

Facilitation Of Antigen Targeting for Tumor Vaccines Using Artificial Intelligence

Jiaming Cui *

School of Zhengzhuo Middle School, Zhengzhou, Henan, China

* Corresponding Author Email: cuijiaming1117@outlook.com

Abstract. Vaccine, a traditional treatment to fight against infected pathogens, is now applied into the therapy of tumor. For the development of tumor vaccines, the critical challenge is the selection of tumor-specific antigens that could distinguish normal and malignant cells. The human genome is vast. A single tumor may be accompanied by dozens or even hundreds of gene mutations, making it extremely difficult to select the appropriate mutation that corresponds to an ideal targeting antigen. Leveraging artificial intelligence and big data models to accelerate antigen selection is crucial for advancing tumor vaccine research. This article will explore the principles of tumor vaccines and antigen screening, delve into cutting-edge AI-assisted technologies in tumor vaccine development, and discuss how AI addresses technological gaps in this field. It will also provide insights into the potential impact of AI on advancing biological research, particularly in the context of current priorities in tumor clinical treatment.

Keywords: Tumor vaccines; Antigen Selection; Artificial Intelligence; Neoantigen for tumors.

1. Introduction

A tumor refers to a mass of tissue formed by the abnormal proliferation of cells potentially under the influence of carcinogenic factors. Most tumors are caused by genetic mutations that lead to uncontrolled and vigorous growth of cells, which can possess autonomy and continue to increase stably even after the initial carcinogenic factors have been removed. Tumor vaccine is an effective method for treating tumors. They involve introducing tumor antigens into the patient's body through tumor cells, tumor-associated proteins or peptides, or genes expressing tumor antigens to overcome tumor-induced immunosuppression, enhance immunogenicity, and activate the patient's immune system [1][2]. This induces both cellular and humoral immune responses, thereby facilitating controlling or eliminating the tumor. Antigen selection is a key process for tumor vaccine engineering and development.

Currently, the majority of selected antigens are tumor-associated antigens (TAAs), which are abnormally expressed in tumor cells but also present in normal cells. Since TAAs are self-antigens, patients may have central tolerance to these antigens. The activated immune system targeting TAAs might inadvertently damage normal cells. Tumor-specific antigens (TSAs), on the other hand, are expressed only in tumor cells. Vaccines that utilize TSAs have high immunogenicity and can minimally affecting normal cells, thus making TSAs "ideal antigens." However, screening for TSAs is difficult. Key challenges include the low targeting efficiency for T cells, the multi-specificity of T cell receptors (TCRs) for antigen peptides, and weak binding affinity between TCRs and peptide-major histocompatibility complexes (pMHCs). Traditional methods of identifying and screening these antigens are time-consuming, labor-intensive, costly, and can prone to significant errors.

With the advancement of computer technology and artificial intelligence (AI), AI algorithms are being applied more frequently to aid tumor vaccine development in multiple aspects such as neoantigen prediction, nucleic acid modification, sequence optimization, and the screening and optimization of vaccine candidates. Compared to traditional methods for antigen selection, AI has numerous advantages including higher efficiency and accuracy. This article reviews recent methods on leveraging AI models to enhance efficiency and accuracy of the antigen selection process of the development of tumor vaccines.

2. Principles of Antigen Selection

The principles of antigen selection are foundational to the development of effective cancer vaccines. When tumor antigens are introduced into a patient's body, they serve as a beacon for the immune system, prompting it to recognize and attack tumor cells. The goal is to trigger a robust and targeted immune response, specifically designed to identify and eliminate cancerous cells while sparing healthy ones [4].

The ideal antigens for cancer vaccines are those that are universally present across all cancer cells of a particular type but absent in normal, healthy cells. This selectivity ensures that the immune response is focused on destroying the tumor without causing collateral damage to the body's normal tissues, thereby minimizing potential side effects. Tumor-specific antigens (TSAs), which are unique to cancer cells, are particularly valuable because they reduce the risk of an autoimmune response, where the immune system might mistakenly attack normal cells.

The challenge in antigen selection lies in identifying these unique markers. Cancer cells often exhibit a wide variety of mutations, making it difficult to pinpoint which antigens will effectively stimulate the immune system. Advanced techniques, including those powered by artificial intelligence, are increasingly being used to overcome this challenge, allowing researchers to select the most promising antigens for vaccine development.

