CRISPR/Cas-9 Genetic Editing of 'ROS1' gene Using Computational tool

Manav Goenka
Department of Biotechnology
Techno India University, West Bengal
manav.gnk@gmail.com

Aniket De
Department of Biotechnology
Techno India University, West Bengal
aniketde9@gmail.com

Arup Ratan Biswas*
Corresponding Author:
Department of Chemistry,
Techno India University, West Bengal
hod.chemisty@technoindiaeducation.com

***Abstract*:**

One of the most promising genome-editing tools available now is the CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein-9) system, discovered in 2012 by Nobel Prize-winning Laureates Jennifer Doudna and Emmanuelle Charpentier. It is a bacterial defensive mechanism that exhibits the cleavage of genomic DNA at the desired location, resulting in the exit of the old genes and the induction of a new set of genes. The accuracy, precision or fidelity of the genetic cut depends on the target and the proto-spacer adjacent motif (PAM) sequences. The target sequence is 20 bases long and belongs to a particular CRISPR locus on a crRNA array. The Cas9 protein recognises the PAM sequence (5'-NGG-3') by selecting the correct location of base-pair bonds within the target sequence on the host genome. Assembling the nucleotide sequence related to PAM and target sequence into a plasmid and then transfecting the plasmid into a cell shows that Cas9 with the help of a crRNA detected the correct sequence within a host cell. This resulted in a single or double-stranded break at the appropriate location in the DNA, thereby working as a molecular scissor and performing a genetic cut. We choreographed this tool in achieving the information related to the generation of the PAM sequence and the off-target sites associated with the ROS1 gene responsible for lung cancer prognosis. To achieve the same, we investigated the said gene on preformed online software tools available like that of CCTop and SYNTHEGO to generate the best possible target sequence along with their appropriate guide RNAs. Furthermore, an approach has been made to establish the protein characteristics related to the generation of hydropathy index and the polarity.

Experimental Design: The amino acid sequence of the ROS1 protein with its accession number was obtained from Expasy. The FASTA file of the nucleotide sequence of each amino acid was scanned batch mode in SYNTHEGO and CCTop computational tools respectively.

Result: We have identified the top 4 best gRNA sequences based on the highest SYNTHEGO scores range 0.98 to 0.99. Furthermore, we used CCTop to break down the entire sequence into several target sequences and also to find out a guideRNA corresponding to each target sequence and accordingly we have identified the 4 best target and guide RNA sequences with highest efficacy score.

Conclusion: Our manuscript is aimed at showcasing the best target sequence and guideRNA sequence complimentary to the target sequence utilizing the model software like that of SYNTHEGO and CCTop.

Keywords— CRISPR/Cas9; gRNA; molecular scissor; ROS1 gene; SYNTHEGO and CCTop computational tool.

I. INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR) – CRISPR-associated protein 9 (Cas9) system is a bacterial defence mechanism against phage infection. The system is a component of the adaptive immunity in bacteria against viruses and plasmids. This method has had successful applications in biological systems ranging from yeasts to rodents and mammals and thus, has intentionally been used as a powerful RNA-guided DNA targeting platform for genome editing, transcriptional perturbation, epigenetic modulation, and genome imaging. [1] This technology allows precise manipulation of any genomic sequence specified by a short stretch of guide RNA, allowing elucidation of gene function involved in disease development and progressions, correction of disease-causing mutations, and inactivation of activated oncogenes or activation of deactivated cancer suppressor genes when utilizing a fusion protein of nuclease-deficient Cas9 and effector domain. [2, 3] CRISPR based genome-wide screens can be leveraged using single-guide RNA (sgRNA) libraries for the identification of drug-target or disease-resistance genes, such as novel tumour suppressors or oncogenes, and to quickly assess drug targets [4, 5].

CRISPR/Cas9 endonuclease system is currently targeted as a molecular surgery tool to achieve success in cancer treatment. Cancers are mostly related with genetic alteration and mismatches in cell cycle checkpoints. The tumour suppressor genes and proteins play a crucial role in controlling the cell cycle. Mutation in any of the checkpoints or the tumour suppressor gene may change the scenario and may cause a chaotic situation to an individual's life, causing a clinical manifestation leading to cancer. Briefly, Cas9 locates specific 20-base-pair (bp) target sequences within the genomes that are billions of base pairs long and subsequently induces sequence-specific double-stranded DNA (dsDNA) cleavage.[3] In this manuscript we are highlighting the genetic changes associated with the ROS1 oncogene enhancement, also known as MCF-3, an integral membrane protein. Studies revealed rearrangements of the ROS1 protein tyrosine kinase is an extremely rare event in the case of non-small cell lung cancer (NSCLC) cases. [6] Amplification leads to higher rates of ROS1-mRNA and protein synthesis. Furthermore, as ROS1 gene encodes a transmembrane protein of the same name, we envisaged the characteristic properties of the ROS1 protein so as to get the holistic picture of the gene as well as the protein associated with lung carcinoma. Our main approach is to highlight the CRISPR/Cas9 technology as a novel genetic cutting tool that may be employed in editing the target region of the ROS1 gene and at the same time, to give an overall view of the

physical properties associated with the protein encoded by the ROS1 gene.

