# Designing sgRNA Sequences for Therapeutic CRISPR Applications Targeting T790M and L858R Mutations in Lung Adenocarcinoma

Darren Mark Effendy[1], Amanda Kaela Affandy[2], Claire Handoko[3]

[1]Penabur Junior College Kelapa Gading, Jakarta, Indonesia (darrenmarkeffendy@gmail.com)

[2]SMA NationalHigh Jakarta, Jakarta, Indonesia (amandakaffandy@gmail.com)

[3]Bina Bangsa School PIK, Indonesia (claire.handoko.2009@gmail.com)

## *ABSTRACT*

*The formation of non-small lung cancer, particularly adenocarcinoma, can be attributed to several key mutations within the* Tyrosine Kinase *domain of the Epidermal Growth Factor Receptor* (EGFR) *gene. Said mutations result in a receptor that remains activated despite the absence of its ligand, continuously activating downstream oncogenic pathways, such as* MAPK *and* PI3K/AKT*, which in turn lead to uncontrolled division and autophagy. Two mutations in particular–*exon 19 deletion (ΔE746-A750) *(44%), and* L858R *(41%)–account for nearly all cases of EGFR-caused Adenocarcinoma. At the same time, secondary mutations such as* T790M *confer resistance to first-generation tyrosine kinase inhibitors (TKIs). The use of the CRISPR process, along with the Cas9 enzyme, shows significant promise in cancer therapy. By employing bioinformatic platforms such as Benchling and CHOPCHOP, our team hypothesized that it is possible to design a single-guided RNA (sgRNA) that guides the Cas9 endonuclease towards specific oncogenes, allowing it to cleave these genes, thereby enabling targeted gene 'knock-outs'. However, we found that this approach faces numerous limitations, including low on-target specificity, which compromises safety and precision.*

***Keywords:*** *Adenocarcinoma, CRISPR Cas9, sgRNA, EGFR, Bioinformatic approach.*

## *INTRODUCTION*

Lung cancer remains the leading global cause of cancer-related mortality (1, 2), often due to late-stage diagnosis and sedentary lifestyles. In 2022, the global incidence of lung cancer was approximately 2.5 million new cases, with 717,211 cases (45.6% of male patients) and 541,971 cases (59.7% of female patients) identified as Lung Adenocarcinoma. Therefore, Adenocarcinoma is most prevalent in the lungs. (2).

Adenocarcinoma manifests as a malignant tumor of glandular epithelial cells. It develops when these specialized cells undergo unregulated cell division due to the expression of certain oncogenes– genes which promote cancer when mutated or overexpressed. (3, 4, ).

Adenocarcinoma can affect many organs, including the lungs, the prostate, the pancreas, the breasts, the colon, or the stomach. However, adenocarcinoma is most predominant in the lungs, forming a strain of non-small cell lung cancer (NSCLC) (5).

Adenocarcinoma is a multifaceted disease influenced by environmental, lifestyle, and genetic factors. A study conducted at Dr.Soetomo District General Hospital, Surabaya, in 2018 (6), Lung Adenocarcinoma in Indonesia is mostly prevalent in male smokers aged over 50, with a history of heavy smoking, particularly those who smoked filter cigarettes. In contrast, another study conducted in the United States concluded that factors related to digestive and metabolic health, such as gastroesophageal reflux disease (GERD), diabetes mellitus (DM), high body mass index (BMI), and obesity, are common conditions observed in patients with Lung and Esophageal Adenocarcinoma (7).

Over time, Adenocarcinoma typically spreads in two pathways (4). Invasive adenocarcinoma involves local spread to surrounding tissues, while metastatic adenocarcinoma occurs when the cancer cells enter the lymphovascular system and form secondary tumors in other parts of the body, far from the original organ, marking a more advanced stage of the cancer. In the case of lung adenocarcinoma, these patterns of spread do not appear to be directly influenced by EGFR mutation type (8). However, in 5 instances involving non-smokers, a diffuse, nodular pattern of cancer spread was observed, described as "miliary pulmonary metastases" (9). Despite these findings, a significant correlation between the type of EGFR mutation and metastatic behaviour has not been consistently shown.

The epidermal growth factor receptor (EGFR), a member of the membrane-bound tyrosine kinase family, is the principal gene associated with link mutations in adenocarcinoma (10). It is activated upon the binding of its mitogen ligand EGF. This activation results in the activation of two relevant pathways: mitogen-activated protein kinase (MAPK) and the downstream PI3K/AKT/mTOR pathway. MAPK is involved in the activation of transcription factors for several genes involved in mitosis (11), whereas PI3K/AKT/mTOR leads to the inhibition of autophagy (12,13,14) and the increase of cell mass (15).

