TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

A Machine Learning Approach to Genome Editing Techniques

Dishani Haldar[1], Divyasree G[2], Divya M[3], Mouli Dudekula[4], Naveen Chandra Gowda[5]
[1,2,3,4] Students, B.Tech.in Computer Science & Engineering, REVA University, Bengaluru, Karnataka, India
[5]Assistant Professor, REVA University, Bengaluru, Karnataka, India

Article Info
Volume 83
Page Number: 4647-4652
Publication Issue:
May-June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

In the current era, the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated system (Cas) is developed and used as the most adaptable instrument for genetic administration application. It is encompass of repetitive bases followed by short fragments of DNA. The manipulation of targeted genes and genomic regions that are balancing to a programmable single guide RNA (sgRNA),but the effectiveness of the sgRNA is not properly defined for the target site so unintended off-targets might be cleaved. Modernistic methods for sgRNA designs are based on predicting the off-targets for a sgRNA using basic sequence features. We present a summary and relative analysis of algorithms based on machine learning approaches which will be more impressive and predictable method for predicting susceptibility of a genomic site to be cleaved by a given sgRNA. We will show that the predictions are more accurate and then validate the occurrence of bulges.

Keywords: *CRISPR, CRISPR/Cas, sgRNA, DNA*

1. Introduction

Considering with the Immune System, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) is considered to be more flexible. This method explains the detail about the possible occurrence of bulges and consolidate a wide scope of highlights wrapping the particular genomic content and the features of the sgRNA and pairwise similarity between the sgRNA and the genomic target. [16].

Gene editing is to roll out a particular improvement in the gene sequence. The Gene editing is led utilizing enzymes and nuclease that are being built to focus on a particular DNA succession where they will play out the DNA strands cutting, empowering the expulsion of existing DNA and the addition of substitution DNA.CRISPR-Cas9 is functional with allowing the removal and insertion of DNA in the desired locations [19].

Genome editing technique uses Meganuclease, ZFNs, and TALEN, CRISPR provides a new method for genetic administration are to assist in the implementing of desired traits by customizing endogenous genes, because of the tranquilize use and cost-efficiency of CRISPR. [17].

Machine Learning is a study of logical algorithms based on mathematical model that systems used to perform unique tasks without having to use certain instructions, depending on the patterns and respective inference values. Machine learning in this field of genomics are impacting how the genetic research is being conducted, gene sequencing, gene editing etc [9].

The change of the CRISPR-Cas9 framework as a genome altering procedure has the capacity to update the targeted genes and genomic regions that are reciprocal to a programmable single guide RNA (sgRNA). We will present an objective appraisal, an algorithm inside machine learning structure that predicts a genomic site to be separated by a given sgRNA. This evaluation will do the forecast progressively precise and the learning procedure gives translation with respect to designs that control the activity of the CRISPR-Cas9 framework [18].

2. Literature Survey

2.1 Related Work

Many authors proposed the work on the semi-considered problem in different ways. Few of those are summarised in the Table 1. It also shows the drawbacks still found each solution they proposed.

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

2.2 Gaps Identified

Based on the drawbacks identified in the Table 1, we summerise the gaps as,

i) In CRISPR/Cas9, Cas9 can just genetic sequence arrangements of around 20 bases in length, implying that more extended groupings can't be focused on.

ii) In CRISPR/Cas9 system as a genome editing technique has ability to manipulate targeted genes, there is huge risk of gene mutation in this editing technique.