Despite of its full potential, this genetic cutting tool has its own demerits to the point that it cannot be used in full therapeutic measures since it is indeed a matter of further investigation regarding the precision and accuracy of cutting the target sequence and not generating the off targets, which may cause genetic damage instead of it achieving the goal. Even though of its demerits it is now the buzz of frontier research and had opened a new dimension in the field of cancer biology research in the name of a new discipline termed as CRISPR biology. This manuscript stands over the others since it highlights the latest tool ever developed in the treatment of any form of cancer and here to specify is the lung cancer.

II. ROS1 ROLE AND PATHWAY IN LUNG CANCER

ROS1 or, c-ros oncogene 1 is a part of the receptor tyrosine kinase family and is located in chromosome 6. [7] The gene is analogous to the ALK protein and till date, no receptor or ligand has been determined for the wild-type ROS1 gene. [8]. The main cause for lung carcinoma is fusion of the ROS1 protein with other molecule genes such as CD74, EZR, MSN or FIG, leading to a permanently "on" phosphorylation cascade. This fusion leads to auto-phosphorylation of ROS1 and involving the SHP, MEK, ERK, STAT3, and AKT molecules.

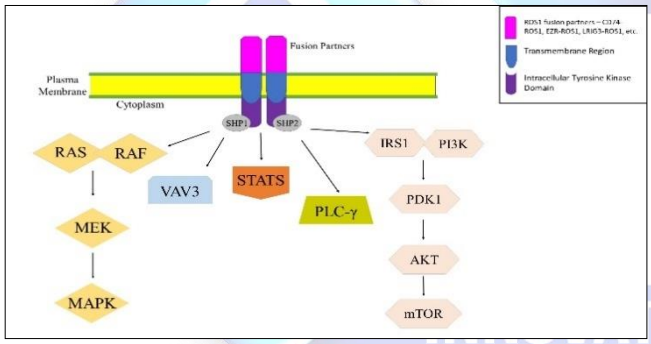

Figure 1: ROS1 signalling pathways. [9] ROS1 fuses with a nearby gene and leads to phosphorylation of the SHP molecules attached in the intracellular TK domain. This phosphorylation leads to further activation of the various other oncogenic pathways.

III. MECHANISM OF CRISPR/CAS-9 ACTION

The CRISPR/Cas system is RNA mediated and relies on small RNA sequences (approx. 20-22 nucleotides long) for detection and silencing of foreign DNA in a site-specific manner. They use a non-specific endonuclease to cut a genomic sequence. A small guide RNA (gRNA) guides the Cas protein to a specific site. [9] The Cas protein in an endonuclease which by definition means that it cuts a specific stretch of nucleotides within the nucleic acid. It is guided by a short nucleotide guide RNA (or gRNA) which are approximately 20-22 nucleotides long, to locate the complementary protospacer DNA target in a genome.

The defense mechanism of CRISPR/Cas9 involves 3 distinct steps [10]: adaptation of the CRISPRs, genesis of crRNA and lastly, silencing of the foreign DNA. In the adaptation step, the Cas operon transcribes the cas1-cas2 complex which chooses a portion of the foreign DNA to integrate into the host's CRISPR arrangement. This copy is called the spacer sequence and the protospacer selected by the adaptation machinery is usually compatible with the PAM sequence of the silencing machinery. This sequence is integrated to the immediate downstream to the leader sequence with a record of the previous infections. [11] This is followed by the step of crRNA genesis where the crRNAs are transcribed and matured. In numerous organisms, continuous production of crRNA and Cas9 proteins takes place, operating in a 'surveillance mode'. In specific strains of E. coli, foreign presence also triggers an elevation in expression of the complex. [10] Finally, the crRNAs are loaded onto the final effector complex and guided to the invading DNA by the recognition of a PAM sequence. The Cas9 complex then cuts the double strand specifically 3 base pairs upstream to the PAM site.

