Proceedings of the 4th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/60/20241415

Application of CRISPR combined with deep learning in cancer detection and therapy

Rongye Zhou

College of Informatics, Huazhong Agricultural University, Wuhan, China

zhou21@webmail.hzau.edu.cn

Abstract. Cancer is a huge health issue around the world, and genetic abnormalities are one of the leading causes. Gene editing has gained importance in cancer research with the advent of CRISPR technology, particularly the CRISPR/Cas system, which is utilized in cancer detection and treatment because of its high specificity and sensitivity. Meanwhile, as a branch of machine learning, deep learning has shown great potential in cancer detection and treatment by constructing and training multi-layer neural networks. Deep learning algorithms increase the accuracy of early cancer diagnosis by detecting gene mutations and expression patterns linked to cancer. This article reviews the most recent applications of CRISPR combined with deep learning in cancer detection and treatment. CRISPR technology has shown outstanding performance in nucleic acid testing, virus detection, and protein detection; In terms of treatment, it can be utilized to remove cancer-related genes by gene editing , enhance immune cell function, and improve the efficiency of immunotherapy. Deep learning techniques play an important role in cancer diagnosis, prognosis prediction, and CRISPR targeted and off target prediction. Combining CRISPR and deep learning is expected to improve cancer detection and treatment methods, providing new directions for future research.

Keywords: CRISPR/Cas9, Deep Learning, Cancer, Gene Editing, SgRNA.

1. Introduction

Cancer is a class of diseases distinguished by the proliferation and spread of aberrant cells which can not be controlled. Cancer is produced by abnormal genetic modifications such as proto-oncogene activation, tumor suppressor gene inactivation, and the accumulation of various genetic abnormalities. Facing a severe global health problem posed by cancer, understanding the mechanisms of genomics, cellular, and microenvironmental changes in cancer formation is crucial for its prevention, detection, and treatment. Since the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology, It is commonly used for gene editing in biology, medicine, and agriculture. Proteins of CRISPR and CRISPR-associated (Cas) are core compositions of the adaptive immune system of ancient bacterial and have now developed into powerful gene editing tools. The adaptive immune response of CRISPR/Cas9 is generated by bacteria over time that can resist both invading viruses and foreign DNA. The CRISPR/Cas9 system which is classified as a type II CRISPR system comprises a single guide RNA (sgRNA) and the Cas9 nuclease. This system induces double-strand breaks (DSBs) at particular DNA target sites through the sgRNA, which guides the Cas9 nuclease to the targeted DNA sequence. These DSBs are typically repaired by one of two mechanisms: homology-

DOI: 10.54254/2753-8818/60/20241415

directed repair if there are available homologous sequences, or non-homologous end-joining if the homologous sequences are absent. CRISPR/Cas9 gene-editing technology allows specific DNA modifications in target genes, enabling various aspects of genome editing in cells, such as immunotherapy and knockout of drug-resistant genes, advancing cancer diagnosis and treatment.

Deep learning is a sub-domain of machine learning. It constructs and trains multi-layer neural networks to solve complex tasks by simulating the way the human brain processes data and creates patterns. Deep learning adjusts network weights through forward and backward propagation algorithms by designing and training multi-layer neural networks, achieving automatic feature extraction and pattern recognition of complex data. In cancer detection and treatment, deep learning shows tremendous potential. Tumors can be automatically identified and classified through analyzing medical images with convolutional neural networks (CNNs), which improve the accuracy of early can diagnosis. Deep learning can identify cancer-associated gene mutations and expression patterns by processing large-scale genomic data, providing precise molecular diagnostics. This paper aims to investigate the most recent uses of deep learning in cancer detection and treatment with CRISPR, providing a reference for future research on enhancing diagnosis and treatment approaches in cancer.

2. Application of CRISPR in Cancer Detection

CRISPR/Cas-mediated nucleic acid detection methods are simpler and faster than traditional PCR approaches. Notably, CRISPR/Cas12a and CRISPR/Cas13a show excellent specificity and sensitivity. When paired with isothermal nucleic acid amplification techniques, CRISPR/Cas-based technologies have detection sensitivity comparable to PCR. Furthermore, the CRISPR/Cas system's ability to discriminate single-base mismatches results in extremely precise detection. These are the reasons why CRISPR performs well and be suitable for cancer detection.