3. Key Considerations for Antigen Selection

The selection of antigens for tumor vaccines is a critical step that significantly influences the efficacy and safety of the resulting treatment. Several key considerations guide this process, each addressing different aspects of how antigens interact with the immune system and the broader implications for patient outcomes. These considerations include the type of antigen, its stability, immunogenicity, and safety.

3.1. Antigen Type

The type of antigen selected for a tumor vaccine plays a pivotal role in determining the specificity and effectiveness of the immune response.

Tumor-Associated Antigens (TAAs): TAAs are proteins that are overexpressed in tumor cells but are also present, albeit at lower levels, in normal cells. While TAAs are more readily identifiable and have been widely studied, their presence in normal tissues poses a significant risk of autoimmune reactions. This is because the immune system may not only attack the tumor cells but also the normal cells that express these antigens, leading to potentially harmful side effects. For example, certain antigens overexpressed in melanoma or breast cancer might be shared by healthy tissues, increasing the risk of an autoimmune response.

Tumor-Specific Antigens (TSAs): TSAs, also known as neoantigens, are derived from mutations that occur exclusively in tumor cells. These antigens are not found in normal cells, making them ideal targets for tumor vaccines. The immune system can be trained to recognize these antigens as foreign, thereby focusing its attack solely on the tumor cells. This specificity reduces the likelihood of damaging healthy tissues and results in a more targeted immune response. However, identifying TSAs is challenging due to the unique nature of the mutations in each patient's tumor, necessitating advanced techniques such as next-generation sequencing and bioinformatics analysis to accurately detect and select these antigens.

3.2. Antigen Stability

The stability of an antigen is another crucial factor in vaccine design. Stable antigens are those that can persist within cells for extended periods, ensuring continuous expression and presentation to the immune system. This prolonged presence allows for sustained immune stimulation, which is essential for developing a robust and durable immune response. If an antigen degrades too quickly or is not consistently expressed, the immune system may not have sufficient time to mount an effective

response, reducing the vaccine's efficacy. Therefore, selecting antigens with high intracellular stability is key to achieving long-term protection against the tumor.

3.3. Immunogenicity

Immunogenicity refers to the antigen's ability to provoke an immune response [3]. Strongly immunogenic antigens are more effective at activating the immune system, leading to a more powerful and sustained attack on the tumor cells. The level of immunogenicity is influenced by several factors, including the antigen's structure, how it is presented to the immune system, and the presence of adjuvants that enhance the immune response. Selecting antigens with high immunogenicity is essential for developing a vaccine that can effectively stimulate the body's defenses and achieve meaningful clinical outcomes.

3.4. Safety

Safety is the paramount concern in vaccine development. While it is essential to activate the immune system against the tumor, it is equally important to avoid triggering severe side effects. The selected antigens must minimize harm to the patient by reducing the risk of off-target effects, such as autoimmunity or excessive inflammation. This is particularly critical when using TAAs, as their expression in normal tissues can lead to unintended immune attacks on healthy cells. Safety considerations also involve ensuring that the immune response is appropriately regulated to prevent overstimulation, which could cause systemic reactions.

4. The Role of Artificial Intelligence and Computational Algorithms for Antigen Selection

The development of artificial intelligence and big data technologies has provided more sophisticated tools for antigen screening. AI techniques can be employed to address the challenges of traditional methods, improving efficiency in various stages of tumor vaccine development.

4.1. Detection of Neoantigens

A tumor may harbor dozens or even hundreds of gene mutations, making the selection of an appropriate mutation as an antigen extremely challenging. Detecting neoantigens using traditional methods is particularly difficult, but AI technology can quickly and accurately identify tumor-specific targets, thereby screening for ideal neoantigens.

The Spliced Neoantigen Finder (SNAF) algorithm was developed by Professor Nathan Salomonis and Dr. Guangyuan Li at Cincinnati Children's Hospital Medical Center. This model integrates advanced deep learning models (DeepImmuno) and probabilistic algorithm models (BayesTS). The former predicts the immunogenicity of neoantigens, while the latter ranks the tumor specificity of these neoantigens. This approach enables the identification, prioritization, and interpretation of different classes of spliced neoantigens, initiating the SNAF-T workflow (predicting T-cell antigens) and the SNAF-B workflow (predicting B-cell antigens). This comprehensive method not only improves the accuracy of neoantigen identification but also prioritizes potential targets, providing a scientific basis for the development of personalized and shared cancer vaccines. The procedure of the SNAF algorithm is as follows:

- ① Data Input: Collect and preprocess the patient's RNA-seq data.
- ② T-Cell Antigen Prediction: Predict HLA-binding peptides and analyze tumor specificity.
- ③ B-Cell Antigen Prediction: Identify transmembrane proteins, analyze tumor specificity, and validate structures.
- ④ Experimental Validation: Validate using mass spectrometry, immunofluorescence, and functional assays.
- ⑤ Results Output: Rank neoantigens and visualize the results.
- ⑥ Application Translation: Design vaccines and develop immunotherapies.