Mutations within the EGFR-TK gene, specifically its tyrosine kinase domain, are strongly correlated with the formation of adenocarcinoma. Two mutations, exon 19 deletion (ΔE746-A750) (44%) and L858R (41%), account for the vast majority of activating mutations, whereas G719X (Glycine 719 → any other amino acid) (~4%) and S7681 (2%) more rarely behave as activating mutations. (16)

L858R is a missense mutation where the Leucine in position 858 (acg) mutates into arginine (atg). Located in the activation loop of the EGFR-TK domain, the substitution results in a permanently activated (expressed) state without the presence of a bonded ligand (17). This ligand-independent activation of the EGFR receptor results in the continuous activation of the MAPK and PI3K/AKT pathways. Such activation results in uncontrolled, continuous cell division and autophagy, which are correlated with the development and behaviour of the adenocarcinoma and cancer as a whole.

A common secondary point mutation, T790M compounds the therapeutic resistance of

Adenocarcinoma, by inhibiting the effectiveness of first-generation EGFR tyrosine kinase inhibitors (TKIs). Roughly 50%–60% of secondary resistance to first-generation EGFR TKIs is due to EGFR T790M, which results from a c.2369C>T substitution in exon 20, substituting threonine with methionine at position 790. Importantly, this mutation doesn't interfere with CRISPR binding, and so allele-specific editing is still possible (18).

The clustered regularly interspaced palindromic repeats (CRISPR) system is a gene-editing technology that allows the modification of sections of DNA sequences (19, 20), positioning itself as a promising alternative approach to cancer therapy. This process involves the Cas9 endonuclease enzyme, which cuts the DNA at a particular section, and a guide specific to the CRISPR Cas9 process called single-guide RNA (sgRNA). sgRNA is a synthetic fusion of CRISPR RNA (crRNA), which matches the target DNA sequence, and trans-activating CRISPR RNA (tracrRNA), which attaches to Cas9 and activates it. SgRNA guides the Cas9 enzyme to the intended section, ensuring onsite cleavage of DNA, creating a double-stranded break (DSB) and activating the cell's DNA repair mechanism, which primarily occurs through non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathways. To perform a gene 'knock-out', the less-precise NHEJ is used to repair the break inaccurately, leading to the deactivation of the oncogene, and thus the suppression of malignant features.

Just as with exon 19 deletion, L858R and T790M have high prevalence; however, unlike it, they are point mutations. In this case, CRISPR can be used to deactivate these point mutations through targeted 'knockouts'. Consequently, we hypothesize that the design and optimization of sgRNA (in terms of specificity and efficiency (21, 22) ), leading the Cas9 endonuclease, can be done and evaluated using bioinformatic models.

## *Methodology*

This study employed a bioinformatic pipeline to design sgRNAs, aiming to identify the most efficient and specific variants for the L858R and T790M mutations.

### *2.1 FASTA retrieval*

The canonical FASTA sequence of the EGFR-TK gene was first extracted from NCBI's genomic database. Due to alternative splicing resulting in variants of the primary transcript, the canonical transcript (NM_005228.5) was chosen because it is the predominant isoform and exhibits conservation across species (23).

### *2.2 Gene structure retrieval*

Before proceeding to sgRNA design, the mutations were first located within the CDS. L858R and T790M mutations are located within exons 21 and 20, respectively. To verify exon-intron architecture, gene prediction was initially conducted by uploading the primary transcript (NM_005228.5) into Fgenesh v.26.

Fgenesh is part of the Softberry suite of analysis tools. It primarily serves as a gene prediction model; our team chose it due to its de novo and ab initio mechanisms. De novo implies the construction of gene models without any prior reference annotation. In contrast, ab initio, correlatively, means gene prediction is based solely on DNA sequence, using statistical models such as the Hidden Markov Models (HMM) (24).

These predictions, reported by Fgenesh v.26, were benchmarked against NCBI's curated annotated reference to further confirm the genetic architecture.

Subsequently, the precise loci of the target mutations were located. The primary transcript (NM_005228.5) was modified in *silico* using bioinformatic software Benchling to introduce the target mutations, allowing the formation of complementary sgRNAs.

### *2.3 Guide Design*

The modified transcript was imported into Benchling, a cloud-based genomic editing platform. A region of interest, with a length of 60 bases, was selected to screen potential sgRNA candidates upstream of PAMs. Similarly, we evaluated results from Benchling by simultaneously uploading the modified transcript into CHOPCHOP v.3 (25), a CRISPR/TALEN guide RNA design tool.

sgRNAs in both computational models were subsequently designed with the default parameters set by Benchling.

### 2.4 Sequence Alignment

To evaluate off-target sites, the most optimal sgRNAs were imported into the Nucleotide BLAST.

