Algorithms used in CRISPR, A Gene-Editing Technology

Team Members:

Marium Salman

Masooma Raza

Laiba Khalid

Marwah Tariq

Supervisor:

Miss Hafiza Anisa Ahmed

Algorithms used in CRISPR, A Gene-Editing Technology

ABSTRACT:

A gene editing technology, while focusing on CRISPR/Cas9 technology that is enhanced by the essential role of algorithms in improving accuracy and proficiency by determining the possibilities and challenges related with CRISPR gene editing. The outbreak of genomic sequencing technologies combined with current developments in gene editing techniques has raised the potentials of genomic uses in many organisms in which these studies were previously not eagerly available or conceivable. These developments play a big role for agricultural and medical applications, it examines the possible drawbacks as well as off-target effects and algorithms limitations. It contributes to the existing discussion on the progress and application of CRISPR/Cas9 technology growing rapidly in the field of genetic engineering.

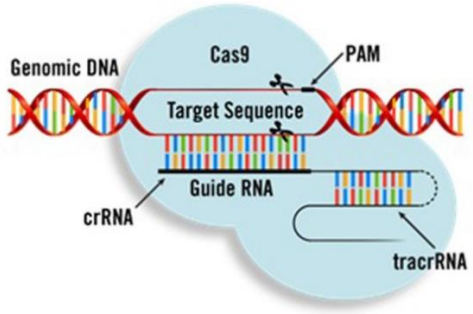

Fig.1

Main components of CRISPR

INTRODUCTION:

Significant advancements have been made in biotechnology in recent years, and the branch of genetic engineering is advancing at an unprecedented pace, yielding numerous advantages. Genome editing technology has revolutionized genetic and biological re-search via the novel ability to precisely manipulate and modify the genomes of living organisms. Genome editing technology emerged in the 1990s, and various methods have since been developed for targeted gene editing. In general, three systems each with their own advantages have been widely used in cells and animals, including transcription activator-like effector nuclease (TALEN), zinc finger nuclease (ZFN), and clustered regularly interspaced short palindromic repeats (CRISPR) (1). CRISPR – clustered regularly interspaced short palindromic repeats – were first discovered in the sequences of DNA from Escherichia coli bacteria and described in 1987 by Ishino et al. from Osaka University (Japan) (2). The name CRISPR refers to the unique organization of short, partially repeated DNA sequences found in the genomes of prokaryotes. CRISPR and its associated protein (Cas-9) is a method of adaptive immunity in prokaryotes to defend themselves against viruses or bacteriophages (3). CRISPR-Cas9 is a unique technology that enables geneticists and medical researchers to edit parts of the genome by removing, adding or altering sections of

the DNA sequence. It is currently the simplest, most versatile and precise method of genetic manipulation (4). There were ways to edit the genomes of some plants and animals before the CRISPR method was unveiled in 2012 but it took years and cost hundreds of thousands of dollars. CRISPR has made it cheap and easy. CRISPR technology also has the potential to transform medicine, enabling us to not only treat but also prevent many diseases (5).

Algorithms play a crucial role in advancing gene editing studies, particularly in the context of CRISPR/Cas9 technology. CRISPR/Cas9 is more suitable for application compared to other gene editing technologies and has several important advantages (6). They were developed using diverse computational approaches based on different datasets and quality designs, their performance could differ significantly when used in different experimental situations (7). Algorithms offer several advantages, including the identification of potential editing sites with high specificity and accuracy, the design and optimization of guide RNAs (gRNAs) for maximum efficiency, prediction of experimental outcomes, data analysis to gain insights into biological processes, and the development of novel gene editing tools and technologies. In particular, the ease with which CRISPR-Cas9 can be configured to recognize new genomic sequences has driven a revolution in genome editing that has accelerated scientific breakthroughs and discoveries in disciplines as diverse as synthetic biology, human gene therapy, disease modeling, drug discovery, neuroscience, and the agricultural sciences (8). For instance, algorithms have been used to identify off-target effects in human cells, design optimized gRNAs for cancer therapy, predict outcomes of plant gene editing experiments, analyze data to unravel gene function, and create more efficient CRISPR-Cas9 variants. CRISPR/Cas9 can be used to operate gene appearance, correct mutations, or disrupt cancer-related genes, making it a versatile tool for various applications in cancer research and therapy (9). These applications collectively enhance the precision, efficiency, and safety of gene editing while fostering the development of innovative therapeutic approaches and a deeper understanding of biological processes. The advancements in genome editing techniques have opened up new doors for what genome editing can do to address issues in medicine, agriculture, and beyond. CRISPR has completely revolutionized what genome editing can mean for our future by increasing the speed and breadth of science (10).

