A Programmable RNA Knockdown Using CRISPR-Cas13 Ribonuclease Towards Gene Therapy

SAIFULLAH

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

A Programmable RNA Knockdown Using CRISPR-Cas13 Ribonuclease Towards Gene Therapy

SAIFULLAH

Supervisor: Professor Dr. Toshifumi Tsukahara

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
[Materials Science]
September, 2021

Abstract

Recent approvals in gene therapy have paved the road for an extensive second upsurge of therapies and forefront the groundwork for next-generation treatment strategies. CRISPR-Cas effectors have flourished as an overwhelming tool that can potentially endeavor future genetic medicine. In this doctoral thesis, I investigated the therapeutic potential of anaplastic lymphoma kinase (ALK) and developed optimal parameters for programmable RNA knockdown using CRISPR-Cas13a ribonucleoprotein. Furthermore, I applied the optimized protocol for echinoderm microtubule-associated protein-like 4 (EML4)-*ALK* transcript knockdown as a proof-of-principle for RNA-based cancer therapy. More explicitly, this dissertation consists of five chapters. Where, chapter one summarizes the recent development in RNA-targeted therapeutics including RNAi and CRISPR-Cas systems, and discusses the landscape of *EML4-ALK*-positive lung cancer treatment. This summarization of recent work will assist to understand the recent advancement of RNA-targeted CRISPR-Cas technologies and underpinning the contemporary *ALK*-positive cancer therapeutics.

In the second chapter, I explored the clinical outcome and associated genes of *ALK* expression using integrative bioinformatics. ALK is a tyrosine kinase receptor that is genetically altered in several cancers, such as non-small cell lung cancer (NSCLC), melanoma, lymphoma, and other tumors. Although *ALK* is associated with various cancers, the relationship between *ALK* expression and patient prognosis in different cancers is poorly understood. Here, I show a correlation between *ALK* expression and its clinical outcome in patients with lung adenocarcinoma (LUAD), melanoma, ovarian carcinoma (OV), diffuse large B-cell lymphoma (DLBC), acute myeloid leukemia (AML), and breast cancer (BC) using different computational assessments. I analyzed *ALK* transcriptional expression, patient survival rate, genetic alteration, protein network, and gene and microRNA (miRNA) co-expression. I found that deregulated expression of *ALK* is associated with a high mortality rate in *ALK*-positive cancers. I identified 214 missense mutations, 24 truncating mutations, seven fusions, and two in-frame mutations, with the highest alteration of *ALK* in melanoma. I further showed that 17 genes and 19 miRNAs were exclusively co-expressed and found that *EML4* was the most positively correlated gene. The gene ontology and signaling pathways of the genes co-expressed with *ALK* involved in these six cancers were also identified. My findings offer a basis for *ALK* as a prognostic biomarker and therapeutic target in cancers, which will potentially contribute to precision oncology.

In the third chapter, I developed optimal parameters for programmable and effective RNA knockdown using marker genes including firefly luciferase and mCherry transcripts. RNAi technology has noteworthy potential as a future medicine and could ideally be used to knock down disease-related RNAs. However, owing to frequent off-target effects, limited accessibility of nuclear transcripts, and low efficiency, the medical application of the technique remains challenging. Here, I first evaluated the stability of the Cas13a transcript and guide RNA. Next, I optimized the Cas13a and guide RNA expression vectors to achieve effective knockdown of firefly luciferase (FLuc) transcript, used as a target RNA. The knockdown specificity of Cas13a on target-search was next examined. I found that the 1:3 molar ratio concentration of Cas13a and guide RNA vector is preferable for effective knockdown than the vector amount. Based on the Cas13a selectivity results, I observed restricted endonuclease activity in 3′ crRNA–gRNA orientation. I also found the highest activity between 24–30 bp long gRNAs with limited mismatch tolerance. Cas13a could effectively knock

down FLuc luminescence (70–76%), and mCherry fluorescence (72%). Accordingly, Cas13a has strong potential for use in RNA knockdown and regulation.

Next, chapter fourth showed the feasibility of Cas13a ribonuclease in downregulation of oncogenic driver *EML4-ALK* expression in human disease model lung cancer cells. ALK tyrosine kinase inhibitors provoke a significant anti-tumor response; however, they inevitably succumb to the acquired resistance. Therefore, an alternative therapeutic strategy that limits *ALK* over-activation is necessary for the treatment of this lung cancer. Here, I show that the CRISPR-Cas13a effector possesses effective knockdown potency for oncogenic driver *EML4-ALK* transcript in lung cancer cells. I found the *EML4* transcript was not substantially expressed but *ALK* expressed 80–100 fold higher in *ALK*-positive lung cells compared to a non-fusion transcript in HEK293T cells. I also found that the EML4-ALK oncoproteins were robustly down-regulated (>80%) by employing Cas13a in those lung cancer cells based on western blot results. Consequently, the tyrosine kinase phosphorylation (50–70%) and cell growth (up to 40%) were inhibited. Overall the obtained data demonstrated that the CRISPR-Cas13a protein downregulated the ALK expression in the lung cancer cells. Thus, CRISPR-Cas13a mediated *EML4-ALK* RNA knockdown devises a potential therapeutic strategy for treating *ALK*-positive lung cancer.

Finally, chapter V recapitulates the total work and discusses the benefits, challenges, and future directions. In conclusion, Cas13a has strong potential for use in RNA regulation and could contribute to the development of next-generation genetic medicine.

Keywords: Prognosis of *ALK* expression; RNA knockdown; CRISPR-Cas13a; FLuc transcript; *EML4-ALK*-positive lung cancer.

Table of Content

Index	Page No.
Abstract……	i-ii
List of Tables and Figures……..	viii-xi
List of Abbreviation……	xii-xiii
Chapter I: General introduction: Programmable RNA knockdown technologies in therapeutics and the landscape of *EML4-ALK*-positive lung cancer treatment	
1. Outline…….	1
Part-I: Programmable molecular scissors for RNA engineering in therapeutics	
1.1. Current scenario of conventional therapy versus gene therapy...	1
1.2. Programmable RNA knockdown strategies in therapeutics……	3
1.2.1. RNAi……	3
1.2.1.1. Antisense Oligonucleotides (ASOs)……..	3
1.2.1.2. micro-RNA (miRNA)……	3
1.2.1.3. small interfering RNA (siRNA)…….	4
1.2.1.4. Strengths and challenges of RNAi	5
1.2.2. CRISPR-Cas system: discovery, diversity, and abundance…..	5
1.2.2.1 Discovery of CRISPR: From yogurt to novel prize……	5-7
1.2.2.2. Diversity and abundance……	7
1.2.2.3. RNA targeted CRISPR effectors……..	8-13
1.2.2.3.1. Type III……	9
1.2.2.3.2. Cas9 (Type II)……...	9
1.2.2.3.3. Cas12g (Type V-G)……..	11
1.2.2.3.4. Cas13 (Type VI)……...	11
1.2.2.3.5. Strengths and challenges of CRISPR-Cas……	12
Part-II: Molecular genetics and therapeutic trends of the EML4-ALK-positive lung cancer	
1.3. EML4 in human……..	14
1.4. ALK in human……..……...	14

1.5. Molecular genetics of *EML4-ALK* in lung cancer……. 16
1.6. Molecular signaling pathways……. 17
1.7. Contemporary treatment systems for *ALK*-positive lung cancer…….. 17
1.7.1. Rationale for ALK inhibitor…… 17
1.7.1.1. First generation……..... 17
1.7.1.2. Second generation……. 17
1.7.1.3. Third generation……..... 18
1.7.2. Chemotherapy…….. 18-20
Part-III: Problem identification and justification, and objectives
1.8. Problem identification and justification……... 20
1.8.1. Why is a new ALK-positive lung cancer treatment strategy necessary?..... 20-22
1.8.2. Which can be a future suitable strategy for ALK-positive cancer treatment?..... 23
1.9. Objective of this study…….. 24
1.10. Bibliography…….. 25-32

Chapter II: Exploration of clinical consequence and associated genes of *ALK* expression in cancers using integrative bioinformatics
2. Outline……..... 33
2.1. Introduction……..……. 34
2.2. Materials and Methods…….…… 35-37
2.2.1. Experimental design……... 35
2.2.2. Expression profile of ALK across human normal and cancer tissues……. 36
2.2.3. Significant ALK expression variation in different cancer types…….. 36
2.2.4. Prognostic investigation of ALK mRNA expression in patients with cancer…… 36
2.2.5. Evaluation of ALK alteration and associated cancers…… 36
2.2.6. ALK protein network analysis and clinical significance in cancers……. 37

2.2.7. Profiling of genes and non-coding RNAs co-occurred and co-expressed with ALK…… 37
2.2.8. Elucidation of Gene ontologies (GOs) and signaling pathways of ALK and its correlated genes……... 37
2.3. Results……. 38-53
2.3.1. Expression profile of ALK across human normal and cancer tissues…….. 38
2.3.2. Significant variation in ALK expression in different cancer types…… 40
2.3.3. Prognostic investigation of ALK mRNA expression in patients with cancer…….. 42
2.3.4. Evaluation of ALK alteration in associated cancers……. 44
2.3.5. ALK protein network analysis and clinical significance in cancers…….. 46
2.3.6. Profiling of genes and micro RNAs co-occurred and co-expressed with ALK……. 48
2.3.7. Gene ontologies and signaling pathway elucidation of ALK and its correlated genes…… 52
2.4. Discussion…… 54-57
2.5. Concluding remark……..... 57
2.6. Bibliography……... 57-65
2.7. Supplementary…… 65-68

Chapter III: Optimization of CRISPR-Cas13a mediated effective RNA knockdown activity using luciferase as a target transcript

3. Outline…….……. 69
3.1. Introduction…….…….. 70
3.2. Materials and Methods……. 71-78
3.2.1. Experimental design……... 71
3.2.2. Plasmid vectors…….. 72
3.2.3. Guide RNA design……. 73
3.2.4. Sanger sequencing…….. 75
3.2.5. Cell culture and transfection…….. 75

3.2.6. RNA stability assay…… 75
3.2.7. Total RNA extraction……. 76
3.2.8. cDNA synthesis…….. 76
3.2.9. RT-PCR quantification……... 76
3.2.10. SYBR green-based qPCR……..……... 77
3.2.11. Luciferase luminometry assay…….. 77
3.2.12. Data analysis……. 78
3.3. Results……. 78-85
3.3.1. Analysis of RNA stability…….. 78
3.3.2. Optimizing the quantity of Cas13a for efficient RNA knockdown…… 80
3.3.3. Optimal condition for RNA knockdown and preference of 5′ crRNA for Cas13a activity……... 82
3.3.4. Selectivity of Cas13a in guide RNA design…… 83
3.3.5. Validation of effective knockdown: mCherry…… 85
3.4. Discussion…… 86-88
3.5. Concluding remark……..... 88
3.6. Bibliography……... 88-91
3.7. Supplementary……... 92

Chapter IV: The CRISPR-Cas13a gene-editing system downregulated the oncogenic driver *EML4-ALK* expression in *ALK*-positive lung cancer cells

4. Outline…… 93
4.1. Introduction…… 94
4.2. Materials and Methods……. 95-99
4.2.1. Experimental outline…….. 95
4.2.2. Expression vector and guide RNA design……. 95
4.2.3. DNA Sequencing……... 96
4.2.4. Cell culture and transfection…….. 97
4.2.5. Total RNA extraction……. 97
4.2.6. cDNA synthesis…… 97

4.2.7. SYBR green-based qPCR…….. 98
4.2.8. MTT assay……. 98
4.2.9. Control drug……... 98
4.2.10. Western blot analyses……..…… 98
4.2.11. Computational assessment and data analysis……... 99
4.3. Results……. 99-108
4.3.1. Confirmation of *EML4-ALK* in lung cancer cells 99
4.3.2. Expression analysis and clinical significance of *EML4-ALK* transcript…… 101
4.3.3. CRISPR-Cas13a downregulated the target ALK in a programmable manner……... 103
4.3.4. Programmable knockdown of oncogenic EML4-ALK v1 oncoprotein in H3122 lung cancer cells……. 105
4.3.5. Programmable downregulation of oncogenic EML4-ALK v3a/b expression in H2228 lung cancer cells……. 107
4.4. Discussion……... 109-110
4.5. Concluding remark…… 111
4.6. Bibliography……... 111-114
4.7. Supplementary…… 115

Chapter V: Final discussion
5.1. Summary…… 116-120
5.2. Future direction……..... 120-121

Achievement……. 122-123

Acknowledgement…… 124

List of Tables and Figures

List of Tables

Index	Page No.
Chapter II: Exploration of clinical consequence and associated genes of *ALK* expression in cancers using integrative bioinformatics	
Table 1. Significant co-occurring protein partners of ALK gene signature were obtained using cBioPortal for cancer genomics…….....	48
Supplementary table S1. Relationship between anaplastic lymphoma kinase (ALK) expression and survival in various cancers…….....	68
Chapter III: Optimization of CRISPR-Cas13a mediated effective RNA knockdown activity using luciferase as a target transcript	
Table 1. Oligo sequences used for PCR amplification reaction……..…	73
Table 2. crRNA and guide RNA sequences……	74
Chapter IV: The CRISPR-Cas13a gene-editing system downregulated the oncogenic driver *EML4-ALK* expression in *ALK*-positive lung cancer cells	
Table 1. crRNA and guide RNA sequences……	96
Table 2. Genetic fusion of *ALK* with other mRNAs involved in different types of cancers, based on data retrieved from the TCGA PanCancer Atlas using cBioPortal (10,953 patients/10,967 samples)…	103

List of Figures

Index	Page No.
Chapter I: General introduction: Programmable RNA knockdown technologies in therapeutics and the landscape of *EML4-ALK*-positive lung cancer treatment	
Figure 1. Mechanism of RNAi mediated ASO, miRNA, and siRNA biogenesis in therapeutics……...…….	4
Figure 2. Road to the success of CRISPR engineering, from yogurt to novel prize…….……	6
Figure 3. General structural overview of CRISPR-Cas classification…..	7

Figure 4. Structural module of RNA cleavage CRISPR-Cas and CRISPR array...……...……. 8
Figure 5. The mood of action of all CRISPR-Cas effectors which are involved in RNA degradation……...…… 10
Figure 6. Generalized mechanism of CRISPR-Cas13 mediated target RNA knockdown…...……... 13
Figure 7. Functional domain and motifs in EML4-ALK oncofusion….. 15
Figure 8. Molecular genetic variants and frequency of EM4-ALK oncofusion……...……. 16
Figure 9. Commonly used inhibitors and chemotherapeutic molecules for EML4-ALK positive lung cancers…….. 17
Figure 10. The death rate for all cancer types across the world estimated by WHO in 2020…….……. 20
Figure 11. ALK-dependent acquired resistance generated by inhibitors 22

Chapter II: Exploration of clinical consequence and associated genes of *ALK* expression in cancers using integrative bioinformatics

Figure 1. Expression profile of anaplastic lymphoma kinase (ALK) across human normal and cancer tissues…….. 39
Figure 2. Anaplastic lymphoma kinase (ALK) expression pattern in various cancer types……...…….. 41
Figure 3. Anaplastic lymphoma kinase (ALK) expression and clinical prognosis in six different cancers retrieved from PrognoScan microarray cancer database…….. 43
Figure 4. Data on genetic alteration of anaplastic lymphoma kinase (ALK) and consequent cancers in patients derived from The Cancer Genome Atlas (TCGA) PanCancer Atlas database through cBioPortal.. 45
Figure 5. Protein network of anaplastic lymphoma kinase (ALK) and their clinical significance in six cancers……..…….…… 47
Figure 6. Co-expression analysis of anaplastic lymphoma kinase (ALK) and its 17 mutually associated functional protein partner genes in six cancer types.……... 49
Figure 7. Hierarchical clustering and similarity matrix analysis of

ALK and the expression of correlated genes in six cancer types.....….... 50
Figure 8. MicroRNAs (miRNAs) co-expressed with ALK and its mutually exclusive genes in humans…… 51
Figure 9. Gene ontology and pathway analyses of anaplastic lymphoma kinase (ALK) and its co-related functional protein partners.. 53
Figure S1. Anaplastic lymphoma kinase (ALK) expression and clinical prognosis consequence in non-small-cell lung carcinoma (NSCLC) and breast cancer……. 65
Figure S2. Oncoprint of anaplastic lymphoma kinase (ALK) and common genes…...……..……. 66
Figure S3. MicroRNA (miRNA) network with anaplastic lymphoma kinase (ALK) and 17 common target genes……...…….. 67

Chapter III: Optimization of CRISPR-Cas13a mediated effective RNA knockdown activity using luciferase as a target transcript

Figure 1. Analysis of RNA stability using a transcription inhibitor…… 79
Figure 2. Quantitative analysis of Cas13a and guide RNA……...….. 81
Figure 3. RNA knockdown and preference of Cas13a functionality for 5′ crRNA–guide RNA……...……... 83
Figure 4. Effect of guide RNA design on Cas13a……... 84
Figure 5. mCherry mRNA knockdown by Cas13a……..……... 85
Figure S1. pcDNA3-Cas13a construct map, RT-PCR gel image of Cas13a mRNA and crRNA, and Phylogenetic tree of Luciferin……..... 92

Chapter IV: The CRISPR-Cas13a gene-editing system downregulated the oncogenic driver *EML4-ALK* expression in *ALK*-positive lung cancer cells

Figure 1. Validation of *EML4-ALK* oncofusion in H3122 and H2228 lung cancer cells…… 100
Figure 2. *EML4* and *ALK* mRNA abundance in HEK293T and lung cancer cells, and the clinical significance of ALK expression in lung adenocarcinoma (LUAD)……...…….. 102
Figure 3. CRISPR-Cas13a downregulated the target *ALK* in a

programmable manner……..…….. 104
Figure 4. Programmable knockdown of oncogenic EML4-ALK oncoprotein in H3122 lung cancer cells…….. 106
Figure 5. Programmable knockdown of oncogenic EML4-ALK oncoprotein in H2228 lung cancer cells……...…… 108
Figure S1. Cytotoxicity and specificity of Cas13a in H3122 cells…….. 115

Chapter V: Final discussion

Figure 1. Graphical representation of entire work for Chapter II…… 117
Figure 2. State-of-art of the proposed therapeutic for EML4-ALK-positive lung cancer cells described in Chapter IV……..…… 119

List of Abbreviation

Short form	*Full form*
RNA	Ribonucleic acid
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPR-Cas	CRISPR associated effector
Cas13	CRISPR-associated protein 13
ALK	Anaplastic lymphoma kinase
EML4	Echinoderm microtubule-associated protein-like 4
miRNA	micro RNA
crRNA	CRISPR RNA
gRNA	Guide RNA
FLuc	Firefly luciferase
bp	base pair
FDA	The United States Food and Drug Administration
DNA	Deoxyribonucleic acid
RNAi	RNA interference
siRNA	small interfering RNA
shRNA	short-hairpin RNA
ASO	Antisense oligonucleotide
dsRNA	Double stranded RNA
mRNA	messenger RNA
RISC	RNA induced silencing complex
saRNA	small activating RNA
Cas9	CRISPR-associated protein 9
PAM	Protospacer adjacent motif
DR	Direct repeat
HEPN	Higher Eukarytoes and Prokaryotes Nucleotide
HELP	hydrophobic EML protein
TK	Tyrosine Kinase
NSCLC	Non-small-cell lung carcinoma
LUAD	Lung Adenocarcinoma
PFS in aspect of CRISPR-Cas13	Protoscpacer flanking site
PFS in aspect of cancer	Progression free survival
WHO	World Health Organization
OV	Ovarian carcinoma
DLBC	Diffused large B-cell lymphoma
AML	Acute myeloid leukemia

BC	Breast cancer
TCGA	The Cancer Genome Atlas
KEGG	Kyoto Encyclopedia of Genes and Genomes
PANTHER	protein analysis through evolutionary relationships
HR	Hazard Ratio
P < 0.05	Statistically significant alpha value
HEK 293	Human embryonic kidney 293
A, C, G or T	Nucleotide
et al.	et alliori (and others)
mRNA	Messenger Ribonucleic acid
IncRNA	Long non-coding Ribonucleic acid
PCR	Polymerase Chain Reaction
qPCR	Quantitative PCR
RT-PCR	Reverse Transcriptase PCR
e.g.	Exempli gratia
˚C	Degree Celsius
h	hour
min	Minute
s	Second
kb	Killo base pair
MS	Microsoft
$MgCl_2$	Magnesium chloride
etc.	Et cetera
dNTP	Deoxynucleotide Tri-phosphate
µL	Micro liter
µm	Micro miter
mL	Mili liter
pmol	Pico-mole
ng	Nano-gram
M	Molar
M	DNA ladder marker
SDS	Sodium dodecyl sulfate
g	gravitational force
R^2	Correlation coefficient
~	Near about
NCBI	National Center for Biotechnology Information
PAGE	Polyacrylamide gel electrophoresis
NEB	New England Biolabs
FBS	Fetal Bovine Serum
DMEM	Dulbecco's Modified Eagle's Medium
EDTA	Ethylenediaminetetraacetic acid

Chapter I

General Introduction: Programmable RNA knockdown technologies in therapeutics and the landscape of *EML4-ALK*-positive lung cancer treatment

Graduate School of Advanced Science and Technology
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (JAIST), JAPAN

Chapter I

General introduction: Programmable RNA knockdown technologies in therapeutics and the landscape of *EML4-ALK*-positive lung cancer treatment

1. Outline

Recent appreciation of two gene therapy medicines by FDA (United State of America Food and Drug Administration) has paved the way for an embracing second rise of therapies and forefront the basic research for next-generation technologies. CRISPR-Cas effectors have thrived as a devastating genome editing tool that can potentially endeavor future genetic medicine. In this general introduction, I summarize the recent development in RNA targeted therapeutics including RNAi and CRISPR-Cas systems, and discuss insight into wide potential and challenges. Furthermore, the contemporary therapeutic systems of *EML4-ALK*-positive cancer are comprehensively reviewed and illustrate which CRISPR-Cas system is the most suitable for next-generation RNA targeted therapeutics. This general introduction will help to understand the recent advancement of RNA-targeted CRISPR-Cas technologies and the foundation for contemporary *EML4-ALK*-positive cancer therapeutics.

Keywords: RNA knockdown; RNAi; CRISPR-Cas; *EML4-ALK* fusion gene.

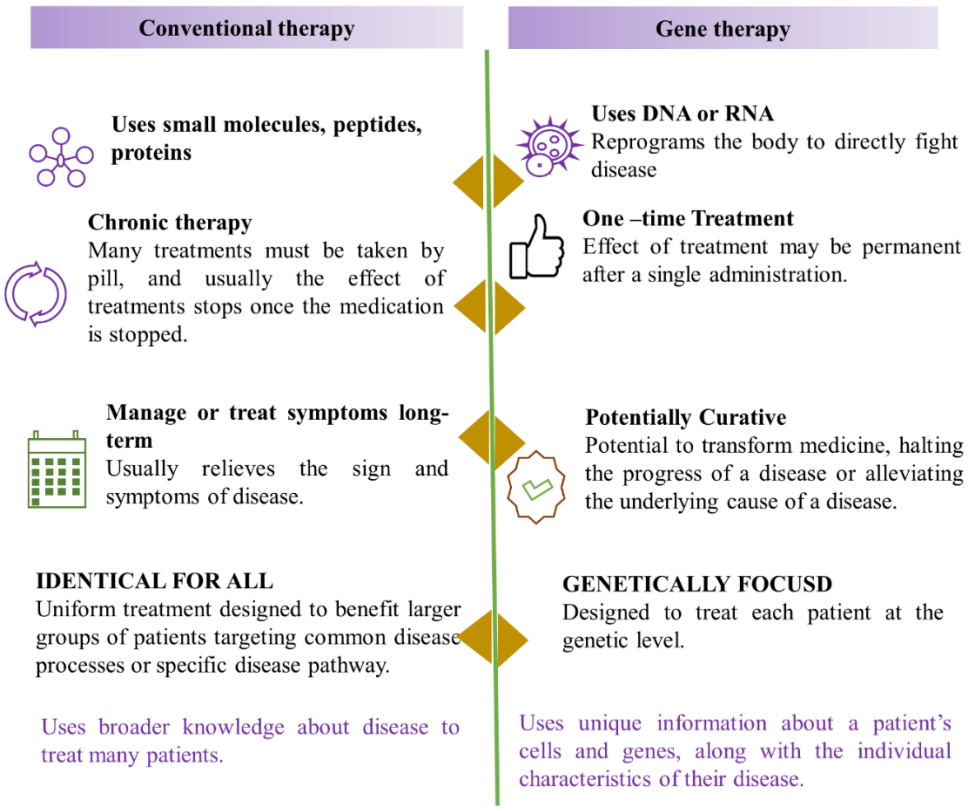
Part-I: Programmable molecular scissors for RNA engineering in therapeutics

1.1. Current scenario of conventional therapy versus gene therapy

Existing pharmaceutical companies have been manufacturing small-molecule therapeutics to develop drugs for many decades. These small-molecules bind, ordinarily in an antagonistic manner, to compartments of target proteins like receptors, enzymes, or other proteins. Which hinders a biological pathway resulting in a therapeutic effect for disease [1]. These small-molecule therapeutics (describe as conventional therapy hereafter) have great advantages including ease to produce and drug administration, favorable pharmacokinetics, and the capability to go through the pore of the cellular membrane. However, their extraordinary potentials are limited due to dependence on the druggability of the target [1,2]. As of 2017, approved drugs targeted only 667

proteins among about 3000 druggable targets of 20,000 human proteins, which indicated that the vast majority of proteins remain undruggable [2,3].

In addition to limited druggability, conventional therapy is identical for a large group of patients targeting the common disease. While the inter-individual difference in drug-metabolizing enzymes or transporters varies drug efficacy and safety [4-6]. Therefore, tailored-based medicine has become indispensable for accomplishing the best drug response and confirming optimal care. Recent advances in gene therapy have revolutionized a promising therapeutic option for precise and personalized treatment of diverse life-threatening diseases (infographic Table below) [7]. Gene therapy is the transfer of nucleic acid information, usually in the form of DNA or RNA in a vector system, to a target cell to modify a phenotype for therapeutic purposes. An example of such therapeutics is the use of RNAi (RNA interference) related gene therapy like small interfering RNA (siRNA) [8], short-hairpin RNA (shRNA), micro RNA (miRNA) [9], an antisense oligonucleotide (ASO) [10] and recently included CRISPR-Cas system [11]. Although clinical progress of RNAi has faced noteworthy hurdles in terms of immunogenicity and effectiveness for a couple of decades, in recent years, the field has gained successful momentum with the approval of two commercial RNAi therapeutics called patisiran and givosiran [12,13].



1.2. Programmable RNA knockdown strategies in therapeutics

Programmable RNA knockdown tools can be classified into two broad groups such as RNAi and CRISPR-Cas systems. Below I discuss both of the therapeutic tools.

1.2.1. RNAi

In 2006 Dr. Fire and Dr. Mello won the novel prize in physiology for identifying double-stranded RNAs (dsRNA) as the representative agents for post-transcriptional gene knockdown in *C. elegans*, which they termed as RNAi [14]. Later it was discovered that 21-22 nucleotide dsRNA (siRNA) could silence a gene expression in mammalian cells without stimulating non-specific interferon reaction [15]. Likewise, ASO, shRNA, and miRNA are major tools used as RNA therapeutic.

1.2.1.1. Antisense Oligonucleotides (ASOs)

ASOs are short-sized, synthetic, single-stranded complementary oligonucleotides of a target mRNA that capable to alter protein translation through several different strategies. Primarily ASOs are DNA-based which form a DNA-RNA hybrid and serve as a substrate for RNase H (Figure 1A) [1]. Besides the inhibitory translation of a target mRNA, ASO also alters splicing or miRNA function, by Watson-Crick base pairing to target splice sites or miRNA respectively [16,17].

1.2.1.2. micro-RNA (miRNA)

miRNAs are short, synthetic, double-stranded RNAs (dsRNA) and mimic a naturally expressing miRNA that can knock down or repress translation of target mRNA in a sequence-specific way [1]. Primary transcript (pri-miRNAs) is cleaved by Drosha (RNase III) in the nucleus and form 3′ overhang (pre-miRNA) [18]. Next, the pre-miRNA is exported to the cytoplasm with the help of exportin 5. In the cytoplasm, the stem-loop is cleaved by the Dicer (RNase III) and loaded into the endogenous RNA-induced silencing complex (RISC) (including Argonaute protein) [17]. Finally, RISC cleaves the target mRNA or repress translation (Figure 1B) [17,18].

1.2.1.3. small interfering RNA (siRNA)

As described above for miRNA, another type of RNAi that follows almost similar RNAi pathway is siRNA/shRNA [15]. Which are 20-25 bp short, synthetic, dsRNA that can degrade the expression of target mRNA in a sequence-specific manner by activating the RISC (Figure 1C) [18]. As a consequence of the trivial distinction between miRNAs and siRNAs pathways, siRNA can elicit more effective knockdown than miRNAs whereas one miRNA can target several mRNAs at the same time [1,19]. Therefore, both have divergent functions in clinical practices. In addition to these familiar tools in RNA therapeutic, saRNA (small activating RNA) and RNA aptamers have been used at low frequency in therapeutics [1].

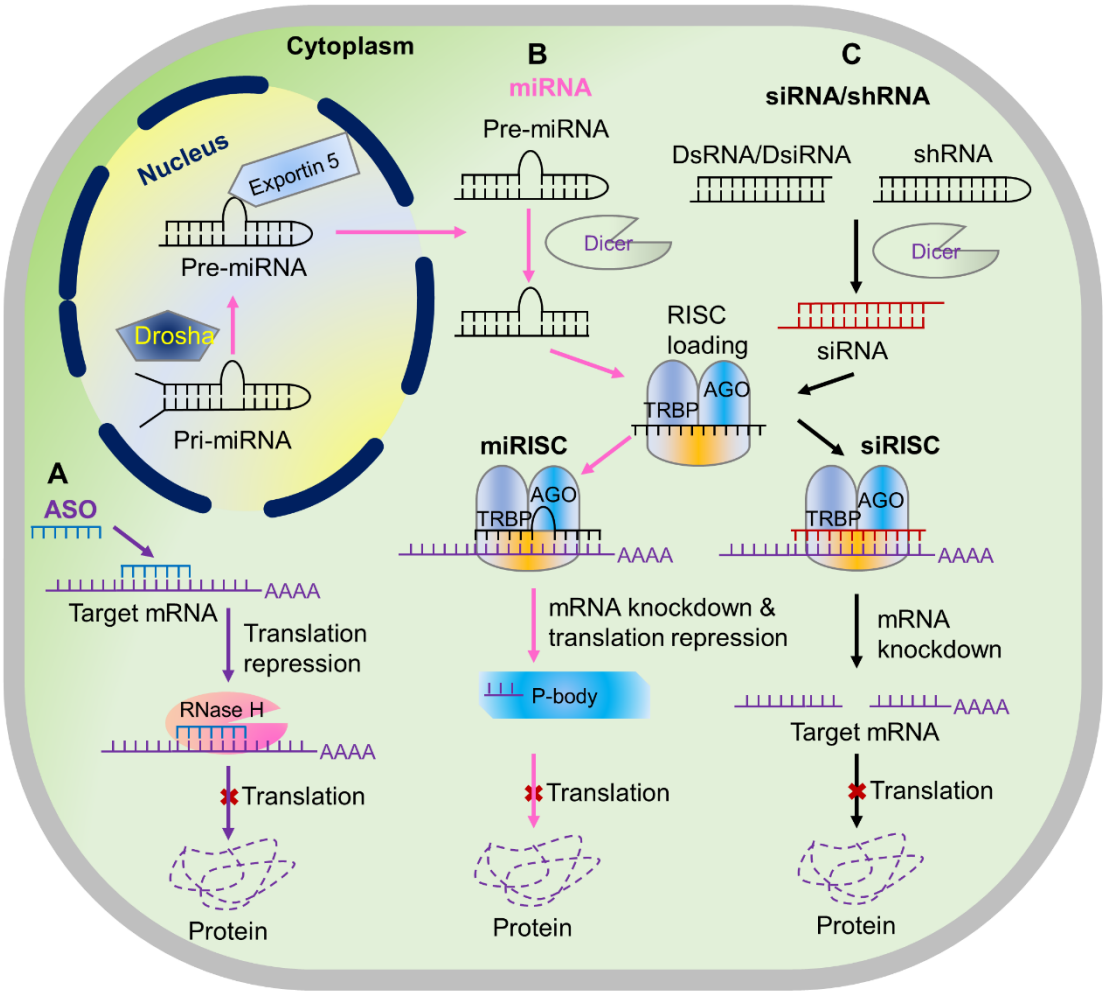


Figure 1. Mechanism of RNAi mediated ASO, miRNA, and siRNA biogenesis in therapeutics. A, ASO mediated translation blocking of a target mRNA (or gene) by endogenous RNase H nuclease [1]. **B**, miRNA mediated target mRNA knockdown and translation repression; Initial miRNA transcript (pri-miRNA) is cleaved in nucleus by

Drosha to produce 3′ overhang (pre-miRNA) [18]. Which is then exported to the cytoplasm and cleaved again by Dicer and loaded into RISC (RNA-induced silencing complex). The activated RISC knockdown the target-mRNA and inhibits the translation. TRBP, transactivation-response-RNA-binding protein; AGO, Argonaute. **C**, siRNA/shRNA mediated target RNA knockdown modulated by RISC [18].

1.2.1.4. Strengths and challenges of RNAi

RNAi has distinct strengths over small-molecule therapeutic (conventional) and monoclonal antibody drugs such as i) RNAi functions based on Watson-Crick base pairing with target mRNA, whereas conventional therapeutic require the druggable target of a particular protein. As a result, there are many proteins related to the diseases that are undruggable by conventional small-molecule drugs but theoretically, RNAi can target all mRNAs [18]. ii) RNAi uses endogenous nuclease, so only one factor (for example dsRNA) is needed for a programmable knockdown. iii) It has high potency [19]. As well as v) Stable and long-term silencing is achievable using shRNA or by a series of transfections.

Although RNAi possesses excellent benefits in gene therapy development, several challenges hinder its extraordinary clinical use including i) higher off-target effect in transcript knockdown [20], ii) limited accessibility to target nuclear transcripts, iii) cost associated with chemical modification for siRNA/ASO [18], iv) transient knockdown with a single transfection, therefore multiple transfections are needed for an effective knockdown [21], v) difficulty to knockdown transcripts which have high turnover [19,21].

1.2.2. CRISPR-Cas system: discovery, diversity, and abundance

1.2.2.1 Discovery of CRISPR: From yogurt to novel prize

Bacteria and archaea employ Clustered Regularly Interspaced Short Palindromic Repeat-associated (CRISPR-Cas) endonuclease protein to a target foreign genetic element for cleavage (like viral nucleic acid) with the help of CRISPR RNA (crRNA) to function as an adaptive immune response [22]. In 2020, Dr. E Charpentier and Jennifer A. Doudna shared a novel prize in chemistry for the extraordinary discovery

of a genome-editing tool (CRISPR-Cas9). The journey behind this outstanding discovery had been started by Japanese researchers who identified a series of short uninterrupted repeats alienated with short sequences in the genetic element of *E. coli* in 1987 (Figure 2) [23]. After 10 years, Mojica and his colleagues identified that short uninterrupted repeats had been omnipresent in all bacteria using bioinformatics analysis and later the term CRISPR had introduced [24,25].

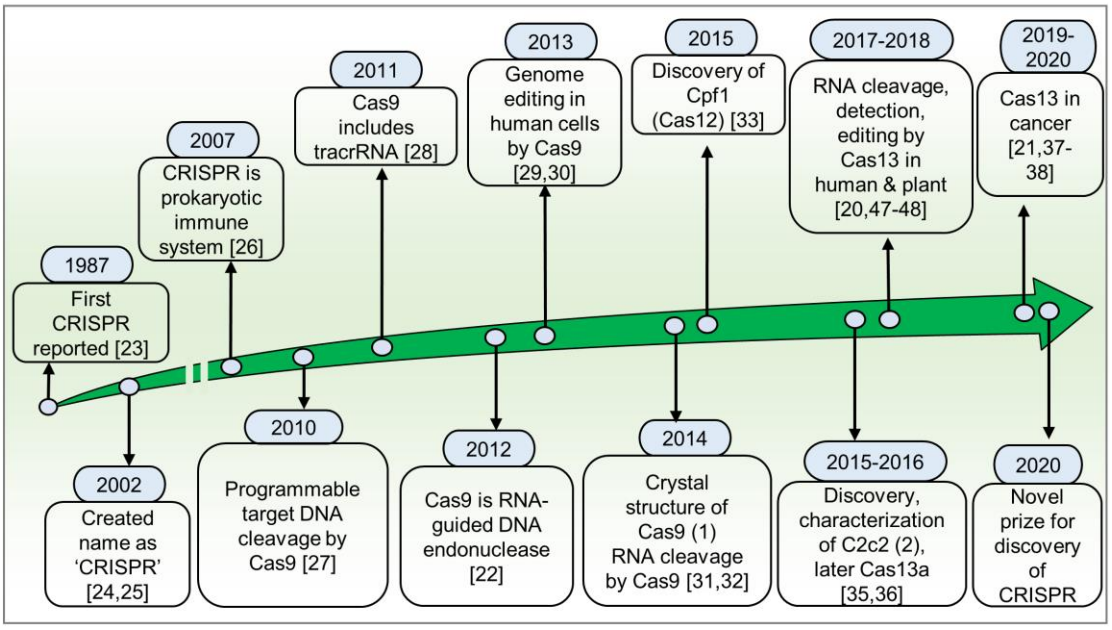


Figure 2. Road to the success of CRISPR engineering, from yogurt to novel prize.

In 2007, first experimentally proved that the CRISPR-Cas involved in adaptive immunity in lactic acid bacteria (*Streptococcus thermophilus*) against lytic phage [26]. Later, the Cas9 was characterized (former name Cas5, Csx12, or Csn1) and demonstrated the RNA-mediated DNA cleavage [27]. Soon after, Charpentier and colleagues discovered an unknown molecule called tracrRNA, which is a part of Cas9 for DNA cleavage [28]. By 2012, Dr. Charpentier and Dr. Doudna purified and characterized the Cas9 and showed *in vitro* target DNA cleavage [22]. After this *in vitro* findings, use of Cas9 has exploded. The principle was later proved in human cells [29,30] and crystal structured was uncovered [31,32]. Next, different new CRISPR-Cas proteins (like CRISPR-Cas12) were discovered [33], and revolutionized in various biotechnological applications [34]. Lately researcher found Cas13 (formerly C2c2) single CRISPR effector endonuclease which able to cleave target RNA in bacteria and human cells [35,36], and also in human cancer cells [21,37,38]. Since E Charpentier

and JA. Doudna invented the CRISPR/Cas9 endonuclease in 2012, use of CRISPR has surged in many biotechnological applications, therefore, their work has been awarded novel prize in 2020.

1.2.2.2. Diversity and abundance

CRISPR-Cas endonucleases are classified into two major groups such as Class 1 and Class 2 based on evolutionary relationship and single/multi-subunit effector protein [39]. The detailed classification with the segmental structure of the network has shown in Figure 3 [39,40]. Class 1 CRISPR-Cas group has multiple effector modules that make crRNA-binding machinery and function together for cleaving a target nucleic acid [40]. This group is sub-categorized into three types (Type I, Type III, and Type IV). Whereas the Class 2 group has a single, multi-domain effector protein and functions similarly to Class 1. Which is also subdivided into three types such as Type II (Cas9), Type V (Cas12), and Type VI (Cas13). Furthermore, these six types have been grouped into various categories based on the origin of species.

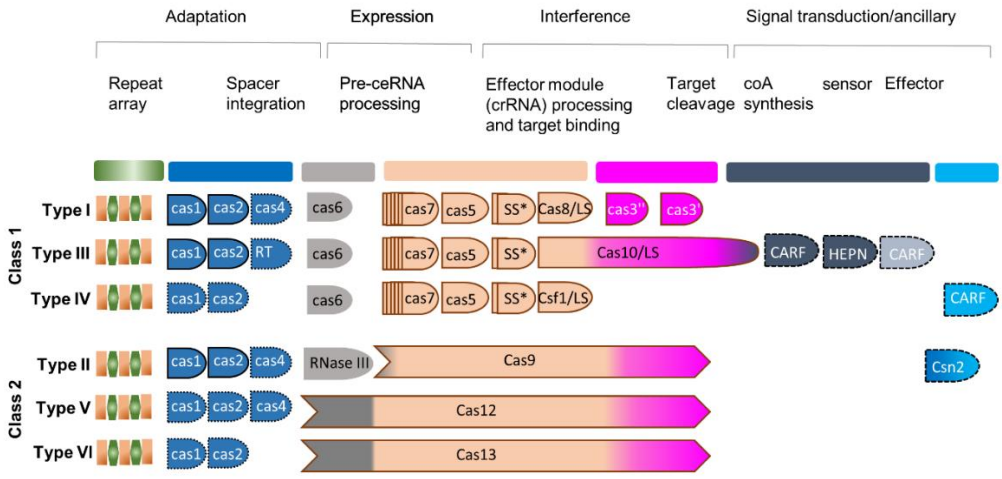


Figure 3. General structural overview of CRISPR-Cas classification [40]. Major classes of CRISPR-Cas effectors; Class 1 contains multi-subunit effectors for a catalytic activity and CRISPR array for guiding the effector to a target site. In contrast, Class 2 possesses a single effector associated with multi-domains accessory protein. Similarly, it consists of a CRISPR array. Each major class is sub-divided into three types; All sub-classes are categorized into three main functional domains such as Adaptation, Expression, and Interference. One additional functional domain is signal transduction.

1.2.2.3. RNA targeted CRISPR effectors

Based on CRISPR-Cas effector functionality on target nucleic acid (DNA/RNA or not defined yet), so far there are four CRISPR-Cas effectors have been reported that able to cleave target specific RNA molecules like Type III A-C (multi-subunit effector), Cas9, Cas12g, and Cas13 (single subunit effectors) (Figure 4) [40]. The basic principle, mode of action, strength and challenges are described below in the therapeutic point of view.

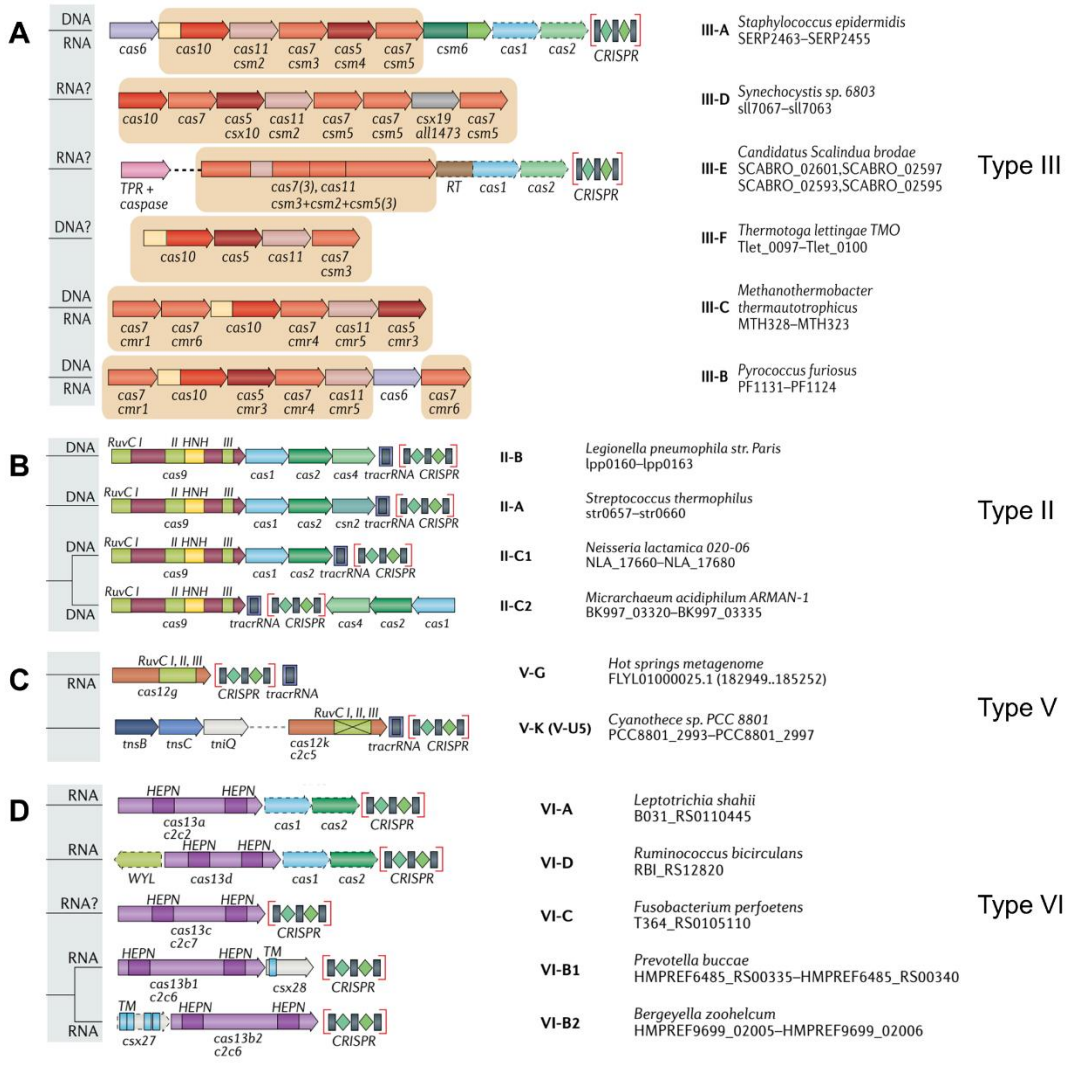


Figure 4. Structural module of RNA cleavage CRISPR-Cas and CRISPR array [40]. A, details structural module of CRISPR-Cas effector and CRISPR array of Class 1 Type III; where Type III-A to Type III-D are acted on RNA and Type III-E to Type III-F are not yet concluded (indicated as '?' sign). **B**, Class 2 Type II CRISPR-Cas effector and CRISPR array. **C**, Class 2 Type V CRISPR-Cas effector CRISPR array. **D**, Class 2 Type VI CRISPR-Cas effector and CRISPR array.

