Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.32136

The Critical Role of Bioinformatics in Tumor Gene Editing Research: Progress and Prospects

Xucan Liu

Department of Bioinformatics, Southern Medical University, Guangzhou, China
3247030010@i.smu.edu.cn

Abstract. The high level of genetic heterogeneity that exists within malignant tumors acts as a limiting factor to the effectiveness of traditional treatment methods. Gene editing technologies, for example, CRISPR, present new typical pathways for the purpose of precise intervention. And the efficient and safe application of these technologies depends on the situation where massive omics data is systematically analyzed by making use of bioinformatics. Bioinformatics, which serves as the core link between tumor data and precise editing, permeates the entire process from the discovery of targets and the design of protocols to the evaluation of functions. This review places its focus on the role of bioinformatics in two key stages, namely the stage before editing and the stage after editing. In the stage before editing, in the process of integrating database resources such as TCGA and DepMap and combining them with algorithms such as synthetic lethality prediction, key targets are screened. Tools such as CRISPick are used to design gRNAs, and potential off - target risks are assessed. In the stage after editing, in the process of analyzing CRISPR screening data using workflows such as MAGeCK, functional genes are identified, and the editing efficiency and genome - wide off - target effects are assessed. Overall, bioinformatics offers an integrated solution that ranges from the discovery of targets and the design of tools to the assessment of risks and the analysis of efficacy. This significantly enhances the accuracy and efficiency of tumor gene editing research. In the future, in the situation where single - cell, spatial multi - omics, and deep learning models are in an integrated development state, bioinformatics will further speed up the advancement of tumor gene editing towards precise spatiotemporal regulation and clinical translation.

Keywords: Bioinformatics, Tumor, Gene editing, CRISPR/Cas9

1. Introduction

Cancer represents a significant public - health problem that poses a threat to human health across the globe. When we refer to the statistics provided by the World Health Organization, it is found that on a global scale, nearly ten million people lose their lives due to cancer each year. The process of its occurrence and development is closely associated with gene mutations as well as regulatory abnormalities [1]. Among these cases, around 5% - 10% of cancers are directly driven by the situation where hereditary gene mutations take place. This provides an important starting point for the process of conducting early risk assessment and performing precise intervention [2]. In essence,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.32136

cancer is a complex disease which involves multiple steps and multiple genes. The core molecular mechanism of it lies in the situation where the functions of key genes acquire activation or inactivation. Specifically, when oncogene mutations occur, they often show narrow - range hotspot clusters. For example, codons 12 and 13 of KRAS lead to the abnormal activation of signaling pathways. In the case of tumor suppressor genes, they often present as widely distributed nonsense mutations or deletions. Take RB1 and PTEN as examples; these cause imbalances in the cell cycle and DNA repair. Moreover, some gene mutations have significant tumor subtype specificity. For instance, the mutation region of EGFR shows a significant difference between glioblastoma and lung adenocarcinoma [3].

In the recent years that have passed, the gene - editing technology has been undergoing a process of rapid development. This development has given rise to entirely new strategies to be used for the treatment of cancer. By the time when it reached the end of the year 2023, the very first human therapy which was based on CRISPR technology got approval for its marketing. This event marked the official entry of CRISPR genome editing into the stage of clinical translation [4]. The CRISPR/Cas9 system guides the Cas9 nuclease via sgRNA to generate DNA double-strand breaks (DSBs) near specific PAM sequences, and relies on non-homologous end joining (NHEJ) or homologous recombination repair (HDR) mechanisms to achieve gene knockout, insertion, or repair, thereby achieving precise gene regulation [5]. With its advantages of ease of operation, high efficiency, and strong scalability, CRISPR/Cas9 has been widely used in research fields such as tumor driver gene screening, disease model construction, and optimization of tumor immunotherapy [5].

