastatic behaviors and poor prognosis in lung cancer. *J. Exp. Clin. Cancer Res.* 34 (December), 157.
- Zimpfer, A., Maruschke, M., Rehn, S., Kundt, G., Litzenberger, A., Dammert, F., et al., 2014. Prognostic and diagnostic implications of epithelial cell adhesion/activating molecule (EpCAM) expression in renal tumours: a retrospective clinicopathological study of 948 cases using tissue microarrays. *BJU Int.* 114 (2), 296–302.

Circulating tumor cells versus circulating tumor DNA in lung cancer – which one will win?

Silvia Calabuig-Fariñas^{1,2*}, Eloísa Jantus-Lewintre^{1,3*}, Alejandro Herreros-Pomares^{1,3}, Carlos Camps^{1,4,5}

¹Molecular Oncology Laboratory, General University Hospital Research Foundation, University General Hospital of Valencia, Valencia, Spain;

²Department of Pathology, Universitat de València, Valencia, Spain; ³Department of Biotechnology, Universitat Politècnica de València, Valencia, Spain; ⁴Department of Medicine, Universitat de València, Valencia, Spain; ⁵Department of Medical Oncology, University General Hospital of Valencia, Valencia, Spain

Contributions: (I) Conception and design: S Calabuig-Fariñas, E Jantus-Lewintre, C Camps; (II) Administrative support: None; (III) Provision of study materials or patients: S Calabuig Fariñas, E Jantus-Lewintre, A Herreros-Pomares; (IV) Collection and assembly of data: S Calabuig-Fariñas, A Herreros-Pomares; (V) Data analysis and interpretation: S Calabuig Fariñas, E Jantus-Lewintre, A Herreros-Pomares; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*These authors contributed equally to this work.

Correspondence to: Carlos Camps, MD, PhD. Associate Professor, Department of Medicine, Universitat de València, Av. de Blasco Ibáñez, 15, 46010 València, Valencia, Spain. Email: camps_car@gva.es.

Abstract: Liquid biopsies appear to be a reliable alternative to conventional biopsies that can provide both precise molecular data useful for improving the clinical management of lung cancer patients as well as a less invasive way of monitoring tumor behavior. These advances are supported by important biotechnological developments in the fields of circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA). Analysis of CTCs and ctDNA may be useful in treatment selection, for response monitoring, and in studying resistance mechanisms. This review focuses on the most recent technological achievements and the most relevant clinical applications for lung cancer patients in the CTC and ctDNA fields, highlighting those that are already (or are close to) being implemented in daily clinical practice.

Keywords: Lung cancer, circulating tumor cell (CTC); circulating tumor DNA (ctDNA); biomarker; liquid biopsy

Submitted Jul 06, 2016. Accepted for publication Sep 14, 2016.

doi: 10.21037/tlcr.2016.10.02

View this article at: <http://dx.doi.org/10.21037/tlcr.2016.10.02>

Introduction

Lung cancer is the leading cause of cancer-related mortality among men and women worldwide, with more than 1.8 million estimated new cases each year (1). Despite advances in biomedical research and improvements in both the diagnostic tools and therapeutic options that have become available over the past few decades, lung cancer still has a 5-year overall survival (OS) rate of 18% for all stages (2). The main reason for this poor outcome for this cancer type is late diagnosis: a high percentage of the patients are diagnosed at advanced stages when curative surgery is no longer possible. The clinical management of lung cancer in advanced stages is also changing; better understanding and descriptions of the molecular abnormalities present in lung

cancer have opened up new therapeutic options in specific disease subsets.

Lung cancer: driver alterations, predictive biomarkers, and intratumor heterogeneity

The development of a new generation of molecular techniques has led to substantial advances in the knowledge of cancer genomes, and specifically in lung cancer, facilitating the discovery of oncogenic-driver mutations/alterations that cause aberrant signaling and proliferation in certain tumor subsets. These findings have allowed the development of new treatment strategies based on molecular targets and thus have reshaped tumor classification from classical histological

methods towards molecular pathology approaches. Indeed, non-small cell lung cancer (NSCLC) is one of the most genomically diverse tumor types and there is a wide variety of molecularly-defined patient subsets characterized by specific driver-mutations, involving different genes such as epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), or V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) (3-7) among others.

The identification of driver mutations located in the tyrosine kinase domain of EGFR as the primary oncogenic event in a subset of lung adenocarcinomas (8-13) led to the development of specific tyrosine-kinase inhibitors (TKIs) for this receptor, resulting in a dramatic change in the treatment of advanced NSCLC patients harboring EGFR mutations. The use of EGFR-TKIs has produced a particularly large increase in progression-free survival (PFS) with a negligible toxicity profile as well as an increase in OS to more than 24 months (11,14-18). Unfortunately, the effect of TKIs is limited over time because of the emergence of drug resistance. A number of molecular mechanisms underlying acquired resistance to EGFR-TKIs have been reported, including the secondary EGFR Thr790Met (T790M) mutation found in around 50% of patients, loss of the EGFR mutant allele, MET amplification, hepatocyte growth factor (HGF) overexpression, phosphatase and tensin homolog (PTEN) downregulation, epithelial-mesenchymal transition (EMT), BRAF mutations, and other mechanisms (19-22). Resistance is frequently related to the emergence of a clone thought to be initially present at a low percentage in the tumor, thus emphasizing the role of intratumour heterogeneity as one way of explaining resistance mechanisms to targeted agents (23,24).

Other driver mutations in lung cancer have also been targeted by these new drugs but, again, the emergence of resistance is a common event in patients treated with targeted-therapies (25,26). In addition, metastases are responsible for roughly 90% of cancer-associated patient mortality, through largely unknown mechanisms. Therefore, future studies must aim to directly analyze metastatic cells in order to better understand the mechanisms of cancer spread (27). However, metastasis biopsy is an invasive procedure limited to certain locations, and additionally, recent work has shown that different metastatic sites harbor different genomic aberrations and so biopsy of only one or two accessible metastases may not be enough to represent the whole tumor genome (24,28). Finally, solid tumors also exhibit temporal heterogeneity, evolving over time under selection-pressure with treatment (29-32). Even though,

and considering the heterogeneity of the tumors, it becomes difficult to have a complete scenario of the whole tumor based on the information obtained from small biopsies, and in several cases from a restringed number of tumor cells which, in turn, could lead to erroneous therapeutic decisions.

Since the beginning of the era of targeted therapies, there has been a clear need to understand the mechanisms of resistance and therefore rebiopsies at the time of clinical progression or the emergence of treatment resistance were gradually incorporated into clinical practice. Considering all the above-mentioned facts about lung cancer: presence of driver mutations, tumor heterogeneity, tumor dynamics, drug sensitivity, drug resistance, it is of remarkable importance the development of a non-invasive way to obtain this valuable biologic information, which includes the ability to diagnose, prognose and monitor lung cancer evolution. At this point, liquid biopsies seem to be the approach that covers all these requirements. In lung cancer, blood samples are the most explored type of liquid biopsy, and have been used to improve the diagnosis as well as for the searching prognostic and predictive biomarkers. Liquid biopsies have several advantages: (I) they allow rapid biomarker assessment in lung cancer patients for whom solid biopsies are impossible because of restricted or extremely risky access possibilities; (II) they can easily be repeated during cancer patient follow-up in order to control treatment efficiency; and/or (III) they can be used to detect genomic alterations occurring as result of resistance to therapy.

Liquid biopsies in lung cancer

In lung cancer liquid biopsies, blood samples are mainly used as a sample source for analyzing circulating tumor cells (CTCs) or circulating tumor DNA (ctDNA), in addition to other biomarkers of interest, such as circulating microRNAs, circulating RNA, platelets, plasma/serum metabolites, or exosomes (*Figure 1*). To explore the data available regarding the above in this review we used MEDLINE to perform a comprehensive literature search for original and review articles related to the terms “liquid biopsy” and “CTCs”, “ctDNA” and “NSCLC”, or “lung cancer”. Furthermore, to complete our search we also reviewed the bibliographies of these articles; only those published in English and in peer-reviewed journals were included. Our purpose was to describe the current contribution of CTC and ctDNA detection and analysis

