ESA

EXPLORING SCIENCE ACADEMIC CONFERENCE SERIES

PROCEEDINGS OF THE 2026 INTERNATIONAL CONFERENCE ON

Life Sciences and Health Engineering

IC-LSHE 2026

International Conference on Life Sciences and Health Engineering

Vol. 12 (2026)

EDITOR-IN-CHIEF

Prof. Svetlana V. Androsova

Vol. 12 · 2026

E-ISSN 3105-0522

P-ISSN 3105-0514

PROCEEDINGS

PROCEEDINGS OF THE 2026 INTERNATIONAL CONFERENCE ON LIFE SCIENCES AND HEALTH ENGINEERING (IC-LSHE 2026)

VOL. 12

Editor-in-Chief

Prof. Svetlana V. Androsova

IC-LSHE Conference Organizing Committee. “Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)”. *Exploring Science Academic Conference Series*, vol. 12, Feb. 2026, https://doi.org/10.70267/qaxwxh71

VOL. 12 (2026): PROCEEDINGS OF THE 2026 INTERNATIONAL CONFERENCE ON LIFE SCIENCES AND HEALTH ENGINEERING (IC-LSHE 2026)

February 26-28, 2026, Oxford, UK
Exploring Science Academic Conference Series
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI Prefix: 10.70267 Email: secretary@zeupconference.com

PUBLISHED BY: ZEUS PRESS

Address: Flat/Rm 91_3 9/F Chinachem Golden Plaza No.77 Mody Road Tsim Sha Tsui KL, Hong Kong

Individual paper REPRINTS may be ordered at: sherlene@zeupconference.com

Preface

It is with great pleasure that we present the proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026). The central theme of IC-LSHE 2026, "Converging Technologies for Transformative Healthcare," reflects our commitment to fostering interdisciplinary dialogue at the intersection of life sciences, medicine, engineering, and data science. We believe that the most significant advancements in healthcare emerge when traditional disciplinary boundaries dissolve, and we are proud to have provided a global platform for this important exchange.

A total of 68 submissions were received from authors across multiple countries and institutions. Each submission underwent a rigorous double-blind peer review process, where each paper was evaluated by independent experts based on criteria including originality, significance, methodology, clarity, and relevance to the conference themes. Following this comprehensive evaluation process, 17 papers were selected for inclusion in these proceedings.

We extend our sincere gratitude to all reviewers for their dedication and expertise in maintaining the high standards of this publication, and to all authors for contributing their valuable research to this conference. We hope these proceedings will serve as a catalyst for further collaboration and innovation within the global life sciences and health engineering community.

We look forward to the continued growth and impact of IC-LSHE in shaping the future of healthcare.

Prof. Svetlana V. Androsova

Editor-in-Chief

IC-LSHE 2026

Contents

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260101

Predictive and Analytical Methods for Determining Viral Variation: Evolution and Frontier Breakthroughs

Mingshuo Zhang*

Department of Biotechnology, Yanbian University, Jilin, China

**Corresponding author: Mingshuo Zhang*

Abstract

The continuous mutation of viruses poses a persistent and severe challenge to global public health security, disease prevention and control, and biomedical research and development. The accurate prediction and in-depth analysis of viral mutation trends and their potential impacts are of critical strategic importance for constructing effective epidemic early warning systems and guiding the targeted development of vaccines and drugs. In recent years, the rapid development of artificial intelligence and high-throughput sequencing technologies has significantly enhanced our ability to monitor and infer viral evolutionary paths. In particular, the application of machine learning models in predicting mutation hotspots and functional impacts has opened new avenues for research. Moreover, interdisciplinary approaches combining genomics, structural biology, and computational modeling provide deeper insights into the mechanisms driving viral evolution. This paper aims to systematically review the evolution of methods for viral mutation prediction and analysis, delve into frontier technological breakthroughs in the field, objectively analyze the advantages and limitations of existing approaches, and provide an outlook on future directions. Furthermore, the integration of real-time surveillance data with predictive models is emphasized as a key factor in improving the timeliness and accuracy of mutation alerts. The goal is to offer a comprehensive and valuable reference for scientific research and practical applications in related fields, thereby assisting human society in better responding to the complex problems arising from viral variation.

Keywords

sequence alignment, phylogenetic tree, statistical model

1. Introduction

Since the global pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), although its severity and mortality rates have declined in some regions, the increasing emergence of new variants has led to frequent “breakthrough infections”-the phenomenon of reinfection in vaccinated or previously infected individuals. Rapid mutation of the virus not only presents significant obstacles to existing vaccines and therapeutic interventions but also markedly increases the potential risk of future pandemics. Consequently, the forward-looking prediction of viral mutations has become exceptionally important.

However, existing models for predicting SARS-CoV-2 mutations still face numerous challenges. A core difficulty lies in effectively integrating the regularity and inherent stochasticity of viral mutations while satisfying the real-world demands for minimal data and rapid response. From an evolutionary biology perspective, viral evolution is often characterized by “few-site mutations” and “rare beneficial mutations.” This implies that most mutations may be neutral or even deleterious, with only a very small fraction conferring advantages in aspects such as transmissibility or immune evasion, thereby prevailing through natural selection. The inability to accurately anticipate these critical evolutionary directions could result in greater harm to society.

To increase the accuracy of viral mutation prediction and promote the development of new models and methodologies, this paper consolidates multiple research findings to systematically summarize and review the current state of research. By synthesizing the experiences and lessons from existing methods, this paper hopes to provide clear guidance and a solid foundation for subsequent studies, ultimately contributing to the resolution of the global challenge of viral mutation. In recent years, data-driven approaches, particularly those combining large-scale genomic sequencing data with complex computational models such as machine learning, have demonstrated immense potential. Such methods can continuously monitor and analyze viral genomic sequences to capture early signals of potential high-risk variants, thereby securing valuable time for public health responses. In the future, the integration of multidimensional data and the development of more advanced artificial intelligence models will be crucial directions for the field.

2. Evolution of Traditional Methods

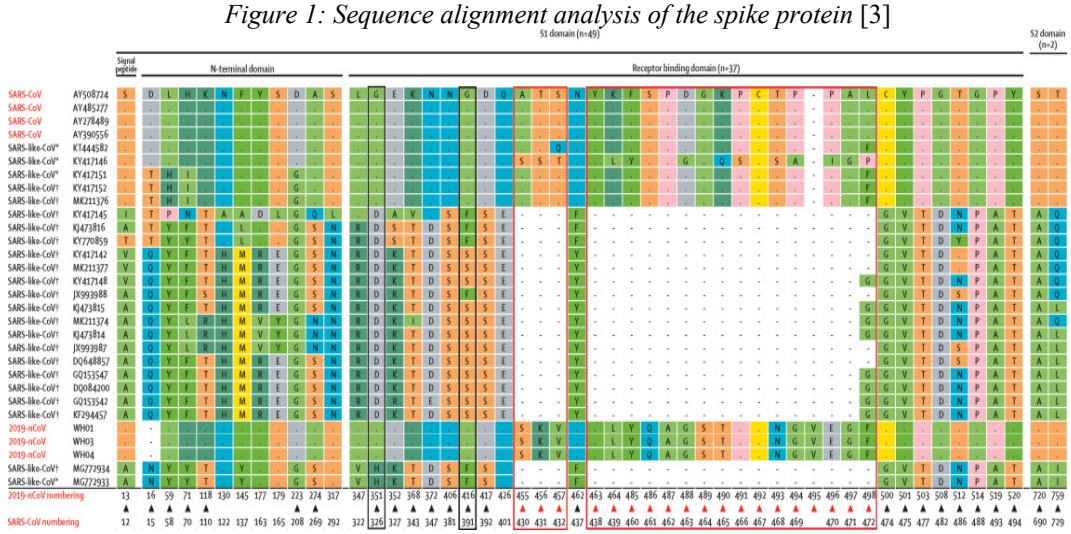
2.1 Analysis Methods Based on Sequence Alignment

One of the most widely used methods in viral analysis is the Basic Local Alignment Search Tool (BLAST). BLAST actually uses a local alignment algorithm that finds the matching segments with the highest similarity between two sequences. Its simple working flow includes the following steps.

It is used to discover the new coronavirus. In the whole process, researchers first acquire the complete genomic sequence of SARS-CoV-2 and then submit it as a query on the online website of the National Center for Biotechnology Information (NCBI) through the online BLAST tool. After that, the system will complete the following steps of algorithms to quickly identify similar sequences in the database. Finally, similar sequence alignment results can be obtained. The result will mark the similarity score, query coverage, E value (expected value), percent identity, etc., to represent the statistical similarity between them [1].

As shown in Figure 1, researchers can also use BLAST to perform comparative genomic analysis of coronavirus disease 2019 (COVID-19). Specifically, the ribonucleic acid genome of SARS-CoV-2 was compared with the RNA genomes of other human coronaviruses. Through the comparison of BLAST results, this study identified the highly conserved genomic regions of different coronaviruses and predicted the degree of viral immune escape [2].

Figure 1: Sequence alignment analysis of the spike protein [3]



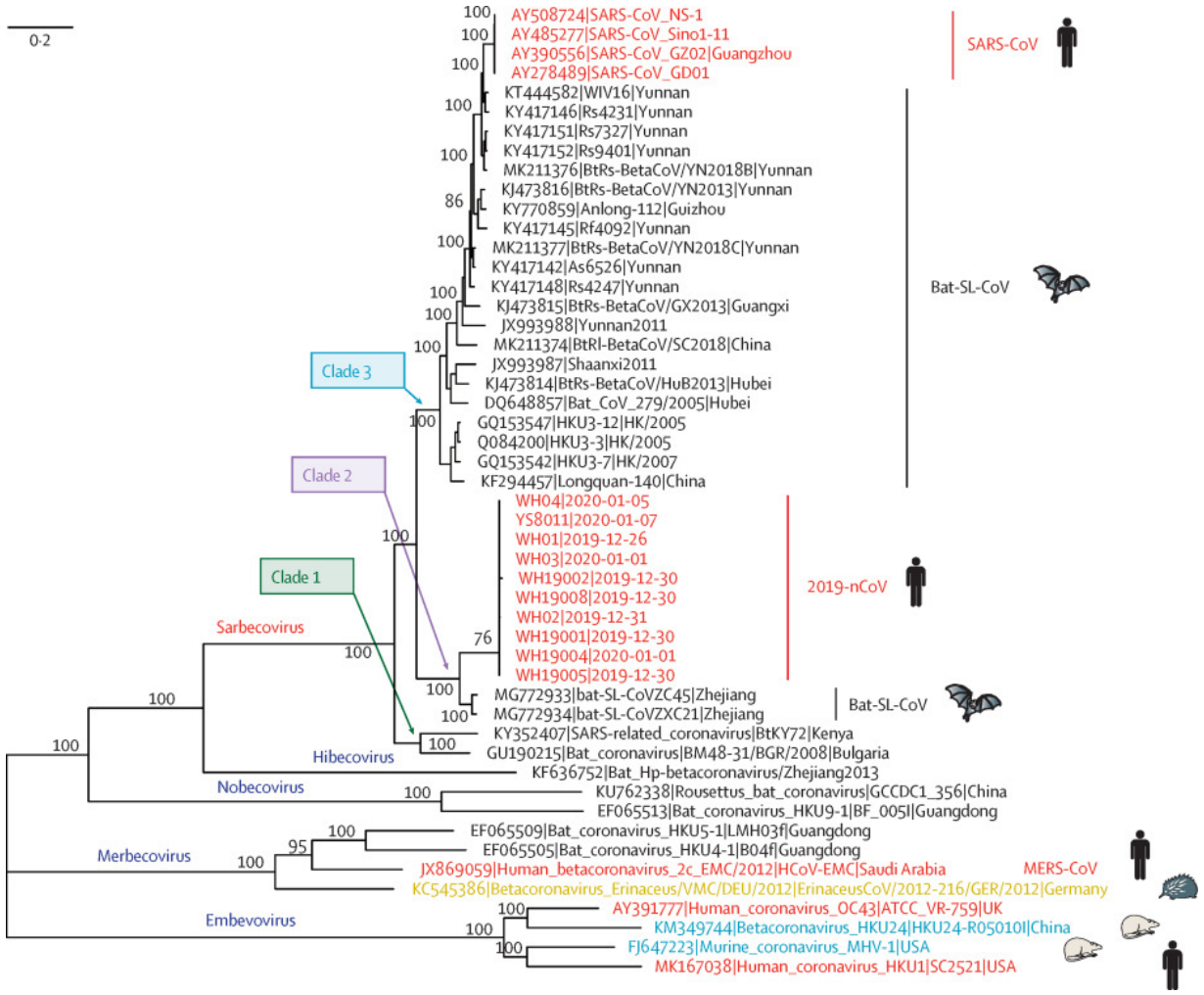
2.2 Phylogenetic Analysis

Phylogenetic analysis is a method used to identify the evolutionary path of a virus and predict its future behavior by visualizing the evolutionary relationships among different species (or other types of populations). A phylogenetic tree is a diagram showing the evolutionary relationships among various groups of organisms-be they species, subspecies, or even genes. The method principle is that the more similar the genes of the virus are, the closer they will be on the tree and the more likely they are to have diverged from a more recent common ancestor. There are generally three main steps involved in the construction of phylogenetic trees: sample collection and sequencing, multiple sequence alignment and tree inference.

This approach was applied in early studies of SARS-CoV-2 in Iran to determine the origin of this virus and the pathway of transmission, identification and typing of variant sites and inference of evolutionary relationships and divergence times. In this study, the authors constructed a phylogenetic tree on the basis of alignment of viral genomes from Iranian samples to a set of global reference genomes (e.g., all downloaded reference genomes from the GISAID database). Multiple nucleotide variant sites were subsequently obtained by aligning these sample sequences to a reference genome (e.g., Wuhan-Hu-1), as shown in Figure 2.

The topology of the obtained tree clearly shows that viral strains with specific mutations are clustered in one another and thus generate different evolutionary clades or lineages. Phylogenetic analysis can be applied to the temporal data of sample collection, and time-scaled Bayesian phylogenetic analysis can be used to reconstruct the divergence times of different lineages of viruses and the age of time to the most recent common ancestor (TMRCA) to understand the virus's evolutionary history and the transmission timeline of this virus in the human population [4].

Figure 2: Analysis of the phylogenetic tree of coronaviruses [3]



2.3 Statistical Models

The prediction and analysis of virus changes are not just hypothetical. They can be exact matches. Two methods are called maximum likelihood and Bayesian inference.

During the COVID-19 pandemic, scientists have studied virus family trees. This was very important. It helps track the virus in real time. It also helps predict dangerous new variants several weeks before normal health checks find them. By looking at the virus's family tree, researchers can see how the virus moves. They can also see how it changes from person to person. This gives real clues. These clues can help guide public health choices.

CoVerage is an automated tool. It was made by the Helmholtz Centre for Infection Research in Germany. It keeps watching how the SARS-CoV-2 virus has changed over time. It learns from how different versions of the virus spread among people. The system uses a math method named Bayesian inference. This method turns growth patterns into numbers. These numbers show how the virus spreads. In simple terms, raw gene data are used to obtain clear results. Public health workers can use these results directly [5].

3. Frontier Technology Breakthroughs

3.1 Multi-Omics Fusion Analysis

This method mixes different kinds of data. It uses data from genomics, transcriptomics, proteomics, and other similar fields. The goal is to learn how mutations truly work at a tiny level. For example, scientists can look at changes in the spike protein of a virus. They can also see how these changes affect RNA messages. Then, they run experiments to determine how tightly the virus binds. Together, this allows them to see more than just one type of data. However, this approach is not perfect. It's like trying to combine many different puzzles at once. Additionally, you need very powerful computers to make it work. In addition, sometimes, the results can surprise you.

3.2 Machine Learning and Deep Learning

Machine learning models, such as random forest and support vector machines (SVMs), have been used to forecast the antigenic changes of influenza viruses, but the performance of these models relies largely on feature engineering. On the other hand, deep neural networks (DNNs), long short-term memory (LSTM) networks, transformers and other deep learning approaches are capable of capturing complicated features automatically from raw data and have been applied to predict conformational changes in the SARS-CoV-2 S protein, for example. However, these models usually exhibit "black-box" characteristics and require large amounts of training datasets.

In a recent study entitled "Quantitatively characterizing the host adaptation process of SARS-CoV-2 variants through a deep learning-based prediction model" published in Cell Discovery, a team from the Chinese People's Liberation Army Center for Disease Control and Prevention, the Chinese Academy of Medical Sciences, Southeast University, and other institutions developed a model labeled ARNLE that can quantify the host adaptation process of SARS-CoV-2 variants and can accurately classify the viral host with high precision [6].

3.3 Molecular Dynamics (MD) Simulation

Molecular dynamics simulations are used to simulate the structural dynamics of viruses at the atomic level after mutation. The application is the study of mutations in the spike (S) protein of SARS-CoV-2 and how they affect the binding of antibodies (microlevel functional effects).

In 2022, scientists in Turkey employed computational methods such as molecular dynamics (MD) simulations to study 13 new SARS-CoV-2 mutations discovered in Turkey. Researchers have successfully characterized and studied the effects of newly discovered mutations on protein dynamics. This study offers an important theoretical basis for understanding the molecular mechanisms of viral variation and demonstrates the strong potential of MD simulations for rapidly assessing the potential effects of newly emerging mutations.

Effect: This method is mainly used to study the effects of new mutations on the structure, stability, and binding affinity of viral proteins to target host cell surface proteins such as angiotensin-converting enzyme 2 [7].

4. Future Directions

4.1 Fusion of Interdisciplinary Methods

This idea is about bringing together different areas, such as biology, computer science, mathematics, and physics. It uses the best parts of each method. For example, deep learning is good at finding features, and dynamic simulations can model objects at the molecular level. Combining these factors helps make better predictions for virus mutations. Mathematics can also help improve algorithms for integrating multiomics data. This makes data usage more efficient.

4.2 Real-time Dynamic Monitoring and Prediction Systems

Leveraging the power of rapid genetic sequencing, large-scale data systems, and machine learning, this study proposes a framework capable of tracking and forecasting viral mutations in real time. Such a system enables continuous collection and interpretation of global data on viral evolution, improving our capacity to recognize emerging patterns in how viruses change. Crucially, it supports public health decision-making by delivering timely, science-backed recommendations for outbreak response-for example, issuing early warnings when new variants emerge that demonstrate heightened transmissibility or immune evasion.

At the University of Nevada, Las Vegas, a research team developed an innovative tool named independent component analysis of variants, which integrates multifaceted data to improve the detection and surveillance of SARS-CoV-2 variants. Their approach involves regular wastewater sampling and analysis across diverse urban and rural settings. This offers a unique window into how the virus evolves within different communities. By applying independent component analysis (ICA), a computational method designed to separate blended signals, researchers can isolate and track specific mutation signatures-even within highly complex wastewater samples containing mixed variant populations. Between late 2021 and 2023, this system consistently identified major variants, including Delta, Omicron, and several recombinant forms [8].

4.3 Deep Fusion of Multiomics and AI

The continuous mutation of viruses poses a persistent and severe challenge to global public health security, disease prevention and control, and biomedical research and development. The accurate prediction and in-depth analysis of viral mutation trends and their potential impacts are of critical strategic importance for constructing effective epidemic early warning systems and guiding the targeted development of vaccines and drugs. This paper aims to systematically review the evolution of methods for viral mutation prediction and analysis, delve into frontier technological breakthroughs in the field, objectively analyze the advantages and limitations of existing approaches, and provide an outlook on future directions. The goal is to offer a comprehensive and valuable reference for scientific research and practical applications in related fields, thereby assisting human society in better responding to the complex problems arising from viral variation.

Since the global pandemic of SARS-CoV-2, although its severity and mortality rates have declined in some regions, the incessant emergence of new variants has led to frequent “breakthrough infections”-the phenomenon of reinfection in vaccinated or previously infected individuals. Rapid mutation of the virus not only presents significant obstacles to existing vaccines and therapeutic interventions but also markedly increases the potential risk of future pandemics. Consequently, the forward-looking prediction of viral mutations has become exceptionally important.

However, existing models for predicting SARS-CoV-2 mutations still face numerous challenges. A core difficulty lies in effectively integrating the regularity and inherent stochasticity of viral mutations while satisfying the real-world demands for minimal data and rapid response. From an evolutionary biology perspective, viral evolution is often characterized by “few-site mutations” and “rare beneficial mutations.” This implies that most mutations may be neutral or even deleterious, with only a very small fraction conferring advantages in aspects such as transmissibility or immune evasion, thereby prevailing through natural selection.

The inability to accurately anticipate these critical evolutionary directions could result in greater harm to society.

To increase the accuracy of viral mutation prediction and promote the development of new models and methodologies, this paper consolidates multiple research findings to systematically summarize and review the current state of research. By synthesizing the experiences and lessons from existing methods, this article hopes to provide clear guidance and a solid foundation for subsequent studies, ultimately contributing to the resolution of the global challenge of viral mutation. In recent years, data-driven approaches, particularly those combining large-scale genomic sequencing data with complex computational models such as machine learning, have demonstrated immense potential. Such methods can continuously monitor and analyze viral genomic sequences to capture early signals of potential high-risk variants, thereby securing valuable time for public health responses. In the future, the integration of multidimensional data and the development of more advanced artificial intelligence models will be crucial directions for the field.

5. Conclusion

This paper comprehensively surveys the evolution of predictive and analytical approaches for viral variation, ranging from conventional techniques such as sequence alignment, phylogenetic analysis, and statistical modeling to recent breakthroughs, including multiomics integration, machine and deep learning, and molecular dynamics simulations. This paper outlines the core principles, applications, strengths, and constraints of each method, offering a holistic perspective on the current state of viral mutation research.

The ongoing emergence of novel viral variants, such as SARS-CoV-2, highlights the urgent need for accurate and timely prediction of viral evolution. Although traditional approaches have established an essential foundation, the sheer complexity and speed of viral mutation demand more advanced tools. Emerging technologies-especially those combining large-scale genomic data with sophisticated computational models-provide unmatched potential for detecting subtle evolutionary patterns and anticipating the rise of high-risk variants.

Moving forward, the field is set to progress through the integration of interdisciplinary strategies, the creation of real-time dynamic prediction systems, and deeper merging of multiomics data with artificial intelligence. These synergistic approaches will refine our molecular-level insight into viral adaptation, increase the accuracy of forecasting models, and yield practical guidance for public health decision-making. Ultimately, enhancing our ability to predict and analyze viral changes will strengthen global health security, speed up the development of precise vaccines and antivirals, and improve our capacity to curb the effects of future pandemics. Such scientific advancements are indispensable for humanity's proactive stance against the enduring threat of viral evolution.

References

[1] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. BLAST+: architecture and applications. BMC Bioinformatics. 2009, 10(1), p. 421. https://doi.org/10.1186/1471-2105-10-421.

[2] Hussein, M., Andrade dos Ramos, Z., Berkhout, B. and Herrera-Carrillo, E. In Silico Prediction and Selection of Target Sequences in the SARS-CoV-2 RNA Genome for an Antiviral Attack. Viruses. 2022, 14(2), p. 385. https://doi.org/10.3390/v14020385.

[3] Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020, 395(10224), pp. 565-574. https://doi.org/10.1016/S0140-6736(20)30251-8.

[4] d'Alessandro, M., Bergantini, L., Cameli, P., Curatola, G., Remediani, L., Sestini, P. and Bargagli, E. Peripheral biomarkers' panel for severe COVID-19 patients. Journal of Medical Virology. 2021, 93(3), pp. 1230-1232. https://doi.org/10.1002/jmv.26577.

[5] Norwood, K., Deng, Z.-L., Reimering, S., Robertson, G., Foroughmand-Araabi, M.-H., Goliaei, S., Hölzer, M., Klawonn, F. and McHardy, A. C. In silico genomic surveillance by CoVerage predicts and characterizes SARS-CoV-2 variants of interest. Nature Communications. 2025, 16(1), p. 6281. https://doi.org/10.1038/s41467-025-60231-4.

[6] Li, C., Chen, L. and Lan, T. Artificial intelligence (AI) reveals the pandemic potential and host adaptation of SARS-CoV-2 variants. Cell. 2024, 2(15), pp. 1152–1166.

[7] Unlu, S., Uskudar-Guclu, A. and Cela, I. The impacts of 13 novel mutations of SARS-CoV-2 on protein dynamics: In silico analysis from Turkey. Human Gene. 2022, 33, p. 201040. https://doi.org/https://doi.org/10.1016/j.humgen.2022.201040.

[8] Zhuang, X., Vo, V., Moshi, M. A., Dhede, K., Ghani, N., Akbar, S., Chang, C.-L., Young, A. K., Buttery, E., Bendik, W., et al. Early detection of emerging SARS-CoV-2 Variants from wastewater through genome sequencing and machine learning. Nature Communications. 2025, 16(1), p. 6272. https://doi.org/10.1038/s41467-025-61280-5.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260102

Advances in the Influence of Wnt on Mesenchymal Stem Cell Differentiation

Jiaying Sun*

School of Pharmacy, East China University of Science and Technology, Shanghai, 201200, China

**Corresponding author: Jiaying Sun*

Abstract

Mesenchymal stem cells (MSCs) possess self-renewal ability and high differentiation potential. It can be used to construct in vitro models to assist with pathological and pharmacological research, as well as applications in tissue engineering. The core process in the osteogenic differentiation of MSCs is osteogenesis, which is the process of bone tissue formation primarily carried out by osteoblasts. Chondrocytes and skeletal muscle cells are also involved in the process of bone injury repair. As a group of highly conserved signaling proteins, Wnt family proteins affect the direction of MSC differentiation through autocrine or paracrine mechanisms. This review summarizes the influence of Wnt in differentiation process, as well as the relevant factors in the Wnt signaling pathway for directed differentiation during cell culture, aims to provide new solutions and guidance for precise treatment.

Keywords

Wnt, mesenchymal stem cells, directed differentiation

1. Introduction

Growth factors are a class of polypeptide or protein biological signaling molecules that can regulate cell growth, division, differentiation and metabolism, utilized to induce the differentiation of mesenchymal stem cells (MSCs). Osteogenesis-related growth factors include VEGF, BMP and Wnt family proteins.

The osteogenic differentiation of MSCs holds great potential for applications in the medical field. Statistical data show that from 1990 to 2021, the incidence and prevalence of spinal fractures in China have continued to rise, making fracture treatment and related rehabilitation increasingly important[1]. Currently, common treatment such as internal fixation screws often pose difficulties in removal after surgery and may lead to complications such as bone necrosis. In addition, after fracture surgery, limb movement is restricted and the risk of muscle atrophy increases, which in turn raises the risk of joint wear and reinjury. Using organoid technology to culture and differentiate MSCs in vitro with osteogenic factors can effectively shorten the rehabilitation period and reduce the need for secondary surgeries, offering a new strategy for fracture treatment and bone defects[2]. Besides, the osteogenic differentiation of MSCs in vitro also holds significant scientific value for the study of calcification-related diseases. A research team from Shandong First Medical University investigated the molecular mechanisms underlying the development of calcific aortic valve disease by inducing

osteogenesis in human valvular interstitial cells in vitro[3]. This review elucidates the mechanisms by which they regulate the osteogenic differentiation of MSCs, as well as the facilitative effects of related activators observed during in vitro culture and differentiation. Studies have shown that mesenchymal progenitor cells in skeletal muscles near bones mediate initial fibrosis in damaged areas and participate in the formation of cartilage and bone, thereby playing a key role in bone regeneration; clinically, it has also been found that complete coverage by skeletal muscle can improve bone healing. Xing and others found that skeletal muscle is involved in chondrogenesis and endochondral ossification during the early stages of callus formation[4]. Irisin, a peptide fragment secreted synthetically by skeletal muscle, has also been shown to promote osteogenic differentiation, either exogenously or endogenously, by inhibiting lipogenic gene expression and activating the classical Wnt pathway[5]. Therefore, this review also includes the discussion of skeletal muscle directional differentiation.

The core process in the osteogenic differentiation of MSCs is osteogenesis. Osteogenesis is the process of bone tissue formation primarily carried out by osteoblasts. Mesenchymal cells, under the influence of growth factors, TGF-β, and signaling pathways, first differentiate into osteoprogenitor cells, and then further differentiate into osteoblasts. Osteoblasts migrate to bone formation sites and secrete bone matrix, constituting the process of bone formation[6].

2. Wnt Proteins and Corresponding Signaling Pathways

Wnt proteins are a class of highly conserved signaling proteins that play an important role in regulating development and modulating diseases such as cancer. In pluripotent MSCs, Wnts attach to the cell surface or extracellular matrix, and function through autocrine or paracrine mechanisms[7]. Figure 1 shows the three-dimensional structure of protein Wnt-5a, a typical factor in the family.

Figure 1: The three - dimensional structure of Protein Wnt-5a[8]

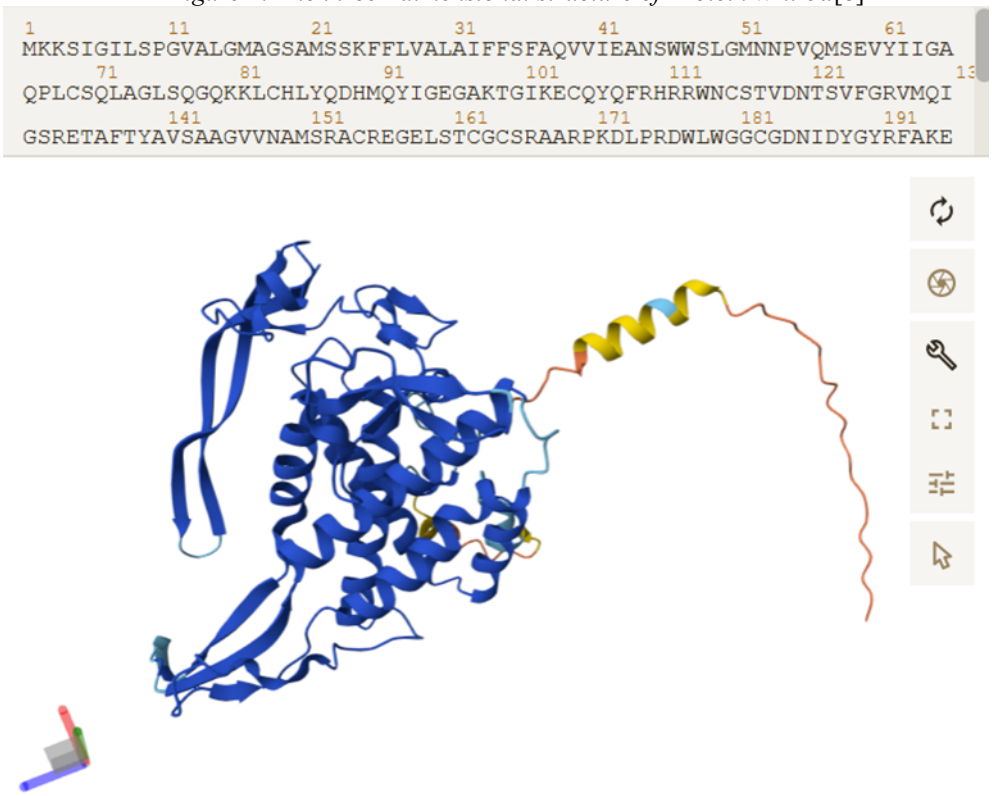
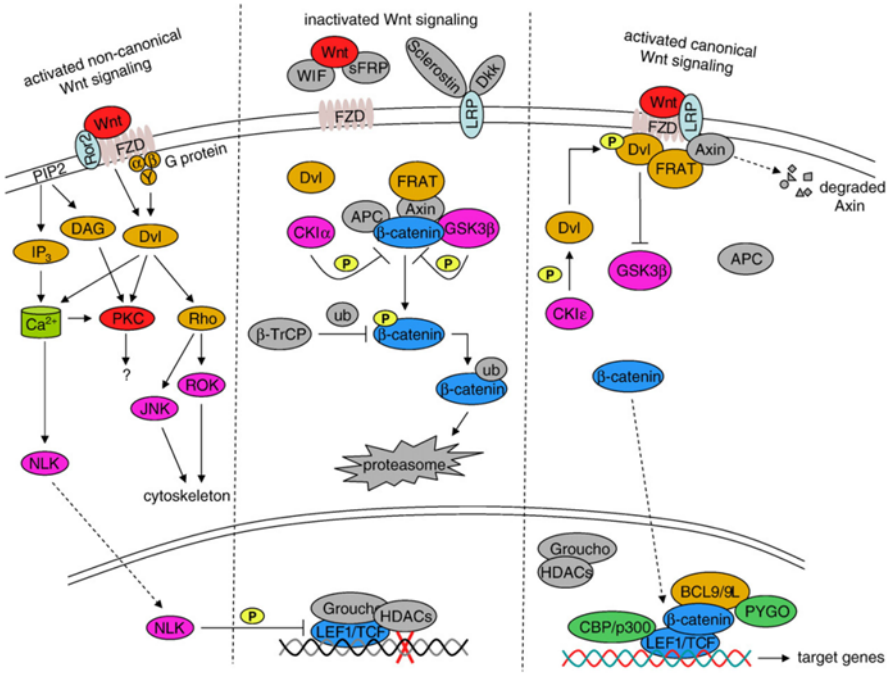


Figure 2: Schematic representations of the canonical and non - canonical Wnt pathway[7]



The Wnt signaling pathway is divided into canonical one and non-canonical one, with β-catenin serving as the hallmark protein of the canonical Wnt pathway. The canonical Wnt pathway is activated when Wnts bind to receptors Fzd and LRP5/6, resulting in decreased phosphorylation of β-catenin to prevent its binding modification by ubiquitin molecules and delivery to the proteasome, which then translocates into the nucleus to activate the transcription of specific genes (Figure 2).

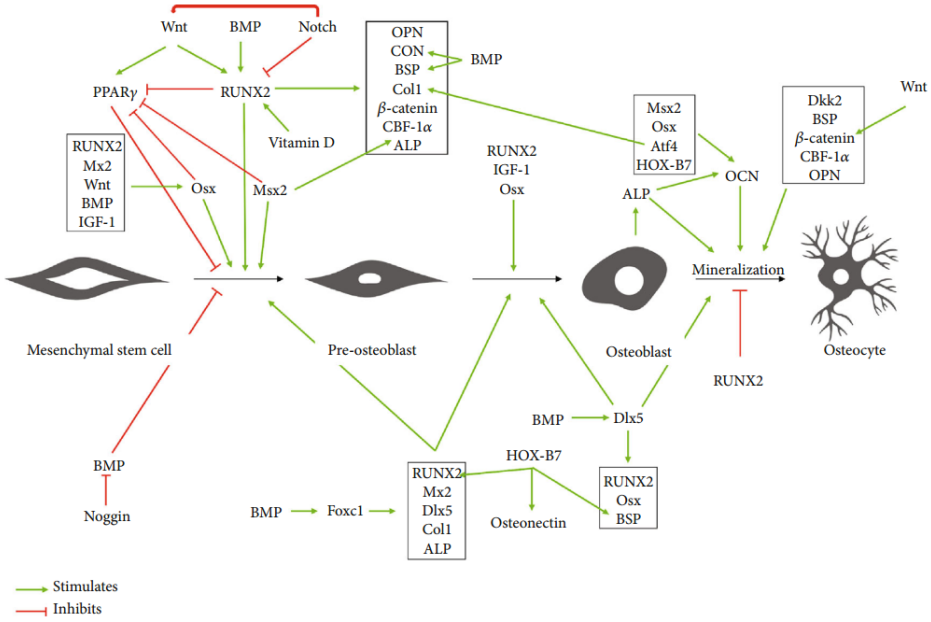
3. Osteogenesis-Related MSCs Differentiation

3.1 Osteogenic Differentiation

During the differentiation of osteoblasts, Wnt proteins influence cell fate by regulating the expression of a series of transcription factors, among which Runx2 and Osterix are two key transcription factors. Activation of the Wnt signaling pathway can promote the expression of Runx2 and Osterix, thereby facilitating the differentiation of osteoblasts into osteocytes and initiating the synthesis and deposition of bone matrix[6].

In the regular bone formation process, the impact of the canonical Wnt pathway on osteogenesis depends on the developmental stage of the target cells. The canonical pathway stimulates differentiation of MSCs while inhibiting the terminal differentiation of mature osteoblasts. Within the non-canonical signaling pathways, Wnt11 is upregulated during both osteogenic and chondrogenic differentiation of MSCs. Since non-canonical Wnt signaling is dynamically related to calcium ion signaling, promoting the osteogenic differentiation of osteoprogenitor cells by regulating the activation and expression of key transcription factors such as Runx2 and Osterix. In addition, by modulating calcium ion concentration, it can indirectly upregulate essential osteogenic genes like Runx2 and COL1A1, directly promoting bone matrix synthesis and mineralization[7]. Figure 3 displays effects of multiple bone differentiation-related biomolecules on various stages of osteogenic differentiation of MSCs.

Figure 3: The effects of bone differentiation-related biomolecules on the osteogenesis process[9]



3.2 Chondrogenic Differentiation

As for chondrogenesis, Wnt signaling exerts an inhibitory effect, corresponding to the upregulation of β-catenin in osteoprogenitor cells and its downstream regulation in chondroprogenitor cells. During this induction process, Sox9 is a specific transcription factor essential for chondrocyte differentiation and can suppress the transcriptional activity of Runx2 and β-catenin. The canonical Wnt pathway can regulate Sox9 to promote chondrogenic differentiation during the early stages of MSC differentiation. In the later stages, the antagonism between Sox9 and β-catenin has a significant effect[10].

3.3 Myogenesis

Wnt proteins serves as important stimulators of myogenesis in MSCs. The canonical Wnt pathway can induce myogenesis by activating the gene expression of myogenic regulatory factors (MRFs)[11]. Wnts directly affect myosatellite cells during the early stages of myogenesis. After the activation of Wnt, precursor cells shift from proliferation toward myogenic differentiation[12]. Conversely, inhibition of this pathway results in precursor cells differentiating toward adipocyte[13].

4. Critical Factors for In Vitro Culture and Differentiation of MSCs

4.1 In Vitro Differentiation of MSCs

At present, in vitro differentiation of MSCs is mainly used to construct in vitro models for providing medical strategies and tissue transplantation therapy[9]. The heterogeneity exhibited during the in vitro expansion and cultivation of MSCs leads to differences in cell morphology and physiological functions while the aging process is triggered, which affects the clinical outcomes of tissue transplantation. It has been found in research that specific culture conditions, such as hypoxic and low-temperature environments, are beneficial for MSCs to maintain their conformation and slow down the acceleration of aging caused by in vitro culture (Figure 4). Moreover, modifying and enhancing certain signaling pathways in MSCs also facilitates the expression of chemokines, thereby promoting the homing of MSCs to damaged areas for repair.

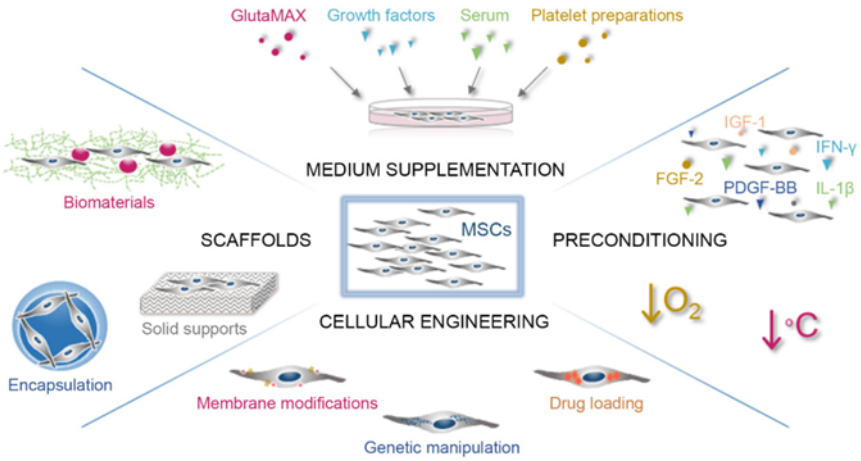


Figure 4 Various methods to promote the in vitro culture and differentiation of mesenchymal stem cells[14]

4.2 In Vitro Osteogenic Differentiation

In the laboratory, adding follicle-stimulating hormone β-subunit (FSHβ) to the culture medium can enhance the BMP9-induced transcription of cyclin D1,then augment BMP9-induced Wnt/β-catenin signaling, ultimately strengthening the osteogenic effect induced by BMP9[15].

4.3 In Vitro Chondrogenic Differentiation

Hypoxia-inducible factor (HIF-1α) can be added to the culture medium for in vitro chondrogenic differentiation. MSCs differentiate into chondrocytes in an avascular environment. HIF-1α functions to promote Sox9 gene expression under hypoxic conditions and simultaneously reduce protease-mediated degradation, thereby maintaining Sox9 mRNA levels, stimulating the expression of Sox5 and Sox6, and upregulating the expression of type II collagen and proteoglycans[16]. It has also been shown that HIF-1α has a role in co-regulating Inhibin β-A (INHBA), belonging to Sox9-independent genes, with HIF-2α to promote chondrogenic differentiation of MSCs under hypoxic conditions[17]. Another optional ingredient is lithium chloride (LiCl) .The effect of LiCl is to activate β-catenin, which promotes rapid proliferation of MSCs in the early stage and regulates the upregulation of Sox9 expression to enhance chondrogenic differentiation of MSCs. However, since β-catenin also promotes osteogenic differentiation, creating a strong antagonistic effect with Sox9 at a later stage of differentiation, excessive addition of LiCl causes MSCs to tend to differentiate into osteoblasts. Thus, the dosage of this inducer must be strictly controlled[18].

4.4 In Vitro Myogenic Differentiation

Researches have proved that Adipose-Derived Mesenchymal Stem Cells (ADSCs) have potential for myogenic differentiation both in vivo and in vitro [19]. It was found that the medium under low calcium concentration (calcium concentration <0.002 mmol/mL) promoted ADSC proliferation and differentiation in the direction of myogenesis. The potential mechanism may be the inhibitory effect of high calcium concentration on MyoG presence. Since the Wnt pathway exists to upregulate the expression of MRFs, the mechanism may be related to the inhibition of the Wnt pathway[20]. Through this process, MSCs differentiate into myosatellite cells, which differentiate into myofibroblasts and myoblasts in turn.

5. Discussion

In the in vitro culture of MSCs related to osteogenesis, although the proteins within the pathway and their interactions has already been proven for a long time, the effects of activators or inhibitors on the final

differentiation outcome have remained elusive due to the interactions between signaling pathways and the varying influences these pathways exert on MSCs at different stages of differentiation. At present, research methods on the effects of activators or inhibitors on in vitro differentiation remain limited to dosage titration, with feasibility judged solely by calcium deposition assays.

Since the Wnt signaling pathway relies on cellular transcription mechanisms, the author believes that chromatin immunoprecipitation followed by sequencing (ChIP-seq) technology can be used to study the specific binding sites of activators or inhibitors. ChIP-seq technology specifically enriches DNA fragments bound by target proteins through chromatin immunoprecipitation, enabling the identification of genomic loci and providing genome-wide information on interactions with histones, transcription factors, and other key regulators. In previous studies, researchers have used ChIP-seq technology to demonstrate that Sox transcription factors are essential for the formation of Wnt enhanceosomes by β-catenin, which itself lacks intrinsic DNA-binding activity, and for the recruitment of these complexes to different lineage-specific WNT-responsive enhancers (WREs), thereby regulating transcription[21]. The ChIP-seq technology also be applied to demonstrate that Sox transcription factors are essential for the formation of Wnt enhanceosomes by β-catenin, which itself lacks intrinsic DNA-binding activity, and for the recruitment of these complexes to different lineage-specific WREs, thereby regulating transcription[22]. It illustrates that ChIP-seq can more accurately determine the effects of added activators or inhibitors on transcription by pinpointing the binding sites of specific proteins or transcription factors.

There is a competitive relationship among the several differentiation pathways of MSCs, and the previously mentioned osteogenic and chondrogenic differentiation is a typical example. Osteoporosis, which is often seen in the elderly population, is also associated with a dysregulation of the competitive balance between the osteogenic and lipogenic directions of MSCs[23], which is often accompanied by lipid metabolism-related diseases[24]. ChIP-seq assist in exploring the interactions between differentiation directions by uncovering potential targets of activators or inhibitors in the transcriptional landscape, and provide new targets for disease-related drugs while improving the efficiency of in vitro differentiation. This provides new ideas and directions for the treatment of bone-related diseases.

6. Conclusion

As one of the key signaling pathways affecting cell differentiation, the Wnt pathway plays an extremely important role in the in vitro culture and differentiation of MSCs. This review starts by exploring the mechanisms through which the Wnt pathway affects MSCs, and discusses activators that can be utilized in the culture environment for osteogenic, chondrogenic, and myogenic differentiation. In addition, the review introduces the emerging ChIP-seq technology which holds significant value for identifying the specific transcriptional targets of signaling pathways involved in cell differentiation, assisting stem cells in osteogenesis-related directed differentiation cultures, shows the potential of application in tissue engineering techniques and establishment of in vitro models.

References

[1] Wei, Y. C., Xie, Y. C., Xuan, A. W., Lian, Y., Wang, Z. and Yu, H. L. Analysis of the burden of vertebral column fracture in china from 1990 to 2021 and trend forecast to 2035. Chinese Journal of Spine and Spinal Cord. 2025, 35(6), pp. 631-638. https://doi.org/10.3969/j.issn.1004-406X.2025.06.09.

[2] Luo, Z. J., Yang, L. and Wang, D. Organoids: novel strategies for locomotor system repair and reconstruction. Journal of Air Force Medical University. 2025, 46(8), pp. 981-985. https://doi.org/10.13276/j.issn.2097-1656.2025.08.001.

[3] Xing, K., Zheng, Q., Sun, J. S., Liu, X. L. and Wang, Z. J. Comparison of osteogenic induction methods for human valve interstitial cells in vitro. Journal of Shandong University(Health Sciences). 2025, 63(06), pp. 11-18. https://doi.org/10.6040/j.issn.1671-7554.0.2024.0701.

[4] Julien, A., Kanagalingam, A., Martínez-Sarrà, E., Megret, J., Luka, M., Ménager, M., Relaix, F. and Colnot, C. Direct contribution of skeletal muscle mesenchymal progenitors to bone repair. Nature Communications. 2021, 12(1), p. 2860. https://doi.org/10.1038/s41467-021-22842-5.

[5] Liu, D. F. and Tian, B. F. Research progress and application prospects of myogenic factors involved in bone and fat formation. Advances in Clinical Medicine. 2023, 13, p. 18842. https://doi.org/10.12677/ACM.2023.13122651.