1.2.2.3.1. Type III

Type III CRISPR-Cas systems belong to class 1. They are categorized into six subtypes such as Type III-A, III-B, III-C, III-D, III-E, and III-F [40]. Type III-A, III-B, and III-C have been reported as RNA cleavage function. The unique feature among all subtypes is the presence of a Cas10 nuclease except in Type III-E. The generalized mechanism for Type III-A has drawn in Figure 5A [41-43]. CRISPR array is transcribed into pre-crRNA and then becomes mature with the help of Cas6 [44]. Later, it goes under secondary processing which is unknown yet. There are several subunits of Type III-A (Cas10, Cas11, Cas7, Cas5, and Csx1) combined to cleave nucleic acid guided by mature crRNA. Type III-A cleaves target (specific 6-nucleotide) RNA performed by Cas7 (Figure 5A: ☑ symbol) and two nonspecific nucleic acids: ssDNA (single-stranded DNA) and RNA depend on Cas10 activity (Figure 5A: ☒ symbol) [41,42]. Cyclic oligos produced by Palm domain of Cas10 are allosterically activate Csm6/Csx1 which is involved directly in nonspecific RNA knockdown [45]. Multiple nuclease activity and especially nonspecific cleavage function raise a question for suitability in the therapeutic purpose.

1.2.2.3.2. Cas9 (Type II)

CRISPR-Cas9 is a class 2 single effector programmable endonuclease. The discovery of CRISPR-Cas9 endonuclease has further pushed the advance of RNA therapeutics to the front position. Cas9 nuclease is used for genomic DNA editing but O'Connell et al have shown that Cas9 also effectively knock down target specific single-stranded RNA (ssRNA) molecule [22,46]. Cas9 endonuclease binds to target RNA only when spacer (guide RNA) matches and the PAM (protospacer adjacent motif) is presented in *trans* as separate DNA oligomers (PAMers) (Figure 5B) [46]. PAMers means PAM: 5′ NGG, and oligos that are complementary to target RNA. When PAMers bind to target RNA then the Cas9 is directed to target RNA molecule following guide-RNA and cleaves it. That ultimately hinders the protein translation. Cas9 is a single CRISPR-Cas effector and has higher potency over multi-subunit effectors like Type III CRISPR-Cas nuclease. Nevertheless, it holds the ability to interact with nonspecific DNA [46], theoretically leading to unintended permanent error at the genomic level.

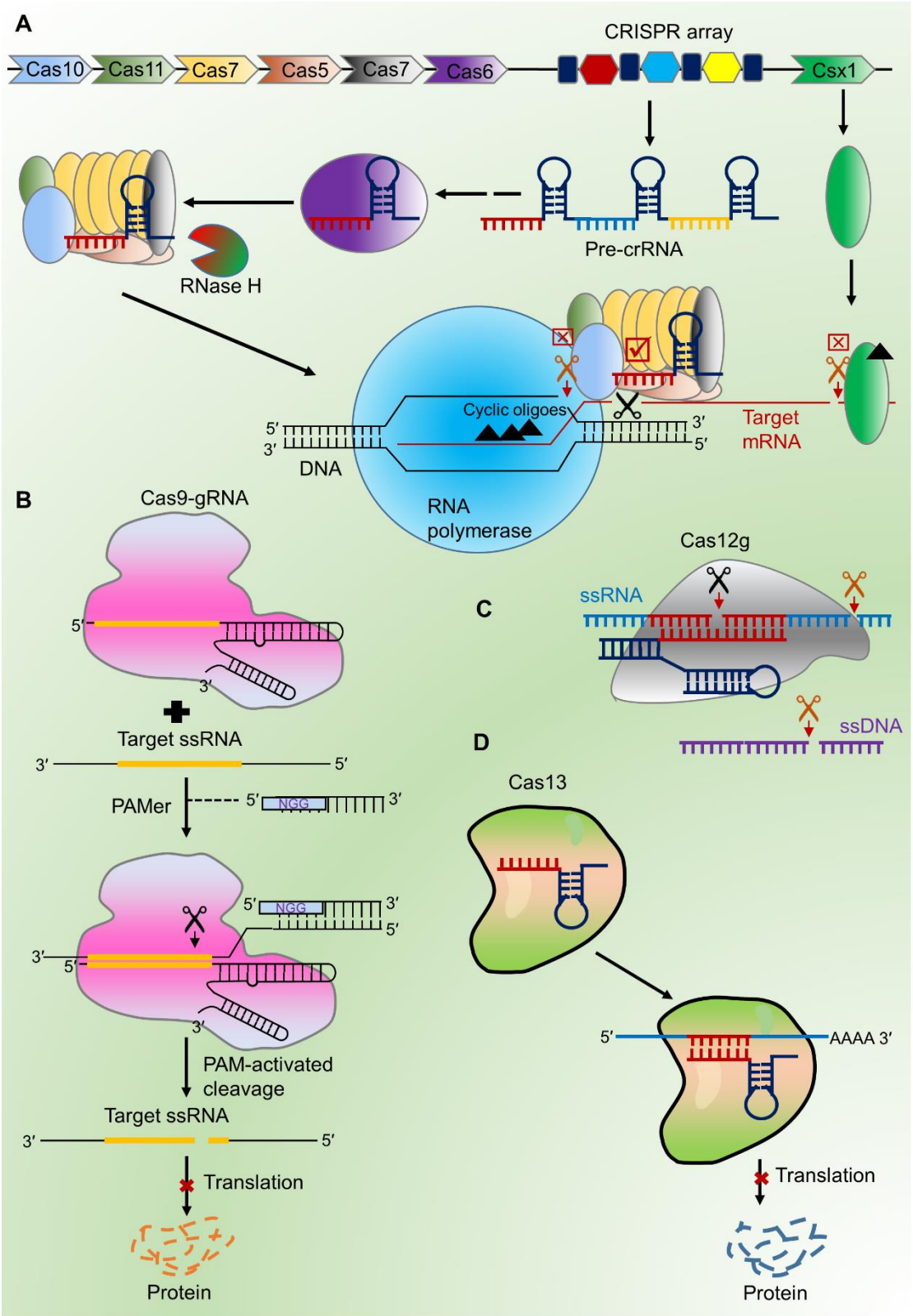


Figure 5. The mood of action of all CRISPR-Cas effectors which are involved in RNA degradation. A, Mechanism of Type III-A CRISPR-Cas mediated target nucleic acid cleavage [41-43]; Type III-A cleaves three nucleic acids such as target RNA (indicated as ☑ symbol), non-target ssDNA and ssRNA (marked as ☒ symbol). **B**,

Cas9 mediated RNA cleavage activity [46]; This system works only when the PAMer oligos hybridize with target RNA, then guide RNA starts search-target and hybridize flanking by PAMer. Later, it cleaves target RNA, this is called PAM activated RNA cleavage. **C**, similar with Type III-A, CRISPR-Cas12g also cleaves three nucleic acids including one target RNA, and two non-target cleavages (ssRNA and ssDNA) [47]. **D**, Strategy how Cas13 cleaves target RNA molecule [36]. The detailed strategy has shown in Figure 6.

1.2.2.3.3. Cas12g (Type V-G)

Recently identified Cas12g (Type V-G) has shown the potentiality of RNA cleavage [47]. Cas12g is a class 2 Type V CRISPR-Cas effector. It is comparatively simpler and smaller in size (767 aa) than other nucleases. Also, no PAM sequence is required for interference activity [47]. However, Cas12g possesses analogous function like Type III-A. It shows target RNA degradation and non-target cleavage of ssDNA and ssRNA (Figure 5C) [47].

1.2.2.3.4. Cas13 (Type VI)

Type VI Cas proteins (Cas13) are also simple structure modules and have two HEPN (higher eukaryotes and prokaryotes nucleotide-binding) catalytic domains [35]. These domains are located at both terminal ends and cleave target RNA upon binding. There are four subtypes of Cas13 based on the origin, protein length, and structural orientation such as Cas13a (formerly known as C2c2 or Type VI-A) from *Leptotrichia wadei* or *L. shahii*, Cas13b (C2c6 or Type VI-B) from *Prevotella buccae* or *Prevotella sp.* P5-125, Cas13c (C2c7 or Type VI-C) from *Fusobacterium perfoetens*, and Cas13d (CasRx or Type VI-D) from *Ruminococcus flavefaciens* XPD3002 (Figure 4) [40,48,49].

In the structural organization of Cas13, as like to Cas9 and Cas12, the Cas13a adopts a bi-lobed globular protein comprising one crRNA Recognition lobe (REC), and one Nuclease lobe (NUC) [50]. The REC lobes contain an N-terminal domain (NTD), and a Helical-1 domain whereas NUC lobes have two HEPN domains unlike RuvC in Cas9 and Cas12, and Helical-2 and -3/linker [50]. In the case of crRNA, the crRNAs of Cas13 are the relatively simple structure of direct repeat (DR), which is flanked by

5′ or 3′ about 20–30 bp long ‘spacer’ nucleotide (guide RNA) [51]. The crRNAs are transcribed into pre-crRNAs that require further processing to become functionally mature crRNAs [52].

In mechanism, DR or stem-loop of mature crRNA are recognized and bound to REC lobes in Cas13 with high affinity and specificity [51]. The spacer of mature crRNA (especially seed region) helps in the target-search procedure and stabilized the preliminary hybridization of the ‘spacer’ to the target RNA in the close vicinity of the PFS (protospacer flanking site) site (Figure 5D and Figure 6) [49,51]. Upon the stable hybridization between Cas13 bounded crRNA-spacer and target-RNA, a large conformational change is adopted by crRNA and Cas13a. More specifically, Helical-2 domains spin away from the HEPN-2 domain, which helps to rotate the HEPN-1 domain, coming to the two halves of the HEPN active site residing in HEPN-1 and -2 close enough together to make the catalytic site for RNA cleavage [53]. Thus, the active catalytic site cleaves the target RNAs. The advantages of Cas13 over other methods are the simple structure of CRISPR-Cas effector and crRNA architecture, limited PFS preference. The RNA level manipulation is less error-prone than DNA.

1.2.2.3.5. Strengths and challenges of CRISPR-Cas systems

Overall, the CRISPR-Cas effector systems are advantageous over RNAi therapeutic such as i) No need for multiple rounds of transfection for sufficient knockdown [21]. ii) As RNAi uses endogenous nucleases, knockdown of a nuclear transcript is challenging but CRISPR-Cas mediated system can overcome it easily by tagging a localization signal with the Cas protein [20,21,48]. iii) It has higher effectiveness [20,48]. iv) Similar to RNAi, theoretically CRISPR can easily target a gene of interest. v) Stable and long-term silencing using these techniques. vi) Comparatively low cost than RNAi techniques. However, several concerns should be addressed for getting its amazing clinical application including i) reduction of off-target effect and ii) like traditional gene therapy, CRISPR-Cas systems also have a concern about the immune-toxicity [54].

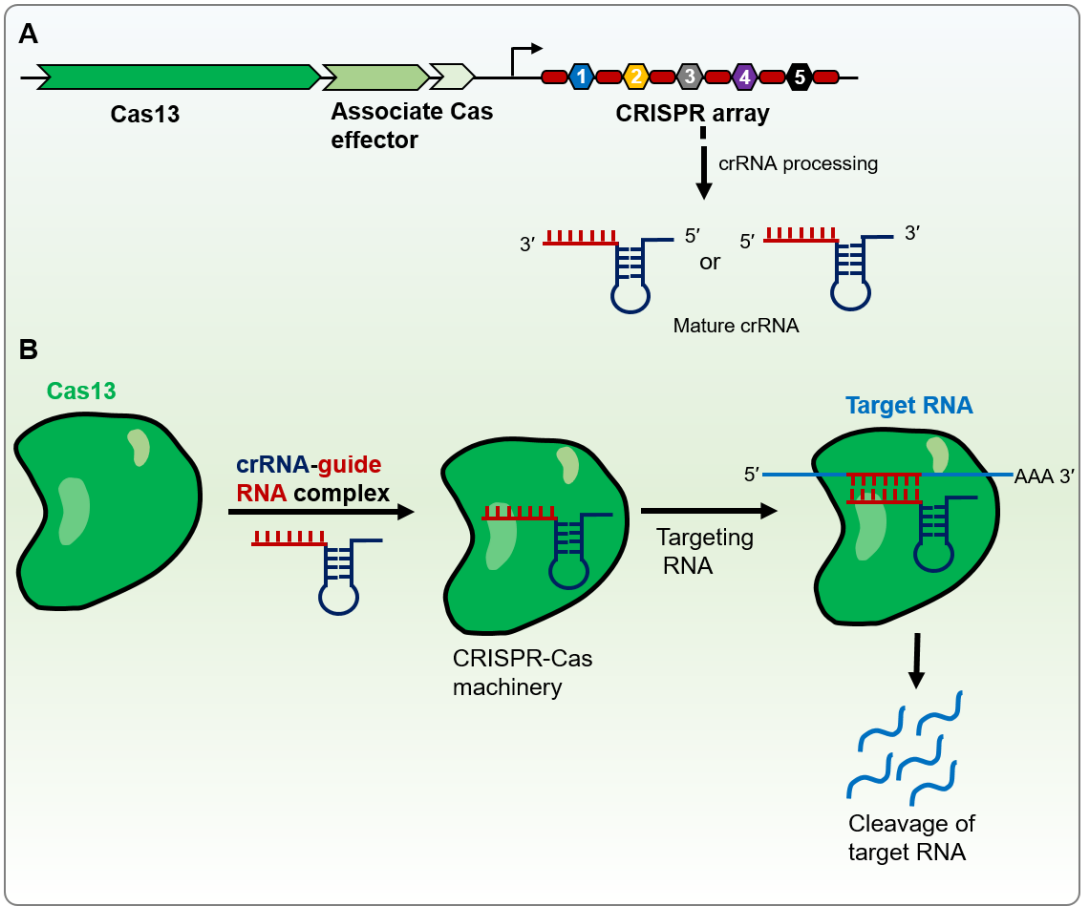


Figure 6. Generalized mechanism of CRISPR-Cas13 mediated target RNA knockdown. A, Schematic of CRISPR-Cas endonuclease gene and CRISPR RNA (crRNA) array; initially crRNA is transcribed into pre-crRNA and later become mature crRNA by secondary processing. **B**, CRISPR-Cas13 mediated target RNA knockdown; the mature crRNA binds efficiently to REC lobe in Cas13 that allow Cas13 protein to search a target RNA with the help of spacer (red part in mature crRNA). Once, it hybridizes with target RNA, a conformation change happens in Helical-2 and HEPN domains in the Cas13 that activates the catalytic site in HEPN domains. Finally, the active site cleaves a target RNA of interest.

General introduction part-II: Molecular genetics and therapeutic trends of the *EML4-ALK*-positive lung cancer

1.3. EML4 in human

Echinoderm microtubule-associated protein (EMAP) was first identified as a major element of microtubule development in sea urchin eggs and categorized as a first member of the EMPA-like (EML) protein family [55]. Since then, several variants have been determined across the animals including humans, most of which are associated with microtubule assembly during mitosis [56]. There are six EML proteins express in humans, EML1 to EML6, which are classified into two subfamilies based on their protein domain organization [57]. EML1 to EML4 consists of the N-terminal coiled-coil domain, followed by a hydrophobic EML protein (HELP) domain and variable repeats of tryptophan-aspartic acid (WD) (Figure 7A) [58,59]. These domains are alienated by unstructured basic linker region. In contrast, the absence of N-terminal coiled-coil domain in EML5 to EML6 proteins and instead, contain three repeats of the HELP-WD domains [57,58].

1.4. ALK in human

Anaplastic Lymphoma Kinase (ALK) was first reported as a fusion partner with nucleophosmin (NPM) resulted in over-activation of kinase activity found in non-Hodgkin's lymphoma [60,61]. ALK belongs to the insulin receptor kinase superfamily, which comprised of N-terminal extracellular region including two MAM (meprin A5 protein and protein tyrosine phosphatase Mu) domains, one (LDLa) low-density lipoprotein receptor domain class A domain close to a Gly-rich region ahead of transmembrane (TM) helix (Figure 7B) [58,62]. The C-terminal intracellular region has a tyrosine kinase (TK) domain [58]. ALK expressed the most in brain tissues, therefore, it most likely involves the growth of the nervous system [63,64]. ALK knockout in mice resulted in no major defects in viability except behavioral abnormalities [63].

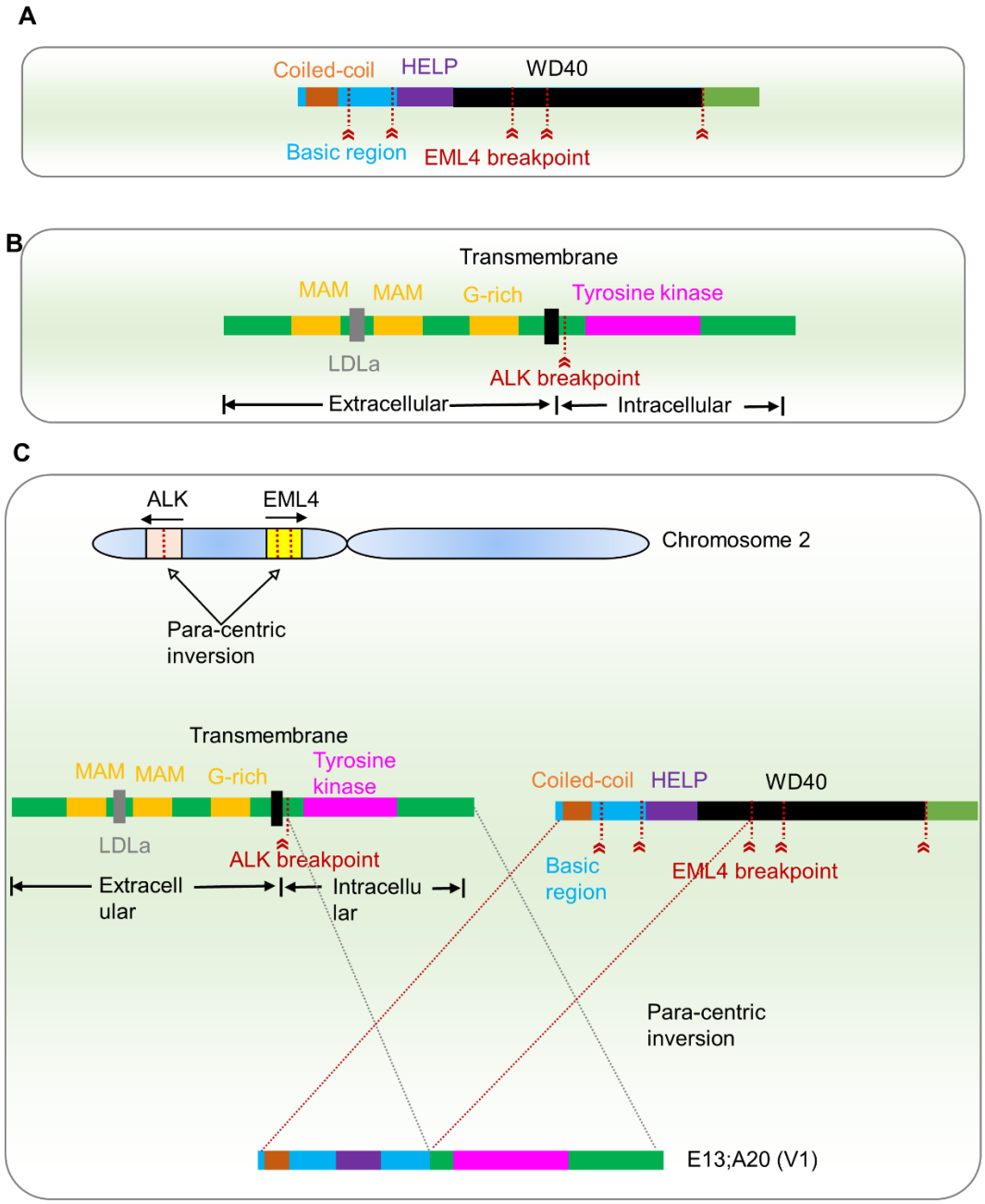


Figure 7. Functional domain and motifs in EML4-ALK oncofusion. A, Graphical representation of functional domain and motifs in *EML4* gene; Coiled-coil and HELP (hydrophobic EML protein) domains are connected by unstructured basic region. Red dot lines indicated that the breakpoint when EML4 translocated with other fusion partners. **B**, ALK is a receptor protein that contains three regions including intracellular, transmembrane, and extracellular. Intracellular regions have one functional tyrosine kinase (TK) domain whereas the extracellular region possesses a G-rich region, LDLa (low-density lipoprotein receptor domain class A), and MAM (meprin A5 protein and protein tyrosine phosphatase Mu) domains. **C**, How *EML4* and *ALK* make a fusion gene through para-centric inversion.

1.5. Molecular genetics of *EML4-ALK* in lung cancer

Non-small cell lung cancer (NSCLC) responsible for 85% of lung cancer, and most of these are lung adenocarcinoma (LUAD) [65,66]. One of the most common types of oncofusion in LUAD is EML4-ALK which drives the constitutive activation of ALK and thus increased cell survival, cell growth, and proliferation [56]. The *EML4* and *ALK* genes are located in reverse orders within the short arm of chromosome 2, and the paracentric inversion of this chromosomal region [inv(2)(p21p23] results in a fusion oncogene so-called *EML4-ALK*, discovered in Japanese lung cancer patients in 2007 (Figure 7C) [67]. So far, fifteen variants of *EML4-ALK* have been reported [56,58].

In all cases, the breakpoint within the *ALK* is close to the 50 end of exon 20 [56,58]. In this way, the TK-containing intracellular domain is only included in the EML4-ALK fusion variants. The breakpoint within the *EML4* gene is inconstant. The most frequent variants is V1 (*EML4* exon13;*ALK* exon 20) (36%), and then V3a (E6;A20) and V3b (E6ins33;A20) (29%) [56,58]. The shortest version is V5 (E2;A20) (2%) (Figure 8) [68]. Although breakpoint within the *EML4* gene is a more variable but unique feature is the presence of coiled-coil domain in all variants. This domain is recently identified as trimerization domain (TD) responsible for ALK overactivation through oligomerization and auto-phosphorylation [21,56,58].

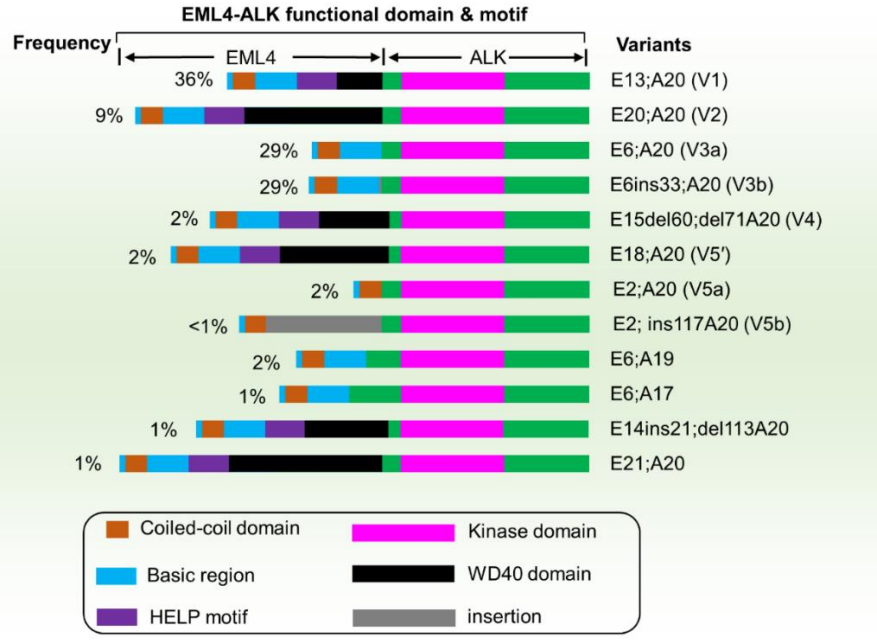


Figure 8. Molecular genetic variants and frequency of *EM4-ALK* oncofusion [58,68].

1.6. Molecular signaling pathways

To elucidate the signaling pathways, retrospective studies have used cells derived from NSCLC patients or disease model cells expressing *EML4-ALK* variants like H3122 (expressing variant V1), H2228 (harboring variant V3a/b) [69]. The ALK inhibitor (For example, Crizotinib or TAE684) had used in these cell lines to provide the effect on phosphorylation of signaling partners such Akt, ERK, and STAT3 [70]. In this way, *EML4-ALK* oncofusion is supposed to lead to the constitutive activation of downstream oncogenic signaling through several pathways by activating phosphorylation such as RAS/RAF/MEK/ERK, JAK/STAT, and PI3K/Akt [71].

1.7. Contemporary treatment systems for ALK-positive lung cancer

Patients harboring *EML4-ALK* have divergent pathological characteristics comprising young age of onset, light or no smoking history, and adenocarcinoma in histology [69,72]. Cytotoxic platinum or pemetrexed based chemotherapy is often used for patients with advanced stages of the disease [73]. With the advancement of molecular pathobiology of *EML4-ALK*-positive lung cancer, these days ALK inhibitor-based personalized therapy has proven promising results in prolonging patient survival.

1.7.1. Rationale for ALK inhibitor

1.7.1.1. First generation

An ATP analog tyrosine kinase inhibitor of ALK, called crizotinib (structure in Figure 9A), was first discovered and later approved (first generation) by the FDA in 2011 as oral medicine for *ALK*-positive NSCLC patients [74]. The crizotinib not only inhibits ALK but also MET and ROS1driven tumors [75,76]. Shaw et al found crizotinib is superior in terms of progression-free survival (PFS) and drug response rate over standard first-line pemetrexed-plus-platinum chemotherapy in 347 patients with previously untreated *ALK*-positive NSCLC [77].

1.7.1.2. Second generation

Regrettably, after a promising response to treatment with crizotinib, patients invariably succumb to acquired resistance within one year [78,79]. Further studies identified the reason and classified it into two roads for acquired resistance such as ALK-dependent

or ALK-independent pathways. In the ALK-dependent pathways, secondary mutations for example L1196M, C1156Y, G1269A appear in the pocket of tyrosine kinase at the crizotinib binding site [79,80]. In contrast, ALK-independent pathway associated with alternative pathway activation for cell proliferation and growth including KRAS, EGFR, cMET, and AXL or conversion into small cell lung cancer [81]. To circumvent this resistance, second-generation ALK inhibitors like ceritinib, alectinib were discovered and approved by FDA, and they showed three-fold longer PFS over crizotinib (structure in Figure 9A) [81]. As a consequence, alectinib is broadly used as first-line therapy for *ALK*-positive NSCLC patients [82,83].

1.7.1.3. Third generation

However, as observed in crizotinib, some patients with ALK-positive lung cancer treated with a second-generation inhibitor like alectinib acquired resistance due to the occurrence of mutation particularly G1202R and I1171N/S/T in the ALK kinase domain [79,84]. To get rid of acquired resistance from second-generation ALK inhibitor, a third-generation inhibitor called lorlatinib was developed and accepted by FDA (structure in Figure 9A) [85-87]. The lorlatinib exhibited an extraordinary response rate (47%) and PFS (7.3 months) in ALK-positive lung cancer patients who previously received a minimum of one inhibitor [86,88]. Nevertheless, recently some compound mutations that induce resistance against lorlatinib have been reported [89,90]. Fascinatingly, acquired resistance by lorlatinib can be re-sensitized by combination therapy with first or second-generation inhibitors [91]. During the writing of this thesis, Mizuta et al reported another new inhibitor called gilteritinib that able to overcome the resistance of lorlatinib [81].

1.7.2. Chemotherapy

Single-agent chemotherapy often shows lower PFS over combination chemotherapy. A retrospective study found that single-agent pemetrexed chemotherapy resulted in 5.5 months PFS whereas platinum/pemetrexed combine regimen exhibited higher PFS (7.3 months) [92]. Surprisingly, the majority of studies found that inhibitors possess dominance features in response rate and PFS over conventional platinum or pemetrexed based chemotherapy for *ALK*-positive cancer. For instance, 7.7 months PFS and 65% drug response rate in crizotinib while 3 months PFS and 20% response rate in

pemetrexed/docetaxel were reported (structure in Figure 9B) [77,93]. Other studies reported 16.5 months and 8.1 PFS for ceritinib and chemotherapy (cisplatin or carboplatin plus pemetrexed) respectively [94].

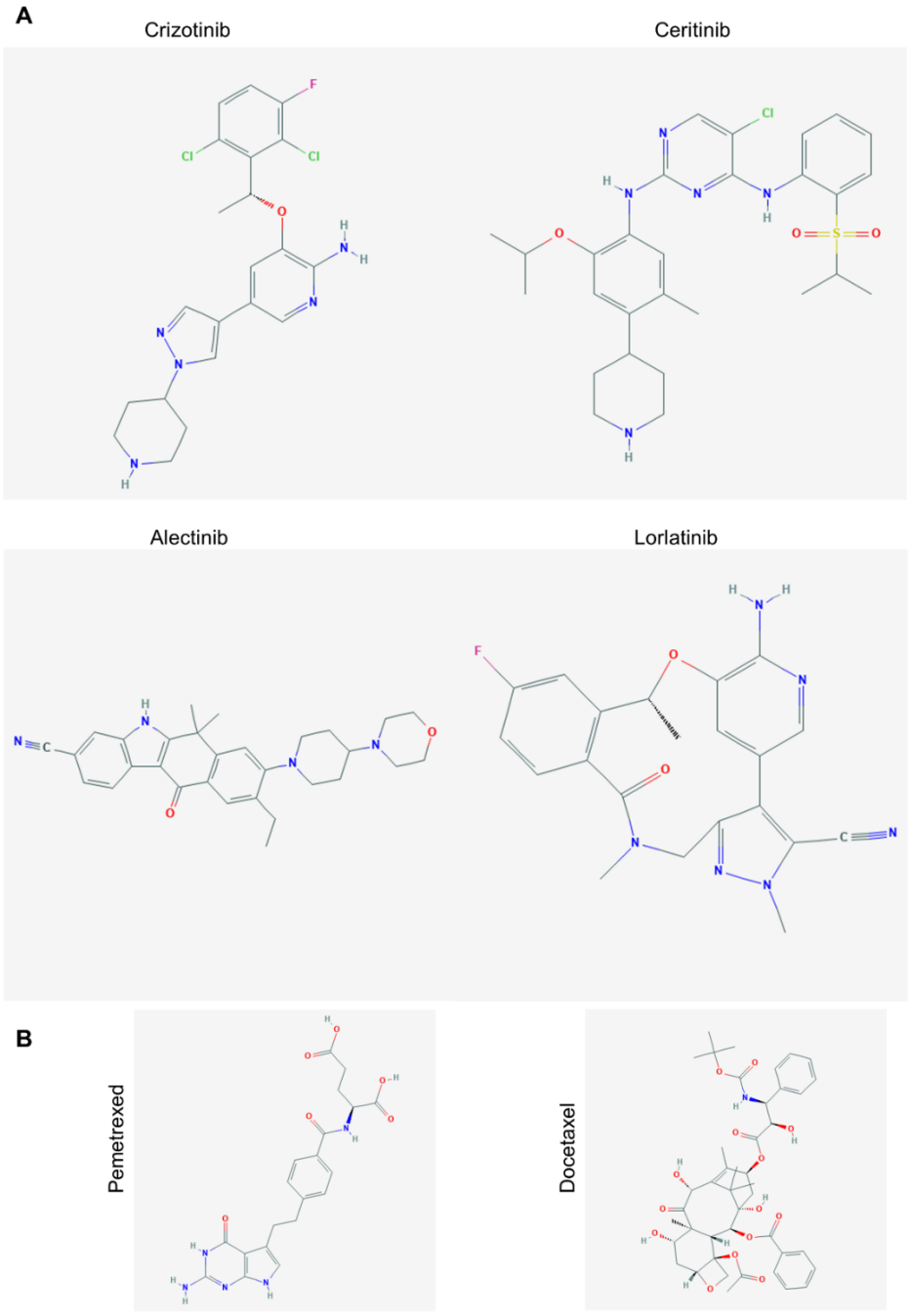


Figure 9. Commonly used inhibitors and chemotherapeutic molecules for EML4-ALK positive lung cancers. A, Inhibitors used for ALK-positive NSCLC approved by FDA; Crizotinib, first-generation inhibitor; Ceritinib and Alectinib, second-generation inhibitor; Lorlatinib, third-generation ALK inhibitor. **B**, Molecules often used as chemotherapy for ALK-positive cancers such as pemetrexed and docetaxel. All molecular structures are derived from PubChem (NCBI) [95].

General introduction part-III: Problem identification and justification, and objectives

1.8. Problem identification and justification

1.8.1. Why is a new *ALK*-positive lung cancer treatment strategy necessary?

According to the World Health Organization (WHO) report 2020, death of lung cancer is the highest among all cancers (Figure 10) [96]. *ALK*-positive lung cancer is one of the important cancers which is not correlated with environmental factors like smoking [56]. The therapeutic system available these days for this *ALK*-positive cancer can be categorized into two groups like ALK-inhibitor and chemotherapy.

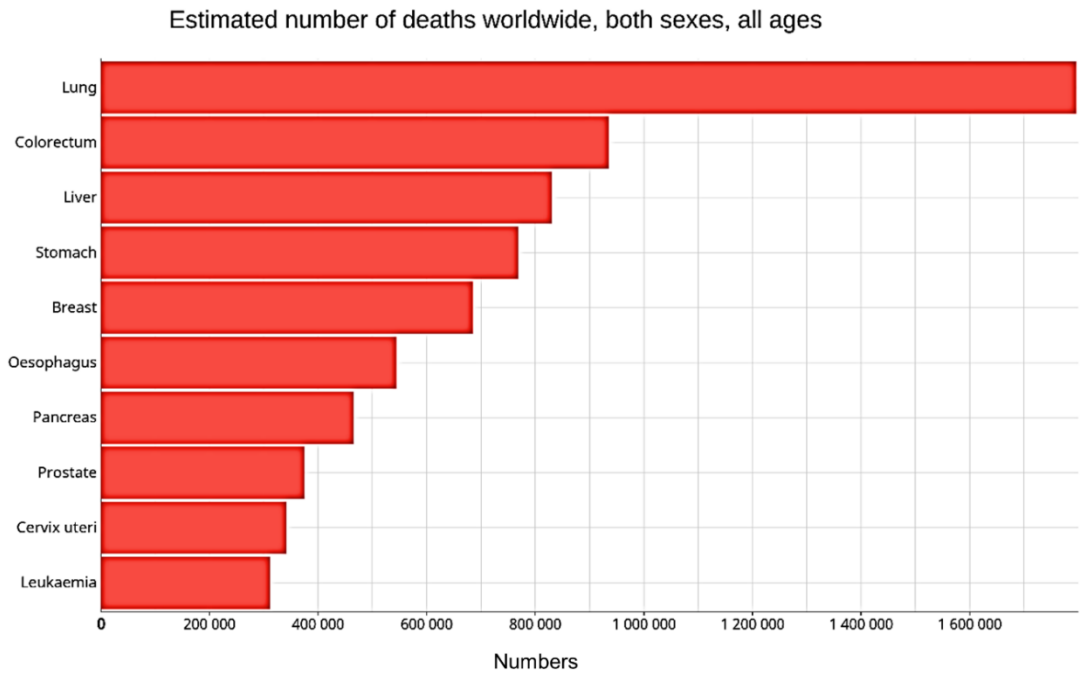


Figure 10. The death rate for all cancer types across the world estimated by WHO in 2020 [96].

The inhibitor-mediated systems hold promising therapeutic outcomes. But the most hurdle is the patient experience acquired resistance due to mutation in the kinase domain of ALK or bypass activation of cell proliferating pathways. In particular, crizotinib generates L1196M, C1156Y, and G1269A mutations whereas alectinib induces G1202R and I1171N/S/T mutations in the kinase domain of ALK (Figure 11) [79,80,84]. Patients treated with a third-generation inhibitor like lorlatinib also succumb to resistance [89,90]. Another challenge for the use of inhibitors is the associated adverse side effect such as diarrhea, vomiting, nausea, and an elevated level of alanine aminotransferase [94].

Contrarily, the chemotherapeutic system possesses lower efficacy and shorter progression-free survival (PFS) than inhibitors [77,93]. The chemotherapy system is highly toxic to cells, and show adverse effect including nausea, vomiting, and anemia [94]. Although the potency is low, platinum/pemetrexed combined chemotherapy can be administrated for the *ALK*-positive NSCLC once the patient becomes resistant to past inhibitor treatment [73]. In line with the evidence of retrospective studies for the development of a new inhibitor and later facing a challenge for acquired resistance, it is denoting that alternative study is strongly needed for ALK-positive lung cancer.

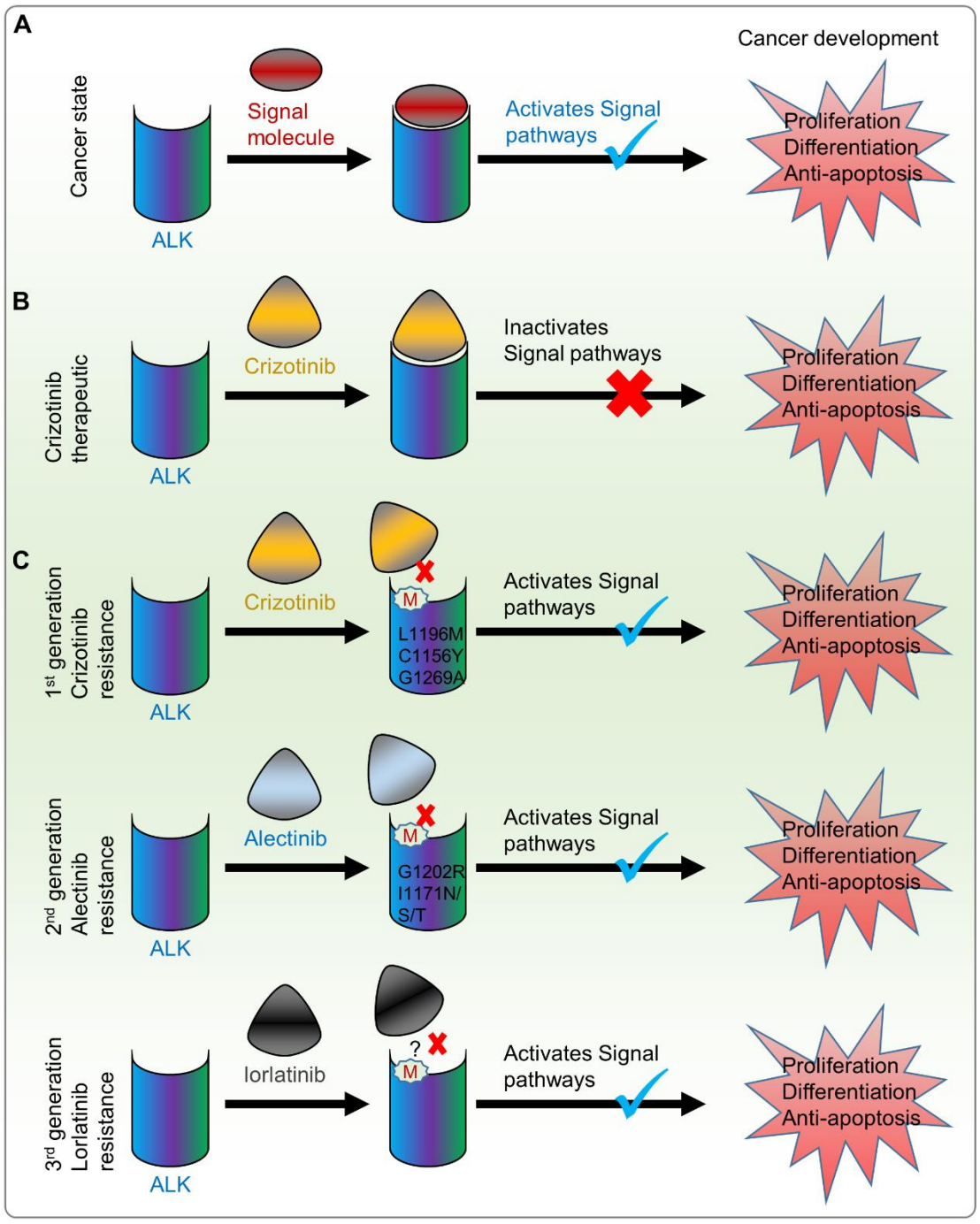


Figure 11. ALK-dependent acquired resistance generated by inhibitors. A, Cancer state of ALK-positive lung cancer where signal molecule binds to ALK receptor to activate the downstream signal. **B**, Crizotinib mediated therapeutic strategy where crizotinib competes with signal molecule and binds to the pocket of TK domain, by which downstream signal is hindered. **C**, Inhibitors mediated acquired resistance through ALK-dependent pathway; mutation has occurred in the binding site of ALK, thereby inhibitors cannot attach to the pocket of ALK. Which ultimately activates the downstream signaling.

1.8.2. Which can be a future suitable strategy for *ALK*-positive cancers treatment?

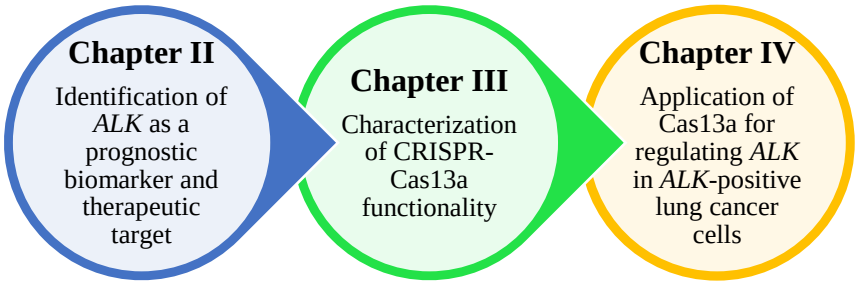
Now question arising which type of therapeutic method fits the most. As small-molecule inhibitor and chemotherapy (conventional therapy) have been facing challenges, the gene therapy-based technique could be a potential alternative. Based on the recent evidence for the successful RNAi therapeutic approval by FDA for some genetic diseases [1], the CRISPR-Cas13 based therapeutic strategy is being proposed in this study. This is because the CRISPR-Cas13 possesses superiority over available RNAi tools including a) higher gene knockdown efficiency [20], b) single-dose is sufficient for gene knockdown [21], c) chemical modification is not needed that reduce the cost, d) higher programmable manner knockdown, e) the most prominently, lower off-target transcript degradation [20]. The use of Cas13 is also advantageous over Cas9 such as a) Cas9 holds the ability to interact with non-specific DNA when it uses for nucleic acid engineering either in DNA or RNA level manipulation, which can introduce a permanent error in the genome [21,51], b) RNA has a very short half-life and expresses as tissue-specific manner. Therefore, even though non-specific targeting happens in the case of Cas13, the effect will be transient, unlike DNA-based gene engineering.

1.9. Objective of this study

In the course of my doctoral research, I have identified *ALK* as a prognostic biomarker and therapeutic target using bioinformatics analysis and developed a potential therapeutic strategy for the *ALK*-positive lung cancer cells using the CRISPR-Cas13a system in disease model human cells. The principal objectives of the current doctoral thesis are given below:

- Identification of *ALK* expression as a prognostic biomarker and therapeutic target in *ALK*-positive cancers using integrative bioinformatics approach.
- To break down of firefly luciferase RNAs using the CRISPR-Cas13a machinery as a target model for the further application in disease-relevant RNAs, which helps to know how is the efficiency, specificity, and functionality of Cas13a ribonuclease on a target-specific RNA.
- Application of CRISPR-Cas13a machinery in downregulation of *EML4-ALK* fusion RNAs for the novel treatment system of Non-small-cell lung carcinoma (NSCLC).

These principle objectives are split into three experimental chapters which are depicted as infographic flowchart.