Table 1: Survey of different mechanisms under CRISPR

Author & year	Problem Considered	Solution Proposed	Dataset considered	Advantages	Disadvantages
Johns Hopkins, et.al, March 2018 [1]	genetically engineer malaria-resistant mosquitoes	Researcher have invented malaria resistant mosquitoes by deleting a gene called FREP1, which causes malaria in the mosquito gut.	anopheles mosquito	genetically engineer malaria-resistant mosquitoes	The malaria resistant mosquito parasites are unable to survive for a long period of time to mature to a stage at which it spreads as a danger to human.
Luke Dormehl, et.al, August 2017 [2]	Genetically engineered animals as a organ donor for human beings	Researcher have discovered that pig is suitable for organ transmission as it's organ are similar to human beings in both size and anatomy.	Pig	Using this CRISPR genome editing technique and porcine somatic cell nuclear transfer they have successfully generated viable pigs that are PERV - inactivated.	In the organ transmission there are certain precautions has to be taken else dynamic immune response as such complement activation, coagulation, thrombosis can happen.
Luke Dormehl, et.al, July 2017 [3]	Genetically encode a GIF in the virus genome using CRISPR.	They have invented E.coli bacteria can store images and movies in it's DNA, so using that sequence in virus to encode a GIF.	E.coli	Using this genome editing technique the in bacteria could be used to capture complex information with a time in living bacteria.	Using this CRISPR technology to create live cells that can record biological or environmental signals.
Stephen Long, et.al, March 2018[5]	Genetically improved plants that makes more productivity at using less water.	The genetic engineered CRISPR process is involved in regulating the amount of photosynthetic gene produced in a plant by integrating additional copies of the developed gene into it's DNA.	tobacco crops	The engineered plants gives more productivity, and ever increasing demands of water using irrigation. The demands will be more with increasing productivity.	The large improvements has been done in the crops but the breeding has not improved.
Dyllan Furness, et.al, April 2018[6]	world's first' CRISPR-powered disease detection	Genetically evolved CRISPR to remarkable disease detection, detects biomarkers that is associated with	tweaks to the genomes of myriad plants and animals	The detection of organisms that contains DNA or RNA, infectious disease even cancers can be targeted.	The detection of molecules disposable paper test strip, which will be able to detect multiple test

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

		diseases.			on one strip.
Michael Le Page, et.al, November 2018[4]	Genetically invented gene-edited babies	The experiment was to create HIV resistant children by deleting both the copies of CCR5 gene.	Human embryo gene.	The body cells should resist HIV infection.	This genome editing technique causes mutation in the DNA and that causes manipulation of other genes.
Luca Pinello, et.al, July 2016 [3]	The CRISPR should quantify and visualized the NHEJ and HDR mutagenesis profile.	The genetically discovered CRISPResso classifies any mutation overlapping the window around the expected cleavage site.	Bacterial DNA sequence.	The sequencing of amplified genomic region allows quantitative and sensitive detection of target mutation.	presence of sequencing errors.
F. Ann Ran, et.al, July 2015[7]	Multiplex Genome Engineering	The genetically evolved heterologous expression can achieve target cleavage of chromosomes.	Streptococcus pyogenesSF370 type II	The minimal three-component system for a efficient RNA-guided genome alteration in cells.	In the prediction all RNA designs couldn't felicitate cleavage of their genomic target.
Martin Jinek, et.al, August 2012[8]	A genetically engineered dual RNA-guided DNA	A two-RNA structure directs an endonuclease to cleave target DNA.	Streptococcus pyogenes	The main featurerequired for cite specific endonuclease that catalysed the DNA cleavage could be captured in a single chimerical RNA.	the plasmid DNA was restricted and mapped with AvrII following Cas9 cleavage.
Wiedenheft B, et.al, February 2012[9]	RNA-guided gene silencing systems in bacteria and archaea.	CRISPR-mediated immune system is based on small RNAs for sequence-explicit detection and silencing of foreign nucleic acids, including viruses and plasmids.	Eukaryotic RNA sequence	A small RNAs are used to find and destroy foreign nucleic acids	Proposed system is in eukaryotic RNA sequence.
Max Delbrück, et.al, October 2016[10]	Efficient CRISPR observation in mouse cells	A genetically discovered mouse model that carries Cas9 protein.	Mouse DNA sequence	In the genetically modified primary cells the efficient inactivation of genes.	the B cells of mouse cannot be cultivated for any length of time, because they do not survive long outside their natural environment
Degao Liu, et.al, September 2019[11]	A cell mediated targeted mutagenesis for genomics research in c3 and C4 plants.	The evolved CRISPR tool is efficient in creating biallelic index mutation to reveal the roles of blue light receptor in	C_3 and C_4 plants DNA sequence.	CRISPR/Cas9 technology implement efficient targeted index mutation for rapid generation of desirable mutants in	The main appearance of blue light receptor in the CAM has not been fully detected yet.