[6] Shi, Y. The investigation of energy metabolism in osteoblasts and osteoclasts. West China Journal of Stomatology. 2021, 39(5), pp. 501-509. https://doi.org/10.7518/hxkq.2021.05.002.

[7] Ling, L., Nurcombe, V. and Cool, S. M. Wnt signaling controls the fate of mesenchymal stem cells. Gene. 2009, 433(1), pp. 1-7. https://doi.org/10.1016/j.gene.2008.12.008.

[8] Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P. and Avsec, Ž. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science, 381(6664), p. eadg7492. https://doi.org/10.1126/science.adg7492.

[9] Mollentze, J., Durandt, C. and Pepper, M. S. An In Vitro and In Vivo Comparison of Osteogenic Differentiation of Human Mesenchymal Stromal/Stem Cells. Stem Cells International. 2021, 2021(1), p. 9919361. https://doi.org/10.1155/2021/9919361.

[10] Chen, Z. H., Zhao, G. L., Shi, J. S., Jiang, L. and Xia, J. Regulation mechanism of Sox9 on chondrocyte differentiation and matrix production. Fudan University Journal of Medical Sciences. 2019, 46(6), pp. 824-828. https://doi.org/10.3969/j.issn.1672-8467.2019.06.018.

[11] von Maltzahn, J., Chang, N. C., Bentzinger, C. F. and Rudnicki, M. A. Wnt signaling in myogenesis. Trends in Cell Biology. 2012, 22(11), pp. 602-609. https://doi.org/10.1016/j.tcb.2012.07.008.

[12] Zheng Qi, Z. Q., Sui MengHua, S. M. and Ling YingHui, L. Y. The role of key signaling pathways in the proliferation and differentiation of skeletal muscle satellite cells. Acta Veterinaria et Zootechnica Sinica. 2017, 48(11), pp. 2005–2014.

[13] Shen, L., Geng, B. and Zhou, S. Effect of WNT2B on adipocytogenic differentiation of human bone marrow mesenchymal stem cells. Chinese Journal of Experimental Surgery. 2022, 39(1), pp. 108-111.

[14] Olmedo-Moreno, L., Aguilera, Y., Baliña-Sánchez, C., Martín-Montalvo, A. and Capilla-González, V., 2022. Heterogeneity of In Vitro Expanded Mesenchymal Stromal Cells and Strategies to Improve Their Therapeutic Actions. *Pharmaceutics.*

[15] Cheng, Y. L., Zhao, Y. X., Zhang, R. N. and Liu, H. The study of the Wnt/β-catenin on FSHβ enhanced BMP9-induced mesenchymal stem cells osteogenesis. Chinese Journal of Osteoporosis. 2020, 26(12), pp. 1773-1778. https://doi.org/10.3969/j.issn.1006-7108.2020.12.010.

[16] Zhou, J. J., Yu, G. R., Cao, C. F., Chen, X. Q., Pang, J. H. and Shi, J. X. SymbolDirectional differentiation of tissue engineered chondrocytes. Journal of Clinical Rehabilitative Tissue Engineering Research. 2010, 14(33), pp. 6194-6197. https://doi.org/10.3969/j.issn.1673-8225.2010.33.028.

[17] Lafont, J. E., Talma, S., Hopfgarten, C. and Murphy, C. L. Hypoxia Promotes the Differentiated Human Articular Chondrocyte Phenotype through SOX9-dependent and -independent Pathways *. Journal of Biological Chemistry. 2008, 283(8), pp. 4778-4786. https://doi.org/10.1074/jbc.M707729200.

[18] Cheng, P. Effect and molecular mechanisms of s signaling pathway on chondrogenic differentiation of bone mesenchymal stem cells. Doctoral Dissertation, Wuhan: Huazhong University of Science and Technology, 2015.

[19] Torres-Torrillas, M., Rubio, M., Damia, E., Cuervo, B., del Romero, A., Peláez, P., Chicharro, D., Miguel, L. and Sopena, J. J. Adipose-Derived Mesenchymal Stem Cells: A Promising Tool in the Treatment of Musculoskeletal Diseases. International Journal of Molecular Sciences. 2019, 20(12), p. 3105. https://doi.org/10.3390/ijms20123105.

[20] Chen, H., Ke, C., Zhang, Z. K., Li, L. Q. and Lin, M. Effects of calcium concentration on proliferation and myogenic differentiation of adipose stem cells. Progress in Modern Biomedicine. 2022, 22(7), pp. 1206-1210. https://doi.org/10.13241/j.cnki.pmb.2022.07.002.

[21]Nakato, R. and Sakata, T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. Methods. 2021, 187, pp. 44-53. https://doi.org/10.1016/j.ymeth.2020.03.005.

[22] Mukherjee, S., Luedeke, D. M., McCoy, L., Iwafuchi, M. and Zorn, A. M. SOX transcription factors direct TCF-independent WNT/β-catenin responsive transcription to govern cell fate in human pluripotent stem cells. Cell Reports. 2022, 40(8). https://doi.org/10.1016/j.celrep.2022.111247.

[23] Du, G., Cheng, X., Zhang, Z., Han, L., Wu, K., Li, Y. and Lin, X. TGF-Beta Induced Key Genes of Osteogenic and Adipogenic Differentiation in Human Mesenchymal Stem Cells and MiRNA–mRNA Regulatory Networks. Frontiers in Genetics. 2021, 12, p. 759596. https://doi.org/10.3389/fgene.2021.759596.

[24] Wang, L. M., Liu, K. J., Liu, Z. X. and Zhou, Y. The relationships between lipid metabolism-related diseases and osteoporosis. Chinese Journal of Osteoporosis. 2021, 27(05), pp. 767-770+775. https://doi.org/10.3969/j.issn.1006-7108.2021.05.027.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output ofthe'science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260103

Hi-C Scaffolding Challenges in Non-Model Lepidoptera

Xueting Wang*

College of Life Sciences, Tianjin Normal University, Tianjin, China

**Corresponding author: Xueting Wang*

Abstract

The limited availability of high-quality genomic resources for non-model Lepidoptera severely hampers research on their evolutionary adaptations and restricts their application in biodiversity conservation and agroforestry. This study systematically addresses a critical bottleneck-quality degradation in Hi-C scaffolding-that undermines the structural integrity of Lepidoptera genome assemblies. The findings show that intrinsic genomic features, such as high heterozygosity and abundant repetitive sequences, coupled with technical noise in Hi-C data and algorithmic limitations, frequently lead to large-scale structural errors (including misjoins, inversions, and translocations) that are not easily detected by conventional assembly metrics. To address these challenges, the paper proposes an integrated framework that combines haplotype-resolved assembly, manual curation, and multi-dimensional evaluation. This framework integrates optimized experimental and computational workflows, stringent quality control procedures, and standardized evaluation criteria to establish a comprehensive quality assurance system. Additionally, this paper emphasizes the importance of community collaboration and data transparency in fostering reproducible and scalable genomic research. Together, these advances are expected to significantly enhance the reliability and applicability of genomic resources for non-model Lepidoptera, providing a robust genetic foundation for evolutionary biology, comparative genomics and integrated pest management.

Keywords

non-model lepidoptera, Hi-C scaffolding, genome assembly, quality assessment, haplotype phasing

1. Introduction

Non-model lepidopteran species constitute vital components of global biodiversity and provide substantial research value across diverse fields, including agriculture, biomedicine, biomimetic materials, and ecological conservation. Nevertheless, genome assembly for these organisms has historically been impeded by their distinctive genomic characteristics-notably high heterozygosity and abundant repetitive sequences-along with persistent technical barriers. Current genomic research on non-model Lepidoptera is rapidly advancing, with technological innovations shifting the focus from foundational genome sequencing toward comprehensive functional and mechanistic interpretation. For instance, phylogenomic demonstrated that even analyses incorporating hundreds of genes remain constrained by early transcriptomes and draft genomes-characterized by fragmented contigs and compositional bias-hindering resolution of key evolutionary relationships within lineages such as Gelechioidea[1]. These findings underscore the critical necessity for high-quality,

chromosome-level genome assemblies. Consequently, the research paradigm is evolving from basic sequence assembly toward an integrated framework combining precise genomic data with multi-omics approaches to elucidate underlying biological mechanisms.

Long-read sequencing technologies, notably PacBio (HiFi and CLR) and Oxford Nanopore Technologies (ONT), coupled with chromatin conformation capture techniques such as Hi-C/Omni-C, have substantially enhanced the efficiency and cost-effectiveness of generating high-quality genome assemblies. These advancements have facilitated pivotal shifts in genomic research: from single-genome analyses to pangenomic investigations elucidating adaptive evolution across populations; from static sequence assembly to dynamic regulatory mapping via integration with transcriptomic and other multi-omics data; and from correlation-based inference toward causal validation utilizing gene-editing technologies like CRISPR/Cas9. Recent systematic evaluations provide comprehensive comparisons of these technologies and their associated bioinformatic workflows [2, 3].

This study addresses critical limitations in genome assemblies for non-model Lepidoptera, focusing specifically on challenges encountered during Hi-C scaffolding. Key issues include interference from tandem repeat arrays, optimization of proximity ligation efficiency parameters and the requirement for manual curation of chromatin interaction maps. To resolve these challenges, this paper proposes an integrated strategy combining long-read sequencing (PacBio HiFi or ONT) with Hi-C scaffolding using established pipelines such as chromap+YaHS or BWA+HapHiC, enhanced by manual refinement using Juicebox. This approach demonstrably improves the completeness and accuracy of genome assemblies for non-model lepidopteran species [4].

2. Challenges and Limitations in Hi-C Scaffolding for Non-Model Lepidoptera Genomes

While long-read sequencing and Hi-C scaffolding have become the “gold standard” for high-quality genome assembly, yielding advanced evaluation metrics such as Scaffold N50 and BUSCO scores, these continuity indicators often obscure substantial structural deficiencies, undermining the reliability of downstream biological analyses [5]. A major concern is that high continuity metrics may coexist with high structural error rates, posing a significant limitation to current assembly evaluation frameworks. For instance, in the benchmarking study, the genome assembled from ONT data achieved an impressive Scaffold N50 of 17.48 Mb and a BUSCO completeness of 94.8%, yet contained 852 structural errors. This discrepancy underscores the inadequacy of conventional metrics in detecting misassemblies related to sequence order and orientation at the chromosomal level.

The persistence of errors-such as misjoins, inversions, and translocations-compromises the integrity of genome-based studies, particularly in comparative genomics and gene family analyses. These issues are especially pronounced in non-model Lepidoptera, due to their high heterozygosity and repetitive content. Manual curation, using tools like Juicebox, and advanced evaluation methods, such as EagleC, are essential for ensuring assembly accuracy, even when standard metrics suggest high continuity and completeness [6].

Achieving accurate chromosome-scale assemblies is primarily hindered by inherent limitations in the Hi-C scaffolding process. First, tandem repetitive regions constitute a major impediment to correct scaffold construction. For instance, Zhang et al.’s effort to assemble the first telomere-to-telomere genome of the silkworm demonstrated that the complex “TTAGG” telomeric repeats prevented the generation of single contigs spanning three chromosomes, necessitating gap filling using external reference sequences. This underscores that highly repetitive regions-such as telomeres and centromeres-disrupt Hi-C interaction signals and frequently lead to chimeric scaffolds.

Second, parameterizing proximity ligation thresholds in scaffolding algorithms entails a fundamental compromise. Walden et al. explicitly illustrated this dilemma: their Omni-C contact maps revealed pervasive off-diagonal signal blocks and anomalous interaction patterns within scaffolds. These artifacts highlight the challenge for algorithms in balancing sensitivity and specificity-excessively lenient parameters induce misjoins, while overly stringent thresholds result in premature chromosomal fragmentation.

These challenges necessitate substantial reliance on manual curation within contemporary genome assembly workflows, highlighting a fundamental limitation in achieving fully automated processing at critical stages. As noted by Zhang et al., “manual correction using Juicebox, while effective for error resolution, lacks

standardized protocols." Similarly, Walden et al. emphasized that "most assemblies would benefit from additional manual curation." These observations collectively confirm that human intervention remains indispensable in current scaffolding practices.

A particularly illustrative example is Zhang's comparative analysis, which identified a megabase-scale inversion on chromosome 24 of the previously published P50T-SilkBase reference genome-an error unequivocally revealed by Hi-C interaction maps. This case underscores a critical concern: even publicly accessible reference genomes may harbor undetected large-scale structural errors. Rectifying such inaccuracies still relies on researchers' empirical judgment and laborious manual curation, which not only introduces subjectivity but also impedes the scalable, reproducible utilization of genomic data.

3. Analysis of the Root Causes of Quality Defects and Technical Route Optimization

The assembly of high-quality genomes presents significant challenges for non-model Lepidoptera species, which frequently exhibit extreme genomic characteristics including large genome sizes, high repetitive content, and substantial heterozygosity. Although the integrated approach of PacBio HiFi sequencing with Hi-C scaffolding represents the current "gold standard," the resulting assemblies often contain undetected structural errors that compromise their biological utility. A comprehensive investigation into the origins of these deficiencies is imperative, both for assessing the reliability of extant genomic resources and for informing the design of future sequencing endeavors targeting this ecologically and economically important insect group.

The structural errors in Hi-C scaffolding mentioned above do not exist in isolation, and their root causes can be systematically analyzed from three aspects: intrinsic genomic features, technical data attributes, and algorithm limitations. The primary challenges in Hi-C scaffolding arise from both the intrinsic characteristics of Hi-C data and the limitations inherent in current computational algorithms. Hi-C technology infers linear genomic organization from three-dimensional chromatin proximity data; however, experimental artifacts-including uneven cross-linking efficiency and restriction enzyme bias-significantly compromise the signal-to-noise ratio, obscuring the distinction between genuine chromatin interactions and random collisions.

These data limitations are further amplified by variations in algorithmic approaches. Global partitioning methods, such as 3D-DNA, effectively correct large-scale interchromosomal translocations but frequently fail in highly repetitive regions, exemplified by centromeres in lepidopteran genomes. Signal depletion in these regions often results in assembly breaks or chimeric joins. Conversely, graph-based approaches like SALSA offer enhanced flexibility but exhibit high sensitivity to connection thresholds. Suboptimal parameterization in complex genomes can consequently introduce small-scale inversions or translocations.

Benchmark studies consistently indicate that no single algorithm resolves all structural errors, underscoring the necessity for specialized post-assembly validation tools such as EagleC to correct such errors. This highlights the inherent limitations of fully automated scaffolding pipelines in addressing genomic complexity, particularly for non-model organisms possessing intricate genomic architectures.

Hi-C scaffolding challenges frequently arise from upstream sequencing quality. PacBio HiFi sequencing delivers high accuracy, producing reliable contigs and minimizing structural errors. Conversely, Oxford Nanopore Technologies (ONT) offers longer reads and higher throughput but exhibits higher raw error rates, which can result in uncorrected sequence errors within initial assemblies. When these erroneous contigs are utilized for Hi-C scaffolding, algorithms may inadvertently incorporate these errors into chromosomal structures. Consequently, minor contig errors can escalate into significant chromosomal defects, as demonstrated in case studies.

A key challenge in genome assembly resides in the context-dependent performance of diverse software tools. Systematic benchmarking studies on model species, such as the silkworm, have revealed that even when processing identical datasets, assemblers like HiFiASM and NextDenovo generate contigs exhibiting divergent error profiles. This observation underscores a fundamental challenge: no universally optimal assembler exists, as tool efficacy is heavily contingent upon data type and intrinsic genomic characteristics. The current lack of comprehensive benchmarking across the field compels most non-model genome projects to rely on empirical tool selection. This approach invariably introduces software-specific artifacts, thereby compromising data reliability and reusability.

Current technological approaches present several trade-offs. While the "PacBio HiFi + Hi-C" combination remains the gold standard, its high cost limits large-scale species surveys. Oxford Nanopore technology provides a scalable alternative, but its effectiveness hinges on careful evaluation of the error rate-to-cost ratio. Algorithmically, the context-dependence of tools and the lack of systematic benchmarking present significant bottlenecks in enhancing data quality. Therefore, advancing comprehensive benchmarking initiatives and developing next-generation algorithms capable of integrating multi-source data and automating error correction are crucial steps to overcome current limitations and ensure that genomic resources provide reliable biological insights--This issue suggests that optimizing the technical route requires personalized selection of assembly tools based on the genomic characteristics of the target species, such as heterozygosity and repeat sequence proportion, rather than adopting a one size fits all universal process.

4. An Integrated Approach to Constructing High-Quality Genomes of Non-Model Lepidoptera

The challenges in Hi-C scaffolding of non-model Lepidoptera genomes arise from high heterozygosity, abundant repetitive sequences, experimental noise, and algorithmic limitations. These factors amplify noise levels within chromatin proximity signals, resulting in large-scale structural errors in assemblies. Such inaccuracies are often obscured by high contiguity metrics yet compromise biological validity. Addressing this necessitates a systematic, integrated approach optimizing technological, evaluative, and community standards, thereby advancing Lepidoptera genomics from qualitative assessment towards quantitative precision.

The transition from conventional hybrid genome assembly to haplotype-resolved assembly constitutes a significant methodological advancement in genomic reconstruction. This paradigm shift necessitates a strategic reorientation of assembly strategies, prioritizing high-fidelity long-read sequencing technologies such as PacBio HiFi, in conjunction with advanced phasing-enabled assemblers like hifiasm [7]. These tools enable the precise resolution of highly heterozygous genomes into two complete haplotypes during the initial assembly phase, thereby providing purified input data for subsequent Hi-C scaffolding.

Implementing independent Hi-C scaffolding for each haplotype utilizing specialized workflows-such as chromap coupled with yahs-effectively mitigates allelic interference and substantially reduces the risk of large-scale misassemblies. This approach has been empirically validated in pioneering initiatives, including the Human Pangenome Reference Consortium, establishing novel benchmarks for the production of high-quality, allele-aware reference genomes [8].

The systematic application of this methodology to non-model Lepidoptera species addresses assembly fragmentation and structural inaccuracies induced by high heterozygosity, transitioning genomic data quality from merely "usable" to scientifically "reliable." Furthermore, the targeted incorporation of characterized telomeric sequences, such as the "TTAGG" repeat motif, during the assembly process enhances chromosomal end completeness, laying a robust foundation for achieving authentic telomere-to-telomere genome assemblies.

Developing a multidimensional quality assessment framework that surpasses conventional metrics is essential for ensuring the credibility of genomic assemblies and is a crucial step toward standardizing the field. While traditional indicators such as Scaffold N50 and BUSCO scores assess assembly contiguity and gene completeness, they exhibit systematic limitations in detecting chromosomal-scale structural errors. Therefore, establishing a more rigorous and comprehensive quality evaluation protocol is critical, incorporating mandatory validation steps, such as Hi-C contact map-based automated assessments using deep learning tools like EagleC, combined with long-read reconciliation analysis. At the same time, chromosome mounting verification can be supplemented with LACHESIS software to further reduce misjudgment rates through multi-tool cross-validation.

Hi-C-based evaluation enables the quantitative and objective identification of cryptic structural variations, such as inversions and translocations, which manifest as characteristic anomalous patterns in the interaction matrix. Simultaneously, long-read reconciliation verifies assembly authenticity at the sequence level by assessing alignment concordance between raw long reads and the assembled genome, providing a complementary evaluation of both accuracy and completeness. This integrated "multi-layered validation" framework should become the new standard for defining "reference-grade" genomes, effectively addressing the current overemphasis on contiguity at the expense of structural accuracy in genomic quality assessment.

To promote sustainable progress across the field, it is essential to establish a more open and collaborative community ecosystem. We propose launching a "Benchmark Dataset" initiative specifically targeting non-model Lepidoptera genomes. This program will systematically collect and curate comprehensive datasets from species representing varying levels of genomic complexity, including HiFi/ONT long-read sequences and Hi-C interaction data. These meticulously curated benchmark resources will provide a transparent and equitable platform for evaluating diverse assembly algorithms and workflows.

Standardized datasets will facilitate the establishment of standardized methodologies and the iterative refinement of computational tools, enabling researchers to evaluate and select approaches based on consistent criteria. Concurrently, we advocate for implementing transparent data release practices, mandating the submission of essential raw data-including Hi-C interaction matrices-alongside publications. This requirement enhances the rigor of peer review and maximizes the long-term utility of datasets,as demonstrated in recent high-quality chromosome-level assemblies such as that of the Atlas moth [9], thereby reducing redundant research efforts and resource expenditure arising from data inaccessibility. This commitment to open collaboration will steer Lepidoptera genomics toward a more reproducible and cumulative research paradigm, establishing a foundation for sustained progress within the field [10].Collectively, the systematic implementation of haplotype-resolved assembly strategies, the establishment of rigorous multidimensional quality assessment protocols, and the promotion of open community collaboration will directly address the core challenges in Hi-C scaffolding for non-model Lepidoptera genomes. This integrated approach will establish a comprehensive genomic framework essential for deciphering the exceptional diversity and evolutionary history of this extensive insect group, ushering in a transformative phase for Lepidoptera genomics.

Future research should prioritize three areas: first, developing more advanced algorithms capable of autonomously detecting and correcting structural errors; second, establishing comprehensive evaluation standards that account for both sequence accuracy and biological validity; and third, fostering broader data sharing through standardized formats and repositories. Collectively, these efforts will contribute to a more efficient and sustainable genomics research ecosystem for Lepidoptera, advancing insights into insect evolution and biodiversity conservation.

5. Conclusion

The assembly of high-quality genomes for non-model Lepidoptera species remains a substantial challenge, with the primary bottleneck arising from quality deficiencies in Hi-C scaffolding. This study rigorously examines the underlying causes of these limitations: the interplay between genomic features-notably elevated heterozygosity and abundant repetitive elements-and technical artifacts inherent in Hi-C data, exacerbated by algorithmic constraints. Collectively, these factors induce pervasive structural errors undetectable by conventional assembly metrics.

To address this multifaceted challenge, we propose an integrated solution comprising three core components: optimized technical workflows incorporating haplotype-resolved assembly strategies, expert-guided manual curation, and multidimensional evaluation frameworks utilizing three-dimensional genomic evidence. This approach establishes a robust quality assurance system for genome assembly projects.

Prospectively, fostering community collaboration through standardized benchmark datasets and promoting data transparency will be imperative for the sustainable advancement of this field. These initiatives will enhance the reliability of genomic resources, accelerate understanding of Lepidoptera evolution and biology, and ultimately advance biodiversity conservation and pest management efforts.

This research delineates a clear technical roadmap and quality standards for non-model Lepidoptera genome studies, propelling the field from merely "obtaining genomes" toward systematically "acquiring high-quality genomic resources." Addressing Hi-C scaffolding quality issues not only bolsters the reliability of comparative genomics and evolutionary biology research but also furnishes a more precise genetic foundation for practical applications, including pest management and biodiversity conservation.

It is crucial to acknowledge current solution limitations, particularly the reliance on expert knowledge during manual curation and the substantial computational demands of haplotype-resolved assembly strategies-the reliance on expert experience for manual calibration directly limits the efficiency of this approach in large-scale genome assembly of non model Lepidoptera species; However, high computing demands have raised the

technical threshold for small and medium-sized laboratories, which is not conducive to the popularization of technology. Future advancements in sequencing technologies and algorithmic innovations-especially deeper integration of artificial intelligence into genome assembly pipelines-are anticipated to reduce dependence on manual intervention, enabling more efficient and precise automated assembly processes.

Furthermore, enhancing international collaborations and establishing rigorous quality control frameworks will propel non-model Lepidoptera genomic research into a new phase. Such concerted initiatives will provide enhanced technical support for elucidating the evolutionary mechanisms underpinning insect diversity, thereby enabling the development of more effective conservation and management strategies for this ecologically pivotal group.

References

[1] Kawahara, A. Y., Storer, C., Carvalho, A. P. S., Plotkin, D. M., Condamine, F. L., Braga, M. P., Ellis, E. A., St Laurent, R. A., Li, X., Barve, V., et al. Evolutionary genomics of the Lepidoptera: From silkmoths to pest species. Annual Review of Entomology. 2019, 64, pp. 257-276. https://doi.org/10.1146/annurev-ento-011118-112424.

[2] Zhang, T., Xing, W., Wang, A., Zhang, N., Jia, L., Ma, S. and Xia, Q. Comparison of long-read methods for sequencing and assembly of lepidopteran pest genomes. International Journal of Molecular Sciences. 2023, 24(1), p. 649. https://doi.org/10.3390/ijms24010649.

[3] Walden, K. K. O., Cao, Y., Fields, C. J., Hernandez, A. G., Rendon, G. A., Robinson, G. E., Skinner, R. K., Stein, J. A. and Dietrich, C. H. High-quality genome assemblies for nine non-model North American insect species representing six orders (Insecta: Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Neuroptera). Molecular Ecology Resources. 2024, 24(8), p. e14010. https://doi.org/https://doi.org/10.1111/1755-0998.14010.

[4] Kim, S.-R., Kwak, W., Kim, H., Caetano-Anolles, K., Kim, K.-Y., Kim, S.-B., Choi, K.-H., Kim, S.-W., Hwang, J.-S., Kim, M., et al. Genome sequence of the Japanese oak silk moth, Antheraea yamamai: the first draft genome in the family Saturniidae. GigaScience. 2018, 7(1), p. gix113. https://doi.org/10.1093/gigascience/gix113.

[5] Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017, 356(6333), pp. 92-95. https://doi.org/10.1126/science.aal3327.

[6] Wang, X., Luan, Y. and Yue, F. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. Science Advances. 2022, 8(24), p. eabn9215. https://doi.org/10.1126/sciadv.abn9215.

[7] Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. and Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature methods. 2021, 18(2), pp. 170-175. https://doi.org/10.1038/s41592-020-01056-5.

[8] Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021, 592(7856), pp. 737-746. https://doi.org/10.1038/s41586-021-03451-0.

[9] Li, H., Liu, C., Zhang, Y., Wang, X., Xiang, Z., Xia, Q., Miao, X. and Dai, F. A chromosome-level genome assembly of the Atlas moth (*Attacus atlas*) provides insights into its gigantism and conservation. Nature Ecology & Evolution. 2023, 7(9), pp. 1452-1464. https://doi.org/10.1038/s41559-023-02132-7.

[10] Van Oosterhout, C., Breden, F., Hohenlohe, P. A., Garner, B., Hand, B. K., Harrisson, K. A. and Funk, W. C. Genomic erosion in threatened species: A practical guide for conservation planners. Trends in Genetics. 2023, 39(4), pp. 281-297. https://doi.org/10.1016/j.tig.2022.12.007.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260204

The Application and Progress of Multimodal Models in the Medical Field

Qiyuan Zhang*

Zhongnan University of Economics and Law, Department of College of Information Engineering, Wuhan, China

**Corresponding author: Qiyuan Zhang*

Abstract

With the rapid development of artificial intelligence, especially generative models and multimodal large models, medical artificial intelligence has gradually moved from the early era of single-modal image recognition and text classification to the era of multimodal modeling. Medical data is inherently multimodal, including various modalities such as images, clinical texts, structured data, and genetic and signal information. For instance, in the diagnosis of lung diseases, multimodal models can integrate chest CT images with patients' electronic health records (EHR) and medical history texts to quickly locate the lesions and generate preliminary diagnostic suggestions; in the orthopedic treatment scenarios, the models can combine X-ray images with surgical record texts to assist doctors in formulating personalized surgical plans. How to effectively integrate these modalities, and conduct tasks such as diagnostic assistance, report generation, multi-round questioning, pathological explanation and reasoning, has been a focus of research in recent years. This paper systematically reviews the development path of medical multimodal models, summarizes the changes in the capabilities of mainstream methods and the limitations of datasets, and looks forward to the challenges and future trends for practical deployment in clinical settings.

Keywords

multimodal data fusion, healthcare, artificial intelligence, clinical implementation, cross-modal modeling

1. Introduction

In modern medical practice, digital diagnosis and treatment technologies have facilitated the rapid accumulation of massive heterogeneous clinical data, including imaging (such as CT, MRI, etc.), free text (medical records, consultation opinions), structured data (signs and test values in electronic health records), genomic information, and physiological signals (electrocardiograms, blood oxygen curves), etc. Each data modality carries unique clinical semantics. For instance, imaging reveals the shape of lesions, while genomic information indicates genetic characteristics. The combination of these two can overcome the limitations of individual data and provide support for precise diagnosis and personalized treatment, which is a key step in transitioning from empirical medicine to precision medicine.

However, cross-modal modeling and inter-modal reasoning have long been challenges for medical AI. Different modalities have significant differences in semantics, structure, and scale: images convey information through visual features, while texts describe diseases in natural language, and semantic systems are difficult to unify; structured data formats are well-organized, but images and texts have no fixed structure, and pre-processing methods are difficult to be universally applicable; genomic data reaches the GB level, while medical record texts are only the KB level, the disparity in data volume makes storage and computing difficult to coordinate, and these problems severely restrict the application of medical AI in cross-modal integration, and technological breakthroughs are needed to advance it.

JN Acosta pointed out that with the accumulation of biomedical data and the decline in sequencing costs, the foundation for the development of multimodal AI is gradually becoming complete. Multimodal AI has great application, challenges, and opportunities in the health field [1]. Mesko B focused on the attention brought by ChatGPT at the end of 2022 to large language models (LLMs) and the early single-modal limitations. Due to the medical multimodal characteristics, he explored the potential of LLMs to develop towards multimodality[2]. Michael Moor proposed Med-Flamingo, a multimodal few-shot learner applicable to the medical field, which solved the problem of data scarcity and took a crucial step towards the development of medical generative visual language models (VLMs) in the multi-modal direction[3]. Multimodal large language models (MLLMs) perform well in general multimodal tasks, but in the medical field, due to training costs, data, and other limitations, there are also difficulties in visual localization tasks. Jinlong He proposed the PFMVG model for this[4]. Fenglin Liu proposed a multimodal, multi-domain, and multilingual basic model suitable for solving rare or newly emerging diseases and non-English language problems[5].

“Multimodal Data Processing” as a research field can be traced back to the human-computer interaction (HCI) studies in the 1980s. After entering the medical field, the development of this area has gone through stages from early text-image matching and shallow feature fusion, to recent systematic improvements such as multimodal large language models (MLLMs), generative models (such as Diffusion), and general cross-modal agents (such as XGeM).

Although existing research has made significant progress in the technical path, there are still issues such as geographical limitations of data sets, insufficient model interpretability, and templateization of text generation. Based on this, this paper focuses on the development and application of medical multimodal models, using “fusion stage” as the core classification basis to sort out mainstream technical methods, systematically summarizing the performance characteristics, advantages and disadvantages, and data set limitations of each method, while analyzing the challenges faced in clinical implementation, and finally looking forward to future technical optimization directions, providing theoretical references for the transformation of medical multimodal models from “method innovation” to “clinical practicality”.

2. Overview of Mainstream Methods

2.1 Early Fusion

In the practice of multimodal data fusion in the medical field, early fusion is one of the core strategies, focusing on the initial stage of the data processing flow. Specifically, it involves systematic integration of heterogeneous modal data such as medical images, electronic medical records, clinical narratives, and physiological signals at the pre-stage of feature extraction and encoding. It constructs a multimodal fusion feature space to input the integrated data into a deep learning model for end-to-end learning. The core logic is to enable the model to fully explore the fine-grained correlations between cross-modal data at the early stage of information processing, accurately capture the complementary knowledge of different modalities, and provide comprehensive multimodal information support for clinical decision-making tasks such as disease diagnosis and prognosis prediction [6]. It directly concatenates the original features of different modalities (such as pixel values of images and word vectors of text) into a single vector and inputs it into a deep learning model (such as convolutional neural network CNN or fully connected network). Shared neural network layers (such as CNN or Transformer’s encoding layer) are used to uniformly encode the multimodal data and generate cross-modal feature representations.

2.2 Mid-Level Fusion

Mid-level Fusion refers to a strategy where features are independently extracted from each modality data to form an intermediate layer representation (such as high-order visual features of images, semantic vectors of text), and then fused through a cross-modal interaction mechanism. Its core lies in balancing modality specificity and correlation, supporting complex clinical tasks such as diagnostic assistance and report generation. The intermediate features of each modality are weighted and fused through an attention module (such as multi-head attention mechanism), highlighting key information. The features of different modalities are mapped to a unified semantic space (such as through fully connected layers or contrastive learning), and then fused.

2.3 Late Fusion

In the field of multimodal data fusion in healthcare, Late Fusion is a strategy where the fusion is carried out after each modality data has been independently processed and generates decision results. That is, models are first constructed for different modalities such as medical images and text medical records, and prediction results are output. Then, these results are integrated through methods such as weighted fusion and voting mechanisms to form the final decision. Its core lies in preserving the independent processing links of each modality data, reducing cross-modal interference, and simultaneously enhancing the decision robustness through the integration at the result level, providing comprehensive support based on independent judgments of multiple modalities for clinical tasks [7].

3. Challenges and Prospects

3.1 Challenges

The current application of multimodal models in the medical field has made certain progress in terms of technical paths and practical task empowerment, but still faces many global and deep-seated challenges.

3.1.1 Challenges at the Data Level

The challenges at the application level lie in the data level. The current layout and configuration of medical data are far from meeting the continuous optimization requirements of multimodal models. There are bottlenecks that urgently need to be overcome in aspects such as the balance of data sources, the completeness of sharing mechanisms, and the standardization of annotation systems. This largely restricts the further improvement and wide applicability of model performance. Although the medical field has abundant data, high-quality, large-scale, and multimodal correlated data sets are scarce. The scale of publicly available medical image data sets is much smaller than that of general data sets, and there is a significant difference in volume. The data volume and disease type distribution in different regions and medical institutions are uneven. For example, data is scarce in remote areas, and rare disease data is even more scarce, resulting in insufficient training data for model training, making it difficult to learn comprehensive features and limiting the generalization ability [8]. Multimodal data covers a large amount of sensitive information of patients, and the difficulty of privacy protection during data sharing and use is high. Once leaked, it will cause serious damage to patients' rights and interests. Moreover, multimodal data increases the risk of patient re-identification. Although privacy protection technologies are developing, they still face many challenges, such as the balance between data encryption and model training efficiency.

3.1.2 Challenges at the Model Level

In terms of the development dimension of the model itself, there is still considerable room for improvement in the core capabilities of multimodal models in medical scenarios. The interpretability issue of the model has always been a key bottleneck hindering its deep implementation in clinical practice. Medical decisions require high interpretability to gain clinical trust, but most current multimodal models operate in a "black box" manner, making it difficult to clearly explain the decision-making process and basis. When diagnosing diseases, doctors cannot understand why the model reaches a certain conclusion, making it difficult to form a reasoning logic that conforms to the rigorous requirements of the medical industry. This makes it difficult for the model to fully gain the widespread trust and recognition of medical professionals when assisting clinical decision-

making [9]. At the same time, in terms of the quality control of generated content, the model lacks stable and unified standards. Some generated results have a gap from clinical actual needs and cannot fully adapt to diverse medical application scenarios. The model's robustness is insufficient. The clinical environment is complex and variable, and data has noise and missing values. The existing models have poor adaptability to data changes and environmental interference. When encountering incomplete or interfered data cases, the model performance is likely to drop significantly, affecting the stability of diagnosis. For example, in emergency scenarios, some patient check data is missing, and the model's diagnostic accuracy is severely affected. The lack of deep multimodal fusion leads to insufficient exploration of complex and deep correlations between different modalities. In cases of inconsistent text and images, the accuracy of judgment is low. In the medical field, this is manifested as the inability to precisely integrate information when integrating data such as images, text, and genes, affecting the diagnostic accuracy [10].

3.1.3 Challenges at the Application Level

Furthermore, from the perspective of the overall application ecosystem, the integration of multimodal models with the existing medical system is not yet deep enough. There are many areas that require coordination and improvement, such as the connection with clinical work processes, the collaboration of technical applications among different medical institutions, and the compatibility of cross-scenario applications. The lack of industry standards, the absence of unified industry standards for multimodal data formats, fusion methods, and verification processes, and the difficulty in comparing and integrating models developed by different institutions all hinder the promotion and application of the technology. Each medical institution and research team develops models according to their own standards, resulting in poor interoperability between models and limiting multi-center research and large-scale clinical applications. The low clinical adaptability of the models makes it difficult to connect with existing hospital information systems (HIS), laboratory information systems (LIS), etc., and the integration with clinical actual work processes is low. This leads to inconvenience for doctors when using the models, increases the workload, and reduces the enthusiasm for applying the models. These problems collectively prevent the application effect of multimodal models in the medical field from being fully exerted and from realizing its inherent value potential.

3.2 Prospects

Looking ahead, the development prospects of multimodal models in the medical field are extremely promising, and they will undoubtedly play an increasingly important role in promoting the intelligent transformation of the medical industry. In terms of technological innovation, in the future, continuous in-depth exploration will be carried out to enhance the generalization ability of the models. By building a more comprehensive and diverse technical system, the limitations of current models in application scenarios and data dependence will be broken, enabling the models to better adapt to different medical environments and different disease types, and achieving wider clinical applicability.

In terms of optimizing the core capabilities of the model, interpretability and security will become the key research directions in the future. By continuously innovating technical methods and improving the reasoning mechanism and security guarantee system of the model, the model can not only possess efficient auxiliary diagnostic capabilities but also present the decision basis in a clearer and more understandable way, effectively enhancing the trust of medical workers in the model, and laying a solid foundation for the large-scale application of the model in clinical practice.

From the perspective of industry application, in the future, multimodal models will further integrate deeply with all aspects of the medical industry, playing a significant role in various fields such as disease diagnosis, treatment plan formulation, medical resource allocation, medical education and research. Through the collaborative development with the existing medical system, the service process will be continuously optimized, and the quality and efficiency of medical services will be improved, providing strong support for achieving higher-level medical and health services, and promoting the medical industry to move towards a more intelligent, precise and efficient direction.

4. Conclusion

This article focuses on the application and progress of multimodal models in the medical field, systematically reviewing their development path. Based on the "integration stage" as the core classification criterion, the mainstream methods are classified into three categories: early, mid-term, and late integration. The characteristics, advantages and disadvantages, as well as data set limitations of each method are analyzed. Research has shown that multimodal models can integrate various multimodal medical data such as images and text, and have demonstrated practical value in scenarios like lung disease diagnosis and orthopedic treatment. However, their clinical implementation faces multiple challenges at the levels of data, models, and applications. In the future, it is necessary to focus on the collaborative optimization of "data-model-application", to break through data bottlenecks, enhance model interpretability and security, and deepen integration with the medical system, in order to facilitate the development of the medical industry towards intelligence and precision, and provide technical support for improving medical service quality.

References

[1] Acosta, J. N., Falcone, G. J., Rajpurkar, P. and Topol, E. J. Multimodal biomedical AI. Nature Medicine. 2022, 28(9), pp. 1773-1784. https://doi.org/10.1038/s41591-022-01981-2.

[2] Meskó, B. The Impact of Multimodal Large Language Models on Health Care's Future. Journal of Medical Internet Research. 2023, 25, p. e52865. https://doi.org/10.2196/52865.

[3] Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., Zakka, C., Reis, E. P. and Rajpurkar, P., 2023. Med-Flamingo: a Multimodal Medical Few-shot Learner. *In:* Stefan, H., Antonio, P., Divya, S., et al. (eds.) *Proceedings of the 3rd Machine Learning for Health Symposium.* Proceedings of Machine Learning Research: PMLR.

[4] He, J., Li, P., Liu, G. and Zhong, S. Parameter-Efficient Fine-Tuning Medical Multimodal Large Language Models for Medical Visual Grounding. In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, 2025; pp. 1-5. https://doi.org/10.1109/ISBI60581.2025.10981029.

[5] Liu, F., Li, Z., Yin, Q., Huang, J., Luo, J., Thakur, A., Branson, K., Schwab, P., Yin, B., Wu, X., et al. A multimodal multidomain multilingual medical foundation model for zero shot clinical diagnosis. npj Digital Medicine. 2025, 8(1), p. 86. https://doi.org/10.1038/s41746-024-01339-7.

[6] Gadzicki, K., Khamsehashari, R. and Zetzsche, C. Early vs Late Fusion in Multimodal Convolutional Neural Networks. In 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 2020; pp. 1-6. https://doi.org/10.23919/FUSION45008.2020.9190246.

[7] Escalante, H. J., Hérnadez, C. A., Sucar, L. E. and Montes, M., 2008. Late fusion of heterogeneous methods for multimedia image retrieval. *Proceedings of the 1st ACM international conference on Multimedia information retrieval.* Vancouver, British Columbia, Canada: Association for Computing Machinery.

[8] Liu, F., Zhu, T., Wu, X., Yang, B., You, C., Wang, C., Lu, L., Liu, Z., Zheng, Y., Sun, X., et al. A medical multimodal large language model for future pandemics. npj Digital Medicine. 2023, 6(1), p. 226. https://doi.org/10.1038/s41746-023-00952-2.

[9] AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R. and Sheikh, J. Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook. Journal of Medical Internet Research. 2024, 26, p. e59505. https://doi.org/10.2196/59505.

[10] Sun, K., Xue, S., Sun, F., Sun, H., Luo, Y., Wang, L., Wang, S., Guo, N., Liu, L., Zhao, T., et al. Medical multimodal foundation models in clinical diagnosis and treatment: Applications, challenges, and future directions. Artificial Intelligence in Medicine. 2025, 170, p. 103265. https://doi.org/10.1016/j.artmed.2025.103265.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260105

Innovative Explorations of Microbial Synthetic Biology in Spermidine Production

Yitong Zhou*

Aulin College, Northeast Forestry University, Harbin, 150040, China

**Corresponding author: Yitong Zhou*

Abstract

Spermidine, a polyamine with bioactivities such as anti-aging and cardiovascular protection, holds broad application prospects in the fields of biomedicine and health products. It can delay the aging process by regulating cellular mechanisms and simultaneously exert a positive protective effect on cardiovascular health, thus attracting widespread attention from the industrial community. However, traditional production methods have long restricted the industrialization of spermidine. The chemical synthesis method has issues including insufficient safety of raw materials and catalysts, numerous by-products, low purity, and environmental pollution. The natural extraction method, on the other hand, results in high product costs due to the low content of active ingredients in raw materials and low extraction efficiency, making it difficult to meet market demand. The emergence of microbial synthetic biology technology has brought a new breakthrough to spermidine production. Through metabolic engineering of microbial chassis, optimization of strain performance, and combination with adaptive fermentation processes, the yield and purity of spermidine have been significantly improved. At present, relevant enterprises have promoted industrialization attempts, and costs have been reduced to a certain extent. Nevertheless, challenges still exist in aspects such as strain stability, economics of large-scale production, and industry certification. Continuous technological improvement is required in the future to promote its large-scale application.

Keywords

spermidine, microbial synthetic, biocomputing, metabolic network

1. Introduction

Spermidine is a polyamine endogenously present in human cells, which participates in fundamental physiological processes including cell proliferation and differentiation. In recent years, its potential in the biomedical field has become increasingly prominent. Endowed with core bioactivities such as anti-aging and cardiovascular protection, it is recognized within the industry as “the next 100-billion-level bioactive molecule” and has garnered substantial attention in the research and development of products for chronic disease intervention and functional restoration.

Nevertheless, literature surveys indicate that exorbitant production costs remain a critical constraint to its large-scale application. Conventional production approaches either depend on high-risk chemical feedstocks

or are constrained by the low extraction efficiency of natural raw materials. This leads to prohibitively high prices of end products, rendering them inaccessible for broader application scenarios [1].

Against this backdrop, microbial synthetic biology technology, characterized by its designability, environmental friendliness, and high efficiency, offers a novel avenue to overcome the cost bottleneck. Through the engineering of microbial metabolic pathways, the directed biosynthesis of spermidine can be achieved, circumventing the limitations of traditional processes. This also serves as the core rationale for the initiation of the present study.

2. Dilemmas in Traditional Production Methods

Preliminary research investigations indicate that traditional spermidine production relies on two primary approaches: wheat germ extraction and chemical synthesis.

The former method yields a product with a purity of merely 1% and contains gluten allergens, accompanied by high costs for separation and purification. While the latter method achieves improved purity, its production cost remains persistently high due to complex operational procedures and expensive reagents.

The global spermidine market has long been monopolized by foreign enterprises. For instance, Sigma-Aldrich prices its 100-gram reagent at over ten thousand yuan, which severely restricts the scope of its application.

3. Technological Breakthroughs in Microbial Spermidine Synthesis

3.1 Metabolic Network Reconstruction Driven by Biocomputing

Metabolic network reconstruction driven by biocomputing leverages computational tools such as bioinformatics analysis, metabolic flux simulation, and molecular structure prediction to accurately identify bottleneck nodes in the spermidine synthesis pathway. These nodes include feedback inhibition of rate-limiting enzymes, insufficient precursor supply, and accumulation of product toxicity.

Subsequently, combined with synthetic biology techniques—such as site-directed mutagenesis, gene editing, and gene tandem expression—targeted engineering is performed. Ultimately, a closed-loop metabolic system encompassing “carbon source uptake-precursor synthesis-product generation-secretion and efflux” is constructed.

The engineering work on Saccharomyces cerevisiae conducted by Qin Jiufu’s team [2]serves as a typical practice of this strategy. Its core focuses on three key modules: “enhancement of precursor supply, alleviation of feedback inhibition, and optimization of product transport”. The technical details and implementation process of each module are as follows:

3.1.1 Precursor Supply Enhancement: Systematic Optimization of the Ornithine Synthesis Pathway

Ornithine serves as the core precursor for spermidine synthesis. Specifically, ornithine is decarboxylated to form putrescine, and putrescine then combines with decarboxylated S-adenosylmethionine to generate spermidine. However, the ornithine synthesis pathway (glutamate → ornithine) in Saccharomyces cerevisiae faces two major issues: feedback inhibition of rate-limiting enzymes and insufficient metabolic flux.

The research team achieved a two-fold increase in precursor supply through a three-step strategy of “computational localization - enzyme molecular modification - gene expression regulation”.

3.1.2 Biological Computation for Bottleneck Localization: Identifying Rate-Limiting Enzymes and Regulatory Nodes

First, metabolic flux analysis (MFA) was used to track the flow of carbon source (glucose) in the yeast metabolic network. Detection results showed that the conversion rate of glutamate to ornithine was only 31%, which was much lower than the theoretical value (over 80%). Through enzyme kinetic simulation, the key rate-limiting enzyme in the pathway was identified as N-acetylglutamate kinase (ArgB). The activity of this

enzyme is strongly inhibited by the end product ornithine, with an inhibition constant Ki = 0.32 mM—meaning that even a very low concentration of ornithine can reduce ArgB activity by 50% [3].

