ISSN No:-2456-2165

Role of CRISPR/Cas9 in Genetic Manipulation of ROS1 and EGFR Genes using Synthego Platform

Manav Goenka[1] Aniket De[1], and Arup Ratan Biswas[2]
[1]3RDYear Student, Department of Biotechnology, Techno India University, Kolkata, India; [2]Assistant Professor and Head, Department of Chemistry, Techno India University, Kolkata, India

Abstract:- Mutations and fusions in kinase enzymes are often observed in cancer prognosis. The growth and survival of tumor cells depend on the activation of kinase enzymes which when activated unrestrained can lead to the uncontrolled division of malignant lung cells. Thus, their inhibition is viewed as a promising and effective anti-cancer therapy. ROS1 and EGFR are two tyrosine kinases that have been explored as the genes responsible for Non-Small Cell Lung Cancer (NSCLC). By interrupting the unchecked division of these genes, the development of malignant lung cancer cells can be blocked. The results show 4 of the top-line RNAs for altering the gene quality as well as the target sequences relevant to cleavage by that gRNA, 4 for each gene. We propose a genetic approach of controlling the ROS1 and EGFR genes guided by CRISPR/Cas-9 to guarantee fewer symptoms and an increasingly powerful treatment, by the use of computational tools.

Keywords:- *CRISPR; Cas-9; Lung cancer; EGFR; ROS1; Synthego.*

I. INTRODUCTION

Invent of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 has taken the medical research worldwide to a newer height, particularly in fighting cancer and more deadly diseases that are currently engulfing the mankind at large. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 is a technology evolved from type II immune system of archaebacteria. Exploring the functionality of Cas9 protein leads the researcher to a site-specific targeted genetic manipulation. It is a robust technology that is used to make genetic cuts and edits in the DNA. According to WHO[1], Lung cancer is an alarming threat to mankind with about 2.09 million cases and 1.78 million deaths worldwide. Although Lung cancer is usually more predominant in males as compared to females, this demographic has been shifting dramatically [2] since the last few decades. Although, several cases have been reported in non-smokers [3] and caused due to genetic factors such as inheritance and a mutation in some specific genes.Most of these genes are involved in the regulation of gene activity, cell proliferation, differentiation, and apoptosis.[4] The occurrence of lung cancer is governed by the mutation of epithelial cells present in the lungs, caused by inhalation of harmful, toxic substances such as nickel, cadmium, benzene, etc,[5] causing the cells to cluster together and hence forming a tumor which affects the other organs while being transported by the lymph fluid. There are various types of lung cancers such as adenocarcinoma, squamous cell carcinoma, combined small cell carcinoma, large cell carcinoma, and oat cell carcinoma which are caused by a mutation in genes: ROS1[6][7] and EGFR[8][9]. Since the remediation of lung cancer surgery is restricted to only those carcinomas which have not spread beyond the lungs i.e., up to stages I, II, and III, and is followed by chemotherapy and radiation therapy to kill the cancer cells; we elucidate a genome manipulation-based approach towards Lung Cancer using CRISPR/Cas-9 Technology to edit the ROS1 and EGFR genes using the most effective guide RNA sequences to control and reduce the effects of mutation leading to cancer itself.This paper provides data for future researchers and is approved towards the CRISPR/Cas9 system.

II. ROLE OF ROS1 IN LUNG CANCER

The ROS1 (c-ros oncogene 1) gene is located in chromosome 6 and included in the receptor tyrosine kinase family. [10] It is structurally analogous to the ALK protein. In the human cell, no ligand-receptor nor any physiological role for the wild-type ROS1 has been determined to date.[6]Abnormal cell growth is triggered by the fusion of the ROS1 gene with a nearby gene(Figure 1), leading it to be persistently stuck in the “on” position.[11] The patients who test positive for ROS1 fusion tend to be younger than the general average with a median age of 50.5 years,[12] non-smokers with a diagnosis of adenocarcinoma. These fusions have been the target for many drug studies with most of the drugs being testes only for ROS1-positive non-small cell lung carcinoma (NSCLC).