1.10. Bibliography

1. Dammes, N.; Peer, D. Paving the Road for RNA Therapeutics. *Trends in pharmacological sciences* **2020**, *41*, 755-775, doi:10.1016/j.tips.2020.08.004.
2. Santos, R.; Ursu, O.; Gaulton, A.; Bento, A.P.; Donadi, R.S.; Bologa, C.G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T.I., et al. A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* **2017**, *16*, 19-34, doi:10.1038/nrd.2016.230.
3. Relling, M.V.; Evans, W.E. Pharmacogenomics in the clinic. *Nature* **2015**, *526*, 343-350, doi:10.1038/nature15817.
4. Anguela, X.M.; High, K.A. Entering the Modern Era of Gene Therapy. *Annual Review of Medicine* **2019**, *70*, 273-288, doi:10.1146/annurev-med-012017-043332.
5. Saifullah; Fuke, S.; Nagasawa, H.; Tsukahara, T. Single nucleotide recognition using a probes-on-carrier DNA chip. *BioTechniques* **2019**, *66*, 73-78, doi:10.2144/btn-2018-0088.
6. Saifullah; Tsukahara, T. Genotyping of single nucleotide polymorphisms using the SNP-RFLP method. *Bioscience trends* **2018**, *12 3*, 240-246, doi:https://doi.org/10.5582/bst.2018.01102.
7. Ledford, H. Gene-silencing technology gets first drug approval after 20-year wait. *Nature* **2018**, *560*, 291-292, doi:10.1038/d41586-018-05867-7.
8. Zhang, L.; Zhang, T.; Wang, L.; Shao, S.; Chen, Z.; Zhang, Z. In vivo targeted delivery of CD40 shRNA to mouse intestinal dendritic cells by oral administration of recombinant Sacchromyces cerevisiae. *Gene Therapy* **2014**, *21*, 709-714, doi:10.1038/gt.2014.50.
9. Hanna, J.; Hossain, G.S.; Kocerha, J. The Potential for microRNA Therapeutics and Clinical Research. *Frontiers in Genetics* **2019**, *10*, doi:10.3389/fgene.2019.00478.
10. Kinali, M.; Arechavala-Gomeza, V.; Feng, L.; Cirak, S.; Hunt, D.; Adkin, C.; Guglieri, M.; Ashton, E.; Abbs, S.; Nihoyannopoulos, P., et al. Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *The Lancet Neurology* **2009**, *8*, 918-928, doi:https://doi.org/10.1016/S1474-4422(09)70211-X.
11. Hirakawa, M.P.; Krishnakumar, R.; Timlin, J.A.; Carney, J.P.; Butler, K.S. Gene editing and CRISPR in the clinic: current and future perspectives. *Biosci Rep* **2020**, *40*, BSR20200127, doi:10.1042/BSR20200127.
12. Titze-de-Almeida, S.S.; Brandão, P.R.P.; Faber, I.; Titze-de-Almeida, R. Leading RNA Interference Therapeutics Part 1: Silencing Hereditary Transthyretin Amyloidosis, with a Focus on Patisiran. **2020**, *24*, 49-59, doi:10.1007/s40291-019-00434-w.
13. de Paula Brandão, P.R.; Titze-de-Almeida, S.S.; Titze-de-Almeida, R. Leading RNA Interference Therapeutics Part 2: Silencing Delta-Aminolevulinic Acid Synthase 1, with a Focus on Givosiran. **2020**, *24*, 61-68, doi:10.1007/s40291-019-00438-6.
14. Fire, A.; Xu, S.; Montgomery, M.K.; Kostas, S.A.; Driver, S.E.; Mello, C.C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **1998**, *391*, 806-811, doi:10.1038/35888.

15. Dykxhoorn, D.M.; Palliser, D.; Lieberman, J. The silent treatment: siRNAs as small molecule drugs. *Gene Therapy* **2006**, *13*, 541-552, doi:10.1038/sj.gt.3302703.
16. Stenvang, J.; Petri, A.; Lindow, M.; Obad, S.; Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* **2012**, *3*, 1, doi:10.1186/1758-907X-3-1.
17. Broderick, J.A.; Zamore, P.D. MicroRNA therapeutics. *Gene Therapy* **2011**, *18*, 1104-1110, doi:10.1038/gt.2011.50.
18. Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic siRNA: state of the art. *Signal Transduction and Targeted Therapy* **2020**, *5*, 101, doi:10.1038/s41392-020-0207-x.
19. Aagaard, L.; Rossi, J.J. RNAi therapeutics: Principles, prospects and challenges. *Advanced Drug Delivery Reviews* **2007**, *59*, 75-86, doi:https://doi.org/10.1016/j.addr.2007.03.005.
20. Abudayyeh, O.O.; Gootenberg, J.S.; Essletzbichler, P.; Han, S.; Joung, J.; Belanto, J.J.; Verdine, V.; Cox, D.B.T.; Kellner, M.J.; Regev, A., et al. RNA targeting with CRISPR–Cas13. *Nature* **2017**, *550*, 280-284, doi:10.1038/nature24049.
21. Saifullah; Sakari, M.; Suzuki, T.; Yano, S.; Tsukahara, T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. *International Journal of Molecular Sciences* **2020**, *21*, 8904.
22. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816-821, doi:10.1126/science.1225829.
23. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* **1987**, *169*, 5429-5433, doi:10.1128/jb.169.12.5429-5433.1987.
24. Mojica, F.J.M.; Díez-Villaseñor, C.; Soria, E.; Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* **2000**, *36*, 244-246, doi:https://doi.org/10.1046/j.1365-2958.2000.01838.x.
25. Jansen, R.; Embden, J.D.A.v.; Gaastra, W.; Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **2002**, *43*, 1565-1575, doi:https://doi.org/10.1046/j.1365-2958.2002.02839.x.
26. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D.A.; Horvath, P. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **2007**, *315*, 1709-1712, doi:10.1126/science.1138140.
27. Garneau, J.E.; Dupuis, M.-È.; Villion, M.; Romero, D.A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadán, A.H.; Moineau, S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **2010**, *468*, 67-71, doi:10.1038/nature09523.
28. Deltcheva, E.; Chylinski, K.; Sharma, C.M.; Gonzales, K.; Chao, Y.; Pirzada, Z.A.; Eckert, M.R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **2011**, *471*, 602-607, doi:10.1038/nature09886.

29. Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-programmed genome editing in human cells. *eLife* **2013**, *2*, e00471, doi:10.7554/eLife.00471.
30. Mali, P.; Yang, L.; Esvelt, K.M.; Aach, J.; Guell, M.; DiCarlo, J.E.; Norville, J.E.; Church, G.M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339*, 823-826, doi:10.1126/science.1232033.
31. Nishimasu, H.; Ran, F.A.; Hsu, Patrick D.; Konermann, S.; Shehata, Soraya I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156*, 935-949, doi:10.1016/j.cell.2014.02.001.
32. Jinek, M.; Jiang, F.; Taylor, D.W.; Sternberg, S.H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S., et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **2014**, *343*, 1247997, doi:10.1126/science.1247997.
33. Zetsche, B.; Gootenberg, Jonathan S.; Abudayyeh, Omar O.; Slaymaker, Ian M.; Makarova, Kira S.; Essletzbichler, P.; Volz, Sara E.; Joung, J.; van der Oost, J.; Regev, A., et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163*, 759-771, doi:10.1016/j.cell.2015.09.038.
34. Doudna, J.A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096, doi:10.1126/science.1258096.
35. Shmakov, S.; Abudayyeh, Omar O.; Makarova, Kira S.; Wolf, Yuri I.; Gootenberg, Jonathan S.; Semenova, E.; Minakhin, L.; Joung, J.; Konermann, S.; Severinov, K., et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* **2015**, *60*, 385-397, doi:10.1016/j.molcel.2015.10.008.
36. Abudayyeh, O.O.; Gootenberg, J.S.; Konermann, S.; Joung, J.; Slaymaker, I.M.; Cox, D.B.T.; Shmakov, S.; Makarova, K.S.; Semenova, E.; Minakhin, L., et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016**, *353*, aaf5573, doi:10.1126/science.aaf5573.
37. Wang, Q.; Liu, X.; Zhou, J.; Yang, C.; Wang, G.; Tan, Y.; Wu, Y.; Zhang, S.; Yi, K.; Kang, C. The CRISPR-Cas13a Gene-Editing System Induces Collateral Cleavage of RNA in Glioma Cells. *Advanced Science* **2019**, *6*, 1901299, doi:https://doi.org/10.1002/advs.201901299.
38. Gao, J.; Luo, T.; Lin, N.; Zhang, S.; Wang, J. A New Tool for CRISPR-Cas13a-Based Cancer Gene Therapy. *Molecular Therapy - Oncolytics* **2020**, *19*, 79-92, doi:10.1016/j.omto.2020.09.004.
39. Makarova, K.S.; Wolf, Y.I.; Alkhnbashi, O.S.; Costa, F.; Shah, S.A.; Saunders, S.J.; Barrangou, R.; Brouns, S.J.J.; Charpentier, E.; Haft, D.H., et al. An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **2015**, *13*, 722-736, doi:10.1038/nrmicro3569.
40. 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., et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **2020**, *18*, 67-83, doi:10.1038/s41579-019-0299-x.
41. Benda, C.; Ebert, J.; Scheltema, Richard A.; Schiller, Herbert B.; Baumgärtner, M.; Bonneau, F.; Mann, M.; Conti, E. Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Molecular Cell* **2014**, *56*, 43-54, doi:https://doi.org/10.1016/j.molcel.2014.09.002.

42. Taylor, D.W.; Zhu, Y.; Staals, R.H.J.; Kornfeld, J.E.; Shinkai, A.; van der Oost, J.; Nogales, E.; Doudna, J.A. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science* **2015**, *348*, 581-585, doi:10.1126/science.aaa4535.
43. Burmistrz, M.; Krakowski, K.; Krawczyk-Balska, A. RNA-Targeting CRISPR–Cas Systems and Their Applications. *International Journal of Molecular Sciences* **2020**, *21*, 1122.
44. Nickel, L.; Ulbricht, A.; Alkhnbashi, O.S.; Förstner, K.U.; Cassidy, L.; Weidenbach, K.; Backofen, R.; Schmitz, R.A. Cross-cleavage activity of Cas6b in crRNA processing of two different CRISPR-Cas systems in Methanosarcina mazei Gö1. *RNA Biology* **2019**, *16*, 492-503, doi:10.1080/15476286.2018.1514234.
45. Niewoehner, O.; Garcia-Doval, C.; Rostøl, J.T.; Berk, C.; Schwede, F.; Bigler, L.; Hall, J.; Marraffini, L.A.; Jinek, M. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **2017**, *548*, 543-548, doi:10.1038/nature23467.
46. O'Connell, M.R.; Oakes, B.L.; Sternberg, S.H.; East-Seletsky, A.; Kaplan, M.; Doudna, J.A. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **2014**, *516*, 263-266, doi:10.1038/nature13769.
47. Yan, W.X.; Hunnewell, P.; Alfonse, L.E.; Carte, J.M.; Keston-Smith, E.; Sothiselvam, S.; Garrity, A.J.; Chong, S.; Makarova, K.S.; Koonin, E.V., et al. Functionally diverse type V CRISPR-Cas systems. *Science* **2019**, *363*, 88-91, doi:10.1126/science.aav7271.
48. Cox, D.B.T.; Gootenberg, J.S.; Abudayyeh, O.O.; Franklin, B.; Kellner, M.J.; Joung, J.; Zhang, F. RNA editing with CRISPR-Cas13. *Science* **2017**, *358*, 1019-1027, doi:10.1126/science.aaq0180.
49. Konermann, S.; Lotfy, P.; Brideau, N.J.; Oki, J.; Shokhirev, M.N.; Hsu, P.D. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **2018**, *173*, 665-676.e614, doi:https://doi.org/10.1016/j.cell.2018.02.033.
50. Knott, G.J.; East-Seletsky, A.; Cofsky, J.C.; Holton, J.M.; Charles, E.; O'Connell, M.R.; Doudna, J.A. Guide-bound structures of an RNA-targeting A-cleaving CRISPR–Cas13a enzyme. *Nature Structural & Molecular Biology* **2017**, *24*, 825-833, doi:10.1038/nsmb.3466.
51. O'Connell, M.R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *Journal of Molecular Biology* **2019**, *431*, 66-87, doi:https://doi.org/10.1016/j.jmb.2018.06.029.
52. East-Seletsky, A.; O'Connell, M.R.; Knight, S.C.; Burstein, D.; Cate, J.H.D.; Tjian, R.; Doudna, J.A. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **2016**, *538*, 270-273, doi:10.1038/nature19802.
53. Liu, L.; Li, X.; Ma, J.; Li, Z.; You, L.; Wang, J.; Wang, M.; Zhang, X.; Wang, Y. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **2017**, *170*, 714-726.e710, doi:https://doi.org/10.1016/j.cell.2017.06.050.
54. 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., et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine* **2019**, *25*, 249-254, doi:10.1038/s41591-018-0326-x.
55. Suprenant, K.A.; Dean, K.; McKee, J.; Hake, S. EMAP, an echinoderm microtubule-associated protein found in microtubule-ribosome complexes. *Journal of Cell Science* **1993**, *104*, 445-450.

56. Sabir, S.R.; Yeoh, S.; Jackson, G.; Bayliss, R. EML4-ALK Variants: Biological and Molecular Properties, and the Implications for Patients. *Cancers* **2017**, *9*, 118.
57. Suprenant, K.A.; Tuxhorn, J.A.; Daggett, M.A.; Ahrens, D.P.; Hostetler, A.; Palange, J.M.; VanWinkle, C.E.; Livingston, B.T. Conservation of the WD-repeat, microtubule-binding protein, EMAP, in sea urchins, humans, and the nematode C. elegans. *Development genes and evolution* **2000**, *210*, 2-10, doi:10.1007/pl00008183.
58. Bayliss, R.; Choi, J.; Fennell, D.A.; Fry, A.M.; Richards, M.W. Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. *Cellular and molecular life sciences : CMLS* **2016**, *73*, 1209-1224, doi:10.1007/s00018-015-2117-6.
59. Eichenmüller, B.; Everley, P.; Palange, J.; Lepley, D.; Suprenant, K.A. The Human EMAP-like Protein-70 (ELP70) Is a Microtubule Destabilizer That Localizes to the Mitotic Apparatus *. *Journal of Biological Chemistry* **2002**, *277*, 1301-1309, doi:10.1074/jbc.M106628200.
60. Cohen, J. Response. *Science* **1995**, *267*, 316-316, doi:10.1126/science.7824924.
61. Shiota, M.; Fujimoto, J.; Semba, T.; Satoh, H.; Yamamoto, T.; Mori, S. Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* **1994**, *9*, 1567-1574.
62. Bossi, R.T.; Saccardo, M.B.; Ardini, E.; Menichincheri, M.; Rusconi, L.; Magnaghi, P.; Orsini, P.; Avanzi, N.; Borgia, A.L.; Nesi, M., et al. Crystal Structures of Anaplastic Lymphoma Kinase in Complex with ATP Competitive Inhibitors. *Biochemistry* **2010**, *49*, 6813-6825, doi:10.1021/bi1005514.
63. Bilsland, J.G.; Wheeldon, A.; Mead, A.; Znamenskiy, P.; Almond, S.; Waters, K.A.; Thakur, M.; Beaumont, V.; Bonnert, T.P.; Heavens, R., et al. Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **2008**, *33*, 685-700, doi:10.1038/sj.npp.1301446.
64. Vernersson, E.; Khoo, N.K.S.; Henriksson, M.L.; Roos, G.; Palmer, R.H.; Hallberg, B. Characterization of the expression of the ALK receptor tyrosine kinase in mice. *Gene Expression Patterns* **2006**, *6*, 448-461, doi:https://doi.org/10.1016/j.modgep.2005.11.006.
65. Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.; Chirieac, L.R.; D'Amico, T.A.; DeCamp, M.M.; Dilling, T.J.; Dobelbower, M., et al. Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN* **2017**, *15*, 504-535, doi:10.6004/jnccn.2017.0050.
66. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* **2008**, *83*, 584-594, doi:10.4065/83.5.584.
67. Soda, M.; Choi, Y.L.; Enomoto, M.; Takada, S.; Yamashita, Y.; Ishikawa, S.; Fujiwara, S.-i.; Watanabe, H.; Kurashina, K.; Hatanaka, H., et al. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **2007**, *448*, 561-566, doi:10.1038/nature05945.
68. Noh, K.W.; Lee, M.S.; Lee, S.E.; Song, J.Y.; Shin, H.T.; Kim, Y.J.; Oh, D.Y.; Jung, K.; Sung, M.; Kim, M., et al. Molecular breakdown: a comprehensive view of anaplastic lymphoma kinase (ALK)-rearranged non-small cell lung cancer. *The Journal of pathology* **2017**, *243*, 307-319, doi:10.1002/path.4950.

69. Koivunen, J.P.; Mermel, C.; Zejnullahu, K.; Murphy, C.; Lifshits, E.; Holmes, A.J.; Choi, H.G.; Kim, J.; Chiang, D.; Thomas, R., et al. EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2008**, *14*, 4275-4283, doi:10.1158/1078-0432.ccr-08-0168.
70. Sabir, S.R.; Yeoh, S.; Jackson, G.; Bayliss, R. EML4-ALK Variants: Biological and Molecular Properties, and the Implications for Patients. *Cancers (Basel)* **2017**, *9*, doi:10.3390/cancers9090118.
71. Hallberg, B.; Palmer, R.H. The role of the ALK receptor in cancer biology. *Annals of Oncology* **2016**, *27*, iii4-iii15, doi:https://doi.org/10.1093/annonc/mdw301.
72. Katayama, R.; Khan, T.M.; Benes, C.; Lifshits, E.; Ebi, H.; Rivera, V.M.; Shakespeare, W.C.; Iafrate, A.J.; Engelman, J.A.; Shaw, A.T. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proceedings of the National Academy of Sciences* **2011**, *108*, 7535-7540, doi:10.1073/pnas.1019559108.
73. Lin, J.J.; Schoenfeld, A.J.; Zhu, V.W.; Yeap, B.Y.; Chin, E.; Rooney, M.; Plodkowski, A.J.; Digumarthy, S.R.; Dagogo-Jack, I.; Gainor, J.F., et al. Efficacy of Platinum/Pemetrexed Combination Chemotherapy in ALK-Positive NSCLC Refractory to Second-Generation ALK Inhibitors. *Journal of Thoracic Oncology* **2020**, *15*, 258-265, doi:10.1016/j.jtho.2019.10.014.
74. Malik, S.M.; Maher, V.E.; Bijwaard, K.E.; Becker, R.L.; Zhang, L.; Tang, S.W.; Song, P.; Liu, Q.; Marathe, A.; Gehrke, B., et al. U.S. Food and Drug Administration Approval: Crizotinib for Treatment of Advanced or Metastatic Non–Small Cell Lung Cancer That Is Anaplastic Lymphoma Kinase Positive. *Clinical Cancer Research* **2014**, *20*, 2029-2034, doi:10.1158/1078-0432.ccr-13-3077.
75. Christensen, J.G.; Zou, H.Y.; Arango, M.E.; Li, Q.; Lee, J.H.; McDonnell, S.R.; Yamazaki, S.; Alton, G.R.; Mroczkowski, B.; Los, G. Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma. *Molecular cancer therapeutics* **2007**, *6*, 3314-3322, doi:10.1158/1535-7163.mct-07-0365.
76. Yasuda, H.; de Figueiredo-Pontes, L.L.; Kobayashi, S.; Costa, D.B. Preclinical Rationale for Use of the Clinically Available Multitargeted Tyrosine Kinase Inhibitor Crizotinib in ROS1-Translocated Lung Cancer. *Journal of Thoracic Oncology* **2012**, *7*, 1086-1090, doi:https://doi.org/10.1097/JTO.0b013e3182570919.
77. Shaw, A.T.; Kim, D.-W.; Nakagawa, K.; Seto, T.; Crinó, L.; Ahn, M.-J.; De Pas, T.; Besse, B.; Solomon, B.J.; Blackhall, F., et al. Crizotinib versus Chemotherapy in Advanced ALK-Positive Lung Cancer. *New England Journal of Medicine* **2013**, *368*, 2385-2394, doi:10.1056/NEJMoa1214886.
78. Camidge, D.R.; Doebele, R.C. Treating ALK-positive lung cancer—early successes and future challenges. *Nature Reviews Clinical Oncology* **2012**, *9*, 268-277, doi:10.1038/nrclinonc.2012.43.
79. Gainor, J.F.; Dardaei, L.; Yoda, S.; Friboulet, L.; Leshchiner, I.; Katayama, R.; Dagogo-Jack, I.; Gadgeel, S.; Schultz, K.; Singh, M., et al. Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer discovery* **2016**, *6*, 1118-1133, doi:10.1158/2159-8290.cd-16-0596.

80. Choi, Y.L.; Soda, M.; Yamashita, Y.; Ueno, T.; Takashima, J.; Nakajima, T.; Yatabe, Y.; Takeuchi, K.; Hamada, T.; Haruta, H., et al. EML4-ALK Mutations in Lung Cancer That Confer Resistance to ALK Inhibitors. *New England Journal of Medicine* **2010**, *363*, 1734-1739, doi:10.1056/NEJMoa1007478.
81. Mizuta, H.; Okada, K.; Araki, M.; Adachi, J.; Takemoto, A.; Kutkowska, J.; Maruyama, K.; Yanagitani, N.; Oh-hara, T.; Watanabe, K., et al. Gilteritinib overcomes lorlatinib resistance in ALK-rearranged cancer. *Nature communications* **2021**, *12*, 1261, doi:10.1038/s41467-021-21396-w.
82. Camidge, D.R.; Dziadziuszko, R.; Peters, S.; Mok, T.; Noe, J.; Nowicka, M.; Gadgeel, S.M.; Cheema, P.; Pavlakis, N.; de Marinis, F., et al. Updated Efficacy and Safety Data and Impact of the EML4-ALK Fusion Variant on the Efficacy of Alectinib in Untreated ALK-Positive Advanced Non–Small Cell Lung Cancer in the Global Phase III ALEX Study. *Journal of Thoracic Oncology* **2019**, *14*, 1233-1243, doi:https://doi.org/10.1016/j.jtho.2019.03.007.
83. Nakagawa, K.; Hida, T.; Nokihara, H.; Morise, M.; Azuma, K.; Kim, Y.H.; Seto, T.; Takiguchi, Y.; Nishio, M.; Yoshioka, H., et al. Final progression-free survival results from the J-ALEX study of alectinib versus crizotinib in ALK-positive non-small-cell lung cancer. *Lung Cancer* **2020**, *139*, 195-199, doi:https://doi.org/10.1016/j.lungcan.2019.11.025.
84. Yanagitani, N.; Uchibori, K.; Koike, S.; Tsukahara, M.; Kitazono, S.; Yoshizawa, T.; Horiike, A.; Ohyanagi, F.; Tambo, Y.; Nishikawa, S., et al. Drug resistance mechanisms in Japanese anaplastic lymphoma kinase-positive non-small cell lung cancer and the clinical responses based on the resistant mechanisms. *Cancer Sci* **2020**, *111*, 932-939, doi:10.1111/cas.14314.
85. Shaw, A.T.; Felip, E.; Bauer, T.M.; Besse, B.; Navarro, A.; Postel-Vinay, S.; Gainor, J.F.; Johnson, M.; Dietrich, J.; James, L.P., et al. Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial. *The Lancet Oncology* **2017**, *18*, 1590-1599, doi:https://doi.org/10.1016/S1470-2045(17)30680-0.
86. Solomon, B.J.; Besse, B.; Bauer, T.M.; Felip, E.; Soo, R.A.; Camidge, D.R.; Chiari, R.; Bearz, A.; Lin, C.-C.; Gadgeel, S.M., et al. Lorlatinib in patients with ALK-positive non-small-cell lung cancer: results from a global phase 2 study. *The Lancet Oncology* **2018**, *19*, 1654-1667, doi:https://doi.org/10.1016/S1470-2045(18)30649-1.
87. Zou, Helen Y.; Friboulet, L.; Kodack, David P.; Engstrom, Lars D.; Li, Q.; West, M.; Tang, Ruth W.; Wang, H.; Tsaparikos, K.; Wang, J., et al. PF-06463922, an ALK/ROS1 Inhibitor, Overcomes Resistance to First and Second Generation ALK Inhibitors in Preclinical Models. *Cancer Cell* **2015**, *28*, 70-81, doi:https://doi.org/10.1016/j.ccell.2015.05.010.
88. Mogenet, A.; Tomasini, P.; Greillier, L.; Barlesi, F. Lorlatinib: an additional option for ALK-positive non-small cell lung cancer? *Translational Lung Cancer Research* **2019**, S383-S386.
89. Okada, K.; Araki, M.; Sakashita, T.; Ma, B.; Kanada, R.; Yanagitani, N.; Horiike, A.; Koike, S.; Oh-hara, T.; Watanabe, K., et al. Prediction of ALK mutations mediating ALK-TKIs resistance and drug re-purposing to overcome the resistance. *EBioMedicine* **2019**, *41*, 105-119, doi:10.1016/j.ebiom.2019.01.019.
90. Recondo, G.; Mezquita, L.; Facchinetti, F.; Planchard, D.; Gazzah, A.; Bigot, L.; Rizvi, A.Z.; Frias, R.L.; Thiery, J.P.; Scoazec, J.-Y., et al. Diverse

Resistance Mechanisms to the Third-Generation ALK Inhibitor Lorlatinib in ALK-Rearranged Lung Cancer. *Clinical Cancer Research* **2020**, *26*, 242-255, doi:10.1158/1078-0432.ccr-19-1104.

91. Shaw, A.T.; Bauer, T.M.; de Marinis, F.; Felip, E.; Goto, Y.; Liu, G.; Mazieres, J.; Kim, D.-W.; Mok, T.; Polli, A., et al. First-Line Lorlatinib or Crizotinib in Advanced ALK-Positive Lung Cancer. *New England Journal of Medicine* **2020**, *383*, 2018-2029, doi:10.1056/NEJMoa2027187.
92. Shaw, A.T.; Varghese, A.M.; Solomon, B.J.; Costa, D.B.; Novello, S.; Mino-Kenudson, M.; Awad, M.M.; Engelman, J.A.; Riely, G.J.; Monica, V., et al. Pemetrexed-based chemotherapy in patients with advanced, ALK-positive non-small cell lung cancer. *Annals of Oncology* **2013**, *24*, 59-66, doi:https://doi.org/10.1093/annonc/mds242.
93. Solomon, B.J.; Mok, T.; Kim, D.-W.; Wu, Y.-L.; Nakagawa, K.; Mekhail, T.; Felip, E.; Cappuzzo, F.; Paolini, J.; Usari, T., et al. First-Line Crizotinib versus Chemotherapy in ALK-Positive Lung Cancer. *New England Journal of Medicine* **2014**, *371*, 2167-2177, doi:10.1056/NEJMoa1408440.
94. Soria, J.-C.; Tan, D.S.W.; Chiari, R.; Wu, Y.-L.; Paz-Ares, L.; Wolf, J.; Geater, S.L.; Orlov, S.; Cortinovis, D.; Yu, C.-J., et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. *The Lancet* **2017**, *389*, 917-929, doi:10.1016/S0140-6736(17)30123-X.
95. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B., et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **2020**, *49*, D1388-D1395, doi:10.1093/nar/gkaa971.
96. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians n/a*, doi:https://doi.org/10.3322/caac.21660.

Chapter II

Exploration of clinical consequence and associated genes of ALK expression in cancers using integrative bioinformatics

Graduate School of Advanced Science and Technology
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (JAIST), JAPAN

Chapter II

Exploration of clinical consequence and associated genes of *ALK* expression in cancers using integrative bioinformatics

2. Outline

Anaplastic lymphoma kinase (ALK) is a tyrosine kinase receptor that is genetically altered in several cancers, such as non-small cell lung cancer (NSCLC), melanoma, lymphoma, and other tumors. Although *ALK* is associated with various cancers, the relationship between *ALK* expression and patient prognosis in different cancers is poorly understood. Here, I show a correlation between *ALK* expression and its clinical outcome in patients with lung adenocarcinoma (LUAD), melanoma, ovarian carcinoma (OV), diffuse large B-cell lymphoma (DLBC), acute myeloid leukemia (AML), and breast cancer (BC) using different computational web portals. I analyzed *ALK* transcriptional expression, patient survival rate, genetic alteration, protein network, and gene and microRNA (miRNA) co-expression. I found higher *ALK* expression compared to normal tissues in LUAD, melanoma, and OV, which are associated with poor patient survival rates. In contrast, lower transcriptional expression deteriorates the survival rate of patients with DLBC, AML, and BC. I identified 214 missense mutations, 24 truncating mutations, seven fusions, and two in-frame mutations, with the highest alteration of *ALK* in melanoma. I further showed that 17 genes and 19 miRNAs were exclusively co-expressed and found that echinoderm microtubule-associated protein-like 4 (EML4) was the most positively correlated gene (log odds ratio >3). The gene ontology and signaling pathways of the genes co-expressed with *ALK* involved in these six cancers were also identified. My findings offer a basis for *ALK* as a prognostic biomarker and therapeutic target in cancers, which will potentially contribute to precision oncology and help clinicians regarding treatment options.

Keywords: *ALK* expression, cancers, integrative bioinformatics, patient prognosis, multi-omics analysis, co-expression, signaling pathway

2.1. Introduction

Anaplastic lymphoma kinase (ALK) is a transmembrane tyrosine kinase receptor belonging to the insulin receptor superfamily [1]. Genetic aberrations of *ALK*, including gene fusions, translocations, or inversions with different gene partners, have been identified in many solid cancers. For instance, *EML4-ALK* oncofusion in NSCLC [2,3], nucleophosmin (NPM)-ALK fusion oncogene in anaplastic large cell lymphoma (ALCL) [4], truncating *ALK* in melanoma [5,6], and translocation of *ALK* in ovarian serous carcinoma [7-9]. *ALK* translocations also appear in DLBC [10,11], AML [12,13], BC [14,15], and renal cell carcinomas [16] at low rates. The resultant X-ALK proteins comprise the C-terminal part of the entire intracellular tyrosine kinase domain and the N-terminal part of the fusion proteins, usually the di- or trimerization domain [17].

Once the fusion protein is activated by di- or trimerization, ALK plays a primary role in constitutive autophosphorylation, leading to the activation of downstream signaling and subsequent arrest of cell proliferation and growth. There are four main signaling pathways downstream of kinase-activated ALK: proto-oncogene protein p21/extracellular-signal-regulated kinase (RAS/ERK), phosphoinositol-3 kinase (PI3K)/Akt, Janus kinase/signal transducer and activator of transcription (JAK/STAT), and phospholipase Cγ (PLCγ) [18,19]. The favorably used pathways show a discrepancy depending on the ALK isoform type (wild type or mutated, fused with a protein partner, and full-length or truncated), and subcellular localization [5,20,21].

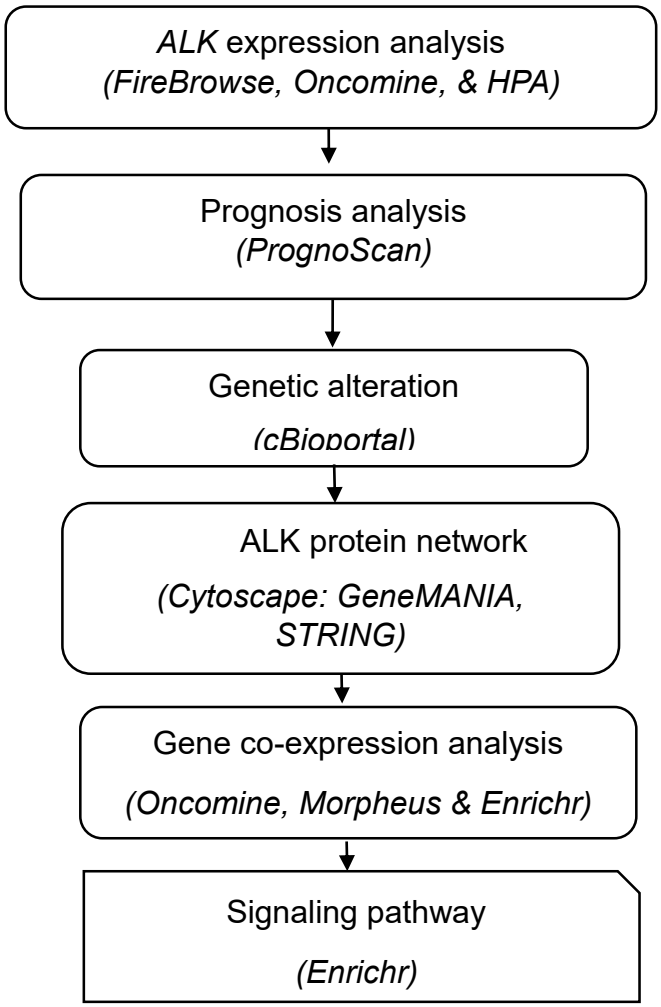
Multi-omics data along with clinical information have revolutionized the field of medicine and biology for a comprehensive understanding of disease genotypes and phenotypes for human health [22-24]. One-dimensional omics data for cancer studies provide only limited information regarding the etiology of oncogenesis and cancer progression [22]. In contrast, multi-omics data can provide a greater understanding of the prognosis and predictive precision of disease phenotypes, thereby improving therapeutic and prevention measurements [25,26]. The prediction of cancer prognosis is a profound point of interest for clinicians, patients with cancer, and healthcare professionals because it facilitates all kinds of decisions concerning patient care and clinical treatment [27,28]. Therefore, prognostic biomarkers have been employed to precisely select patient subgroups using various treatment approaches [29,30].

In this study, I comprehensively analyzed *ALK* expression and its clinical significance in LUAD, melanoma, OV, DLBC, AML, and BC patients using publicly available multi-platform datasets. I evaluated gene and mRNA expression patterns, patient prognosis, genetic alteration analysis, protein-protein interaction network, gene co-expression, microRNA (miRNA) co-expression analysis, gene ontology, and signaling pathways in human LUAD, melanoma, OV, DLBC, AML, and BC. The combined data provide supporting evidence for the use of *ALK* as a prognostic biomarker and therapeutic target.

2.2. Materials and Methods

2.2.1. Experimental design

This study was performed according to the following schematic experimental design:



2.2.2. Expression profile of *ALK* across human normal and cancer tissues

ALK mRNA and protein expression data in normal human tissues were obtained from the Human Protein Atlas, normal tissue immunohistochemistry, Expression Atlas data, and RNA-seq expression data using the Open Targets Genetics Portal (https://genetics.opentargets.org) [31]. Gene expression patterns across normal and cancer tissues were determined using the FireBrowse (http://firebrowse.org/) datasets [32]. The *ALK* query in the FireBrowse database was performed using default settings.

2.2.3. Significant *ALK* expression variation in different cancer types

The expression levels of *ALK* gene in lung cancer (LUAD, LUSC), skin cancer (CM), ovarian cancer (OSC), blood cancer (DLBC, AML), and BC and normal tissue counterparts were assessed using Oncomine (https://www.oncomine.org) [33] web portal. The following parameters were considered to generate a graph from Oncomine data: *p*-value, <0.05; fold change, 2; gene rank, top 10%. The Oncomine server covers microarray data from 86,733 tumors and 12,764 normal tissues. The immunohistochemistry (IHC) analysis was conducted using Human Protein Atlas (HPA) database (http://www.proteinatlas.org) [34]. The default parameters were considered for IHC images.

2.2.4. Prognostic investigation of *ALK* mRNA expression in patients with cancer

The biological relationship between *ALK* expression and prognosis of patients with different cancers was assessed using PrognoScan cancer microarray datasets (http://www.prognoscan.org) [35,36] with a Cox *p*-value <0.05. PrognoScan provides an effective platform for assessing potential cancer biomarkers and therapeutic targets.

2.2.5. Evaluation of *ALK* alteration and associated cancers

The TCGA PanCancer Atlas datasets (10,953 patients/10,967 samples) were used to evaluate the genetic alterations of *ALK* in associated cancers, and cBioPortal cancer genomics data (https://www.cbioportal.org) were used to evaluate the RNA-seq-based mRNA expression levels [37,38].

2.2.6. ALK protein network analysis and clinical significance in cancers

Protein-protein interaction analysis with the target gene was performed using GeneMANIA and STRING platforms with Cytoscape software version 3.5.2 [32]. The predicted 30 proteins were then queried using the cBioPortal platform to identify genetic alterations and clinical significance in lung, skin, ovary, blood, and breast cancers.

2.2.7. Profiling of genes and non-coding RNAs co-occurred and co-expressed with *ALK*

Genes exclusively co-occurring with the target *ALK* gene were analyzed using the cBioPortal platform based on *q*-value ranking (<0.05). The co-expression value of the co-occurring genes was obtained using the Oncomine microarray datasets. The co-expression heatmap was generated using the GraphPad Prism software version 8. Based on the co-expression values, the influence of one gene on the expression of another gene was evaluated using a similarity matrix and hierarchical clustering tools via the Morpheus server [39]. The miRNAs co-expressed with *ALK* and common genes were obtained from the Enrichr web portal (https://maayanlab.cloud /Enrichr/) [40,41] and scrutinized using Cytoscape_v3.8.2 software.

2.2.8. Elucidation of Gene ontologies (GOs) and signaling pathways of *ALK* and its correlated genes

The GOs and signaling pathways of *ALK* with co-occurring genes and relevant bar plots were derived from the Enrichr platform [40,41]. Enrichr is an integrative web-based gene-list enrichment analysis tool that compares different genomic datasets. The input genes were classified into biological processes, molecular functions, and cellular components according to the GO terms. Similarly, signaling pathways were defined using Reactome 2016, PANTHER 2016, and KEGG 2019. A Cox *p*-value <0.05 was considered statistically significant for both GO and signaling pathway bar plots.

2.3. Results

2.3.1. Expression profile of *ALK* across human normal and cancer tissues

I observed *ALK* RNA and protein expression in various human normal tissues using the Open Target Genetics platform [31]. Both RNA and protein expression levels were the highest in the brain. The bar graph illustrates that *ALK* expression is higher in the skin, reproductive, and lung tissues than in the blood and breast tissues (Figure 1A). Next, the FireBrowse dataset [32] was used to examine *ALK* transcript expression in multiple human cancers. The results showed an almost similar pattern of *ALK* gene expression in the corresponding cancer tissues (Figure 1B), which primarily suggests that *ALK* is involved in the tumorigenesis of the skin, ovary, lungs, blood, and breast cancers.

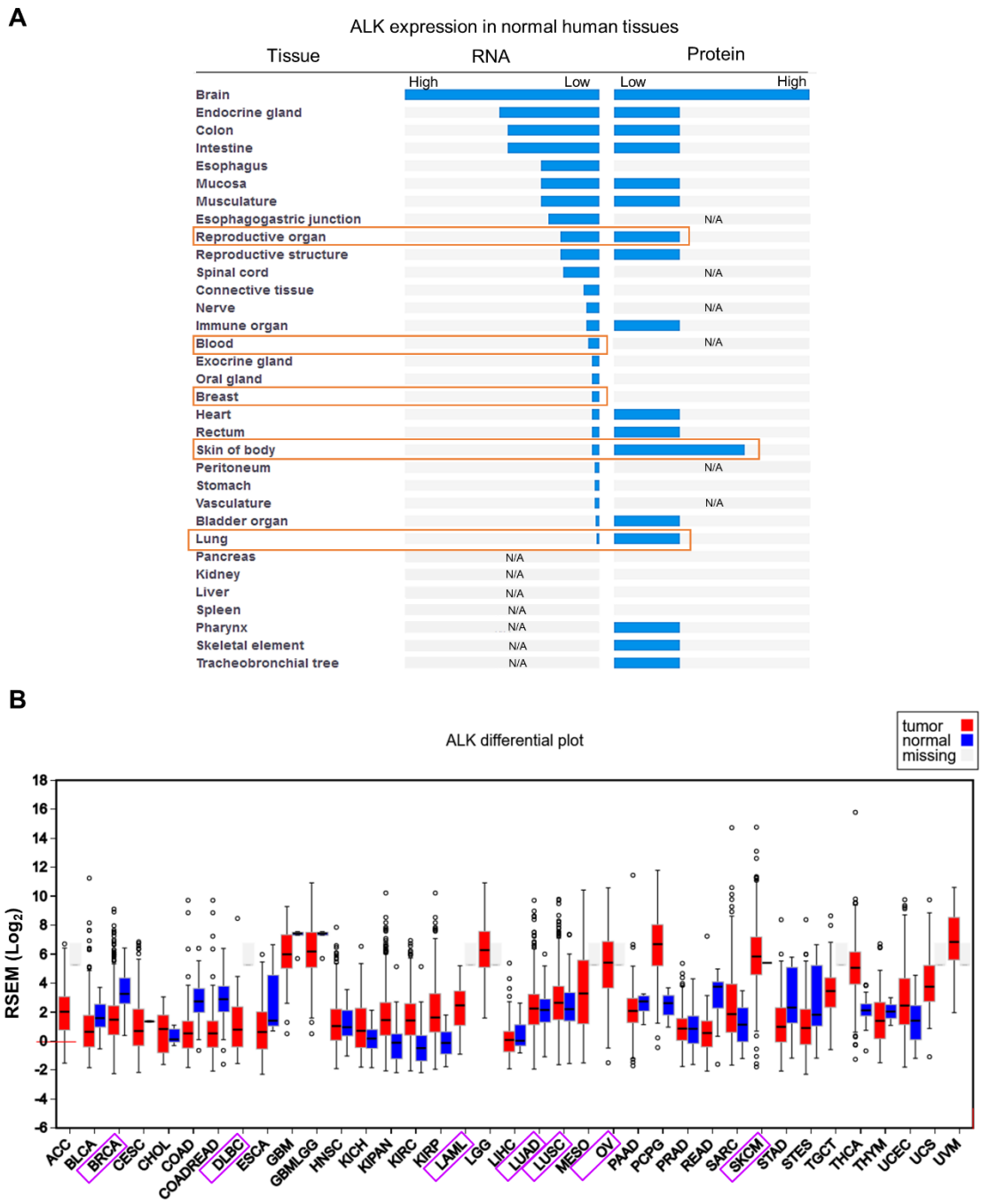


Figure 1. Expression profile of anaplastic lymphoma kinase (ALK) across human normal and cancer tissues. A, Outline of *ALK* mRNA and protein expression in normal human tissues, based on Human Protein Atlas (HPA) normal tissue immunohistochemistry, Expression Atlas data, and RNA-seq expression data derived from Open Targets Genetics platform. **B**, RSEM (RNA-Seq by Expectation-Maximization) RNA-Seq expression profiles for each cancer and corresponding normal tissue were obtained from The Cancer Genome Atlas (TCGA) obtained using the FireBrowse datasets. The boxes denote the median and the 25th and 75th % dots

symbolize outliers. The red boxes represent tumor tissues, while the blue boxes represent corresponding normal tissues. The gray boxes signify that no normal samples exist for that disease cohort.

2.3.2. Significant variation in ALK expression in different cancer types

For further examination, interactive web portals Oncomine [33] and HPA [34] were used for assessing the ALK expression in lung, skin, ovary, blood, and breast cancers. As shown in Figure 2A–F, a significantly higher expression of *ALK* mRNA was observed in the lung (LUSC and LUAD), skin (CM), and ovary (OSC) and lower expression in the blood (DLBC and AML) and breast cancer (IBC) than in the healthy controls. The results indicated strong evidence of upregulation of *ALK* in LUSC, LUAD, CM, and OSC and downregulation in DLBC, AML, and IBC tissues compared to their corresponding normal tissues. Next, an elevated expression of ALK was confirmed at the protein level in lung (LUAD), skin (melanoma), and ovarian cancer (Figure 2G–I). Which suggested a consistency between RNA and protein expression of ALK in lung (LUAD), skin (melanoma), and ovarian cancer.

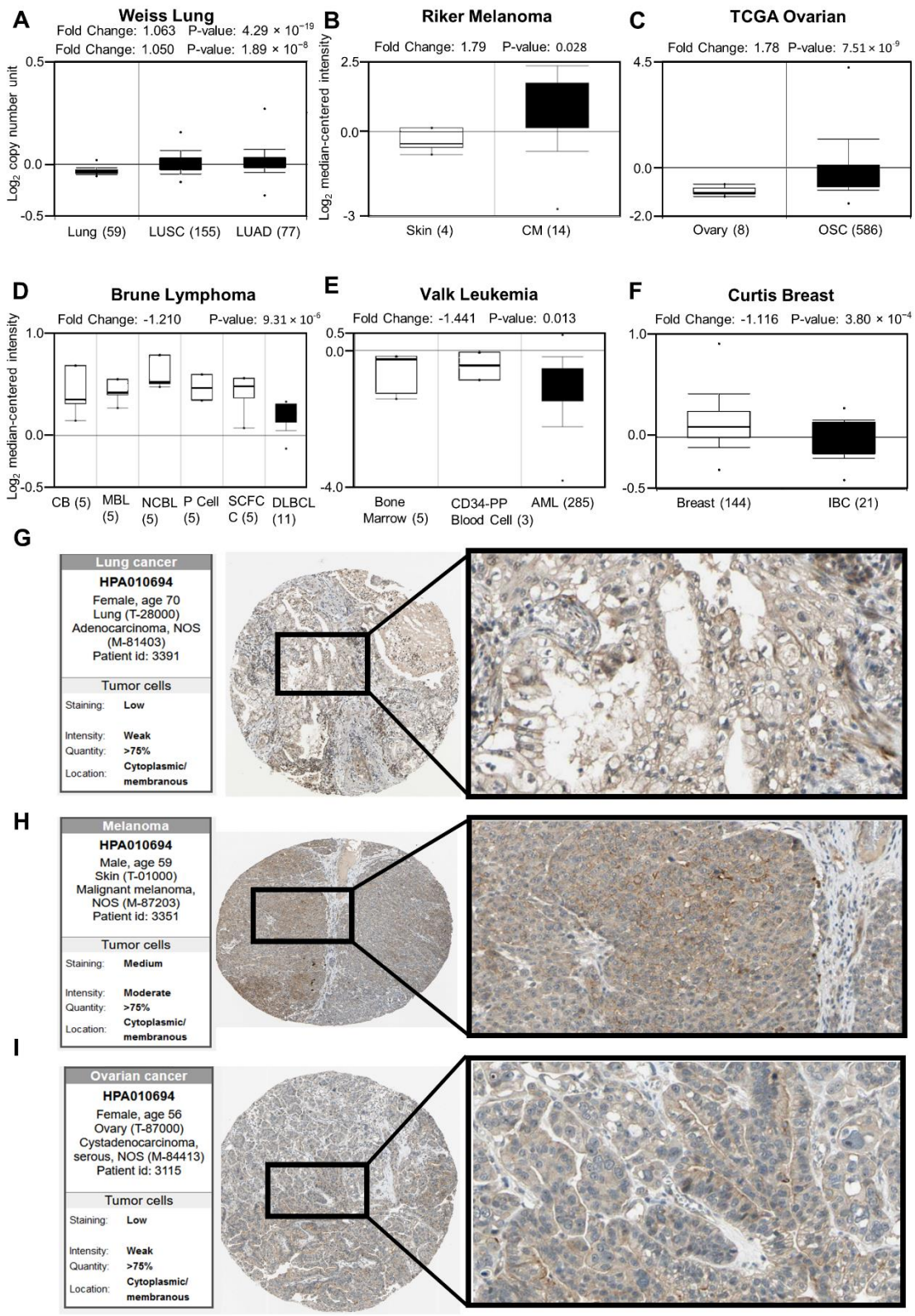


Figure 2. Anaplastic lymphoma kinase (ALK) expression pattern in various cancer types. A–**F**, *ALK* expression in six cancer types derived from Oncomine cancer microarray database. The left side box plot represents *ALK* expression in normal tissue, while the right side box represents cancer tissue. Statistically significant difference

between normal and cancer tissues is indicated by $p < 0.05$. LUSC, Lung squamous cell carcinoma; LUAD, lung adenocarcinoma; CM, Cutaneous melanoma; OSC, Ovarian serous carcinoma CB, Centroblast; MBL, Memory B-lymphocyte; NCBL, Naive Pre-germinal Center B-lymphocyte; P cell, Plasma Cell; SCFC, Small Cleaved Follicle Center Cell; DLBCL, Diffuse Large B-Cell Lymphoma; CD-34 PP, CD34-Positive Peripheral Blood Cell; AML, Acute Myeloid Leukemia and IBC, Invasive Breast Carcinoma. *ALK* was substantially upregulated in lung, skin, and ovary cancers. In contrast, *ALK* expression was considerably downregulated in blood and breast cancers. The sample number is indicated inside the brackets. **G**–I, Immunohistochemistry (IHC) images of ALK protein expression in lung (LUAD), skin (melanoma), and ovary (ovarian cancer) derived from HPA (Human Protein Atlas). The brown staining denotes the antibody is conjugated with its corresponding antigen.