Algorithms are invaluable in gene editing, yet they face significant challenges. Numerous studies have reported unexpected genomic mutation and mosaicism following the use of CRISPR/Cas nucleases, and it is currently unclear how prevalent these disadvantageous events are and how robust and sensitive the strategies to detect these unwanted events may be (11). One prominent issue is the occurrence of off-target effects, where unintended edits to DNA happen outside the target region. CRISPR/Cas9 could be used for targeted manipulation of the mammalian genome, reports emerged that its use carries a risk of unintended mutagenesis at closely matched genomic sequences (12). While algorithms aim to identify and reduce these effects, they are not infallible; for instance, one study found that an algorithm could only detect 70% of actual off-target edits. A major concern for implementing CRISPR/Cas9 for gene therapy is the relatively high frequency of off-target effects (OTEs), which have been observed at a frequency of ≥50% (13). Moreover, the accuracy and specificity of gene editing algorithms heavily rely on the quality of their training data, and if this data is flawed or incomplete, the algorithm's predictions may not align with real-world results. Gene Editing offers promising possibilities for disease prevention and infertility

treatment. However, the effects of modified genes within the gene pool of a given population are unforeseeable and uncontrollable (14). Over prediction is another concern, as algorithms, trained under ideal conditions, can sometimes exaggerate the accuracy and specificity of editing outcomes. Insufficient validation of algorithm-designed guide RNAs can lead to problems like off-target effects and reduced editing efficiency, underlining the importance of rigorous validation procedures. However, lessons learned from traditional gene therapy should prompt greater caution in moving forward with CRISPR systems to avoid adverse events and setbacks to the development of what may be a unique clinically beneficial technology (15). Transparency is vital, yet some studies lack clarity about the algorithms used, hindering the assessment of their accuracy and reliability. Additionally, issues like small sample sizes, lack of diversity in training data, complex algorithms, and ethical concerns add further complexity to the development and application of gene editing algorithms, highlighting the need for thorough evaluation and ethical consideration in this cutting-edge field.

In our broadly study of CRISPR/Cas9 technology and gene editing algorithms, we highlight the crucial impact of CRISPR on genetic engineering that allows detailed modification of genomes. The usefulness of CRISPR has enhanced in numerous scientific disciplines, from synthetic biology to human gene treatment and agriculture. In this situation, the essential role of algorithms becomes obvious as they identify specific DNA targets, improving guide RNAs, predicting experimental results, and analyzing data for understanding of biological developments. Applications variety from cancer treatment to agricultural sciences and drug discovery that showcase the flexibility of this technology. It also discusses about the challenges related with CRISPR/Cas9 technology. It highlights the requirements for severe validation procedures and transparency in algorithm design to report challenges like off-target effects, reduced editing efficiency and overprediction. Our research paper navigates the complexity of gene editing algorithms by offering understandings into their strengths, restrictions and the vital role they play in determining the future of genetic engineering.

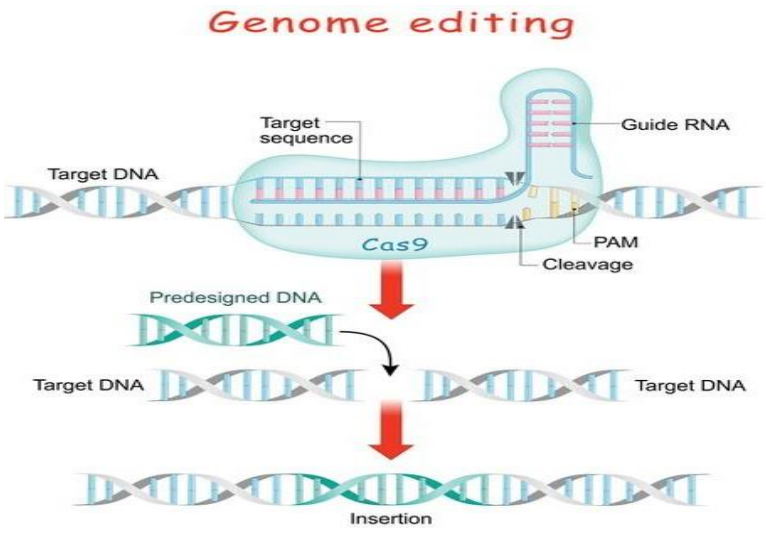

Fig.2

How Crispr Gene-Editing Works

LITERATURE REVIEW:

CRISPR/Cas9 edits genes by precisely cutting DNA and then allowing natural DNA repair processes to take over. It consists of Cas9 enzyme and a guide RNA. CRISPR genome editing allows scientists to quickly create cell and animal models, which researchers can use to research into diseases like cancer and mental illness. In addition, CRISPR is now being developed as a rapid diagnostic.

Sunil Bodapati et al., (2020) discuss about using Clustered regularly interspaced palindromic repeats (CRISPR) technology for genetics studies. The main focus is the analysis of different algorithms such as Redundant siRNA activity (RSA), MAGeCK Robust Ranking Algorithm (RRA), HiTSelect, MAGeCK Maximum Likelihood Estimation (MLE), BAGEL, CRISPhieRmix, CERES, JACKS, and a simple t-test for CRISPR-based genetic screens. Overall, the article suggests that MAGeCK RRA is a best choice for most CRISPR screens(16).

Emmanuelle Charpentier and Jennifer Doudna developed the CRISPR-associated nuclease9 (CRISPR/Cas9), a gene editing technology that provided new tools for precise gene editing. They said that it is possible to target any genomic locus virtually using only a complex nuclease protein with short RNA as a site-specific endonuclease. From which they discovered that Cancer is caused by genomic changes in tumor cells. CRISPR/Cas9 can be used in the field of cancer research to edit genomes for exploration of the mechanisms of tumorigenesis and development. However, several challenges remain before this technology can be used in the clinical treatment of cancer safely and efficiently. Moreover, with the CRISPR/Cas9 technology in cancer research, diagnosis, and treatment, continuous efforts are needed to overcome the targeted challenges in the future (17).