A. PATHWAY OF ROS1 FUNCTION

The ROS1 gene transcribes for a protein consisting of an intracellular Tyrosine-Kinase domain, a small transmembrane region, and nine repeated fibronectin-like motifs that constitute the extracellular ligand-binding domain.[10][11] It takes part in phosphorylation events, interacting with a variety of proteins and can thus lead to the activation of various respective oncogenic pathways.[13] Auto-phosphorylation of the ROS1 gene and the phosphorylation of the SHP, MEK, ERK, STAT3, and AKT molecules are activated by the fusion of ROS1 with other genes. These functions of ROS1 are pharmacologically blocked by the use of drugs.[14] A more precise pathway has been described below in **Figure 1**.

ISSN No:-2456-2165

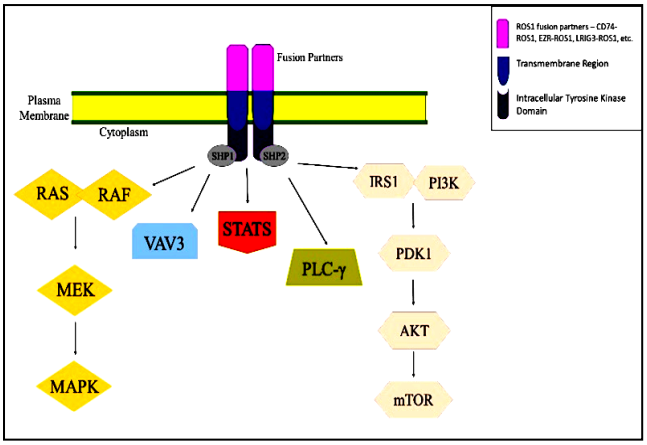

FIG 1: – ROS1signaling pathways. [12][13][14] ROS1 fuses with a nearby gene and leads to phosphorylation of the SHP molecules attached in the intracellular TK domain. This phosphorylation leads to further activation of the various other oncogenic pathways.

III. ROLE OF EGFR IN LUNG CANCER

The EGFR (Epidermal Growth Factor Receptor) gene codes for making the epidermal growth factor receptor, which is a transmembrane glycoprotein undergoing conformational changes to assist in autophosphorylation and the MAPK pathway.[9]Overexpression or mutation of the EGFR gene is a major factor leading to NSCLC. 90% of the known EGFR mutations occur as frame-shift mutations in Exon 19 or point mutations in Exon 21[15][16] resulting in the continuous activation of the various signal transduction pathways, and thus leading to the various tumorigenic pathways such as proliferation, apoptosis, and angiogenesis disregarding the presence of an extracellular ligand.[17]

A. PATHWAY OF EGFR FUNCTION[18][19]

The EGFR pathway plays a crucial role in regulating growth, survival, proliferation, differentiation, and cell to cell communication in mammalian cells. 15 members belonging to the EGF ligand family have been identified as the input signals and induces the homodimerization (EGFR – EGFR) and heterodimerization (EGFR – HER2). This dimerization causes transphosphorylation on numerous tyrosine residues which lead to phosphotyrosine-binding adaptors linking to phosphorylated receptors, through SOS (a guanine nucleotide exchange protein) and RAS which is a small GTP-binding protein, to a linear cascade culminating in ERK1 and ERK2. Various transcription factors are then activated when they translocate to the nucleus.

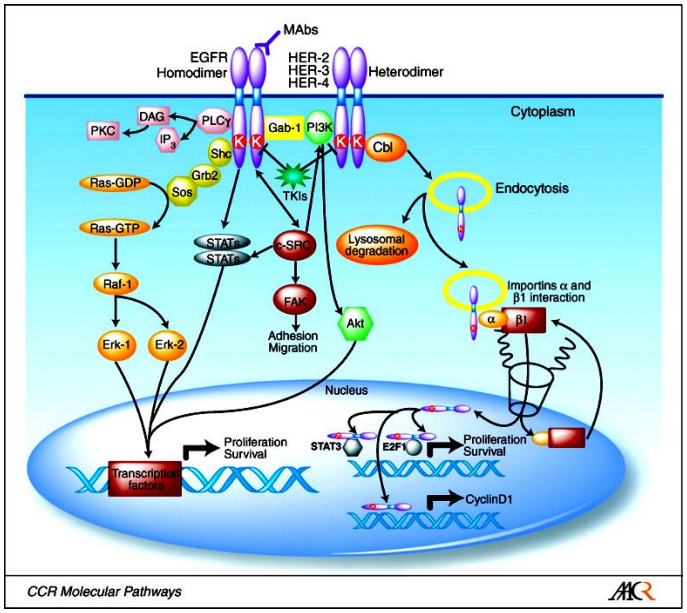

FIG 2 :– EGFR signaling pathways as described by Scaltriti et.al. [20]

