UNIVERSIDAD DE GRANADA

Programa de doctorado en Bioquímica y Biología Molecular (B16.56.1)

PhD dissertation

Study of novel biomarkers and molecular therapies against lung cancer

PhD candidate:

Alberto M. Arenas Molina

PhD supervisor:

Pedro P. Medina Vico, PhD

Editor: Universidad de Granada. Tesis Doctorales
Autor: Alberto Manuel Arenas Molina
ISBN: 978-84-1195-499-0
URI: https://hdl.handle.net/10481/96715

“You should enjoy the little detours to the fullest.
Because that's where you will find things
more important than what you want.”

Yoshihiro Togashi

Table of contents

Table of contents 7
List of abbreviations 11
Abstract 15
Resumen 17
Objectives 21
Objetivos 22
Chapter 1. Introduction 25
1.1 Lung cancer in the global health landscape 25
1.1.1 Lung cancer is a preventable global health emergency 25
1.1.2 Classification of lung cancer 27
1.1.3 Current therapies against lung cancer 27
1.2 The molecular origins of cancer 29
1.2.1 From normal to tumor: the tumorigenesis process 29
1.2.2 Driver oncogenes and tumor suppressor genes in NSCLC 30
1.3 Non-coding RNAs: the forgotten ones 34
1.3.1 Micro-RNAs 34
1.3.2 Long non-coding RNAs 39
1.3.3 Other non-coding RNAs 43
1.4 CRISPR/Cas9: a new weapon against cancer 46
1.4.1 The CRISPR/Cas system: discovery, function, and variants 46
1.4.2 CRISPR/Cas9 applications in biomedicine and cancer 49
1.4.3 The ever-growing CRISPR toolbox 51
1.4.4 Limitations and concerns of CRISPR in today's medicine 54
Chapter 2. Materials and methods 59
2.1 Molecular biology methods 59
2.1.1 DNA extraction and PCR 59
2.1.2 RNA extraction 59
2.1.3 RT-qPCR 60
2.1.4 Protein extraction and Western blot 60

2.1.5 Vector cloning 61
2.1.6 Sanger sequencing 62
2.1.7 T7 endonuclease assay 62
2.1.8 Biotin pull-down 62
2.2 Cell experiments *in vitro* and *in vivo* 64
2.2.1 Cell culture 64
2.2.2 Cell viability assays 64
2.2.3 Colony assay 65
2.2.4 Scratch wound-healing assay 65
2.2.5 Tumor cell-derived xenografts in mice 65
2.3 Gene delivery methods 66
2.3.1 Transfection of plasmids and miRNA mimics 66
2.3.2 RNPs delivery 66
2.3.3 Lentivirus preparation 66
2.3.4 Lentiviral infection and clone selection 67
2.3.5 Adenoviral infection 67
2.4 Screenings and high-throughput methods 68
2.4.1 LncRNA array profiling 68
2.4.2 RNA-sequencing 68
2.4.3 CRISPR-screening 69
2.5 Bioinformatic methods, statistics, and ethics 73
2.5.1 Omics databases and bioinformatic analysis tools 73
2.5.2 Analysis of TCGA-LUAD data 73
2.5.3 Statistical analysis 74
2.5.4 Ethics for human patients and animal procedures 74
Chapter 3. Characterization of novel lncRNA biomarkers for LUAD 77
3.1 Background 77
3.2 Results 79
3.2.1 Selection of a candidate lncRNA 79
3.2.2 DLG2-AS1 is downregulated in LUAD patients 79
3.2.3 Phenotypical assays in DLG2-AS1 restoration cell models 80
3.2.4 DLG2-AS1 does not act as a cis-regulator for DLG2 81
3.2.5 DLG2-AS1 shows a good potential as a biomarker for LUAD 84

3.3 Discussion 86
Chapter 4. Validation of the oncogenic potential of a mutated miRNA 91
4.1 Background 91
4.1.1 Mutations in miRNA seed regions are rare, yet existing 91
4.1.2 The seed region of miR-133b is mutated in LUAD. 92
4.2 Results 93
4.2.1 Detection of miR-133b mutation at gDNA and RNA levels 93
4.2.2 Mutant miR-133b has an oncogenic role in LUAD overexpression cell models 94
4.2.3 Experimental validation of predicted miR-133b targets 97
4.2.4 Discovery of novel experimental targets of miR-133b$^{WT/MUT}$ 98
4.3 Discussion 103
Chapter 5. CRISPR-based molecular therapy against *KRAS*-mutant LUAD 109
5.1 Background 109
5.2 Results 111
5.2.1 Optimization of delivery and design of sgRNAs 111
5.2.2 KRAS$^{G12C/G12D}$ sgRNAs are specific and do not target KRAS^{WT} 112
5.2.3 CRISPR-KRAS therapy impairs cell viability of KRAS-mutant cell lines 114
5.2.4 Development of an inducible murine model for CRISPR-KRAS therapy 115
5.3 Discussion 118
Chapter 6. CRISPR-screening for the detection of collateral dependencies 125
6.1 Background 125
6.1.1 PKP1 is an oncogene and potential therapeutic target in LUSC 125
6.1.2 CRISPR-screening: a molecular tool for the discovery of collateral dependencies 126
6.2 Results 128
6.2.1 Library titration and Cas9 activity checking 128
6.2.2 CRISPR-screening and NGS sample preparation 129
6.2.3 Quality check of NGS data 130
6.2.4 PKP1-KO cells show a collateral dependency on mitochondrial function 131

6.3 Discussion 134
Conclusions 139
Conclusiones 140
Supplementary material 143
References 151
Annex 177
Publications during the PhD 179
Funding and fellowships 181
Copyright permissions 182

List of abbreviations

AAV	Adeno-associated virus
AdV	Adenoviral particle
AGO	Argonaute protein
AKT	Protein kinase B
ASO	Antisense oligonucleotide
AUC	Area under curve
bp	Base pair
CD	Collateral dependency
cDNA	Copy DNA
ceRNA	Competing endogenous RNA
CFU	Colony forming units
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
DGCR8	DiGeorge syndrome critical region 8
DLG2-AS1	Disks large homolog 2, antisense 1
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DSB	Double-strand break
EGFR	Epithelial growth factor receptor
EMA	European Medicines Agency
EMT	Epithelial-mesenchymal transition
EV	Empty vector
FDA	U.S. Food and Drug Administration
FBS	Fetal bovine serum
GAP	GTPase-activating protein
gDNA	Genomic DNA
GDP	Guanosine diphosphate
GEF	Guanosine exchange factor
GSEA	Gene set enrichment analysis
GTP	Guanosine triphosphate
HDR	Homology-directed repair
HiFi-Cas9	High-fidelity Cas9
HRP	Horseradish peroxidase
ICE	Inference of CRISPR edits
kb	Kilobase
kDa	Kilodalton
KRAS	Kristen-rat sarcoma viral proto-oncogene
LCLC	Large cell lung cancer
lncRNA	Long non-coding RNA
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
LV	Lentiviral particle

List of abbreviations

MAPK	Mitogen-activated protein kinases
MEF	Mouse embryonic fibroblast
miRNA	Micro-RNA
mRNA	Messenger RNA
MRP	Mitochondrial ribosome protein
MTERF4	Mitochondrial termination factor 4
mTOR	Mammalian target of rapamycin
MUT	Mutant
ncRNA	Non-coding RNA
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NSCLC	Non-small cell lung cancer
ORF	Open reading frame
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PI3K	Phosphoinositide 3-kinase
piRNA	PIWI-interacting RNA
PRC2	Polycomb repressive complex 2
qPCR	Quantitative polymerase chain reaction
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RNP	Ribonucleoprotein
ROC	Receiver operating characteristic
rRNA	Ribosomal RNA
RT-qPCR	Reverse-transcription, quantitative polymerase chain reaction
SCLC	Small cell lung cancer
SDS-PAGE	Sodium dodecyl sulfate - protein acrylamide gel electrophoresis
sgRNA	Single guide RNA
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SNV	Single nucleotide variation
SpCas9	*Streptococcus pyogenes* CRISPR-associated protein 9
TKI	Tyrosine kinase inhibitor
tracrRNA	*Trans*-activating CRISPR-RNA
tRNA	Transference RNA
UTR	Untranslated region
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHO	World's Health Organization
WT	Wildtype

Abstract / Resumen

Abstract

Cancer is a global health emergency nowadays, and lung cancer is the subtype with the highest mortality overall. Understanding the molecular mechanisms that originate tumors has been the focus of cancer research in the last decades. This has allowed us to discover novel biomarkers for improving the diagnosis and the prognosis of patients, and to design novel targeted therapeutic strategies. Despite the many advances, lung cancer still accounts for millions of deaths worldwide every year, and the 5-year survival rate remains at around 20%. There is therefore an urgency to continue studying this disease, and to develop novel biomarkers and targeted therapies to improve the diagnosis, prognosis, and treatment of lung cancer patients. In this thesis, we have aimed to increase our understanding of the molecular biology of lung cancer from two different angles:

1. Identification of novel non-coding RNAs as lung cancer biomarkers.

The identification of frequently dysregulated or mutated genes in cancer is of the utmost importance for the discovery of novel diagnostic/prognostic biomarkers, and potential novel targets for molecular therapy. However, to date most of the research has been focused on the protein-coding part of the genome, often overlooking non-coding RNAs, such as micro-RNAs (miRNAs) and long non-coding RNAs (lncRNAs). Non-coding RNAs exert key roles in gene expression regulation and appear altered in several diseases, including lung cancer. Therefore, in this thesis we have focused on discovering altered lncRNAs and miRNAs in lung adenocarcinoma (LUAD) patients, to characterize them and validate their oncogenic potential and possible use as biomarkers.

Thus, we have discovered a lncRNA, *DLG2-AS1*, which we found to be downregulated in a cohort of 65 LUAD patient samples (44/65), compared to their paired, adjacent non-tumor tissues. *DLG2-AS1* showed an area under curve (AUC) of 0.726, which represents a good biomarker potential compared to other already validated lncRNA and protein biomarkers. We also aimed to demonstrate the tumor suppressor role of *DLG2-AS1*, and its *cis*-regulatory role over the overlapping gen *DLG2* in LUAD cell lines. However, we observed no phenotypical differences upon its overexpression, nor did we find a correlation between *DLG2-AS1* and *DLG2* expression in our models.

Furthermore, we also studied the oncogenic potential of a miRNA, *miR-133b*, which was found somatically mutated in its seed region in a previous miRNA study in LUAD patients. The overexpression of mutant and wildtype versions of *miR-133b* in LUAD cell lines showed a strong oncogenic effect of mutant *miR-133b* in the A549 and H2126 cell lines, compared to its wildtype version. A subsequent RNA sequencing analysis provided us with a series of downregulated cancer-related genes (e.g., *TAGLN2, and PTBP* for wildtype; *QKI*, and *RHOQ* for mutant)

which could explain the phenotype we observed if they are proven targets for *miR-133b* in future experimental validations.

2. Development of novel molecular therapies based on the CRISPR/Cas9 system.

On the other hand, some frequently mutated genes can serve not only as biomarkers, but also as targets for molecular therapies against cancer, such as the oncogene *KRAS*, which is the most frequently mutated driver oncogene in LUAD. For years, mutant *KRAS* has been an elusive target for therapy until just recently, with two novel drugs targeting $KRAS^{G12C}$ mutant tumors having reached the clinic since 2021: sotorasib and adagrasib. However, these drugs have faced some difficulties in late-phase clinical trials, as most of the patients develop resistances and eventually relapse in a one-year time. Therefore, novel strategies for targeting *KRAS*-mutant tumors need to be explored, and the CRISPR/Cas9 technology might be the solution. CRISPR/Cas9 can specifically delete mutant alleles of a target gene without affecting the wildtype, non-tumor one, therefore selectively eliminating tumor cells harboring such mutations. We aimed to optimize the design and delivery, and to test the specificity and efficiency of a novel molecular therapy against $KRAS^{G12C/G12D}$ LUAD using the CRISPR technology and a high-fidelity version of Cas9 (CRISPR-KRAS therapy). After assessing the specificity of the designed single-guide RNAs, our CRISPR-KRAS therapy successfully caused a ~70% reduction in cell viability in *KRAS*-mutant LUAD cell lines, without affecting *KRAS*-wildtype cells.

Finally, the cellular mechanisms by which tumor cells become resistant to targeted therapies against oncogenes can be studied by performing a high-throughput CRISPR screening. This allows us to discover the collateral dependencies that arise in the surviving cells after knocking out an oncogene. To apply this cutting-edge technology, we performed a CRISPR-screening on knockout (KO) clones previously generated in our laboratory for a squamous cell lung cancer (LUSC) oncogene, *PKP1*. After the screening, we discovered that the most essential genes for *PKP1*-KO clones included a large number of genes implicated in mitochondrial translation and transcription termination, such as *MTERF4* and several mitoribosome proteins. This opens a new line of research for a future targeted co-therapy against LUSC, which could involve PKP1 inhibition along with some mitochondrial translation inhibitor antibiotics, to enhance its effect and prevent resistances.

Resumen

Actualmente el cáncer es una emergencia sanitaria global, y el cáncer de pulmón el subtipo con una mayor mortalidad. Entender los mecanismos moleculares que originan los tumores ha sido el foco de la investigación oncológica en las últimas décadas. Esto nos ha permitido descubrir nuevos biomarcadores para mejorar el diagnóstico y el pronóstico de los pacientes, y diseñar nuevas estrategias terapéuticas dirigidas. A pesar de los muchos avances, el cáncer de pulmón es aún responsable de millones de muertes cada año en el mundo, y la tasa de supervivencia a 5 años se mantiene en torno al 20%. Hay, por tanto, una urgencia en continuar estudiando esta enfermedad, y en desarrollar nuevos biomarcadores y terapias dirigidas que mejoren el diagnóstico, pronóstico y tratamiento de los pacientes. En esta tesis se ha tratado de incrementar el conocimiento sobre la biología molecular del cáncer de pulmón desde dos ángulos distintos:

1. Identificación de nuevos ARN no codificantes como biomarcadores de cáncer de pulmón.

La identificación de genes frecuentemente desregulados o mutados en cáncer es de vital importancia para el descubrimiento de nuevos biomarcadores diagnósticos/pronósticos, así como potenciales nuevas dianas terapéuticas. Sin embargo, la mayoría de los estudios hasta la fecha se han centrado en la parte codificante del genoma, a menudo olvidando los ARN no codificantes, tales como micro-ARN (miRNA) y ARN largos no codificantes (lncRNA). Los ARN no codificantes ejercen funciones reguladoras de la expresión génica muy importantes y aparecen alterados en numerosas enfermedades, incluido el cáncer de pulmón. Por tanto, en esta tesis nos hemos centrado en hallar lncRNA y miRNA alterados en pacientes con adenocarcinoma de pulmón (LUAD), para caracterizarlos y validar su potencial oncogénico y posible uso como biomarcadores.

Así, hemos descubierto un lncRNA*, DLG2-AS1*, que aparece regulado a la baja en muestras tumorales de una cohorte de 65 pacientes de LUAD (44/65), comparado con sus respectivas muestras pareadas no tumorales. *DLG2-AS1* mostró un área bajo la curva (AUC) de 0,726; lo que representa un buen potencial como biomarcador comparado con otros lncRNA y proteínas biomarcadores ya validados. También intentamos demostrar el papel como supresor tumoral de *DLG2-AS1* y su papel regulador en cis sobre el gen solapante *DLG2* en líneas celulares de LUAD, pero no conseguimos observar ninguna diferencia fenotípica tras su sobreexpresión, ni vimos una correlación entre los niveles de expresión de *DLG2-AS1* y *DLG2* en nuestros modelos.

Además, estudiamos el potencial oncogénico de un miRNA, *miR-133b*, el cual se halló mutado somáticamente en su región semilla en un estudio previo de

miRNA en pacientes con LUAD. La sobreexpresión de las versiones mutante y silvestre de *miR-133b* en líneas celulares de LUAD nos mostraron un fuerte efecto oncogénico de *miR-133b* mutante en las líneas A549 y H2126 comparado con su versión silvestre. Un análisis de secuenciación de ARN posterior nos reveló una serie de genes relacionados con cáncer (p. ej., *TAGL2 y PTBP1* para el silvestre*; QKI* y *RHOQ* para el mutante) regulados a la baja que podrían explicar el fenotipo observado si demuestran ser dianas de *miR-133b* en futuras validaciones experimentales.

2. Desarrollo de nuevas terapias moleculares basadas en el sistema CRISPR/Cas9.

Por otro lado, algunos genes frecuentemente mutados en cáncer pueden servir no sólo como biomarcadores, sino también como dianas terapéuticas contra el cáncer, como ocurre con *KRAS*, el oncogén más frecuentemente mutado en LUAD. *KRAS* mutante ha sido hasta hace muy poco una diana elusiva, con tan solo dos fármacos dirigidos contra $KRAS^{G12C}$ aprobados desde 2021: el sotorasib y el adagrasib. Sin embargo, estos fármacos se han enfrentado a diversas dificultades en los últimos ensayos clínicos, ya que la mayoría de los pacientes acababan desarrollando resistencias y sufriendo una recidiva en un periodo de un año. Por ello, se necesita seguir explorando nuevas estrategias terapéuticas contra tumores *KRAS*-mutantes, y la tecnología de edición génica CRISPR/Cas9 podría ser la solución. CRISPR/Cas9 puede de forma específica eliminar los alelos mutantes de un gen diana sin afectar a su versión silvestre no tumoral. Así, de forma altamente selectiva, se pueden eliminar las células tumorales que posean dicha mutación. En esta tesis se trató de optimizar el diseño y la entrega, y medir la especificidad y eficiencia de una nueva terapia molecular contra $KRAS^{G12C/G12D}$ usando la tecnología CRISPR y una versión de alta fidelidad del enzima Cas9 (terapia CRISPR-KRAS). Tras comprobar la especificidad de los ARN guías diseñados, nuestra terapia CRISPR-KRAS consiguió reducir la viabilidad celular hasta alrededor de un 70% en líneas celulares de LUAD *KRAS* mutantes, sin afectar a las células *KRAS* silvestres.

Por último, los mecanismos celulares por los cuales los tumores se vuelven resistentes a las terapias dirigidas contra oncogenes pueden estudiarse por medio de un cribado CRISPR (*CRISPR-screening*). Esta técnica nos permite conocer las dependencias colaterales que aparecen en las células supervivientes tras eliminar un determinado oncogén. Para aplicar esta tecnología puntera, realizamos un cribado CRISPR sobre clones "*knockout*" (KO) que habíamos generado previamente en el laboratorio para un oncogén (*PKP1*) en cáncer de pulmón epidermoide (LUSC). Tras el cribado, descubrimos que entre los genes más esenciales para los clones *PKP1*-KO se hallaban un gran número de genes implicados en la traducción y terminación de la transcripción mitocondriales, como *MTERF4* y numerosas proteínas mitorribosomales. Esto abre una nueva línea de investigación para una futura co-terapia dirigida contra el LUSC, que involucraría un inhibidor de PKP1 junto con algún antibiótico inhibidor de la traducción mitocondrial, con el objetivo de aumentar su efecto y evitar la aparición de resistencias.

Objectives / Objetivos

Objectives

The general objective of this thesis was to identify, validate, and develop novel biomarkers and molecular therapeutic strategies against lung cancer. The specific objectives were:

Objective 1. **To detect and validate altered non-coding RNAs in lung cancer patients that could serve as novel biomarkers for the disease.**

1.1 To identify and characterize a dysregulated long non-coding RNA in lung adenocarcinoma patient samples as a novel candidate biomarker for the disease.

1.2 To validate the oncogenic driver potential of a mutated micro-RNA in lung adenocarcinoma patients.

Objective 2. **To assess the therapeutic potential of a CRISPR-based molecular therapy against lung cancer oncogenes.**

2.1 To optimize and assess the specificity and efficiency of a CRISPR/Cas9-based strategy to specifically target *KRAS*-mutant (G12C/G12D) lung adenocarcinoma cells.

2.2 To perform a CRISPR-screening to study the collateral dependencies arisen after the knockout of the *PKP1* oncogene in lung squamous cell carcinoma.

Objetivos

El objetivo general de esta tesis ha sido identificar, validar y desarrollar nuevos biomarcadores y estrategias de terapia molecular contra el cáncer de pulmón. Los objetivos específicos fueron:

Objetivo 1. **Detectar y validar ARN no codificantes alterados en pacientes con cáncer de pulmón que puedan servir como nuevos biomarcadores para la enfermedad.**

1.1 Identificar y caracterizar un ARN largo no codificante desregulado en muestras de pacientes con adenocarcinoma de pulmón como candidato a ser un nuevo biomarcador para la enfermedad.

1.2 Validar el potencial como gen conductor del cáncer de un micro-ARN mutado en pacientes con adenocarcinoma de pulmón.

Objetivo 2. **Estudiar el potencial terapéutico de una terapia molecular basada en CRISPR contra oncogenes en cáncer de pulmón.**

2.1 Optimizar y estudiar la especificidad y eficiencia de una estrategia basada en CRISPR/Cas9 para eliminar de forma específica células de adenocarcinoma de pulmón *KRAS* mutantes (G12C/G12D).

2.2 Realizar un cribado CRISPR para estudiar las dependencias colaterales que aparecen tras la eliminación del oncogén *PKP1* en cáncer de pulmón epidermoide.

Chapter 1.
Introduction

Chapter 1. Introduction

This first chapter is meant to serve as a general introduction to the topics that will be discussed in this PhD dissertation. In addition, for each results chapter (Chapters 3 to 6) a brief, more specific background section will also be included to contextualize the obtained results with the state-of-the-art and the previous research carried out in our group. This introduction begins by highlighting the impact of lung cancer in nowadays' global health landscape. Following that, it explains the molecular bases of cancer, with a focus on concepts such as oncogenes/tumor suppressor genes, and driver mutations. Then, it discusses the roles of non-coding RNAs as potential driver elements in tumorigenesis. Finally, it details the functioning of the CRISPR/Cas9 gene editing technology, and some of its applications to today's biomedicine and the development of novel therapies against cancer.

1.1 Lung cancer in the global health landscape

1.1.1 Lung cancer is a preventable global health emergency

Cancer is the common term for a group of diseases, also known as neoplasms or malignant tumors, which consist in the rapid, uncontrolled growth of aberrant cells in our bodies, leading to the invasion of adjacent tissues and organs, which ultimately causes their malfunction. Cancer usually begins as a mass of aberrant cells forming a primary tumor, which then grows and causes local invasion of healthy tissues, and eventually a metastasis process might occur, when cancer cells gain the potential to spread to distal places in the body, affecting the normal function of several organs, and finally leading to the death of the patient. According to the World Cancer Report 2020 from the World's Health Organization (WHO), before the COVID-19 pandemics, cancer was the leading cause of death in most of the high-income countries in the world, overtaking cardiovascular diseases (Figure 1) (Wild et al., 2020).

Cancer can be roughly classified according to the primary tumor initiation site. Lung cancer is the second most prevalent cancer in both sexes, after breast cancer, and it stands out as the top cancer type in terms of mortality, accounting for around 1.8 million deaths worldwide in 2020 (Figure 2). The primary cause for lung cancer is active or passive tobacco smoking, accounting for 63% of all lung cancer deaths (Forouzanfar et al., 2016). Current lung cancer incidence is a clear reflection of the smoking trends in the last 20-30 years. The "smoking epidemics", as it has been named, is slowly declining in high-income countries, but starting to hit regions with previously low incidence of lung cancer, mainly Asian and Pacific countries (Dai et al., 2022). Following the trends of the smoking epidemics, in a few

decades this will result in a rising number of lung cancer cases in such regions, first in men, then in women (Thun et al., 2012). These alarming numbers add up to the poor prognosis that lung cancer patients have nowadays, as the overall 5-year survival rate remains at around 20% (Siegel et al., 2023). Therefore, lung cancer is a global health emergency nowadays, and several efforts are being made (i) towards the prevention of new cases by promoting healthier lifestyles and minimizing exposure to carcinogens (i.e., smoking habitude), and (ii) towards investing in cancer research to improve the diagnosis and prognosis, and to develop novel treatments for these patients.

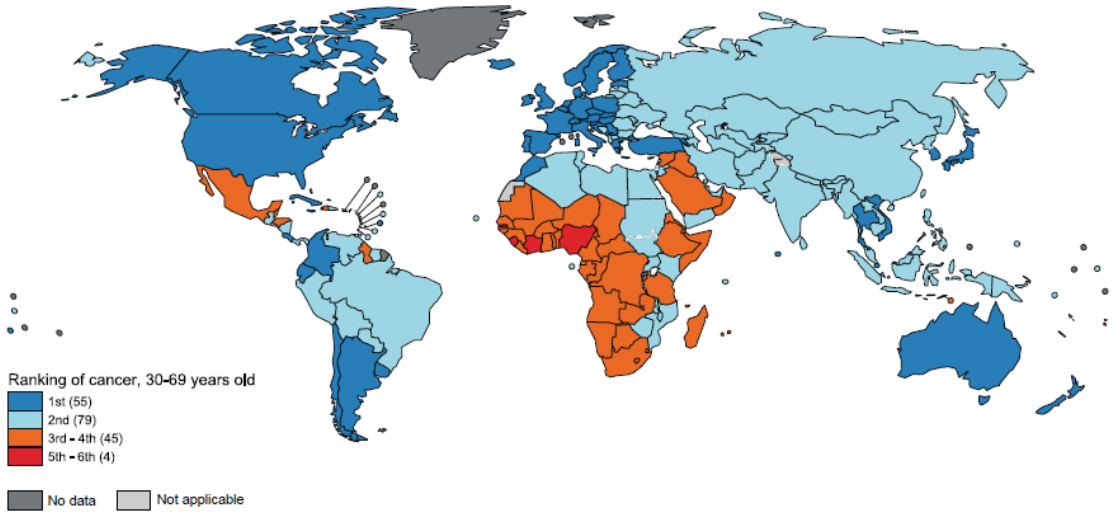


Figure 1. Ranking of cancer as the leading cause of premature death (ages 30-69) in the world. Figure extracted from the WHO's World Cancer Report 2020 (Wild et al., 2020) *with copyright permission of the International Agency for Research on Cancer (IARC)/WHO.*

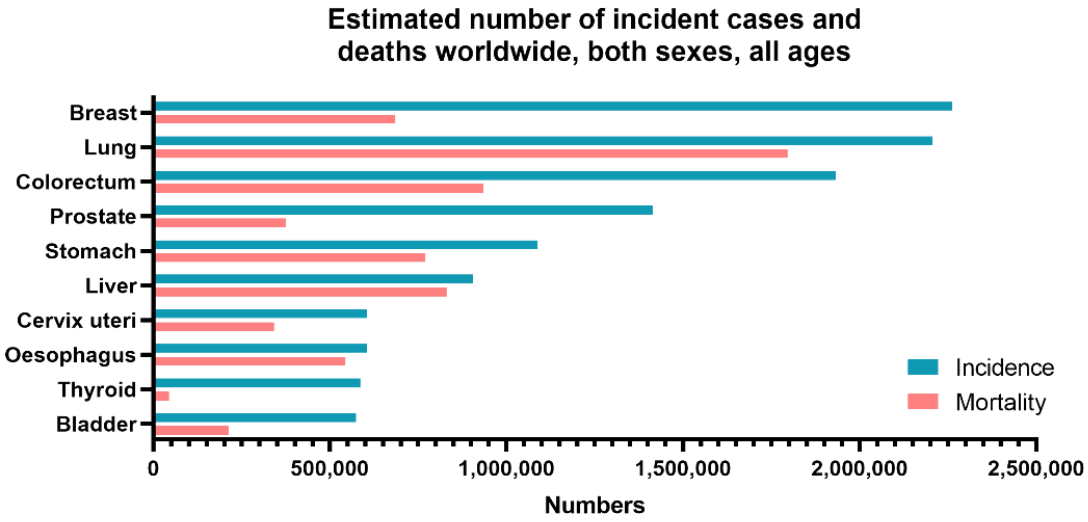


Figure 2. Number of estimated global new cancer cases and deaths in 2020. Graph generated with data from the Global Cancer Observatory (https://gco.iarc.fr/).

1.1.2 Classification of lung cancer

Lung cancer has been traditionally divided in two main types, according to the histomorphological features that the tumors showed upon surgery: non-small cell lung cancer (NSCLC), the main subtype, accounting for 80-85% of all lung cancers; and small cell lung cancer (SCLC, 10-15%) (Barta et al., 2019). In addition, there are some other rare types of lung tumors (<5% of all lung cancer cases), such as lung carcinoid tumors (Barta et al., 2019), and some newly defined tumor subtypes whose distinction from others is "omics"-based, and they can now be more accurately classified thanks to the advances in molecular genetics from the past decade (Nicholson et al., 2022).

Following the classical, histomorphological classification, NSCLC can be further subdivided into different subtypes, the biggest three being lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and large cell lung carcinoma (LCLC). LUAD is the most common subtype of lung cancer overall (40-50% of total), and the most frequent in non-smoker patients. LUSC is the second most common (~20% of total), and the most related to a smoking history. Finally, LCLC is a much rarer (>5% of total) and faster-growing cancer than LUAD or LUSC (Figure 3) (M. Wang et al., 2021).

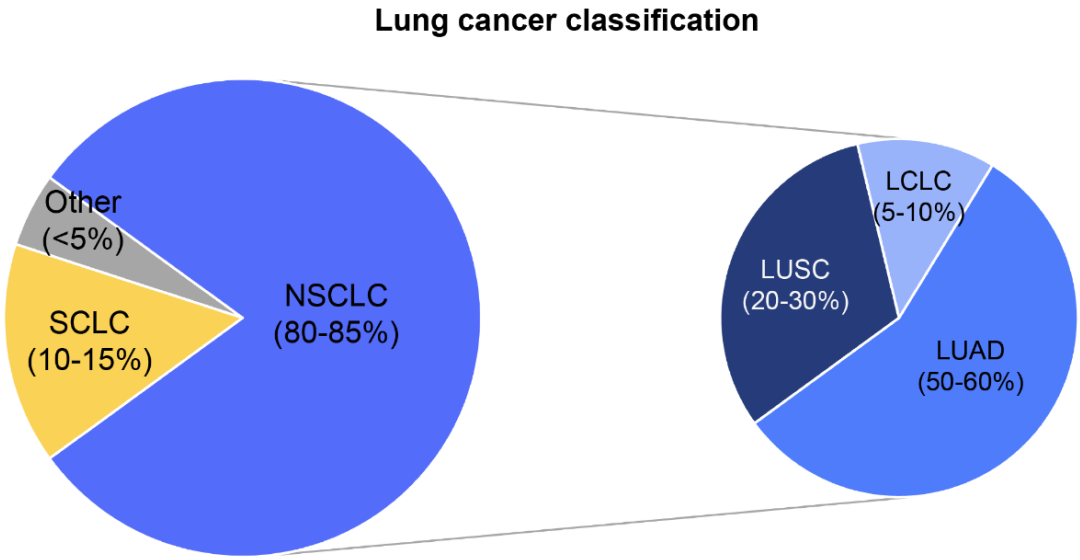


Figure 3. Lung cancer classification. NSCLC – non-small cell lung cancer; SCLC – small cell lung cancer; LUAD – lung adenocarcinoma; LUSC – lung squamous cell carcinoma; LCLC – large cell lung carcinoma. Data obtained from (Barta et al., 2019; M. Wang et al., 2021).

1.1.3 Current therapies against lung cancer

The first-line treatment for lung cancer patients depends on the specific lung cancer subtype, but it is mostly surgery, when possible, and a combination of chemotherapy and radiotherapy. SCLC has a much faster growth rate and more risk to metastasize to distal organs, therefore chemotherapy is urgent in SCLC

patients, who tend to be very responsive to it (H. Zhao et al., 2018). For NSCLC, although it is less aggressive, it is also diagnosed at later stages, therefore tumor resection is the immediate option for most NSCLC patients. Some studies have shown the benefit of administering adjuvant chemotherapy after the surgery, improving the 5-year survival rate (Pignon et al., 2008). For NSCLC patients in which surgery is not possible, the recommended treatment is a combination of radiotherapy and double chemotherapy of a platin-based antineoplastic and a second drug (Hirsch et al., 2017).

Advances in molecular medicine have led to the creation of targeted therapies, offering a more precise and personalized approach to cancer treatment. These therapies provide an alternative to the invasive traditional treatments such as surgery, chemotherapy, and radiotherapy, which often come with significant side effects for patients. For example, bevacizumab is a monoclonal antibody targeting the vascular endothelial growth factor (VEGF), which is often overexpressed in tumor cells and promotes angiogenesis, one of the cancer hallmarks (Sandler et al., 2006).

In lung cancer, the most used targeted therapies approved for the clinic are epithelial growth factor receptor tyrosine kinase inhibitors (EGFR TKIs), such as erlotinib or gefitinib (Hirsch et al., 2017; Mayekar & Bivona, 2017). The use of EGFR TKIs as a general adjuvant therapy has not been found to be effective for early-stage, unselected lung cancer patients (Hirsch et al., 2017), but they are a highly promising treatment option for EGFR-mutated and advanced-stage patients. Indeed, several clinical trials have demonstrated the better performance of EGFR TKIs primary treatment for EGFR-mutant patients in terms of objective response rate, progression-free survival, and overall quality of life compared to chemotherapy (Hirsch et al., 2017). Several more other targeted therapies are approved for its clinical use, such as crizotinib and alectinib for *ALK* rearrangements-driven tumors; or dabrafenib and trametinib for *BRAF*-mutant NSCLC (Mayekar & Bivona, 2017). More research and clinical trials are underway developing more of these precision medicine strategies, as the main problem of targeted therapies is that almost all tumors eventually develop a resistance against them, which makes patients relapse in a 1-2 years' time (Hirsch et al., 2017; Sequist et al., 2011).

Finally, in recent years a lot of focus has also been put on immunotherapy, which consists in the development of drugs that stimulate the patient's immune system to attack cancer cells. For instance, some drugs have reached the clinical use as second-line treatments for lung cancer, such as nivolumab, which blocks the receptor of the immune suppressor protein PD-1 (programmed cell death protein 1), thus activating the patient's own T-cells for enhanced cancer cell recognition and elimination (Hirsch et al., 2017).

1.2 The molecular origins of cancer

Cancer is a disease greatly dependent on environmental factors, with 44.4% of total cancer deaths being attributable to avoidable risk conducts, such as smoking, alcohol use, an unhealthy diet, or a sedentary lifestyle (Tran et al., 2022). Despite the great influence of external factors, the molecular origins of cancer are intrinsically genetic. Cancer appears fundamentally due to the accumulation of genomic abnormalities (i.e., mutations, translocations, amplifications, chromosomal rearrangements…) that confer the cells a proliferative advantage over their neighbors, which ultimately leads to the process of tumorigenesis.

1.2.1 From normal to tumor: the tumorigenesis process

The tumorigenesis process is very complex, as it usually involves the co-occurrence of several of these advantageous aberrations in a single cancer clone (Campbell et al., 2020), enabling it for what are called "the cancer hallmarks". This includes removing constraints on cell growth, promoting angiogenesis, scaping from the immune system vigilance, invading through tissue barriers, or spreading to other organs, to name a few (Hanahan, 2022). All these cancer hallmark acquisitions are not controlled by one specific cellular program. Instead, there is a vast collection of potential cellular abnormalities that cancer cells can choose from to create their own unique combinations (Campbell et al., 2020). Thus, although tumors may share similar macroscopic and clinical features, the cellular abnormalities that cause them are highly diverse, and this is the cause for the great heterogeneity that cancer presents between different patients, and why it is virtually unfeasible to develop a universal cure for it (Campbell et al., 2020; Marusyk & Polyak, 2009).

Tumorigenesis follows a dynamic that resembles classical Darwinian evolution (Campbell et al., 2020; Marusyk & Polyak, 2009). It all starts with a single cell clone gaining a certain proliferative advantage that allows it to expand and overgrow its neighboring cells. The more clonal expansions a cancer cell has, the more genomic instability it generates, which leads to further secondary or tertiary somatic aberrations that accumulate in the cells' DNA, ultimately resulting in different cancer subclone populations within the same tumor mass (Marusyk & Polyak, 2009). As happens in classical species evolution, not all genomic abnormalities, mainly mutations, are advantageous for the cells. Quite the contrary, most mutations (97%) are under neutral selection, or a few of them are directly wiped out of the population because they are disadvantageous (Mcfarland et al., 2017). The neutral or slightly deleterious mutations that do not confer a proliferative advantage are called "passenger" mutations, whereas the less

frequent, advantageous ones, which are selected in the cancer cell population and drive the tumorigenesis process are called "driver" mutations (Vogelstein et al., 2013).

Driver mutations often appear in hotspots across the genome of tumor cells, with a mean of 2-8 driver mutations per tumor (Vogelstein et al., 2013). Genes which harbor driver mutations and are therefore crucial for tumorigenesis are called "driver genes". Importantly, not all driver genes harbor driver mutations, as some may be altered at an expression level by other mechanisms, such as epigenetic regulation (Vogelstein et al., 2013). The term "epi-driver" has been proposed for such cases, for example, the tumor suppressor gene *SMARCA2*, whose sequence is rarely mutated in NSCLC cell lines and patients, but its expression is often lost due to promoter hypermethylation (Peinado et al., 2020; J. Wu et al., 2019).

1.2.2 Driver oncogenes and tumor suppressor genes in NSCLC

Driver genes can have different roles during cancer development. On the one hand, *oncogenes* promote tumorigenesis and are often found overexpressed or harboring activating mutations in cancer cells. On the other hand, *tumor suppressor genes* usually control normal cell growth and proliferation. Therefore, inactivating mutations in their sequence, or a downregulation of their expression might lead to tumorigenesis.

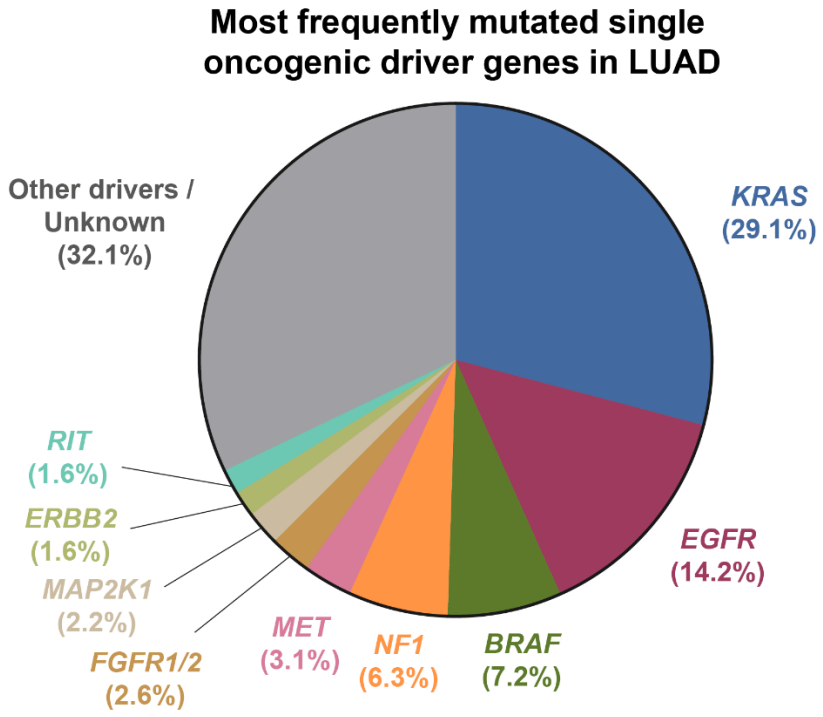


Figure 4. Most frequently mutated single oncogenic driver genes in lung adenocarcinoma (LUAD). Graph adapted from (Skoulidis & Heymach, 2019).

In LUAD, the major form of lung cancer, the most frequently mutated genes are *TP53* (46% of patients), *KRAS* (33%), *STK11* (17%), *KEAP1* (17%), and *EGFR* (14%) (Collisson et al., 2014). *TP53* (Tumor Protein 53), *STK11* (Serine/Threonine Kinase 11), and *KEAP1* (Kelch-like ECH-Associated Protein 1) are critical tumor suppressor genes that appear altered not only in LUAD, but also in the vast majority of malignancies, often co-occurring with other driver mutations. *KRAS* and *EGFR* are proto-oncogenes whose mutations are mutually exclusive in LUAD, and that are considered single oncogenic driver genes (Figure 4), meaning that the mutation of one of them is sufficient to trigger tumorigenesis (Skoulidis & Heymach, 2019). While driver mutations in tumor suppressor genes like *TP53* are very common and useful as prognostic biomarkers, they are also more challenging to target clinically (Collisson et al., 2014). Conversely, single oncogenic drivers such as *KRAS* (more frequently mutated in smoker patients and Caucasian populations) and *EGFR* (more common in non-smokers and Asian populations) are very valuable prognostic biomarkers and promising therapeutical targets for the clinical care of LUAD patients (Cheng et al., 2016; Marks et al., 2008):

- **The *KRAS* oncogene**

KRAS (Kirsten Rat Sarcoma Viral Proto-oncogene) is a member, along NRAS and HRAS, of the Ras family of proteins. The Ras-family genes were firstly identified as human homologs to some sarcoma-related genes in rats (Pulciani et al., 1982). Nowadays, we know that Ras-proteins are a family of ~21 kDa membrane proteins which present guanosine triphosphatase (GTPase) activity. Ras-family proteins are involved in an important plethora of cell pathways that control cell proliferation, differentiation, and apoptosis, such as the MAPK (Mitogen-Activated Protein Kinases), and the PI3K/AKT/mTOR (Phosphoinositide 3-Kinase / Protein Kinase B / Mammalian Target of Rapamycin) signaling cascades (Riely et al., 2009). Specifically, KRAS acts a molecular on/off switch: when bound to GTP, the protein is at its active form, allowing the signal to go downstream, and thus promoting cell proliferation (Figure 5). When the phosphate from GTP is hydrolyzed, either by GTPase-activating proteins (GAPs) or intrinsic GTPase activity from KRAS, this results in the GDP-bound, inactive form of KRAS, blocking the downstream signaling. The action of guanosine exchange factors (GEFs), such as SOS1/2, which are activated by the upstream signal of receptors tyrosine kinase, promote the exchange of GDP for GTP, thus returning KRAS to its active state (Karachaliou et al., 2013).

KRAS is one of the most frequently mutated genes not only in NSCLC patients, but also in many other types of cancer, such as pancreatic adenocarcinoma (mutated in 88% of patients), colorectal adenocarcinoma

(50%), and multiple myeloma (18%) (Prior et al., 2020). The most frequently mutated position falls in codon 12 (G12), where a glycine residue is usually replaced by a cysteine (G12C mutation – 42% of all *KRAS* mutations in NSCLC), a valine (G12V mutation – 21%), or an aspartic acid residue (G12D mutation – 17%) (Karachaliou et al., 2013). These aminoacidic changes impede GTPase activity by preventing the interaction with GAPs, thus blocking KRAS mainly at the GTP-bound, active state of protein. This causes an over-activation of the proliferation pathways where KRAS is involved, leading to excessive cell proliferation, apoptosis inhibition and, ultimately, the tumorigenesis process (Trahey & Mccormick, 1987).

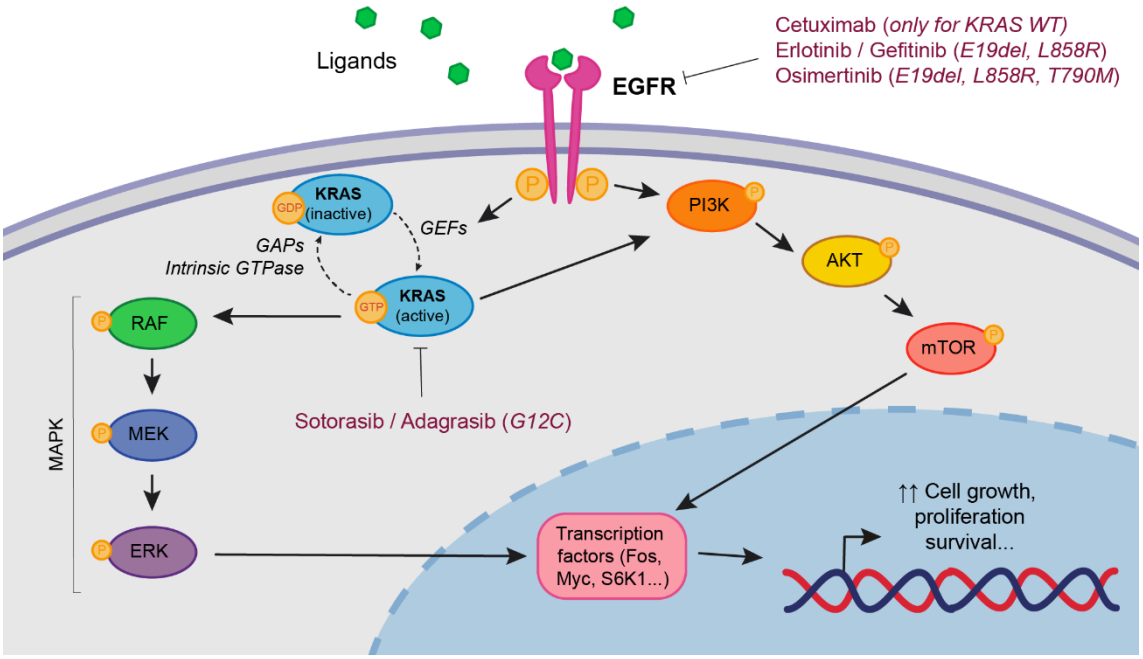


Figure 5. Simplified schematics of the EGFR and KRAS signaling cascade interacting with the MAPK (RAF-MEK-ERK) and PI3K/AKT/mTOR pathways. Targeted FDA-approved therapies against EGFR/KRAS, with the specific targeted mutations in parentheses, are also shown.

As for its clinical value, *KRAS* serves as a prognostic biomarker for EGFR-blocking immunotherapies such as cetuximab. Cetuximab is used as a first-line treatment only for *KRAS*-wildtype colorectal and head-and-neck cancer patients, either as a monotherapy or in combination with irinotecan (Heinemann et al., 2009; Martinelli et al., 2009). In NSCLC, however, its use is not yet approved by the U.S. Food and Drug Administration (FDA), and the correlation between the mutational status of *KRAS* and the response to EGFR-blocking therapies is not clear (Román et al., 2018). Regarding its direct therapeutical value, *KRAS*-mutant tumors frequently experiment the phenomenon that is known as "oncogenic addition", which makes *KRAS*-mutant

tumor cells extremely dependent on mutant *KRAS* expression, and consequently, they are more sensitive to its inhibition, compared to their non-tumor, *KRAS*-wildtype counterparts (Singh & Settleman, 2009). Therefore, mutant *KRAS* has been for a long time a coveted target for therapy. For years, KRAS has been considered an "undruggable" target protein because of its small size, relatively smooth and shallow surface, and the lack of drug binding pockets except for GDP/GTP binding (Nagasaka et al., 2021). Nevertheless, in the last 5 years two compounds have been tested in clinical trials and approved by the FDA for its use in *KRAS*-mutant NSCLC: sotorasib (AMGEN, approved in May 2021), and adagrasib (Mirati Therapeutics, approved in December 2022) (Yun et al., 2023). The benefits and limitations of these two drugs targeting G12C mutant KRAS will be thoroughly discussed later in Chapter 5.

- **The *EGFR* oncogene**

EGFR (also known as *HER1*, or *ERBB1*) is one of the four members of the human epidermal growth factor (EGF) transmembrane receptor family. EGFR operates as an inactive monomer that dimerizes with itself or with another member of its family in response to EGF binding (Mass, 2004). The dimerization triggers the activation of a complex downstream signaling network involving the MAPK and/or PI3K cascade, which ultimately promotes cell proliferation, differentiation and invasivity.

The most frequent *EGFR* mutations are a deletion in exon 19 (E19del), and the substitution of a leucine for an arginine in exon 21 (L858R). Both mutations are TKI-sensitizing, thus erlotinib and gefitinib are a first line treatment for these patients (A. R. Li et al., 2008). However, after treatment with EGFR TKIs, almost all patients eventually develop a resistance and relapse, mostly because of secondary EGFR mutations appearing in the surviving cell population. The most common resistance mechanism is the T790M mutation, which confers resistance to first- and second-generation EGFR TKIs, such as erlotinib or afatinib (Hirsch et al., 2017). Therefore, the FDA and the European Medicines Agency (EMA) have approved the use of a third-generation, irreversible EGFR inhibitor, osimertinib, which targets both the T790M and the previous TKI-sensitizing mutations (Hirsch et al., 2008).

In conclusion, it is evident that determining the patient's *KRAS* and *EGFR* mutational status before initiating a treatment is of extreme clinical relevance in NSCLC. Nevertheless, over a third of patients still lack a well-stablished oncogenic driver gene (Figure 4), limiting their therapeutic options to the general, untargeted treatments. Thus, the identification of novel driver genes is of the utmost importance for the development of novel diagnostic and prognosis biomarkers, and therapeutic strategies against cancer, as we will further discuss in the following sections.

1.3 Non-coding RNAs: the forgotten ones

The central dogma of biology defined by Francis Crick in 1958 (Crick, 1958) states that the genetic information of organisms lies encoded into the DNA, which is then transcribed to messenger RNA (mRNA), and finally mRNAs are translated into proteins, which exert different functions (structural, enzymatic, metabolic…) in the cells. This laid the foundation for molecular biology research during the next 60 years, expanding the knowledge about the concepts of gene expression, transcription, and translation as we know them today. Thus, we now estimate that only ~1.1% of the human DNA actually translates to protein, what we know as the protein-coding genome (Andrades, 2022; Frankish et al., 2019). Most research up to date looking for driver alterations in cancer has focused on this protein-coding part of the human genome (Martínez-Jiménez et al., 2020). However, the complete catalogue of single driver genes responsible for tumorigenesis is still not known for a large fraction of patients (Figure 4), which leaves these patients without specific diagnostic/prognostic biomarkers, or a target for a precision therapy Therefore, an interesting approach that has been explored in the past decades is interrogating the other ~98.9% of the genome, the non-coding part, in the search for novel driver alterations.

The non-coding genome, unjustly referred to at first as "junk DNA", is constituted by a vast number of elements, including intronic regions, untranslated regions (UTRs) proximal to protein-coding genes, promoter and enhancer regions, pseudogenes, repetitive sequences (retrotransposons, α-satellite regions…), and non-coding RNAs (ncRNAs), which are DNA sequences that are transcribed to RNA, but are not translated into proteins at the ribosomes (Palazzo & Gregory, 2014). Until relatively recently, ncRNAs have been overlooked in massive genomic studies looking for driver genes, as the implications of their alterations in cell biology were and are largely unknown. Nowadays, we know that ncRNAs can be important regulatory elements of gene expression, and therefore their alterations can potentially affect the development of several diseases, including cancer (Slack & Chinnaiyan, 2019). For their classification, ncRNAs can be roughly divided into two categories according to their size: long ncRNAs (>200 bp), and short ncRNAs (<200 bp), which include micro-RNAs and other regulatory and structural ncRNAs.

1.3.1 Micro-RNAs

Among the rest of short ncRNAs, micro-RNAs (miRNAs) stand out as one of the most studied because of their widespread role in gene expression regulation, and their implications in disease. They are defined as short (~22 bp), highly conserved ncRNAs that regulate gene expression by binding to the mRNA of their targets, usually impeding its translation (Bartel, 2004).

a. Biogenesis and functions of miRNAs

The canonical biogenesis process of miRNAs in cells involves a sequential series of processing steps from the primary transcript in the nucleus, until the mature miRNA binds to its target in the cytoplasm (Figure 6). First, miRNAs are transcribed in the cell nucleus by the RNA polymerase II or III, forming a primary miRNA transcript, or pri-miRNA. The exact length and sequence of the pri-miRNA is not well-defined, and it depends on the specific miRNA locus, but they all have in common a stemloop structure with a hairpin end, and a central mismatch in the double-strand RNA region. The pri-miRNA is cleaved by a protein complex called Microprocessor, which is formed by a ribonuclease III enzyme (Drosha), and two accessory DiGeorge syndrome critical region 8 RNA-binding proteins (DGCR8). The microprocessor cleaves the pri-miRNA at the basal junction, generating a shorter, ~70 bp miRNA precursor, known as the pre-miRNA. The pre-miRNA is then exported out of the nucleus via the exportin-5/RanGTP complex (Bartel, 2004; O'Brien et al., 2018).