However, the high heterogeneity of tumor genomes, the complexity of editing target selection, and the control of off-target risks severely restrict the precise application of gene editing technology. Bioinformatics plays a crucial role in the systematic integration, modeling, prediction, and result analysis of high-dimensional omics data, particularly in target screening, editing protocol design, and efficacy and safety assessment [6]. Based on this, this article systematically reviews the core applications and latest advancements of bioinformatics in the entire process of tumor gene editing, providing a reference for promoting the development of this field towards precision and clinical application.

2. Navigation before editing: target discovery and strategy design

2.1. From databases to potential targets

The act of discovering tumor gene editing targets highly relies on the situation where there are high - quality multi - omics data resources and the process of carrying out systematic bioinformatics analysis. The Cancer Genome Atlas (TCGA) project is one of those tumor databases that are used most widely. It is in the process of integrating genomic data, transcriptomic data, proteomic data, as well as epigenetic data from thousands of samples that cover various cancer types. It systematically describes the process of how the mutation profiles, the copy number variations, and the expression aberrations of tumor - related genes occur, and this lays the foundation for the act of identifying driver genes and key regulatory factors [7]. The Cancer Dependence Map (DepMap) then further combines the process of conducting CRISPR gene perturbation screening, the process of performing drug sensitivity analysis, and the multi - omics features to systematically reveal the situation where tumor cells have a dependence on specific genes. This platform currently covers around 2000 tumor models, more than 30 tumor lineages, and 1150 whole - genome CRISPR screening datasets. It has successfully carried out the act of identifying various tumor vulnerabilities, including the lineage -

DOI: 10.54254/2753-8818/2026.32136

specific dependence, the tumor suppressor gene deletion - related dependence, and the synthetic lethal dependence. And it has propelled targets such as PRMT5 and WRN into the process of clinical research [8].

At the level of conducting data analysis, the analysis of differential expression serves as a frequently employed strategy for the purpose of carrying out initial target screening. In the situation where transcriptional levels between tumor and normal tissues are compared by making use of tools such as Limma and DESeq2, genes that are significantly aberrantly expressed can be rapidly located, which helps to narrow down the scope of potential targets [9]. After that, in the process of conducting pathway enrichment analysis (for example, GO and KEGG enrichment) by using tools like Metascape and DAVID, candidate genes are mapped to specific biological processes and signaling pathways. This allows for the prioritization of the screening of key nodes that are involved in the processes of tumorigenesis and development at the functional level [10]. The act of predicting synthetic lethal relationships offers an important approach for the discovery of targets that have low toxicity and high specificity. By using databases and algorithms such as CERAL and SynLethDB, genes that have synthetic lethal relationships to oncogenes can be systematically identified. These genes are not essential in normal cells, but they become factors that tumor cells are dependent on for survival under specific mutational backgrounds. As a result, the accuracy and safety of target selection are improved [11].

2.2. Designing precise editing tools

After the process of target identification takes place, the design of guide RNA (gRNA) turns out to be a crucial step in the situation where one aims to let the efficient and specific gene editing be achieved. Online tools such as CRISPick are capable of carrying out a systematic prediction of potential gRNA sequences on the basis of the target gene sequence and the recognition characteristics of Cas protein, and they can assess the gene editing efficiency of these sequences through a comprehensive scoring model. High - quality gRNAs usually have a situation where they have a perfect match with the target sequence, an appropriate GC content (which is in the range from 40% to 60%), and a low tendency for the formation of secondary structure. At the same time, they avoid the act of targeting regions that are outside of key functional exons. By doing so, they can improve the success rate of editing as well as the biological relevance.