Further transcriptome association analysis revealed that glutamate dehydrogenase (encoded by the gdh1 gene) can catalyze the production of glutamate (providing substrates for ornithine synthesis), but its natural expression level is relatively low, resulting in insufficient glutamate supply. Verification via computational tools (e.g., COBRA Toolbox metabolic network model) confirmed that the feedback inhibition of ArgB and the low expression of gdh1 are the core bottlenecks in ornithine synthesis, which require priority modification. Verification via computational tools (e.g., COBRA Toolbox metabolic network model) confirmed that the feedback inhibition of ArgB and the low expression of gdh1 are the core bottlenecks in ornithine synthesis, which require priority modification.

3.1.3 Site-Directed Saturation Mutagenesis for ArgB Modification: Relieving Feedback Inhibition and Enhancing Enzyme Activity

To address the feedback inhibition of ArgB, site-directed saturation mutagenesis was applied to modify the enzyme's "allosteric regulatory domain"—a key region where ornithine binds and inhibits ArgB activity. Mutating critical amino acids in this domain can reduce the enzyme's affinity for ornithine while preserving or even enhancing its catalytic activity. The specific implementation steps are as follows:

First, homology modeling and molecular docking (using the AutoDock Vina tool) were employed to simulate the binding process between ornithine and ArgB. The results revealed that V126 (valine) and F189 (phenylalanine) in ArgB's allosteric regulatory domain are key ornithine-binding sites; the side chains of these two amino acids fix ornithine via hydrophobic interactions and hydrogen bonds, triggering conformational changes in the enzyme and thus inhibiting its activity.

Next, saturation mutagenesis primers were designed and a mutant library was constructed. For the V126 and F189 sites, primers containing degenerate codons were designed (e.g., the "NNS" degenerate codon was used for the V126 site, which can encode 20 natural amino acids, and the same for the F189 site) to ensure all possible amino acid mutations were covered at each site. Using the Saccharomyces cerevisiae argB gene as a template, PCR amplification was performed with the degenerate primers to obtain argB fragments containing the mutation sites. These fragments were then inserted into an expression vector (e.g., pRS426 plasmid) carrying the strong promoter Ptef1 to construct an ArgB mutant library.

Subsequently, the mutant library was transformed into an ArgB-deficient yeast strain (ΔargB), and target mutants were screened through "ornithine-deficient medium screening + enzyme activity determination". In the first step, only strains with normal ArgB activity could grow on medium without ornithine, enabling the initial elimination of inactive mutants. In the second step, ArgB activity in the surviving strains was detected using an enzyme activity assay kit, and the inhibition rate of ornithine on the enzyme was measured via high-performance liquid chromatography (HPLC). Finally, the "V126A/F189L double mutant" was screened out.

For the verification of mutant efficacy, results showed that after modification, the feedback inhibition coefficient of ArgB toward ornithine decreased from 0.32 to 0.07—meaning a much higher concentration of ornithine is required to inhibit enzyme activity, and the inhibitory effect is reduced by over 80%. Additionally, the specific enzyme activity (catalytic efficiency per unit mass of enzyme) of the mutant ArgB increased by 1.8 times, indicating that the same amount of enzyme can catalyze more glutamate to produce ornithine.

3.1.4 Gene Tandem Expression: Enhancing the Conversion Rate of Glutamate to Ornithine

To address the issue of insufficient glutamate supply, the "mutant argB gene" and "glutamate dehydrogenase gene (gdh1)" were subjected to tandem expression with the following specific design: the constitutive promoter Ptef1— the one with the highest activity in Saccharomyces cerevisiae and the promoter of translation elongation factor 1α— was selected to ensure efficient transcription of both genes; the mutant argB and gdh1 were tandemly linked via Overlap Extension PCR without inserting a terminator in between, which ensures both genes are transcribed on the same mRNA; the tandem gene fragment was inserted into a yeast integrative vector (e.g., pUG6) and integrated into the rDNA locus (a multi-copy site enabling high-copy gene expression) of the yeast genome through homologous recombination. After modification, the intracellular conversion rate of glutamate to ornithine increased from 31% to 68%, and metabolic flux analysis showed that

the flux of the ornithine synthesis pathway reached 12.6 mmol/(g·h), providing an adequate precursor for spermidine synthesis [4].

3.2 Multi-level Relief of Feedback Inhibition: Breaking the "Product Inhibition Curse" in Spermidine Synthesis

In the spermidine synthesis pathway, in addition to the feedback inhibition from the precursor ornithine, the end product spermidine inhibits synthesis through two mechanisms: ① directly inhibiting the key enzyme S-adenosylmethionine decarboxylase (SpeD); ② forming "metabolic coupling inhibition" with SAM synthetase (Sam2) (SpeD activity depends on S-adenosylmethionine (SAM) generated by Sam2, while spermidine inhibits both simultaneously). The research team completely relieved this inhibition through a multi-level strategy combining "molecular structure simulation-guided mutagenesis and epigenetic regulation". The specific process is as follows:

Bioinformatics calculations identified the inhibition sites of SpeD. The three-dimensional structure of SpeD was resolved using X-ray Crystallography (with a resolution of 2.3 Å), and combined with molecular dynamics simulations (using GROMACS software), the binding mechanism between spermidine and SpeD was clarified: spermidine binds to the "pocket adjacent to the active center" of SpeD through hydrogen bonds and electrostatic interactions, causing conformational changes in the active center that prevent substrate SAM from binding. Y213 (tyrosine) and R267 (arginine) within the pocket are key binding sites — the hydroxyl group of Y213 forms a hydrogen bond with the amino group of spermidine, and the guanidino group of R267 forms an electrostatic interaction with the carboxyl group of spermidine.

Alanine Scanning Mutagenesis was used to screen for inhibition-resistant mutants: this technique replaces candidate sites such as Y213 and R267 one by one with alanine (which has a methyl side chain and no hydrogen bond/electrostatic interaction capabilities) to evaluate the impact of mutations on enzyme activity and inhibition sensitivity. Mutant construction: Site-directed mutagenesis primers were designed for each site, PCR amplified the speD mutant genes, and inserted into expression vectors containing the Ptef1 promoter.

Functional verification: Enzyme activities of each mutant were measured under different spermidine concentrations (0.1-5 mM): the wild-type SpeD showed 50% enzyme activity reduction at 0.15 mM spermidine (Ki=0.15 mM); the Y213A mutant (Y213 replaced with alanine) had a Ki increased to 2.3 mM (requiring 15-fold higher spermidine concentration for enzyme inhibition) while retaining 92% of its enzyme activity (not significantly inactivated by the mutation); although the R267A mutant also relieved inhibition, it only retained 40% of enzyme activity. Therefore, the Y213A mutant was ultimately selected [5].

3.3 CRISPR-dCas9 Epigenetic Regulation: Relieving Metabolic Coupling Between SpeD and Sam2

S-adenosylmethionine synthetase (Sam2) is responsible for generating S-adenosylmethionine (SAM), the substrate of SpeD. However, spermidine inhibits the transcription of the sam2 gene through an "unknown regulatory pathway", leading to insufficient SAM supply and subsequent indirect inhibition of SpeD activity (metabolic coupling inhibition). The research team employed the CRISPR-dCas9 system (inactivated Cas9 with no DNA cleavage activity, serving only as a "targeting vector") for epigenetic regulation to relieve this coupling, with details as follows:

3.3.1 Target Selection: Promoter Regulatory Region of the Sam2 Gene

Chromatin Immunoprecipitation Sequencing (ChIP-seq) revealed that spermidine induces transcriptional repressors (e.g., Mig1) to bind to the -300~-200 bp region of the sam2 promoter, thereby inhibiting transcription. Consequently, this region was selected as the targeting site for dCas9.

3.3.2 Construction of the CRISPR-dCas9 Activation System

sgRNA design: Based on the targeting region of the sam2 promoter, 2 sgRNAs (to ensure targeting specificity) were designed and cloned into a vector containing the U6 promoter (a dedicated promoter for sgRNA).

dCas9 fused with activator: dCas9 was fused with a transcriptional activator of Saccharomyces cerevisiae (e.g., Gal4 activation domain) to construct a "dCas9-activator" expression cassette. This cassette is controlled by the Pgal1 promoter and can be induced to express by galactose.

Transformation and induction: The sgRNA vector and the dCas9-activator vector were co-transformed into yeast strains. After adding galactose to induce expression, the dCas9-activator binds to the inhibitory region of the sam2 promoter in a targeted manner, competitively displacing transcriptional repressors and recruiting transcriptional machinery (e.g., RNA polymerase) to promote sam2 transcription.

3.3.3 Effect of Modification

The mRNA expression level of the sam2 gene increased by 2.5-fold, and the production rate of decarboxylated SAM (the substrate of SpeD) increased by 2.1-fold. When the intracellular spermidine concentration reached 1.8 g/L (far exceeding the inhibition threshold of 1.2 g/L in traditional processes), no synthesis stagnation was observed, completely relieving the metabolic coupling inhibition.

3.4 Enhancement of Product Transport System: Constructing a "Synthesis-Transport Coupling" Module to Alleviate Product Toxicity

Intracellular accumulation of spermidine exerts toxicity on yeast cells (e.g., disrupting cell membrane integrity and inhibiting the respiratory chain). In traditional processes, when the intracellular concentration exceeds 1.2 g/L, the cell growth rate decreases by 50%. The research team achieved efficient extracellular secretion of spermidine through the strategy of "transporter screening - signal peptide optimization - coupling module design", with the specific process as follows:

3.4.1 Screening of Specific Transporters via Comparative Transcriptomics

Transporters associated with spermidine synthesis were screened using comparative transcriptomic analysis (RNA-seq): the experimental group was set as a "high-spermidine-producing strain" (with optimized precursors and relieved feedback inhibition) and the control group as a wild-type strain; transcriptomic data of the two groups were analyzed by the DESeq2 tool, which revealed that the mRNA levels of three transporters (Tpo2, Tpo3, Qdr2) were significantly upregulated in the experimental group, with Pearson correlation coefficients of 0.89, 0.76, and 0.68 respectively, highly consistent with the expression trend of the spermidine synthesis gene speE; for functional verification, overexpression strains of the three transporters were constructed by inserting the tpo2, tpo3, and qdr2 genes into vectors containing the Ptef1 promoter respectively, and the spermidine concentration in the fermentation supernatant was detected via HPLC— the Tpo2-overexpressing strain showed an extracellular secretion level 4.2 times that of the control group and the highest transport specificity for spermidine (its transport efficiency was 2.7 times that of putrescine and 1.3 times that of spermine), thus confirming Tpo2 as an "efficient and specific spermidine efflux transporter".

3.4.2 Signal Peptide Optimization: Enhancing the Secretion Efficiency of Tpo2

Tpo2 is a membrane protein, and its N-terminal native signal peptide guides the protein to localize to the cell membrane but with relatively low efficiency. To optimize its localization efficiency, the research team replaced the native signal peptide: they selected the α-factor signal peptide— which has the highest secretion efficiency in Saccharomyces cerevisiae, is naturally used for secretory proteins such as invertase, and can efficiently guide proteins to the cell membrane via the endoplasmic reticulum-Golgi pathway— then replaced Tpo2's native signal peptide with this 19-amino-acid α-factor signal peptide by PCR to construct the "α-factor signal peptide-Tpo2" fusion gene; efficacy verification via fluorescent labeling (GFP fusion) showed that after optimization, Tpo2's localization efficiency on the cell membrane increased by 2.3-fold, the extracellular secretion rate of spermidine increased by 3.7-fold, and the intracellular accumulation decreased by 42%.

3.4.3 Design of the "Synthesis-Transport Coupling" Module: Achieving Dynamic Balance

To avoid "mismatch between synthesis rate and transport rate" (e.g., intracellular accumulation caused by faster synthesis than transport, or raw material waste caused by faster transport than synthesis), the research team designed a gene tandem module mediated by a self-cleaving peptide. The module structure involves tandemly connecting the spermidine synthase gene (speE, responsible for spermidine production) with the optimized tpo2 gene via the P2A self-cleaving peptide (derived from foot-and-mouth disease virus, which

automatically cleaves into two independent proteins after translation) to form the "speE-P2A-tpo2" cassette. For promoter control, the inducible promoter Pgal10 (repressed by glucose and induced by galactose) is used to ensure the module is only activated during the spermidine synthesis phase. As for the dynamic balance mechanism, when the spermidine synthesis rate increases (with enhanced speE expression), tpo2 achieves "equivalent expression" through the synchronous translation and cleavage of P2A, and the transport system is activated simultaneously to ensure timely extracellular efflux of the product; conversely, when the synthesis rate decreases, the transport rate also decreases accordingly to avoid energy waste.

3.5 Verification of Synergistic Effects: Testing the Industrial Potential in a 5L Fermenter

After modifying the above three modules (precursor optimization, feedback derepression, and transport enhancement), the research team verified their synergistic effects through 5L fed-batch fermentation. The specific parameters and results are as follows: For fermentation conditions, glucose was used as the sole carbon source (initial concentration of 40 g/L, maintained at 5-10 g/L via feeding), pH was controlled at 6.0, dissolved oxygen was maintained at 30%, and the fermentation cycle was 36 hours. Key results showed that the spermidine yield reached 2.3 g/L, the raw material cost was reduced by 68% compared with the traditional process (which requires exogenous addition of methionine and putrescine), and the product yield was increased to 0.057 g/g glucose. The core innovation lies in the construction of a closed-loop metabolic network of "carbon source uptake → glutamic acid → ornithine → spermidine → extracellular secretion"; all synthesis units rely on the yeast's own metabolism without any exogenous precursors, completely solving the cost and efficiency problems of the traditional process.

4. Discovery of the Extracellular Secretion Mechanism

In the innovation of the spermidine production process, the Qin Jiufu team not only achieved a breakthrough in metabolic network reconstruction but also accomplished a leap from the traditional model to an innovative system in the extraction and purification stage. Through bioinformatics-driven mining and application of transporters, the team transformed the spermidine extraction model from "juice squeezing"-like cell disruption extraction to "tap water collection"-like continuous secretion collection. This revolutionized the downstream process and brought comprehensive optimizations in impurity removal rate, production cost, and production efficiency.

In the stage of bioinformatics screening and functional verification of transporters, the team established a multi-dimensional screening system. Based on the Saccharomyces cerevisiae genome database, 23 candidate members belonging to the polyamine transporter family (PAT) were first screened from 127 potential transporters through sequence homology analysis (with an E-value threshold set at 1e-5). Subsequently, using co-expression analysis of transcriptome data, it was found that when the expression of the spermidine synthesis gene speE was upregulated, the mRNA levels of three transporters—Tpo2, Tpo3, and Qdr2—increased synchronously (Pearson correlation coefficients were 0.89, 0.76, and 0.68, respectively). To verify their functions, the team constructed overexpression strains of these three genes and detected the spermidine concentration in the fermentation supernatant via high-performance liquid chromatography (HPLC). The results showed that the extracellular secretion of the Tpo2-overexpressing strain was the highest, reaching 4.2 times that of the control group. Moreover, its transport specificity experiment indicated that the transport efficiency of this protein for spermidine was 2.7 times that of putrescine and 1.3 times that of spermine, confirming it as a high-efficiency and specific spermidine efflux transporter.

This innovation in the extraction model brought significant process advantages. The traditional "juice squeezing"-like process requires cell disruption via high-pressure homogenization or ultrasonic treatment (with a disruption rate of approximately 85%). While releasing intracellular products, it also introduces impurities such as cell debris, nucleic acids, and proteins into the extract, significantly increasing the load of subsequent purification steps (e.g., ion exchange chromatography, gel filtration). Comparative data showed that the concentration of protein impurities in the extract of the traditional process was as high as 12.6 g/L, whereas after adopting the Tpo2-mediated "tap water collection" model, the concentration of protein impurities in the fermentation supernatant was only 0.8 g/L, representing a 93.6% increase in impurity removal rate. This change reduced the number of purification steps from 4 to 2 (only requiring nanofiltration decolorization and

crystallization), which not only shortened the process cycle (from 24 hours to 8 hours) but also reduced the purification cost by 62%.

In addition, the continuous secretion model also achieved a leap in production efficiency. The traditional process is limited by the cell growth cycle and requires centralized harvesting when cells reach the stationary phase, with the problem of product accumulation inhibition (when the intracellular spermidine concentration exceeds 1.2 g/L, the cell growth rate decreases by 50%). In contrast, the "tap water collection" model maintains the intracellular spermidine concentration stably at 0.3–0.5 g/L through real-time product efflux, relieving the inhibitory effect. Cells can maintain the metabolic activity of the logarithmic growth phase, enabling continuous fed-batch fermentation. Practical operation data showed that the fermentation productivity under this model reached 0.064 g/(L·h), which was 106% higher than the 0.031 g/(L·h) of traditional batch fermentation. Furthermore, the fermentation cycle could be extended to 72 hours without a significant decrease in productivity, significantly improving equipment utilization.

The innovation of the Qin Jiufu team is not only a breakthrough at the technical level but also establishes a green production model integrating "synthesis-secretion-purification". By combining bioinformatics screening with metabolic engineering modification, the team realized the transformation of products from intracellular accumulation to continuous extracellular secretion. This provides a highly valuable reference paradigm for the extraction processes of other microbial secondary metabolites, especially demonstrating great application potential in reducing industrial production costs and improving product purity [5].

5. Conclusion

In summary, spermidine, as a natural polyamine in the human body, has become a highly promising "100-billion-level bioactive molecule" in the biomedical field due to its prominent activities in anti-aging, cardiovascular protection, and other areas. It also demonstrates significant value in the research and development of chronic disease intervention and functional repair products. However, its large-scale application has long been hindered by the drawbacks of traditional production methods—high-risk chemical synthesis and low-efficiency natural extraction together drive up costs, making it difficult for products to achieve widespread use.

Microbial synthetic biology technology, with the advantages of designability, environmental friendliness, and high efficiency, enables the targeted synthesis of spermidine by modifying microbial metabolic pathways. This effectively avoids the shortcomings of traditional processes and provides a key solution to break through the cost bottleneck. This technical route not only meets the industrialization needs of spermidine but also confirms the necessity of this study starting from "solving application pain points," laying a core direction for subsequent research on large-scale production and application of spermidine.

References

[1] Mohaqiq, M., Palou, R., Li, R., Zhang, G., Vaishnav, H., Parweez, F., Moyana, T., Carragher, D., Cameron, D. W., Ramsay, T., et al. Engineered spermidine-secreting *Saccharomyces boulardii* ameliorates colitis and colon cancer in mice. Scientific Reports. 2025, 15(1), p. 31959. https://doi.org/10.1038/s41598-025-16736-5.

[2] University, S. Sichuan University's achievement: Synthetic biology enables large-scale production of the "longevity molecule" spermidine. Available from: https://reportify.cn/social-media/692823570141898 (accessed 8 January 2026).

[3] Gou, L., Wang, J., Liu, D., Wu, S., Zhou, X., Fan, T.-P. and Cai, Y. Multi-strategy metabolic reprogramming of *Serratia marcescens* HBQA7 for high-efficiency spermidine biosynthesis via the carboxyaminopropylagmatine pathway. Chemical Engineering Journal. 2025, 522, p. 167413. https://doi.org/https://doi.org/10.1016/j.cej.2025.167413.

[4] Zhao, Z., Zou, D., Ji, A., Wu, Y., Guo, A. and Wei, X. Rational design of S-adenosylmethionine decarboxylase SpeD and spermidine synthase SpeE for green synthesis of spermidine. International

Journal of Biological Macromolecules. 2025, 316, p. 144680. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2025.144680.

[5] Zhao, Z., Wu, Y., Fan, S., Li, Z., Zou, D., Guo, A. and Wei, X. Biosynthesis of the functional component spermidine from *Bacillus amyloliquefaciens* by iterative integration expression. ACS Synthetic Biology. 2025, 14(5), pp. 1745-1755. https://doi.org/10.1021/acssynbio.5c00087.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260106

Comparative Study of Early Diabetes Risk Stratification Based on Machine Learning Algorithms

Tianze Zhang*

School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

**Corresponding author: Tianze Zhang*

Abstract

Objective: This study aims to systematically compare the performance of multiple machine learning models in diabetes risk prediction and identify key risk factors, thereby providing data-driven decision support for early diabetes screening. **Methods:** Using the UCI Pima Indians Diabetes dataset, five models-logistic regression, K-nearest neighbors, support vector machine, decision tree, and random forest-were trained and evaluated. Model performance was comprehensively assessed via metrics including AUC-ROC, precision, and recall, with feature importance analysis employed to elucidate core diabetes risk factors. **Results:** The random forest model demonstrated superior performance across multiple metrics (AUC = 0.8167). Plasma glucose was consistently identified as the strongest predictor, with body mass index (BMI) and age also emerging as significant contributors. **Conclusion:** The random forest model exhibits robust performance and effective capture of feature interactions, making it well-suited for early diabetes prediction with considerable potential for clinical application.

Keywords

diabetes prediction, machine learning, feature importance, random forest, model comparison

1. Introduction

1.1 Diabetes Background

Diabetes, a chronic metabolic disorder affecting populations worldwide, impacted 537 million adults in 2021 according to the International Diabetes Federation (IDF), with projections indicating a rise to 783 million by 2045 [1]. Its complications include cardiovascular disease, retinopathy, and other conditions. Early prediction can reduce healthcare costs by more than 30% (WHO, 2022) [2].

1.2 Limitations of Traditional Prediction Methods

Current clinical standards, such as the oral glucose tolerance test (OGTT) and HbA1c measurement, have notable drawbacks: invasive procedures result in low patient adherence [3]; traditional risk assessment tools (e.g., the FINDRISC questionnaire) achieve accuracy rates of only 65–70% [4]; and these approaches fail to account for nonlinear feature interactions (e.g., the synergistic effect between BMI and blood pressure).

1.3 Application Potential of Machine Learning

Supervised learning algorithms have demonstrated substantial potential in medical prediction tasks, with advantages in three key areas: effective handling of high-dimensional clinical data (e.g., the eight heterogeneous features in the Pima dataset), capability to capture complex nonlinear relationships and feature interactions (visualizable through advanced interpretability methods such as SHAP), and superior predictive performance in comparable studies-for instance, the XGBoost model achieved an AUC of 0.89 in Alghamdi et al. (2022), underscoring the practical value of machine learning in diabetes risk prediction.

2. Dataset Description

2.1 Data Source

The Pima Indians Diabetes Database from the UCI Machine Learning Repository was utilized [5]. The data were collected from Pima Indian heritage women aged 21 years and older in the Phoenix area of Arizona, United States. This population exhibits a type 2 diabetes incidence rate exceeding the global average by more than 50% (NIH, 2019).

2.2 Samples and Features

The dataset [6] includes 768 samples, comprising 268 individuals with diabetes (positive class, 34.9%) and 500 without (negative class). The diabetes prediction models were constructed using eight clinical features: number of pregnancies (Pregnancies, reflecting hormone-related risk in females), plasma glucose concentration during an oral glucose tolerance test (Glucose), diastolic blood pressure (BloodPressure), triceps skinfold thickness (SkinThickness, indicative of body fat percentage), 2-hour serum insulin level (Insulin), body mass index (BMI), diabetes pedigree function (DiabetesPedigreeFunction, quantifying familial genetic risk), and age (Age). The target variable, Outcome, is binary and labeled according to WHO diagnostic criteria (fasting plasma glucose $\geqslant$ 7.0 mmol/L or 2-hour OGTT glucose $\geqslant$ 11.1 mmol/L), with 1 denoting "diagnosed with diabetes within 5 years" and 0 denoting "no diagnosis."

2.3 Key Feature Distribution Description

Statistical analysis of the Pima Indians Diabetes dataset reveals substantial heterogeneity in demographic, physiological, and metabolic characteristics. Demographically, the number of pregnancies exhibits a right-skewed distribution (skewness = 0.90), with 75% of individuals reporting six or fewer pregnancies. Age displays a distinctive bimodal distribution, with a primary peak in the 24–29-year reproductive age group and a secondary peak in the 41–45-year high-risk segment. Among key physiological indicators, plasma glucose concentration is markedly higher in the positive group (mean = 141.26 mg/dL) than in the negative group (mean = 109.98 mg/dL), with Kolmogorov–Smirnov test confirming significant differences in distribution ($D = 0.46$, $p < 1e\text{-}16$). BMI analysis indicates a significantly higher obesity prevalence in the positive group (62.3%) compared to the negative group (38.7%) ($\chi^2 = 34.21$, $p = 4.5e\text{-}9$). Metabolically, insulin levels are exponentially distributed (skewness = 2.51), with the majority of samples (68.2%) below 100 μU/mL. Skin thickness is strongly correlated with BMI ($r = 0.54$). Notably, blood pressure (24.3%) and skin thickness (29.1%) contain substantial proportions of zero values. These feature patterns provide critical data foundations and quality considerations for subsequent machine learning modeling.

Table 1: Feature Correlation Analysis

Feature Pair	Correlation Coefficient	Clinical Significance
Age – Pregnancies	**0.54**	Reflects cumulative reproductive effect
Glucose – Outcome	**0.47**	Core diagnostic indicator
SkinThickness – BMI	**0.54**	Markers of body fat percentage
Insulin – SkinThickness	**0.44**	Indicator of insulin resistance

Table 2: Descriptive Statistics of Features in the Diabetes Dataset

Feature	Mean ± SD	Minimum	25th Percentile	Median	75th Percentile	Maximum
Pregnancies	3.85 ± 3.37	0	1	3	6	17
Glucose	120.89 ± 31.97	0	99	117	140.25	199
BloodPressure	69.11 ± 19.36	0	62	72	80	122
SkinThickness	20.54 ± 15.95	0	0	23	32	99
Insulin	79.80±115.24	0	0	30.5	127.25	846
BMI	31.99+7.88	0	27.3	32	36.6	67.1
DiabetesPedigreeFunction	0.47 ± 0.33	0.078	0.244	0.372	0.626	2.42
Age	33.24 ± 11.76	21	24	29	41	81

The feature distribution analysis of the Pima Indians Diabetes dataset (Figure 1) shows distinct distributional characteristics and clinical implications across the eight key features. Pregnancies, insulin levels, and DiabetesPedigreeFunction exhibit pronounced right-skewed distributions, indicating that most samples cluster at lower values with a minority of high-value outliers. Glucose concentration approximates a normal distribution with a peak in the 100–125 mg/dL range-close to the prediabetes diagnostic threshold-thus carrying significant clinical warning value. Blood pressure and BMI display relatively uniform distributions, reflecting the continuous nature of these physiological parameters in the population. Age is predominantly distributed between 20 and 40 years with a right-skewed pattern, consistent with the study's focus on women aged 21 and older. Notably, skin thickness shows a high concentration of values in the 0–40 mm range with many zero entries, providing clear guidance for subsequent data cleaning and outlier handling. These distributional patterns not only highlight the unique characteristics of the study population but also inform critical decisions in feature engineering and model selection: right-skewed features may require transformation, zero values necessitate appropriate imputation strategies, and near-normally distributed features are well-suited for direct inclusion in linear modeling frameworks.

Figure 1: Histograms of Feature Distributions

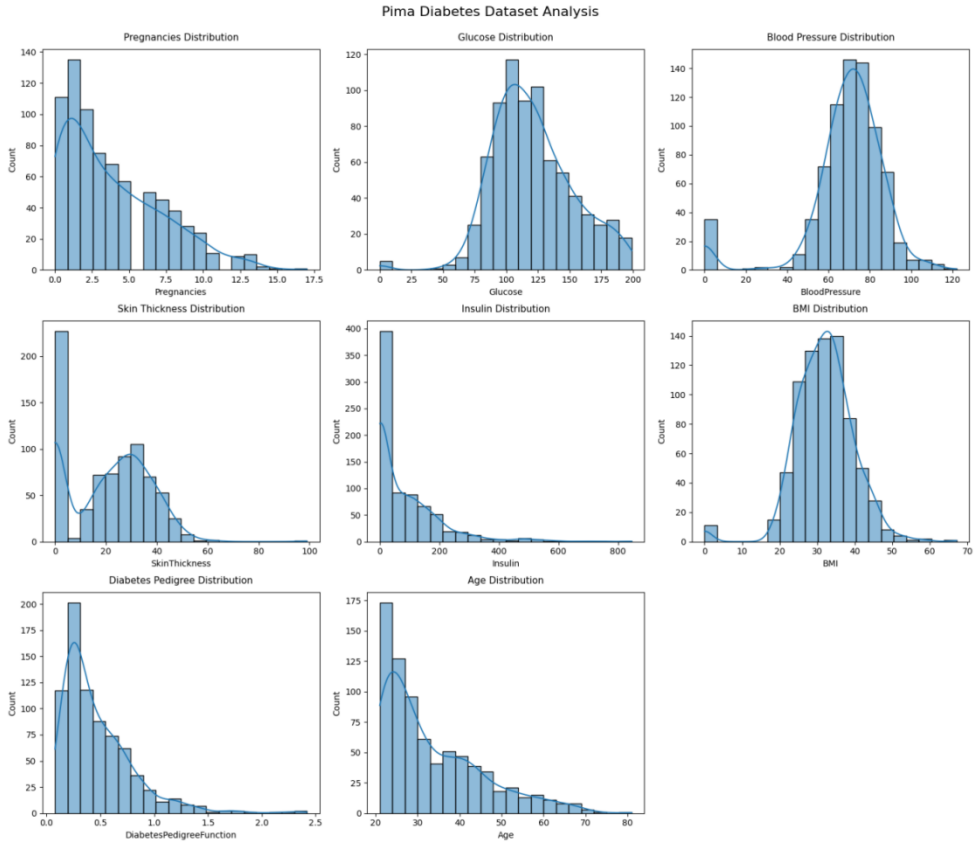


Figure 2: Bar Chart of the Binary Distribution of the Target Variable

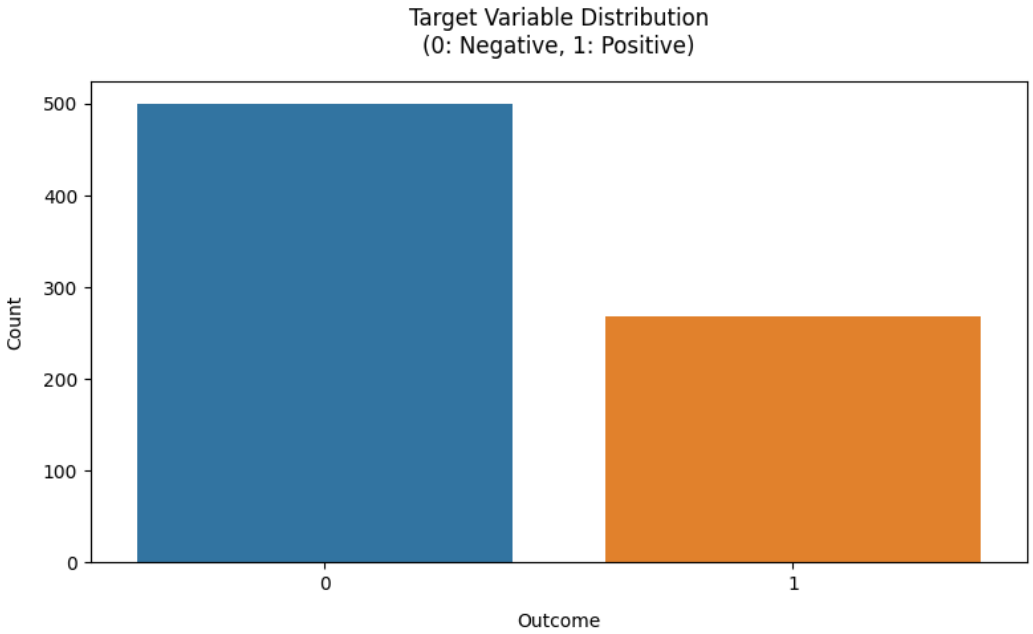
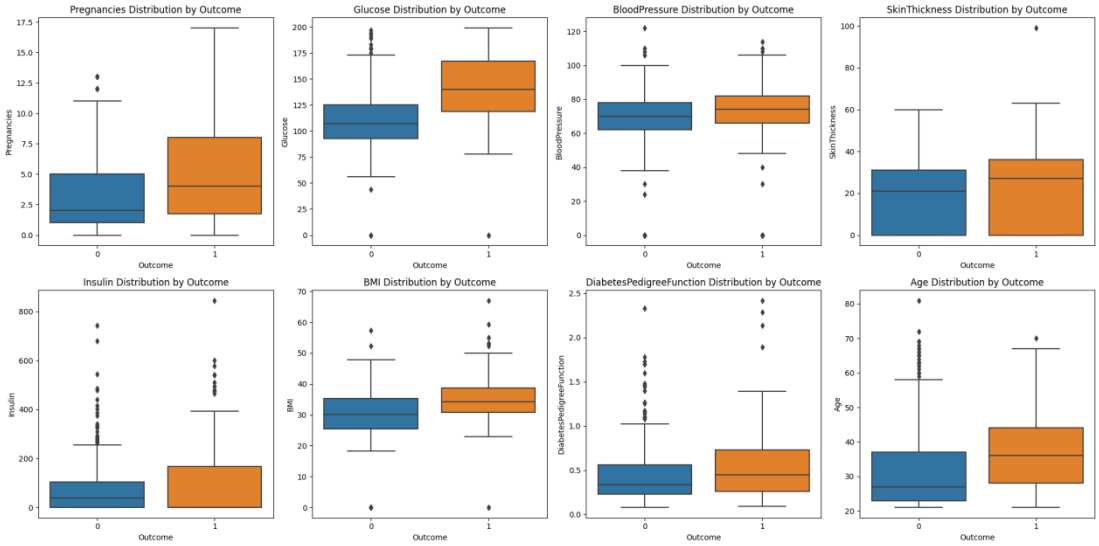


Figure 3: Boxplots of Features Stratified by Target Variable



(Illustrating Distributional Differences Across Groups)

The figures presents boxplots illustrating the distributional differences in the eight key features of the diabetes dataset-Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, and Age-between the non-diabetic (Outcome = 0) and diabetic (Outcome = 1) groups. The results show that Glucose exhibits the most pronounced association with diabetes, with a markedly higher median and overall level in the diabetic group than in the non-diabetic group. Additionally, features such as BMI, Age, and Pregnancies generally display higher medians or broader ranges in the diabetic cohort. Although Insulin shows distributional differences between groups, its variability is relatively high. These inter-group differences visually highlight the associations between physiological indicators and diabetic status, providing a data-driven basis for constructing subsequent diabetes prediction models.

Figure 4: Correlation Heatmap

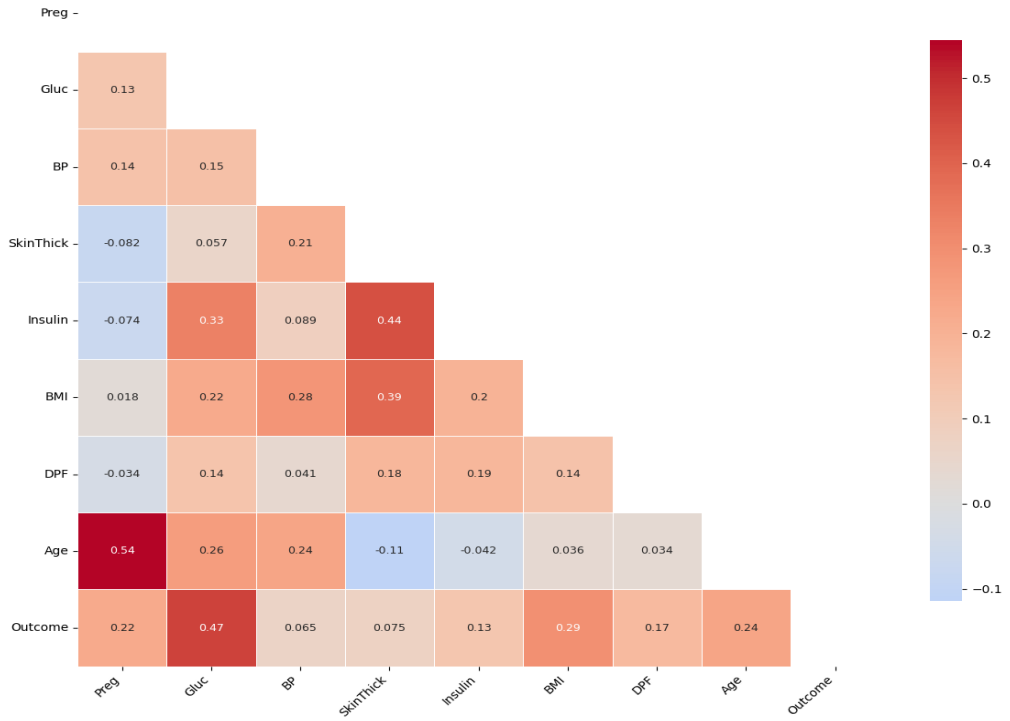


Figure 4 displays the correlation heatmap for the diabetes dataset, where the intensity of color and numerical annotations represent the Pearson correlation coefficients between each feature (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age) and the target variable (Outcome: presence or absence of diabetes). The results reveal that Glucose exhibits the strongest correlation with Outcome (r = 0.47), followed by BMI (r = 0.29) and Age (r = 0.24), all showing positive associations. Significant inter-feature correlations include Insulin with SkinThickness (r = 0.44), BMI with SkinThickness (r = 0.39), and Age with Pregnancies (r = 0.54), reflecting the intrinsic physiological relationships between features. Overall, most features demonstrate low-to-moderate positive correlations with Outcome, providing a correlation-based foundation for subsequent feature selection and model development in diabetes prediction.

3. Data Preprocessing

3.1 Missing Value Handling

Missing values in the dataset are encoded as "0," predominantly within physiological features. Statistical analysis identified the following features as containing implausible zero values (inconsistent with medical knowledge, as physiological measures such as glucose and blood pressure cannot be zero): Glucose, BloodPressure, SkinThickness, Insulin, and BMI. Identification Method: Each of the aforementioned feature columns was scanned to count the occurrences of zero values (results summarized in Table 3).

Table 3: Summary of Implausible Zero Values in Features

Feature	Number of Zero Values
Glucose	5
BloodPressure	35
SkinThickness	227
Insulin	374
BMI	11

Imputation Strategy: All zero values in the aforementioned features were uniformly replaced with **NaN** (Not a Number) to explicitly designate them as missing, thereby preventing their misinterpretation as valid physiological measurements during modeling.

Missing Value Imputation Strategy and Implementation Method: Median imputation using SimpleImputer (strategy='median'). Rationale: The median is robust to outliers, which makes it suitable for physiological variables with extreme values (e.g., Insulin exhibits strong right-skewness, where the mean is heavily influenced by outliers); it preserves central tendency, is computationally efficient, and is well-suited for medium-sized datasets; it aligns with standard preprocessing practices in medical data analysis and avoids sample loss due to missingness. Procedure: The imputer was fitted exclusively on the training set (imputer. fit (X_train [physiological_features])) to prevent information leakage from the test set. Imputation was applied separately to both training and test sets (imputer.transform(X_train) and imputer.transform(X_test)), ensuring distributional consistency.

3.2 Dataset Splitting

Splitting Method Stratified sampling was employed (implemented via train_test_split with default stratification) to divide the dataset into training and test sets. Parameters:

Test set proportion: test_size=0.2 (80% training, 20% testing) ; Random seed: random_state=42 for reproducibility, Rationale: An 80:20 split balances training effectiveness and testing reliability in a small dataset (n = 768); a fixed random seed ensures experimental reproducibility, aiding model tuning and comparison; stratified sampling preserves the proportional distribution of the target variable (Outcome) across both subsets, reducing sampling bias.

3.3 Feature Scaling

Scaling Method: Standardization (StandardScaler) was used to transform features into a distribution with a mean of 0 and a standard deviation of 1, using the formula:

$$(x_{\text{scaled}}=\frac{x-\mu}{\sigma}) \tag{1}$$

where (μ) is the feature mean and (σ) is the feature standard deviation.

Rationale: Standardization is well-suited for approximately normally distributed features (e.g., Glucose, BMI), enhancing convergence speed and accuracy in linear models (e.g., logistic regression, SVM); it preserves the shape of the feature distribution and outlier information, consistent with the variability of physiological indicators; and it ensures compatibility with distance-based models (e.g., KNN) by preventing features with disparate scales (e.g., Age ranging from 0–100 vs. DiabetesPedigreeFunction ranging from 0–2.42) from dominating model decisions.

Procedure: The scaler was fitted exclusively on the training set (scaler.fit(X_train)) to compute the mean((μ)) and standard deviation((σ))from training data only. The training set mean ((μ))and standard deviation ((σ))were used to scale both the training and test sets (scaler.transform(X_train) and scaler.transform(X_test)). Key Principle: The test set must not be used to fit the scaler to prevent data leakage, as it represents “unseen” data, and its distribution must not influence the training process.

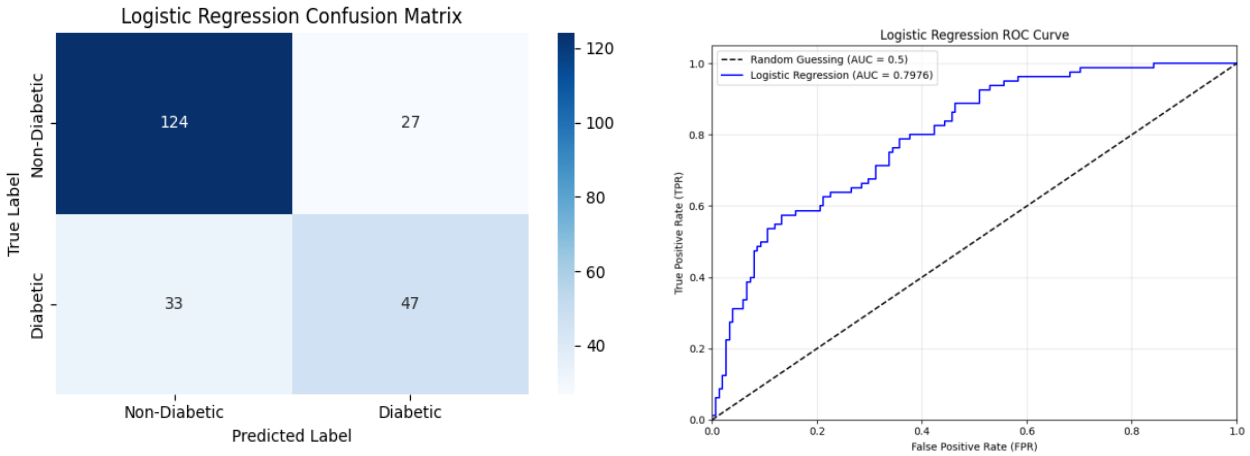
4. Model Selection Rationale

4.1 Logistic Regression Model

4.1.1 Confusion Matrix Results

The confusion matrix for the logistic regression model shows that, among samples with a true label of non-diabetes, the model correctly predicted non-diabetes in 124 cases and incorrectly predicted diabetes in 27 cases. Among samples with a true label of diabetes, the model incorrectly predicted non-diabetes in 33 cases and correctly predicted diabetes in 47 cases.

Figure 5: Confusion Matrix and ROC Curve for Logistic Regression



4.1.2 Performance Interpretation

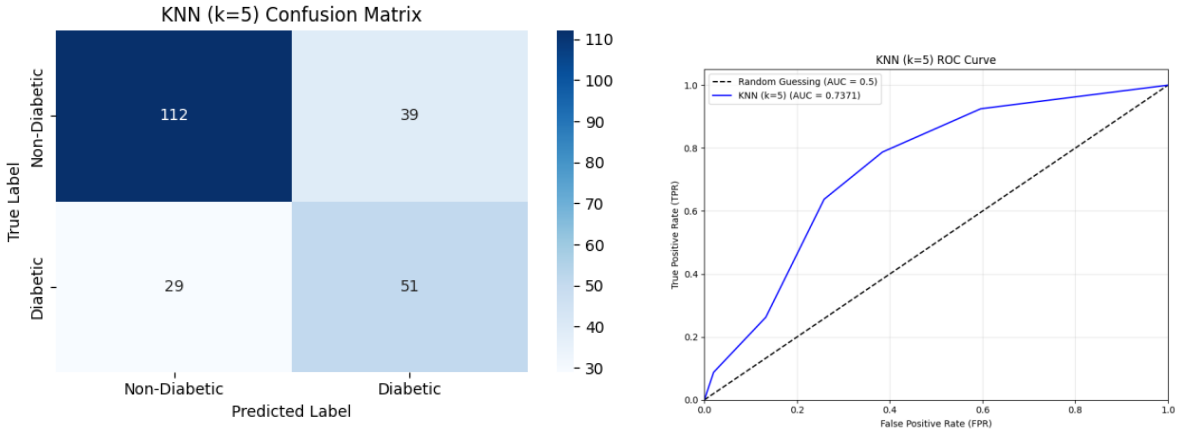
The ROC curve for the logistic regression model lies well above the random-guessing baseline, with an **AUC of 0.7975**. The curve rises steeply at low false positive rates, indicating strong discriminatory power, particularly for non-diabetic samples (evidenced by the high number of correct non-diabetes predictions). However, the curve flattens in later segments, and the classification report reveals lower precision and recall for the diabetic class. This suggests frequent misclassification of diabetic cases, likely due to weak linear separability between diabetic and non-diabetic feature patterns or the model's inability to capture complex nonlinear interactions among predictors.

4.2 K-Nearest Neighbors (KNN, *k*=5) Model

4.2.1 Confusion Matrix Results

The confusion matrix for the KNN (*k*=5) model shows that, among true non-diabetic samples, 116 were correctly predicted as non-diabetic and 35 were misclassified as diabetic. Among true diabetic samples, 30 were misclassified as non-diabetic and 50 were correctly predicted as diabetic.