## *Results and Discussion*

Gene prediction is done to validate exon-intron architecture and the location of point mutations T790M and L858R. In parallel, sgRNA design is critical for accurate guidance of the Cas9 endonuclease, which underpins the effectiveness of the CRISPR process as a whole.

### *3.1 Gene prediction verification*

The results of genomic prediction from Fgenesh are as follows (further details in ***Table 1***):

*FGENESH 2.6 Prediction of potential genes in Homo_sapiens genomic DNA*

*Seq name: test sequence*

*Length of sequence: 192612*

*Number of predicted genes 2: in +chain 2, in -chain 0.*

*Number of predicted exons 40: in +chain 40, in -chain 0.*

*Positions of predicted genes and exons: Variant 1 from 1, Score:267.378564*

***Table 1.*** *FGenesh exon prediction. The table reports the two exons predicted by FGenesh. Predicted exons are from the primary transcript* (NM_005228.5).

| Gene | Strand | Start | End | # Exons | TSS (bp) | PolyA (bp) |
|---|---|---|---|---|---|---|
| 1 | + | 262 | 15,786 | 6 | – | 15,786 |
| 2 | + | 31,102 | 186,883 | 34 | 31,102 | 186,883 |

The National Center for Biotechnology Information (NCBI) contains a curated annotated reference for the EGFR-TK gene. Multiple discrepancies, such as the individual length and number of exons (40 reported by Fgenesh's prediction against 28 in the NCBI reference)

Although reliable, FGenesh is not perfect. It is most likely that a portion of the "40" exons predicted are false positives (26), which can be potentially due to the misinterpretation of

sequence patterns from ab initio predictions. NCBI, on the other hand, cross-references with other databases.

Consequently, our team chose to proceed with NCBI's expert-curated sequences due to the already extensive review and support from experimental data.

### *3.2 Computational models involved.*

Our team utilized Benchling primarily for its cloud sharing ability, allowing multiple individuals to work simultaneously on one project. Conversely, CHOPCHOP was employed for its quick and clean interface, suitable for validating sgRNA candidates from Benchling.

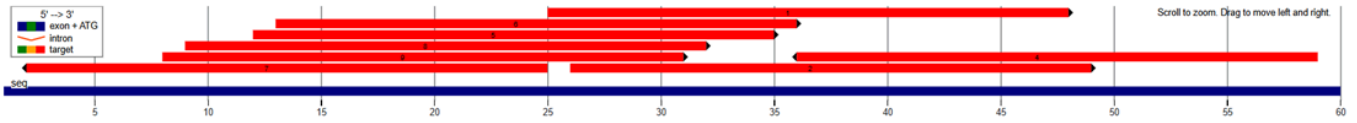
Both computational models identified an identical sgRNA. Six and nine potential sgRNA candidates were designed for T790M and L858R, respectively ***(Table 2,3 & Fig. 1).*** A distinction, however, would be CHOPCHOP's evaluation of specificity, which is done through a mismatch (MM) sequence. For instance, MM1 denotes a potential off-target site that possesses a single nucleotide mismatch with the sgRNA sequence, thereby increasing the risk of off-target cleavage. This distinctive feature allows for a comparative evaluation with sequence alignment results in BLAST.

***Table 2.*** *T790M sgRNA candidates (benchling)**.** The table reports six sgRNA designed using Benchling from the 60-nucleotide interest region.*

| Position | Strand | Sequence | PAM | Specificity score | Efficiency score |
|---|---|---|---|---|---|
| 2610 | -1 | tgagctgcacggtggaggtg | agg | 31.88425 | 64.16179 |
| 2615 | -1 | catgatgagctgcacggtgg | agg | 40.15848 | 55.79479 |
| 2618 | -1 | ctgcatgatgagctgcacgg | tgg | 42.33117 | 79.76679 |
| 2621 | -1 | gagctgcatgatgagctgca | cgg | 35.61432 | 71.37097 |
| 2643 | 1 | atcatgcagctcatgcccttt | cgg | 39.54848 | 51.38578 |
| 2654 | 1 | catgcccttcggctgcctcc | tgg | 33.69811 | 35.54578 |

***Table 3.*** *L858R sgRNA candidates (benchling). The table reports nine sgRNA designed using Benchling from the 60-nucleotide interest region.*