Off - target effects are among the primary risks that play a role in limiting the clinical application of gene editing. In the process of trying to deal with this situation, bioinformatics tools can carry out the task of predicting off - target sites by means of whole - genome sequence alignment. Cas - OFFinder is capable of searching for genomic regions that have 1 - 3 base mismatches with gRNA sequences. Then, in the situation where it is combined with chromosomal localization and gene function annotation, it can conduct the process of grading off - target risk assessment [12]. BLAST tools are able to quickly perform the action of screening for highly homologous sequences. In order to further reduce the off - target risk, studies usually combine multiple optimization strategies. These strategies include taking the action of selecting gRNAs that have higher specificity scores, employing a dual - gRNA synergistic cleavage strategy, or using high - fidelity Cas protein variants [13]. As time has passed and reached the present day, bioinformatics tools have evolved from simply relying on sequence matching to precise prediction systems. These systems integrate molecular mechanisms and deep learning. The CRISPT model constructs RNA - DNA interaction fingerprints through the process of molecular dynamics simulations. Then, in the situation where it is combined with AI algorithms, it can elevate the accuracy of off - target prediction to new levels. Its developed sgRNA optimization module has successfully achieved the goal of reducing the off - target risk of

DOI: 10.54254/2753-8818/2026.32136

disease - related gene editing, such as that of PCSK9 and BCL11A [14]. The CRISPR - MCA model, which incorporates an ESB data balancing strategy, shows superior prediction performance when compared to traditional methods on complex datasets that contain insertions and deletions. It can achieve the process of precise localization of off - target binding and cleavage sites across the entire genome [15].

By means of the process of applying bioinformatics prediction and experimental validation in a combined way, the situation where the occurrence of non - specific editing events can be minimized can be achieved while the process of ensuring editing efficiency is going on.

3. Decoding after editing: functional screening and efficacy analysis

3.1. Interpreting high-throughput screening experiments

The core of CRISPR whole-genome screening technology lies in using lentiviral vectors to introduce a genome-wide sgRNA library into Cas9-expressing cells. Guided by the sgRNA, specific cleavage of target genes is achieved, and combined with phenotypic screening and high-throughput sequencing analysis, the functional genes influencing key phenotypes such as tumor cell survival, proliferation, or drug resistance are systematically located. Compared to the studies of traditional single - gene, this technology has the ability to carry out an analysis on the function of tens of thousands of genes within the process of a single experiment. By doing so, it offers a highly efficient way for the discovery of tumor drivers as well as therapeutic targets [16]. For instance, in the case of TP53 - mutant acute myeloid leukemia, the XPO7 - NPAT survival axis was successfully identified. And in small cell lung cancer, HDAC3 was discovered to be a radiosensitization target. These two examples both show the application value of this technology in complex tumor models [17].

In the recent few years, the screening technology of single - cell CRISPR has come to be an important direction of development within this field. Technologies like CROP - seq and Perturb - seq directly carry out the coupling of genetic perturbation with the sequencing of single - cell transcriptome. This kind of operation not only conducts the identification of key genes that have an impact on cell survival or proliferation but also systematically conducts the analysis of the remodeling effects that gene perturbation has on the global transcriptional network. In this way, it directly makes the connection between genotype and complex phenotypes such as cell state transitions and differentiation trajectories. In the situation where single - cell CRISPR screening provides its help, researchers are able to analyze the process of how the deletion of specific genes drives T cell exhaustion or promotes macrophage polarization in the process of being in the tumor microenvironment. This provides a new dimension of analysis for the study of the regulatory mechanisms of tumor immunity [18].

As high - throughput screening technologies are undergoing the process of popularization, bioinformatics algorithms have carried out a systematic iteration in the process of analyzing massive biological data. Gradually, they have formed a technology system that takes CRISPR screening, single - cell multi - omics integration, deep learning empowerment, and petabyte - scale genome retrieval as its center. In the field where the analysis of CRISPR screening data is being carried out, the MAGeCK series of algorithms has been continuously upgraded. It has evolved from conducting basic statistical analysis to developing into MAGeCK - VISPR. In this way, it has let an integrated workflow of standardized quality control, multi - condition comparison, and visualization be realized. Moreover, it has introduced maximum likelihood estimation to adapt to complex experimental designs. After that, MAGeCKFlute has further integrated functional enrichment and pathway analysis modules. As a result, it has achieved end - to - end analysis that goes from raw

DOI: 10.54254/2753-8818/2026.32136

data to biological interpretation [19,20]. Gscreend effectively performs the task of correcting for systematic biases that are caused by cellular heterogeneity. It does so by introducing a skewed normal distribution to model the empty distribution and then combining it with slice analysis strategies [21].

