Brendan Ready
COMP 662
Module 1 Assignment: Research Paper

How Big Data is Being Utilized in CRISPR Gene Editing

Abstract

CRISPR and the Cas9 enzyme allow researchers to utilize any cell's own gene repair mechanism to add, delete, or modify the genes of any living organism, similar to how a programming language is used to edit a database. Big data can be used to construct a database of any living organism's genome, and relationships between genes and their expressive traits can be mapped. The most promising use of CRISPR combined with big data is in the field of medicine, where genome sequencing and comparative analysis using algorithms can now be quickly and inexpensively done to gain insights into what is causing a disease and finding possible cures to that disease. Genetic engineering can then be tailored to an individual patient to treat the specific causes of their disease. However, there are many limitations and risks to using CRISPR and big data often fails to show what is causing more complex diseases.

Introduction

A genome is similar to a database that stores genetic material in the form of DNA. Thus, information about our genes can be compiled into datasets that can be analyzed. With the advent of CRISPR gene editing, we are now able to edit that genomic database. DNA is composed of a double helix of four base pairs of nucleotides: thymine, guanine, cytosine, and adenine. These nucleotides repeat in certain sequences. Among these repeated sequences are spacers, which are specific

pieces of genes. Genes are similar to key/value pairs in the codebase of a database in how they control the expression of certain traits. CRISPR is like a programming language that can edit DNA's nucleotide sequences and specific genes. It uses a protein called CAS9 that acts like a molecular scissors and can be programmed to cut the DNA exactly where we want. Cas9 was discovered in bacteria that was using it as a way to defend itself against viruses. Cas9 carries an RNA guide that is a one sided strand of certain nucleotide base pairs that can be custom built to match a specific DNA sequence. This RNA guide binds to a targeted sequence of DNA in a genome. When this RNA guide lines up with a targeted strand of DNA, the Cas9 enzyme cuts it at that exact location. Once the DNA is cut, scientists can hack the repair process of DNA to introduce template DNA that guides what new nucleotide base pairs are inserted in the gap. Thus, using CRISPR, scientists can now change an organism's DNA by adding, removing, or altering genetic material at specific locations within the genome (GetSmarter, 2019). With CRISPR, scientists are now able to quickly and easily manipulate an entire genome.

Body

Being able to edit a genome in the lab is one thing, but the question now becomes how to use it for practical applications in the real world. This is where big data proves to be most helpful. Random genes could be inserted into random locations in a genome, and random genes could be deleted from a genome, but this Frankenstein approach would produce random and unwanted results. Big data allows researchers to quickly store and analyze large sets of data. Today, next-generation sequencing can sequence an entire genome in 24 hours for little more than a thousand dollars. Now that

scientists are quickly able to perform genome sequencing, large data sets of a species' genome can be produced and data mined to perform targeted research into genetic patterns. Machine learning can be applied to these large multi-dimensional datasets and predictive models can be built to produce predictable outcomes. The final outcome using CRISPR would ideally heal a specific disease or create a desired outcome in the form of genetic engineering. For example, a cancer causing mutation in one's genes could be remedied so that the risk of developing that cancer would be mitigated. CancerLinQ (Cancer Learning Intelligence Network for Quality) is a big data program that has sequenced the genomes of 1,500 people's cancers to find the DNA mutations driving them, and to find which mutations correlated with which outcomes (Mosaic, 2018). This has revealed complex patterns of the genetic causes of cancers, and can find the exact genes that are responsible for causing a particular tumor. Another example is that of using genetic engineering to increase the size of fruits and vegetables or increase crop yields. One such project in genetic engineering is CropsOS, which uses machine learning and predictive analytics to create a gene editing system to improve the traits of crops such as flavor, nutrient density, and sustainability (Sterling, 2017).

The most promising use of using big data and CRISPR is in the field of medicine. Big data is being used to increase our understanding of the genetic causes of many diseases. In studies, electronic health records are being linked to genetic data that has been gathered for over a million Americans that utilizes entire genome sequences, proteins, metabolites, RNA, DNA, and behavioral traits (GetSmarter, 2019). Big data can be used to perform comparative analysis between an individual and the entire

population. Researchers can then identify variations in one's genes and harmful mutations. Moreover, we can compare the entire genome of a parent and child to look for genetic differences and perform a more targeted approach to identify unique changes in the genes of a child or a combination of genetic traits that have caused a disease. Even more targeted approaches can be done by comparing the genomes of patients with similar medical histories. Artificial intelligence can be used to analyze the genomic patterns from patients who appear to have the same disease, and it can pinpoint mutations, variations, and changes in the DNA of only a small number of cells (Sandoz, Nd).The results can reveal complex patterns and interactions of DNA, which can then be used to identify medical treatments using CRISPR.