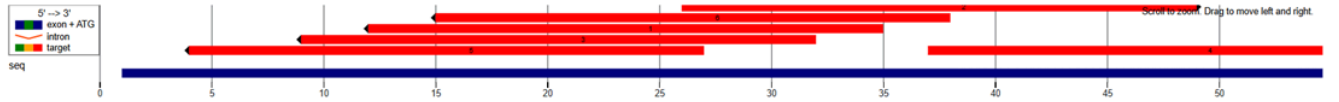
| Position | Strand | Sequence | PAM | Specificity score | Efficiency score |
|---|---|---|---|---|---|
| 2809 | -1 | tctgtgatcttgacatgctg | cgg | 38.86458 | 61.77917 |
| 2826 | 1 | catgtcaagatcacagattt | tgg | 32.40663 | 34.0963 |
| 2827 | 1 | atgtcaagatcacagatttt | ggg | 29.24152 | 35.76113 |
| 2830 | 1 | tcaagatcacagattttggg | cgg | 54.55215 | 64.04463 |
| 2831 | 1 | caagatcacagattttgggc | ggg | 38.81625 | 57.7261 |
| 2843 | 1 | ttttgggcgggccaaactgc | tgg | 55.55276 | 35.20674 |
| 2843 | -1 | ctcttccgcacccagcagtt | tgg | 39.9886 | 32.90189 |
| 2844 | 1 | tttgggcgggccaaactgct | ggg | 44.85912 | 51.41534 |
| 2849 | 1 | gcgggccaaactgctgggtg | cgg | 38.2449 | 51.86638 |




Download results: Please select one

View in UCSC genome browser

| Rank | Target sequence | Genomic location | Strand | GC content (%) | Self-complementarity | MM0 | MM1 | MM2 | MM3 | Efficiency |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TTTTGGGCGGGCCAAACTGCTGG | seq:26 | + | 60 | 1 | 0 | 1 | 0 | 1 | 35.21 |
| 2 | TTTGGGCGGGCCAAACTGCTGGG | seq:27 | + | 60 | 1 | 0 | 1 | 1 | 1 | 51.42 |
| 3 | GCGGGCCAAACTGCTGGGTGCGG | seq:32 | + | 70 | 0 | 0 | 1 | 0 | 10 | 51.87 |
| 4 | CTCTTCCGCACCCAGCAGTTTGG | seq:37 | - | 60 | 1 | 1 | 0 | 0 | 12 | 0.00 |
| 5 | TCAAGATCACAGATTTTGGGCGG | seq:13 | + | 40 | 0 | 0 | 1 | 1 | 13 | 64.04 |
| 6 | CAAGATCACAGATTTTGGGCGGG | seq:14 | + | 45 | 0 | 1 | 0 | 0 | 17 | 57.73 |
| 7 | TCTGTGATCTTGACATGCTGCGG | seq:3 | - | 45 | 0 | 1 | 0 | 0 | 19 | 0.00 |
| 8 | ATGTCAAGATCACAGATTTTGGG | seq:10 | + | 30 | 0 | 1 | 0 | 2 | 29 | 35.76 |
| 9 | CATGTCAAGATCACAGATTTTGG | seq:9 | + | 35 | 0 | 1 | 0 | 4 | 27 | 34.10 |




Download results: Results table as .tsv

View in UCSC genome browser

| Rank | Target sequence | Genomic location | Strand | GC content (%) | Self-complementarity | MM0 | MM1 | MM2 | MM3 | Efficiency |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | ATCATGCAGCTCATGCCCTTCGG | seq:27 | + | 50 | 1 | 0 | 1 | 0 | 11 | 51.39 |
| 3 | CATGATGAGCTGCACGGTGGAGG | seq:10 | - | 60 | 0 | 0 | 1 | 1 | 14 | 55.79 |
| 4 | CATGCCCTTCGGCTGCCTCCTGG | seq:38 | + | 70 | 1 | 1 | 0 | 0 | 25 | 0.00 |
| 1 | CTGCATGATGAGCTGCACGGTGG | seq:13 | - | 60 | 0 | 0 | 1 | 0 | 7 | 79.77 |
| 6 | GAGCTGCATGATGAGCTGCACGG | seq:16 | - | 55 | 1 | 0 | 1 | 1 | 33 | 71.37 |
| 5 | TGAGCTGCACGGTGGAGGTGAGG | seq:5 | - | 65 | 0 | 1 | 0 | 0 | 31 | 64.16 |

***Fig.1.*** *T790M (Top) and L858R (Bottom) sgRNA candidates (CHOPCHOP). The images report CHOPCHOP's identical designs for the sgRNAs from the 60-nucleotide interest region.*

### *3.3 Suboptimal sgRNA for L858R and T790M*

During an in *silico* trial, the most optimal sgRNAs are selected primarily by specificity and efficiency evaluation (21, 22, 27). Specificity is derived from off-target scores and quantifies how uniquely the sgRNA targets the intended DNA sequence, hence how unlikely it is that an off-target site will be edited. Conversely, the efficiency score presents how well the sgRNA can lead the Cas9 endonuclease to induce targeted DNA cleavage and cause a double-strand break (DSB). It is important to note that a higher magnitude of these scores indicates better results.