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

		CAM plants.		the model .	
Qing-hui Yu, et.al, September 2017[12]	The genetically induced target mutation furthermore, quality substitution to produce long timeframe of realistic usability in tomato.	The genome editing technique method used for ALC gene mutation and replacement in tomato.	Agrobacterium tumefaciens RNA sequence.	The crop breed efficiency and adaptability of multiplying provides a reasonable expectation towards breeding goals.	The mutation rate is very less in this experiment.
Khaoula Belhaj, et.al, April 2015[13]	Editing plant genomes with CRISPR/Cas9	CRISPR/Cas9 is an RNA guided DNA endonuclease inherent to prokaryotic systems.	Plant genome.	In this genetically evolved plants homozygous knockout change can be delivered in a solitary age.	An independently created guide RNA with which it frames a composite, that increases chance of mutation.
Cara L Mortimer, et.al, April 2015[14]	Induced transgene utilizing viral vectors from transient to stable articulation	The genetically invented viral vector amplify the transgene to develop into robust transient platform.	Virus RNA sequence .	The induced hyper-expression of the vectors provide both activation and amplification of heterologous transgene expression.	The initiation of mRNA intensification and significant level protein gathering can causes several changes in plants.

3. Proposed System

The CRISPR-Cas9 framework introduced many machine learning algorithms for genome editing techniques which will be more accurate cleavage efficiency, the dimensional structure and acerbity of the entire genomic site and also the PAM region.

The machine learning approach to this genome editing techniques requires the assembly of a training dataset, and the fusion of a collection of features which can be used to predict and analyse the cleavage efficiencies. The dataset will enhance the learning process, as a collection of uncleaved sites were assembled that range from specific to the target site to those that are specific to the sgRNA.

The specific workflow for the gene editing system is proposed in a flowchart follows

- According to the workflow given below, for the implementation of the proposed system the target site has to be selected.
- The replacement site and codon optimised variant should be selected.
- sgRNA and cas9 protein will form complex vector.
- Further the complex vector will be delivered to the specific site of the target sequence.
- Sequencing will be there to detect target mutation and off-target effect.

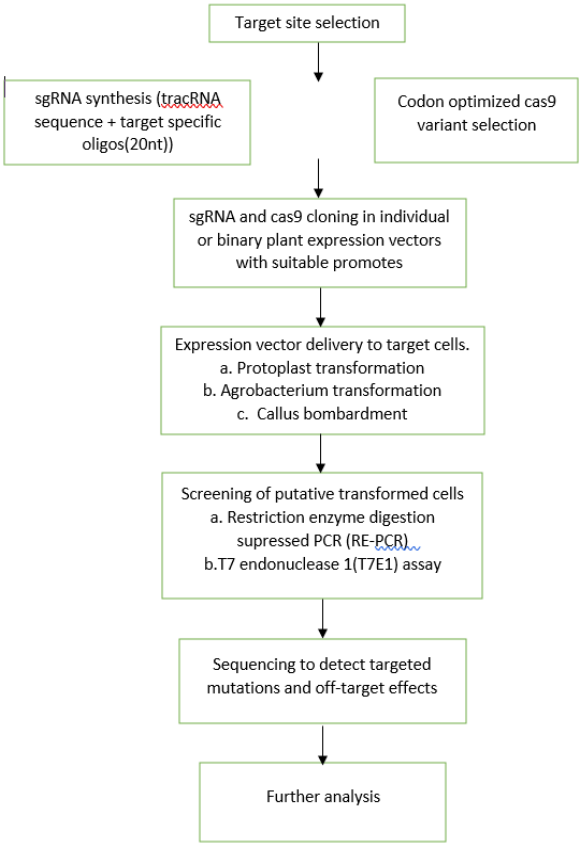


Figure 1: System Flowchart

TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

4. Implementation

We made use of the Hardware modules are Wifi-Module, Desktop Monitor, connecting cable, computer minimum 8 gb RAM and i7 processor and cloud server. The software modules are python 3.7 for writing scripts, Anaconda Environment, Azimuth, design module, gRNA module.

The gene dataset considered for this project is genome dataset.

The implementation procedure in this genome editing techniques follows machine learning approach. It follows the following steps :