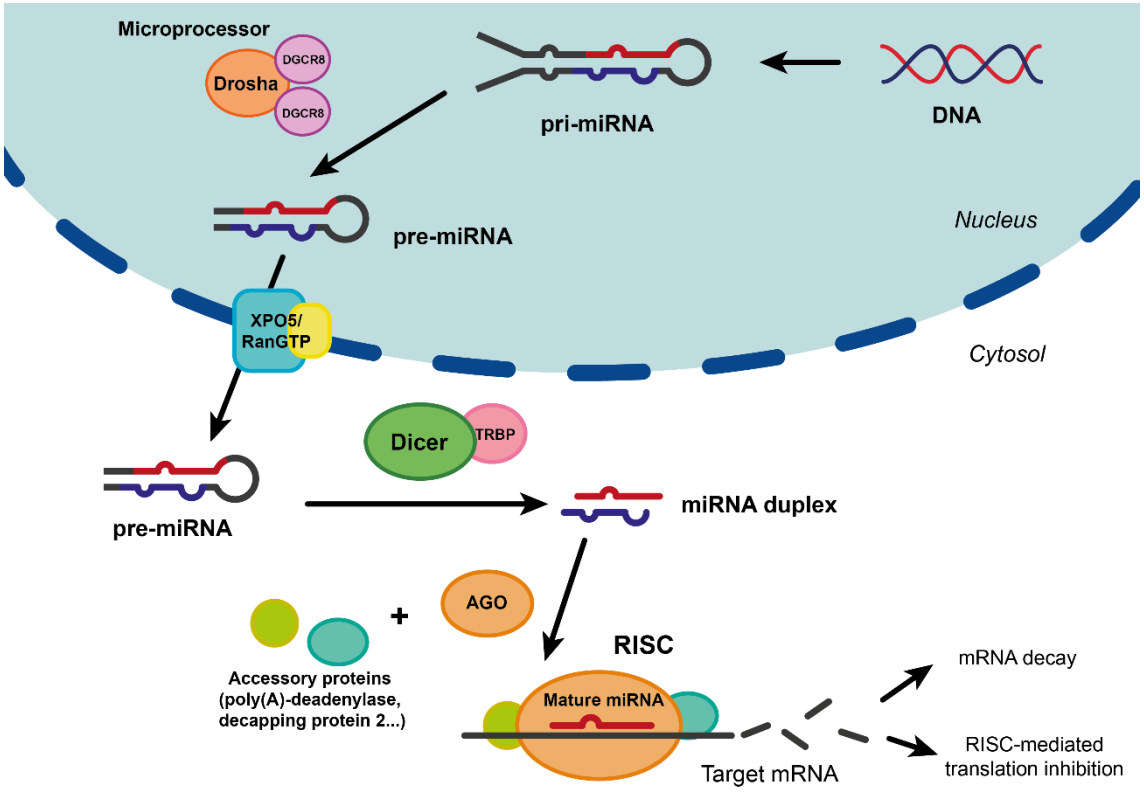


Figure 6. Schematics of canonical miRNA biogenesis.

In the cytoplasm, the pre-miRNA undergoes further processing by the ribonuclease Dicer with the help of the *trans*-activation-responsive RNA-binding protein (TRBP). This processing results in cleavage at the apical junction of the loop, generating a mature miRNA duplex comprised of two complementary strands, the 5p and the 3p strands. The miRNA duplex is then loaded onto one of

the Argonaute family of proteins (AGO1, AGO2, AGO3, and AGO4), thus forming the RNA-induced silencing complex (RISC) (Bartel, 2004; O'Brien et al., 2018). Then, one of the miRNA duplex strands (the "passenger" strand) is degraded, leaving the other one, or "guide" strand, loaded onto RISC. The mechanism that determines which strand is preferred over the other is complex and appears to be influenced by the cellular microenvironment. In certain cases, there is a clear preference for either the 5p or the 3p strand, while in others the strand selection appears to be random and follows a 50-50 ratio (O'Brien et al., 2018).

Finally, the mature miRNA bound to RISC is able to silence the expression of its targets by complementary binding to the mRNA. The binding sites of a miRNA are generally located in the 3'-UTR of mRNAs, but in some uncommon cases they might also appear in the 5'-UTR or the coding sequence (O'Brien et al., 2018). What ultimately determines the specificity of a miRNA is the seed region, defined as the nucleotides 2-8 in the 5' end of the mature miRNA. The complementarity of the seed region with the target mRNA can be perfect, or most times at least one bulge is formed by an unfavorable GU pair (O'Brien et al., 2018). In either case, this seed binding is usually sufficient for the miRNA to exert its function, although some additional pairing at the 3' end of the miRNA can help the stability and specificity of the binding (Broughton et al., 2016).

There are several ways of action of the RISC when bound to the target mRNA, and therefore mechanisms of miRNA regulation:

1. Inducing mRNA decay: if AGO2 is present in RISC, and if there is perfect complementarity, AGO2's endonuclease activity is induced and the mRNA is therefore cleaved and degraded, preventing it from translation (Jo et al., 2015). If the complementarity is not perfect, or some of the other Argonaute family members are present (AGO1, AGO3, AGO4), the RISC recruits several accessory proteins such as poly(A)-deadenylases or the decapping protein 2 (DCP2), which remove the 3'-poly(A) tail and the 5'-cap, leaving the mRNA exposed and ready for degradation by cytoplasmatic exonucleases (Orang et al., 2014).

2. RISC-mediated translation inhibition: AGO also acts as a competitor for some translation initiation factors, such as eIF4E, eIF4G, or PABP, directly inhibiting ribosomal translation (Orang et al., 2014).

3. Nuclear miRNA regulation: some miRNAs are found enriched, and there is evidence that RISC can be imported into the nucleus, where it can affect nuclear mRNA levels (O'Brien et al., 2018). RISC has also been documented to have non-canonical functions, affecting the transcription of some genes by interaction with the chromatin or the nascent transcript, although the full extent of this regulation and the precise way in which it operates are still not fully understood (Stavast & Erkeland, 2019).

4. RISC-mediated upregulation: although the general way of action of miRNAs is downregulating target genes, some cases have been reported of miRNAs enhancing the translation of target mRNAs. This is done via AGO2 recruiting FXR1 (Fragile-X Mental Retardation Protein 1), which binds to AU-rich elements in the 3'-UTR and interacts with the ribosome, promoting the translation. However, it appears that this mechanism is very context-dependent, and it only occurs in certain conditions, such as a nutrient starvation or in quiescent cells (O'Brien et al., 2018; Orang et al., 2014).

b. MiRNA roles in cancer

Because of their role as key regulators of gene expression, miRNAs are candidates for the maintenance of cell proliferation balance, and therefore they are found frequently altered in cancer (Medina & Slack, 2008). The first studies described a differential expression of some miRNAs between tumor and non-tumor tissues (J. Lu et al., 2005). In addition, some components of the miRNA biogenesis machinery have also been documented to have a role in tumorigenesis. Such is the case for some mutations found in Drosha, DGCR8, or Dicer (Kumar et al., 2007). This is in agreement with the downregulation of Dicer described in some tumors, which poses it as a tumor suppressor gene (Karube et al., 2005). MiRNAs themselves can have different roles in tumorigenesis. Several examples have been documented of miRNAs acting with either an oncogenic (oncomiRs), or a tumor suppressor role (tumor suppressor miRNAs) (Medina & Slack, 2008).

OncomiRs are frequently found overexpressed in cancer cells, and they can exert their role by silencing tumor suppressor genes or indirectly inducing the expression of other oncogenes. Some of the most studied examples of oncomiRs are:

1. *miR-17-92a* cluster: also known as the oncomiR-1 cluster, it was one of the first oncomiRs described. This polycistronic cluster of 7 miRNAs (*miR-17-5p*, *miR-17-3p*, *miR-18*, *miR-19a*, *miR-19b-1*, *miR-20*, and *miR-92-1*) is located in chromosome 13q31, a genomic region that is frequently amplified in B-cell lymphomas (He et al., 2005). This amplification results in the overexpression of the *miR-17-92a* cluster, which interacts with the oncogene *MYC* in a positive feedback loop, enhancing its tumor-promoting effects (He et al., 2005; O'donnell et al., 2005).

2. *miR-21*: another well-known example is *miR-21*, which stands out as the most overexpressed miRNA across several subtypes of cancer (Volinia et al., 2010). *MiR-21* promotes tumorigenesis by targeting important tumor suppressor proteins such as PTEN (Phosphatase and Tensin Homolog), and high levels of *miR-21* in tissue, plasma, and serum samples serve as a bad prognosis biomarker for NSCLC patients (Bica-Pop et al., 2018). Besides, *miR-21* was one of the first examples of "oncomiR addiction": similarly to

the oncogenic addiction that some tumors experiment with mutant *KRAS* (Singh & Settleman, 2009), some tumors are completely dependent on miR-21 expression for their development, and its inhibition resulted in complete tumor regression (Medina et al., 2010).

3. *miR-155*: it is another of the most important oncomiRs, which is found overexpressed in several hematological malignancies and solid tumors (Rupaimoole & Slack, 2017). Inducible models of lymphoma in mice have been stablished by just overexpressing *miR-155*, and tumor regression is observed upon its withdrawal (Babar et al., 2012). Among *miR-155*'s downregulated targets we find several tumor suppressor genes involved in angiogenesis, apoptosis, and cell cycle, such as the von Hippel-Lindau tumor suppressor (*VHL*), and the tumor protein p53 inducible nuclear protein 1 (*TP53INP1*) (Rupaimoole & Slack, 2017).

Conversely, examples of well-stablished tumor suppressor miRNAs are the *let-7* family, *miR-15a/16-1*, and the *miR-34* family:

1. *Let-7* family: the *let-7* family of miRNAs, which is comprised in humans by 12 genes scattered across different chromosomes, were first described as regulators of temporal differentiation in *C. elegans*, and it was the first example of conserved miRNAs in humans (Pasquinelli et al., 2000). *Let-7* is downregulated in several tumor types compared to non-tumor tissues, and it acts as a tumor suppressor by silencing several oncogenic targets such as RAS-family proteins (*NRAS*), and cell cycle-related proteins like cyclin A2 (*CCNA2*) (Johnson et al., 2007).

2. *MiR-15a/16-1*: they were the first tumor suppressor miRNAs described. Initially, it was discovered that the 13q14 region is recurrently lost in B-cell leukemia, even though no protein-coding genes were annotated in that region. Two miRNA genes, *miR-15a* and *miR-16-1*, were later discovered in such locus (Calin et al., 2002), which have a strong tumor suppressor role by targeting the B-cell leukemia oncogene *BCL2* (Cimmino et al., 2005).

3. *MiR-34* family: the *miR-34* family is comprised of *miR-34a, b* and *c*. Their expression is positively correlated and controlled by p53 (He et al., 2007), and the reintroduction of *miR-34a* in neuroblastoma showed a tumor suppressor effect by inducing apoptosis (Welch et al., 2007).

Finally, some therapeutical approaches taking advantage of the miRNAs' role in cancer have been attempted, either by restoring the expression of lost tumor suppressor miRNAs with miRNAs mimics; or by blocking the action of oncomiRs using antagonic molecules known as antagomiRs (Rupaimoole & Slack,

2017). For instance, MRX34 (Hong et al., 2020) and TargomiRs (van Zandwijk et al., 2017) are *miR-34a*/*miR-16a* mimics that reached phase-I clinical trials for advanced solid tumors and NSCLC, respectively. Another example is cobomarsen, an antagomiR for *miR-155* that reached phase-II clinical trials for mycosis fungoides, a special type of cutaneous T-cell lymphoma (Seto et al., 2018). Although these first clinical trials had to be terminated because of low response rates, immune-related adverse effects, or business reasons, they paved the way for future miRNA-based therapies against cancer, and hopefully the latest advances in RNA delivery will allow the development of more of these RNA-based drugs until their application in the clinic (Arenas et al., 2022).

1.3.2 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are defined as RNA transcripts longer than 200 bp that do not encode for proteins. LncRNAs can possess all the elements that a protein-coding gene has except for a stable open reading frame (ORF), including promoter and enhancer regions, UTRs, introns, alternative splicing elements, and polyadenylation signals (Bhat et al., 2016). As happens with miRNAs, lncRNAs are critical gene expression regulation elements and have significant implications in various pathologies, including cancer.

a. Functions of lncRNAs

Because of the broadness in their definition, lncRNAs can exert a wide variety of functions in terms of gene expression regulation, both at nuclear and cytoplasmatic levels. The versatility of lncRNAs as regulatory molecules comes from (i) their sequence, (ii) their wide length span, from 200 bp up to ~100 kilobases; and (iii) their ability to bind DNA, proteins, and other RNAs.

LncRNA sequences are overall less conserved than protein-coding genes. However, their promoters and enhancers have a higher conservation rate, hinting to the importance of lncRNA transcription (Ginn et al., 2020). As a matter of fact, some lncRNAs exert their function as *cis*-regulatory elements by enhancing or suppressing the expression of neighboring genes through their own transcription, even though the lncRNA transcript molecule itself may not be required for this regulation. This has been demonstrated by studies in which the full transcription of some lncRNAs was blocked by insertion of premature polyadenylation signals, while their promoter regions were left untouched. The authors then observed how the regulation of neighboring genes persisted, despite the absence of the lncRNA transcript (Engreitz et al., 2016; Paralkar et al., 2016).

The variable length of lncRNA also allows them to adopt numerous secondary, and even tertiary structures, which have been related to their function and their ability to bind DNA, RNA, and proteins. LncRNAs usually exhibit low conservation across different species, displaying only “patches” of conserved bases

scattered across large, seemingly unconstrained sequences. As a result, it appears that evolution has preserved certain domains within lncRNAs that affect their structure, with the sequence itself assuming a secondary role (Zampetaki et al., 2018). For example, such is the case for the lncRNA *MEG3* (Maternally-Expressed Gene 3), a tumor suppressor lncRNA which acts by stimulating p53-mediated transcription. Three functional domains (M1, M2, and M3) are found in *MEG3* sequence, of which M2 and M3 are necessary for p53 activation. The substitution of half of the M2 domain by an artificial sequence that rendered a similar secondary structure kept its full functionality in activating p53 (X. Zhang et al., 2010), thus proving that some lncRNAs' function is dependent on its structure, not their sequence.

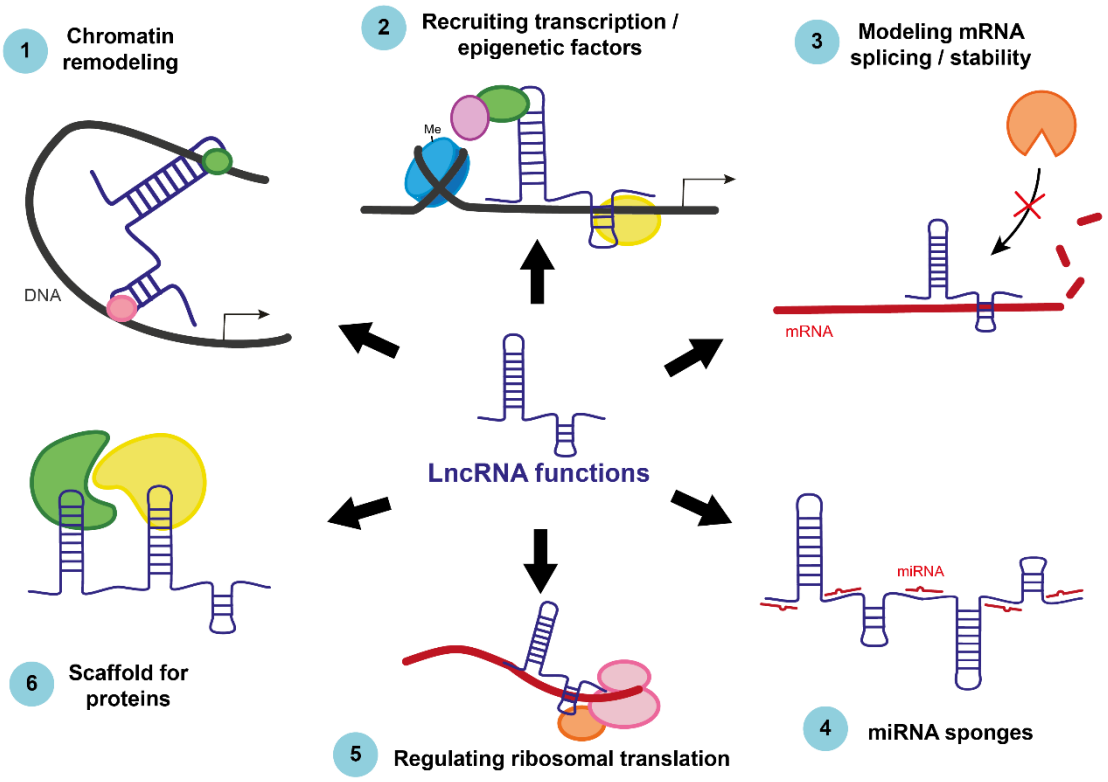


Figure 7. Summary of different lncRNA functions.

The exact function and way of action for the majority of annotated human lncRNAs remains largely unknown, as this is a relatively new field of study and the mechanistical study of lncRNA is complex, as will be further discussed in Chapter 3. Some of the functions attributed to lncRNAs are (Figure 7):

1. Chromatin remodeling: in the nucleus, lncRNAs can recruit chromatin remodeling complexes, changing the architecture of the chromatin and its condensation state (Ginn et al., 2020; Peinado et al., 2018). This is the case for one archetypical lncRNA, *XIST* (X-Inactive Specific Transcript), which binds to the X-chromosome and recruits the polycomb repressive

complex 2 (PRC2). PRC2 is a chromatin remodeling complex that condensates the chromatin, thus inactivating the X-chromosome, which allows the X-chromosome dosage compensation in women (J. T. Lee & Bartolomei, 2013).

2. Epigenetic and transcriptional regulation: likewise, lncRNAs can serve as scaffolds between the promoter region of target genes and demethylases or certain transcription factors, enhancing their expression. Conversely, they can also act as decoys for such proteins, sequestering transcription factors and repressing the target's expression (Peinado et al., 2018). In addition, as aforementioned, some lncRNAs act in *cis*, regulating the expression of neighboring genes by means of their own transcription (Kornienko et al., 2013).

3. Regulation of splicing and stability of mRNAs: some lncRNAs can interact with the primary transcripts of mRNA and modulate alternative splicing (Bhat et al., 2016; Ginn et al., 2020). This is the case for one of the first studied lncRNAs, *MALAT-1* (Metastasis-Associated Lung Adenocarcinoma Transcript 1), which has been shown to interact with splicing factors, altering their distribution, and thus altering the splicing of mRNAs precursors (Tripathi et al., 2010). Other lncRNAs also interact with mature mRNAs, regulating their stability by activating or blocking their degradation by SMD (Staufen1-Mediated mRNA Decay) (Peinado et al., 2018).

4. Competing endogenous RNAs (ceRNAs): lncRNAs may act as ceRNAs by binding to other regulatory non-coding RNAs, mainly miRNAs, sequestering them, and impeding their function. For example, *NEAT1* (Nuclear Paraspeckle Assembly Transcript 1) is an oncogenic lncRNA in NSCLC that promotes cell growth by acting as a "miRNA-sponge" for *miR-377-3p*. The miRNA *miR-377-3p* is normally suppressing its target, the oncogenic transcription factor E2F3, whose expression is restored upon overexpression of *NEAT1* in NSCLC, which brings down the levels of free *miR-377-3p* and initiates the tumorigenesis process (Ginn et al., 2020).

5. Translation regulation: lncRNAs can also interact with the ribosome, enhancing or suppressing translation. For example, *lincRNA-p21* interacts with its target mRNA by base complementarity and induces ribosome "drop-off", which slows down translation (Yoon et al., 2012). Another type of lncRNAs, such as *Uchl1-AS*, contain SINEB2 sequences that overlap those of their corresponding sense genes, allowing their association under certain stress conditions, which enhances interaction with polysomes, inducing the translation of the sense gene. This has been proposed to be a mechanism for induction of gene expression

using "SINEUPs", a new class of artificial lncRNAs containing these SINEB2 sequences (Zucchelli et al., 2015).

6. Scaffold for proteins: some cytoplasmatic lncRNAs can also regulate the stability and interaction of proteins, by promoting their ubiquitination and subsequent degradation, or sequestering some of them and impeding the formation of protein complexes (Schmitt & Chang, 2016).

b. LncRNAs in lung cancer

Similarly to miRNAs, dysregulation of lncRNA expression has been found to be involved in the development of different human cancers. There are several examples of lncRNAs involved in tumorigenesis, enabling cancer cells for cancer hallmarks such as inducing angiogenesis, evading growth suppressors, sustaining proliferative signaling, inducing migration and metastasis, or suppressing apoptosis (Gutschner & Diederichs, 2012). In the specific case of NSCLC, hundreds of lncRNAs have been associated with tumor development through gene expression microarrays and massive parallel RNA sequencing of lung tumor tissues and paired adjacent non-tumor tissues (Loewen et al., 2014). In the last decade, several studies have aimed to validate such dysregulated lncRNAs as oncogenic/tumor suppressors with functional assays (Ginn et al., 2020).

Two well-known examples of oncogenic lncRNAs in NSCLC are *MALAT-1*, and *HOTAIR* (HOX Transcript Antisense RNA). *MALAT-1* expression is upregulated in metastatic NSCLC patients compared to those that do not present metastasis (Ji et al., 2003), and it has been reported to be involved in tumorigenesis through different mechanisms, such as promoting epithelial-mesenchymal transition (EMT), enabling migration and metastasis, and activating the AKT/mTOR axis (Ginn et al., 2020). On the other hand, *HOTAIR* is overexpressed in NSCLC compared to normal lung tissues, and it can function as a chromatin remodeling regulator by recruiting PCR2 and silencing gene expression, and also as a sponge for several miRNAs (Loewen et al., 2014). This all results in an increase of cell migration, growth, proliferation, drug resistance, invasion, and metastasis (Gupta et al., 2010; Loewen et al., 2014).

Growing evidence indicates that lncRNAs can also function as tumor suppressor genes, such as *TUSC7* (Tumor Suppressor Candidate 7), and *GAS5* (Growth Arrest Specific 5). *TUSC7* (also known as *LOC285194*) is downregulated in NSCLC tissues and cell lines, and its overexpression is shown as a good prognostic biomarker for this disease. In addition, restoration of *TUSC7* in NSCLC cell lines reduced proliferation, enhanced apoptosis, and diminished cell migration, apparently because of its interaction with p53 (Z. Wang et al., 2016; Zhou et al., 2019). As another example, *GAS5* is able to post-transcriptionally regulate p53, p21, and E2F1, thus inhibiting tumor growth and promoting apoptosis in NSCLC cells (X. Shi et al., 2015).

Regarding their clinical applications, different studies have found that lncRNAs appear as novel and valuable tools for cancer prognosis, diagnosis, and treatment. LncRNAs can be detected in body fluids, such as blood or saliva, due to their capability of being secreted into the extracellular matrix by tumor cells, either directly, as circulating lncRNAs, or by previous encapsulation into exosomes, apoptotic bodies, or lipoprotein complexes (T. Shi et al., 2016). A notable example is *PCA3* (Prostate Cancer Antigen 3), an FDA-approved urine diagnostic and prognostic biomarker for prostate cancer, which even outperforms the widely used prostate-specific antigen (PSA) in terms of sensitivity and specificity (Bolha et al., 2017). As for lung cancer, in addition to the aforementioned *MALAT-1* and *GAS5*, some lncRNAs have been identified as stable blood biomarkers for NSCLC diagnosis (*XIST*, and *HIF1A-AS1*); poor prognosis biomarkers (*CCAT2*, and *CARLo-5*); and potential targets for molecular therapy (*H19*) (T. Lu et al., 2018). Furthermore, a novel antisense oligonucleotides (ASOs) therapy against *MALAT-1* was found to be effective in reducing the extravasation and lung nodule formation capacity of lung cancer cells, both *in vivo* and *in vitro* (Gutschner et al., 2013).

1.3.3 Other non-coding RNAs

In addition to miRNAs and lncRNAs, there are some other categories of ncRNAs that also have been studied for their regulatory functions, and their potential role in cancer development.

a. Transfer and ribosomal RNAs

Transfer RNAs (tRNAs) are small (70-100 bp), conserved, non-coding RNA molecules that adopt a characteristic cloverleaf shape, and present several nucleotide modifications. They play a crucial role during protein translation, as they are in charge of loading the 20 different amino acids, each corresponding with its matching 3-nucleotide codon in the mRNA, and transferring them to the nascent polypeptide chain at the ribosome. The study of tRNAs deregulations in diseases, either by changes in their expression, relative abundance, or modification state, is still in its early stages. In the context of cancer, an overall overexpression of tRNAs has been described, but it remains unclear whether this is a cause or a consequence of the higher protein synthesis and metabolic rates observed in tumor cells (M. Santos et al., 2019).

On the other hand, ribosomal RNAs (rRNAs) are highly conserved, structural components of the ribosomes. In eukaryotes, the rRNAs found in cytosolic ribosomes are the 18S in the small 40S subunit, and the 5S, 5.8S, and 28S in the large 60S subunit. These rRNA genes are transcribed by the RNA polymerase I, and they undergo a complex and very tightly-regulated maturation process before they form part of ribosomes (Pecoraro et al., 2021). Because of their greater protein synthesis requirements, tumor cells frequently activate different cell pathways with the aim to increase the synthesis and processing of rRNAs, and can

therefore be a target for molecular therapies. Indeed, this is the case for the oncogene *MYC*, which upon overexpression, it boosts all the steps of rRNA biosynthesis and maturation through diverse molecular mechanisms, for example, activating the whole RNA polymerase I machinery (Gaviraghi et al., 2019).

b. Circular RNAs

Circular RNAs (circRNAs) are non-coding, single strand RNA molecules that appear as a covalently closed loop, instead of a linear structure. Some circRNAs originate from introns that are excised during mRNA processing in the nucleus, while others come from skipped exons resulting from alternative splicing events (Dong et al., 2021). CircRNAs can function as ceRNAs similarly to lncRNAs, by acting as miRNA sponges, sequestering them, and impeding their action. In addition, thanks to their capability to bind to RNA and proteins, some circRNAs have also been implicated as direct regulators of protein translation, transcription by RNA polymerase II, splicing, and RNA maturation (Dong et al., 2021).

Because of their closed structure, circRNAs are much more stable and resistant to degradation by RNA exonucleases than linear RNA molecules. In addition, they have been detected as either circulating circRNAs or inside exosomes in human blood plasma, urine, and other body fluids, which makes them promising molecules for their use as biomarkers (Vo et al., 2019). In fact, several circRNAs have been described as altered in various types of cancer, and depending on which miRNAs they sponge, they can have oncogenic or tumor suppressor roles (R. M. Santos et al., 2020). For example, *circFGFR1* is upregulated in NSCLC, where it acts as a sponge for a miRNA, *miR-381-3*, which leads to upregulation of a chemokine receptor (CXCR4), resulting in tumor cell progression. Besides, a high expression of *circFGR1* also works as a prognosis biomarker for resistance to anti-PD1-based immunotherapy (Santos et al., 2020).

c. Small interfering RNAs

Small interfering RNAs (siRNAs) are very similar to miRNAs in terms of both their biogenesis and function. The main differences are (i) that siRNAs are originated from a long, double-strand RNA molecule instead of a pre-miRNA with a hairpin structure, and (ii) that siRNAs always require perfect complementarity to their target mRNAs, thus inducing AGO2's endonuclease cleavage of the mRNA, whereas miRNAs may have unpaired nucleotides, which triggers other ways of action as discussed in the previous sections. This results in siRNAs being much more specific, and usually targeting just one mRNA, while one single miRNA can regulate multiple different targets at the same time (Lam et al., 2015).

Although siRNAs were first thought of having an ectopic origin, such as viral infections, some endogenous siRNAs have also been found in the genome of animal cells, mostly related to repetitive elements like transposons, where they appear to

have a role in genome defense by silencing such genomic mobile elements (Piatek & Werner, 2014; Sontheimer & Carthew, 2005). The role of human endogenous siRNAs in cancer, however, has been mostly overshadowed by a plethora of studies about the use of artificial siRNAs as tools for gene silencing in basic and translational studies (Lam et al., 2015).

d. PIWI-interacting RNAs

Another similar element to siRNAs and miRNAs are PIWI-interacting RNAs (piRNAs). They are named after their binding to PIWI-domain proteins, which are a subfamily of the Argonaute protein family that lacks endonuclease activity. Unlike miRNAs and siRNAs, piRNAs are transcribed from intergenic clustered regions, and they do not undergo Dicer processing. When loaded onto PIWI-proteins, piRNAs function mainly by silencing transposable elements of the genome, although they have also been reported to silence protein-coding genes, and they appear to be essential for germline development across different animal species (X. Wang et al., 2022).

Several piRNAs have been found altered in cancer and show potential as biomarkers for the disease. For example, *piR-55490* is downregulated in lung cancer tissues and cells, and it has a tumor suppressor role by binding to the 3'-UTR of *mTOR*, thus suppressing the AKT/mTOR pathway (S. Chen, Ben, et al., 2021).

e. Other small non-coding RNAs

In the nucleus, two more types of small ncRNAs can be found: small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs), both having functions related to RNA processing. Many alterations of snRNAs have been related to cancer. Among several other examples, the spliceosome-related snRNA *7SK* is downregulated in tongue, breast, blood, and colon cancer, and has been found to have tumor suppressor capabilities when restored in cell lines (Guglas et al., 2022).

As for snoRNAs, they are an abundant and very stable family of RNA transcripts which are found in the nucleolus forming ribonucleoprotein complexes. Although initially thought to have a housekeeping function in the cells, mostly related to rRNA modification and processing, several examples of oncogenic or tumor suppressor snoRNAs have been described, and there are distinct snoRNA signatures that serve as biomarkers to tell apart NSCLC and non-tumor samples (Thorenoor & Slaby, 2015).

1.4 CRISPR/Cas9: a new weapon against cancer

Since their biotechnological application developed by Jennifer Doudna and Emmanuelle Charpentier in 2012, the CRISPR/Cas9 gene edition system has revolutionized the field of molecular biology thanks to its incredible potential to specifically knockout or edit the desired target genes. Previous gene-editing systems, such as zinc-finger and TAL-effector nucleases, have been left outdated because of the greater performance of CRISPR/Cas9 in terms of specificity against a determined DNA locus, and its ease of design and synthesis (Doudna & Charpentier, 2014). As such, applications of CRISPR/Cas9 in biomedicine have flourished in the last decade, to the point that Doudna and Charpentier were awarded in 2020 with the Nobel Prize in Chemistry for their contribution to this groundbreaking discovery.

1.4.1 The CRISPR/Cas system: discovery, function, and variants

The earliest descriptions of CRISPR date back to the late 1980s, where the first studies analyzing the genome of bacteria and archaea described a mysterious, clustered, regularly interspaced, palindromic repeats (CRISPR) region, whose function was unknown (Ishino et al., 2018). More than one decade later, the independent work of Mojica and two other groups shed light on the origins of those CRISPR regions, as they were similar to sequences of bacteriophages and archaeal viruses (Mojica et al., 2005). The parallel work on CRISPR-associated (Cas) proteins allowed to define the CRISPR/Cas system as a sort of primitive "immune system" for bacteria and archaea, protecting them against the infection of bacteriophages or other viruses (Ishino et al., 2018).

a. CRISPR/Cas as a primeval immune system in bacteria and archaea

The way it operates, the CRISPR/Cas system enables bacteria and archaea to "remember" past infections, thus developing a sort of immunity against future infections by the same virus. In brief, the process can be divided in three phases: adaptation, RNA expression and processing, and interference (Figure 8A). In the adaptation phase, upon infection, a series of Cas-proteins (Cas1, Cas2…) can digest the viral genetic material into small fragments, which are then stored in the bacterial genome as new spacers in the CRISPR array. In the RNA expression and processing phase, the CRISPR locus is transcribed, and after processing by more accessory proteins (e.g., Cas6, RNAse-III), short CRISPR-RNAs (crRNAs) are generated. Finally, in the interference phase, those crRNAs can be then loaded onto a Cas-endonuclease, the most widely known being Cas9, which directed by complementarity of the crRNA, can then bind to specific regions in the viral genome, activate their endonuclease activity, produce double-strand breaks (DSBs), and therefore digest the viral genome, preventing new infections (Horvath & Barrangou, 2010; Ishino et al., 2018).

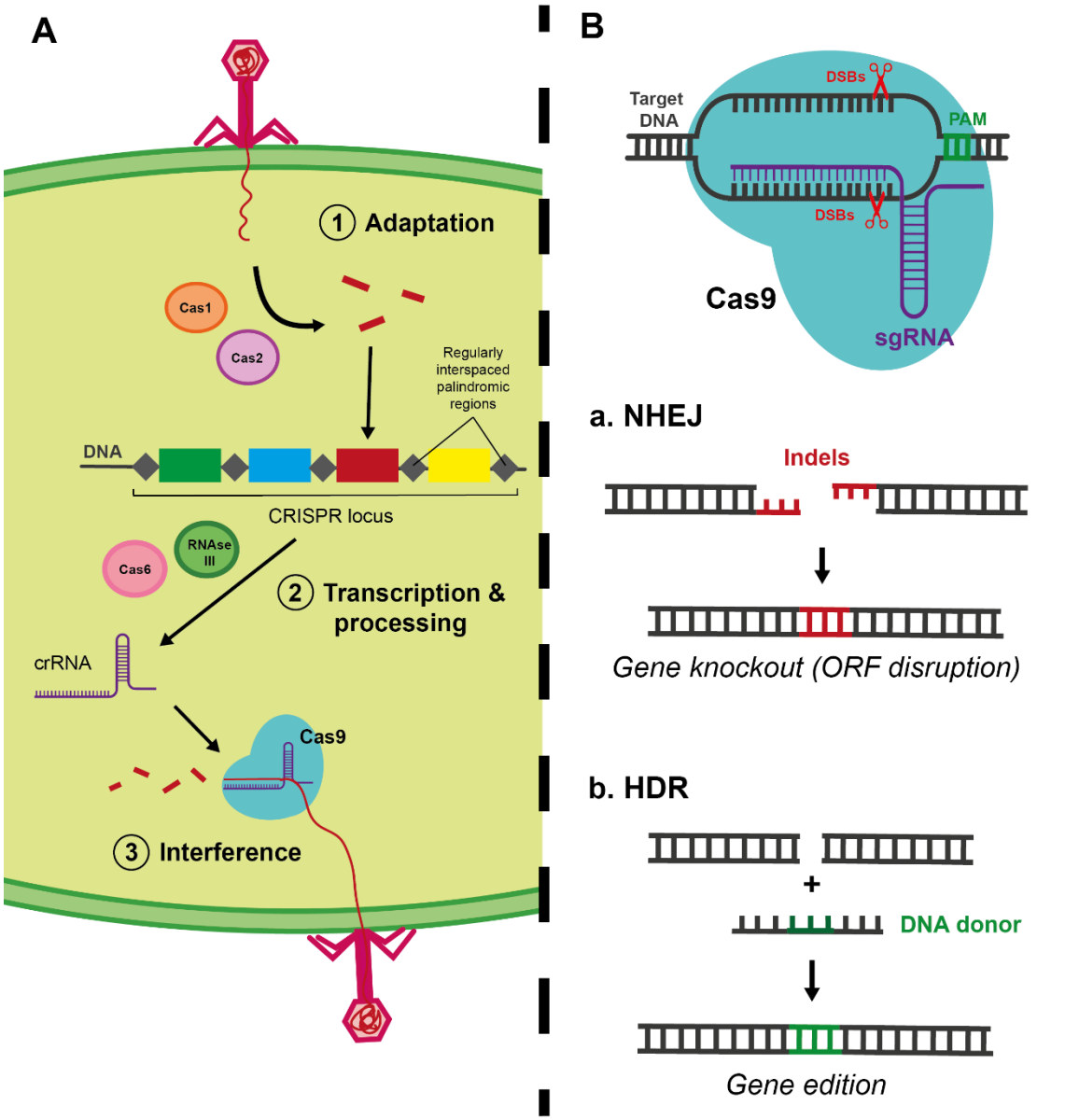


Figure 8. (A) Function of CRISPR/Cas9 as a primitive immune system for bacteria and archaea. (B) CRISPR/Cas9 application in molecular biology. Directed by a single-guide RNA (sgRNA) and a protospacer adjacent motif (PAM), Cas9 generates double-strand breaks (DSBs) in the target DNA sequence, thus inducing either non-homologous end joining (NHEJ), which results in specific gene knockouts by ORF-disrupting indels; or homology-directed repair (HDR) favored by the addition of a DNA donor molecule, which enables specific gene edition.

b. Classification of CRISPR/Cas systems

Because it is an antiviral defensive mechanism conserved across several species of microorganisms (up to ~90% of archaea and ~40% of bacteria) (Horvath & Barrangou, 2010), various types of CRISPR/Cas systems have been found in different organisms. Thus, CRISPR/Cas systems are classified into two major groups (Makarova et al., 2019):

i. Class 1 systems: they are the most common, found in both bacteria and archaea. Class 1 CRISPR/Cas systems possess an effector complex formed by multiple Cas-proteins that bind together with the crRNA, recognize the target, and cleave it.

ii. Class 2 systems: they are only found in some bacteria, but because of their simplicity, they are the ones that have been mostly explored for biotechnological applications. Instead of a multiprotein complex, they have a single, multidomain effector protein, the most common being Cas9, but there are others such as Cas12 or Cas13. In addition, they usually require a secondary RNA molecule for the processing of crRNAs, known as the *trans*-activating crRNA (tracrRNA). The tracrRNA is complementary to the repeat sequences of the crRNA, and it binds together Cas9 and the crRNA, which triggers the crRNA processing and promotes the endonuclease activity of Cas9 directed in their target sequence.

Doudna and Charpentier were the first to discover the possibility of fusing together the crRNA and the tracrRNA in a single-guide RNA (sgRNA) molecule that directs Cas9 to their specific targets (Jinek et al., 2012). Later, Feng Zhang's group was pioneer in developing a protocol for the design of artificial sgRNAs to induce DSBs in the desired DNA locus of Cas9-expressing human cells (Ran et al., 2013). The exact cleavage site, and therefore the specificity of Cas9, depends on the complementarity of the sgRNA with the target sequence, and on the presence of certain conserved protospacer adjacent motifs (PAMs), juxtaposed to the complementary region in the target DNA (Jinek et al., 2012). Every Cas protein variant has a specific PAM sequence, which is necessary for Cas binding to the target DNA. The PAM sequence also defines the exact location of the cleavage site, which is usually 3-4 nucleotides upstream of the PAM. For example, the PAM of the widely used *Streptococcus pyogenes*' Cas9 (SpCas9) is 5'-NGG-3' (N = any base), whereas the PAM for Cas12a (formerly known as Cpf1) is a longer, T-rich motif, 5'-TTTV-3' (V = any base but thymine) (Anders et al., 2014; Bandyopadhyay et al., 2020).

In the past years, several variants of Cas9 have appeared, usually by mutagenesis and directed evolution of bacterial SpCas9. By inducing certain mutations or deletions in the minimal SpCas9 sequence, we can change the detected PAM sequences, and the efficiency of Cas9's endonuclease activity (or even suppress it), thus increasing the versatility of CRISPR/Cas9 to target whichever gene we desire (Kleinstiver et al., 2015; Vakulskas et al., 2018). Furthermore, different applications of other class 2 CRISPR/Cas systems have been developed, taking advantage of the different molecules they target. For instance, the Cas13 family targets single or double-strand RNA molecules instead of DNA,

which has some interesting applications that will be further detailed in later sections.

1.4.2 CRISPR/Cas9 applications in biomedicine and cancer

The applications of CRISPR/Cas9 in molecular biology mostly come from its capacity to induce DSBs in the desired DNA locus in a very specific and efficient manner. Guided by a customized sgRNA, designed specifically against a certain gene, Cas9's endonuclease activity induces DSBs at such location. Then, cells are able to detect those DSBs and try to repair them thanks to their intrinsic DNA damage repair mechanisms. There are two competing DSBs repair mechanisms, and either one or the other are favored depending on the cell status (Figure 8B):

i. Non-homologous end joining (NHEJ): generally, throughout the whole cell cycle the main DNA damage repair mechanism that cells undergo is non-homologous end joining (NHEJ). NHEJ is induced when a DSBs is detected, and there is no DNA template that can be used to repair the damage. In this case, the cell tries to repair the gap by holding both ends together and directly ligating them. However, this repair mechanism is very error-prone, as it randomly introduces small insertions or deletions (indels) of nucleotides at the cleavage site (Chapman et al., 2012). If the breaking point falls into a coding sequence, those indels might potentially alter the ORF of the sequence, thus obtaining a non-functional protein, which could be useful as a knockout (KO) model of the target gene.

ii. Homology-directed repair (HDR): only during the S/G2 phases, right after the DNA synthesis, cells have another DNA molecule to use as a template to repair the DNA damage without the risk of introducing indels. This mechanism is termed homology-directed repair (HDR), and it is favored with the presence of another, partially complementary DNA molecule that acts as a donor, albeit with a lower efficiency than NHEJ (25% vs 75%) (H. Yang et al., 2020). By using an artificial donor DNA template which is complementary to the flanks of the break, we can generate knock-in models that express a gene of interest, or correct a certain mutation, thus editing the targeted gene (Song & Stieger, 2017). Some small molecules that inhibit NHEJ mediators can be used to favor HDR over NHEJ when inducing Cas9 edition, but the efficiency still remains low and there are toxicity issues, as NHEJ is critical for the maintenance of genome stability in cells (H. Yang et al., 2020).

a. CRISPR/Cas9 as a tool for gene therapy of genetic diseases

In biomedicine, CRISPR/Cas9 has been widely applied to generate cell and animal models for the study of human diseases. The specificity and efficiency of CRISPR allows us to obtain stable-over-time, gain-or-loss models of genes, which are excellent tools to study their implications and functions in diseases, and how to cure them. In addition, CRISPR has supposed a revolution in the context of gene therapy against previously incurable genetic diseases. Indeed, several clinical trials are being carried out for patients with hematological diseases such as sickle cell disease and β-thalassemia, which are caused by inactivating mutations in the hemoglobin β gene (Ahumada-Ayala et al., 2023). In one of the concluded ones, by performing an *ex vivo* CRISPR/Cas9 editing of two patients' hematopoietic stem cells, fetal hemoglobin was successfully re-expressed, and after an autologous graft of such edited cells, these two patients' symptoms remitted in just 30 days after the transplant, without needing further blood transfusions (Frangoul et al., 2021).

Another important milestone of CRISPR/Cas9 in therapeutics was a phase-I study for patients with transthyretin amyloidosis, a neurodegenerative disease caused by the accumulation of misfolded transthyretin (TTR) in nervous and cardiac tissues. Six patients were intravenously injected with lipid nanoparticles containing SpCas9 and a sgRNA against the *TTR* gene. One month after the treatment, the levels of TTR were reduced up to 87%, with only some mild adverse effects in half of the patients (Gillmore et al., 2021). Moreover, several other diseases have been approached with CRISPR/Cas9 strategies in preclinical and clinical studies, including Duchenne's muscular dystrophy, type 1 diabetes mellitus, and even infectious diseases such as human immunodeficiency virus (HIV) infection, and the coronavirus disease 2019 (COVID-19) (Ahumada-Ayala et al., 2023).

b. CRISPR/Cas9 applications in cancer

Regarding cancer research, thanks to CRISPR/Cas9 we have a plethora of cell and murine models for the basic research of oncogenes and tumor suppressor genes, and preclinical testing of drugs and molecular treatments. In terms of clinical applications in patients, most approaches up to date have focused on using CRISPR/Cas9 for immunotherapy and chimeric antigen receptor (CAR) T cell generation.

CAR-T cells are immune cells from the own patients engineered to express ligands for certain tumor antigens, such as CD40 or CD19. Thus, after an autologous transplant of CRISPR-modified CAR-T cells, tumor cells that present said target antigens in their surface are eliminated (Azangou-Khyavy et al., 2020; Z. Liu et al., 2023). Another application of CRISPR in CAR-T cell immunotherapy is desensitizing CAR-T cells to immune checkpoint inhibitors that tumor cells often overexpress for immune evasion, such as the programmed cell death protein ligand 1 (PD-L1). For example, in a phase-I clinical trial for NSCLC (NCT02793856),

12 patients were infused with autologous CAR-T cells knocked out for the PD-1 gene, thus preventing immune evasion without any safety issues (Bender et al., 2021). CRISPR has also been used to increase safety and performance of CAR-T cells, for example, by knocking-out the CD7 receptor that induces fratricidal activity among CAR-T cells, and also to generate "off-the-shelf", universal CAR-T cells with modified HLAs to avoid graft-versus-host-disease (Z. Liu et al., 2023; Maldonado-Pérez et al., 2022). Moreover, in addition to CAR-T cell immunotherapy, CRISPR has been employed to engineer more tumor-selective and efficient oncolytic viruses, which preferably infect and replicate within tumor cells and constitute a novel line of immunotherapy (Azangou-Khyavy et al., 2020).

Finally, a few approaches have also been attempted for the application of CRISPR/Cas9 as a direct therapeutic strategy, though only in mice models. For instance, one study was able to successfully target and correct *TP53*-mutant tumors in mice by using adeno-associated viruses (AAVs) injected intravenously and loaded with an inducible cassette of Cas9, two sgRNAs flanking the *TP53* mutation, and a cDNA donor with *TP53* wildtype sequence (Chira et al., 2018).

1.4.3 The ever-growing CRISPR toolbox

CRISPR/Cas' functionality is not limited to generating knockout/knock-in models by inducing DSBs and repairing them by NHEJ/HDR. One discovery that broadened the applications of CRISPR is the catalytically inactive, "dead" Cas9 (dCas9), which harbors two silencing mutations preventing the DNA endonuclease activity, while maintaining the specificity of its sgRNA-driven DNA sequence recognition (L. S. Qi et al., 2013). Thus, by fusing dCas9 to an effector module that instead of inducing DSBs fulfills another function, the applications of CRISPR in molecular biology are countless (**Table 1**).

For example, dCas9 has been applied to develop CRISPR tools for studying the epigenetic regulation of some genes. One of the most studied epigenetic regulation mechanisms is the hypermethylation of DNA at the CpG islands found in promoter regions of genes, which is associated with gene silencing (Portela & Esteller, 2010). However, the study of hypermethylation at specific loci is challenging because chemical demethylating agents, such as decitabine, have genome-wide effects. By fusing a methylase (DNA-methyltransferase 3A, DNMT3A) or a demethylase (ten-eleven translocation methylcytosine dioxygenase 1, TET1CD) domain to sgRNA-driven dCas9, we can direct the epigenetic modification to a concrete part of the genome, thus removing the background effect of other genes (Smith et al., 2022).

Another application is coupling dCas9 to a series of transcription factor recruiters (such as the VPR domain) or repressor domains (e.g., the Krüppel associated box, or KRAB) which results in CRISPR-activating (CRISPRa) or inhibiting (CRISPRi) technologies, respectively. These two applications present several advantages over classical overexpression of plasmids for gain-of-function

models, or CRISPR-KO and miRNA mimics/siRNAs for loss-of-function models. CRISPRa achieves overexpression levels within the physiological range, and because it acts at a transcription level, it does not overlook transcription regulation events that could be crucial to determine the function of certain genes, especially for some ncRNAs (Morelli et al., 2021). On the other hand, CRISPRi is better than CRISPR-KO and miRNA mimics/siRNAs in some instances, as the KO of a gene is often associated with more intracellular bypass mechanisms to overcome its loss, thus raising more clonal-specific events, whereas the downregulation by CRISPRi is reversible and leaves the genomic gene sequence unperturbed (L. S. Qi et al., 2013). Besides, as opposed to what happens with siRNAs or miRNA mimics, CRISPRi downregulation is less artificial and less prone to secondary effects derived from overexpressing an interfering RNA molecule way beyond its physiological levels, thus obtaining more biologically relevant results.

Table 1. The CRISPR toolbox: summary of different CRISPR applications in molecular biology.

CRISPR technology	Cas variant + complements	Application	Reference
CRISPR-KO	Cas9 + sgRNA	Gene KO by indel induction after DSB and NHEJ	(Ran et al., 2013)
CRISPR-knock-in	Cas9 + sgRNA + Donor DNA	Gene edition by HDR after DSB	(Chira et al., 2018)
CRISPR-DiR	dCas9-DNMT/TET1 + sgRNA	Targeted demethylation/methylation	(Smith et al., 2022)
CRISPRi	dCas9-KRAB + sgRNA	Gene expression inhibition	(L. S. Qi et al., 2013)
CRISPRa	dCas9-VPR + sgRNA	Gene overexpression	(Morelli et al., 2021)
Base editors	dCas9-cytidine/adenosine deaminase + sgRNA	Gene edition (just one nucleotide)	(Jeong et al., 2020)
Prime editing	Nickase-Cas9-RT + pegRNA	Gene edition without DSB	(Anzalone et al., 2019)
FiCAT	dCas9 + piggyBac transposase	Gene edition of long inserts (up to ~8 kb)	(Pallarès-Masmitjà et al., 2021)
DETECTR	Cas12a + sgRNA + cleavage reporter	Nucleic acid diagnostics (DNA)	(J. S. Chen et al., 2018)
SHERLOCK	Cas13a + sgRNA + cleavage reporter	Nucleic acid diagnostics (RNA)	(Kellner et al., 2019)

DSB – Double-strand break; NHEJ – Non-homologous end joining; HDR – Homology-directed repair; dCas9 – nuclease dead Cas9; DNMT – DNA methyltransferase 1; TET1 – Tet methylcytosine dioxygenase 1; KRAB – Krüppel associated box; VPR – V64-p65-Rta domain; RT – reverse-transcriptase; pegRNA – prime editing guide RNA.

Both CRISPRa and CRISPRi, along with classical CRISPR-KO, have been implemented in high-throughput screenings (CRISPR-screenings) as a way to detect genome-wide targets for genes involved in diseases, perform high-throughput cancer functionality studies, discover drug-sensitizing or resistance genes, or study the appearance of vulnerabilities and collateral dependencies after the application of a targeted therapy, as will be further discussed in Chapter 6 (Kurata et al., 2018; Thomsen & Mikkelsen, 2019).

For the purpose of gene edition, CRISPR-based base editors have also been developed, which allow correction of single-nucleotide variants (SNVs) or edition of a specific nucleotide. By fusing a cytidine or adenosine deaminase to dCas9, C→T and A→C transitions can be induced at specific positions, respectively (Jeong et al., 2020). A more versatile approach is prime editing, which can be used to also edit longer indels, and it is less limited by a PAM presence right next to a determined base transition. Prime editing consists of fusing a reverse-transcriptase (RT) to a modified Cas9 that only produces single-strand DNA breaks (Cas9 nickase). By using a longer sgRNA that includes the complementary RNA sequence to the desired edit (prime editing guide RNA, or pegRNA), edition can be achieved with a higher efficiency than by inducing HDR (Anzalone et al., 2019). Another approach is the find and cut-and-transfer (FiCAT) technology, which combines a dCas9 with an engineered piggyBac transposase, thus allowing the insertion of sequences up to 8 kb with double the efficiency than conventional HDR (Pallarès-Masmitjà et al., 2021).

Lastly, CRISPR also has applications in the field of nucleic acid diagnostics, thanks to the promiscuous nuclease activity of some Cas variants upon binding to their target, namely Cas12a and Cas13. By designing nucleic acid-based, quenched fluorescent reporters that emit a signal only after being cleaved, Cas12a/13 can be used to quickly and specifically detect a concrete sequence (i.e., a viral RNA or DNA genome) in a human serum, urine, or saliva sample. Cas12a was first described by Doudna's research group, it cleaves DNA, and it is applied in a technology they named DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter) (J. S. Chen et al., 2018). Conversely, Cas13 detects and cleaves RNA, and it was applied by Zhang and colleagues for its diagnostic use in a technology named SHERLOCK (Specific High sensitivity Enzymatic Reporter unLOCKing), which they used to detect the infection by Zika virus in human samples (Kellner et al., 2019). Both DETECTR and SHERLOCK require a previous isothermal amplification of the sample (in addition to a reverse-transcription in the case of SHERLOCK) and present several advantages over other diagnostic methods such as a quantitative PCR (qPCR), as it is faster and can be used at point-of-care since it does not require expensive and specialized instruments like a thermocycler. Because of this, several studies have tried to use DETECTR, SHERLOCK, and some other similar technologies for the detection of SARS-CoV-2 during the COVID-19 pandemics, and as a preparation for massive diagnostics in potential future viral epidemics (W.-F. Tang et al., 2023).