Dedicated algorithms have come into existence in the situation where they are needed to address the high - dimensional and sparse characteristics of single - cell CRISPR screening data. cMAGeCK takes the action of extending traditional screening analysis to the single - cell level. It then constructs a systematic framework for sgRNA allocation, then cell phenotype grouping, and then gene effect scoring [22]. Mixscape is able to effectively distinguish between real perturbation effects and intrinsic state differences in heterogeneous cell populations by means of a perturbation response scoring model [23]. SCREE offers a one - stop analysis workflow that supports the integration of RNA, then ATAC, and then proteomics multi - omics. It has the efficient processing capabilities for millions of single - cell data points [24]. At the same time, deep learning technology is further causing paradigm shifts in screening to take place. The self - supervised model CP - DINO, which is based on a contrastive learning strategy, can carry out the prediction of gene function using cell morphology features without the need for manual annotation. This significantly improves the efficiency of unbiased screening [25]. For the retrieval needs of petabyte - scale biological sequences, MetaGraph makes use of annotated de Bruin graphs to construct indexes. This enables it to achieve low - cost, high - efficiency whole - genome retrieval and significantly reduce the storage and computational burdens [26].

Overall, related algorithms are showing a trend where they are going through a process of evolving. That is, in the situation where they are being transformed from statistical models to deep learning, from single - mode learning to multimodal integration, and from efficiency improvement to cost optimization. This kind of evolving process provides strong support for going to deal with the issues of heterogeneity, high noise, as well as high dimensionality that exist in large - scale screening data.

3.2. Evaluating editing efficiency and safety

Accurately carrying out an assessment of the editing efficiency and the off - target risk of the CRISPR - Cas9 system is a crucial pre - condition that is necessary for its clinical translation. Editing efficiency usually means the proportion of cells in which events like insertion, deletion, or substitution take place at the target site. In the meantime, safety assessment places its focus on potential off - target effects that occur across the whole genome. Bioinformatics tools play a central and important role in the process of processing, predicting, and interpreting the editing results, and they offer important evidence that can be used for the optimization of high - precision and low - risk gene editing systems.

Currently, the process of conducting an efficiency assessment that is based on targeted amplicon deep sequencing has turned into a standard method. In the situation where PCR amplification and high - throughput sequencing are carried out on the target site region of the edited cell population, and then combined with the process of bioinformatics comparative analysis, it becomes possible to accurately calculate the occurrence proportion of various editing events. Tools like CRISPResso2 are able to efficiently make a distinction between real editing signals and background noise, and then carry out an analysis of the distribution characteristics of different indel types and allele editing states. This provides support for achieving a deeper understanding of the functional consequences of the editing results [27].

DOI: 10.54254/2753-8818/2026.32136

In the process of conducting off - target assessment, a technical system which combines the process of computational prediction and the process of experimental detection has gradually come into being. The intracellular double - strand break capture technology has the ability to directly locate off - target sites in a real editing environment. For example, Tracking - seq, through the process of tracking single - stranded DNA that is bound to RPA, is suitable for low - cell - input samples. When it is combined with strand - specific alignment and noise reduction algorithms, it can cover multiple tools such as Cas9, base editors (BE), and Prime editors, and this can significantly improve the sensitivity of off - target detection [28]. In the case of dealing with the DSB - free nature of base editors, CHANGE - seq - B optimizes the traditional CHANGE - seq by integrating a dedicated signal recognition algorithm with epigenetic features. This allows for the accurate differentiation between spontaneous mutations and BE - induced editing, and its detection efficiency is significantly higher than that of traditional methods [29].