IV. CRISPR/CAS9 TECHNOLOGY

The type II CRISPR-Cas systems, which provide bacteria with adaptive immunity to viruses and plasmids have given rise to CRISPR-Cas9 technology. The CRISPR-associated protein Cas9 is an endonuclease that with the aid of a guide sequence within an RNA duplex forms base pairs with DNA target sequences thus enabling Cas9 to introduce a site-specific double-strand break in the DNA.[21]Despite its advancements and promises; off-target effects, mosaicism, and limited ways of visualizing the embryos are some of its limitations. However, laboratories across the globe are now using this technology for innovative approaches in the field of biology.[22]

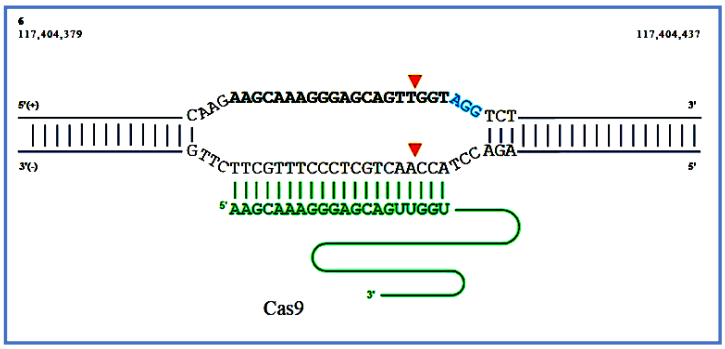

FIG 3 :– Synthego output example with PAM sequence, Guide sequence, and the target region

V. COMPUTATIONAL TOOL BASED APPROACH

The system of activity of CRISPR was improved with present-day computational instruments as well as its plausible results were likewise anticipated all the more precisely. The scope of information sources of info and the joining of a lot of highlights can be utilized to anticipate cleavage efficiencies. An AI calculation depends on a scope

of information sources of info and the joining of a lot of highlights that can be utilized to anticipate cleavage efficiencies. Analysts began executing AI calculations utilizing computational instruments to develop a progressively exact cleavage result and disposing of the off-target bad marks which is that it divides askew DNA. They would break down a portion of the most important and reliable CRISPR AI systems that are eligible for usage and assess their validity by looking at their yields for our desired outcomes. Preparation of the single reference RNA for the Cas9 protein is assumed to be the basic reason for the poor mark of the Askew cleavage. Synthego [24] is an online instrument concocted by MIT and considered as the most innovative solution out of the well-known analytical methodsdueto its willingness to take into account DNA bulges, which are sometimes ignored by other devices. This has had a significant impact on improving accuracy because DNA bulges arevery common phenomena that tend to hinder the desired result of our DNA manipulation.

A. Experimental Setup in Synthego

We analyzed our genesROS1 and EGFR, for lung cancerusing Synthego's Knockout Guide Design with the followinginputs:

(I) ROS1

1) Genome – Homo sapiens – Ensembl GRCh38 (GenomeReference Consortium Human Build 38.
2) Gene – ROS1 – 6098 ENSG00000047936ROS proto-oncogene 1, receptor tyrosine kinase.
3) Nuclease – SpCas9 – *Streptococcus pyogenes*.

(II) EGFR

1) Genome – Homo sapiens – Ensembl GRCh38 (Genome Reference Consortium Human Build 38.
2) Gene – EGFR – 1956 ENSG00000146648 ROS epidermal growth factor receptor.
3) Nuclease – SpCas9 – *Streptococcus pyogenes*.