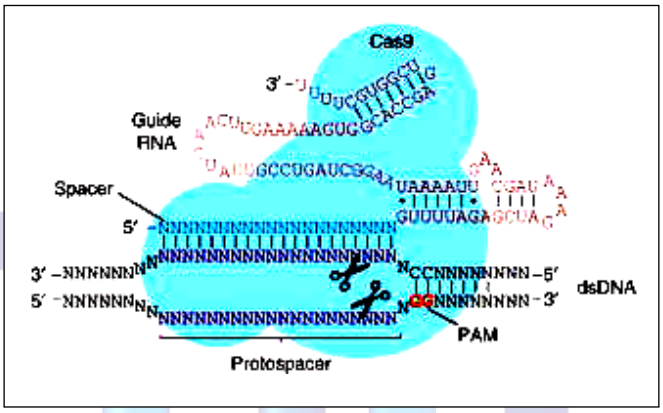

Figure 2: CRISPR/Cas-9 showing as molecular scissor

IV. RESEARCH OBJECTIVE

Our research is focused in elucidating the target sequence associated with the ROS1 gene responsible for the clinical manifestation of lung cancer or more specifically, NSCLC that may be possible recognized by the PAM sequence corresponding to the target sequence and employing the computational tool we have focused to elucidate the corresponding PAM and target sequence related to the ROS1 gene. At the same time, we also elucidated the protein properties in terms of its hydropathy index and polarity related to the ROS1 protein. Our manuscript gives the necessary information regarding further work in synthesizing the software based generated PAM sequence and transplanting the same in the plasmid vector and to check outthe interference of the generated PAM sequence with the single guide RNA for target identification and finally cessation of the target sequence and establishing the cut sequence with modified sequence which shall normalize the function of the ROS1 gene or rather deactivate the over expression of the ROS1 gene and thereby would diminish the cancer prognosis.

V. RESEARCH METHOD

The research method or the experimental set up for this manuscript is divided in to two parts.

Part-A mainly emphasizes the protein properties elucidated as protscale graphs related to hydropathy index and the polarity.

Part-B mainly emphasizes the computational tool-based approach undertaken to identify the target sequence, PAM sequence and SgRNA obtained from different online computational tool like that of CCTop and SYNTHEGO.

A. Elucidation of Protein Properties

A.1. Elucidation of Protscale Graphs

In order to properly understand and work with the ROS1 protein, we first examined and analyzed the chemical properties of the protein starting from its accession number, proteomics, properties such as hydropathy index and polarity

A.1.1. Hydropathy Index

An amino acid's hydropathy index is a number reflecting the sidechain's hydrophobic or hydrophilic properties [13]. The greater the number, the higher is the hydrophobic character of the amino acid. This property was proposed in 1982 by Jack Kyte and Russel F. Doolittle. [14]. The most hydrophobic amino acids are considered to be Isoleucine and Valine. Arginine and lysine are the most hydrophilic ones. It is very important in the composition of proteins; hydrophobic amino acids appear to be central (in terms of the 3-dimensional form of the protein [15]) while hydrophilic amino acids are more generally located on the protein surfaces. We have identified the amino acid sequences that show hydrophobic properties and it has been represented in **Figure 3**.

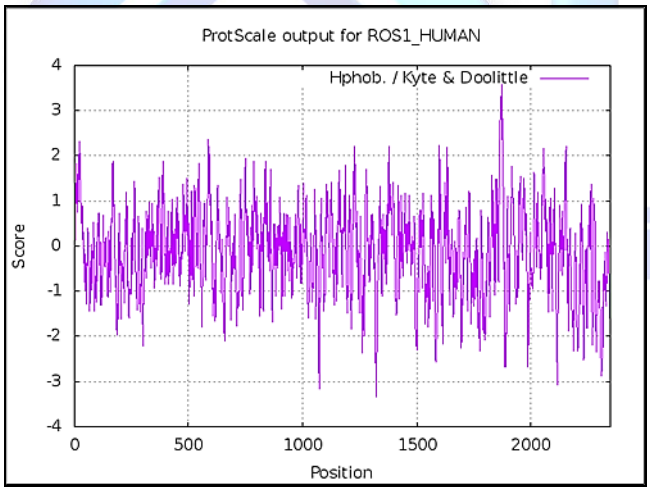

Figure 3: The hydrophobic property of ROS1 protein

A.1.2. Polarity

Polarity of a protein is the resultant of the electronegativity difference between the bonded atoms. A protein may be termed as being polar or nonpolar depending on the distribution of charges in its amino acid sequence. It is generally observed that amino acids with polar side groups are present on the protein surface while the non-polar amino acids constitute the interior core of the proteins. Polar amino acids tend to be hydrophilic with their non-polar counterparts being generally hydrophobic. J M Zimmerman [16] in 1968 attempted to use statistical methods by taking into consideration polarity and bulkiness of the protein as factors to determine the individual amino acid role within the protein configuration. We have identified the amino acid sequences that show polar properties and it has been represented in **Figure 4.**