1.4.4 Limitations and concerns of CRISPR in today's medicine

As we have discussed, CRISPR is a revolutionary technology with a wide range of applications that has the potential to cure or treat genetic disorders, such as cancer. However, so far, the applications of CRISPR in cancer therapy have been limited and mostly restricted to CAR-T cell engineering. This is because of some concerns and limitations associated with the use of CRISPR/Cas9 directly in human patients, which include:

1. Off-target effects: one of the main concerns of using CRISPR/Cas9 in humans is the appearance of off-target effects if CRISPR/Cas9 cleaves the DNA at unwanted locations. Depending on the affected locus, this could lead to severe side effects and dramatical consequences, especially if this off-target edition occurs in germline cells, as these changes would be transferred to future generations (Ahumada-Ayala et al., 2023). To prevent this, it is necessary to exhaustively assess the specificity of the sgRNAs before implementing them into any clinical therapy. Also, as previously explained, several Cas9 variants are being developed, such as the high-fidelity Cas9 (HiFi-Cas9), which has fewer off-targets than regular SpCas9 (Vakulskas et al., 2018). Another approach is reducing the time of exposure to Cas9 edition, either by using an inducible system, or by modifying Cas9 to have a shorter lifespan (Chira et al., 2018).

2. Delivery: a key step in every therapeutical strategy is the delivery to target cells. The delivery of CRISPR therapies can be either local or systemic. Local administration is easier, but only suitable for certain tumor types, such as melanomas. Systemic delivery can reach otherwise inaccessible tumors, but it makes it difficult to direct the expression of CRISPR/Cas9 to tumor cells and therefore increases the risk of off-target effects in non-tumor cells (Arenas et al., 2022). Several deliveries strategies for CRISPR/Cas9, either in the form of plasmids, mRNA, or as ribonucleoparticles (RNPs) have been attempted. These include viral vectors (lentivirus, adenovirus, AAVs…), lipid-based nanoparticles, microinjection, or extracellular vesicles, each one of them with their own pros and cons in terms of efficiency and safety (Yip, 2020).

3. Immunogenicity: because of its bacterial origin, some elements of the CRISPR/Cas9 system might trigger immune-related adverse effects in their therapeutic applications. For instance, out of a random cohort of 22 blood donors, around 60-70% were found to have antibodies or T-cells reactive to SpCas9 (Charlesworth et al., 2019). Moreover, some secondary structures of sgRNAs could also be recognized by the immune system and initiate an immune response. Further studies are necessary to evaluate to which extent the immunogenicity of CRISPR/Cas9 can hinder its clinical applications (Azangou-Khyavy et al., 2020).

4. *TP53*-dependence: p53, also known as the "guardian" of the genome, is a very important tumor suppressor protein encoded by the *TP53* gene, which is implicated in DNA-damage response, apoptosis, and cell cycle control. Its inactivating mutation is a common feature across several types of tumors, and therefore a driver oncogenic event. Some authors have proved that Cas9-derived DSBs trigger a DNA damage response mediated by p53 that leads to cell apoptosis. However, *TP53*-mutant clones (which are relatively common in tumors) might survive and be specifically selected within the tumor population. Therefore, these authors recommend to carefully monitor the p53 status before and after edition (Azangou-Khyavy et al., 2020; Enache et al., 2020).

5. Ethical concerns: the therapeutical use of CRISPR is widely accepted and has proven to be key for the cure of some genetic diseases. Nevertheless, ethical concerns may arise when gene editing is used for eugenic purposes, which involve modifying genes to enhance or introduce desirable traits in the human species. In 2018, it was noteworthy the case of one scientist who acknowledged having applied CRISPR/Cas9 to twin embryos before their development with the aim of making them immune to HIV infection. This act was highly irresponsible and condemned by the international scientific community, as the mosaicism and potential off-target effects resulting from the editing may have fatal consequences on the health of these twins in the future, who will need to be monitored for the rest of their lives. Nowadays, most international guidelines and expert bioethics committees only allow the use of CRISPR on somatic cells with a therapeutic purpose, whereas its application in germline cells or viable embryos pre-implantation is strictly forbidden (Ahumada-Ayala et al., 2023).

In conclusion, while CRISPR/Cas9 holds tremendous potential for treating genetic diseases in human patients, it also presents some limitations and ethical concerns that are being addressed before its widespread application in clinics. Furthermore, some socioeconomic implications should also be considered, such as the lax regulations of CRISPR use in some developing countries, the equitable distribution of this technology, and the potential for it to exacerbate already existing inequalities. As CRISPR/Cas9 research continues to advance, it will be crucial for us scientists, clinicians, and politicians to engage in dialogue and ethical reflection to ensure that the technology is used safely and responsibly.

Chapter 2.

Materials and methods

Chapter 2. Materials and methods

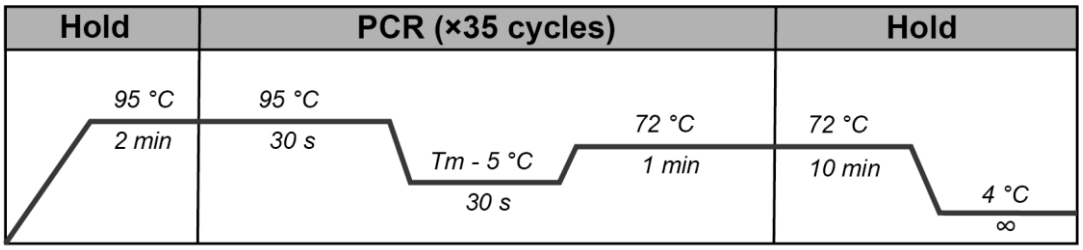
This chapter will compile the general methodology and materials employed throughout this thesis. The methods are divided in molecular biology methods (general DNA, RNA, and protein protocols), cell experiments *in vitro* and *in vivo*, gene delivery methods, screenings and high-throughput methods, and bioinformatic methods, statistics and ethics.

2.1 Molecular biology methods

2.1.1 DNA extraction and PCR

First, cell pellets were obtained by centrifugation of trypsinized cells at 300 ×*g* for 5 min. Total genomic DNA (gDNA) was extracted from the cell pellets using the QuickExtract™ DNA Extraction Solution (Lucigen, #QE09050) following manufacturer's guidelines. A polymerase chain reaction (PCR) was afterwards performed to amplify concrete regions within that gDNA, using specific forward and reverse primers for each gene that are compiled in **Supplementary Table 1**.

For regular PCR amplification, the DreamTaq Green PCR Master Mix (ThermoFisher Scientific, #K1081) was used, following the manufacturers' recommendations. In brief, 50-100 ng of DNA were mixed with the DreamTaq reaction buffer and the corresponding primers, and incubated in a thermocycler following the program:



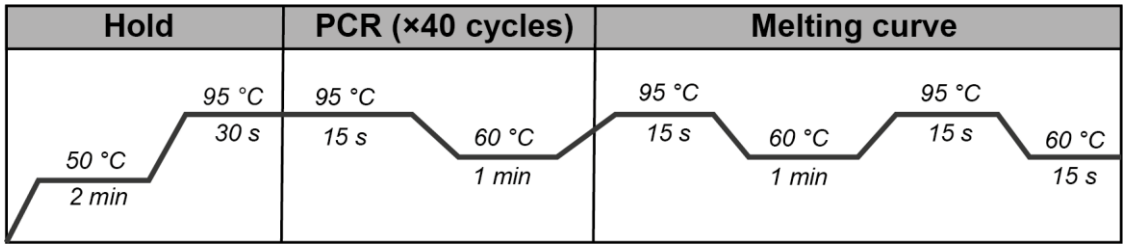
2.1.2 RNA extraction

Before reverse transcription quantitative PCR (RT-qPCR), total RNA was extracted from dry cell pellets using TRIzol (TRI Reagent®, Sigma-Aldrich, #93289), following the manufacturer's recommended protocol, and inside a clean air hood to prevent RNase contamination. RNA concentration and purity (260/280 and 260/230 ratios) were then measured using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, ND-2000). To avoid DNA

contamination, extracted RNA was digested with DNase I (Invitrogen, #18068015) following the recommended protocol. RNA was then stored at -20 °C for early use, or at -80 °C for long-term storage.

2.1.3 RT-qPCR

Total RNA (2 µg) was reverse-transcribed to copy DNA (cDNA) using RevertAid RT kit (ThermoFisher Scientific, #K1691). A SYBR Green quantitative PCR reaction was then performed from 25-100 ng of cDNA template, using KAPA SYBR® FAST (Merck, #SFUKB), in 96-well plates, and measured at the QuantStudio™ 3 Real-Time PCR System (ThermoFisher Scientific, #A28567). The next program was followed:



Relative expression was then calculated using the ΔΔCt method, using a reference gene to normalize the gene expression among samples (*GAPDH* for protein-coding genes, *U1 snRNA* for lncRNAs, and *SNORD44* for miRNAs).

Because of their short length, for RT-qPCR of miRNAs a polyadenylation strategy was followed as described by (R. Shi & Chiang, 2005). Briefly, a polyadenylation step using Poly(A) Polymerase Tailing kit (Epicentre, #PAP5104H) was added before reverse-transcription. Then, the reverse-transcription was performed substituting the random hexamer primer included in the RevertAid RT kit for an adapter-oligo-dT primer (see sequence in **Supplementary Table 1**). This allows the elongation of the miRNA with a poly(A) tail and an adapter sequence, so that in the qPCR the forward primer matches the sequence of the miRNA, whereas the reverse primer is universal and complementary to the adapter sequence.

2.1.4 Protein extraction and Western blot

For protein extraction, cell pellets were resuspended in RIPA lysis buffer containing phosphatase (7 mM sodium orthovanadate) and protease inhibitors (0.2 mM phenylmethylsulfonyl fluoride (PMSF), and Pierce™ Protease Inhibitor Minitablets (ThermoFisher Scientific, #A32955), according to manufacturer's instructions). Cell suspensions were incubated at 4 °C for 20 min, and then

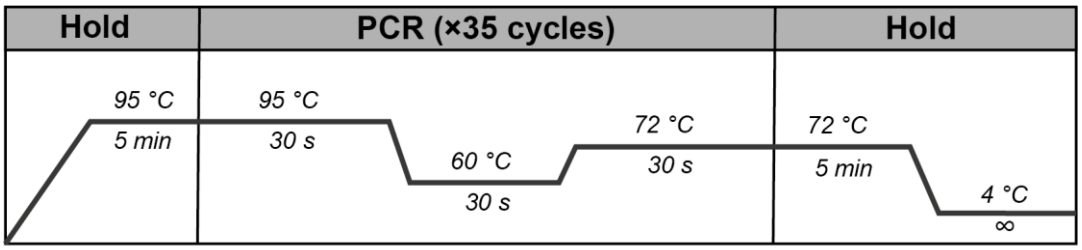
centrifuged at 4 °C, 16,000G for 15 min. The protein-containing supernatant fraction was collected, and the protein concentration was quantified by Bradford (VWR, #A6932) at 600 nm in a GloMax® Microplate Reader (Promega, #GM3000).

For the detection of specific proteins by Western blot, 30-50 µg of extracted protein were loaded into a sodium dodecyl sulfate (SDS) protein acrylamide gel electrophoresis (SDS-PAGE) to separate proteins according to their molecular weight. Proteins were then transferred to a polyvinylidene difluoride (PVDF) membrane, and then the membrane was blocked with phosphate-buffered saline (PBS) + 5% powder milk + 0.1% Tween-20. The membrane was then incubated with the primary antibodies overnight at 4 °C with constant shaking. The next day, the membrane was washed for 10 min three times in PBS + 0.1% Tween-20, and then incubated with the secondary antibodies labeled with horseradish peroxidase (HRP). All the used antibodies are listed in **Supplementary Table 3**.

Protein specific bands were revealed using SuperSignal™ West Femto substrate (ThermoFisher Scientific, #A43841) and an ImageQuant™ LAS-4000 imaging system. Lastly, Western blot bands were digitalized and then analyzed by densitometry using the ImageJ software and normalizing with the expression of the protein taken as loading control (β-actin or α-tubulin).

2.1.5 Vector cloning

The empty vector (EV) pLVX-IRES-zsGreen1 (Clontech, #632187) was used as the backbone to generate a vector to introduce the lncRNA *DLG2-AS1* into the cells. Two oligonucleotides containing overhanging restriction sites for EcoRI and XbaI were designed (sequences shown in **Supplementary Table 1**). To amplify the insert, a PCR was performed using Phusion High-Fidelity DNA Polymerase (ThermoScientific, #F530S) because of its better performance in terms of less error introduction compared to other polymerases. The reactions were set using the manufacturer's recommendations and the following program:



Then, the PCR product was run and the ~200 bp insert was purified from an 1.5% agarose gel using a GenElute™ Gel Extraction Kit (Sigma-Aldrich, #NA1111). The purified insert was then double-digested along with the EV using EcoRI and

XbaI (ThermoScientific, #ER0271 and #ER0681). The cut insert and EV were ligated using T4 ligase (New England Biolabs, #M0202) and then the resulting plasmid was introduced into chemically competent bacteria (*E. coli* DH5-α) for its amplification. All reactions were set up following the manufacturers' recommendations.

2.1.6 Sanger sequencing

PCR reaction products or plasmid digestions were run in a 1-2% agarose gel, then the band corresponding to the expected size of the amplicon was excised, and the DNA was extracted using a GenElute™ Gel Extraction Kit (Sigma-Aldrich, #NA1111). Purified DNA was measured using a NanoDrop 2000 device (ThermoFisher Scientific, ND-2000), and then ~200 ng of DNA were mixed with a sequencing primer and shipped to STABVIDA (Caparica, Portugal) for Sanger sequencing. Results were downloaded and analyzed using the SnapGene Viewer (v4.3.11) software.

2.1.7 T7 endonuclease assay

To analyze the efficiency of the different sgRNAs designed for CRISPR/Cas9 edition, a T7-endonuclease I assay was performed (New England Biolabs, #M0302). In brief, gDNA was extracted from CRISPR/Cas9 edited cells, and a PCR was performed to amplify the *KRAS* region (primer sequences found in **Supplementary Table 1**). Then, a heteroduplex reaction and a T7 endonuclease digestion were carried out following the manufacturer's recommendations, and the digestion product was run in an agarose gel electrophoresis. The T7-endonuclease only cleaves the DNA if it detects mismatches in the formed heteroduplexes, which should only be present if there has been Cas9 edition. Therefore, a double band should be expected when edition is successful, whereas a single band is representative for the lack of edition.

2.1.8 Biotin pull-down

To detect direct interaction between miRNAs and their mRNA targets, a biotin pull-down experiment was carried out to first prove that biotinylated miRNAs were being loaded onto the RISC. First, we performed an *in vitro* RNA synthesis of the miRNAs using biotin-labelled deoxyribonucleotides (IDT, #232691864/#232691865), the T7 RNA polymerase (Promega, #P2077), and a DNA template containing the T7 RNA polymerase promoter (5'-TAATACGACTCACTATAGGG-3') followed by the specific miRNA sequence. The transcription product was treated with DNase I (Invitrogen, #18068015) and the RNA was purified using the RNeasy Mini kit (Qiagen, # 74104).

For the pull-down, the protocol described by (Gerber et al., 2006) was followed, with some modifications. In brief, 2-3 million cells were lysed using a home-made IP buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 5% glycerol, 0.5% Triton X-100, 1 mM DTT, 0.1 mg/mL heparin, 0.1 mg/mL tRNA, and 2 protease inhibitor tablets). Then, the lysates were incubated along with the biotinylated miRNAs and streptavidin-labelled Dynabeads™ M-280 (ThermoFisher Scientific, #11205D). Finally, after several washes, the proteins bound to the streptavidin beads were eluted, run in an SDS-PAGE, and blotted against the AGO2 protein.

2.2 Cell experiments *in vitro* and *in vivo*

2.2.1 Cell culture

All tumor cell lines used were purchased from the American Type Culture Collection (ATCC). Cells were kept frozen in cryovials in liquid nitrogen (-196 °C) suspended in fetal bovine serum (FBS) + 10% dimethyl sulfoxide (DMSO). RAS-less mouse embryonic fibroblasts (MEFs) models of human $KRAS^{WT}$ (#RPZ25854), $KRAS^{G12C}$ (#RPZ26186), and $KRAS^{G12D}$ (#RPZ26198) were kindly donated by the RAS initiative team at the Frederik National Laboratory for Cancer Research (https://www.cancer.gov/research/key-initiatives/ras/outreach/reference-reagents/cell-lines).

For daily 2D culture, cells were kept in T-75 culture flasks or P-100 dishes for a maximum time of 2 months, changing the medium and splitting the cells by trypsinization every 3-5 days depending on their confluence state. Cells were cultured under standard conditions (at 37 °C, in a 5% CO_2 atmosphere) using DMEM High Glucose medium (Biowest, #L0104) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S); or RPMI 1640 medium (Biowest, #L0501) supplemented with 10% FBS, 1% P/S, and 1% L-glutamine; depending on the requirements of each cell line (see **Supplementary Table 4**).

For spheroid 3D culture, 2,000 cells/well were plated in ultra-low attachment 96-well plates (Corning Costar ®, #3474). Cells were mixed with 50 µL Matrigel (Corning, #354248) in RPMI 1640 complete growth medium in a 1:1 proportion to establish the scaffold system for the spheroid suspensions. Then, 100 µL of RPMI medium was added on top of each well to reach a final volume of 200 µL. The wells were refilled once a week with fresh medium to compensate the liquid evaporation.

2.2.2 Cell viability assays

For regular cell viability assays with resazurin, 2,000 cells/well were plated on 96-well plates, and then 20 µL of resazurin (0.1 M resazurin sodium salt powder (Sigma-Aldrich, #R7017) diluted 1:200 with fresh medium) were added at 0, 24, 48, and 72 hours. Cells were left in the incubator at 37 °C for 4-6 hours after adding the resazurin, then the reaction was stopped by adding 30 µL of 3% SDS, and fluorescence was measured at 590 nm using a GloMax® equipment. Fluorescence measure results were normalized by the fluorescence at t=0 h. All the experiments were performed in triplicates or quadruplicates.

For CellTiter-Glo® cell viability assays (Promega, #G7571), 3,000 cells/well were plated in 96-well culture plates. Cell viability was measured 7 (2D culture) or 12 days (3D culture) later according to the manufacturer's instructions. Briefly,

100 µL of CellTiter-Glo reagent were added; vigorously mixed for 5 min, ensuring the dissolution of the spheroid matrix in the case of 3D cultures; and then placed on a plate shaker for 25 min to ensure complete cell lysis prior to assessment of the luminescent signal at 525 nm in a GloMax® equipment.

2.2.3 Colony assay

Cells were plated in 6-well plates at very low density (1,000-3,000 cells/well), and they were kept in incubation for 1-2 weeks at 37 °C, depending on the cell growth rate. Medium was renewed once a week. After that time, cells were fixed with a fixating and staining solution (PBS + 13% formaldehyde + 0.1% crystal violet) to visualize the colonies. Pictures were taken using a scanner, and the ImageJ software was used to analyze the images and count the number of colonies.

2.2.4 Scratch wound-healing assay

Cells were cultured in 6-well plates until a 100% confluence was reached. Then, a scratch with a sterile pipette tip was made and the regular media were substituted by starvation medium (regular medium supplemented with 1% FBS instead of 10% FBS). Pictures were taken at 0 and 24 hours after the scratch was made, they were analyzed using the Measure Wound Healing Tool from the software ImageJ, and the percentage of closed wound after 24 hours was calculated.

2.2.5 Tumor cell-derived xenografts in mice

Around 1x10^6 tumor cells/xenograft were collected by centrifugation and resuspended in 200 µL of 1:1 RPMI:Matrigel (Corning, #354248). Then, 8-week-old female NOD Scid Gamma (NSG) mice were anesthetized in a sealed chamber using 2.5% isoflurane in O_2 at 0.2 L/min, and the cells were injected subcutaneously into the rear flanks of the mice. Tumor xenografts were left to grow for 2 weeks until a small bulge was visible in the flanks of the mice, and then doxycycline treatment was started as described in (Cawthorne et al., 2007). In brief, doxycycline was administered orally *ad libitum* mixed with jelly (2 mg/mL), inside small cups of ~15 mL that were placed in the mice cage every two days for a week. Tumor growth was then evaluated twice a week by measuring tumor sizes, and the tumor volume was calculated using the formula $V = \frac{ab^2}{2}$, where a is the tumor's longer axis, and b the shorter one. Mice were euthanized and the experiment concluded when one of the flank tumors reached 3,000 m^3 in volume, or more than 21 days had passed.

2.3 Gene delivery methods

2.3.1 Transfection of plasmids and miRNA mimics

Cells were seeded (250,000 cells/well) in 6-well plates and cultured until a cell confluence of 60-70% was reached. Different transfection reagents were employed, always following the manufacturer's recommended reaction settings: TransIT (Mirus Bio, #MIR-6000) for regular transfection of plasmids, and lipofectamine RNAimax (ThermoFisher Scientific, #13778075) for miRNA mimics (Merck, #900010/20/30). Cells were incubated in reduced serum OptiMEM™ medium (ThermoFisher Scientific, #31985062) with the plasmid and the transfection reagent for 24 hours, and then transfection medium was replaced by fresh medium. Overexpression of the desired gene was then confirmed by RT-qPCR 48 or 72 hours after transfection, and by daily observation under a fluorescence microscope when expressing a reporter fluorescent protein (zsGreen1/EGFP/mCherry).

2.3.2 RNPs delivery

CRISPR RNA (crRNA) for *KRAS*^{G12C} or *KRAS*^{G12D}, *trans*-activating crRNA (tracrRNA) and a high-fidelity version of Cas9 (HiFi-Cas9) were purchased from Integrated DNA Technologies (IDT), and then properly conjugated to form ribonucleoproteins (RNPs) according to the manufacturer's instructions.

RNPs were delivered into the cells either by regular lipofection using CRISPRMAX™ transfection reagent (ThermoFisher Scientific, #CMAX00001), or by electroporation using an Amaxa 4D-Nucleofector (Lonza, #AAF-1003), testing different nucleofection programs (EH-158, EN-138, CM-130, or CA-137). All procedures were performed following the manufacturer's recommended protocols.

2.3.3 Lentivirus preparation

Lentiviral particles (LVs) were prepared using HEK-293T cells and the packaging plasmids VSV-G (Addgene, #8454) and psPAX2 (Addgene, #12260). First, HEK-293T were kept in culture in T-175 flasks and then co-transfected with the two packaging plasmids and the desired lentiviral plasmid, using LipoD293™ (SignaGen Laboratories, #SL100668). LVs were collected 2 (1st pick-up) and 3 days (2nd pick-up) after transfection, filtering the old media with 0.45 μm filters. For LV concentration, we performed an ultracentrifugation for 2 hours, at 4 °C and 70,000 ×*g*.

For lentiviral titration, we used K562 cells because of their easiness to be infected. Cells were infected with 5, 10, and 50 μL of the LV to be titrated, and then

incubated at 37 °C for 5 days. Finally, a FACS Aria cytometer was used to measure the percentage of fluorescent green positive cells, and the titer of the virus was calculated taking a fluorescence percentage < 20%, and applying the following formula:

$$PFU\ \left(\frac{LV}{mL}\right) = \frac{\left(\% \cdot 100{,}000\ \frac{LV}{100}\right) \cdot 1000\ \frac{\mu L}{mL}}{x\ \mu L}$$

2.3.4 Lentiviral infection and clone selection

The day before infection, ~250,000 cells/well were plated in a 6-well plate. The next day, or when the cell confluency reached 60-70%, the calculated amount of LV considering the chosen multiplicity of infection (MOI) was added to the wells mixed with 1 mL of fresh media. For inducible HiFi-Cas9 LVs infection, a low MOI (0.3-1) was used to ensure a single integration. Cells were incubated with the LV for 5 hours, and then the virus was removed and replenished with fresh medium.

For selection of successfully infected cells, cells were either incubated in medium with puromycin (1 µg/mL) for at least one week, or sorted by green fluorescence using a FACS Aria cytometer, depending on the reporter gene included in the lentiviral vector. Infected cells were regularly checked for green fluorescence under the microscope to check that the lentiviral integration had not been lost.

2.3.5 Adenoviral infection

Adenoviral vectors (AdVs) were designed to contain HiFi-Cas9 + both *KRAS*$^{G12C/G12D}$ sgRNAs + EGFP (AdV-Cas9), or just EGFP as a control (AdV-GFP). The plasmid cloning and the viral packaging were performed by VectorBuilder (Guangzhou, China), obtaining ultra-purified, high-titer AdVs (>10^{12} viral particles/mL).

For AdVs infection, ~300,000 cells were infected with a MOI = 1000 viral particles/cell. Cells were collected and mixed with AdVs and then left for incubation at 37 °C in a 1.5 mL tube. After 2 h, cells were replated in 2D or 3D 96-well plates for cell viability assays.

2.4 Screenings and high-throughput methods

2.4.1 LncRNA array profiling

LncRNA expression profiles were studied in LUAD patient samples paired with their adjacent, non-tumor samples by RT-qPCR, using the Human LncRNA Profiler Array kit (System Biosciences, #RA900A-1). This commercial kit includes the primers for the detection of a panel of 90 disease-related lncRNAs in humans (https://www.systembio.com/lncrna-profiler-qpcr-array-kit-human), along with 5 reference genes (*18S rRNA*, *RNU43*, *GAPDH*, *LAMIN A/C*, and *U6*) for normalization of expression results, and a negative reaction control.

2.4.2 RNA-sequencing

a. RNA sample preparation and sequencing

RNA sequencing (RNAseq) was performed to discover genes or pathways altered after the overexpression of wildtype (WT) or mutant (MUT) *miR-133b*. After transfection of two cell lines (A549 and H2126) with scramble, miR-$133b^{WT}$, or $miR\text{-}133b^{MUT}$ mimics, cell pellets were freshly obtained and flash-frozen in liquid nitrogen. RNA was immediately extracted from frozen pellets using a miRVana™ miRNA isolation kit (ThermoFisher Scientific, #AM1560). Three replicates were performed for every condition of the experiment.

Before shipping the samples in dry ice to GENEWIZ (Leipzig, Germany), the concentration, purity, and integrity of extracted RNA were checked using an Agilent TapeStation Instrument. An RNA Integrity Number (RIN) threshold of at least 8 was considered optimal for RNA sequencing. Finally, a strand-specific RNAseq, with poly(A) selection and 20-30 million reads per sample was performed.

b. Quality check and initial data filtering

'FASTQC' (v0.11.9) was used for the initial quality check (QC) of raw FASTQ files downloaded after completion of the sequencing. Reads were aligned to the GRCh30.d1.vd1 human genome from GDC, and BAM files were obtained using *STAR* (v2.6.0b). BAM general QC was performed using '*qualimap'* (v2.2.1), and the read counts per gene were calculated using '*HTSEq'* (v2.0.2) with the following parameters: *htseq-count -m intersection-nonempty -f bam -t exon -i gene_id -r pos -s reverse.*

Data were first filtered by removing expression counts from unaligned reads (*__not_aligned*), non-unique alignments (*__alignment_not_unique*), low alignment quality (*__too_low_aQual*), ambiguous association to more than one gene

(__*ambiguous*), and not associated to any feature defined in GENCODE (__*no_feature*). Genes without >=10 reads in at least 2 samples were filtered out.

c. Expression data analysis

For the analysis of differential expression, '*DESeq2*' (Bioconductor 3.15) with independent filtering was used, specifying a threshold of *alpha = 0.01*, and *abs(logFC) > log_2(1.5).* Graphs were plotted using *ggplot2* R package.

d. Pathway analysis

First, low-expression filters were applied, and genes were ranked by their shrunken logFC. The mean of logFC was taken for duplicate gene names corresponding to different Ensembl genes IDs. Then, '*fgsea()*' was used against the whole MSigDB database. Also, '*collapsePathways()*' was used to simplify the list of pathways and remove the redundant ones, pre-filtering the list of pathways by *padj < 0.05*.

2.4.3 CRISPR-screening

A CRISPR-screening was performed to discover which collateral dependencies appear after knocking out the *PKP1* oncogene in the LUSC cell line SK-MES-1. The top candidate genes could be a source of resistance and/or potential co-targets of a future molecular therapy against *PKP1* in LUSC.

In brief, a lentiviral, whole genome-targeting library of sgRNAs is used to infect the cell lines, and they are kept in culture for growth selection for 21 days. After that time, cells are collected, the DNA is extracted, and next generation sequencing (NGS) is performed to analyze the enrichment or loss of sgRNAs. Those sgRNAs that are depleted from the KO populations, but not in the control, are essential genes for the KO cell, and therefore, potential collateral dependencies or vulnerabilities that arise after targeting the KO gene in a molecular therapy (Figure 9).

a. Lentiviral library preparation and titration

Prior to the screening, lentivirus containing the CRISPR/Cas9 system were delivered to the 3 cell lines (parental SK-MES-1, *PKP1*-KO1, and *PKP1*-KO2) and selected after infection to ensure stable SpCas9 expression. The human CRISPR Knockout Pooled Library Brunello (Addgene, #73178) was expanded in HEK-293T cells and packaged into lentiviral particles as described by (Thomsen & Mikkelsen, 2019). The Brunello library contains 76,441 sgRNAs targeting 19,114 human genes

(4 sgRNAs/gene), and its integrity was regularly checked by sequencing after every expansion.

For the titration of the library prior to its use, 100,000 cells/well were plated in 6-well plates, and then serial dilutions of the virus ranging from 10^{-2} to 10^{-6} were prepared. Two days after infection, puromycin (2 ng/µL) was added to the medium for selection of successfully infected cells. Cells were left in culture for 2 weeks, and then 0.6% methylene blue in methanol was added to visualize the colonies. The number of colonies was then counted, and the colony forming units (CFU) per mL of virus was calculated.

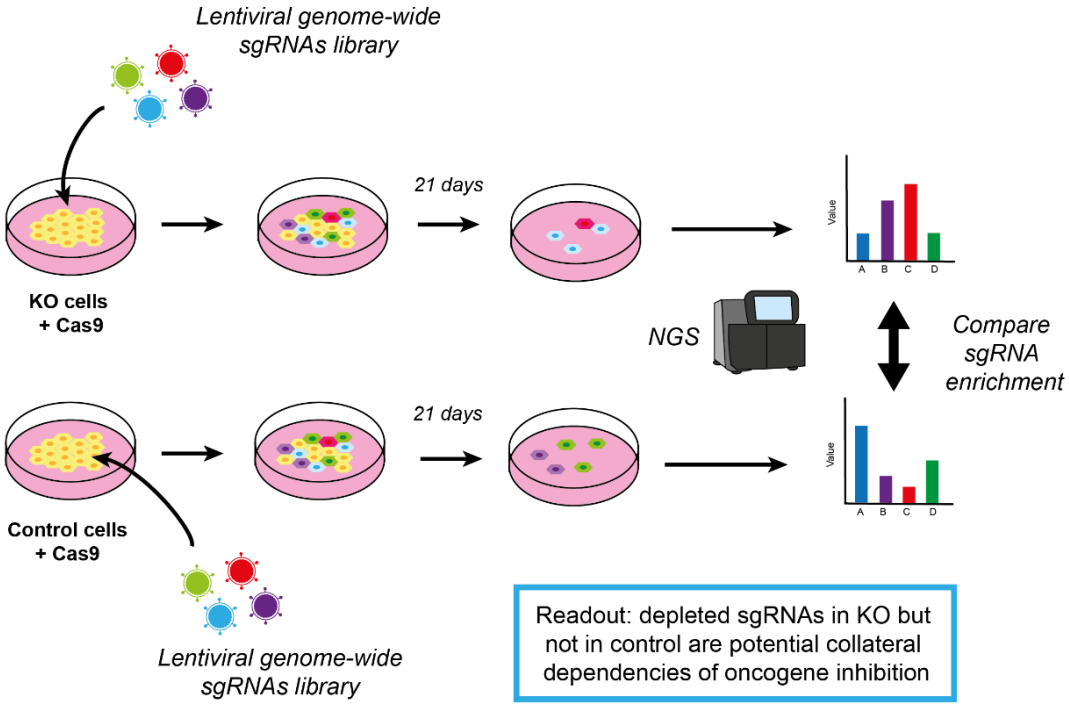


Figure 9. Schematics of a genome-wide CRISPR-screening protocol for the detection of collateral dependencies appearing after the KO of an oncogene under study.

b. Lentiviral library infection and selection

To ensure enough coverage of the screening (at least 500 cells/sgRNA), 80 million SK-MES-1 control, KO-1 and KO-2 cells were plated in 20 T-175 flasks per condition. One day after seeding, cells were infected with the Brunello library (MOI=0.5) and three days later puromycin (2 ng/µL) was added to the growth medium for selection of positively infected cells.

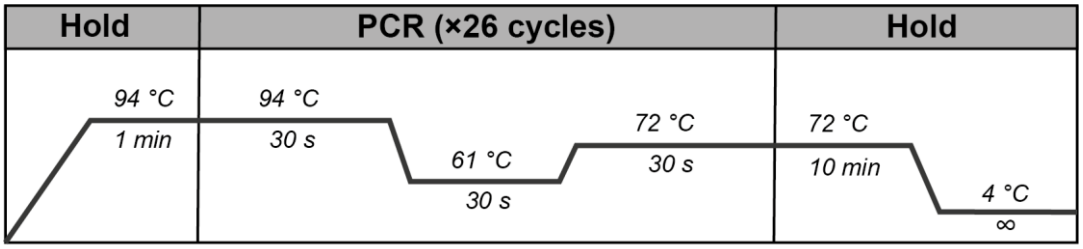
Cells were trypsinized, counted, and 80 million cells per condition were replated into T-175 flasks every 2-3 days, keeping the puromycin selection until day 10 after infection, when half of the cells (40 million) were pelleted (t=7 days

post puromycin selection), and the rest were kept in culture for another 14 days (until t=21 days post puromycin selection).

c. gDNA extraction and NGS sample preparation

Cell pellets were resuspended in 12 mL of lysis buffer (10 mM Tris-HCl, 1 mM EDTA, 0.5% SDS, 150 mM NaCl, pH=10.5) and 60 µL of proteinase K (Sigma-Aldrich, #539480) were added. The lysis reaction was incubated overnight at 55°C, and then an ethanol/salt DNA precipitation was performed.

Next, to amplify the sgRNA region before PCR, a large-scale PCR protocol optimized by Dr. Emil A. Thomsen from Giehm's lab was used. Briefly, it was estimated that approximately 26 PCR reactions/sample using 10 µg of gDNA as a template are necessary to maintain coverage. Reactions were set up using Ex-Taq® DNA polymerase (Takara Bio, #RR01CM), and a set of 22 forward primers, each one containing a different stagger region, are used to increase library diversity and to prevent sequencing errors during NGS. The next PCR cycling protocol was followed:



An aliquot of one PCR reaction/sample was run in a 1% agarose gel to check the proper functioning of the PCR by appearance of a sole band. Then, all the reactions from the same sample were pooled together and a magnetic bead DNA purification was performed using HighPrep™ PCR Clean-Up system (Magbio, #AC-60050), following the manufacturer's recommended protocol. Finally, purified DNA concentration was measured in a Qubit™ fluorometer (ThermoFisher Scientific, #Q33327) before shipping the samples to the NGS facility.

d. NGS analysis

Library preparation and NGS (Illumina NovaSeq 6000, 650-800M reads) were performed at MOMA NGS Core Center (Department of Molecular Medicine, Aarhus University Hospital, Denmark). The initial quality check was performed using '*FASTQC*' *(v0.11.9)*. Next, '*cutadapt*' (v2.10) in Python (v3.8.5) was used to trim down the reads to the sgRNA region, and then '*bowtie*' (v1.3.0) was used to map the reads to all the sgRNAs of the Brunello library. Finally, the statistical

analysis was performed using the *JACKS* tool (Allen et al., 2019), and further network analysis was performed using '*FDRnet*' (L. Yang et al., 2021) and the STRING database.

2.5 Bioinformatic methods, statistics, and ethics

2.5.1 Omics databases and bioinformatic analysis tools

The following databases and online tools were employed for the purposes thereafter detailed:

- UCSC Genome Browser (https://genome.ucsc.edu/): primer design, *in silico* PCR, genome loci and gene annotations visualization.

- Ensembl v109 (https://www.ensembl.org/): download of cDNA sequences, identification of isoforms and transcripts.

- TGCA/cBioportal (https://www.cbioportal.org/): download of RNAseq data from external LUAD cohorts.

- miRbase v22.1 (https://www.mirbase.org/): sequence and annotation retrieval of mature and pre-miRNAs.

- TargetScan v8.0 (https://www.targetscan.org/) and miRDB (https://mirdb.org/): prediction of miRNA targets.

- DepMap portal (https://depmap.org/portal/ccle/): includes the Cancer Cell Line Encyclopedia (CCLE), containing gene expression and mutation data for most cancer cell lines.

- Synthego's Inference of CRISPR Edits (ICE) (https://ice.synthego.com): analysis of CRISPR gene edition efficiency from Sanger sequencing files.

- STRING v11.5 (https://string-db.org/): obtention of a map of protein-protein interactions.

- Metascape v3.5 (https://metascape.org/): meta-analysis of a list of genes, including pathway analysis and protein interactions.

2.5.2 Analysis of TCGA-LUAD data

Raw RNA-Sequencing (RNA-Seq) gene expression counts were downloaded from the TCGA-LUAD project (n = 533) using the Genomic Data Commons Data Transfer Tool (v 1.4.0) (May 9th, 2019). RNA-Seq data were analyzed using the R package '*edgeR*' (R version 3.5.3, Bioconductor version 3.8). Data were normalized using the trimmed mean of M-values method, and the counts per million (CPM) were extracted for *DLG2-AS1* (ENSG00000274006.1) and for *DLG2* (ENSG00000150672.15).

2.5.3 Statistical analysis

GraphPad Prism 9.0.0's two-tailed Student's t-test considering the mean values, the standard deviation, and the number of replicates (at least, three) was used to determine whether the difference between two sets of normally distributed data is significant (*p*-value < 0.5).

Correlation between the expression of *DLG2* and *DLG2-AS1* was calculated using Pearson's correlation test. For the ROC curves, we fitted logistic regression models to predict the classification of samples as "tumor" or "normal" based on gene expression, using the R package '*pROC*' (R v3.5.3).

2.5.4 Ethics for human patients and animal procedures

RNA samples from LUAD patients were obtained from the Basque Biobank (Bilbao, Spain). Participants provided written consent in accordance with the procedures of the Declaration of Helsinki and the institutional and national guidelines. Seventy tumor samples were taken from primary malignant LUAD tumors, as well as their adjacent non-tumor tissues, based on macroscopic examination by trained pathologists. The study was approved by the Ethics Committee (CEI Granada), Department of Health, Andalusian Regional Government, and by the Basque Foundation for Health Innovation and Research, Spain.

The animal research procedures were performed following the European Directive 2010/63/EU. The PhD candidate who performed the experiments is habilitated for mice and rats care, euthanasia, and experimentation (FELASA functions A, B, and C, capacitation certificate #EXP-001057).

Chapter 3.
Characterization of novel lncRNA biomarkers for LUAD

Chapter 3. Characterization of novel lncRNA biomarkers for LUAD

This chapter will address Objective 1.1. Here, we aimed to characterize lncRNAs whose expression is dysregulated in LUAD patients and could therefore be used as diagnostic or prognostic biomarkers for the disease. Most of the contents from this Chapter are compiled in our publication: "*DLG2-AS1 as a novel biomarker for lung adenocarcinoma.*" (Arenas et al., 2020).

3.1 Background

Because most research up to date searching for novel biomarkers and altered genes in cancer has focused on the protein-coding part of the genome, in our group we aimed to discover and characterize novel non-coding elements that appear altered in LUAD patients. In here, we focused on lncRNAs, which as previously discussed in Chapter 1, are greatly versatile molecules with an important role in gene expression regulation. Overall, lncRNAs exhibit a range of qualities that makes them interesting candidates as biomarkers for cancer, and, in some instances, they even surpass protein biomarkers or protein-coding mRNAs in terms of:

1. Stability: lncRNAs show high stability when circulating in body fluids, especially if they are encapsulated into exosomes or lipoproteins (Bolha et al., 2017). In addition, they can withstand harsh conditions, such as several freeze-thaw cycles, and up to 24 h at room temperature, without affecting their measured expression levels by RT-qPCR in plasma samples from lung cancer patients (Yuan et al., 2020).

2. Specificity: overall, lncRNA expression patterns display a higher tissue- and cell type-specificity than protein-coding genes (C. Jiang et al., 2016), which makes them useful biomarkers for specific tissues or cancer types.

3. Quantifiability: lncRNAs can be easily quantified using RT-qPCR, a well-stablished and relatively low-cost detection method. Also, some panels of disease-related lncRNAs and microarrays are already commercially available for a higher throughput (P. Qi et al., 2016).

4. Accessibility: lncRNAs dysregulation in primary tumors is mirrored in the circulating lncRNA levels in several body fluids, including whole blood, plasma, urine, and saliva. This reduces the need for invasive procedures such as biopsies and allows for frequent monitoring of cancer progression (Bolha et al., 2017).

There are several examples of clinically relevant lncRNAs as cancer biomarkers, such as *PCA3* (prostate cancer), *MALAT-1* (lung cancer), *HOTAIR* (oral squamous cell carcinoma), and *H19* (gastric cancer) (Bolha et al., 2017). Of note is the case of *PCA3*, which was approved by the FDA in 2012 for its use in patients under suspicion of requiring a biopsy for the diagnosis of prostate cancer (G. L. Lee et al., 2011). Therefore, in this chapter, we will address the discovery and characterization of novel lncRNAs as potential biomarkers for LUAD.

3.2 Results

3.2.1 Selection of a candidate lncRNA

First, to select a lncRNA candidate dysregulated in LUAD, an initial RT-qPCR screening was performed using a commercial panel of disease-related human lncRNAs (Human LncRNA Profiler® kit from System Biosciences). With such panel, we measured the expression of 90 disease-related lncRNAs in 5 initial LUAD samples from patients, each one with its paired, adjacent, non-tumor tissue. We found 3 lncRNAs (*E2F4-antisense*, *DLG2-AS1*, and *lincRNA-SFMBT2*) which were downregulated (tumor/normal fold change < 0.66) in all of the 5 patient samples (Figure 10). *DLG2-AS1* was selected as our main candidate for further validation because no cancer-related studies were published about it and, moreover, it was the candidate that presented the best amplification results after running the qPCR products in an agarose gel (Supplementary Figure 1).

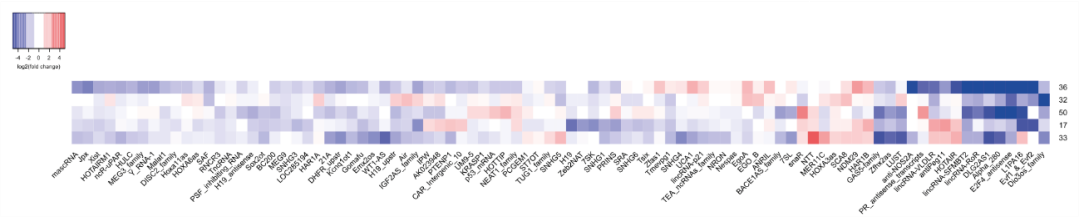


Figure 10. Heatmap of the expression of 90 lncRNAs included in the Human LncRNA Profiler ® kit (System Biosciences) in an initial screening of 5 LUAD patients.

3.2.2 DLG2-AS1 is downregulated in LUAD patients

After designing a new pair of specific oligonucleotides for *DLG2-AS1*, we performed RT-qPCR on the rest of our 70 patients cohort to measure the expression levels of *DLG2-AS1* in LUAD and paired, non-tumor RNA samples. Five patient samples (#19, #28, #33, #34, and #41) were discarded due to low quality amplification, possibly because of RNA degradation. The expression of *DLG2-AS1* in the remaining tumor samples (n=65) were lower than those observed in the paired, non-tumor samples (p<0.0001) (Figure 11A). Out of the 65 patients, 67.7% of them (44/65) presented *DLG2-AS1* downregulation in the tumor sample (Figure 11B), thus showing, for the first time, that *DLG2-AS1* is downregulated in LUAD patients, which could suggest a tumor suppressor role of this lncRNA in LUAD (Arenas et al., 2020).

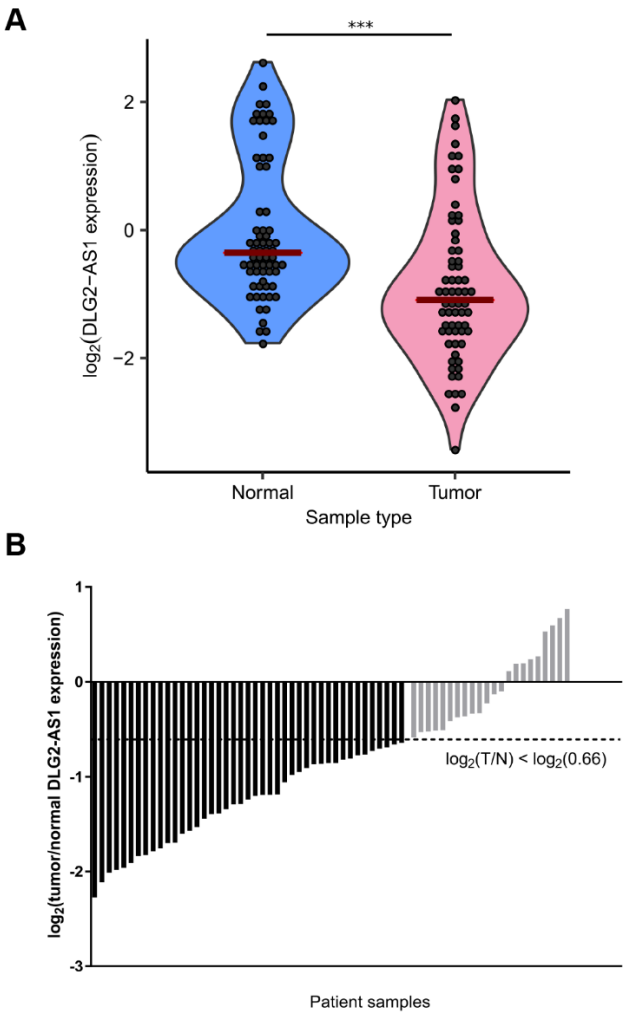


*Figure 11. (A) DLG2-AS1 expression in normal and tumor paired samples. (B) Tumor/normal fold change (FC) of DLG2-AS1 expression in the 65 LUAD patients. Shown in a darker color are the patients who presented DLG2-AS1 downregulation (FC < 0.66). *** = p-value < 0.001.*

3.2.3 Phenotypical assays in DLG2-AS1 restoration cell models

To further explore the tumor suppressor role hypothesis, we tried to perform some functional studies on the *DLG2-AS1* gene to unravel whether it has any cancer-related functions. We first generated a vector construction containing the *DLG2-AS1* gene sequence (pLVX-DLG2AS1-IRES-zsGreen1), and we used the same plasmid without the *DLG2-AS1* sequence as an empty vector (EV) control (pLVX-IRES-zsGreen, Clonetech #632187). The pLVX-DLG2AS1-IRES-zsGreen1 lentiviral vector contains the sequence of the *DLG2-AS1* gene under the control of the cytomegalovirus (CMV) strong promoter, followed by an internal ribosome entry site (IRES) that links the expression of *DLG2-AS1* to the expression of a green fluorescent protein (zsGreen1) used as a reporter gene for selection of successfully infected cells.

We first delivered both vectors to the A549 LUAD cell line by transient transfection, and then checked by RT-qPCR that the overexpression of *DLG2-AS1* was successful. Although *DLG2-AS1* overexpression was evident, the *DLG2-AS1* expression levels decayed over time in 72 hours (Figure 12A). Thus, we considered that a lentiviral, integrative delivery was perhaps more adequate to conduct long-term experiments such as a colony assay. Therefore, we generated LVs, and we infected three LUAD cell lines (A549, H1944, and H23) to obtain three *DLG2-AS1* stable-over-time overexpression models (Figure 12B).

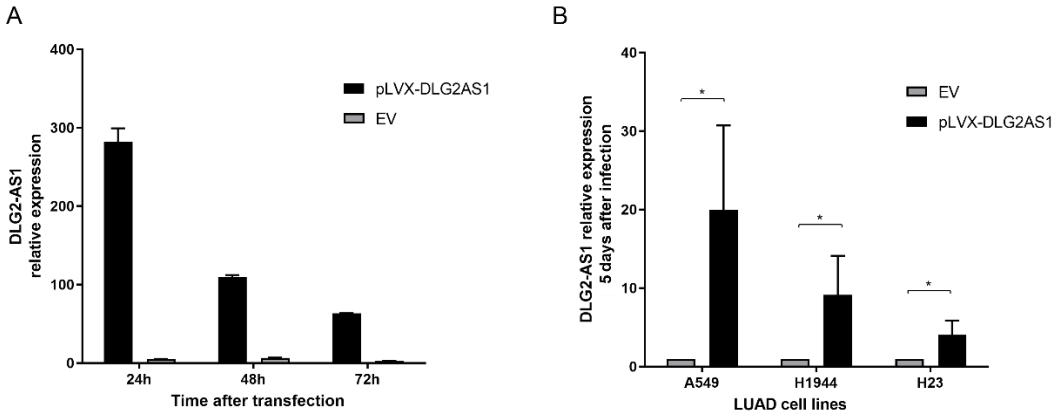


*Figure 12. DLG2-AS1 overexpression measurement by RT-qPCR in a transient model (A), and an integrative, stable-over-time one (B). Error bars in (A) represent technical replicates and not biological ones, therefore no statistical test was performed. * = p-value < 0.05*

Once the models were obtained, we used them to perform resazurin, cell colony, and scratch wound-healing assays to test if *DLG2-AS1* restoration was affecting cell viability, clonogenicity, or migration, respectively. Unfortunately, according to our results in these restoration models, no statistically significant phenotypical changes were observed for any of these cancer hallmarks, as there were no differences between EV- and pLVX-DLG2AS1-infected cells in terms of cell viability, clonogenicity, or migration (

Figure 13).

3.2.4 DLG2-AS1 does not act as a cis-regulator for DLG2

Next, due to previous evidence showing that some antisense lncRNAs may exert their function by up- or downregulating their overlapping protein-coding genes (Villegas & Zaphiropoulos, 2015), we studied whether there was any correlation between the expression of *DLG2-AS1* and the expression of its overlapping protein-coding gene, *DLG2*. First, we studied the expression levels of *DLG2-AS1* and *DLG2* in a set of 20 matched tumor-normal patients, finding

no significant correlation (Pearson correlation r = 0.378, p = 0.1, n = 20) (Figure 14A). To corroborate this result, we also analyzed an external dataset of 12 tumor-normal paired samples from LUAD available at the TCGA data portal (https://portal.gdc.cancer.gov). Analysis of this external dataset further confirmed the lack of correlation between *DLG2-AS1* and *DLG2* in patient samples (Pearson correlation r = 0.16, p = 0.62, n = 12) (Figure 14B).