Through a series of precise computational and experimental steps, the SNAF algorithm efficiently identifies and validates potential neoantigen targets in cancer. These targets can be used not only for designing personalized cancer vaccines but also to provide a scientific foundation for developing broad-spectrum immunotherapies. As technology continues to advance, SNAF is expected to play a significant role in more types of cancer, offering new hope and treatment opportunities for cancer patients.

4.2. Antigen Prioritization and Screening

The application of AI in tumor vaccine antigen screening has made significant progress. Traditional antigen screening methods typically require extensive experimentation and time, whereas AI technology can rapidly and accurately identify antigens with potential anti-tumor activity from vast amounts of biological data through deep learning and big data analysis.

Professor Liu Qi from Tongji University proposed the pTuneos framework, short for prioritizing tumor neoantigens from next-generation sequencing data, designed to effectively predict neoantigens from high-throughput sequencing data and evaluate and rank their true immunogenicity [5]. This framework is divided into several key components (as shown in figure 1):

① Data Preprocessing:

Quality Control: Ensure the quality of sequencing data.

Alignment to Genome: Align exome sequencing and transcriptome sequencing data to a reference genome.

Mutation and Expression Profiling: Calculate gene mutation and expression profiles.

HLA Typing: Determine the patient's HLA type.

② Candidate Neoantigen Identification:

Based on Mutation and Expression: Identify potential neoantigens using gene mutation and expression profile information.

HLA Compatibility: Predict which neoantigens are likely to be presented by the patient's HLA molecules, considering HLA typing.

③ Machine Learning Model Training:

Immunoactivity Data: Train machine learning models using immunoactivity data obtained from in vitro experiments on neoantigens.

Handling Data Imbalance: Address the issue of imbalance between positive and negative samples.

MHC Complex and TCR Recognition: The model can screen for neoantigens that can be presented by MHC complexes and recognized by T-cell receptors.

④ Neoantigen Ranking Optimization:

Immunogenicity Scoring Function: Construct a scoring function based on various features involved in the generation, presentation, and recognition of neoantigens.

Ranking Optimization: Use the scoring function to optimize the ranking of candidate neoantigens, evaluating their true immunogenicity.