B. Synthego Output

The output from Synthego gives us a clear comparative study of the possibleguides after running them through (I) Knockout guide structure(ii) Verifying sgRNA plan and (iii) ICE Analysis. Thiseffective apparatus recommends to us the best gRNA grouping relying upon the genome of use and the quality that we are attempting to control and can be thus used to configure the data directRNA. It likewise gives us a visual interface on each gRNAsuccession's on track versus the off-target score and positions themfrom the most noteworthy effectiveness to least for that specificquality. One can likewise arrange the gRNA groupings onlinefrom Synthego to be conveyed to their lab.

I. Best guides for Breast Cancer CRISPR Control

TABLE 1 attached below contains the list of all possible guideRNA combinations to bring in the 4 best resultspossible for using CRISPR to mediate a clinical trial for the EGFR gene. **TABLE 2** attached below contains the list of all possible guideRNA combinations to bring in the 4 best results possible for using CRISPR to mediate a clinical trial for the ROS1 gene.The results show 4 top-rated guideRNAs for editing ROS1 and EGFR genes, the target sequences that are relevant for cleavage bythat gRNA, also 4. It also shows the respective protein-codinggenes for that sequence, the chromosome number in parallelalong with the cute site and the PAM region. A comprehensiveschematic representation of the mechanism, site, sequence,and the PAM is also available for reference.The gRNAs are also available for purchase directly whichis a huge advantage of the website for planning and carryingout the experiments.

Table 1:- *Analysis of EGFR gene using SYNTHEGO software*

UUACUCGUGCCUUGGCAAAC		
Best off-target sites	***Chr no'***	***PAM***
TTGCTCGTTCCTGGGCAAAC		AGG
TTACTCATTCCTTGGCAGAC	22	TGG
TTATGCTTCCCTTGGCAAAC	18 3	AGG
TTAGTCCTGCCTAGGCAGAC	3	TGG
CUUUUUCUUCCAGUUUGCCA		
Best off-target sites	***Chr no'***	***PAM***
CTTTTTCTTCCAGTTTCCTA		AGG
CTTTTTCCTCCTGTTTGCCT	8	TGG
CTTTGTTCTCCAGTTTGCCA	2	TGG
TTTTGTCTTCTAGTTTGCCA	2 15	GGG
UGAGCUUGUUACUCGUGCCU		
Best off-target sites	***Chr no'***	***PAM***
TGAGCTTGTTACTCGTGCCT		TGG
AGAGCTTGTTACTTGTGCCC	7	TGG
TGATCTTGTTCCTCCTGACT	9 X	GGG
TGAGCTTGGCACTGGAGCCT	1	AGG
GAGUAACAAGCUCACGCAGU		
Best off-target sites	***Chr no'***	***PAM***
GAGTAACAAGCTCACGCAGT		TGG
GACTGACACGCTCACGCAGT	7 19	GGG
GAGCACCAAGCTCAAGCAGG	22 HSCHR20_1_CTG3	TGG
GAGTAAGAAGCTCAGGCTGA		AGG

Table 2: Analysis of ROS1 gene using SYNTHEGO software

AAGCAAAGGGAGCAGUUGGU		
Best off-target sites	***Chr no'***	***PA***

		M
AAGCAGAGGGAGGAGCTGGT AACCAAAGGGATCAGTGGGT AAGCAAAAGGGGAAGTTGGT AAGCATAGGCAGCAGTGGGT	HSCHR22_1_CTG5 13 8 2	GGG GGG TGG GGG
CUUCCAAUGGAAGAAGCAAA		
Best off-target sites	***Chr no'***	***PAM***
CTTCCAACGGAAGAAGAAAA CATCCAAAGGAATAAGCAAA GTTCCAATGGAATTAGCAAA CTCCCAATGGAGGAAGCAGA	 8 5 7 4	AGG GGG CGG AGG
UGAGCUUGUUACUCGUGCCU		
Best off-target sites	***Chr no'***	***PAM***
CTTCCAACGGAAGAAGAAAA CATCCAAAGGAATAAGCAAA GTTCCAATGGAATTAGCAAA CTCCCAATGGAGGAAGCAGA	 8 5 7 4	AGG GGG CGG AGG
GAGUAACAAGCUCACGCAGU		
Best off-target sites	***Chr no'***	***PAM***
GCTTCCAAGGGAAGAAGCCA GTTTTCAAAGGAAGAAGCAA GCTTCCGATGGGGGAAGCAA GCTTTCAATGTAAAAAGCAA	 22 2 7 4	GGG AGG GGG AGG

PAM is one of the most crucial factors during the cleavage mechanism. The guideRNA recognizes the PAM site for cleavage so any error in the PAM site can lead to the entire experiment being unsuccessful.