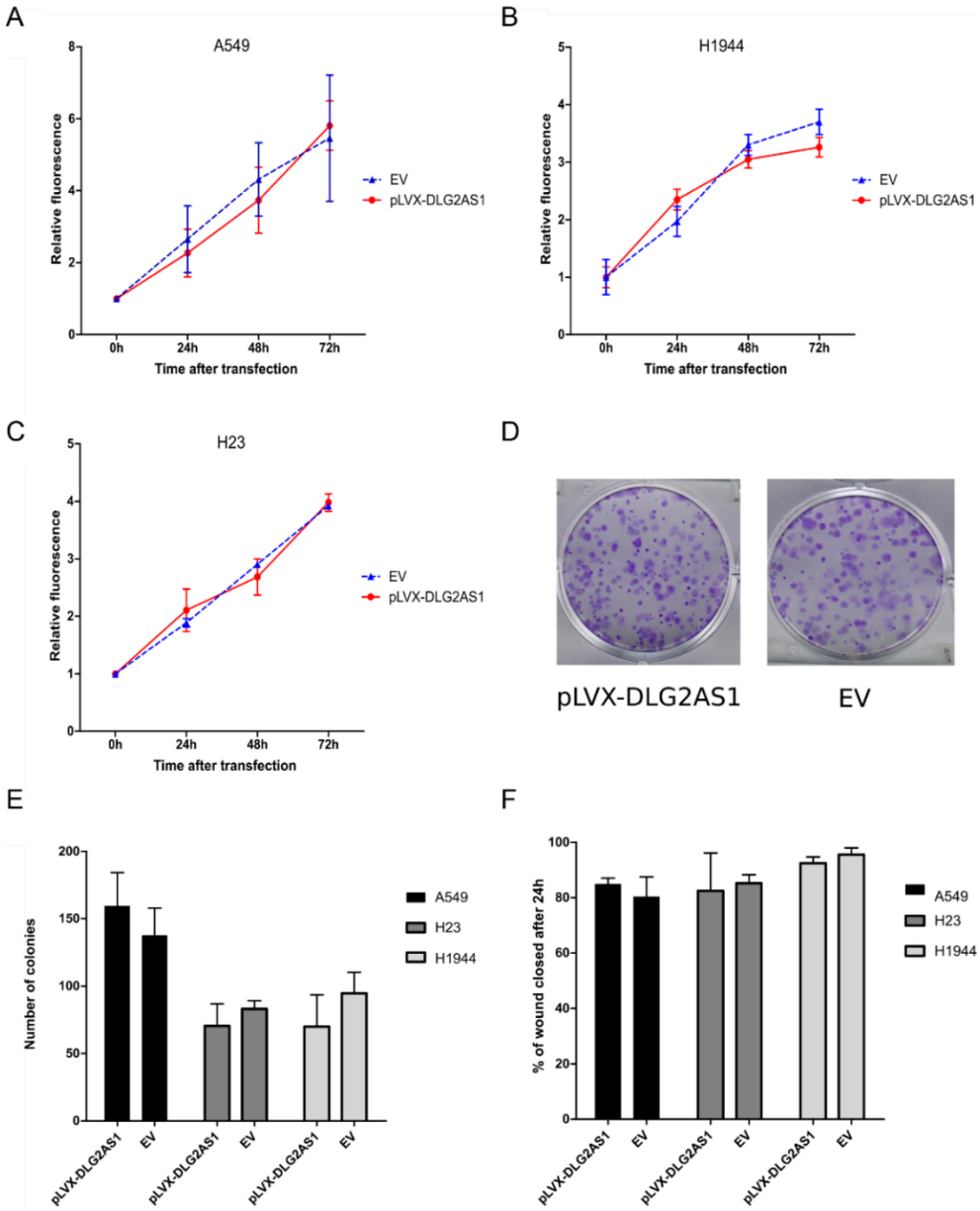


Figure 13. Functional assays performed in three LUAD cell lines (A549, H1944, and H23), comparing DLG2-AS1 restoration (pLVX-DLG2AS1) versus empty vector control (EV). We observed no significant differences between pLVX-DLG2AS1 and EV cells when measuring cell viability by resazurin assay (A, B, C), cell clonogenicity by a colony assay (D, E), and cell migration by a scratch-wound healing assay (F).

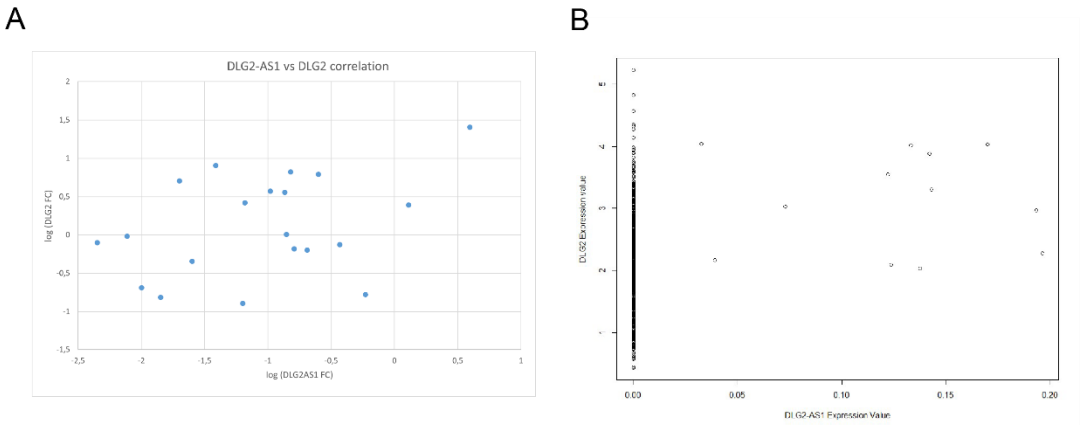


Figure 14. Correlation plots of DLG2-AS1 (X-axis) and DLG2 (Y-axis) in (A) a subset of our 65 LUAD patient samples (n=20), and (B) the LUAD patients from TCGA portal which presented some detectable DLG2-AS1 expression (n=12).

Furthermore, we also measured *DLG2* expression by RT-qPCR in the *DLG2-AS1* restoration models we had previously generated, and we observed that the mRNA levels of *DLG2* were unaffected by *DLG2-AS1* overexpression (Figure 15A). Finally, we also considered that some lncRNAs may regulate gene expression at a translational level by interacting with the ribosome (Zucchelli et al., 2015). Such lncRNAs might not affect the mRNA expression levels of the genes they are regulating, but directly their protein expression levels. To fully discard this option, we measured the expression of the DLG2 protein by Western blot after introducing a plasmid containing *DLG2-AS1* (pLVX-DLG2AS1) or the empty vector (EV). However, we observed no significant differences in DLG2 protein levels between pLVX-DLG2AS1 and EV-transfected cells (Figure 15B). Therefore, we ruled out *DLG2-AS1* mRNA expression as a regulator of DLG2 expression in LUAD patient samples.

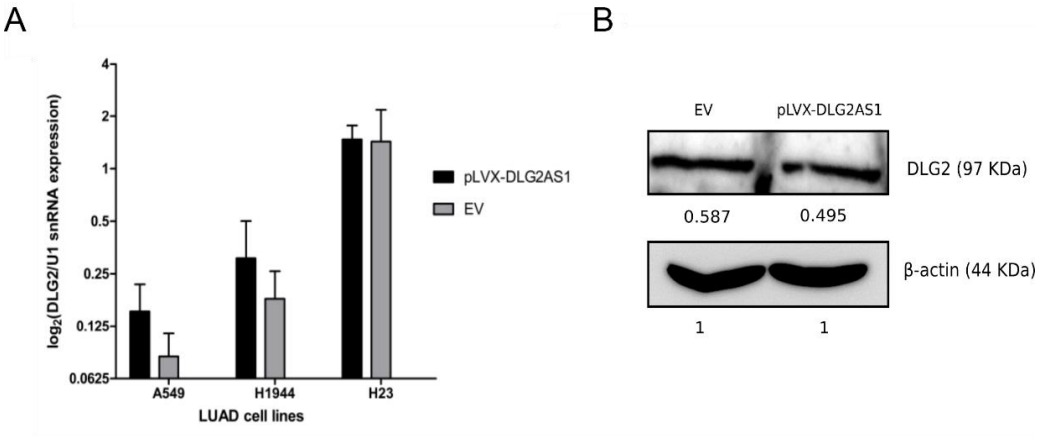


Figure 15. DLG2 expression in the DLG2-AS1 overexpression models of LUAD cell lines (A549, H1944, H23), both at RNA levels by RT-qPCR (A), and protein levels by Western blot (B).

3.2.5 DLG2-AS1 shows a good potential as a biomarker for LUAD

Lastly, we assessed the usefulness of *DLG2-AS1* expression as a biomarker for the classification of samples as tumor or normal. Using the data from our patient cohort (n = 65 tumor-normal pairs), a receiver operating characteristic (ROC) curve was generated. For each possible threshold value of *DLG2-AS1* expression, samples were classified as tumor or normal based on whether *DLG2-AS1* expression was below or above the threshold. Sensitivity and specificity of the classification were assessed for each possible threshold value, and they were plotted yielding the ROC curve. Then, the area under curve (AUC) was calculated as a measure of the idoneity of *DLG2-AS1* expression to distinguish between tumor and normal patients. The AUC of *DLG2-AS1* was 0.726 (95% CI = 0.638–0.815) showing a threshold of −0.916 for $\log_2$ (Delta-Cq), an optimal specificity of 80%, and a sensitivity of 60% (Figure 16A). To compare its specificity and sensitivity with other well-known cancer biomarkers (C. H. Wu & Hwang, 2019) and cancer-related lncRNAs (T. Lu et al., 2018), we calculated the ROC curves of the *EGFR* and *TP53* genes, as well as the oncogenic lncRNAs *MALAT-1* and *NEAT1*, using TCGA-LUAD data. We obtained AUC values of 0.539, 0.703, 0.542, and 0.573, respectively, which are considerably lower than the AUC values obtained for *DLG2-AS1* (Figure 16B).

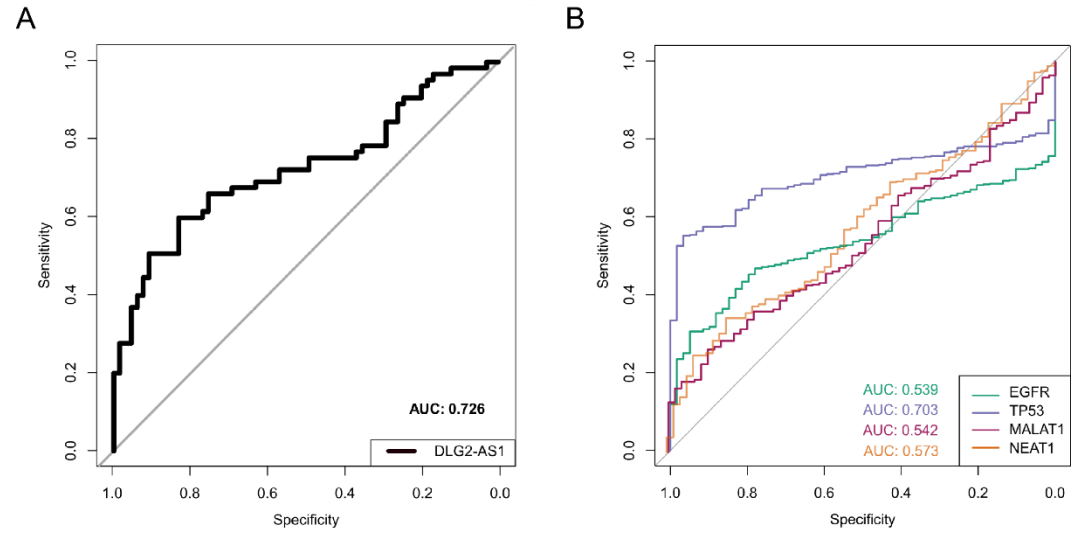


Figure 16. (A) Receiver operating characteristic (ROC) curve and area under curve (AUC) value of DLG2-AS1 expression in our cohort of LUAD patients. (B) ROC curves and AUC values of other LUAD biomarkers (EGFR, TP53, MALAT-1, and NEAT1), obtained from TCGA data.

Furthermore, to put in context the value of *DLG2-AS1* as a LUAD lncRNA biomarker, we downloaded gene expression data from TCGA and classified the patients according to their *MALAT-1* or *NEAT1* expression, two of the most studied lncRNAs in lung cancer, which are validated as LUAD biomarkers (Ji et

al., 2003; Yu et al., 2017). We calculated the fold change (FC) between the tumor and the non-tumor sample of 57 patients where both of those data were available in TCGA (a similar cohort size to our 65 LUAD paired samples), and we classified the patients based on whether they presented *MALAT-1* or *NEAT1* upregulation (FC >1.5), no change in expression (0.66 < FC < 1.5), or downregulation (FC < 0.66), in a similar manner as we did for *DLG2-AS1* (**Table 2**).

Table 2. Expression status of DLG2-AS1, MALAT-1, and NEAT1 in our patient cohort (n=65, DLG2-AS1) or TCGA (n=57, MALAT-1 and NEAT1).

Gene	Expression status	% of patients
DLG2-AS1	Upregulated	4.6
	No change	27.7
	Downregulated	67.7
MALAT-1	Upregulated	21.1
	No change	42.1
	Downregulated	36.8
NEAT1	Upregulated	28.1
	No change	24.5
	Downregulated	47.4

Patients were classified according to their tumor/non-tumor fold change (FC) expression of the genes: upregulated = FC>1.5; no change = 0.66<FC<1.5; downregulated = FC<0.66.

Taking all this into consideration, *DLG2-AS1* proved to be a better biomarker than the lncRNAs *MALAT-1* and *NEAT1* in cohorts of similar size. Together with the rest of results presented in this section, *DLG2-AS1* is proposed as a novel lncRNA tumor biomarker for LUAD (Arenas et al., 2020).

3.3 Discussion

There is a growing body of evidence suggesting that the dysregulation of lncRNAs is associated with several human diseases, including lung cancer, where lncRNAs play a significant role in diagnosis, prognosis, and treatment (T. Lu et al., 2018). In fact, several studies have proven the value of lncRNAs as cancer biomarkers in the clinic. *PCA3* is used as a biomarker for prostate cancer, complementing the PA test in patients under the suspicion of requiring a biopsy (G. L. Lee et al., 2011), and serving also as a prognosis biomarker, since its expression levels correlate with tumor aggressiveness (Bolha et al., 2017). For NSCLC, current biomarkers used in the clinic are mostly tumor-secreted proteins, such as CEA, NSE, CA125, or CYFRA21-1. Despite their high sensitivity, these protein biomarkers display overall a low specificity (P. Qi et al., 2016). In comparison, the expression signature of three lncRNAs, *SPRY4-IT1*, *ANRIL*, and *NEAT1*, showed an AUC of 0.876, with an 82.8% sensitivity and a 92.3% specificity. In addition, the circulating levels of these lncRNAs correlated well with tumor size, with higher levels indicating a higher tumor burden (Hu et al., 2016). Therefore, the identification and research of cancer-associated lncRNAs is critical for understanding the roles of lncRNAs in the carcinogenesis and improving the current clinic.

For the first time, in here we propose the lncRNA *DLG2-AS1* as a potential biomarker for LUAD (Arenas et al., 2020). *DLG2-AS1*, also known as *AP001825.2* (NCBI Gene ID: 100302690), is an antisense lncRNA within the first intron of the overlapping protein-coding gene *DLG2* (Disks Large Homologue 2). *DLG2-AS1* and *DLG2* are mainly expressed in brain adult tissues, as DLG2 is a MAGUK family protein for which a role in NMDA-receptor assembly has been proposed (E. Kim et al., 1996). DLG2 is a homologous protein to Drosophila's dlg-A, which is considered a tumor suppressor protein (Mazoyer et al., 1995). In humans, a DLG2 isoform was overexpressed in renal oncocitoma (Zubakov et al., 2006). Regarding *DLG2-AS1* expression, one study showed a downregulation in brain tissues from patients with schizophrenia (Polesskaya et al., 2003). However, before our study, no information of *DLG2-AS1/DLG2* in lung cancer was published, which makes our results on *DLG2-AS1* downregulation in LUAD patients a novel discovery (Arenas et al., 2020).

Gene expression regulation roles of lncRNAs depend on their molecular way of action, including interfering with transcription, serving as ceRNAs, mRNA maturation, mRNA stability or translation, among others (Peinado et al., 2018). LncRNAs may act regulating in *cis* the expression of neighboring genes, or in *trans* modulating distant gene expression. Antisense lncRNAs like *DLG2-AS1* are transcribed from the opposite strand of other genes, and they usually regulate in *cis* the expression of their overlapping protein-coding genes (Villegas & Zaphiropoulos, 2015). However, according to our results, there is no correlation on the expression of *DLG2-AS1* and *DLG2,* neither in

patient samples nor in our *DLG2-AS1* restoration cell models. In external data from TCGA (n=582 patients), *DLG2-AS1* expression was very low and only detectable in 12 patients. Therefore, we could not perform a correlation analysis for *DLG2* and *DLG2-AS1* in a larger cohort. Although our results do not support the modulating role in *cis* of *DLG2-AS1* over *DLG2* expression, other regulating roles in *trans* could be discovered in future mechanistic studies, as was done with other *trans*-acting lncRNAs like *HOTAIR*, *NEAT1*, or *Braveheart* (Zampetaki et al., 2018). The methodology for these mechanistical studies, however, is complex. For example, *HOTAIR* forms an intricate secondary structure that was determined by chemical probing (Selective 2'-Hydroxyl analyzed by Primer Extension, or SHAPE), and thus, some binding domains to PRC2 were discovered (Somarowthu et al., 2015). *NEAT1*'s function as a structural scaffold for nuclear paraspeckles was also discovered by SHAPE probing, combined with complex computational analyses (Y. Lin et al., 2018). Finally, *Braveheart* is a lncRNA that promotes cardiovascular lineage commitment thanks to its binding to a zinc finger transcription factor, which was discovered after selective CRISPR deletions of G-rich motifs in its sequence, followed by immunoprecipitation of bound factors (Z. Xue et al., 2016).

One important limitation of our study is the use of ectopic overexpression models for the study of a lncRNA's function. As discussed in Chapter 1, sometimes lncRNAs exert their function by activating or silencing in *cis* neighboring genes upon their own transcription, independently from the lncRNA transcript itself (Engreitz et al., 2016; Paralkar et al., 2016). If this were the case for *DLG2-AS1*, neither an ectopic plasmid overexpression, or a lentiviral integration model of *DLG2-AS1*'s RNA molecule would be helpful to unravel their functions and phenotypical effects. This could explain the lack of phenotype, as well as the lack of correlation between *DLG2* and *DLG2-AS1* expression observed when overexpressing *DLG2-AS1* in LUAD cell lines. In general, lncRNAs possess unique biological properties and roles that make their functional study challenging (Morelli et al., 2021). For *trans*-acting lncRNAs, finding the interactions of a lncRNA and performing complex mechanistical studies can be daunting for non-specialized laboratories, as the previous examples of *HOTAIR*, *NEAT1*, and *Braveheart*'s characterizations show (Y. Lin et al., 2018; Somarowthu et al., 2015; Z. Xue et al., 2016). For *cis*-acting and overlapping lncRNAs within protein-coding genes, it is difficult to genetically manipulate their expression without affecting the neighboring genes. One of the technologies that has allowed the study of lncRNAs within their genomic context is the CRISPR genome editing technology. The CRISPR technology has been used to provide evidence of lncRNAs acting in *cis* unlinking their function from the RNA transcript, for example, by inserting premature polyadenylation sites, or substituting the lncRNA exons by a

reporter gene (Mattick et al., 2023). In addition, the CRISPRa technology is a great tool for gain-of-function studies of lncRNAs because of two main reasons: (i) it overexpresses a target gene by activating transcription directly at the genomic locus, therefore conserving all the *cis*-regulatory implications of lncRNA expression (Morelli et al., 2021), and (ii) it works better with lowly expressed genes, as it is usually the case for lncRNAs, allowing high overexpression levels within the physiological range.

Because none of our phenotypical assays (cell viability, clonogenicity, and migration) displayed a positive result, we could not demonstrate *DLG2-AS1*'s role as a tumor suppressor lncRNA in LUAD. However, in order to totally discard *DLG2-AS1* from being a *bona fide* tumor suppressor gene, it would be necessary to check if *DLG2-AS1* is involved in other cancer-related processes, such as angiogenesis, immune evasion, or metastasis (Hanahan, 2022). The further study of these complex cancer hallmarks would require animal models, which unfortunately were beyond the scope of our study.

Finally, we were interested in determining whether *DLG2-AS1* can serve as a good diagnostic biomarker for LUAD patients. The ROC curves that we conducted to analyze the diagnostic power of *DLG2-AS1* showed that *DLG2-AS1* has a relatively high diagnostic value for LUAD patients (AUC = 0.726) in comparison with other broadly studied and validated LUAD biomarkers (*EGFR, TP53*), as well as other lncRNAs with proven clinical value (*MALAT-1*, *NEAT1*). In different cohorts to ours, *MALAT-1* displayed AUCs of 0.79 and 0.703 for circulating and exosomal *MALAT-1*, respectively (T. Lu et al., 2018). Also, for comparison, the FDA-approved *PCA3* has an overall AUC of 0.75 (Bolha et al., 2017). All these results suggest a potential use of *DLG2-AS1* as a LUAD biomarker for the diagnosis of the disease. Nevertheless, further analyses, such as its detection by RT-qPCR in exosomes or by liquid biopsy, would be necessary to completely validate its clinical use as a LUAD biomarker, as has been done with other lncRNAs. This is in fact one limitation of the use of lncRNAs as biomarkers for diseases: despite their high tissue-specificity, they are usually very low-expressed, and sometimes RT-qPCR or microarrays are not sensitive enough for the detection of circulating lncRNAs in blood plasma (T. Shi et al., 2016). Because of this, it is proposed that some lncRNA markers might be better used in combination with other lncRNAs or traditional protein markers, as this typically yields a better diagnostic power, with higher sensitivity, specificity, and AUC values (T. Lu et al., 2018). This could open a new research line for *DLG2-AS1*, aiming to detect it in exosomes or as a circulating lncRNA in blood plasma, and testing its performance as a LUAD biomarker in combination with others.

Chapter 4.
Validation of the oncogenic potential of a mutated miRNA

Chapter 4. Validation of the oncogenic potential of a mutated miRNA

This chapter will address Objective 1.2. In here, we tried to validate the oncogenic driver potential of a miRNA, *miR-133b*, which was found mutated in LUAD patient samples, thus being another possible biomarker for early diagnosis or prognosis, or a target for therapy. In section 4.1 we provide a brief background on the previous work performed in our research group until the discovery of the mutation of *miR-133b* in LUAD patients. Then, we proceed with the novel results we obtained in section 4.2, and a discussion on the matter in section 4.3.

4.1 Background

4.1.1 Mutations in miRNA seed regions are rare, yet existing

Traditionally, most research searching for genomic variants in cancer has been focused on the protein-coding genome, while often overlooking non-coding regions (Tan, 2020). The reasons for this are, among others, the poor annotation of many non-coding genes, the lack of knowledge about the implications of mutations in non-coding regions, and the higher cost of whole-genome sequencing (WGS) compared to whole-exome sequencing (WES) (Tan, 2020). Therefore, our group aimed to address this knowledge gap by identifying, validating, and characterizing novel mutations in non-coding regions of the genome. To achieve this goal, we performed a targeted deep-sequencing aiming for non-coding RNA regions, including miRNAs, lncRNAs, intronic splice regions, proximal promoters, and 3'-UTRs. We analyzed a cohort of 70 LUAD primary tumor samples, 27 of them with their paired, non-tumor adjacent tissue samples; along with 37 LUAD-annotated cell lines, and WES and WGS data we downloaded from the TCGA-LUAD project (n=582 patients). Thus, we obtained a number of mutated candidates for further experimental validation.

As it was already discussed in Chapter 1, miRNAs are important players in gene expression regulation, and their expression is frequently altered in cancer. However, because of their short length and their high sequence conservation, mutations are very rare in miRNA genes, especially within the mature seed region that determines its specificity and the binding to the target mRNA (Bartel, 2004). For instance, in a pan-cancer analysis of more than 2,600 tumors, only one miRNA, *miR-142*, was found recurrently mutated at the seed region in B-cell non-Hodgkin lymphoma (Rheinbay et al., 2020), as also shown by a previous study (Kwanhian et al., 2012). Therefore, despite their rareness, mutations in the seed region of mature

miRNAs can be drivers of disease, as is the example of *miR-142* in lymphoma (Kwanhian et al., 2012), or some seed-mutated miRNAs in other diseases apart from cancer, such as *miR-96* in progressive hearing loss (Mencía et al., 2009), or *miR-204* in retinal dystrophy (Conte et al., 2015).

4.1.2 The seed region of miR-133b is mutated in LUAD.

From our variant calling in non-coding regions, we found that, indeed, mutations in the seed region of miRNAs were very rare. We discovered just five variants, two in cell lines and three in primary tumors, which affected highly conserved nucleotides within the seed region of a miRNA. The one that presented the highest conservation score, and that was predicted to have the most deleterious effect was a variant affecting *miR-133b* (chr6:52148992_G>T) (Andrades, 2022).

The mutation of *miR-133b* was detected in the tumor sample of one patient from our paired LUAD cohort (n=27), whereas the corresponding non-tumor sample harbored the wildtype version of *miR-133b*. This indicates that the mutation is not germline but somatic, suggesting a higher likelihood of a potential oncogenic role. An analysis of additional *miR-133b* somatic variants in data from the International Cancer Genome Consortium (ICGC) revealed two variants in the *miR-133b* seed region in chronic lymphocytic leukemia and colon adenocarcinoma samples, and 17 more variants across the *MIR133B* gene in other types of cancer, including one in LUAD, albeit not in the mature miRNA (Andrades, 2022).

As it will be more extensively reviewed in the discussion section from this Chapter (section 4.3), *miR-133b* is overall considered a tumor suppressor miRNA dysregulated in lung cancer and several other cancer subtypes. This, together with the variants that we and others found in its sequence, proposes mutant *miR-133b* as a biomarker for some of these pathologies, or even as a potential therapeutic target (D. Li et al., 2017). As such, we proceeded with the validation of the oncogenic potential of this novel mutation of *miR-133b* in LUAD.

4.2 Results

4.2.1 Detection of miR-133b mutation at gDNA and RNA levels

First, we aimed to experimentally validate the *miR-133b* mutations found in our targeted sequencing. Because neither PCR nor qPCR are sensitive enough to detect a one-nucleotide change, we decided to instead perform Sanger sequencing of a specific PCR product from the *MIR133B* locus (Figure 17A). Thus, after designing primers for the genomic DNA sequence of *miR-133b* (primers miR133b_gen_Fw+Rv in **Supplementary Table 1**), we performed a PCR using the patient samples' gDNA, and extracted the amplified DNA for Sanger sequencing. The sequencing data we obtained revealed that, indeed, the *miR-133b* mutation (chr6:52,148,992; G>T) was present only in the tumor sample from the patient, whereas the non-tumor, paired sample presented the wildtype sequence of *miR-133b* (Figure 17B).

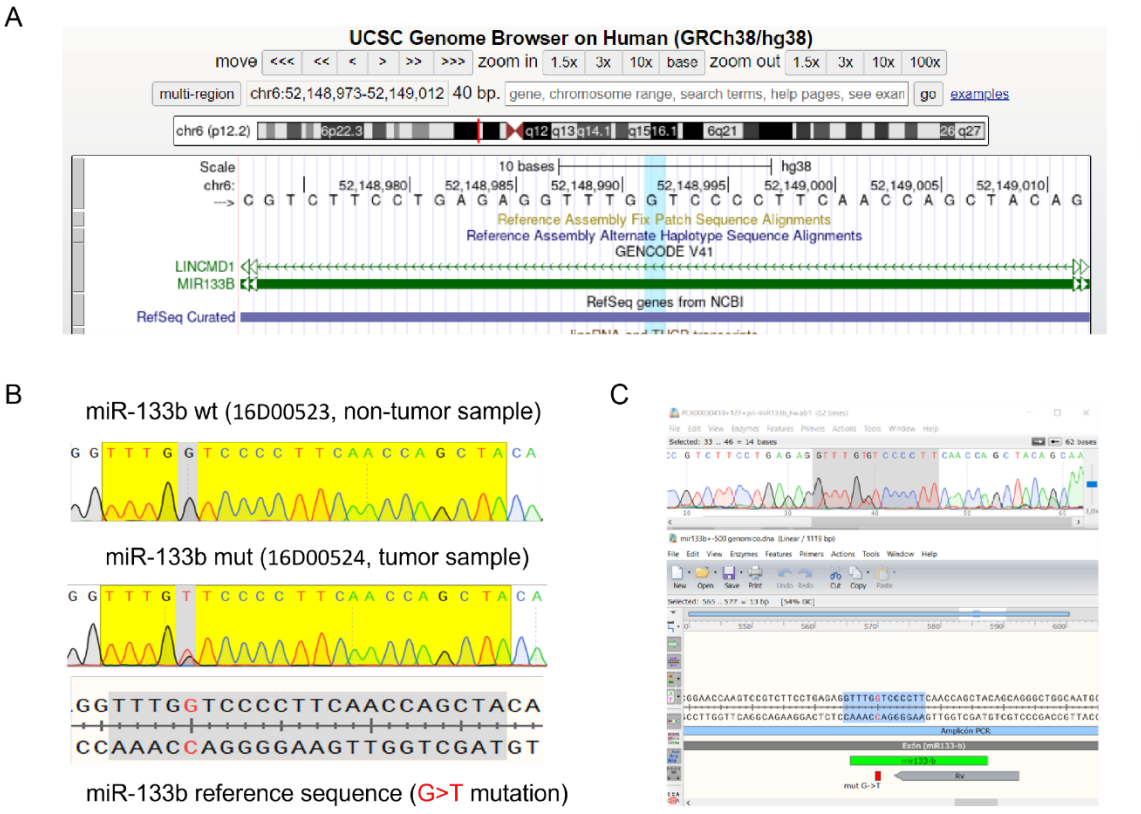


Figure 17. (A) Screenshot of the UCSC Genome Browser depicting the genomic coordinates of the miR-133b mutation (G>T). (B) Sanger sequencing of genomic DNA from the control non-tumor (16D00523) and tumor sample (16D00524) of the patient harboring the miR-133b mutation. (C) Sanger sequencing of the pri-miR-133b, showing the mutation at an RNA level.

However, just detecting the mutation at a genomic DNA level is not enough to ensure that the mutant version of *miR-133b* is being expressed at an RNA level.

Therefore, we next tried to Sanger sequence the primary RNA transcript of *miR-133b* by designing primers that amplify a wider region (±500 bp) up- and downstream from the mature miRNA sequence, which we considered to be the pri-*miR-133b* (primers pri-miR133b-Fw+Rv in **Supplementary Table 1**). Also, instead of using gDNA, we performed the PCR using cDNA obtained after reverse-transcribing the RNA from the patient samples. Again, the sequencing data showed us that the mutation is present only in the tumor sample, and not in the non-tumor one, which proved to us that the *miR-133b* mutation is present at both gDNA and RNA levels (Figure 17C).

4.2.2 Mutant miR-133b has an oncogenic role in LUAD overexpression cell models

Once the mutation was validated, we proceeded to analyze its oncogenic potential by overexpressing both the wildtype (*miR-133b*^{WT}) and mutant (*miR-133b*^{MUT}) versions of *miR-133b* in LUAD cell lines, in order to study phenotypical differences between the two of them. We designed *miR-133b*$^{WT/MUT}$ mimics (Merck, sequences shown in **Supplementary Table 2**) and we delivered them to three LUAD cell lines: H2126, A549, and H1650, which presented high, average, and low expression of *miR-133b* according to the CCLE database, respectively (Figure 18).

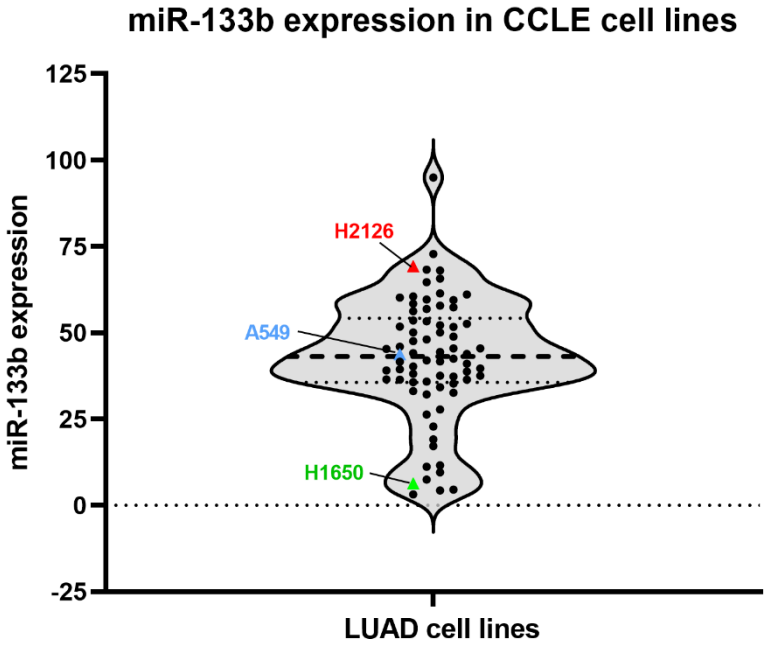


Figure 18. Expression of miR-133b in lung adenocarcinoma (LUAD) cell lines according to CCLE database. Selected cell lines are marked in color: H2126 (high expression), A549 (average), and H1650 (low).

We first checked that overexpression of *miR-133b*$^{WT/MUT}$ mimics was successful by measuring *miR-133b* RNA levels by RT-qPCR (Figure 19). Of note, because *miR-133b*^{WT} and *miR-133b*^{MUT} differ just in the first nucleotide, a simple RT-qPCR is not able to distinguish between the two of them, so both are measured indistinctly using the same primer.

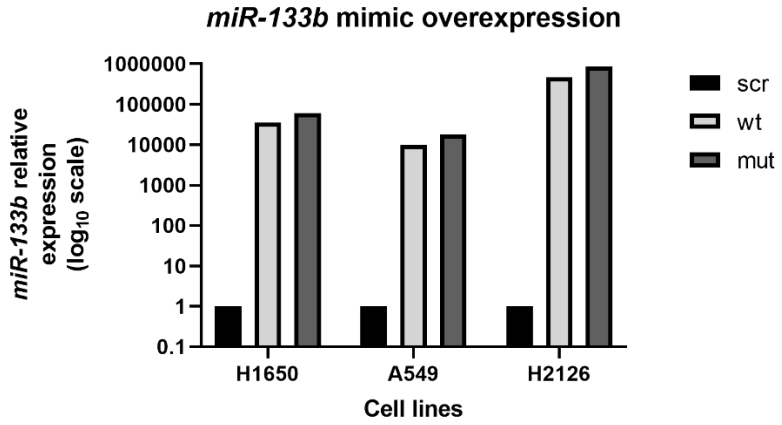


Figure 19. Measure of miR-133b overexpression by RT-qPCR after transfection with miR-133b$^{WT/MUT}$ *mimics. Because it was measured just once as a check for successful overexpression, no statistical analyses were performed.*

Then, we conducted resazurin and colony assays with the transfected cell lines to test any phenotypical changes in cell viability or clonogenicity. In terms of cell viability, we did not obtain any significant differences for the A549 and H2126 cell lines (Figure 20A-B), and only the H1650 cell line showed a significant reduction of cell viability when transfected with *miR-133b*^{WT} compared to *miR-133b*^{MUT} (Figure 20C).

As for the colony assays, in this case the H1650 cell line formed very small and irregular colonies, and it did not display any phenotypical difference (Figure 21A). However, we observed a clear tumorigenic effect of *miR-133b*^{MUT} in the H2126 cell line, as *miR-133b*^{WT} did not show any phenotypical differences compared to the scramble, whereas *miR-133b*^{MUT} greatly increased the number of colonies formed (Figure 21B). Finally, A549 transfected with *miR-133b*^{WT} showed a huge decrease in the number of colonies compared to the scramble, whereas *miR-133b*^{MUT} just brought down the number of colonies very slightly, as if the mutation was "impairing" the tumor suppressor effect of *miR-133b* (Figure 21C). All this showed, for the first time, an oncogenic effect of *miR-133b*^{MUT}.

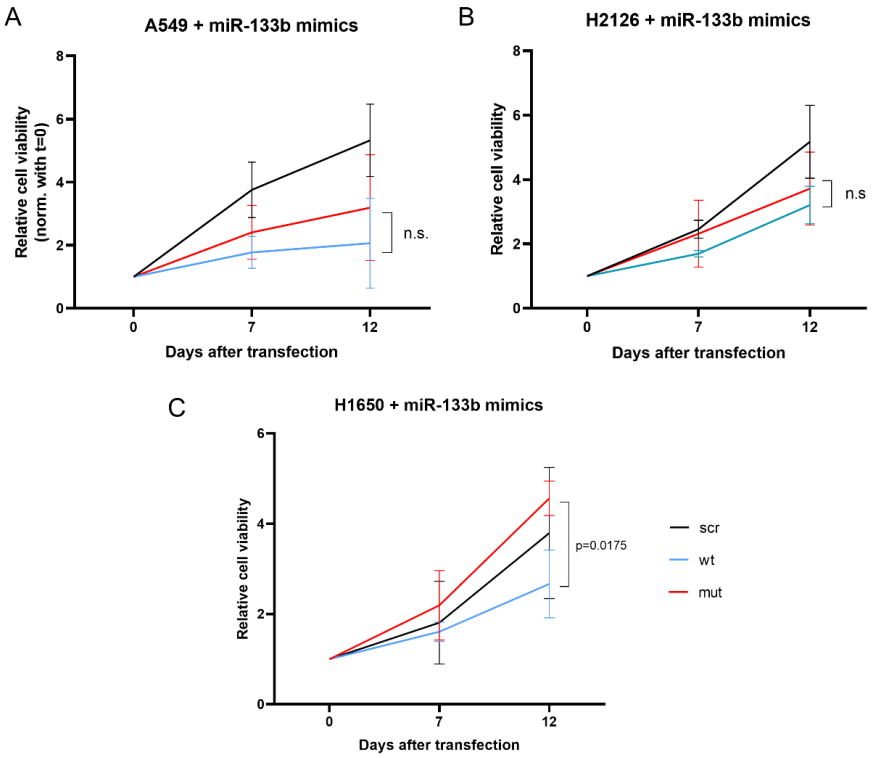


Figure 20. Cell viability assays of A549 (A), H2126 (B), and H1650 (C) cell lines transfected with miR-133b^{WT} mimic, miR-133b^{MUT} mimic, or the scramble control (scr).

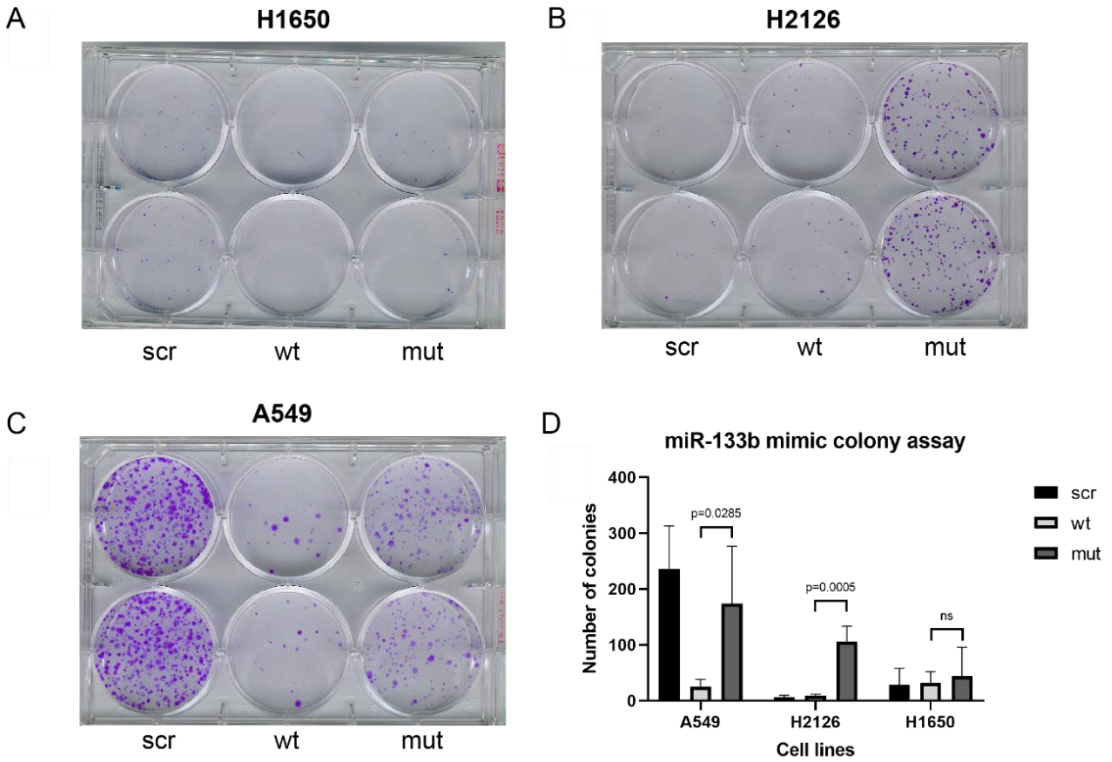


Figure 21. Representative colony assay pictures for A549 (A), H2126 (B), and H1650 (C). (D) Total count of colonies (n=4). T-test statistical significance was considered when p-value < 0.05.

4.2.3 Experimental validation of predicted miR-133b targets

Once we had demonstrated that the *miR-133b* mutation had an oncogenic role, we aimed to look for downstream pathways and targets of both *miR-133b*^{WT} and *miR-133b*^{MUT} that could explain the observed phenotype. We figured out two hypotheses, according to the results of our phenotypic assays: either (i) *miR-133b*^{MUT} targets a tumor suppressor gene, hence its oncogenic effect, which would fit with what we observed in the H2126 colony assays (Figure 21B); or (ii) *miR-133b*^{WT} targets an oncogene which *miR-133b*^{MUT} does not, therefore the oncogene downregulation is lost after the mutation, which could explain the results observed in the H1650 cell viability assay (Figure 20C) and the A549 colony assay (Figure 21C).

Because both previous research (L. Liu et al., 2012) and bioinformatic predictions from our lab (Andrades, 2022) had identified the oncogene *EGFR* as a *miR-133b*^{WT} target, we initially selected it as our first candidate target along with *DICER1*, a tumor suppressor gene that was a predicted target for *miR-133b*^{MUT}, but not *miR-133b*^{WT}. We initially measured the protein expression levels of EGFR and DICER1 by Western blot 48 hours after overexpressing *miR-133b*$^{WT/MUT}$ in LUAD cell lines, thus testing our two hypotheses. However, we could not see any consistent changes in any of the candidate genes in our cell models (Figure 22), so we proceeded to search for novel experimental targets that could explain the observed phenotype.

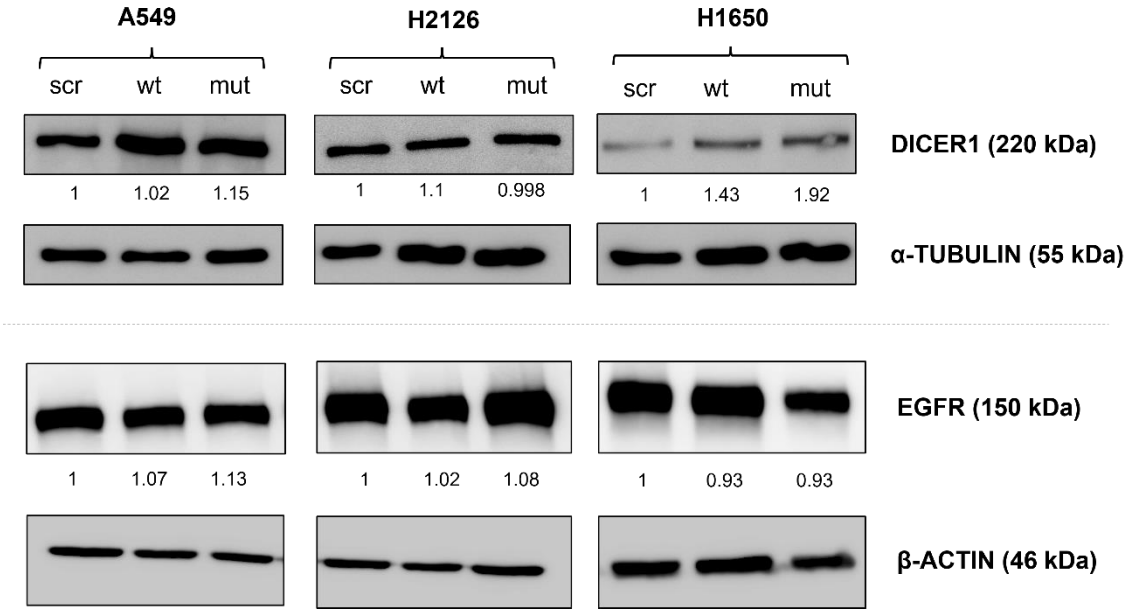


Figure 22. Representative images of Western blots of predicted miR-133b$^{WT/MUT}$ *targets. DICER1 and α-TUBULIN were run in 6% acrylamide SDS-PAGE, whereas EGFR and β-TUBULIN were run in 8% acrylamide SDS-PAGE. Densitometry measures were normalized by the respective reference gene (α-TUBULIN or β-ACTIN) and the scramble control (scr) condition.*

4.2.4 Discovery of novel experimental targets of miR-133b$^{WT/MUT}$

To identify new genes or pathways that are targeted by *miR-133b$^{WT/MUT}$*, we aimed to perform a biotin pull-down to detect direct miRNA-mRNA binding, and an RNAseq for the detection of both directly and indirectly altered downstream targets.

Before the pull-down, we first needed to test whether the biotin-labeled *miR-133b$^{WT/MUT}$* mimics that are necessary for the procedure are actually loaded onto the RISC. Thus, we performed a biotin pull-down (see section 2.1.8) to check if we were able to detect the AGO2 protein, a member of RISC, bound to our biotinylated miRNA mimics. We first tried to synthesize the biotinylated mimics ourselves, which presented the advantage of labeling several nucleotides within the mimic sequence, thus increasing pull-down efficiency afterwards. However, we failed to detect any RNA after the *in vitro* synthesis, possibly because of the short length of the miRNA mimic, which prevents the T7 RNA polymerase from properly binding to its promoter and start transcription. Therefore, to continue with the experiment, we ordered some customized commercial miRNA mimics (Integrated DNA Technologies - IDT, # 232691864/5) which were already labeled, albeit with a single biotin molecule at the 5'-end. Unfortunately, according to our pull-down results, these commercial mimics were either not being loaded onto RISC, or just one biotin label was not enough to yield a detectable pulled-down amount of miRNA mimic (Figure 23). Thus, we had to halt the pull-down experiment until this technical issue is solved.

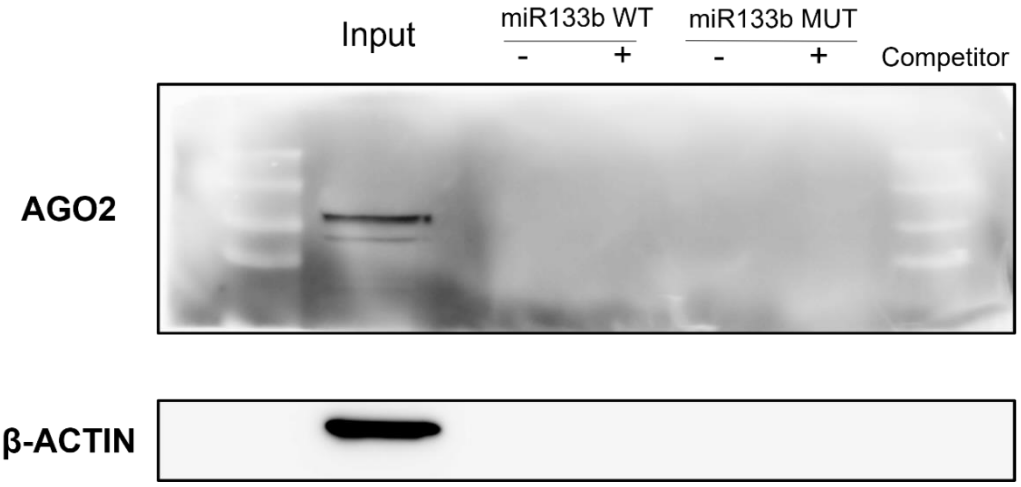


Figure 23. Biotin pull-down of biotinylated miR-133b mimics. An AGO2 band would mean that the miRNA mimic is being loaded onto RISC. The corresponding unbiotinylated miRNA mimic acts as a competitor and can show whether the binding is specific.

Parallelly, in order to identify new direct or indirect pathways or genes altered after *miR-133b$^{WT/MUT}$* overexpression, we performed an RNAseq using the two cell lines where we observed a stronger phenotypic effect of *miR-133b$^{WT/MUT}$*

overexpression (A549 and H2126). The RNA sample preparation, sequencing and data analysis were performed as previously described in section 2.4.2. The initial QC analysis revealed that the sequencing data were of great quality. The number of aligned reads was within the range of ~50-70 million reads, and both the principal component (PC) analysis and sample-to-sample distance plot showed the three replicates of each condition clustered very closely and distinctly from other conditions (Figure 24).

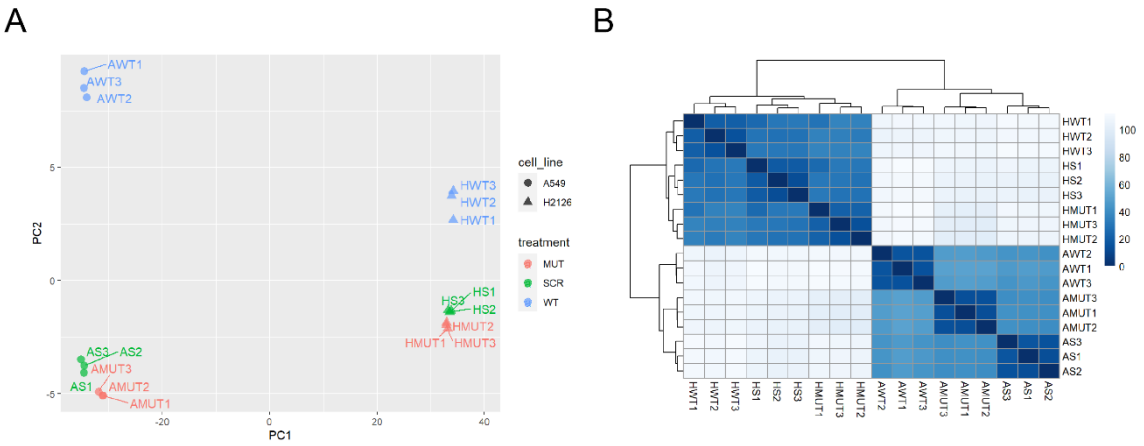


Figure 24. Quality control analysis of RNA-sequencing data. (A) Principal component (PC) analysis across samples. (B) Sample-to-sample distance plot. Sample labels: A/H (A549/H2126), S/WT/MUT (scramble/wildtype/mutant), 1/2/3 (# of replicate).

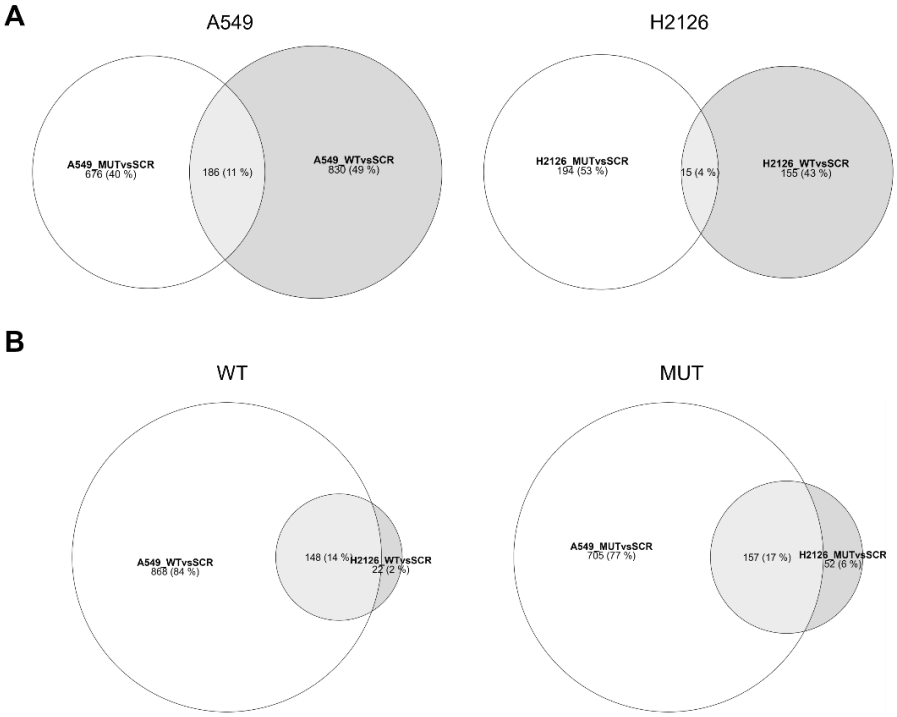


Figure 25. Euler diagrams of commonly downregulated genes between WT and MUT (A), and between the two cell lines (B).