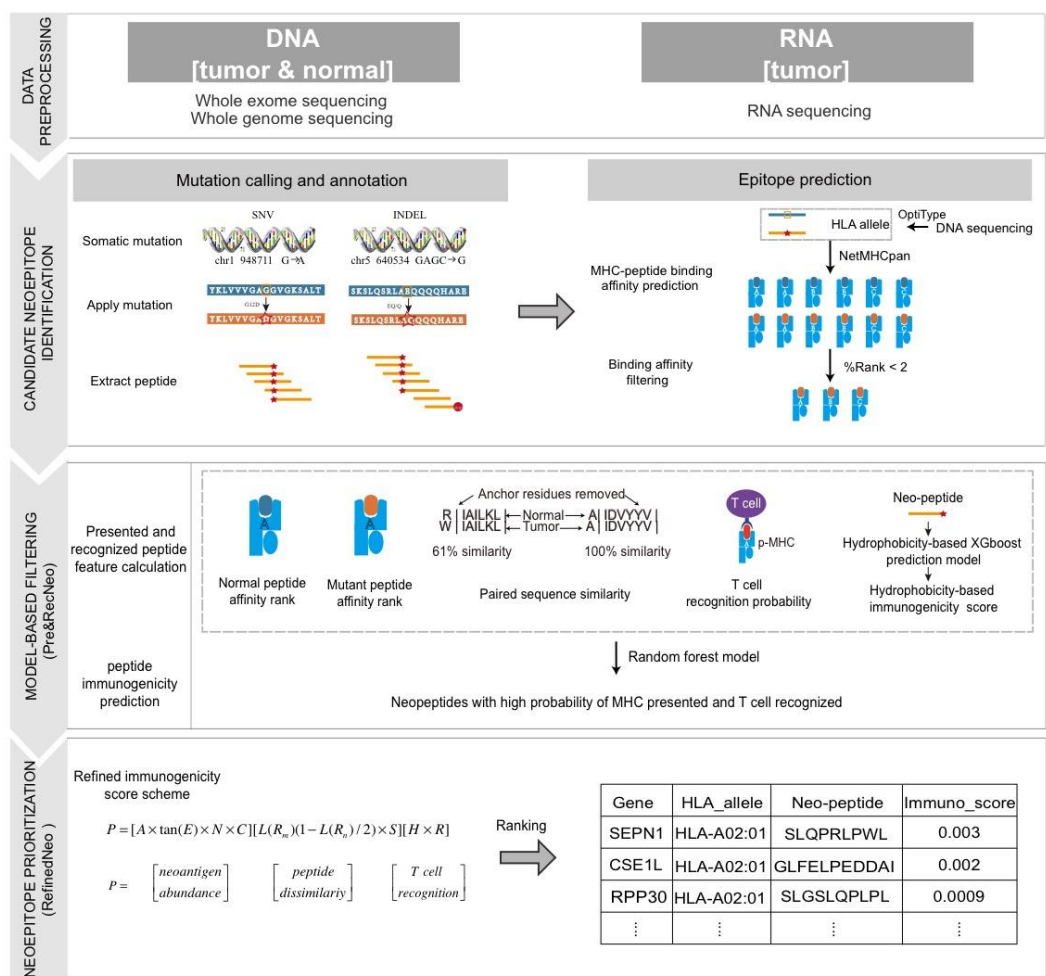


Fig. 1 pTuneos Framework

This framework leverages machine learning and AI to rapidly screen a large number of candidates neoantigens, allowing for a more accurate assessment of their true immunogenicity. This aids researchers in identifying the most promising antigens for the development of cancer vaccines.

4.3. Analysis and Prediction of Viral Genomes

Bioinformaticians often need to quickly and accurately identify viral sequences from vast amounts of genomic data. Traditional methods of viral detection may be inefficient and insufficiently sensitive to new or unknown viruses. Therefore, developing new machine-learning methods to improve the accuracy and efficiency of virus detection is crucial. By applying deep learning and pattern recognition to extensive viral genome data, AI can accurately predict viral mutations and evolutionary trends. This capability helps scientists rapidly identify key protein targets of viruses, facilitating the swift development of related vaccines.

For example, a research team at Beihang University developed a new machine learning ensemble method called VirusImmu, designed to predict the immunogenicity of viral antigens. This method has demonstrated efficiency in analyzing large datasets, such as viral proteomes, and significantly reduces the experimental costs of developing new vaccine candidates. It was successfully applied in identifying linear B-cell epitopes of the African swine fever virus (ASFV), testing the affinity of the antigen B-cell epitope to the antibody using various methods and validated through in vitro indirect ELISA. (Figure 2)

① Epitope Prediction: Researchers used VirusImmu to predict several B-cell linear epitope candidates from the pp220 polyprotein of ASFV.

② Immunogenicity Assessment: VirusImmu predicted that 8 out of 12 epitopes had antigenicity. Indirect ELISA confirmed that 7 of these linear B-cell epitopes specifically and dose-dependently reacted with ASFV serum IgG antibodies.

③ Vaccine Development Support: The predictions from VirusImmu assist in identifying epitopes that can induce strong immune responses in pigs, providing a valuable tool for vaccine development.

④ Linear B-Cell Epitope Identification: Identified linear B-cell epitopes against ASFV were validated through in vitro indirect ELISA in the laboratory.