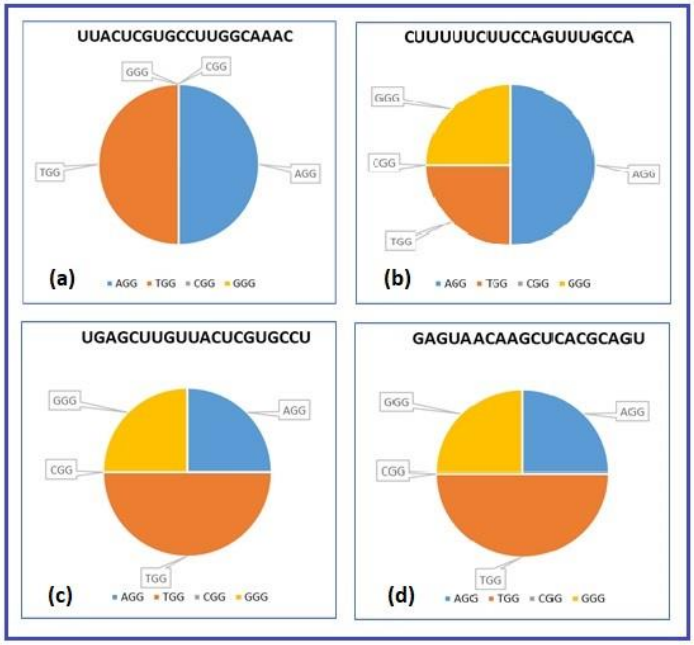

FIG 4 - PAM ratio for the off-target sites for the EGFR gene
(a)UUACUCGUGCCUUGGCAAAC
(b)CUUUUUCUUCCAGUUUGCCA
(c)UGAGCUUGUUACUCGUGCCU
(d)GAGUAACAAGCUCACGCAGU

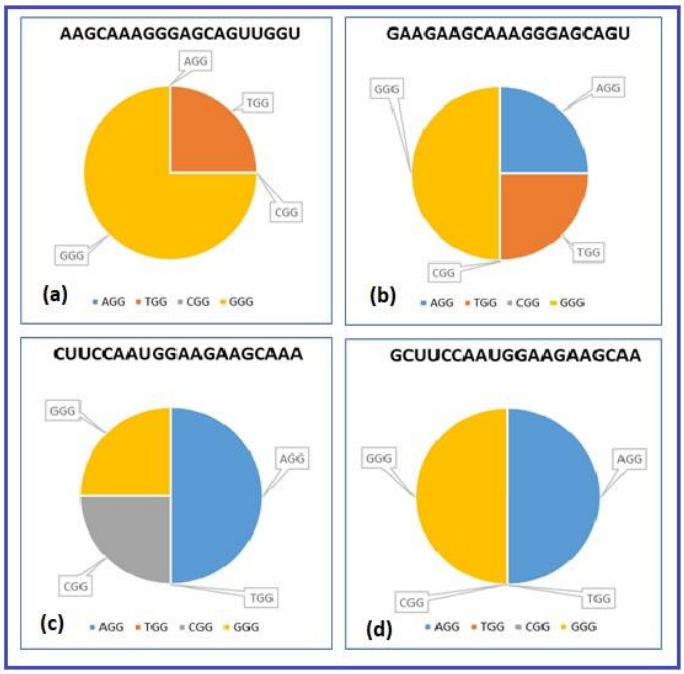

FIG 5 - PAM ratio for the off-target sites for the ROS1 gene (a) AAGCAAAGGGAGCAGUUGGU(b)GAAGAAGCAAAG GGAGCAGU(c) CUUCCAAUGGAAGAAGCAAA(d) GCUUCCAAUGGAAGAAGCAA