We obtained a large number of downregulated genes in the A549 and H2126 cell lines after overexpressing *miR-133b*^{WT} or *miR-133b*^{MUT}, compared to the scramble. Interestingly, we detected very few downregulated genes in common between the WT and MUT conditions in both cell lines, but a great overlap of detected targets between the two cell lines (Figure 25). Furthermore, as depicted in the volcano plots, there were more downregulated genes than upregulated genes, which we expected given the overall gene expression repressive function of miRNAs (Figure 26).

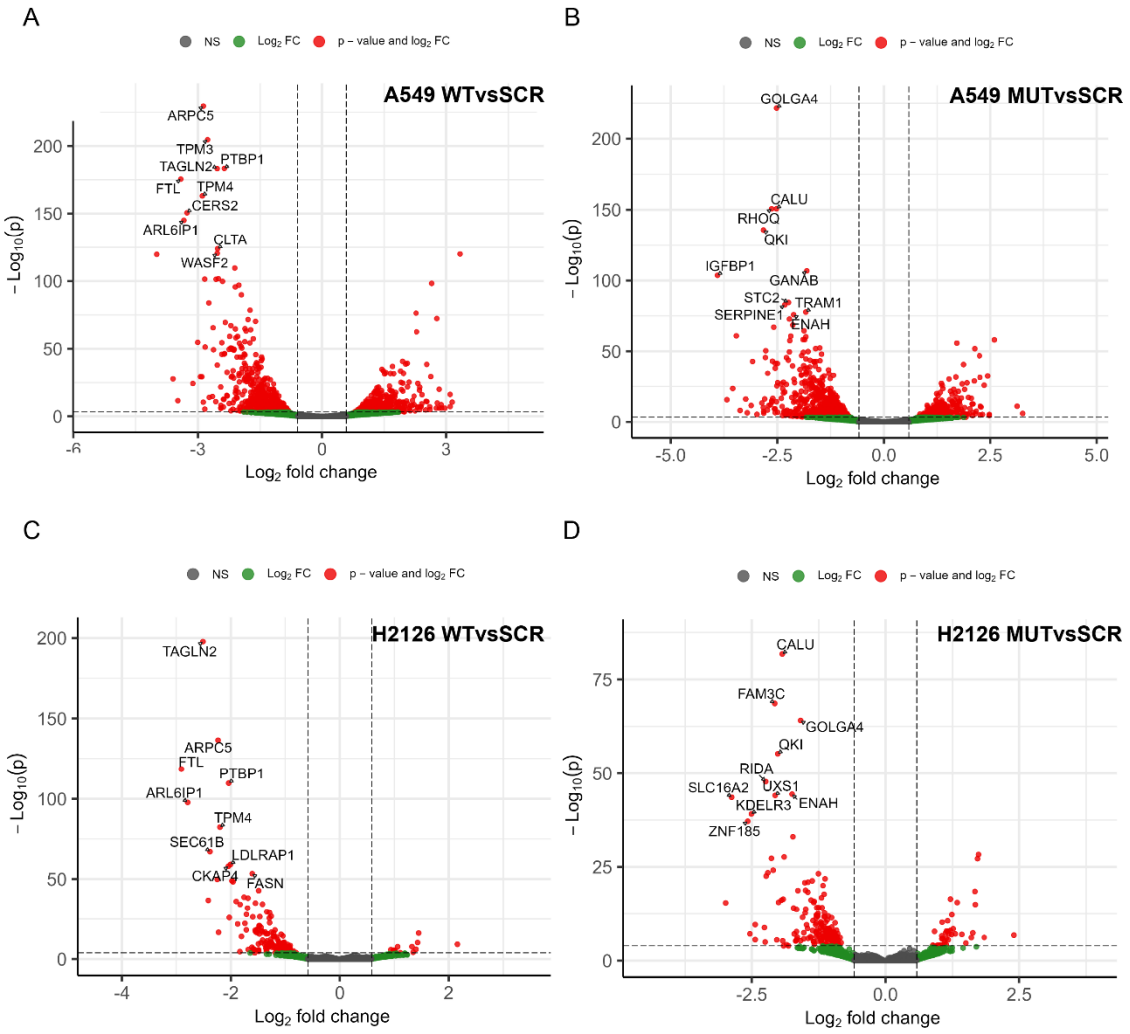


Figure 26. Volcano plots of RNA-sequencing data. The top 10 downregulated genes, ranked by p-value, are shown for every condition: A549 WTvsSCR (A), A549 MUTvsSCR (B), H2126 WTvsSCR (C), H2126 MUTvsSCR (D).

Moreover, we performed a gene set enrichment analysis (GSEA) for both cell lines. The top common downregulated pathways for *miR-133b*^{WT} were related

to cytoskeleton, focal adhesion, and secretion, whereas *miR-133b*^{MUT} pathways were mostly related to vesicle trafficking (Figure 27).

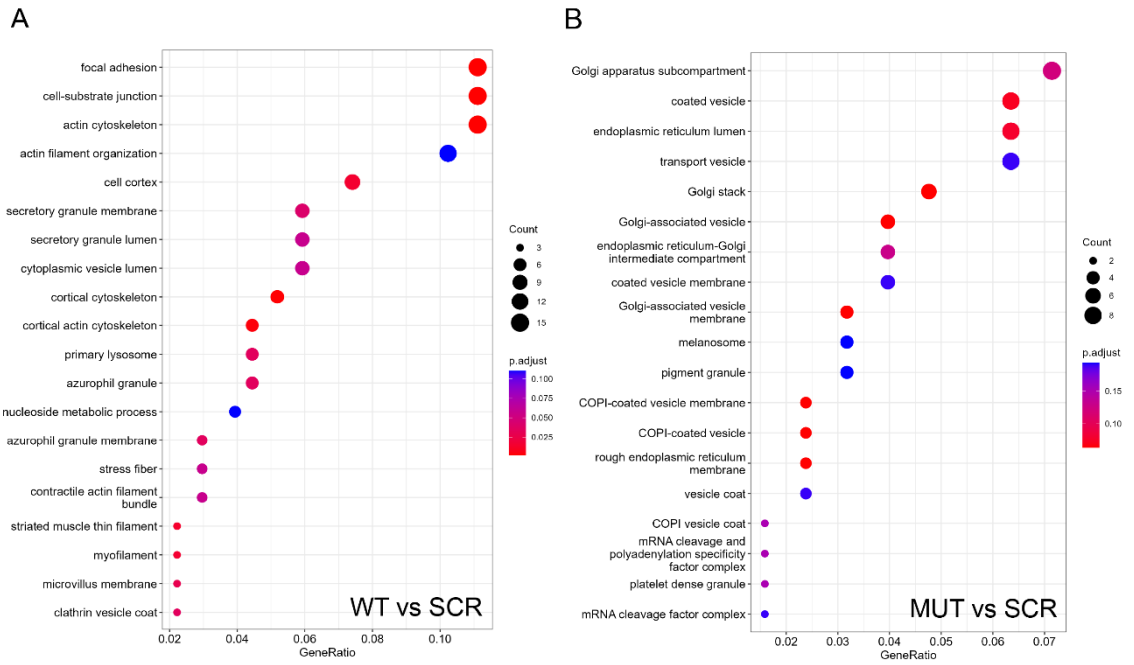


Figure 27. Gene set enrichment analysis (GSEA) of the common top downregulated pathways in both A549 and H2126 cell lines.

Finally, to choose our final candidates for future validation, we selected the top 10 most-downregulated genes, ordered by *p*-value ranked product of the two cell lines, and crossed those data with target predictions of the miRDB and TargetScan databases and literature information on experimental validations and oncogenic/tumor suppressor roles of the candidates for both *miR-133b*^{WT} (**Table 3**) and *miR-133b*^{MUT} (**Table 4**).

Table 3. Top 10 common miR-133b^{WT} target candidates from RNA-sequencing

Rank	Gene	Oncogene?	Predicted *miR-133b*^{WT} target but not *miR-133b*^{MUT}?	Validated *miR-133b*^{WT} target according to literature?	References
1	*ARPC5*	Yes	Yes	No	(Huang et al., 2021)
2	*TAGLN2*	Yes	Yes	Yes	(Z. Li et al., 2021; Y. Tang et al., 2019; F. Zhao et al., 2019)

3	*PTBP1*	Yes	Yes	Yes	(Z. Li et al., 2021; Sugito et al., 2017; Sugiyama et al., 2016)
4	*FTL*	?	Yes	Yes*	(Z. Li et al., 2020)
5	*TPM4*	Yes	No	Yes*	(Caporali et al., 2021)
6	*TPM3*	Yes	No	No	(S. Chen, Shen, et al., 2021)
7	*ARL6IP1*	Yes	No	No	(Guo et al., 2010)
8	*CERS2*	?	No	No	(Aldoghachi et al., 2019; Fan et al., 2015)
9	*CKAP4*	Yes	Yes	No	(Kimura et al., 2016)
10	*CLTA*	?	Yes	No	(Tsygankova & Keen, 2019)

**Only in one study. A "?" mark denotes conflicting reports on the exact tumorigenic role of the gene.*

Table 4. Top 10 common miR-133b^{MUT} target candidates from RNA-sequencing

Rank	Gene	Tumor suppressor gene?	Predicted *miR-133b^{MUT}* target but not *miR-133b^{WT}*?	References
1	*CALU*	No, oncogene	No	(Y. Yang et al., 2021)
2	*GOLGA4*	No	No	-
3	*QKI*	Yes	Yes*	(Cao et al., 2021)
4	*FAM3C*	No, oncogene	No	(M. Shi et al., 2018)
5	*RHOQ*	Yes	No	(Satoh et al., 2022)
6	*ENAH*	No, oncogene	Yes*	(D. D. Wang et al., 2017)
7	*GANAB*	Yes*	No	(Chiu et al., 2011)
8	*RIDA*	No	No	-
9	*UXS1*	No	No	-
10	*PITPNB*	No	No	-

**Only in one study / predicted by just one database.*

4.3 Discussion

According to the miRbase database (https://www.mirbase.org/), the *miR-133* family of miRNAs comprises *miR-133a-1* (chr18:21825698-21825785; accession #MI0000450); *miR-133a-2* (chr20:62564912-62565013, accession #MI0000451), which shares the same mature sequence than *miR-133a-1*, but is found in a different chromosome; and *miR-133b* (chr6:52148923-52149041; accession #MI0000822), which differs in just one nucleotide from the others. Initially, it was thought that these miRNAs were muscle-specific and related to skeletal muscle development (J. F. Chen et al., 2006). However, neither the deletion of *miR-133b*, nor the overexpression of *miR-133a* had an effect in terms of muscle regeneration, muscle function, or molecular changes in mice, possibly because of a compensation effect by the other *miR-133* member, or the action of the clustered *miR-1/miR-206* pair (Boettger et al., 2014; Deng et al., 2011).

The first association of *miR-133b* with cancer was done by Bandrés et al. (2006), who found that *miR-133b* was one of the most downregulated miRNAs in both colorectal cancer cell lines and patient samples (Bandrés et al., 2006). A later lung cancer study showed that *miR-133b* was the most downregulated miRNA in 10 lung tumor samples compared to adjacent, non-tumor tissue. They also demonstrated that *miR-133b* targets two antiapoptotic proteins, BCL2L2 and MCL-1, and that the overexpression of *miR-133b* sensitized NSCLC cell lines to an increased apoptosis response after treatment with gemcitabine (Crawford et al., 2009). This suggested, for the first time, a tumor suppressor role of *miR-133b* in lung cancer. Several later studies pointed in the same direction: *miR-133b* has a tumor suppressor role in lung cancer by interacting with antiapoptotic and pro-migration/invasion proteins such as the EGFR pathway or FSCN1 (Pan et al., 2017). Furthermore, *miR-133b* has been extensively studied in several other subtypes of cancer, where it also has been found mostly downregulated, and it appears to have a tumor suppressor role via interaction with different pro-tumorigenic mechanisms, including apoptosis inhibition, sustained cell proliferation, the Warburg effect, angiogenesis, migration, and invasion (D. Li et al., 2017).

Interestingly, one of the first studies of *miR-133b* in NSCLC was the one conducted by Liu et al. (2012), who showed that *EGFR* is targeted by *miR-133b*, enhancing apoptosis and drug response, and inhibiting cell invasion (L. Liu et al., 2012). Although previous work from our group also found *EGFR* as a predicted target of *miR-133b*^{WT}, but not *miR-133b*^{MUT} (Andrades, 2022), we failed to replicate these results in terms of EGFR protein downregulation after transfection with *miR-133b* mimics, nor was *EGFR* detected as a hit in our RNAseq analysis.

We demonstrated with our phenotypical assays that *miR-133b*^{MUT} has oncogenic potential in terms of cell clonogenicity in at least two LUAD cell lines.

We observed two distinct phenotypes which could be explained by different mechanisms of action of *miR-133b*^{MUT}: either by directly downregulating a tumor suppressor gene, or by interrupting the tumor suppressor activity of *miR-133b*^{WT}. Therefore, we believe that the genetic context of the tumor cell is crucial to predict the impact of *miR-133b* mutations, which could also explain why we were not able to detect EGFR downregulation in our miRNA mimic overexpression models. For future studies, it would be interesting to explore the effect of *miR-133b*^{MUT} in other cancer hallmarks, such as apoptosis inhibition, migration, or invasiveness, and also to assess whether this oncogenic potential is replicated in *in vivo* models.

To understand the different phenotypical effects that we observed depending on the cell line, it would be necessary to identify which targets of *miR-133b*$^{WT/MUT}$ are responsible for such phenotypes. There are several possibilities when trying to experimentally find the targets of a concrete miRNA. The most used, high-throughput methods can be roughly divided in (i) methods that detect expression changes after overexpression of the miRNA under study (such as RNAseq or proteomics), and (ii) methods which aim to detect direct binding between miRNAs and their target mRNAs (Akbari Moqadam et al., 2012). RNAseq has the advantage of being a widely used technique with a large number of service providers and standardized analysis tools, but it is not able to distinguish between direct and indirect effects of the miRNA overexpression. Therefore, a later validation would be necessary to demonstrate that the miRNA is effectively targeting the gene, for example, by a 3'-UTR luciferase assay (Akbari Moqadam et al., 2012). Additionally, some miRNAs do not induce the degradation of the target mRNA. Instead, they exert their function by directly interfering with ribosomal translation, resulting in alterations in the protein levels of the targeted mRNA (Orang et al., 2014). In those cases, an RNAseq could not detect such targets, as the mRNA expression levels would not be affected. Conversely, proteomics has the advantage of showing the final protein level changes resulting from miRNA inhibition, but it also detects indirect targets, it is not such a well-optimized technique as RNAseq, and the protein fold changes that a single miRNA can generate are often minimal, as some authors report (Baek et al., 2008). Finally, direct miRNA-mRNA interaction detection techniques such as biotin pull-down (Gerber et al., 2006), CLIP-seq (cross-linking, immunoprecipitation, and sequencing) (Imig et al., 2014), or CLASH (cross-linking, ligation, and sequencing of hybrids) (Helwak & Tollervey, 2014) often have the limitation of being more complex, expensive, and requiring more specific resources such as biotin-labeled miRNA mimics, radioactivity detection equipment, or transgenic cell line clones expressing a tagged AGO protein, respectively.

Our attempt to perform a biotin pull-down had to be halted because of technical difficulties when trying to synthesize the biotin-labeled miRNA mimics, and after checking that the already-biotinylated commercial miRNA mimics were not being loaded onto RISC. In the future, we will explore other alternatives for

direct detection of mRNA targets, such as a recently released miR-CLIP-seq commercial kit (e.g., miR-eCLIP® from EclipseBio), or a collaboration with laboratories specialized in CLASH.

On the other hand, the RNAseq results showed us interesting candidates for further studying the implications of *miR-133b* mutations in tumorigenesis. Of note, we first noticed the low number of common targets between *miR-133b*^{WT} and *miR-133b*^{MUT}, which agrees with both the bioinformatic predictions and our initial hypothesis that a mutation of a miRNA seed region can dramatically change its potential targets and downstream pathways. Also, contrarily to what we were expecting because of the different phenotypical effects observed in the A549 and the H2126 cell lines, we found a lot of overlapping targets between the two cell lines, meaning that probably the different phenotypes observed in the A549 and the H2126 are a result of different genetic backgrounds and dependencies on *miR-133b* and its downstream targets, instead of different pathways being activated in each model.

Among the *miR-133b*^{WT} top downregulated targets, we discovered transgelin-2 (*TAGLN2*) and polypyrimidine tract binding protein 1 (*PTBP1*). These two genes have been recurrently reported as targets of *miR-133b* in previous research, further reinforcing the reliability of our RNA sequencing results. For instance, both genes were found as targets of *miR-133b* in cardiomyocytes, where they are responsible for the induction of cardiac fibrosis and apoptosis after doxorubicin treatment (Z. Li et al., 2021). Interestingly, these two genes have a connection to muscle development. TAGLN2 is a canonical marker for smooth muscle (Tsuji-Tamura et al., 2021), while PTBP1 is a splice factor necessary for the generation of muscle-specific isoforms of pyruvate kinases (Sugiyama et al., 2016). This supports the initial findings that the miR-133 family, also known as myomiRs, plays a specific role in muscle development. Our GSEA data further reinforces this, as we discovered that the pathways most strongly affected by *miR-133b* regulation are related to the cytoskeleton machinery, a crucial component in muscle contraction. In addition to muscle-related functions, *miR-133b* and its targets also have implications in cancer development. For example, in both bladder cancer and esophageal squamous cell carcinoma, *miR-133b* exerts a tumor suppressor role by downregulating *TAGLN2* expression (Y. Tang et al., 2019; F. Zhao et al., 2019). Similarly, *PTBP1* can also act as an oncogene silenced by *miR-133b* in gastric cancer (Sugiyama et al., 2016) and rhabdomyosarcoma (Sugito et al., 2017).

As for *miR-133b*^{MUT} top candidate targets, the most relevant for our study were quaking RNA binding protein (*QKI*), and Ras homolog family member Q (*RHOQ*). There is growing evidence that *QKI* is a tumor suppressor in multiple types of cancer, including oral, prostate, lung, colon, and breast carcinomas (Cao et al., 2021; de Miguel et al., 2016; W. Lu et al., 2014; G. Yang et al., 2010; Y. Zhao et al., 2014). Therefore, its targeting by *miR-133b*^{MUT} could explain the oncogenic phenotype that we observed in our cell models. In a similar manner, it was found

that *RHOQ* functions as tumor suppressor gene in LUAD, as its knockdown led to an increase in EMT, and patients with high levels of *RHOQ* expression had better prognosis (Satoh et al., 2022).

To continue advancing on this project, it would be necessary to individually validate the selected candidates by analyzing expression changes after *miR-133b*$^{WT/MUT}$ overexpression, both by RT-qPCR or Western blot. Next, we should implement a luciferase assay testing whether the 3'-UTR of the selected genes is able to bring down the luciferase signal *in vitro*. Thus, we could demonstrate that the repression of the candidate genes is because of direct binding of *miR-133b*$^{WT/MUT}$ to the mRNA. Finally, the implication of these targets in *miR-133b*-derived tumorigenesis can be definitively proven by using phenotypical assays to check whether knocking down *TAGLN2/PTBP1* or *QKI/RHO* phenocopies the results observed from overexpressing *miR-133b*^{WT} or *miR-133b*^{MUT}, respectively.

Finally, regarding the biomarker potential of miR-133b, future research should aim to optimize its detection by non-invasive methods and with single-base resolution to differentiate between *miR-133b*^{WT} and *miR-133b*^{MUT}. This has been already done for the detection of other disease-related miRNAs (e.g., *miR-122* in human serum samples) by using dynamic chemistry labeling (DCL) probes, which can be implemented for single-base resolution (López-Longarela et al., 2020; Robles-Remacho et al., 2023).

Chapter 5.

CRISPR-based molecular therapy against KRAS-mutant LUAD

Chapter 5. CRISPR-based molecular therapy against *KRAS*-mutant LUAD

This chapter will address Objective 2.1. In here, we aimed to design and optimize a CRISPR/Cas9-based molecular therapy targeting G12C/G12D *KRAS*-mutant LUAD (CRISPR-KRAS therapy). To do that, we optimized the sgRNA design and delivery, assessed its specificity and efficiency *in vitro*, and attempted an *in vivo* inducible approach of our CRISPR-KRAS strategy.

5.1 Background

As discussed in Chapter 1, *KRAS* stands out as the most frequently mutated oncogene in LUAD, and it serves as an important diagnostic and prognostic marker for the disease, as well as a long-wanted target for therapy. In LUAD, the most frequent mutations of *KRAS* fall in codon 12, (G12C for smoker patients and G12D for non-smoker patients) (Karachaliou et al., 2013). Codon 12 mutations prevent GTPase activity by GAPs, and thus block the KRAS protein mostly into its GTP-bound active state, leading to an overactivation of KRAS' downstream pathways (MAPK, PI3K/AKT/mTOR…), which ultimately results in tumorigenesis, a more aggressive disease, and poor prognosis of the patients (Trahey & Mccormick, 1987).

However, although the importance of *KRAS* mutations in several human cancers has been well-established, to date, no effective anti-cancer therapies specifically targeting *KRAS* mutations have reached the clinic, except for two recently developed $KRAS^{G12C}$ inhibitors: sotorasib and adagrasib. The way these compounds work is by covalently binding to the cysteine residue present in $KRAS^{G12C}$, and by positioning themselves into an allosteric pocket (switch-II pocket) only available when $KRAS^{G12C}$ is bound to GDP (inactive state). Therefore, upon binding of these compounds, nucleotide exchange is impeded and $KRAS^{G12C}$ is locked into its GDP-bound, inactive form, thus preventing overactivation of downstream pathways (Goebel et al., 2020).

Despite the initial enthusiasm over these two G12C inhibitors, the latest clinical trials have revealed that their performance falls short of expectations, mainly because of acquired resistance and adaptation mechanisms that arise in patients within a short time after treatment. One of the main sources of resistance is the overexpression of $KRAS^{G12C}$ to compensate for the $KRAS^{G12C}$-GDP blockade (Awad et al., 2021). A potential way to circumvent this bypass mechanism would be to completely ablate $KRAS^{G12C}$ at a genomic level, thus preventing rewiring and feedback loops. However, specificity is here a concern, since wildtype *KRAS*

(*KRAS*^{WT}) is a critical gene for non-tumor cells, and *KRAS*$^{G12C/G12D}$ only differ from the wildtype version in one nucleotide.

In this regard, the CRISPR/Cas9 technology stands out as a specific and efficient way to knock out any desired gene. By designing a specific sgRNA against the mutant version of *KRAS* (*KRAS*^{MUT}), Cas9 is directed to the target DNA and induces DSBs. These breaks trigger a DNA damage response and an error-prone repair mechanism (NHEJ), which results in random insertions and deletions (indels). Such indels can cause frameshifts and premature stop codons, thus altering the ORF and potentially rendering the target *KRAS*^{MUT} non-functional (Chapman et al., 2012).

Thus, here we aimed to design a CRISPR/Cas9-based strategy against *KRAS*^{MUT} (CRISPR-KRAS) that is summarized in Figure 28. Briefly, we designed specific sgRNAs against both *KRAS*^{G12C} and *KRAS*^{G12D}, and used a high-fidelity version of Cas9 (HiFi-Cas9) to disrupt *KRAS*^{MUT} in tumor cells without affecting *KRAS*^{WT} in surrounding non-tumor cells. As will be later discussed in section 5.3, our CRISPR-KRAS therapy presents some advantages over the KRAS^{G12C} inhibitors (sotorasib and adagrasib), which makes it an interesting strategy for future clinical approaches against *KRAS*-mutant LUAD.

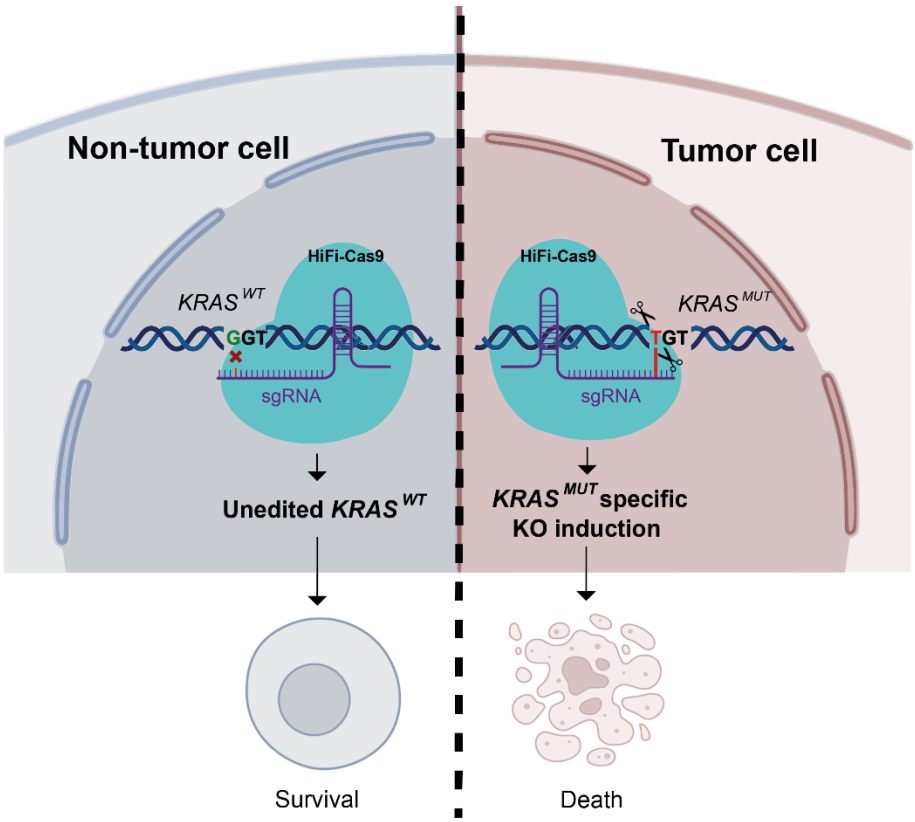


Figure 28. Schematics of our CRISPR-KRAS strategy, using HiFi-Cas9 and KRAS$^{G12C/G12D}$ *sgRNAs targeting specifically the mutant versions of KRAS (KRAS*^{MUT}*), without affecting wildtype KRAS (KRAS*^{WT}*).*

5.2 Results

5.2.1 Optimization of delivery and design of sgRNAs

Our first approach was testing whether the Cas9 technology is specific enough to detect just a single nucleotide change between a mutant sequence and its wildtype version. To do that, we designed a plasmid containing SpCas9 (pSpCas9(BB)-2A-GFP, Addgene #48138), where we subcloned sgRNAs against either $KRAS^{G12C}$ (c34G>T), or against $EGFR^{L858R}$ (c.2573T>G), which represent two of the most important oncogenic mutations in LUAD. Interestingly, while the *KRAS* mutation is upstream of a PAM site, the *EGFR* mutation itself generates a PAM which is absent in the wildtype sequence. We transfected a panel of wildtype/mutant LUAD cell lines with the plasmids, and then we conducted a T7-endonuclease assay to test SpCas9 edition, which is represented by a double band. We observed how edition was specific for the *EGFR* mutation, as only the mutant cell line (H1975) showed edition. Conversely, *KRAS* edition was present in both cell lines, meaning that SpCas9+sgRNA was not specific at recognizing and cleaving $KRAS^{G12C}$ (Figure 29).

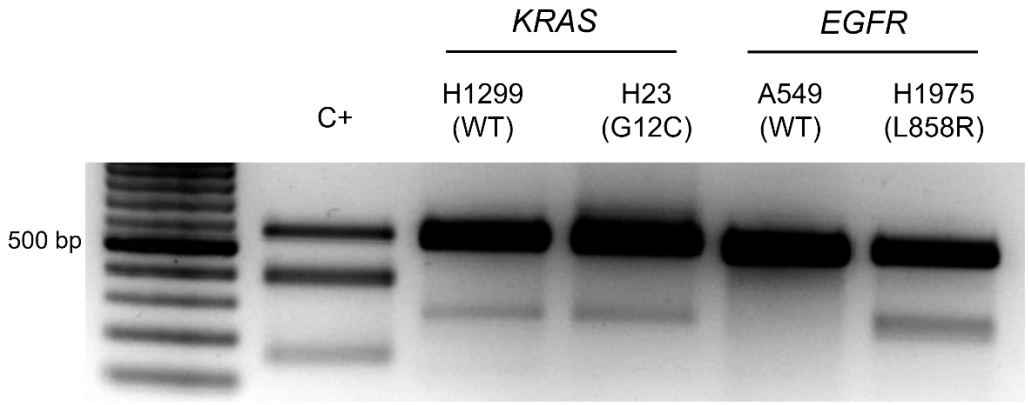


Figure 29. T7-endonuclease assay of a $KRAS^{WT}$ (H1299), $KRAS^{G12C}$ (H23), $EGFR^{WT}$ (A549), and $EGFR^{L858R}$ (H1975) LUAD cell lines after transfection with the corresponding plasmid (SpCas9 + KRAS/EGFR sgRNA). A double band represents edition in the targeted locus (mutant KRAS/EGFR).

To address the issue of $KRAS^{WT}$ off-targets, we considered three different approaches: (i) delivering Cas9+sgRNAs by RNPs instead of plasmid expression, as RNPs are less prone to induce off-targets (Yip, 2020); (ii) changing the SpCas9 for a high-fidelity version of Cas9 (HiFi-Cas9), which was reported to induce less off-targets when delivered in the form of RNPs (Vakulskas et al., 2018); and (iii) designing other sgRNAs against a second available PAM nearby the mutation site.

In order to optimize the delivery strategy for the RNPs, we tested both regular lipofection and nucleofection using different Amaxa 4D-Nucleofector™ programs (EH-158, EN-138, CM-130, and CA-137). According to our results, both delivery strategies worked well, as we could observe edition afterwards (Supplementary Figure 2). Nevertheless, due to the higher reagent consumption,

and a higher loss of cell viability after nucleofection, we decided to continue with regular lipofection for the rest of the experiments. In addition, we designed more specific sgRNAs against *KRAS*^{G12C} and *KRAS*^{G12D}, testing two different PAMs (shown in orange in Figure 30A) that were available in the *KRAS* sequence right downstream of the mutation site: PAM 1 (P1, AGG) and PAM 2 (P2, TGG). The sequences of the designed sgRNAs are detailed in Figure 30A.

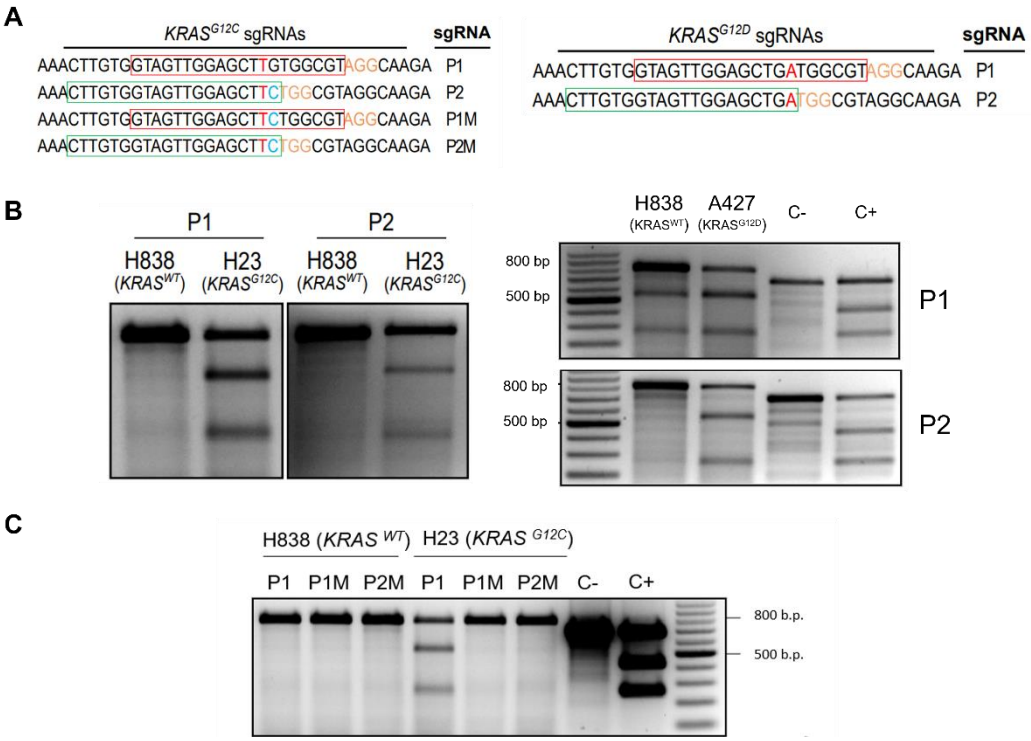


Figure 30. (A) Sequence of designed sgRNAs against KRAS$^{G12C/G12D}$*. In orange, the available PAMs next to the mutation site (in red). In blue, the deliberately introduced mismatches for further assessing the specificity of sgRNAs. (B) T7-endonuclease assays testing the specificity of the newly designed sgRNAs against KRAS*^{G12C} *and KRAS*^{G12D}*, using PAM1 (P1) and PAM2 (P2). (C) T7-endonuclease assay testing the specificity of HiFi-Cas9 with the sgRNAs containing additional mismatches (P1M and P2M). C-: negative control; C+: positive control.*

5.2.2 KRAS$^{G12C/G12D}$ *sgRNAs are specific and do not target KRAS*^{WT}

To test the specificity of our CRISPR system against both *KRAS*^{G12C} and *KRAS*^{G12D}, we transfected *KRAS*^{WT} (H838), a *KRAS*^{G12C} (H23), and a *KRAS*^{G12D} (A427) LUAD cell lines with RNPs formed by the more specific HiFi-Cas9, and our sgRNAs against both PAMs. Then, we again conducted a T7-endonuclease assay to check the edition. If unedited, a single, ~800 bp band should appear, corresponding to the uncut *KRAS* amplicon. However, if edition has occurred, two shorter fragments should appear at ~550 bp and ~250 bp. We observed how the G12D_ P1 was not specific and could not discriminate between *KRAS*^{G12D} and *KRAS*^{WT}. However, G12D_P2, and both G12C_P1 and G12C_P2 showed great specificity, as edition was

only detected in the cell lines harboring the *KRAS* mutation matching to each sgRNA (Figure 30B).

To further assess the specificity of HiFi-Cas9, we designed two more sgRNAs for $KRAS^{G12C}$ (G12C_P1M and G12C_P2M) with an intentional mismatch (represented in blue in Figure 30A), to test whether the edition efficiency was affected by it. Indeed, upon the introduction of a single mismatch, HiFi-Cas9+sgRNA does not recognize its target, and therefore edition does not occur (Figure 30C), proving again its great specificity.

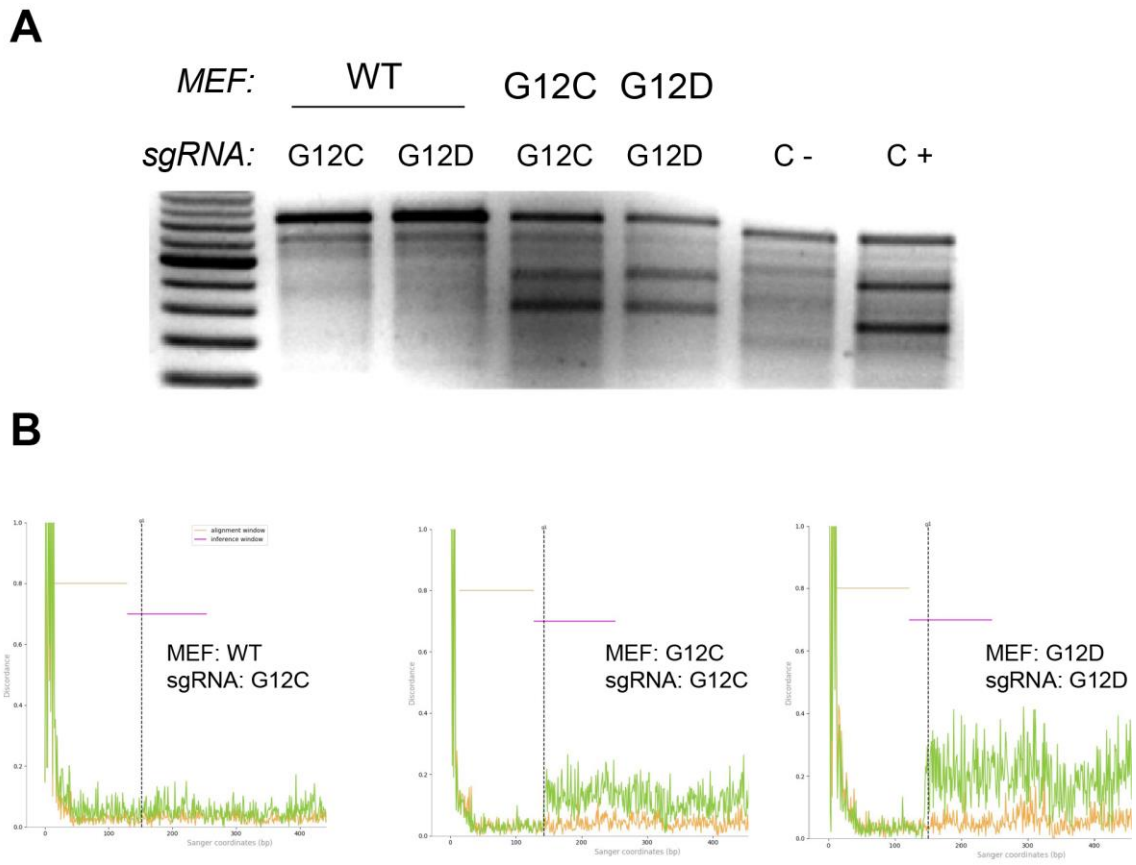


Figure 31. (A) T7-endonuclease assay of MEF models with human $KRAS^{WT}$, $KRAS^{G12C}$, and $KRAS^{G12D}$, treated with specific sgRNAs. C-: negative control; C+: positive control. (B) ICE analysis of Sanger sequenced files from MEFs targeted with sgRNAs. The dotted vertical line denotes the DNA cut site, after which the sequence discordance begins if editing occurs.

Finally, to ensure that the observed specificity was not dependent on the genomic context, but solely on the *KRAS* mutational status, we conducted an additional T7-endonuclease assay on an isogenic panel of MEF models, in which only the *KRAS* variants ($KRAS^{WT}$, $KRAS^{G12C}$, and $KRAS^{G12D}$) were different. Neither of the sgRNAs targeting mutant *KRAS* caused DNA editing in $KRAS^{WT}$ MEFs, and editing was only observed when the corresponding sgRNA-KRAS-mutant was delivered into the MEFs harboring the matching *KRAS* mutation (Figure 31A). These results were further validated through analysis of Sanger sequenced

samples using the ICE tool from Synthego, which confirmed the high accuracy of our system observed with T7 endonuclease assays (Figure 31B).

5.2.3 CRISPR-KRAS therapy impairs cell viability of KRAS-mutant cell lines

Once we confirmed that our sgRNAs were specific and did not target $KRAS^{WT}$, we decided to test the efficiency of our HiFi-Cas9 + $KRAS^{G12C/G12D}$ sgRNA system (CRISPR-KRAS) as a therapeutic approach against *KRAS*-mutant LUAD cell lines. To do that, and setting our eyes in a more translational approach for further *in vivo* applications, we designed adenoviral vectors (AdVs) containing either an empty vector with just GFP as a control (AdV-GFP), or the HiFi-Cas9 sequence + a multiplex cassette including both sgRNAs against $KRAS^{G12C}$ and $KRAS^{G12D}$ (AdV-Cas9). We infected H358 ($KRAS^{G12C}$), A427 ($KRAS^{G12D}$), and H838 ($KRAS^{WT}$) cells with both AdVs, and then measured cell viability using CellTiter-Glo® in both 2D and 3D cultures (Figure 32).

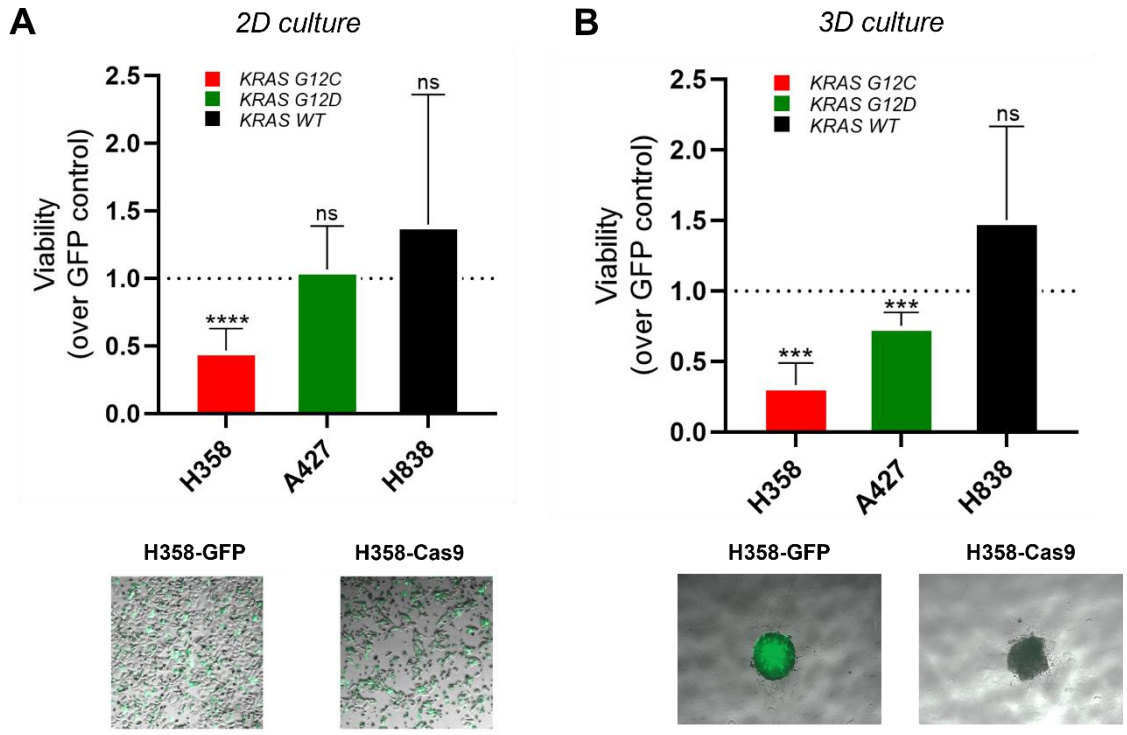


*Figure 32. Cell viability assays of CRISPR-KRAS treated LUAD cells in 2D (A) and 3D culture (B). Representative images of successfully AdV-infected cells are shown below, where a modest AdV transduction efficacy (~50%) can be observed. **** = p-value < 0.0001; *** = p-value<0.001; ns = p-value >0.5.*

Despite the modest transduction efficiency of AdVs compared to RNPs, we reported a reduction in cell viability from around 50% in 2D to nearly 70% in 3D

for H358, and from staying unchanged in 2D to nearly 25% in 3D for A427, while the viability of H838 (wildtype control) remained unchanged in both culture systems.

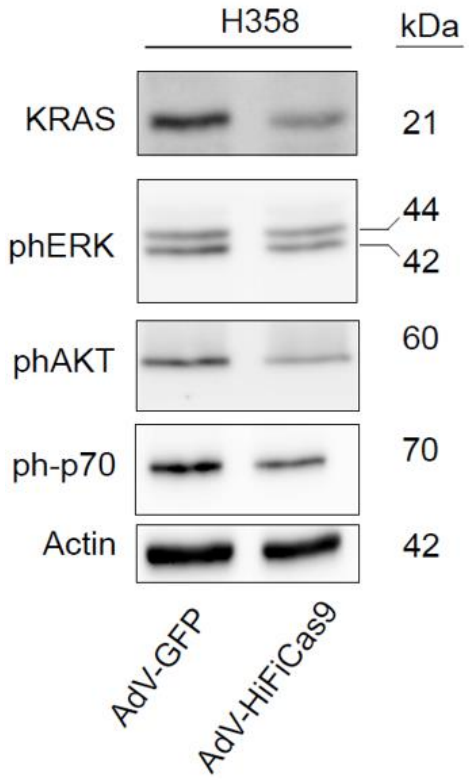


Figure 33. Western blot protein analysis of KRAS and its downstream targets after CRISPR-KRAS treatment of the H358 LUAD cell line.

To further demonstrate that the reduction of cell viability was due to our CRISPR-KRAS therapy disrupting mutant KRAS, we extracted protein from treated cells and conducted a Western blot to monitor the levels of KRAS and a series of downstream targets of the KRAS-signaling cascade (phosphorylated ERK (ph-ERK) from the MAPK pathway, and phosphorylated AKT (ph-AKT) and p70S6 from the mTOR pathway). We observed how the overall protein levels of both KRAS and its downstream targets were downregulated in the AdV-Cas9 treated cells compared to the GFP control (Figure 33), thus proving that our strategy is efficient at disrupting KRAS expression.

5.2.4 Development of an inducible murine model for CRISPR-KRAS therapy

Finally, we attempted a first *in vivo* approach for our CRISPR-KRAS therapy by using a doxycycline (dox) inducible model. First, we obtained dox-inducible clones of the H1792 ($KRAS^{G12C}$), A427 ($KRAS^{G12D}$), and H838 ($KRAS^{WT}$) cell lines. To do that, we first delivered a lentiviral plasmid (pLVX-TRE-HiFiCas9-Puro,

VectorBuilder #VB181129-1029urq) containing HiFi-Cas9 under the control of a tetracycline response element (TRE) promoter, and puromycin resistance for selection of successfully infected cells. By using a low MOI (0.3-0.5), we obtained a low number of transduced cells, but ensured a single integration which prevents expression leaking thereafter. Next, we treated the surviving clones with doxycycline, and selected those clones that did not exhibit leaking of HiFi-Cas9 expression upon dox-treatment (Supplementary Figure 3). Then, we transduced those clones with LVs carrying a second plasmid (pLVX-dsgRNA-EGFP, VectorBuilder #VB181004-1106yfp) which contained both sgRNAs against *KRAS*$^{G12C/G12D}$ and EGFP as a reporter, and we sorted the final clones by green fluorescence.

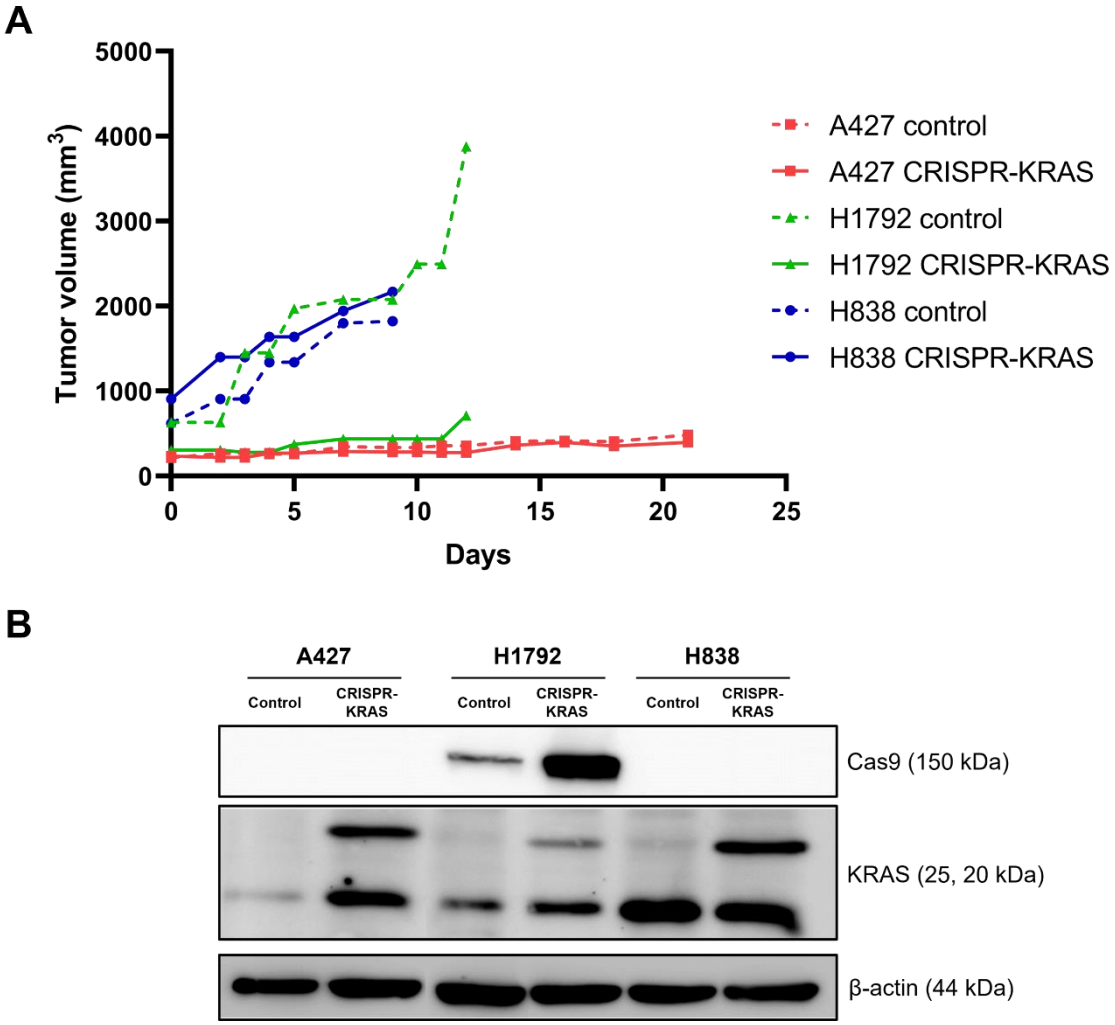


Figure 34. Dox-inducible CRISPR-KRAS xenografts in mice. (A) Measurements of tumor volumes over time. (B) Western blot measuring HiFi-Cas9 induction and KRAS disruption in final extracted tumors. Only one xenograft per condition was conducted, so no statistical tests were performed.

After expanding the final clones (dox-inducible HiFi-Cas9 + dual sgRNAs) *in vitro*, we introduced subcutaneous cell-derived xenografts into NSG mice of either

the final clones (CRISPR-KRAS), or the clones without the dual sgRNAs as a control. Mice were then treated with doxycycline for one week as detailed in section 2.2.5. The tumor growth rate was monitored during that time taking measures of the bulges in the flanks of mice, the tumor volumes were calculated, and, at the end of the experiment, tumors were collected, measured, and protein was extracted for Western blot analysis.

We observed how the H838 ($KRAS^{WT}$), as expected, did not show any differences in tumor growth, whereas the H1792 ($KRAS^{G12C}$) presented a strong tumor growth reduction when the CRISPR-KRAS therapy is induced by doxycycline, compared to the control (Figure 34A). However, the A427 ($KRAS^{G12D}$) did not show any differences in tumor growth and, overall, it did not develop substantial, measurable tumors over time. This could be attributed to an inaccurate injection of the xenograft rather than the ineffectiveness of the CRISPR-KRAS therapy. Moreover, after performing a Western blot analysis to monitor HiFi-Cas9 induction and KRAS disruption, only the H1792 showed a strong HiFi-Cas9 expression, whereas both H838 and A427 seem to have lost it. In addition, KRAS expression was consistent among all CRISPR-KRAS treated conditions, which is opposite to what we observed in the *in vitro* AdV experiments (Figure 34B).

5.3 Discussion

It is an undeniable fact that CRISPR/Cas9 has revolutionized the fields of biomedicine and biotechnology as an efficient and specific technology to induce KOs and gene edition at the desired DNA locus. Therefore, in the last decade, there has been a boom of preclinical and clinical studies applying CRISPR/Cas9 to cure previously untreatable diseases. Cancer is no exception to this, and due to the clinical importance of *KRAS*-mutant tumors, several approaches using CRISPR/Cas to target them have been developed in the last 5 years (**Table 5**).