Figure 6: Confusion Matrix and ROC Curve for KNN



4.2.2 Performance Interpretation

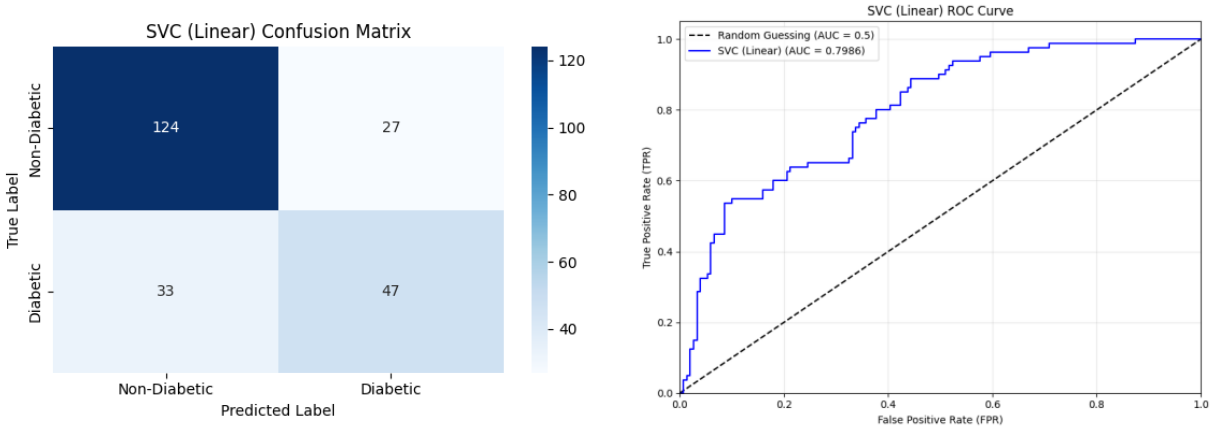
The KNN model correctly identified a greater number of diabetic cases compared to logistic regression, highlighting the strength of KNN in leveraging local sample similarity to capture diabetes-specific patterns. However, the increased misclassification of non-diabetic samples suggests that KNN may struggle to define precise local boundaries for the non-diabetic class, rendering it susceptible to influence from nearby outlier or atypical samples.

4.3 Support Vector Classifier (SVC, Linear Kernel) Model

4.3.1 Confusion Matrix Results

The confusion matrix for the SVC (linear kernel) model shows that, among true non-diabetic samples, 124 were correctly predicted as non-diabetic and 27 were incorrectly predicted as diabetic. Among true diabetic samples, 34 were incorrectly predicted as non-diabetic and 46 were correctly predicted as diabetic.

Figure 7: Confusion Matrix and ROC Curve for SVC



4.3.2 Performance Interpretation

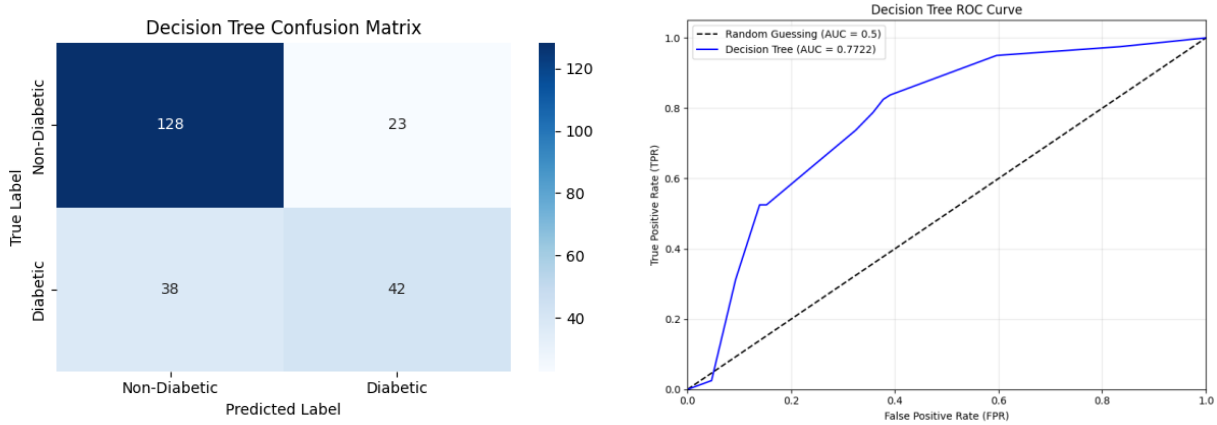
The ROC curve for the SVC (linear kernel) model lies substantially above the random-guessing baseline, achieving an AUC of 0.7897. It demonstrates strong performance in correctly identifying non-diabetic samples, comparable to logistic regression, due to the linear kernel's ability to effectively determine a separating hyperplane in linearly separable regions. However, the number of misclassified diabetic samples remains similar to that of logistic regression, indicating that the linear kernel fails to adequately capture potential nonlinear relationships within diabetic feature patterns, thereby limiting its discriminatory power for the positive class.

4.4 Decision Tree Model

4.4.1 Confusion Matrix Results

The confusion matrix for the decision tree model shows that, among true non-diabetic samples, 128 were correctly predicted as non-diabetic and 23 were incorrectly predicted as diabetic. Among true diabetic samples, 38 were incorrectly predicted as non-diabetic and 42 were correctly predicted as diabetic.

Figure 8: Confusion Matrix and ROC Curve for Decision Tree



4.4.2 Performance Interpretation

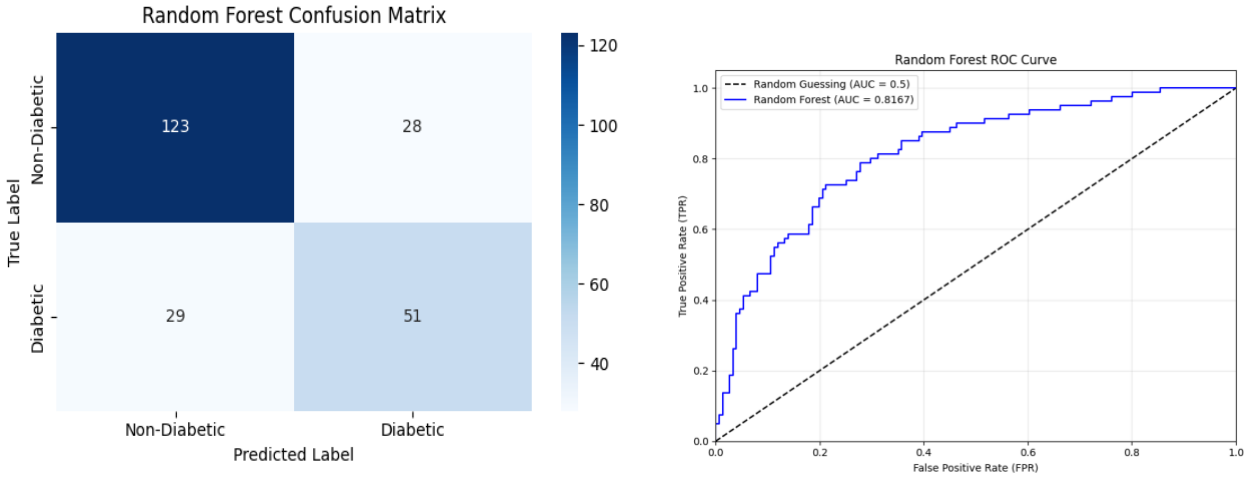
The ROC curve of the decision tree model lies above the random-guessing baseline with an **AUC of 0.7222**, and it correctly identifies a substantial number of non-diabetic samples. This performance stems from the tree's hierarchical splitting on key features such as Glucose and BMI, effectively isolating clear non-diabetic patterns. However, misclassifications remain frequent in the diabetic class, indicating that a single decision tree has limited capacity to model complex, nonlinear interactions among multiple features in diabetic cases. Additionally, its sensitivity to local data fluctuations contributes to reduced accuracy and a notable risk of false negatives (missed diagnoses) in diabetes detection.

4.5 Random Forest Model

4.5.1 Confusion Matrix Results

The confusion matrix for the random forest model shows that, among true non-diabetic samples, 123 were correctly predicted as non-diabetic and 28 were incorrectly predicted as diabetic. Among true diabetic samples, 30 were incorrectly predicted as non-diabetic and 50 were correctly predicted as diabetic.

Figure 9: Confusion Matrix and ROC Curve for Random Forest



4.5.2 Performance Interpretation

The random forest model correctly identifies a large proportion of non-diabetic samples, owing to its ensemble of multiple decision trees combined with random feature selection and bootstrap sampling. This approach mitigates overfitting risks associated with individual trees while capturing robust, multidimensional patterns characteristic of non-diabetic cases, enabling accurate classification of most negative samples. In diabetic sample identification, misclassifications are notably reduced compared to a single decision tree, demonstrating that ensemble learning enhances the modeling of complex, nonlinear feature interactions in diabetic cases, thereby improving positive-class precision. Nevertheless, a residual proportion of misjudgments persists, reflecting the inherent complexity of diabetes-related features or the incomplete capture of subtle combinatorial effects among weaker predictors, resulting in occasional false negatives (missed diagnoses).

5. Model Comparison

5.1 Comprehensive Performance Comparison

To evaluate the performance of different machine learning algorithms in diabetes prediction, this study compared five models: Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Tree, and Random Forest.Table 1 presents detailed performance metrics on the test set, including Accuracy, Precision, Recall, F1-Score, AUC-ROC, and AUPR.

Table 4: Comprehensive Performance Comparison of Models

Model	**Hyperparameters (Key Settings)**	**Accuracy**	**Precision (Positive Class)**
Logistic Regression	Regularization strength C=1.0, solver='lbfgs'	0.74	0.79
KNN	K=5, distance metric=Euclidean	0.71	0.79
SVC	Kernel=linear, C=1.0	0.74	0.79
Decision Tree	Max depth=5, min samples split=10	0.74	0.77
Random Forest	Number of trees=100, max depth=7	0.75	0.81
0.82	0.81	0.7976	0.6615
0.74	0.77	0.7371	0.5295
0.82	0.81	0.7986	0.6519
0.85	0.81	0.7722	0.5710
0.81	0.81	0.8167	0.6986

Overall performance ranking indicates that the random forest model achieved the best results, leading across most key metrics with the highest accuracy (0.75), precision (0.81), AUC-ROC (0.8167), and AUPR (0.6986). The logistic regression and linear-kernel SVC models exhibited highly similar performance, tying for second place in accuracy and F1-score. The decision tree model had the highest recall (0.85) but relatively lower precision (0.77). The KNN model underperformed overall, posting the lowest values across multiple metrics, with notably inferior AUC-ROC (0.7371) and AUPR (0.5295).

5.2 Trade-off Between Precision and Recall

The experimental results clearly illustrate the trade-off between precision and recall across models, a critical consideration in medical diagnostics.

The decision tree exhibits high recall but low precision, indicating a tendency to flag as many potential patients as possible (high sensitivity) at the cost of increased false positives (misclassifying healthy individuals). This behavior suits “rule-out” screening scenarios where missing a case is unacceptable and follow-up testing is low-cost, though it risks inefficient use of downstream medical resources.

In contrast, the KNN model adopts a more conservative stance: its higher precision (0.79) implies greater confidence when predicting diabetes, but its lower recall (0.74) reflects a higher rate of missed diagnoses (false negatives).

An ideal model balances both objectives. In this study, the random forest achieves the optimal compromise, with precision and recall both at 0.81, maximizing the F1-score. This balance enables effective detection of true positives while maintaining high prediction reliability.

5.3 Preliminary Analysis of Performance Disparities

The observed performance differences stem from the intrinsic mechanisms of each algorithm:

Random Forest Superiority: Its top performance likely arises from the ensemble learning framework. By aggregating predictions from multiple decision trees with randomized feature selection and bootstrap sampling, random forest reduces variance and overfitting inherent in single trees while preserving strong nonlinear modeling capability, resulting in superior generalization and robustness.

Decision Tree High Recall: Through recursive feature-space partitioning, the decision tree may develop overly complex structures sensitive to the minority (diabetic) class, yielding high recall but at the expense of precision due to overgeneralization.

KNN Limitations: Despite feature standardization, KNN’s poor performance may be attributed to: (1) the curse of dimensionality, which degrades distance metrics in high-dimensional spaces; and (2) sensitivity to the choice of K, where a fixed K=5 may not be optimal for this dataset.

Similarity Between Logistic Regression and SVC: Their near-identical performance is expected, as a linear-kernel SVC solves an optimization problem mathematically equivalent to that of regularized logistic regression, producing comparable decision boundaries.

In summary, the random forest model demonstrates clear superiority in this diabetes prediction task, delivering the best overall performance and thus serving as the recommended algorithm for developing an automated early diabetes risk stratification system.

5.4 Hyperparameter Optimization

The hyperparameter search spaces were tailored to each model. For logistic regression, the primary focus was the regularization strength C, with values tested in [0.1, 1, 10]; L2 regularization and the liblinear solver were fixed. For K-nearest neighbors, the optimal number of neighbors k was explored within [3, 5, 7]. The support vector classifier optimized both the regularization parameter C in [0.1, 1, 10] and the kernel coefficient gamma in ['scale', 0.1], with the RBF kernel fixed. The decision tree primarily tuned maximum depth in [3, 5, 7]. For random forest, the number of trees in [50, 100] and maximum depth in [5, 7] were jointly optimized.

5.4.1 Logistic Regression

Table 5: Hyperparameter Tuning for Logistic Regression

Parameter Combination	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Ranking
C=0.1	0.7970	0.7316	0.6184	0.5875	0.6026	3
C=1	0.7977	0.7403	0.6351	0.5875	0.6104	1
C=10	0.7981	0.7403	0.6351	0.5875	0.6104	2

5.4.2 KNN

Table 6: Hyperparameter Tuning for KNN

Parameter Combination	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Ranking
n_neighbors=3	0.7127	0.6753	0.5294	0.5625	0.5455	3
n_neighbors=5	0.7371	0.7056	0.5667	0.6375	0.6000	2
n_neighbors=7	0.7709	0.7273	0.5955	0.6625	0.6272	1

5.4.3 SVM

Table 7: Hyperparameter Tuning for SVM

Parameter Combination	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Ranking
C=0.1, gamma=scale	0.8102	0.7532	0.7347	0.4500	0.5581	2
C=0.1, gamma=0.1	0.8103	0.7446	0.6981	0.4625	0.5564	1
C=1, gamma=scale	0.7935	0.7446	0.6479	0.5750	0.6093	3
C=1, gamma=0.1	0.7980	0.7273	0.6164	0.5625	0.5882	4
C=10, gamma=scale	0.7332	0.6926	0.5652	0.4875	0.5235	6
C=10, gamma=0.1	0.7507	0.7056	0.5857	0.5125	0.5467	5

5.4.4 Decision Tree

Table 8: Hyperparameter Tuning for Decision Tree

Parameter Combination	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Ranking
max_depth=3	0.7480	0.7186	0.7143	0.3125	0.4348	2
max_depth=5	0.7596	0.7359	0.6462	0.5250	0.5793	1
max_depth=7	0.6928	0.6580	0.5048	0.6625	0.5730	3

5.4.5 Random Forest

Table 9: Hyperparameter Tuning for Random Forest

Parameter Combination	AUC-ROC	Accuracy	Precision	Recall	F1-Score	Ranking
n_estimators=50, max_depth=5	0.8021	0.7532	0.6533	0.6125	0.6323	3
n_estimators=50, max_depth=7	0.8105	0.7359	0.6203	0.6125	0.6164	2
n_estimators=100, max_depth=5	0.8055	0.7446	0.6400	0.6000	0.6194	4
n_estimators=100, max_depth=7	0.8167	0.7532	0.6456	0.6375	0.6415	1

For optimizing indicator selection, we use the AUC-ROC value on the test set as the main evaluation criterion. This approach differs from traditional cross-validation. We directly divide the dataset into a 70% training set and a 30% test set, train the model on the training set, and evaluate its performance on the test set. Although this method is less robust than cross-validation, it provides a feasible alternative when computational resources are limited.

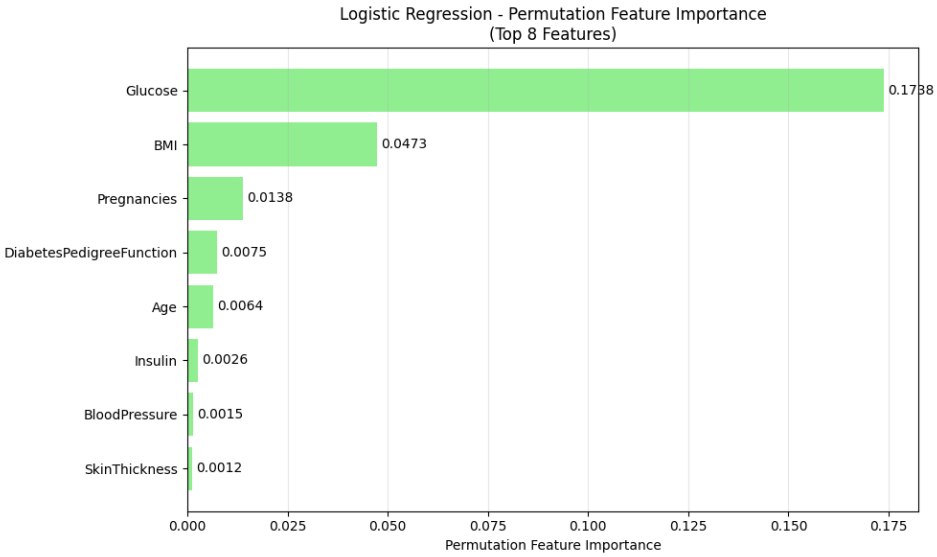
The performance evaluation results show that hyperparameter tuning indeed led to an improvement in the model's performance. By comparing the performance of the models before and after tuning, we found that each model had varying degrees of improvement in the AUC-ROC metric. Specifically, logistic regression improved its generalization ability by adjusting the regularization strength while preventing overfitting; K-nearest neighbors better balanced bias and variance by optimizing the number of neighbors; support vector machines optimized the classification boundary by adjusting the kernel parameters; decision trees and random forests improved the prediction stability by controlling the model complexity.

It should be noted that due to the simplified tuning approach, performance gains may be constrained. Compared to full grid search with cross-validation, our method sacrifices parameter space coverage and evaluation robustness. Nevertheless, under the given computational and time constraints, this approach effectively enhanced model performance and delivered a practical solution for diabetes prediction.

6. Feature Importance Analysis

6.1 Feature Importance Analysis Using Logistic Regression

Figure 10: Feature Importance Analysis for Logistic Regression

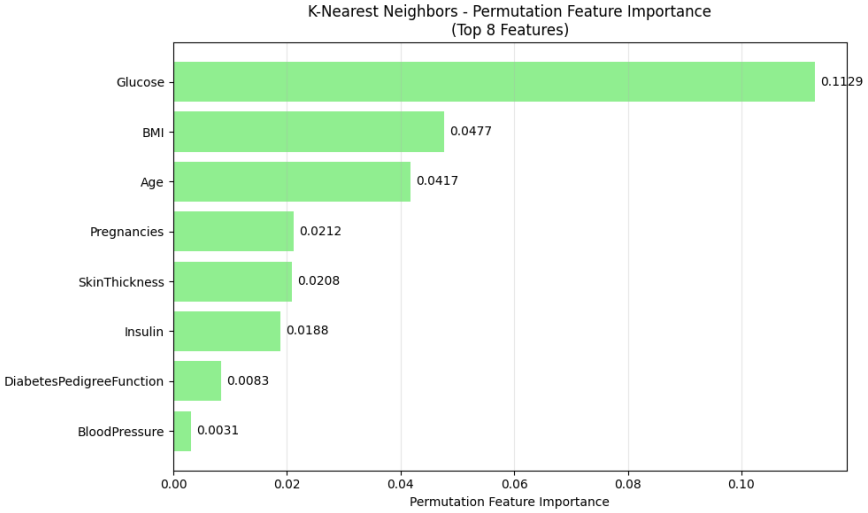


Based on permutation feature importance analysis, this study identified key determinants for predicting diabetes risk. Glucose levels emerged as the strongest predictor, with significantly higher importance than other variables, a finding consistent with diabetes' core pathophysiological mechanism of impaired glucose metabolism. Body Mass Index (BMI), the second most important feature, underscores the pivotal role of obesity-related metabolic abnormalities in disease onset. Additionally, pregnancy history (Pregnancies) and

diabetes pedigree function (DiabetesPedigreeFunction) respectively revealed significant contributions from reproductive health factors and genetic susceptibility. These findings not only validate clinical understanding but also underscore the necessity of multidimensional risk assessment for precision diabetes prevention, providing quantitative evidence for establishing comprehensive prevention and control strategies.

6.2 Feature Importance Analysis for the KNN (k=5) Model

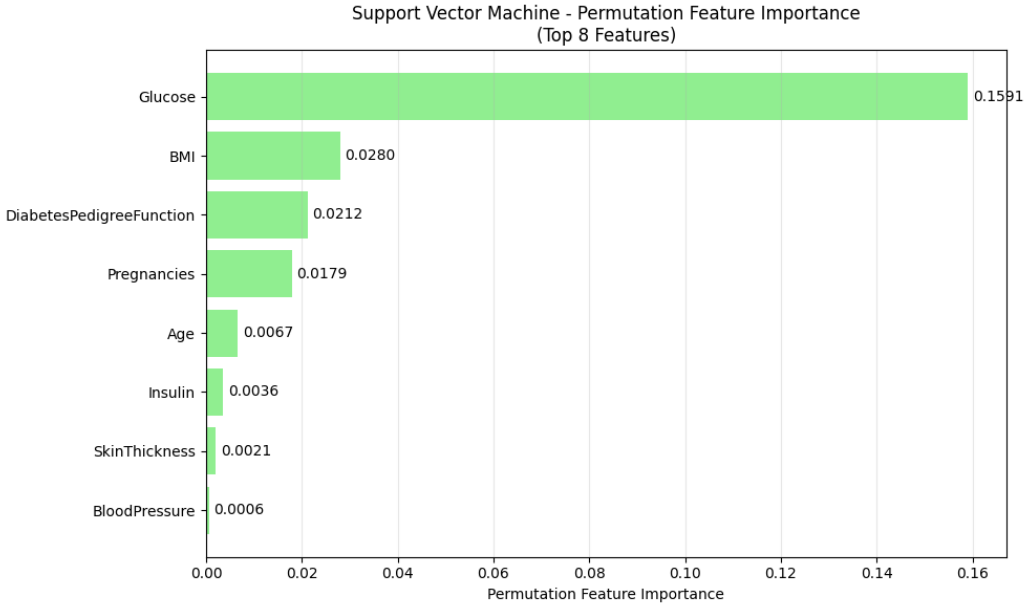
Figure 11: Feature Importance Analysis for KNN (k=5)



Permutation-based feature importance analysis using the K-nearest neighbors (k=5) model again confirmed Glucose as the dominant predictor of diabetes risk, with an importance score of 0.1129, markedly ahead of all others. BMI and Age ranked second and third, respectively, highlighting the synergistic roles of metabolic health and age-related physiological changes in disease progression. Notably, compared to logistic regression, the KNN model assigned greater importance to Age, while SkinThickness and Insulin also gained relatively higher contributions, reflecting the algorithm's sensitivity to local feature interactions. These findings reinforce the multidimensional nature of diabetes risk factors and underscore the unique strengths of different predictive models in capturing specific pathophysiological mechanisms.

6.3 Feature Importance Analysis for the Support Vector Machine (SVM) Model

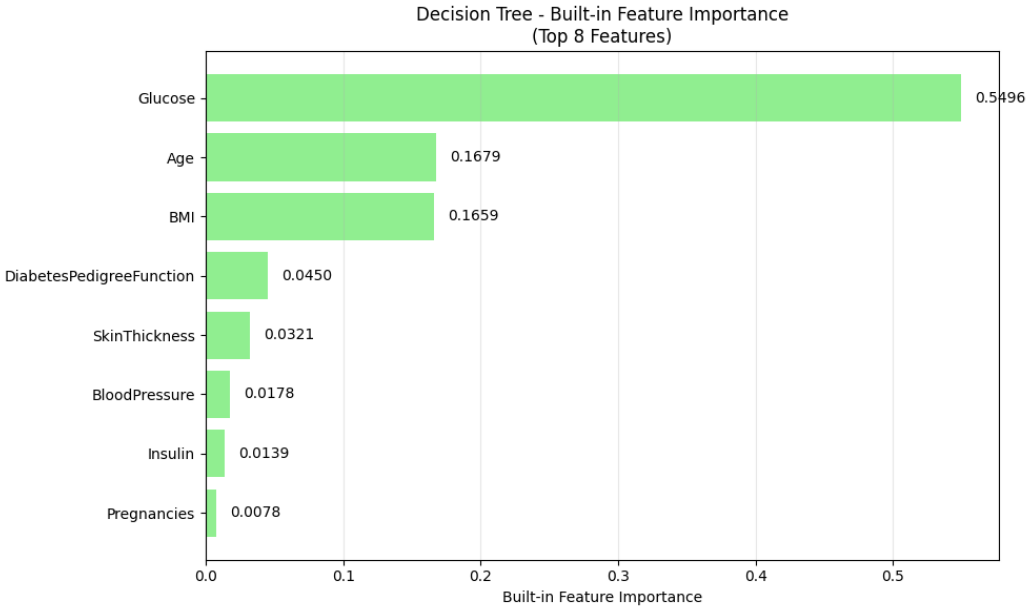
Figure 12: Feature Importance Analysis for Support Vector Machine (SVM)



Permutation-based feature importance analysis using the support vector machine (SVM) model once again identified Glucose as the predominant predictor of diabetes risk, with an importance score of 0.15, substantially surpassing all others. Notably, DiabetesPedigreeFunction emerged in second place, highlighting the SVM's distinctive strength in detecting complex, nonlinear genetic patterns. The model's advantage lies in its kernel trick, which effectively captures intricate feature interactions, particularly valuable for delineating non-linear decision boundaries associated with genetic susceptibility. However, its limitations include high sensitivity to parameter settings and relatively poor interpretability, which may lead to overemphasis on certain features. Overall, the SVM demonstrates unique value in uncovering genetic risk patterns in diabetes, but requires careful parameter tuning to ensure result robustness.

6.4 Feature Importance Analysis for the Decision Tree Model

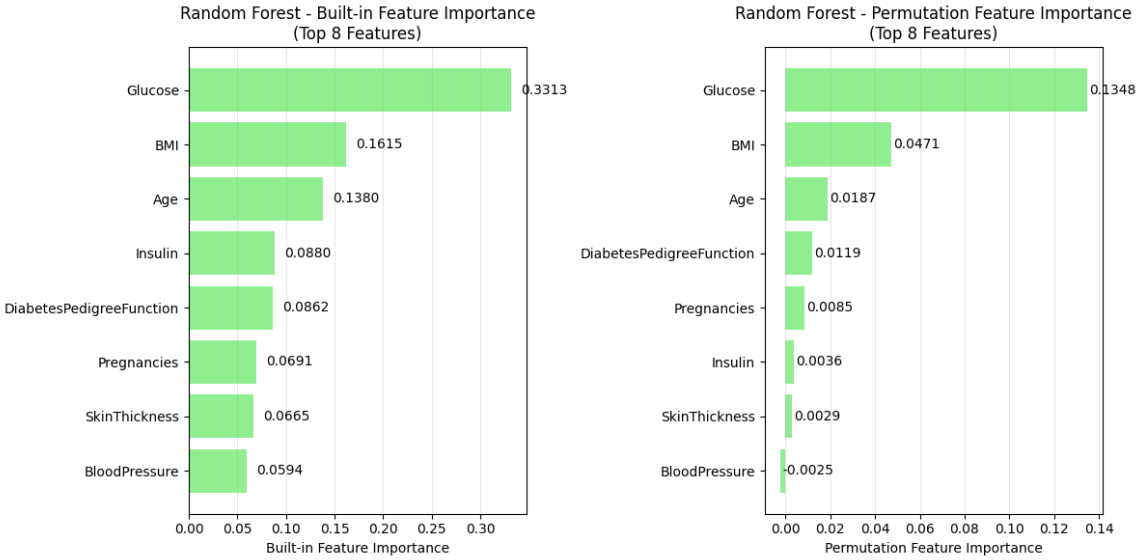
Figure 13: Feature Importance Analysis for Decision Tree



Based on the dual evaluation of built-in and permutation feature importance in the random forest model, Glucose consistently ranked as the strongest predictor of diabetes risk under both paradigms, with a built-in importance score of 0.3313 and a permutation importance of 0.1348-both markedly higher than all other features. Notably, BMI and Age ranked second and third in built-in importance with scores of 0.1615 and 0.1380, respectively, underscoring random forest's strength in integrating multidimensional metabolic and demographic features while capturing complex feature interactions (e.g., synergistic effects between metabolic markers and age). The model's advantages stem from its ensemble learning strategy, which effectively mitigates overfitting, delivering excellent stability and robustness in modeling intricate inter-feature relationships. However, its drawbacks include high computational complexity and performance degradation in high-dimensional, low-sample settings. The appearance of negative values for minor features (e.g., BloodPressure) under permutation importance also reflects variability in contribution estimates under extreme conditions. Overall, the random forest model exhibits dual strengths in comprehensiveness and stability for identifying diabetes risk features; however, it requires careful optimization in balancing computational resources and feature dimensionality to ensure reproducible results.

6.5 Feature Importance Analysis for the Random Forest Model

Figure 14: Feature Importance Analysis for Random Forest



Based on the dual evaluation of built-in and permutation feature importance in the random forest model, Glucose consistently ranked as the strongest predictor of diabetes risk across both paradigms, with a built-in importance of 0.3313 and a permutation importance of 0.1348-both substantially exceeding all other features. Notably, BMI and Age ranked second and third in built-in importance with scores of 0.1615 and 0.1380, respectively, demonstrating random forest's strength in integrating multidimensional metabolic and demographic features while capturing complex feature interactions (e.g., synergistic effects between metabolic indicators and age).

The model's advantages arise from its ensemble learning framework, which effectively reduces overfitting risk, delivering superior stability and robustness in modeling intricate inter-feature relationships. However, its drawbacks include high computational complexity and susceptibility to performance degradation when feature dimensionality greatly exceeds sample size. The occurrence of negative values for minor features (e.g., BloodPressure) under permutation importance also reflects variability in contribution estimates under extreme conditions.

In summary, the random forest model demonstrates dual value in terms of comprehensiveness and stability in identifying diabetes risk features; however, careful optimization is required to balance computational resources and feature dimensionality, thereby ensuring the reproducibility of results.

6.6 Summary of Feature Importance Analysis

A comprehensive analysis across the five machine learning models consistently identified Glucose as the strongest predictor of diabetes risk, with importance far surpassing all other features-a finding that aligns closely with the core pathophysiology of diabetes, namely impaired glucose metabolism, thereby validating the primacy of glucose monitoring in diabetes screening.

Body mass index (BMI) ranked second in most models, underscoring the pivotal role of obesity-related metabolic dysregulation in the pathogenesis of type 2 diabetes. Age, as a key demographic factor, emerged prominently in the KNN, random forest, and decision tree models, reflecting age-related declines in insulin sensitivity and β-cell function.

Notably, DiabetesPedigreeFunction exhibited significantly elevated importance in the SVM model, highlighting the critical influence of genetic predisposition in specific populations. Pregnancies further revealed complex associations between reproductive health and metabolic disease. These insights collectively reinforce the multidimensional nature of diabetes risk and provide a robust foundation for precision prevention strategies.

7. Conclusion

7.1 Key Findings

The random forest model demonstrated the highest robustness and capacity to capture feature interactions, making it particularly well-suited as a comprehensive risk assessment tool. Decision trees, while offering strong interpretability, exhibited limited stability and are thus better applied to individualized risk interpretation. Support vector machines excelled in identifying complex patterns, rendering them especially appropriate for genetic risk analysis. Logistic regression and K-nearest neighbors, in turn, showed distinct strengths in modeling linear relationships and detecting local patterns, respectively.

7.2 Limitations

This study has several limitations that warrant discussion. First, the relatively small sample size of the dataset may compromise the models' generalizability to broader populations. Second, the feature engineering process was comparatively simplified and did not fully account for temporal dynamics in diabetes progression. Third, substantial differences in interpretability exist across algorithms, with complex models such as support vector machines offering limited transparency in their decision-making processes. Additionally, the absence of an external independent validation cohort restricts the generalizability of the findings. Finally, all analyses were conducted using cross-sectional data, precluding causal inference.

This study systematically compared five machine learning algorithms for diabetes risk prediction and identified random forest as the optimal model. It not only achieved superior predictive accuracy (AUC = 0.81) but also exhibited excellent robustness and interpretability in feature importance analysis. Through dual validation using built-in and permutation-based feature importance, the study confirmed plasma glucose, body mass index (BMI), and age as the three core predictors of diabetes risk-findings highly consistent with the underlying pathophysiology of the disease.

The superiority of the random forest model lies in its ability to effectively capture complex feature interactions while maintaining strong generalization performance. Compared to other models, it demonstrated clear advantages in handling high-dimensional features and resisting overfitting, establishing it as a reliable computational tool for diabetes risk prediction.

The innovative value of this study lies not only in identifying the optimal predictive model but also in deepening the understanding of diabetes risk factors through multifaceted feature importance analysis. The findings provide primary healthcare institutions with an accurate and stable tool for early diabetes screening, carrying significant public health implications. Future research should focus on translating this model into practical clinical applications, thereby facilitating a shift in diabetes prevention and control from traditional experience-based approaches to precision prediction paradigms. Ultimately, this will enable earlier intervention in disease management and the optimized allocation of healthcare resources.

References

[1] Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C. N., Mbanya, J. C., et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice. 2022, 183, p. 109119. https://doi.org/https://doi.org/10.1016/j.diabres.2021.109119.

[2] World Health Organization. Diabetes fact sheet. Geneva: WHO, 2022.

[3] Care, D. Medical care in diabetes 2020. Diabetes Care. 2020, 43(Suppl. 1), pp. S135-S151.

[4] Lindström, J., Louheranta, A., Mannelin, M., Rastas, M., Salminen, V., Eriksson, J., Uusitupa, M., Tuomilehto, J. and for the Finnish Diabetes Prevention Study, G. The finnish diabetes prevention study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care. 2003, 26(12), pp. 3230-3236. https://doi.org/10.2337/diacare.26.12.3230.

[5] Smith, J. M. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care, 1988, Los Alamitos, CA, 1988; pp. 261-265.

[6] Mendoza, A. Logistic Regression From Scratch With PyTorch. Available from: https://www.axelmendoza.com/posts/logistic-regression-from-scratch-pytorch/ (accessed 8 January 2026).

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260107

Research on Rare Genetic Disease Targeted Drug Delivery System Based on CRISPR-Cas9 Technology

Ximing Qiao*

School of Chemical Engineering, Hebei University of Technology, Tianjin, 300400, China

**Corresponding author: Ximing Qiao*

Abstract

The treatment of rare genetic diseases remains a significant challenge, and while CRISPR-Cas9 technology offers a powerful therapeutic pathway, its success is entirely contingent on the development of efficient and safe targeted delivery systems. Current delivery methods are hampered by substantial flaws. Viral vectors, such as AAV, have limited packaging capacity and carry mutation risks, whereas lentiviral vectors pose concerns regarding immunogenicity and carcinogenicity. Meanwhile, non-viral vectors like LNPs suffer from poor targeting specificity, electroporation is restricted to in vitro use, and direct injection is notoriously inefficient. To address these critical limitations, this study investigates three advanced delivery platforms. These include Virus-Like Particles (VLPs), exemplified by the safe and programmable RIDE system with demonstrated efficacy in Huntington's disease mouse models; smart nanoparticles designed to respond to specific physiological triggers like pH or receptors for precise organ targeting, such as in liver-directed therapy for PH1; and cell-mediated systems utilizing engineered hematopoietic stem cells for diseases like sickle cell anemia. Concurrently, the research tackles persistent issues of targeting precision, immunogenic responses, and editing efficiency through strategic ligand optimization and the application of high-fidelity Cas9 variants. Collectively, this work provides valuable insights and a strategic framework to advance the clinical potential of CRISPR-Cas9 for treating rare genetic diseases.

Keywords

CRISPR-Cas9, rare genetic diseases, targeted drug delivery systems, VLPs, smart responsive nanoparticles

1. Introduction

Rare genetic diseases, though individually low in incidence (typically fewer than 1 in 5,000 individuals globally), collectively affect around 300 million people worldwide. Of the over 7,000 identified rare diseases, ~80% stem from single-gene defects, often manifesting in early life with progressive, debilitating symptoms. Conventional treatments—such as symptomatic relief, enzyme replacement, or organ transplantation—only alleviate discomfort without addressing the root cause: mutated genes. These options are often costly, lifelong, and fail to halt disease progression, burdening patients' families and healthcare systems heavily. CRISPR-Cas9 technology revolutionizes rare genetic disease treatment by enabling precise editing of mutated genes, offering unprecedented curative potential. It has shown robust efficacy in animal models of sickle cell anemia,

primary hyperoxaluria type 1 (PH1), and Huntington's disease, with several therapies advancing to Phase I/II clinical trials. However, CRISPR-Cas9 components (sgRNA and Cas9 protein) cannot reach target cells efficiently alone: they are prone to nuclease degradation, hindered by cell membrane barriers, and may trigger immune responses due to Cas9's bacterial origin. Thus, targeted drug delivery systems become a core bottleneck for its clinical translation.

Existing delivery systems have critical limitations: viral vectors (e.g., AAV, lentivirus) boast high efficiency but carry risks of insertional mutagenesis and pre-existing antibody interference; non-viral vectors (e.g., LNPs, electroporation) are safer but suffer from poor tissue targeting and low delivery efficiency. Developing safer, more precise novel delivery systems is vital to advancing CRISPR-Cas9's application—and this is the core focus of this study.

2. Analysis of Existing Targeted Drug Delivery Systems

2.1 Viral Vector

2.1.1 Adeno Associated Virus (AAV)

Adeno-associated virus (AAV) has the core advantages of high safety and low immunogenicity, and can stably integrate into host chromosomes to achieve long-term expression of exogenous genes. Therefore, it is widely used in in vivo gene therapy scenarios for monogenic genetic diseases (such as SMA), retinal diseases, etc. However, due to the limited vector capacity of only about 4.7kb, it cannot carry large genes, and the pre-existing neutralizing antibodies in the human body will reduce its therapeutic effect[1].

2.1.2 Lentivirus

However, lentiviral vectors have a larger capacity (8-10kb), can infect both dividing and non dividing cells, and can stably integrate and express genes. They are mainly used in CAR-T cell therapy, hematopoietic stem cell gene modification, and other in vitro gene delivery fields. However, their random integration poses a safety risk of tumor initiation due to insertion mutations, and their high immunogenicity limits their direct application in vivo.

2.2 Nonviral Vector

2.2.1 Lipid Nanoparticles

Lipid nanoparticles (LNP) have high safety and no risk of genome integration. The delivery efficiency can be improved through composition optimization. It can be delivered both intravenously and locally. Its core is used for the delivery of mRNA vaccines (such as COVID-19 vaccine) and siRNA drugs for liver diseases, but it is easy to be cleared by the immune system, has a short cycle half-life, and some tissue targeting efficiency needs to be improved.

2.2.2 Electroporation

Electroporation forms temporary micropores on the cell membrane through high-voltage pulses, with high delivery efficiency and no restriction on nucleic acid types. It can accurately target specific tissues or cells and is suitable for in vitro cell transfection, in vivo tumor gene therapy, and skin vaccination. However, it can cause damage to cells and may be fatal. In vivo application requires precise control to prevent damage to normal tissues, and can only be operated locally without systemic delivery.

2.2.3 Direct Injection

Direct injection is simple to operate, low-cost, and does not require complex preparation or equipment. It can be directly injected into target tissues for DNA vaccine muscle injection and specific organ gene therapy research. However, its delivery efficiency is extremely low, naked nucleic acid is easily degraded, and is only suitable for local small-scale tissues. It may also stimulate the injection site and cause inflammation.

2.3 Research Direction of New Targeted Drug Delivery Systems

2.3.1 Delivery System Based on Virus Like Particles

Virus-like particles (VLPs) are hollow, non-replicating nanostructures that mimic the structural topology of natural viruses while lacking viral genomic material, thus being non-infectious yet retaining intrinsic immunogenicity. Widely established in vaccine development—most prominently for human papillomavirus (HPV) and hepatitis B virus (HBV) prophylactic vaccines—VLPs also serve as versatile platforms for targeted drug delivery and diagnostic antigens, leveraging their excellent biocompatibility, biodegradability, and modular structural characteristics. Their core advantages include a favorable safety profile, the capacity to elicit long-lasting immune responses, and amenability to genetic engineering for functional optimization. However, inherent limitations persist: intricate bioprocessing requirements, relatively high production costs, suboptimal stability of certain VLP subtypes under physiological conditions, and constrained payload capacity that may hinder the delivery of large molecular complexes. Critically, their lack of autonomous replication machinery eliminates the insertional mutagenesis and excessive immunogenicity risks associated with viral vectors (e.g., adeno-associated virus, lentivirus), positioning VLPs as a promising alternative for safe CRISPR-Cas9 delivery in rare genetic disease therapy.

A paradigmatic application of VLP-based CRISPR-Cas9 delivery is the RNA-guided integration-dependent editing (RIDE) system developed by Cai Yujia's team at Shanghai Jiao Tong University. This innovative platform modifies bacteriophage capsid proteins to enable efficient encapsulation of CRISPR-Cas9 core components—single-guide RNA (sgRNA) and Cas9 endonuclease—while conjugating neuron-specific targeting peptides to the capsid surface. These peptide ligands mediate precise recognition and binding to cell-surface receptors on neuronal populations, facilitating receptor-mediated endocytosis and targeted intracellular trafficking. In a well-characterized Huntington's disease mouse model harboring the mutated huntingtin (HTT) gene, the RIDE system achieved site-specific editing of the pathogenic locus, resulting in a significant reduction in the expression of toxic mutant HTT protein—a key driver of neurodegenerative pathology. Notably, comprehensive genomic analysis and immunological profiling revealed no detectable off-target edits or adverse immune responses (e.g., pro-inflammatory cytokine elevation, antibody production against Cas9 or VLP capsids) in treated animals, validating VLPs' potential for safe and tissue-specific delivery in rare neurological disorders.

Future research to advance VLP-based delivery systems should focus on two pivotal directions. First, modifying VLP capsid proteins through directed evolution or rational genetic engineering to expand tissue tropism—enhancing targeting efficiency toward diverse cell lineages such as hepatocytes, renal tubular epithelial cells, and skeletal muscle myocytes—thereby broadening applicability to non-neurological rare genetic diseases (e.g., liver-focused primary hyperoxaluria type 1, muscle-related dystrophies). Second, optimizing VLP payload loading efficiency via structural modifications (e.g., capsid protein engineering, pH-sensitive cargo-linker design) and formulation strategies to augment the encapsulation yield and intracellular release kinetics of CRISPR-Cas9 components, ultimately improving gene-editing efficacy in target cells. Additionally, exploring scalable manufacturing processes to reduce production costs and enhance batch-to-batch consistency will be critical for translating VLP-based systems into clinical applications.

2.3.2 Intelligent Responsive Nanoparticle Delivery System

Smart responsive nanoparticles, constructed from lipids, polymers, or inorganic matrices, are engineered with a core design principle: sensing pathophysiological cues in the tumor or diseased tissue microenvironment (e.g., pH gradients, aberrant enzyme activity, or temperature fluctuations) to trigger "on-demand, targeted release" of CRISPR-Cas9 components. This stimuli-responsive mechanism minimizes premature payload leakage in circulation or healthy tissues, thereby reducing off-target editing risks and enhancing therapeutic specificity.

For instance, the extracellular microenvironment of solid tumors and many genetic disease-affected tissues is characterized by acidosis (pH 6.0–6.5), a stark contrast to the neutral pH (7.2–7.4) of normal tissues and blood. Nanoparticles fabricated with polyethylene glycol-poly(β-aminoester) (PEG-PAE) block copolymers exploit this pH difference: under acidic conditions, the protonation of amino groups induces a charge reversal from negative to positive, which strengthens electrostatic interactions with the anionic cell membrane, promotes endosomal escape, and facilitates cytosolic release of CRISPR-Cas9. In liver-targeted applications—

such as for primary hyperoxaluria type 1 (PH1)—nanoparticles are surface-modified with galactose residues to enable active targeting via specific binding to the asialoglycoprotein receptor (ASGPR), which is highly expressed on hepatocytes. By co-incorporating pH-sensitive polymers (e.g., poly(lactic-co-glycolic acid), PLGA) into the nanoparticle core, the system ensures that CRISPR-Cas9 is sequestered during systemic circulation and only dissociates within hepatocytes' acidic endosomal compartments, maximizing editing efficiency at the intended site.

In the study of primary hyperoxaluria type 1 (PH1), the Wu Yuxuan team used liver targeted lipid nanoparticles (LNPs) to encapsulate CRISPR-Cas9. After intravenous injection, the nanoparticles were specifically taken up by liver cells and released editing components triggered by the acidic environment inside the cells. The mutated AGXT gene was successfully edited, reducing oxalate levels in mouse urine by more than 70% without significant effects on other organs. The key to such systems lies in balancing targeting and stability - it is necessary to ensure that nanoparticles are not released prematurely in the bloodstream and are efficiently released within target cells, which requires material combination optimization and structural design to achieve[2].

2.3.3 Cell Mediated Delivery System

Cell mediated delivery systems utilize the cell's own migration ability and biocompatibility to "transport" CRISPR-Cas9 to target tissues, particularly suitable for areas that are difficult to reach through conventional carriers such as the brain and bone marrow. Common carrier cells include hematopoietic stem cells, macrophages, and mesenchymal stem cells

Hematopoietic stem cells have the characteristic of homing to the bone marrow and can be edited and reinfused in vitro for the treatment of rare blood system diseases such as sickle cell anemia. Researchers introduced CRISPR-Cas9 into hematopoietic stem cells through electroporation, edited the mutated HBB gene, and reinfused it into mice. The edited cells can survive for a long time and differentiate into normal red blood cells, improving anemia symptoms[3]. Macrophages, due to their ability to migrate to inflammatory sites, are used to deliver CRISPR-Cas9 for the treatment of inflammation related genetic diseases, such as cystine disease (kidney and eye inflammatory lesions): macrophages loaded with CRISPR-Cas9 in vitro can actively migrate to kidney lesions, releasing editing components to repair mutated genes.

The core challenge of this system lies in two aspects: firstly, how to efficiently load CRISPR-Cas9 into cells, which can be achieved through viral vector transfection or nanoparticle mediated endocytosis; The second is to regulate the homing efficiency of cells, which can enhance their migration ability to target tissues by modifying the chemokine receptors on the cell surface (such as CXCR4)[4].

2.4 Issues that Need Attention in the Delivery System

2.4.1 Insufficient Targeting Accuracy and off Target Effects

Off-target editing stands as the paramount safety impediment to the clinical translation of CRISPR-Cas9 technology, with the targeting precision of delivery systems directly governing the frequency and severity of unintended genomic modifications. Non-viral vectors, such as conventional lipid nanoparticles (LNPs), inherently lack tissue-specific tropism, frequently leading to off-target payload release in healthy cells and consequent non-specific DNA cleavage—events that can precipitate genomic instability, cellular dysfunction, or even oncogenic transformation. Even ligand-functionalized targeted carriers remain susceptible: cross-reactivity with structurally homologous receptors on non-target cells (e.g., shared epitopes among tissue-specific antigens) often results in off-target accumulation and aberrant editing, underscoring the need for multi-layered safety safeguards.