2.3.3. Prognostic investigation of *ALK* mRNA expression in patients with cancer

The correlation between the level of *ALK* expression and the survival rate of patients with lung, skin, ovarian, blood, and breast cancers was determined using the PrognoScan dataset [35,36] with significant Cox *p*-values (<0.05). The analysis showed that *ALK* overexpression is negatively proportional to the survival rate of patients with LUAD (Jacob-00182-UM dataset), melanoma (GSE19234), and ovarian cancers (GSE9891) with increased risk association (hazard ratio, HR, >1) (Figure 3A–C and G, Table S1). In contrast, the differential expression of *ALK* was associated with a higher death rate in DLBCL (E-TABM-346), AML (GSE8970), and BC (GSE9893) with lower hazard ratios (HR, >0.5) (Figure 3D-F and G). In addition, the dataset GSE8894 showed that lower expression of *ALK* deteriorates the survival rate of patients with NSCLC, whereas higher expression leads to lower relapse-free survival with higher risk (HR, >1) in breast cancers, as shown in Figure S1 and Table S1. Therefore, the obtained data suggested that the deregulated expression of *ALK* could elicit a poor patient prognosis.

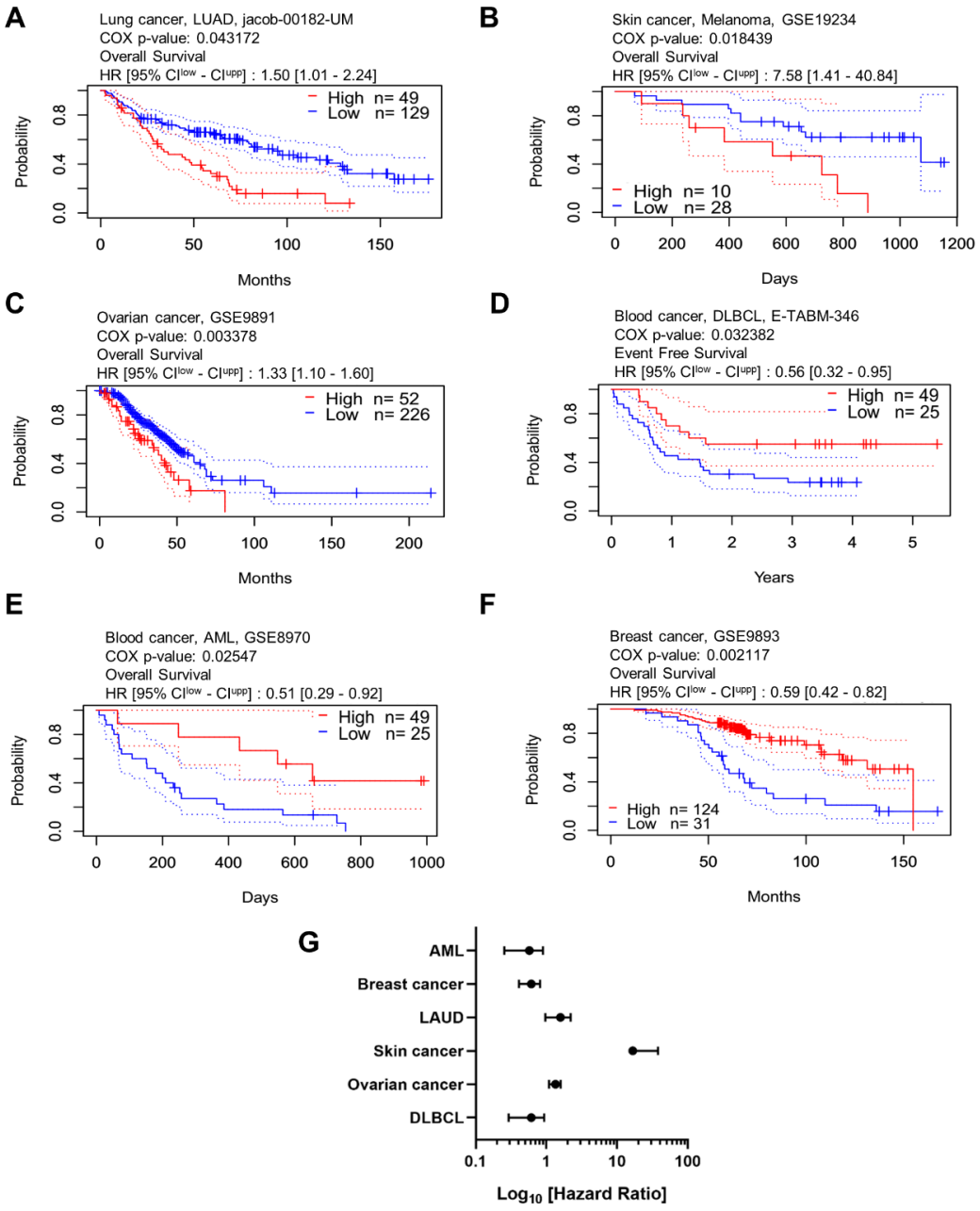


Figure 3. Anaplastic lymphoma kinase (ALK) expression and clinical prognosis in six different cancers retrieved from PrognoScan microarray cancer database. Kaplan–Meier patient survival estimate of lung adenocarcinoma **A**, melanoma **B**, ovarian cancer **C**, blood cancer: DLBCL and AML **D**–**E**, and breast cancer **F** for *ALK* expression. The survival curve was determined as the threshold of Cox *p*-value <0.05 and *p*-value <0.01. The red line denotes high expression, while the blue line denotes low expression. The dotted line symbolizes maximum and minimum values of the survival average. HR, hazard ratio; CI, confidence interval; n, number of patients. **G**, The statistically significant hazard ratio for my six different cancers was determined from Figure A–F and expressed as a forest plot.

2.3.4. Evaluation of *ALK* alteration in associated cancers

Data on genetic alterations in the *ALK* gene and their relevance and expression in different cancers were obtained from the cBioPortal server [37,38]. Based on the results, a total of 247 alterations were identified from one to 1620 amino acid sequences (Figure 4A). The highest alteration type was missense mutations (214 mutations) followed by 24 truncating mutations (including nonsense, frameshift deletion, frameshift insertion, and splicing), seven gene fusions, and two in-frame mutations. Alteration frequency was the highest in melanoma (14% in 444 cases) and the lowest in AML (1% in 200 cases) (Figure 4B). No alteration was found in DLBC. As depicted in Figure 4C, profiling of mutated RNA expression shows that missense mutations are omnipresent in each cancer type (highest in melanoma), seven oncofusions, nine truncating mutations, three splicings, and one in-frame mutation. In contrast, the frequency of copy number gain was the highest, which was distributed in each of the cancers, followed by copy number amplification, deep deletion, shallow deletion, and diploid. Consequently, these results suggest that the overexpression or downregulation of *ALK* in LUAD, skin melanoma, ovarian cancer or AML, DLBC, and BC, respectively, is correlated with mutations and copy number alterations.

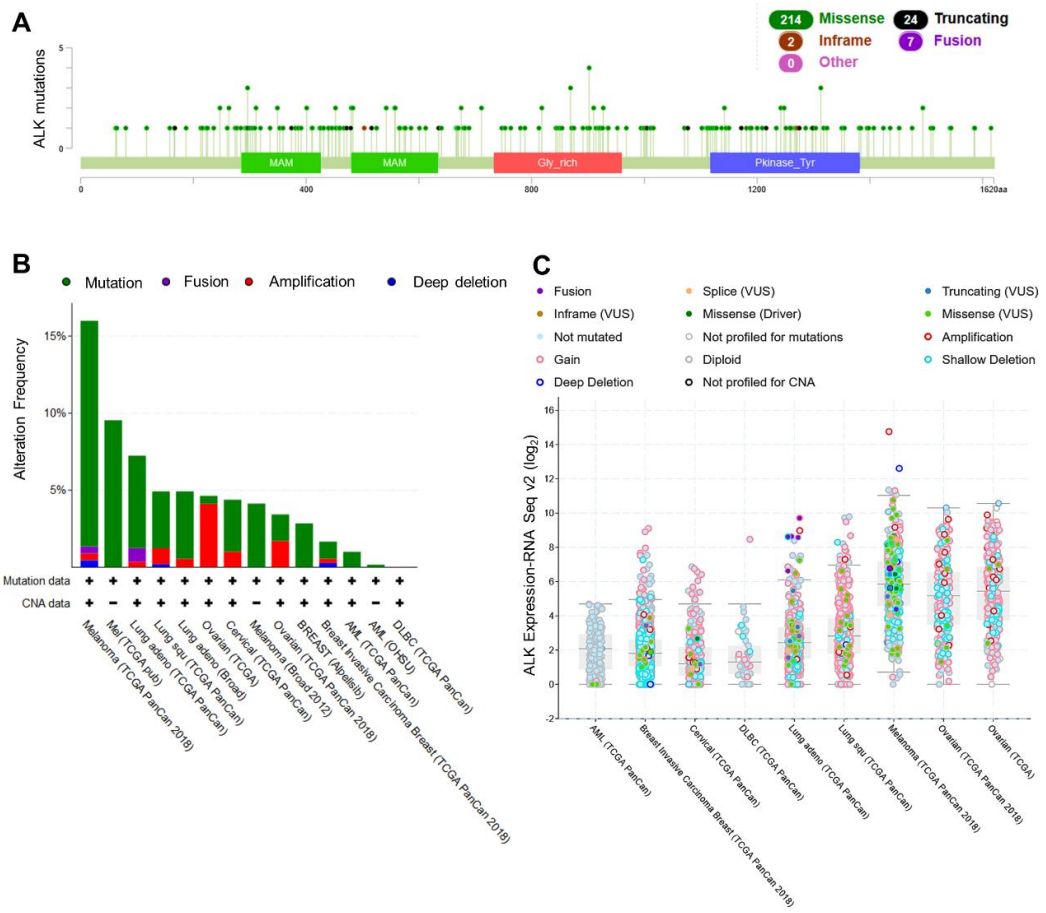


Figure 4. Data on genetic alteration of anaplastic lymphoma kinase (ALK) and consequent cancers in patients derived from The Cancer Genome Atlas (TCGA) PanCancer Atlas database through cBioPortal. A, Schematic of ALK protein and its associated functional domains with the physical location of mutations. A total of 247 alterations were found from one to 1620 amino acid sequences, where the highest alteration type was a missense mutation (214) and the lowest was an in-frame mutation (2). **B**, The alteration frequencies of ALK and their associated cancers. The Y-axis shows the alteration frequency percentage, and the X-axis represents the type of cancers generated from the alterations, including mutation and copy number alterations (CAN). The highest alteration frequency was conveyed in melanoma (gene-altered 14% in 444 cases). **C**, RNA seq v2 results of genetically altered *ALK* mRNA expression in 14 cancer studies obtained from the cBioPortal web. The mutations included missense mutations in all cancer types (green dots), seven oncofusion in LUAD and melanoma (violet dots), nine truncating (gray dots), three slicings (light brown dots), and one in-frame mutation (deep brown dot). The frequency of copy number gain was the highest and distributed in all cancer types, followed by copy number amplification, deep deletion, shallow deletion, and diploid as indicated in different colors.

2.3.5. ALK protein network analysis and clinical significance in cancers

To understand the significance of the diseases and predict the association between genotype and phenotype, GeneMANIA and STRING servers through Cytoscape_v3.8.2 software [32] were used to construct a protein network. Based on these results, the predicted 20 protein partners from the GeneMANIA server were EPHB3, MDK, AMELY, PTPRB, PTPRZ1, SEC16B, EPHA1, PTPRG, SEC16A, JAK3, MMP13, FBXO24, PTN, ZC3HC1, PDLIM3, MYBPC2, PTPRJ, KRT74, MTIF2, and SHC3 ([1] see the footnote for abbreviation) by considering physical interactions, co-expression, co-localization, genetic interactions, and shared protein domains (Figure 5A). Ten proteins were identified in the STRING server, including EML4, FRS3, PIK3CA, TNFRSF8, JAK3, PLCG1, KRAS, NPM1, HRAS, and JAK2 ([2] see the footnote) (Figure 5B). Next, the genetic alteration frequency of ALK and the 30 interacting protein partners were analyzed using cBioPortal in the six cancer types. The highest alteration frequency was observed in skin melanoma (70%), followed by LUAD (65%), IBC (58%), and OC (57%), and the lowest was in AML (36%) (Figure 5C). I found that these alterations in 17 mutual (Table 1) proteins decreased the patient's disease-free survival compared to the unaltered group in all the six cancer types (Figure 5D). These findings indicate that the identified ALK and its protein partners might be associated with LUAD, SM, OC, DLBC, AML, and BC. Furthermore, genetic alterations in these proteins can change the clinical outcomes.

[1] EPHB3 (Ephrin type-B receptor 3), MDK (Midkine), AMELY (Amelogenin Y-Linked), PTPRB (Receptor-type tyrosine-protein phosphatase beta), PTPRZ1 (Receptor-type tyrosine-protein phosphatase zeta), SEC16B (SEC16 Homolog B, Endoplasmic Reticulum Export Factor), EPHA1 (Ephrin Type-A Receptor 1), PTPRG (Receptor-Type Tyrosine-Protein Phosphatase Gamma), SEC16A (SEC16 Homolog A, Endoplasmic Reticulum Export Factor), JAK3 (Janus Kinase 3), MMP13 (Matrix Metallopeptidase 13), FBXO24 (F-Box Protein 24), PTN (Pleiotrophin), ZC3HC1 (Zinc Finger C3HC-Type Containing 1), PDLIM3 (PDZ And LIM Domain 3), MYBPC2 (Myosin Binding Protein C2), PTPRJ (Protein Tyrosine Phosphatase Receptor Type J), KRT74 (Keratin 74), MTIF2 (Mitochondrial Translational Initiation Factor 2), and SHC3 (SHC Adaptor Protein 3).

[2] EML4 (Echinoderm Microtubule-Associated Protein-Like 4), FRS3 (Fibroblast Growth Factor Receptor Substrate 3), PIK3CA (Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha), TNFRSF8 (Tumor Necrosis Factor Receptor Superfamily Member 8), JAK3 (Janus Kinase 3), PLCG1 (Phospholipase C Gamma 1), KRAS (Kirsten Rat Sarcoma Viral Oncogene Homolog), NPM1 (Nucleophosmin 1), HRAS (Harvey Rat Sarcoma Viral Oncogene Homolog), and JAK2 (Janus Kinase 2).

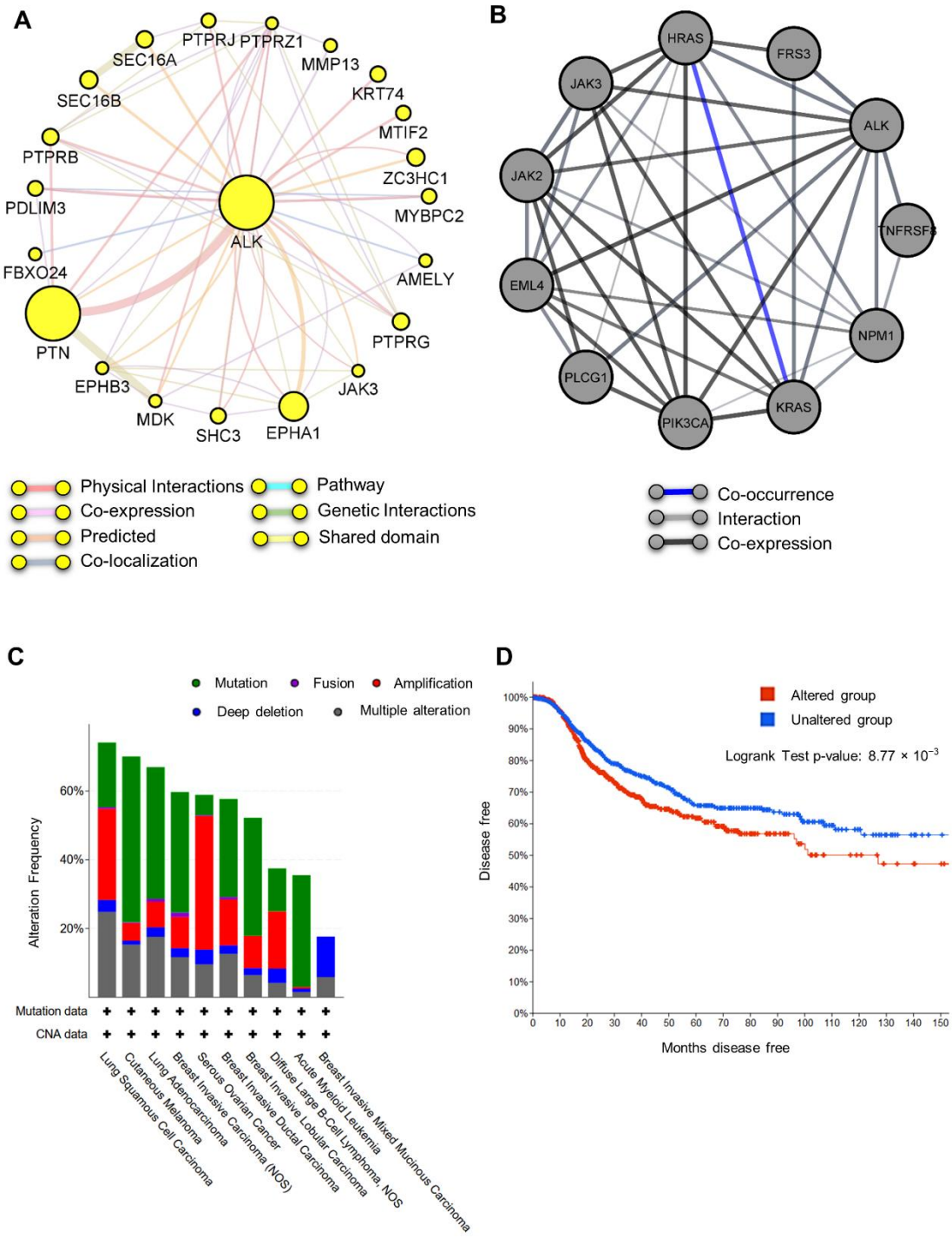


Figure 5. Protein network of anaplastic lymphoma kinase (ALK) and their clinical significance in six cancers. A–**B**, Predicted functional protein partners of ALK generated by considering physical interaction, co-expression, co-localization, genetic interactions, shared protein domains from GeneMANIA and STRING web servers through Cytoscape_v3.8.2 software. **C**, The genetic alteration frequency of 30 gene signatures in Figure A-B was generated for lung, melanoma, ovarian, DLBC, AML, and breast cancers using the cBioPortal platform. The highest alteration percentage was

reported in lung cancer followed by melanoma and breast cancer, while the lowest was in AML. **D**, Kaplan–Meier patient survival estimation of patients with genetically altered and unaltered ALK and 17 partner proteins (Table 1) in six types of cancers generated from cBioPortal. Patients with the altered forms of *ALK* and the 17 correlated genes showed significantly less survival than those with the unaltered forms.

2.3.6. Profiling of genes and micro RNAs co-occurred and co-expressed with ALK

I recognized the mutually exclusive co-occurrence genes from the 30 ALK protein partners identified using cBioPortal [37,38]. I found 17 genes that exclusively co-occur with *ALK* based on *q*-values (<0.05), which are listed in Table 1. I found that *EML4* was the most mutually exclusive gene (odds ratio > 3, *q*-value < 0.001) among all genes, whereas ZC3HC1 had the lowest exclusivity (odds ratio 1.33, q-value 0.036). I also observed that genetic alterations were the highest in the *EPHB3* gene (11%).

Table 1. Significant co-occurring protein partners of *ALK* gene signature were obtained using cBioPortal for cancer genomics.

	Protein partner	Alteration (%)	Log_2 Odds Ratio	*q*-value	Tendency
1	EML4	2.1	>3	<0.001	Co-occurrence
2	KRT74	1.8	2.718	<0.001	
3	SEC16A	3	2.611	<0.001	
4	TNFRSF8	2.5	2.441	<0.001	
5	MTIF2	2.1	2.357	<0.001	
6	PTPRJ	2.6	2.312	<0.001	
7	PTPRB	9	2.185	<0.001	
8	PTPRG	4	2.173	<0.001	
9	PLCG1	2.6	2.111	<0.001	
10	EPHA1	4	1.848	<0.001	
11	JAK2	4	1.812	<0.001	
12	SHC3	1.8	1.722	0.02	
13	FBXO24	3	1.674	0.002	
14	JAK3	4	1.655	0.002	
15	MYBPC2	2.8	1.599	0.008	
16	EPHB3	11	1.413	<0.001	
17	ZC3HC1	2.6	1.336	0.036	

To observe co-expression, I collected the fold change values of *ALK* and identified 17 genes from the microarray datasets of the Oncomine server [33]. The

expression analysis showed that all the genes are either overexpressed or downregulated in LUAD, CM, OSC, DLBC, AML, and BC cancers, as depicted in a heatmap (Figure 6). Noticeably, MTIF2 and EPHB3 are overexpressed in these six cancers, whereas JAK2 and SHC3 are downregulated in all cancers except DLBC and AML, respectively. The results indicated that *ALK* and the identified genes are associated with these cancers either positively or negatively. Furthermore, to determine the impact of each gene expression on the expression of another gene, I used a similarity matrix analysis of the genes with their co-expression values using Morpheus [39]. As shown in Figure 7, the expression of one gene affects that of another gene. This suggests that the identified co-expressed genes of *ALK* might be associated with the progression of the six cancer types included in my study.

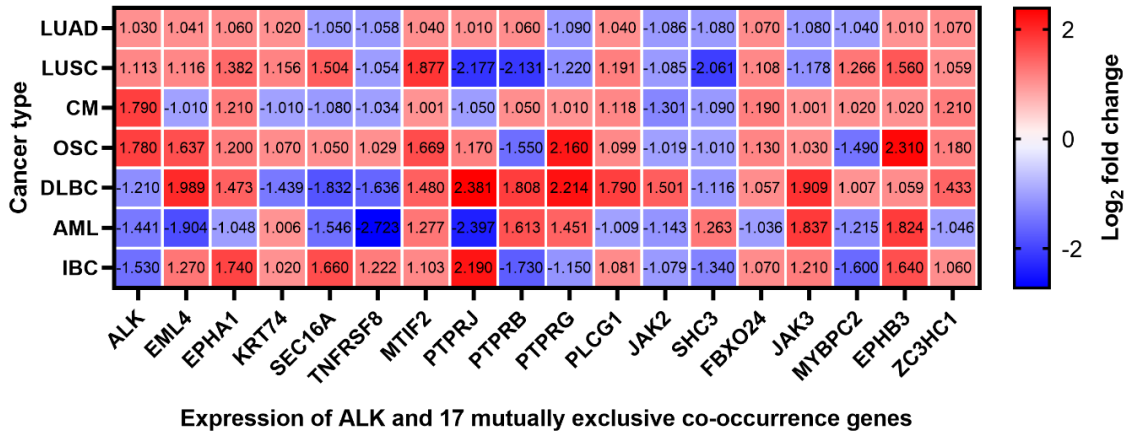


Figure 6. Co-expression analysis of anaplastic lymphoma kinase (ALK) and its 17 mutually associated functional protein partner genes in six cancer types. The heat map was generated from the $\log_2$ fold change expression value of *ALK* and mutually exclusive protein partner genes retrieved from the cancer microarray database (Oncomine).

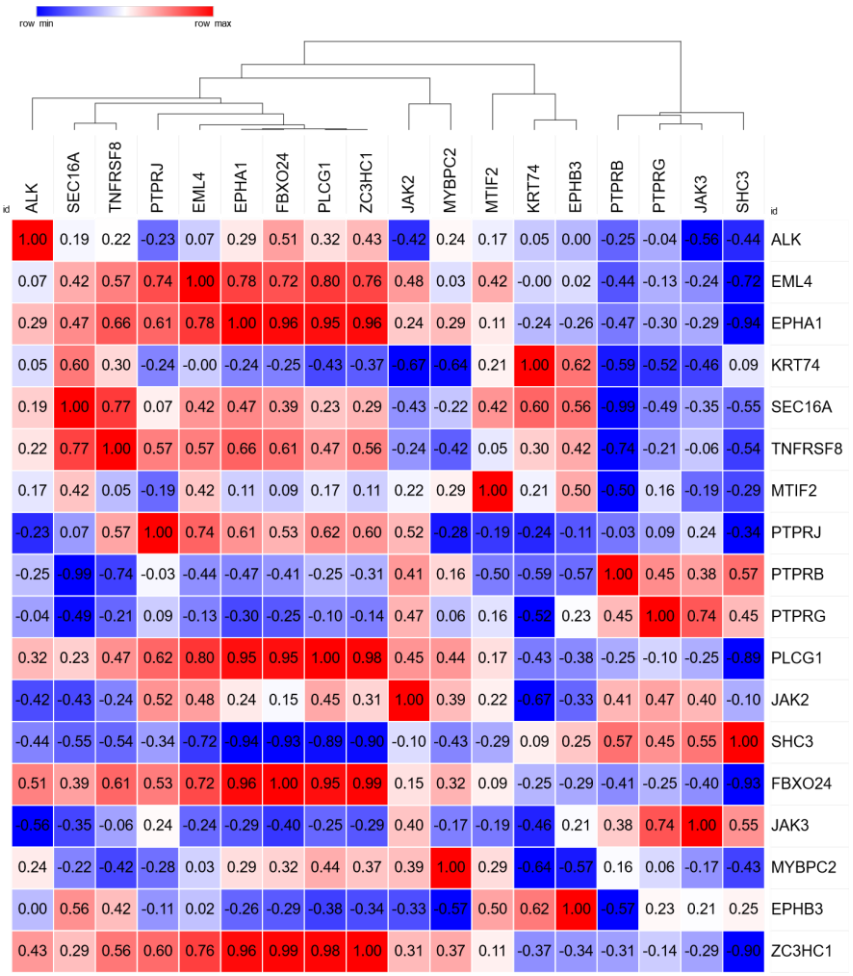


Figure 7. Hierarchical clustering and similarity matrix analysis of *ALK* and the expression of correlated genes in six cancer types. The heat map was generated from the $\log_2$ fold change expression value of *ALK* and common genes retrieved from Oncomine. The outcome of the correlation was visualized by the similarity matrix and hierarchical clustering tools in the Morpheus server.

Further, I investigated whether the microRNAs (miRNAs) were co-expressed with *ALK* and 17 common genes using the Enrichr server [40,41] based on *p*-value ranking (<0.05). I identified 19 out of 382 substantially co-expressed miRNAs in humans, as shown in Figure 8 (green node) and Supplementary Figure S3. These

findings indicated that the identified miRNAs might contribute to LUAD, SM, OC, DBLC, AML, and BC development along with *ALK* and common genes.

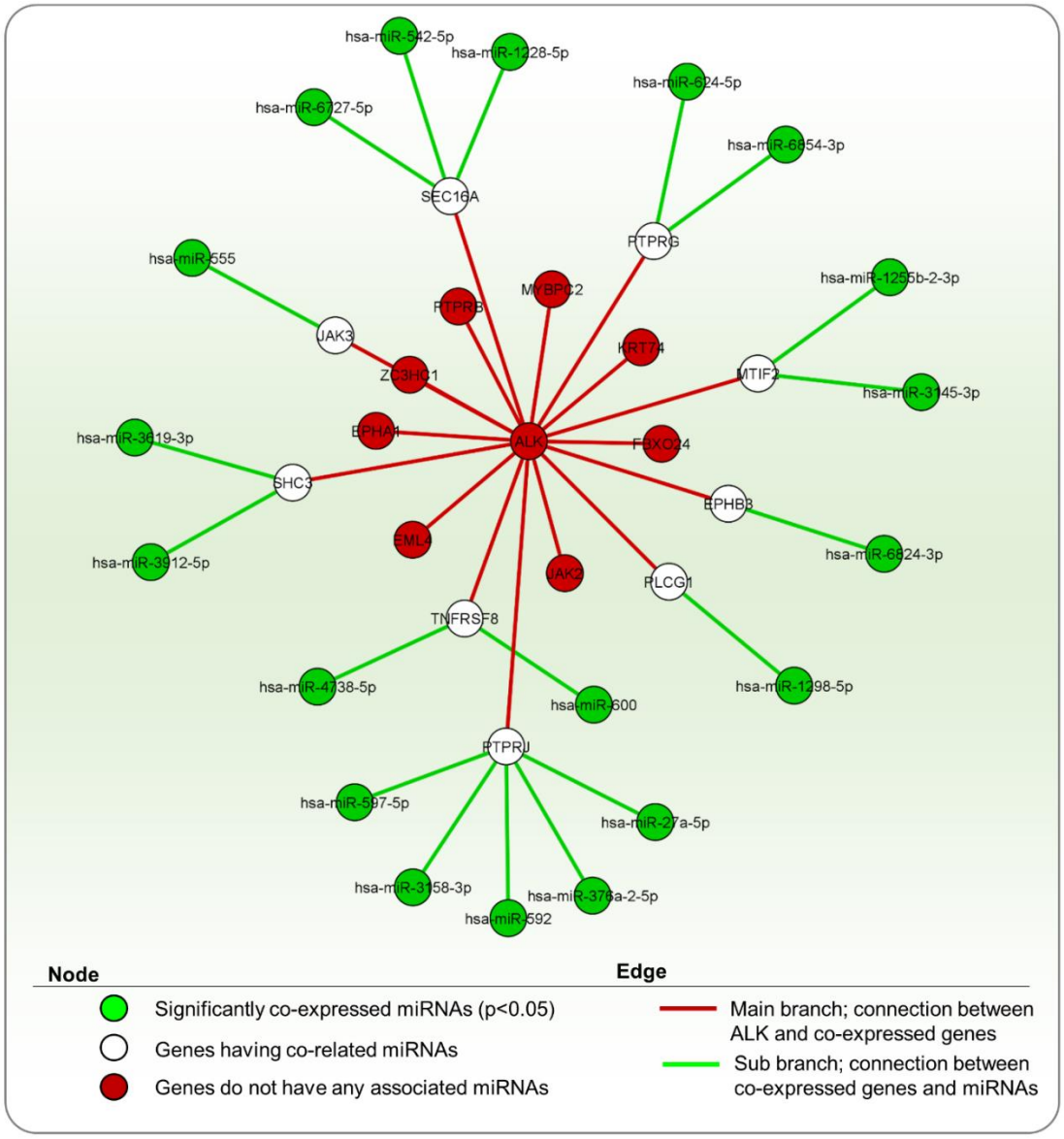


Figure 8. MicroRNAs (miRNAs) co-expressed with *ALK* and its mutually exclusive genes in humans. The miRNAs were identified from the Enrichr web portal and the interactions with the common genes were analyzed using Cytoscape_v3.8.2 software. The green nodes indicate significantly enriched terms ($p<0.05$). The white and red nodes denote genes with or without miRNA co-expression, respectively. The complete list of miRNAs with associated genes is shown in Supplementary Figure S3.

2.3.7. Gene ontologies and signaling pathway elucidation of *ALK* and its correlated genes

Based on the expression profile of *ALK* and its mutual genes, I determined gene ontological features and enrichment pathways using the Enrichr server [40,41] for integrative analysis of the progression of LUAD, SM, OC, DLBC, AML, and BC in humans. The resulting gene ontology (GO) terms mainly included the transmembrane receptor protein tyrosine kinase signaling pathway, protein autophosphorylation, negative regulation of cell communication (Figure 9A), protein kinase binding, protein tyrosine kinase activity, MAPK activity, growth factor receptor binding (Figure 9B), an integral component of the plasma membrane, specific granule membrane, and endosome lumen activity (Figure 9C). For pathway identification, I considered the results from three datasets, as shown in 9D-F. For the Reactome 2016 datasets, axon guidance, ERK activation, VEGFR2 mediated cell proliferation, and FCERI-mediated MAPK activation were considered (Figure 9D). Analysis of data from PANTHER 2016 showed the involvement of PDGF signaling, JAK/STAT signaling, angiogenesis, EGFR signaling, chemokine and cytokine signaling, histamine H1 receptor and interferon-gamma signaling, and axon guidance pathways (Figure 9E). Finally, KEGG 2019 suggested pathways related to non-small cell lung cancer, Th1, Th2, and T17 cell differentiation, pathways in cancer, axon guidance, chemokine and prolactin signaling, adherens junction, glioma, and ErbB signaling (Figure 9F). Therefore, these pathways might be involved in LUAD, SM, OC, DLBC, AML, and BC tumorigenesis.

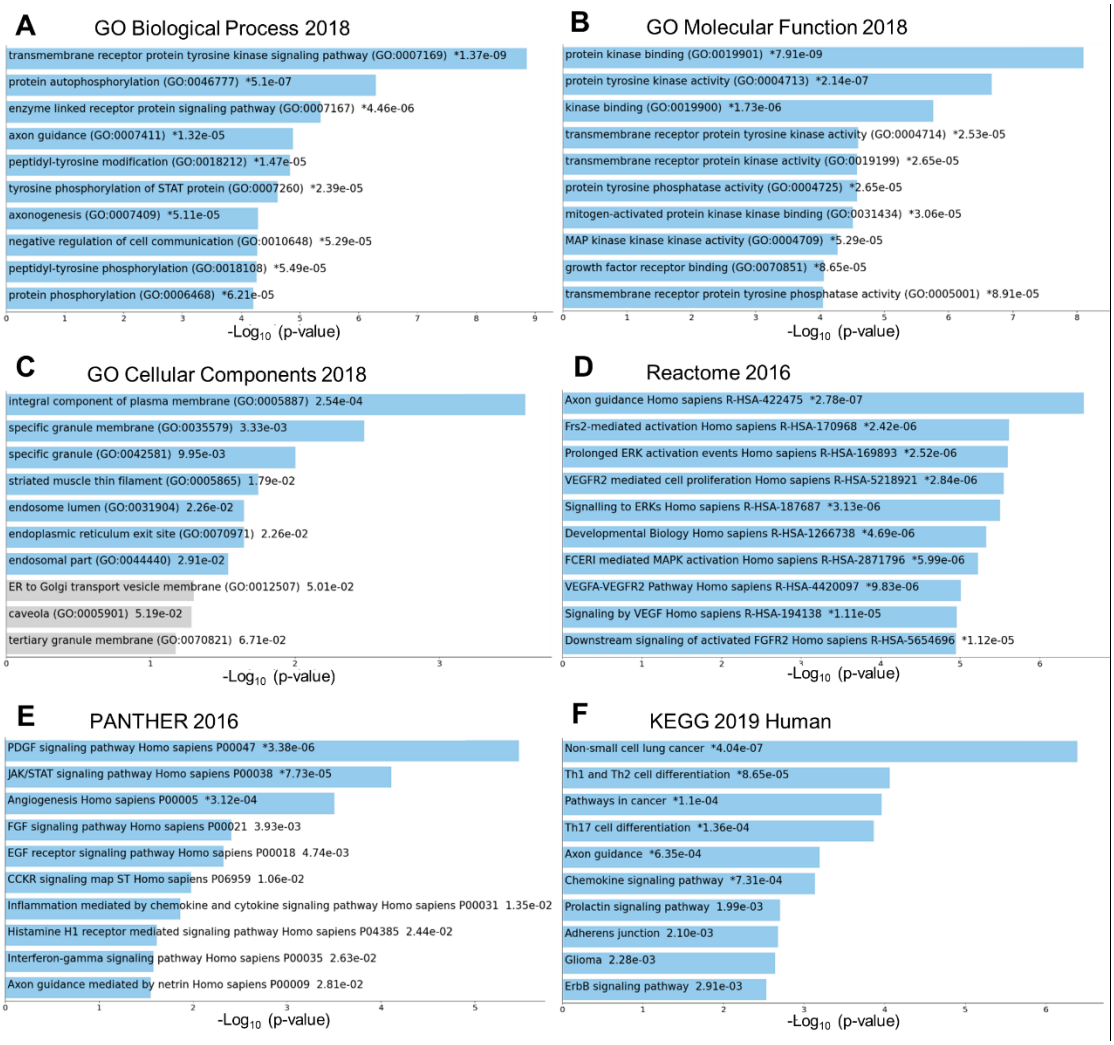


Figure 9. Gene ontology and pathway analyses of anaplastic lymphoma kinase (ALK) and its co-related functional protein partners. A, Biological process; **B**, Molecular function; **C**, Cellular components; **D**, Reactome; **E**, PANTHER; **F**, KEGG pathways of ALK and 18 partner proteins were obtained from the Enrichr web tool. The bar chart shows the top 10 enriched terms in the chosen library, along with their corresponding *p*-values. Colored bars correspond to terms with significant *p*-values (<0.05). * indicates statistically significant values with $p < 0.05$.

2.4. Discussion

In this study, I used the Open Targets Genetics portal and observed that *ALK* is highly expressed in lung, skin, and ovary normal tissues compared to blood and breast tissues (Figure 1A). Using FireBrowse, Oncomine, and HPA, I showed that the *ALK* expression at the mRNA and protein levels had similar expression patterns in their respective cancer tissues, such as upregulation in LUAD, melanoma, OC, and downregulation in DLBC, AML, and BC (Figure 1B and Figure 2). This indicates an association between *ALK* expression and the progression of these six cancers. This finding was consistent with the biological fact that *ALK* expression has shown upregulated in lung cancer [2,42], melanoma [43], OC [7,9], and downregulated in DLBC [44], AML [45], and BC [45] tissue samples.

Prediction of a disease phenotype based on specific gene expression substantially contributes to patient care with different treatment strategies. Therefore, it is necessary to have a strong prognostic index that facilitates the treatment of LUAD, melanoma, OC, DLBC, AML, and BC. Based on the prognostic analysis results, I found that the overexpression of *ALK* positively correlates with the progression of LUAD, melanoma, and OC with higher hazard risk (HR>1), whereas the downregulation of *ALK* is negatively correlates with the survival rate of patients with DLBC, AML, and BC with relatively lower hazard risk (Figure 3). Previous studies on LUAD and melanoma showed that *EML4-ALK* rearrangement is associated with a poorer disease-free survival rate than patients with *ALK*-negative cancers [46-48]. However, patients with *ALK* rearrangement had a better prognosis for ovarian cancer than those without *ALK* rearrangement [49]. In line with my study, a previous study in BC revealed that patients with ALK protein overexpression exhibited poor overall survival or relapse-free survival compared to those with low *ALK* expression [50] (Figure S1). All these studies except BC stated the patient survival rate based on *ALK*-positive or *ALK*-negative samples, whereas my study showed prognosis based on *ALK* expression level. So far, I have not found prognostic studies on DLBC and AML based on ALK expression patterns. It is worthwhile to conduct further studies on overexpression or downregulation of *ALK* for long-term follow-up and detailed molecular analysis to identify *ALK* as a prognostic biomarker for patients with LUAD, melanoma, OC, DLBC, AML, and BC.

This divergent *ALK* expression in the six cancers might be attributed to the many mutations and copy number alterations observed in this study (Figure 4). Cancer development is a chain of histopathological procedures that are influenced by four important factors, such as somatically acquired genetic, transcriptomic, proteomic, and epigenetic changes [51]. Among my queried patients (4865), *ALK* was altered in 5% of patients (242) with a somatic mutation frequency of 4.3%. In total, missense, truncating, fusion, and in-frame mutations, splicing, copy number gain, amplification, deep deletion, shallow deletion, and diploid were identified (Figure 4). Several studies have also shown genetic alterations, including mutations and fusions in *ALK* [2,52-58]. In contrast to my findings, Cheng *et al*. found three recurrent hotspots (R1275Q/*/L, F1245C/I/L/V, and F1174L/V) in the tyrosine kinase domain and one (R395H/C) in the MAM domain in a population-scale cohort of different tumor samples [59]. However, I did not identify any cancer hotspots in the queried TCGA samples.

Protein-protein interactions have been used for the analysis of biological processes, disease progression, and downstream signaling and have also been targeted for drug development [60,61]. In the present study, I identified 30 proteins (10 from STRING and 20 from GeneMANIA) using Cytoscape software. These 30 protein signatures showed an alteration frequency of up to 65% in melanoma and lung cancers. PIK3CA had the highest alteration frequency (22%), and HRAS is the least frequent protein (1.8%) (Figure S2). Similarly, Millis *et al*. revealed that PIK3CA has a mutational frequency of 13% in solid tumors [62]. Another study showed that this gene is amplified by 33-42% in lung cancer and 15-27% in ovarian cancer [63]. Hobbs *et al.* also reported that HRAS is the least frequent (4%) protein compared to other isoforms such as KRAS (85%) in human cancers [64]. These observations suggest the involvement of PIK3CA, HRAS, and other partners in various cancers.

Furthermore, I identified 17 genes exclusively co-occurring with 30 genes using cBioPortal datasets (Table 1). Based on my analysis, I found that *EML4* was the most correlated gene. Soda *et al.* (2007) first reported the fusion of *EML4* with *ALK*, an oncogenic driver leading to NSCLC, which was later identified in colorectal cancer and BC [2,14]. Other genes, such as *EPHB3*, *PTPRJ*, *PTPRB*, and *PLCG1* in lung cancer [65-68], *JAK3* in melanoma [69], *EPHA1* in OC [70], *TNFRSF8* (*CD30*) in DLBC [71],

and *PLCG1*, *PTPRG* in BC [68,72], have been reported. However, I did not find any relationship for 6 out of 17 genes, including *KRT74*, *MTIF2*, *SHC3*, *FBXO24*, *MYBPC2*, and *ZC3HC1* in my six cancers. However, their co-expression with *ALK* was confirmed in my co-expression analysis (Figures 6 and 7). Moreover, I found that *ALK* is upregulated in LUAD, CM, and OSC and downregulated in DLBC, AML, and IBC. This observation is consistent with that of previous studies [3,8,43-45]. Among the identified miRNAs, hsa-miR-1228, hsa-miR-600, and hsa-miR-542-5p were associated with lung cancer [73-75], hsa-miR-3158 and hsa-miR-3145 were found in melanoma [76], and hsa-miR-3619-3p, hsa-miR-4738, and hsa-miR-1298-5p were found in BC [77-79]. Some miRNAs were reported in different cancers or functions rather than the cancer type stated here. For example, hsa-miR-27a and hsa-miR-6727-5p were reported in cervical cancer [80,81], hsa-miR-597-5p in cutaneous squamous cell carcinoma [82], hsa-miR-592 in colorectal cancer [83], hsa-miR-624-5p in osteosarcoma [84], and hsa-miR-555 exhibited antiviral properties against poliovirus [85]. To date, five other miRNAs, hsa-miR-376a-2-5p, hsa-miR-6824-3p, hsa-miR-1255b-2-3p, hsa-miR-3912-5p, and hsa-miR-6854-3p, have not been identified in the queried cancers.

I also evaluated the GO enrichment pathways of biological processes, molecular functions, and cellular components for the functional analysis of *ALK* and its correlated genes. Importantly, the receptor tyrosine kinase (RTK) signaling activity was the most relevant in all three GO enrichment pathways (Figure 9 A-C). RTKs, such as vascular endothelial growth factor receptors (VEGFRs), fibroblast growth factor receptors (FGFRs), platelet-derived growth factor receptors (PDGFRs), or ALK, arbitrate the major signaling pathways that are convoluted in cell proliferation, differentiation, migration, and cell growth, and their dysregulation is a crucial feature of human cancers, such as lung and breast cancers [86,87]. ALK is activated via dimerization resulting from autophosphorylation and tyrosine activation of the four substrates of ALK, including mitogen-activated protein kinases (MAPK), phosphatidylinositol 3-kinase (PI3K), phospholipase C-γ (PLC γ), and Janus kinase (JAK) [18,19,88]. I also identified similar signaling pathways in the three signaling pathways databases: Reactome, PANTHER, and KEGG (Figure 9D-F). These pathways are not only associated with NSCLC and BC but also trigger the progression of melanoma, ovarian cancer, DLBC, and AML tumorigenesis [89-93]. Further studies are required based on gene silencing

and overexpression to identify concrete pathways involved in the progression of ALK-mediated melanoma, OC, DLBC, and AML.