Table 5. Summary of studies targeting KRAS using CRISPR/Cas-based technologies.

CRISPR/Cas system	**Tumor subtype**	***KRAS* variant(s)**	**Reference**
HiFi-Cas9	Lung	G12C, G12D	Álvarez-Pérez et al. (under review)
SpCas9 + Base editors	Pancreas, colorectal, and lung	G12D, G12S, G13D	(Sayed et al., 2022)
SpCas9	Pancreas	G12D	(Ischenko et al., 2021)
SpCas9	Pancreas	G12D	(McAndrews et al., 2021)
dCas9-HDAC1	Colorectal and lung	Whole *KRAS*	(J. Liu et al., 2021)
SpCas9	Colorectal	Whole *KRAS*	(Wan et al., 2020)
CasRX (Cas13d)	Pancreas	G12C, G12D	(W. Jiang et al., 2020)
SpCas9 + dCas9-KRAB	Lung	G12S	(Gao et al., 2020)
SpCas9	Pancreas	G12D	(Lentsch et al., 2019)
SpCas9	Colorectal	G12D, G12V, G13D	(W. Kim et al., 2018)
Cas13a	Pancreas, colorectal, and lung	G12D	(X. Zhao et al., 2018)
SpCas9	Colorectal and pancreas	G12D, G12V	(W. Lee et al., 2018)
SpCas9	Lung	Whole *KRAS*	(S. M. Kim et al., 2018)
SpCas9	Pancreas	Whole *KRAS*	(Muzumdar et al., 2017)

Adapted and updated from (Bender et al., 2021).

Most of these previous attempts to target $KRAS^{MUT}$ tumors using CRISPR systems were restricted in specificity and did not provide adequate evidence of

specific $KRAS^{MUT}$ targeting, since they either aimed to target all *KRAS* variants, or inadequately evaluated $KRAS^{WT}$ targeting (S. M. Kim et al., 2018; W. Kim et al., 2018; Lentsch et al., 2019). In contrast, our study thoroughly evaluated the specificity of our CRISPR-KRAS system by (i) interrogating the specificity of HiFi-Cas9 with the introduction of additional mismatches in the sgRNA sequence, (ii) examining edition in $KRAS^{WT}$ cell lines, and (iii) checking the specificity of both $KRAS^{G12C}$ and $KRAS^{G12D}$-targeting systems in isogenic MEF models.

Moreover, the CRISPR-KRAS strategy developed as part of this thesis constitutes a proof-of-concept for the design of a multiplexed system that can be programmed to target two different mutations, either in *KRAS* or in other genes. In fact, CRISPR-targeting of oncogenic mutations has vast potential, as some authors have proven by analyzing SNVs in the 20 most mutated driver genes in cancer, and checking whether they fall either in the PAMs, or into the seed sequence next to a PAM. About half of the SNVs fitted these criteria, and therefore they are targetable by one of the three CRISPR/Cas system analyzed, which could be a potential tool to specifically eliminate tumor cells harboring such SNVs (Gao et al., 2020). It would be interesting to see which ones of those SNVs generate new PAMs in the sequence of the driver gene, since, according to our data, this results in a better specificity, as we saw with the *EGFR* sgRNAs we designed as a first approach, compared to the *KRAS* ones.

Our cell viability results indicate that disruption of either $KRAS^{G12C}$ or $KRAS^{G12D}$ driver mutations inhibits cell growth in LUAD cell lines *in vitro*, with no cell toxicity effects or editing observed in $KRAS^{WT}$ cells. As previously pointed out by some authors (Janes et al., 2018), $KRAS^{MUT}$ dependency may be underestimated in 2D culture, which is why we also conducted 3D cell viability assays, and our next step should be to escalate the AdV delivery system for an *in vivo* administration into mice xenografts. Indeed, our results support this idea, as the 3D settings showed a higher *KRAS* dependency and greater reduction of cell viability when inhibiting both $KRAS^{G12C}$ and $KRAS^{G12D}$. Furthermore, the observed decrease in viability correlates with the deactivation of KRAS-dependent pathways, evidenced by the decrease in the phosphorylation of ERK (MAPK pathway), and p70S6K (PI3K/AKT/mTOR pathway). We did not observe, however, a consistent change in AKT phosphorylation as reported in previous studies (Canon et al., 2019).

Regarding the clinical implications of our results, *KRAS* mutations are especially prevalent in two of the deadliest cancers: LUAD (32% of patients, with G12C as the most frequent variant), and pancreatic ductal adenocarcinoma (88%, mostly G12D) (Prior et al., 2020). For this reason, the development of antineoplasic therapies targeting mutant *KRAS* has been long pursued. After decades of research efforts, sotorasib (AMG-510, sold under the name Lumakras®, from AMGEN), was the first specific and irreversibly $KRAS^{G12C}$ inhibitor approved in 2021 for NSCLC patients that had received at least one prior systemic therapy (Canon et al., 2019). Just recently, in December 2022, another $KRAS^{G12C}$ inhibitor, adagrasib (MRTX849, commercialized as Krazati®, from Mirati Therapeutics), was approved by the FDA

for the treatment of KRAS^{G12C} NSCLC patients previously treated with chemotherapy and anti-PD-1/PD-L1 therapy, and it is currently in phase-I/II clinical trials (KRYSTAL-1, NCT03785249) (Jänne et al., 2022). Although to date only KRAS^{G12C} inhibitors have reached the clinics, there are several studies underway for drugs targeting G12D and other *KRAS* mutations, albeit still in preclinical stages (Nagasaka et al., 2021). The high reactivity of the cysteine residue in KRAS^{G12C} has favored the development of irreversible inhibitors for it instead of other, more frequent *KRAS* mutations. For example, the lower reactivity of the aspartic acid in KRAS^{G12D} makes it more difficult to target, and the compounds in development, e.g., MRTX1133, are reversible inhibitors that do not bind covalently to it and are therefore less efficient than sotorasib or adagrasib (Hallin et al., 2022).

Despite the excitement surrounding these two KRAS^{G12C} inhibitors, the most recent clinical trials have shown that their effectiveness has not lived up to expectations. Overall, their clinical benefit as second-line treatment appears to be lower (objective response rate of 36-43%) than what has been observed in other targeted therapies, such as EGFR TKIs (60-70%) (Gourd, 2020; Nagasaka et al., 2021). In one of the latest phase-III clinical trial in KRAS^{G12C} NSCLC patients (CodeBreaK 200, NCT04303780), whilst doubling the response rate to the anticancer drug docetaxel, only 28% of patients presented an objective response to sotorasib, and overall survival was not improved compared to docetaxel (de Langen et al., 2023). As such, currently these drugs are not ready to become first-line agents for the treatment of *KRAS*-mutant lung cancer.

The reason behind the limitations of the clinical results so far could be the adaptation mechanisms and the resistances acquired after the blockage of KRAS^{G12C}, frequently due to the appearance of other secondary *KRAS* mutations or *KRAS^{G12C}* allele amplifications. The secondary mutations can disable the intrinsic GTPase activity of KRAS, definitely blocking it into the GTP-bound form, and thus rendering the inhibitors useless and perpetuating the active state and the oncogenic signaling (Awad et al., 2021). Furthermore, cells adapt to the blockade in the GDP-form by overexpressing *KRAS^{G12C}*, and overactivating SOS1/2 and other GEFs, which promotes that *de novo* synthesized KRAS binds to GTP and remains in its active form as a compensatory mechanism (J. Y. Xue et al., 2020).

As we have seen, both adaptive and most of the described resistance mechanisms to current KRAS^{G12C} inhibitors rely on the relentless expression of *KRAS^{G12C}*. Therefore, future molecular therapies should aim to also block the production of *de novo* KRAS^{G12C}. Indeed, this is one of the biggest advantages of our CRISPR-KRAS strategy over KRAS^{G12C} inhibitors like sotorasib. Because our CRISPR-KRAS therapy ablates *KRAS^{G12C}* at DNA level, it would be preferable and could circumvent many of the aforementioned resistance and adaptation mechanisms. In addition, the ease of design of multiplexed sgRNAs translates into a

greater versatility, allowing us to simultaneously target several *KRAS*-mutant variants for which no drugs are still available, or even other oncogenic driver genes that act as synthetic lethalities for *KRAS*-mutant tumors.

However, as the rest of CRISPR/Cas9-based therapies, our CRISPR-KRAS strategy still cannot outperform the drugs applied in clinics in terms of safety and delivery. As discussed earlier in Chapter 1, the two main problems that CRISPR/Cas9 technologies need to face before its clinical use are off-target effects and safe delivery options. Although we have thoroughly assessed the specificity of our sgRNAs and the HiFi-Cas9, off-target DSBs with unpredictable consequences cannot be ruled out after long-term exposure to our CRISPR-KRAS therapy. Some strategies that could be followed to reduce HiFi-Cas9 exposure time are using an inducible system or customizing the Cas9 with a ubiquitination tag to reduce its lifespan (Chira et al., 2018). In our attempt to generate a dox-inducible xenograft mice model for CRISPR-KRAS, despite the apparent good results for the H1792 tumors, there seem to be issues with the inducible HiFi-Cas9 and the disruption of KRAS when applied *in vivo*. As this experiment was merely an initial approach, additional replicates in more mice should be performed to conduct a proper statistical analysis and to verify whether the observed differences are due to the action of our CRISPR-KRAS therapy, or just an artefact of the experiment.

The systemic delivery of the CRISPR/Cas9 system is another handicap to solve. Both Cas9 and sgRNAs can be delivered either in the form of DNA plasmids, mRNAs, or RNPs. Nucleic acid versions of Cas9+sgRNAs, either as DNA or mRNAs, are usually delivered by viral vectors. Integrative ones, such as LVs, are more efficient, but raise concerns regarding uncontrolled integration sites. AdVs and AAVs are recommended for *in vivo* approaches as they are not integrative, which is why we used them for the cell viability assays, with the aim to further escalate them in the future to administration in mice. However, their use in human patients entails the drawbacks of the immunogenicity of AdVs, and the very limited packaging capacity of AAVs (Arenas et al., 2022). Conversely, RNPs show the highest efficiency because of their faster degradation, which reduces exposure time and prevents off-targets (Yip, 2020). This agrees with our results, as by changing from using plasmids to deliver our CRISPR-KRAS therapy *in vitro* with RNPs greatly improved specificity and prevented editing of $KRAS^{WT}$. However, RNPs are much more expensive to produce on a large scale for *in vivo* administration in patients, and they are difficult to package into viral vectors, which are the ones showing the highest efficiency in delivery to cells *in vivo*. Finally, other delivery methods have been developed, such as extracellular vesicles or lipid-based nanoparticles, although they often require extensive optimization and need further improvement before their clinical applications (Yip, 2020).

In here, we have shown the optimization and the first steps of developing our CRISPR-KRAS strategy. According to the results presented in this thesis, and further experiments performed *in vivo* with preclinical models of patient-derived

xenografts (not included in this thesis, but compiled in our publication Álvarez-Pérez et al. "*High-fidelity Cas9-mediated targeting of KRAS driver mutations restrains lung cancer in preclinical models*.", currently under review), our CRISPR-KRAS therapeutical strategy shows great potential for the treatment of *KRAS*-mutant LUAD both *in vitro* and *in vivo*, and it better addresses KRAS oncogenic inhibition than sotorasib.

In conclusion, our CRISPR-KRAS strategy presents a proof of concept for specifically and efficiently disrupting $KRAS^{G12C/G12D}$ in tumor cells without affecting non-tumor, $KRAS^{WT}$ cells. This shows greater versatility and could potentially circumvent the problems with resistances to $KRAS^{G12C}$ inhibitors like sotorasib. Further research is still necessary to overcome the off-target and delivery issues that hinder the clinical use of CRISPR/Cas9-based therapies. The forthcoming advancements in delivery systems will be critical in determining if our CRISPR-KRAS strategy finally has a clinical future.

Chapter 6.
CRISPR-screening for the detection of collateral dependencies

Chapter 6. CRISPR-screening for the detection of collateral dependencies

This chapter will address Objective 2.2. In here, we aimed to perform a CRISPR-screening to unravel the collateral dependencies arising after the knockout of an oncogene in a LUSC model, which are potential sources of resistance to targeted molecular therapies. The work performed in this Chapter was carried out during an international fellowship at Prof. Jacob Giehm Mikkelsen's research group at the Aarhus University, Denmark.

6.1 Background

6.1.1 PKP1 is an oncogene and potential therapeutic target in LUSC

Lung squamous cell carcinoma (LUSC) is the second most frequent form of lung cancer after LUAD, adding up to around 20% of all lung cancers (Barta et al., 2019). LUSC possesses distinct genetic particularities when compared to LUAD. For instance, mutations in the most common driver genes of LUAD, *KRAS* and *EGFR*, are typically absent in LUSC (Niu et al., 2022; Rekhtman et al., 2012). Thus, many targeted therapies developed against LUAD, such as those targeting activating mutations in EGFR (i.e., TKIs), are largely ineffective in LUSC patients (Giaccone, 2005; Kwak et al., 2010; Rosell et al., 2009). Besides, some agents used as treatment for LUAD (e.g., bevacizumab or pemetrexed) are not indicated for LUSC patients, in whom they may cause adverse effects more frequently (Azzoli et al., 2009; Scagliotti et al., 2008). Although some advances have been made in the past years, an effective targeted therapy for LUSC remains to be found, and there is still a long way to go to improve the prognosis of these patients (Niu et al., 2022).

Our laboratory recently discovered a novel oncogenic protein called plakophilin-1 (PKP1) that promotes LUSC (Martin-Padron et al., 2020), thus opening a new line of research aimed at developing new therapeutic strategies targeting PKP1 in LUSC. Plakophilin-1 is an 83 kDa protein encoded by the *PKP1* gene that is located on human chromosome 1, and it belongs to the armadillo repeat family of proteins called plakophilins. Plakophilins are normally expressed in stratified and single-layered skin epithelial cells, being key proteins in the function of the desmosomes, which allow cell-to-cell adherence. Prior to our work, it was described that mutations in *PKP1* cause a lower number and a worse constitution of desmosomal structures, which phenotypically translates into a loss of the integrity of the epidermis, generating a pathology known as skin fragility syndrome (McGrath et al., 1997). Besides, *PKP1* has been described as a potential tumor suppressor protein associated with metastasis in melanoma (H. Z. Wang et

al., 2019), prostate cancer (C. Yang et al., 2015), head and neck cancer (Teh et al., 2011), esophageal adenocarcinoma (Kaz et al., 2012) and, interestingly, even some lung cancer cell lines (Haase et al., 2019).

Contrarily to the previous evidence of *PKP1* acting as a tumor suppressor, we (Angulo et al., 2008) and others (J. Liu et al., 2018) observed that *PKP1* is among the most overexpressed genes in squamous-type lung tumors, and therefore decided to functionally study the contribution that *PKP1* exerted in the development of this subtype of lung cancer. To do this, we developed models of loss-and-gain function of *PKP1* which made us conclude that *PKP1* plays mainly an oncogenic role in LUSC, and that its inhibition reduces tumor proliferation in both *in vitro* and *in vivo* models. On the other hand, when overexpressing *PKP1* in LUSC cell lines, we observed an increase in tumor proliferation and survival (Martin-Padron et al., 2020). Deepening into the molecular mechanisms of *PKP1*'s oncogenic activity, we further discovered that *PKP1* is capable of increasing the translation of the *MYC* oncogene in a feedforward loop manner. Thus, PKP1 binds to the 5'-UTR of *MYC* and interacts with the poly(A) binding protein (PABP), promoting *MYC* ribosomal translation. At the same time, in the nucleus, MYC can bind to the *PKP1* promoter region and activate *PKP1* transcription, thus closing a feedforward regulation loop (Boyero et al., 2022).

These discoveries nominate PKP1 as a new post-transcriptional regulator of *MYC*, and thus, PKP1 becomes a potential new antitumor therapeutic target for LUSC. However, the appearance of resistances to a future *PKP1*-targeting therapy may hinder the whole process and therefore, a study of the vulnerabilities that appear after *PKP1* inhibition is essential so that future targeted therapies against *PKP1* in LUSC are successful.

6.1.2 CRISPR-screening: a molecular tool for the discovery of collateral dependencies

Targeted therapies face one big challenge in cancer therapeutics, which is the appearance of resistances. Tumor cells often find resistances to targeted therapies by rewiring their intracellular pathways and developing non-mutational bypass mechanisms (Garraway & Jänne, 2012). These bypass pathways that are able to sustain cancer cell survival after inhibition of a driver oncogene are defined as collateral dependencies (CDs) (Lou et al., 2019). Identifying such CDs is a key step on developing a novel targeted therapy, as they will allow us to anticipate and circumvent resistances that arise after the treatment, for example, by designing a double therapy co-targeting the desired gene and its CDs simultaneously.

One methodology to detect mechanisms of drug resistance and CDs of inhibited oncogenes are CRISPR/Cas9 loss-of-function library screenings. They consist of generating a lentiviral library containing genome-wide sgRNAs, infecting a treated and a control cell pool, and then analyzing by NGS the differences on sgRNA enrichment between the treated and the control conditions (Thomsen &

Mikkelsen, 2019). Since their first applications in the mid-2010s, CRISPR library screenings have proven to be a powerful tool for detecting phenotype-related genes, as well as finding out genetic dependencies given a certain gene's inhibition (CRISPRi/KO libraries) or activation (CRISPRa libraries) (Kurata et al., 2018). Thus, in order to design an effective PKP1-inhibiting therapy in the future, we considered relevant using a CRISPR loss-of-function library screening on *PKP1*-KO cell models to discover which other genes or cell pathways might be involved in the resistances appearing after depleting *PKP1* from tumor cells.

Prof. Mikkelsen's research group at the Aarhus University, Denmark, has wide experience on the study of delivery vectors for gene therapy, such as modified lentivirus, HIV-derived vectors, or the PiggyBac transposon (Thomsen et al., 2022; Wolff et al., 2021; Wolff & Mikkelsen, 2022). To date, they are one of the few European laboratories specialized in the application of lentiviral CRISPR libraries, and they have a well-stablished and optimized pipeline for the analysis of the results, hence our choice for collaborating with them in this work.

6.2 Results

6.2.1 Library titration and Cas9 activity checking

The Brunello lentiviral library was previously amplified and packaged as described in (Thomsen & Mikkelsen, 2019). Prior to its use, it was necessary to titrate it specifically for the three cell lines we were going to perform the CRISPR-screening on: the parental SK-MES-1 LUSC cell line as a control of wildtype *PKP1* expression, and two SK-MES-1 KO clones for the *PKP1* gene that were previously generated in (Martin-Padron et al., 2020) (KO-1 and KO-2). The obtained viral titers were 3.6×10^6 CFU/mL for control, 1.17×10^7 CFU/mL for KO-1, and 8.45×10^6 CFU/mL for KO-2 (Figure 35).

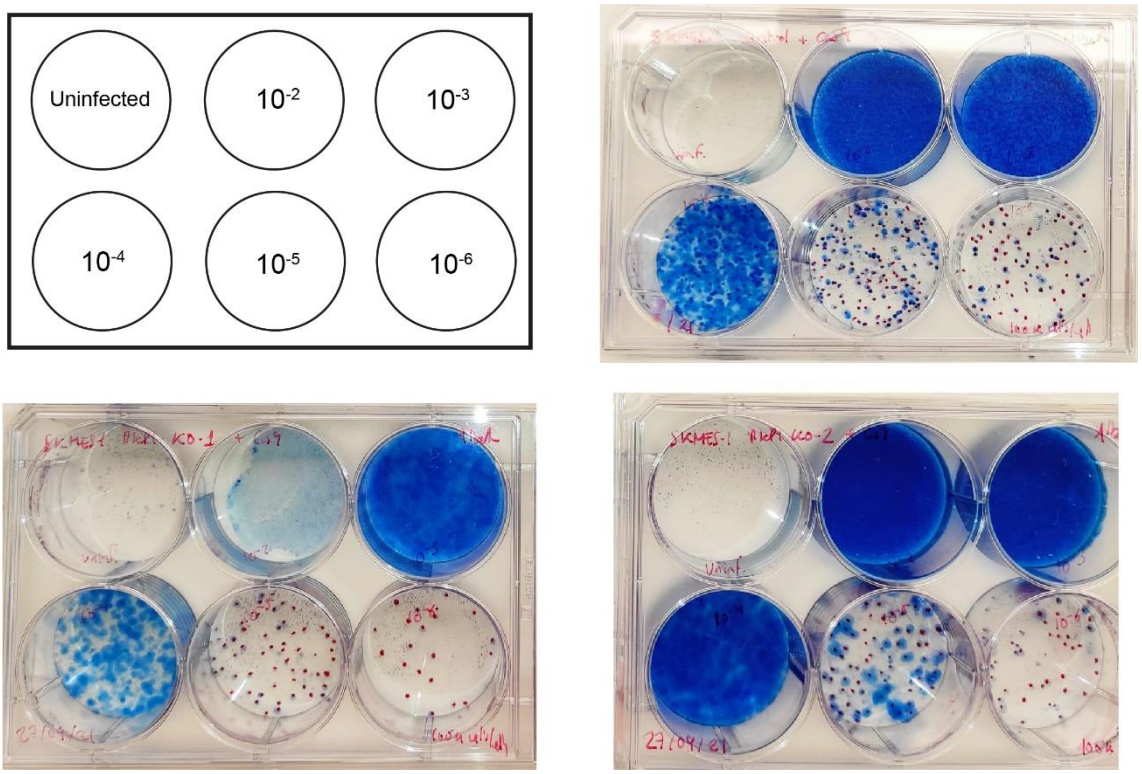
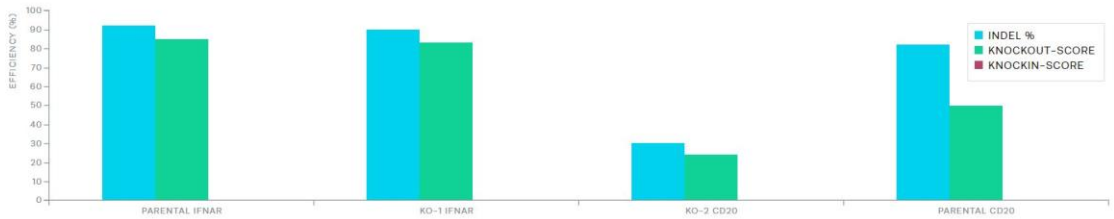


Figure 35. CRISPR-screening lentiviral library titration. Serial dilutions of the virus were prepared as shown in the diagram, and the number of colonies in the 10^{-5} and the 10^{-6} wells of the control (upper right), KO-1 (lower left), and KO-2 (lower right) clones were counted for the calculation of the library titer in CFU/mL.

In addition, an initial Cas9 activity check was also required to ensure the Cas9 delivered to our three cell models was properly working. To do that, the Cas9-expressing control, KO-1, and KO-2 cells were infected with lentivirus containing a sgRNA against the *IFNAR* gene. This sgRNA was chosen because in previous experiments at Giehm's group it displayed high efficiency, and after we checked in the DepMap database that the SK-MES-1 cell line has basal expression

of *IFNAR*. After infection and puromycin selection, gDNA was extracted, PCRs amplifying *IFNAR* were run, and their products analyzed by Sanger sequencing. An ICE analysis was then performed on the sequencing data, revealing that both control and KO-1 showed great Cas9 efficiency, with >90% indel rate for the *IFNAR* gene. However, KO-2 did not show any *IFNAR* amplification, so a second high-efficiency sgRNA was tested, this time against the *CD20* gene, which also showed basal expression in the SK-MES-1 cell line according to DepMap. After the ICE analysis, we discovered that Cas9 was also properly working and inducing indels in the KO-2 clone, albeit with a lower efficiency than for *IFNAR* (~30% indel rate for *CD20*) (Figure 36).



SAMPLE	GUIDE TARGET	PAM SEQUENCE	INDEL %	MODEL FIT (R^2)	KNOCKOUT-SCORE	
Parental IFNAR	TAGATGACAACTTTATCCTG	AGG	92	0.94	85	View Details
KO-1 IFNAR	TAGATGACAACTTTATCCTG	AGG	90	0.95	83	View Details
KO-2 CD20	GGATCATCAGAAGACCCCCC	AGG	30	0.94	24	View Details
Parental CD20	GGATCATCAGAAGACCCCCC	AGG	82	0.9	50	View Details

Figure 36. Cas9 activity checking by indel analysis using Synthego's Interference of CRISPR Edits (ICE) tool.

6.2.2 CRISPR-screening and NGS sample preparation

The three cell lines (control, KO-1, and KO-2) were infected with the Brunello library and kept in growth for 21 days, as detailed in section 2.4.3.b and Figure 9. After some troubleshooting with the amount of cells plated, and cells not growing enough to ensure complete coverage of the library for the KO-1 condition (at least, 40 million cells per condition), we obtained one replicate of the KO-1 clone with suspected compromised library coverage (A7 and A21 samples, for 7-days and 21-days pellets, respectively), two adequate coverage KO-2 clone replicates (B7, B21 for replicate 1; and Y7, Y21 for replicate 2), and two adequate coverage control replicates (C7, C21 for replicate 1; and Z7, Z21 for replicate 2).

Once all the pellets were obtained, the gDNA was extracted and the targeted regions for NGS were amplified by PCR as described in section 2.4.3.c. An electrophoresis in a 1% agarose gel was performed before shipping the samples to the NGS facility for checking that the PCRs were properly working, and that we were amplifying the sgRNA regions, which should render a band of ~375 bp (Supplementary Figure 4). After this check, PCR products were purified and sent to

NGS. An additional sample of the library plasmids (pP) was also sent for sequencing as a quality control for later data analysis.

6.2.3 Quality check of NGS data

Upon receiving the NGS results, a quality control was performed to check the correct performance of the screening and the sequencing. In summary, the sequencing coverage was fine for all the samples except for the library plasmid control, as the expected number of reads was <40M (**Table 6**). Therefore, a second sample of the library plasmid (pP2) had to be sent for sequencing. The percentage of lost sgRNAs after '*cutadapt*' filtering and '*bowtie*' mapping to the Brunello library was, as expected, higher in the t=21 days samples than in the t=7 days samples. However, the Z21 sample showed a relatively low amount of lost sgRNAs (0.5% at day 7 vs 0.6% at day 21), meaning that it could have been subjected to a lower selective pressure.

Table 6. Statistics of NGS data results.

Condition	Sample label	Total reads (millions)	*'cutadapt'* filtered (%)	*'bowtie'* mapped (%)	Lost sgRNAs (%)
KO-1	A7	55.7	96.5	84.7	1
	A21	56.5	96.3	84.8	2.4
KO-2	B7	49	98.1	88.7	0.7
	B21	51	97.2	86.6	2.5
	Y7	40.6	98.1	88.5	0.5
	Y21	56.5	97.8	87.3	1.9
Control	C7	42.8	95.7	86	0.8
	C21	40.2	97.9	87	3.3
	Z7	52.8	96.4	84	0.5
	Z21	70.9	96.6	84.2	0.6
Library plasmids	pP	30.7	98	89.7	0.2
	pP2	38.1	98.2	88.7	0.1

Sample labeling: 7 – 7 days pellet; 21 – 21 days pellet; pP – Library plasmids controls; A – KO-1; B, Y – KO-2, replicates 1 and 2; C, Z – parental control, replicates 1 and 2.

Further analyses were made to calculate the coverage per sgRNA, and the cumulative distribution function (CDF) of the sgRNA ranking of every sample. The expected sgRNA coverage was ~500 reads/sgRNA, and all the samples were distributed around such threshold (Figure 37A). However, the C21 and Z21 showed a big difference of coverage, when they were supposed to be replicates of the same condition. In addition, after calculating the AUC of the CDF, we saw how the t=21 samples showed a higher depletion of genes termed as "essential" for cells, represented by a higher AUC value (Figure 37B). However, again we

observed a difference in the depletion of essential genes among the C21 and Z21, as the Z21 showed an apparent lower selective pressure.

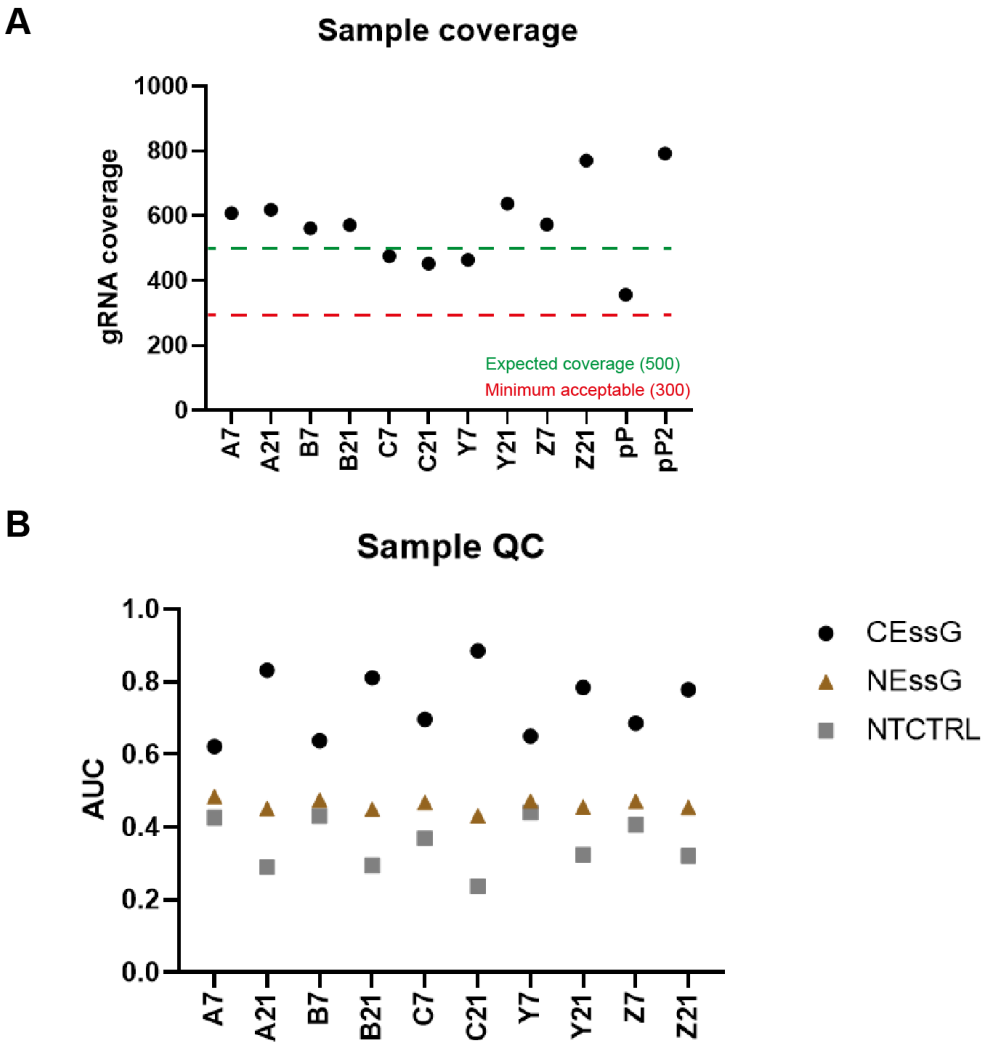


Figure 37. Quality check analysis of the coverage per sgRNA (A), and the area under curve (AUC) from the cumulative distribution function of sgRNA ranking (B). A higher AUC represents a higher depletion of essential genes. For comparison, Trevor Hart's genesets are also shown: CEssG (Core Essential Genes), NEssG (Non-Essential Genes), and NTCTRL (Non-Targeting Control).

6.2.4 PKP1-KO cells show a collateral dependency on mitochondrial function

After the quality check, we performed a *JACKS* analysis that assigns a depletion score to every sgRNA based on the number of final reads compared to the previous timepoint. The more negative this score is, the more depleted the sgRNA appears in a sample, and therefore the more "essential" such gene is for that condition. We represented all the sgRNA in a dot plot, grouping all the KO conditions in the Y axis, and the controls in the X axis (Figure 38). Our collateral

dependencies, and thus possible mechanisms of resistance to *PKP1* inhibition, are those sgRNAs that present a very negative score in the Y axis, but a "neutral" score (between -0.5 and 0.5) in the X axis. Individual dot plots for some of the top scoring sgRNAs can also be seen in Supplementary Figure 5.

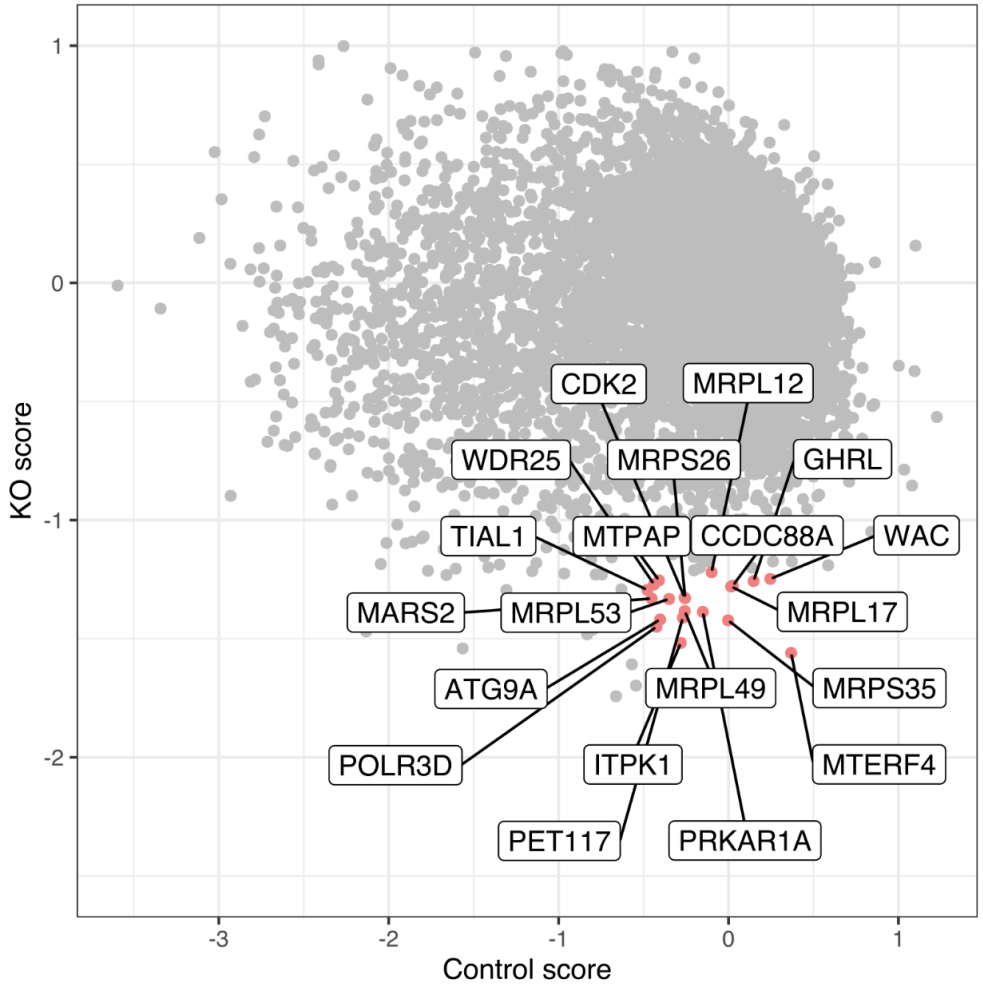


Figure 38. Dot plot representing the JACKS depletion score of all sgRNAs after the CRISPR-screening. Represented in red with a label are the top 20 genes with a lowest KO score, and a neutral control score (between -0.5 and 0.5).

After taking the top 100 scoring genes (lowest KO score, neutral control score between -0.5 and 0.5), we also performed a protein interaction analysis with STRING and Metascape to see which pathways were enriched and could possibly be implicated in collateral dependencies of cells after *PKP1* inhibition. The STRING protein interaction network analysis defined two main clusters. The first cluster was mainly composed by structural mitochondrial ribosome proteins (MRPs), such as *MRPS35*, *MRPS26*, *MRPL49*, or *MRPL53*; along with some components of the electron transport chain complex II, such as *SDHC* and *SDHA* (succinate dehydrogenase cytochrome subunits C and A) (Figure 39A). The second cluster included genes involved in mitochondrial tRNA processing and methylation (e.g.,

MARS2, a methionine-tRNA ligase), and overall, a lot of genes participating in mitochondrial mRNA processing, such as *MTPAP* (mitochondrial poly-A polymerase), and the top candidate in terms of *JACKS* score, *MTERF4* (mitochondrial termination factor 4) (Figure 39B).

Finally, the pathway enrichment analysis performed in Metascape showed that the most common affected pathways where those related to mitochondrial gene expression, and processing of mitochondrial mRNA (Figure 39C). Taken together, these results demonstrate an increased dependence of SK-MES-1 cells on mitochondrial function-related genes when *PKP1* is depleted from them. This suggests some kind of mitochondrial function vulnerability that could be exploited as a future co-therapy along with *PKP1* inhibition in LUSC, as will be further discussed in the next section.

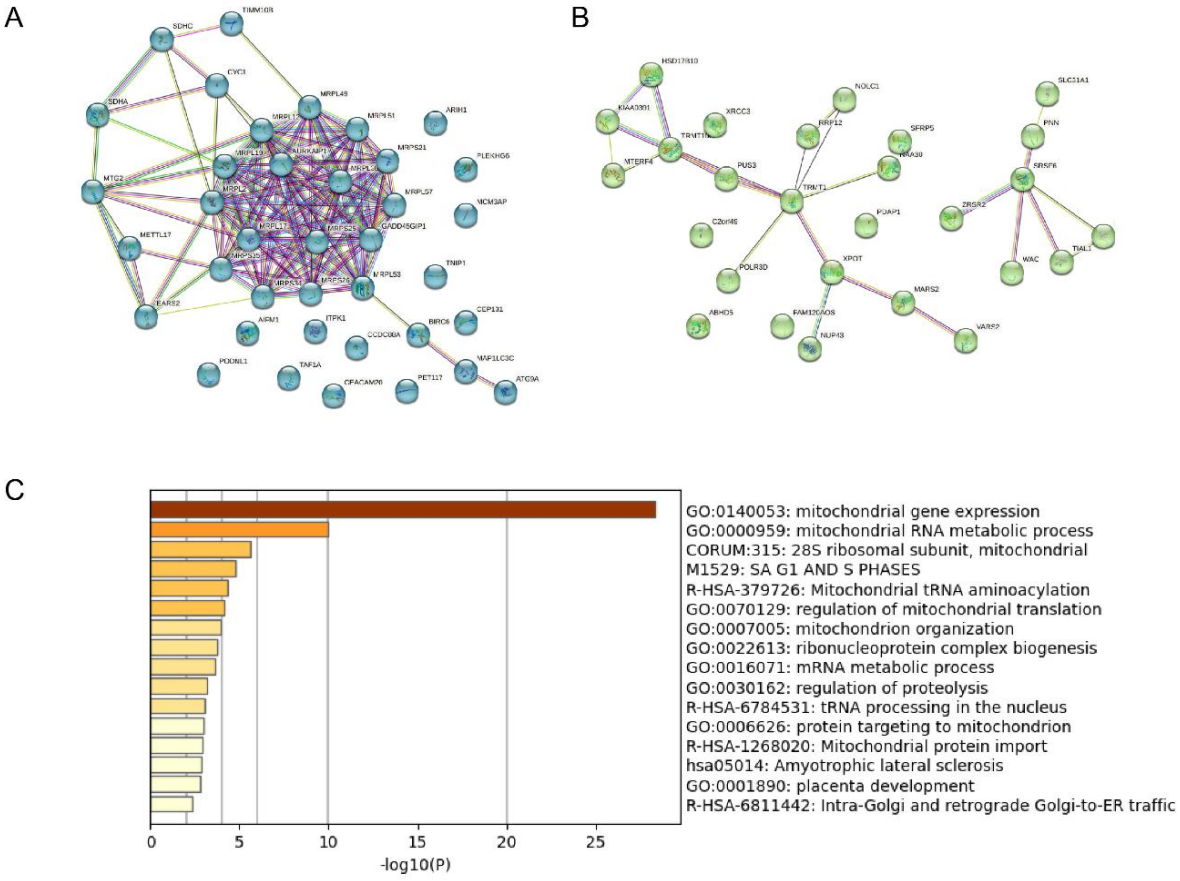


Figure 39. Protein network interaction clusters (A, B) and pathway enrichment analysis (C) obtained from the STRING and Metascape tools.

6.3 Discussion

The advent of targeted molecular therapies has advanced cancer treatment by circumventing the many secondary effects often associated with conventional therapies, such as chemotherapy or radiotherapy. This improvement in treatment has led to an overall enhancement in the quality of life for cancer patients. Nevertheless, targeted therapies almost always encounter the challenge of resistances, as evidenced by clinical cases of EGFR TKIs and the latest $KRAS^{G12C}$ inhibitors, where most of the patients suffer relapses within a short period of time, typically less than a year (de Langen et al., 2023; Hirsch et al., 2017). Studying the CDs that arise after the administration of a targeted therapy can help us to explain why these resistances occur, and how to prevent them. For this matter, CRISPR-screening is a very useful, high-throughput technique, which has been used, for example, to reveal CDs inherent to $KRAS^{G12C}$ inhibition with drugs, and therefore, potential co-therapies that augment the response to the treatment (Lou et al., 2019).

Our first idea was to perform a CRISPR-screening with the aim of finding CDs appearing in resistant cells to our CRISPR-KRAS therapy (see Chapter 5). However, *KRAS* is a broadly studied oncogene for which CRISPR-screenings have already been performed, and therefore CDs of its inhibition have already been found in other elements of the RAS pathway (*SHP2*, receptor tyrosine kinases), the mTOR pathway, and cell cycle regulators (*CDK4*, *CCND1*) (Kwan et al., 2022; Lou et al., 2019). Moreover, a KO clone for *KRAS* after administering our CRISPR strategy could not be obtained in any of the cell lines we tested in Chapter 5. Therefore, in order to learn this cutting-edge technique and acquire novel and valuable data, we decided to perform a CRISPR-screening to uncover the CDs of another less studied oncogene in lung cancer, *PKP1*, which we were researching in our group and had showed us promising results. We had previously generated two stable *PKP1*-KO clones from a prior publication (Martin-Padron et al., 2020), allowing us to use them along with the parental LUSC cell line SK-MES-1 as a control for the screening.

Our CRISPR-screening encountered some difficulties on the way. Since it was our first time performing a CRISPR-screening on our cell models, we had to first optimize the cell plating and growth to maintain the three conditions running in parallel for 21 days. This was challenging since the parental cells did not have the same growth rate as the KO clones. Moreover, we had issues with the KO-1 replicates, as cells did not grow enough to reach the necessary 40 million to ensure adequate coverage. Consequently, we were concerned that the quality of our screening might be compromised. However, the quality check analysis did not show any major problems with the KO-1 samples (A7 and A21). Instead, the main problem with our CRISPR-screening was the lack of correlation among some of the replicates (control replicates C and Z). It appeared that one of the replicates had undergone lower selective pressure, resulting in less sgRNA depletion. This difference was likely due to the second replicate cells (Z) not growing as much as

in the first replicate (C) during the 21-days selection. A lower number of cell divisions means less sgRNA depletion, and consequently, less screening power to discover potential hits. However, this did not entirely compromise the screening, as combining all the KO and control replicates still allowed us to obtain true hits. We tested all possible comparisons between individual and gathered samples, and we found that the top genes were not changing significantly. Therefore, we decided to focus on the comparison between all the control and all the KO samples, regardless of their selective pressure. This comparison would contain the most information and have the least number of false positives and clonal-specific hits.

Our results showed a strong dependence of *PKP1*-KO clones on several genes involved in mitochondrial ribosome structure, transcription, tRNA modification, and translation, suggesting an overall dependence on the mitochondrial activity when *PKP1* is depleted from cells. This is an unprecedented finding, as the role of plakophilins in tumorigenesis depends on their intracellular localization, and the function of *PKP1* in mitochondria is still unknown (Wolf & Hatzfeld, 2010). PKP1 is primarily found at the desmosomes along with the rest of plakophilins, where they have a structural and tumor suppressor role by promoting cell-to-cell adhesion and preventing migration, which might be the explanation for the less metastatic capacity of LUSC compared to LUAD (Martin-Padron et al., 2020). Once phosphorylated, PKP1 can be liberated from desmosomes and subsequently detected free in both the cytoplasm and the nucleus. The role of nuclear PKP1 remains unclear, but cytoplasmatic PKP1 has been proposed to have an oncogenic role by promoting the translation of some oncogenes (Wolf & Hatzfeld, 2010). This is exemplified by the feedforward regulation loop between *PKP1* and *MYC* that we discovered in LUSC (Boyero et al., 2022). Future research lines should aim to determine *PKP1*'s role in the mitochondria, to shed some light on the reasons behind the mitochondrial vulnerabilities appearing after *PKP1* knockout.

The mitochondrial genome is composed by 37 genes: 2 rRNAs, 22 tRNAs, and 13 proteins that are all part of the oxidative phosphorylation chain (Taanman, 1999). The rest of proteins necessary for mitochondrial transcription and translation are encoded by nuclear genes, and thus need to be imported from the cytosol. Mitochondrial ribosomes, or mitoribosomes, are formed by two subunits composed by the 12S and the 16S mitochondrial-encoded rRNAs, alongside 82 accessory mitochondrial ribosomal proteins (MRPs) that are encoded in the nucleus and imported via TIM/TOM into the mitochondria (De Silva et al., 2015; Lopez Sanchez et al., 2021). For the assembly of mitoribosomes, the two rRNAs (12S and 16S) are transcribed in the mitochondria, and then undergo several modifications (methylation, pseudouridylation…) (D'Souza & Minczuk, 2018) for which some nuclear transferase enzymes are also necessary to be imported from outside. Finally, inside the mitochondria, the modified rRNAs bind together with the MRPs to form the two subunits of mitoribosomes (Lopez Sanchez et al., 2021).

Many of the top scoring genes depleted in our CRISPR-screening were part of the MRP family, which include accessory mitoribosome assembly factors, such

as GTPases or RNA helicases. MRPs have a variety of roles apart from their structural and translation-related function in mitoribosomes, including modifying gene expression and genomic stability, and forming oxidative phosphorylation enzyme complexes that produce reactive oxygen species (ROS) (X. Lin et al., 2022). Some MRP signatures have even been related to cancer, as is the case for one of the top hits in our CRISPR-screening, *MRPS35*, which along four other MRPs appears upregulated in breast cancer and acts a biomarker for poor prognosis in patients (X. Lin et al., 2022). Other top hits include *PET117*, a cytochrome C oxidase chaperone involved in the electron transport chain, and the top candidate in terms of *JACKS* score, *MTERF4*. *MTERF4* belongs to the MTERF family of proteins, which are transcription factors involved in mitochondrial transcription (Cámara et al., 2011). However, *MTERF4* itself has a unique and key role in allowing the assembly of mitoribosomes, as it binds to the 16S mitochondrial rRNA, and promotes its methylation by the RNA methyltransferase NSUN4 (Yakubovskaya et al., 2012). Without this methylation, the mitoribosomal subunits are unable to assemble together, and therefore mitochondrial translation is heavily impaired, as shown by *MTERF4* knockout models in mice (Cámara et al., 2011).

Taken together, all these results seem to hint that *PKP*1-KO cells become more dependent on mitochondrial translation than *PKP1*-expressing cells, thus opening a new line of research in *PKP1*-inhibiting co-therapies for LUSC and possible mechanisms of resistance. Our group is currently studying therapeutic strategies against *PKP1* in LUSC. One approach could be using drugs to block its phosphorylation, either directly or indirectly, to prevent its release into the cytoplasm. Thus, the oncogenic role of cytoplasmatic PKP1 is prevented, while we stimulate the tumor suppressor role of desmosomal PKP1 (Wolf & Hatzfeld, 2010). Our findings about the mitochondrial vulnerabilities of *PKP1*-KO cells could be implemented into this future molecular therapy against *PKP1* in LUSC, for example, as a combination therapy to prevent resistances. Because of the bacterial origin of mitochondria, several antibiotics can be used to inhibit mitochondrial translation or transcription. Some long-known mitoribosome translation inhibitors include D-chloramphenicol, tigecycline, actinonin, and several other other antibiotics (tetracycline, lincomycin, erythromycin, neomycin…) (Borst & Grivell, 1971; Richter et al., 2015; L. Zhang et al., 2005). Some of these drugs have been proven to be effective as treatment against certain types of cancer highly dependent on mitochondrial activity, such as acute myoid leukemia (Škrtić et al., 2011). Moreover, for these cancers there are also compounds used for specific inhibition of mitochondrial transcription, such as 2-C-methyladenosine (Bralha et al., 2015).

As a future perspective to continue with this project, first, it would be necessary to individually validate the top hits of our CRISPR-screening, for example, by using siRNAs against the candidate gene and *PKP1* simultaneously in *PKP1*-expressing cells, and then checking for synergy effects. In addition, a more translational approach would be to use cytoplasmatic PKP1-inhibiting drugs in combination with some of the aforementioned mitochondrial translation inhibiting antibiotics, such as tigecycline, and measure the efficiency of the combination treatment for different LUSC models both *in vitro* and *in vivo*.

Conclusions / Conclusiones

Conclusions

1. The lncRNA *DLG2-AS1* is consistently downregulated in LUAD patients. Although its restoration in LUAD cell lines did not show any phenotypical effects in terms of cell viability, clonogenicity or migration, some other tumor suppressor roles for *DLG2-AS1* in LUAD cannot be ruled out.

2. *DLG2-AS1* is not a *cis*-regulator of its overlapping gene *DLG2*, but other regulatory functions in *trans* could be possible and up for further mechanistical studies.

3. *DLG2-AS1* displayed good potential as a biomarker for LUAD compared to other well-stablished lncRNA and protein-coding biomarkers, although its detection through non-invasive methods needs to be studied before its clinical application.

4. The miRNA *miR-133b* has a seed region mutation in a LUAD patient, which exerts a strong oncogenic effect when reproduced in LUAD cell lines by overexpression of a miRNA mimic.

5. Wildtype and mutant *miR-133b* have different sets of downregulated genes that are predicted targets of the miRNA and support an oncogenic effect of the mutation. However, it is still necessary to functionally validate these target predictions.

6. The designed sgRNAs for our CRISPR-KRAS therapy display high specificity and are able to discriminate between wildtype and mutant *KRAS*, inducing edition only in the $KRAS^{G12C}$ and $KRAS^{G12D}$ alleles.

7. Our CRISPR-KRAS therapy successfully impairs tumor cell viability in *KRAS*-mutant LUAD cell lines after administration with adenovirus containing HiFi-Cas9 and specific sgRNAs against $KRAS^{G12C}$ and $KRAS^{G12D}$.

8. A CRISPR-screening over *PKP1*-KO clones showed collateral dependencies related to mitochondrial transcription and translation. This could be exploited in a future *PKP1*-targeting therapy by using mitochondrial function inhibitor antibiotics as a co-therapy to improve the efficiency and prevent resistances.

Conclusiones

1. El lncRNA *DLG2-AS1* se encuentra consistentemente regulado a la baja en pacientes con LUAD. Aunque su restauración en líneas celulares de LUAD no mostró ningún efecto fenotípico en términos de viabilidad, clonogenicidad o migración celular, no se puede descartar algún otro papel como supresor tumoral de *DLG2-AS1* en LUAD.

2. *DLG2-AS1* no es un regulador en cis de su gen solapante *DLG2*, aunque otras funciones reguladoras en trans podrían ser posibles y sujeto de futuros estudios mecanísticos.

3. *DLG2-AS1* mostró un buen potencial como biomarcador para LUAD comparado con otros lncRNA y genes codificantes de proteína ampliamente validados como biomarcadores, aunque su detección por técnicas no invasivas debe ser estudiada antes de su aplicación clínica.

4. El miRNA *miR-133b* posee una mutación en su región semilla en un paciente de LUAD, la cual ejerce un fuerte efecto oncogénico cuando es reproducida por sobreexpresión de un miRNA mímico en líneas celulares de LUAD.

5. Las versiones silvestre y mutante de *miR-133b* generan diferentes conjuntos de genes regulados a la baja que son dianas predichas del miRNA y respaldan nuestra hipótesis sobre el potencial oncogénico de la mutación. Sin embargo, es aún necesaria la validación funcional de estas predicciones de dianas.