Xingliang Ma et al. designed a robust CRISPR/Cas9 vector system for multiplex genome editing in monocot and dicot plants. SgRNA and off-target prediction algorithms can be used to do editing or make changes in the genome. In this research they designed PCRbased procedures to rapidly generate multiple sgRNA expression cassettes, which can be assembled into the binary CRISPR/Cas9 vectors in one round of cloning by Golden Gate ligation or Gibson Assembly. The system utilizes a plant codon optimized Cas9 gene, for convenient and high-efficiency multiplex genome editing in monocot and dicot plants that can help us to develop crops that are more resistant to diseases, pests and environmental stresses (18).

James Zou developed a new algorithm SPROUT, it helps to predict errors in CRISPR-Cas9 gene editing, addressing accuracy and safety concerns. Using machine learning, it forecasts potential mistakes during CRISPR edits, aiding precise gene modifications. By analyzing thousands of CRISPR edits in human immune cells, SPROUT identifies error-prone DNA sequences and estimates the risk and scale of unintended DNA. SPROUT is integrated into research labs worldwide, aiding scientists in identifying error-prone DNA sequences for safer and more precise CRISPR-Cas9 gene editing in various disease studies (19).

Xi Xiang et al. developed a method to improve CRISPR-Cas9 gene editing by selecting the best guide RNA (gRNA) molecules. Selecting the right gRNA is challenging but crucial for accurate gene editing. Their approach combines data, deep learning, and molecular interactions to predict efficient gRNAs. They used a dataset of over 10,000 gRNAs to create a precise model, potentially

enhancing the precision of CRISPR-based gene editing. Researchers worldwide are using this method to improve CRISPR-Cas9 gene editing, enhancing the accuracy of genetic modifications for potential therapeutic advancements in diseases like cancer (20).

Kevin Bloh et al. introduces DECODR, a software tool for analyzing CRISPR-edited DNA sequences. It addresses the challenges of analyzing diverse CRISPR-induced sequence changes. DECODR accurately detects indels of any size and identifies inserted and deleted bases. The study outlines its methods, including DNA extraction and sequencing. DECODR excels in analyzing clonal and bulk cell populations, including complex compound indels. It outperforms existing tools like TIDE and ICE, making it a valuable resource for precise CRISPR gene editing analysis, advancing genetic research. The DECODR software provides valuable assistance in today's world by enhancing our ability to precisely analyze and understand genetic modifications resulting from CRISPR gene-editing experiments. This aids researchers in advancing gene therapy, biotechnology, and genetic engineering applications, ultimately contributing to the development of innovative treatments for genetic diseases and improved crop production, among other important scientific and medical advancements (21).

Comprehensive Overview of CRISPR Analysis Algorithms and Tools

CRISPR-Cas9 sgRNA Design and Off-Target Considerations: Insights and Algorithmic Approach:

To use the sgRNA-Cas9 system for genome editing, the first step is to identify a specific DNA sequence called a PAM within the target region. The sgRNA sequence, which is made up of 20 bases, is located right before the PAM on the same DNA strand. It's important to avoid off-target activity because Cas9 can unintentionally cut DNA at sites with similar sequences. Different studies have shown that the sgRNA-Cas9 system can tolerate a certain number of mismatches with the target sequence, but the level of off-target activity depends on the number and positions of these mismatches. Designing sgRNAs within the first half of the CDS sequence and avoiding consecutive runs of four T or four A bases are common practices. Searching for potential off-target sites can be time-consuming, so researchers often use indexed databases that focus on variations in the seed sequence to make the process more efficient.

The algorithm computes the number of base mismatches, including specific bulge penalties. In the default setting, a bulge penalty equals three base mismatches (two if inside the seed sequence), RNA bulge extension penalty equals one base mismatch, and DNA bulge extension penalty equals two base mismatches. By shifting the 5′ fragment one base to the left, an alignment with one base mismatch and one DNA bulge can be achieved. It also considers special cases, such as the number of mismatches and indels within the seed sequence. If a specific alignment has a DNA bulge and base mismatch inside the seed sequence, it is not considered a homology to the reference. It computes five cases: H, L1, R1, L2, R2. If d is found to be a homology to r in one case, further computation is unnecessary (22).

