# Chapter 9: Summary and Future Implications

## 9.1 Summary

Genome editing by targeted nucleases like CRISPR-Cas9 has shown significant potential in the recent years. This approach is efficient in curing genetic diseases, chromosomal rearrangements, combating viral diseases, enhancing crop production etc. Cas9 mediated genome engineering has enhanced the generation of transgenic animal models, to engineer Embryonic and induced pluripotent stem cell disease models or in gene corrections. The effectiveness of CRISPR-Cas based genome editing makes it possible to target multiple genes simultaneously. This will help to identify genes that play a crucial function in the desired phenotype. Cas9 can be repurposed (dCas9) and can be used to activate or repress a particular gene. dCas9 will only be able to create a single-stranded nick at the target site so by fusing it with activator and repressor domains. In no time after its potential for genome editing identified ample amount of data has been generated.

Despite ample experimental studies on CRISPR-Cas based genome editing repositories for incorporating this data are lacking. We have developed "CrisprGE- a central hub of CRISPR based genome editing", a manually curated repository encompassing comprehensive details of 4680 gene sequences targeted by CRISPR-Cas approach against 32 model organisms. The database further assists the users with browsing, searching, and advanced search facilities. This database is linked with some external resources like KEGG and UniProt for a better understanding of the genes reported. Analysis tools like BLAST NTdb, BLAST CrisprGE, and CRISPR mapper help users to align their desired sequences against CrisprGE data. This database is freely accessible at http://crdd.osdd.net/servers/crisprge/.

CRISPR mediated editing always requires single guide RNA (sgRNA) that contains a sequence complementary to the target site. Selection of effective sgRNA is a requisite step in the development of sgRNA-based therapeutics. In this regard, few algorithms are available for identification and prediction of sgRNA. But an algorithm that can predict qualitative as well as the quantitative efficacy of a sgRNA is lacking. Therefore, in this endeavor we have developed "ge-CRISPR- an integrated Pipeline for the prediction and analysis of sgRNAs genome editing efficiency for CRISPR-Cas system". ge-CRISPR is a SVM based algorithm integrated in two pipelines that can predict qualitative as well as

quantitative efficiency respectively of an individual sgRNA along with its associated off-targets. This tool is developed using experimentally validated sgRNA from different studies. We anticipate that this tool will be useful in predicting highly efficacious sgRNA with minimum or no off-targets for the better specificity of genome editing by CRISPR-Cas. The web server is available at http://bioinfo.imtech.res.in/manojk/gecrispr/.

In order to minimize the deleterious effects of CRISPR-Cas i.e. off-targets, few approaches have been employed like using dCas9 or fusing it with FokI nucleases. We have developed sgRNAx a tool for the extraction of sgRNAs on the basis of PAM. We have incorporated all the available PAM sequences and Cas9 orthologs reported in the literature. This tool helps in designing sgRNAs with wild-type Cas9 and also to design sgRNA pairs with other Cas variants (dCas9 and dCas9-FokI) and PAM sequences. This tool is freely accessible at http://bioinfo.imtech.res.in/manojk/sgRNAx/.

Recently, a new class of proteins has been identified that has the capability to inhibit the activity of CRISPR-Cas machinery. These proteins are termed as Anti-CRISPR proteins. Reports have suggested that transfecting cells with Anti-CRISPR proteins after few hours of sgRNA: Cas9 transcripts will help to block the off-target activity of CRISPR editing. In this arena, we have developed AntiCrispr- an algorithm for the prediction and analysis of Anti-CRISPR proteins. We have used machine learning techniques like SVM and WEKA for developing models on various features like the amino acid composition, dipeptide composition, terminal residues composition and their hybrids. ACPs were analyzed using amino acid composition, their positional preference, and motif analysis. All constructed models are integrated into AntiCrispr web server. This tool also helps to analyze the physicochemical properties of an individual query protein. This web server is available at http://bioinfo.imtech.res.in/manojk/AntiCrispr/.

CRISPR-Cas9 genome editing holds a great potential in overcoming viral infections. Various wet lab experiments have been conducted successfully to edit viruses like HIV, EBV, HPV, HCV etc. In this regard, we have developed CrisprVir- a repository of sgRNAs against human viruses. Data in the repository is organized in a very simple manner on the basis of DNA, RNA, and retroviruses. Further viruses are categorized on the basis of family they belong. Individual entry of a sgRNA harbor important information like sgRNA sequence, PAM, its start and end coordinates, GC-AT content gene region and its associated efficiency for editing. The database is further linked with

useful external links like NCBI, taxonomy. Further for predicting off-targets CrisprVir is integrated with OFF-TarPred that can help users to find off-targets associated with their sequence of interest in desired genomes. This repository is very accessible at http://bioinfo.imtech.res.in/manojk/CrisprVir/.

### 9.2 Future implications

Genome editing by CRISPR-Cas has been going strong with each passing day. Apart from editing, data on other important application of CRISPR is being generated. Therefore, our foremost priority will be to update the CrisprGE database accordingly. Subsequently, we have trained our prediction algorithm ge-CRISPR using sequence features; we will implement other aspects of sgRNA features to further enhance the applicability of our tool. As data in this field is generating in a rapid rate, we will be focused on integrating new high throughput data on CRISPR and will update our predictive models on the new data to further improve the performance of the predictive models. Further, we will integrate new parameters like Cas9 orthologs and new PAM sequences in the sgRNAx tool to cover broad aspects in designing proficient sgRNAs. In case of Anti-CRISPR protein prediction algorithm, we will update our datasets with the new Anti-CRISPR proteins if available and will integrate them into the predictive models. We will also analyze various aspects of Anti-CRISPR proteins like analysis of Anti-CRISPR genes etc. Lastly, we have created a repository of predicted sgRNAs against human viruses. We will try to experimentally test these predicted sgRNAs in some of these viruses like Dengue etc.