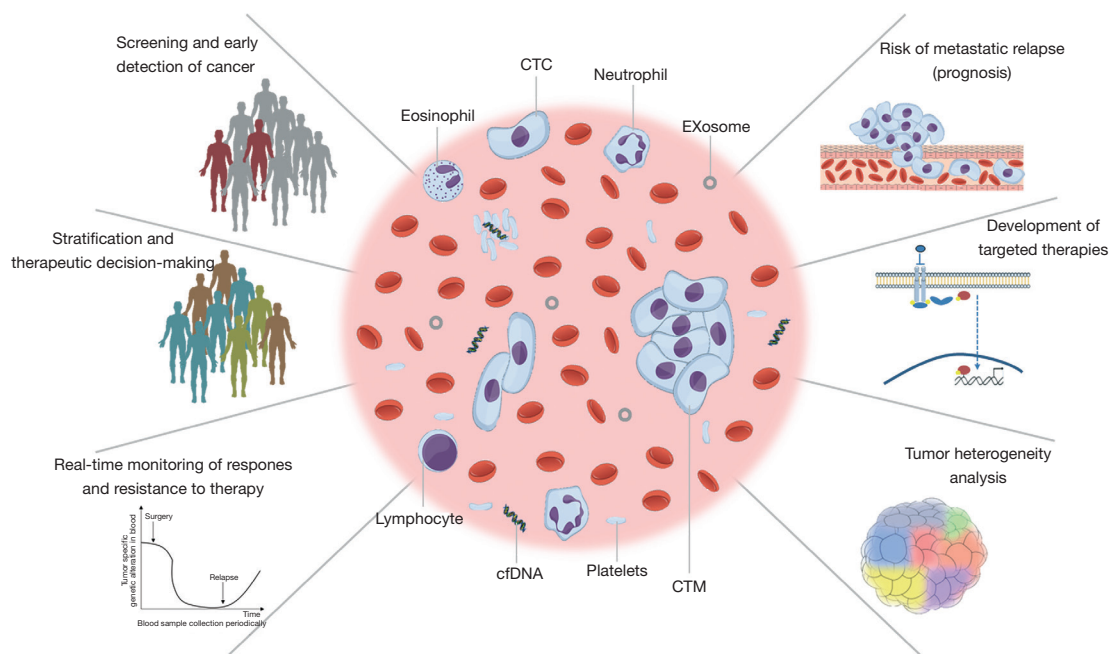


Figure 1 The potential clinical benefits of CTC and ctDNA analyses in cancer care. CTC, circulating tumor cell; cfDNA, circulating free DNA; CTM, circulating tumor microemboli.

in lung cancer patients and to compare the advantages and disadvantages of these two approaches.

CTCs versus ctDNA: which one will win?

There are an increasing number of scientific reports pointing out the advantages of, and difficulties in, both detecting mutations and isolating CTCs and ctDNA in both metastatic and non-metastatic lung cancer (*Table 1*). The discrepancies in sensitivity, reproducibility, and concordance with tissue biopsies are likely due to the different approaches and methodologies used as well as their clinical settings.

CTCs

CTCs are tumor cells from solid tumors that spread via blood and/or lymphatic vessels. CTCs are shed into the vasculature from primary tumors and are postulated to contain subpopulations of cells with the potential to spread and initiate distant metastases (33). They were observed for the first time by Thomas Ashworth in 1869 in the blood of a man with metastatic cancer (34), but they only became relevant in modern cancer research a couple of decades ago with the demonstration of their presence early in the course of malignant disease (35). Several models have been

suggested to describe the dissemination process whereby tumor cells leave the primary tumor to colonize distant organs, either when they become competent to metastasize or because of physical tumor extension (27,36).

CTCs can circulate in the bloodstream of lung cancer patients as single cells or as aggregates known as circulating tumor microemboli (CTM) (37-39). In this regard, the phenotype of single or aggregated CTCs can be different and so may present different levels of potential aggressiveness (37,38,40,41). Similar to single migratory mesenchymal-like CTCs, CTMs appear to be enriched in mesenchymal markers, an indicator of increased potential plasticity, which in turn seems to be related to more aggressive behavior, thus supporting their role in both tumor cell dissemination and the initiation of metastatic outgrowth (38,42-44). The presence of CTMs has been reported as a negative prognostic factor in lung cancer patients (38,40,45).

Isolation and detection of CTCs

Although many technologies have been developed over the past few years to detect and isolate CTCs in the peripheral blood of NSCLC patients (44,46-48), this task remains challenging (*Figure 2*). In advanced lung cancer patients,

Table 1 Advantages and limitations of CTCs and ctDNA liquid biopsy assays

	Advantages	Limitations
CTCs	<p>Constitute a minimally-invasive procedure with high specificity</p> <p>Address intra-tumor heterogeneity and supplies an adequate reflection of the tumor</p> <p>Allow structural evaluation of cancer phenotype</p> <p>Give information about tumor progression and metastasis</p> <p>Permit different <i>in vitro</i> and <i>in vivo</i> assays</p> <p>Make molecular characterization of the disease possible</p> <p>Offer the use of immunolabeling techniques</p> <p>Provide complementary information to ctDNA</p> <p>Could facilitate therapeutic decision-making</p>	<p>Prospective collection is required for targeting pre-analytical differences</p> <p>Constitutes a small and fragile population</p> <p>Analytical methods with very high sensitivity and specificity are needed</p> <p>False-negative and false-positive errors</p> <p>Low signal-to-noise ratio, mostly in early-stage disease</p> <p>Heterogeneity of CTCs complicates identification</p> <p>Disparity of techniques used for CTC isolation</p> <p>—</p> <p>—</p>
ctDNA	<p>Constitutes a minimally-invasive procedure with high specificity</p> <p>Addresses intra-tumor heterogeneity and supplies an adequate reflection of the tumor</p> <p>Extremely high sensitivity for detection of cancer burden even after curative care</p> <p>Might predict acquired treatment resistance</p> <p>Could facilitate therapeutic decision-making</p>	<p>Prospective collection is required for targeting pre-analytical differences</p> <p>False-negative and false-positive errors</p> <p>Low signal-to-noise ratio, mostly in early-stage disease</p> <p>Lack of standardization of pre-analytical conditions</p> <p>No protein or functional studies available</p>

CTCs, circulating tumor cells; ctDNA, circulating tumor DNA.

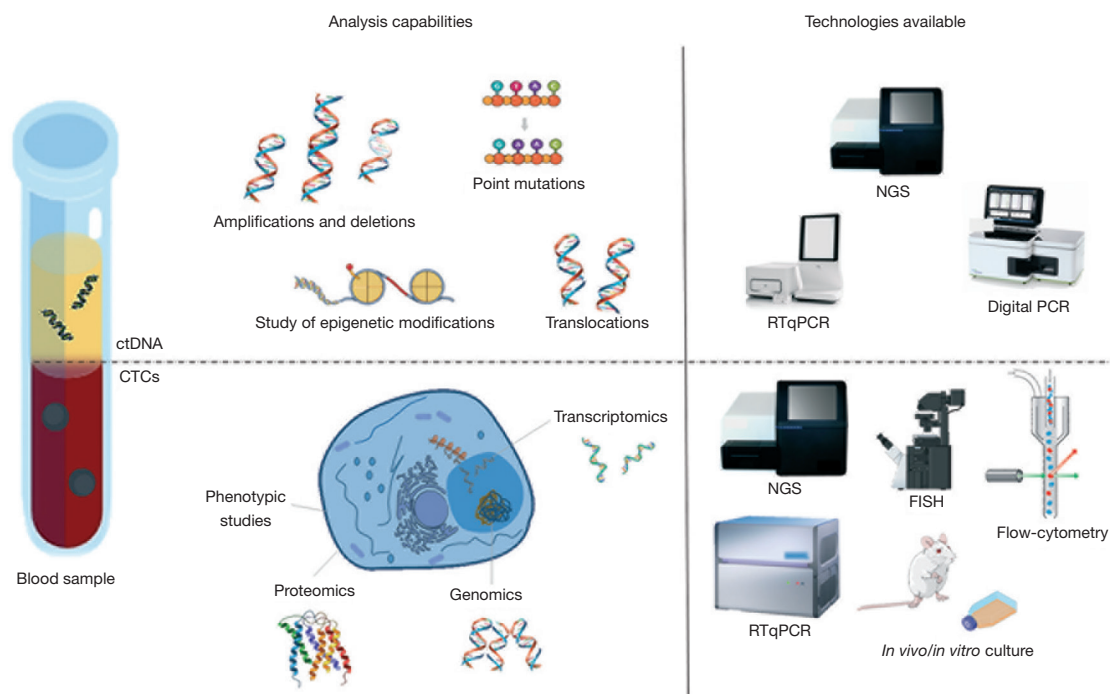


Figure 2 Comparison of the analysis capability and the technologies available for CTCs and ctDNA. CTC, circulating tumor cell; cfDNA, circulating free DNA; NGS, next generation sequencing; RTqPCR, real-time quantitative PCR; FISH, fluorescence in situ hybridization.

CTCs are rare cells present in very low concentrations in the bloodstream; approximately 1 mL of whole peripheral blood contains 1–10 CTCs against a background of 10^6 – 10^7 nucleated blood cells and around 10^9 red blood cells (49). Therefore, to reach the extreme sensitivity required to detect CTCs, an enrichment step is often required to increase their concentration before trying to detect or capture them. In this context, different methods for cellular enrichment, characterization, and identification of lung CTCs have been developed, including CTC microchips, filtration devices, quantitative reverse-transcription PCR assays, automated microscopy systems, etc. (46,50,51).