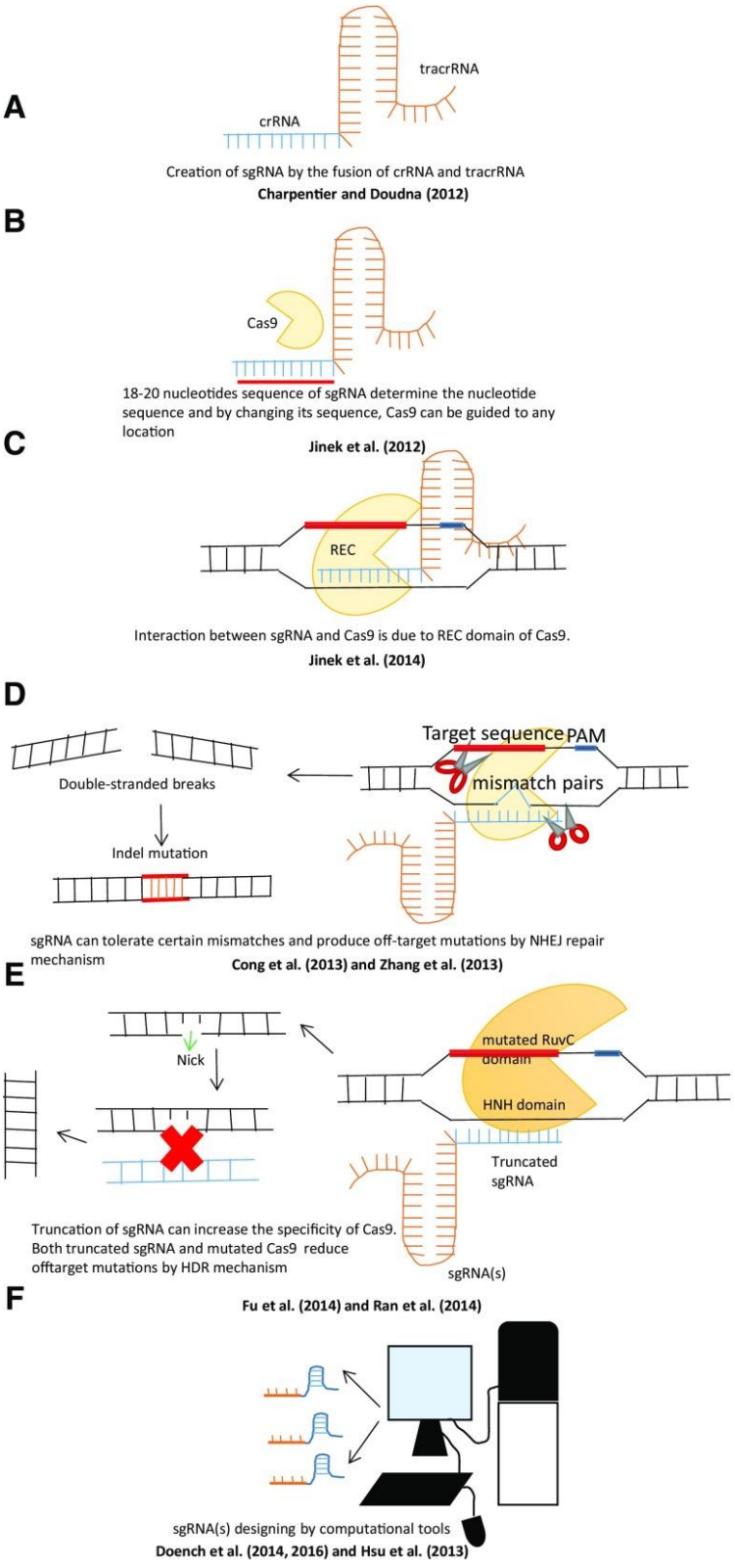

Fig.3

sgRNA Designing By Computational Tools

Efficiency Prediction Algorithms:

Sprout:

As a simple and programmable nuclease-based genome editing tool, the CRISPR/Cas9 system has been widely used in target-gene repair and gene-expression regulation. The DNA mutation generated by CRISPR/Cas9-mediated double-strand breaks determines its biological and phenotypic effects. Experiments have demonstrated that CRISPR/Cas9-generated cellular-repair outcomes depend on local sequence features (23).

Although such an RNA-guided system permits precise genome editing, it presents a great challenge for off-target effects. Studies have demonstrated that Cas9 can tolerate a number of base pair (bp) mismatches, and incorrectly cleave sites do not fully complement the sgRNA (off-target sites). Off-target effects can lead to potential side effects, which will hinder the development and clinical applications of CRISPR system (24). However, existing prediction methods rely on manually constructed features or insufficiently detailed prediction labels. They cannot satisfy clinical-level-prediction accuracy, which limit the performance of these models to existing knowledge about CRISPR/Cas9 editing (25).

To address this challenge, researchers have developed machine learning algorithms to predict CRISPR editing outcomes. SPROUT (Specific Predictions of Repair Outcomes for CRISPR Treatments) is one such algorithm that utilizes a gradient boosting approach to analyze the sequence of the editing site and other genomic factors to predict the likelihood of indel (insertion or deletion) mutations and the overall efficiency of CRISPR editing. By predicting potential editing outcomes, SPROUT can help researchers select the most effective gRNA sequences and optimize CRISPR editing strategies.

Decodr:

During CRISPR-directed gene editing, multiple gene repair mechanisms interact to produce a wide and largely unpredictable variety of sequence changes across an edited population of cells. Shortcomings inherent to previously available proposal-based insertion and deletion (indel) analysis software necessitated the development of a more comprehensive tool that could detect a larger range and variety of indels while maintaining the ease of use of tools currently available. To that end, we developed Deconvolution of Complex DNA Repair (DECODR) (26).