6. Los sgRNA diseñados para nuestra terapia CRISPR-KRAS muestran alta especificidad y son capaces de discriminar entre las versiones silvestre y mutante de *KRAS*, editando solamente los alelos $KRAS^{G12C}$ y $KRAS^{G12D}$.

7. Nuestra terapia CRISPR-KRAS reduce con éxito la viabilidad celular en líneas celulares de LUAD tras la administración de adenovirus que contienen la HiFi-Cas9 y sgRNA específicos contra $KRAS^{G12C}$ y $KRAS^{G12D}$.

8. El cribado CRISPR de clones KO para *PKP1* mostró dependencias colaterales relacionadas con la transcripción y traducción mitocondriales. Esto podría ser aprovechado en una futura terapia molecular dirigida contra *PKP1* usando antibióticos inhibidores de la función mitocondrial como co-terapia para mejorar la eficiencia y prevenir resistencias.

Supplementary material

Supplementary material

Supplementary Table 1. List of used oligonucleotides and their sequences.

Oligo name	5'-3' Sequence	Tm (°C)	Use
AAAA-EcoRI-DLG2-AS1	AAAAGAATTCTTAACTTGTTAATCA	56.2	Cloning
DICER-1_qFw	GAAGAATCAGCCTCGCAAC	62.6	qPCR
DICER-1_qRv	TGTGGGCAAATCAAAACGAAC	66.7	qPCR
DLG2_qFw	GAAGACCTCATTCTTTCCTATG	58.3	qPCR
DLG2_qRv	CGGCTTCTATAAACTTGTGC	59	qPCR
DLG2-AS1_qFw	ATCCGGATGTGAGGTTATAAT	58.7	qPCR
DLG2-AS1_qRv	AATCCAGATCCCAAGACTTC	59.5	qPCR
EGFR_ex3_qFw	GGAGGTGGCTGGTTATGTCC	65.6	qPCR
EGFR_ex3_qRv	TTCTCATGGGCAGCTCCTTC	66.7	qPCR
GAPDH_qFw	GAAGGTGAAGGTCGGAGTC	61.3	qPCR
GAPDH_qRv	GAAGATGGTGATGGGATTTC	60.8	qPCR
KRAS_Fw	CACGTCTGCAGTCAACTGGA	65.4	PCR + Sequencing
KRAS_Rv	TCCGGTAGTTGTAGGTTCTCT	59.3	PCR
miR133b_gen_Fw	TCGTTGTTCATGGGTATTGGTTT	65.9	PCR
miR133b_gen_Rv	TTGCTCACTTTTGACTCCAG	60.6	PCR
miR133b_mut_Fw	TTCCCCTTCAACCAGCTA	60.8	qPCR
miR133b_seq_Fw	GCTTGAGACACACCAAGATA	57.8	Sequencing
miR133b_wt_Fw	GTCCCCTTCAACCAGCTA	60.3	qPCR
pri-miR133b_Fw	TCAGAAGAAAGATGCCCCCTG	67.5	PCR + Sequencing
pri-miR133b_Rv	TGCTGTAGCTGGTTGAAGGG	65.3	PCR
SNORD44_qFw	GTCTTAATTAGCTCTAACTGACT	52.3	qPCR
TTTT-XbaI-DLG2-AS1	TTTTTCTAGACCAGATGGTCAGTGA	64.7	Cloning
U1_snRNA_qFw	GGGAGATACCATGATCACGAAGGT	68	qPCR
U1_snRNA_qRv	CCACAAATTATGCAGTCGAGTTTCCC	70.6	qPCR
Universal_Reverse_qPCR	GCGAGCACAGAATTAATACGAC	62.5	qPCR

Supplementary Table 2. Sequences of miR-133b mimics.

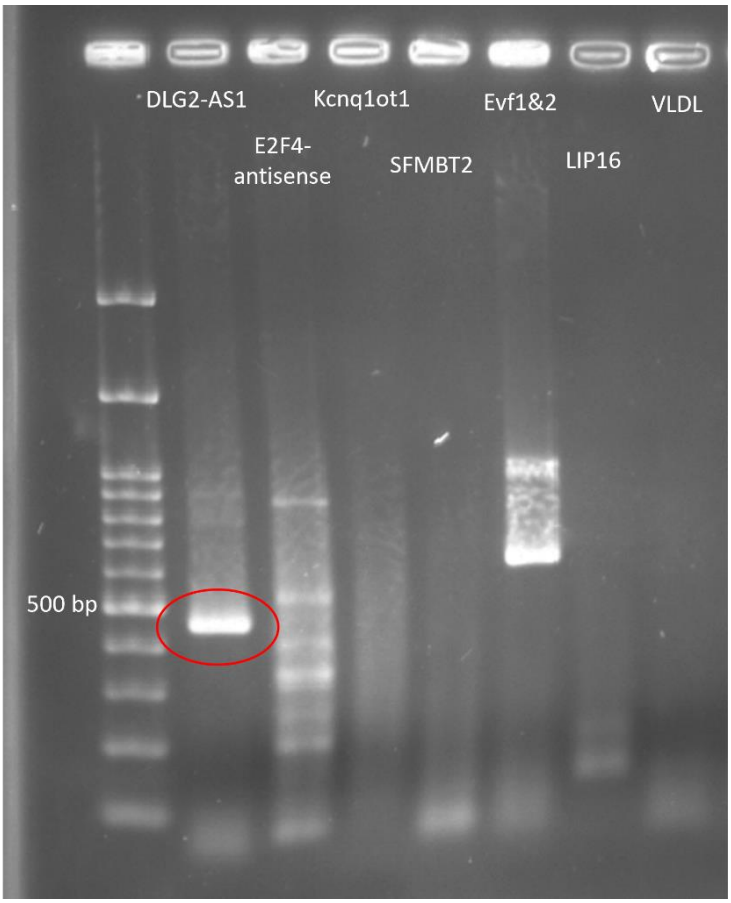
Mimic	Strand	Sequence (5'-3')
miR-133b^{WT}	Sense	UUUGGUCCCCUUCAACCAGCUA
	Antisense	GCUGGUUGAAGGGGACCAAAUU
miR-133b^{MUT}	Sense	UUUGUUCCCCUUCAACCAGCUA
	Antisense	GCUGGUUGAAGGGGAACAAAUU
Scramble (negative control)	Sense	UUCUCCGAACGUGUCACGU
	Antisense	ACGUGACACGUUCGGAGAA

Supplementary Table 3. List of used primary and secondary antibodies.

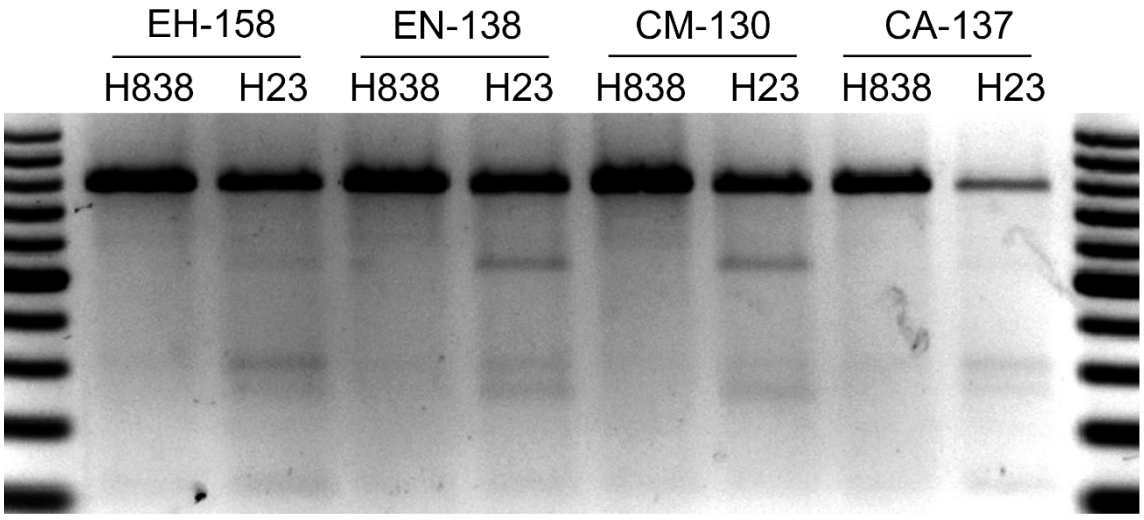
Protein	Supplier (#reference)	Dilution
AGO2	Sigma-Aldrich (#SAB4200085)	1:1500
Cas9	Cell Signaling (#14697)	1:1000
DICER1	Cell Signaling (#5362T)	1:500
DLG2	Alomone Labs (#APZ-002)	1:1000
EGFR	Cell Signaling (#4267T)	1:5000
KRAS	Abcam (#ab180772)	1:200
HRP-Mouse IgG	Dako (#P0447)	1:5000
ph-AKT	Cell Signaling (#4060T)	1:1000
ph-ERK	Cell Signaling (#4370T)	1:1000
ph-p70S6	Santa Cruz (#SC-8416)	1:1000
HRP-Rabbit IgG	Dako (#P0448)	1:2000
α-TUBULIN	Santa Cruz (#sc-23948)	1:10000
β-ACTIN	Sigma-Aldrich (#A5441)	1:20000

Supplementary Table 4. List of used cell lines and their corresponding culture media.

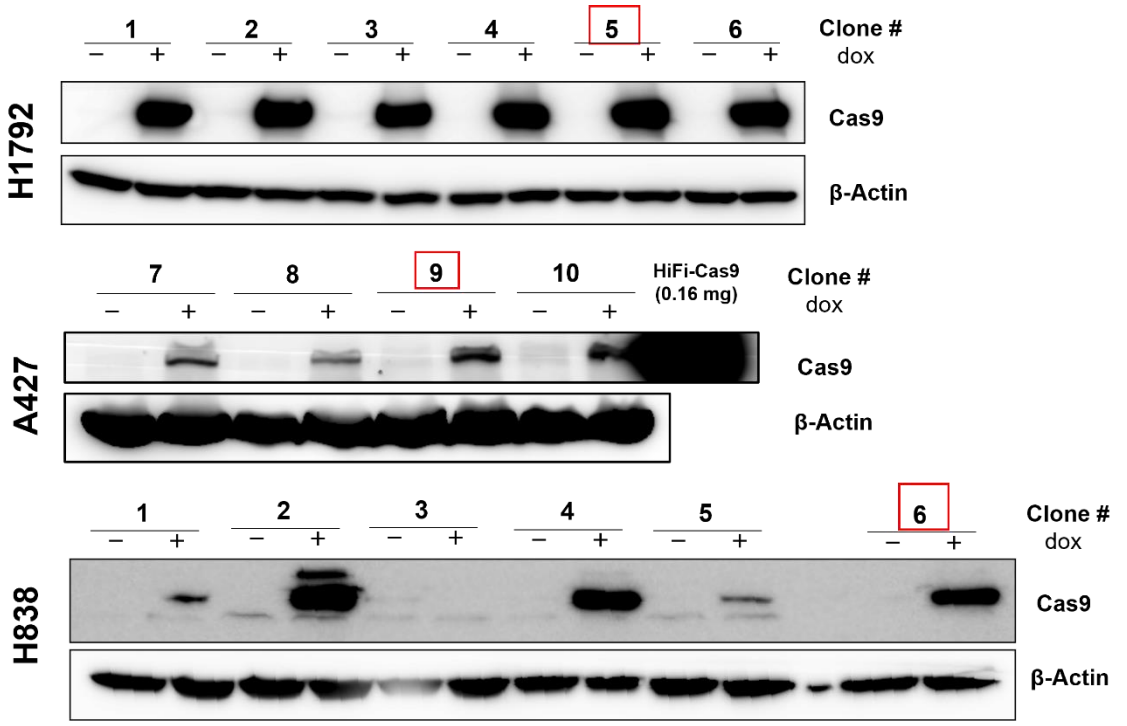
Cell line	Cell type	Culture medium
A427	Lung adenocarcinoma	RPMI 1640
A549	Lung adenocarcinoma	DMEM
HEK-293T	Human embryonic kidney	DMEM
K562	Chronic myeloid leukemia	RPMI 1640
MEF	Mouse embryonic fibroblasts	DMEM
NCI-H1299	Lung adenocarcinoma	RPMI 1640
NCI-H1650	Lung adenocarcinoma	RPMI 1640
NCI-H1792	Lung adenocarcinoma	RPMI 1640
NCI-H1944	Lung adenocarcinoma	RPMI 1640
NCI-H1975	Lung adenocarcinoma	RPMI 1640
NCI-H2122	Lung adenocarcinoma	RPMI 1640
NCI-H2126	Lung adenocarcinoma	RPMI 1640
NCI-H23	Lung adenocarcinoma	RPMI 1640
NCI-H358	Lung adenocarcinoma	RPMI 1640
NCI-H838	Lung adenocarcinoma	RPMI 1640
SK-MES-1	Lung squamous cell carcinoma	DMEM



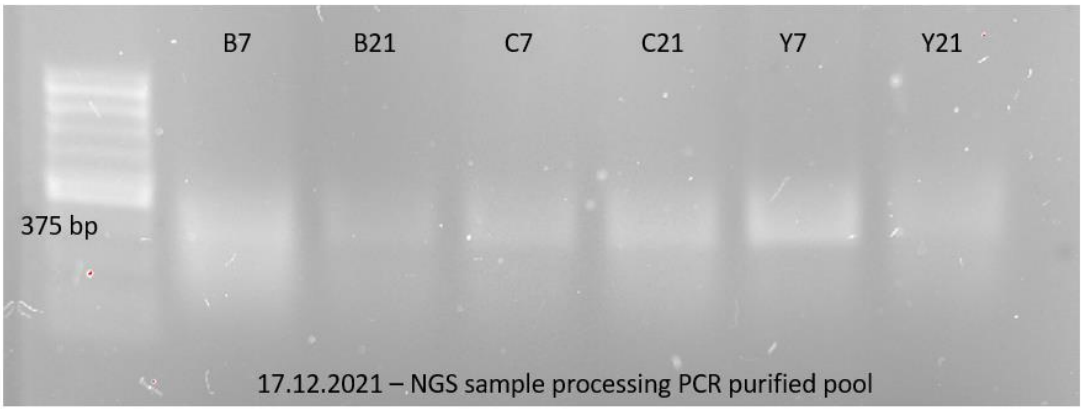
Supplementary Figure 1. Agarose electrophoresis gel of qPCR products from one of the patient samples (33S), showing that DLG2-AS1 is the lncRNA candidate which presents a more clear, unique band.



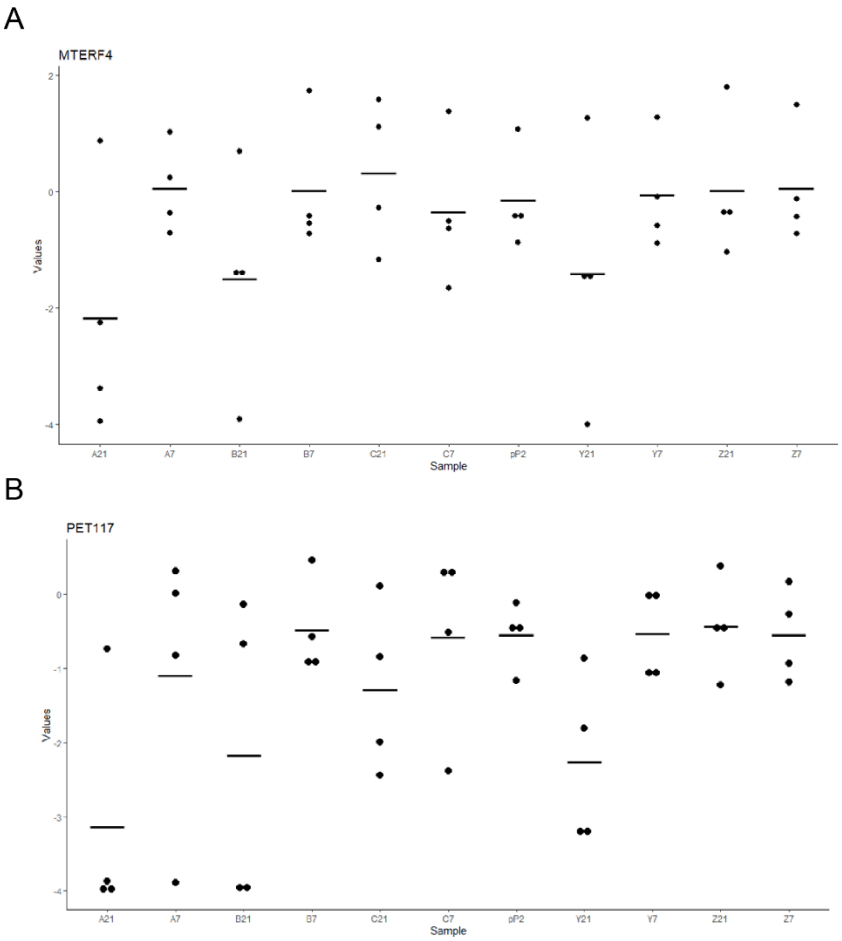
Supplementary Figure 2. T7-endonuclease assay testing different HiFiCas9+sgRNA RNP nucleofection programs (EH-158, EN-138, CM-130, and CA-137) in two different LUAD cell lines, H838 ($KRAS^{WT}$), and H23 ($KRAS^{G12C}$). Edition can be observed (double bands) and is specific in the KRAS mutant cell line for all programs.



Supplementary Figure 3. Western blot results from H1792, A427, and H838 clones expressing doxycycline-inducible HiFi-Cas9. In red squares, the clones without leaking selected for the next lentiviral infection.



Supplementary Figure 4. Agarose gel electrophoresis of the pool of PCR reactions before NGS.



Supplementary Figure 5. Individual dot plot of JACKS depletion score of two of the top scoring sgRNAs, MTERF4 (mitochondrial termination factor 4), and PET117 (cytochrome C oxidase chaperone). Each dot represents one of the four sgRNAs that the Brunello library has for every gene.

References

References

References

Ahumada-Ayala, M., Aguilar-López, R., González-Stoylov, N., Palacio-Sosa, E., Cervantes-Barragán, D. E., & Fernández-Hernández, L. (2023). Editing the Human Genome with CRISPR/Cas: A Review of its Molecular Basis, Current Clinical Applications, and Bioethical Implications. *Revista de Investigacion Clinica; Organo Del Hospital de Enfermedades de La Nutricion*, *75*(1). https://doi.org/10.24875/RIC.22000252

Akbari Moqadam, F., Pieters, R., & Den Boer, M. L. (2012). The hunting of targets: challenge in miRNA research. *Leukemia*, *27*(1), 16–23. https://doi.org/10.1038/leu.2012.179

Aldoghachi, A. F., Baharudin, A., Ahmad, U., Chan, S. C., Ong, T. A., Yunus, R., Razack, A. H., Yusoff, K., & Veerakumarasivam, A. (2019). Evaluation of CERS2 Gene as a Potential Biomarker for Bladder Cancer. *Disease Markers*, *2019*. https://doi.org/10.1155/2019/3875147

Allen, F., Behan, F., Khodak, A., Iorio, F., Yusa, K., Garnett, M., & Parts, L. (2019). Jacks: Joint analysis of CRISPR/Cas9 knockout screens. *Genome Research*, *29*(3), 464–471. https://doi.org/10.1101/GR.238923.118/-/DC1

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, *513*(7519), 569–573. https://doi.org/10.1038/nature13579

Andrades, Á. (2022). *Bioinformatic approaches for the discovery of non-coding alterations in cancer* [Dissertation]. University of Granada.

Angulo, B., Suarez-Gauthier, A., Lopez-Rios, F., Medina, P. P., Conde, E., Tang, M., Soler, G., Lopez-Encuentra, A., Cigudosa, J. C., & Sanchez-Cespedes, M. (2008). Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. *The Journal of Pathology*, *214*(3), 347–356. https://doi.org/10.1002/PATH.2267

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Arenas, A. M., Andrades, A., Patiño-Mercau, J. R., Sanjuan-Hidalgo, J., Cuadros, M., García, D. J., Peinado, P., Rodriguez, M. I., Baliñas-Gavira, C., Álvarez-Perez, J. C., & Medina, P. P. (2022). Chapter 12 - Opportunities of miRNAs in cancer therapeutics. In M. Negrini, G. A. Calin, & C. M. Croce (Eds.), *MicroRNA in Human Malignancies* (pp. 153–164). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-822287-4.00015-3

Arenas, A. M., Cuadros, M., Andrades, A., García, D. J., Coira, I. F., Rodríguez, M. I., Baliñas-Gavira, C., Peinado, P., Álvarez-Pérez, J. C., & Medina, P. P. (2020). LncRNA DLG2-AS1 as a Novel Biomarker in Lung Adenocarcinoma. *Cancers*, *12*(8), 1–10. https://doi.org/10.3390/CANCERS12082080

References

Awad, M. M., Liu, S., Rybkin, I. I., Arbour, K. C., Dilly, J., Zhu, V. W., Johnson, M. L., Heist, R. S., Patil, T., Riely, G. J., Jacobson, J. O., Yang, X., Persky, N. S., Root, D. E., Lowder, K. E., Feng, H., Zhang, S. S., Haigis, K. M., Hung, Y. P., … Aguirre, A. J. (2021). Acquired Resistance to KRAS G12C Inhibition in Cancer . *New England Journal of Medicine*, *384*(25), 2382–2393. https://doi.org/10.1056/NEJMOA2105281

Azangou-Khyavy, M., Ghasemi, M., Khanali, J., Boroomand-Saboor, M., Jamalkhah, M., Soleimani, M., & Kiani, J. (2020). CRISPR/Cas: From Tumor Gene Editing to T Cell-Based Immunotherapy of Cancer. *Frontiers in Immunology*, *11*, 2062. https://doi.org/10.3389/FIMMU.2020.02062/BIBTEX

Azzoli, C. G., Baker, S., Temin, S., Pao, W., Aliff, T., Brahmer, J., Johnson, D. H., Laskin, J. L., Masters, G., Milton, D., Nordquist, L., Pfister, D. G., Piantadosi, S., Schiller, J. H., Smith, R., Smith, T. J., Strawn, J. R., Trent, D., Giaccone, G., & Kimmel, S. (2009). American Society of Clinical Oncology Clinical Practice Guideline Update on Chemotherapy for Stage IV Non-Small-Cell Lung Cancer. *J Clin Oncol*, *27*, 6251–6266. https://doi.org/10.1200/JCO.2009.23.5622

Babar, I. A., Cheng, C. J., Booth, C. J., Liang, X., Weidhaas, J. B., Saltzman, W. M., & Slack, F. J. (2012). Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), E1695–E1704. https://doi.org/10.1073/PNAS.1201516109

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64–71. https://doi.org/10.1038/nature07242

Bandrés, E., Cubedo, E., Agirre, X., Malumbres, R., Zárate, R., Ramirez, N., Abajo, A., Navarro, A., Moreno, I., Monzó, M., & García-Foncillas, J. (2006). Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Molecular Cancer*, *5*, 29. https://doi.org/10.1186/1476-4598-5-29

Bandyopadhyay, A., Kancharla, N., Javalkote, V. S., Dasgupta, S., & Brutnell, T. P. (2020). CRISPR-Cas12a (Cpf1): A Versatile Tool in the Plant Genome Editing Tool Box for Agricultural Advancement. *Frontiers in Plant Science*, *11*, 1589. https://doi.org/10.3389/FPLS.2020.584151

Barta, J. A., Powell, C. A., & Wisnivesky, J. P. (2019). Global Epidemiology of Lung Cancer. *Annals of Global Health*, *85*(1). https://doi.org/10.5334/AOGH.2419

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, *116*(2), 281–297. https://doi.org/10.1016/S0092-8674(04)00045-5

Bender, G., Fahrioglu Yamaci, R., & Taneri, B. (2021). CRISPR and KRAS: a match yet to be made. *Journal of Biomedical Science*, *28*(1). https://doi.org/10.1186/S12929-021-00772-0

Bhat, S. A., Ahmad, S. M., Mumtaz, P. T., Malik, A. A., Dar, M. A., Urwat, U., Shah, R. A., & Ganai, N. A. (2016). Long non-coding RNAs: Mechanism of action and functional utility. *Non-Coding RNA Research*, *1*(1), 43. https://doi.org/10.1016/J.NCRNA.2016.11.002

Bica-Pop, C., Cojocneanu-Petric, R., Lorand Magdo, ·, Raduly, L., Gulei, D., & Berindan-Neagoe, I. (2018). Overview upon miR-21 in lung cancer: focus on NSCLC. *Cellular and*

Molecular Life Sciences, *75*(19), 3539–3551. https://doi.org/10.1007/s00018-018-2877-x

Boettger, T., Wüst, S., Nolte, H., & Braun, T. (2014). The miR-206/133b cluster is dispensable for development, survival and regeneration of skeletal muscle. *Skeletal Muscle*, *4*(1). https://doi.org/10.1186/S13395-014-0023-5

Bolha, L., Ravnik-Glavač, M., & Glavač, D. (2017). Long Noncoding RNAs as Biomarkers in Cancer. *Disease Markers*, *2017*. https://doi.org/10.1155/2017/7243968

Borst, P., & Grivell, L. A. (1971). Mitochondrial ribosomes. *FEBS Letters*, *13*(2), 73–88. https://doi.org/10.1016/0014-5793(71)80204-1

Boyero, L., Martin-Padron, J., Fárez-Vidal, M. E., Rodriguez, M. I., Andrades, Á., Peinado, P., Arenas, A. M., Ritoré-Salazar, F., Alvarez-Perez, J. C., Cuadros, M., & Medina, P. P. (2022). PKP1 and MYC create a feedforward loop linking transcription and translation in squamous cell lung cancer. *Cellular Oncology (Dordrecht)*, *45*(2), 323–332. https://doi.org/10.1007/S13402-022-00660-1

Bralha, F. N., Liyanage, S. U., Hurren, R., Wang, X., Son, M. H., Fung, T. A., Chingcuanco, F. B., Tung, A. Y. W., Andreazza, A. C., Psarianos, P., Schimmer, A. D., Salmena, L., & Laposa, R. R. (2015). Targeting mitochondrial RNA polymerase in acute myeloid leukemia. *Oncotarget*, *6*(35), 37216. https://doi.org/10.18632/ONCOTARGET.6129

Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W., & Pasquinelli, A. E. (2016). Pairing Beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular Cell*, *64*(2), 320. https://doi.org/10.1016/J.MOLCEL.2016.09.004

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., & Croce, C. M. (2002). Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15524–15529. https://doi.org/10.1073/PNAS.242606799

Cámara, Y., Asin-Cayuela, J., Park, C. B., Metodiev, M. D., Shi, Y., Ruzzenente, B., Kukat, C., Habermann, B., Wibom, R., Hultenby, K., Franz, T., Erdjument-Bromage, H., Tempst, P., Hallberg, B. M., Gustafsson, C. M., & Larsson, N. G. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metabolism*, *13*(5), 527–539. https://doi.org/10.1016/j.cmet.2011.04.002

Campbell, P. J., Getz, G., Korbel, J. O., Stuart, J. M., Jennings, J. L., Stein, L. D., Perry, M. D., Nahal-Bose, H. K., Ouellette, B. F. F., Li, C. H., Rheinbay, E., Nielsen, G. P., Sgroi, D. C., Wu, C. L., Faquin, W. C., Deshpande, V., Boutros, P. C., Lazar, A. J., Hoadley, K. A., … Zhang, J. (2020). Pan-cancer analysis of whole genomes. *Nature*, *578*(7793), 82–93. https://doi.org/10.1038/s41586-020-1969-6

Canon, J., Rex, K., Saiki, A. Y., Mohr, C., Cooke, K., Bagal, D., Gaida, K., Holt, T., Knutson, C. G., Koppada, N., Lanman, B. A., Werner, J., Rapaport, A. S., San Miguel, T., Ortiz, R., Osgood, T., Sun, J. R., Zhu, X., McCarter, J. D., … Lipford, J. R. (2019). The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature*, *575*(7781), 217–223. https://doi.org/10.1038/S41586-019-1694-1

Cao, Y., Chu, C., Li, X., Gu, S., Zou, Q., & Jin, Y. (2021). RNA-binding protein QKI suppresses breast cancer via RASA1/MAPK signaling pathway. *Annals of Translational Medicine*, *9*(2), 104–104. https://doi.org/10.21037/ATM-20-4859

Caporali, S., Calabrese, C., Minieri, M., Pieri, M., Tarantino, U., Marini, M., D'ottavio, S., Angeletti, S., Mauriello, A., Cortese, C., Bernardini, S., & Terrinoni, A. (2021). The miR-133a, TPM4 and TAp63γ Role in Myocyte Differentiation Microfilament Remodelling and Colon Cancer Progression. *International Journal of Molecular Sciences*, *22*(18). https://doi.org/10.3390/IJMS22189818

Cawthorne, C., Swindell, R., Stratford, I. J., Dive, C., & Welman, A. (2007). Comparison of Doxycycline Delivery Methods for Tet-Inducible Gene Expression in a Subcutaneous Xenograft Model. *Journal of Biomolecular Techniques : JBT*, *18*(2), 120. /pmc/articles/PMC2062538/

Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell*, *47*(4), 497–510. https://doi.org/10.1016/J.MOLCEL.2012.07.029

Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., Vakulskas, C. A., Collingwood, M. A., Zhang, L., Bode, N. M., Behlke, M. A., Dejene, B., Cieniewicz, B., Romano, R., Lesch, B. J., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Weinberg, K. I., & Porteus, M. H. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine*, *25*(2), 249. https://doi.org/10.1038/S41591-018-0326-X

Chen, J. F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., Conlon, F. L., & Wang, D. Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics*, *38*(2), 228. https://doi.org/10.1038/NG1725

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, *360*(6387), 436–439. https://doi.org/10.1126/SCIENCE.AAR6245

Chen, S., Ben, S., Xin, J., Li, S., Zheng, R., Wang, H., Fan, L., Du, M., Zhang, Z., & Wang, M. (2021). The biogenesis and biological function of PIWI-interacting RNA in cancer. *Journal of Hematology & Oncology*, *14*(1), 93. https://doi.org/10.1186/S13045-021-01104-3

Chen, S., Shen, Z., Gao, L., Yu, S., Zhang, P., Han, Z., & Kang, M. (2021). TPM3 mediates epithelial-mesenchymal transition in esophageal cancer via MMP2/MMP9. *Annals of Translational Medicine*, *9*(16), 1338–1338. https://doi.org/10.21037/ATM-21-4043

Cheng, T. Y. D., Cramb, S. M., Baade, P. D., Youlden, D. R., Nwogu, C., & Reid, M. E. (2016). The international epidemiology of lung cancer: Latest trends, disparities, and tumor characteristics. *Journal of Thoracic Oncology*, *11*(10), 1653–1671. https://doi.org/10.1016/J.JTHO.2016.05.021

Chira, S., Gulei, D., Hajitou, A., & Berindan-Neagoe, I. (2018). Restoring the p53 'Guardian' Phenotype in p53-Deficient Tumor Cells with CRISPR/Cas9. *Trends in Biotechnology*, *36*(7), 653–660. https://doi.org/10.1016/j.tibtech.2018.01.014

Chiu, C. C., Lin, C. Y., Lee, L. Y., Chen, Y. J., Lu, Y. C., Wang, H. M., Liao, C. T., Chang, J. T. C., & Cheng, A. J. (2011). Molecular chaperones as a common set of proteins that regulate the invasion phenotype of head and neck cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *17*(14), 4629–4641. https://doi.org/10.1158/1078-0432.CCR-10-2107

Cimmino, A., Calin, G. A., Fabbri, M., Iorio, M. V, Ferracin, M., Shimizu, M., Wojcik, S. E., Aqeilan, R. I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.-G., Kipps, T. J., Negrini, M., & Croce, C. M. (2005). miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13944–13949. www.pnas.orgcgidoi10.1073pnas.0506654102

Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., Chmielecki, J., Beer, D. G., Cope, L., Creighton, C. J., Danilova, L., Ding, L., Getz, G., Hammerman, P. S., Hayes, D. N., Hernandez, B., Herman, J. G., Heymach, J. V., Jurisica, I., Kucherlapati, R., … Cheney, R. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, *511*(7511), 543–550. https://doi.org/10.1038/nature13385

Conte, I., Hadfield, K. D., Barbato, S., Carrella, S., Pizzo, M., Bhat, R. S., Carissimo, A., Karali, M., Porter, L. F., Urquhart, J., Hateley, S., O'Sullivan, J., Manson, F. D. C., Neuhauss, S. C. F., Banfi, S., & Black, G. C. M. (2015). MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(25), E3236–E3245. https://doi.org/10.1073/PNAS.1401464112

Crawford, M., Batte, K., Yu, L., Wu, X., Nuovo, G. J., Marsh, C. B., Otterson, G. A., & Nana-Sinkam, S. P. (2009). microRNA 133B targets prosurvival molecules MCL-1 and BCL2L2 in lung cancer. *Biochemical and Biophysical Research Communications*, *388*(3), 483. https://doi.org/10.1016/J.BBRC.2009.07.143

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, *12*, 138–163.

Dai, X., Gakidou, E., & Lopez, A. D. (2022). Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. *Tobacco Control*, *31*, 129–137. https://doi.org/10.1136/tobaccocontrol-2021-056535

de Langen, A. J., Johnson, M. L., Mazieres, J., Dingemans, A.-M. C., Mountzios, G., Pless, M., Wolf, J., Schuler, M., Lena, H., Skoulidis, F., Yoneshima, Y., Kim, S.-W., Linardou, H., Novello, S., van der Wekken, A. J., Chen, Y., Peters, S., Felip, E., Solomon, B. J., … CodeBreaK 200 Investigators. (2023). Sotorasib versus docetaxel for previously treated non-small-cell lung cancer with KRASG12C mutation: a randomised, open-label, phase 3 trial. *The Lancet*. https://doi.org/10.1016/S0140-6736(23)00221-0

de Miguel, F. J., Pajares, M. J., Martínez-Terroba, E., Ajona, D., Morales, X., Sharma, R. D., Pardo, F. J., Rouzaut, A., Rubio, A., Montuenga, L. M., & Pio, R. (2016). A large-scale analysis of alternative splicing reveals a key role of QKI in lung cancer. *Molecular Oncology*, *10*(9), 1437–1449. https://doi.org/10.1016/J.MOLONC.2016.08.001

De Silva, D., Tu, Y. T., Amunts, A., Fontanesi, F., & Barrientos, A. (2015). Mitochondrial ribosome assembly in health and disease. *Cell Cycle (Georgetown, Tex.)*, *14*(14), 2226–2250. https://doi.org/10.1080/15384101.2015.1053672

Deng, Z., Chen, J. F., & Wang, D. Z. (2011). Transgenic overexpression of miR-133a in skeletal muscle. *BMC Musculoskeletal Disorders*, *12*. https://doi.org/10.1186/1471-2474-12-115

Dong, H., Zhou, J., Cheng, Y., Wang, M., Wang, S., & Xu, H. (2021). Biogenesis, Functions, and Role of CircRNAs in Lung Cancer. *Cancer Management and Research*, *13*, 6651–6671. https://doi.org/10.2147/CMAR.S324812

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213). https://doi.org/10.1126/SCIENCE.1258096

D'Souza, A. R., & Minczuk, M. (2018). Mitochondrial transcription and translation: overview. *Essays in Biochemistry*, *62*(3), 309. https://doi.org/10.1042/EBC20170102

Enache, O. M., Rendo, V., Abdusamad, M., Lam, D., Davison, D., Pal, S., Currimjee, N., Hess, J., Pantel, S., Nag, A., Thorner, A. R., Doench, J. G., Vazquez, F., Beroukhim, R., Golub, T. R., & Ben-David, U. (2020). Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nature Genetics*, *52*(7), 662–668. https://doi.org/10.1038/S41588-020-0623-4

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription, and splicing. *Nature*, *539*(7629), 452. https://doi.org/10.1038/NATURE20149

Fan, S. H., Wang, Y. Y., Lu, J., Zheng, Y. L., Wu, D. M., Zhang, Z. F., Shan, Q., Hu, B., Li, M. Q., & Cheng, W. (2015). CERS2 suppresses tumor cell invasion and is associated with decreased V-ATPase and MMP-2/MMP-9 activities in breast cancer. *Journal of Cellular Biochemistry*, *116*(4), 502–513. https://doi.org/10.1002/JCB.24978

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Biryukov, S., Brauer, M., Cercy, K., Charlson, F. J., Cohen, A. J., Dandona, L., Estep, K., Ferrari, A. J., Frostad, J. J., Fullman, N., Godwin, W. W., Griswold, M., Hay, S. I., Kyu, H. H., Larson, H. J., Lim, S. S., … Zhu, J. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1659–1724. https://doi.org/10.1016/S0140-6736(16)31679-8

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., de Montalembert, M., Rondelli, D., … Corbacioglu, S. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine*, *384*(3), 252–260. https://doi.org/10.1056/NEJMOA2031054

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Sala, S. C., Chrast, J., Cunningham, F., Domenico, T. Di, Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*, 767. https://doi.org/10.1093/nar/gky955

Gao, Q., Ouyang, W., Kang, B., Han, X., Xiong, Y., Ding, R., Li, Y., Wang, F., Huang, L., Chen, L., Wang, D., Dong, X., Zhang, Z., Li, Y., Ze, B., Hou, Y., Yang, H., Ma, Y., Gu, Y., & Chao, C. C. (2020). Selective targeting of the oncogenic KRAS G12S mutant allele by

CRISPR/Cas9 induces efficient tumor regression. *Theranostics*, *10*(11), 5137–5153. https://doi.org/10.7150/thno.42325

Garraway, L. A., & Jänne, P. A. (2012). Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discovery*, *2*(3), 214–226. https://doi.org/10.1158/2159-8290.CD-12-0012

Gaviraghi, M., Vivori, C., & Tonon, G. (2019). How Cancer Exploits Ribosomal RNA Biogenesis: A Journey beyond the Boundaries of rRNA Transcription. *Cells*, *8*(9). https://doi.org/10.3390/CELLS8091098

Gerber, A. P., Luschnig, S., Krasnow, M. A., Brown, P. O., & Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(12), 4487–4492. https://doi.org/10.1073/PNAS.0509260103

Giaccone, G. (2005). Epidermal growth factor receptor inhibitors in the treatment of non-small-cell lung cancer. *Journal of Clinical Oncology*, *23*(14), 3235–3242. https://doi.org/10.1200/JCO.2005.08.409

Gillmore, J. D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M. L., Seitzer, J., O'Connell, D., Walsh, K. R., Wood, K., Phillips, J., Xu, Y., Amaral, A., Boyd, A. P., Cehelsky, J. E., McKee, M. D., Schiermeier, A., Harari, O., Murphy, A., … Lebwohl, D. (2021). CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *New England Journal of Medicine*, *385*(6), 493–502. https://doi.org/10.1056/NEJMOA2107454

Ginn, L., Shi, L., Montagna, M. La, & Garofalo, M. (2020). LncRNAs in Non-Small-Cell Lung Cancer. *Non-Coding RNA*, *6, 25*. https://doi.org/10.3390/ncrna6030025

Goebel, L., Müller, M. P., Goody, R. S., & Rauh, D. (2020). KRasG12C inhibitors in clinical trials: a short historical perspective. *RSC Medicinal Chemistry*, *11*(7), 760. https://doi.org/10.1039/D0MD00096E

Gourd, E. (2020). Overall survival with osimertinib in untreated NSCLC. *The Lancet. Oncology*, *21*(1), e15. https://doi.org/10.1016/S1470-2045(19)30778-8

Guglas, K., Kozłowska-Masłoń, J., Kolenda, T., Paszkowska, A., Teresiak, A., Bliźniak, R., & Lamperska, K. (2022). Midsize noncoding RNAs in cancers: a new division that clarifies the world of noncoding RNA or an unnecessary chaos? *Reports of Practical Oncology and Radiotherapy*, *27*(6), 1077. https://doi.org/10.5603/RPOR.A2022.0123

Guo, F., Liu, Y., Li, Y., & Li, G. (2010). Inhibition of ADP-ribosylation factor-like 6 interacting protein 1 suppresses proliferation and reduces tumor cell invasion in CaSki human cervical cancer cells. *Molecular Biology Reports*, *37*(8), 3819–3825. https://doi.org/10.1007/S11033-010-0037-Y

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Tsai, M. C., Hung, T., Argani, P., Rinn, J. L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R. B., Van De Vijver, M. J., Sukumar, S., & Chang, H. Y. (2010). Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, *464*(7291), 1071. https://doi.org/10.1038/NATURE08975

Gutschner, T., & Diederichs, S. (2012). The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biology*, *9*(6), 703. https://doi.org/10.4161/RNA.20481

Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Groß, M., Zörnig, M., MacLeod, A. R., Spector, D. L., & Diederichs, S. (2013). The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research*, *73*(3), 1180. https://doi.org/10.1158/0008-5472.CAN-12-2850

Haase, D., Cui, T., Yang, L., Ma, Y., Liu, H., Theis, B., Petersen, I., & Chen, Y. (2019). Plakophilin 1 is methylated and has a tumor suppressive activity in human lung cancer. *Experimental and Molecular Pathology*, *108*, 73–79. https://doi.org/10.1016/J.YEXMP.2019.04.001

Hallin, J., Bowcut, V., Calinisan, A., Briere, D. M., Hargis, L., Engstrom, L. D., Laguer, J., Medwid, J., Vanderpool, D., Lifset, E., Trinh, D., Hoffman, N., Wang, X., David Lawson, J., Gunn, R. J., Smith, C. R., Thomas, N. C., Martinson, M., Bergstrom, A., … Christensen, J. G. (2022). Anti-tumor efficacy of a potent and selective non-covalent KRASG12D inhibitor. *Nature Medicine*, *28*(10), 2171–2182. https://doi.org/10.1038/s41591-022-02007-7

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

He, L., He, X., Lim, L. P., De Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., Jackson, A. L., Linsley, P. S., Chen, C., Lowe, S. W., Cleary, M. A., & Hannon, G. J. (2007). A microRNA component of the p53 tumour suppressor network. *Nature*, *447*. https://doi.org/10.1038/nature05939

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., & Hammond, S. M. (2005). A microRNA polycistron as a potential human oncogene. *Nature*, *435*(7043), 828. https://doi.org/10.1038/NATURE03552

Heinemann, V., Stintzing, S., Kirchner, T., Boeck, S., & Jung, A. (2009). Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR. *Cancer Treatment Reviews*, *35*(3), 262–271. https://doi.org/10.1016/j.ctrv.2008.11.005

Helwak, A., & Tollervey, D. (2014). Mapping the miRNA interactome by crosslinking ligation and sequencing of hybrids (CLASH). *Nature Protocols*, *9*(3), 711. https://doi.org/10.1038/NPROT.2014.043

Hirsch, F. R., Scagliotti, G. V, Mulshine, J. L., Kwon, R., Curran, W. J., Wu, Y.-L., & Paz-Ares, L. (2017). Lung cancer: current therapies and new targeted treatments. *The Lancet*, *389*(10066), 299–311. https://doi.org/10.1016/S0140-6736(16)30958-8

Hirsch, F. R., Spreafico, A., Novello, S., Wood, M. D., Simms, L., & Papotti, M. (2008). The prognostic and predictive role of histology in advanced non-small cell lung cancer: A literature review. *Journal of Thoracic Oncology*, *3*(12), 1468–1481. https://doi.org/10.1097/JTO.0b013e318189f551

Hong, D. S., Kang, Y. K., Borad, M., Sachdev, J., Ejadi, S., Lim, H. Y., Brenner, A. J., Park, K., Lee, J. L., Kim, T. Y., Shin, S., Becerra, C. R., Falchook, G., Stoudemire, J., Martin, D., Kelnar, K., Peltier, H., Bonato, V., Bader, A. G., … Beg, M. S. (2020). Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *British Journal of Cancer*, *122*(11), 1630. https://doi.org/10.1038/S41416-020-0802-1

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of Bacteria and Archaea. *Science*, *327*(5962), 167–170. https://doi.org/10.1126/SCIENCE.1179555/SUPPL_FILE/167_THUMB.JPG

Hu, X., Bao, J., Wang, Z., Zhang, Z., Gu, P., Tao, F., Cui, D., & Jiang, W. (2016). The plasma lncRNA acting as fingerprint in non-small-cell lung cancer. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine*, *37*(3), 3497–3504. https://doi.org/10.1007/S13277-015-4023-9

Huang, S., Li, D., Zhuang, L. L., Sun, L., & Wu, J. (2021). Identification of Arp2/3 Complex Subunits as Prognostic Biomarkers for Hepatocellular Carcinoma. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/FMOLB.2021.690151

Imig, J., Brunschweiger, A., Brümmer, A., Guennewig, B., Mittal, N., Kishore, S., Tsikrika, P., Gerber, A. P., Zavolan, M., & Hall, J. (2014). miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19–miR-106a interaction. *Nature Chemical Biology 2014 11:2*, *11*(2), 107–114. https://doi.org/10.1038/nchembio.1713

Ischenko, I., D'Amico, S., Rao, M., Li, J., Hayman, M. J., Powers, S., Petrenko, O., & Reich, N. C. (2021). KRAS drives immune evasion in a genetic model of pancreatic cancer. *Nature Communications*, *12*(1). https://doi.org/10.1038/S41467-021-21736-W

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *Journal of Bacteriology*, *200*(7). https://doi.org/10.1128/JB.00580-17

Janes, M. R., Zhang, J., Li, L. S., Hansen, R., Peters, U., Guo, X., Chen, Y., Babbar, A., Firdaus, S. J., Darjania, L., Feng, J., Chen, J. H., Li, S., Li, S., Long, Y. O., Thach, C., Liu, Y., Zarieh, A., Ely, T., … Liu, Y. (2018). Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell*, *172*(3), 578-589.e17. https://doi.org/10.1016/J.CELL.2018.01.006

Jänne, P. A., Riely, G. J., Gadgeel, S. M., Heist, R. S., Ou, S.-H. I., Pacheco, J. M., Johnson, M. L., Sabari, J. K., Leventakos, K., Yau, E., Bazhenova, L., Negrao, M. V., Pennell, N. A., Zhang, J., Anderes, K., Der-Torossian, H., Kheoh, T., Velastegui, K., Yan, X., … Spira, A. I. (2022). Adagrasib in Non–Small-Cell Lung Cancer Harboring a KRAS G12C Mutation . *New England Journal of Medicine*, *387*(2), 120–131. https://doi.org/10.1056/NEJMOA2204619

Jeong, Y. K., Song, B., & Bae, S. (2020). Current Status and Challenges of DNA Base Editing Tools. *Molecular Therapy : The Journal of the American Society of Gene Therapy*, *28*(9), 1938–1952. https://doi.org/10.1016/J.YMTHE.2020.07.021

Ji, P., Diederichs, S., Wang, W., Boïng, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., Thomas, M., Berdel, W. E., Serve, H., & Muïler-Tidow, C. (2003). MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, *22*, 8031–8041. https://doi.org/10.1038/sj.onc.1206928

Jiang, C., Li, Y., Zhao, Z., Lu, J., Chen, H., Ding, N., Wang, G., Xu, J., & Li, X. (2016). Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget*, *7*(6), 7120–7133. https://doi.org/10.18632/oncotarget.6859

Jiang, W., Li, H., Liu, X., Zhang, J., Zhang, W., Li, T., Liu, L., & Yu, X. (2020). Precise and efficient silencing of mutant KrasG12D by CRISPR-CasRx controls pancreatic cancer progression. *Theranostics*, *10*(25), 11507. https://doi.org/10.7150/THNO.46642

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816. https://doi.org/10.1126/SCIENCE.1225829

Jo, M. H., Shin, S., Jung, S. R., Kim, E., Song, J. J., & Hohng, S. (2015). Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Molecular Cell*, *59*(1), 117–124. https://doi.org/10.1016/J.MOLCEL.2015.04.027

Johnson, C. D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., Wang, X., Shelton, J., Shingara, J., Chin, L., Brown, D., & Slack, F. J. (2007). The let-7 MicroRNA Represses Cell Proliferation Pathways in Human Cells. *Cancer Res*, *67*(16), 7713–7735. https://doi.org/10.1158/0008-5472.CAN-07-1083

Karachaliou, N., Mayo, C., Costa, C., Magrí, I., Gimenez-Capitan, A., Molina-Vila, M. A., & Rosell, R. (2013). KRAS mutations in lung cancer. *Clinical Lung Cancer*, *14*(3), 205–214. https://doi.org/10.1016/j.cllc.2012.09.007

Karube, Y., Tanaka, H., Osada, H., Tomida, S., Tatematsu, Y., Yanagisawa, K., Yatabe, Y., Takamizawa, J., Miyoshi, S., Mitsudomi, T., & Takahashi, T. (2005). Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Science*, *96*(2), 111–115. https://doi.org/10.1111/J.1349-7006.2005.00015.X

Kaz, A. M., Luo, Y., Dzieciatkowski, S., Chak, A., Willis, J. E., Upton, M. P., Leidner, R. S., & Grady, W. M. (2012). Aberrantly methylated PKP1 in the progression of Barrett's esophagus to esophageal adenocarcinoma. *Genes Chromosomes Cancer*, *51(4)*, 384–393. https://doi.org/10.1002/gcc.21923

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., & Zhang, F. (2019). SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature Protocols*, *14*(10), 2986–3012. https://doi.org/10.1038/s41596-019-0210-2

Kim, E., Cho, K. O., Rothschild, A., & Sheng, M. (1996). Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins. *Neuron*, *17*(1), 103–113. https://doi.org/10.1016/S0896-6273(00)80284-6

Kim, S. M., Shin, S. C., Kim, E. E., Kim, S. H., Park, K., Oh, S. J., & Jang, M. (2018). Simple in vivo gene editing via direct self-assembly of Cas9 ribonucleoprotein complexes for cancer treatment. *ACS Nano*, *12*(8), 7750–7760. https://doi.org/10.1021/ACSNANO.8B01670

Kim, W., Lee, S., Kim, H. S., Song, M., Cha, Y. H., Kim, Y. H., Shin, J., Lee, E. S., Joo, Y., Song, J. J., Choi, E. J., Choi, J. W., Lee, J., Kang, M., Yook, J. I., Lee, M. G., Kim, Y. S., Paik, S., & Kim, H. (2018). Targeting mutant KRAS with CRISPR-Cas9 controls tumor growth. *Genome Research*, *28*(3), 374–382. https://doi.org/10.1101/gr.223891.117

Kimura, H., Fumoto, K., Shojima, K., Nojima, S., Osugi, Y., Tomihara, H., Eguchi, H., Shintani, Y., Endo, H., Inoue, M., Doki, Y., Okumura, M., Morii, E., & Kikuchi, A. (2016). CKAP4 is a Dickkopf1 receptor and is involved in tumor progression. *The Journal of Clinical Investigation*, *126*(7), 2689–2705. https://doi.org/10.1172/JCI84658

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J. R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, *11*(1), 1–14. https://doi.org/10.1186/1741-7007-11-59

Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R., & Jacks, T. (2007). Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genetics*, *39*(5), 673–677. https://doi.org/10.1038/ng2003

Kurata, M., Yamamoto, K., Moriarity, B. S., Kitagawa, M., & Largaespada, D. A. (2018). CRISPR/Cas9 library screening for drug target discovery. *Journal of Human Genetics*, *63*, 179–186. https://doi.org/10.1038/s10038-017-0376-9

Kwak, E. L., Bang, Y.-J., Camidge, D. R., Shaw, A. T., Solomon, B., Maki, R. G., Ou, S.-H. I., Dezube, B. J., Jänne, P. A., Costa, D. B., Varella-Garcia, M., Kim, W.-H., Lynch, T. J., Fidias, P., Stubbs, H., Engelman, J. A., Sequist, L. V., Tan, W., Gandhi, L., … Iafrate, A. J. (2010). Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer. *New England Journal of Medicine*, *363*(18), 1693–1703. https://doi.org/10.1056/NEJMOA1006448

Kwan, A. K., Piazza, G. A., Keeton, A. B., & Leite, C. A. (2022). The path to the clinic: a comprehensive review on direct KRASG12C inhibitors. *Journal of Experimental & Clinical Cancer Research : CR*, *41*(1), 27. https://doi.org/10.1186/S13046-021-02225-W