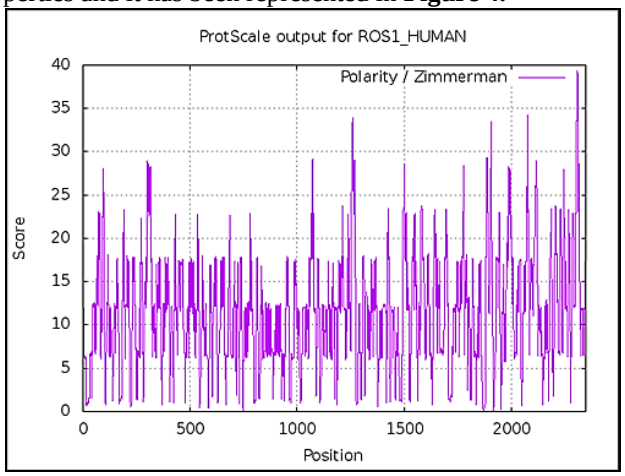

Figure 4: The polarity behaivour of ROS1 protein

B. Computation tool based approach to identify target sequence for CRISPR

At first, we worked out the amino acid sequence of the ROS1 protein with its accession number from ExPASy after which we individually looked at the nucleotide sequence of each amino acid and converted that into a FASTA file. Now this FASTA file was scanned in batch mode in two different software – SYNTHEGO and CCTop and the following factors were common for both the software before analysis.

Genome – Homo sapiens – Ensembl GRCh38 (Genome Reference Consortium Human Build 38).
Nuclease – SpCas9 – Streptococcus pyrogenes.

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. The calculation is based on information that has been extracted through various sources and the amalgamation of all this information can be utilized by the AI to predict 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 breakdown 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. Of the well-known analytical methods, SYNTHEGO [17] and CCTop [18, 19] are considered as the most innovative solutions 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.

C. Synthego Output

The output from Synthego 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 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 direct RNA. 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 specific quality. One can likewise arrange the gRNA groupings online from Synthego to be conveyed to their lab.

C.1. Experimental Setup in Synthego

We analysed the gene ROS1, for lung cancer using Synthego's Knockout Guide Design with the following inputs:

(I) ROS1

1) Genome – Homo sapiens – Ensembl GRCh38 (Genome Reference Consortium Human Build 38).

2) Gene – ROS1 – 6098 ENSG00000047936 ROS proto-oncogene 1, receptor tyrosine kinase.

3) Nuclease – SpCas9 – Streptococcus pyogenes.

D. CCTop Output

As for CCTop, it's not yet known to take into account the bulges and loops totally while analyzing the sequence, however, the output presented by CCTop is more detailed and organized when it comes to actual experimentation. The output is given by breaking down the entire sequence into several target sequences and suggesting a guideRNA corresponding to that. It is sorted according to the target sequence and the varying efficacy score that depends on their off-target activity and is presented with its oligo-pair extension coordinates, PAM, gene name of the corresponding sequence, and the gene id giving a higher control to the experimentation carrier.[20].

VI. RESULT

TABLE- 1 attached below contains the list of all possible guideRNA's where there is the possibility of mimicking genetic editing. The best possible guide RNA's (gRNA) for maximum Cas-9 activity are:AAGCAAAGGGAGCAGUUGGU,GAAGAAGCAAAGGG AGCAGU,CUUCCAAUGGAAGAAGCAAA,GCUUCCAAUGG AAGAAGCAA. The results show 4 top-rated guideRNAs for editing ROS1, the target sequences, the respective protein-coding genes for that sequence, the chromosome number in parallel along with the cute site and the PAM region. The identification of the gRNA sequence with possible off target sites located in a specific chromosome and with specific PAM region as obtained while running the Synthego software would help in designing the specific primer for the wet lab experimentation. The statistical analysis of the PAM ratio for the four gRNA sequences is shown in **Figure 5.**

TABLE 1: SYNTHEGO SOFTWARE ANALYSIS FOR ROS1

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

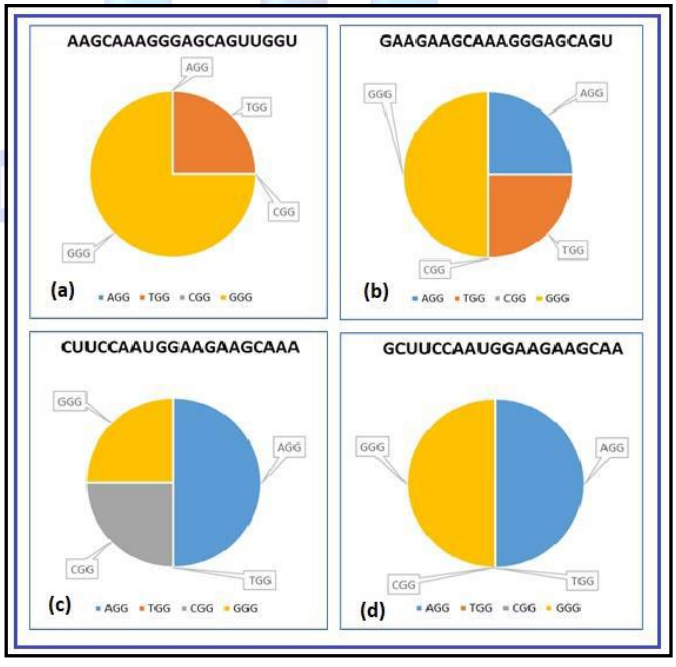

Figure 5: PAM ratio of the ROS1 gene

We have identified the 4 best target sequences and guides based on the highest efficacy score as shown below as showcased in **Figures 6, 7 and 8.**