It utilizes a novel deconvolution algorithm to accurately identify and quantify insertion or deletion (indel) mutations resulting from the CRISPR editing process. DECODR can handle indels of any size and can analyze data from both single and multi-guide CRISPR experiments. DECODR does not rely on predefined proposal sequences, allowing it to detect a wider range of indels and provide more comprehensive results. Additionally, DECODR can analyze data from both clonally expanded and bulk cell populations, making it a versatile tool for various CRISPR editing studies.

DECODR has been shown to be an accurate and efficient tool for analyzing CRISPR editing outcomes, providing valuable insights into the efficiency and specificity of different gRNA sequences and CRISPR editing strategies. It has been successfully used in a variety of studies to assess the impact of CRISPR editing on gene expression, protein function, and cellular phenotypes.

However, while many tools are available, assessment of their application scenarios and performance benchmarks are limited. Moreover, new deep learning tools have been explored lately for gRNA efficiency prediction (27).

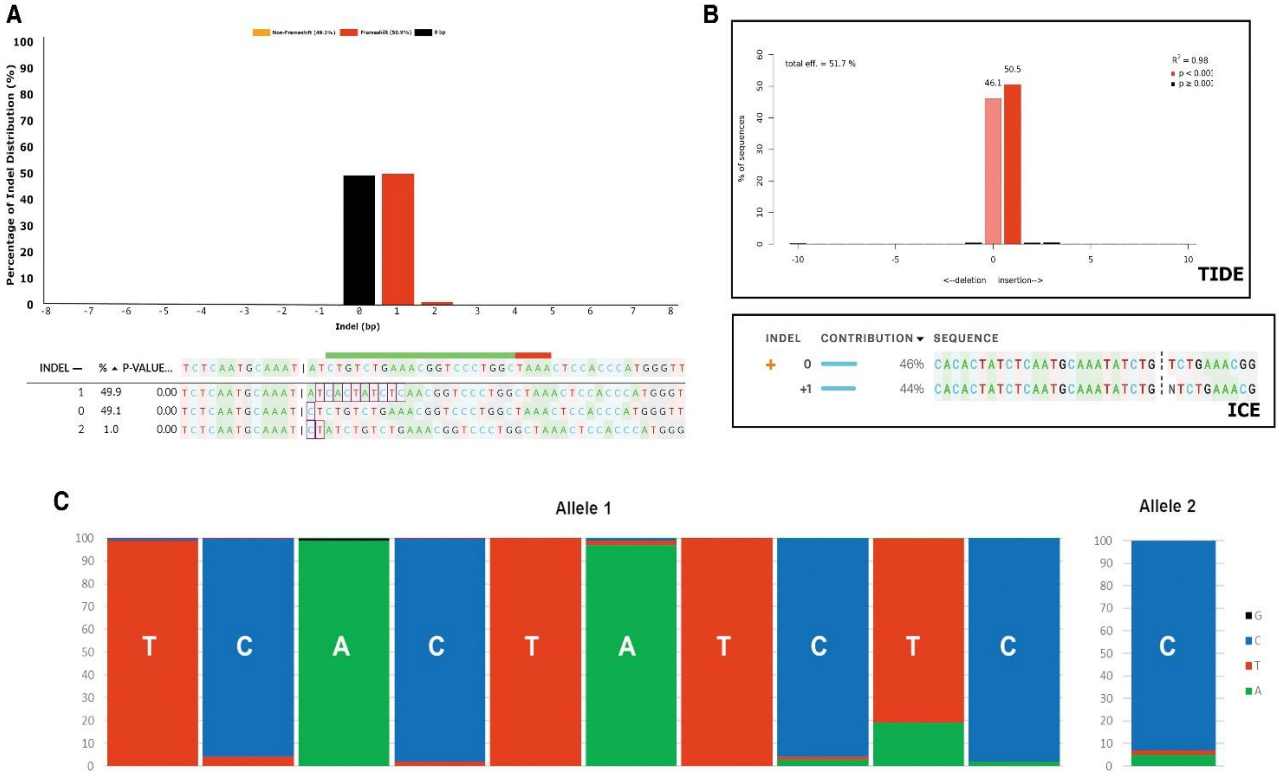

Fig.4

DECODR Efficiency in Analyzing Poly-Allelic DNA Sequencing Samples

Analysis of CRISPR Screens Algorithms:

Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK): MAGeCK is to identify essential sgRNAs, genes and pathways from CRISPR/Cas9 knockout screens. In control of the false discovery rate (FDR) and its high sensitivity, it performs better current computational methods. As well as its results are also strong across different sequencing depths and numbers of sgRNAs per gene. The algorithm is able to complete both positive and negative selection screens at the same time, and detect statistically significant effect and marker genes also knowns as cell type-specific (CTS) essential genes and pathways. The MAGeCK algorithm for CRISPR-based genetic screens includes; Variance Estimation, Data Normalization, Gene Ranking, Statistical Testing, Significance Calculation, Ranking of sgRNAs, Pathway Analysis, Robust Ranking Aggregation (RRA) (28).

Robust Ranking Algorithm (RRA): Robust Ranking Algorithm is a part of the MAGeCK it is specifically designed for CRISPR knockout screens. This specialization enhances its relevance and effectiveness in this context. It takes raw sgRNA read counts as input by providing a clear analysis

based on the primary experimental data. This can be beneficial for users who prefer working directly with raw counts. Using a modified robust ranking algorithm, it combines guide-level p values at the gene level. This method is designed to handle variations in the data by giving more reliable ranking of genes based on their possible significance (29).