CTC detection or capture methods can be broadly categorized as: (I) label-dependent, based on positive enrichment involving cell surface markers such as epithelial cell adhesion molecule (EpCAM); or (II) label-independent, based on negative selection, such as size, or other differential biophysical properties of CTCs. Besides these two main approaches, other strategies include direct CTC imaging and functional assays (46,47,52–54). The first label-dependent studies tried to detect lung CTCs based on the assumption that circulating tumoral cells maintain the same characteristics as their tissue of origin, therefore most lung CTC categorization was based on the expression of epithelial-specific markers such as cytokeratin (CK) or intermediate filaments (IF) (41,55–60). Therefore, positive enrichment methods define lung CTCs as nucleated cells present in the bloodstream that express epithelial CKs and EpCAM and do not express the white blood cell surface antigen CD45 (59,61,62).

One of these methodologies is the CellSearch[®] system (Veridex, Raritan, NJ, USA) which has been approved by the Food and Drug Administration (FDA) for monitoring metastatic breast cancer (63), castration-resistant prostate cancer (64), and colon cancer patients (65). It has also been shown to be of prognostic significance for small cell lung cancer (SCLC) but the test has not yet been FDA-approved for this cancer type (40,66,67). The method is based on an initial enrichment of EpCAM positive cells followed by immunofluorescent staining using epithelial markers (CK 8, 18, 19), a leucocyte marker (CD45), and 4', 6-diamidino-2-phenylindole (DAPI) for nuclear staining. The CellSearch[®] definition of CTCs is any intact EpCAM+/CK+/CD45–cell at least 4 μ m in size and with a nucleus occupying at least 50% of the cytoplasm (41,55–61). Using EpCAM-dependent assays, CTCs can be detected in approximately 20–40% of NSCLC patients (41,55–58,60). Unfortunately, technologies relying on EpCAM positive selection cannot

detect CTCs that have undergone EMT or any cancer stem cells that have not yet started epithelial differentiation. In lung cancer, CK-negative CTCs, which potentially represent tumor cells undergoing EMT, have also been identified. Consequently, the use of EpCAM as a positive selection marker should be carefully considered when trying to detect CTCs in NSCLC patients. Unfortunately, so far no reliable surface markers, which distinguish lung CTCs from normal epithelial cells and can be used for such label-dependent approaches, have yet been identified.

Label-independent approaches to CTC detection in lung cancer are an attractive alternative. One such method, Isolation by Size of Epithelial Tumor cell (ISET[®], developed by Rarecells Company, France), used for cells isolated from lung cancer patients, involves a CTC enrichment step based on size by using a filtration device followed by cytological characterization. This system has been used to detect lung CTCs in both metastatic and non-metastatic NSCLC patients and has shown increased sensitivity in a wider range of patients compared to label-dependent methods such as CellSearch[®] (41,55,68–75).

Another direct technology, known as ScreenCell[®], is also based on the size of CTCs but, in addition, allows their isolation so that they can be subjected to further morphological and molecular studies (76,77). Interestingly, it has been shown that lung CTCs isolated with different systems, can be cultured *in vitro* which is of particular interest for generating *in vitro* and *in vivo* models. As a first step towards this goal, data has already been generated for successful short-term cultures (up to 28 days) of CTCs isolated from patients with lung cancers (78–80). Such model systems could be used to study drug susceptibility or to generate sufficient numbers of cells for systematic deep analysis of their molecular profiles or biological behavior (52,81). Several recent studies have reported the development of mouse xenografts generated directly from CTCs or from breast, colorectal, prostate, hepatocellular, small cell lung, or gastric cancer CTC cultures (82–87). In particular, CTCs enriched from blood samples from SCLC patients were subcutaneously implanted into immunocompromised mice as CTC-derived explants (CDX); the CTCs were tumorigenic at densities of more than 53 CTCs/1 mL of blood, however, such large numbers of CTCs are not always obtained from advanced patients, thus highlighting one of the biggest challenges associated with these approaches (86).

Current models generated either *in vitro* or *in vivo* are also potentially limited if the clones they are grown from

do not accurately reflect the true heterogeneity of the tumor (e.g., there may be a selective advantage for highly aggressive clones). Furthermore, *in vivo* xenograft models do not recapitulate tumor-host interactions that may play a role in drug resistance. Direct comparison between label-dependent and label-independent CTC isolation methods shows that both approaches have pros and cons. Label-dependent methods are more specific but they are rendered ineffective when antigen expression is lost in certain CTC subpopulations, and the cells become less viable after isolation. On the other hand, label-independent approaches are less specific but do not depend on CTC phenotype, and seem to better preserve CTC viability for downstream applications. There are currently many other technological developments focused on exhaustive lung CTC characterization in the pipeline at several diagnostic companies.

CTCs: clinical applications

CTC analysis is considered an interesting approach for early diagnosis, prognosis assessment, prediction of treatment efficacy, and early detection of lung cancer relapse. The most relevant lung cancer CTC studies are summarized in *Table 2*.

Larger numbers of CTCs are recovered from SCLC patients than from NSCLC patients (40,56). In addition, some researches have reported a positive association between the number of CTCs and clinical stage or the presence of distant metastases in primary lung cancer (56,57,73,88,93), whereas other studies failed to find any significant differences (50,92). Regarding other clinical applications, the majority of articles on CTCs in lung cancer focus on their prognostic role. In a large population of NSCLC patients, one group showed that isolation of more than 50 CTCs per 10 mL sample (using ISET[®]) is of prognostic value and is associated with shorter OS and disease-free survival (DFS) (89). Similar results were reported by Krebs *et al.* showing that PFS and OS was significantly better in advanced NSCLC patients with fewer CTCs (41). However, in another NSCLC patient cohort it was reported that although the median survival time tends to be shorter in CTC-positive than in CTC-negative patients, the difference was not significant (92).

In SCLC a significant association between higher numbers of CTCs and shorter survival has been described, and at least one study has reported that CTCs are a better predictor of survival than disease stage and tumor response

(40,66,90). Moreover, a reduction in the number of CTCs after chemotherapy was also significantly associated with better outcomes in SCLC (66,94). Regarding the role of CTCs as biomarkers for therapeutic monitoring in NSCLC, comparisons between studies performed on samples collected before and after chemotherapy have consistently found that survival rates were significantly worse for patients with CTC counts that remained positive during treatment (95,96). In a group of patients treated with erlotinib and pertuzumab, the authors found that the decrease in CTC count upon treatment were significantly associated with disease response (91).

One of the main difficulties of working with CTCs in the field of lung cancer is their use as a theranostic tool for detecting somatic mutations (97). However, in 2008 Maheswaran and Sequist identified the presence of EGFR-activating mutations in 11 out of 12 (92%) of CTCs isolated from EGFR-mutated patients. During follow-up the authors detected the T790M mutation (which confers drug resistance) in CTCs collected from patients who progressed to TKI treatment (98). Moreover, *EGFR* mutations in CTCs from NSCLC patients were recently successfully assessed using sensitive next generation sequencing (NGS) (59). Similarly, in 2012 Paul Hofman's group published results from ALK-specific fluorescence in situ hybridization (FISH) analyses performed on archived lung cancer patient CTC samples. Their blind analysis of CTCs and corresponding tumor tissue showed a perfect match: 5 patients positive for ALK rearrangement in both CTC and tumor tissue were found while 82 were negative for this mutation in both CTC and tumor tissue (99).

In summary, there are currently 343 studies registered on "ClinicalTrials.gov" concerning CTCs and lung cancer, but despite the numerous scientific publications on this topic, these cells are still not used in routine clinical practice. This can be explained by the large number of methods available for CTC detection, and by the difficulty of selecting a reliable lung CTC marker. Despite the efforts made by the scientific community in the CTC field to try to improve lung cancer management, the analytical specificity and clinical utility of these methods must still be demonstrated in large prospective multicenter studies in order to obtain the evidence required for their introduction into the daily management of lung cancer patients.