Both values are correlated with GC content; GC content underlines what percentage do guanine and cytosine nucleotides make of the guide's composition. G-C pairs form strong triple hydrogen bonds, which allow for increased stability when the sgRNA binds to the DNA strand. Too little will result in weaker base pairing, causing premature dissociation; conversely, too much hinders strand separation. Hence, there needs to be an ideal amount (40% to 60%) for an optimal result.

Taking these parameters into account, out of the fifteen designed sgRNAs (***Table 2,3 & Fig. 1),*** the most 'optimal' sgRNAs are "ctgcatgatgagctgcacgg" and "tcaagatcacagattttggg" for T790M and L858R, respectively.

Although the Efficiency scores (~80) are considered sufficient (28), the specificity scores of both the 'optimal' sgRNAs are suboptimal at best.

Low specificity can be attributed to a wide range of factors, including but not limited to: similarity with other sequences within the human genome assembly *(GRCh38)* and unsuitable GC percentage (29).

### *3.4 Possible Clinical Consequences of Targeting Mutations with Designed sgRNAs*

Results of sequence alignment using Nucleotide BLAST ***(Table 4,5)*** reveal various other genes within the human genome, which, due to being precise matches to designed sgRNAs, could be potential off-target sites.

Although the majority of these predicted "off-target' sites are variants of the EGFR-TK genes and their associated exons, non-EGFR-TK binding regions were also detected.

Among these, are lengths of the SHROOM 3 (NM_020859.4) for T790M and PDPK1 gene (NM_001261816.2) for L858R's sgRNA

SHROOM3 is associated with podocyte formation within the kidney (30), whilst PDPK1 is involved in the activity of the following enzymes: 3-phosphoinositide-dependent protein kinase, phospholipase activator & binding. Hence, PDPK1 is a critical gene for intra- and extracellular cell signalling (31).

Despite no known direct interaction between EGFR and SHROOM3 or PDPK1 ***(Fig.2)***, the designed sgRNA exhibited sequences complementary to regions within SHROOM3 and PDPK1. This is possible due to molecular cross-talk through intermediary genes such as LRIG1 or CDH1 (SHROOM3) and PIK3CA or CRK (PDPK1) involved within different signal transduction pathways between the genes.

Due to such perfect alignments, it is probable that the chosen sgRNA leads the Cas9 endonuclease to cleave these genes. This results in the deactivation (knock-out) of genes essential to metabolic processes, hence serious medical consequences.

***Table 4*** *T790M sgRNA sequence alignment (Nucleotide BLAST). The table reports the first five potential off-target risks associated with the designed sgRNA, sorted by E-value.*

| query acc.ver | subject acc.ver | Identity (%) | alignment length | mismatches | E value | Bit score |
|---|---|---|---|---|---|---|
| Query_4252805 | MT010322.1 | 100 | 20 | 0 | 0.021 | 40.1 |
| Query_4252805 | AC092463.5 | 100 | 16 | 0 | 5.2 | 32.2 |
| Query_4252805 | AC007064.27 | 100 | 16 | 0 | 5.2 | 32.2 |
| Query_4252805 | AL135786.17 | 100 | 16 | 0 | 5.2 | 32.2 |
| Query_4252805 | AC005301.22 | 100 | 16 | 0 | 5.2 | 32.2 |

***Table 5*** *L858R sgRNA sequence alignment (Nucleotide BLAST). The table reports the first five potential off-target risks associated with the designed sgRNA, sorted by E-value.*

| query acc.ver | subject acc.ver | Identity (%) | alignment length | mismatches | E value | Bit score |
|---|---|---|---|---|---|---|
| Query_293022 9 | HM437235.1 | 100 | 20 | 0 | 0 | 1 |
| Query_293022 9 | XM_054357417.1 | 100 | 20 | 0 | 0 | 1 |
| Query_293022 9 | XM_047419953.1 | 100 | 20 | 0 | 0 | 1 |
| Query_293022 9 | BC094761.1 | 100 | 20 | 0 | 0 | 1 |
| Query_293022 9 | NM_001346899.2 | 100 | 20 | 0 | 0 | 1 |