VI. CONCLUSION

Computational methodology towards life sciences might be a key factor in accomplishing the culminated mechanical utility to make the idea of DNA control proficient and fit to be moved from work area to bed for

down to earth executions. CRISPR has brought us ever so close to altering hereditary material as per our will cost-viably, with less intricacy and lesser negative marks than its predecessors, for example, ZFNs, TALENs, RNAi[25], and so forth.Combined with the online tool, Synthego, our paper uses the strength of Artificial Intelligence and Machine Learning to pinpoint some of the best sequence targets to aide Lung Cancer treatment using the CRISPR/Cas-9 mechanism. Increased efficacy and low off-target cleavage rates will pave way for CRISPR human clinical trials and could potentially open doors to higher grade genetic control in CRISPR babies [26] or higher organisms. Due to the COVID-19 pandemic and the unavailability of laboratory facilities, we opted to use the Computational Tools available at our disposal.We are working on the further up-gradation of CRISPR/Cas9 in Lung Cancer.

ACKNOWLEDGEMENTS

The Authors acknowledge the director and vice-chancellor of Techno India University to provide us the opportunity to work on the cutting edge technology of CRISPR/Cas-9 through computational modeling. Our sincere grattitude to the chancellor of Techno India University, Prof. (Dr.) Gautam Roychoudhury for providing us the opportunity to work in this key issue of scientific temperament.

REFERENCES

[1]. Who.int. 2020. *Cancer*. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/cancerLast Accessed September 6, 2020.
[2]. Barrera, R. and Morales Fuentes, J., 2012. Lung cancer in women. *Lung Cancer: Targets and Therapy*, p.79.DOI: 10.2147/LCTT.S37319
[3]. Couraud, S., Zalcman, G., Milleron, B., Morin, F. and Souquet, P., 2012. Lung cancer in never smokers – A review. European Journal of Cancer, 48(9), pp.1299-1311. DOI: 10.1016/j.ejca.2012.03.007
[4]. Liu, G., Pei, F., Yang, F., Li, L., Amin, A., Liu, S., Buchan, J. and Cho, W., 2017. Role of Autophagy and Apoptosis in Non-Small-Cell Lung Cancer. International Journal of Molecular Sciences, 18(2), p.367. DOI: 10.3390/ijms18020367
[5]. Besaratinia, A. and Pfeifer, G., 2008. Second-hand smoke and human lung cancer. The Lancet Oncology, 9(7), pp.657-666. DOI: 10.1016/S1470-2045(08)70172-4
[6]. Bubendorf, L., Büttner, R., Al-Dayel, F., Dietel, M., Elmberger, G., Kerr, K., López-Ríos, F., Marchetti, A., Öz, B., Pauwels, P., Penault-Llorca, F., Rossi, G., Ryška, A. and Thunnissen, E., 2016. Testing for ROS1 in non-small cell lung cancer: a review with recommendations. *Virchows Archiv*, 469(5), pp.489-503.DOI: 10.1007/s00428-016-2000-3
[7]. Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., MacNeill, J., Ren, J., Yuan, J., Bakalarski, C., Villen, J., Kornhauser, J., Smith, B., Li, D., Zhou, X., Gygi, S., Gu, T., Polakiewicz, R., Rush, J. and Comb, M., 2007. Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. *Cell*, 131(6), pp.1190-1203.DOI: 10.1016/j.cell.2007.11.025
[8]. Al Olayan, A., Al Hussaini, H. and Rahman Jazieh, A., 2012. The roles of epidermal growth factor receptor (EGFR) inhibitors in the management of lung cancer. *Journal of Infection and Public Health*, 5, pp.S50-S60.DOI: 10.1016/j.jiph.2012.09.004
[9]. Ellison, G., Zhu, G., Moulis, A., Dearden, S., Speake, G. and McCormack, R., 2020. *EGFR Mutation Testing In Lung Cancer: A Review Of Available Methods And Their Use For Analysis Of Tumour Tissue And Cytology Samples*.DOI: 10.1136/jclinpath-2012-201194
[10]. Nagarajan, L., Louie, E., Tsujimoto, Y., Balduzzi, P., Huebner, K. and Croce, C., 1986. The human c-ros gene (ROS) is located at chromosome region 6q16----6q22. *Proceedings of the National Academy of Sciences*, 83(17), pp.6568-6572.DOI: 10.1073/pnas.83.17.6568
[11]. Lung Cancer Foundation of America. 2020. What Is ROS1-Positive In Lung Cancer And How Is It Treated?. LCFA. [online] Available at: https://lcfamerica.org/lung-cancer-info/types-lung-cancer/ros1-positive-lung-cancer/#:~:text=ROS1%2Dpositive%20lung%20cancer%20occurs,which%20is%20what%20cancer%20isAccessed September 6, 2020.
[12]. Rothschild, S., 2015. Therapies in Non-Small Cell Lung Cancer—Beyond EGFR and ALK. *Cancers*, 7(2), pp.930-949.DOI: 10.3390/cancers7020816
[13]. Chin, L., Soo, R., Soong, R. and Ou, S., 2012. Targeting ROS1 with Anaplastic Lymphoma Kinase Inhibitors: A Promising Therapeutic Strategy for a Newly Defined Molecular Subset of Non–Small-Cell Lung Cancer. *Journal of Thoracic Oncology*, 7(11), pp.1625-1630.DOI: 10.1097/JTO.0b013e31826baf83
[14]. Davies, K. and Doebele, R., 2013. Molecular Pathways: ROS1 Fusion Proteins in Cancer. *Clinical Cancer Research*, 19(15), pp.4040-4045.DOI: 10.1158/1078-0432.CCR-12-2851
[15]. Gupta, R., Dastane, A., Forozan, F., Riley-Portuguez, A., Chung, F., Lopategui, J. and Marchevsky, A., 2008. Evaluation of EGFR abnormalities in patients with pulmonary adenocarcinoma: the need to test neoplasms with more than one method. *Modern Pathology*, 22(1), pp.128-133. DOI: 10.1038/modpathol.2008.182
[16]. Su, J., Zhong, W., Zhang, X., Huang, Y., Yan, H., Yang, J., Dong, Z., Xie, Z., Zhou, Q., Huang, X., Lu, D., Yan, W. and Wu, Y., 2017. Molecular characteristics and clinical outcomes of EGFR exon 19 indel subtypes to EGFR TKIs in NSCLC patients. Oncotarget, 8(67), pp.111246-111257.DOI: 10.18632/oncotarget.22768
[17]. Castellanos, E., Feld, E. and Horn, L., 2017. Driven by Mutations: The Predictive Value of Mutation Subtype in EGFR -Mutated Non–Small Cell Lung