ctDNA

The field of ctDNA analysis originally started almost 70 years

Table 2 Diagnostic and prognostic impact of CTC analysis in lung cancer

Study	Patients included	Histology	Stage	Technology	Sensitivity (%)	Diagnostic and/or prognostic significance	Refs.
Stenel et al. [2003]	62	NSCLC	I-IIIa	Ficoll-Hypaque	18	Disseminated cancer cells in the PV blood are detectable in patients with operable NSCLC and are associated with a poor clinical outcome	(88)
Tanaka et al. [2009]	125	NSCLC	I-IV	CellSearch	30-71	CTC count is a useful surrogate prognostic marker, but insufficient for discriminating malignant and non-malignant patients	(56)
Hou et al. [2009]	97	SCLC	IIB-IV	ISET/CellSearch	85	M30, M65, lactate dehydrogenase, and CTC number are prognostic for patient survival. Detection of CTM in 32% of patients	(39)
Hofman et al. [2011]	208	NSCLC	I-III	ISET	49	The presence and level of 50 or more CNHCs are associated with worse survival	(89)
Hofman et al. [2011]	210	NSCLC	I-IV	ISET/CellSearch	39-69	The presence of CTCs detected by CellSearch and ISET shows the best correlation with shorter DFS at a univariate and multivariate level	(55)
Krebs et al. [2011]	101	NSCLC	III-IV	CellSearch	21	PFS and OS are significantly higher for patients with ≤ 5 CTCs before chemotherapy	(41)
Hiltermann et al. [2012]	59	SCLC	I-IV	CellSearch	73	Absolute CTCs after one cycle of chemotherapy in patients with SCLC is the strongest predictor for response on chemotherapy and survival	(66)
Hou et al. [2012]	97	SCLC	I-IV	ISET/CellSearch	85	Pretreatment CTCs, change in CTC number after one cycle of chemotherapy, CTM, and apoptotic CTCs are independent prognostic factors	(40)
Naito et al. [2012]	51	SCLC	I-IV	CellSearch	68	Patients with ≥ 8 CTCs had worse survival than those with < 8 CTCs at baseline, posttreatment, or relapse	(90)
Hofman et al. [2012]	250	NSCLC	I-IV	ISET/Cytology	49	49% of patients showed CNHCs, corresponding to malignant (41%), uncertain malignant (6%), and benign cells (2%)	(71)
Punnoose et al. [2012]	41	NSCLC	III-IV	CellSearch/RT-PCR	78	There is a correlation between decreases in CTC counts and response to radiography by either FDG-PET or RECIST in advanced NSCLC patients	(91)
Paillet et al. [2013]	18	NSCLC	IIB-IV	CellSearch/ISET/RT-PCR	100	ALK rearrangement can be detected in the CTCs of patients with ALK-positive NSCLC, enabling both diagnostic testing and monitoring of crizotinib treatment	(72)
Funaki et al. [2013]	130	NSCLC	I-IV	Rosette Sep	74	DFS rate was significantly worse in patients where CTMs were detected	(45)
Juan et al. [2014]	37	NSCLC	IIB-IV	CellSearch	24	No significant differences in PFS and OS were observed between patients with or without CTCs at baseline	(92)
Muñelo-Romay et al. [2014]	43	NSCLC	IIB-IV	CellSearch	42	Patients with ≥ 5 CTCs at baseline presented worse PFS and OS. In addition, patients with increased levels of CTCs during the treatment showed lower OS and PFS rates	(61)
Carlsson et al. [2014]	129	NSCLC	I-IV	HD-CTC assay	50	The presence of CTMs with clinical and imaging data significantly discriminates diagnostic accuracy in NSCLC patients	(37)
Chudasama et al. [2015]	20	NSCLC	III-IV	Screen Cell	25	An increase in CTCs following EC is observed in 75% of patients, which may have implications for tumor dissemination	(77)

NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; CTCs, circulating tumor cells; CTM, circulating tumor microemboli; DFS, disease-free survival; PFS, progression-free survival; OS, overall survival; CNHCs, circulating non-hematological cells; PV, pulmonary vein; CK, cytokeratin; EC, endobronchial cryotherapy; NS, non-specific.

ago (100); higher levels of circulating free DNA (cfDNA) were identified in cancer patients compared to healthy controls, suggesting that this correlated with malignancy and tumor stage (101-103). To date, two main mechanisms for releasing ctDNA have been postulated: “passive” and “active”. The passive mechanism involves the release of nucleic acids into the bloodstream, either directly from apoptotic and necrotic tumor cells or indirectly by necrotic tumor cells engulfed by macrophages (104). Data about the size distribution of cfDNA has revealed an enrichment in 150–180 bp fragments which reflects the nucleosomal pattern of DNA fragmentation characteristic of apoptotic processes (105-107). In contrast, fragments of tumoral nucleic acid can also be actively released into the circulation by living cells. One potential explanation hypothesizes that cancer cells release nucleic acids to transform the targeted recipient cells at distant locations, although the mechanisms are not completely understood (108). It is important to consider that ctDNA may represent a small proportion of the total cfDNA, at levels corresponding to one genome equivalent in 5 mL of plasma (0.01% allele fraction), and thus it may be undetectable with routine sampling (103,109). Apart from this, ctDNA levels can vary according to tumor burden and stage, anatomical proximity to vasculature, and biological features like apoptotic rate and metastatic potential (110,111).

Detection and quantification of ctDNA

The most common sample source used for ctDNA isolation is plasma collected in standard EDTA tubes. However, considering the low percentage of ctDNA present within total isolated cfDNA, it is important to control both the analytical and pre-analytical steps that can significantly affect ctDNA detection in blood samples (112). Plasma samples should be processed and stored immediately after blood collection to avoid increases in genomic DNA released from white blood cell lysis that could modify the relative levels of ctDNA. Therefore, the uses of standardized protocols in conjunction with specialized preservative-containing tubes (e.g., Streck Cell-Free DNA BCT) are strongly recommended (113).

The amounts of ctDNA present in lung cancer patient samples give important diagnostic and prognostic information about the disease (114,115). However, the most important advantage of this technology is that it enables such analyses via easily obtained, minimally-invasive samples which are likely to reflect any genomic abnormalities present in the original neoplasm, giving

insights into the types of mutations, indels, chromosomal rearrangements, chromosomal region gains or losses, and epigenetic modifications present (54,116-118). Given the small proportion of ctDNA present in the total cfDNA samples obtained, it is important to select the correct methods for its isolation and analysis; several highly sensitive techniques have been developed for the latter, ranging from PCR-based to more complex methodologies using NGS (summarized in *Figure 2*).

In lung cancer a variety of methods have been used for ctDNA analysis, many of them based on real-time PCR, although these approaches are more applicable when a limited number of loci are evaluated. Such systems include peptide nucleic acid (PNA) or locked nucleic acid (LNA) mutant-enriched PCR (ME-PCR) (119-121), amplification-refractory mutation system (ARMS) (122), digital PCR (including droplet-based systems) (123), and the beads, emulsification, amplification, and magnetics (BEAMing) system (124,125). Moreover, recently developed NGS technologies have also shown that it is possible to detect many cancer-associated mutations in single lung cancer patient blood samples (126,127). There are also protocols specifically intended to improve the sensitivity of NGS ctDNA sample analysis; these include tagged-amplicon deep sequencing (TAm-Seq) (128), Safe-Sequencing System (Safe-SeqS) (129), and cancer personalized profiling by deep sequencing (CAPP-seq) (109,130,131), among others.

ctDNA: clinical applications in lung cancer

The clinical applications of ctDNA can be divided into two main categories: (I) quantification of circulating DNA for early diagnosis, prognosis, and response prediction; and (II) analysis of ctDNA in order to profile and characterize molecular tumor alterations (*Table 3*).

Lung cancer patients have increased plasma cfDNA levels compared with control individuals (142,143) and the amount of cfDNA has been associated with tumor stage and burden in lung cancer (109). However, there are data demonstrating that this absolute cfDNA amount is limited as diagnostic tool in the absence of contextual knowledge of any associated tumor mutations (114). High levels of cfDNA have been reported as an indicator of poor outcome in lung cancer patients (115,138,142), but in other studies pretreatment levels of total cfDNA were not prognostic of survival (144,145). One possible restriction of these approaches is that cfDNA is also present at high levels in the blood of patients with benign diseases such as hepatic