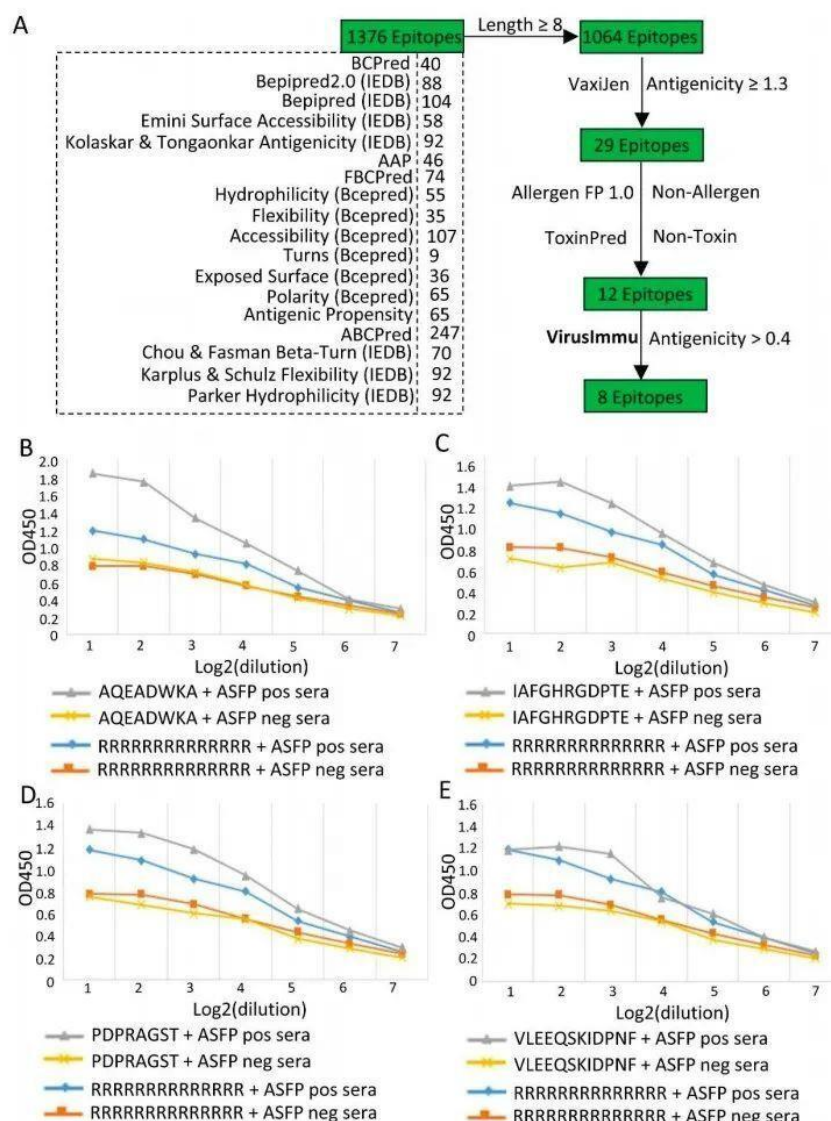


Fig. 2 Affinity of Antigen's B-cell Epitope and Antibodies

VirusImmu, as a novel machine learning ensemble method, shows great potential in predicting the immunogenicity of viral protein fragments, helping accelerate the vaccine development process.

4.4. Optimizing Clinical Trial Design for Vaccines

AI can assist scientists in predicting and evaluating the reactions and effectiveness of vaccines in the human body by simulating large-scale experimental data. This allows for the early identification of potential safety issues and side effects, enabling the optimization of trial designs.

For example, Data Mining for Trial Feasibility is a method that uses data mining techniques to evaluate and optimize clinical trial designs. This approach helps researchers assess the feasibility,

risks, and potential success probability of trials before they begin. In clinical trial feasibility analysis, data mining can:

- ① **Historical Data Analysis:** Analyze past clinical trial data to identify factors contributing to success or failure, providing references for new trial designs.
- ② **Patient Cohort Identification:** Identify patient groups meeting specific criteria by mining electronic health records (EHR) and other medical databases, predicting recruitment difficulty.
- ③ **Risk Assessment:** Identify potential risk factors and adverse events to help researchers evaluate trial safety.
- ④ **Resource Optimization:** Analyze trial execution across different locations, times, and resource allocations to optimize resource distribution.
- ⑤ **Predictive Model Construction:** Build predictive models to estimate trial success rates, costs, and timelines.

The application of data mining in clinical trial feasibility analysis optimizes the design of cancer vaccine clinical trials, reduces potential risks, and provides scientific support for trial management and decision-making, ultimately improving trial success rates. As technology advances and its application scenarios expand, data mining will play an increasingly important role in the clinical trial field.