In summary, this study shows the possible relationship between *ALK* expression and the prognosis of patients with LUAD, melanoma, OC, DLBC, AML, and BC using publicly available multiplatform bioinformatics datasets. My data showed that the genotypic expression of *ALK* is substantially correlated with the clinical phenotype in my six queried cancers. My study may help clinicians with precision clinical care and researchers for developing cancer therapy for *ALK*-positive cancers. Additional therapy for *ALK* in patients with *ALK*-positive melanoma, OC, DLBC, AML, and BC might increase the survival rate. However, further experimental studies are required to confirm that *ALK* is a prognostic biomarker.

2.5. Concluding remark

In this study, to determine *ALK* as a potential prognostic biomarker in *ALK*-positive LUAD, melanoma, OC, DLBC, AML, and BC progression, I assessed *ALK* expression, genetic mutations, protein-protein interaction networks, correlated genes, and miRNAs, and the prognostic values using various publicly available bioinformatics datasets. I conclude that both upregulation of *ALK* in LUAD, melanoma, OV or downregulation in DLBC, AML, and BC progression negatively correlated with patient survival. Furthermore, I revealed the possible gene ontology and signaling pathways associated with *ALK* and its expression in the progression of these six cancers. These pathways may serve as potential targets to inhibit the development of cancers. Collectively, *ALK* could be an effective prognostic biomarker and a potential therapeutic target to control *ALK*-positive cancers. Further studies are required to validate my outcomes through *in vitro* experiments.

2.6. Bibliography

1. Golding, B.; Luu, A.; Jones, R.; Viloria-Petit, A.M. The function and therapeutic targeting of anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC). *Molecular Cancer* **2018**, *17*, 52, doi:10.1186/s12943-018-0810-4.
2. Soda, M.; Choi, Y.L.; Enomoto, M.; Takada, S.; Yamashita, Y.; Ishikawa, S.; Fujiwara, S.-i.; Watanabe, H.; Kurashina, K.; Hatanaka, H., et al. Identification

of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **2007**, *448*, 561-566, doi:10.1038/nature05945.

3. Saifullah; Sakari, M.; Suzuki, T.; Yano, S.; Tsukahara, T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. *International Journal of Molecular Sciences* **2020**, *21*, 8904.
4. Morris, S.; Kirstein, M.; Valentine, M.; Dittmer, K.; Shapiro, D.; Saltman, D.; Look, A. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **1994**, *263*, 1281-1284, doi:10.1126/science.8122112.
5. Wiesner, T.; Lee, W.; Obenauf, A.C.; Ran, L.; Murali, R.; Zhang, Q.F.; Wong, E.W.P.; Hu, W.; Scott, S.N.; Shah, R.H., et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **2015**, *526*, 453-457, doi:10.1038/nature15258.
6. Cesi, G.; Philippidou, D.; Kozar, I.; Kim, Y.J.; Bernardin, F.; Van Niel, G.; Wienecke-Baldacchino, A.; Felten, P.; Letellier, E.; Dengler, S., et al. A new ALK isoform transported by extracellular vesicles confers drug resistance to melanoma cells. *Molecular Cancer* **2018**, *17*, 145, doi:10.1186/s12943-018-0886-x.
7. Ren, H.; Tan, Z.-P.; Zhu, X.; Crosby, K.; Haack, H.; Ren, J.-M.; Beausoleil, S.; Moritz, A.; Innocenti, G.; Rush, J., et al. Identification of Anaplastic Lymphoma Kinase as a Potential Therapeutic Target in Ovarian Cancer. *Cancer Research* **2012**, *72*, 3312-3323, doi:10.1158/0008-5472.can-11-3931.
8. Tang, S.; Yang, F.; Du, X.; Lu, Y.; Zhang, L.; Zhou, X. Aberrant Expression of Anaplastic Lymphoma Kinase in Ovarian Carcinoma Independent of Gene Rearrangement. *International Journal of Gynecological Pathology* **2016**, *35*, 337-347, doi:10.1097/pgp.0000000000000260.
9. Hui, B.; Zhang, J.; Shi, X.; Xing, F.; Shao, Y.W.; Wang, Y.; Zhang, X.; Wang, S. EML4-ALK, a potential therapeutic target that responds to alectinib in ovarian cancer. *Japanese Journal of Clinical Oncology* **2020**, *50*, 1470-1474, doi:10.1093/jjco/hyaa156.
10. Reichard, K.K.; McKenna, R.W.; Kroft, S.H. ALK-positive diffuse large B-cell lymphoma: report of four cases and review of the literature. *Modern Pathology* **2007**, *20*, 310-319, doi:10.1038/modpathol.3800742.
11. Morgan, E.A.; Nascimento, A.F. Anaplastic Lymphoma Kinase-Positive Large B-Cell Lymphoma: An Underrecognized Aggressive Lymphoma. *Advances in Hematology* **2012**, *2012*, 529572, doi:10.1155/2012/529572.
12. Maxson, J.E.; Davare, M.A.; Luty, S.B.; Eide, C.A.; Chang, B.H.; Loriaux, M.M.; Tognon, C.E.; Bottomly, D.; Wilmot, B.; McWeeney, S.K., et al. Therapeutically Targetable ALK Mutations in Leukemia. *Cancer Research* **2015**, *75*, 2146-2150, doi:10.1158/0008-5472.can-14-1576.
13. Maesako, Y.; Izumi, K.; Okamori, S.; Takeoka, K.; Kishimori, C.; Okumura, A.; Honjo, G.; Akasaka, T.; Ohno, H. inv(2)(p23q13)/RAN-binding protein 2 (RANBP2)–ALK fusion gene in myeloid leukemia that developed in an elderly woman. *International Journal of Hematology* **2014**, *99*, 202-207, doi:10.1007/s12185-013-1482-x.
14. Lin, E.; Li, L.; Guan, Y.; Soriano, R.; Rivers, C.S.; Mohan, S.; Pandita, A.; Tang, J.; Modrusan, Z. Exon Array Profiling Detects EML4-ALK Fusion in Breast, Colorectal, and Non–Small Cell Lung Cancers. *Molecular Cancer Research* **2009**, *7*, 1466-1476, doi:10.1158/1541-7786.mcr-08-0522.

15. Hanna, M.G.; Najfeld, V.; Irie, H.Y.; Tripodi, J.; Nayak, A. Analysis of ALK gene in 133 patients with breast cancer revealed polysomy of chromosome 2 and no ALK amplification. *SpringerPlus* **2015**, *4*, 439, doi:10.1186/s40064-015-1235-9.
16. Jeanneau, M.; Gregoire, V.; Desplechain, C.; Escande, F.; Tica, D.P.; Aubert, S.; Leroy, X. ALK rearrangements-associated renal cell carcinoma (RCC) with unique pathological features in an adult. *Pathology - Research and Practice* **2016**, *212*, 1064-1066, doi:https://doi.org/10.1016/j.prp.2016.07.015.
17. Aubry, A.; Galiacy, S.; Allouche, M. Targeting ALK in Cancer: Therapeutic Potential of Proapoptotic Peptides. *Cancers* **2019**, *11*, 275.
18. Della Corte, C.M.; Viscardi, G.; Di Liello, R.; Fasano, M.; Martinelli, E.; Troiani, T.; Ciardiello, F.; Morgillo, F. Role and targeting of anaplastic lymphoma kinase in cancer. *Molecular Cancer* **2018**, *17*, 30, doi:10.1186/s12943-018-0776-2.
19. Hrustanovic, G.; Olivas, V.; Pazarentzos, E.; Tulpule, A.; Asthana, S.; Blakely, C.M.; Okimoto, R.A.; Lin, L.; Neel, D.S.; Sabnis, A., et al. RAS-MAPK dependence underlies a rational polytherapy strategy in EML4-ALK–positive lung cancer. *Nature Medicine* **2015**, *21*, 1038-1047, doi:10.1038/nm.3930.
20. Chiarle, R.; Voena, C.; Ambrogio, C.; Piva, R.; Inghirami, G. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature Reviews Cancer* **2008**, *8*, 11-23, doi:10.1038/nrc2291.
21. Gouzi, J.Y.; Moog-Lutz, C.; Vigny, M.; Brunet-de Carvalho, N. Role of the subcellular localization of ALK tyrosine kinase domain in neuronal differentiation of PC12 cells. *Journal of Cell Science* **2005**, *118*, 5811-5823, doi:10.1242/jcs.02695.
22. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* **2020**, *14*.
23. Rafi, J.H.; Jafar, T.; Pathan, M.T.; Reza, R.; Islam, S.; Sourna, I.J.; Alam, R.; Samad, A.; Ahammad, F. High expression of bone morphogenetic protein 1 (BMP1) is associated with a poor survival rate in human gastric cancer, a dataset approaches. *Genomics* **2021**, *113*, 1141-1154, doi:https://doi.org/10.1016/j.ygeno.2020.11.012.
24. Samad, A.; Jafar, T.; Rafi, J.H. Identification of angiotensin-converting enzyme 2 (ACE2) protein as the potential biomarker in SARS-CoV-2 infection-related lung cancer using computational analyses. *Genomics* **2020**, *112*, 4912-4923, doi:https://doi.org/10.1016/j.ygeno.2020.09.002.
25. Yan, J.; Risacher, S.L.; Shen, L.; Saykin, A.J. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Briefings in Bioinformatics* **2017**, *19*, 1370-1381, doi:10.1093/bib/bbx066.
26. Hasin, Y.; Seldin, M.; Lusis, A. Multi-omics approaches to disease. *Genome Biology* **2017**, *18*, 83, doi:10.1186/s13059-017-1215-1.
27. Hagerty, R.G.; Butow, P.N.; Ellis, P.M.; Dimitry, S.; Tattersall, M.H.N. Communicating prognosis in cancer care: a systematic review of the literature. *Annals of Oncology* **2005**, *16*, 1005-1053, doi:https://doi.org/10.1093/annonc/mdi211.
28. Rabin, B.A.; Gaglio, B.; Sanders, T.; Nekhlyudov, L.; Dearing, J.W.; Bull, S.; Glasgow, R.E.; Marcus, A. Predicting Cancer Prognosis Using Interactive Online Tools: A Systematic Review and Implications for Cancer Care Providers.

Cancer Epidemiology Biomarkers & Prevention **2013**, *22*, 1645-1656, doi:10.1158/1055-9965.epi-13-0513.

29. Huang, S.; Yee, C.; Ching, T.; Yu, H.; Garmire, L.X. A Novel Model to Combine Clinical and Pathway-Based Transcriptomic Information for the Prognosis Prediction of Breast Cancer. *PLOS Computational Biology* **2014**, *10*, e1003851, doi:10.1371/journal.pcbi.1003851.
30. Huang, S.; Chong, N.; Lewis, N.E.; Jia, W.; Xie, G.; Garmire, L.X. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome Medicine* **2016**, *8*, 34, doi:10.1186/s13073-016-0289-9.
31. Ghoussaini, M.; Mountjoy, E.; Carmona, M.; Peat, G.; Schmidt, E.M.; Hercules, A.; Fumis, L.; Miranda, A.; Carvalho-Silva, D.; Buniello, A., et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic acids research* **2021**, *49*, D1311-d1320, doi:10.1093/nar/gkaa840.
32. FireBrowse. Broad Institute TCGA Genome Data Analysis Center (2016): Firehose stddata__2016_01_28 run. Broad Institute of MIT and Harvard. doi:10.7908/C11G0KM9. **2016**, 2498-2504.
33. Rhodes, D.R.; Kalyana-Sundaram, S.; Mahavisno, V.; Varambally, R.; Yu, J.; Briggs, B.B.; Barrette, T.R.; Anstet, M.J.; Kincead-Beal, C.; Kulkarni, P., et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia (New York, N.Y.)* **2007**, *9*, 166-180, doi:10.1593/neo.07112.
34. Uhlen, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhori, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F., et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507, doi:10.1126/science.aan2507.
35. Mizuno, H.; Kitada, K.; Nakai, K.; Sarai, A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC medical genomics* **2009**, *2*, 18, doi:10.1186/1755-8794-2-18.
36. Goto, Y.; Zeng, L.; Yeom, C.J.; Zhu, Y.; Morinibu, A.; Shinomiya, K.; Kobayashi, M.; Hirota, K.; Itasaka, S.; Yoshimura, M., et al. UCHL1 provides diagnostic and antimetastatic strategies due to its deubiquitinating effect on HIF-1α. *Nature communications* **2015**, *6*, 6153, doi:10.1038/ncomms7153.
37. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E., et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling* **2013**, *6*, pl1-pl1, doi:10.1126/scisignal.2004088.
38. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E., et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2012**, *2*, 401-404, doi:10.1158/2159-8290.cd-12-0095.
39. Morpheus. https://software.broadinstitute.org/morpheus.
40. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **2013**, *14*, 128, doi:10.1186/1471-2105-14-128.
41. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A., et al.

Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **2016**, *44*, W90-W97, doi:10.1093/nar/gkw377.

42. To, K.-F.; Tong, J.H.M.; Yeung, K.S.F.; Lung, R.W.M.; Law, P.P.Y.; Chau, S.L.; Kang, W.; Tong, C.Y.K.; Chow, C.; Chan, A.W.H., et al. Detection of ALK Rearrangement by Immunohistochemistry in Lung Adenocarcinoma and the Identification of a Novel EML4-ALK Variant. *Journal of Thoracic Oncology* **2013**, *8*, 883-891, doi:https://doi.org/10.1097/JTO.0b013e3182904e22.
43. Busam, K.J.; Vilain, R.E.; Lum, T.; Busam, J.A.; Hollmann, T.J.; Saw, R.P.M.; Coit, D.C.; Scolyer, R.A.; Wiesner, T. Primary and Metastatic Cutaneous Melanomas Express ALK Through Alternative Transcriptional Initiation. *Am J Surg Pathol* **2016**, *40*, 786-795, doi:10.1097/PAS.0000000000000611.
44. Brune , V.; Tiacci , E.; Pfeil , I.; Döring , C.; Eckerle , S.; van Noesel , C.J.M.; Klapper , W.; Falini , B.; von Heydebreck , A.; Metzler , D., et al. Origin and pathogenesis of nodular lymphocyte–predominant Hodgkin lymphoma as revealed by global gene expression analysis. *Journal of Experimental Medicine* **2008**, *205*, 2251-2268, doi:10.1084/jem.20080809.
45. Valk, P.J.M.; Verhaak, R.G.W.; Beijen, M.A.; Erpelinck, C.A.J.; van Doorn-Khosrovani, S.B.v.W.; Boer, J.M.; Beverloo, H.B.; Moorhouse, M.J.; van der Spek, P.J.; Löwenberg, B., et al. Prognostically Useful Gene-Expression Profiles in Acute Myeloid Leukemia. *New England Journal of Medicine* **2004**, *350*, 1617-1628, doi:10.1056/NEJMoa040465.
46. Shi, J.; Gu, W.; Zhao, Y.; Zhu, J.; Jiang, G.; Bao, M.; Shi, J. Clinicopathological and Prognostic Significance of EML4-ALK Rearrangement in Patients with Surgically Resected Lung Adenocarcinoma: A Propensity Score Matching Study. *Cancer Manag Res* **2020**, *12*, 589-598, doi:10.2147/CMAR.S229217.
47. Sheen, Y.-S.; Liao, Y.-H.; Lin, M.-H.; Chen, J.-S.; Liau, J.-Y.; Tseng, Y.-J.; Lee, C.-H.; Chang, Y.-L.; Chu, C.-Y. A clinicopathological analysis of 153 acral melanomas and the relevance of mechanical stress. *Scientific Reports* **2017**, *7*, 5564, doi:10.1038/s41598-017-05809-9.
48. Cao, S.; Nambudiri, V.E. Anaplastic Lymphoma Kinase in Cutaneous Malignancies. *Cancers* **2017**, *9*, 123.
49. Bi, R.; Bai, Q.; Zhu, X.; Tu, X.; Cai, X.; Jiang, W.; Xu, X.; Tang, S.; Ge, H.; Chang, B., et al. ALK rearrangement: a high-frequency alteration in ovarian metastasis from lung adenocarcinoma. *Diagnostic Pathology* **2019**, *14*, 96, doi:10.1186/s13000-019-0864-7.
50. Siraj, A.K.; Beg, S.; Jehan, Z.; Prabhakaran, S.; Ahmed, M.; R.Hussain, A.; Al-Dayel, F.; Tulbah, A.; Ajarim, D.; Al-Kuraya, K.S. ALK alteration is a frequent event in aggressive breast cancers. *Breast Cancer Research* **2015**, *17*, 127, doi:10.1186/s13058-015-0610-3.
51. Okayama, H.; Kohno, T.; Ishii, Y.; Shimada, Y.; Shiraishi, K.; Iwakawa, R.; Furuta, K.; Tsuta, K.; Shibata, T.; Yamamoto, S., et al. Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas. *Cancer Research* **2012**, *72*, 100-111, doi:10.1158/0008-5472.can-11-1403.
52. Katayama, R.; Friboulet, L.; Koike, S.; Lockerman, E.L.; Khan, T.M.; Gainor, J.F.; Iafrate, A.J.; Takeuchi, K.; Taiji, M.; Okuno, Y., et al. Two Novel ALK Mutations Mediate Acquired Resistance to the Next-Generation ALK Inhibitor Alectinib. *Clinical Cancer Research* **2014**, *20*, 5686-5696, doi:10.1158/1078-0432.ccr-14-1511.

53. Sasaki, T.; Koivunen, J.; Ogino, A.; Yanagita, M.; Nikiforow, S.; Zheng, W.; Lathan, C.; Marcoux, J.P.; Du, J.; Okuda, K., et al. A Novel ALK Secondary Mutation and EGFR Signaling Cause Resistance to ALK Kinase Inhibitors. *Cancer Research* **2011**, *71*, 6051-6060, doi:10.1158/0008-5472.can-11-1340.
54. Heuckmann, J.M.; Hölzel, M.; Sos, M.L.; Heynck, S.; Balke-Want, H.; Koker, M.; Peifer, M.; Weiss, J.; Lovly, C.M.; Grütter, C., et al. ALK Mutations Conferring Differential Resistance to Structurally Diverse ALK Inhibitors. *Clinical Cancer Research* **2011**, *17*, 7394-7401, doi:10.1158/1078-0432.ccr-11-1648.
55. Ai, X.; Niu, X.; Chang, L.; Chen, R.; Ou, S.-H.I.; Lu, S. Next generation sequencing reveals a novel ALK G1128A mutation resistant to crizotinib in an ALK-Rearranged NSCLC patient. *Lung Cancer* **2018**, *123*, 83-86, doi:https://doi.org/10.1016/j.lungcan.2018.07.004.
56. Bresler, Scott C.; Weiser, Daniel A.; Huwe, Peter J.; Park, Jin H.; Krytska, K.; Ryles, H.; Laudenslager, M.; Rappaport, Eric F.; Wood, Andrew C.; McGrady, Patrick W., et al. ALK Mutations Confer Differential Oncogenic Activation and Sensitivity to ALK Inhibition Therapy in Neuroblastoma. *Cancer Cell* **2014**, *26*, 682-694, doi:https://doi.org/10.1016/j.ccell.2014.09.019.
57. Li, Y.Y.; Hanna, G.J.; Laga, A.C.; Haddad, R.I.; Lorch, J.H.; Hammerman, P.S. Genomic Analysis of Metastatic Cutaneous Squamous Cell Carcinoma. *Clinical Cancer Research* **2015**, *21*, 1447-1456, doi:10.1158/1078-0432.ccr-14-1773.
58. Wellstein, A.; Toretsky, J.A. Hunting ALK to feed targeted cancer therapy. *Nature Medicine* **2011**, *17*, 290-291, doi:10.1038/nm0311-290.
59. Chang, M.T.; Asthana, S.; Gao, S.P.; Lee, B.H.; Chapman, J.S.; Kandoth, C.; Gao, J.; Socci, N.D.; Solit, D.B.; Olshen, A.B., et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature Biotechnology* **2016**, *34*, 155-163, doi:10.1038/nbt.3391.
60. Wells, J.A.; McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **2007**, *450*, 1001-1009, doi:10.1038/nature06526.
61. Laraia, L.; McKenzie, G.; Spring, David R.; Venkitaraman, Ashok R.; Huggins, David J. Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. *Chemistry & Biology* **2015**, *22*, 689-703, doi:10.1016/j.chembiol.2015.04.019.
62. Millis, S.Z.; Jardim, D.L.; Albacker, L.; Ross, J.S.; Miller, V.A.; Ali, S.M.; Kurzrock, R. Phosphatidylinositol 3-kinase pathway genomic alterations in 60,991 diverse solid tumors informs targeted therapy opportunities. *Cancer* **2019**, *125*, 1185-1199, doi:10.1002/cncr.31921.
63. Arafeh, R.; Samuels, Y. PIK3CA in cancer: The past 30 years. *Seminars in Cancer Biology* **2019**, *59*, 36-49, doi:https://doi.org/10.1016/j.semcancer.2019.02.002.
64. Hobbs, G.A.; Der, C.J.; Rossman, K.L. RAS isoforms and mutations in cancer at a glance. *Journal of Cell Science* **2016**, *129*, 1287-1292, doi:10.1242/jcs.182873.
65. Li, G.; Ji, X.-D.; Gao, H.; Zhao, J.-S.; Xu, J.-F.; Sun, Z.-J.; Deng, Y.-Z.; Shi, S.; Feng, Y.-X.; Zhu, Y.-Q., et al. EphB3 suppresses non-small-cell lung cancer metastasis via a PP2A/RACK1/Akt signalling complex. *Nature communications* **2012**, *3*, 667, doi:10.1038/ncomms1675.
66. D'Agostino, S.; Lanzillotta, D.; Varano, M.; Botta, C.; Baldrini, A.; Bilotta, A.; Scalise, S.; Dattilo, V.; Amato, R.; Gaudio, E., et al. The receptor protein

tyrosine phosphatase PTPRJ negatively modulates the CD98hc oncoprotein in lung cancer cells. *Oncotarget* **2018**, *9*.

67. Qi, Y.; Dai, Y.; Gui, S. Protein tyrosine phosphatase PTPRB regulates Src phosphorylation and tumour progression in NSCLC. *Clinical and Experimental Pharmacology and Physiology* **2016**, *43*, 1004-1012, doi:https://doi.org/10.1111/1440-1681.12610.
68. Jang, H.-J.; Suh, P.-G.; Lee, Y.J.; Shin, K.J.; Cocco, L.; Chae, Y.C. PLCγ1: Potential arbitrator of cancer progression. *Advances in Biological Regulation* **2018**, *67*, 179-189, doi:https://doi.org/10.1016/j.jbior.2017.11.003.
69. Miao, W.; Li, L.; Liu, X.; Qi, T.F.; Guo, L.; Huang, M.; Wang, Y. A Targeted Quantitative Proteomic Method Revealed a Substantial Reprogramming of Kinome during Melanoma Metastasis. *Scientific Reports* **2020**, *10*, 2485, doi:10.1038/s41598-020-59572-5.
70. CUI, Y.; WU, B.; FLAMINI, V.; EVANS, B.A.J.; ZHOU, D.; JIANG, W.G. Knockdown of EPHA1 Using CRISPR/CAS9 Suppresses Aggressive Properties of Ovarian Cancer Cells. *Anticancer Research* **2017**, *37*, 4415-4424.
71. Salas, M.Q.; Climent, F.; Tapia, G.; DomingoDomènech, E.; Mercadal, S.; Oliveira, A.C.; Aguilera, C.; Olga, G.; Moreno Velázquez, M.; Andrade-Campos, M., et al. Clinicopathologic features and prognostic significance of CD30 expression in de novo diffuse large B-cell lymphoma (DLBCL): results in a homogeneous series from a single institution. *Biomarkers* **2020**, *25*, 69-75, doi:10.1080/1354750X.2019.1691656.
72. Du, Y.; Grandis, J.R. Receptor-type protein tyrosine phosphatases in cancer. *Chinese Journal of Cancer* **2015**, *34*, 61-69, doi:10.5732/cjc.014.10146.
73. Chen, D.; Ma, W.; Ke, Z.; Xie, F. CircRNA hsa_circ_100395 regulates miR-1228/TCF21 pathway to inhibit lung cancer progression. *Cell Cycle* **2018**, *17*, 2080-2090, doi:10.1080/15384101.2018.1515553.
74. Okada, K.; Araki, M.; Sakashita, T.; Ma, B.; Kanada, R.; Yanagitani, N.; Horiike, A.; Koike, S.; Oh-hara, T.; Watanabe, K., et al. Prediction of ALK mutations mediating ALK-TKIs resistance and drug re-purposing to overcome the resistance. *EBioMedicine* **2019**, *41*, 105-119, doi:10.1016/j.ebiom.2019.01.019.
75. Yamaguchi, G.; Takanashi, M.; Tanaka, M.; Fujita, K.; Ohira, T.; Kuroda, M.; Ikeda, N. Isolation of miRNAs that target EGFR mRNA in human lung cancer. *Biochemical and Biophysical Research Communications* **2012**, *420*, 411-416, doi:https://doi.org/10.1016/j.bbrc.2012.03.008.
76. Stark, M.S.; Tyagi, S.; Nancarrow, D.J.; Boyle, G.M.; Cook, A.L.; Whiteman, D.C.; Parsons, P.G.; Schmidt, C.; Sturm, R.A.; Hayward, N.K. Characterization of the Melanoma miRNAome by Deep Sequencing. *PLOS ONE* **2010**, *5*, e9685, doi:10.1371/journal.pone.0009685.
77. Lou, W.; Liu, J.; Ding, B.; Xu, L.; Fan, W. Identification of chemoresistance-associated miRNAs in breast cancer. *Cancer Manag Res* **2018**, *10*, 4747-4757, doi:10.2147/CMAR.S172722.
78. Persson, H.; Kvist, A.; Rego, N.; Staaf, J.; Vallon-Christersson, J.; Luts, L.; Loman, N.; Jonsson, G.; Naya, H.; Hoglund, M., et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. *Cancer Res* **2011**, *71*, 78-86, doi:10.1158/0008-5472.can-10-1869.

79. Zhang, J.; Hu, D. miR-1298-5p Influences the Malignancy Phenotypes of Breast Cancer Cells by Inhibiting CXCL11. *Cancer Manag Res* **2021**, *13*, 133-145, doi:10.2147/CMAR.S279121.
80. Fang, F.; Huang, B.; Sun, S.; Xiao, M.; Guo, J.; Yi, X.; Cai, J.; Wang, Z. miR-27a inhibits cervical adenocarcinoma progression by downregulating the TGF-βRI signaling pathway. *Cell Death & Disease* **2018**, *9*, 395, doi:10.1038/s41419-018-0431-2.
81. Liu, H.; Pei, G.; Song, M.; Dai, S.; Wang, Y. Influence of hsa-miR-6727-5p on the proliferation, apoptosis, invasion and migration of Caski, Hela and SiHa cervical cancer cells. *Journal of B.U.ON. : official journal of the Balkan Union of Oncology* **2017**, *22*, 973-978.
82. Sand, M.; Bechara, F.G.; Gambichler, T.; Sand, D.; Bromba, M.; Hahn, S.A.; Stockfleth, E.; Hessam, S. Circular RNA expression in cutaneous squamous cell carcinoma. *Journal of Dermatological Science* **2016**, *83*, 210-218, doi:https://doi.org/10.1016/j.jdermsci.2016.05.012.
83. Fu, Q.; Du, Y.; Yang, C.; Zhang, D.; Zhang, N.; Liu, X.; Cho, W.C.; Yang, Y. An oncogenic role of miR-592 in tumorigenesis of human colorectal cancer by targeting Forkhead Box O3A (FoxO3A). *Expert opinion on therapeutic targets* **2016**, *20*, 771-782, doi:10.1080/14728222.2016.1181753.
84. Luo, Y.; Liu, W.; Tang, P.; Jiang, D.; Gu, C.; Huang, Y.; Gong, F.; Rong, Y.; Qian, D.; Chen, J., et al. miR-624-5p promoted tumorigenesis and metastasis by suppressing hippo signaling through targeting PTPRB in osteosarcoma cells. *Journal of experimental & clinical cancer research : CR* **2019**, *38*, 488, doi:10.1186/s13046-019-1491-6.
85. Shim, B.-S.; Wu, W.; Kyriakis, C.S.; Bakre, A.; Jorquera, P.A.; Perwitasari, O.; Tripp, R.A. MicroRNA-555 has potent antiviral properties against poliovirus. *Journal of General Virology* **2016**, *97*, 659-668, doi:https://doi.org/10.1099/jgv.0.000372.
86. Lemmon, M.A.; Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **2010**, *141*, 1117-1134, doi:https://doi.org/10.1016/j.cell.2010.06.011.
87. Katayama, R.; Lovly, C.M.; Shaw, A.T. Therapeutic Targeting of Anaplastic Lymphoma Kinase in Lung Cancer: A Paradigm for Precision Cancer Medicine. *Clinical Cancer Research* **2015**, *21*, 2227-2235, doi:10.1158/1078-0432.ccr-14-2791.
88. Regad, T. Targeting RTK Signaling Pathways in Cancer. *Cancers* **2015**, *7*, 1758-1784.
89. Inamdar, G.S.; Madhunapantula, S.V.; Robertson, G.P. Targeting the MAPK pathway in melanoma: why some approaches succeed and other fail. *Biochem Pharmacol* **2010**, *80*, 624-637, doi:10.1016/j.bcp.2010.04.029.
90. Ediriweera, M.K.; Tennekoon, K.H.; Samarakoon, S.R. Role of the PI3K/AKT/mTOR signaling pathway in ovarian cancer: Biological and therapeutic significance. *Seminars in Cancer Biology* **2019**, *59*, 147-160, doi:https://doi.org/10.1016/j.semcancer.2019.05.012.
91. Gritsina, G.; Xiao, F.; O'Brien, S.W.; Gabbasov, R.; Maglaty, M.A.; Xu, R.-H.; Thapa, R.J.; Zhou, Y.; Nicolas, E.; Litwin, S., et al. Targeted Blockade of JAK/STAT3 Signaling Inhibits Ovarian Carcinoma Growth. *Molecular Cancer Therapeutics* **2015**, *14*, 1035-1047, doi:10.1158/1535-7163.mct-14-0800.
92. Matthews, J.M.; Bhatt, S.; Patricelli, M.P.; Nomanbhoy, T.K.; Jiang, X.; Natkunam, Y.; Gentles, A.J.; Martinez, E.; Zhu, D.; Chapman, J.R., et al. Pathophysiological significance and therapeutic targeting of germinal center

kinase in diffuse large B-cell lymphoma. *Blood* **2016**, *128*, 239-248, doi:10.1182/blood-2016-02-696856.

93. Milella, M.; Kornblau, S.M.; Estrov, Z.; Carter, B.Z.; Lapillonne, H.; Harris, D.; Konopleva, M.; Zhao, S.; Estey, E.; Andreeff, M. Therapeutic targeting of the MEK/MAPK signal transduction module in acute myeloid leukemia. *J Clin Invest* **2001**, *108*, 851-859, doi:10.1172/JCI12807.

2.7. Supplementary

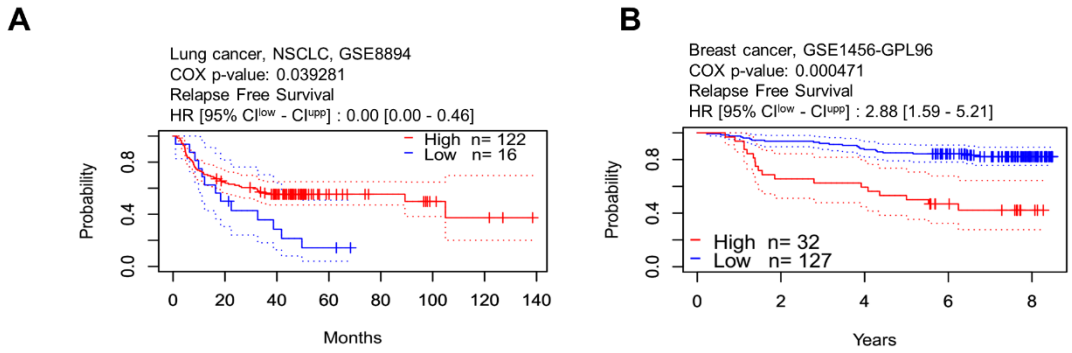


Figure S1. Anaplastic lymphoma kinase (ALK) expression and clinical prognosis consequence in non-small-cell lung carcinoma (NSCLC) and breast cancer. A, Kaplan–Meier patient survival estimate of NSCLC for ALK expression. **B**, Kaplan–Meier patient survival estimate of ALK-positive patients with breast cancer. The survival curve was determined based on the threshold of cox *p*-value <0.05, *p* <0.01. The red line denotes high expression while the blue line denotes low expression. The dotted line symbolizes maximum and minimum values of the survival average. HR, hazard ratio; CI, confidence interval; n, number of patients.

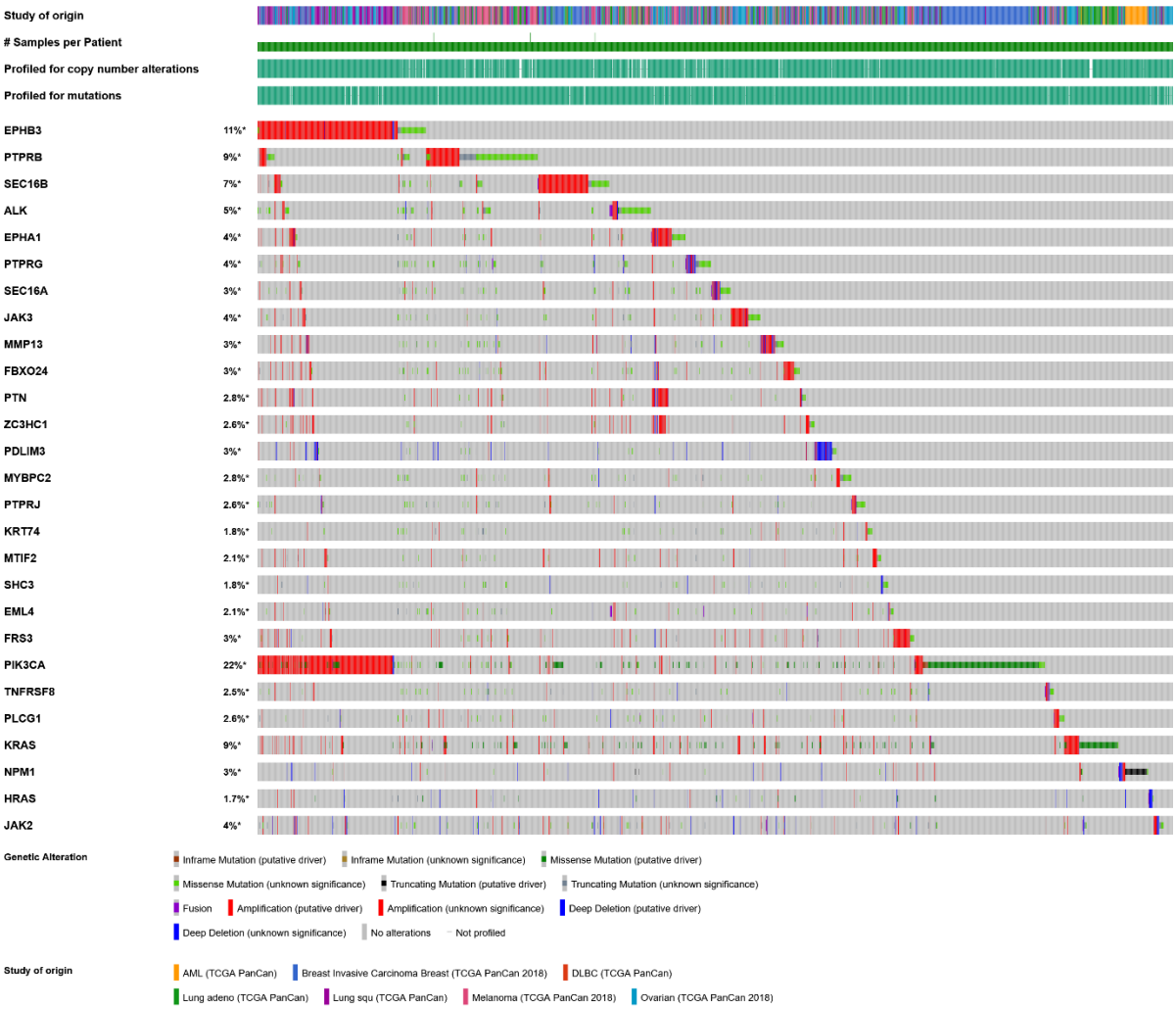


Figure S2. Oncoprint of anaplastic lymphoma kinase (ALK) and common genes. The genetic alteration of *ALK* and 17 common genes are evaluated using cBioPortal in my six queried cancers. The highest alteration is in PIK3CA (22%) followed by EPHB3 (11%), KRAS, and PTPRB (9%), while the lowest is in HRAS (1.7%).

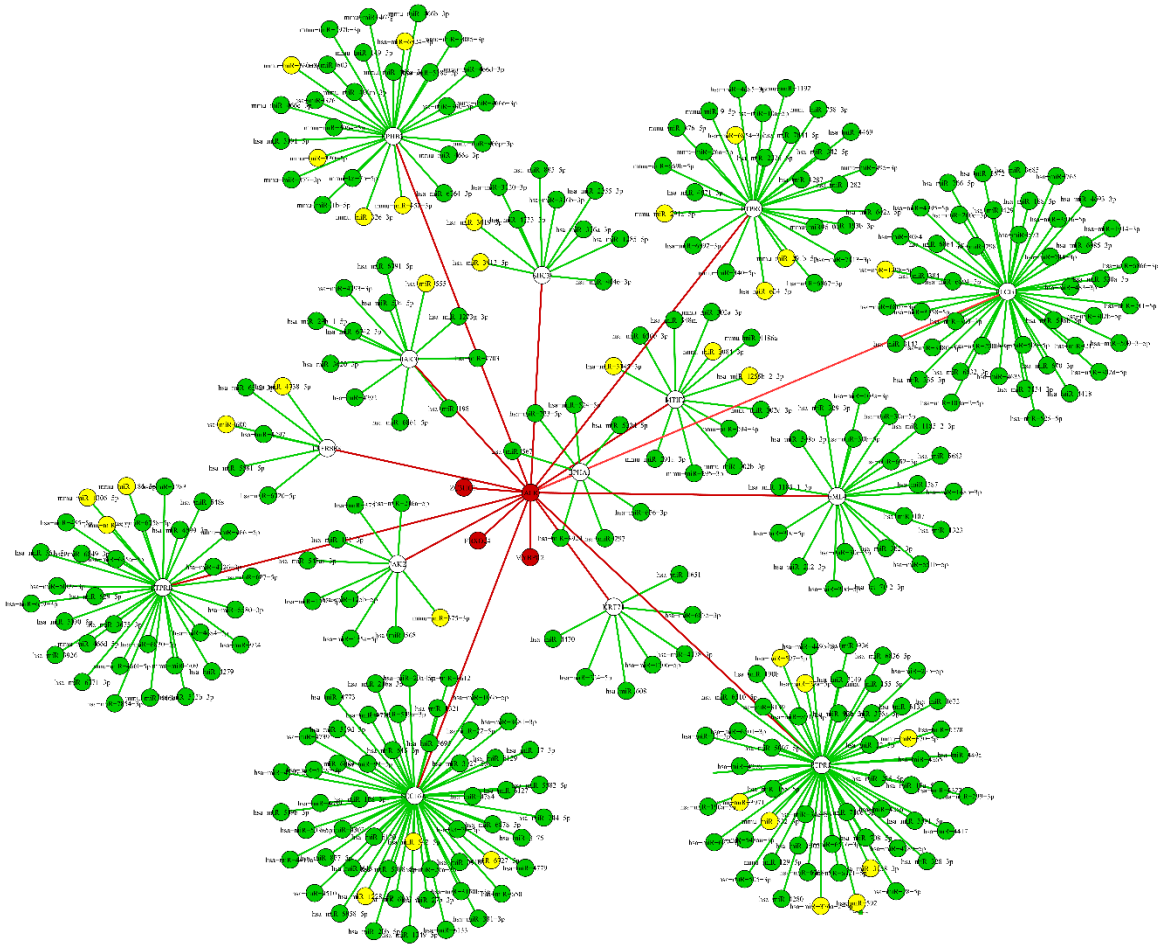


Figure S3. MicroRNA (miRNA) network with anaplastic lymphoma kinase (ALK) and 17 common target genes. Computational prediction of miRNAs co-expressed with *ALK* and common genes in mammals obtained from Enrichr and analyzed by Cytoscape_v3.8.2 software. In the top panel, the white (co-expressed with miRNA) and red (not expressed with any miRNA) nodes indicate target genes (*ALK* and common genes). The yellow nodes indicate miRNAs with significant enrichment terms ($p<0.05$), whereas green nodes denote no significant co-expression with target genes ($p>0.05$).

Supplementary table S1. Relationship between anaplastic lymphoma kinase (ALK) expression and survival in various cancers.

Cancer type	Subtype	Dataset	Endpoint	Array type	Probe id	N	Minimum *p*-value	Cox *p*-value	HR [95% ci^{low} - ci^{upp}]
Lung	NSCLC	GSE8894	Relapse-free survival	HG-U133_Plus_2	208211_s_at	138	0.021281	0.039281	0.00 [0.00 - 0.46]
	Adenocarcinoma	jacob-00182-UM	Overall survival	HG-U133A	208212_s_at	178	0.000015	0.043172	1.50 [1.01 - 2.24]
Skin	Melanoma	GSE19234	Overall survival	HG-U133_Plus_2	208211_s_at	38	0.008206	0.018439	7.58 [1.41 - 40.84]
Ovarian		GSE9891	Overall survival	HG-U133_Plus_2	208212_s_at	278	0.001838	0.003378	1.33 [1.10 - 1.60]
Brain	Astrocytoma	GSE4271-GPL96	Overall survival	HG-U133A	208211_s_at	77	0.002484	0.080639	1.30 [0.97 - 1.74]
	Glioblastoma	GSE7696	Overall survival	HG-U133_Plus_2	208212_s_at	70	0.013752	0.133669	0.76 [0.54 - 1.09]
Breast		GSE1456-GPL96	Relapse-free survival	HG-U133A	208211_s_at	159	0.000001	0.000471	2.88 [1.59 - 5.21]
		GSE9893	Overall survival	MLRG Human 21K V12.0	5325	155	0.000009	0.002117	0.59 [0.42 - 0.82]
		GSE1379	Relapse-free survival	Arcturus 22k	4477	60	0.004926	0.022674	1.75 [1.08 - 2.83]
		GSE6532-GPL570	Relapse-free survival	HG-U133_Plus_2	208211_s_at	87	0.002255	0.032822	5.98 [1.16 - 30.92]
Colorectal		GSE17536	Disease-free survival	HG-U133_Plus_2	208212_s_at	145	0.012683	0.192777	3.77 [0.35 - 41.19]
		GSE17537	Overall survival	HG-U133_Plus_2	208212_s_at	55	0.050980	0.19	6.31 [0.40 - 100.26]
Blood	AML	GSE8970	Overall survival	HG-U133A	208212_s_at	34	0.007686	0.025477	0.51 [0.29 - 0.92]
	DLBCL	E-TABM-346	Event-free survival	HG-U133A	208211_s_at	53	0.021253	0.032382	0.56 [0.32 - 0.95]
Prostate		GSE16560	Overall survival	6K DASL	DAP1_1042	281	0.006261	0.112844	1.03 [0.93 - 1.12]

Chapter III

Optimization of CRISPR-Cas13a mediated effective RNA knockdown activity using luciferase as a target transcript

Graduate School of Advanced Science and Technology
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (JAIST), JAPAN

Chapter III

Optimization of CRISPR-Cas13a mediated effective RNA knockdown activity using luciferase as a target transcript

3. Outline

RNAi technology has noteworthy potential as a future medicine and could ideally be used to knock down disease-related RNAs. However, owing to frequent off-target effects, limited accessibility of nuclear transcripts, and low efficiency, the medical application of the technique remains challenging. Here, I first evaluated the stability of Cas13a transcript and guide RNA. Next, I optimized the Cas13a and guide RNA expression vectors to achieve effective knockdown of firefly luciferase (FLuc) transcript, used as a target RNA. The knockdown specificity of Cas13a on target-search was next investigated. As examined by FLuc assay, Cas13a revealed activity only toward the orientation of the 5′ crRNA–guide RNA complex orientation. The nuclease activity of Cas13a was the highest for guide RNAs 24–30 bp in length, with comparatively low mismatch tolerance. Cas13a could effectively knock down FLuc luminescence (70–76%) and mCherry fluorescence (72%). Thus, Cas13a has strong potential for use in RNA regulation and could contribute to the development of personalized genetic medicine.

Keywords: RNA stability; RNA knockdown; CRISPR-Cas13a; guide RNA design; firefly luciferase; mCherry.

3.1. Introduction

RNA interference (RNAi) technology has gained wide recognition as a technique for knocking down target-specific mRNAs; therefore it has been appreciated for elucidation of cellular pathways, factors associated in diseases, and potential medicine [1-4]. However, many problems should be addressed before this tool can be used in clinical application, including off-target effects, the limited ability to target nuclear mRNAs, the necessity for several rounds of transfection to succeed effective knockdown, and the expenses associated with creating chemical modifications in siRNA or constructing viral vectors for shRNA [1,5-8]. Hence, substitutive methods for effective knockdown of RNA remain limited.

DNA manipulation tools such as Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 enable scientists to dissect the functions of particular genetic elements [9-11], and such tools can also bind and cleave RNA in a programmable manner with the help of PAMer oligoes [12]. One concern about using Cas9 in genomic or transcriptomic engineering is that it holds the ability to interact with non-targeted genomic DNA, theoretically leading to unintentional permanent mutations in the genome [13]. In light of that consideration, the discovery of CRISPR-Cas13 has uncovered the door for transcriptomic engineering. The CRISPR-Cas13 is a type VI RNA-guided RNA-targeting CRISPR-associated Cas effector ribonuclease **[**14-18**]**. Cas13 can only target RNA, has unlocked new opportunities in the field of RNA regulation and medicine that are not accessible when using RNAi or DNA-targeting CRISPR technologies.