Table 3 Studies of the mutation status in cfDNA samples from NSCLC patients

Study	Patients included	Stage	Technology	Sample	Sensitivity (%)	Specificity (%)	Conclusions	Refs.
Rosell <i>et al.</i> [2009]	164	I-IV	ME-PCR	Serum	59	NA	Ninety-seven of 164 patients carried mutations (del19 and L858R) in 64 and 33, respectively	(119)
Taniguchi <i>et al.</i> [2011]	44	III-IV	BEAMing	Plasma	73	NA	From 44 patients that were confirmed to have activating mutations in their primary tumor, 32 mutations were detected in the plasma DNA	(125)
Goto <i>et al.</i> [2012]	194	III-IV	ARMS	Serum	43	100	Fewer patients were EGFR mutation positive when assessed using pretreatment cfDNA (23.7%) vs. tumor tissue-derived DNA (61.5%)	(132)
Huang <i>et al.</i> [2012]	822	III-IV	DHPLC	Plasma	70	80	Concordance of EGFR mutations between tissue and plasma samples was 77%. In patients with two or more lines of EGFR-TKI therapy, EGFR mutation status in plasma samples was a predictor for PFS	(133)
Zhao <i>et al.</i> [2013]	111	I-IV	ME-PCR	Plasma	36	95.5	The sensitivity was 10% in early-stage patients and 56% in advanced-stage patients. Patients with poorly differentiated tumors showed the highest sensitivity (77.8%)	(134)
Kim <i>et al.</i> [2013]	60	III-IV	PNA/LNA PCR	Plasma	17	100	There was no statistically significant difference in clinical parameters between patients with EGFR mutations in plasma and those without EGFR mutations	(121)
Kim <i>et al.</i> [2013]	57	III-IV	PNA/LNA PCR	Serum	73	91	The status of EGFR and KRAS mutation in serum was not prognostic in patients with advanced NSCLC	(135)
Wang <i>et al.</i> [2014]	134	III-IV	ARMS	Plasma	22	97	Patients with high TGF- β 1 levels showed shorter OS and worse response to EGFR-TKIs than patients with low TGF- β 1 levels	(136)
Couraud <i>et al.</i> [2014]	107	I-IV	Ion Torrent PGM	Plasma	58	87	In tumor DNA, 50 mutations (36 EGFR, 5 ERBB2, 4 KRAS, 3 BRAF, and 2 PIK-3CA) were identified, of which 26 were detected in cfDNA	(127)
Weber <i>et al.</i> [2014]	199	II-IV	Cobas® EGFR blood test	Plasma	61	96	Patients with activating EGFR mutations in plasma DNA had a longer PFS than patients with wild-type status	(137)
Karachaliou <i>et al.</i> [2015]	97	III-IV	PNA/LNA PCR	Plasma	78	NA	Median OS was shorter in patients with the L858R mutation in cfDNA than in those with the exon 19 deletion	(138)
Thress <i>et al.</i> [2015]	38	IV	Cobas®/Therascreen®/DDPCR/BEAMing	Plasma	78-100	93-100	For the T790M mutation, the sensitivity and specificity were 73% and 67%, and 81% and 58% with the Cobas®Mutation Test and the BEAMing dPCR, respectively	(139)
Wei <i>et al.</i> [2016]	50	IV	DDPCR	Plasma	76	88	Concurrent mutant T790M DNA detection of lung cancer patients at baseline achieved 82% concordance with matched tissue analysis	(140)
Newman <i>et al.</i> [2016]	142	III-IV	CAPP-Seq	Plasma	92	100	In silico elimination of highly stereotypical background artifacts with a molecular barcoding strategy improves the sensitivity of cancer profiling by CAPP-Seq	(131)
Zheng <i>et al.</i> [2016]	25	III-IV	DDPCR	Plasma	81	100	In patients receiving 2 nd line or later TKI treatment, the T790M cfDNA positive group had a significantly shorter OS than the negative group	(141)

cfDNA, circulating free DNA; ctDNA, circulating tumor DNA; TKI, tyrosine-kinase inhibitor; PFS, progression-free survival; OS, overall survival; ME-PCR, mutant-enriched PCR; BEAMing, beads, emulsification, amplification, and magnetics; DHPLC, denaturing high performance liquid chromatography; ARMS, amplification-refractory mutation system; PNA/LNA PCR, peptide nucleic acid/locked nucleic acid PCR; DDPCR, droplet digital PCR; CAPP-Seq, cancer personalized profiling by deep sequencing.

disorders, diabetes, cardiovascular diseases, non-neoplastic lung diseases, or infections (75).

Regarding the prognostic information provided by ctDNA, monitoring tumor-specific alterations present in ctDNA isolated from plasma from early stage NSCLC patients following surgical resection identified patients with residual disease and was able to detect disease recurrence (109,145). However, despite the reporting of some controversial results, when *KRAS* mutations in ctDNA were assessed as a prognostic marker in NSCLC patients (135,146,147) *KRAS* status in plasma ctDNA was associated with poor tumor response to EGFR-TKIs in NSCLC patients and so it could be used as a predictive marker for selecting appropriate NSCLC patients for such treatments (148-150). Is in this latest aspect, the presence of circulating DNA containing tumor-specific sequences, where we find the most widespread and important applications of ctDNA.

Several reports have analyzed the concordance between genomic alterations (such as *EGFR* mutations) present in lung cancer tissues and corresponding ctDNA samples (119,125,132-134,137): depending on the technology used, the agreement ranges between 60% to more than 90% (122,123,125,138,139,149,151,152). The EURTAC trial was one of the first to explore the feasibility of using ctDNA as a surrogate for tumor biopsy and to correlate mutations in plasma with PFS and OS (120). Since then, several other clinical trials in lung cancer have incorporated the analysis of plasma as a sample source for studying genomic tumor alterations (138).

In the context of metastatic disease the use of ctDNA is particularly useful for patients with tumors that are difficult to biopsy, those with contraindications for biopsy procedures, or where tumor samples have been exhausted; in these cases the possibility of determining the presence of genomic tumor alterations in ctDNA have brought forward the prospect of implementing precision oncology. In addition, ctDNA can be used for real-time monitoring of therapeutic responses to targeted-agents (132,136,152-154) as a valid surrogate for the current use of invasive rebiopsies. In this respect, a number of research groups have recently shown that ctDNA analysis can sensitively and specifically detect T790M clones early, i.e., before therapy or their emergence during EGFR-TKI treatments, demonstrating that this approach represents also an elegant way to overcome the problems arising from tumor heterogeneity(19,29,140,141,155).

For broader applicability, ctDNA analysis platforms focus on not only maximizing analytical sensitivity, but also on providing sufficient genomic coverage to be

able to analyze multiple molecular markers in the same sample, thus providing the possibility of anticipating the molecular alterations expected as tumors evolve. Therefore, ongoing and future prospective studies should aim to test if treatment strategies informed by the unique data provided by ctDNA could yield superior clinical outcomes compared to tissue-based approaches.

The war: strengths, and limitations

There are many studies aiming to detect and/or characterize CTCs or ctDNA in lung cancer patients; the question is which of these two approaches will become the eventual gold standard for managing these patients in the era of precision oncology. In this “war” the usefulness of CTCs for *ex vivo* models, including in functional studies such as cultures, mouse xenografts, or real-time *in vitro* assays for drug sensitivity evaluation, is undisputed. CTC enumeration as a prognostic biomarker in lung cancer research has not been as successful as it was in breast, prostate, and colon cancers for which there is a FDA-approved CTC method; even so, the adoption of this approach in these cancers in daily oncological practice remains low.

Reports on comparative mutation analyses of CTCs and cfDNA have shown an interesting relationship between them in cancer patient blood samples (57,91). Maheswaran *et al.* analyzed *EGFR* mutations in CTCs and ctDNA obtained from the same NSCLC patients and determined that genotyping seemed to be more sensitive in CTCs than in cfDNA and that the associated CTC quantification provided an important context in which to interpret these genotyping results (98). Thus, with the emergence of extremely sensitive technologies, complete genomic and transcriptomic profiles, drug sensitivity testing in CTC-derived cell cultures or in single cells might soon become a reality. Until now, the use of ctDNA has remained limited to research scenarios. However, an *EGFR* plasma test (TheraScreen[®] *EGFR* plasma PCR kit) has recently been approved in Europe and China for screening advanced NSCLC patients where it is impossible to obtain a tumor biopsy, allowing subsequent treatment with gefitinib in appropriate cases. Hence, new perspectives for implementing ctDNA in clinical settings are starting to open up (122).

There is mounting scientific evidence supporting the use of ctDNA for profiling and characterizing lung tumor molecular alterations as well as for monitoring therapies and identifying mutations associated with acquired drug

resistance (91,118,119,122,126,144,149). In this context ctDNA, rather than CTC analysis, is more appealing because plasma samples can be collected and analyzed without requiring prior enrichment and there is no need to isolate a rare cell population. Although pre-analytical conditions for ctDNA analysis must be further standardized, it seems that ctDNA, therefore, is likely to be the preferred option for genotyping and treatment-response monitoring. However, one important limitation of working with these samples is that *in situ* and morphological analyses using FISH and ICC (of particular interest in lung cancer for assessing *ALK* or *ROS1* status) cannot be performed with these samples (49,156). Another drawback of ctDNA is that, because of the high sensitivity of the different methodologies used, it also detects clinically irrelevant molecular changes.

In order to fully incorporate liquid biopsies into clinical practice some critical points must still be addressed: (I) a consensus is required regarding the best matrix of detection (CTC or ctDNA) for each required application; (II) a consensus regarding the ideal technical approach for each application is mandatory; (III) the pre-analytical phase should be standardized to obtain robust and reproducible results; (IV) investment and uptake of the currently available techniques is required in order to bring down prices which presently limit accessibility to patients.

Conclusions

Liquid biopsy has great potential for the management of lung cancer patients. Despite the numerous techniques and experimental approaches that have been established in this field, the common objective of all of them is to develop a useful, sensitive, specific, and real-time prognostic, predictive, and monitoring system using minimally-invasive samples, which can be easily transferred into the clinical practice. From our point of view, ctDNA analysis should be chosen for analysis of mutations, copy number aberrations, and DNA methylation changes, whereas CTC analysis provides the unique opportunity to study whole cells, thus allowing DNA, RNA, and protein-based molecular profiling, as well as use *in vivo* studies. It is likely that both CTCs and ctDNA will have complementary roles as cancer biomarkers and might be used in parallel for earlier lung cancer diagnosis, prediction of treatment responses, or detection of disease progression. Taking all of these arguments into account, we consider the real victory in this “war” is the genuine possibility these technologies create for

translating the concept of precision oncology into clinical practice. Liquid biopsies represent an important advance in the management of lung cancer in which CTCs and ctDNA are both expected to play complementary roles based on their relative strengths and limitations.