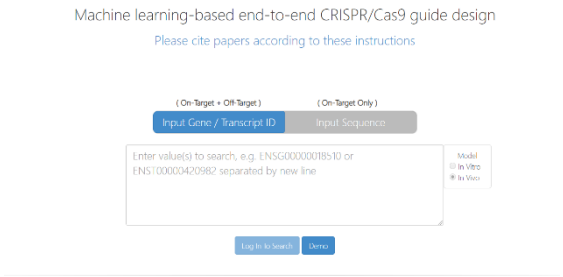
- **Data assembly:** The preparation dataset for this information model was unite from information acquired utilizing a few genome-wide fair strategies.
- **Pairwise arrangement to account for bulges:** In initial stage, the perception is that the pairings of the guide RNA and the comparing genomic locales contains a surpassing huge number of crisscrosses.
- **A machine learning algorithm for predicting cleavage propensity:** The Data model is explicitly learning dependent on a relapse model utilizing the Random Forest calculation, and further permits the variety of cleavage effectiveness.
- **Assessing algorithm performance:** The presentation of Data Model utilizing two cross-approval method forecast . A sgRNA out methodology the examples of a solitary guide RNA were barred and utilized as a test ser in every cycle. The calculation, will get prepared on the staying of the information and used to anticipate the cleavage probabilities.
- **Identifying a succinct set of influential feature:** The calculation will process the comparing commitment of the tried highlights to the relapse model, named as highlight significance.

5. Results

- CRISPR/cas9 offers accurate results that correlate strongly with NGS-based analysis.
- CRISPR/cas9 supports a wide variety of edits including indels, multi-guide fragment deletions, single-nucleotide substitutions, and small and medium HDR insertions.

6. Outcomes

- The “Google Search” option in the frontend of the website helps in providing the previous works that has to done by the biologists from different places.
- The “Watch Video” shows the animated mechanism of how the artificial RNA works.
- “Home” icon guides the user to go back to the home page of the website.
- “Services” shows the progress of the work that’s been done by the various scientists.
- “Contact” provides the details to communicate with the website owner.
- FAQ displays the Frequently Asked Questions by different users that have login.
- “Explore More” icon leads to the working mechanism of the CRISPR/Cas9 genome editing.



Advantages and Application:

- Gene Silencing.
- Homology-directed repair(HDR).
- DNA-free CRISPR-Cas9 gene editing.
- Transient gene silencing or transcriptional repression.
- Embryonic stem cell and transgenic animals.
- Transient activation of endogenous genes.
- Pooled genome-scale knockout screening.

7. Conclusion

We have prepared one machine learning framework that will be more accurate cleavage efficiency, the dimensional structure and rigidity of the entire genomic site as well as the PAM region. various available single guide RNA (sgRNA) is used to make the versatile “CRISPR-Cas9 model”, it will help the user to get more productive and up-to-date information about the disorders and what are the possibilities to resolve and be aware about their time spending and various other similar things.

References

[1] CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection, PLOS PATHOGENS,March 8, 2018

[2] Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors, the NEW ENGLAND JOURNAL of MEDICINE,Babak J. Orandi,March 10, 2016

[3] The CRISPR-Cas systems were selectively inactivated during evolution of Bacillus cereus

May-June 2020
ISSN: 0193-4120 Page No. 4647-4652

group for adaptation to diverse environments The ISME Journal04 March 2018

[4] World's first gene-edited babies announced by a scientist in China 26 November 2018

[5] Improving photosynthesis and crop productivity by accelerating recovery from photoprotection Science 18 Nov 2016:

[6] New startup promises 'world's first' CRISPR-powered disease detection By Dyllan Furness april 26,2018 Digital trends

[7] Multiplex Genome Engineering Using CRISPR/Cas Systems available in PMC 2016 Oct 11

[8] A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity Science 17 Aug 2012:

[9] RNA-guided genetic silencing systems in bacteria and archaea Nature. 2012 Feb 15;482(7385):331-8

[10] Genome editing: Efficient CRISPR experiments in mouse cells Max Delbrück Center for Molecular Medicine October 25, 2016

[11] CRISPR/Cas9-mediated targeted mutagenesis for functional genomics research of crassulacean acid metabolism plants Journal of Experimental Botany, Volume 70, Issue 22, 15 November 2019

[12] CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines Scientific Reports volume 7, Article number: 11874 (2017)

[13] Editing plant genomes with CRISPR/Cas9 ,The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UKApril 2015

[14] Updates in inducible transgene expression using viral vectors: from transient to stable expression Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Queensland 4000, Australia April 2015

[15] Application of CRISPR/Cas9 genome editing to the study and treatment of disease.Arch Toxicol. 2015 Jul;89(7):1023-34.

[16] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339: 819–23. pmid:23287718

[17] Jun Wang, Xiuqing Zhang, Lixin Cheng & Yonglun Luo (2019): An overview and metanalysis of machine and deep learning-based CRISPR gRNA design tools, RNA Biology,

[18] Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems.