

Elucidative PAM/Target Sequence for CRISPR/Cas-9 Activity in Breast Cancer Using a Computational Approach

Aniket De

2ND Year Student, Department of Biotechnology Techno India University
Kolkata, India

Dr. Arup Ratan Biswas*

Assistant Professor and Head, Department of Chemistry
Techno India University
Kolkata, India

Abstract:- Directed genomic area for release and modification are indicated by the nearness of nuclease explicit PAM sequence. By hindering the signs, Herceptin can slow or stop bosom malignant growth development. The outcomes show 4 top of the line direct RNAs for altering HER2 quality, the objective arrangements that are pertinent for cleavage by that gRNA, likewise 4. The central clarification for the terrible sign of cockeyed cleavage is accepted to be in the arranging of the single guide RNA for the Cas-9 protein. This is extremely basic for researchers to know while planning guide RNA. Albeit the two guys and females are inclined to bosom malignant growth, it is more probable for females to build up this sort of disease. Be that as it may, with such a large number of HER2 receptors, bosom malignant growth cells can assemble an excessive number of development signals. The most widely recognized qualities engaged with bosom malignancy incorporate HER2, BRCA1, and BRCA2. In this manner, we propose the genomic approach of controlling the HER2 quality guided by CRISPR/Cas-9 to guarantee lesser symptoms and increasingly powerful treatment.

Keywords:- CRISPR; Cas-9; Breast cancer; Herceptin; Synthego; Guide RNA.

I. INTRODUCTION

Breast Cancer has been identified as the second most common form of cancer after Skin Cancer and an estimate of 268,000 cases in the United States alone. Although both males and females are prone to breast cancer, it is more likely for females to develop this type of cancer. The prime cause of any type of cancer is uncontrolled cell growth and in the case of breast cancer is the abnormally rapid division of the healthy breast cells which leads to the formation of a lump. These infected cells later spread throughout the body via the lymph nodes present in the breast. There are various types of Breast Cancer such as Angriosarcoma, Ductal carcinoma in situ, Inflammatory breast cancer, Invasive lobular carcinoma, Lobular carcinoma in situ, Recurrent breast cancer and so on and each type is caused by a significant mutation in a relative gene. The most common genes involved in breast cancer include HER2[1], BRCA1, and BRCA2. Interestingly, clinicians have identified that 5-10% of the cases of breast cancer may be due to the result of mutated gene inheritance which increases the likelihood of breast cancer. Scientists have also pinpointed some non-

inheritable risk factors of breast cancer which include natural aging, any clinical biopsy that indicated LCIS, a family history of breast cancer, exposure to radiation, obesity, the introduction of menopause at a later age than normal, conceiving at an age beyond 30 years, excessive alcohol consumption, etc. Some identifiable symptoms of breast cancer include a thick lump around the breast, a sudden alteration in the structure of the breast, an inverted nipple, peeling or scaling of areolar tissues around the nipple, redness of the skin. While modern-day medicine has certainly improved the lifestyle and ensured a much higher rate of survival of breast cancer patients, it is still at a lack once cancer has spread. Estrogen blocking medications and Chemotherapy have certainly paved the way for better treatment but this often harms the surrounding cells causing a long term effect. In this paper, we elucidate a genome manipulation based approach towards Breast Cancer using CRISPR/Cas-9 Technology to edit the Herceptin gene using the most effective guide RNA sequences to control and reduce the effects of mutation leading to cancer itself.

II. ROLE OF HERCEPTIN IN BREAST CANCER

Herceptin is a monoclonal antibody that can treat early stages of breast and lung cancer by targeting an area of interest. However, Herceptin treatment is only useful if the cancer is HER2-positive. Herceptin has been approved by the FDA for various functions such as hindering in the metastasis and reducing the growth of breast cancer in the body, often used in along with chemotherapy as an addition to block the chances of recurrence, it can even be combined with other antibodies such as Perjeta and Taxotere to treat early stages of HER2-positive cancer. Cancer cells are growing in an uncontrolled mode. Herceptin works on the cancer cell surface by blocking the chemical signals that might stimulate this uncontrolled growth. Genes are like instruction manuals telling every cell in your body how to grow, what sort of cell to become, and how to act. Genes do this by ordering the cell to make special proteins that cause some activity — such as cell growth, rest, or repair. Some cancer cells have gene abnormalities that tell the cell how much and how quickly to grow[2].

The cancer cells sometimes have too many abnormal copies of those genes. Doctors refer to it as "over-expression" when there are too many copies of these genes. With some forms of gene overexpression, cancer cells make up too many of the proteins that control cell growth and division, causing cancer to grow and spread. Some breast cancer cells make too many copies of a specific

gene, known as HER2, (overexpress). The gene HER2 produces a protein known as the receptor HER2. HER2 receptors on the surface of all cells are like ears or antennae. These HER2 receptors receive signals stimulating cell growth and multiplication. But, with too many HER2 receptors, breast cancer cells can gather too many growth signals. This causes them to start growing too much and to multiply too quickly. It is said that the breast cancer cells which overexpress the HER2 gene are HER2-positive. Herceptin works by attaching itself to the HER2 receptors on the surface of carcinoma cells and by stopping them from receiving growth signals. By blocking the signals, Herceptin can slow or halt breast cancer growth. Herceptin is an example of a controlled drug for the immune system.

Besides blocking HER2 receptors, Herceptin can also help combat breast cancer by alerting the immune system to destroy cancer cells on which it is attached. It is also important to note that Herceptin is not to be used by everyone because certain test criteria need to match which would confirm whether the patient should use it or not. Two of these tests are ImmunoHistoChemistry (IHC) and Fluorescence In Situ Hybridization (FISH) where the IHC measures surface amount of HER2 proteins on the target region using chemical stains and any score beyond 3.0+ would indicate the presence of HER2 positive cancer[3] and would allow the use of Herceptin whereas the FISH test uses synthesized fluorescent labels which can get attached to the HER2 proteins and start emitting fluorescence upon coming in contact but this method is time-consuming and more expensive which is why IHC is the common test although, FISH is more accurate. Lastly, Herceptin medication also causes side effects such as nausea, fever, infections, insomnia, headache, rashes, coughing, etc. Thus, we propose the genomic approach of manipulating the HER2 gene guided by CRISPR/Cas-9 to ensure lesser side effects and more effective treatment.

III. CRISPR-CAS CLEAVAGE MECHANISM

The CRISPR Cas9 protein functions as a pair of molecular scissors that can cut DNA at different positions and thereby alter the gene at the appropriate sites. The Cas9 protein is attached to the artificially generated RNA guide by mixing the chemically synthesized trans-activating CRISPR RNA (tracrRNA) with a chemically synthesized CRISPR RNA (crRNA) or a single RNA guide (sgRNA) consisting of both tracrRNA and crRNA as a single build. The RNA guidance also includes a different set of guidelines that instructs the Cas9 to the particular location where the DNA fragment will be cut off. Upon injecting the Cas9 into the genome, it searches the whole section for the intended site, and the others are called off-target sites[4].

When comparing the sequence with the gRNA, it unwinds the portion of the DNA and tests if the sequence is complementary to that of the guide RNA if it is not complimentary to start scanning Cas9, however, if the strand is complementary to all the bases of the guide RNA, then Cas9 cuts the DNA at that particular spot. In order to fix the harm done by the break, the cell also attempts to attach them directly together which causes frameshift mutation and causes gene expression disruption[5]. In addition, the cell can also seek to repair harm without

inducing mutation, so that can be carried out using a specific matching sequence, scientists may create the required genetic modifications for the same sequence and a cell tries to insert into the DNA portion a portion of the provided section that triggers the requisite alteration.

A. PAM Recognition

The neighboring protospacer motif (or PAM for short) is a small DNA sequence of 2 to 6 base pairs in length that must be present in the DNA area that the CRISPR / Cas 9 endonuclease method requires to target for cleavage. PAM's position is located 3 to 4 downstream nucleotides from the cut spot. The presence of nuclease-specific PAM sequence[6] shall specify the targeted genomic location for editing and alteration. CRISPR controlled version of a single genomic position is limited by the existence of PAM sequences unique to nuclease. From *Streptococcus Pyrogenes* it is evident that Cas-9 recognizes sequence 5'-nGG-3' without the presence of the GG sequence. The existence of a G-consensus sequence contributes to a disturbance of the endonuclease function of the Cas-9 enzyme. Analysis data suggests that Cas-9 endonuclease does not recognize a single PAM, but it recognizes multiple PAM sequences originating from various bacterial organisms, ultimately allowing the researcher select a wide variety of nuclease for CRISPR action, thus creating multi-targets and editing facilities.

B. Single guide RNA (SgRNA) and its need

Being a pair of molecular scissors, the Cas-9 protein needs an indicator as to which portion of the DNA sequence to cleave and this is where the usage of single guide RNA (SgRNA) falls in. As the Cas9 cuts and stores the viral DNA for repeated CRISPR, it removes the PAM sequence to ensure that the bacterial genome is not viewed as a target[7]. This is really relevant for scientists to learn while developing the RNA guide. The RNA guide is a chimerical complex consisting of the trans-activating CRISPR RNA scaffold sequence and the custom-designed CRISPR RNA sequence. Nevertheless, the Cas-9 SgRNA may be known to be almost a required evil, since studies have shown that it affects cleave performance explicitly and is also liable for off-target cleavage.

C. Demerits of CRISPR/Cas-9

The effect of askew can alter the efficiency potential which can result in genomic precarity, preclude it from being prepared which extends to the clinical process. The synthesis of CRISPR RNA (CrRNA) and trans-acting RNA (TrRNA) is also known as an illusory RNA. The 20 base nucleotides of SgRNA are analogous to the objective DNA of Trinucleotide Enthusiasm known as Protospacer Adjacent Motif (PAM) contiguous to a structure that is, for the most part, NGG (where N can be Adenine, Guanine, Cytosine, or Thymine)[8]. Specific inquiries have confirmed that the PAM could indeed be NRG (where R is Adenine or Guanine). Although DNA complimentary is highly directly based on base pair rule (A-T and G-C), the distal portion of the PAM still has an off-focus case with 3' to 5' discrepancies. Wherever X20 is CrRNA, SgRNA 5' PAM 3' contains each non-seed and seed grouping with a variety of investigations revealing that the seed arrangement (for the first half consisting of 10-12 base sets

adjacent to the PAM) controls the lucidity of polymer radio- controlled endonucleases (RGENs) and is considered to be additional vast than non-seed successions.

IV. COMPUTATIONAL TOOL BASED APPROACH

The negative marks are a prime explanation representing the wastefulness of CRISPR in its space. Be that as it may, with present-day computational instruments, the system of activity of CRISPR was improved as well as its plausible results were likewise anticipated all the more precisely. 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. The essential bad mark of Cas- 9 is that it divides askew DNA thus to counter that, analysts began executing AI calculations utilizing computational instruments to develop a progressively exact cleavage result and disposing of the off-target bad marks. 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. The basic reason for the poor mark of the Askew cleavage is assumed to be the preparation of the single reference RNA for the Cas9 protein. Of the well-known analytical methods, Synthego[9] is considered as the most innovative solution because of 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 are very common phenomena that tend to hinder the desired result of our DNA manipulation.

A. Experimental Setup in Synthego

We analyzed our respective gene ERBB2, for breast cancer, using Synthego’s Knockout Guide Design with the following inputs:

- Genome – *Homo sapiens* – Ensembl GRCh38 (Genome Reference Consortium Human Build 38).
- Gene – ERBB2 – 2064 ENSG00000141736 erb-b2 receptor tyrosine kinase 2.
- Nuclease – SpCas9 – *Streptococcus pyrogenes*.

B. Synthego Output

The output from Synthego, an online instrument concocted by MIT, gives us a clear comparative study of the possible guides after running them through (i) Knockout guide structure (ii) Verifying sgRNA plan and (iii) ICE Analysis. This effective apparatus can be used to configure the data direct RNA as it recommends to us the best gRNA grouping relying upon the genome of use and the quality that we are attempting to control. It likewise gives us a visual interface on each gRNA succession's on track versus off-target score and positions them from the most noteworthy effectiveness to least for that specific quality. One can likewise arrange the gRNA groupings online from Synthego to be conveyed to their lab.

C. Best guides for Breast Cancer CRISPR Control

The TABLE1 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.

The results show 4 top-rated guideRNAs for editing HER2 gene, the target sequences that are relevant for cleavage by that gRNA, also 4. It also shows the respective protein-coding genes for that sequence, the chromosome number in parallel along with the cut site and the PAM region. A comprehensive schematic representation of the mechanism, site, sequence, and the PAM is also available for reference.

The gRNAs are also available for purchase directly which is a huge advantage of the website for planning and carrying out your experiments.

CUGCACCUCCUGGAUAUCCU		
<i>Best off-target sites</i>	<i>Chr no'</i>	<i>PAM</i>
CTGGATCTCCTGGAAATCCT	17	TGG
CAGCACATCTTGGATATCCT	17	GGG
CTCCACCTCCTGGGTCTCCT	19	GGG
CTGCACCTGCTGGAGCTCCT	1	GGG
UCCCCCUCCAGGAUAUCC		
<i>Best off-target sites</i>	<i>Chr no'</i>	<i>PAM</i>
CTCCACCTCCCAGGATATCC	8	AGG
TTCCCCCTACCAGTATATCT	12	GGG
TTCCCCCTCCAAAAATATCC	2	TGG
TGCCCCCTCCCAGGTTATGC	13	TGG
CCCCUCCAGGAUAUCCAGG		
<i>Best off-target sites</i>	<i>Chr no'</i>	<i>PAM</i>
CCCCTCCCAGAATAGCCAGG	10	AGG
GCCCTCCCAGGTTATCCAGG	12	TGG
TGCCTCCCAGGCTATCCAGG	19	AGG
CCCCTCCCAGGATATGCAAC	5	AGG
GCACGUAGCCUGCACCUC		

<i>Best off-target sites</i>	<i>Chr no'</i>	<i>PAM</i>
GCACGTAGCACTGCAACTCC	15	AGG
GCACGTGGCCCTGCAGCACC	15	AGG
GCAGACAGCCCTGCACTCC	16	GGG
TCACATAGCCCAGCACCTCC	16	TGG

Table 1:- Synthego Software Analysis

An interesting thing to note here would be the PAM ratios for the guideRNA targeting. 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.

Let's look at the PAM ratio with the help of pie charts in the figures below.