Maximum Likelihood Estimation (MLE): Maximum Likelihood Estimation another component of MAGeCK. MLE is designed to address the challenge of estimating gene effects across CRISPR knockout screens that span multiple conditions like different cell lines or drug treatments. This makes it useful and relevant to experiments with diverse settings. It clearly combines variable sgRNA knockout efficiencies. This is critical because not all sgRNAs have the same efficiency in making gene knockout and accounting for this variability improves the accuracy of the estimation. It requires a design matrix that allow users to specify which counts come from specific conditions. This is advantageous when dealing with complex experimental designs, as it allows for a more detailed analysis by considering the contributions of different conditions. It provides estimates of gene effect size that is helpful for understanding the biological significance of identified genes (30).

High-Throughput Screen Selection Pipeline (HiTSelect): It analyze pipeline for thoroughly selecting screens and identifying functionally related genes and pathways by addressing off-target effects, monitoring for variance in both efficiency of gene silencing and sequencing depth of coverage while integrating appropriate metadata. It is implemented for pathway exploration and data visualization. It is designed for inclusive HiT selection in high-complexity CRISPR screens, including genetic interaction, gene network visualization, efficient annotation, robust statistics, and gene appearance integration. It outperforms similar algorithms (RIGER and RSA) in sensitivity, specificity, and reduced off-target effects. It also offers exclusive tools for analysis and metadata integration (31).

Bayesian Analysis of Gene Essentiality (BAGEL): Bayesian Analysis of Gene Essentiality (BAGEL) method is for measuring gene essentiality in CRISPR knockout screens. It is precisely personalized for evaluating data from high-throughput genetic screens and the goal is to control which genes are essential for cell feasibility or capability. For analyzing pooled library CRISPR screens BAGEL is an enhanced version of their previous Bayesian classifier. It includes a more robust substantial performance enhancements, Bayesian Analysis, statistical model, reference sets and a better user interface (32). There is also a new version of BAGEL which is BAGEL2, which is an upgraded model that provides a more dynamic range of Bayes Factors for detection of tumor suppressor gene, a multi-target improvement that decreases false positives from off-target CRISPR guide RNA (33).

Computational Technique for Gene Dependency Analysis (CERES): A computational technique for the analysis of gene dependency levels from CRISPR-Cas9 essentiality screens while accounting for the copy-number-specific effect which includes correcting the analysis to reflect variations in the number of gene copies in different cells. Its application helps out in decreasing false-positive results and provides more precise estimates of sgRNA activity (34). It normalizes the data in multiple cell lines, which is mainly important in cancer research, cell lines may take diverse genetic backgrounds and copy number variation (CNV) profiles. This normalization helps

to separate the precise effects of individual genes. The advanced dataset is a powerful tool for understanding genetic dependencies in many types of cancer. (35).

Joint Analysis of CRISPR/Cas9 Knockout Screens (JACKS): Joint Analysis of CRISPR/Cas9 Knockout Screens (JACKS) is a Bayesian method that jointly analyzing numerous CRISPR screens. JACKS evaluates reproducible gRNA efficiencies that improves gene essentiality computation and proceed over present methods. It copies log2 fold changes of gRNA read counts among control conditions and treatment as a creation of treatment-dependent gene essentiality and treatment-independent gRNA efficiency. It beats current single-screen analysis approaches such as MeanFC, MAGeCK, BAGEL, ScreenBEAM, and PBNPA by improving accuracy for a large number of cell lines tested with minor errors on average contrast to different methods (36).

CRISPR-based Genomic Screen Analysis Tool (CRISPhieRmix): It precisely designed for CRISPR-based genomic screens. CRISPhieRmix uses a classified combination model to estimate variable sgRNA efficiencies and a longer-tailed null distributions also with findings in genetic screens. It computes false discovery rates (FDRs) by allowing the subsequent probability that each gene is nonessential. It considers all possible combination distributions to get the final FDRs. This method displays important enhancements in the analysis of CRISPR interface and CRISPR activation screens since it can differentiate genes with varying guide efficiencies (37). CRISPhieRmix grant for more correct and dynamic inferences in significant pooled CRISPRi/a screens.

Redundant siRNA Activity (RSA): Redundant siRNA Activity estimates the joint effect of multiple siRNAs directing the similar gene. It identifies significant genes in RNA interference loss-of-function screens by ranking all targeting guides initially through reducing log fold change between the initial and final condition. It allows for rare off-target guides with high effect sizes to be relegated compared to guides by measuring the statistical significance of a gene having highly ranked guides (38).

DISCUSSION:

The research explores the CRISPR/Cas9 technology and the essential role of algorithms in increasing its accuracy and efficiency. The use of CRISPR technology has changed genetic engineering which enables researchers to exactly operate and adapt genomes. The worth of this technology spreads across several scientific disciplines within a range from synthetic biology to human gene treatment and agriculture.