Sequence: GGTCTGAGGCTGTTCATCAGTGG
Efficacy score by CRISPRater: 0.90 HIGH
Oligo pair fwd: TAGGTCTGAGGCTGTTCATCAG rev: AAACCTGATGAACAGCCTCAGA
Top 20 offtarget sites out of 52 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr6:117328760-117328782	+	0	GGTCTGAG[GCTGTTCATCAG]	TGG	0	E	ROS1	ENSG00000047936
chr7:121158410-121158432	-	3	GGTGAGAA[GCTGTTCATCAG]	CGG	8767	I	CPED1	ENSG00000106034
chr15:89665083-89665105	+	4	CATCAGAG[GATGTTCATCAG]	AGG	0	E	PLIN1	ENSG00000166819
chr3:40642966-40642988	+	4	GGCCTTAC[CCTGTTCATCAG]	TGG	37539	-	RP11-528N21.1	ENSG00000226302
chr1:99227627-99227649	-	4	GGTCGGCT[TCTGTTCATCAG]	AGG	36304	-	PLPPR4	ENSG00000117600
chr8:120130004-120130026	+	4	GGACAGAT[GCAGTTCATCAG]	AGG	4664	I	COL14A1	ENSG00000187955
chr3:45774545-45774567	+	4	GCTAAGAG[GCTGGTCATCAG]	GGG	1194	I	SLC6A20	ENSG00000163817
chr3:194801434-194801456	-	3	GATCTGAG[TCTGTCCATCAG]	GGG	19266	-	AC090505.6	ENSG00000237222
chr3:100076445-100076467	-	4	GATGTGAA[GCTGTCCATCAG]	AGG	813	I	FILIP1L	ENSG00000168386
chr3:70603752-70603774	+	4	GATCTCAG[GCACTTCATCAG]	TGG	NA	-	NA	NA
chr12:124921046-124921068	-	4	AGTCTCAG[GGTGGTCATCAG]	AGG	3678	-	UBC	ENSG00000150991
chr2:66135294-66135316	-	4	TGTCTCAG[GTTGCTCATCAG]	GGG	50655	-	AC074391.1	ENSG00000204929
chr10:122192654-122192676	+	3	GGACAGAG[GCTGTTCTTCAG]	CGG	0	E	TACC2	ENSG00000138162
chr5:92573429-92573451	+	4	GAGCTGAT[GCTGTTGATCAG]	GGG	13396	I	RP11-133F8.2	ENSG00000249776
chr13:94697730-94697752	+	4	TGACTGAG[GGTGTTTATCAG]	TGG	1471	-	RN7SL585P	ENSG00000274168
chr14:97337267-97337289	+	3	GGTCTCAA[GCTGTTCCTCAG]	TGG	NA	-	NA	NA
chr11:47655149-47655171	+	4	CGTCTCAG[GCTTTCCATCAG]	AGG	4420	-	AGBL2	ENSG00000165923
chr9:132543502-132543524	+	3	GGGCAGAG[GCTGTTCACCAG]	GGG	455	I	CFAP77	ENSG00000188523
chr22:50246350-50246372	-	4	AGTCTGTG[GCTGAACATCAG]	GGG	0	E	HDAC10	ENSG00000100429

Figure 6: The identified target sequence for CRISPR/Cas-9 activity against ROS1 gene with efficacy score of 0.90 – output obtained from using CCTop software