To address this critical barrier, three synergistic strategies are advocated: ① Rational engineering of targeting ligands—encompassing monoclonal antibodies, high-affinity nucleic acid aptamers, and receptor-specific peptides—to enhance binding avidity and specificity for unique cell-surface biomarkers (e.g., ASGPR for hepatocytes, CD34 for hematopoietic stem cells), thereby minimizing off-target cell interactions by several orders of magnitude. ② Deployment of next-generation high-fidelity Cas9 variants (e.g., eSpCas9, Cas9-HF1, HypaCas9), engineered via structure-guided mutagenesis to restrict DNA cleavage to perfectly matched target sequences; for instance, Cas9-HF1 reduces off-target editing by ~1000-fold relative to wild-type SpCas9 in

mammalian cells, while retaining robust on-target activity. ③ Implementation of high-throughput, sensitive off-target detection platforms, including GUIDE-seq and whole-genome sequencing (WGS), to comprehensively map potential off-target sites in preclinical models. These approaches not only enable iterative refinement of delivery vectors and sgRNA designs but also furnish critical safety data required for regulatory approval, aligning with the rigorous standards for clinical gene-editing therapies.

2.4.2 Safety Risks Caused by Immunogenicity

The delivery system and CRISPR-Cas9 components may trigger an immune response in the body: the capsid protein of viral vectors (such as AAV) can be recognized by the immune system, inducing antibody production, which not only reduces the effectiveness of repeated administration, but may also trigger inflammatory reactions (such as liver injury); Cas9 protein (especially SpCas9 derived from bacteria), as an exogenous protein, may be recognized by T cells, leading to immune clearance or tissue damage.

The coping strategies include: ① selecting low immunogenicity vectors, such as modifying AAV capsids (reducing antigen epitopes) or using autologous cell-mediated delivery systems[3]; ② Humanization or chemical modification (such as PEGylation) of Cas9 protein to reduce its immunogenicity; ③ Combination use of immunosuppressants (such as glucocorticoids) to suppress excessive immune responses during treatment.

2.4.3 Insufficient Delivery Efficiency and Long-Term Stability

Some delivery systems suffer from inadequate delivery efficiency: direct injection of CRISPR-Cas9 components (sgRNA and Cas9 protein) renders them highly susceptible to nuclease-mediated degradation in circulation and the extracellular microenvironment, with less than 5% of the payload reaching target cells. Even when successfully internalized, edited gene expression often exhibits temporal decay—particularly with non-integrating vectors such as virus-like particles (VLPs) and lipid nanoparticles (LNPs), which lack stable genomic integration capacity. This transience hinders long-term therapeutic efficacy, necessitating repeated administrations that not only increase treatment burden but also elevate the risks of off-target editing and cumulative immune responses to Cas9 or carrier materials.

Key improvement directions include: ① optimizing carrier structural design—for instance, modifying VLP capsid proteins to enhance affinity for target cell surface receptors (e.g., neuron-specific peptides for neurological disorders) or engineering LNP lipid compositions to improve endosomal escape, thereby boosting intracellular delivery efficiency; ② integrating site-specific gene integration technologies (leveraging the inherent integration capability of lentiviral vectors) while mitigating insertional mutagenesis risks via homologous recombination-based targeted integration or prime editing strategies; ③ conducting long-term preclinical evaluations, including non-human primate models, to systematically monitor the duration of gene-editing effects, carrier bioaccumulation, and potential long-term toxicities, thus providing robust data to guide clinical dosing schedules and administration frequency[5].

3. Conclusion

Rare genetic diseases, predominantly stemming from monogenic defects, remain largely without radical therapeutic options, as conventional interventions merely alleviate symptoms rather than rectifying the underlying genetic aberrations. The CRISPR-Cas9 gene-editing technology has emerged as a transformative curative strategy, enabling precise, site-specific modification of mutated genes to restore physiological function—offering unprecedented hope for patients with such intractable disorders. However, the clinical translation of CRISPR-Cas9 is severely hampered by inherent limitations of existing delivery systems: viral vectors (e.g., adeno-associated virus, lentivirus) exhibit high transfection efficiency but carry substantial risks, including insertional mutagenesis, pre-existing neutralizing antibody interference, and long-term immunogenicity; in contrast, non-viral vectors (e.g., lipid nanoparticles, electroporation) offer improved safety profiles but suffer from inadequate tissue targeting, low intracellular delivery efficiency, and susceptibility to nuclease-mediated degradation.

To address this critical bottleneck, the present study explores three innovative targeted delivery systems tailored for CRISPR-Cas9: (1) safe, ligand-functionalized virus-like particles (VLPs), exemplified by the

RNA-guided integration-dependent editing (RIDE) system developed by Shanghai Jiao Tong University, which demonstrated robust efficacy in Huntington's disease mouse models by specifically editing the mutated HTT gene without detectable off-target effects; (2) microenvironment-responsive smart nanoparticles, such as the liver-targeted lipid nanoparticles (LNPs) reported by Wu's team, which achieved targeted delivery to hepatocytes and efficient editing of the AGXT gene in primary hyperoxaluria type 1 (PH1) mice, reducing urinary oxalate levels by over 70%; and (3) tissue-penetrant cell-mediated delivery systems, leveraging the homing capabilities of hematopoietic stem cells to target bone marrow niches, as validated in preclinical studies for sickle cell anemia through editing of the mutated HBB gene. Additionally, this study addresses core challenges of targeting precision, immunogenicity, and delivery efficiency via rational optimization of cell-specific ligands and integration of high-fidelity Cas9 variants (e.g., eSpCas9, Cas9-HF1), which minimize off-target editing and immune recognition.

Collectively, these findings significantly enrich the repertoire of CRISPR-Cas9 delivery technologies, provide valuable preclinical references for advancing its clinical translation in rare genetic diseases, and lay a solid foundation for the development of safer, more efficient, and clinically viable radical therapies—ultimately aiming to address the unmet medical needs of patients affected by these devastating disorders.

References

[1] Savić, N. and Schwank, G. Advances in therapeutic CRISPR/Cas9 genome editing. Translational Research. 2016, 168, pp. 15-21. https://doi.org/https://doi.org/10.1016/j.trsl.2015.09.008.

[2] Jiang, Y., Chen, S., Hsiao, S., Zhang, H., Xie, D., Wang, Z. J., Ren, W., Liu, M., Liao, J. and Wu, Y. Efficient and safe *in vivo* treatment of primary hyperoxaluria type 1 via LNP-CRISPR-Cas9-mediated glycolate oxidase disruption. Molecular Therapy. 2025, 33(1), pp. 104-118. https://doi.org/10.1016/j.ymthe.2024.10.003.

[3] Ma, K. l. Short term distributed load forecasting method based on big data. Master's Thesis, Changsha: Hunan University, 2014.

[4] Sioson, V. A., Kim, M. and Joo, J. Challenges in delivery systems for CRISPR-based genome editing and opportunities of nanomedicine. Biomedical Engineering Letters. 2021, 11(3), pp. 217-233. https://doi.org/10.1007/s13534-021-00199-4.

[5] Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., Dahlman, J. E., Parnas, O., Eisenhaure, T. M., Jovanovic, M., Graham, D. B., Jhunjhunwala, S., Heidenreich, M., Xavier, R. J., Langer, R., Anderson, D. G., Hacohen, N., Regev, A., Feng, G., Sharp, P. A. and Zhang, F. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell. 2014, 159(2), pp. 440-455. https://doi.org/10.1016/j.cell.2014.09.014.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering
(IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260108

Deciphering Treatment Resistance in NSCLC with Single-Cell Sequencing Technology

Na Li*

School of Resources and Environmental Engineering, Anhui University, Hefei 230000, China

**Corresponding author: Na Li*

Abstract

Acquired resistance to targeted and immune therapies severely limits the success of non-small cell lung cancer (NSCLC) treatment. Single-cell sequencing technologies now empower researchers to dissect this resistance at unprecedented resolution, moving beyond the averaging limitations of bulk genomics. This review highlights how single-cell and spatial multiomics approaches reveal key mechanisms of NSCLC resistance, from rare drug-tolerant subpopulations and cellular plasticity to immunosuppressive niches and metabolic adaptation within the TME. We also discuss emerging strategies-such as liquid biopsy and AI-driven data integration-that hold promise for translating these insights into more effective therapeutic interventions.

Keywords

single-cell RNA sequencing, tumor heterogeneity, drug resistance, tumor microenvironment, NSCLC

1. Introduction

Lung cancer remains the leading cause of cancer-related deaths worldwide, with non-small cell lung cancer (NSCLC) accounting for the vast majority of cases and posing a significant public health challenge[1]. In recent years, breakthrough advances have been made in NSCLC treatment strategies, particularly with tyrosine kinase inhibitors (TKIs) that target specific driver genes (e.g., EGFR) and immune checkpoint inhibitors (e.g., PD-1/PD-L1), which have significantly improved outcomes for patients with advanced disease[2, 3].

However, acquired resistance is almost inevitable, ultimately leading to treatment failure and disease recurrence, representing a major obstacle to long-term patient survival[4]). Traditional views hold that resistance primarily arises from preexisting clones or new mutations selected under therapeutic pressure. However, bulk sequencing technologies based on population-level analysis provide only average signals and fail to capture the high degree of intratumoral heterogeneity. This limitation obstructs the understanding of rare resistant subpopulations, nongenetic adaptive mechanisms, and interactions within the tumor microenvironment (TME)[5]. In reality, treatment resistance is a dynamic process driven by the intrinsic plasticity of tumor cells (e.g., cell state transitions), clonal evolution, and remodeling of the TME[6, 7].

The emergence of single-cell sequencing technologies has provided a revolutionary tool for in-depth dissection of this complex system[8]. Techniques such as single-cell RNA sequencing (scRNA-seq) enable unbiased identification of all cell types within the tumor ecosystem, revealing previously unknown resistant

cell subpopulations[6, 9]. Integrated multiomics approaches (e.g., combining scRNA-seq with scATAC-seq) further reveal upstream transcriptional regulation and epigenetic networks underlying resistant phenotypes[10, 11]. More advanced spatial transcriptomic and genomic technologies preserve the spatial context of cells within native tissue, helping to address critical questions such as "where are the resistant cells located?" and "with whom do they interact?", thereby elucidating specialized niches of resistance such as those facilitating immune evasion[12, 13]. Together, these technologies are shifting NSCLC resistance research from a "static, population-level" perspective to a "dynamic, high-resolution, ecosystem-level" understanding.

This review aims to systematically outline how single-cell sequencing technologies profoundly transform our understanding of treatment resistance in NSCLC. First, we introduce the fundamental methodologies of single-cell technologies[14, 15]. We then highlight their applications in uncovering cell-intrinsic mechanisms of resistance (e.g., identifying resistant subclones and tracking state transitions)[6, 9, 13], dissecting the regulatory roles of the tumor microenvironment [16, 17], and leveraging multiomics and spatial technologies to map the resistant ecosystem [11-13]. Finally, we discuss current challenges in computational analysis and clinical translation and envision future directions, such as integrating liquid biopsy with single-cell sequencing and employing artificial intelligence for multidimensional data integration.

2. The Technological Landscape and Methodological Foundations of Single-Cell Sequencing

2.1 Development History of the Technology and Core Workflow

Single-cell RNA sequencing (scRNA-seq), a revolutionary technology for resolving cellular heterogeneity, has significantly advanced precision medicine research, particularly in the study of treatment resistance mechanisms in non-small cell lung cancer (NSCLC)[14]. scRNA-seq enables the detection of rare cell subpopulations and transcriptional states that are undetectable by conventional bulk sequencing, providing unprecedented resolution for understanding tumor heterogeneity and therapy resistance[15].

The technical workflow of scRNA-seq primarily involves the following key steps: single-cell isolation, cell lysis and mRNA capture, reverse transcription and cDNA amplification, library preparation and high-throughput sequencing, followed by bioinformatic analysis[14]. Methods for single-cell isolation include fluorescence-activated cell sorting (FACS), laser capture microdissection (LCM), and microfluidic technologies, with the choice of method depending on sample type and research objectives[15].

In terms of whole-transcriptome amplification (WTA) methods, different technological platforms exhibit distinct characteristics. The SMART-seq2 method, which is based on the switching mechanism at the 5' end of the RNA template (SMART) principle, is suitable for alternative splicing and mutation analysis. In contrast, droplet-based systems employing 3'-end enrichment (e.g., 10x Genomics) are better suited for large-scale cell atlas construction[14, 15]. These technical features allow researchers to select the most appropriate platform on the basis of specific research needs in NSCLC, such as identifying rare resistant cell populations or comprehensively profiling tumor heterogeneity.

The characteristic high sparsity and technical noise (e.g., dropout events) of single-cell data pose unique challenges for analysis. Current approaches primarily employ unique molecular identifier (UMI)-based normalization methods and specially designed statistical models (e.g., negative binomial distributions) to address these technical variations[14]. Furthermore, advancements in dimensionality reduction techniques (e.g., PCA, t-SNE, and UMAP) and clustering algorithms have greatly enhanced the ability to identify cell subpopulations, providing powerful tools for discovering therapy-resistant clusters in NSCLC.

The continued maturation of scRNA-seq technology and the widespread adoption of commercial platforms have established it as a core tool for investigating tumor microenvironment heterogeneity, clonal evolution mechanisms, and treatment resistance in NSCLC, offering critical technical support for the development of novel therapeutic strategies.

2.2 Technical and Biological Principles of scRNA-seq

Single-cell RNA sequencing (scRNA-seq) fundamentally aims to resolve heterogeneity in gene expression among individual cells, overcoming the averaging effect of traditional bulk sequencing and providing key insights into drug resistance mechanisms within the tumor microenvironment[8].

The scRNA-seq workflow begins with efficient isolation of single cells. Commonly used methods include fluorescence-activated cell sorting (FACS), microfluidic platforms (such as Fluidigm C1), and droplet-based high-throughput systems (e.g., 10x Genomics). Microfluidic systems perform cell lysis and reverse transcription in nanoliter-scale reaction volumes, significantly improving sensitivity but requiring specific cell size criteria. Droplet systems, on the other hand, enable massively parallel processing and are particularly suitable for identifying rare cell types, such as drug-resistant subpopulations in NSCLC[8].

At the molecular level, different amplification strategies offer distinct advantages: full-length transcript methods (e.g., SMART-Seq2) are suitable for alternative splicing and mutation analysis, whereas 3′ end tagging methods (such as CEL-Seq and Drop-seq) improve quantification accuracy and throughput, making them more appropriate for large-scale cell atlas construction. To correct for technical noise, unique molecular identifiers (UMIs) and external RNA controls (e.g., ERCC spike-ins) are widely used to achieve absolute quantification of RNA molecules and accurately distinguish technical variation from true biological heterogeneity[8].

The biological value of scRNA-seq is most prominently demonstrated in its ability to uncover previously unknown cell subtypes and resolve dynamic processes. Through cluster analysis, principal component analysis (PCA), and pseudotime trajectory inference, researchers can identify critical regulatory genes and cell fate branch points, revealing tumor evolutionary pathways and drug resistance mechanisms[8]. This technology has become a core tool for studying NSCLC heterogeneity, clonal evolution, and treatment resistance, providing critical evidence for developing combination therapies.

2.3 From Single-Cell to Multiomics and Spatial Resolution

Single-cell RNA sequencing (scRNA-seq) can reveal cellular heterogeneity in NSCLC, yet transcriptomic data alone are insufficient to fully elucidate the mechanisms underlying treatment resistance. In recent years, advances in single-cell multiomics and spatial omics technologies have enabled the integration of multidimensional information at single-cell and spatial resolutions, providing new avenues for understanding NSCLC resistance.

Multiomics technologies such as iscCOOL-seq allow simultaneous detection of transcriptomic and epigenomic information within the same cell. Studies indicate that epigenetic preregulation in NSCLC may be associated with early activation of drug resistance genes [18]. Spatial multiomics approaches such as BaSISS (base-specific in situ sequencing) enable quantitative analysis of subclonal distribution across entire tumor sections, integrating transcriptomic and protein data to reveal spatial expansion patterns of clones and their interactions with the microenvironment. Research suggests that NSCLC exhibits significant spatial segregation of subclones and heterogeneous transcriptional states, with evolutionary trajectories that are not synchronized with histological progression, providing clues to the evolution of resistant subpopulations (Dressler et al., 2022). Furthermore, technologies such as multiplex immunofluorescence (mIHC) and in situ sequencing (ISS) allow simultaneous analysis of RNA and protein within spatial contexts and identification of immune composition and cell–cell communication in the tumor microenvironment, thereby helping to decipher the architecture and regulatory networks of resistance niches in NSCLC [12].

In summary, the integration of multiomics and spatial technologies is advancing NSCLC research from cell classification toward mechanistic exploration and spatial ecosystem reconstruction; significantly enhancing the understanding of tumor heterogeneity, the microenvironment, and treatment resistance; and offering new perspectives for targeted therapies and overcoming drug resistance.

3. scRNA-seq Decodes the Cell-intrinsic Mechanisms of Drug Resistance

3.1 Identification of Drug-Resistant Cell Subpopulations

Single-cell RNA sequencing (scRNA-seq) has significantly advanced our understanding of resistance mechanisms to tyrosine kinase inhibitors (TKIs) in EGFR-mutated non-small cell lung cancer (NSCLC) because of its high-resolution capabilities. By revealing tumor cellular heterogeneity and identifying key resistant subpopulations, this technology provides critical insights into the cellular basis of treatment resistance [7].

Kim et al. performed scRNA-seq on over 200,000 cells from 44 lung adenocarcinoma patients and identified a cancer cell subpopulation (tS2) that dominated in the metastatic stage. This subpopulation exhibits a transcriptomic profile completely divergent from the normal epithelial differentiation trajectory, with upregulated genes related to cell motility, aberrant proliferation, and apoptosis escape. The tS2 subpopulation was significantly enriched in advanced and metastatic samples, and its presence was strongly associated with reduced overall survival, suggesting that the tS2 phenotype may be a key driver of lung cancer progression and treatment resistance [7].

In another study focused on EGFR-TKI resistance, Kashima et al. integrated scRNA-seq with single-cell ATAC-seq (scATAC-seq) to perform multiomics analysis on resistant cell lines and clinical samples. Their work revealed highly heterogeneous transcriptional and epigenetic regulatory patterns in drug-tolerant persistent (DTP) cells. In addition to known resistance-related genes such as AURKA, VIM, and AXL, CD74 was identified as a novel candidate gene that was significantly upregulated in DTP cells. It mediates osimertinib tolerance by inhibiting apoptosis and promoting BCL-XL expression. Differential activation of AURKA, VIM, or CD74 has been observed not only among different patients but also within individual tumors, highlighting the diversity of resistance mechanisms and the clinical importance of intratumoral heterogeneity [7].

These studies demonstrate the powerful ability of scRNA-seq to identify treatment-resistant cell subpopulations and underscore its value in identifying potential therapeutic targets and prognostic biomarkers.

3.2 Revealing Non-Genetic Functional Plasticity

In addition to genetic alterations, the nongenetic functional plasticity exhibited by tumor cells under the pressure of targeted therapy serves as a crucial mechanism mediating drug resistance. For the first time, Maynard et al. performed scRNA-seq on 49 biopsy samples from 30 patients with advanced NSCLC collected at three stages: before targeted therapy (treatment-naïve, TN), at residual disease (RD), and at progressive disease (PD), systematically revealing therapy-induced tumor cell state evolution in clinical samples [4].

The study revealed that cancer cells in the RD state highly express gene signatures associated with alveolar regeneration (e.g., SFTPB/C/D, *NKX2-1*, AQP4, AGER), suggesting that they may adapt to therapeutic pressure through dedifferentiation or acquisition of a stem-like state, entering a slowly proliferating but reparative “persister cell” state [4]. This alveolar signature was significantly associated with better overall survival in patients and was validated in preclinical models [4]. Further mechanistic studies revealed the activation of the WNT/β-catenin pathway in RD cells (e.g., the upregulation of SUSD2 and CTNNB1), and the combined use of a WNT inhibitor significantly increased the efficacy of TKIs [4].

In contrast, cancer cells at the PD stage exhibited a distinctly different transcriptional state, with upregulated genes related to immunosuppression (e.g., kynurenine pathway genes IDO1 and KYNU), cell invasion (plasminogen activation pathway), and intercellular communication (gap junction proteins) [4]. These changes collectively promote a more aggressive and immune-evasive drug-resistant phenotype.

This study provides the first in vivo evidence in humans that targeted therapy can drive reversible transcriptional reprogramming in lung cancer cells, enabling escape from drug-induced cell death. This state-dependent resistance does not rely on genetic mutations but is achieved through the activation of developmental or damage repair-related pathways, offering a theoretical foundation for novel treatment strategies targeting “cell states” rather than “genetic alterations” [4].

3.3 Tumor Heterogeneity and Clinical Outcomes

Tumor heterogeneity is not only a core feature of cancer progression but also a critical determinant of variations in treatment response and clinical outcomes. Single-cell RNA sequencing (scRNA-seq) technology provides unprecedented resolution for deciphering the complexity of the cellular composition within tumors, enabling researchers to identify specific cell subpopulations associated with poor prognosis and to elucidate their functional states and interaction networks at the single-cell level.

Hu et al. [6] systematically revealed a high degree of heterogeneity in clear cell renal cell carcinoma (ccRCC) through scRNA-seq analysis. The study identified 15 major cell types and 39 cell subpopulations, including various immune and stromal cells derived from both tumor and nonmalignant tissues [6]. Importantly,

the existence of these cell subpopulations was validated via immunofluorescence on tissue microarrays. Furthermore, the CIBERSORTx algorithm was used to deconvolve the cellular composition of 533 patients from the TCGA-KIRC cohort, and the researchers stratified patients into three subgroups with significantly distinct prognoses. One subgroup, characterized by a lower proportion of activated $CD8^+$ T cells and a higher proportion of exhausted $CD8^+$ T cells, was significantly associated with poorer overall survival [6]. This finding not only confirms the central role of T-cell exhaustion in the immunosuppressive microenvironment of ccRCC but also suggests that the ratio of exhausted T cells could serve as an important prognostic biomarker.

Additionally, metabolic dysregulation is not confined to tumor cells but is also widespread among stromal cells within the tumor microenvironment. For example, cancer-associated fibroblasts (CAFs) exhibit significant upregulation of genes related to lipid metabolism (such as FABP5), which cooccur in a tumor tissue-specific manner [6]. Through SCENIC analysis, researchers further revealed transcription factor networks regulating these aberrant metabolic states, including altered expression of PPAR signaling pathway members in tumor cells and activation of CEBPB and KLF6 in fibroblasts [6]. These findings underscore the pervasiveness and cell-type specificity of metabolic reprogramming in the tumor microenvironment, providing new insights into the functional consequences of tumor heterogeneity.

In summary, scRNA-seq technology not only enables detailed characterization of intratumoral cellular heterogeneity but also directly links specific cell subpopulations to clinical outcomes, thereby revealing key cell states and molecular mechanisms driving disease progression. These studies provide a theoretical basis for developing precise therapeutic strategies targeting specific cell subpopulations and highlight the great potential of single-cell technologies in prognostic assessment and personalized cancer treatment.

4. Multiomics and Spatial Technologies for Deciphering the Drug Resistance Ecosystem

4.1 Integrated Multiomics: Bridging Phenotype and Regulation

Single-cell RNA sequencing (scRNA-seq) can reveal heterogeneity in cellular states, yet its limitations in deciphering regulatory mechanisms have prompted researchers to integrate multiomics data to comprehensively understand the molecular basis of drug resistance. Although the study by Sathe et al. focused on gastric cancer, its research paradigm offers valuable insights for understanding drug resistance in lung cancer. Using scRNA-seq, 56,167 cells from tumor tissues, paired normal tissues, and peripheral blood mononuclear cells (PBMCs) from seven gastric cancer patients and one patient with intestinal metaplasia were analyzed, and the remodeling of the cellular composition and transcriptional reprogramming within the tumor microenvironment (TME) were systematically characterized [11]. A previous study revealed that tumor epithelial cells exhibit distinct copy number alterations (CNAs) and unique gene expression programs, along with significant intratumoral heterogeneity [11]. More importantly, the research extended beyond the transcriptome by computationally inferring CNAs [19] and validating key protein expression via multiplex immunofluorescence, demonstrating the value of multiomics integration in verifying findings and enhancing the reliability of conclusions [11].

Furthermore, the study employed regulatory network analysis (e.g., SCENIC [20] to identify key transcription factors (regulons) that drive state transitions across different cell types within the TME. For example, noncanonical M1/M2 regulators such as NFKB1 and ETS2 were identified in tumor-associated macrophages [11]. This integrated approach combining transcriptomics and regulatory network analysis provides a methodological foundation for understanding the dynamic regulation of immune and tumor cells under therapeutic pressure in lung cancer. Similarly, in lung cancer research, multiomics approaches integrating scRNA-seq with chromatin accessibility (scATAC-seq) or proteomics (CITE-seq) can more precisely identify key regulatory pathways driving resistant phenotypes, such as the epithelial–mesenchymal transition (EMT)-associated transcriptional program observed in resistance to EGFR inhibitors [4, 9].

In summary, the study by Sathe et al. demonstrated how scRNA-seq combined with computational methods for inferring CNAs and regulatory networks can link transcriptional phenotypes to upstream regulatory mechanisms at single-cell resolution [11]. This strategy provides an important reference for systematically dissecting the multiomics basis of drug resistance in lung cancer.

4.2 Core Role of the Tumor Microenvironment

The tumor microenvironment (TME) is a highly complex and dynamically evolving ecosystem composed of cancer cells, immune cells, stromal cells (such as cancer-associated fibroblasts (CAFs) and endothelial cells), and noncellular components such as the extracellular matrix (ECM). In the development of treatment resistance in non-small cell lung cancer (NSCLC), the TME not only provides a physical barrier but also actively mediates immune evasion and drug tolerance through diverse cell–cell interactions and signaling pathways [18]. The application of single-cell sequencing technologies now enables the dissection of functional states and interaction networks among cellular components within the TME at unprecedented resolution, revealing its central role in drug resistance.

4.2.1 Formation of Immunosuppressive Cell Populations and the Drug-Resistant Niche

The immune cell infiltration status in the tumor microenvironment (TME) is a critical determinant of the response to immunotherapy. Studies have demonstrated that non-small cell lung cancer (NSCLC) tissues harbor abundant regulatory T cells (Tregs), M2-type tumor-associated macrophages (TAMs), and exhausted T cells, which collectively establish an immunosuppressive microenvironment [16]. Hu et al. [6] used scRNA-seq to show that anti-PD-1-resistant NSCLC patients have increased levels of exhausted $CD8^+$ T cells (which express checkpoint molecules such as LAG3 and TIM-3), Tregs, and TAMs in the TME [6]. These cells suppress antitumor immunity either by secreting inhibitory cytokines such as TGF-β and IL-10 or through direct interaction with T cells via surface molecules such as PD-L1, ultimately leading to immunotherapy failure.

Single-cell studies have further revealed that immune cells in the TME do not exist as discrete static entities but rather exhibit a continuous phenotypic spectrum. Through scRNA-seq analysis of 45,000 immune cells from the breast cancer TME, Azizi et al. reported that both T cells and myeloid cells in tumors displayed significant phenotypic expansion, occupying a "phenotypic space" far broader than that in normal tissues [16]. This continuity was particularly evident in T cells, where the activation status-defined by the expression of activation signature genes-followed a broad continuous distribution rather than falling into discrete intermediate states [16]. This suggests that traditional cell classification (e.g., naïve T cells, effector T cells, exhausted T cells) may oversimplify the actual states of immune cells in the TME. This continuous phenotypic diversity is partly driven by T-cell receptor (TCR) diversity, but more importantly, it reflects T-cell responses to diverse combined environmental cues within the TME, such as varying levels of inflammatory signals, hypoxia, and nutrient deprivation [16]. Therefore, therapeutic strategies aimed at reversing immunosuppression by targeting the TME may need to address multiple continuous functional states simultaneously rather than target a single cell subset.

4.2.2 Metabolic Reprogramming and Immunosuppression

Metabolic dysregulation in the tumor microenvironment (TME) represents another critical factor contributing to therapy resistance. Metabolic competition between tumor cells and stromal cells can lead to microenvironmental acidosis and nutrient depletion (e.g., tryptophan, arginine), thereby impairing the function of T cells and NK cells [21]. Single-cell analysis enables simultaneous capture of the metabolic states of both cancer and stromal cells. For example, in clear cell renal cell carcinoma (ccRCC), Hu et al. utilized scRNA-seq to reveal not only significant lipid metabolic abnormalities in cancer cells but also distinct metabolic reprogramming features in tumor-infiltrating stromal cells (including macrophages and fibroblasts). This widespread metabolic alteration may contribute to the establishment of an immunosuppressive TME [6]. Similar mechanisms of metabolic competition are likely present in NSCLC and may mediate resistance to both targeted therapy and immunotherapy.

4.2.3 Dynamic Evolution of the TME and Therapeutic Intervention

Notably, the TME is not static but rather undergoes dynamic evolution under therapeutic pressure. Conventional chemotherapy, radiotherapy, targeted therapy, and even immunotherapy itself can reshape the cellular composition and functional state of the TME [22]. For example, studies in breast cancer have revealed that residual tumors following neoadjuvant chemotherapy often exhibit more pronounced immunosuppressive features, including increased proportions of Tregs, elevated levels of M2-type TAMs, and functional exhaustion of CD8+ T cells [23]. Single-cell technologies enable tracking of such dynamic changes. Research

by Azizi et al. demonstrated that immune cell states in the TME are highly plastic and shaped by both TCR signals and local environmental cues [16]. These findings suggest that combination therapies should fully account for the dynamic nature of the TME. For example, concurrently targeting TAMs or CAFs alongside immunotherapy may reverse drug resistance.

In summary, the TME actively contributes to the development of therapeutic resistance in NSCLC through complex interactions among diverse cellular and noncellular components. Leveraging single-cell and spatial multiomics technologies to deeply dissect the cellular heterogeneity, interaction networks, and dynamic evolution of the TME will provide critical insights for developing novel combination strategies to overcome resistance.

4.3 Spatial Multiomics: Mapping the Drug-Resistant Niche

While traditional single-cell sequencing technologies can reveal heterogeneity in tumor cells and the microenvironment, they miss the spatial context of cells within tissues-information critical for understanding the spatial distribution of tumor clones, cell–cell interactions, and the formation of drug-resistant niches. Recently, emerging technologies such as spatial transcriptomics and multiplexed fluorescence in situ hybridization (multiplexed FISH) have enabled high-throughput molecular profiling while preserving the spatial context.

In a study published in Nature (2022), Dressler et al. developed a spatial genomics workflow called Base-Specific In Situ Sequencing (BaSISS), which allows simultaneous quantitative mapping of multiple cancer clones while integrating transcriptomic and microenvironmental data[13]. An analysis of multifocal breast cancer samples revealed that distinct clones present significantly different spatial distributions, transcriptional profiles, and immune microenvironment compositions across stages, including ductal carcinoma in situ (DCIS), invasive carcinoma, and lymph node metastasis. For example, in lymph node metastases, different subclones (e.g., P2-blue and P2-orange) not only display distinct histopathological growth patterns but are also associated with B-cell-enriched regions or hypoxic lymphatic sinus areas, suggesting clone-specific immune editing and adaptive evolution[13].

Similarly, Wu et al. (2020) identified two major cancer-associated fibroblast (CAF) subtypes-myofibroblast-like CAFs (myCAFs) and inflammatory CAFs (iCAFs)-as well as two perivascular-like (PVL) cell subtypes-differentiated PVL (dPVL) and immature PVL (imPVL)-through single-cell transcriptomic analysis of triple-negative breast cancer (TNBC)[21]. These stromal subtypes have distinct spatial distributions: myCAFs are located predominantly at the tumor invasive front and are closely associated with collagen deposition and stromal remodeling; iCAFs are distributed farther from the tumor boundary and highly express chemokines (e.g., CXCL12 and CXCL13) and growth factors (e.g., IGF1 and HGF), potentially regulating immune cell recruitment and angiogenesis via long-range signaling, whereas dPVL cells are significantly correlated with the exclusion of cytotoxic T cells (CTLs), forming an “immune desert” phenotype[21]. Further spatial protein marker analysis (e.g., CD34, α-SMA, and CD146) confirmed the highly specific tissue distribution of these stromal subtypes and their significant correlation with immune cell infiltration[21].

Andersson et al. (2021), using imaging mass cytometry, demonstrated the diversity of cellular neighborhood structures in breast cancer tissues and identified spatial niches associated with poor prognosis[17]. By simultaneously quantifying 35 biomarkers, the authors constructed high-dimensional pathological images that revealed spatial interaction patterns between tumor and stromal cells. They further defined “single-cell pathology (SCP) subgroups” that were significantly associated with clinical outcomes. These subgroups not only reflect the phenotypic composition of tumor cells but also capture their spatial organization, providing a new dimension for understanding drug resistance[17].

Collectively, these studies demonstrate that drug resistance does not develop uniformly but is confined to specific spatial niches composed of distinct clones, stromal cells, and immune cells. Spatial multiomics technologies enable precise mapping of these regions, offering unprecedented insights into resistance mechanisms and informing the development of niche-targeted therapeutic strategies.

5. Discussion

While single-cell sequencing has significantly advanced our understanding of treatment resistance in NSCLC, its inherent limitations continue to challenge the robustness of research conclusions. Tissue dissociation may induce transcriptional stress responses, distorting the in vivo cellular state [14], but the low capture efficiency of rare resistant subpopulations limits the comprehensive identification of key resistance events. Moreover, high data sparsity and batch effects significantly hinder cross-time-point and cross-sample comparisons, and current annotations of cell clusters remain heavily reliant on prior assumptions, introducing considerable subjectivity.

Single-cell studies are inherently observational-they can reveal correlations between resistance and molecular phenotypes but cannot establish causality. Therefore, any inferred resistance-associated cell subpopulation or biomarker must undergo rigorous functional validation [9], a process that is both complex and costly, substantially limiting its clinical translation.

Although this technology shows promise for translational applications, it still faces multiple bottlenecks. For example, dynamic monitoring based on circulating tumor cells (CTCs) remains constrained by detection sensitivity and insufficient technical standardization. Candidate biomarkers derived from single-cell data-such as intermediate EMT-state cells or specific TAM subtypes-require further validation through spatial multiplex fluorescence techniques and large clinical cohorts before they can serve as clinically useful indicators. Similarly, most mechanism-informed combination therapies (e.g., those that target AXL or the CCL2/CCR2 axis) [6] remain in the preclinical stage, with their actual efficacy and applicable patient populations still awaiting confirmation.

Future progress will rely heavily on technological innovation and multidimensional data integration. Although artificial intelligence and generative models hold promise for predicting clonal evolution or virtually screening drug combinations, their predictive reliability depends critically on data quality and algorithmic interpretability. Multiomics integration (e.g., transcriptomic, epigenomic, and proteomic analyses) could theoretically uncover multilayered regulatory networks underlying resistance [6], yet major challenges remain in terms of technical compatibility and analytical harmonization. Furthermore, functional validation platforms such as organoids combined with CRISPR screening offer new pathways for target verification [4], although their throughput and clinical relevance still require improvement.

In summary, single-cell sequencing has increased resistance to single-cell resolution. However, only by systematically addressing multiple bottlenecks-including technical noise, data analysis limitations, and experimental validation-can we deepen our mechanistic understanding of the resistance ecosystem and closely integrate these insights with those of clinical practice to realize the full potential of this technology in precision medicine.

6. Conclusion

Single-cell sequencing technology has increased research on drug resistance in NSCLC to single-cell resolution, profoundly revealing the critical roles of tumor heterogeneity, cellular plasticity, and the TME in treatment resistance [4, 7]. By identifying rare drug-resistant subpopulations, deciphering nongenetic functional state transitions, and characterizing immunosuppressive niches, this technology provides unprecedented insights into the mechanisms underlying drug resistance [9, 12].

However, the technology still faces multiple challenges. Technically, issues such as tissue dissociation-induced stress responses, low capture efficiency of rare cell subpopulations, high data sparsity, and batch effects constrain the accuracy and reproducibility of the data [14]. Analytically, studies remain largely observational and unable to establish causality directly, and cell type annotation still relies heavily on empirical knowledge, introducing subjectivity. From a translational perspective, most candidate biomarkers and mechanism-informed combination therapies remain in the preclinical validation stage, and their clinical utility urgently requires confirmation through large-scale cohorts and functional experiments [6].

Future breakthroughs will depend on synergistic progress in three key areas: first, technological innovation and integration, necessitating improvements in sample processing protocols; second, the development of highly sensitive liquid biopsy techniques; and third, the convergence of single-cell multiomics with spatial technologies [13]. Second, intelligent computing and data analysis, leveraging artificial intelligence and generative models to enhance the reliability of predicting clonal evolution and virtually screening drug combinations, while also improving algorithmic interpretability. Third, functional validation systems should be strengthened by the use of platforms such as organoid models and CRISPR screening to efficiently verify candidate targets and mechanistic hypotheses [4].

In conclusion, only by systematically addressing the bottlenecks ranging from technical noise and data analysis limitations to clinical validation can the profound insights generated by single-cell sequencing be translated into precise diagnostic and therapeutic strategies that guide clinical practice, ultimately realizing its full potential for improving outcomes in NSCLC patients.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021, 71(3), pp. 209-249. https://doi.org/https://doi.org/10.3322/caac.21660.

[2] Soria, J.-C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K. H., Dechaphunkul, A., Imamura, F., Nogami, N. and Kurata, T. Osimertinib in untreated EGFR-mutated advanced non–small-cell lung cancer. New England Journal of Medicine. 2018, 378(2), pp. 113-125. https://doi.org/10.1056/NEJMoa1713137.

[3] Brahmer, J. R., Lee, J.-S., Ciuleanu, T.-E., Bernabe Caro, R., Nishio, M., Urban, L., Audigier-Valette, C., Lupinacci, L., Sangha, R. and Pluzanski, A. Five-year survival outcomes with nivolumab plus ipilimumab versus chemotherapy as first-line treatment for metastatic non–small-cell lung cancer in CheckMate 227. Journal of Clinical Oncology. 2023, 41(6), pp. 1200-1212. https://doi.org/10.1200/JCO.22.01503.

[4] Maynard, A., McCoach, C. E., Rotow, J. K., Harris, L., Haderk, F., Kerr, D. L., Yu, E. A., Schenk, E. L., Tan, W., Zee, A., et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell. 2020, 182(5), pp. 1232-1251.e22. https://doi.org/10.1016/j.cell.2020.07.017.

[5] Stuart, T. and Satija, R. Integrative single-cell analysis-ProQuest. Available from: https://wvpn.ahu.edu.cn/https/77726476706e6973746856265737421e7e056d2372267416b0d9ab8d6562c38/docview/2210427092?pq-origsite=wos&accountid=31106&sourcetype=Scholarly%20Journals (accessed 8 January 2026).

[6] Hu, J., Chen, Z., Bao, L., Zhou, L., Hou, Y., Liu, L., Xiong, M., Zhang, Y., Wang, B. and Tao, Z. Single-cell transcriptome analysis reveals intratumoral heterogeneity in ccRCC, which results in different clinical outcomes. Molecular Therapy. 2020, 28(7), pp. 1658-1672. https://doi.org/10.1016/j.ymthe.2020.04.023.

[7] Kim, N., Kim, H. K., Lee, K., Hong, Y., Cho, J. H., Choi, J. W., Lee, J.-I., Suh, Y.-L., Ku, B. M. and Eum, H. H. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nature communications. 2020, 11(1), p. 2285. https://doi.org/10.1038/s41467-020-16164-1.

[8] Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. and Teichmann, S. A. The technology and biology of single-cell RNA sequencing. Molecular cell. 2015, 58(4), pp. 610-620. https://doi.org/10.1016/j.molcel.2015.04.005.

[9] Kashima, Y., Shibahara, D., Suzuki, A., Muto, K., Kobayashi, I. S., Plotnick, D., Udagawa, H., Izumi, H., Shibata, Y. and Tanaka, K. Single-cell analyses reveal diverse mechanisms of resistance to EGFR tyrosine kinase inhibitors in lung cancer. Cancer research. 2021, 81(18), pp. 4835-4848. https://doi.org/10.1158/0008-5472.CAN-20-2811.

[10] Gu, C., Liu, S., Wu, Q., Zhang, L. and Guo, F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell research. 2019, 29(2), pp. 110-123. https://doi.org/10.1038/s41422-018-0125-4.

[11] Sathe, A., Grimes, S. M., Lau, B. T., Chen, J., Suarez, C., Huang, R. J., Poultsides, G. and Ji, H. P. Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. Clinical Cancer Research. 2020, 26(11), pp. 2640-2653. https://doi.org/10.1158/1078-0432.CCR-19-3231.

[12] Ji, A. L., Rubin, A. J., Thrane, K., Jiang, S., Reynolds, D. L., Meyers, R. M., Guo, M. G., George, B. M., Mollbrink, A. and Bergenstråhle, J. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. cell. 2020, 182(2), pp. 497-514. e22. https://doi.org/10.1016/j.cell.2020.05.039.

[13] Lomakin, A., Svedlund, J., Strell, C., Gataric, M., Shmatko, A., Rukhovich, G., Park, J. S., Ju, Y. S., Dentro, S. and Kleshchevnikov, V. Spatial genomics maps the structure, nature and evolution of cancer clones. Nature. 2022, 611(7936), pp. 594-602. https://doi.org/10.1038/s41586-022-05425-2.

[14] Haque, A., Engel, J., Teichmann, S. A. and Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome medicine. 2017, 9(1), p. 75. https://doi.org/10.1186/s13073-017-0467-4.

[15] Liang, J., Cai, W. and Sun, Z. Single-cell sequencing technologies: current and future. Journal of Genetics and Genomics. 2014, 41(10), pp. 513-528. https://doi.org/10.1016/j.jgg.2014.09.005.

[16] Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V. and Setty, M. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell. 2018, 174(5), pp. 1293-1308. e36. https://doi.org/16.10.1016/j.cell.2018.05.060.

[17] Jackson, H. W., Fischer, J. R., Zanotelli, V. R., Ali, H. R., Mechera, R., Soysal, S. D., Moch, H., Muenst, S., Varga, Z. and Weber, W. P. The single-cell pathology landscape of breast cancer. Nature. 2020, 578(7796), pp. 615-620. https://doi.org/10.1038/s41586-019-1876-x.

[18] Li, J. J., Tsang, J. Y. and Tse, G. M. Tumor microenvironment in breast cancer-updates on therapeutic implications and pathologic assessment. Cancers. 2021, 13(16), p. 4233. https://doi.org/10.3390/cancers13164233.

[19] Müller, S., Cho, A., Liu, S. J., Lim, D. A. and Diaz, A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. Bioinformatics. 2018, 34(18), pp. 3217-3219. https://doi.org/10.1093/bioinformatics/bty316.

[20] Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P. and Aerts, J. SCENIC: single-cell regulatory network inference and clustering. Nature methods. 2017, 14(11), pp. 1083-1086. https://doi.org/10.1038/nmeth.4463.

[21] Wu, S. Z., Roden, D. L., Wang, C., Holliday, H., Harvey, K., Cazet, A. S., Murphy, K. J., Pereira, B., Al-Eryani, G. and Bartonicek, N. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. The EMBO journal. 2020, 39(19), p. e104063. https://doi.org/10.15252/embj.2019104063.

[22] Barker, H. E., Paget, J. T., Khan, A. A. and Harrington, K. J. The tumour microenvironment after radiotherapy: mechanisms of resistance and recurrence. Nature Reviews Cancer. 2015, 15(7), pp. 409-425. https://doi.org/10.1038/nrc3958.

[23] Park, Y. H., Lal, S., Lee, J. E., Choi, Y.-L., Wen, J., Ram, S., Ding, Y., Lee, S.-H., Powell, E. and Lee, S. K. Chemotherapy induces dynamic immune responses in breast cancers that impact treatment outcome. Nature communications. 2020, 11(1), p. 6175. https://doi.org/10.1038/s41467-020-19933-0.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering
(IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260109

Progress in Single-Cell RNA Sequencing Data Analysis and Its Applications in the Study of Tumor Heterogeneity

Qingsong Cai*

College of Life Sciences, Sichuan University, Chengdu, Sichuan, 610000, China

**Corresponding author: Qingsong Cai*

Abstract

Tumors are heterogeneous diseases, and their transition to malignancy is often subtle; however, advanced computational tools can elucidate their development and drug resistance. Single-cell RNA sequencing offers high-resolution insights into cellular and molecular changes, enhancing our understanding of cancer dynamics. This review summarizes the important technological breakthroughs since the birth of single-cell sequencing technology and the recent research results concerning tumor heterogeneity and "precision medicine" through single-cell RNA sequencing. Overall, the future of single-cell RNA sequencing in tumor treatment is undoubtedly promising.

Keywords

application, research progress, single-cell sequencing, tumor heterogeneity

1. Introduction

Tumors are systemic diseases and major global challenges that have posed a threat to human health since the early twenty-first century. They undergo critical transformations—from premalignant to malignant states: from locally contained to metastatic disease and from treatment-responsive to treatment-resistant forms [1]. The complexity of treating tumors is exacerbated by their heterogeneity and the intricate composition of the tumor immune microenvironment (TIME), which significantly impacts tumorigenesis, progression, invasion, metastasis, and drug resistance [2-5]. Advances in sequencing technology have facilitated the analysis of data from individual cancer cells, ushering in the era of "precision medicine", which relies on precise molecular data [6, 7]. Single-cell RNA sequencing (scRNA-seq) offers considerable advantages over bulk RNA sequencing (RNA-seq), which provides only limited insights into tumor heterogeneity and average gene expression values of a sample. scRNA-seq is particularly advantageous because of its ability to characterize the frequency of cell types within each sample and to elucidate relationships between cells or cell types and its high-resolution cracking of the TIME [8, 9]. Owing to these advantages, it is significantly superior to previous sequencing techniques in advancing our understanding of the human biology of various cells, especially immune and tumor cells, and since the first single-cell mRNA sequencing in 2009 and the first human cancer cell sequencing in 2011, single-cell sequencing analysis based on next-generation sequencing has gained

attention and developed rapidly [5]. ScRNA-seq is becoming a popular tool in cancer research to explain disease heterogeneity [10]. ScRNA-seq provides biological information on individual tumor cells, analyzes determinants of gene expression heterogeneity within tumors, and determines the molecular basis of the formation of many tumors [11].