CRISPR-Cas13 can be categorized into four subtypes (Cas13a–Cas13d) based on the origin of their Cas-effectors [19]. All the subtypes hold two endonucleases that are essential for optimal interference. One endonuclease is needed for crRNA maturation from the pre-crRNA processing, while other nuclease activity is conducted by two HEPN (higher eukaryotes and prokaryotes nucleotide) domains for target RNA cleavage [13]. Cas13 system consists of two components: the programmable RNA-targeting ribonuclease Cas13 and a 64-66 nucleotide long CRISPR RNA (crRNA) composed of direct repeat stem-loop and spacer RNA, guiding Cas13 to the target RNA by means of the protospacer-flanking site (PFS) [20]. Retrospective studies reported

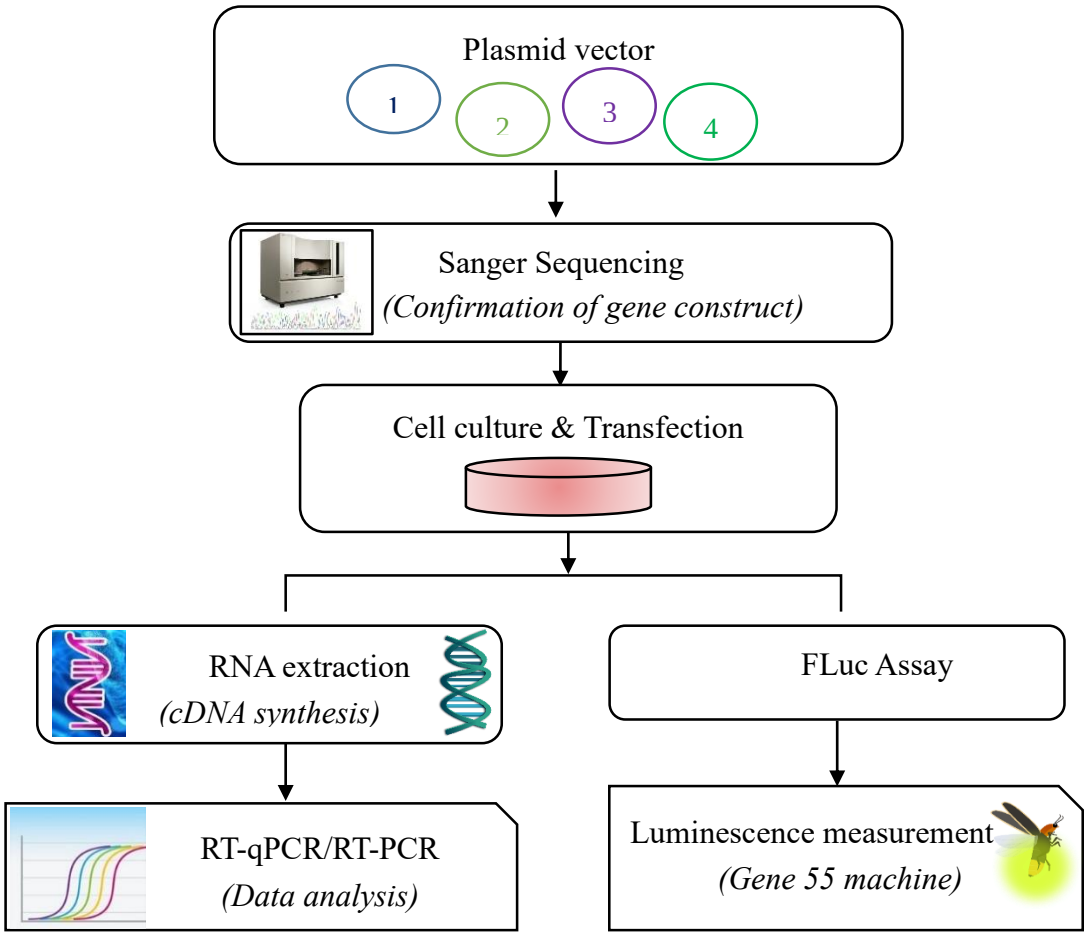
that Cas13 system has minimal PFS limitation and that target sites should be reachable upon Cas13 binding [15-17]. Beyond these basic parameters, we currently lack information about Cas13 mRNA and crRNA stability, finest schemes for effective target RNA knockdown based on target-transcript variability.

In this study, I examined CRISPR-Cas13a and crRNA–guide RNA from the standpoint of stability, effective knockdown efficiency by titration of Cas13 and guide RNA vectors, orientation of crRNA–guide RNA complex, and interference specificity on guide RNA design using firefly luciferase as a model target. Next, I used the programmable knockdown capacity of the Cas13a machinery to target mCherry transcript in HEK293T cells, with the goal of functional activity on different target transcript providing proof-of-principle of an effective knockdown tool.

3.2. Materials and Methods

3.2.1. Experimental design

This study was designed according to the following schematic experimental diagram:



3.2.2. Plasmid vectors

pGL3-Promoter Vector (FLuc) (GenBank Accession Number: U47298) and *Renilla* luciferase pRL-TK Vector (Accession Number: AF025846) were bought from Promega (Madison, WI, USA). pRSET-B mCherry was a kindly gifted by Dr. Hidekazu Tsutsui, JAIST, Japan. mCherry DNA was clone by PCR and inserted into pCS2+ vector at the *Bam*HI and *Eco*RI sites. Plasmid LwCas13a-msfGFP was obtained from Addgene (plasmid #91902) [15]. The Cas13a-GS-msfGFP region (insert) was cloned into the pcDNA3-EGFP (Addgene) by homology DNA assembly cloning. An *Xho*I recognision site, a Kozak sequence, and an NLS were introduced at the 5′ site, and an NLS, 2 × Myc tag, TAA stop codon, and *Xba*I restriction sequence were introduced at the 3′ site. Oligo sequences used for mCherry, Cas13a, and pcDNA3 amplifications are listed in Table 1 [21].

PCRs were done in a total volume of 25 µL, containing 10 µL Milli-Q water, 12.5 µL KOD One master mix (TOYOBO, Osaka, Japan), 1 µL (5 ng/µL) template DNA, and 0.75 µL (10 pmol) of each oligo. PCR settings were as follows: heat up at 95 °C for 3 min, 18 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 5 s; amplification at 68 °C for 1 min, and final amplification at 72 °C for 7 min. Next, the vector PCR products were digested with *Dpn*I: 25 µL PCR product, 9.5 µL water, and 0.5 µL *Dpn*I (NEB, Ipswich, MA, USA). PCR products were electrophoresed on 1% agarose gels and purified using the Gel PCR purification kit (MACHERY-NAGEL GmbH & Co. KG, Germany). The DNA assembly reaction contained 4 µL HiFi assembly master mix (NEB), 1 µL vector PCR product, and 3 µL insert PCR product (1:3). The final vector sequence was validated by Sanger sequencing. The vector map is presented in Figure S1A.

Table 1. Oligo sequences used for PCR amplification reaction

Target	Oligos (5' to 3')
	mCherry
mCherry_Fwd	cccggatccCACCATGGTGAGCAAGGGCGA
mCherry_Rev	ggggaattcTTACTTGTACAGCTCGTCCATGC
	pcDNA3-Cas13a
Cas13a_Fwd	gggctcgagTAGAGCGCTGCCACCATGAA
Cas13a_Rev	gggtctagattacatttcattcaagtcctcttcagaaatgagcttttgctccatttcattc aagtcctcttcagaaatgagcttttgctcggtaccTTTCTTCTTCTTAGCC TGTCCAGC
pcDNA3_Fwd	aggacttgaatgaaatgtaatctagAGGGCCCTATTCTATAGTGTCACC
pcDNA3_Rev	cttcatggtggcagcgctctactcgagCTCGAGCGGCCGCCAGTGTGAT

3.2.3. Guide RNA design

The guide RNAs were designed as the reverse complement of the target cDNA sequence to lodge the Cas13a machinery. The sequences were nominated as follows: PFS on target A or T; GC content 40–60%, off-target effects reduced by the BLAST tool in NCBI, and RNA secondary structure examined using the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) to avoid self-complementarity. Following design, the cDNA sequences of crRNA (direct repeat) and guide RNA (spacer) with 5′ *Bam*HI and 3′ *Eco*RI overhangs (Fwd strand 5′ GATCC, 3′ G; Rev strand 5′ AATTC, 3′ G) on both strands were purchased as oligos from Eurofins Genomics, Tokyo, Japan. The oligos were annealed in a PCR machine. Reactions contained total volume of 30 µL, including 15 µL Milli-Q water, 7 µL of 2 × ligation buffer (Promega), and 4 µL (10 pmol) of each oligo. Annealing was as follows: 95 °C for 5 min, lessened 7 °C every 5 min to 32 °C, 25 °C for 10 min, 4 °C for ∞ min. The annealed DNAs were then ligated with *Bam*HI- and *Eco*RI-digested pCS2+ vector, and transformed into *E. coli* DH5 alpha competent cells. Each construct was confirmed by restriction digestion and Sanger sequencing. Sequences of guide RNAs and crRNAs are listed in Table 2 [21].

Table 2. crRNA and guide RNA sequences

Target RNA name	cDNA of guide RNA (5' to 3')	Length (bp)	Ref.
Non-targeting guide	TTTACAACGTCGTGACTGGGAAAACCCT	28	[16]
crRNA (or DR)	GATTTAGACTACCCCAAAAACGAAGGG GACTAAAAC	36	[16]
Luc guide-1	CGAGAATCTCACGCAGGCAGTTCTATGA GG	30	custm
Luc guide-2	GTAATCCTGAAGGCTCCTCAGAAACAGC TC	30	custm
	Length guide		
Luc 12 guide	GTAATCCTGAAG	12	custm
Luc 18 guide	GTAATCCTGAAGGCTCCT	18	custm
Luc 24 guide	GTAATCCTGAAGGCTCCTCAGAAA	24	custm
Luc 28 guide	GTAATCCTGAAGGCTCCTCAGAAACAGC	28	custm
Luc 30 guide	GTAATCCTGAAGGCTCCTCAGAAACAGC TC	30	custm
Luc 40 guide	GTAATCCTGAAGGCTCCTCAGAAACAGC TCTTCTTCAAAT	40	custm
Luc 50 guide	GTAATCCTGAAGGCTCCTCAGAAACAGC TCTTCTTCAAA TCTATACATTA	50	custm
	Mismatch guide		
Luc WT	GTAATCCTGAAGGCTCCTCAGAAACAGC	28	custm
Luc M1	ATAATCCTGAAGGCTCCTCAGAAACAGC	28	custm
Luc M2	GCAATCCTGAAGGCTCCTCAGAAACAGC	28	custm
Luc M3	GTGATCCTGAAGGCTCCTCAGAAACAGC	28	custm
Luc M4	GTAATCCTGAAGGATCCTCAGAAACAGC	28	custm
Luc M5	GTAATCCTGAAGGCCCCTCAGAAACAGC	28	custm
Luc M6	GTAATCCTGAAGGCTACTCAGAAACAGC	28	custm
Luc M7	GTAATCCTGAAGGCTCCTCAGAAACGGC	28	custm
Luc M8	GTAATCCTGAAGGCTCCTCAGAAACAAC	28	custm
Luc M9	GTAATCCTGAAGGCTCCTCAGAAACAGA	28	custm
	mCherry		
mCherry guide	CGAAGTTCATCACGCGCTCCCACTTGAA	28	custm

3.2.4. Sanger sequencing

Sanger sequencing was done using a 3130XL Genetic Analyzer (Applied Biosystems, Japan) as following my previous protocol [22]. In some cases, 500–1000 ng samples of plasmids with particular primers were sent to a company (Eurofins Genomics) for DNA sequencing.

3.2.5. Cell culture and transfection

HEK293T cells (RIKEN Cell Bank, Japan) were cultured in Dulbecco's modified Eagle's medium (DMEM, Nacalai Tesque, Kyoto, Japan) supplemented with 10% fetal bovine serum (COSMO Bio, Brazil) at 37 °C in a humidified incubator under an atmosphere containing 5% CO_2. Unless otherwise noted, cells were plated in 48-well plates at 6×10^4 cells/well; cell counts were determined using a hemocytometer. The next day, cells were transfected with the indicated vectors using PEI reagent.

3.2.6. RNA stability assay

To monitor the expression of Cas13a mRNA and crRNA–guide RNAs, 2.5×10^3 HEK293T cells were seeded per well; transfections were performed on the following day. Total RNA was extracted after 12, 24, 48, 72, and 96 h, and subjected to cDNA synthesis and RT-PCR. For half-life measurements, 6×10^3 HEK293T cells were seeded per well. The next day, 375 ng each of Cas13a and crRNA–guide RNA vector were transfected separately. Culture media was replaced the following day. On the third day, for time point 0 (t0), cells were washed with D-PBS (Nacalai Tesque), lysed with 100 µL TRIzol reagent (Invitrogen), and then kept in a −80°C freezer. For the remaining time points (2.5, 5, 7.5, 10, and 15 h), cells were simultaneously treated with 5 µg/mL ActD (final concentration). Total RNA was isolated from the cells and subjected to cDNA synthesis and qRT-PCR. The C_T value was normalized against human *18S rRNA* and the corresponding value at t0. Half-life was estimated from t0-normalized data.

The half-life of a transcript was calculated as follows: half-life = ln(2)/kdecay [23]. This equation could be fit to the data using nonlinear regression (exponential one-phase decay; least square), which is available in GraphPad Prism.

3.2.7. Total RNA extraction

Cells were washed quickly with D-PBS and lysed by addition of 100 µL/well TRIzol reagent (Invitrogen). The lysed samples were homogenized by vortexing for 2–3 min. Next, 25 µL chloroform was added to the solution and mixed by hand for 15 s. The sample was centrifuged at 12, 000 × *g* for 15 min at 4 °C, and about 50 µL supernatant was transferred to a new tube. Next, 62.5 µL of 100% isopropanol was added to the supernatant, mixed by inversion, and incubated at room temperature for 10 min. The samples were centrifuged at 12,000 × *g* for 10 min at 4 °C, and the supernatant was discarded. The pellet was washed with 200 µL of 80% ethanol, and the supernatant was discarded. The pellet was then air-dried at room temperature for 5–10 min and dissolved in Milli-Q water.

3.2.8. cDNA synthesis

cDNA synthesis was performed in a total volume of 20 µL. Reaction mixtures contained 100 ng RNA, 0.5 µL of each oligo (dT) primer and random primer, 1 µL of 10 mM dNTPs, and water to a final volume of 14 µL. After mixing, the sample was incubated at 65 °C for 5 min and cooled on ice for 5 min. Next, 4 µL of 5 × buffer, 1 µL of ReverTra Ace (TOYOBO), and 1 µL of 0.1 M DTT were added to the solution, followed by mixing. Incubation was performed in a thermocycler (GeneAmp PCR System 9700, Applied Biosystems, Singapore) as follows: 50 °C for 30 min and 55 °C for 30 min for synthesis, 70 °C for 15 min for inactivation, and 4 °C hold.

3.2.9. RT-PCR quantification

PCR reactions were performed in a total volume of 20 µL on a thermocycler (GeneAmp PCR System 9700). Reaction mixtures contained 9.3 µL of H_2O, 1 µL cDNA template, 4 µL of 5 × Green GoTaq Flexi buffer (Promega), 2 µL of 25 mM $MgCl_2$, 2 µL of 2 mM dNTPs, 0.1 µL Taq polymerase (Promega), and 0.8 µL (10 pmol) of each primer. PCR conditions were as follows: pre-heating at 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 90 s, and final extension at 72 °C for 7 min. *GAPDH* was used as a loading control. Two-microliter aliquots of each PCR product was loaded on 6% polyacrylamide gels. The ImageJ software was used to quantify band intensities [24].

3.2.10. SYBR green-based qPCR

SYBR Green–based qPCR was performed in a total volume of 10 µL on an MX 3000P STRATAGENE Multiplex Quantitative PCR System (Agilent Technologies, USA). The reactions consisted of 3.5 µL water, 5 µL TB Green, 0.2 µL of 50 × ROX reference dye II (TaKaRa Bio, Japan), 0.5 µL cDNA template, and 0.8 µL (10 pmol) of each primer; sequences of all primers are provided in Table S4. qPCR conditions were as follows: pre-heating at 95 °C for 3 min, 45 cycles of 95 °C for 15 s, 55 °C for 10 s, and 72 °C for 15 s, and a final cycle of 95 °C for 1 min, 55 °C for 30 s, and 95 °C for 30 s. The C_T values were normalized against *GAPDH* or *18S rRNA* expression using the $2^{-\Delta\Delta C_T}$ method [25].

3.2.11. Luciferase luminometry assay

Bioluminescence of luciferase was measured according to the Dual-Luciferase Reporter (DLR) Assay System (Promega) protocol with minor modifications. Briefly, HEK239T cells were seeded at 6×10^4 cells/well of a 48-well, and cotransfected with firefly luciferase (as a target) and *Renilla* luciferase (as an internal control) at a 1:10 ratio with Cas13a and guide RNA vectors. Forty-eight hours after transfection, the media were removed, and the cells were washed with D-PBS. The cells were partially lysed in a 50 µL/well of 1 × passive lysis buffer (PLB) (Promega, USA) and incubated for 15 min at room temperature with shaking. Next, the plate was kept on ice, and luminescence was measured in each well. First, a blank was measured: 10 µL of 1 × PLB was mixed into 30 µL luciferase substrate I (LARII) in a measuring tube, firefly relative light unit (RLU) was observed on a Gene Light 55 (Microtech Nichion Co., Ltd., Japan), 30 µL Stop and Glo reagent was added to the same tube, and *Renilla* RLU was observed. Second, the sample was measured: 2 µL cell lysate and 8 µL of 1 × PLB were mixed into 30 µL LARII, and then the RLU values were observed as described above for the blank. The blank value was subtracted from all sample values. The condition of the machine was set to dual-luciferase assay mode: first delay time, 3 s; first count time, 6 s; second delay time, 3 s; second count time, 6 s; repeat time, 1. Finally, the data were normalized as described for the genetic reporter assay [26].

3.2.12. Data analysis

Statistical analyses were performed in MS Excel (version 2016) or GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA, USA). All quantitative data are represented as means ± SEM. A two-tailed Student's t-test was used to evaluate the significance of comparisons between two groups. A p-value of 0.05 was considered statistically significant.

3.3. Results

3.3.1. Analysis of RNA stability

mRNA turnover is an important determinant of the abundance of cellular mRNA and, in turn, the level of protein or functional RNA. I first examined the stability of Cas13a mRNA and crRNA–guide RNA to illustrate their decay rates and expression pattern. Both plasmids were transfected independently into HEK293T cells, and expression started to reduce 48 h after transfection (Figure 1A–B, Figure S1B). To define the decay rates, I assessed the RNA abundance at various time points after treatment with the transcriptional inhibitor actinomycin D (ActD). RNA expression was determined in the lack of ActD treatment (t0) and after treatment. RNA abundance at various time points was firstly normalized against a reasonably stable housekeeping gene, human *18S rRNA*, rather than *GAPDH* [25], however inhibition was not linear (Figure 1C–E). Since ActD can inhibit both RNA pol II and III, I successively normalized against t0. To calculate the RNA decay rate, I tailored an exponential one-phase decay model to the normalized expression data. The half-lives of Cas13a and the crRNA–guide RNA were determined as 5.82 and 7.23 h, separately (Figure 1F–G). These data will be worthwhile in future experiments containing Cas13a.

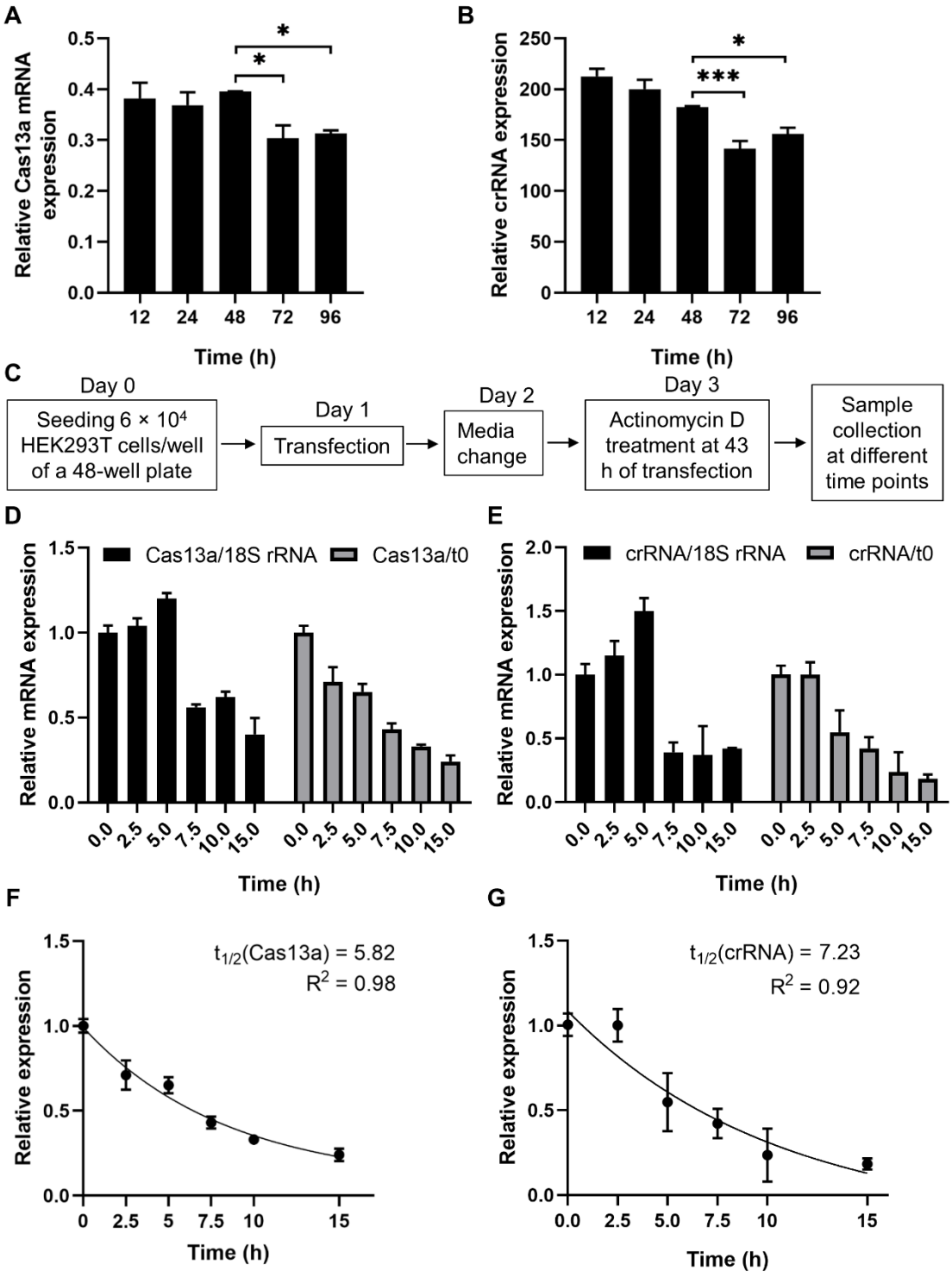


Figure 1. Analysis of RNA stability using a transcription inhibitor [21]. A–B, RT-PCR quantification of Cas13a and crRNA–guide RNA transfected into HEK293T cells. Total RNA was isolated after 12, 24, 48, 72, or 96 h, and subjected to cDNA synthesis and RT-PCR. Data were normalized against *GAPDH*. **C**, Schematic of treatment with the transcription inhibitor actinomycin D (ActD) for half-life analysis. **D**, qRT-PCR of Cas13a mRNA abundance after treatment with

ActD for the indicated times. Cas13a mRNA levels were normalized against the corresponding levels of *18S rRNA* and against time point 0 (i.e., no ActD treatment). **E**, qRT-PCR result of crRNA–guide expression after treatment with ActD for the indicated times. mRNA levels were normalized as in D. **F**, Exponential one-phase decay model of Cas13a mRNA fitted to a time course of transcriptional inhibition, normalized against the t0 time point. **G**, One-phase decay model of crRNA–guide RNA fitted to a time course of transcriptional inhibition, normalized against the t0 time point. All data represent mean values ± SEM from three individual experiments. qRT-PCR results represent three biological replicates and two technical replicates. Statistical significance was determined by Student's t-test. *** $p < 0.001$, * $p < 0.05$.

3.3.2. Optimizing the quantity of Cas13a for efficient RNA knockdown

I next inspected the ability of the programmable ribonuclease Cas13a to knock down firefly luciferase (FLuc) transcript in HEK293T cells. First, I optimized the amounts of Cas13a and guide RNA for knockdown of FLuc (where *Renilla* luciferase was used as an internal control). A lower amount (59 fmol/mL) of Cas13a was not adequate to decline luciferase luminescence (Figure 2C), but a higher amount (177 fmol/mL) substantially diminished the level of luciferase RNA compared to CasControl (68% for guide-1 as well as 76% for guide-2). Excess Cas13a (250.75 fmol/mL) radically exaggerated overall luciferase luminescence. I found analogous luminescence activity when I titrated the guide RNA (Figure 2D). Hereafter, I used 177 fmol/mL Cas13a quantity with a 1:3 molar ratio of guide RNA in successive knockdown experiments. It is important to note that this concentration was not optimal for every CRISPR-Cas13a experiments, and merely provides an approximation of the amount of DNA that is needed for adequate Cas13a activity. Optimization of the absolute and relative concentrations of Cas13a and guide RNA amount is essential before starting a new experiment.

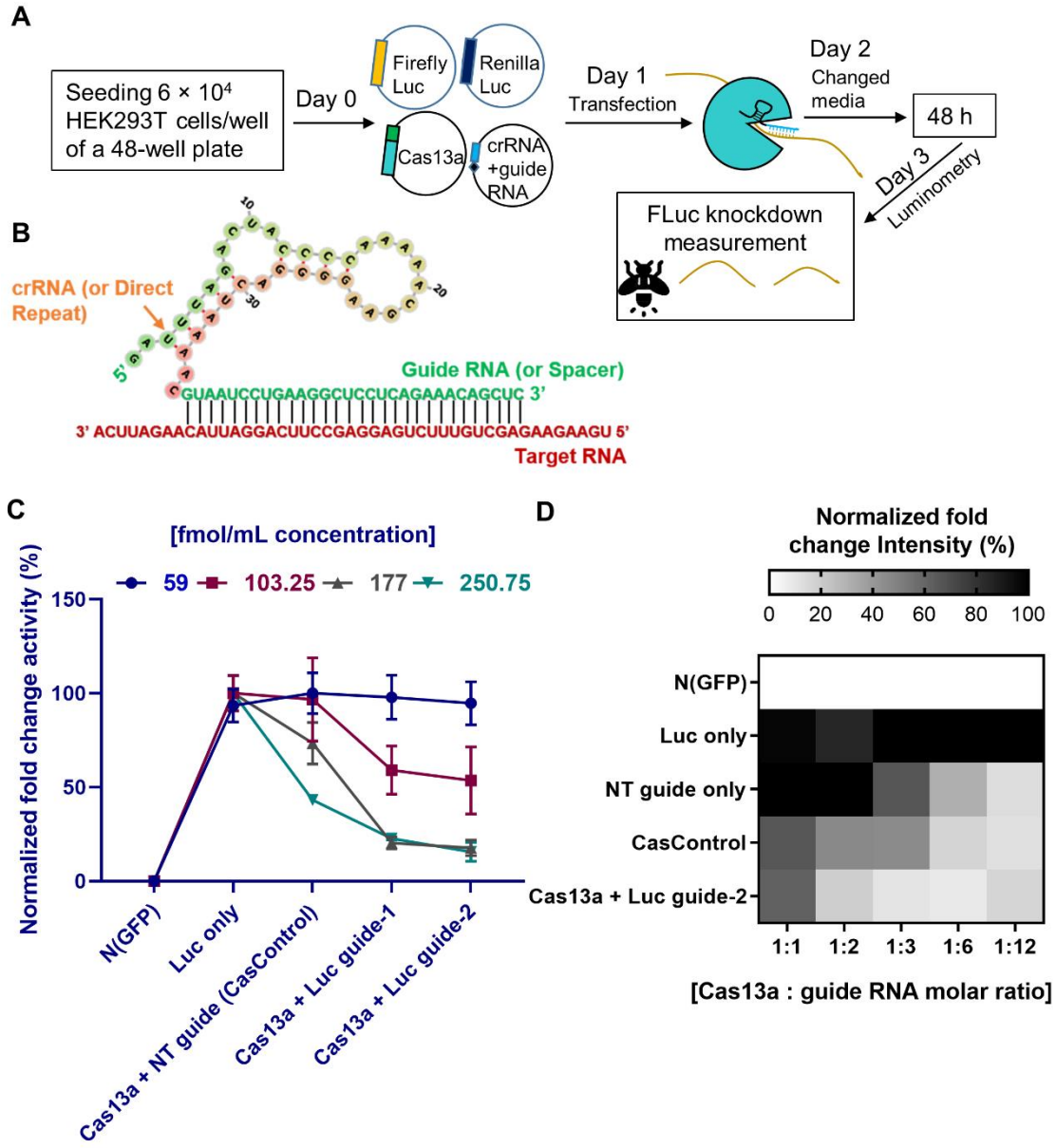


Figure 2. Quantitative analysis of Cas13a and guide RNA [21]. **A**, Experimental outline. **B**, Schematic of the strategy for design of the crRNA–guide RNA complex targeting a specific RNA. **C**, Quantity of Cas13a plasmid required for sufficient knockdown: line graph shows four concentrations (59, 103.25, 177, and 250.75 fmol/mL) of Cas13a vector. N (GFP), negative control; Luc only, luciferase expression without Cas13a; CasControl, Cas13a with nontarget (NT) guide RNA; Cas13a + Luc guide-1/2, Cas13a with target guide-1/2. The FLuc reporter was used as a target transcript while RLuc was used as an internal control. In this experiment, 177 fmol/mL Cas13a yielded sufficient knockdown (68% and 76% in guide-1 and -2), although a small reduction was also observed in CasControl (177: CasControl);

D, Quantity of crRNA–guide RNA plasmid required for knockdown. Heat map shows five different amounts of guide RNAs with 177 fmol/mL Cas13a (1:1, 1:2, 1:3, 1:6, and 1:12). Left side: N (GFP), negative control; Luc only, luciferase expression without Cas13a; NT guide only, nontarget guide without Cas13a; CasControl, Cas13a with nontarget (NT) guide; and Cas13a + Luc guide-2, Cas13a with target guide-2. The target transcript and internal control were used as C. Where a 1:3 (Cas13a: guide RNA) molar ratio was considered satisfactory. All data represent mean values ± SEM from three individual experiments.

3.3.3. Optimal condition for RNA knockdown and preference of 5′ crRNA for Cas13a activity

To determine the effect of Cas13a, crRNA, and experimental integrity, I transfected singular factors in the presence or absence of FLuc, using the optimized Cas13a and guide RNA quantities defined above. No luminescence activity was detected in the absence of FLuc (Figure 3A). On the other hand, luminescence was diminished 70% and 76% relative to CasControl in FLuc guide-1 as well as -2, separately, in the presence of FLuc. However, I also observed a decrease when Cas13a vector alone was used, possibly due to an effect of Cas13a ribonuclease on cellular metabolic load. Afterward, I sought to define which RNA strand/direction of the crRNA–guide RNA complex lead to greater Cas13a knockdown activity. Ribonuclease activity showed a substantial preference for the 5′ crRNA–guide RNA orientation (63% and 67% interference for guide-1 and -2, respectively) over the 3′ crRNA–guide RNA orientation (6% and 24% knockdown; guide-1 and -2) relative to CasControl (Figure 3B). Given that the crRNA scaffold is significant for the effective knockdown activity of Cas13a, afterward I used 5′ crRNA–guide RNA in subsequent experiments.

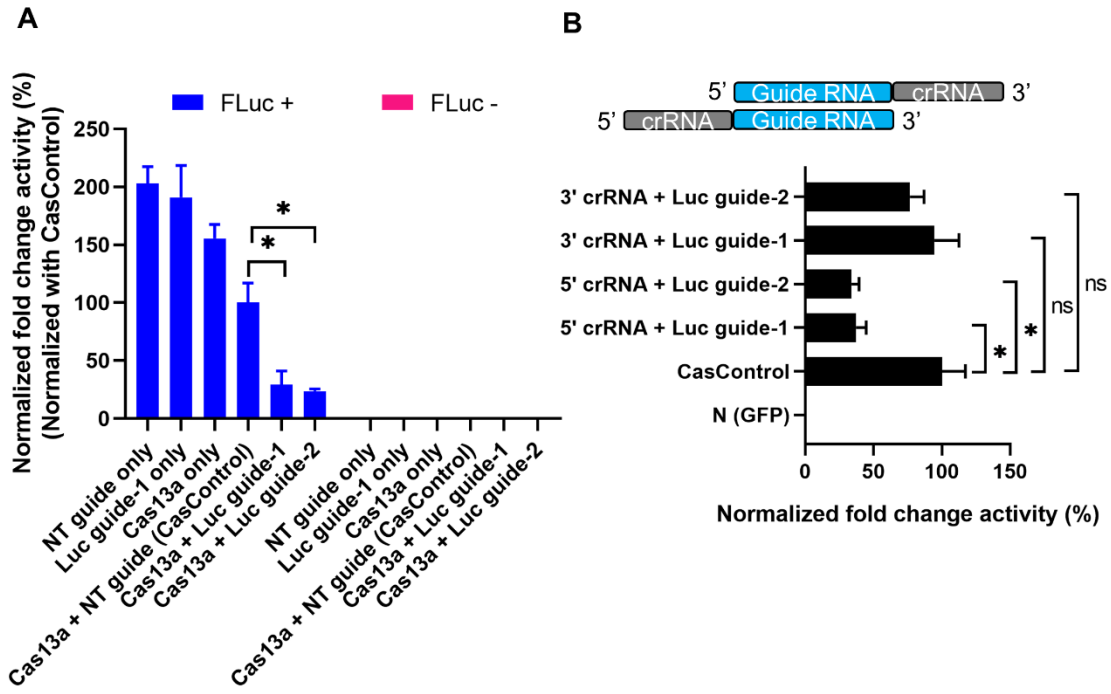


Figure 3. RNA knockdown and preference of Cas13a functionality for 5′ crRNA–guide RNA [21]. **A**, Knockdown of FLuc: FLuc +, presence of firefly luciferase during transfection; CasControl only, nontargeting guide only; Luc guide-1 only, guide RNA for target FLuc; Cas13a only, Cas13a plasmid without any guide; Cas13a + CasControl, Cas13a plasmid and nontargeting guide plasmid transfection; Cas13a + Luc guide-1/2, targeted guide-1/2 and Cas13a. The FLuc was used as a target transcript while RLuc was used as an internal control. **B**, Preference of Cas13a functionality for 5′ crRNA–guide RNA complex: N (GFP), negative control; CasControl, nontargeting guide RNA; 5′/3′crRNA + Luc, Cas13a with 5′ or 3′ crRNA–guide RNA target (FLuc) guide-1/2. The target transcript and internal control were same as A. All data represent mean values ± SEM from three individual experiments. Statistical significance was measured by Student's t-test. * $p < 0.05$, ns, not significant.

3.3.4. Selectivity of Cas13a in guide RNA design

To address the knockdown specificity of Cas13a on target-search guided by guide RNA, I first investigated the effect of guide RNA length on Cas13a nuclease activity using 12, 24, 28, 30, 40, and 50 bp long guide RNAs (in other words, spacer RNAs)

(see Figure 2B and 4A). As presented in Figure 4A, Cas13a was not capable to achieve interference with a 12 bp guide, but prominently decreased luciferase luminescence with a 24, 28, or 30 bp guide (> 70%); its activity reduced with longer guide RNAs (40 or 50 bp).

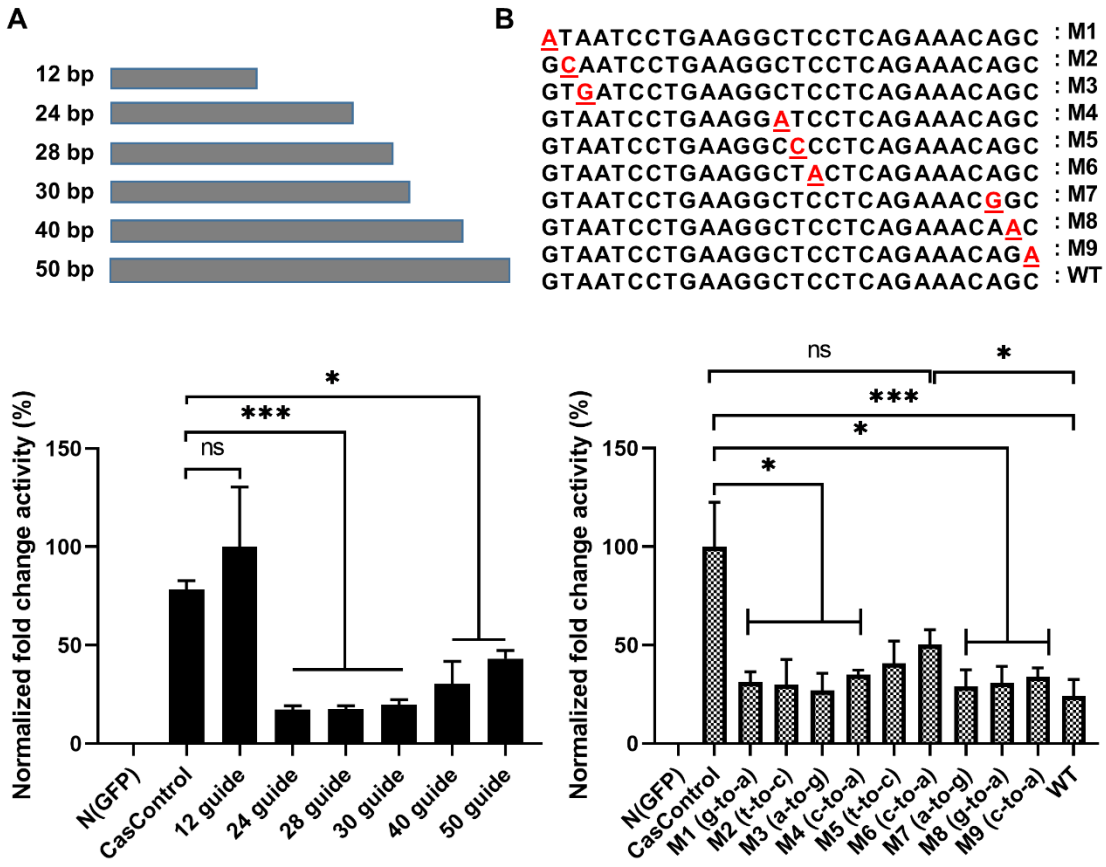


Figure 4. Effect of guide RNA design on Cas13a [21]. A, Length of guide RNA: 12, 24, 28, 30, 40, and 50 bp guide RNAs were tested; knockdown started at a length of 24 bp. The FLuc was used as a target transcript while RLuc was used as an internal control to obtain the guide RNA length limit. **B**, Sensitivity of Cas13a to a single–base mismatch in the guide RNA: M1–M3, mismatch at the 5′ end; M4–M6, mismatch in the middle seed region; M7–M9, a mismatch at the 3′ end. Red letter indicates mismatch nucleotide. Cas13a activity was sensitive to mismatches at all sites, especially in the middle region. Values are normalized against CasControl and expressed as percentages. The target reporter transcript and internal control were the same as A. All data represent mean values ± SEM from three individual experiments. Statistical significance was determined by Student's t-test. *** $p < 0.001$, * $p < 0.05$. ns, not significant.

To further examine the reliance of Cas13a activity on guide RNA design, I inserted a single–base pair mutation in the 5′, middle, or 3′ region of a 28 bp guide RNA. Relative to CasControl, luciferase luminescence was suppressed 68–70%, 58%, and 76% by guide RNAs with a mutation in the 5′ or 3′ region, a mutation in the middle region, and no mutation, respectively (Figure 4B: M1–M3, 5′ site; M4–M6, middle region; M7–M9, 3′ site). In consequence, Cas13a does not tolerate a mismatch in the guide RNA, predominantly in the middle region. These results will support in application of Cas13a to knockdown of disease-relevant RNAs.

3.3.5. Validation of effective knockdown: mCherry

I guessed that the endonuclease activity of Cas13a would show a discrepancy among target RNAs. With this in mind, I targeted mCherry transcripts in HEK293T cells, and found that mCherry fluorescence was reduced up to 72% (Figure 5). Which demonstrating that Cas13a has potential to knock down a RNA of interest effectively and programmable in manner.

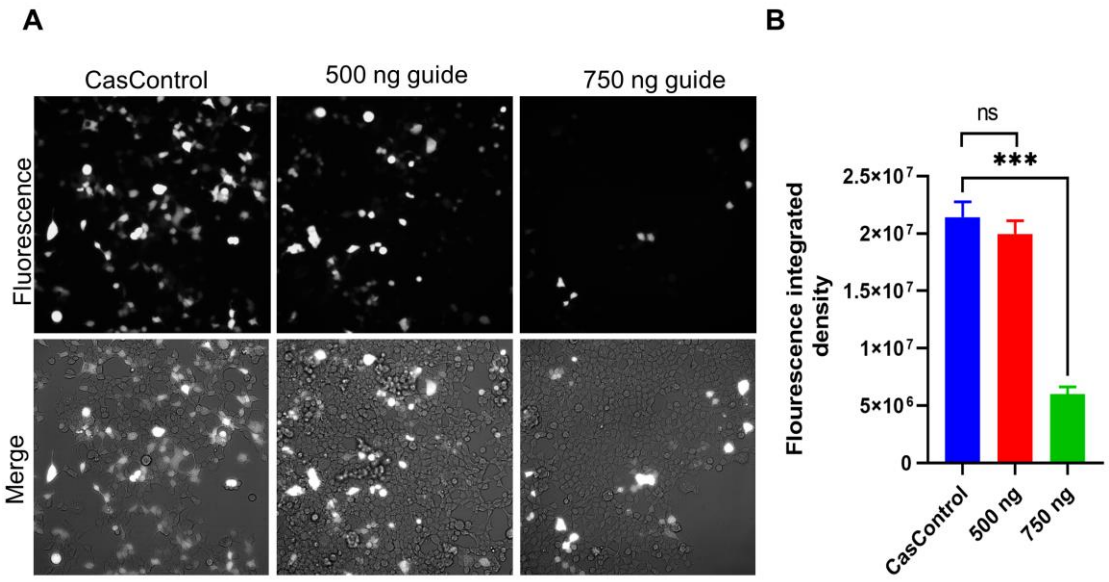


Figure 5. mCherry mRNA knockdown by Cas13a [21]. A, Fluorescence images (20x) of mCherry knockdown at two guide RNA concentrations; **B**, Graphical representative of A. Statistical significance was determined by Student's t-test. ***$p < 0.001$.

3.4. Discussion

I observed that the RNA abundance of Cas13a and crRNA–guide RNA started to reduce 48 h after transfection in HEK293T cells. The half-lives of Cas13a mRNA and crRNA–guide RNA were 5.82 and 7.23 h, respectively (Figure 1). Retrospective study reported that expression of Cas13a mRNA losses 2 h after transfection in human glioma cells [27], however, those authors did not tag Cas13a with msfGFP, which may elucidate the short expression time. As previously described, msfGFP upturns the stability of Cas13a, which I used in my study [15,16]. The authors of the glioma study examined expression levels but did not enumerate the mRNA decay rate for Cas13a or crRNA–guide RNA.

Abudayyeh and colleagues discovered CRISPR-Cas13a and comprehensively studied the effector in bacterial, plant, and human cells [14-16]. They used 150 ng of Cas13a and 300 ng of guide RNA vectors for *Cypridina* luciferase RNA interference in HEK293FT cells of a 96-well plate (40% or 72–75% reduction). Our optimization experiments revealed that 177 fmol/mL Cas13a vector (equivalent to 187.5 ng in a 96-well plate) and a 1:3 molar ratio of vector-to-guide RNA were adequate to achieve 70–76% knockdown of FLuc transcript (Figure 2 & 3A). When I used double the concentration of guide RNA vectors, I saw knockdown in both the control and target samples (Figure 2D; 1:6 and 1:12 molar ratio), potentially due to off-target properties or an overload of guide RNA. My target RNA was firefly luciferase RNA, which is 27–29% alike to the *Gaussia* and *Cypridina* luciferases used by Abudayyeh et al. (Figure S1C). Konermann et al. used 200 ng each of Cas13d and guide RNA vector to knock down mCherry transcripts (87–92%) in HEK293FT cells [17], almost similar to the concentrations used in this study. I believe that it is better to titrate the guide RNA vector concentration based on the molar ratio rather than the concentration of vector, as Cas13a is large, but the guide RNA is very short. For instance, If I add 300 ng guide RNA with 150 ng Cas13a, the concentration of guide RNA would be too high, and the extra foreign RNA could intensify the cellular metabolic load. Hence, titration of Cas13a and guide RNA vectors before execution a new experiment is essential for obtaining high effectiveness of knockdown.

From the standpoint of Cas13a functionality, I found that it was best for the crRNA (direct repeat) to be 5′ of the guide RNA (spacer) (Figure 3B). This is similar

to previous studies of Cas13a, Cas13c, and Cas13d [14–17,27]. Contrarily, the optimal relative position is altered in the Cas13b system, which prefers the crRNA to be 3′ of the guide RNA [13,16,28]. The stem–loop structure of crRNA and their orientation are important for the activity of Cas13a, and the ideal orientation differs among CRISPR-Cas effector types.