5. Conclusion

In conclusion, the integration of artificial intelligence (AI) in tumor vaccine development is an exciting and rapidly advancing field with immense potential to revolutionize cancer treatment. AI has already begun to make significant strides in several key areas of tumor vaccine research, including neoantigen prediction, antigen prioritization and screening, viral genome analysis, B-cell epitope discovery, and the optimization of clinical trial design. These advancements are paving the way for more personalized and effective cancer immunotherapies, positioning AI as a critical driving force behind the next generation of cancer treatments.

One of the most promising applications of AI in tumor vaccine development is neoantigen prediction. Neoantigens, which are unique to tumor cells, represent ideal targets for cancer vaccines because they are not present in normal cells, thus reducing the risk of autoimmunity. However, identifying these neoantigens is a complex task due to the vast number of genetic mutations in tumors. AI algorithms, particularly those utilizing deep learning and machine learning techniques, can sift through massive amounts of genetic data to accurately predict which mutations will produce immunogenic neoantigens. This capability not only accelerates the development of personalized cancer vaccines but also enhances their efficacy by targeting the most relevant antigens.

In addition to neoantigen prediction, AI is playing a pivotal role in antigen prioritization and screening. The process of selecting the most promising antigens for vaccine development has traditionally been time-consuming and resource-intensive, often requiring extensive experimental validation. AI models, however, can analyze and prioritize potential antigens based on various factors, such as immunogenicity, stability, and safety, significantly streamlining the development pipeline. By doing so, AI not only reduces the time and cost associated with vaccine development but also increases the likelihood of clinical success.

Furthermore, AI's application in viral genome analysis and prediction is crucial for developing vaccines against virus-associated cancers and emerging viral threats. AI tools can rapidly analyze viral genetic data to identify mutations, predict evolutionary trends, and pinpoint key protein targets for vaccine development. This capability is particularly valuable in responding to rapidly evolving viruses, where timely vaccine development is critical for effective public health interventions.

B-cell epitope discovery is another area where AI is making significant contributions. B-cell epitopes are regions of an antigen that are recognized by the immune system's B-cells, triggering an antibody response. Identifying these epitopes is essential for designing effective vaccines. AI-driven

methods can predict B-cell epitopes with high accuracy, facilitating the design of vaccines that elicit strong and specific immune responses.

Moreover, AI is transforming the way clinical trials are designed and conducted. By simulating large-scale experimental data, AI can help researchers predict how vaccines will perform in human populations, identify potential safety issues, and optimize trial designs to increase their chances of success. This application not only enhances the efficiency of clinical trials but also ensures that they are more ethically and economically viable.

Despite these advances, the application of AI in tumor vaccine development is not without challenges. Developing robust AI models requires vast amounts of high-quality data, including detailed information on protein structures and amino acid sequences. However, obtaining and sharing such data is often difficult and expensive, particularly in the highly competitive pharmaceutical industry. Additionally, current AI models, while fast and accurate, may struggle with standardization and dependence on specific data sets, potentially limiting their broader applicability.

Looking ahead, it is essential to address these challenges to fully realize the potential of AI in tumor vaccine development. As AI technologies continue to evolve, we anticipate the emergence of more sophisticated and versatile AI-based tools that can overcome current limitations. These tools will likely lead to the development of more personalized and effective cancer vaccines, tailored to the unique genetic profiles of individual patients, thereby offering new hope and improved outcomes for cancer patients worldwide. The future of cancer immunotherapy is bright, and AI is poised to play a central role in shaping it.

References

- [1] Wu Y, Zhuang J, Qu Z, et al. Advances in immunotyping of colorectal cancer [J]. *Frontiers in Immunology*, 2023,14.
- [2] Yu L, Zeng X, Hu X, et al. Advances and challenges in clinical applications of tumor cell-derived extracellular vesicles [J]. *Colloids and Surfaces B: Biointerfaces*, 2024,234:113704.
- [3] Bug D. Natural immunogenic I-Ab-restricted antigen human immune system. *Immunol Case Rep*. 2023;6(4):162
- [4] Li J, Liu F, Chen Y, et al. Dynamic Analysis of a Model on Tumor-Immune System with Regulation of PD-1/PD-L1 and Stimulation Delay of Tumor Antigen [J]. *Qualitative Theory of Dynamical Systems*, 2022,21(3):90.
- [5] Pang Z, Lu M, Zhang Y, et al. Neoantigen-targeted TCR-engineered T cell immunotherapy: current advances and challenges [J]. *Biomarker Research*, 2023,11(1):104.