In this review, we describe the recent progress in scRNA-seq and summarize its application in the study of tumor heterogeneity, with a focus on how this technology can impact precision medicine in the future.

2. Advances in scRNA-seq

Single-cell RNA sequencing (scRNA-seq) has revolutionized the study of cellular heterogeneity, providing unprecedented high-resolution analytical capability for dissecting the tumor microenvironment and revealing the molecular diversity of cancer cells. The standard workflow mainly includes single-cell isolation, reverse transcription of RNA into cDNA, template preamplification, and high-throughput sequencing analysis [12].

Owing to the extremely low amount of RNA in individual cells, efficiently amplifying limited genetic material has become a core technical challenge in scRNA-seq [13]. Currently, commonly used amplification strategies include polymerase chain reaction (PCR)-based amplification (e.g., SMART-Seq2) and in vitro transcription-based amplification (e.g., CEL-Seq2). For cell isolation, mainstream methods include microfluidic systems, droplet-based technologies, and microwell arrays [14]. Although early representative platforms such as Fluidigm C1 could obtain high-quality gene expression data, constraints in throughput and cost limitations have led to their predominant application in small-scale sample studies [15-17].

The emergence of droplet-based technologies (e.g., Drop-seq and inDrop) marked a significant breakthrough in scRNA-seq. In 2017, 10× Genomics integrated related technologies to launch a commercial platform that greatly increased cell throughput and significantly reduced costs, thereby making large-scale tumor studies feasible [18-21].

Particularly critical is that technical approaches capable of obtaining full-length transcripts, such as SMART-Seq2, demonstrate unique advantages in alternative splicing analysis, allele-specific expression detection, and RNA editing identification owing to their comprehensive coverage of cDNA sequences, greatly facilitating in-depth exploration of tumor heterogeneity [22-24]. Subsequent systems such as SMART-seq3 and BD Rhapsody have further achieved remarkable improvements in detection sensitivity and scalability [10, 25, 26].

In recent years, scRNA-seq technology has continued to advance in terms of increased throughput, multimodal integration, and spatial resolution. For example, the Stereo-cell technology released in 2025, which is based on high-density DNA nanoball arrays, enables in situ transcriptome capture without relying on droplet microfluidics, simultaneously integrating transcriptome, protein signals, and morphological information. This provides a novel perspective for multidimensional analysis of tumor heterogeneity and tissue architecture [27].

3. Some New and In-Depth Applications in Studying Tumor Heterogeneity

Tumor heterogeneity, including inter- and intratumor heterogeneity, is a key feature of malignant tumors and refers to molecular biological or genetic alterations in daughter cells during the proliferation of tumor cells, which is a significant obstacle to cancer treatment and research [28]. Recognizing tumor heterogeneity is key to further understanding and treating cancer, and the use of scRNA-seq for research has reached a consensus. scRNA-seq has been used to promote cancer diagnosis and prognosis prediction, increase comprehension of disease progression and cancer metastasis, and guide treatment [29-33]. In recent years, single-cell sequencing techniques have often been combined with other techniques, such as spatial transcriptomics, to deepen our understanding of tumor heterogeneity, from the tumor cells themselves to the various components of the TIME.

A study published in the *Journal of Cellular and Molecular Medicine* integrated single-cell and bulk RNA sequencing data with machine learning algorithms to construct an 11-gene (including ANXA3, APOE, CENPA, CKS1B) Prostate Cancer Meta-Program (PCMP) signature model. This model effectively predicts prognosis and revealed that the PCMP signature promotes epithelial cell malignant transformation by regulating the cell cycle and oxidative phosphorylation pathways. Experimental validation has shown that knocking down CENPA or CKS1B significantly inhibits PC3 cell proliferation [34].

Research published in *JECCR* by a team from Shanghai Pharmaceuticals in collaboration with Wenzhou Medical University combined single-cell sequencing with deep learning to develop a new method for accurately identifying tumor-reactive CD8+ T cells directly from tumor-infiltrating lymphocytes (TILs) without prior knowledge of neoantigens. The AUC of this deep learning model reached 0.958 in the training set. The study revealed that more than one-third of TR T cells were distributed in the GZMK+ effector memory cell subset, challenging the traditional belief that "TR cells exist only in exhausted populations". All types of TR CD8+ T cells commonly downregulated genes related to the mitochondrial respiratory chain pathway (e.g., MT-ATP6 and MT-ND1), suggesting that impaired mitochondrial function is a key feature. This study identified TIGIT-NECTIN2 as a previously overlooked important immune checkpoint in pancreatic cancer. TCR analysis revealed that the "clonal expansiveness" of the TCR alone could not determine tumor reactivity, and some TR CD8+ T cells simultaneously expressed multiple different TCR α/β chains. These multi-TCR combinations demonstrated stronger tumor-killing capacity. Clinical translational analysis revealed that pancreatic cancer patients with a high proportion of TR CD8+ T cells had a better prognosis after receiving neoadjuvant immunotherapy (e.g., GVAX vaccine ± nivolumab), suggesting that TR CD8+ T cells could serve as biomarkers for predicting immunotherapy efficacy [35].

A study published in *Hereditas* revealed the heterogeneity of cancer-associated fibroblasts (CAFs) in the glioblastoma (GBM) microenvironment by integrating scRNA-seq and bulk RNA-seq data and identified 5 CAF subsets, 3 of which were significantly associated with prognosis. Researchers constructed a three-gene risk model based on LITAF, OSMR, and TCF12, which showed good predictive performance in the TCGA and CGGA cohorts (1-year AUC=0.74). High-risk patients had a lower response rate to PD-L1 inhibitors, suggesting that CAFs might mediate immune escape through the IL6-JAK-STAT3 pathway [36].

A research team from Tongji University published a study in *Nature Cancer* 7. They integrated scRNA-seq data from over 4 million single cells from 746 samples across 36 cancer types in public datasets, along with spatial transcriptomics data from 62 samples of 6 cancer types, to establish a pancancer atlas of tumor microenvironment heterogeneity—the TabulaTIME. This atlas defines six major cell lineages and 56 cell subtypes in the TME. CTHRC1 is a marker for extracellular matrix-related CAFs, which are enriched in various cancer types and may be located at the interface between malignant and normal areas, preventing immune cell infiltration. Furthermore, SLPI+ macrophages exhibit a profibrotic phenotype and colocalize with CTHRC1+ CAFs to form unique spatial ecotypes [37].

The study of tumor heterogeneity has revealed that modern medicine has a strong pursuit of "precision medicine", and single-cell sequencing technology, which focuses on the subtle differences between cells, undoubtedly plays an important role in this process. Therefore, deciphering tumor heterogeneity is not only fundamental to understanding cancer biology but also critical for developing effective targeted therapies and overcoming treatment resistance.

4. Conclusions

Since its introduction in 2009, single-cell sequencing technology has developed rapidly, and its powerful analytical capabilities have provided unprecedented help for tumor research. It provides high-resolution transcript sequencing, analyzes intra- and intertumor heterogeneity, and studies the tumor immune microenvironment, making it an increasingly indispensable tool for studying tumor initiation, progression, treatment response, and drug resistance.

However, it is important to recognize that scRNA-seq still has many drawbacks. First, scRNA-seq can capture only a transient "snapshot" of the transcriptome, suffering from transcriptional bursting and insensitivity to low-abundance transcripts. More critically, it fails to provide multiomics information (e.g., genomic variations, protein expression, and epigenetic states), leading to a fragmented and incomplete interpretation of cellular states. Second, the technology heavily relies on live cells. The tissue dissociation process (enzymatic digestion, mechanical force) can induce stress responses, kill cells, degrade RNA, and completely obliterate crucial spatial and morphological contexts. Furthermore, not all cells are captured equally, and differences in platforms or processing dates introduce stubborn batch effects, compromising data quality and comparability. Third, the technology remains expensive, time-consuming, and complex in terms of data analysis, hindering its application in large-scale clinical cohorts. Individual patient differences and the use of varying platforms across studies further limit the generalizability and reliability of findings.

Consequently, current discoveries require cautious validation in clinical trials, and there is still a gap before routine clinical diagnostic application [5, 38].

Therefore, future research directions should be carried out in the direction of higher throughput, higher sensitivity, and lower cost technologies (such as the development of stereo-cells to trillion-level cell fluxes [27]) and should be more deeply integrated with other technologies, such as spatial multiomics, which will help breakdown technical barriers. In addition, in this era of increasingly advanced artificial intelligence technology worldwide, AI and machine learning will play important roles in building cell RNA libraries so that single-cell sequencing technology will play a greater role in cell type identification, gene regulation, disease state prediction, drug response prediction and neoantigen identification. In terms of applications, single-cell sequencing will be used more deeply to promote "precision medicine", such as discovering more biomarkers of disease and tracking tumor evolution at the cellular level to find evidence of drug resistance. We believe that single-cell sequencing technology can be better developed in the future to help humans find cures for more diseases.

References

[1] Rozenblatt-Rosen, O., Regev, A., Oberdoerffer, P., Nawy, T., Hupalowska, A., Rood, J. E., Ashenberg, O., Cerami, E., Coffey, R. J., Demir, E., et al. The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. Cell. 2020, 181(2), pp. 236-249. https://doi.org/10.1016/j.cell.2020.03.053.

[2] Vitale, I., Shema, E., Loi, S. and Galluzzi, L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nature Medicine. 2021, 27(2), pp. 212-224. https://doi.org/10.1038/s41591-021-01233-9.

[3] Liu, J., Xu, T., Jin, Y., Huang, B. and Zhang, Y. Progress and Clinical Application of Single-Cell Transcriptional Sequencing Technology in Cancer Research. Frontiers in Oncology. 2021, 10, p. 593085. https://doi.org/10.3389/fonc.2020.593085.

[4] Erfanian, N., Derakhshani, A., Nasseri, S., Fereidouni, M., Baradaran, B., Jalili Tabrizi, N., Brunetti, O., Bernardini, R., Silvestris, N. and Safarpour, H. Immunotherapy of cancer in single-cell RNA sequencing era: A precision medicine perspective. Biomedicine & Pharmacotherapy. 2022, 146, p. 112558. https://doi.org/10.1016/j.biopha.2021.112558.

[5] Lei, Y., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., Yu, X. and Shi, S. Applications of single-cell sequencing in cancer research: progress and perspectives. Journal of Hematology & Oncology. 2021, 14(1), p. 91. https://doi.org/10.1186/s13045-021-01105-2.

[6] Amelio, I., Bertolo, R., Bove, P., Buonomo, O. C., Candi, E., Chiocchi, M., Cipriani, C., Di Daniele, N., Ganini, C., Juhl, H., et al. Liquid biopsies and cancer omics. Cell Death Discovery. 2020, 6(1), p. 131. https://doi.org/10.1038/s41420-020-00373-0.

[7] Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L. and Siu, L. L. Molecular profiling for precision cancer therapies. Genome Medicine. 2020, 12(1), p. 8. https://doi.org/10.1186/s13073-019-0703-1.

[8] Bühler, M. M., Martin-Subero, J. I., Pan-Hammarström, Q., Campo, E. and Rosenquist, R. Towards precision medicine in lymphoid malignancies. Journal of Internal Medicine. 2022, 292(2), pp. 221-242. https://doi.org/https://doi.org/10.1111/joim.13423.

[9] Ding, J., Sharon, N. and Bar-Joseph, Z. Temporal modelling using single-cell transcriptomics. Nature Reviews Genetics. 2022, 23(6), pp. 355-368. https://doi.org/10.1038/s41576-021-00444-7.

[10] Li, P.-H., Kong, X.-Y., He, Y.-Z., Liu, Y., Peng, X., Li, Z.-H., Xu, H., Luo, H. and Park, J. Recent developments in application of single-cell RNA sequencing in the tumour immune microenvironment and

cancer therapy. Military Medical Research. 2022, 9(1), p. 52. https://doi.org/10.1186/s40779-022-00414-y.

[11] Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D. and Zhang, H. RNA sequencing: new technologies and applications in cancer research. Journal of Hematology & Oncology. 2020, 13(1), p. 166. https://doi.org/10.1186/s13045-020-01005-x.

[12] Gawad, C., Koh, W. and Quake, S. R. Single-cell genome sequencing: current state of the science. Nature Reviews Genetics. 2016, 17(3), pp. 175-188. https://doi.org/10.1038/nrg.2015.16.

[13] Picelli, S. Single-cell RNA-sequencing: The future of genome biology is now. RNA Biology. 2017, 14(5), pp. 637-650. https://doi.org/10.1080/15476286.2016.1201618.

[14] Zweifel, P., Felder, S. and Meiers, M. Ageing of population and health care expenditure: a red herring? Health Economics. 1999, 8(6), pp. 485-496. https://doi.org/https://doi.org/10.1002/(SICI)1099-1050(199909)8:6<485::AID-HEC461>3.0.CO;2-4.

[15] Potter, S. S. Single-cell RNA sequencing for the study of development, physiology and disease. Nature Reviews Nephrology. 2018, 14(8), pp. 479-492. https://doi.org/10.1038/s41581-018-0021-7.

[16]Xin, Y., Kim, J., Ni, M., Wei, Y., Okamoto, H., Lee, J., Adler, C., Cavino, K., Murphy, A. J., Yancopoulos, G. D., et al. Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proceedings of the National Academy of Sciences. 2016, 113(12), pp. 3293-3298. https://doi.org/10.1073/pnas.1602306113.

[17] DeLaughter, D. M. The Use of the Fluidigm C1 for RNA Expression Analyses of Single Cells. Current Protocols in Molecular Biology. 2018, 122(1), p. e55. https://doi.org/https://doi.org/10.1002/cpmb.55.

[18] Klein, Allon M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, David A. and Kirschner, Marc W. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell. 2015, 161(5), pp. 1187-1201. https://doi.org/10.1016/j.cell.2015.04.044.

[19] Macosko, Evan Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, Allison R., Kamitaki, N., Martersteck, Emily M., et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015, 161(5), pp. 1202-1214. https://doi.org/10.1016/j.cell.2015.05.002.

[20] Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017, 8(1), p. 14049. https://doi.org/10.1038/ncomms14049.

[21] Papalexi, E. and Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature Reviews Immunology. 2018, 18(1), pp. 35-45. https://doi.org/10.1038/nri.2017.76.

[22] Ding, S., Chen, X. and Shen, K. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Communications. 2020, 40(8), pp. 329-344. https://doi.org/https://doi.org/10.1002/cac2.12078.

[23] Hochgerner, H., Lönnerberg, P., Hodge, R., Mikes, J., Heskol, A., Hubschle, H., Lin, P., Picelli, S., La Manno, G., Ratz, M., et al. STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. Scientific Reports. 2017, 7(1), p. 16327. https://doi.org/10.1038/s41598-017-16546-4.

[24] Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtukova, I., Loring, J. F., Laurent, L. C., et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature Biotechnology. 2012, 30(8), pp. 777-782. https://doi.org/10.1038/nbt.2282.

[25] Fan, H. C., Fu, G. K. and Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. Science. 2015, 347(6222), p. 1258367. https://doi.org/10.1126/science.1258367.

[26] Shum, E. Y., Walczak, E. M., Chang, C. and Christina Fan, H., (2019). Quantitation of mRNA Transcripts and Proteins Using the BD Rhapsody™ Single-Cell Analysis System. In Suzuki, Y. (ed.) *Single Molecule and Single Cell Sequencing*. Singapore: Springer Singapore, pp. 63-79.

[27] Liao, S., Zhou, X., Liu, C., Liu, C., Hao, S., Luo, H., Hou, H., Liu, Q., Zhang, Z., Xiao, L., et al. Stereo-cell: Spatial enhanced-resolution single-cell sequencing with high-density DNA nanoball-patterned arrays. Science, 389(6762), p. eadr0475. https://doi.org/10.1126/science.adr0475.

[28] Liu, J., Dang, H. and Wang, X. W. The significance of intertumor and intratumor heterogeneity in liver cancer. Experimental & Molecular Medicine. 2018, 50(1), pp. e416-e416. https://doi.org/10.1038/emm.2017.165.

[29] Ruan, H., Wang, Z., Zhai, Y., Xu, Y., Pi, L., Zheng, J., Zhou, Y., Zhang, C., Huang, R., Chen, K., et al. Single-cell transcriptome analysis of diffuse large B cells in cerebrospinal fluid of central nervous system lymphoma. iScience. 2021, 24(9), p. 102972. https://doi.org/10.1016/j.isci.2021.102972.

[30] Wang, R., Dang, M., Harada, K., Han, G., Wang, F., Pool Pizzi, M., Zhao, M., Tatlonghari, G., Zhang, S., Hao, D., et al. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. Nature Medicine. 2021, 27(1), pp. 141-151. https://doi.org/10.1038/s41591-020-1125-8.

[31] Borcherding, N., Voigt, A. P., Liu, V., Link, B. K., Zhang, W. and Jabbari, A. Single-Cell Profiling of Cutaneous T-Cell Lymphoma Reveals Underlying Heterogeneity Associated with Disease Progression. Clinical Cancer Research. 2019, 25(10), pp. 2996-3005. https://doi.org/10.1158/1078-0432.CCR-18-3309.

[32] Davis, R. T., Blake, K., Ma, D., Gabra, M. B. I., Hernandez, G. A., Phung, A. T., Yang, Y., Maurer, D., Lefebvre, A. E. Y. T., Alshetaiwi, H., et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. Nature Cell Biology. 2020, 22(3), pp. 310-320. https://doi.org/10.1038/s41556-020-0477-0.

[33] Cao, X., Wang, Y., Zhang, W., Zhong, X., Gunes, E. G., Dang, J., Wang, J., Epstein, A. L., Querfeld, C., Sun, Z., et al. Targeting macrophages for enhancing CD47 blockade–elicited lymphoma clearance and overcoming tumor-induced immunosuppression. Blood. 2022, 139(22), pp. 3290-3302. https://doi.org/10.1182/blood.2021013901.

[34] Wu, J., Chen, Z., Wu, W., Qin, J., Zhong, R., Meng, J., Yin, Y., Guo, P. and Fan, S. Single-Cell and Bulk RNA Sequencing Highlights Intra-Tumoral Heterogeneity and Malignant Progression Mechanisms in Prostate Cancer. Journal of Cellular and Molecular Medicine. 2025, 29(16), p. e70806. https://doi.org/https://doi.org/10.1111/jcmm.70806.

[35] Sun, H., Shi, C., Fang, G., Guo, Q., Du, Z., Chen, G., Wu, Y., Chen, Z.-S., Hua, J., Zhang, Y., et al. Functional tumor-reactive CD8 + T cells in pancreatic cancer. Journal of Experimental & Clinical Cancer Research. 2025, 44(1), p. 253. https://doi.org/10.1186/s13046-025-03517-1.

[36] Zhang, W., Li, Y., Li, C. and Huang, Q. Integrated single-cell and bulk transcriptomic profiling reveals cancer-associated fibroblast heterogeneity in glioblastoma and establishes a clinically actionable prognostic model and preliminary experimental validation. Hereditas. 2025, 162(1), p. 173. https://doi.org/10.1186/s41065-025-00548-8.

[37] Han, Y., Zhang, L., Sun, D., Cao, G., Wang, Y., Yue, J., Hu, J., Dong, Z., Li, F., Li, T., et al. Spatiotemporal analyses of the pan-cancer single-cell landscape reveal widespread profibrotic ecotypes

associated with tumor immunity. Nature Cancer. 2025, 6(11), pp. 1880-1898. https://doi.org/10.1038/s43018-025-01039-5.

[38] Huang, D., Ma, N., Li, X., Gou, Y., Duan, Y., Liu, B., Xia, J., Zhao, X., Wang, X., Li, Q., et al. Advances in single-cell RNA sequencing and its applications in cancer research. Journal of Hematology & Oncology. 2023, 16(1), p. 98. https://doi.org/10.1186/s13045-023-01494-6.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260210

Single-cell RNA sequencing reveals the dynamic evolution of the immune microenvironment in coronary artery atherosclerosis

Sixuan Xiang*

West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, China

**Corresponding author: Sixuan Xiang, ORCID: 0009-0001-9994-3329*

Abstract

Coronary artery disease (CAD), which is driven primarily by atherosclerosis, represents a major global health burden. This review explores the dynamic evolution of the CAD immune microenvironment through single-cell RNA sequencing (scRNA-seq), revealing cellular heterogeneity and interactions. On the basis of the American Heart Association's histological classification of atherosclerotic lesions, we systematically summarize discoveries supported by scRNA-seq across disease stages (types I-V) from early monocyte recruitment and lipid-resistant macrophage subpopulations (e.g., CD52-hi macrophages) in initial lesions to intensified inflammation involving T-cell activation, NK cell differentiation, and macrophage polarization in fatty streaks and intermediate lesions and late-stage fibrosis and T-cell clonal expansion leading to plaque instability. Key findings highlight immune imbalance exacerbated by comorbidities such as diabetes and systemic lupus erythematosus (SLE), with potential biomarkers (e.g., JUN) and therapeutic targets (e.g., the CCL4–CCR5 axis and SPP1+ macrophages). Methodological limitations, such as the lack of spatial information and challenges in inferring causal relationships, are discussed. Future prospects, such as spatial transcriptomics integration, multiomics approaches, and AI-assisted precision medicine, are also proposed. This review highlights the transformative role of scRNA-seq in advancing CAD pathology toward precision diagnostics and therapies.

Keywords

single-cell RNA sequencing, coronary artery disease, immune microenvironment, atherosclerosis, cellular heterogeneity

1. Introduction

Coronary artery disease (CAD) is one of the leading causes of death worldwide[1, 2]. The global burden of CAD is reflected not only in high mortality rates but also in high morbidity and healthcare costs. Many CAD patients require long-term medication therapy, interventional treatments (such as percutaneous coronary intervention (PCI)), or surgical treatments (such as coronary artery bypass grafting (CABG)) [2, 3]. These treatment methods not only impose financial burdens on patients but also place enormous pressure on healthcare systems. Therefore, reducing the global burden of CAD requires collective efforts worldwide,

including raising public health awareness, improving lifestyles, strengthening early screening and diagnosis, and developing more effective treatment methods.

In recent years, a growing body of research has demonstrated that the immune microenvironment plays a crucial role in the occurrence and development of CAD[4-6]. Atherosclerosis, the primary pathological mechanism underlying CAD, is recognized as a chronic inflammatory disease [7]. In its early stages, immune cells such as monocytes and T cells infiltrate the arterial intima [8], where inflammatory factors affect them to activate vascular endothelial cells, promoting lipid deposition and oxidation that lead to plaque formation. Multiple immune cell subsets, including macrophages, T cells, B cells, and natural killer cells, exist in CAD patients [8]. These immune cells exacerbate vascular inflammatory responses and promote the formation and rupture of atherosclerotic plaques by releasing inflammatory factors and cytotoxic molecules.

During the development of coronary atherosclerotic disease (CAD), the composition and function of resident immune cells play crucial roles in plaque formation, stabilization, and rupture. However, traditional bulk transcriptomics and histology have difficulty differentiating cell subpopulations and their transitions at the single-cell level. Single-cell RNA sequencing (scRNA-seq) provides individual-cell resolution, making it possible to analyze the heterogeneity and functional differentiation of immune cells and their cellular interactions [9].

On the basis of the AHA's classification of atherosclerosis progression and further research [10-14], this paper systematically reviews the major biological discoveries, methodological limitations of scRNA-seq in CAD research, and technologies and research methods that are expected to solve these problems, aiming to identify verifiable and translatable directions for future research.

2. Single-Cell RNA Sequencing Technologies

To understand how single-cell RNA sequencing (scRNA-seq) shapes the current understanding of the immune microenvironment of CAD, this section outlines the key sample preparation, sequencing platforms, and commonly used analysis procedures and focuses on the application of this technology in CAD research. The first two steps of scRNA-seq involve separating individual cells and constructing sequencing libraries. The methods for cell separation include mechanical separation, enzymatic separation, and flow cytometry (FACS) or microfluidic sorting [15, 16]. The methods for library construction mainly include the Smart-seq series, the 10x Genomics platform, and Drop-seq, among which each method has its own characteristics and is suitable for different experimental needs [17-19]. Luecken et al. provided a tutorial of workflows for single-cell RNA-seq analysis [20], including preprocessing and downstream analysis. Raw data are processed to obtain count matrices, and quality control, normalization, dimensionality reduction, visualization and other steps are usually performed to provide a reliable, low-dimensional, and comparable gene expression matrix. After preprocessing, researchers can conduct clustering, trajectory inference, differential expression and other analytic steps. With various protocols and analysis tools, scRNA-seq reveals great advantages in the inference of cell types, functional states, and interaction networks in CAD. Moreover, this technology can also be integrated with multiomics (e.g., CITE-seq, scATAC-seq), providing a more comprehensive perspective.

Single-cell RNA sequencing (scRNA-seq) has shifted from static bulk RNA analysis, which masks cellular heterogeneity by averaging signals, to single-cell resolution, revealing dynamic immune trajectories and subpopulations in the coronary artery disease (CAD) immune microenvironment. These findings provide detailed insights into atherosclerosis progression across pathological stages. Overall, tools based on scRNA-seq provide a robust framework for researching the immune dynamics of CAD, as detailed in the following sections.

3. Main Body

3.1 Type I: Initial Lesion

In this stage, the immune environment is characterized mainly by mild inflammation, with the accumulation of isolated macrophages. Its major activities include endothelial injury, initial recruitment of monocytes and preliminary formation of foam cells.

In the coronary vascular wall, scRNAA-seq, along with snATAC-seq, has revealed 14 different cell types, including smooth muscle smooth muscle cells (SMCs), endothelial cells, fibroblasts, macrophages, T cells and other types of cells [21]. Analysis at the mononuclear level has greatly improved the ability to distinguish rare cell types. In the development of arterial atherosclerosis, an increased proportion of immune cells indicates their participation in the early stage of lesion development[21]. To understand how these cells contribute to disease, further research has investigated the gene regulatory mechanisms affected by CAD risk variants. The study revealed that the specific chromatin accessible region of these cell types may be affected by CAD risk genes[21]. In support of this result, the enrichment of crucial TFs (e.g., TCF21, PRDM16, and TBX2) [21] in several specific cell types (e.g., SMCs and fibroblasts) suggests their elusive roles in the early initiation or maintenance of cellular phenotype transformation in disease.

Kirichenko et al. revealed the increased proinflammatory activation of monocytes in diabetes mellitus and suggested that it is the cause of accelerated atherogenesis. These findings suggest that the development of atherosclerosis may be affected by comorbidities [22]. Indeed, monocyte recruitment is one of the main pathological manifestations in this stage, as explored in another study investigating monocyte heterogeneity in the initiation of artery atherosclerosis [23]. Using advanced single-cell technologies such as CITE-seq and single-cell RNA sequencing, this study identified novel monocyte subpopulations—such as MHCIIhi, IFN-responsive, and monocyte–platelet aggregates—and their associations with CAD risk factors such as hyperlipidemia, smoking, and race. These subpopulations exhibit distinct transcriptional and phenotypic profiles, with specific alterations linked to CAD risk states, suggesting their potential as biomarkers or therapeutic targets for CAD.

The preliminary formation of foam cells is another pathological manifestation of the initial lesion. After being recruited under the vascular endothelium, monocytes, crucial infiltrating cells, differentiate into macrophages [24] and accumulate, providing the basis for foam cell formation. Using scRNA-seq and multitomic approaches, Jiang et al. [25] discovered that CD52-hi macrophages, which are enriched in the genetic heritability of CAD, accumulate significantly fewer lipoproteins with activated lipid metabolism pathways. It has a strong ability to remove and process lipoproteins in the early stage of arterial atherosclerosis. The mechanisms may involve the upregulation of certain lipid transport proteins and metabolic enzymes or the downregulation of pathways leading to lipid accumulation to resist lipid overload. This ability can be helpful in slowing down or prohibiting the formation and development of plaques.

In the initial lesion stage, scRNA-seq and snATAC-seq identify diverse cell types within the microenvironment and regulatory impacts from CAD risk genes. Single-cell RNA sequencing revealed the heterogeneity of monocytes and the lipid resistance mechanism of CD52hi macrophages, laying the foundation for atherosclerosis. Although the inflammation was mild, risk factors such as diabetes exacerbated monocyte activation, suggesting that prevention strategies should target initial recruitment. Discoveries at this stage provided molecular insights into the transition to fatty streaks.

3.2 Type II Fatty Streaks

In this stage, along with intimal thickening, multilayered foam cells, and lipid deposition, the inflammatory microenvironment intensifies, with T cells and B cells beginning to participate, and proinflammatory factors increase.

Dendritic cells and macrophages play critical roles in promoting inflammation in the development of arterial atherosclerosis. Dendritic cells present antigens to T lymphocytes to bridge innate and adaptive immune responses. Macrophages exhibit polarization during atherosclerosis, with M1-type macrophages and M2-type macrophages showing opposite responses to inflammation [26]. Although M1/M2 polarization provides a useful framework for understanding the function of macrophages, in the complex microenvironment of atherosclerosis, the behavior of macrophages is far from simple binary opposition. They may exhibit a continuous spectrum or a mixed phenotype. Its function may be highly dependent on dynamic changes in local signals (such as oxidized lipids, cytokines, and hypoxia). In addition, in the same study, T cells and macrophages were linked to the release of DAMPs and PRR activation in artery atherosclerosis, but the causal relationship is still unknown. Research by Hortsmann revealed enrichment of proliferation-associated pathways in several subtypes of NK cells and T cells and pathways linked with immune cell activation, signaling through immune receptors, and differentiation in T and B lymphocytes [27]. The detection of

restricted T-cell receptor sequences in multiple adaptive immune cell subtypes from CAD patients points to enhanced cellular clonality and antigen specificity. These findings indicate the specific differentiation of T cells in CAD [27]. These results validated the immunological expansion in the fatty streak stage at the molecular level.

The research team of Li selected a crucial C1 RACK1+ NK cell subgroup, which has differential potential in the formation of coronary plaques according to CytoTrace [28]. Although NK cells do not participate in lipid deposition directly, the C1 RACK1+ NK cell subset may indirectly regulate the lipid uptake ability of macrophages or smooth muscle cells by secreting cytokines or chemokines, thus altering the speed of lipid deposition. Research in this area could provide a key mechanistic basis for understanding the inflammatory cascade of atherosclerosis and for developing antiplaque therapeutic strategies targeting NK cell function.

In this stage, more macrophages generally transform into foam cells by engulfing lipids, revealing the core characteristics of plaque formation. Research has also revealed that the activation and differentiation of T cells, especially CD4+ T cells, regulate inflammatory reactions. The number of recruited cells in the Type I stage further increases, and these cells differentiate [24]. For example, SMCs display significant phenotypic plasticity in the development of artery atherosclerosis and can differentiate into fibromyocytes, which are associated with specific TFs (e.g., AP-1, TCF21) and genes (TNFRSF11B, FN1) [21]. In the same study, some TFs (e.g., PU.1/SPIB, IRF) and CAD risk gene loci (e.g., GWAS SNPs) were significantly enriched in macrophages during the progression of artery atherosclerosis [21]. These regulatory relationships between cells and their regulators are critical for understanding pathological mechanisms and immune reactions in artery atherosclerosis.

Another study [29]explored the characteristics of peripheral circulating immune cells and identified JUN as a highly active regulon within monocyte clusters via scRNA-seq. Further analysis revealed that JUN expression was significantly upregulated in the CAD group. On the basis of these findings, JUN may serve as a potential diagnostic biomarker and predictor for CAD, suggesting that the JUN signaling pathway is a possible therapeutic target. This study reveals the significant activity of the JUN regulon in single cells of CAD patients through advanced single-cell analysis technology, which is a very promising finding. However, correlation analysis alone is not sufficient to establish causality. Is the upregulation of JUN a driving factor for the initiation of CAD or a secondary reaction in the atherosclerotic inflammatory environment? Moreover, does the sample cohort of the study cover patients of different stages and subtypes of CAD to ensure the general applicability of JUN as a biomarker? These issues require further large-scale and more refined stratified research to be confirmed.

The analysis of scRNA-seq data during the fatty streak stage revealed an increase in the inflammatory response, including the differentiation of natural killer (NK) cells, macrophage polarization, and an increase in biomarkers such as JUN. These results emphasize how immune cells evolve from initial accumulation to more specialized functions driven by local signaling. This process contributes to the development of more complex intermediate plaques.

3.3 Type III: Intermediate

This stage focuses on the formation of extracellular lipid pools and the complexity of plaques, characterized by an increase in cell interactions within the immune microenvironment, such as the migration of macrophages and smooth muscle cells.

During this pathological period, the complexity of plaques is affected by various factors. Single-cell RNA sequencing (scRNA-seq) and a series of related technologies have revealed high expression of CCL4 in T cells, monocytes and macrophages and increased expression of its receptor CCR5 in atherosclerotic plaques compared with controls [30]. These results suggest that blocking the CCL4–CCR5 interaction is a promising therapeutic target for CAD. However, the effect of blocking the interaction between the CCL4 ligand and its receptor on immune responses in people with CAD still needs to be investigated.

The effect of comorbidities on plaques deserves more attention. Unlike the role of the cytokine IL-1 in the pathogenesis of coronary artery disease (CAD) observed in the general population, the results of this transcriptomic study indicate that type I interferon signaling plays a crucial role in cardiovascular inflammation [31]. When CAD is associated with different diseases, the main pathological factors may change, and the study

of these changes has important clinical significance. Similarly, another study on circadian rhythm preliminarily revealed a strong correlation between elevated CRD scores, increased immunoinflammatory activation, and reduced fibrosis [32]. This association suggests a potential mechanism that could drive the unstable transformation of atherosclerotic plaques.

In addition, the complexity of plaques could be associated with the polarization of cells, such as type II macrophages [26]. For example, a study investigating specific aortic valve diseases via paired scRNA-seq revealed both pro- and anti-inflammatory roles of T cells, and if similar dual-role cells can be found in atherosclerosis, it will help researchers understand the transition in plaque stability and be beneficial for timely and effective treatment.

In this stage, single-cell RNA sequencing (scRNA-seq) revealed an increase in cellular interactions, such as high expression of the CCL4–CCR5 axis in T cells and macrophages and the effects of comorbidities (such as SLE) and circadian rhythm disorders. These findings reveal a critical transition from the peak of inflammation to structural complexity, suggesting potential therapeutic targets.

3.4 Type IV: Advanced Atheroma

This pathological type includes the dominant formation of the lipid core and initial development to the mature covering stage of the fibrous cap, accompanied by the transition from the peak inflammatory response of the immune microenvironment to chronic inflammation. The processes of proliferation and fibrosis around the lipid core accelerate, and regulatory T cells increase their ability to stabilize plaques. This stage is characterized by both the risk of plaque expansion and potential stability evolution features.

SMCs and macrophages play important roles in fibrosis[21], and further studies investigating their associated subgroups and regulatory factors are needed.

Research by Fu et al. [33] revealed that SPP1+ macrophages accumulate in the perivascular adipose tissue (PVAT) of atherosclerotic coronary arteries and exacerbate fibrosis by promoting the migration and proliferation of fibrofatty progenitor cells through OPN-CD44/integrin interactions. The degree of this fibrosis is positively correlated with the severity of coronary artery narrowing, indicating that SPP1+ macrophages in coronary PVAT may play a significant role in the progression of coronary artery atherosclerosis. Single-cell RNA-seq has, for the first time, accurately identified the SPP1+ macrophage subpopulation in atherosclerotic plaques, revealing the specific enrichment of these cells in perivascular adipose tissue (PVAT). Traditional methods have difficulty distinguishing the heterogeneity of macrophage subtypes, but this technology clearly identifies their profibrotic phenotype through gene expression characteristics (such as high expression of SPP1 and MMP9), filling the gap in the understanding of cell functional differentiation in the inflammatory microenvironment.

As a hallmark of CAD, which is explicitly described as primarily resulting from coronary atherosclerosis, ischemic cardiomyopathy (ICM) studies of SMC subgroups can reveal important SMC remodeling at the fibrosis stage of artery atherosclerosis. The critical C6 S100A4+ SMC subpopulation [34], identified through scRNA-seq analysis, interacts with endothelial cells via the PTN-NCL pathway, influencing disease progression. Key transcription factors such as KLF2, FOS, FOSB, and JUNB were identified in this subpopulation, offering potential insights for preventing fibrosis progression. Another study confirmed that previously validated CAD risk sites (such as LMOD1) have specific regulatory specificity in SMCs. Research shows that the regulation of LMOD1 in smooth muscle cells may participate in the mechanism of CHD risk by affecting cell function, particularly aspects such as vascular remodeling and contraction ability [21], which may be related to changes in vascular function in late-stage lesions.

At the gene level, “SMARCA4, translated into protein BRG1, showed significant differences and was found to promote fibroblast proliferation and migration in in vivo experiments.” [35] This discovery holds dual significance for understanding the formation and stability of the ‘fibrous cap’ in atherosclerosis. In the early stages, the moderate proliferation and migration of fibroblasts contribute to the formation of the fibrous cap, which wraps around the lipid core and plays a role in stabilizing the plaque. However, excessive and sustained activation can lead to pathological fibrosis, vascular wall sclerosis, and lumen stenosis. Therefore, the profibrotic effect of SMARCA4/BRG1 is a ‘double-edged sword’, and its specific role depends on the stage of the disease and the degree of activation. These findings suggest that treatment strategies targeting this

pathway need to have precise temporal specificity. Compared with traditional bulk sequencing, which can only extract the average signal of cell populations, this highlights the extraordinary value of scRNA-seq technology in analyzing cellular heterogeneity and identifying key driver genes. Future research should focus on spatial transcriptomics to locate the specific spatial positions of these highly expressed SMARCA4 fibroblasts within the plaque (whether they are in the core of the fibrous cap or near the area of immune cell infiltration), as well as ATAC-seq to directly confirm that BRG1 indeed alters the chromatin accessibility of these cells, which will lay a solid foundation for the development of novel therapeutic strategies for atherosclerosis on the basis of epigenetic regulation.

During this stage, scRNA-seq identified fibrosis processes driven by subtypes such as SPP1+ macrophages and S100A4+ SMCs, revealing the role of genes such as SMARCA4 in cell proliferation. These insights clarify the evolution from chronic inflammation to a stable but progressive pathological period, providing a foundation for understanding late instability and emphasizing the necessity of multiomics integration.

3.5 Type V: Complicated Lesion

In this stage, an acute inflammation outbreak in the immune microenvironment occurs, with platelets and coagulation factors participating. This late type is characterized by plaque rupture, hemorrhage, and thrombosis, leading to acute events.

According to previous studies [24], M1 macrophages (proinflammatory macrophage subtypes) and specific T-cell subtypes (e.g., M1 macrophages). Th1 and Th17 cells) are more active during this period. These cells secrete large amounts of proinflammatory cytokines, exacerbating the inflammation and instability of plaques. In contrast, the functional state of immune suppressor cells such as regulatory T cells (Tregs) may be affected, and their ability to suppress inflammation may weaken, further worsening the proinflammatory process. Unstable plaques are at risk of rupture under the influence of reactions to self-antigens, the cytotoxic pathway, and the exhaustion pathway of CD8+ T cells [36].

HF (heart failure) is an important pathological change that exacerbates myocardial atherosclerosis. Using scRNA-seq, Merten et al. conducted a series of studies [37-39] focused on the activation of T cells and monocytes and the expansion of circulating T cells in the HF. Single-cell RNA sequencing of peripheral immune cells from healthy and HF donors also revealed a significant increase in CD4+ and CD8+ T cells. This study, together with previous studies [36], demonstrated the significant impact of CD8+ T cells in the late stage of atherosclerosis. Upon exposure to the secretome of HFs, endothelial cells increase their ability to adhere to cocultured monocytes and downregulate the level of the checkpoint inhibitor PD-L1 [37, 39], indicating a reduction in the degree of autoimmune reactions by T cells. These findings suggest that endothelial cells are not only targets of inflammatory responses but also active participants in immunity. By downregulating PD-L1, endothelial cells accelerate the recruitment of monocytes and the activation of T cells. In the future, it can be hypothesized that targeting PD-L1 expression in endothelial cells (such as through anti-inflammatory cytokines or epigenetic regulation) can restore immune homeostasis.

scRNA-seq in the complex lesion stage highlights acute inflammation outbreaks, such as the autoimmune response and the clonal expansion of T cells in HFs. These late mechanisms reveal the risk of plaque rupture due to immune imbalance, emphasizing the need for targeted intervention in advanced CAD stages.

4. Discussion and Future Prospects

Using single-cell RNA sequencing technology, researchers have revealed multiple key findings in the immune microenvironment of CAD patients according to the pathological classification framework. In the early stage (types I--II), they explored the lipid resistance of cells such as CD52hi macrophages and the heterogeneity of monocytes [23, 25]. In the middle stage (Type III), T cells are activated by DAMPs, the polarization of macrophages, and the differentiation subtypes of NK cells [26-28]. In the late stage (types IV-V), the field focused on macrophage-driven fibrosis and the expansion of T-cell clones [33, 36-39] and emphasized the dynamic changes across stages (such as from initial inflammation to unstable transformation) and overall patterns (such as immune imbalance amplifying comorbidity effects[22, 24, 31, 32]).

While scRNA-seq has revolutionized CAD research, several limitations exist. (1) Lack of spatial information: Although scRNA-seq captures heterogeneity, it ignores tissue localization, leading to an

incomplete understanding of local mechanisms such as PVAT fibrosis [33] or endothelial interactions [34]. (2) Insufficient causal validation: Since scRNA-seq is a technology that relies on complex data analysis to reveal the characteristics of single-cell transcriptomes, many of its findings (such as JUN upregulation[29] and the CCL4–CCR5 axis [30]) remain at the level of association and cannot distinguish causality, especially with the limitation of a small sample size in generalization. (3) Lack of integration of comorbidities: Although it touches on diabetes and SLE[22, 24, 31], it lacks multiomics validation, resulting in insufficient comprehensiveness of immune changes.

Considering these limitations, several directions for future research should be explored further. (1) Integrating spatial transcriptomics with scRNA-seq can supplement positional information, further analyzing cellular interactions such as those in the cores of plaques. (2) Single-cell multiomics (e.g., scATAC-seq + scRNA-seq) can elucidate gene regulation, such as by verifying the role of TFs (such as TCF21 and JUNB[21, 34]) in heterogeneous subtypes. (3) AI-assisted analysis can be used for early intervention (e.g., through machine learning [21] for personalized prediction of disease trajectories). (4) In clinical translation, targeted immunotherapies for subtypes (such as SPP1+ macrophages [33]and S100A4+ SMCs [34]) or circulating biomarkers (such as JUN[29]) can be developed.

References

[1] Abbas, A., Raza, A., Ullah, M., Hendi, A. A., Akbar, F., Khan, S. U., Zaman, U., Saeed, S., Ur Rehman, K., Sultan, S., et al. A Comprehensive Review: Epidemiological Strategies, Catheterization and Biomarkers used as a Bioweapon in Diagnosis and Management of Cardio Vascular Diseases. Curr Probl Cardiol. 2023, 48(7), p. 101661. https://doi.org/10.1016/j.cpcardiol.2023.101661.

[2] Zaman, M. A. Coronary Artery Disease (CAD) and its interventional treatments. International Journal on Heart and Vascular system. 2023, 3(1), pp. 1-3.

[3] Melidoniotis, E., Kalogeropoulos, K., Tsatsaris, A., Zografakis-Sfakianakis, M., Lazopoulos, G., Tzanakis, N., Anastasiou, I. and Skalidis, E. Geospatial epidemiology of coronary artery disease treated with percutaneous coronary intervention in Crete, Greece. Geospatial Health. 2024, 19(1), p. 1251.

[4] Han, H., Du, R., Cheng, P., Zhang, J., Chen, Y. and Li, G. Comprehensive Analysis of the Immune Infiltrates and Aberrant Pathways Activation in Atherosclerotic Plaque. Frontiers in Cardiovascular Medicine. 2021, 7, p. 602345. https://doi.org/10.3389/fcvm.2020.602345.

[5] Liu, C., Liu, J., Zhang, Y., Wang, X. and Guan, Y. Immune-related potential biomarkers and therapeutic targets in coronary artery disease. Frontiers in Cardiovascular Medicine. 2023, 9, p. 1055422. https://doi.org/10.3389/fcvm.2022.1055422.

[6] He, T., Muhetaer, M., Wu, J., Wan, J., Hu, Y., Zhang, T., Wang, Y., Wang, Q., Cai, H. and Lu, Z. Immune Cell Infiltration Analysis Based on Bioinformatics Reveals Novel Biomarkers of Coronary Artery Disease. Journal of Inflammation Research. 2023, 16(null), pp. 3169-3184. https://doi.org/10.2147/JIR.S416329.

[7] Kott, K. A., Chan, A. S., Vernon, S. T., Hansen, T., Kim, T., de Dreu, M., Gunasegaran, B., Murphy, A. J., Patrick, E., Psaltis, P. J., et al. Mass cytometry analysis reveals altered immune profiles in patients with coronary artery disease. Clinical & Translational Immunology. 2023, 12(11), p. e1462. https://doi.org/10.1002/cti2.1462.

[8] Jansen, M. F., Hollander, M. R., van Royen, N., Horrevoets, A. J. and Lutgens, E. CD40 in coronary artery disease: a matter of macrophages? Basic Research in Cardiology. 2016, 111(4), p. 38. https://doi.org/10.1007/s00395-016-0554-5.

[9] Li, X. and Wang, C.-Y. From bulk, single-cell to spatial RNA sequencing. International Journal of Oral Science. 2021, 13(1), p. 36. https://doi.org/10.1038/s41368-021-00146-0.

[10] Stary, H. C. Natural history and histological classification of atherosclerotic lesions: an update. Arteriosclerosis, Thrombosis, and Vascular Biology. 2000, 20(5), pp. 1177-1178. https://doi.org/10.1161/01.ATV.20.5.1177.

[11] Stary, H. C., Blankenhorn, D. H., Chandler, A. B., Glagov, S., Insull Jr, W., Richardson, M., Rosenfeld, M., Schaffer, S., Schwartz, C. and Wagner, W. A definition of the intima of human arteries and of its atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Arteriosclerosis and Thrombosis: A Journal of Vascular Biology. 1992, 12(1), pp. 120-134. https://doi.org/10.1161/01.atv.12.1.120.

[12] Stary, H. C., Chandler, A. B., Dinsmore, R. E., Fuster, V., Glagov, S., Insull Jr, W., Rosenfeld, M. E., Schwartz, C. J., Wagner, W. D. and Wissler, R. W. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis: a report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation. 1995, 92(5), pp. 1355-1374. https://doi.org/10.1161/01.cir.92.5.1355.