To comprehend Cas13a's target RNA binding and cleavage functions, I investigated the length constraints for guide RNAs in human HEK293T cells. I found that a 24–30 bp guide RNA was adequate for maximal FLuc transcript knockdown (> 75%). The cleavage activity was decreased when the guide was both in shorter (12 bp) or longer (40 or 50 bp) (Figure 4A). A analogous observation (21/24 bp length) was reported in a previous study that used an *in vitro* experiment [15]. However, with one exemption observed in human cells, previous study showed that shortening a guide RNA below 28 bp diminished Cas13a ribonuclease activity [16]. Contrarily, I found the highest activity between 24 and 30 bp. The reason for this inconsistency might be related to variations in experimental handling, target genes, or cell types. Cas13a ribonuclease activity was never observed with 40 or 50 bp long guide RNAs.

Later, I observed that ribonuclease activity of Cas13a was sensitive to mutations between the guide RNA and target RNA molecules (Figure 4B). Previous studies of Cas13a and Cas13b mediated RNA interference in HEK293FT cells, using either crRNA spacers or target transcript plasmids containing single or double mutations, reported mutation sensitivity, particularly in the middle region of the guide RNA (seed region) [15,16]. I also detected such a sensitivity in the middle portion of three guide RNAs (Figure 4B; M4-M6 guide RNAs). This sensitivity property broadens the applicability of Cas13a to target mRNA, especially for knockdown of disease-relevant transcripts. Recently one *in vitro* study on Cas9 and Cas12a found that the tolerance of mispairing between target DNA and guide RNA is surprisingly high and substantially influenced by polarity of the target DNA [29]. Which could produce unintended long-lasting error through collateral cleavage in human genome [29]. Similarly, Cas13 also showed collateral cleavage activity in vitro assay [14,30]. Interestingly, this collateral activity of Cas13 has not been discovered in either human or plant cell lines [15,16]. Further experiment is required to comprehend the Cas13′s collateral activity in the border milieu of immunity.

Cas13a activity varies among target RNA transcripts [15-17]. In this study, I was able to diminish mCherry transcript expression up to 72% using Cas13a (Figure 5), whereas another study reported 92% knockdown of mCherry using a different CRISPR effector (Cas13d) [17].

3.5. Concluding remark

In this study, I determined the half-life of CRISPR Cas13a mRNA and crRNA in human cells. Subsequently, I successfully knocked down firefly luciferase in HEK293T cells (>70%) after optimizing the concentrations of Cas13a mRNA and guide RNA. In addition, I evaluated the effects of the orientation of the crRNA–guide RNA complex, length, and mismatches on Cas13a activity. Based on our successful robust knockdown of marker genes such as firefly luciferase, mCherry in HEK293T cells, I speculated that this approach could be used to knock down other transgenes or genes of interest (or RNA of interest, e.g., COVID-19 RNA or noncoding RNA) that affect the biological functions of cells.

3.6. Bibliography

1. Lieberman, J. Tapping the RNA world for therapeutics. *Nature Structural & Molecular Biology* **2018**, *25*, 357-364, doi:10.1038/s41594-018-0054-4.
2. Manfredsson, F.P.; Lewin, A.S.; Mandel, R.J. RNA knockdown as a potential therapeutic strategy in Parkinson's disease. *Gene therapy* **2006**, *13*, 517-524, doi:10.1038/sj.gt.3302669.
3. Kim, D.H.; Rossi, J.J. RNAi mechanisms and applications. *BioTechniques* **2008**, *44*, 613-616, doi:10.2144/000112792.
4. Mroweh, M.; Decaens, T.; Marche, P.N.; Macek Jilkova, Z.; Clément, F. Modulating the Crosstalk between the Tumor and Its Microenvironment Using RNA Interference: A Treatment Strategy for Hepatocellular Carcinoma. *International Journal of Molecular Sciences* **2020**, *21*, 5250.

5. Yin, W.; Rogge, M. Targeting RNA: A Transformative Therapeutic Strategy. *Clinical and translational science* **2019**, *12*, 98-112, doi:10.1111/cts.12624.
6. Aagaard, L.; Rossi, J.J. RNAi therapeutics: principles, prospects and challenges. *Adv Drug Deliv Rev* **2007**, *59*, 75-86, doi:10.1016/j.addr.2007.03.005.
7. Pecot, C.V.; Calin, G.A.; Coleman, R.L.; Lopez-Berestein, G.; Sood, A.K. RNA interference in the clinic: challenges and future directions. *Nat Rev Cancer* **2011**, *11*, 59-67, doi:10.1038/nrc2966.
8. Kamiya, Y.; Takeyama, Y.; Mizuno, T.; Satoh, F.; Asanuma, H. Investigation of Strand-Selective Interaction of SNA-Modified siRNA with AGO2-MID. *International Journal of Molecular Sciences* **2020**, *21*, 5218.
9. Doudna, J.A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096, doi:10.1126/science.1258096.
10. Hsu, P.D.; Lander, E.S.; Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **2014**, *157*, 1262-1278, doi:10.1016/j.cell.2014.05.010.
11. Filippov-Levy, N.; Reich, R.; Davidson, B. The Biological and Clinical Role of the Long Non-Coding RNA LOC642852 in Ovarian Carcinoma. *International Journal of Molecular Sciences* **2020**, *21*, 5237.
12. O'Connell, M.R.; Oakes, B.L.; Sternberg, S.H.; East-Seletsky, A.; Kaplan, M.; Doudna, J.A. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **2014**, *516*, 263-266, doi:10.1038/nature13769.
13. O'Connell, M.R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. *Journal of Molecular Biology* **2019**, *431*, 66-87, doi:https://doi.org/10.1016/j.jmb.2018.06.029.
14. Abudayyeh, O.O.; Gootenberg, J.S.; Konermann, S.; Joung, J.; Slaymaker, I.M.; Cox, D.B.T.; Shmakov, S.; Makarova, K.S.; Semenova, E.; Minakhin, L., et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016**, *353*, aaf5573, doi:10.1126/science.aaf5573.

15. Abudayyeh, O.O.; Gootenberg, J.S.; Essletzbichler, P.; Han, S.; Joung, J.; Belanto, J.J.; Verdine, V.; Cox, D.B.T.; Kellner, M.J.; Regev, A., et al. RNA targeting with CRISPR–Cas13. *Nature* **2017**, *550*, 280-284, doi:10.1038/nature24049.
16. Cox, D.B.T.; Gootenberg, J.S.; Abudayyeh, O.O.; Franklin, B.; Kellner, M.J.; Joung, J.; Zhang, F. RNA editing with CRISPR-Cas13. *Science* **2017**, *358*, 1019-1027, doi:10.1126/science.aaq0180.
17. Konermann, S.; Lotfy, P.; Brideau, N.J.; Oki, J.; Shokhirev, M.N.; Hsu, P.D. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **2018**, *173*, 665-676.e614, doi:10.1016/j.cell.2018.02.033.
18. Burmistrz, M.; Krakowski, K.; Krawczyk-Balska, A. RNA-Targeting CRISPR–Cas Systems and Their Applications. *International Journal of Molecular Sciences* **2020**, *21*, 1122.
19. 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., et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **2020**, *18*, 67-83, doi:10.1038/s41579-019-0299-x.
20. Granados-Riveron, J.T.; Aquino-Jarquin, G. CRISPR–Cas13 Precision Transcriptome Engineering in Cancer. *Cancer research* **2018**, *78*, 4107-4113, doi:10.1158/0008-5472.can-18-0785.
21. Saifullah; Sakari, M.; Suzuki, T.; Yano, S.; Tsukahara, T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. *International Journal of Molecular Sciences* **2020**, *21*, 8904, doi:https://doi.org/10.3390/ijms21238904.
22. Saifullah; Tsukahara, T. Genotyping of single nucleotide polymorphisms using the SNP-RFLP method. *Biosci Trends* **2018**, *12*, 240-246, doi:10.5582/bst.2018.01102.

23. Lugowski, A.; Nicholson, B.; Rissland, O.S. Determining mRNA half-lives on a transcriptome-wide scale. *Methods (San Diego, Calif.)* **2018**, *137*, 90-98, doi:10.1016/j.ymeth.2017.12.006.

24. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **2012**, *9*, 671-675, doi:10.1038/nmeth.2089.

25. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **2001**, *25*, 402-408, doi:10.1006/meth.2001.1262.

26. Schagat, T.; Paguio, A.; Kopish, K. Normalizing genetic reporter assays: approaches and considerations for increasing consistency and statistical significance. *Cell Notes* **2007**, *17*, 9-12.

27. Wang, Q.; Liu, X.; Zhou, J.; Yang, C.; Wang, G.; Tan, Y.; Wu, Y.; Zhang, S.; Yi, K.; Kang, C. The CRISPR-Cas13a Gene-Editing System Induces Collateral Cleavage of RNA in Glioma Cells. *Adv Sci (Weinh)* **2019**, *6*, 1901299-1901299, doi:10.1002/advs.201901299.

28. Smargon, A.A.; Cox, D.B.T.; Pyzocha, N.K.; Zheng, K.; Slaymaker, I.M.; Gootenberg, J.S.; Abudayyeh, O.A.; Essletzbichler, P.; Shmakov, S.; Makarova, K.S., et al. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular Cell* **2017**, *65*, 618-630.e617, doi:10.1016/j.molcel.2016.12.023.

29. Hewes, A.M.; Sansbury, B.M.; Barth, S.; Tarcic, G.; Kmiec, E.B. gRNA Sequence Heterology Tolerance Catalyzed by CRISPR/Cas in an In Vitro Homology-Directed Repair Reaction. *Molecular therapy. Nucleic acids* **2020**, *20*, 568-579, doi:10.1016/j.omtn.2020.03.012.

30. East-Seletsky, A.; O'Connell, M.R.; Knight, S.C.; Burstein, D.; Cate, J.H.; Tjian, R.; Doudna, J.A. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **2016**, *538*, 270-273, doi:10.1038/nature19802.

3.7. Supplementary

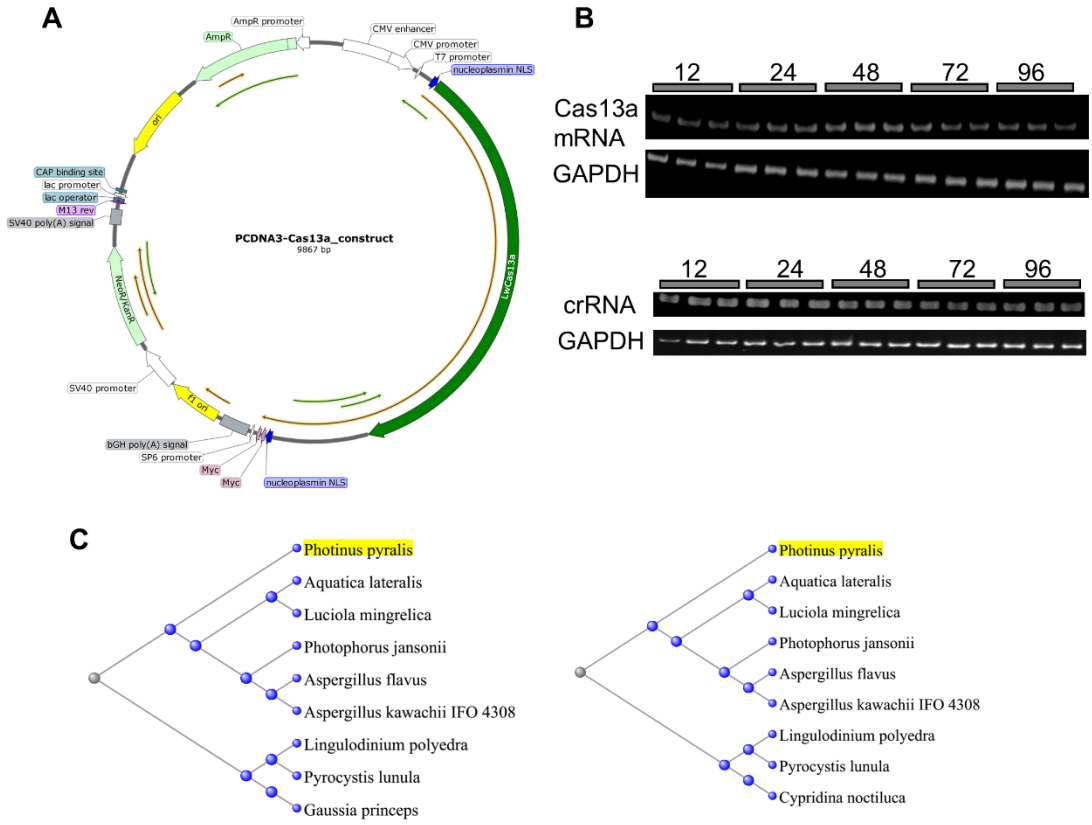


Figure S1. pcDNA3-Cas13a construct map, RT-PCR gel image of Cas13a mRNA and crRNA, and Phylogenetic tree of Luciferin [21]. A, Circular map of pcDNA3-Cas13a construct vector plasmid. **B**, RT-PCR gel images of Cas13a and crRNA guide. **C**, Phylogenic tree of Luciferin originated from various species. Identity of Firefly luciferase protein with Gaussia and Cypridinia is 29.03% and 27.45% respectively.

Chapter IV

The CRISPR-Cas13a gene-editing system downregulated the oncogenic driver *EML4-ALK* expression in *ALK*-positive lung cancer cells

Graduate School of Advanced Science and Technology
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (JAIST), JAPAN

Chapter IV

The CRISPR-Cas13a gene-editing system downregulated the oncogenic driver *EML4-ALK* expression in *ALK*-positive lung cancer cells

4. Outline

Anaplastic Lymphoma Kinase (ALK) tyrosine kinase inhibitors provoke a significant anti-tumor response; however, they inevitably succumb to the acquired resistance. Therefore, an alternative therapeutic strategy that limits *ALK* over-activation is necessary for the treatment of this lung cancer. Here, I show that the CRISPR-Cas13a effector possesses effective knockdown potency for oncogenic driver *EML4-ALK* transcript in lung cancer cells (H3122 and H2228). I found the *EML4* transcript was not substantially expressed but *ALK* expressed 100 fold and 80 fold higher in H3122 and H2228 cells separately compare to a non-fusion transcript in HEK293T cells. I also found that the EML4-ALK oncoproteins were robustly down-regulated (>80%) by employing Cas13a in those lung cancer cells based on western blot results. Consequently, the tyrosine kinase phosphorylation (50–70%) and cell growth (up to 40%) were inhibited. Overall the obtained data demonstrated that the CRISPR-Cas13a protein downregulated the ALK expression in the lung cancer cells. This study suggests that CRISPR-Cas13a mediated *EML4-ALK* RNA knockdown devises a potential therapeutic strategy for treating *ALK*-positive lung cancer.

Keywords: CRISPR-Cas13a; EML4-ALK; RNA knockdown; cell viability; H3122 and H2228 cells.

4.1. Introduction

Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer, and the majority of these are adenocarcinoma [1,2]. One of the most common types of oncofusion in lung adenocarcinoma is EML4-ALK which leads to the constitutive activation of ALK and thus increased cell survival, cell growth, and proliferation [3,4]. The *EML4* and *ALK* genes are located 12 Mb apart on chromosome 2 (2p21 and 2p23), but are oriented in different directions [5]. The genes can be fused through paracentric inversion, resulting in the development of lung adenocarcinoma [5].

At least 15 variants of oncogenic fusion have been identified to date [3]. All variants contain the basic region of EML4, which is required for the activation of ALK through oligomerization and autophosphorylation [4,5]. Upon activation, ALK plays a primary role in the development of lung adenocarcinoma, as well as colorectal and breast cancers [6]. The fusion variant of the highest frequency is *EML4-ALK* v1 [3], which is present in the H3122 cell model. To date, inhibitors like crizotinib have demonstrated significant antitumor activity against these cancers, but eventual resistance limits the therapeutic benefits [7-9]. Therefore, a new therapeutic strategy is desired for the treatment of lung adenocarcinoma.

Clustered regularly interspaced short palindromic repeats (CRISPR) associated protein called Cas9 has led to tremendous progress in biology [10]. As the functional diversity of CRISPR-Cas protein is further discovered, RNA-targeting is emerging as a powerful mood of CRISPR-based gene engineering. This is because RNA engineering has no risk of unintended everlasting mutation at genomic DNA. In light of that concern, RNA-targeting CRISPR-Cas13 [11-14] system has gained attention in RNA therapeutic which can only target RNA. Recently, two groups successfully downregulated the targeted oncogenes including *KRAS*-G12D mRNA and *TERT* mRNA in human pancreatic and hepatoma cells, respectively using the CRISPR-Cas13a [15,16]. However, there is no evidence about the *EML4-ALK* transcript knockdown in *ALK*-positive lung cancer.

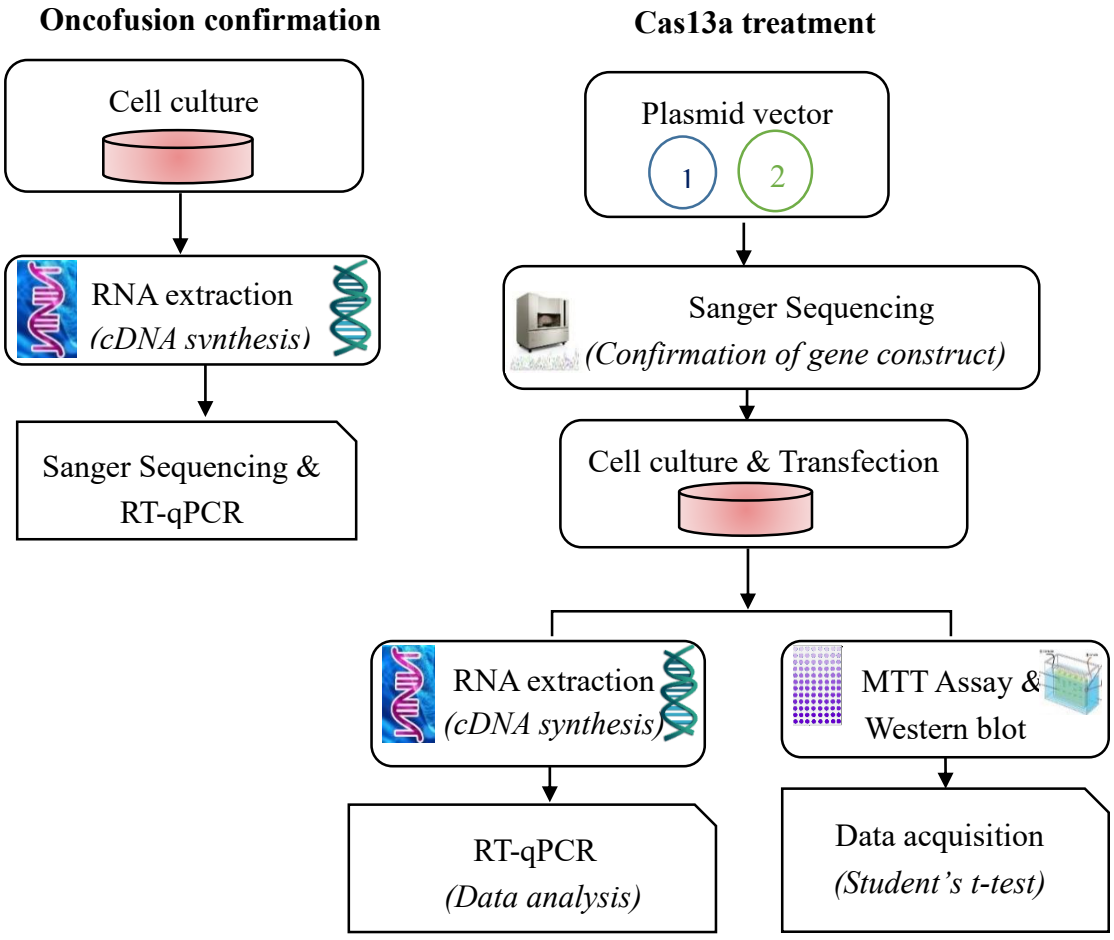
In this study, I used the programmable knockdown capacity of the Cas13a machinery to target *ALK* RNA in *EML4-ALK* fusion transcripts in H3122 and H2228

lung cancer cells, with the goal of decreasing cell viability and providing proof-of-principle of a novel therapeutic approach.

4.2. Materials and Methods

4.2.1. Experimental outline

This study was conducted as following experimental outline:



4.2.2. Expression vector and guide RNA design

Plasmid LwCas13a-msfGFP was obtained from Addgene (plasmid #91902) [11]. The guide RNA was prepared as the reverse complement of the target *EML4-ALK* cDNA site to employ the Cas13a machinery. The following parameters were considered for the design: PFS (protospacer flanking site) on-target and off-target A or T; GC content 40–60%, off-target effects reduced by the BLAST tool in NCBI. Later, the cDNA

sequences of guide RNAs with 5′ *Bam*HI and 3′ *Eco*RI overhangs (Fwd strand 5′ GATCC, 3′ G; Rev strand 5′ AATTC, 3′ G) on both strands were purchased as oligos from Eurofins Genomics, Tokyo, Japan. The oligos were diluted into Milli-Q water at 10 pmol concentration and then annealed in a thermocycler. The reactions were executed in a total volume of 20 µL containing the 10 µL of each forward and reverse oligos. Annealing condition was as follows: 95 °C for 5 min, reduced 7 °C every 5 min to 32 °C, 25 °C for 10 min, and 4 °C for ∞ min. The annealed fragment was later ligated with *Bam*HI- and *Eco*RI-digested pCS2+ expression vector, and cloned into *E. coli* DH5α competent cells. The final construct was endorsed by restriction digestion and Sanger sequencing as well. Sequences of guide RNAs are listed in Table 1 [17].

4.2.3. DNA Sequencing

DNA sequencing was conducted using a 3130XL Genetic Analyzer (Applied Biosystems, Japan) as following my previous protocol [18]. In some cases, 500–800 ng samples of plasmids with specific primers were sent to Eurofins Genomics, Japan company for DNA sequencing.

Table 1. crRNA and guide RNA sequences

Target RNA name	cDNA of guide RNA (5’ to 3’)	Length (bp)	Ref.
Non-targeting guide	TTTACAACGTCGTGACTGGGAAAACCCT	28	[11]
crRNA (or DR)	GATTTAGACTACCCCAAAAACGAAGGG GACTAAAAC	36	[11,19]
	EML4-ALK guide		
ALK guide	CTGGCAGCAATGTCTCGGTGGATGAAGT GG	30	Custom
FBP guide (H3122_v1)	GCTTCCGGCGGTACACTTTAGGTCC	25	Custom
FBP guide (H2228_v3a)	TCCGGCGGTACACTTGGTTGATGATGAC	28	Custom
FBP guide (H2228_v3b)	TTTTCGCGAGTTGACATTTTTGCTTGG	27	Custom

4.2.4. Cell culture and transfection

H3122 and H2228 cells were kindly provided by Dr. Jeffrey A Engelman (Novartis Institutes for Biomedical Research, Cambridge, MA, USA) [20,21]. Both cells were cultured in RPMI 1640 (Roswell Park Memorial Institute Medium 1640, Nacalai Tesque) medium supplemented with 10% FBS in a humidified incubator consisted of 5% CO_2 and 37 °C temperature. Both cells were transfected using the Lipofectamine 3000 Reagent (Invitrogen, Carlsbad, CA, USA). Before the day of transfection, all cells were counted by hemocytometer and seeded at a specific number based on an experimental type.

4.2.5. Total RNA extraction

Firstly, cells were washed rapidly with PBS and lysed by adding 250 µL/well of 24-well plate TRIzol (Invitrogen). The lysed sample was pipetted and then vortexed for 2–3 min for homogenization. Secondly, 50 µL chloroform was added to the reaction and mixed by hand for 15 s. The samples were centrifuged at 12,000 × *g* for 15 min, and the supernatant was taken to a new tube. Next, 125 µL of 100% isopropanol was added to the supernatant, mixed and incubated at room temperature for 10 min. The solution was centrifuged at 12,000 × *g* for 10 min at 4 °C, and the supernatant was thrown out. The pellet was washed with 500 µL of 75% ethanol, and the supernatant was discarded. The pellet was then air-dried for 5–10 min and dissolved in Milli-Q water and stored at -80 °C freezer for further experiment.

4.2.6. cDNA synthesis

cDNA was synthesized in a total volume of 20 µL. Reaction consisted of 100 ng RNA, 0.5 µL of 20 bp oligo (dT) primer, 0.5 µL of random primer, 1 µL of 10 mM dNTPs, and water up to a final volume of 14 µL. After mixing, the sample was incubated at 65 °C for 5 min and cooled on ice for 3 min. Later, 4 µL of 5 × buffer, 1 µL of ReverTra Ace (TOYOBO), and 1 µL of 0.1 M DTT were added to the solution. The mixture was then incubated as follows: 50 °C for 30 min and 55 °C for 30 min for cDNA synthesis, 70 °C for 15 min for inactivation of the enzyme.

4.2.7. SYBR green-based qPCR

SYBR Green-based qPCR was executed in a total volume of 10 μL on an MX 3000P STRATAGENE Multiplex Quantitative PCR System (Agilent Technologies, USA). The reactions contained 3.5 μL water, 0.2 μL of 50 × ROX reference dye II (TaKaRa Bio, Japan), 5 μL TB Green, 0.5 μL cDNA template, and 0.8 μL (10 pmol) of each primer. qPCR was performed as follows: pre-heating at 95 °C for 3 min, 45 cycles of 95 °C for 15 s, 55 °C for 15 s, and 72 °C for 15 s, and a final cycle of 95 °C for 1 min, 55 °C for 30 s, and 95 °C for 30 s. The C_T values were normalized against *GAPDH* expression using the $2^{-\Delta\Delta C}{}_T$ method [22].

4.2.8. MTT assay

H3122 cells and H2228 cells were seeded at 5000 and 9000 cells respectively per well in a 96-well plate. The next day, Cas13a with guide RNA vectors (or crizotinib) were transfected. Cell viability was measured at 48 and 72 h of post-transfection by MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (Invitrogen) reagent according to the manufacturer's protocol. Absorbance was quantified with a ThermoFisher Spectrophotometer 1000 (Molecular Devices, Inc.) at a wavelength of 540 nm.

4.2.9. Control drug

Crizotinib (PF-02341066, Selleck Chemicals) was used as a positive control drug. Crizotinib was dissolved in DMSO and then diluted into water. The IC_{50} value for both cells was determined by MTT assay. The IC50 value is shown in Figure 4B.

4.2.10. Western blot analyses

Cells were lysed in buffer comprising 150 mM NaCl, 0.5% deoxycholic acid sodium salt, 1% NP-40, 0.1% SDS, 25 mM Tris-HCL (pH 7.4), supplemented with protease inhibitor cocktail set III (Wako Pure Chemical Industries, Japan) and PhosphoStop (Roche). Proteins were separated by 9% and 11% SDS-PAGE for blotting with ALK (tALK and pALK) and β-actin antibodies, respectively. Afterward, proteins were transferred onto a PVDF membrane (Millipore, Bedford, MA, USA). Next, ECL Select detection reagent (GE Healthcare, Italy) was used to detect the proteins. The anti-ALK

antibody (Cell Signaling Technology) was used at a dilution of 1:2000. Anti-pALK antibody (Cell Signaling Technology) was used at a dilution of 1:1000. Anti-β–actin antibody (Proteintech) was used at a dilution of 1:2000.

4.2.11. Computational assessment and data analysis

Clinical significance of *ALK* expression and patient prognosis in lung adenocarcinoma (LUAD) was analyzed from the PrognoScan cancer microarray datasets [23]. The genetic fusions of ALK with other partner proteins involved in different types of cancers were achieved from the TCGA PanCancer Atlas through cBioPortal [24,25]. Statistical analyses were executed in MS Excel (version 2016) or GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA, USA). All quantitative data are represented as means ± SEM. A two-tailed Student's t-test was used to estimate the significance (<0.05) of comparisons between the two groups.

4.3. Results

4.3.1. Confirmation of *EML4-ALK* in lung cancer cells

The existence of *EML4-ALK* oncofusion in H3122 and H2228 cells is confirmed by RT-PCR, direct sequencing, and qRT-PCR expression analyses. As shown in Figure 1A–C, the *EML4-ALK* v1 is present in H3122 cells whereas the *EML4-ALK* v3a/b fusion oncogene is in H2228 lung cancer cells. Based on the Sequencing results I confirmed that there is a 33 bp intron sequence between *EML4* and *ALK* in *EML4-ALK* v3b oncofusion. This intron sequence can be a target guide RNA site for the knockdown, which potentially neither destroys the original *EML4* nor *ALK* transcript. The existence of these three oncofusions in both cells was further confirmed by comparing a nonfusion transcript expressed in HEK293T cells (Figure 1D). This expression analysis was performed by qRT-PCR analysis by designing the fusion breakpoint-specific primer. These results suggested that oncofusion is existed in both cells. This further indicates that there is no cross-contamination of *EML4-ALK* variants in the cell lines.

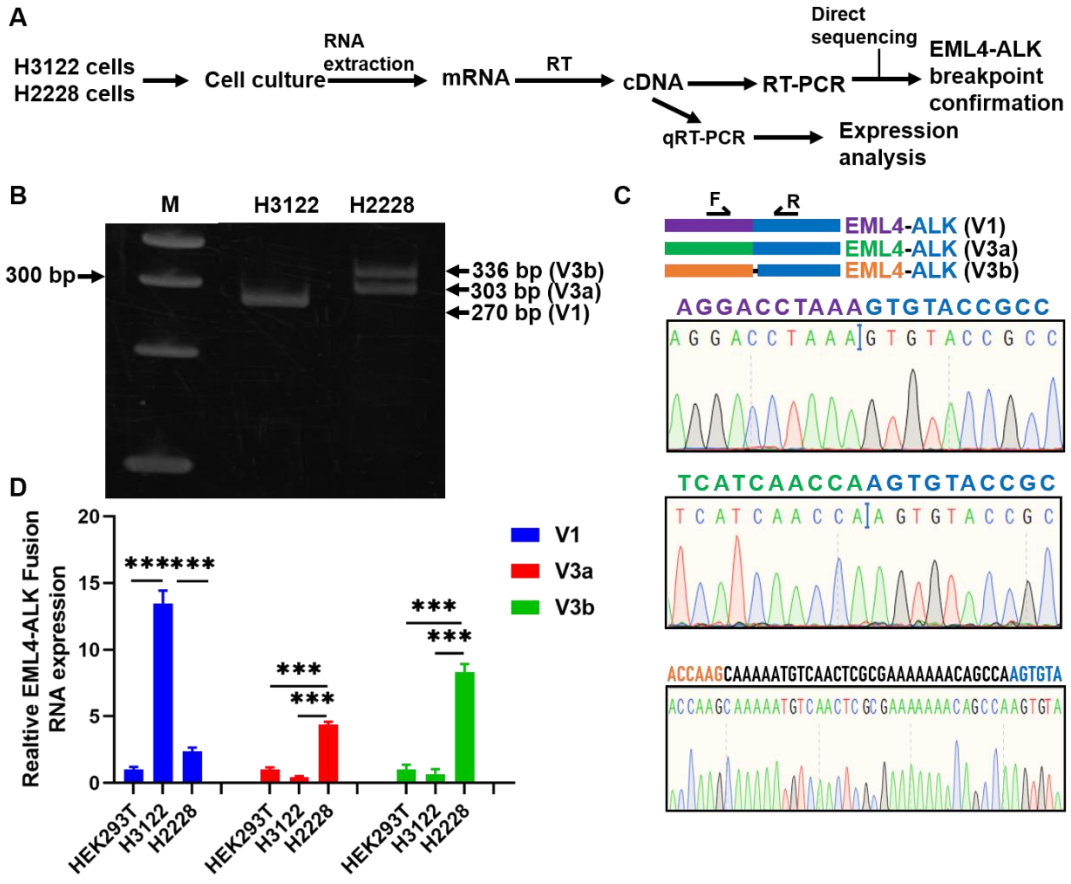


Figure 1. Validation of *EML4-ALK* oncofusion in H3122 and H2228 lung cancer cells. A, Schematic of *EML4-ALK* fusion breakpoint confirmation methods. **B**, RT-PCR analysis using polyacrylamide gel electrophoresis (PAGE); M, 100 bp marker; H3122 cell harboring *EML4-ALK* v1 fusion confirmed at 270 bp RT-PCR product length whereas H2228 cells harboring *EML4-ALK* v3a and *EML4-ALK* v3b fusions identified at 336 bp and 303 bp RT-PCR product lengths. **C**, Confirmation of RT-PCR product using the sequencing analysis; the PAGE bands are isolated and confirmed the fusion breakpoints in all three variants. **D**, Further confirmation of *EML4-ALK* oncofusion using the qRT-PCR expression analysis; HEK293T was used as a nonfusion *ALK* control; *GAPDH* was used as an internal control. qRT-PCR results represent three biological replicates and two technical replicates. Statistical significance was determined by Student's t-test. *** $p < 0.001$.

4.3.2. Expression analysis and clinical significance of *EML4-ALK* transcript

The oncofusion consisted of two genes called *EML4* and *ALK*. Which gene functionally active was investigated by expression analysis using the qRT-PCR method. Based on the results, the *EML4* gene is not differentially expressed in H3122 and H2228 cells as compared to the nonfusion transcript in HEK293T cells (Figure 2A). On the other hand, *ALK* is expressed 100 fold and 80 fold higher in H3122 and H2228 cells respectively as compared to HEK293T cells (Figure 2B). This result initially indicates that the *ALK* plays the important role in tumorigenesis.

Furthermore, I have checked the clinical significance of *ALK* expression in lung adenocarcinoma using the PrognoScan microarray cancer database. As shown in Figure 2C, higher expression of *ALK* is inversely proportional to patient survival rate. Besides, *ALK* not only fuses with *EML4* but also forms oncofusion with other mRNA partners such as *STRN*, *TPM1*, *PPP4R3B*, *ACTG2*, *STK39*, *KCNQ5*, *KCNQ5*, *MALAT1*, *GTF2IRD1*, *GALNT14*, and *SCEL*, which, according to data derived from the TCGA PanCancer Atlas through cBioPortal (10,953 patients/10,967 samples), contribute to the growth of eight other tumor types (Table 2) [17]. Therefore, regulation of *ALK* mRNA signifies a potential therapeutic strategy.

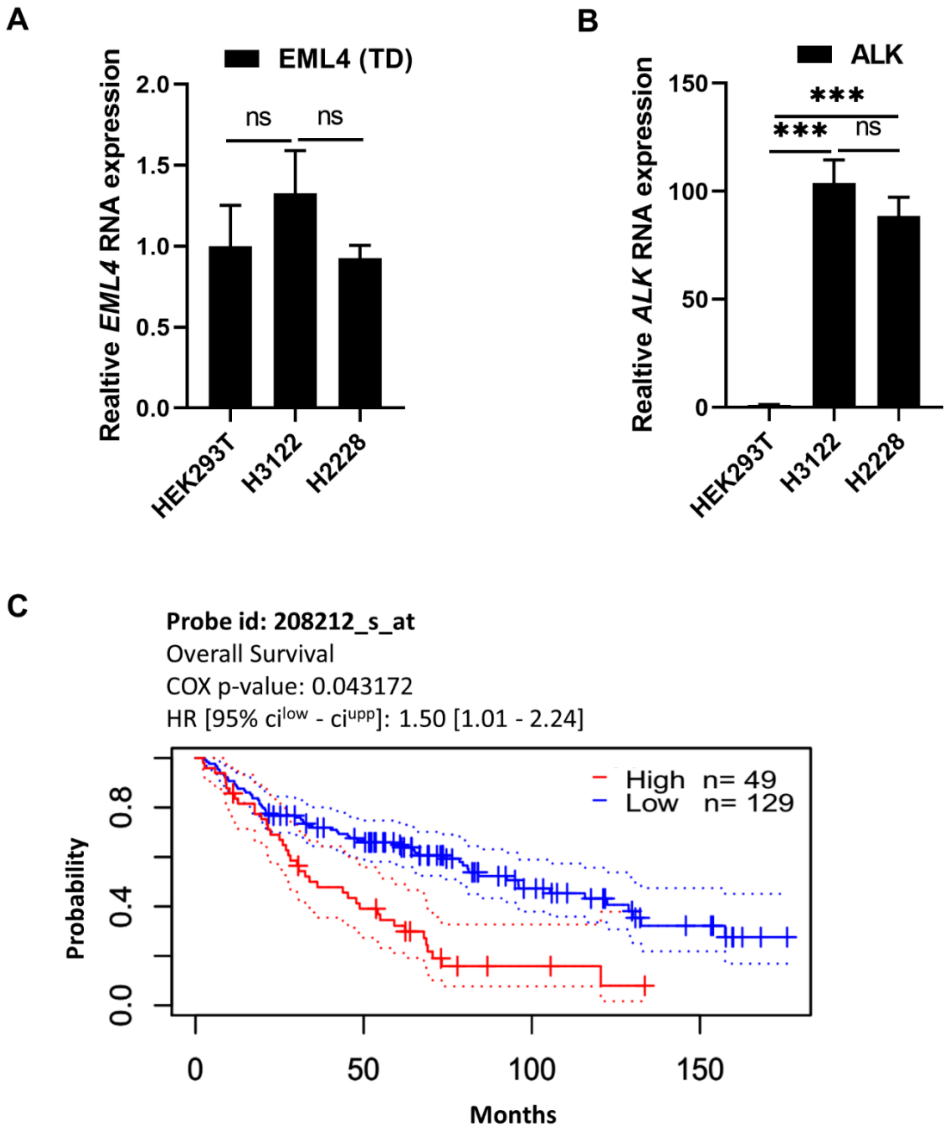


Figure 2. *EML4* and *ALK* mRNA abundance in HEK293T and lung cancer cells, and the clinical significance of *ALK* expression in lung adenocarcinoma (LUAD). **A-B**, mRNA levels (assessed by qRT-PCR) of *EML4* and *ALK* in HEK293T, H3122, and H2228 cells. *GAPDH* was used as an internal control gene. **C**, Kaplan–Meier analysis of patient overall survival in LUAD. Data were derived from the PrognoScan cancer microarray datasets, with COX p-value < 0.05. The prognostic curve distinguishes LUAD patients with high (red) and low (blue) *ALK* expression. The dotted lines designate the maximum and minimum values of the survival average. qRT-PCR results represent three biological replicates and two technical replicates. Statistical significance was determined by Student's t-test. *** $p < 0.001$. ns, not significant.

Table 2. Genetic fusion of *ALK* with other mRNAs involved in different types of cancers, based on data retrieved from the TCGA PanCancer Atlas using cBioPortal (10,953 patients/10,967 samples).

Fusion	Copy number	Cancer type	Patients (*n*)
EML4-ALK	diploid	Lung adenocarcinoma	103
EML4-ALK	gain	Lung adenocarcinoma	49
EML4-ALK	diploid	Papillary renal cell carcinoma	18
EML4-ALK	shallowdel	Papillary thyroid cancer	18
STRN-ALK	shallowdel	Papillary renal cell carcinoma	44
STRN-ALK	deepdel	Papillary thyroid cancer	7
TPM1-ALK	diploid	Bladder urothelial carcinoma	89
PPP4R3B-ALK	shallowdel	Rectal adenocarcinoma	136
ACTG2-ALK	amp	Leiomyosarcoma	51
ALK-STK39	shallowdel	Cutaneous melanoma	171
KCNQ5-ALK		Cutaneous melanoma	506
MALAT1-ALK	diploid	Papillary thyroid cancer	31
GTF2IRD1-ALK	diploid	Papillary thyroid cancer	12
ALK-GALNT14	gain	Uterine serous carcinoma/uterine papillary serous carcinoma	124
ALK-SCEL	gain	Uterine endometrioid carcinoma	156

4.3.3. CRISPR-Cas13a downregulated the target *ALK* in a programmable manner

When I designed the guide RNA at 3′ G or C (on-target) PFS (protospacer flanking site), the target transcript *ALK* was not knocked down as relative to non-target guide RNA (CasControl) (Figure 3A-B). Later, I have found that both 3′ and 5′ A or T PFS showed significant *EML-ALK* transcript knockdown (Figure 3C). This variation may come from the PFS preference, GC content, or RNA secondary structure. Therefore, I considered the A or T PFS at both on-target and off-target PFS sequences, 40–60% GC content, and minimal RNA secondary structure for further experiment.

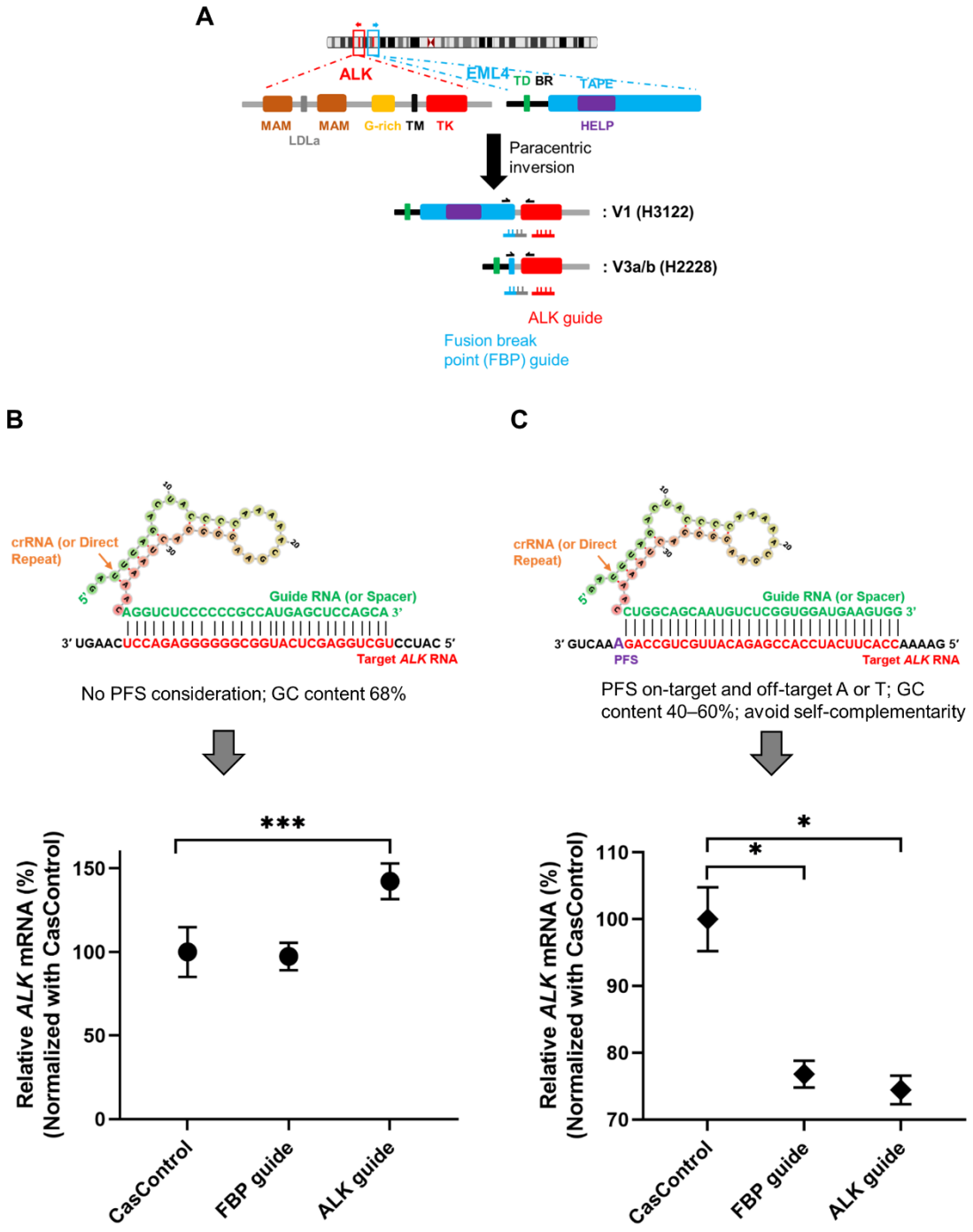


Figure 3. CRISPR-Cas13a downregulated the target *ALK* in a programmable manner. A, Graphical representation of EML4-ALK v1 and EML4-ALK v3a/b oncofusion with the associated functional domain and guide RNA location in H3122 and H2228 cells. Left side: TD, trimerization domain; TAPE, tandem atypical propeller domain; HELP, hydrophobic motif in EML proteins; and TK, tyrosine kinase domain. FBP guide, *EML4-ALK* target guide at the fusion breakpoint (FBP) of *EML4* and *ALK*;

ALK guide, *EML4-ALK* target guide in the TK domain. **B**, Guide RNA design without considering PFS (protospacer flanking site) sequence. Cas13a did not downregulate the *ALK*. CasControl, nontarget guide RNA. **C**, Guide RNA design with 3′- and 5′- A or T PFS sequence while Cas13a downregulated the *ALK* RNA. qRT-PCR results represent three biological replicates and two technical replicates. Statistical significance was determined by Student's t-test. *** $p < 0.001$. * $p < 0.05$.