Furthermore, when it comes to structural variations like chromosomal translocations and large deletions, in the situation where PEM - Q is dealing with such cases, it constructs a unified workflow for the analysis of whole - genome DNA repair products. This workflow has the ability to systematically conduct a process of capturing various types of variants. These variants include small indels, large deletions, vector integration, and chromosomal translocations. The sensitivity and consistency of this workflow are superior to those of traditional methods. This superiority provides a key means to comprehensively conduct an assessment of gene editing safety [30]. In the process of considering computational prediction models, the CCLMoff framework, which is based on deep learning, undertakes a process of integrating large - scale off - target detection data together with RNA language models. By doing so, it makes the highly accurate and generalized prediction of off - target sites possible. This framework maintains excellent performance even when there is a situation where additional epigenetic information is absent [31].

In summary, at present, the process of off - target assessment has developed into a situation where a complete technological chain has been formed. This technological chain consists of experimental detection and then computational prediction and then multi - omics validation. This formation of the technological chain is in a state where it is laying a solid foundation for the safety assessment of basic research on tumor gene editing and also for the clinical translation of such basic research.

4. Conclusion and perspectives: toward an intelligent and personalized future

Bioinformatics has permeated the entire process of tumor gene editing, from target identification to protocol design and experimental validation, becoming a core support for improving research efficiency and accuracy, and driving the formation of a data-driven closed-loop optimization system. In the time to come, the process of the deep integration of multiple disciplines will carry out a further promotion of the transformation of tumor treatment in the direction of intelligence and personalization. To begin with, in the situation where artificial intelligence and gene editing are combined, it will bring about a significant improvement in design efficiency and predictive capabilities. In the process of building deep learning models based on large - scale CRISPR screening and editing data, these models can predict editing efficiency, product type, as well as off - target risks in a more accurate way, and carry out an optimization of gRNA design. This will drive the editing strategies from being experience - driven to being model - driven, and accelerate the iteration and optimization of protocols [32]. Secondly, the integration of multi - omics will turn into an important foundation for precise targeting. By carrying out a fusion of genomic, transcriptomic, epigenomic, and single - cell and spatial multi - omics data, the heterogeneity of the tumor microenvironment can be analyzed in a systematic manner. And then, key regulatory networks that

DOI: 10.54254/2753-8818/2026.32136

drive clonal evolution and immune escape can be identified. In this way, high - resolution maps can be provided for the precise targeting of editing sites [33]. Finally, at the level of clinical translation, integrated bioinformatics platforms will offer support for the rapid analysis of patients' multi - omics data. They will conduct a screening of personalized and actionable targets, and carry out a dynamic assessment of post - treatment clonal changes and immune responses. This will provide a basis for the decision - making in the development of personalized gene - editing treatment plans [34]. Overall, bioinformatics, when combined with artificial intelligence and multi - omics technologies, will keep on driving tumor gene editing towards a treatment that is efficient, safe, and precise.