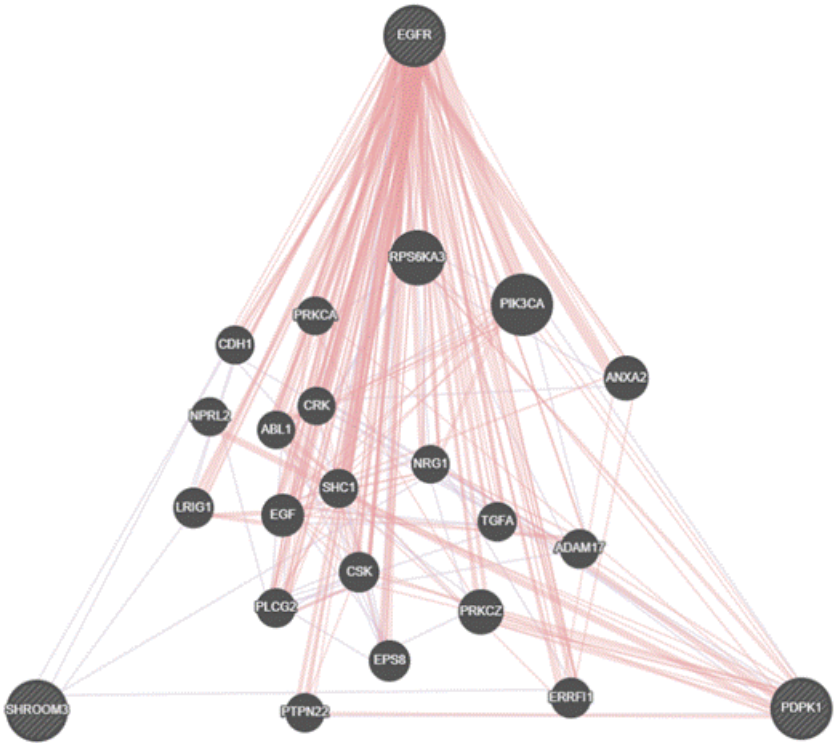



***Fig.2.*** *Gene Interaction network between SHROOM3, PDPK1, and EGFR (Genemania). The image reports the physical interactions, co-expression, and co-localization of the genes.*

### *3.5 Limitations and recommendations*

As previously discussed, a major limitation of this research is the specificity of the designed sgRNAs. Significant modification of the sgRNA must be conducted before they are considered clinically viable (32). We suggest research into several methods to increase specificity, including: truncation and extension (33), high fidelity Cas9 variants (such as SpCas9-HF1) (34), dCas9-FokI fusion nucleases (21), and Cas9 Nickase mutants (35)

Truncation or expansion of the sgRNA to 17 - 18 BP is known to increase the specificity of the sgRNA (31) significantly.

Dimeric RNA-guided FokI-dCas9 nuclease (RFN) contains a FokI cleavage domain, which is attached to the Cas9 (21), since the FokI functions only as a dimer, two adjacent half-sites are required, doubling the target length and increasing fidelity. Comparatively, A Cas9 mutant, Cas9 nickase (Cas9n) D10A mutant requires two sgRNAs to nick both strands within the dsDNA, thereby increasing specificity.

However, in the case that the sgRNA does lead the Cas9 endonuclease towards the intended mutations (T790M & L858R), several issues could arise after cleaving. Primarily, the knockout of these genes results in currently unknown consequences (36)

### *Conclusion*

Our team performed sgRNA design, evaluation, and optimization with the assistance of bioinformatic models such as Benchling and CHOPCHOP. Results displayed a total of 15 sgRNAs for L858R and T790M mutations with suboptimal specificity scores. This indicated the presence of numerous off-target sites and the necessity for further optimization of designed sgRNAs.

### *Acknowledgements*

The authors extend their immense gratitude to Juan Bahrian for his invaluable support and guidance in the development of this paper.

### *References*

1. World Health Organization. *Lung Cancer*. 26 June 2023, https://www.who.int/news-room/fact-sheets/detail/lung-cancer.
2. World Health Organization. *Global Lung Cancer Incidence According to Subtype: New Study Highlights Rising Adenocarcinoma Rates Linked to Air Pollution*. 3 Feb. 2025, https://www.iarc.who.int/wp-content/uploads/2025/02/pr359_E.pdf.
3. Cleveland Clinic Medical Professional. *Oncogenes*. Cleveland Clinic, 24 Feb. 2025, https://my.clevelandclinic.org/health/body/24949-oncogenes.
4. Cleveland Clinic. *Adenocarcinoma Cancers: What Is Adenocarcinoma?* 25 June 2024, https://my.clevelandclinic.org/health/diseases/21652-adenocarcinoma-cancers.
5. Lung Cancer Foundation of America. *Adenocarcinoma*. n.d., https://lcfamerica.org/about-lung-cancer/diagnosis/types/adenocarcinoma/#:~:text=Adenocarcinoma.
6. Laily, Luluk Lady., Santi Martini, Kurnia Dwi Artanti, and Sri Widati. “Risk Factors of Lung Adenocarcinoma in Patients at Dr. Soetomo District General Hospital Surabaya in 2018.” *Indonesian Journal of Public Health*, vol. 13, no. 2, 2018, https://e-journal.unair.ac.id/IJPH/article/view/13393/pdf.
7. Agrawal, Kriti., Ronald J Markert, and Sangeeta Agrawal. “Risk Factors for Adenocarcinoma and Squamous Cell Carcinoma of the Esophagus and Lung.” AME Medical Journal, vol. 3, 2018, p. 35. https://doi.org/10.21037/amj.2018.02.04
8. Beypinar, Ismail., Hacer Demir, Murat Araz, and Mukremin Uysal. “The Relationship between EGFR Mutation and Metastasis Pattern in Lung Adenocarcinoma.” *Journal of Oncological Sciences*, vol. 5, no. 2, 2019, pp. 61–64. https://doi.org/10.1016/j.jons.2019.08.002.
9. Laack, Eckart et al. “Miliary Never-Smoking Adenocarcinoma of the Lung: Strong Association with Epidermal Growth Factor Receptor Exon 19 Deletion.” *Journal of*