4.3.4. Programmable knockdown of EML4-ALK v1 oncoprotein in H3122 lung cancer cells

So far, more than 15 variants of *EML4-ALK* fusions have been identified. The highest frequency is *EML4-ALK* v1 (E13; A20) and then *EML4-ALK* v3a/b [3,4]. Hence, I targeted the *ALK* in the *EML4-ALK* v1 fusion oncogene in the H3122 cell line model for human lung adenocarcinoma. Firstly, I observed the cytotoxicity effect of Cas13a in H3122 cells. Based on the MTT assay result there is no significant cytotoxicity of Cas13a compared to transfection reagent control (Figure S1A). Next, I investigated the knockdown of EML4-ALK v1 using the Cas31a. Based on the western blot results as shown in Figure 4A–D, both phosphorylated ALK (pALK) and total ALK (tALK) were >50% and >80% downregulated respectively after LwaCas13a treatment at both guide RNAs relative to CasControl. Which was also similar degree with the positive control drug: crizotinib. This robust knockdown of oncogenic driver ALK resulted in 34–42% inhibition of cell growth (determined by MTT assay) at 48 and 72 h post-transfection relative to CasControl (Figure 4E). Furthermore, I found Cas13a is specific enough to bind target fusion RNA compared to a fusion control (Figure S1B–E). Taken together, the pALK and tALK downregulations, and cell viability results indicated that the CRISPR-Cas13a system effectively and specifically downregulated the fusion oncogene *EML4-ALK* in the H3122 lung cancer cells.

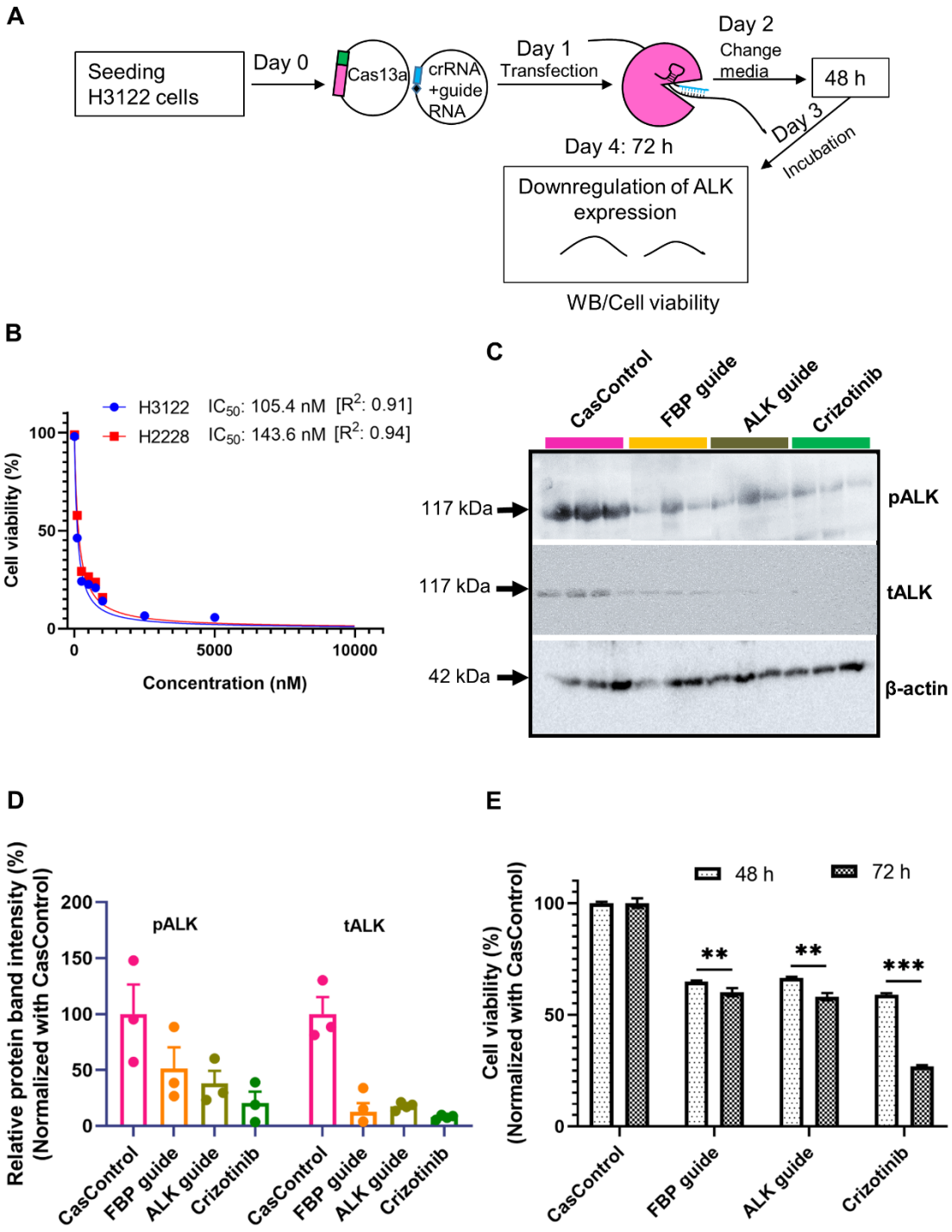


Figure 4. Programmable knockdown of oncogenic EML4-ALK oncoprotein in H3122 lung cancer cells. A, Schematic of CRISPR-Cas13a-mediated therapeutic strategy; WB, western blot. **B**, Determination of IC_{50} values of crizotinib for H3122 and H2228 cells as a positive control. **C**, Western blot analysis showing the representative images of phosphorylated ALK (pALK) and total ALK (tALK) in H3122 lung cancer cells after 72 h post-transfection of Cas13a and specific guide RNA

vectors; CasControl, non-targeting guide RNA; FBP guide, *EML4-ALK* fusion breakpoint target guide; ALK guide, ALK tyrosine kinase target guide; Crizotinib, a control drug; β-actin was used as a loading control. **D**, Graphical depiction of the western blot in Figure C. **E**, Growth-inhibition (measured by MTT assay) response to 48 and 72 h of Cas13a treatment in H3122 cells. All data represent the mean values ± standard error of the mean (SEM) from three or four individual experiments. *** $p < 0.001$. ** $p < 0.01$.

4.3.5. Programmable downregulation of EML4-ALK v3a/b oncoprotein expression in H2228 lung cancer cells

To check the reproducibility of *EML4-ALK* knockdown using Cas13a, I targeted later the *EML4-ALK* v3a/b oncofusion gene expressing in H2228 cells. Where I found a similar degree of pALK and tALK protein knockdown based on western blot analysis (Figure 5A–C). The growth inhibition response after Cas13a treatment was also considerable at both 48 and 72 h time points according to the cell viability results (Figure 5D). These observations indicated that Cas13a downregulated the EML4-ALK v3a/b fusion protein in H2228 lung cancer cells.

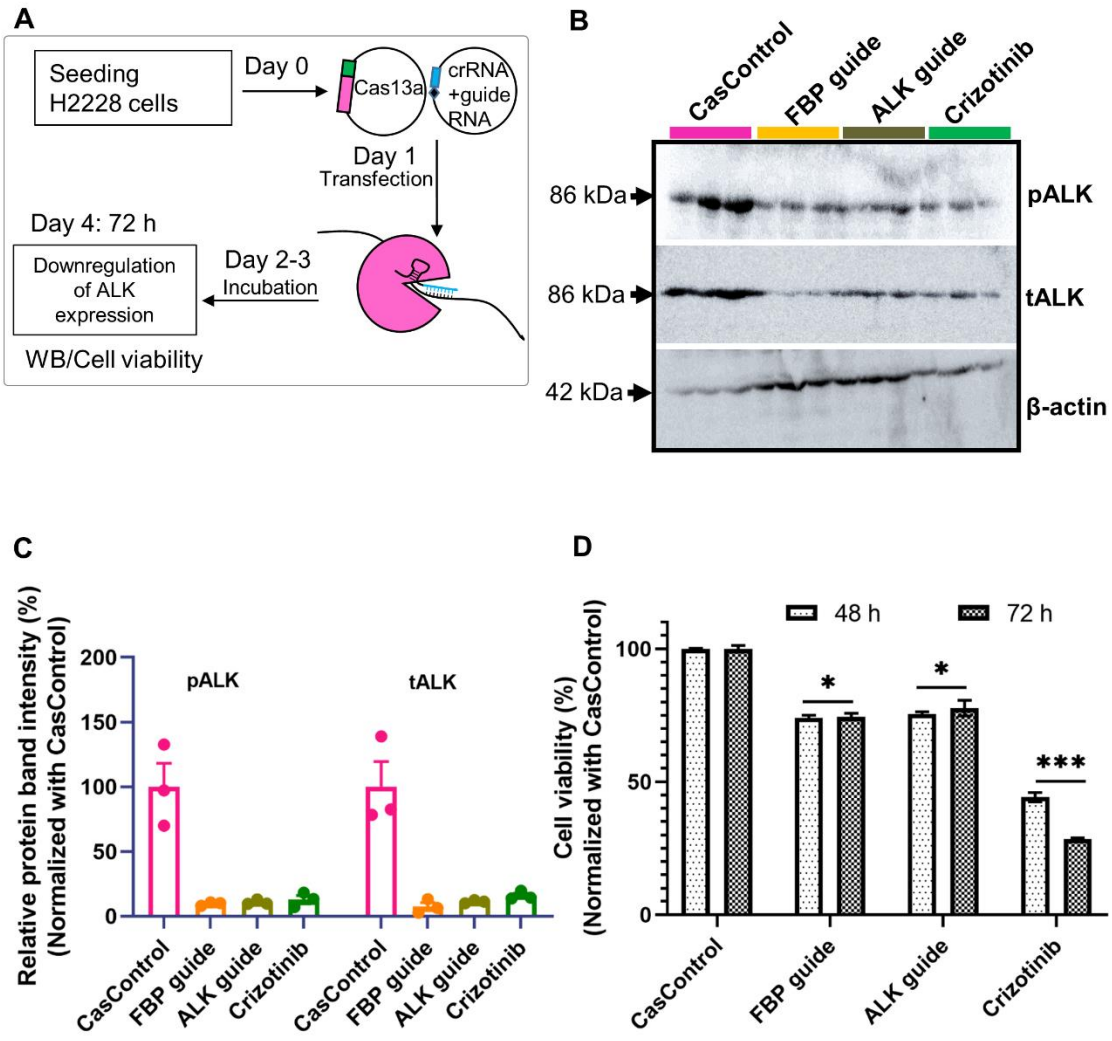


Figure 5. Programmable knockdown of oncogenic EML4-ALK oncoprotein in H2228 lung cancer cells. A, Schematic of CRISPR-Cas13a-mediated therapeutic strategy; WB, western blot. **B**, Western blot analysis showing the representative images of phosphorylated ALK (pALK) and total ALK (tALK) in H2228 lung cancer cells after 72 h post-transfection of Cas13a and specific guide RNA vectors; CasControl, non-targeting guide RNA; FBP guide, *EML4-ALK* fusion breakpoint target guide; ALK guide, ALK tyrosine kinase target guide; Crizotinib, a control drug; β-actin was used as a loading control. **C**, Graphical depiction of the western blot in Figures B. **D**, Growth-inhibition (measured by MTT assay) response to 48 and 72 h of Cas13a treatment in H2228 cells. All data represent the mean values ± standard error of the mean (SEM) from three individual experiments. *** $p < 0.001$. * $p < 0.05$.

4.4. Discussion

In this study, the *EML4-ALK* fusion variants were confirmed in H3122 and H2228 cells using RT-PCR, Sanger sequencing, and qRT-PCR analyses. The identification and expression of the target gene are essential before a knockdown study [26]. Besides, sometimes cell lines might have unintended contamination with other cells due to improper handling during the cell culture. Hence, I have confirmed both fusion breakpoint sequences and their expression. The results indicated that both cell lines were in perfect condition.

Information from the PrognoScan database indicated that higher expression of *ALK* is significantly correlated with lower patient survival in lung adenocarcinoma (Figure 2C), suggesting that if we could decrease *ALK* expression, patient survival rate might be improved. This is because, inhibition of EML4-ALK with an ALK tyrosine kinase inhibitor like TAE684, or via knockdown using RNAi resulted in the retraction of downstream signaling and induction of apoptosis primarily through RAS-MAPK signaling in *EML4-ALK*-positive cells [27-29] However, cancer eventually develops resistance to inhibitors [27,28], thereby limiting clinical benefits. On the other hand, RNAi has a higher off-target effect, and limited accessibility to nuclear transcript [11,19]. Therefore, I used the Cas13a-mediated RNA knockdown approach to regulate the expression of the *EML4-ALK* oncofusion transcript in this study.

Although initial data reported that LwaCas13a contains 3′ non-G PFS preference in *in vitro* assay but later the same author proved that it does not exhibit any PFS restriction for target RNA knockdown in human cells [11,19,30]. However, I found 3′ A or T PFS showed the highest *ALK* knockdown (Figure 3). Several factors including PFS preference, GC and Tm values, RNA secondary structure might contribute to target RNA-guide RNA hybridization. Therefore, further study is needed to conclude this discrepancy about whether A or T is preferable PFS for *ALK* knockdown or not.

Katayama *et al.* obtained robust ALK protein knockdown and around 50% cell viability using siRNA in H3122 cells and H3122 CR (crizotinib resistance) cells [31]. At this point, I confirmed 50–70% pALK and over 80% tALK knockdown based on western blot analyses and up to 40% cell growth inhibition (determined by MTT assay

in H3122 cells) (Figure 4–5). In contrast, one study succeeded to reduce viability by over 60% using the siRNA technique [32]. The initial response of crizotinib to cell viability (69–74%) is greater than Cas13a-mediated RNA knockdown (34–42%) in this study (Figure 6A) and other TKI studies [8,27,32]. Previous studies have shown that the autophosphorylation of ALK activates downstream signaling [3,4]. Based on my pALK results, I speculate that the reduced phosphorylation can diminish the activation of downstream signaling and subsequent cell survival. The efficiency of EML4-ALK knockdown and reduction in cell viability that I observed in this study indicated that Cas13a-mediated targeting of the *EML4-ALK* transcript not only knocked down ALK expression but also changed cell physiology, as reflected by reductions in cell viability at two different time points (Figure 4–5). This leads me to conclude that RNA-mediated gene knockdown using CRISPR-Cas13a is effective, programmable, low-cost, and long-lasting.

Despite the noteworthy advantages of Cas13a, it has some shortcomings. Cas13a still retains some off-target effects [11] and PFS preference [26], and has lower efficiency than counterparts such as Cas13d [13], but it is still superior to RNAi from the standpoints of specificity and efficiency [11,12]. Off-target effects must be circumvented, as even a single nonspecific knockdown could change an important function in the target cells. Hence, alternative RNA-guided CRISPR or other RNA cleavage systems are still desirable to resolve these issues and empower the development of therapeutics based on this technology.

In this study, I have not yet investigated apoptosis and cell proliferation. I will conduct these phenomena to explain whether *ALK* knockdown leads to anti-proliferative or apoptotic properties or both of them near future. Even though ALK is an orphan receptor, knockdown can decline the expression of ordinary ALK (nonfusion) functions. My fusion-guided RNA (FBP guide) can solve this concern because it only binds to the fusion breakpoint of the *EML4-ALK* transcript. A more lucid guide design targeted at inhibiting expression of the *EML4-ALK* mRNA has potential as an RNA-based therapy.

4.5. Concluding remark

In summary, the CRISPR-Cas13a system significantly downregulated total ALK and phosphorylated ALK in both EML4-ALK-positive lung cancer cells (H3122 and H2228). The knockdown of the oncofusion gene *EML4-ALK* resulted in alteration of cell physiology, reflected by a decrease in viability. Overall the obtained data demonstrated that *EML4-ALK* transcript knockdown using CRISPR-Cas13a inhibited *EML4-ALK*-positive lung cancer cell survival. This study offers a basis for an effective therapeutic strategy for treating patients with *ALK*-positive lung adenocarcinoma using the CRISPR-Cas13a system.

4.6. Bibliography

1. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* **2008**, *83*, 584-594, doi:10.4065/83.5.584.
2. Saifullah, S.; Sakari, M.; Suzuki, T.; Yano, S.; Tsukahara, T. P28-11 The CRISPR-Cas13a gene-editing system underlies a potential therapeutic strategy for EML4-ALK-positive lung cancer cells. *Annals of Oncology* **2021**, *32*, S347, doi:https://doi.org/10.1016/j.annonc.2021.05.737.
3. Bayliss, R.; Choi, J.; Fennell, D.A.; Fry, A.M.; Richards, M.W. Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. *Cellular and molecular life sciences : CMLS* **2016**, *73*, 1209-1224, doi:10.1007/s00018-015-2117-6.
4. Sabir, S.R.; Yeoh, S.; Jackson, G.; Bayliss, R. EML4-ALK Variants: Biological and Molecular Properties, and the Implications for Patients. *Cancers (Basel)* **2017**, *9*, 118, doi:10.3390/cancers9090118.
5. Soda, M.; Choi, Y.L.; Enomoto, M.; Takada, S.; Yamashita, Y.; Ishikawa, S.; Fujiwara, S.-i.; Watanabe, H.; Kurashina, K.; Hatanaka, H., et al. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **2007**, *448*, 561-566, doi:10.1038/nature05945.
6. Lin, E.; Li, L.; Guan, Y.; Soriano, R.; Rivers, C.S.; Mohan, S.; Pandita, A.; Tang, J.; Modrusan, Z. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Molecular cancer research : MCR* **2009**, *7*, 1466-1476, doi:10.1158/1541-7786.mcr-08-0522.
7. Golding, B.; Luu, A.; Jones, R.; Viloria-Petit, A.M. The function and therapeutic targeting of anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC). *Molecular cancer* **2018**, *17*, 52, doi:10.1186/s12943-018-0810-4.

8. Katayama, R.; Khan, T.M.; Benes, C.; Lifshits, E.; Ebi, H.; Rivera, V.M.; Shakespeare, W.C.; Iafrate, A.J.; Engelman, J.A.; Shaw, A.T. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 7535-7540, doi:10.1073/pnas.1019559108.
9. Zou, H.Y.; Li, Q.; Lee, J.H.; Arango, M.E.; McDonnell, S.R.; Yamazaki, S.; Koudriakova, T.B.; Alton, G.; Cui, J.J.; Kung, P.P., et al. An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer research* **2007**, *67*, 4408-4417, doi:10.1158/0008-5472.can-06-4443.
10. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **2012**, *337*, 816-821, doi:10.1126/science.1225829.
11. Abudayyeh, O.O.; Gootenberg, J.S.; Essletzbichler, P.; Han, S.; Joung, J.; Belanto, J.J.; Verdine, V.; Cox, D.B.T.; Kellner, M.J.; Regev, A., et al. RNA targeting with CRISPR-Cas13. *Nature* **2017**, *550*, 280-284, doi:10.1038/nature24049.
12. Abudayyeh, O.O.; Gootenberg, J.S.; Essletzbichler, P.; Han, S.; Joung, J.; Belanto, J.J.; Verdine, V.; Cox, D.B.T.; Kellner, M.J.; Regev, A., et al. RNA targeting with CRISPR–Cas13. *Nature* **2017**, *550*, 280-284, doi:10.1038/nature24049.
13. Konermann, S.; Lotfy, P.; Brideau, N.J.; Oki, J.; Shokhirev, M.N.; Hsu, P.D. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **2018**, *173*, 665-676.e614, doi:10.1016/j.cell.2018.02.033.
14. Tambe, A.; East-Seletsky, A.; Knott, G.J.; Doudna, J.A.; O'Connell, M.R. RNA Binding and HEPN-Nuclease Activation Are Decoupled in CRISPR-Cas13a. *Cell reports* **2018**, *24*, 1025-1036, doi:10.1016/j.celrep.2018.06.105.
15. Gao, J.; Luo, T.; Lin, N.; Zhang, S.; Wang, J. A New Tool for CRISPR-Cas13a-Based Cancer Gene Therapy. *Molecular Therapy - Oncolytics* **2020**, *19*, 79-92, doi:10.1016/j.omto.2020.09.004.
16. Zhao, X.; Liu, L.; Lang, J.; Cheng, K.; Wang, Y.; Li, X.; Shi, J.; Wang, Y.; Nie, G. A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment. *Cancer letters* **2018**, *431*, 171-181, doi:10.1016/j.canlet.2018.05.042.
17. Saifullah; Sakari, M.; Suzuki, T.; Yano, S.; Tsukahara, T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. *International Journal of Molecular Sciences* **2020**, *21*, 8904, doi:https://doi.org/10.3390/ijms21238904.

18. Saifullah; Tsukahara, T. Genotyping of single nucleotide polymorphisms using the SNP-RFLP method. *Bioscience trends* **2018**, *12*, 240-246, doi:10.5582/bst.2018.01102.
19. Smargon, A.A.; Cox, D.B.T.; Pyzocha, N.K.; Zheng, K.; Slaymaker, I.M.; Gootenberg, J.S.; Abudayyeh, O.A.; Essletzbichler, P.; Shmakov, S.; Makarova, K.S., et al. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular Cell* **2017**, *65*, 618-630.e617, doi:10.1016/j.molcel.2016.12.023.
20. Fukuda, K.; Takeuchi, S.; Arai, S.; Katayama, R. Epithelial-to-Mesenchymal Transition Is a Mechanism of ALK Inhibitor Resistance in Lung Cancer Independent of ALK Mutation Status. **2019**, *79*, 1658-1670, doi:10.1158/0008-5472.can-18-2052.
21. Koivunen, J.P.; Mermel, C.; Zejnullahu, K.; Murphy, C.; Lifshits, E.; Holmes, A.J.; Choi, H.G.; Kim, J.; Chiang, D.; Thomas, R., et al. EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2008**, *14*, 4275-4283, doi:10.1158/1078-0432.ccr-08-0168.
22. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **2001**, *25*, 402-408, doi:10.1006/meth.2001.1262.
23. Mizuno, H.; Kitada, K.; Nakai, K.; Sarai, A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Medical Genomics* **2009**, *2*, 18, doi:10.1186/1755-8794-2-18.
24. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E., et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2012**, *2*, 401-404, doi:10.1158/2159-8290.cd-12-0095.
25. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E., et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **2013**, *6*, pl1, doi:10.1126/scisignal.2004088.
26. Han, H. RNA Interference to Knock Down Gene Expression. *Methods in molecular biology (Clifton, N.J.)* **2018**, *1706*, 293-302, doi:10.1007/978-1-4939-7471-9_16.
27. Abudayyeh, O.O.; Gootenberg, J.S.; Konermann, S.; Joung, J.; Slaymaker, I.M.; Cox, D.B.T.; Shmakov, S.; Makarova, K.S.; Semenova, E.; Minakhin, L., et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016**, *353*, aaf5573, doi:10.1126/science.aaf5573.
28. Katayama, R.; Shaw, A.T.; Khan, T.M.; Mino-Kenudson, M.; Solomon, B.J.; Halmos, B.; Jessop, N.A.; Wain, J.C.; Yeo, A.T.; Benes, C., et al. Mechanisms

of acquired crizotinib resistance in ALK-rearranged lung Cancers. *Science translational medicine* **2012**, *4*, 120ra117, doi:10.1126/scitranslmed.3003316.

29. Takezawa, K.; Okamoto, I.; Nishio, K.; Jänne, P.A.; Nakagawa, K. Role of ERK-BIM and STAT3-survivin signaling pathways in ALK inhibitor-induced apoptosis in EML4-ALK-positive lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2011**, *17*, 2140-2148, doi:10.1158/1078-0432.ccr-10-2798.
30. Gootenberg, J.S.; Abudayyeh, O.O. Nucleic acid detection with CRISPR-Cas13a/C2c2. **2017**, *356*, 438-442, doi:10.1126/science.aam9321.
31. Katayama, R.; Khan, T.M.; Benes, C.; Lifshits, E.; Ebi, H.; Rivera, V.M.; Shakespeare, W.C.; Iafrate, A.J.; Engelman, J.A.; Shaw, A.T. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proceedings of the National Academy of Sciences* **2011**, *108*, 7535-7540, doi:10.1073/pnas.1019559108.
32. Hrustanovic, G.; Olivas, V.; Pazarentzos, E.; Tulpule, A.; Asthana, S.; Blakely, C.M.; Okimoto, R.A.; Lin, L.; Neel, D.S.; Sabnis, A., et al. RAS-MAPK dependence underlies a rational polytherapy strategy in EML4-ALK–positive lung cancer. *Nature Medicine* **2015**, *21*, 1038-1047, doi:10.1038/nm.3930.

4.7. Supplementary

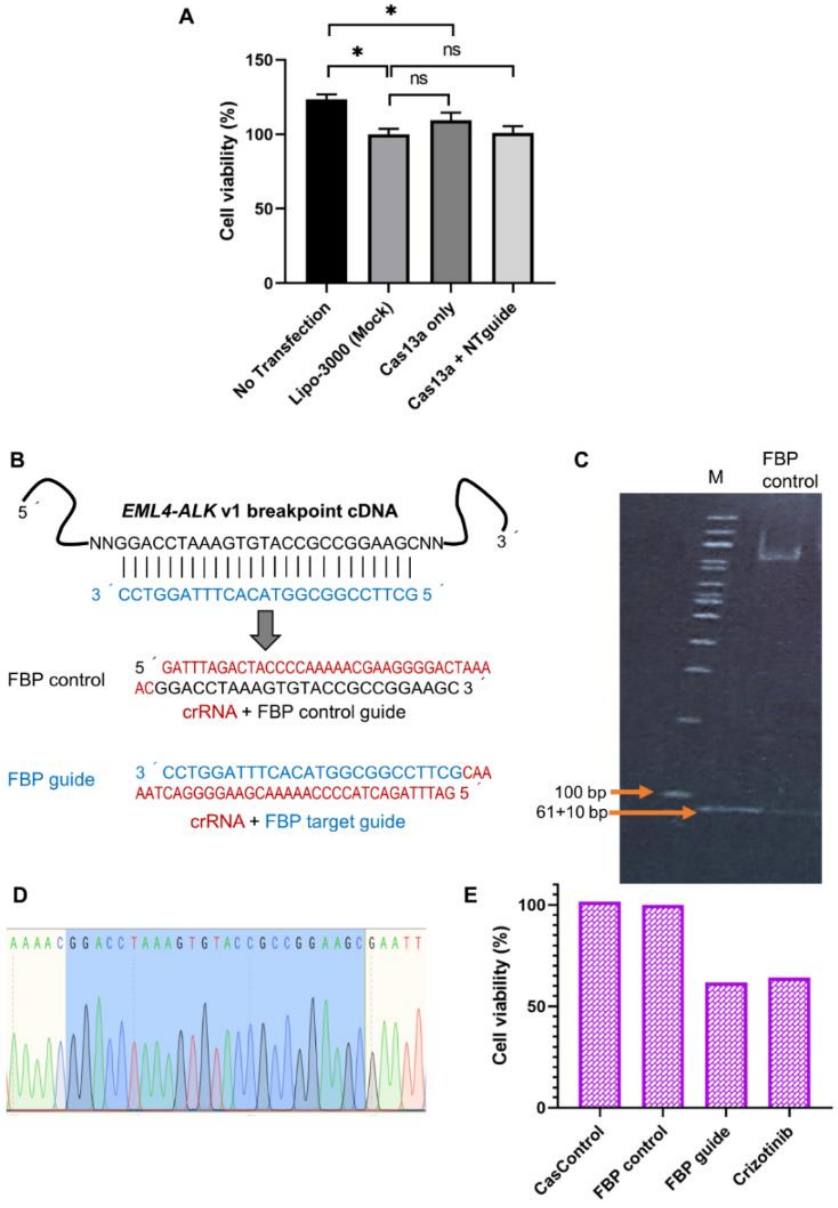


Figure S1. Cytotoxicity and specificity of Cas13a in H3122 cells. A, Cytotoxicity (determined by MTT assay) analysis of Cas13a after 48h of post-transfection using Lipofectamine 3000 (Invitrogen) reagent; Lipo-3000 (mock), lipofectamine-3000 reagent control; Cas13a only, Cas13a vector expression; Cas13a + NT guide, Cas13a vector and nontarget guide vector. **B**, Strategy of specificity test using fusion breakpoint (FBP) guide design. **C**–**D**, Confirmation of FBP control guide insertion into pCS2+ only vector using restriction digestion (*Bam*HI and *Eco*RI) and Sanger sequencing respectively. **E**, Cell viability (determined by MTT assay) result of FBP control and FBP target guide.

Chapter V

Final discussion

Graduate School of Advanced Science and Technology
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (JAIST), JAPAN

Chapter V
Final discussion

5.1. Summary

In short, this dissertation presents an effective and programmable RNA knockdown method using the CRISPR-Cas13a and addresses the feasibility of the system in marker gene knockdown and *ALK*-positive cancer therapy application.

Chapter II identified the clinical significance, genetic alteration, and co-expressive coding and non-coding genes of *ALK* transcript expression in cancer development using integrative bioinformatics. In this study, I found that the deregulated expression of *ALK* significantly altered the patient survivability in six different cancers. This suggested that tight regulation of *ALK* expression is indispensable. Moreover, I have identified that *ALK* is exclusively co-expressed with 17 coding genes and 19 miRNAs. Furthermore, I investigated the cellular pathways involved in ALK-mediated cell proliferation, migration, and survival. Collectively, the obtained data demonstrated that *ALK* with identified coding and non-coding genes are associated with multi-cancer development. In conclusion, precision medicine is required for patients with cancer harboring abnormal *ALK*. The entire work is shown in Figure 1 with integrative bioinformatics tools.

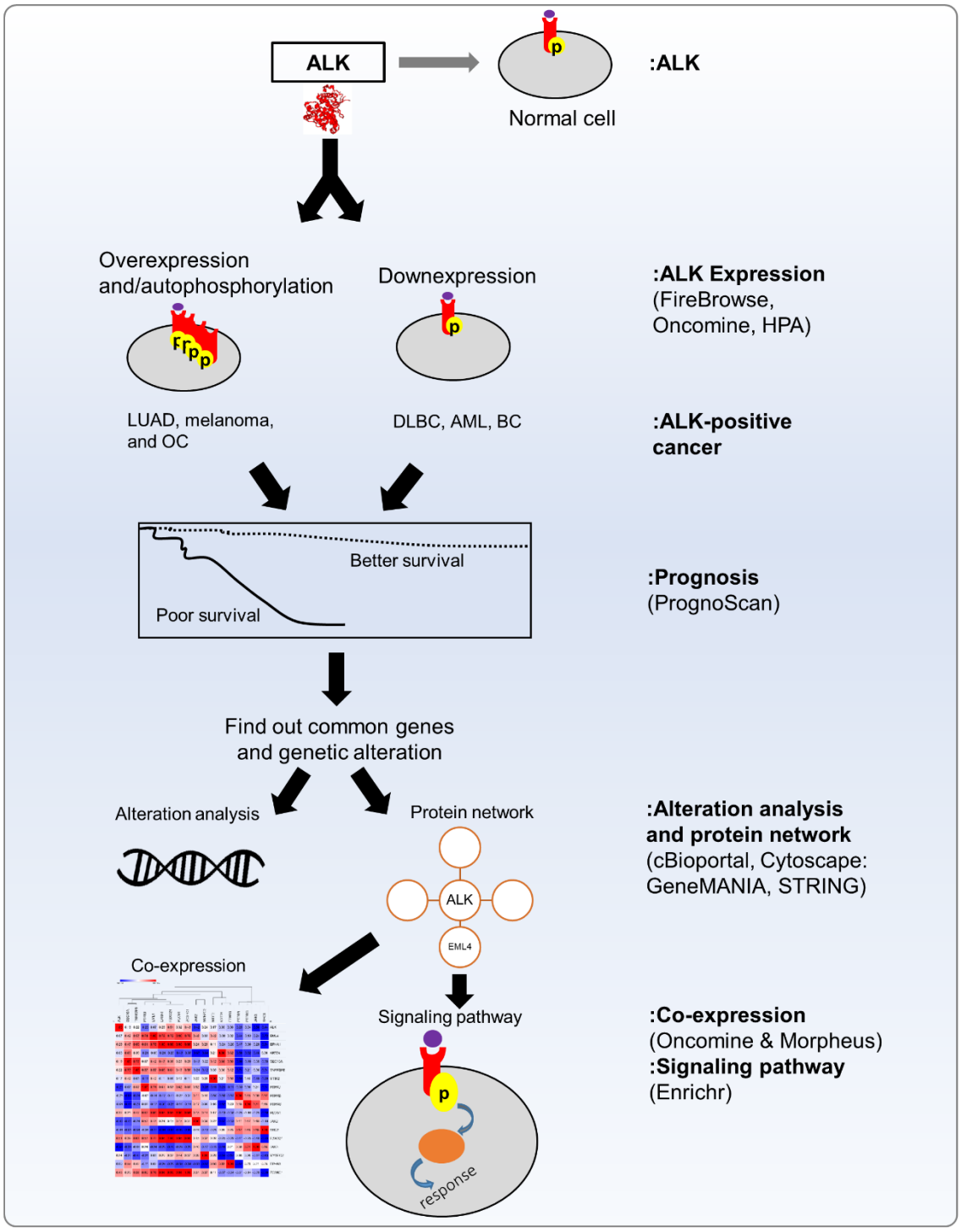


Figure 1. Graphical representation of entire work for Chapter II

In **Chapter III**, I employed the CRISPR-Cas13a system and optimized the conditions to achieve programmable knockdown of marker genes including firefly luciferase and mCherry in HEK293T cells. I have shown an in-depth analysis of using CRISPR-Cas13a technology in place of RNAi technology as RNAi is noted to have frequent off-target effects, low efficiency, and limited accessibility of nuclear transcripts. I have conducted several experiments to determine Cas13a mRNA and guide RNA stability, effective Cas13a/guide RNA concentrations, and optimal guide RNA-crRNA orientation, length, and mismatch tolerance. I have found that it is preferable to titrate the guide RNA vector concentration based on the molar ratio rather than the amount of vector. All initial experiments are carried out in HEK293T cells knocking down firefly luciferase activity through mRNA targeting. I have established optimal targeting parameters in this system before also testing the knockdown efficiency of mCherry in HEK293T cells.

In eukaryotic and prokaryotic systems, the CRISPR-Cas9 system has been very useful for DNA editing with precision but is unable to directly edit RNA, thus, the CRISPR-Cas13a system was recently developed and has been shown to effectively knockdown RNA expression in mammalian cells. In **Chapter IV**, as a proof-of-principle in gene therapy, I have used the optimized protocols to successfully knock down an oncofusion protein EML4-ALK in the human lung adenocarcinoma H3122 and H2228 cell lines, resulted in reduced cell viability in these cells. I found that tyrosine phosphorylation was significantly reduced, which is one of the important features for activating the downstream signaling for cellular overgrowth in lung adenocarcinoma. Taken together, I have demonstrated the feasibility of Cas13a as RNA-targeted gene therapy in a disease model human lung cancer which can be drawn as depicted in Figure 2.

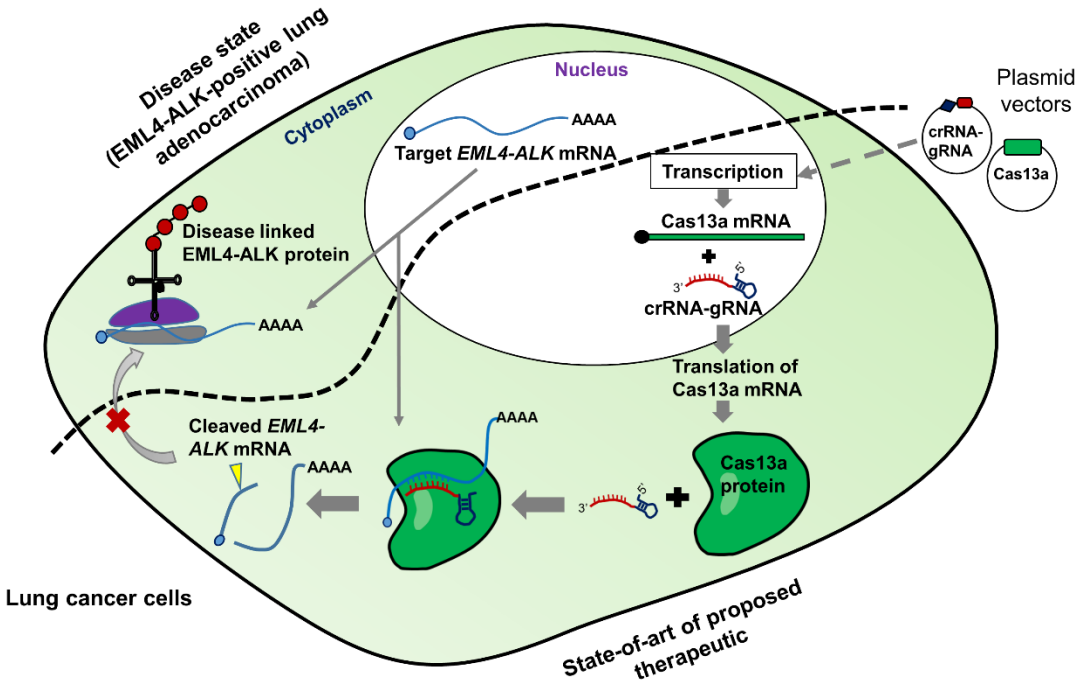


Figure 2. State-of-art of the proposed therapeutic for EML4-ALK-positive lung cancer cells described in Chapter IV. The black dash line separates the disease state of EML4-ALK-positive lung cancer and my proposed therapeutic strategy using Cas13a ribonuclease. In the disease state, the oncofusion *EML4-ALK* transcript is being transcribed in the nucleus and then undergoes translation into the oncoprotein receptor (diseases associated) in the cytoplasm. On the other hand, the therapeutic state comprises the internalization of vectors for Cas13a and guide RNA. The cellular system then transcribes and translates the vector's DNA into functionally active Cas13a endoribonuclease. With the guidance of guide RNA, the Cas13a cleaves the target *EML4-ALK* transcript that ultimately hinders the translation of diseased-linked EML4-ALK oncoprotein receptor.

The new data I have contributed to the molecular biology field including i) tight regulation is needed for ALK expression ii) I developed the finest scheme for CRISPR-Cas13a functionality, iii) specifically, it is better to titrate the guide RNA vector amount based on the molar ratio rather than the amount of vector, iv), Importantly, as a proof of principle of novel therapeutic approach, I used the optimized protocols to magnificently knock down an oncofusion protein called EML4-ALK in the human lung adenocarcinoma H3122 and H2228 cell lines.

5.2. Future direction

Based on the obtained data from this dissertation, I believe that CRISPR-Cas13a mediated RNA targeting-based gene therapy system will be a next-generation treatment strategy especially for incurable diseases and undruggable target diseases. I propose that the Cas13a is a more effective, low-cost, and programmable option for RNA knockdown as opposed to RNAi technology. Furthermore, based on the recent successful approval of RNAi in gene therapy and the recent addition of clinical studies of Cas9 effector, I speculate that this Cas13 ribonuclease could be used as RNA-targeted gene therapy. Also, I believe this system will appear soon under clinical trial. However, the following issue should be considered to get the extraordinary clinical benefit of this CRISPR-Cas effector:

Guide design and PFS preference

One of the important criteria for successful programmable RNA knockdown using Cas13a is guide design. As I have shown in Chapter III and IV that we should consider the length (24–30 bp), mismatch (no mismatch allowed), GC content (40–60%), Tm value (55–65 °C), secondary structure (should be less complex), and minimal similarity with unintended transcripts (<40%), and PFS preference. Cas13a and Cas13b have

shown PFS preference but a recent study showed that Cas13d does not retain PFS preference. Therefore, Cas13d can be an option in therapeutic.

Cancer therapy

In the cancer study, Cas13a mediated gene therapy would be a powerful gene therapy technique only when a single oncogene is solely responsible for the development of a cancer. If multiple genes are associated with the development of a cancer, then this method may not be a suitable option. In my study, the growth inhibition response of H2228 cells after Cas13a treatment is relatively lower than H3122 cells. I anticipate that as the HELP domain is absent in EML4-ALK v3a/b expressing in H2228 cells, the HELP domain or other unknown gene may involve in cancer recurrence.

Immunization

As other gene therapy tools including RNAi or Cas9 are challenging the immunization issues, this system also has a concern for cellular immunization. In light of this consideration, an immunosuppressive method or alternative RNA knockdown technology needs to develop. For instance, if we could use our cellular endonuclease in a programmable manner and effective way, that can be a great alternative.

Collateral cleavage

There are contradictory findings for the collateral activity of Cas13a. The group who identified the Cas13a initially found that Cas13a shows collateral activity in bacteria in an *in vitro* study. Interestingly, the same group later interrogated that the collateral activity is absent in human and plant cells. However, another group found that Cas13a has shown collateral cleavage in glioma cells. Likewise, the Cas13b and Cas13d also showed collateral cleavage function. Further studies are needed to conclude this disputed issue. This collateral cleavage is important because it exerts the off-target effect which is not desirable for molecular therapy except for cancer therapy.

Achievements

❖ **Original Research Article**

1. **Saifullah**, Sakari M, Suzuki T, Yano S, Tsukahara T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. Int J Mol Sci. 2020 Nov 24;21(23):8904. doi: 10.3390/ijms21238904. PMID: 33255340; PMCID: PMC7727695.
2. **Saifullah**, and Tsukahara T. Integrated analysis of *ALK* higher expression in human cancer and downregulation in LUAD using RNA molecular scissor (2021) (*Manuscript prepered for Lung Cancer*).
3. **Saifullah**, and Tsukahara T. Integrated analysis of the clinical consequence and associated gene expression of *ALK* in *ALK*-positive human cancers (2021) (*Manuscript prepered for Heliyon*).

❖ **Abstract publication (International Conference)**

1. **Saifullah**, Sakari M, Suzuki T, Yano S, Tsukahara T. The CRISPR-Cas13a gene editing system underlies a potential therapeutic strategy for *EML4-ALK*-positive lung cancer cells. JSMO February 2021 virtual congress, **Annals of Oncology**, Volume 32, Supplement 4, Page S347, 2021, doi: 10.1016/j.annonc.2021.05.737.
2. **Saifullah**, Sakari M, Suzuki T, Tsukahara T. RNA guided CRISPR-Cas protein downregulates the oncogenic driver ALK expression in human lung cancer cell. European Society of Human Genetics Conference. Virtual; August 28–31, 2021. (The abstract will publish in the ‘‘European Journal of Human Genetics (Nature) journal’’).
3. **Saifullah**, Sakari M, Tsukahara T. Programmable RNA knockdown using CRISPR-Cas13a and therapeutic target for ALK-rearranged lung cancer. MBSJ 2020 (online)
4. **Saifullah**, Sakari M, Tsukahara T. Application of CRISPR-Cas13 machinery in disease treatment. MBSJ 2019. Fukuoka, Kyushu. (Poster)

❖ **Awards**

- ✓ Ministry of Education Culture Sports, Science and Technology (MEXT), Japan (Monbukagakusho Scholarship) 2018-2021 for doctoral study.
- ✓ JAIST Research Grant-2020, for attending JSMO February 2021 conference.

Achievements

(Continued)

The credentials based on master's research are given below:

❖ **Original Research Article**

- Saifullah, Fuke S, Nagasawa H, Tsukahara T. Single nucleotide recognition using a probes-on-carrier DNA chip. BioTechniques (2019). (DOI: 10.2144/btn-2018-0088 PMID: 30744407)

- Saifullah, Tsukahara T. Genotyping of single nucleotide polymorphisms using the SNP-RFLP method. BioScience Trends 12(3), 240-246 (2018). (DOI: 10.5582/bst.2018.01102, PMID: 30012914)

❖ **International Conference (Abstract)**

- Saifullah, Nagasawa H, Tsukahara T. Proposing a probes-on-carrier based oligo-DNA microarray method for SNPs detection. JAIST Japan-India Symposium on Materials Science (JISMS), March 5-6, 2018 (Poster).

❖ **Awards**

- **Excellent Student Award-2018** from School of Materials Science, JAIST
- Ministry of Education Culture Sports, Science and Technology (MEXT), Japan (Monbukagakusho Scholarship) 2016-2018 for Master of Science study.

Acknowledgements

First and foremost, I would like to express my profound gratitude to the Supreme being, creator **Allah**, whose mercy keeps me alive and enable me to complete this thesis successfully.

My regard, gratitude and appreciation go to my supervisor **Professor Toshifumi Tsukahara**, Area of Bioscience and Biotechnology, Japan Advanced Institute of Science and Technology (JAIST), for his constant supervision, advices, constructive guidance, generosity, encouragement to pursue new work, and freedom to generate idea and execution during my doctoral course.

Besides my supervisor, I would like to acknowledge my second supervisor **Associate Professor Takumi Yamaguchi** and minor research advisor **Associate Professor Hidekazu Tsutsui**, Area of Bioscience and Biotechnology, JAIST, for their constructive comments for writing my dissertation.

Next, I would like to extend my gratitude to Dr. Matomo Sakari, Prof. Fujimoto Lab (especially Assist. Prof. Dr. Shigetaka Nakamura), Prof. Tsutsui Lab, JAIST and Professor Suzuki lab, Cancer Research Institute, Kanazawa University and Dr. Yoshitsugu Aoki and Dr. Norio Motohashi, National Center for Neurology and Psychiatry, Tokyo for their contribution in technical support and experimental skill development in my doctoral study. Moreover, I would like to express my appreciation to Dr. Thoufic Anam Azad, Dr. Umme Qulsum, Dr. Sonali Bhakta, John Munene, Jiarui Li, Wenhao Jin, Ruchika Mishra, Tanaka san, Saitake san, Mitsuki Furuya, Tetsuto Tohama, Junling Mo, Heeraman, Tomoko Onishi, Zhaoyuan Dai, Chisato Okhudaira, Wakana Chinen for their cooperation throughout my thesis.

My appreciation is also extended to all current and former lab members in our lab who helped me either directly or indirectly during my research work.

Finally, I am really grateful to my parents (**Nazrul Islam** and **Airun Nesa**) whose continuous inspiration and emotional support always kept my sights confidently set on my research. I am also grateful to my wife, **Sumaya Sultana** for her inspiration and continuous support. I would not be able to pay off her suffering and sacrifice during my research work. This work would not have been completed without them. Accordingly, I would like to dedicate this work to them.

September, 2021. **The Author**