Sequence: GGAGGGTGGAAGTTGGGTGGGGG
Efficacy score by CRISPRater: 0.90 HIGH
Oligo pair fwd: TAGGAGGGTGGAAGTTGGGTGG rev: AAACCCACCCAACTTCCACCCT
Top 20 offtarget sites out of 52 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr11:71066131-71066153	+	3	TGAGTGGG[GAAGTTGGGTGG]	GGG	9006	I	SHANK2	ENSG00000162105
chr4:152429356-152429378	+	4	GTTAGGGG[GAAGTTGGGTGG]	TGG	16750	I	FBXW7	ENSG00000109670
chr7:99385288-99385310	+	4	GGGCAGGG[GAAGTTGGGTGG]	GGG	390	I	ARPC1B	ENSG00000130429
chr3:151151573-151151595	+	3	GGAGGCCT[GAAGTTGGGTGG]	GGG	4566	I	MED12L	ENSG00000144893
chr16:20451916-20451938	+	3	GGGGGCTG[GTAGTTGGGTGG]	TGG	147	I	ACSM2A	ENSG00000183747
chr16:12105866-12105888	+	4	TGAGTGGA[GAAGTTGGGTGG]	GGG	10559	I	RP11-276H1.2	ENSG00000261293
chr17:74429179-74429201	+	4	GGGTGGAG[GTAGTTGGGTGG]	TGG	1712	I	GPRC5C	ENSG00000170412
chr12:113387089-113387111	+	4	GGTAGGTG[CCAGTTGGGTGG]	TGG	0	E	PLBD2	ENSG00000151176
chr7:44700673-44700695	+	4	GGTGAGTA[GAAGTTGAGTGG]	AGG	404	I	OGDH	ENSG00000105953
chr6:72643107-72643129	+	2	TGAGGGTG[GAAGTTGGGAGG]	AGG	2233	I	KCNQ5-IT1	ENSG00000233844
chr20:44102267-44102289	+	2	TGAGGGTG[GAAGTTGGGAGG]	AGG	9406	-	JPH2	ENSG00000149596
chr4:44567445-44567467	+	2	TGAGGGTG[GAAGTTGGGAGG]	AGG	33565	-	RP11-500G9.1	ENSG00000251159
chr4:171725717-171725739	+	2	TGAGGGTG[GAAGTTGGGAGG]	AGG	86515	-	GALNTL6	ENSG00000174473
chr4:163943587-163943609	+	2	GTAGGGTG[GAAGTTGGGAGG]	AGG	10405	I	RP11-606P2.1	ENSG00000177803
chr8:43518841-43518863	+	3	GGGGGTTG[GAAGTTGGGAGG]	TGG	4420	-	SNX18P27	ENSG00000253418
chrX:141894921-141894943	+	3	GGTGGGAG[GAAGTTGGGAGG]	TGG	107	I	MAGEC3	ENSG00000165509
chr7:83198781-83198803	+	4	AAAGGTTG[GAAGTTGGGAGG]	AGG	35851	-	PCLO	ENSG00000186472
chrX:56682837-56682859	+	3	GGATGGTT[GAAGTTGGGAGG]	AGG	46400	-	LINC01420	ENSG00000204272
chr17:47695562-47695584	+	4	GAAGTGGG[GAAGTTGGGCGG]	GGG	0	E	TBKBP1	ENSG00000198933

Figure 7: The identified target sequence for CRISPR/Cas-9 activity against ROS1 gene with efficacy score of 0.90 – output obtained from using CCTop software.

Sequence: CCCCTGGTCAGAGCCCTCAGTGG
Efficacy score by CRISPRater: 0.86 HIGH
Oligo pair with 5' extension fwd: TAggCCCCTGGTCAGAGCCCTCAG rev: AAACCTGAGGGCTCTGACCAGGGG
Oligo pair with 5' substitution fwd: TAggCCTGGTCAGAGCCCTCAG rev: AAACCTGAGGGCTCTGACCAGG
Top 20 offtarget sites out of 52 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr19:48805417-48805439	+	3	CCCTTCCT[CAGAGCCCTCAG]	TGG	1078	I	BCAT2	ENSG00000105552
chr7:47103915-47103937	+	4	CCACTCCA[CAGAGCCCTCAG]	TGG	24787	-	AC004870.3	ENSG00000229192
chr5:27073623-27073645	+	4	AACCTGGA[CAGAGCCTTCAG]	GGG	35037	I	CDH9	ENSG00000113100
chr1:162366773-162366795	+	4	ACCCTATT[CAGAGCCTTCAG]	TGG	109	I	NOS1AP	ENSG00000198929
chr22:18554657-18554679	+	4	CCCTAGGA[CAGAGCCTTCAG]	GGG	2084	I	PI4KAP1	ENSG00000274602
chr22:21494183-21494205	+	4	CCCTAGGA[CAGAGCCTTCAG]	GGG	2089	I	PI4KAP2	ENSG00000183506
chr2:63354002-63354024	+	4	TCCCTGAA[CAGAGCCTTCAG]	GGG	5771	I	WDPCP	ENSG00000143951
chr1:56917296-56917318	+	3	TCCCTGGA[CAGAGCCCCCAG]	GGG	247	I	C8A	ENSG00000157131
chr9:32651165-32651187	+	3	TCCCTGGG[CAGAGCCCACAG]	GGG	2480	-	RP11-555J4.4	ENSG00000223440
chr1:145849342-145849364	+	3	CCCTTGGC[CAGAGCCCCCAG]	AGG	0	E	PIAS3	ENSG00000131788
chr12:103899979-103900001	+	4	GCCCCTGT[CAGAGCCCACAG]	AGG	2848	I	TTC41P	ENSG00000214198
chr10:106866136-106866158	+	4	TTCCTGGA[CAGAGCCCCCAG]	GGG	36463	I	SORCS1	ENSG00000108018
chr8:2852812-2852834	+	4	CAACTGGA[CAGAGCCCCCAG]	AGG	30330	-	GS1-57L11.1	ENSG00000253853
chr18:76166797-76166819	+	4	CTCTTGGC[CAGAGCCCACAG]	AGG	6485	-	RP11-94B19.6	ENSG00000266743
chr3:179137499-179137521	+	4	CCCTGGGA[CAGAGCCCCCAG]	AGG	4628	I	RP11-360P21.2	ENSG00000229102
chr3:112093144-112093166	+	4	CCCCACTT[CAGAGCCCCCAG]	AGG	194	I	C3orf52	ENSG00000114529
chr5:88070484-88070506	+	4	TCCCTGAA[CAGAGCCCCCAG]	GGG	NA	-	NA	NA
chr12:47740993-47741015	+	4	CCCCTCAG[CAGAGCCCACAG]	TGG	0	E	RP1-197B17.3	ENSG00000257433
chr16:34396737-34396759	+	4	CCCAAGCT[CAGAGCCCTGAG]	TGG	NA	-	NA	NA