Acknowledgements

Funding: SC is supported by a grant from the Spanish Ministry of Economy and Competitiveness and the National Program for Research Aimed at the Challenges of Society, *RETOS-Colaboración* [RTC-2014-1532-1]. Funding was also provided from the *Red Temática de Investigación Cooperativa en Cáncer* (RTICC), *Instituto de Salud Carlos III* (ISCIII), the Spanish Ministry of Economy and Competitiveness, and the European Regional Development Fund (ERDF) “*Una manera de hacer Europa*” [RD12/0036/0025].

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Torre LA, Bray F, Siegel RL, et al. Global Cancer Statistics, 2012. *CA Cancer J Clin* 2015;65:87-108.
2. Siegel RL, Miller KDK, Jemal A. Cancer Statistics, 2016. *CA Cancer J Clin* 2016;66:7-30.
3. Ladanyi M. Targeted therapy of cancer: new roles for pathologists. *Mod Pathol* 2008;21 Suppl 2:S1.
4. Hirsch FR, Wynes MW, Gandara DR, et al. The tissue is the issue: personalized medicine for non-small cell lung cancer. *Clin Cancer Res* 2010;16:4909-11.
5. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol* 2011;12:175-80.
6. Kwak EL, Bang Y-J, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693-703.
7. Slebos RJ, Kibbelaar RE, Dalesio O, et al. K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N Engl J Med* 1990;323:561-5.
8. Kosaka T, Yatabe Y, Endoh H, et al. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res* 2004;64:8919-23.
9. Tokumo M, Toyooka S, Kiura K, et al. The relationship

- between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res* 2005;11:1167-73.
10. Sordella R, Bell DW, Haber DA, et al. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 2004;305:1163-7.
 11. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129-39.
 12. Sequist L V, Joshi VA, Janne PA, et al. Epidermal growth factor receptor mutation testing in the care of lung cancer patients. *Clin Cancer Res* 2006;12:4403s-4408s.
 13. Paez J, Sellers WR. PI3K/PTEN/AKT pathway. A critical mediator of oncogenic signaling. *Cancer Treat Res* 2003;115:145-67.
 14. Sequist L V, Martins RG, Spigel D, et al. First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. *J Clin Oncol* 2008;26:2442-9.
 15. Mok TS, Wu Y-L, Thongprasert S, et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. *N Engl J Med* 2009;361:947-57.
 16. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012;13:239-46.
 17. Keedy VL, Temin S, Somerfield MR, et al. Epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J Clin Oncol* 2011;29:2121-7.
 18. Fukuoka M, Wu YL, Thongprasert S, et al. Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol* 2011;29:2866-74.
 19. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2005;352:786-92.
 20. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* 2009;28 Suppl 1:S24-31.
 21. Lee JK, Shin J-Y, Kim S, et al. Primary resistance to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in patients with non-small-cell lung cancer harboring TKI-sensitive EGFR mutations: an exploratory study. *Ann Oncol* 2013;24:2080-7.
 22. Zhou W, Ercan D, Chen L, et al. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* 2009;462:1070-4.
 23. Aparicio S, Caldas C. The Implications of Clonal Genome Evolution for Cancer Medicine. *N Engl J Med* 2013;368:842-51.
 24. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883-92.
 25. Shaw AT, Kim D-W, Mehra R, et al. Ceritinib in ALK-Rearranged Non-Small-Cell Lung Cancer. *N Engl J Med* 2014;370:1189-97.
 26. Shu CA, Rizvi NA. Into the Clinic With Nivolumab and Pembrolizumab. *Oncologist* 2016;21:527-8.
 27. Valastyan S, Weinberg RA. Tumor Metastasis : Molecular Insights and Evolving Paradigms. *Cell* 2011;147:275-92.
 28. Chong CR, Janne PA. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med* 2013;19:1389-400.
 29. Piotrowska Z, Niederst MJ, Karlovich CA, et al. Heterogeneity Underlies the Emergence of EGFR T790M Wild-Type Clones Following Treatment of T790M-Positive Cancers with a Third-Generation EGFR Inhibitor. *Cancer Discov* 2015;5:713-22.
 30. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 2012;12:323-34.
 31. Burrell RA, McGranahan N, Bartek J, et al. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013;501:338-45.
 32. Piotrowska Z, Sequist LV. Treatment of EGFR-Mutant Lung Cancers After Progression in Patients Receiving First-Line EGFR Tyrosine Kinase Inhibitors : A Review. *JAMA Oncol* 2016;2:948-54.
 33. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nat Rev Cancer* 2008;8:329-40.
 34. Ashworth TR. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. *Aust Med J* 1869;14:146-9.
 35. Alix-Panabières C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. *Clin Chem* 2013;59:110-8.
 36. Kim MY, Oskarsson T, Acharyya S, et al. Tumor Self-Seeding by Circulating Cancer Cells. *Cell*

- 2009;139:1315-26.
37. Carlsson A, Nair VS, Luttggen MS, et al. Circulating tumor microemboli diagnostics for patients with non-small-cell lung cancer. *J Thorac Oncol* 2014;9:1111-9.
 38. Hou J-MM, Krebs M, Ward T, et al. Circulating Tumor Cells as a Window on Metastasis Biology in Lung Cancer. *Am J Pathol* 2011;178:989-96.
 39. Hou J-M, Greystoke A, Lancashire L, et al. Evaluation of circulating tumor cells and serological cell death biomarkers in small cell lung cancer patients undergoing chemotherapy. *Am J Pathol* 2009;175:808-16.
 40. Hou JM, Krebs MG, Lancashire L, et al. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. *J Clin Oncol* 2012;30:525-32.
 41. Krebs MG, Sloane R, Priest L, et al. Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J Clin Oncol* 2011;29:1556-63.
 42. Chen X-X, Bai F. Single-cell analyses of circulating tumor cells. *Cancer Biol Med* 2015;12:184-92.
 43. Lim LS, Hu M, Huang MC, et al. Microsieve lab-chip device for rapid enumeration and fluorescence in situ hybridization of circulating tumor cells. *Lab Chip* 2012;12:4388-96.
 44. O'Flaherty JD, Gray S, Richard D, et al. Circulating tumour cells, their role in metastasis and their clinical utility in lung cancer. *Lung Cancer* 2012;76:19-25.
 45. Funaki S, Sawabata N, Abulaiti A, et al. Significance of tumour vessel invasion in determining the morphology of isolated tumour cells in the pulmonary vein in non-small-cell lung cancer. *Eur J Cardiothorac Surg* 2013;43:1126-30.
 46. Parkinson DR, Dracopoli N, Petty BG, et al. Considerations in the development of circulating tumor cell technology for clinical use. *J Transl Med* 2012;10:138.
 47. Yap TA, Lorente D, Omlin A, et al. Circulating tumor cells: A multifunctional biomarker. *Clin Cancer Res* 2014;20:2553-8.
 48. Pantel K, Alix-Panabières C. Real-time liquid biopsy in cancer patients: Fact or fiction? *Cancer Res* 2013;73:6384-8.
 49. Alama A, Truini A, Coco S, et al. Prognostic and predictive relevance of circulating tumor cells in patients with non-small-cell lung cancer. *Drug Discov Today* 2014;19:1671-6.
 50. Wendel M, Bazhenova L, Boshuizen R, et al. Fluid biopsy for circulating tumor cell identification in patients with early- and late-stage non-small cell lung cancer: a glimpse into lung cancer biology. *Phys Biol* 2012;9:016005.
 51. Andree KC, van Dalum G, Terstappen LWMM. Challenges in circulating tumor cell detection by the CellSearch system. *Mol Oncol* 2016;10:395-407.
 52. Ozkumur E, Shah AM, Ciciliano JC, et al. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 2013;5:179ra47.
 53. Joosse SA, Gorges TM, Pantel K. Biology, detection, and clinical implications of circulating tumor cells. *EMBO Mol Med* 2014;7:1-11.
 54. Ignatiadis M, Lee M, Jeffrey SS. Circulating tumor cells and circulating tumor DNA: Challenges and opportunities on the path to clinical utility. *Clin Cancer Res* 2015;21:4786-800.
 55. Hofman V, Ilie MI, Long E, et al. Detection of circulating tumor cells as a prognostic factor in patients undergoing radical surgery for non-small-cell lung carcinoma: Comparison of the efficacy of the CellSearch Assay™ and the isolation by size of epithelial tumor cell method. *Int J Cancer* 2011;129:1651-60.
 56. Tanaka F, Yoneda K, Kondo N, et al. Circulating tumor cell as a diagnostic marker in primary lung cancer. *Clin Cancer Res* 2009;15:6980-6.
 57. Isobe K, Hata Y, Kobayashi K, et al. Clinical significance of circulating tumor cells and free DNA in non-small cell lung cancer. *Anticancer Res* 2012;32:3339-44.
 58. Hirose T, Murata Y, Oki Y, et al. Relationship of circulating tumor cells to the effectiveness of cytotoxic chemotherapy in patients with metastatic non-small-cell lung cancer. *Oncol Res* 2012;20:131-7.
 59. Marchetti A, Del Grammastro M, Felicioni L, et al. Assessment of EGFR mutations in circulating tumor cell preparations from NSCLC patients by next generation sequencing: Toward a real-time liquid biopsy for treatment. *PLoS One* 2014;9:e103883.
 60. Devriese LA, Bosma AJ, van de Heuvel MM, et al. Circulating tumor cell detection in advanced non-small cell lung cancer patients by multi-marker QPCR analysis. *Lung Cancer* 2012;75:242-7.
 61. Muínelo-Romay L, Vieito M, Abalo A, et al. Evaluation of circulating tumor cells and related events as prognostic factors and surrogate biomarkers in advanced NSCLC patients receiving first-line systemic treatment. *Cancers (Basel)* 2014;6:153-65.
 62. de Wit S, van Dalum G, Lenferink AT, et al. The detection of EpCAM(+) and EpCAM(-) circulating tumor cells. *Sci Rep* 2015;5:12270.