Cancer. *Journal of Thoracic Oncology*, 12(4), pp.612-623. DOI: 10.1016/j.jtho.2016.12.014
[18]. Oda, K., Matsuoka, Y., Funahashi, A. and Kitano, H., 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1(1).DOI: 10.1038/msb4100014
[19]. Yarden, Y. and Shilo, B., 2007. SnapShot: EGFR Signaling Pathway. *Cell*, 131(5), pp.1018.e1-1018.e2.DOI: 10.1016/j.cell.2007.11.013
[20]. Scaltriti, M. and Baselga, J., 2006. The Epidermal Growth Factor Receptor Pathway: A Model for Targeted Therapy. Clinical Cancer Research, 12(18), pp.5268-5272.DOI: 10.1158/1078-0432.CCR-05-1554
[21]. Doudna, J. and Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213), p.1258096. DOI: 10.1126/science.1258096
[22]. Plaza Reyes, A. and Lanner, F., 2017. Towards a CRISPR view of early human development: applications, limitations and ethical concerns of genome editing in human embryos. *Development*, 144(1), pp.3-7. DOI: 10.1242/dev.139683
[23]. De, A. and Biswas, A., 2020. Elucidative PAM/Target Sequence for CRISPR/Cas- 9 Activity in Breast Cancer Using a Computational Approach. *International Journal of Innovative Science and Research Technology*, 5(7), pp.872-876.DOI: 10.38124/IJISRT20JUL757
[24]. Design.synthego.com. 2020. *Synthego*. [online] Available at: https://design.synthego.com/#/Accessed September 6, 2020.
[25]. Gaj, T., Gersbach, C. and Barbas, C., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31(7), pp.397-405.DOI: 10.1016/j.tibtech.2013.04.004
[26]. Sand, M., Bredenoord, A. and Jongsma, K., 2019. After the fact—the case of CRISPR babies. *European Journal of Human Genetics*, 27(11), pp.1621-1624.DOI: 10.1038/s41431-019-0459-5