References

[1] Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clinicians 71, 209–249 (2021).
[2] Pal, M., Das, D. & Pandey, M. Understanding genetic variations associated with familial breast cancer. World J Surg Onc 22, 271 (2024).
[3] Martínez-Jiménez, F. et al. A compendium of mutational cancer driver genes. Nat Rev Cancer 20, 555–572 (2020).
[4] Wong, C. UK first to approve CRISPR treatment for diseases: what you need to know. Nature 623, 676–677 (2023).
[5] Wang, S.-W. et al. Current applications and future perspective of CRISPR/Cas9 gene editing in cancer. Mol Cancer 21, 57 (2022).
[6] Moon, I. et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. Nat Med 29, 2057–2067 (2023).
[7] Zhang, Z. et al. A survey and evaluation of web-based tools/databases for variant analysis of TCGA data. Briefings in Bioinformatics 20, 1524–1541 (2019).
[8] Arafeh, R., Shibue, T., Dempster, J. M., Hahn, W. C. & Vazquez, F. The present and future of the cancer dependency map. Nat Rev Cancer 25, 59–73 (2025).
[9] Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014).
[10] Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523 (2019).
[11] Wang, J. et al. SynLethDB 2.0: A web-based knowledge graph database on synthetic lethality for novel anticancer drug discovery. Database 2022, baac030 (2022).
[12] Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473–1475 (2014).
[13] Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013).
[14] Chen, Q. et al. Genome-wide CRISPR off-target prediction and optimization using RNA-DNA interaction fingerprints. Nat Commun 14, 7521 (2023).
[15] Yang, Y., Zheng, Y., Zou, Q., Li, J. & Feng, H. Overcoming CRISPR-Cas9 off-target prediction hurdles: A novel approach with ESB rebalancing strategy and CRISPR-MCA model. PLoS Comput Biol 20, e1012340 (2024).
[16] Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87 (2014).
[17] Semba, Y., Yamaguchi, T., E Bauer, D. & Ogawa, S. The XPO7-NPAT axis represents key vulnerabilities in TP53-mutated acute myeloid leukemia. Blood 147, 662–674 (2026).
[18] Zhou, P. et al. Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. Nature 624, 154–163 (2023).
[19] Li, W. et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol 16, 281 (2015).
[20] Wang, B. et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nat Protoc 14, 756–780 (2019).
[21] Imkeller, K., Ambrosi, G., Boutros, M. & Huber, W. gscreend: Modelling asymmetric count ratios in CRISPR screens to decrease experiment size and improve phenotype detection. Genome Biol 21, 53 (2020).
[22] Yang, L. et al. scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol 21, 19 (2020).
[23] Yang, B. et al. PerturbDB for unraveling gene functions and regulatory networks. Nucleic Acids Research 53, D1120–D1131 (2025).

DOI: 10.54254/2753-8818/2026.32136

[24] Wei, H., Han, T., Li, T., Wu, Q. & Wang, C. SCREE: A comprehensive pipeline for single-cell multi-modal CRISPR screen data processing and analysis. Briefings in Bioinformatics 24, bbad123 (2023).
[25] Sivanandan, S. et al. A pooled cell painting CRISPR screening platform enables de novo inference of gene function by self-supervised deep learning. Preprint at https: //doi.org/10.1101/2023.08.13.553051 (2023).
[26] Li, W. Petabase-scale sequence search with MetaGraph. Nat Genet 57, 2620–2620 (2025).
[27] Clement, K., Rees, H., C. Canver, M. & M. Gehrke, J. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol 37, 224–226 (2019).
[28] Zhu, M. et al. Tracking-seq reveals the heterogeneity of off-target effects in CRISPR–Cas9-mediated genome editing. Nat Biotechnol 43, 799–810 (2025).
[29] Lazzarotto, C. R. et al. Sensitive and unbiased genome-wide profiling of base-editor-induced off-target activity using CHANGE-seq-BE. Nat Biotechnol https: //doi.org/10.1038/s41587-025-02948-7 (2026) doi: 10.1038/s41587-025-02948-7.
[30] Liu, M. et al. Global detection of DNA repair outcomes induced by CRISPR–Cas9. Nucleic Acids Research 49, 8732–8742 (2021).
[31] Du, W. et al. A versatile CRISPR/Cas9 system off-target prediction tool using language model. Commun Biol 8, 882 (2025).
[32] Wessels, H.-H. et al. Prediction of on-target and off-target activity of CRISPR–Cas13D guide RNAs using deep learning. Nat Biotechnol 42, 628–637 (2024).
[33] Sun, Y. et al. Integrated multi-omics profiling to dissect the spatiotemporal evolution of metastatic hepatocellular carcinoma. Cancer Cell 42, 135-156.e17 (2024).
[34] Launspach, M. et al. Personalized CRISPR knock-in cytokine gene therapy to remodel the tumor microenvironment and enhance CAR T cell therapy in solid tumors. Nat Commun 16, 10987 (2025).