63. Cristofanilli M, Hayes DF, Budd GT, et al. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. *J Clin Oncol* 2005;23:1420-30.
64. de Bono JS, Scher HI, Montgomery RB, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* 2008;14:6302-9.
65. Cohen SJ, Punt CJ, Iannotti N, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol* 2008;26:3213-21.
66. Hiltermann TJN, Pore MM, Van den Berg A, et al. Circulating tumor cells in small-cell lung cancer: A predictive and prognostic factor. *Ann Oncol* 2012;23:2937-42.
67. Krebs MG, Metcalf RL, Carter L, et al. Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat Rev Clin Oncol* 2014;11:129-44.
68. Farace F, Massard C, Vimond N, et al. A direct comparison of CellSearch and ISET for circulating tumour-cell detection in patients with metastatic carcinomas. *Br J Cancer* 2011;105:847-53.
69. Krebs MG, Hou JM, Sloane R, et al. Analysis of circulating tumor cells in patients with non-small cell lung cancer using epithelial marker-dependent and -independent approaches. *J Thorac Oncol* 2012;7:306-15.
70. Mascalchi M, Falchini M, Maddau C, et al. Prevalence and number of circulating tumour cells and microemboli at diagnosis of advanced NSCLC. *J Cancer Res Clin Oncol* 2016;142:195-200.
71. Hofman V, Long E, Ilie M, et al. Morphological analysis of circulating tumour cells in patients undergoing surgery for non-small cell lung carcinoma using the isolation by size of epithelial tumour cell (ISET) method. *Cytopathology* 2012;23:30-8.
72. Pailler E, Adam J, Barthélémy A, et al. Detection of circulating tumor cells harboring a unique ALK rearrangement in ALK-positive non-small-cell lung cancer. *J Clin Oncol* 2013;31:2273-81.
73. Pailler E, Auger N, Lindsay CR, et al. High level of chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell lung cancer. *Ann Oncol* 2015;26:1408-15.
74. Faugeroux V, Pailler E, Auger N, et al. Clinical Utility of Circulating Tumor Cells in ALK-Positive Non-Small-Cell Lung Cancer. *Front Oncol* 2014;4:281.
75. Ilie M, Hofman V, Long E, et al. Current challenges for detection of circulating tumor cells and cell-free circulating nucleic acids, and their characterization in non-small cell lung carcinoma patients. What is the best blood substrate for personalized medicine? *Ann Transl Med* 2014;2:107.
76. Fiorelli A, D'Andrilli A, Anile M, et al. Sequential Bilateral Bronchoscopic Lung Volume Reduction With One-Way Valves for Heterogeneous Emphysema. *Ann Thorac Surg* 2016;102:287-94.
77. Chudasama D, Rice A, Soppa G, et al. Circulating tumour cells in patients with lung cancer undergoing endobronchial cryotherapy. *Cryobiology* 2015;71:161-3.
78. Kolostova K, Spicka J, Matkowski R, et al. Isolation, primary culture, morphological and molecular characterization of circulating tumor cells in gynecological cancers. *Am J Transl Res* 2015;7:1203-13.
79. Zhang Z, Shiratsuchi H, Lin J, et al. Expansion of CTCs from early stage lung cancer patients using a microfluidic co-culture model. *Oncotarget* 2014;5:12383-97.
80. Baccelli I, Schneeweiss A, Riethdorf S, et al. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat Biotechnol* 2013;31:539-44.
81. Maheswaran S, Haber DA. Ex Vivo Culture of CTCs: An Emerging Resource to Guide Cancer Therapy. *Cancer Res* 2015;75:2411-5.
82. Yu M, Bardia A, Aceto N, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* 2014;345:216-20.
83. Cayrefourcq L, Mazard T, Joosse S, et al. Establishment and characterization of a cell line from human circulating colon cancer cells. *Cancer Res* 2015;75:892-901.
84. Rossi E, Rugge M, Facchinetti A, et al. Retaining the long-survive capacity of Circulating Tumor Cells (CTCs) followed by xeno-transplantation: not only from metastatic cancer of the breast but also of prostate cancer patients. *Oncoscience* 2013;1:49-56.
85. Sun Y-F, Xu Y, Yang X-R, et al. Circulating stem cell-like epithelial cell adhesion molecule-positive tumor cells indicate poor prognosis of hepatocellular carcinoma after curative resection. *Hepatology* 2013;57:1458-68.
86. Hodgkinson CL, Morrow CJ, Li Y, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat Med* 2014;20:897-903.
87. Yuan D, Chen L, Li M, et al. Isolation and characterization of circulating tumor cells from human gastric cancer patients. *J Cancer Res Clin Oncol* 2015;141:647-60.
88. Sienel W, Seen-Hibler R, Mutschler W, et al. Tumour cells in the tumour draining vein of patients with non-small cell lung cancer: detection rate and clinical significance. *Eur J*

- Cardiothorac Surg 2003;23:451-6.
89. Hofman V, Bonnetaud C, Ilie MI, et al. Preoperative circulating tumor cell detection using the isolation by size of epithelial tumor cell method for patients with lung cancer is a new prognostic biomarker. *Clin Cancer Res* 2011;17:827-35.
 90. Naito T, Tanaka F, Ono A, et al. Prognostic impact of circulating tumor cells in patients with small cell lung cancer. *J Thorac Oncol* 2012;7:512-9.
 91. Punnoose EA, Atwal S, Liu W, et al. Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. *Clin Cancer Res* 2012;18:2391-401.
 92. Juan O, Vidal J, Gisbert R, et al. Prognostic significance of circulating tumor cells in advanced non-small cell lung cancer patients treated with docetaxel and gemcitabine. *Clin Transl Oncol* 2014;16:637-43.
 93. Hashimoto M, Tanaka F, Yoneda K, et al. Significant increase in circulating tumour cells in pulmonary venous blood during surgical manipulation in patients with primary lung cancer. *Interact Cardiovasc Thorac Surg* 2014;18:775-83.
 94. Bevilacqua S, Gallo M, Franco R, et al. A "live" biopsy in a small-cell lung cancer patient by detection of circulating tumor cells. *Lung Cancer* 2009;65:123-5.
 95. Li J, Shi S-B, Shi W-L, et al. LUNX mRNA-positive cells at different time points predict prognosis in patients with surgically resected nonsmall cell lung cancer. *Transl Res* 2014;163:27-35.
 96. Yamashita J, Matsuo A, Kurusu Y, et al. Preoperative evidence of circulating tumor cells by means of reverse transcriptase-polymerase chain reaction for carcinoembryonic antigen messenger RNA is an independent predictor of survival in non-small cell lung cancer: a prospective study. *J Thorac Cardiovasc Surg* 2002;124:299-305.
 97. Costa DB. Identification of somatic genomic alterations in circulating tumors cells: Another step forward in non-small-cell lung cancer? *J Clin Oncol* 2013;31:2236-9.
 98. Maheswaran S, Sequist L V, Nagrath S, et al. Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. *N Engl J Med* 2008;359:366-77.
 99. Ilie M, Long E, Butori C, et al. ALK-gene rearrangement: A comparative analysis on circulating tumour cells and tumour tissue from patients with lung adenocarcinoma. *Ann Oncol* 2012;23:2907-13.
 100. Mandel P, Metais P. Les acides nucléiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil* 1948;142:241-3.
 101. Gormally E, Caboux E, Vineis P, et al. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat Res* 2007;635:105-17.
 102. Anker P, Mulcahy H, Chen XQ, et al. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev* 1999;18:65-73.
 103. Bettgowda C, Sausen M, Leary RJ, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci Transl Med* 2014;6:224ra24.
 104. Crowley E, Di Nicolantonio F, Loupakis F, et al. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 2013;10:472-84.
 105. Jahr S, Hentze H, Englisch S, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 2001;61:1659-65.
 106. Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 2005;102:16368-73.
 107. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. *Clin Chem* 2015;61:112-23.
 108. Stroun M, Lyautey J, Lederrey C, et al. Alu repeat sequences are present in increased proportions compared to a unique gene in plasma/serum DNA: evidence for a preferential release from viable cells? *Ann N Y Acad Sci* 2001;945:258-64.
 109. Newman AM, Bratman S, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* 2014;20:548-54.
 110. Kiddess E, Jeffrey SS. Circulating tumor cells versus tumor-derived cell-free DNA: rivals or partners in cancer care in the era of single-cell analysis? *Genome Med* 2013;5:70.
 111. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008;14:985-90.
 112. El Messaoudi S, Rolet F, Mouliere F, et al. Circulating cell free DNA: Preanalytical considerations. *Clin Chim Acta* 2013;424:222-30.
 113. Qin J, Williams TL, Fernando MR. A novel blood collection device stabilizes cell-free RNA in blood during sample shipping and storage. *BMC Res Notes* 2013;6:380.
 114. Sozzi G, Conte D, Mariani L, et al. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. *Cancer Res* 2001;61:4675-8.
 115. Sirera R, Bremnes RM, Cabrera A, et al. Circulating DNA