Class II CRISPR/Cas systems include several subtypes, such as type II, V, and VI, each with different Cas proteins (e.g., Cas9, Cas12, Cas13). Cas9 can recognize and cleaves double-stranded DNA at specific protospacer adjacent motif (PAM) sequences; Cas12a can cleave double-stranded DNA containing 5'-TTTN-3' PAM sequences and activate side chain cleavage signal amplification. Cas13a can recognize and cleave RNA at specific PFS sequences. Cas14a can recognize and degrade ssDNA [1]. These characteristics of Cas proteins make class II CRISPR/Cas systems efficient tools for detecting various analytes.

Viral infection is one of the causes of cancer, so nucleic acid detection can be an essential method for cancer detection and routine identification of viral infections and disease progression. The CRISPR/Cas system now has been extensively utilized in the development of various nucleic acid biosensors, taking advantage of its recognition ability for single base mutations and efficient signal amplification characteristics. Researchers have combined CRISPR/Cas9 with the technology of sequence-based amplification of nuclear acid to enhance the specificity of viral nucleic acid detection. For example, CRISPR/Cas12a-assisted isothermal amplification enabled sensitive detection of human papillomavirus (HPV) DNA [2].

Electrochemical biosensor platforms and fluorescence signals are also common methods for target detection. One method that can achieve ultra sensitive detection of target analytes is to integrate isothermal nucleic acid amplification with CRISPR/Cas systems. Additionally, ctDNA detection utilizes the presence of circulating tumor DNA in the blood for cancer diagnosis. Low expression of microRNA (miRNA) is associated with certain cancer types, and the combination of PCR and CRISPR/Cas12a system can detect miRNA [3].

Abnormal protein expression can be related to cancer. Abnormal protein expression can be associated with cancer, as evidenced by higher levels of carcinoembryonic antigen in patients compared to healthy individuals. There are similar situations regarding the levels of prostate specific antigen as well as alpha fetoprotein [4]. Using horseradish peroxidase-labeled detection antibodies to provide colorimetric signals and RNA transcripts can activate the CRISPR/Cas13a system, which improves fluorescence signals for quantitative investigation of target proteins.

DOI: 10.54254/2753-8818/60/20241415

Extracellular vesicles (EVs, membrane vesicles shed by cells) contain nucleic acids, proteins, and lipids in bodily fluids and can reflect the characteristics of parent cells, potentially associating with cancer. Researchers have developed methods to detect nasopharyngeal carcinoma EV proteins using the CRISPR system. High sensitivity was demonstrated in the isothermal detection of multiple EV miRNAs using rolling loop amplification and CRISPR/Cas9 system [5].

3. Methods of CRISPR in Cancer Therapy

CRISPR can edit cancer-related genes and enhance the anti-cancer abilities of immune cells, with immunotherapy emerging as a novel treatment strategy. T-cell killing regulators in cancer cells have been identified through whole-genome screening of co-cultured cancer cells and cytotoxic T cells [5]. Therapy of chimeric antigen receptor (CAR) T cells involves collecting autologous T cells and engineering them to attack cancer antigens ex vivo before injecting them back into the patient. The CRISPR-Cas9 system can also enhance the function of CAR T cells by interrupting the genes that inhibit receptors or signaling molecules. This enables the implementation of CAR T cell therapy for B-cell malignant tumors. CRISPR/Cas9 technology can also eliminate genes encoding inhibitory receptors on T cell surfaces, like programmed death protein 1 (PD-1) and cytotoxic T-lymphocyte antigen 4, improving T cell immunotherapy efficiency [1]. CRISPR-Cas9 edited PD-1 gene knockout have shown success in enhancing T cell effector functions in certain cancer types [6].

In cancer genomics operations, CRISPR/Cas9 can correct genetic abnormalities that control cancer formation and development. One cancer treatment strategy is knocking off chemotherapy resistance genes or genes required for cancer cell survival. The CRISPR-Cas9 system can activate tumor suppressor genes, restrict the proliferation and migration of cancer cells, induce cells apoptosis, and therefore slow the progression of cancer. One example is that the CRISPR/Cas9 system has been used to silence the endogenous cyclin-dependent kinase 11 gene whose abnormal expression is associated with the occurrence of various cancers. This method has been tested in osteosarcoma cell lines and inactivate drug-resistant genes in order to improve chemotherapy efficiency as a potential cancer treatment method.