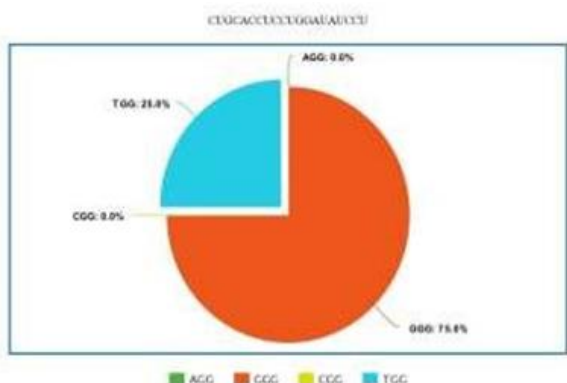


Fig 1:- CUGCACCUCUGGAUAUCCU

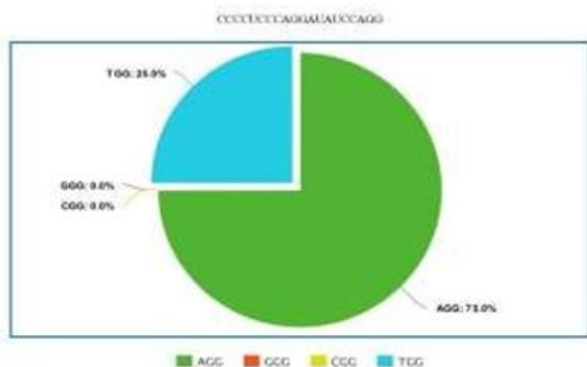


Fig 2:- CCCUCCCCAGGAUAUCCAGG

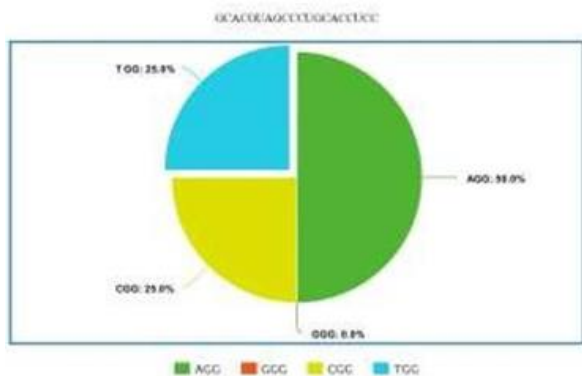


Fig 3:- GCACGUAGCCCUGCACCUCC

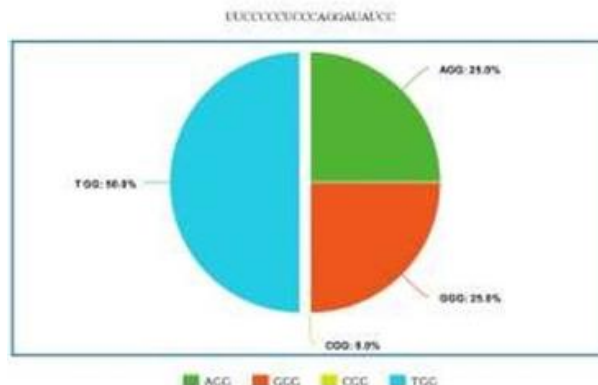


Fig 4:- UUCCCCCUCCCAGGAUAUCC

V. CONCLUSION

After cautious evaluation of all the AI calculations out there for CRISPR Technology, we can learn that computational methodology towards life sciences might be a key factor to accomplishing the culminated mechanical utility that can make the idea of DNA control proficient and fit to be moved from work area to bed for down to earth executions. CRISPR is the nearest we have ever been to alter hereditary material as per our will cost-viably, with less intricacy and lesser negative marks than it's predecessors, for example, ZFNs, TALENs, RNAi[10], and so forth. Our paper uses the strength of AI and Machine Learning combined online tool, Synthego, to pinpoint some of the best sequence targets to aide Breast 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[11] or higher organisms.

REFERENCES

- [1]. Wen, Wenhsiang et al. "Mutations in the Kinase Domain of the HER2/ERBB2 Gene Identified in a Wide Variety of Human Cancers." The Journal of molecular diagnostics: JMD vol. 17,5 (2015): 487-95. doi:10.1016/j.jmoldx.2015.04.003
- [2]. Khalili, P., Arakelian, A., Chen, G. et al. Effect of Herceptin on the development and progression of skeletal metastases in a xenograft model of human breast cancer. Oncogene 24, 6657–6666 (2005). https://doi.org/10.1038/sj.onc.1208790
- [3]. Gown, A. Current issues in ER and HER2 testing by IHC in breast cancer.
- [4]. Doudna, J. A. et. all 2014 Genome editing. The new frontier of genome engineering with CRISPR-Cas9.
- [5]. Mojica, F. J. et. all 2009 Short motif sequences determine the targets of the prokaryotic CRISPR defense system

- [6]. Kleinstiver, B., Prew, M., Tsai, S. et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* 33, 1293–1298 (2015). <https://doi.org/10.1038/nbt.3404>
- [7]. 14 Sep 2015: Prykhozhij SV, Rajan V, Gaston D, Berman JN (2015) Correction: CRISPR MultiTargeter: A Web Tool to Find Common and Unique CRISPR Single Guide RNA Targets in a Set of Similar Sequences. *PLOS ONE* 10(9): e0138634. <https://doi.org/10.1371/journal.pone.0138634>
- [8]. Yanni Lin, Thomas J. Cradick, Matthew T. Brown, Harshavardhan Deshmukh, Piyush Ranjan, Neha Sarode, Brian M. Wile, Paula M. Vertino, Frank J. Stewart, Gang Bao, CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences, *Nucleic Acids Research*, Volume 42, Issue 11, 17 June 2014, Pages 7473–7485, <https://doi.org/10.1093/nar/gku402>
- [9]. <http://design.synthego.com>
- [10]. Gaj, T. et. al 2013 ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.
- [11]. Sand, M., Bredenoord, A.L. & Jongsma, K.R. After the fact—the case of CRISPR babies. *Eur J Hum Genet* 27, 1621–1624 (2019). <https://doi.org/10.1038/s41431-019-0459-5>