- is a useful prognostic factor in patients with advanced non-small cell lung cancer. *J Thorac Oncol* 2011;6:286-90.
116. Ignatiadis M, Dawson SJ. Circulating tumor cells and circulating tumor DNA for precision medicine: dream or reality? *Ann Oncol* 2014;25:2304-13.
 117. Sirera R, Gil M, Blasco A, et al. Retrospective analysis of the prognostic role of p16 protein inactivation in plasma in patients with locally advanced non-small cell lung cancer. *Lung Cancer* 2008;61:104-8.
 118. Bordi P, Del Re M, Danesi R, et al. Circulating DNA in diagnosis and monitoring EGFR gene mutations in advanced non-small cell lung cancer. *Transl Lung Cancer Res* 2015;4:584-97.
 119. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2009;361:958-67.
 120. Rosell R, Molina MA, Serrano MJ. EGFR mutations in circulating tumour DNA. *Lancet Oncol* 2012;13:971-3.
 121. Kim HR, Lee SY, Hyun DS, et al. Detection of EGFR mutations in circulating free DNA by PNA-mediated PCR clamping. *J Exp Clin Cancer Res* 2013;32:50.
 122. Douillard J-Y, Ostoros G, Cobo M, et al. Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* 2014;9:1345-53.
 123. Oxnard GR, Paweletz CP, Kuang Y, et al. Noninvasive detection of response and resistance in egfrmutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin Cancer Res* 2014;20:1698-705.
 124. Diehl F, Li M, He Y, et al. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* 2006;3:551-9.
 125. Taniguchi K, Uchida J, Nishino K, et al. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. *Clin Cancer Res* 2011;17:7808-15.
 126. Murtaza M, Dawson S-J, Tsui DWY, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013;497:108-12.
 127. Couraud S, Vaca-Paniagua F, Villar S, et al. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. *Clin Cancer Res* 2014;20:4613-24.
 128. Martinez P, McGranahan N, Birkbak NJ, et al. Computational optimisation of targeted DNA sequencing for cancer detection. *Sci Rep* 2013;3:3309.
 129. Kinde I, Wu J, Papadopoulos N, et al. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* 2011;108:9530-5.
 130. Lebofsky R, Decraene C, Bernard V, et al. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol* 2015;9:783-90.
 131. Newman AM, Lovejoy AF, Klass DM, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* 2016;34:547-55.
 132. Goto K, Ichinose Y, Ohe Y, et al. Epidermal growth factor receptor mutation status in circulating free DNA in serum: from IPASS, a phase III study of gefitinib or carboplatin/paclitaxel in non-small cell lung cancer. *J Thorac Oncol* 2012;7:115-21.
 133. Huang Z, Wang Z, Bai H, et al. The detection of EGFR mutation status in plasma is reproducible and can dynamically predict the efficacy of EGFR-TKI. *Thoracic Cancer* 2012;3:334-40.
 134. Zhao X, Han RB, Zhao J, et al. Comparison of epidermal growth factor receptor mutation statuses in tissue and plasma in stage I-IV non-small cell lung cancer patients. *Respiration* 2013;85:119-25.
 135. Kim ST, Sung JS, Jo UH, et al. Can mutations of EGFR and KRAS in serum be predictive and prognostic markers in patients with advanced non-small cell lung cancer (NSCLC)? *Med Oncol* 2013;30:328.
 136. Wang S, Han X, Hu X, et al. Clinical significance of pretreatment plasma biomarkers in advanced non-small cell lung cancer patients. *Clin Chim Acta* 2014;430:63-70.
 137. Weber B, Meldgaard P, Hager H, et al. Detection of EGFR mutations in plasma and biopsies from non-small cell lung cancer patients by allele-specific PCR assays. *BMC Cancer* 2014;14:294.
 138. Karachaliou N, Mayo-de las Casas C, Queralt C, et al. Association of EGFR L858R Mutation in Circulating Free DNA With Survival in the EURTAC Trial. *JAMA Oncol* 2015;1:149-57.
 139. Thress KS, Brant R, Carr TH, et al. EGFR mutation detection in ctDNA from NSCLC patient plasma: A cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer* 2015;90:509-15.
 140. Wei Z, Shah N, Deng C, et al. Circulating DNA addresses cancer monitoring in non small cell lung cancer patients for detection and capturing the dynamic changes of the disease. *Springerplus* 2016;5:531.

141. Zheng D, Ye X, Zhang MZ, et al. Plasma EGFR T790M ctDNA status is associated with clinical outcome in advanced NSCLC patients with acquired EGFR-TKI resistance. *Sci Rep* 2016;6:20913.
142. Sozzi G, Conte D, Leon M, et al. Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J Clin Oncol* 2003;21:3902-8.
143. Fournié GJ, Courtin JP, Laval F, et al. Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours. *Cancer Lett* 1995;91:221-7.
144. Li BT, Drilon A, Johnson ML, et al. A prospective study of total plasma cell-free DNA as a predictive biomarker for response to systemic therapy in patients with advanced non-small-cell lung cancers. *Ann Oncol* 2016;27:154-9.
145. Ludovini V, Pistola L, Gregorc V, et al. Plasma DNA, microsatellite alterations, and p53 tumor mutations are associated with disease-free survival in radically resected non-small cell lung cancer patients: a study of the perugia multidisciplinary team for thoracic oncology. *J Thorac Oncol* 2008;3:365-73.
146. Gautschi O, Huegli B, Ziegler A, et al. Origin and prognostic value of circulating KRAS mutations in lung cancer patients. *Cancer Lett* 2007;254:265-73.
147. Camps C, Jantus-Lewintre E, Cabrera A, et al. The identification of KRAS mutations at codon 12 in plasma DNA is not a prognostic factor in advanced non-small cell lung cancer patients. *Lung Cancer* 2011;72:365-9.
148. Wang S, An T, Wang J, et al. Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer. *Clin Cancer Res* 2010;16:1324-30.
149. Mok T, Wu Y-L, Lee JS, et al. Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy. *Clin Cancer Res* 2015;21:3196-203.
150. Douillard J-Y, Ostoros G, Cobo M, et al. First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. *Br J Cancer* 2014;110:55-62.
151. He W, Xu D, Wang Z, et al. Detecting ALK-rearrangement of CTC enriched by nanovelcro chip in advanced NSCLC patients. *Oncotarget* 2016. [Epub ahead of print].
152. Marchetti A, Palma JF, Felicioni L, et al. Early Prediction of Response to Tyrosine Kinase Inhibitors by Quantification of EGFR Mutations in Plasma of NSCLC Patients. *J Thorac Oncol* 2015;10:1437-43.
153. Marcq M, Vallée A, Bizieux A, et al. Detection of EGFR Mutations in the Plasma of Patients with Lung Adenocarcinoma for Real-Time Monitoring of Therapeutic Response to Tyrosine Kinase Inhibitors? *J Thorac Oncol* 2014;9:e49-50.
154. Yang X, Zhuo M, Ye X, et al. Quantification of mutant alleles in circulating tumor DNA can predict survival in lung cancer. *Oncotarget* 2016;7:20810-24.
155. Sorensen BS, Wu L, Wei W, et al. Monitoring of epidermal growth factor receptor tyrosine kinase inhibitor-sensitizing and resistance mutations in the plasma DNA of patients with advanced non-small cell lung cancer during treatment with erlotinib. *Cancer* 2014;120:3896-901.
156. Zhang Z, Ramnath N, Nagrath S. Current Status of CTCs as Liquid Biopsy in Lung Cancer and Future Directions. *Front Oncol* 2015;5:209.

Cite this article as: Calabuig-Fariñas S, Jantus-Lewintre E, Herreros-Pomares A, Camps C. Circulating tumor cells versus circulating tumor DNA in lung cancer—which one will win? *Transl Lung Cancer Res* 2016;5(5):466-482. doi: 10.21037/tlcr.2016.10.02