Moreover, the CRISPR-Cas9 system can explore and intervene in cancer-related epigenetic changes, playing a significant role in tumor-associated gene expression [7]. CRISPR-Cas9 mediates epigenetic alternation and transcriptional regulation, while also requiring site specificity, which can be easily achieved through the use of dead Cas9 (dCas9). Because dCas9 can associate to the object genome DNA sequence, and it can fuse with various transcriptional regulatory domains or bind with epigenetic modification factors, which can affect the proliferation of cancer cells. CRISPR-Cas9 mediated genome modification can regulate specific proteins, influencing cancer cell function and behavior.

The CRISPR/Cas9 system shows tremendous potential in eliminating or inactivating oncogenic viral infections. This technology may directly target and destroy critical viral genes and has been used to treat a variety of human viruses, including hepatitis B virus (HBV), HPV, etc. Knocking down HPV oncogenes E6 and E7 prevents cervical cancer growth; CRISPR-Cas9 technologies specific in HBV can successfully disrupt HBV covalently closed circular DNA [8].

4. Application of Deep Learning in Cancer Detection and Therapy

Traditional machine learning methods are often limited in dealing with the lack of clinical information on specific primary cancer sites when diagnosing cancers of unknown origin. They rely on a few characteristic genes and are difficult to predict more cancer types and subtypes. Deep learning algorithms improve the diagnostic accuracy of primary unknown cancers by utilizing a large number of features of the genome and transcriptome. The Pan-Cancer Analysis of Whole Genomes consortium utilized deep learning models to individually and collectively predict how multiple types of cancer originate, demonstrating the potential of deep learning in diagnosing primary unknown cancers.

To determine the presence of tumor cells in patient samples, it is essential to evaluate cancer-related biomarkers and characterize tumor type, staging, and grading, typically done through histopathology or cytopathology by microscopic observation. Deep learning technologies, particularly deep CNNs based

DOI: 10.54254/2753-8818/60/20241415

on histological images, can automate cancer grading [9]. Semantic segmentation algorithms can be used to detect specific regions in histopathological images, and generative adversarial networks (GANs) can help with object location accuracy, allowing relevant staff to comprehend and process cancer histology images. Emantic segmentation methods can be applied to histopathological images to locate specific regions, and GANs can assist in accurately locating objects, helping relevant personnel understand and process cancer histological images.

Traditional survival prediction methods such as Cox proportional hazards regression have some limitations when applied to genomic and transcriptome data, while deep learning models can improve the accuracy of prognosis prediction by utilizing nonlinear relationships. Cox-nnet is a model that combines Cox regression and neural networks, which can effectively utilize the depth features extracted from hidden layers to predict survival [10]. This model has achieved excellent accuracy on RNA seq data for various types of cancer and successfully identified biological information related to prognosis. The PASNet and Cox PASNet models combine neural networks and biological pathway information to identify important pathways and genes that affect cancer prognosis.

5. Deep Learning Assists CRISPR in Cancer Detection and Treatment

Traditional machine learning models can be used for CRISPR/Cas9 off-target and on-target predictions. However, deep learning, based on new sequence encoding strategies, feature engineering, attention mechanisms, and other technologies and models, offers more possibilities for CRISPR off-target and on-target predictions.

For the situations of off target, some studies have adopted novel deep learning architectures such as CnnCrispr and piCRISPR, relying on global statistical information or physical information features. These models show better predictive performance, while algorithms of traditional machine learning and classification regression do not perform as well as the former. According to a new sequence encoding scheme based on deep neural networks, CNN and feedforward neural network deep learning networks is stable and superior to some methods for advanced off target scoring prediction. Also, the performance of some machine learning classifiers are not as good as deep learning schemes [11].

CNN-based DeepCas9 is an efficient deep learning framework. Ten distinct CRISPR/Cas9 datasets were used in experiments to demonstrate that DeepCas9 is more effective than more conventional machine learning techniques, such as logistic regression and random forests, at predicting targeting activity. RNA guides can be recognized and predicted using DeepSgRNA. To reach the most advanced sgRNA prediction efficiency, it depends on the hierarchical feature generation capacity of CNN. The DeepCRISPR deep learning system can predict the efficacy of sgRNA targeting knockout and off-target cutting in the same time. The researchers used a original single hot coding strategy that combines four-channel sgRNA-DNA sequence encoding, where every feature is considered as an independent channel, with epigenetic feature encoding to achieve better prediction results [12].

Some deep learning models based on attention mechanisms have achieved some promising results. Convolutional, recurrent, attention, and dense layers are combined in the CNN, BLSTM, and attention layers-based CRISPR-IP off-target prediction model to overcome information loss in sequence encoding and learn from local to global features. AttCRISPR improves model interpretability and predictive performance by using spatial and temporal attention modules. CRISPR-ONT and CRISPR-OFFT are used to predict the on-target and off-target activities of sgRNA, respectively, with enhanced interpretability.

Deep learning involves four processes in predicting sgRNA activity. First, compile the data and extract sgRNAs with high cutting efficiency from published publications. Remove superfluous samples for more accurate evaluation. The generated data is divided into training dataset and testing dataset. Second, to represent data. Use numerical encoding to encode these input sequences. Common data encoding techniques include one-hot encoding and word2vec embedding. Third, train deep learning models. Input sgRNAs that have high cutting efficiency to train in deep learning models. Meanwhile, some feature information got from the sequences of sgRNA can also be used as input data to increase prediction accuracy. Finally, evaluate performance. One optimization method is k-fold cross validation,

DOI: 10.54254/2753-8818/60/20241415

which can optimize the hyperparameters of deep learning. To estimate generalization performance, the trained model should be applied to the candidate dataset [12].

The development of prediction tools of CNNs based sgRNA activity has also accelerated the process, as more and more CRISPR genome editing data is available. CNNs are the most frequent type of deep neural network for detecting abstract information in vast datasets. The strategy of weight sharing is used to obtain the hierarchical spatial pattern of the input. The parameter update is implemented using backpropagation algorithms. DeepCRISPR trains an autoencoder based on a deep convolutional denoising neural network through unsupervised learning techniques to learn abstract information of sgRNA sequences [13]. Additionally, the model combines extra epigenetic information to improve predictive capability. Recurrent neural networks (RNN) are artificial neural networks capable of processing variable-length input sequences. Due to their recurrent connections, RNNs are often used for ordered sequence problems. Long short-term memory networks (LSTM) can model hidden layers using memory units, overcoming gradient explosion or vanishing problems. These memory units have the input, output and forget gate, which operate the information flow that aids in predicting network outputs. LSTMs effectively capture dynamic information sequentially to assist in the classification of serialized data through cyclic connections of nodes in the hidden layer. For example, a DeepHF based on BiLSTM achieved the measurement of sgRNA targeting activity of variants and wild-type Streptococcus pyogenes Cas9 in human cells [14]. CNNs are limited to learning local patterns and function best when the input contains some spatially invariant patterns, but RNNs, which are derived from feedforward networks, can retain input data. C-RNNCrispr combines CNN and RNN to complement each other and achieve targeted activity prediction. A bidirectional gated recurrent unit layer is applied to capture the features of the sequence recognized by CNN in forward direction and backward direction .

The design of sgRNA requires finding target sites in the genome, which can be achieved by scanning PAM sequences. For the prediction of target sequences, researchers have developed a model to predict the targeting efficacy of sgRNA. The model first uses Support Vector Machine (SVM) to select the optimal subset from numerous features, and then uses a logistic regression classifier to train the features selected by SVM, thereby generating a model for predicting the targeted efficacy of sgRNA. The development of CRISPRpred has improved the efficiency of predicting targeting activity. WU-CRISPR identifies new structural and sequence features from specific datasets and builds an SVM-based sgRNA potency prediction model. It leverages machine learning approaches for effective sgRNA design and is simple to use [15].

For off target detection, researchers analyzed the off target effect on merged filtered data without considering how PAM works. Instead, they provided a detailed description of various mismatches of sgRNA and DNA, including substitution of PAM, mismatches, insertions, and deletions. They proposed an algorithm called Cutting Frequency Determination (CFD) score to rank potential off target effects. They compared the CFD score with Hsu Zhang's off target effect measurement results and found that the CFD score performed better in various tests. That is to say, their algorithm can effectively prevent high-frequency off target effects [16]. In addition, some tools and algorithms such as CROP-IT, E-CRISPR, and sgRNACas9 can also be used to estimate the potential in off target of sgRNA.

Additionally, some factors influence the performance of CRISPR gene editing, and key factors such as sgRNA length and spacer regions, characteristics of Cpf1 protein, sgRNA sequence constraints, and pol III promoter termination signals need attention.

6. Conclusion

The CRISPR/Cas system and deep learning show significant potential and unique advantages in cancer detection and treatment. CRISPR technology brings revolutionary changes to research and treatment of cancer with its versatility in gene editing, immune therapy, and viral gene targeting. Deep learning algorithms enhance the accuracy of cancer type diagnosis and prognosis prediction by leveraging big data and complex bioinformatics features. Combining CRISPR and deep learning enhances the understanding of cancer molecular mechanisms, improves the precision and efficiency of CRISPR

technology through sgRNA design and targeting prediction models, and lays a solid foundation for personalized treatment and precision medicine.

