COMPUTATIONAL INSIGHTS INTO CRISPR/CAS9 SYSTEM FOR IMPROVED GENOME EDITING

JASPREET KAUR DHANJAL

DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
APRIL 2019

COMPUTATIONAL INSIGHTS INTO CRISPR/CAS9 SYSTEM FOR IMPROVED GENOME EDITING

by

Jaspreet Kaur Dhanjal

Department of Biochemical Engineering and Biotechnology

Submitted

in fulfilment of the requirements for the degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi

April 2019

Certificate

This is to certify that the thesis entitled **'Computational insights into CRISPR/Cas9 system for improved genome editing'** being submitted by **Ms. Jaspreet Kaur Dhanjal** to the Indian Institute of Technology Delhi for the award of the degree of **'Doctor of Philosophy'**, is a record of the bonafide research work carried out by her, which has been prepared under my supervision in conformity with the rules and regulations of the Indian Institute of Technology Delhi. The research reports and the results presented in this thesis have not been submitted for any degree or diploma in any other University or Institute.

Dr. D. Sundar
Professor
Department of Biochemical Engineering and Biotechnology
Indian Institute of Technology Delhi

Acknowledgements

First and foremost I want to thank my supervisor, **Prof. D. Sundar**. I appreciate all his contributions of time and ideas to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am also thankful for the excellent example he has provided as a successful professor and a great human.

Besides my supervisor, I would like to thank the rest of my research committee- Prof. G.P. Agarwal, Prof. Ritu Kulshreshtha and Prof. Samudrala Gourinath (JNU, New Delhi), for their insightful comments and encouragement, and also for the hard questions which incented me to widen my research from various perspectives. I am also grateful to Dr. Manish Agarwal, Administrator for The Central Hybrid Supercomputing Cluster at IIT Delhi, for being very kind, patient, and always willing to support whenever I approached him.

The members of my research group (DAILAB @ IIT Delhi) have contributed immensely to my personal and professional time at the institute. The group has been a source of friendship as well as good advice and collaboration. It was a pleasure to work with all the scholars and the numerous other grad students and interns who have come through the lab. I would like to specially mention Mr. Shashank, Navaneethan, Shreya, Samvit, Yugesh and Vivek for helping me in carrying out my work. Learning from them was a great experience.

Lastly, I would like to thank my family for all their love and encouragement, for my parents and brothers who supported me in all my pursuits.

Jaspreet Kaur Dhanjal

Abstract

In spite of a decade of research in the area of targeted genome editing, which has revolutionized the field of science, technology and medicine, there are still many aspects that limit the widespread use of genome editing tools for targeted cleavage of the region of interest in genomic DNA. Various issues associated with these tools include the lack of specificity and precision, low efficiency in mammalian cells, delivery into the cells and immunogenicity. In this thesis, I have studied the problem of specificity and precision of one of the most recent genome editing tool-CRISPR/Cas9 system.

The ability to direct the CRISPR/Cas9 nuclease to a unique target site within a genome holds great promise in the field of targeted genome engineering. However, CRISPR RNA is reported to bind to other genomic locations that differ from the intended target site by a few nucleotides, demonstrating significant off-target activity. Previous studies have sought to predict CRISPR target sites and their corresponding possible off-targets within a genome that are mainly based on sequence similarity between CRISPR RNA and the genomic DNA. Moreover, the intricacies of RNA-DNA interaction recently reported in literature, namely the mismatch, single guide RNA and DNA bulge tolerance pattern are not yet explored for screening the off-targets. This thesis is an attempt to understand the working of CRISPR/Cas9 system, mainly focusing on the factors governing its on-target and off-target activity. A webserver called *CRISPcut* has been developed for prediction of optimal sgRNAs for targeting the human genome in different cell types. The predictions made by this webserver were further improved using a machine learning-based model that gives the probability of a human genomic region (similar to the target site) to be an off-target location for cleavage. Finally, the applicability of CRISPR/Cas9 system to target cancer-specific cell vulnerabilities for designing new therapeutic interventions was investigated.