Big data has led to the rise of the genome-wide association study, a form of analysis that uses the data collected to form a hypothesis about what is causing a particular disease. The genomes of a large number of people are sequenced and then compared using an algorithm. Such algorithms can be simple or quite complex. Algorithms allow researchers to compare how frequently a certain variant of DNA appears in people who have a certain disease or trait versus people who don't have that disease or trait. If the DNA variant appears alongside a trait or disease, the algorithm flags it as a possible cause (Mosaic, 2018). Big data is even being used to study the 3D shapes of chromosomes and how they affect gene expression (Mosaic, 2018). Genes can fold very differently depending on what type of cells they are located in the body and this can affect gene expression. This can then be taken into account when using CRISPR in different types of cells in order to properly fold proteins when performing genetic engineering. Because genome sequencing has become more affordable and

because CRISPR/Cas9 makes genetic engineering easier and cost-effective, it is hoped that it will lead to the development of precision medicine where customized treatments can be developed for diseases that are based on an individual's unique genetic profile.

Clinical trials are taking place using CRISPR to treat blood disorders, cancers, eye disease, chronic infection, and rare protein-folding diseases. One such trial was recently done using a type of immunotherapy in which patients' own T cells were genetically modified to better detect and kill the cancer causing cells. The therapy involved making four genetic modifications to T cells to give them a claw-like protein that detects and grabs onto certain types of cancer cells, and also removes three genes that can interfere with the T cells ability to kill cancers (National Cancer Institute, 2020). It is also likely that CRISPR will be able to treat sickle-cell disease, which is well-suited for CRISPR gene therapy because it targets a specific type of cell (Shwartz, 2018). Clinical trials at Stanford University are currently taking place to use CRISPR to correct the defective hemoglobin genes that deform red blood cells and cause sickle-cell disease.

Big data analysis is proving extremely useful in finding targeted uses of CRISPR to cure genetic diseases, find drug treatments, and stop viruses. It shows a lot of promise in treating cardiovascular disease. Big data can give us insights into how CVD develops, by analyzing lipid metabolism, electrophysiology, and cardiomyopathies (Krittanawong, et al., 2017). CRISPR can then be used to inactivate or delete certain harmful mutations in the ACE and CAD receptors that cause CVD. Or it can be used to add certain genes to specific metabolic pathways in the genome that prevent CVD. Another promising use case of CRISPR is in the development of new drugs to treat

heart disease by figuring out which genes are involved in transforming human embryonic stem cells into heart muscle cells by turning specific genes on and off repeatedly (Shwartz, 2018). CRISPR is also being used outside the body to treat diseases. This ex vivo gene editing makes it easier for CRISPR to target the exact cells that need to be edited. In February 2019, researchers successfully treated a patient with beta thalassemia using this ex vivo gene editing method. Finally, CRISPR could be used to treat viruses. Research is being done to use it to treat Covid-19 by targeting and degrading the RNA of the virus.

Despite the promise of CRISPR and the use of big data, they have their limitations and there are numerous safety concerns. There are limitations to treating complex diseases like cancers that have numerous environmental as well as genetic causes that vary greatly based on the individual. Conclusions drawn from using big data are never going to be able to reveal all the causes of every type of cancer on earth, and using CRISPR is never going to cure every cancer on earth. Although the immunotherapy therapy using CRISPR was found to be safe, the study found that the treatment had only a small effect on the study participants' cancers, with only 10% of the T cells used in the therapy receiving all four genetic modifications that were attempted using CRISPR (National Cancer Institute, 2020). Some study participants also had some off-target genetic edits in their modified T cells that were harmless, but unintended. It is currently risky to carry out genetic engineering in humans because the CRISPR gene editing method is far from being perfected. Similar to how programming errors can easily be introduced into a code base that causes an entire database to crash, studies have shown that CRISPR can also introduce unwanted changes in a

genome. CRISPR gene editing can cause unpredictable results and unwanted effects that lead to unintended genetic changes. Three studies have shown that large unintended DNA deletions and reshuffling often take place when using CRISPR gene editing (Nature, 2020). These three experiments that used CRISPR have revealed how the gene editing process can make large, unwanted changes to the genome at or near the target site (Nature, 2020). Introducing genetic changes into the body can also lead to a severe immune system response. In 1999 a patient died after his own immune system launched a massive attack against a virus carrying a gene therapy he had received (National Cancer Institute, 2020). Despite its promise in revolutionizing the treatment of diseases, a more realistic outlook shows that it is still too risky to use CRISPR to treat diseases in standard medicine, and it is never going to be a cure-all.