Figure 8: The identified target sequence for CRISPR/Cas-9 activity against ROS1 gene with efficacy score of 0.86 – output obtained from using CCTop software.

VII. DISCUSSION

The manuscript aims at deciphering the latest molecular biology technique in terms of CRISPR/Cas-9 genetic alteration or modification or edition related to ROS1 gene. At the same time, it also envisages the protein characteristics related to the ROS1 protein, a biomarker in lung cancer prognosis. ROS1 is a protein; overexpression relates to the lung carcinoma and has got clinical implication in diagnosing the lung carcinoma other than histochemical or
histopathological or histo-immunological techniques. The ROS1 protein is encoded by the gene termed as "ROS1" gene who's over expression translates it into ROS1 protein. We tried to explore the concept of CRISPR/Cas-9 system to decipher the genetic alteration of the ROS1 gene using computational tool like that of SYNTHEGO and CCTop to generate the target sequence as well as the PAM for each sequence of the ROS1 gene. Our result in terms of ROS1 protein characteristic like that of hydropathy index and polarity envisages that the protein is a transmembrane spanning between the inner and outer domain of the membrane as evident from the Figures 3 and 4. The hydropathy index indicates that the most of the amino acid composition of the ROS1 protein is hydrophobic in nature. Furthermore, we attempted to elucidate the single guide RNA sequence corresponding to the target sequence as referred to here as the DNA site sequence as evident from Table 1. For each SgRNA sequence we attempted to elucidated the possible target sequence and PAM sequence generated from the two computational tools (SYNTHEGO and CCTop) with highest efficacy score as shown in the Figure(s) 6, 7 and 9. The sequence reflected in the manuscript is needed to be processed in the wet lab by transfecting the designed SgRNA sequence in to plasmid and validating the same for precision and accuracy cutting or edition of the said target sequence as recognized by the corresponding PAM sequence.

VII. FUTURE DIRECTIONS:

Our aim in this manuscript was to study the structure and the properties of the ROS1 protein encoded by ROS1 gene as well as to generate the target sequence with appropriate guideRNAs for genetic alteration or cutting or modification employing CRISPR/Cas9 genetic tool. The technique got more relevance and importance after being awarded with Noble prize by the investigators. Future research is needed to achieve the precision and accuracy of identifying the target location among the million-base pair of a gene and cutting the exact target sequence and not generating the off-target sequence is a million-dollar question that needs to be answered or investigated further to achieve its full potential in using as a therapeutic measure against all form of cancer and genetic disorders. It is well known fact that DNA bulges after certain base pairs that may get unrecognized by the Cas-9 system and therefore may generate off target sequence which shall be more detrimental rather than to be useful.[21]

VIII. CONCLUSION

Our paper identifies the best sequences of the ROS1 gene that can be targeted with the highest efficacy and lowest off-target cleavage with the CRISPR system and potentially pave way for higher oncological research for human welfare [22]. We have tried to contribute to the existing knowledge of science using some computational tools to aid the advances with limited resources since we are unable to visit our labs during this pandemic. We aspire to keep working using computational tools and work on the demerits of existing technology to make it desk to bed readily.

ACKNOWELEDGEMENT

The authors of this manuscript would like to shower their heartfelt acknowledgement to the Honorable Chancellor, Professor (Dr.) Goutam Roy Choudhury, of Techno India University, West Bengal for giving the opportunity to work and present the paper in the dynamic field of science that was acclaimed worldwide and awarded with Noble prize for the year 2020 in Chemistry.