[13] Stary, H. C., Chandler, A. B., Glagov, S., Guyton, J. R., Insull Jr, W., Rosenfeld, M. E., Schaffer, S. A., Schwartz, C. J., Wagner, W. D. and Wissler, R. W. A definition of initial, fatty streak, and intermediate lesions of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation. 1994, 89(5), pp. 2462-2478. https://doi.org/10.1161/01.atv.14.5.840.

[14] Virmani, R. Atherosclerosis Pathology. Available from: https://emedicine.medscape.com/article/1612610-overview?form=fpf (accessed 8 January 2025).

[15] Hameed, M., Rai, P., Makris, M. and Weger-Lucarelli, J. Optimized protocol for mouse footpad immune cell isolation for single-cell RNA sequencing and flow cytometry. STAR Protocols. 2023, 4(3), p. 102409. https://doi.org/10.1016/j.xpro.2023.102409.

[16] Jogi, H. R., Smaraki, N., Nayak, S. S., Rajawat, D., Kamothi, D. J. and Panigrahi, M. Single cell RNA-seq: a novel tool to unravel virus-host interplay. VirusDisease. 2024, 35(1), pp. 41-54. https://doi.org/10.1007/s13337-024-00859-w.

[17] Chen, G., Ning, B. and Shi, T. Single-cell RNA-seq technologies and related computational data analysis. Frontiers in Genetics. 2019, 10, p. 317. https://doi.org/10.3389/fgene.2019.00317.

[18] Ding, S., Chen, X. and Shen, K. Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Communications. 2020, 40(8), pp. 329-344. https://doi.org/10.1002/cac2.12078.

[19]Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. Comparative analysis of single-cell RNA sequencing methods. Molecular cell. 2017, 65(4), pp. 631-643. e4. https://doi.org/10.1016/j.molcel.2017.01.023.

[20] Luecken, M. D. and Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology. 2019, 15(6), p. e8746. https://doi.org/10.15252/msb.20188746.

[21] Turner, A. W., Hu, S. S., Mosquera, J. V., Ma, W. F., Hodonsky, C. J., Wong, D., Auguste, G., Song, Y., Sol-Church, K. and Farber, E. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nature Genetics. 2022, 54(6), pp. 804-816. https://doi.org/10.1038/s41588-022-01069-0.

[22] Kirichenko, T., Nedosugova, L., Bochkareva, L., Galstyan, K., Orekhov, A. and Sobenin, I. Monocyte activation in coronary heart disease and type 2 diabetes. Atherosclerosis. 2022, 355, p. 12. https://doi.org/10.1016/j.atherosclerosis.2022.06.263.

[23] Bashore, A. C., Xue, C., Kim, E., Yan, H., Zhu, L. Y., Pan, H., Kissner, M., Ross, L. S., Zhang, H. and Li, M. Monocyte single-cell multimodal profiling in cardiovascular disease risk states. Circulation Research. 2024, 135(6), pp. 685-700. https://doi.org/10.1161/CIRCRESAHA.124.324457.

[24] Hedrick, C. C. Single-cell sleuthing: Cracking the monocyte code for cardiovascular clues. Circulation Research. 2024, 135(6), pp. 701-703. https://doi.org/10.1161/CIRCRESAHA.124.325134.

[25] Jiang, J., Hiron, T. K., Agbaedeng, T. A., Malhotra, Y., Drydale, E., Bancroft, J., Ng, E., Reschen, M. E., Davison, L. J. and O'Callaghan, C. A. A novel macrophage subpopulation conveys increased genetic risk of coronary artery disease. Circulation Research. 2024, 135(1), pp. 6-25. https://doi.org/10.1161/CIRCRESAHA.123.324172.

[26] Li, Y., Shi, H., Zou, Y., Guan, Y., Liu, N. and Liu, B. Unraveling the immune activation mechanisms of DAMPs in coronary artery disease through transcriptomic and single-cell analyses. Cytokine. 2025, 191, p. 156952. https://doi.org/10.1016/j.cyto.2025.156952.

[27] Horstmann, H., Anto Michel, N., Westermann, D., Maerz, W., Zirlik, A. and Wolf, D. Prediction of cardiovascular outcomes by subtle changes in the peripheral immune cell landscape-insights from the LURIC single cell RNA-sequencing study (LuRNA). European Heart Journal. 2024, 45(Supplement_1), p. ehae666. 3699. https://doi.org/10.1093/eurheartj/ehae666.3699.

[28] Li, X., Li, Q., Liu, H., Zhang, Y., Xia, J., Wang, X., Lei, T. and Ma, J. To explore the key subgroup and their immune microenvironment during the formation of coronary plaque with scRNA-seq. Cardiology Research and Practice. 2025, 2025(1), p. 3221767. https://doi.org/10.1155/crp/3221767.

[29] Song, X., Fu, Y., Li, C., Jia, Q., Ren, M., Zhang, X., Bie, H., Zhou, H., Gan, X. and He, S. Single-cell RNA sequencing atlas of peripheral blood mononuclear cells from subjects with coronary artery disease. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2024, 1871(1), p. 119593. https://doi.org/10.1016/j.bbamcr.2023.119593.

[30] Feng, Z., Li, H., Chen, N., Xu, K. and Zhang, B. Deciphering the role of CCL4-CCR5 in coronary artery disease pathogenesis: insights from Mendelian randomization, bulk RNA sequencing, single-cell RNA, and clinical validation. International Journal of Medical Sciences. 2024, 21(14), p. 2683. https://doi.org/10.7150/ijms.99518.

[31] LI, C., Osmani, L., Gu, J., Zhao, H., Kang, I., Bucala, R. and Dong, X. POS1390 single-cell profiling identifies peripheral immune signature of coronary artery disease in sle patients. Annals of the Rheumatic Diseases. 2024, 83, p. 929. https://doi.org/10.1136/annrheumdis-2024-eular.412.

[32] Zhou, Z., Zhang, G., Wang, Z., Xu, Y., Qin, H., Zhang, H., Zhang, P., Li, Z., Xu, S. and Tan, X. Molecular subtypes of ischemic heart disease based on circadian rhythm. Scientific Reports. 2024, 14(1), p. 14155. https://doi.org/10.1038/s41598-024-65236-5.

[33] Fu, M., Shu, S., Peng, Z., Liu, X., Chen, X., Zeng, Z., Yang, Y., Cui, H., Zhao, R. and Wang, X. Single-cell RNA sequencing of coronary perivascular adipose tissue from end-stage heart failure patients identifies *SPP1*+ macrophage subpopulation as a target for alleviating fibrosis. Arteriosclerosis, Thrombosis, and Vascular Biology. 2023, 43(11), pp. 2143-2164. https://doi.org/10.1161/ATVBAHA.123.319828.

[34] Nie, W., Wang, Y., Xiao, Y., Lin, Z., Zhang, J., Zhao, Z. and Wang, Z. Identification of a key smooth muscle cell subset driving ischemic cardiomyopathy progression through single-cell RNA sequencing. Scientific Reports. 2025, 15(1), p. 27331.

[35] Liu, L., Li, C., Yu, L., Wang, Y., Pan, X. and Huang, J. Deciphering the role of SMARCA4 in cardiac disorders: Insights from single-cell studies on dilated cardiomyopathy and coronary heart disease. Cellular Signalling. 2024, 119, p. 111150. https://doi.org/10.1016/j.cellsig.2024.111150.

[36] Iqneibi, S., Saigusa, R., Khan, A., Oliaeimotlagh, M., Armstrong Suthahar, S. S., Kumar, S., Alimadadi, A., Durant, C. P., Ghosheh, Y. and McNamara, C. A. Single cell transcriptomics reveals recent CD8T cell receptor signaling in patients with coronary artery disease. Frontiers in Immunology. 2023, 14, p. 1239148. https://doi.org/10.3389/fimmu.2023.1239148.

[37] Merten, M., Rasper, T., Schuhmacher, B., Cremer, S., Hoffmann, J., Zeiher, A. M., Dimmeler, S. and Abplanalp, W. T. Chronic ischemic heart failure defines immune cell activation signatures from ScRNA-seq and flow cytometry with implications for cardiac resident cells. Circulation. 2022, 146(Suppl_1), pp. A13067-A13067. https://doi.org/10.1161/circ.146.suppl_1.13067.

[38]Abplanalp, W. T., Merten, M., Cremer, S., Rasper, T., Holz, K., Schuhmacher, B., Macinkovic, I., Tombor, L., John, D. and Zeiher, A. M. Abstract P3069: ScRNA-seq and flow cytometry based analysis reveals that chronic heart failure associates with enhanced activation and clonal expansion of T cells with predictive autoreactive capacity. Circulation Research. 2023, 133(Suppl_1), pp. AP3069-AP3069. https://doi.org/10.1161/res.133.suppl_1.P3069.

[39] Merten, M., Rasper, T., Cremer, S., Schuhmacher, B., Tombor, L., Zeiher, A., Hoffmann, J., Abplanalp, W. and Dimmeler, S. Human, mouse scRNA-seq and flow cytometry data in ischemic heart failure models predict immune cell activation signatures with altered endothelial cell cross talk. European Heart Journal. 2022, 43(Supplement_2), p. ehac544. 2965. https://doi.org/10.1093/eurheartj/ehac544.2965.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260111

Determinants of U.S. Health Insurance Charges: Evidence from Multivariate Regression

Pengyuan Qian*

School of Mathematics and Statistics, Ningbo University, Ningbo, 315211, China

**Corresponding author: Pengyuan Qian*

Abstract

U.S. national health expenditures accounted for 17.3% of GDP, raising persistent concerns among insurers, policymakers, and households. Using the Medical Cost Personal Dataset (n=1,338), which includes demographic and lifestyle variables, this study examines the determinants of individual health insurance charges. A multivariate regression framework was used to estimate the effects of age, body mass index (BMI), smoking status, number of children, sex, and region. The results indicate that smoking increased the predicted charges by more than $23,000, while age and BMI also had statistically significant positive effects. In contrast, sex and regional differences were not significant after controlling for other factors. The model explained approximately 74% of the variance in charges. These findings align with prior evidence on lifestyle-related risks and provide support for risk-adjusted premium design, highlighting the policy relevance of smoking cessation and obesity prevention.

Keywords

health insurance charges, multivariate regression, smoking, body mass index, risk-based pricing, risk adjustment

1. Introduction

According to the Centers for Medicare & Medicaid Services (CMS), U.S. national health expenditures reached USD 4.5 trillion in 2022, accounting for 17.3% of gross domestic product (GDP), and are projected to grow at an average annual rate of 5.4% through 2031 [1]. Rising healthcare costs have placed sustained pressure on households, insurers, and policymakers. Health insurance plays a critical role in providing financial protection, yet premium growth and medical charges continue to outpace wage growth and inflation. Identifying the factors that drive variation in individual insurance charges is therefore essential for designing risk-adjusted premium structures and for informing effective public health strategies.

Prior research has consistently highlighted the influence of lifestyle and demographic characteristics on healthcare expenditures. Smoking has long been recognized as a major driver of costs. Evidence from national surveys shows that current smokers face substantially greater annual medical spending than nonsmokers do [2]. Moreover, longitudinal analyses confirm that smoking, along with other behavioral risks, leads to persistently elevated healthcare costs over time [3]. Obesity, which is commonly measured by body mass index

(BMI), is another significant determinant. Recent evidence from the Trøndelag Health Study in Norway demonstrated that higher BMI is consistently associated with greater healthcare expenditures, with costs rising sharply among individuals at the upper end of the BMI distribution [4]. Similar patterns are reported in U.S. data, where obesity has been estimated to account for approximately $147 billion in annual medical spending [5]. Age also exerts a fundamental influence, reflecting the growing burden of chronic conditions and long-term care needs among older populations. CMS data show that individuals aged 65 and older, who constitute 17% of the population, account for 37% of total personal healthcare spending [6], underscoring the fiscal significance of aging.

In addition to individual risk factors, regional and systemic variation also shape expenditure patterns. Comparative evidence from Germany has shown that regional differences in healthcare costs can be largely explained by variations in morbidity and the distribution of medical resources [7]. In the United States, the Dartmouth Atlas has documented large geographic variations in Medicare spending that are not consistently associated with better health outcomes, suggesting that contextual and provider-level factors drive cost disparities [8]. Moreover, population-level estimates indicate that modifiable behavioral risks such as smoking and high BMI account for a large share of total health spending, underscoring the policy importance of preventive strategies and actuarial risk adjustment [9].

Despite these insights, much of the evidence to date is derived from large-scale surveys or aggregate expenditure accounts, which may obscure variation at the individual level. Few studies apply a unified regression framework to jointly examine demographic, behavioral, and regional determinants via microlevel insurance data. This paper addresses that gap by analyzing the Medical Cost Personal Dataset (n = 1,338), a widely used dataset containing individual-level information on health insurance charges and related covariates. A multivariate regression model was applied to quantify the relative contributions of smoking, BMI, and age while also evaluating the additional roles of sex, number of children, and region. The study contributes by (1) confirming the effects of established cost drivers within a transparent regression framework, (2) assessing the significance of additional demographic and contextual variables, and (3) drawing implications for risk-adjusted premium design and public health interventions. In this way, it illustrates the value of accessible microdata for replicable empirical analysis and policy discussion in health economics.

2. Literature Review

2.1 Smoking and Healthcare Expenditures

Smoking is a major driver of excess medical costs because of its association with chronic diseases. U.S. survey data show that smokers face substantially higher annual expenditures than nonsmokers do [2]. National estimates attribute approximately 8–9% of total healthcare spending to smoking, exceeding USD 170 billion annually [10]. International evidence reinforces this burden, with the WHO reporting that smoking accounted for 5.7% of global health spending in 2012 [11]. Longitudinal analyses further indicate that reductions in smoking prevalence translate into measurable cost savings [3]. These findings underscore the economic importance of tobacco control and justify the inclusion of smoking status in expenditure models.

2.2 Body Mass Index and Obesity

Obesity, measured by BMI, is another key determinant of healthcare expenditures. U.S. estimates suggest that obesity adds over USD 140 billion in annual medical costs [5]. Using instrumental variables, Cawley and Meyerhoefer [12] showed that conventional estimates may understate the causal effect, with obesity increasing costs by approximately USD 2,700 per adult per year. More recent analyses confirm a national burden exceeding USD 170 billion annually, with severe obesity driving disproportionate costs [13]. OECD countries also report that overweight- and obesity-related conditions account for a sizable share of health budgets [14]. Collectively, these studies confirm that BMI is a critical predictor of both individual and system-level costs.

2.3 Age and Demographic Determinants

Age is a fundamental predictor, with per capita healthcare spending rising sharply among older adults. In the U.S., individuals aged 65+ account for 37% of total personal health spending, despite representing 17% of the population [6]. However, classic health economics research emphasizes that much of this association

reflects proximity to death rather than age per se [15]. Other demographic factors have mixed effects: sex-related differences often disappear once morbidity is controlled, whereas the number of dependent children tends to increase family expenditures but has a limited impact on per capita charges.

2.4 Regional Variation and Systemic Factors

Geographic variation also contributes to expenditure differences. In Germany, disparities are largely explained by morbidity and resource distribution [7]. In the U.S., Medicare spending varies by more than 40% across regions, yet higher spending does not consistently yield better outcomes [8, 16]. These findings suggest that systemic factors—provider practice styles, resource allocation, and postacute care utilization—are key drivers of regional variation, with implications for both insurers and policymakers.

2.5 Gaps in the Literature

Despite extensive evidence, three gaps remain. First, most analyses rely on aggregate survey or expenditure data, limiting insight into individual-level heterogeneity. Second, relatively few studies have jointly evaluated demographic, behavioral, and regional factors within a unified econometric framework. Third, while the costs of modifiable risks such as smoking and obesity are well established, their interactions with demographic variables and implications for premium design are less frequently addressed. This study contributes by applying multivariate regression to microlevel insurance data, offering a transparent assessment of established determinants and their relative importance.

3. Methodology

3.1 Data Sources

This study employs the Medical Cost Personal Dataset (N=1,338), which is publicly available and widely used in methodological demonstrations of health expenditure modeling. The dataset contains individual-level information on annual medical insurance charges and key demographic and lifestyle variables. While not a nationally representative survey, it provides a transparent microlevel setting that allows for the testing of established hypotheses regarding healthcare cost determinants.

3.2 Variables

The dependent variable is annual health insurance charges (USD). The independent variables include:

Age: measured in years.

Sex: binary indicator (male = 1, female = 0).

Body mass index (BMI): a continuous variable, defined as weight in kilograms divided by height in meters squared.

Children: number of dependent children covered by the policy.

Smoker: binary indicator (smoker = 1, nonsmoker = 0).

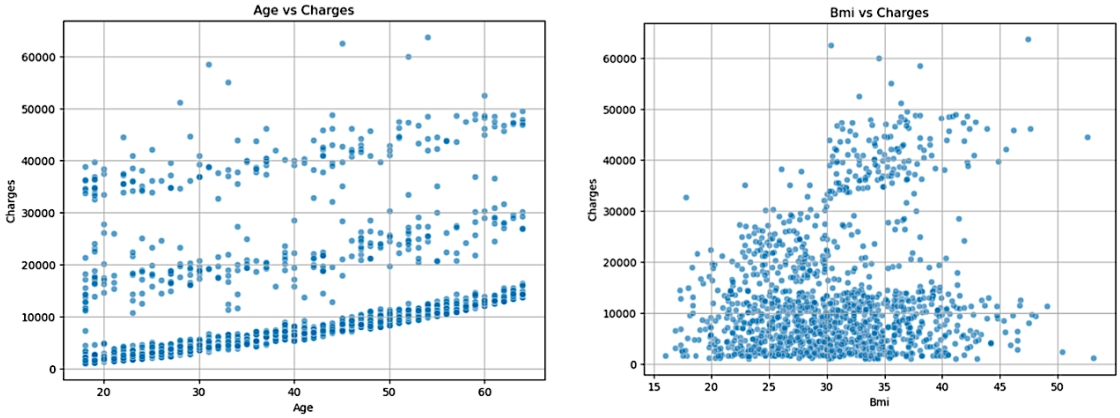
Region: categorical variable for U.S. geographic regions (Northeast, Northwest, Southeast, Southwest).

The selection of these variables is consistent with prior studies that identified demographic and lifestyle factors as key drivers of medical expenditures [2, 4, 17].

3.3 Descriptive Analysis

To provide an initial overview, scatter plots were generated to illustrate the bivariate relationships between charges and two major predictors: age and BMI (Figure 1). Charges increase steadily with age, whereas BMI also has a positive, although more dispersed, association. These patterns provide preliminary justification for including age and BMI as explanatory variables in the regression framework.

Figure 1: Scatter plots of annual health insurance charges by age and BMI



Scatter plot of age vs charges

Scatter plot of bmi vs charges

3.4 Statistical Considerations

Prior to estimation, descriptive statistics and graphical analyses were used to identify potential outliers and explore bivariate associations. Pairwise correlations were examined to assess multicollinearity among predictors, and residual plots were inspected to evaluate normality and heteroskedasticity. While the dataset does not permit advanced robustness checks such as instrumental variable estimation or longitudinal analysis, an ordinary least squares (OLS) framework provides a transparent baseline for assessing relative associations.

The regression model is specified as follows:

$$Charges_i = \beta_0 + \beta_1 Age_i + \beta_2 BMI_i + \beta_3 Children_i + \beta_4 Smoker_i + \beta_5 Sex_i + \beta_6 Region_i + \epsilon_i$$

where $Charges_i$ represents annual medical charges for individual i and where ϵ_i is the error term.

4. Results

4.1 Descriptive Statistics by Region

Table 1 reports summary statistics of annual insurance charges across four U.S. regions. The Southeast region has the highest mean charges (USD 63,770) and moderate variability, indicating both higher overall costs and dispersion. Southwest China has the lowest mean (USD 52,591) but the highest standard deviation, suggesting considerable heterogeneity in expenditures within the region. In comparison, Northeast China and Northwest China display more consistent patterns, with moderately high means and relatively lower standard deviations. These regional differences may reflect variations in population health, provider availability, or local economic conditions.

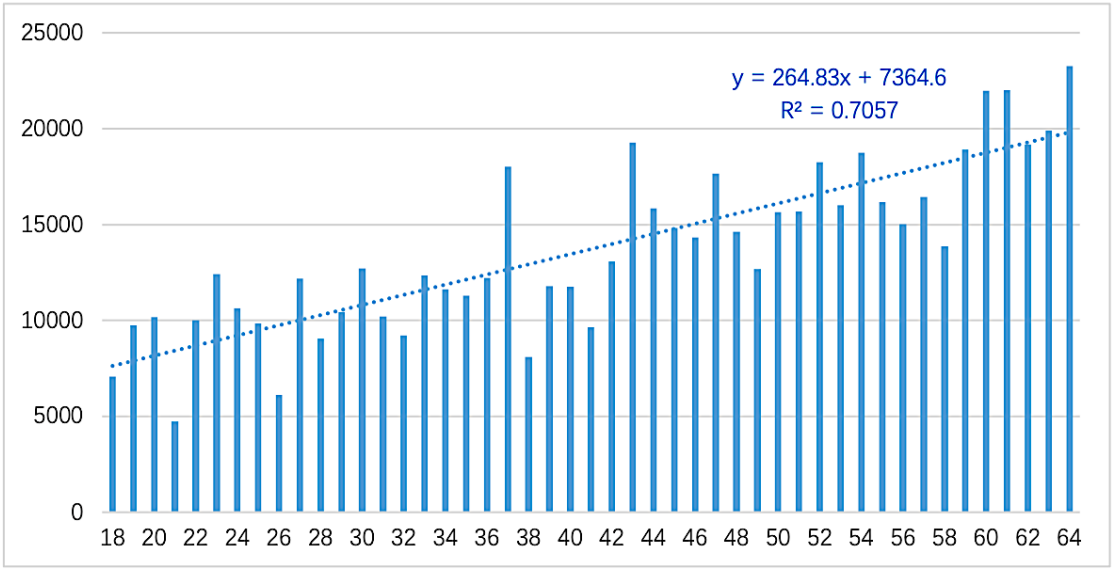
Table 1. Descriptive statistics of annual insurance charges by region

Region	Number	Sum	Mean	Standard deviation
Northeast	324	4343669	58571	11256
Northwest	325	4035712	60021	11072
Southwest	325	4012755	52591	13971
Southeast	364	5363690	63770	11557

4.2 Age and Healthcare Charges

Figure 3 depicts average annual charges by age, with a fitted linear trend line. A positive association is observed, with charges rising by approximately USD 265 per year of age ($R^2 = 0.71$). While the overall trend is increasing, fluctuations are evident, particularly between the ages of 30 and 50, suggesting heterogeneity attributable to unobserved health or behavioral factors.

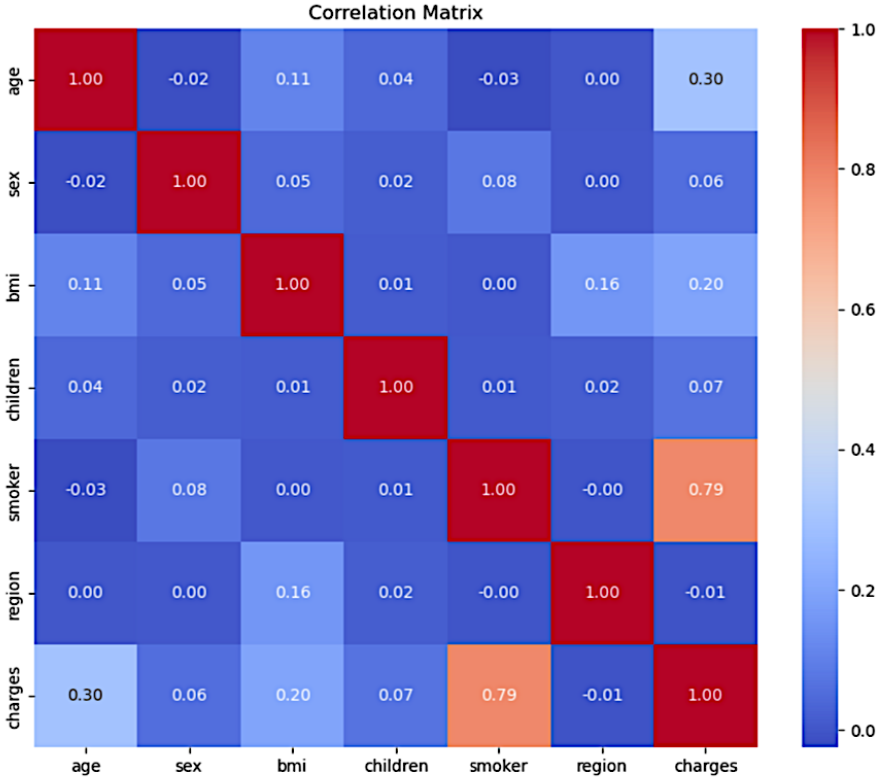
Figure 2: Average annual insurance charges by age with a fitted trend line



4.3 Correlation Analysis

Figure 4 presents the correlation matrix of the key variables. Smoking status had the strongest correlation with charges (r = 0.79), followed by age (r = 0.30) and BMI (r = 0.20). Other variables, such as sex and region, exhibited negligible associations. These patterns provide preliminary evidence that lifestyle-related factors dominate demographic and geographic determinants in shaping insurance costs.

Figure 3: Correlation matrix of demographic, lifestyle, and regional variables



4.4 Regression Analysis

Table 2 reports the F test results for the regression model. The overall F statistic of 508.7 ($p < 0.001$) confirms that the independent variables jointly explain a significant share of the variation in charges. The model achieves an R^2 of 0.742 (adjusted R^2= 0.740), indicating that approximately 74% of the variance in annual insurance charges is explained by the predictors.

Table 2: F test results for the regression model

Source	DF	FValue	P-Value
Model	6	508.7	<.0001
Error	1063		
Total	1069		

Table 3 presents the estimated coefficients. Smoking status has the largest effect, with smokers incurring USD 23,650 higher annual charges than nonsmokers do ($p < 0.001$). Age and BMI also have significant positive effects, with each additional year of age associated with USD 257 higher charges and each unit increase in BMI linked to USD 336 higher charges. The number of children is positively associated with charges (USD 425 per child, $p < 0.01$). In contrast, sex and regional indicators are statistically insignificant, suggesting a limited influence on individual-level variation once behavioral risk is controlled for.

Table 3. Parameter estimates from the multivariate regression model

Variable	Parameter Estimate	Standard Error	T Value	P Value
Intercept	-1.195e+04	1086.94	-10.991	<.0001
Age	257.06	13.45	19.109	<.0001
Sex	-18.79	375.77	-0.05	0.96
BMI	335.78	31.66	10.607	<.0001
Children	425.09	154.43	2.753	0.006
Smoker	2.365e+04	465.25	50.829	<.0001
Region	-271.28	170.37	-1.592	0.112

Figure 4 compares the predicted and actual charges. Most observations lie close to the 45-degree line, confirming good model fit, although larger deviations are visible at higher charge levels. Figure 5 shows the residual distribution, which approximates normality but reveals heavier tails driven by smokers and individuals with elevated BMI.

Figure 4: Actual versus predicted charges from the regression model

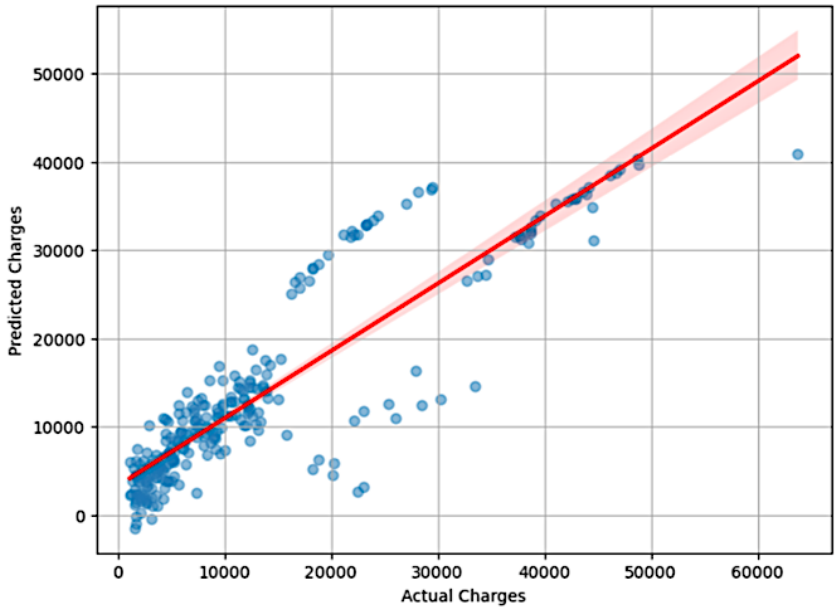
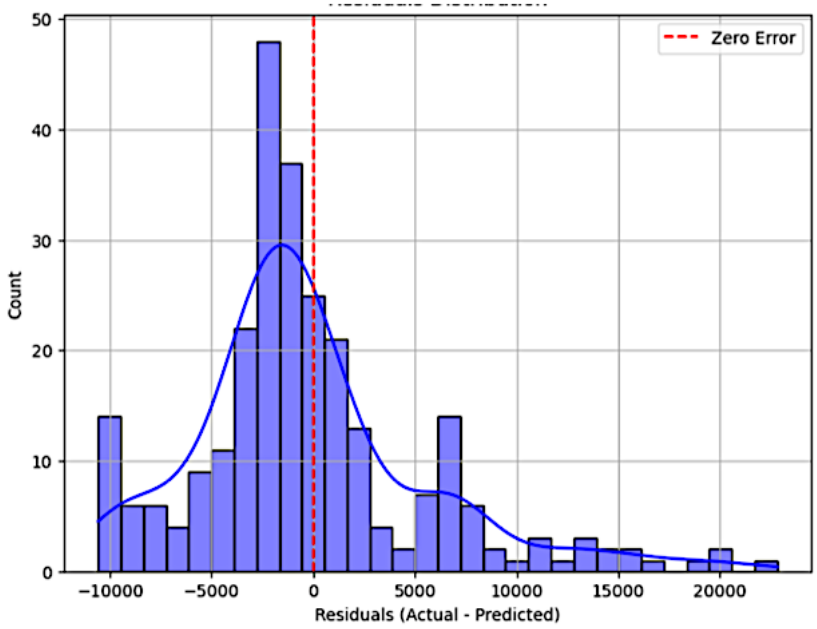


Figure 5: Residual distribution of the regression model



5. Conclusion

This study examined the determinants of U.S. health insurance charges via microlevel data and a multivariate regression framework. The results demonstrate that smoking is by far the most influential predictor of charges, followed by age and BMI, whereas sex and regional indicators are not statistically significant. The model explains approximately 74% of the variance in charges, underscoring the importance of behavioral risk factors in shaping medical expenditures. These findings are consistent with recent evidence showing that smoking continues to impose a substantial burden on U.S. healthcare spending [18, 19] and that obesity-related multimorbidity significantly increases costs across demographic groups [20]. Longitudinal analyses further confirm that lifestyle risks such as smoking and obesity independently drive higher medical expenditures [17], whereas demographic and regional differences explain relatively little variation once risk factors are controlled for [21].

The findings have direct implications for policy and practice. For insurers, incorporating behavioral risk factors such as smoking and BMI into premium design can improve pricing accuracy. For policymakers, targeted interventions aimed at smoking cessation and obesity prevention could substantially reduce overall healthcare costs and improve public health outcomes.

Several limitations should be noted. The dataset, while useful for methodological demonstration, is not nationally representative and lacks detailed clinical information. The cross-sectional design also precludes analysis of dynamic changes over time. Future research should incorporate longitudinal and nationally representative data and explore additional explanatory factors, such as genetic predispositions, socioeconomic status, and detailed behavioral measures.

Overall, the analysis highlights that modifiable lifestyle risks are central drivers of health insurance charges. By prioritizing prevention and integrating risk adjustment mechanisms, both insurers and policymakers can move toward a more equitable and sustainable approach to managing healthcare costs.

References

[1] Centers for Medicare & Medicaid Services (CMS). National health expenditure projections 2022–2031. Baltimore, MD: U.S. Department of Health and Human Services, 2023.

[2] Swedler, D. I., Miller, T. R., Ali, B., Waeher, G. and Bernstein, S. L. National medical expenditures by smoking status in American adults: an application of Manning's two-stage model to nationally representative data. BMJ open. 2019, 9(7), p. e026592. https://doi.org/10.1136/bmjopen-2018-026592.

[3] Lightwood, J. and Glantz, S. A. Smoking behavior and healthcare expenditure in the United States, 1992–2009: panel data estimates. PLoS medicine. 2016, 13(5), p. e1002020. https://doi.org/10.1371/journal.pmed.1002020.

[4] Hansen Edwards, C., Håkon Bjørngaard, J., Minet Kinge, J., Åberge Vie, G., Halsteinli, V., Ødegård, R., Kulseng, B. and Waaler Bjørnelv, G. The healthcare costs of increased body mass index-evidence from The Trøndelag Health Study. Health Economics Review. 2024, 14(1), p. 36. https://doi.org/10.1186/s13561-024-00512-8.

[5] Finkelstein, E. A., Trogdon, J. G., Cohen, J. W. and Dietz, W. Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates: Amid calls for health reform, real cost savings are more likely to be achieved through reducing obesity and related risk factors. Health affairs. 2009, 28(Suppl1), pp. w822-w831. https://doi.org/10.1377/hlthaff.28.5.w822.

[6] Centers for Medicare & Medicaid Services (CMS). National health expenditure data: Personal health care spending by age and gender, 2010–2020. Baltimore, MD: U.S. Department of Health and Human Services, 2020.

[7] Göpffarth, D., Kopetsch, T. and Schmitz, H. Determinants of regional variation in health expenditures in Germany. Health economics. 2016, 25(7), pp. 801-815. https://doi.org/10.1002/hec.3200.

[8] Skinner, J. and Fisher, E. S. Reflections on geographic variations in U.S. health care. Dartmouth Atlas White Paper. Hanover, NH: Dartmouth Institute for Health Policy & Clinical Practice, 2010.

[9] Bolnick, H. J., Hodgson, T. A. and Caplan, C. Health and economic impact of chronic diseases and modifiable risk factors in the United States. Journal of Public Health Management and Practice. 2020, 26(2), pp. 92-99. https://doi.org/10.1097/PHH.0000000000001100.

[10] Xu, X., Bishop, E. E., Kennedy, S. M., Simpson, S. A. and Pechacek, T. F. Annual healthcare spending attributable to cigarette smoking: an update. American journal of preventive medicine. 2015, 48(3), pp. 326-333. https://doi.org/10.1016/j.amepre.2014.10.012.

[11] Goodchild, M., Nargis, N. and d'Espaignet, E. T. Global economic cost of smoking-attributable diseases. Tobacco control. 2018, 27(1), pp. 58-64. https://doi.org/10.1136/tobaccocontrol-2016-053305.

[12] Cawley, J. and Meyerhoefer, C. The medical care costs of obesity: an instrumental variables approach. Journal of health economics. 2012, 31(1), pp. 219-230. https://doi.org/10.1016/j.jhealeco.2011.10.003.

[13] Ward, Z. J., Bleich, S. N., Long, M. W. and Gortmaker, S. L. Association of body mass index with health care expenditures in the United States by age and sex. PloS one. 2021, 16(3), p. e0247307. https://doi.org/10.1371/journal.pone.0247307.

[14] World Health Organization (WHO) and Organisation for Economic Co-operation and Development (OECD). The heavy burden of obesity: The economics of prevention. Paris: OECD Publishing. , 2019.

[15] Zweifel, P., Felder, S. and Meiers, M. Ageing of population and health care expenditure: a red herring? Health economics. 1999, 8(6), pp. 485-496. https://doi.org/10.1002/(SICI)1099-1050(199909)8:6<485::AID-HEC461>3.0.CO;2-4.

[16] Institute of Medicine. Variation in Health Care Spending: Target Decision Making, Not Geography. Washington, DC: National Academies Press, 2013.

[17] Kim, Y. The effects of smoking, alcohol consumption, obesity, and physical inactivity on healthcare costs: a longitudinal cohort study. BMC Public Health. 2025, 25(1), p. 873. https://doi.org/10.1186/s12889-025-xxxx.

[18] Gu, D., Sung, H.-Y., Calfee, C. S., Wang, Y., Yao, T. and Max, W. Smoking-attributable health care expenditures for US adults with chronic lower respiratory disease. JAMA Network Open. 2024, 7(5), pp. e2413869-e2413869. https://doi.org/10.1001/jamanetworkopen.2819203.

[19] Nargis, N., Hussain, A. G., Asare, S., Xue, Z., Majmundar, A., Bandi, P., Islami, F., Yabroff, K. R. and Jemal, A. Economic loss attributable to cigarette smoking in the USA: an economic modelling study. The Lancet Public Health. 2022, 7(10), pp. e834-e843. https://doi.org/10.1016/S2468-2667(22)00020-X.

[20] Ezendu, K., Pohl, G., Lee, C. J., Wang, H., Li, X. and Dunn, J. P. Prevalence of obesity-related multimorbidity and its health care costs among adults in the United States. Journal of Managed Care & Specialty Pharmacy. 2025, 31(2), pp. 179-188. https://doi.org/10.18553/jmcp.2025.31.2.179.

[21] Adjei, N. N., Haas, A., Sun, C. C., Zhao, H., Yeh, P. G., Giordano, S. H., Toumazis, I. and Meyer, L. A. Healthcare Costs in the United States by Demographic Characteristics and Comorbidity Status. Value in Health. 2025, 28(2), pp. 206-214. https://doi.org/https://doi.org/10.1016/j.jval.2024.10.3847.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260112

Pharmacoeconomic Evaluation of an Olanzapine-Containing Quadruple Antiemetic Regimen for the Prevention of CINV: An Efficacy-Cost Analysis

Qiyi Zhang*

Aix-Marseille College, Wuhan University of Technology, Wuhan, 430070, China

**Corresponding author: Qiyi Zhang*

Abstract

This study systematically evaluates the efficacy, safety, and cost-effectiveness of an olanzapine-containing quadruple antiemetic regimen compared with the standard triple regimen for preventing chemotherapy-induced nausea and vomiting (CINV) associated with moderately to highly emetogenic chemotherapy. The aim is to provide evidence-based guidance for optimizing clinical CINV management. The quadruple regimen significantly improved the complete response rate for delayed-phase CINV to 83% – 86% and increased the overall no-nausea rate to 78.7%, demonstrating superior efficacy compared with the standard triple regimen. Adding olanzapine addressed the pharmacological limitations of the standard regimen by enhancing delayed-phase coverage and controlling refractory nausea through multi-receptor antagonism. Pharmacoeconomic analysis indicates that, despite the additional cost of olanzapine, the quadruple regimen reduces the need for rescue medications and complication management costs due to its enhanced efficacy, yielding cost savings of 300 – 500 yuan (RMB) per treatment cycle and a superior cost-effectiveness ratio. Incorporating low-dose olanzapine into the standard triple regimen offers a more effective and cost-efficient approach to preventing CINV in patients receiving moderately to highly emetogenic chemotherapy. Future research should focus on developing individualized predictive models and validating real-world effectiveness in diverse patient populations.

Keywords

chemotherapy-induced nausea and vomiting, olanzapine, antiemetic regimen, cost-effectiveness analysis, pharmacoeconomics

1. Introduction

Chemotherapy is a cornerstone of cancer treatment, significantly enhancing patient survival and quality of life. However, cytotoxic agents, while effective in targeting tumor cells, frequently cause substantial toxicity to healthy tissues. Among the most prevalent adverse effects are gastrointestinal toxicities, with chemotherapy-induced nausea and vomiting (CINV) representing a particularly distressing and common challenge for patients. Without appropriate prophylactic measures, the incidence of CINV can reach 70%–80%, and in cisplatin-based regimens, the incidence may approach 100%. Despite standardized prophylactic regimens,

chemotherapy-induced nausea and vomiting (CINV) affects approximately 60%–80% of patients. Prospective studies indicate that, even with antiemetic agents, CINV incidence ranges from 11.5% to 27.3%. Inadequately controlled CINV can lead to multiple complications: frequent vomiting disrupts normal nutritional intake, severe vomiting may cause esophageal mucosal injury or gastrointestinal bleeding, and dehydration can increase renal burden and exacerbate physical decline. Over time, patients may experience malnutrition, weight loss, and electrolyte imbalances (e.g., hypokalemia, hyponatremia), ultimately compromising immune function. Psychologically, persistent CINV can trigger fear of chemotherapy, anxiety, and depression, significantly impairing quality of life.

Early antiemetics were primarily dopamine D_2receptor antagonists (RAs), such as metoclopramide, which showed modest efficacy in mild cases but caused extrapyramidal side effects at higher doses. Phenothiazines, including chlorpromazine, were subsequently used but their clinical application was limited by frequent adverse effects. In the 1980s, 5-hydroxytryptamine type 3 receptor (5-HT_3R) antagonists were developed, followed in the early 21st century by neurokinin-1 receptor (NK-1R) antagonists. NK-1R antagonists are often used in combination with 5-HT_3R antagonists, resulting in more effective antiemetic regimens that significantly improve control of nausea and vomiting, especially those induced by highly emetogenic chemotherapy [1]. Although progress has been made in managing chemotherapy-induced nausea and vomiting (CINV), notable challenges and unmet clinical needs persist. Some patients experience breakthrough nausea or vomiting with standard prophylactic regimens, and available antiemetic agents may lead to additional side effects or financial strain. To address these challenges, this study examines current antiemetic combinations, evaluating their efficacy and cost-effectiveness to inform clinical decision-making.

2. Fundamentals and Developments in the Field

2.1 Mechanisms of Chemotherapy-Induced Nausea and Vomiting

Chemotherapy-induced nausea and vomiting (CINV) is a complex phenomenon involving multiple factors and mechanisms. Its pathogenesis primarily encompasses three key aspects: direct stimulation of the gastrointestinal tract, activation of the chemoreceptor trigger zone (CTZ), and the influence of emotional factors on the cerebral cortex. Chemotherapeutic agents such as platinum compounds, anthracyclines, and nitrogen mustards can directly irritate the gastrointestinal mucosa, leading to the release of neurotransmitters such as 5-hydroxytryptamine (5-HT) from enterochromaffin cells. These neurotransmitters bind to their respective receptors, activating vagal afferent fibers that relay signals to the vomiting center in the medulla oblongata, thereby triggering the vomiting reflex. Concurrently, chemotherapy drugs and their metabolites can circulate through the bloodstream to affect the CTZ, a region rich in various neurotransmitter receptors, including dopamine and 5-HT receptors. When these receptors are activated by the drugs, neural impulses are transmitted to the vomiting center, resulting in vomiting. Furthermore, psychological elements such as anxiety and fear can modulate the vomiting center via the cerebral cortex, potentially initiating or intensifying symptoms of nausea and vomiting [2].

2.2 History of Antiemetic Drug Development

The evolution of antiemetic drug development can be broadly divided into distinct stages. Prior to the 1970s, effective treatments for chemotherapy-induced nausea and vomiting (CINV) were unavailable, with management relying primarily on agents such as antihistamines, anticholinergics, and phenothiazines. These agents, however, offered limited efficacy and were associated with significant adverse effects. The therapeutic landscape changed significantly in the 1980s with the recognition of 5-HT's pivotal role in CINV pathogenesis. This led to the emergence of the first-generation 5-HT_3 receptor antagonist (5-HT_3RA), ondansetron, which substantially improved antiemetic outcomes. Subsequent clinical introductions of agents like granisetron, tropisetron, and ramosetron further advanced CINV prevention and management [3].

The 21st century marked the advent of the Neurokinin-1 receptor antagonists (NK-1RAs), such as aprepitant. These agents exert their effect by blocking the binding of substance P to the NK-1 receptor and demonstrate particularly strong preventative efficacy against delayed-phase CINV. Current guidelines are based on emetogenic risk stratification: highly emetogenic chemotherapy (HEC) typically recommends a triple regimen (5-HT_3RA + NK-1RA + dexamethasone), while moderately to low emetogenic risk (MEC/LEC)

utilizes appropriately de-escalated antiemetic regimens (see Table 1) [4]. Nevertheless, interpatient variability and differences in drug metabolism still result in substantial heterogeneity in treatment efficacy. A subset of patients continues to experience breakthrough CINV, indicating that existing regimens require further optimization and refinement [1].

Table 1: Recommended Antiemetic Regimens for Chemotherapy with Different Emetogenic Risks

Emetogenic Level	Representative Drugs	Basic Regimen	Optimized Regimen
High Emetogenic Risk	Cisplatin, Carboplatin (AUC ≥ 4)	Triple: 5-HT$_3$ RA + NK-1 RA + Dexamethasone	Quadruple: Basic + Olanzapine 5-10 mg
Moderate Emetogenic Risk	Oxaliplatin, Carboplatin (AUC < 4)	Double: 5-HT$_3$ RA + Dexamethasone	Triple: Add NK-1 RA or Olanzapine
Low Emetogenic Risk	Paclitaxel, Gemcitabine	5-HT$_3$ RA or Dexamethasone [5]	

Note: AUC refers to the Area Under the Concentration-Time Curve, which measures the total drug absorption in the bo dy and serves as a key parameter for evaluating drug exposure.

3. Efficacy and Limitations of Current Standard Regimens

Currently, for highly emetogenic chemotherapy (HEC), the triple regimen consisting of a 5-HT$_3$ receptor antagonist (5-HT$_3$RA), an NK-1 receptor antagonist (NK-1RA), and dexamethasone has become the first-line antiemetic strategy. This approach performs well in controlling acute-phase vomiting, but its complete response rate (CRR) in the delayed phase remains only 65%-70%[1], highlighting a clear shortfall in long-term prevention of nausea and vomiting. The occurrence of breakthrough vomiting is linked to multiple factors, including Metabolic characteristics of the drugs, individual patient variations, and deviations in regimen adherence. For example, NK-1RAs have relatively short half-lives; taking aprepitant as an instance, its half-life is approximately 40 hours, which struggles to cover the peak periods of delayed vomiting, leading to diminished protective effects. Additionally, patient-specific factors such as being female, under 50 years old, or having a history of motion sickness can further reduce the CRR by 20%-30%. On the other hand, in clinical practice, dexamethasone is often dose-reduced or discontinued early due to its numerous adverse effects, which may lead to a 40% decline in the delayed-phase complete response rate [1].