Algorithm's role in CRISPR/Cas9 Technology

Algorithms which are mainly designed for CRISPR/Cas9, plays a crucial role in the success of this gene-editing technology. The capability of algorithms to recognize specific DNA targets, enhance guide RNA design, predict experimental results and observe data which gives a strong support for advancing CRISPR applications. Particularly, CRISPR-Cas9's is to identify new genomic sequences in genome editing and development in various fields such as drug discovery, disease model, agriculture, and neuroscience.

Significance of Predictive Tools

Algorithms including SPROUT, have developed as valuable predictive tools for CRISPR results. It is a machine learning-based algorithm for guessing errors in CRISPR-Cas9 gene editing, increasing accuracy and protection. It utilizes gradient boosting approach which identifies error-prone DNA sequences modifying off-target effects. This analytical ability helps researchers in leading harmless and further accurate CRISPR-Cas9 gene editing, especially in study of diseases.

DECODR another algorithm presents an innovative deconvolution method for analyzing CRISPR editing results. Its adaptability in treatment indels of changing sizes and similarity with both single and multi-guide CRISPR researches places it as a valuable tool for researchers looking for complete understandings into gene editing efficiency and accuracy.

Addressing Challenges and Ethical Reflections

Using CRISPR/Cas9 for gene editing gives challenges like unpredicted genomic changes and off-target effects. Algorithms which are designed to challenge these issues have also their restrictions. The study pressures the requirement for severe justification methods and algorithm design to report difficulties like overprediction, lacking validation of guide RNAs, and potential off-target effects. Ethical thoughts are critical in discussing CRISPR. The impulsive and uncontainable effect of improved genes in populations, highlighting the requirement for careful implementation.

CONCLUSION:

In conclusion, this research explains the critical effect of CRISPR/Cas9 technology on genetic engineering and the vital role algorithms in determining its future. The flexibility of CRISPR, joined with the accuracy presented by algorithms which gives new paths for improvements in medicine and agriculture. The research directs the difficulty of gene editing algorithms, presents understandings into their robustness and limitations. CRISPR technology has vast potential but continuing research and ethical selections are critical. We want invention to overcome challenges and provide responsible use in the rapidly developing field of genetic engineering.

REFERENCES:

(1) Tavakoli, Kamand et al. "Applications of CRISPR-Cas9 as an Advanced Genome Editing System in Life Sciences." BioTech (Basel), vol. 10, no. 3, 2021, article 14. DOI: 10.3390/biotech10030014.

(2) Gostimskaya, Irina. "CRISPR–Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing." Biochemistry (Mosc), vol. 87, no. 8, 2022, pp. 777–788. DOI: 10.1134/S0006297922080090.

(3) Asmamaw, Misganaw, and Belay Zawdie. "Mechanism and Applications of CRISPR/Cas-9-

(4) Mittal, Rama Devi. "Gene Editing in Clinical Practice: Where are We?" Indian Journal of Clinical Biochemistry, vol. 34, no. 1, 2019, pp. 19-25. DOI: 10.1007/s12291-018-0804-4.

(5) Plumer, Brad, et al. "A Simple Guide to CRISPR, One of the Biggest Science Stories of the Decade." Vox, 27 Dec. 2018.

(6) Li, Hongyi et al. "Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances, and prospects." Signal Transduction and Targeted Therapy, vol. 5, no. 1, 2020, article 1. DOI: 10.1038/s41392-019-0109-2.

(7) Chen, Yuhao, and Xiaowei Wang. "Evaluation of Efficiency Prediction Algorithms and Development of Ensemble Model for CRISPR/Cas9 gRNA Selection." Bioinformatics, vol. 38, no. 23, 2022, pp. 5175–5181. https://doi.org/10.1093/bioinformatics/btac681.

(8) Gaj, Thomas et al. "Genome-Editing Technologies: Principles and Applications." Cold Spring Harbor Perspectives in Biology, vol. 8, no. 12, 2016, article a023754. DOI: 10.1101/cshperspect.a023754.

(9) Zhang, Huimin et al. "Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer." Molecular Cancer, vol. 20, 2021, article 126. DOI: 10.1186/s12943-021-01404-5.

(10) Mah, Amanda, and Robert. "Genome Editing Techniques: The Tools That Enable Scientists to Alter the Genetic Code." Bioengineering, 2022.

(11) Davies, Benjamin. "The technical risks of human gene editing." Hum Reprod, vol. 34, no. 11, 2019, pp. 2104–2111. DOI: 10.1093/humrep/dez162.

(12) Fu, Yanfang et al. "High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells." Nat Biotechnol, vol. 31, no. 9, 2013, pp. 822–826. DOI: 10.1038/nbt.2623.

(13) Uddin, Fathema et al. "CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future." Front Oncol, vol. 10, 2020, article 1387. DOI: 10.3389/fonc.2020.01387.

(14) Rubeis, Giovanni, and Florian Steger. "Risks and benefits of human germline genome editing: An ethical analysis." Asian Bioeth Rev, vol. 10, no. 2, 2018, pp. 133–141. DOI: 10.1007/s41649-018-0056-x.

(15) Uddin, Fathema et al. "CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future." Front Oncol, vol. 10, 2020, article 1387. DOI: 10.3389/fonc.2020.01387.

(16) Bodapati, Sunil, et al. "A benchmark of algorithms for the analysis of pooled CRISPR screens." Genome Biology, vol. 21, 2020, article 62. DOI: 10.1186/s13059-020-01972-x.