*Thoracic Oncology*, vol. 6, no. 1, 2011, pp. 199–202. https://doi.org/10.1097/JTO.0b013e3181fb7cf1.
10. EGFR Epidermal Growth Factor Receptor [Homo Sapiens (Human)] – Gene – NCBI." *National Center for Biotechnology Information*, n.d., https://www.ncbi.nlm.nih.gov/gene/1956.
11. Morrison, Deborah K. "MAP Kinase Pathways." *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 11, 2012, a011254. https://doi.org/10.1101/cshperspect.a011254.
12. Fruman, David A., Honyin Chiu, Benjamin D. Hopkins, Shubha Bagrodia, Lewis C. Cantley, and Robert T. Abraham. "The PI3K Pathway in Human Disease." Cell, vol. 170, no. 4, 10 Aug. 2017, pp. 605–635. https://doi.org/10.1016/j.cell.2017.07.029
13. Hemmings, Brian A., and David F. Restuccia. "PI3K-PKB/Akt Pathway." Cold Spring Harbor Perspectives in Biology, vol. 4, no. 9, 2012, a011189. https://doi.org/10.1101/cshperspect.a011189
14. Hsu, Patrick D., et al. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." Nature Biotechnology, vol. 31, no. 9, 2013, pp. 827–832. https://doi.org/10.1038/nbt.2647
15. Rabanal-Ruiz, Yolanda, Elena G. Otten, and Viktor I. Korolchuk. "mTORC1 as the Main Gateway to Autophagy." *Essays in Biochemistry*, vol. 61, no. 6, 2017, pp. 565–584. https://doi.org/10.1042/ebc20170027.
16. Kumar, Amit, Edward T. Petri, Balazs Halmos, and Titus J. Boggon. "Structure and Clinical Relevance of the Epidermal Growth Factor Receptor in Human Cancer." Journal of Clinical Oncology, vol. 26, no. 10, 2008, pp. 1742–1751. https://doi.org/10.1200/JCO.2007.12.1178
17. Yun, Ch., T. J. Boggon, Y. Li, M. S. Woo, H. Greulich, M. Meyerson, and M. J. Eck. "Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity." Cancer Cell, vol. 11, no. 3, 2007, pp. 217–227. Elsevier, https://doi.org/10.1016/j.ccr.2006.12.017
18. Suda, Kenichi, Ryoichi Onozato, Yasushi Yatabe, and Tetsuya Mitsudomi. "EGFR T790M Mutation: A Double Role in Lung Cancer Cell Survival?" Journal of Thoracic Oncology, vol. 4, no. 1, 2009, pp. 1–4. https://doi.org/10.1097/JTO.0b013e3181913c9f
19. Redman, Melody, Andrew King, Caroline Watson, and David King. "What Is CRISPR/Cas9?" Archives of Disease in Childhood – Education and Practice, vol. 101, no. 4, Apr. 2016, pp. 213–15. BMJ Publishing Group, https://doi.org/10.1136/archdischild-2016-310459
20. Hsu, Patrick D., Eric S. Lander, and Feng Zhang. "Development and Applications of CRISPR-Cas9 for Genome Engineering." Cell, vol. 157, no. 6, 2014, pp. 1262–78. Elsevier, https://doi.org/10.1016/j.cell.2014.05.010
21. Wyvekens, Nicolas, Vivek V. Topkar, Chiraag Khayter, J. Keith Joung, and Shengdar Q. Tsai. "Dimeric CRISPR RNA-Guided FOKI-DCAS9 Nucleases Directed by Truncated gRNAs for Highly Specific Genome Editing." Human Gene Therapy, vol. 26, no. 7, 2015, pp. 425–431. https://doi.org/10.1089/hum.2015.084
22. Haeussler, Maximilian, Kerstin Schönig, Hannes Eckert, et al. "Evaluation of Off-Target and On-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR." Genome Biology, vol. 17, 2016, p. 148. https://doi.org/10.1186/s13059-016-1012-2
23. Kim, Boyeon, Yoonjung Kim, Saeam Shin, Seung-Tae Lee, Jae Yong Cho, and Kyung-A Lee. "Application of CRISPR/Cas9-Based Mutant Enrichment Technique to Improve the