सारांश

लक्षित जीनोम संपादन के क्षेत्र में एक दशक के अनुसंधान ने विज्ञान, प्रौद्योगिकी और चिकित्सा के क्षेत्र में क्रांति लाई। इसके बावजूद अभी भी कई पहलू हैं जो जीनोम संपादन उपकरण के व्यापक उपयोग को उद्देशित जीनोमिक डीएनए के लक्षित अनुभेदन के लिए सीमित करते हैं। इन उपकरणों से जुड़े विभिन्न मुद्दों में विशिष्टता और परिशुद्धता की कमी, स्तनधारी कोशिकाओं में कम दक्षता, कोशिकाओं में वितरण और अनियंत्रितता शामिल हैं। इस थीसिस में, मैंने सबसे आधुनिक जीनोम संपादन साधन- क्रिस्पर/कैस9 प्रणाली की विशिष्टता और सटीकता की समस्या का अध्ययन किया है।

जीनोम में एक अद्वितीय लक्ष्य स्थल के लिए क्रिस्पर/कैस9 नुक्लिएस को निर्देशित करने की क्षमता लक्षित जीनोम अभियांत्रिकी के क्षेत्र में अद्‌भुत क्षमता दर्शाती है। यद्यपि क्रिस्पर आरएनए को कुछ न्यूक्लियोटाइड्स द्वारा लक्षित स्थल से भिन्न अन्य जीनोमिक स्थानों से बाँधने की सूचना दी गई है, जो महत्वपूर्ण ऑफ-टार्गेट गतिविधि का प्रदर्शन करते हैं। पिछले अध्ययनों ने एक जीनोम में क्रिस्पर लक्ष्य साइटों और उनके संबंधित संभावित लक्ष्य का अनुमान लगाने का प्रयास किया है, जो मुख्य रूप से क्रिस्पर आरएनए और जीनोमिक डीएनए के बीच अनुक्रम समानता से संबंधित हैं। किंतु हाल ही में साहित्य में सुचना दी गई है कि आरएनए-डीएनए परस्पर क्रिया की पेचीदगियों, अर्थात् बेमेल, एसजीआरएनए और डीएनए उभार सहिष्णुता पैटर्न को ऑफ-टारगेट की जांच के लिए नहीं खोजा गया है। यह थीसिस क्रिस्पर/कैस9 प्रणाली के कार्य को समझने का एक प्रयास है, मुख्य रूप से इसके ऑन-टारगेट और ऑफ-टारगेट गतिविधि को नियंत्रित करने वाले कारकों पर ध्यान केंद्रित कर रहा है। विभिन्न प्रकारों में मानव जीनोम को लक्षित करने के लिए इष्टतम एसजीआरएनए की पूर्वानुमान के लिए क्रिस्पकट नामक एक वेबसर्वर विकसित किया गया है। इस वेबसर्वर द्वारा किये गए अनुमानित परिणाम को एक मशीन लर्निंग-आधारित मॉडल का उपयोग करके और भी बेहतर बनाया गया, जिसने मानव जीनोमिक क्षेत्र (लक्ष्य साइट के समान) को अनुभेदन के लिए एक ऑफ-टारगेट स्थान होने की संभावना बताता है। अंततः नए चिकित्सीय हस्तक्षेपों की रचना करने के लिए कैंसर-विशिष्ट कोशिकाओं की कमियों को लक्षित करने के लिए क्रिस्पर/कैस9 प्रणाली की प्रयोज्यता की जांच की गई है।

Table of Contents

List of figures

List of tables

List of abbreviations

Cas	CRISPR associated protein
CDS	Protein coding sequence
CRISPcut	sgRNA design tool developed as a part of this thesis
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
dCas9	Catalytically dead Cas9
DSB	Double strand break
GARP	Gene Activity Rank Profile
HDR	Homology directed repair
iPSCs	Induced pluripotent stem cells
LINE	Long interspersed nuclear elements
LTR	Long terminal repeats
MMEJ	Microhomology-mediated end joining
NCBI	National Center for Biotechnology Information
NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
RNP	Ribonucleoproteins
sgRNA	Single guide RNA
SINE	Short interspersed nuclear elements
SpCas9	*Streptococcus pyogenes* Cas9
SQL	Structured Query Language
TALE	Transcription activator-like effectors
tracrRNA	Transactivating CRISPR RNA
UTR	Untranslated region
ZFN	Zinc finger nuclease