(17) Zhang, Huimin, et al. "Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer." Molecular Cancer, vol. 20, 2021, article 126. DOI: 10.1186/s12943-021-01404-5.

(18) Ma, Xingliang, et al. "A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants." Molecular Plant, vol. 8, no. 8, 2015, pp. 1274–1284. DOI: 10.1016/j.molp.2015.04.007.

(19) Armitage, Hanae. "CRISPR algorithm predicts how well gene editing will work." Stanford Medicine Magazine, 2019.

(20) Xiang, Xi, et al. "Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning." Nature Communications, vol. 12, 2021, article 3238. DOI: 10.1038/s41467-021-23576-0.

(21) Bloh, Kevin, et al. "Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data." CRISPR Journal. DOI: 10.1089/crispr.2020.0022.

(22) Zhou, Hong, et al. "Whole genome analysis of CRISPR Cas9 sgRNA off-target homologies via an efficient computational algorithm." BMC Genomics, vol. 18, supplement 9, 2017, p. 826. DOI: 10.1186/s12864-017-4212-7.

(23) Liu, Xiuqin, et al. "Predicting CRISPR/Cas9 Repair Outcomes by Attention-Based Deep Learning Framework." Cells, vol. 11, no. 11, 2022, p. 1847. https://doi.org/10.3390/cells11111847.

(24) Chen, Qinchang, et al. "Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction fingerprints." Nature Communications, vol. 14, 2023, Article number 7521. https://doi.org/10.1038/s41467-023-08024-8.

(25) Liu, Xiuqin, et al. "Predicting CRISPR/Cas9 Repair Outcomes by Attention-Based Deep Learning Framework." Cells, vol. 11, no. 11, 2022, article 1847. https://doi.org/10.3390/cells11111847.

(26) Bloh, Kevin, et al. "Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data." CRISPR Journal, vol. 4, no. 1, 2021, pp. 120-131. https://doi.org/10.1089/crispr.2020.0022.

(27) Konstantakos, Vasileios, et al. "CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning." Nucleic Acids Research, vol. 50, no. 7, 22 April 2022, pp. 3616–3637. https://doi.org/10.1093/nar/gkac192. Published: 29 March 2022.

(28) Li, Wei, et al. "MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens." Genome Biology, vol. 15, no. 12, 2014, article number: 554. https://doi.org/10.1186/s13059-014-0554-4. Published: 05 December 2014.

(29) Bodapati, Sunil, et al. "A benchmark of algorithms for the analysis of pooled CRISPR screens." Genome Biology, vol. 21, 2020, article number: 62. https://doi.org/10.1186/s13059-020-01972-x. Published: 09 March 2020.

(30) Bodapati, Sunil, et al. "A benchmark of algorithms for the analysis of pooled CRISPR screens." Genome Biology, vol. 21, 2020, article number: 62. https://doi.org/10.1186/s13059-020-01972-x. Published: 09 March 2020.

(31) Diaz, Aaron A., et al. "HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis." Nucleic Acids Research, vol. 43, no. 3, 2015, e16. DOI: 10.1093/nar/gku1197. Published online 2014 Nov 26. PMCID: PMC4330337. PMID: 25428347.

(32) Hart, Traver, and Jason Moffat. "BAGEL: a computational framework for identifying essential genes from pooled library screens." BMC Bioinformatics, vol. 17, 2016, p. 164. DOI: 10.1186/s12859-016-1015-8. Published online 16 April 2016.

(33) Kim, Eiru, and Traver Hart. "Improved analysis of CRISPR fitness screens and reduced off-target effects with the BAGEL2 gene essentiality classifier." Genome Medicine, vol. 13, 2021, p. 2. DOI: 10.1186/s13073-020-00819-2. Published online 06 January 2021.

(34) Meyers, Robin M., et al. "Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells." Nat Genet, vol. 49, no. 12, 2017, pp. 1779–1784. DOI: 10.1038/ng.3984. Published online 2017 Oct 30. PubMed Central (PMC) ID: PMC5709193. NIH Manuscript Submission (NIHMS) ID: NIHMS910985. PMID: 29083409.

(35) Bodapati, Sunil, et al. "A benchmark of algorithms for the analysis of pooled CRISPR screens." Genome Biology, vol. 21, 2020, article number 62. DOI: 10.1186/s13059-020-01972-x. Published online 09 March 2020.

(36) Allen, Felicity, et al. "JACKS: joint analysis of CRISPR/Cas9 knockout screens." Genome Research, vol. 29, no. 3, 2019, pp. 464-471. DOI: 10.1101/gr.238923.118. Published online 2019 Mar. PMC6396427. PMID: 30674557.

(37) Zhao, Yueshan, et al. "Bioinformatics approaches to analyzing CRISPR screen data: from dropout screens to single-cell CRISPR screens." Quantitative Biology, vol. 10, no. 4, 2022, pp. 307-320. DOI: 10.15302/J-QB-022-0299.

(38) Bodapati, Sunil, et al. "A benchmark of algorithms for the analysis of pooled CRISPR screens." Genome Biology, vol. 21, 2020, article number 62. DOI: 10.1186/s13059-020-01972-x. Published online 09 March 2020.