Besides health concerns, there are many ethical and moral concerns over using CRISPR in human embryos. This is because using genetic engineering in an embryo creates a permanent change to the genome that will be passed down for generations (Nature, 2020). It is still a mystery as to how DNA in human embryos that have been cut by the CRISPR method make genetic repairs. Studies using CRISPR in mouse embryos as well as human embryonic stem cells have demonstrated that editing chromosomes there can cause large, unwanted effects (Nature, 2020). There are also moral concerns over how parents could use genetic engineering to create the perfect child or how governments could use genetic engineering to create super soldiers. If successful, these designer babies with perfect gene traits could give the wealthy a tremendous advantage over everyone else in society. Using CRISPR in an attempt to create the perfect child could lead to many unintended consequences. Genetics is

extremely complex and there's a lot we don't understand about the expression of genes. Rather than making a child smarter, taller, and stronger, using CRISPR in human embryos might in fact disrupt gene functioning and lead to new genetic diseases because of its unpredictability. The human genome is too complex at this point to easily genetic engineer with predictable results.

Conclusion

Using big data techniques alongside CRISPR has numerous limitations and risks. Inherited diseases that affect different cell types in different organs such as cystic fibrosis and many types of cancers that affect different cell types are difficult to analyze and treat due to their complexity. Using CRISPR in human embryos could create unintended and permanent effects that affect future generations. It is unlikely that we will be able to create designer babies anytime soon. It is also unlikely that we will be able to create a cure for every type of cancer, let alone every disease. Because of the complexity of the human genome, there is still a lot we don't understand. Despite collecting and analyzing large amounts of data, the human genome is still very much a mystery and this alone will prevent us from using genetic engineering on humans anytime soon. Genetic engineering has been successfully done on simpler genomes, such as crops, to create desired traits. Big data and CRISPR have also shown promise in solving diseases that have a clear and identifiable genetic cause. It is promising that CRISPR will soon be used to create individually tailored medicine to develop gene therapies using a targeted approach. A realistic assessment of big data and CRISPR show us that they will help us cure many types of disease, but not all of them.

Work Cited

National Cancer Institute. “How CRISPR Is Changing Cancer Research and Treatment.” *National Cancer Institute*, NIH, 27 July, 2020, https://www.cancer.gov/news-events/cancer-currents-blog/2020/crispr-cancer-research-treatment.

Sandoz. “Genomics and big data – unlocking the code to new therapies.” *Sandoz, A Novartis Division,* Sandoz Saudi Arabia, Nd, https://www.sa.sandoz.com/node/21476.

Mosaic. “How big data is changing science.” *Mosaic Science,*Medium, 2 Oct., 2018, https://medium.com/mosaic-science/how-big-data-is-changing-science-97201e911bf0.

GetSmarter. “How Big Data is Changing Genetic Research.” *GetSmarter,* 2 U, Inc., 18 July, 2019, https://www.getsmarter.com/blog/market-trends/how-big-data-is-changing-genetic-research/.

Shwartz, Mark. “Target, delete, repair.” *Stanford Medicine Magazine,* Stanford Medicine, 2018, https://stanmed.stanford.edu/2018winter/CRISPR-for-gene-editing-is-revolutionary-but-it-comes-with-risks.html.

Shwartz, Mark. “The power of CRISPR.” *Stanford Medicine Magazine,* Stanford Medicine, 2018,

https://stanmed.stanford.edu/2018winter/scientists-are-refining-CRISPR-for-wider-applications.html.

Ledford, Heidi. “CRISPR gene editing in human embryos wreaks chromosomal mayhem.” *Nature,* Nature, 25 June, 2020, https://www.nature.com/articles/d41586-020-01906-4.

Henderson, Hope. “CRISPR Clinical Trials: A 2021 Update.” *Perspectives,* Innovative Genomics Institute, 3 March, 2021, https://innovativegenomics.org/news/crispr-clinical-trials-2021/.

Krittanawong, Chayakrit, et al. “Big Data and Genome Editing Technology: A New Paradigm of Cardiovascular Genomics.” *Current Cardiology Reviews,* US National Library of Medicine, Nov. 2017, Vol.13 , Issue 4, pp. 301-304, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5730963/

Uddin, Fathema, et al. “CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future.” *Frontiers in Oncology,* Frontiers, 7 August, 2020, https://www.frontiersin.org/articles/10.3389/fonc.2020.01387/full.

Sterling, John. “Fully enabling genome-editing system for crop improvement launched.” *Genetic Engineering & Biotechnology News,* GEN, Oct., 2017, https://www.genengnews.com/insights/fully-enabling-genome-editing-system-for-crop-improvement-launched/.

Nova, PBS. “The Realities of Gene Editing with CRISPR.” *Nova*, PBS, 9 Sept., 2020, [Video]. YouTube. https://www.youtube.com/watch?v=E8vi_PdGrKg&t=571s.