N.B: *The manuscripts highlight the cutting edge technology of genetic editing as a promising field of discovery in molecular medicine by "***Jennifer Doudna and Emmanuelle Charpentier***"* for *which they together were awarded with* ***Noble Prize in Chemistry*** *for the year 2020. The manuscript is an interface between molecular biology and computer algorithm. The authors of the manuscript are thankful to the editors and the reviewers of the esteemed journal to provide the opportunity to present this unique research article of CRISPR/Cas-9 system interfaced with computer algorithm in order to promote interdisciplinary research as well as to highlight the latest knowhow in the field of molecular biology. The authors feel that few journals provide opportunities to the budding researcher to showcase their talent in the latest cutting edge technology of any discipline of science, for which the authors are thankful to the general administration of the esteemed journal.*

REFERENCES

[1] Zhang, J., Adikaram, P., Pandey, M., Genis, A. and Simonds, W., 2016. Optimization of genome editing through CRISPR-Cas9 engineering. Bioengineered, 7(3), pp.166-174.

[2] Jiang, F., & Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annual review of biophysics, 46, 505–529.

[3] De, A., & Biswas, A. (2020). CRISPR/Cas-9 Genetic Editing of 'Neu' gene in Breast Cancer Prognosis. International Journal for Innovative Research in Multidisciplinary Field.

[4] Dominguez, A. A., Lim, W. A., & Qi, L. S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nature reviews. Molecular cell biology, 17(1), 5–15.

[5] Shalem, O., Sanjana, N. E., & Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nature reviews. Genetics, 16(5), 299–311.

[6] Clavé, S., Gimeno, J., Muñoz-Mármol, A. M., Vidal, J., Reguart, N., Carcereny, E., Pijuan, L., Menéndez, S., Taus, Á., Mate, J. L., Serrano, S., Albanell, J.,

Espinet, B., Arriola, E., & Salido, M. (2016). ROS1 copy number alterations are frequent in non-small cell lung cancer. Oncotarget, 7(7), 8019–8028.

[7] 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.

[8] 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.

[9] Goenka, M., De, A., & Biswas, A. R., Dr. (2020). Role of CRISPR/Cas9 in Genetic Manipulation of ROS1 and EGFR Genes using Synthego Platform. Volume 5 - 2020, Issue 9 - September International Journal of Innovative Science and Research Technology, 5(9), 1080-1085.

[10] McDade, J. (n.d.). Components of CRISPR/Cas9. Retrieved from https://blog.addgene.org/components-of-crispr/cas9-our-new-crispr-101-ebook

[11] Terns, Michael P., and Rebecca M. Terns. "CRISPR-Based Adaptive Immune Systems." Current Opinion in Microbiology, vol. 14, no. 3, June 2011, pp. 321–27.

[12] Hille, Frank, et al. "The Biology of CRISPR-Cas: Backward and Forward." Cell, vol. 172, no. 6, Mar. 2018, pp. 1239–59.

[13] Damodharan, L., & Pattabhi, V. (2004). Hydropathy analysis to correlate structure and function of proteins. Biochemical and biophysical research communications, 323(3), 996–1002.

[14] Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of molecular biology, 157(1), 105–132.

[15] Gallo, E., & Gellman, S.H. (1993). Hydrogen-bond-mediated folding in depsipeptide models of .beta.-turns and .alpha.-helical turns. Journal of the American Chemical Society, 115, 9774-9788.

[16] Zimmerman, J. M., Eliezer, N., & Simha, R. (1968). The characterization of amino acid sequences in proteins by statistical methods. Journal of theoretical biology, 21(2), 170–201.

[17] Design.synthego.com. 2020. Synthego. [online] Available at: https://design.synthego.com/#/

[18] Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PloS one, 10(4), e0124633.

[19] Abadi, S., Yan, W. X., Amar, D., & Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. PLoS computational biology, 13(10), e1005807.

[20] Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J. H., & Heckl, D. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. Nucleic acids research, 46(3), 1375–1385.

[21] De, A., & 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. 872-876.

[22] Annunziato, S., Lutz, C., Henneman, L., Bhin, J., Wong, K., Siteur, B., van Gerwen, B., de Korte-Grimmerink, R., Zafra, M. P., Schatoff, E. M., Drenth, A. P., van der Burg, E., Eijkman, T., Mukherjee, S., Boroviak, K., Wessels, L. F., van de Ven, M., Huijbers, I. J., Adams, D. J., Dow, L. E., … Jonkers, J. (2020). In situ CRISPR-Cas9 base editing for the development of genetically engineered mouse models of breast cancer. The EMBO journal, 39(5), e102169.