Kwanhian, W., Lenze, D., Alles, J., Motsch, N., Barth, S., Döll, C., Imig, J., Hummel, M., Tinguely, M., Trivedi, P., Lulitanond, V., Meister, G., Renner, C., & Grässer, F. A. (2012). MicroRNA-142 is mutated in about 20% of diffuse large B-cell lymphoma. *Cancer Medicine*, *1*(2), 141–155. https://doi.org/10.1002/CAM4.29

Lam, J. K. W., Chow, M. Y. T., Zhang, Y., & Leung, S. W. S. (2015). siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy. Nucleic Acids*, *4*(9), e252. https://doi.org/10.1038/MTNA.2015.23

Lee, G. L., Dobi, A., & Srivastava, S. (2011). Diagnostic performance of the PCA3 urine test. *Nature Reviews Urology*, *8*(3), 123–124. https://doi.org/10.1038/nrurol.2011.10

Lee, J. T., & Bartolomei, M. S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell*, *152*(6), 1308–1323. https://doi.org/10.1016/J.CELL.2013.02.016

Lee, W., Lee, J. H., Jun, S., Lee, J. H., & Bang, D. (2018). Selective targeting of KRAS oncogenic alleles by CRISPR/Cas9 inhibits proliferation of cancer cells. *Scientific Reports*, *8*(1), 1–7. https://doi.org/10.1038/s41598-018-30205-2

Lentsch, E., Li, L., Pfeffer, S., Ekici, A. B., Taher, L., Pilarsky, C., & Grützmann, R. (2019). CRISPR/Cas9-mediated knock-out of krasG12D mutated pancreatic cancer cell lines. *International Journal of Molecular Sciences*, *20*(22). https://doi.org/10.3390/ijms20225706

Li, A. R., Chitale, D., Riely, G. J., Pao, W., Miller, V. A., Zakowski, M. F., Rusch, V., Kris, M. G., & Ladanyi, M. (2008). EGFR Mutations in Lung Adenocarcinomas Clinical Testing

Experience and Relationship to EGFR Gene Copy Number and Immunohistochemical Expression. *J Mol Diagn*, *10*, 242–248. https://doi.org/10.2353/jmoldx.2008.070178

Li, D., Xia, L., Chen, M., Lin, C., Wu, H., Zhang, Y., Pan, S., & Li, X. (2017). miR-133b, a particular member of myomiRs, coming into playing its unique pathological role in human cancer. *Oncotarget*, *8*(30), 50193. https://doi.org/10.18632/ONCOTARGET.16745

Li, Z., Liu, J., Chen, H., Zhang, Y., Shi, H., Huang, L., Tao, J., Shen, R., & Wang, T. (2020). Ferritin Light Chain (FTL) competes with long noncoding RNA Linc00467 for miR-133b binding site to regulate chemoresistance and metastasis of colorectal cancer. *Carcinogenesis*, *41*(4), 467–477. https://doi.org/10.1093/CARCIN/BGZ181

Li, Z., Ye, Z., Ma, J., Gu, Q., Teng, J., & Gong, X. (2021). MicroRNA-133b alleviates doxorubicin-induced cardiomyocyte apoptosis and cardiac fibrosis by targeting PTBP1 and TAGLN2. *International Journal of Molecular Medicine*, *48*(1). https://doi.org/10.3892/IJMM.2021.4958

Lin, X., Guo, L., Lin, X., Wang, Y., & Zhang, G. (2022). Expression and prognosis analysis of mitochondrial ribosomal protein family in breast cancer. *Scientific Reports*, *12*(1), 10658. https://doi.org/10.1038/S41598-022-14724-7

Lin, Y., Schmidt, B. F., Bruchez, M. P., & McManus, C. J. (2018). Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Research*, *46*(7), 3742–3752. https://doi.org/10.1093/NAR/GKY046

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., Kovatich, A. J., Benz, C. C., Levine, D. A., Lee, A. V., Omberg, L., Wolf, D. M., Shriver, C. D., Thorsson, V., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., … Hu, H. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*, *173*(2), 400-416.e11. https://doi.org/10.1016/J.CELL.2018.02.052

Liu, J., Sun, M., Cho, K. B., Gao, X., & Guo, B. (2021). A CRISPR-Cas9 repressor for epigenetic silencing of KRAS. *Pharmacological Research*, *164*. https://doi.org/10.1016/J.PHRS.2020.105304

Liu, L., Shao, X., Gao, W., Zhang, Z., Liu, P., Wang, R., Huang, P., Yin, Y., & Shu, Y. (2012). MicroRNA-133b inhibits the growth of non-small-cell lung cancer by targeting the epidermal growth factor receptor. *FEBS Journal*, *279*(20), 3800–3812. https://doi.org/10.1111/J.1742-4658.2012.08741.X

Liu, Z., Shi, M., Ren, Y., Xu, H., Weng, S., Ning, W., Ge, X., Liu, L., Guo, C., Duo, M., Li, L., Li, J., & Han, X. (2023). Recent advances and applications of CRISPR-Cas9 in cancer immunotherapy. *Molecular Cancer*, *22*(1), 35. https://doi.org/10.1186/S12943-023-01738-6

Loewen, G., Jayawickramarajah, J., Zhuo, Y., & Shan, B. (2014). Functions of lncRNA HOTAIR in lung cancer. *Journal of Hematology and Oncology*, *7*(1), 1–10. https://doi.org/10.1186/S13045-014-0090-4

Lopez Sanchez, M. I. G., Krüger, A., Shiriaev, D. I., Liu, Y., & Rorbach, J. (2021). Human Mitoribosome Biogenesis and Its Emerging Links to Disease. *International Journal of*

Molecular Sciences 2021, Vol. 22, Page 3827, *22*(8), 3827. https://doi.org/10.3390/IJMS22083827

López-Longarela, B., Morrison, E. E., Tranter, J. D., Chahman-Vos, L., Léonard, J. F., Gautier, J. C., Laurent, S., Lartigau, A., Boitier, E., Sautier, L., Carmona-Saez, P., Martorell-Marugan, J., Mellanby, R. J., Pernagallo, S., Ilyine, H., Rissin, D. M., Duffy, D. C., Dear, J. W., & Díaz-Mochón, J. J. (2020). Direct Detection of miR-122 in Hepatotoxicity Using Dynamic Chemical Labeling Overcomes Stability and isomiR Challenges. *Analytical Chemistry*, *92*(4), 3388–3395. https://doi.org/10.1021/ACS.ANALCHEM.9B05449

Lou, K., Steri, V., Ge, A. Y., Hwang, Y. C., Yogodzinski, C. H., Shkedi, A. R., Choi, A. L. M., Mitchell, D. C., Swaney, D. L., Hann, B., Gordan, J. D., Shokat, K. M., & Gilbert, L. A. (2019). KRASG12C inhibition produces a driver-limited state revealing collateral dependencies. *Science Signaling*, *12*(583). https://doi.org/10.1126/SCISIGNAL.AAW9450

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Robert Horvitz, H., & Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. *Nature*, *435*, 834–838. https://doi.org/10.1038/nature03702

Lu, T., Wang, Y., Chen, D., Liu, J., & Jiao, W. (2018). Potential clinical application of lncRNAs in non-small cell lung cancer. *OncoTargets and Therapy*, *11*, 8045. https://doi.org/10.2147/OTT.S178431

Lu, W., Feng, F., Xu, J., Lu, X., Wang, S., Wang, L., Lu, H., Wei, M., Yang, G., Wang, L., Lu, Z., Liu, Y., & Lei, X. (2014). QKI impairs self-renewal and tumorigenicity of oral cancer cells via repression of SOX2. *Cancer Biology & Therapy*, *15*(9), 1174–1184. https://doi.org/10.4161/CBT.29502

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2019). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

Maldonado-Pérez, N., Tristán-Manzano, M., Justicia-Lirio, P., Martínez-Planes, E., Muñoz, P., Pavlovic, K., Cortijo-Gutiérrez, M., Blanco-Benítez, C., Castella, M., Juan, M., Wenes, M., Romero, P., Molina-Estévez, F. J., Marañón, C., Herrera, C., Benabdellah, K., & Martin, F. (2022). Efficacy and safety of universal (TCRKO) ARI-0001 CAR-T cells for the treatment of B-cell lymphoma. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/FIMMU.2022.1011858

Marks, J. L., Broderick, S., Zhou, Q., Chitale, D., Li, A. R., Zakowski, M. F., Kris, M. G., Rusch, V. W., Azzoli, C. G., Seshan, V. E., Ladanyi, M., & Pao, W. (2008). Prognostic and Therapeutic Implications of EGFR and KRAS Mutations in Resected Lung Adenocarcinoma. *Journal of Thoracic Oncology*, *3*(2), 111–116. https://doi.org/10.1097/JTO.0b013e318160c607

Martinelli, E., De Palma, R., Orditura, M., De Vita, F., & Ciardiello, F. (2009). Anti-epidermal growth factor receptor monoclonal antibodies in cancer therapy. *Clinical and Experimental Immunology*, *158*(1), 1. https://doi.org/10.1111/J.1365-2249.2009.03992.X

Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., Gonzalez-Perez, A., & Lopez-Bigas, N. (2020). A compendium of mutational cancer driver genes. *Nature Reviews Cancer*, *20*, 555–572. https://doi.org/10.1038/s41568-020-0290-x

Martin-Padron, J., Boyero, L., Rodriguez, M. I., Andrades, A., Díaz-Cano, I., Peinado, P., Baliñas-Gavira, C., Alvarez-Perez, J. C., Coira, I. F., Fárez-Vidal, M. E., & Medina, P. P. (2020). Plakophilin 1 enhances MYC translation, promoting squamous cell lung cancer. *Oncogene*, *39*(32), 5479–5493. https://doi.org/10.1038/S41388-019-1129-3

Marusyk, A., & Polyak, K. (2009). Tumor heterogeneity: causes and consequences. *Biochimica et Biophysica Acta*, *1805*(1), 105–117. https://doi.org/10.1016/j.bbcan.2009.11.002

Mass, R. D. (2004). The HER receptor family: A rich target for therapeutic development. *International Journal of Radiation Oncology Biology Physics*, *58*(3), 932–940. https://doi.org/10.1016/j.ijrobp.2003.09.093

Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L. L., Chen, R., Dean, C., Dinger, M. E., Fitzgerald, K. A., Gingeras, T. R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J. B., Lee, J. T., … Wu, M. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*, *17*, 1–17. https://doi.org/10.1038/s41580-022-00566-8

Mayekar, M. K., & Bivona, T. G. (2017). Current Landscape of Targeted Therapy in Lung Cancer. *Clinical Pharmacology and Therapeutics*, *102(5)*, 757–764. https://doi.org/10.1002/cpt.810

Mazoyer, S., Gayther, S. A., Nagai, M. A., Smith, S. A., Dunning, A., Van Rensburg, E. J., Albertsen, H., White, R., & Ponder, B. A. J. (1995). A gene (DLG2) located at 17q12-q21 encodes a new homologue of the drosophila tumor suppressor dlg-A. *Genomics*, *28*(1), 25–31. https://doi.org/10.1006/geno.1995.1101

McAndrews, K. M., Xiao, F., Chronopoulos, A., LeBleu, V. S., Kugeratski, F. G., & Kalluri, R. (2021). Exosome-mediated delivery of CRISPR/Cas9 for targeting of oncogenic KrasG12D in pancreatic cancer. *Life Science Alliance*, *4*(9). https://doi.org/10.26508/LSA.202000875

Mcfarland, C. D., Yaglom, J. A., Wojtkowiak, J. W., Scott, J. G., Morse, D. L., Sherman, M. Y., & Mirny, L. A. (2017). Molecular and Cellular Pathobiology The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Research*, *77(18)*, 4763–4772. https://doi.org/10.1158/0008-5472.CAN-15-3283-T

McGrath, J. A., McMillan, J. R., Shemanko, C. S., Runswick, S. K., Leigh, I. M., Lane, E. B., Garrod, D. R., & Eady, R. A. J. (1997). Mutations in the plakophilin 1 gene result in ectodermal dysplasia/skin fragility syndrome. *Nature Genetics*, *17*(2), 240–244. https://doi.org/10.1038/ng1097-240

Medina, P. P., Nolde, M., & Slack, F. J. (2010). OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature*, *467*(7311), 86–90. https://doi.org/10.1038/NATURE09284

Medina, P. P., & Slack, F. J. (2008). MicroRNAs and cancer: An overview. *Cell Cycle*, *7*(16), 2485–2492. https://doi.org/10.4161/CC.7.16.6453

Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L. A., Del Castillo, I., Steel, K. P., Dalmay, T., Moreno, F., & Moreno-Pelayo, M. A. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics*, *41*(5), 609–613. https://doi.org/10.1038/ng.355

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/S00239-004-0046-3

Morelli, E., Gulla', A., Amodio, N., Taiana, E., Neri, A., Fulciniti, M., & Munshi, N. C. (2021). CRISPR Interference (CRISPRi) and CRISPR Activation (CRISPRa) to Explore the Oncogenic lncRNA Network. *Methods in Molecular Biology*, *2348*, 189–204. https://doi.org/10.1007/978-1-0716-1581-2_13/FIGURES/2

Muzumdar, M. D., Chen, P. Y., Dorans, K. J., Chung, K. M., Bhutkar, A., Hong, E., Noll, E. M., Sprick, M. R., Trumpp, A., & Jacks, T. (2017). Survival of pancreatic cancer cells lacking KRAS function. *Nature Communications*, *8*(1), 1–19. https://doi.org/10.1038/s41467-017-00942-5

Nagasaka, M., Potugari, B., Nguyen, A., Sukari, A., Azmi, A. S., & Ou, S. H. I. (2021). KRAS Inhibitors– yes but what next? Direct targeting of KRAS– vaccines, adoptive T cell therapy and beyond. In *Cancer Treatment Reviews* (Vol. 101). W.B. Saunders Ltd. https://doi.org/10.1016/j.ctrv.2021.102309

Nicholson, A. G., Tsao, M. S., Beasley, M. B., Borczuk, A. C., Brambilla, E., Cooper, W. A., Dacic, S., Jain, D., Kerr, K. M., Lantuejoul, S., Noguchi, M., Papotti, M., Rekhtman, N., Scagliotti, G., van Schil, P., Sholl, L., Yatabe, Y., Yoshida, A., & Travis, W. D. (2022). The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. In *Journal of Thoracic Oncology* (Vol. 17, Issue 3, pp. 362–387). Elsevier Inc. https://doi.org/10.1016/j.jtho.2021.11.003

Niu, Z., Jin, R., Zhang, Y., & Li, H. (2022). Signaling pathways and targeted therapies in lung squamous cell carcinoma: mechanisms and clinical trials. *Signal Transduction and Targeted Therapy*, *7*(353). https://doi.org/10.1038/s41392-022-01200-x

O'Brien, J. A., Heyam, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, *1*, 402. https://doi.org/10.3389/fendo.2018.00402

O'donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V, & Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, *435*, 839–843. https://doi.org/10.1038/nature03677

Orang, A. V., Safaralizadeh, R., & Kazemzadeh-Bavili, M. (2014). Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. *International Journal of Genomics*, *2014:970607*. https://doi.org/10.1155/2014/970607

Palazzo, A. F., & Gregory, R. (2014). The Case for Junk DNA. *PLOS Genetics*, *10*(5), e1004351. https://doi.org/10.1371/journal.pgen.1004351

Pallarès-Masmitjà, M., Ivančić, D., Mir-Pedrol, J., Jaraba-Wallace, J., Tagliani, T., Oliva, B., Rahmeh, A., Sánchez-Mejías, A., & Güell, M. (2021). Find and cut-and-transfer (FiCAT)

mammalian genome engineering. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-27183-x

Pan, J. Y., Sun, C. C., Bi, Z. Y., Chen, Z. L., Li, S. J., Li, Q. Q., Wang, Y. X., Bi, Y. Y., & Li, D. J. (2017). miR-206/133b Cluster: A Weapon against Lung Cancer? *Molecular Therapy. Nucleic Acids*, *8*, 442. https://doi.org/10.1016/J.OMTN.2017.06.002

Paralkar, V. R., Taborda, C. C., Huang, P., Yao, Y., Kossenkov, A. V., Prasad, R., Luan, J., Davies, J. O. J., Hughes, J. R., Hardison, R. C., Blobel, G. A., & Weiss, M. J. (2016). Unlinking a lncRNA from its associated cis element. *Molecular Cell*, *62*(1), 104. https://doi.org/10.1016/J.MOLCEL.2016.02.029

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., & Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, *408*(6808), 86–89. https://doi.org/10.1038/35040556

Pecoraro, A., Pagano, M., Russo, G., & Russo, A. (2021). Ribosome Biogenesis and Cancer: Overview on Ribosomal Proteins. *International Journal of Molecular Sciences*, *22*(11), 5496. https://doi.org/10.3390/IJMS22115496

Peinado, P., Andrades, A., Cuadros, M., Rodriguez, M. I., Coira, I. F., Garcia, D. J., Álvarez-Perez, J. C., Baliñas-Gavira, C., Arenas, A. M., Patiño-Mercau, J. R., Sanjuan-Hidalgo, J., Romero, O. A., Montuenga, L. M., Carretero, J., Sanchez-Cespedes, M., & Medina, P. P. (2020). Comprehensive Analysis of SWI/SNF Inactivation in Lung Adenocarcinoma Cell Models. *Cancers*, *12*(12), 1–14. https://doi.org/10.3390/CANCERS12123712

Peinado, P., Herrera, A., Baliñas, C., Martín-Padrón, J., Boyero, L., Cuadros, M., Coira, I. F., Rodriguez, M. I., Reyes-Zurita, F. J., Rufino-Palomares, E. E., Lupiáñez, J. A., & Medina, P. P. (2018). Long Noncoding RNAs as Cancer Biomarkers. In J. Chakrabarti & S. Mitra (Eds.), *Cancer and Noncoding RNAs* (Vol. 1, pp. 95–114). Academic Press. https://doi.org/10.1016/B978-0-12-811022-5.00006-1

Piatek, M. J., & Werner, A. (2014). Endogenous siRNAs, regulators of internal affairs. *Biochemical Society Transactions*, *42*(4), 1174. https://doi.org/10.1042/BST20140068

Pignon, J. P., Tribodet, H., Scagliotti, G. V., Douillard, J. Y., Shepherd, F. A., Stephens, R. J., Dunant, A., Torri, V., Rosell, R., Seymour, L., Spiro, S. G., Rolland, E., Fossati, R., Aubert, D., Ding, K., Waller, D., & Le Chevalier, T. (2008). Lung adjuvant cisplatin evaluation: A pooled analysis by the LACE collaborative group. *Journal of Clinical Oncology*, *26*(21), 3552–3559. https://doi.org/10.1200/JCO.2007.13.9030

Polesskaya, O. O., Haroutunian, V., Davis, K. L., Hernandez, I., & Sokolov, B. P. (2003). Novel putative nonprotein-coding RNA gene from 11q14 displays decreased expression in brains of patients with schizophrenia. *Journal of Neuroscience Research*, *74*(1), 111–122. https://doi.org/10.1002/JNR.10752

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, *28*(10), 1057–1068. https://doi.org/10.1038/nbt.1685

Prior, I. A., Hood, F. E., & Hartley, J. L. (2020). The Frequency of Ras Mutations in Cancer. *Cancer Research*, *80(14)*, 2969–2974. https://doi.org/10.1158/0008-5472.CAN-19-3682

Pulciani, S., Santos, E., Lauver, A. V, Long, L. K., Robbins, K. C., & Barbacid, M. (1982). Oncogenes in human tumor cell lines: Molecular cloning of a transforming gene from human bladder carcinoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2845–2849. https://doi.org/10.1073/pnas.79.9.2845

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, *152*(5), 1173. https://doi.org/10.1016/J.CELL.2013.02.022

Qi, P., Zhou, X. yan, & Du, X. (2016). Circulating long non-coding RNAs in cancer: current status and future perspectives. *Molecular Cancer*, *15*(1). https://doi.org/10.1186/S12943-016-0524-4

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rekhtman, N., Paik, P. K., Arcila, M. E., Tafe, L. J., Oxnard, G. R., Moreira, A. L., Travis, W. D., Zakowski, M. F., Kris, M. G., & Ladanyi, M. (2012). Clarifying the Spectrum of Driver Oncogene Mutations in Biomarker-Verified Squamous Carcinoma of Lung: Lack of EGFR/KRAS and Presence of PIK3CA/AKT1 Mutations. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.*, *18*(4), 1167–1176. https://doi.org/10.1158/1078-0432.CCR-11-2109

Rheinbay, E., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J. M., Juul, R. I., Lin, Z., Feuerbach, L., Sabarinathan, R., Madsen, T., Kim, J., Mularoni, L., Shuai, S., Lanzós, A., Herrmann, C., Maruvka, Y. E., … Zamora, J. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*, *578*(7793), 102–111. https://doi.org/10.1038/S41586-020-1965-X

Richter, U., Lahtinen, T., Marttinen, P., Suomi, F., & Battersby, B. J. (2015). Quality control of mitochondrial protein synthesis is required for membrane integrity and cell fitness. *The Journal of Cell Biology*, *211*(2), 373–389. https://doi.org/10.1083/JCB.201504062

Riely, G. J., Marks, J., & Pao, W. (2009). KRAS mutations in non-small cell lung cancer. *Proceedings of the American Thoracic Society*, *6*(2), 201–205. https://doi.org/10.1513/PATS.200809-107LC

Robles-Remacho, A., Luque-Gonzalez, M. A., López-Delgado, F. J., Guardia-Monteagudo, J. J., Fara, M. A., Pernagallo, S., Sanchez-Martin, R. M., & Diaz-Mochon, J. J. (2023). Direct detection of alpha satellite DNA with single-base resolution by using abasic Peptide Nucleic Acids and Fluorescent in situ Hybridization. *Biosensors and Bioelectronics*, *219*, 114770. https://doi.org/10.1016/J.BIOS.2022.114770

Román, M., Baraibar, I., López, I., Nadal, E., Rolfo, C., Vicent, S., & Gil-Bazo, I. (2018). KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an

old target. *Molecular Cancer*, *17*(1), 1–14. https://doi.org/10.1186/S12943-018-0789-X

Rosell, R., Moran, T., Queralt, C., Cardenal, F., Camps, C., Majem, M., Lopez-Vivanco, G., Isla, D., Provencio, M., Insa, A., Massuti, B., Luis Gonzalez-Larriba, J., Paz-Ares, L., Bover, I., Garcia-Campelo, R., Moreno, M. A., Catot, S., Rolfo, C., Reguart, N., … Taron, M. (2009). Screening for Epidermal Growth Factor Receptor Mutations in Lung Cancer A bs tr ac t. *New England Journal of Medicine*, *361*, 958–967. https://doi.org/10.1056/NEJMoa0904554

Rupaimoole, R., & Slack, F. J. (2017). MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nature Reviews Drug Discovery*, *16*(3), 203–222. https://doi.org/10.1038/nrd.2016.246

Sandler, A., Gray, R., Perry, M. C., Brahmer, J., Schiller, J. H., Dowlati, A., Lilenbaum, R., & Johnson, D. H. (2006). Paclitaxel-Carboplatin Alone or with Bevacizumab for Non-Small-Cell Lung Cancer. *New England Journal of Medicine*, *355*, 2542–2550. https://doi.org/10.1056/NEJMoa061884

Santos, M., Fidalgo, A., Varanda, A. S., Oliveira, C., & Santos, M. A. S. (2019). tRNA Deregulation and Its Consequences in Cancer. *Trends in Molecular Medicine*, *25*(10), 853–865. https://doi.org/10.1016/J.MOLMED.2019.05.011

Santos, R. M., Moreno, C., & Zhang, W. C. (2020). Non-Coding RNAs in Lung Tumor Initiation and Progression. *International Journal of Molecular Sciences*, *21*(8). https://doi.org/10.3390/IJMS21082774

Satoh, K., Sakai, S., & Nishizuka, M. (2022). Knockdown of RhoQ, a member of Rho GTPase, accelerates TGF-β-induced EMT in human lung adenocarcinoma. *Biochemistry and Biophysics Reports*, *32*. https://doi.org/10.1016/J.BBREP.2022.101346

Sayed, S., Sidorova, O. A., Hennig, A., Augsburg, M., Cortés Vesga, C. P., Abohawya, M., Schmitt, L. T., Sürün, D., Stange, D. E., Mircetic, J., & Buchholz, F. (2022). Efficient Correction of Oncogenic KRAS and TP53 Mutations through CRISPR Base Editing. *Cancer Research*, *82*(17), 3002. https://doi.org/10.1158/0008-5472.CAN-21-2519

Scagliotti, G. V., Parikh, P., Von Pawel, J., Biesma, B., Vansteenkiste, J., Manegold, C., Serwatowski, P., Gatzemeier, U., Digumarti, R., Zukin, M., Lee, J. S., Mellemgaard, A., Park, K., Patil, S., Rolski, J., Goksel, T., De Marinis, F., Simms, L., Sugarman, K. P., & Gandara, D. (2008). Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *Journal of Clinical Oncology*, *26*(21), 3543–3551. https://doi.org/10.1200/JCO.2007.15.0375

Schmitt, A. M., & Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, *29*(4), 452–463. https://doi.org/10.1016/J.CCELL.2016.03.010

Sequist, L. V., Waltman, B. A., Dias-Santagata, D., Digumarthy, S., Turke, A. B., Fidias, P., Bergethon, K., Shaw, A. T., Gettinger, S., Cosper, A. K., Akhavanfard, S., Heist, R. S., Temel, J., Christensen, J. G., Wain, J. C., Lynch, T. J., Vernovsky, K., Mark, E. J., Lanuti, M., … Engelman, J. A. (2011). Genotypic and Histological Evolution of Lung Cancers Acquiring Resistance to EGFR Inhibitors. *Science Translational Medicine*, *3*(75), 75ra26. https://doi.org/10.1126/SCITRANSLMED.3002003

Seto, A. G., Beatty, X., Lynch, J. M., Hermreck, M., Tetzlaff, M., Duvic, M., & Jackson, A. L. (2018). Cobomarsen, an oligonucleotide inhibitor of miR-155, co-ordinately regulates multiple survival pathways to reduce cellular proliferation and survival in cutaneous T-cell lymphoma. *British Journal of Haematology*, *183*(3), 428–444. https://doi.org/10.1111/BJH.15547

Shi, M., Duan, G., Nie, S., Shen, S., & Zou, X. (2018). Elevated FAM3C promotes cell epithelial- mesenchymal transition and cell migration in gastric cancer. *OncoTargets and Therapy*, *11*, 8491–8505. https://doi.org/10.2147/OTT.S178455

Shi, R., & Chiang, V. L. (2005). Facile means for quantifying microRNA expression by real-time PCR. *BioTechniques*, *39*(4), 519–524. https://doi.org/10.2144/000112010

Shi, T., Gao, G., & Cao, Y. (2016). Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics. *Disease Markers*, *2016*. https://doi.org/10.1155/2016/9085195

Shi, X., Sun, M., Liu, H., Yao, Y., Kong, R., Chen, F., & Song, Y. (2015). A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Molecular Carcinogenesis*, *54 Suppl 1*(S1), E1–E12. https://doi.org/10.1002/MC.22120

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/CAAC.21763

Singh, A., & Settleman, J. (2009). Oncogenic K-ras "addiction" and synthetic lethality. *Cell Cycle*, *8*(17), 2676–2678. https://doi.org/10.4161/CC.8.17.9336

Skoulidis, F., & Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nature Reviews Cancer*, *19*(9), 495–509. https://doi.org/10.1038/s41568-019-0179-8

Škrtić, M., Sriskanthadevan, S., Jhas, B., Gebbia, M., Wang, X., Wang, Z., Hurren, R., Jitkova, Y., Gronda, M., Maclean, N., Lai, C. K., Eberhard, Y., Bartoszko, J., Spagnuolo, P., Rutledge, A. C., Datti, A., Ketela, T., Moffat, J., Robinson, B. H., … Schimmer, A. D. (2011). Inhibition of mitochondrial translation as a therapeutic strategy for human acute myeloid leukemia. *Cancer Cell*, *20*(5), 674–688. https://doi.org/10.1016/J.CCR.2011.10.015

Slack, F. J., & Chinnaiyan, A. M. (2019). The Role of Non-coding RNAs in Oncology. *Cell*, *179*(5), 1033. https://doi.org/10.1016/J.CELL.2019.10.017

Smith, J., Banerjee, R., Weeks, R. J., & Chatterjee, A. (2022). Editing of DNA Methylation Patterns Using CRISPR-Based Tools. *Methods in Molecular Biology*, *2458*, 63–74. https://doi.org/10.1007/978-1-0716-2140-0_4

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., & Pyle, A. M. (2015). HOTAIR forms an intricate and modular secondary structure. *Molecular Cell*, *58*(2), 353. https://doi.org/10.1016/J.MOLCEL.2015.03.006

Song, F., & Stieger, K. (2017). Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Molecular Therapy. Nucleic Acids*, *7*, 53–60. https://doi.org/10.1016/J.OMTN.2017.02.006

Sontheimer, E. J., & Carthew, R. W. (2005). Silence from within: Endogenous siRNAs and miRNAs. *Cell*, *122*(1), 9–12. https://doi.org/10.1016/J.CELL.2005.06.030

Stavast, C. J., & Erkeland, S. J. (2019). The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. *Cells*, *8*(11), 1465. https://doi.org/10.3390/cells8111465

Sugito, N., Taniguchi, K., Kuranaga, Y., Ohishi, M., Soga, T., Ito, Y., Miyachi, M., Kikuchi, K., Hosoi, H., & Akao, Y. (2017). Cancer-Specific Energy Metabolism in Rhabdomyosarcoma Cells Is Regulated by MicroRNA. *Nucleic Acid Therapeutics*, *27*(6), 365–377. https://doi.org/10.1089/NAT.2017.0673

Sugiyama, T., Taniguchi, K., Matsuhashi, N., Tajirika, T., Futamura, M., Takai, T., Akao, Y., & Yoshida, K. (2016). MiR-133b inhibits growth of human gastric cancer cells by silencing pyruvate kinase muscle-splicer polypyrimidine tract-binding protein 1. *Cancer Science*, *107*(12), 1767–1775. https://doi.org/10.1111/CAS.13091

Taanman, J. W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1410*(2), 103–123. https://doi.org/10.1016/S0005-2728(98)00161-3

Tan, H. (2020). Somatic mutation in noncoding regions: The sound of silence. *EBioMedicine*, *61*, 103084. https://doi.org/10.1016/J.EBIOM.2020.103084

Tang, W.-F., Tran, T. A., Wang, L.-Y., & Horng, J.-T. (2023). SARS-CoV-2 pandemics: An update of CRISPR in diagnosis and host–virus interaction studies. *Biomedical Journal*. https://doi.org/10.1016/J.BJ.2023.02.007

Tang, Y., Liu, J. H., Shi, Z. X., Li, Z., Liu, H. T., & Lu, P. (2019). MicroRNA-133b suppresses cell proliferation and invasion of esophageal squamous cell carcinoma via downregulating TAGLN2 expression. *Zhonghua Zhong Liu Za Zhi [Chinese Journal of Oncology]*, *41*(2), 91–96. https://doi.org/10.3760/cma.j.issn.0253-3766.2019.02.003

Teh, M. T., Ken Parkinson, E., Thurlow, J. K., Liu, F., Fortune, F., & Wan, H. (2011). A molecular study of desmosomes identifies a desmoglein isoform switch in head and neck squamous cell carcinoma. *Journal of Oral Pathology & Medicine*, *40*(1), 67–76. https://doi.org/10.1111/J.1600-0714.2010.00951.X

Thomsen, E. A., & Mikkelsen, J. G. (2019). CRISPR-Based Lentiviral Knockout Libraries for Functional Genomic Screening and Identification of Phenotype-Related Genes. *Methods in Molecular Biology*, *1961*, 343–357. https://doi.org/10.1007/978-1-4939-9170-9_21

Thomsen, E. A., Skipper, K. A., Andersen, S., Haslund, D., Skov, T. W., & Mikkelsen, J. G. (2022). CRISPR-Cas9-directed gene tagging using a single integrase-defective lentiviral vector carrying a transposase-based Cas9 off switch. *Molecular Therapy. Nucleic Acids*, *29*, 563–576. https://doi.org/10.1016/J.OMTN.2022.08.005

Thorenoor, N., & Slaby, O. (2015). Small nucleolar RNAs functioning and potential roles in cancer. *Tumor Biology*, *36*(1), 41–53. https://doi.org/10.1007/S13277-014-2818-8

Thun, M., Peto, R., Boreham, J., & Lopez, A. D. (2012). Stages of the cigarette epidemic on entering its second century. *Tobacco Control*, *21:96e101*. https://doi.org/10.1136/tobaccocontrol-2011-050294

Trahey, M., & Mccormick, F. (1987). A Cytoplasmic Protein Stimulates Normal N-ras p21 GTPase, But Does Not Affect Oncogenic Mutants. *Science*, *238*(4826), 542–545. https://doi.org/10.1126/SCIENCE.2821624

Tran, K. B., Lang, J. J., Compton, K., Xu, R., Acheson, A. R., Henrikson, H. J., Kocarnik, J. M., Penberthy, L., Aali, A., Abbas, Q., Abbasi, B., Abbasi-Kangevari, M., Abbasi-Kangevari, Z., Abbastabar, H., Abdelmasseh, M., Abd-Elsalam, S., Abdelwahab, A. A., Abdoli, G., Abdulkadir, H. A., … Murray, C. J. L. (2022). The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *400*(10352), 563–591. https://doi.org/10.1016/S0140-6736(22)01438-6

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., Blencowe, B. J., Prasanth, S. G., & Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*, *39*(6), 925–938. https://doi.org/10.1016/J.MOLCEL.2010.08.011

Tsuji-Tamura, K., Morino-Koga, S., Suzuki, S., & Ogawa, M. (2021). The canonical smooth muscle cell marker TAGLN is present in endothelial cells and is involved in angiogenesis. *Journal of Cell Science*, *134*(15). https://doi.org/10.1242/jcs.254920

Tsygankova, O. M., & Keen, J. H. (2019). A unique role for clathrin light chain A in cell spreading and migration. *Journal of Cell Science*, *132*(10). https://doi.org/10.1242/JCS.224030

Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., Bode, N. M., McNeill, M. S., Yan, S., Camarena, J., Lee, C. M., Park, S. H., Wiebking, V., Bak, R. O., Gomez-Ospina, N., Pavel-Dinu, M., Sun, W., Bao, G., Porteus, M. H., & Behlke, M. A. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human haematopoietic stem and progenitor cells. *Nature Medicine*, *24*(8), 1216. https://doi.org/10.1038/S41591-018-0137-0

van Zandwijk, N., Pavlakis, N., Kao, S. C., Linton, A., Boyer, M. J., Clarke, S., Huynh, Y., Chrzanowska, A., Fulham, M. J., Bailey, D. L., Cooper, W. A., Kritharides, L., Ridley, L., Pattison, S. T., MacDiarmid, J., Brahmbhatt, H., & Reid, G. (2017). Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *The Lancet. Oncology*, *18*(10), 1386–1396. https://doi.org/10.1016/S1470-2045(17)30621-6

Villegas, V. E., & Zaphiropoulos, P. G. (2015). Neighboring Gene Regulation by Antisense Long Non-Coding RNAs. *International Journal of Molecular Sciences*, *16*(2), 3251. https://doi.org/10.3390/IJMS16023251

Vo, J. N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., Wu, Y. M., Dhanasekaran, S. M., Engelke, C. G., Cao, X., Robinson, D. R., Nesvizhskii, A. I., & Chinnaiyan, A. M. (2019). The Landscape of Circular RNA in Cancer. *Cell*, *176*(4), 869-881.e13. https://doi.org/10.1016/J.CELL.2018.12.021

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science (New York, N.Y.)*, *339*(6127), 1546–1558. https://doi.org/10.1126/SCIENCE.1235122

Volinia, S., Galasso, M., Costinean, S., Tagliavini, L., Gamberoni, G., Drusco, A., Marchesini, J., Mascellani, N., Sana, M. E., Jarour, R. A., Desponts, C., Teitell, M., Baffa, R., Aqeilan, R., Iorio, M. V., Taccioli, C., Garzon, R., Di Leva, G., Fabbri, M., … Croce, C. M. (2010). Reprogramming of miRNA networks in cancer and leukemia. *Genome Research*, *20*(5), 589. https://doi.org/10.1101/GR.098046.109

Wan, T., Chen, Y., Pan, Q., Xu, X., Kang, Y., Gao, X., Huang, F., Wu, C., & Ping, Y. (2020). Genome editing of mutant KRAS through supramolecular polymer-mediated delivery of Cas9 ribonucleoprotein for colorectal cancer therapy. *Journal of Controlled Release*, *322*, 236–247. https://doi.org/10.1016/J.JCONREL.2020.03.015

Wang, D. D., Jin, Q., Wang, L. L., Han, S. F., Chen, Y. B., Sun, G. D., Sun, S. F., Sun, S. W., Wang, T., Liu, F. J., Wang, P., & Shi, B. (2017). The significance of ENAH in carcinogenesis and prognosis in gastric cancer. *Oncotarget*, *8*(42), 72466–72479. https://doi.org/10.18632/ONCOTARGET.19801

Wang, H. Z., Wang, F., Chen, P. F., Zhang, M., Yu, M. X., Wang, H. L., Zhao, Q., & Liu, J. (2019). Coexpression network analysis identified that plakophilin 1 is associated with the metastasis in human melanoma. *Biomedicine & Pharmacotherapy*, *111*, 1234–1242. https://doi.org/10.1016/J.BIOPHA.2018.12.135

Wang, M., Herbst, R. S., & Boshoff, C. (2021). Toward personalized treatment approaches for non-small-cell lung cancer. *Nature Medicine*, *27*, 1345–1356. https://doi.org/10.1038/s41591-021-01450-2

Wang, X., Ramat, A., Simonelig, M., & Liu, M. F. (2022). Emerging roles and functional mechanisms of PIWI-interacting RNAs. *Nature Reviews Molecular Cell Biology*, *24*(2), 123–141. https://doi.org/10.1038/s41580-022-00528-0

Wang, Z., Jin, Y., Ren, H., Ma, X., Wang, B., & Wang, Y. (2016). Downregulation of the long non-coding RNA TUSC7 promotes NSCLC cell proliferation and correlates with poor prognosis. *American Journal of Translational Research*, *8*(2), 680. /pmc/articles/PMC4846917/

Welch, C., Chen, Y., & Stallings, R. L. (2007). MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene*, *26*, 5017–5022. https://doi.org/10.1038/sj.onc.1210293

Wild, C. P., Weiderpass, E., & Stewart, B. W. (2020). *World Cancer Report: Cancer Research for Cancer Prevention*. International Agency for Research on Cancer (IARC). http://publications.iarc.fr/586

Wolf, A., & Hatzfeld, M. (2010). A role of plakophilins in the regulation of translation. *Cell Cycle*, *9*(15), 3045–3050. https://doi.org/10.4161/CC.9.15.12446

Wolff, J. H., Haldrup, J., Thomsen, E. A., Andersen, S., & Mikkelsen, J. G. (2021). piggyPrime: High-Efficacy Prime Editing in Human Cells Using piggyBac-Based DNA Transposition. *Frontiers in Genome Editing*, *3*. https://doi.org/10.3389/FGEED.2021.786893

Wolff, J. H., & Mikkelsen, J. G. (2022). Delivering genes with human immunodeficiency virus-derived vehicles: still state-of-the-art after 25 years. *Journal of Biomedical Science*, *29*(1). https://doi.org/10.1186/S12929-022-00865-4

Wu, C. H., & Hwang, M. J. (2019). Risk stratification for lung adenocarcinoma on EGFR and TP53 mutation status, chemotherapy, and PD-L1 immunotherapy. *Cancer Medicine*, *8*(13), 5850. https://doi.org/10.1002/CAM4.2492

Wu, J., He, K., Zhang, Y., Song, J., Shi, Z., Chen, W., & Shao, Y. (2019). Inactivation of SMARCA2 by promoter hypermethylation drives lung cancer development. *Gene*, *687*, 193–199. https://doi.org/10.1016/J.GENE.2018.11.032

Xue, J. Y., Zhao, Y., Aronowitz, J., Mai, T. T., Vides, A., Qeriqi, B., Kim, D., Li, C., de Stanchina, E., Mazutis, L., Risso, D., & Lito, P. (2020). Rapid non-uniform adaptation to conformation-specific KRASG12C inhibition. *Nature*, *577*(7790), 421. https://doi.org/10.1038/S41586-019-1884-X

Xue, Z., Hennelly, S., Doyle, B., Gulati, A. A., Novikova, I. V., Sanbonmatsu, K. Y., & Boyer, L. A. (2016). A G-rich motif in the lncRNA Braveheart interacts with a zinc finger transcription factor to specify the cardiovascular lineage. *Molecular Cell*, *64*(1), 37. https://doi.org/10.1016/J.MOLCEL.2016.08.010

Yakubovskaya, E., Guja, K. E., Mejia, E., Castano, S., Hambardjieva, E., Choi, W. S., & Garcia-Diaz, M. (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. *Structure (London, England : 1993)*, *20*(11), 1940. https://doi.org/10.1016/J.STR.2012.08.027

Yang, C., Fischer-Kešo, R., Schlechter, T., Ströbel, P., Marx, A., & Hofmann, I. (2015). Plakophilin 1-deficient cells upregulate SPOCK1: implications for prostate cancer progression. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine*, *36*(12), 9567–9577. https://doi.org/10.1007/S13277-015-3628-3

Yang, G., Fu, H., Zhang, J., Lu, X., Yu, F., Jin, L., Bai, L., Huang, B., Shen, L., Feng, Y., Yao, L., & Lu, Z. (2010). RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology*, *138*(1). https://doi.org/10.1053/J.GASTRO.2009.08.001

Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., & Xia, N. (2020). Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *International Journal of Molecular Sciences*, *21*(18), 1–20. https://doi.org/10.3390/IJMS21186461

Yang, L., Chen, R., Goodison, S., & Sun, Y. (2021). An efficient and effective method to identify significantly perturbed subnetworks in cancer. *Nature Computational Science*, *1*(1), 79–88. https://doi.org/10.1038/s43588-020-00009-4

Yang, Y., Wang, J., Xu, S., Shi, F., & Shan, A. (2021). Calumenin contributes to epithelial-mesenchymal transition and predicts poor survival in glioma. *Translational Neuroscience*, *12*(1), 67. https://doi.org/10.1515/TNSCI-2021-0004

Yip, B. H. (2020). Recent Advances in CRISPR/Cas9 Delivery Strategies. *Biomolecules*, *10*(6). https://doi.org/10.3390/BIOM10060839

Yoon, J. H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., Huarte, M., Zhan, M., Becker, K. G., & Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Molecular Cell*, *47*(4), 648. https://doi.org/10.1016/J.MOLCEL.2012.06.027

Yu, X., Li, Z., Zheng, H., Matthew, |, Chan, T. V, William, |, & Wu, K. K. (2017). NEAT1: A novel cancer-related long non-coding RNA. *Cell Proliferation*, *2*. https://doi.org/10.1111/cpr.12329

Yuan, S., Xiang, Y., Guo, X., Zhang, Y., Li, C., Xie, W., Wu, N., Wu, L., Cai, T., Ma, X., Yu, Z., Bai, L., & Li, Y. (2020). Circulating Long Noncoding RNAs Act as Diagnostic Biomarkers in Non-Small Cell Lung Cancer. *Frontiers in Oncology*, *10*, 2609. https://doi.org/10.3389/FONC.2020.537120

Yun, J., Nakagawa, R., & Tham, K. (2023). KRAS-targeted therapy in the treatment of non-small cell lung cancer. *Journal of Oncology Pharmacy Practice*, *29*(2), 422–430. https://doi.org/10.1177/10781552221118848

Zampetaki, A., Albrecht, A., & Steinhofel, K. (2018). Long Non-coding RNA Structure and Function: Is There a Link? *Frontiers in Physiology*, *9*(AUG). https://doi.org/10.3389/FPHYS.2018.01201

Zhang, L., Ging, N. C., Komoda, T., Hanada, T., Suzuki, T., & Watanabe, K. (2005). Antibiotic susceptibility of mammalian mitochondrial translation. *FEBS Letters*, *579*(28), 6423–6427. https://doi.org/10.1016/J.FEBSLET.2005.09.103

Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y., & Klibanski, A. (2010). Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid: Isoform Structure, Expression, and Functions. *Endocrinology*, *151*(3), 939. https://doi.org/10.1210/EN.2009-0657

Zhao, F., Zhou, L. H., Ge, Y. Z., Ping, W. W., Wu, X., Xu, Z. Le, Wang, M., Sha, Z. L., & Jia, R. P. (2019). MicroRNA-133b suppresses bladder cancer malignancy by targeting TAGLN2-mediated cell cycle. *Journal of Cellular Physiology*, *234*(4), 4910–4923. https://doi.org/10.1002/JCP.27288

Zhao, H., Ren, D., Liu, H., & Chen, J. (2018). Comparison and discussion of the treatment guidelines for small cell lung cancer. *Thoracic Cancer*, *9*(7), 769–774. https://doi.org/10.1111/1759-7714.12765

Zhao, X., Liu, L., Lang, J., Cheng, K., Wang, Y., Li, X., Shi, J., Wang, Y., & Nie, G. (2018). A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment. *Cancer Letters*, *431*, 171–181. https://doi.org/10.1016/J.CANLET.2018.05.042

Zhao, Y., Zhang, G., Wei, M., Lu, X., Fu, H., Feng, F., Wang, S., Lu, W., Wu, N., Lu, Z., & Yuan, J. (2014). The tumor suppressing effects of QKI-5 in prostate cancer: a novel diagnostic and prognostic protein. *Cancer Biology & Therapy*, *15*(1), 108–118. https://doi.org/10.4161/CBT.26722

Zhou, H., Chen, A., Shen, J., Zhang, X., Hou, M., Li, J., Chen, J., Zou, H., Zhang, Y., Deng, Q., She, K., Shi, X., & He, J. (2019). Long non-coding RNA LOC285194 functions as a tumor suppressor by targeting p53 in non-small cell lung cancer. *Oncology Reports*, *41*(1), 15–26. https://doi.org/10.3892/OR.2018.6839

Zubakov, D., Stupar, Z., & Kovacs, G. (2006). Differential expression of a new isoform of DLG2 in renal oncocytoma. *BMC Cancer*, *6*(1), 1–8. https://doi.org/10.1186/1471-2407-6-106

Zucchelli, S., Fasolo, F., Russo, R., Cimatti, L., Patrucco, L., Takahashi, H., Jones, M. H., Santoro, C., Sblattero, D., Cotella, D., Persichetti, F., Carninci, P., & Gustincich, S. (2015). SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Frontiers in Cellular Neuroscience*, *9*(MAY), 174. https://doi.org/10.3389/FNCEL.2015.00174

Annex

Annex

Publications during the PhD

*Shared first-authorship is denoted by **

Original articles

1. Álvarez-Pérez JC*, Sanjuán-Hidalgo J*, **Arenas AM***, Andrades A, Medina PP. High-fidelity Cas9-mediated targeting of KRAS driver mutations restrains lung cancer in preclinical models. [*Under review*]

2. **Arenas AM**, Ruiz-Jiménez J, López-Hidalgo JL, Sanjuán-Hidalgo J, Medina PP. Defining a *bona fide* cell model for SMARCA4-deficient, undifferentiated tumor. [*Under review*]

3. Boyero L, Martin-Padron J, Fárez-Vidal ME, Rodriguez MI, Andrades A, Peinado P, **Arenas AM**, Ritoré-Salazar F, Alvarez-Perez JC, Cuadros M, Medina PP. PKP1 and MYC create a feedforward loop linking transcription and translation in squamous cell lung cancer. *Cellular Oncology*. 2022. doi: 10.1007/s13402-022-00660-1.

4. Cuadros M*, García DJ*, Andrades A*, **Arenas AM**, Coira IF, Baliñas-Gavira C, Peinado P, Rodríguez MI, Álvarez-Pérez JC, Ruiz-Cabello F, Camós M, Jiménez-Velasco A, Medina PP. LncRNA-mRNA Co-Expression Analysis Identifies AL133346.1/CCN2 as Biomarkers in Pediatric B-Cell Acute Lymphoblastic Leukemia. *Cancers*. 2020; 12(12):3803. doi: https://doi.org/10.3390/cancers12123803.

5. Peinado P*, Andrades A*, Cuadros M*, Rodriguez MI*, Coira IF, Garcia DJ, Álvarez-Perez JC, Baliñas-Gavira C, **Arenas AM**, Patiño-Mercau JR, Sanjuan-Hidalgo J, Romero OA, Montuenga LM, Carretero J, Sanchez-Cespedes M, Medina PP. Comprehensive Analysis of SWI/SNF Inactivation in Lung Adenocarcinoma Cell Models. *Cancers*. 2020; 12(12):3712. doi: https://doi.org/10.3390/cancers12123712.

6. **Arenas AM***, Cuadros M*, Andrades A, García DJ, Coira IF, Rodríguez MI, Baliñas-Gavira C, Peinado P, Álvarez-Pérez JC, Medina PP. LncRNA DLG2-AS1 as a Novel Biomarker in Lung Adenocarcinoma. *Cancers*. 2020; 12(8):2080. doi: 10.3390/cancers12082080.

7. Cuadros M*, Andrades A*, Coira IF, Rodríguez MI, Álvarez-Pérez JC, Peinado P, **Arenas AM**, García DJ, Jiménez P, Camós M, Jiménez-Velasco A,

Medina PP. Expression of the long non-coding RNA TCL6 is associated with clinical outcome in pediatric B-cell acute lymphoblastic leukemia. *Blood Cancer J.* 2019; 9, 93. doi: 10.1038/s41408-019-0258-9. Impact Factor 2020: 8.023 (Q1).

Reviews and book chapters

1. Andrades A*, Peinado P*, Álvarez-Pérez JC*, Sanjuán-Hidalgo J, García DJ, **Arenas AM**, Matia-González AM, Medina PP. SWI/SNF complexes in hematological malignancies: biological implications and therapeutic opportunities. *Mol Cancer* 2023, 22,39. doi: 10.1186/s12943-023-01736-8.

2. **Arenas AM***, Andrades A*, Patiño-Mercau JR, Sanjuan-Hidalgo J, Cuadros M, García DJ, Peinado P, Rodríguez MI, Baliñas-Gavira C, Álvarez-Pérez C, Medina PP. Opportunities of miRNAs in cancer therapeutics. In "*MicroRNA in Human Malignancies, 1st Edition*"; Negrini M, Calin G, Croce C (Eds); Elsevier, Feb 17, 2022; pp. 153-164. ISBN: 9780128222874

Funding and fellowships

During this thesis, the PhD candidate has been supported by the following fellowships:

- **FPU17/01258** contract (*Programa de ayudas para la formación de profesorado universitario*) from the Spanish Ministry of Universities.

- European Molecular Biology Organization (EMBO) short-term fellowship (**STF_9006**), for a 3-months visit to Prof. Jacob Giehm Mikkelsen's research group at the Aarhus University, Denmark.

In addition, the PhD candidate has been a member of, and this thesis has received funding from the projects SAF2015-67919-R, AECC-2018, B-CTS-480-UGR20, and PID2021-126111OB-I00. The PhD candidate was also granted a project as a co-PI from the University of Granada's *Programa de Proyectos de Investigación Precompetitivos para Jóvenes Investigadores Modalidad 20.b* (PPJIB2020.22). The PhD candidate would also like to thank all the generous donations made to the *Aula de Investigación Contra la Leucemia Infantil "Héroes Contra la Leucemia"*.

Copyright permissions

Chapter 1, Figure 1. Extracted from World Cancer Report 2020. Permission granted by the International Agency for Research on Cancer (IARC) / World Health Organization (WHO), on 06/02/2023 after online request.

Sylvia Lesage<lesages@iarc.who.int> en nombre de
IARC Publications<Publications@iarc.who.int>
lun 06/02/2023 12:00

Para:Alberto Manuel Arenas Molina <alberto.arenas@genyo.es>;

Dear Mr Arenas Molina,

Thank you for your request for permission to reproduce, reprint or translate certain IARC/WHO copyrighted material.

On behalf of the International Agency for Research on Cancer/World Health Organization, we are pleased to authorize your request to reproduce the IARC/WHO materials as detailed in the form below, subject to the terms and conditions of the non-exclusive licence below.

If you have questions regarding this authorization, please contact publications@iarc.fr.

We thank you for your interest in IARC/WHO published materials.

Kind regards,
IARC Permissions team