Clinical Sensitivity of Plasma EGFR Testing in Patients with Non–Small Cell Lung Cancer." Cancer Cell International, vol. 22, no. 1, 2022, p. 82. https://doi.org/10.1186/s12935-022-02504-2
24. Solovyev, Victor, and Alexey Salamov. "Automatic Annotation of Microbial Genomes and Metagenomic Sequences." Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies, edited by Robert W. Li, Nova Science Publishers, 2011, pp. 61–78.
25. Labun, Kornel, Thomas G. Montague, Michael Krause, Yiping N. Torres Cleuren, Håkon Tjeldnes, and Eivind Valen. "CHOPCHOP v3: Expanding the CRISPR Web Toolbox beyond Genome Editing." Nucleic Acids Research, vol. 47, no. W1, 2019, pp. W171–W174. https://doi.org/10.1093/nar/gkz365
26. Solovyev, Victor, Pavel Kosarev, Ivan Seledsov, and Dmitry Vorobyev. "Automatic Annotation of Eukaryotic Genes, Pseudogenes and Promoters." Genome Biology, vol. 7, suppl. 1, 2006, S10. https://doi.org/10.1186/gb-2006-7-s1-s10
27. Uddin, Fathema, Charles M. Rudin, and Triparna Sen. "CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future." Frontiers in Oncology, vol. 10, Aug. 2020, p. 1387. Frontiers Media, https://doi.org/10.3389/fonc.2020.01387
28. Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, et al. "Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9." Nature Biotechnology, vol. 34, no. 2, 2016, pp. 184–191. https://doi.org/10.1038/nbt.3437
29. Konstantakos, Vasileios, Anastasios Nentidis, Anastasia Krithara, and George Paliouras. "CRISPR–Cas9 gRNA Efficiency Prediction: An Overview of Predictive Tools and the Role of Deep Learning." Nucleic Acids Research, vol. 50, no. 7, 2022, pp. 3616–3637. https://doi.org/10.1093/nar/gkac192
30. Paul, Alana, Aidan Lawlor, Kathleen Cunanan, Prabhjot S. Gaheer, Akshat Kalra, Mario Napoleone, Michael B. Lanktree, and David Bridgewater. "The Good and the Bad of SHROOM3 in Kidney Development and Disease: A Narrative Review." Canadian Journal of Kidney Health and Disease, vol. 10, 2023. https://doi.org/10.1177/20543581231212038
31. "PDPK1 3-Phosphoinositide Dependent Protein Kinase 1 [Homo Sapiens (Human)] – Gene – NCBI." NCBI Gene, n.d., https://www.ncbi.nlm.nih.gov/gene/5170
32. Tycko, Joshua, Valerie E. Myer, and Patrick D. Hsu. "Methods for Optimizing CRISPR-CAS9 Genome Editing Specificity." Molecular Cell, vol. 63, no. 3, 2016, pp. 355–70. Elsevier, https://doi.org/10.1016/j.molcel.2016.07.004
33. Wu, Xuebing, Alex J. Kriz, and Phillip A. Sharp. "Target Specificity of the CRISPR-Cas9 System." Quantitative Biology, vol. 2, no. 2, 2014, pp. 59–70. Springer, https://doi.org/10.1007/s40484-014-0030-x
34. Kleinstiver, Benjamin P., Vivek Pattanayak, Michelle S. Prew, Shengdar Q. Tsai, Nhu T. Nguyen, Zongli Zheng, and J. Keith Joung. "High-Fidelity CRISPR–Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects." Nature, vol. 529, no. 7587, 2016, pp. 490–95. Nature Publishing Group, https://doi.org/10.1038/nature16526
35. Ran, F. Ann, Patrick D. Hsu, Changan Lin, Jonathan S. Gootenberg, Silvana Konermann, Andrew E. Trevino, et al. "Double Nicking by RNA-Guided CRISPR CAS9 for Enhanced Genome Editing Specificity." Cell, vol. 154, no. 6, 2013, pp. 1380–89. Elsevier, https://doi.org/10.1016/j.cell.2013.08.021

36. Narasimhan, Vagheesh M., Katherine A. Hunt, Daniel Mason, Christopher L. Baker, Konrad J. Karczewski, Michael R. Barnes, et al. “Health and Population Effects of Rare Gene Knockouts in Adult Humans with Related Parents.” Science, vol. 352, no. 6284, 2016, pp. 474–77. AAAS, https://doi.org/10.1126/science.aac8624