CRISPR still faces a series of challenges in off target effects, delivery efficiency and safety, editing efficiency and cellular adaptability. Deep learning also faces issues such as data heterogeneity, data size, and model interpretability. However, as technology advances and algorithms are optimized, deep learning will play an increasingly crucial role in the accuracy and efficiency of gene editing, expanding the value of application of CRISPR systems in cancer genomics research. Deep learning will help analyze complex cancer biomarkers and treatment response prediction models, thereby promoting the development of personalized medicine. Integrating multimodal biological data such as genome, transcriptome, and proteome can construct more comprehensive cancer detection and treatment models. In the future, CRISPR and deep learning models will be employed for large-scale gene function research and drug screening, expediting the discovery and development of new anti-cancer medications and advancing cancer treatment.

References

[1] Cheng, X., Fan, S., Wen, C., & Du, X. (2020). CRISPR/Cas9 for cancer treatment: technology, clinical applications and challenges. Briefings in functional genomics, 19(3), 209–214. https://doi.org/10.1093/bfgp/elaa001

[2] Katti, A., Diaz, B. J., Caragine, C. M., Sanjana, N. E., & Dow, L. E. (2022). CRISPR in cancer biology and therapy. Nature reviews. Cancer, 22(5), 259–279. https://doi.org/10.1038/s41568-022-00441-w

[3] Hou, J. et al. (2021). Integrating genome-wide CRISPR immune screen with multi-omic clinical data reveals distinct classes of tumor intrinsic immune regulators. Journal for immunotherapy of cancer, 9(2), e001819. https://doi.org/10.1136/jitc-2020-001819

[4] Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nature biotechnology, 33(5), 510–517. https://doi.org/10.1038/nbt.3199

[5] Zhang, X., & Meyerson, M. (2020). Illuminating the noncoding genome in cancer. Nature cancer, 1(9), 864–872. https://doi.org/10.1038/s43018-020-00114-3.

[6] Song, X., Liu, C., Wang, N., Huang, H., He, S., Gong, C., & Wei, Y. (2021). Delivery of CRISPR/Cas systems for cancer gene therapy and immunotherapy. Advanced drug delivery reviews, 168, 158–180. https://doi.org/10.1016/j.addr.2020.04.010

[7] Wang, Z., Liu, N., Feng, F., & Ma, Z. (2015). Synthesis of cadmium, lead and copper alginate nanobeads as immunosensing probes for the detection of AFP, CEA and PSA. Biosensors & bioelectronics, 70, 98–105. https://doi.org/10.1016/j.bios.2015.03.015

[8] Gong, S., Zhang, S., Lu, F., Pan, W., Li, N., & Tang, B. (2021). CRISPR/Cas-Based In Vitro Diagnostic Platforms for Cancer Biomarker Detection. Analytical chemistry, 93(35), 11899–11909. https://doi.org/10.1021/acs.analchem.1c02533

[9] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome medicine, 13(1), 152. https://doi.org/10.1186/s13073-021-00968-x

[10] Ching, T., Zhu, X., & Garmire, L. X. (2018). Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. PLoS computational biology, 14(4), e1006076. https://doi.org/10.1371/journal.pcbi.1006076

[11] He, Y., Zhao, H., & Wong, S. T. C. (2021). Deep learning powers cancer diagnosis in digital pathology. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 88, 101820. https://doi.org/10.1016/j.compmedimag.2020.101820

[12] Zhang, G., Luo, Y., Dai, X., & Dai, Z. (2023). Benchmarking deep learning methods for predicting CRISPR/Cas9 sgRNA on- and off-target activities. Briefings in bioinformatics, 24(6), bbad333. https://doi.org/10.1093/bib/bbad333

[13] Chuai, G. et al. (2018). DeepCRISPR: optimized CRISPR guide RNA design by deep learning. Genome biology, 19(1), 80. https://doi.org/10.1186/s13059-018-1459-4

[14] Wang, D. et al. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nature communications, 10(1), 4284. https://doi.org/10.1038/s41467-019-12281-8

[15] Wong, N., Liu, W., & Wang, X. (2015). WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome biology, 16, 218. https://doi.org/10.1186/s13059-015-0784-0

[16] Hsu, P. D. et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology, 31(9), 827–832. https://doi.org/10.1038/nbt.2647