In recent years, olanzapine, as a multi-receptor antagonist, has been incorporated into antiemetic regimens and demonstrated significant benefits [6]. Studies show that adding low-dose olanzapine (5-10 mg, days 1-4) to the traditional triple regimen to form a quadruple regimen can markedly elevate the delayed-phase CRR to 83%-86%, with the overall no-nausea rate increasing to 78.7%, representing a substantial improvement over the 65.6% seen with the triple regimen [1, 7]. Olanzapine is particularly effective against nausea symptoms that are difficult to control with conventional drugs, providing a promising option for patients with refractory CINV. Its mechanism of action involves blockade of multiple neurotransmitter pathways associated with emesis, exerting potent central inhibitory effects and thereby playing a pivotal role in the management of delayed-phase CINV. Notably, while the quadruple regimen offers remarkable efficacy, olanzapine itself may cause adverse reactions such as somnolence and metabolic disturbances. Therefore, in practical application, it is essential to weigh the benefits against the risks and individualize the dosage and duration of treatment.

4. Cost-Effectiveness Analysis of Different Antiemetic Regimens

Choosing cost-effective regimens is paramount for optimizing patient outcomes and ensuring healthcare system sustainability. In terms of direct drug costs, NK-1 receptor antagonists like aprepitant (125 mg/80 mg, 3-day course) are relatively expensive, with per-cycle costs ranging from approximately 480-580 yuan, and in many regions, they are not covered under medical insurance reimbursement, thereby increasing the economic burden on patients. Among 5-HT$_3$ receptor antagonists, palonosetron varies in price from 60-200 yuan per injection depending on formulation and brand. In contrast, olanzapine (5 mg × 4 days) costs less than 20 yuan, offering a significant price advantage and a favorable cost–benefit profile [1].

Further pharmacoeconomic analyses have shown that, relative to the traditional aprepitant-containing triple regimen, the olanzapine-containing triple regimen (olanzapine + 5-HT$_3$RA + dexamethasone) outperforms the traditional aprepitant-containing triple regimen not only in acute-phase vomiting relief rates (OR = 1.87, 95% CI: 1.08-3.27) but also in delayed-phase nausea control (OR = 2.78, 95% CI: 1.85-4.19), while reducing costs

per treatment cycle by 300-500 yuan [1]. Even when olanzapine is added to form a quadruple regimen, which incurs a modest cost increase, its substantial improvement in antiemetic efficacy leads to reduced rescue medication use (8.5% vs. 18.9%) and lower demands for complication management, ultimately decreasing overall healthcare expenditures [1, 8]. From a pharmacoeconomic perspective, olanzapine-containing regimens not only demonstrate clinical superiority but also alleviate economic pressures on both patients and medical insurance systems, making them particularly advantageous for implementation in resource-limited healthcare settings.

5. Unmet Clinical Needs and Future Research Directions

Although substantial progress has been made in antiemetic therapy, numerous unmet clinical needs persist. First, there is individualized prophylaxis remains insufficient. Current guidelines primarily recommend regimens based on the emetogenic classification of chemotherapy drugs, often overlooking patient-specific biomarker variations, such as plasma substance P concentrations and genetic polymorphisms (e.g., mutations in the 5-HT_3 receptor gene). These factors can influence the efficacy of antiemetic drugs and the risk of adverse reactions, resulting in some patients failing to benefit from standard regimens. In the future, integrating multidimensional data to develop predictive models will be essential for achieving more precise, personalized antiemetic treatments [5].

Second, antiemetic management strategies for special populations lack specificity. For instance, cancer patients with comorbid diabetes or psychiatric disorders may not tolerate high-dose dexamethasone or olanzapine, hightlighting the urgent need for novel non-hormonal, non-sedating antiemetic drugs or regimens to address these patients' unique requirements [8]. Additionally, the accessibility of long-acting antiemetic formulations remains low. For example, the netupitant/palonosetron fixed-dose combination capsule, with a half-life of up to 96 hours and a simplified dosing schedule, can improve patient adherence. However, its high cost limits widespread adoption. Future efforts should explore ways to enhance accessibility through policy adjustments, pricing negotiations, and insurance coverage strategies [4].

Finally, existing pharmacoeconomic evaluations are largely confined to comparisons of direct drug costs, lacking comprehensive consideration of health outcome metrics such as quality-adjusted life years (QALYs). Future studies should adopt a healthcare system perspective, employing more holistic economic evaluation methods to quantify the impact of optimized antiemetic regimens on patients' long-term quality of life and overall healthcare resource utilization [1].

Although antiemetic therapy has made considerable strides, numerous unmet clinical needs remain, and future research should further explore several key areas. First, the refinement of individualized prevention strategies. Current guidelines primarily base recommendations on the emetogenic classification of chemotherapy drugs, without fully incorporating patient-specific biomarker data, such as plasma substance P concentrations or 5-HT_3 receptor gene polymorphisms. Moving forward, efforts could focus on developing predictive models that integrate multi-omics data with clinical characteristics to identify subpopulations that respond best to particular antiemetic agents—for instance, screening patients with superior responses to ondansetron via 5-HT_3 receptor gene polymorphism testing, thereby enabling precision prescribing. Existing studies have already explored guiding antiemetic selection based on phenotypic classifications of the drug-metabolizing enzyme CYP2D6, offering an initial foundation for personalized therapy.

Additionally, in managing special populations, cancer patients with comorbidities like diabetes or psychiatric disorders often cannot tolerate high-dose dexamethasone or olanzapine, highlighting the need for intensified development of non-hormonal agents with minimal central inhibitory effects. For example, research has investigated the substitute potential of low-dose oxcarbazepine or mirtazapine in targeted populations, or examined the value of non-pharmacological approaches, such as behavioral interventions and acupuncture, as adjuncts to antiemetic care. Moreover, the accessibility and applicability of long-acting formulations require further enhancement. Take the netupitant/palonosetron fixed-dose combination capsule: despite its extended half-life and simplified dosing regimen, high costs have hindered widespread adoption. Future initiatives could involve pharmacoeconomic analyses grounded in real-world data to assess the long-term benefit-cost ratio, alongside promoting domestic generic drug development or adjustments in medical insurance policy to enhance clinical accessibility.

Finally, current pharmacoeconomic evaluations tend to emphasize direct drug costs while overlooking systematic assessments of health outcomes. In the future, incorporating comprehensive metrics like quality-adjusted life years (QALYs), from a healthcare system perspective, will be crucial to quantify how optimized antiemetic regimens affect patients' long-term quality of life, treatment adherence, and overall resource utilization—ultimately providing a more robust evidence base for clinical decisions and policy-making.

6. Conclusion

This article systematically reviews strategies for preventing and managing chemotherapy-induced nausea and vomiting (CINV), focusing on the efficacy and cost-effectiveness of a quadruple regimen that adds olanzapine to the standard triple regimen (5-HT_3 receptor antagonist, NK-1 receptor antagonist, and dexamethasone). The findings demonstrate that this quadruple regimen significantly improves outcomes for patients receiving moderately to highly emetogenic chemotherapy, particularly by increasing complete response and no-nausea rates during the delayed phase. These improvements address clinical challenges outlined in the introduction, including nutritional, physiological, and psychological complications arising from inadequate CINV control.

The regimen's significance lies in olanzapine's multi-receptor modulation, which translates into substantial clinical benefits, and its superior cost-effectiveness, as revealed through pharmacoeconomic analysis. By incorporating low-cost olanzapine, the quadruple regimen reduces overall healthcare costs by enhancing efficacy, decreasing breakthrough vomiting, and minimizing the need for rescue medications and complication management. These findings transcend simple drug cost comparisons, providing a robust framework for clinical decision-making that balances efficacy and resource utilization. The regimen offers a promising approach to improving patient quality of life while optimizing healthcare resource allocation, advancing effective and economical CINV management.

Despite these promising results, the study has certain limitations that warrant consideration. The primary conclusions rely on synthesized literature, which may be susceptible to publication bias. Variations in drug dosages, administration protocols, and patient baseline characteristics across studies may also limit result generalizability. Future research should validate the long-term efficacy and safety of the quadruple regimen in diverse real-world populations and develop tailored strategies for specific groups, such as patients with comorbid diabetes, to further enhance supportive care standards in oncology.

References

[1] Shi, Y. F., Chen, W., Wang, H. L., Li, C. Y., Qi, M. Y. and Li, G. H. The evidence-based pharmacoeconomics evaluation of two kind of antiemetics in the prevention of chemotherapy-induced nausea and vomiting. Clinical Medication Journal. 2018, 16(9), pp. 14-18+28. https://doi.org/10.3969/j.issn.1672-3384.2018.09.004.

[2] Zhang, L. X. and Wang, J. P. Wang Jianping. Efficacy and safety analysis of olanzapine in preventing chemotherapy-induced nausea and vomiting. Chinese Hospital Drug Evaluation and Analysis. 2020, 20(5).

[3] Liu, L. Meta-analysis of 5-HT_3 receptor antagonists combined with dexamethasone for the prevention of chemotherapy-induced nausea and vomiting. China Pharmacy. 2019, 30(1).

[4] Chen, M. and Li, W. Advances in the application of NK-1 receptor antagonists in chemotherapy antiemetics. Chinese Journal of New Drugs. 2021, 30(6).

[5] Jiang, W. Q., Ba, Y., Feng, J. F., Shi, Y. X., Zhang, J., Shen, B. and Xing, P. Y. Chinese experts' consensus on the prevention and treatment of nausea and vomiting related to anti-cancer drug treatment (2019version). Chinese Journal of the Frontiers of Medical Science (Electronic Version). 2019, 11(11), pp. 16-26.

[6] Zhao, L. P. Research progress on drug prevention and treatment of chemotherapy-induced nausea and vomiting. China Medical Herald. 2020, 17(20).

[7] Huang, T. and Yang, R. Observation on the efficacy of olanzapine combined with palonosetron in preventing nausea and vomiting induced by highly emetogenic chemotherapy. Practical Oncology Journal. 2019, 34(7).

[8] Sun, Y. and Sun, H. Research progress in supportive care for tumors: antiemetic therapy. Chinese Journal of Clinical Oncology. 2018, 45(10).

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260113

U-Shaped Associations Between Metabolic Indices and Depression in Middle-aged and Older Adults: A Data-Driven Cross-Sectional Study from CHARLS

Yajie Liang*

Medical Department, ShenZhen University, Shenzhen, 518000, China

**Corresponding author: Yajie Liang*

Abstract

Objective: While the relationship between metabolic dysregulation and depression is well-established, its specific nature in the elderly remains obscure, partially due to the limitations of traditional metrics like BMI in distinguishing visceral fat. This study aimed to systematically investigate the associations between novel composite metabolic indices and depressive symptoms in community-dwelling Chinese middle-aged and older adults, with a particular focus on non-linear threshold effects. Methods: This cross-sectional analysis utilized nationally representative data from the China Health and Retirement Longitudinal Study (CHARLS), comprising 10,195 participants aged ≥45 years. Depressive symptoms were assessed using the 10-item Center for Epidemiologic Studies Depression Scale (CES-D-10). Four novel indices were calculated: China Visceral Adiposity Index (CVAI), Lipid Accumulation Product (LAP), Metabolic Score for Insulin Resistance (METS-IR), and Metabolic Score for Visceral Fat (METS-VF). Besides conventional statistics, we employed multivariable logistic regression to decode the independent associations, with a particular focus on identifying non-linear threshold effects. Results: After adjusting for sociodemographic, lifestyle, and clinical confounders, CVAI, LAP, METS-IR, and METS-VF allexhibited significant U-shaped nonlinear associations with depressive symptoms. Contrary to linear expectations, moderate levels of these indices (typically corresponding to the third quartile) were associated with the lowest risk of depression, conferring a protective effect that diminished or disappeared at extremely high or low levels. Conclusion: In Chinese community-dwelling middle-aged and older adults, moderate visceral adiposity or insulin resistance may be associated with a reduced risk of depression, demonstrating a U-shaped threshold effect. These findings underscore the value of utilizing composite metabolic indices as computational biomarkers for precise mental health screening in aging populations.

Keywords

digital economy, labor market structure, skill-biased technological progress

1. Introduction

Depression, characterized by significant and persistent low mood, has emerged as one of the leading causes of disability worldwide, posing a substantial challenge to public health systems [1]. This burden is particularly

acute in China, where an accelerating aging population faces physiological decline, multi-morbidity, and social marginalization [2-5]. While these psychosocial stressors are critical, <growing evidence from psychosomatic medicine suggests> that the onset of depression is not merely psychological but inextricably linked to systemic metabolic homeostasis.

Central to this physiological dysregulation is Insulin Resistance (IR), a core pathology of metabolic syndrome often implicated in depression's pathophysiology. Mechanisms linking IR to depressive symptoms include chronic low-grade inflammation, oxidative stress, and hypothalamic-pituitary-adrenal (HPA) axis dysfunction, all of which impair neuroplasticity and cerebral energy metabolism [6-11].However, capturing such metabolic subtleties in large-scale geriatric epidemiology remains technically challenging. The widely used Body Mass Index (BMI), for instance, fails to distinguish between subcutaneous fat, visceral fat, and muscle mass. This lack of specificity is especially problematic for the Chinese elderly, who frequently exhibit "Metabolically Obese Normal Weight" or "Sarcopenic Obesity"—phenotypes where metabolic dysregulation co-occurs with a normal BMI [12]. Reliance on BMI alone, therefore, risks obscuring the true association between metabolic abnormalities and depressive symptoms

To overcome these methodological limitations, researchers have developed computational biomarkers that integrate anthropometric and biochemical parameters, offering accessible alternatives to the invasive "gold-standard" clamp technique. Foremost among indices targeting visceral adiposity are the China Visceral Adiposity Index (CVAI) and the Lipid Accumulation Product (LAP). The former (CVAI) has demonstrated superior sensitivity in capturing visceral fat heterogeneity and cardiovascular risk in Chinese populations [12-15] , while the latter (LAP) specifically reflects excessive lipid accumulation associated with obstructive sleep apnea and fatty liver disease[16, 17]. Complementing these adiposity markers are indicators of insulin sensitivity, including the Triglyceride-Glucose Index (TyG), the Metabolic Score for Insulin Resistance (METS-IR), and its derivative, the Metabolic Score for Visceral Fat (METS-VF). These surrogate markers have shown excellent efficacy in predicting type 2 diabetes and cardiovascular events[18-25] .

While these metrics have established predictive value in metabolic diseases in the fields of Bioinformatics and Intelligent Computing[26, 27],their specific epidemiological relationship with geriatric depression—particularly among community-dwelling Chinese adults—remains inadequately investigated. To elucidate this relationship, leveraging large-scale cohort data to mine novel biomarkers and construct interpretable, discriminative computational models has emerged as a pivotal approach in computational biology for enabling early disease warning and advancing precision health management. Previous literature has largely been confined to single-indicator analyses, often neglecting potential non-linear dose-response patterns (e.g., U-shaped or J-shaped associations) that may define the "metabolic-depression" interface. Addressing this gap, the current study leverages nationally representative data from the China Health and Retirement Longitudinal Study (CHARLS) to: (1) systematically compare the associations of five novel indices (CVAI, LAP, METS-IR, METS-VF, and TyG) with depressive symptoms; and (2) elucidate potential non-linear threshold effects. Such insights aim to validate metabolic health as computational biomarkers, providing empirical evidence for early screening and precise intervention in geriatric mental health.

2. Materials and Methods

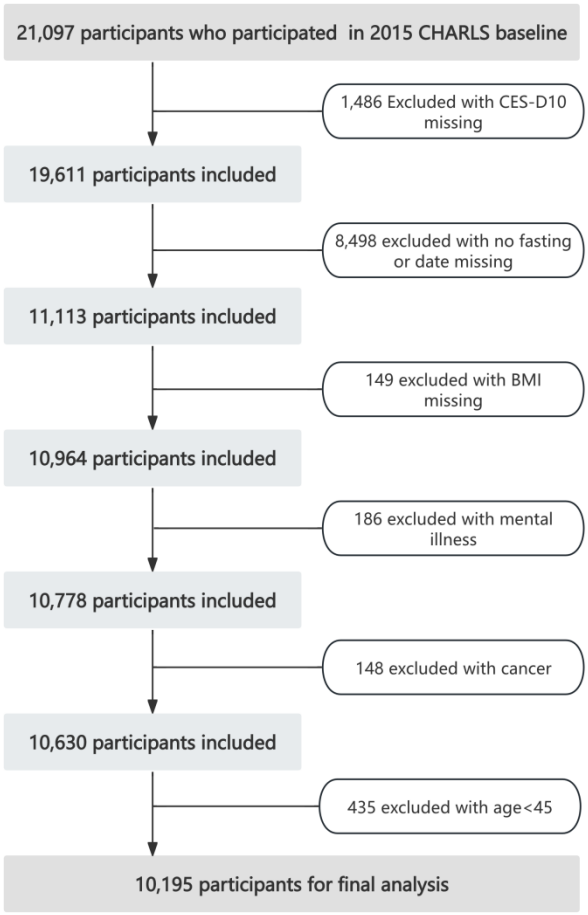
2.1 Data Source and Study Design

This study employed a cross-sectional design using data from the CHARLS database for empirical analysis. CHARLS is a nationally representative longitudinal social survey project designed to comprehensively assess the socioeconomic status and health conditions of the Chinese population aged 45 and above. Its sample covers 150 county-level units and 450 communities/villages across 28 provinces (autonomous regions/municipalities). Although CHARLS includes multiple waves of follow-up data from 2011 to 2020, considering the completeness of core variables and research timeliness, this study primarily utilized data from the 2015 survey for analysis. The project's ethical review was approved by the Biomedical Ethics Committee of Peking University (Approval No.: IRB00001052-11015), and all participants provided written informed consent before enrollment.

2.2 Study Participants

The participant selection process is detailed in Figure 1. Based on the CHARLS 2015 dataset, 21,097 respondents were initially included. To ensure analytical rigor, stringent exclusion criteria were applied to remove ineligible samples: (1) those with missing data for the key variable (CES-D-10 score); (2) those lacking basic parameters required for calculating IR-related indicators (e.g., unclear fasting status, missing BMI data); (3) individuals with a prior diagnosis of severe psychiatric disorders or malignancies and other comorbidities that might interfere with depression risk assessment; and (4) individuals aged <45 years (non-target population). After this strict screening, a total of 10,195 respondents were included in the final analytical cohort for this study.

Figure1: Data Curation and Analytical Cohort Selection Pipeline



2.3 Variable Definitions and Measurements

2.3.1 Assessment of Depressive Symptoms.

Depressive status was assessed using the 10-item Center for Epidemiologic Studies Depression Scale (CES-D-10). This scale has demonstrated good reliability and validity and effectively discriminates between normal mood and clinical depressive states in the Chinese elderly community population[28-30] . The scale comprises 10 items, requiring respondents to report the frequency of specific emotions or behaviors during the past week. Responses are scored on a 4-point scale ranging from 0 (rarely or none of the time, <1 day) to 3 (most or all of the time, 5–7 days), yielding a total score between 0 and 30 [31]. Based on established cut-off values from prior research, a CES-D-10 total score of ≥10 was defined as the presence of depressive symptoms (coded as 1), while a score <10 was defined as the absence of depressive symptoms (coded as 0).

2.3.2 Insulin Resistance Related computational biomarkers

The integration of multiple readily available clinical variables(height, weight, waist circumference (WC), body mass index (BMI), and waist-to-height ratio (WHtR)) and biochemical markers (fasting plasma glucose (FPG), triglycerides (TG), and high-density lipoprotein cholesterol (HDL-C))through specific algorithms enables the construction of composite computational models built upon routinely available clinical metrics. This data-driven approach aims to derive more effective predictive or diagnostic tools than any single parameter alone.

Five computational biomarkers were selected to multidimensionally assess insulin resistance and visceral adiposity: CVAI, LAP, METS-IR, METS-VF, and TyG. The specific formulas for each index are as follows (Note: Units for FPG, TG, and HDL-C are mmol/L):

(1) China Visceral Adiposity Index (CVAI) [15]

male:$CVAI=-267.93+0.68\times age+0.03\times BMI+4.00\times WC+22.00\times \log_{10} TG-16.32\times HDL\_C$;

female:$CVAI=-187.32+1.71\times age+4.23\times BMI+1.12\times WC+39.76\times \log_{10} TG-11.66\times HDL\_C$.

(2) Lipid Accumulation Product (LAP)[32]

male: $LAP=(WC-65)\times TG$;

female: $LAP=(WC-58)\times TG$.

(3) Metabolic Score for Insulin Resistance (METS-IR)[20]_$IR=\ln(2\times FPG+TG)\times BMI/\ln(HDL_C)$

(4) Metabolic Score for Visceral Fat (METS-VF)[22]:

male: $METS\_VF=4.466+0.011*[\ln(METS\_IR)]^3+3.239*[\ln(WHtR)]^3+0.319+0.504*[\ln(age)]$;

female: $METS\_VF=4.466+0.011*[\ln(METS\_IR)]^3+3.239*[\ln(WHtR)]^3+0.594*[\ln(age)]$

(5) Triglyceride-Glucose Index (TyG)[33]:$\ln(TG\times FPG/2)$

2.3.3 Covariates.

To control for potential confounding factors, three categories of covariates were included: sociodemographic characteristics, lifestyle factors, and clinical health status. Sociodemographic variables included age, sex, educational attainment (categorized as below primary school, primary/junior high school, and high school or above), residence (urban/rural), annual household income, and per capita consumption expenditure. Health behavior and clinical characteristics included current smoking status, current alcohol consumption status (both as binary variables: yes/no), as well as BMI and biochemical markers (TG, FPG, HDL-C) obtained through standardized physical examinations.

2.4 Data Analysis Pipeline

All data preprocessing, including cleaning, variable calculation, and management, was conducted using a computational pipeline implemented in R (version 4.3.1). The analysis comprised two main steps: (1) Conventional statistical analysis including descriptive statistics and chi-square tests performed in SPSS 26.0; and (2) Core association analysis utilizing multivariable logistic regression models in R to elucidate the independent associations, with a specialized focus on non-linear pattern discovery by treating metabolic indices as categorical variables (quartiles). Data visualization, including the patient selection flowchart (Figure 1) and comparative box plots (Figure 2), was generated to enhance the interpretability of the complex dataset."

For descriptive statistics, the normality of continuous variables was assessed first. Normally distributed data are presented as mean ± standard deviation, while non-normally distributed data are presented as median (interquartile range, IQR). Categorical variables are presented as frequency (percentage). For group difference analysis, Chi-square tests were used for categorical variables. To determine the independent association between each IR-related index and depressive symptoms, multivariable logistic regression models were constructed, with results expressed as odds ratios (OR) and their 95% confidence intervals (95% CI). To rigorously control for confounding bias, a stepwise adjustment strategy was employed for the models:
Model 1: The basic adjustment model, including only sex, age, educational level, and BMI.

Model 2: Based on Model 1, further adjusted for household income, per capita consumption, smoking, and alcohol consumption status to account for the influence of socioeconomic status and lifestyle factors.
All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant.

3. Results

3.1 Baseline Characteristics

A total of 10,195 participants were included in this study, with a mean age of 60.5 ± 9.5 years, of whom 53.0% (n=5,407) were women. Based on the CES-D-10 criteria, the overall sample was divided into a depressive symptoms group (n=3,359, 32.95%) and a non-depressive symptoms group (n=6,836, 67.05%). The comparative results of baseline characteristics between the two groups are detailed in Table 1.

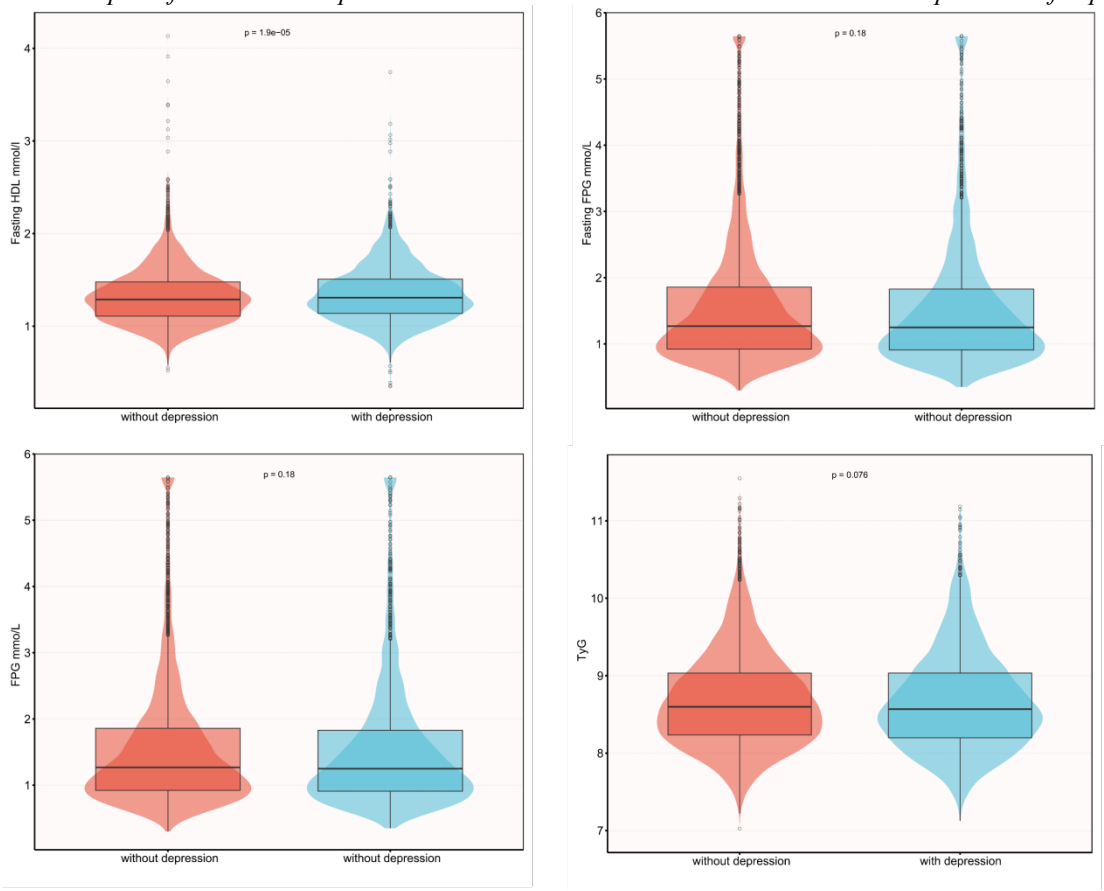
Table1: Baseline characteristics of participants.

Variables	Total (n = 10195)	0 (n = 6836)	1 (n = 3359)	p
WHtR(Mean ± SD)	0.5 ± 0.1	0.5 ± 0.1	0.5 ± 0.1	0.598
BMI (Mean ± SD)	24.5 ± 13.4	24.5 ± 12.3	24.5 ± 15.4	0.765
FPG (Mean ± SD)	5.6 ± 1.7	5.6 ± 1.6	5.6 ± 1.8	0.883
Fasting TG (Mean ± SD)	1.6 ± 1.0	1.6 ± 1.0	1.6 ± 1.0	0.58
Fasting HDL (Mean ± SD)	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	< 0.001
CVAI., Median (IQR)	101.9 (73.5, 129.9)	103.4 (74.5, 131.3)	99.5 (71.0, 127.4)	< 0.001
LAP., Median (IQR)	31.0 (16.7, 55.2)	31.4 (17.1, 55.2)	30.3 (16.0, 55.0)	0.035
METS-IR.(Mean ± SD)	36.5 ± 19.3	36.6 ± 17.5	36.3 ± 22.6	0.493
METS-VF., Median (IQR)	6.6 (6.2, 7.0)	6.6 (6.2, 6.9)	6.6 (6.2, 7.0)	0.743
TyG, Mean ± SD	8.7 ± 0.6	8.7 ± 0.6	8.7 ± 0.6	0.237
Age, Mean ± SD	60.5 ± 9.5	60.3 ± 9.6	60.9 ± 9.2	0.006
Edu, n (%)				< 0.001
High school or above	1011 (9.9)	837 (12.2)	174 (5.2)	
Elementary school	2859 (28.1)	1994 (29.2)	865 (25.8)	
below Elementary school	4312 (42.3)	2508 (36.7)	1804 (53.8)	
Middle school	2007 (19.7)	1494 (21.9)	513 (15.3)	
Gender, n (%)				< 0.001
male	4788 (47.0)	3550 (51.9)	1238 (36.9)	
female	5407 (53.0)	3286 (48.1)	2121 (63.1)	
Habitation, n (%)				< 0.001
Town	3807 (37.3)	2818 (41.2)	989 (29.4)	
Rural	6388 (62.7)	4018 (58.8)	2370 (70.6)	
Total household income, Median (IQR)	4065.0 (1020.0, 27075.0)	5400.0 (1200.0, 32390.0)	2942.5 (885.0, 13975.0)	< 0.001
Per capita household consumption, Median (IQR)	8791.0 (4920.0, 15382.7)	9092.0 (5175.0, 15876.2)	8340.0 (4569.3, 14281.0)	< 0.001
currently drink, n (%)				< 0.001
no	6536 (64.1)	4168 (61)	2368 (70.5)	
yes	3655 (35.9)	2665 (39)	990 (29.5)	
currently smoke, n (%)				< 0.001
no	7362 (72.2)	4819 (70.5)	2543 (75.7)	
yes	2828 (27.8)	2013 (29.5)	815 (24.3)	

Note: WHtR denotes waist-to-height ratio, BMI denotes body mass index, FPG denotes fasting plasma glucose, CVAI de notes China Visceral Adiposity Index, LAP denotes Lipid Accumulation Product, METS-IR denotes Metabolic Score for Insulin Resistance, METS-VF denotes Metabolic Score for Visceral Fat, and TyG denotes Triglyceride-Glucose Index.

Significant differences were observed in sociodemographic and behavioral characteristics between the two groups. The depressive symptoms group had a significantly higher proportion of women, individuals with lower educational attainment, rural residents, and those with lower household income ($P < 0.001$). Notably, smoking and alcohol consumption were more prevalent in the non-depressive symptoms group ($P < 0.001$). In the comparison of metabolic and insulin resistance (IR)-related indices, the depressive symptoms group exhibited a relatively more favorable metabolic profile. Specifically, the CVAI level in this group [99.5 (71.0, 127.4)] was significantly lower than that in the non-depressive group [103.4 (74.5, 131.3), $P < 0.001$]. A similar trend was observed for METS-IR levels [34.5 (29.9, 39.9) vs. 35.4 (30.6, 40.4), $P < 0.001$]. Regarding routine biochemical indicators, fasting plasma glucose (FPG) levels were significantly higher in the non-depressive group ($P = 0.0068$). Although triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and the TyG index did not show statistically significant differences between the groups, data distributions indicated a tendency for the non-depressive group to have higher TG and TyG levels.

Figure2: Box plot of the relationship between conventional biochemical indicators and the presence of depression.



3.2 Univariate Analysis

Table 2 presents the results of the univariate logistic regression analysis examining the impact of each variable on the risk of depressive symptoms. Sociodemographic and lifestyle factors demonstrated significant predictive value. Higher levels of education and household income were associated with a lower risk of depression. The risk of developing depression was 85% higher in women compared to men (OR = 1.85). Interestingly, smoking and alcohol consumption showed a statistically significant protective effect against depressive symptoms in the univariate analysis.

For the IR-related indices (CVAI, LAP, METS-IR, METS-VF), which exhibited a skewed distribution, analysis was performed after grouping the data into quartiles (Q1-Q4).

CVAI: Showed a significant negative trend. As CVAI levels increased, the risk of depression progressively decreased. Compared to the lowest quartile (Q1), individuals in the highest quartile (Q4) had a 20% lower risk of depression.

METS-IR: Similarly demonstrated a protective effect, with the Q4 group showing a 20% reduced risk compared to the Q1 group.

Other Indices: LAP and METS-VF did not show a significant linear trend in the univariate analysis. However, METS-VF indicated a potential nonlinear U-shaped association.

Table 2: Results of Univariate Analysis.

Variable	OR_95CI	P_value
WHtR	0.89 (0.59~1.36)	0.598
BMI	1 (1~1)	0.766
FPG	1 (0.97~1.02)	0.882
Fasting TG	0.99 (0.95~1.03)	0.58
Fasting HDL	1.26 (1.11~1.44)	<0.001
CVAI.	1 (1~1)	0.016
LAP.	1 (1~1)	0.366
METS. IR.	1 (1~1)	0.497
METS. VF.	1 (0.99~1.01)	0.754
TyG	0.96 (0.9~1.03)	0.237
Age	1.01 (1~1.01)	0.006
High school or above	1(Ref)	
below Elementary school	3.46 (2.91~4.12)	<0.001
Elementary school	2.09 (1.74~2.5)	<0.001
Middle school	1.65 (1.36~2)	<0.001
Female	1.85 (1.7~2.01)	<0.001
Rural	1.68 (1.54~1.84)	<0.001
Total household income	1 (1~1)	<0.001
Per capita household consumption	1 (1~1)	0.001
currently drink	0.65 (0.6~0.71)	<0.001
currently smoke	0.77 (0.7~0.84)	<0.001
CVAI		
Q1	1(Ref)	
Q2	0.95 (0.84~1.06)	0.36
Q3	0.86 (0.76~0.96)	0.009
Q4	0.8 (0.71~0.9)	<0.001
LAP		
Q1	1(Ref)	
Q2	0.91 (0.81~1.02)	0.095
Q3	0.85 (0.76~0.96)	0.007
Q4	0.91 (0.81~1.02)	0.101
METS-IR		
Q1		
Q2	0.93 (0.83~1.04)	0.195
Q3	0.79 (0.7~0.89)	<0.001
Q4	0.8 (0.71~0.9)	<0.001
METS-VF		

Q1	1(Ref)	
Q2	0.89 (0.79~1)	0.042
Q3	0.87 (0.77~0.97)	0.017
Q4	1 (0.89~1.13)	0.961

Note: CVAI, LAP, METS-IR, and METS-VF were converted into quartiles for analysis (first quartile [Q1]: lowest 25%; fourth quartile [Q4]: highest 25%).

3.3 Multivariable Logistic Regression Analysis

To control for confounding factors and elucidate the independent effects of each IR-related index, two adjusted models were constructed: Model 1 adjusted for age, sex, education, and BMI; Model 2 further adjusted for income, consumption, smoking, and alcohol consumption.

3.3.1 CVAI and Depressive Symptoms.

As shown in Table 3, in Model 1, CVAI maintained a significant linear protective effect, with the Q4 group showing a 25% reduced risk (P < 0.001). However, in the fully adjusted Model 2, the relationship between CVAI and depression risk transformed into a nonlinear U-shaped curve. The protective effect peaked at the third quartile (Q3) (OR = 0.68, P < 0.001), while this protective effect attenuated in the Q4 group (OR approached 1, though it remained significant with P = 0.039).

Table 3: Results of Multivariable Analysis for CVAI.

CVAI Model 1						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M1. OR (95%CI)	adj. P value
Q1	2534	895 (35.3)	1(Ref)		1(Ref)	
Q2	2534	864 (34.1)	0.95 (0.84~1.06)	0.36	0.85 (0.76~0.96)	0.009
Q3	2534	807 (31.8)	0.86 (0.76~0.96)	0.009	0.76 (0.67~0.86)	<0.001
Q4	2535	768 (30.3)	0.8 (0.71~0.9)	<0.001	0.75 (0.66~0.85)	<0.001
Trend.test	10137	3334 (32.9)	0.92 (0.89~0.96)	<0.001	0.91 (0.87~0.94)	<0.001
CVAI Model 2						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M2. OR (95%CI)	adj. P value
Q1	2534	895 (35.3)	1(Ref)		1(Ref)	
Q2	2534	864 (34.1)	0.95 (0.84~1.06)	0.36	0.77 (0.62~0.95)	0.013
Q3	2534	807 (31.8)	0.86 (0.76~0.96)	0.009	0.68 (0.55~0.84)	<0.001
Q4	2535	768 (30.3)	0.8 (0.71~0.9)	<0.001	0.8 (0.64~0.99)	0.039
Trend.test	10137	3334 (32.9)	0.92 (0.89~0.96)	<0.001	0.92 (0.86~0.99)	0.025

Note: CVAI was converted into quartiles for analysis (first quartile [Q1]: lowest 25%; fourth quartile [Q4]: highest 25%). Model 1 adjusted for age, sex, education, and BMI. Model 2 further adjusted for income, consumption, smoking, and alcohol consumption on the basis of Model 1.

3.3.2 METS-IR and Depressive Symptoms.

As shown in Table 4, METS-IR exhibited a typical U-shaped protective relationship in both models. As the index increased, the depression risk initially decreased and then subsequently increased. In Model 2, the Q3 group demonstrated the strongest protective effect (OR = 0.76, P = 0.012), whereas the associations for the Q2 and Q4 groups did not reach statistical significance. This suggests that a moderate level of METS-IR may be associated with the lowest risk of depression.

Table 4: Results of Multivariable Analysis for METS-IR.

METS-IR Model 1						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M1. OR (95%CI)	adj. P value
Q1	2537	910 (35.9)	1(Ref)		1(Ref)	
Q2	2537	866 (34.1)	0.93 (0.83~1.04)	0.195	0.89 (0.79~1)	0.056
Q3	2537	776 (30.6)	0.79 (0.7~0.89)	<0.001	0.78 (0.69~0.88)	<0.001
Q4	2538	787 (31)	0.8 (0.71~0.9)	<0.001	0.8 (0.7~0.9)	<0.001
Trend.test	10149	3339 (32.9)	0.92 (0.89~0.96)	<0.001	0.92 (0.88~0.96)	<0.001
METS-IR Model 2						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M2. OR (95%CI)	adj. P value
Q1	2537	910 (35.9)	1(Ref)		1(Ref)	

Q2	2537	866 (34.1)	0.93 (0.83~1.04)	0.195	0.91 (0.74~1.11)	0.352
Q3	2537	776 (30.6)	0.79 (0.7~0.89)	<0.001	0.76 (0.62~0.94)	0.012
Q4	2538	787 (31)	0.8 (0.71~0.9)	<0.001	0.99 (0.8~1.23)	0.933
Trend.test	10149	3339 (32.9)	0.92 (0.89~0.96)	<0.001	0.98 (0.91~1.05)	0.555

Note: METS-IR was converted into quartiles for analysis (first quartile [Q1]: lowest 25%; fourth quartile [Q4]: highest 25%). Model 1 adjusted for age, sex, education, and BMI. Model 2 further adjusted for income, consumption, smoking, and alcohol consumption on the basis of Model 1.

3.3.3 LAP and Depressive Symptoms.

As shown in Table 5, although no significance was observed in the univariate analysis, LAP demonstrated a robust U-shaped protective effect after adjusting for covariates. In both Model 1 and Model 2, the Q3 group exhibited the lowest risk of depression (Model 2: OR = 0.72, P = 0.003), suggesting that a moderate level of lipid accumulation may be associated with lower depressive symptoms.

Table 5: Results of Multivariable Analysis for LAP.

LAP Model 1						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M1. OR (95%CI)	adj. P value
Q1	2545	889 (34.9)	1(Ref)		1(Ref)	
Q2	2546	833 (32.7)	0.91 (0.81~1.02)	0.095	0.82 (0.73~0.93)	0.002
Q3	2546	798 (31.3)	0.85 (0.76~0.96)	0.007	0.75 (0.66~0.84)	<0.001
Q4	2546	834 (32.8)	0.91 (0.81~1.02)	0.101	0.78 (0.69~0.89)	<0.001
Trend.test	10183	3354 (32.9)	0.96 (0.93~1)	0.058	0.92 (0.89~0.96)	<0.001
LAP Model 2						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M2. OR (95%CI)	adj. P value
Q1	2545	889 (34.9)	1(Ref)		1(Ref)	
Q2	2546	833 (32.7)	0.91 (0.81~1.02)	0.095	0.79 (0.65~0.97)	0.027
Q3	2546	798 (31.3)	0.85 (0.76~0.96)	0.007	0.72 (0.58~0.89)	0.003
Q4	2546	834 (32.8)	0.91 (0.81~1.02)	0.101	0.78 (0.63~0.97)	0.024
Trend.test	10183	3354 (32.9)	0.96 (0.93~1)	0.058	0.92 (0.86~0.99)	0.021

Note: LAP was converted into quartiles for analysis (first quartile [Q1]: lowest 25%; fourth quartile [Q4]: highest 25%). Model 1 adjusted for age, sex, education, and BMI. Model 2 further adjusted for income, consumption, smoking, and al cohol consumption on the basis of Model 1.

3.3.4 METS-VF and Depressive Symptoms.

As shown in Table 6, the pattern observed for METS-VF was similar to that of CVAI. Model 1 indicated a linear protective association (with a 24% risk reduction in Q4). Model 2, however, revealed a transition to a U-shaped relationship, where the protective effect was most pronounced in the Q3 group (OR = 0.76, P = 0.011). Notably, a very high level (Q4) of METS-VF did not confer additional benefit in terms of risk reduction.

Table 6: Results of Multivariable Analysis for METS-VF.

METS-VF Model 1						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M1. OR (95%CI)	adj. P value
Q1	2534	870 (34.3)	1(Ref)		1(Ref)	
Q2	2534	802 (31.6)	0.89 (0.79~1)	0.042	0.84 (0.74~0.95)	0.005
Q3	2534	790 (31.2)	0.87 (0.77~0.97)	0.017	0.78 (0.69~0.88)	<0.001
Q4	2535	872 (34.4)	1 (0.89~1.13)	0.961	0.76 (0.67~0.87)	<0.001
Trend.test	10137	3334 (32.9)	1 (0.96~1.04)	0.947	0.91 (0.88~0.95)	<0.001
METS-VF Model 2						
Variable	n.total	n.event_%	crude. OR (95%CI)	crude. P value	M2. OR (95%CI)	adj. P value
Q1	2534	870 (34.3)	1(Ref)		1(Ref)	
Q2	2534	802 (31.6)	0.89 (0.79~1)	0.042	0.77 (0.62~0.94)	0.013
Q3	2534	790 (31.2)	0.87 (0.77~0.97)	0.017	0.76 (0.61~0.94)	0.011
Q4	2535	872 (34.4)	1 (0.89~1.13)	0.961	0.83 (0.67~1.04)	0.102
Trend.test	10137	3334 (32.9)	1 (0.96~1.04)	0.947	0.94 (0.88~1.01)	0.108

Note: METS-VF was converted into quartiles for analysis (first quartile [Q1]: lowest 25%; fourth quartile [Q4]: highes t 25%). Model 1 adjusted for age, sex, education, and BMI. Model 2 further adjusted for income, consumption, smoking, and alcohol consumption on the basis of Model 1.

4. Discussion

Utilizing nationally representative data from CHARLS 2015, this study is the first to systematically evaluate the associations between four novel Computational Metabolic Biomarkers (CVAI, LAP, METS-IR, METS-VF) and depressive symptoms among middle-aged and older adults. The core findings diverge significantly from conventional perspectives that view metabolic dysregulation solely as a pathogenic risk. Instead of a linear positive correlation, our multivariable-adjusted analysis revealed that these metabolic indices conferred significant protective effects. Specifically, a distinct U-shaped nonlinear association emerged, wherein moderate levels of CVAI, METS-IR, LAP, and METS-VF (typically the Q3 range) corresponded to the lowest incidence of depression.

This observed "protective threshold" aligns with recent findings by Miyuan Wang et al. [34] and lends empirical support to the "Jolly Fat" hypothesis [35]. Underpinning this hypothesis is the premise that moderate fat reserves may optimize mood regulation through distinct neurobiological pathways. Behaviorally, <dietary intake activates> the brain's reward system, promoting dopamine release to produce anxiolytic and antidepressant effects. Endocrinologically, <adipose tissue functions as an active organ> capable of converting androstenedione into estrone. Elevated estrogen levels, in turn, exert confirmed neuroprotective effects conducive to mood stability [36]. Moreover, from a physiological standpoint, <moderate nutritional reserves> act as a metabolic buffer against chronic wasting diseases in the elderly, thereby mitigating the psychological distress triggered by physical frailty.

Despite these plausible mechanisms, our results stand in contrast to prior literature—particularly studies on LAP [37]and METS-IR[38] often report linear positive correlations with depression. Such discrepancies likely stem from three methodological dimensions. The first is population heterogeneity: <previous research predominantly relied> on Western databases (e.g., NHANES) or younger cohorts (20 years), whereas CHARLS focuses exclusively on Chinese adults aged ≥45. Given that metabolic flexibility and psychological resilience vary significantly across the lifespan, moderate adiposity in the elderly may reflect robust nutritional status rather than pathological burden—a phenomenon known as the "obesity paradox." Second, variation in assessment tools plays a role, as distinct scales (e.g., PHQ-9 vs. CES-D-10) capture different symptom dimensions. Finally, statistical modeling approaches differ; <prior studies on indices like METS-VF> often assumed linearity [39], potentially overlooking the non-linear U-shaped threshold effect characteristic of this specific demographic.

The U-shaped trajectory revealed herein carries substantial public health implications. It suggests that for middle-aged and older adults, <overly stringent weight control or minimal body fat> may not represent the optimal state for mental health. Conversely, maintaining moderate nutritional intake and adiposity

(corresponding to the Q3 level) may confer a protective benefit against depression. These findings do not advocate for pathological obesity, but rather <challenge the "thinner is always better" misconception> in clinical guidance, promoting instead a balanced energy metabolic state.

The strengths of this investigation lie in its large-scale, nationally representative sampling and the innovative comparative analysis of multiple novel IR indices. Future research could leverage more advanced machine learning techniques, such as explainable AI (e.g., SHAP analysis), to automate the detection of non-linear thresholds and improve the predictive power of these indices. Furthermore, integrating our findings with multi-omics data could pave the way for developing more sophisticated computational models for geriatric depression[27]. Interpretations, however, must be tempered by several limitations. First, <the cross-sectional design precludes causal inference>, leaving open the possibility of reverse causation (e.g., depression-induced weight loss). Second, <reliance on self-reported scales> reflects symptom severity rather than a clinical diagnosis. Third, <residual confounding may persist> despite rigorous adjustment for socioeconomic and lifestyle covariates.

5. Conclusions

This study reveals U-shaped associations between novel metabolic indices (CVAI, LAP, METS-IR, METS-VF) and depressive symptoms among Chinese middle-aged and older adults, indicating that moderate metabolic levels may confer mental health benefits. These findings challenge the conventional linear risk perspective and highlight the potential of computational biomarkers for precise depression screening in aging populations. However, the cross-sectional design limits causal inference, and self-reported measures may introduce bias. Future longitudinal studies and advanced modeling approaches are warranted to validate these non-linear patterns and inform targeted interventions for geriatric mental health

References

[1] de Winter, C. F., Bastiaanse, L. P., Hilgenkamp, T. I., Evenhuis, H. M. and Echteld, M. A. Overweight and obesity in older people with intellectual disability. Research in Developmental Disabilities. 2012, 33(2), pp. 398-405. https://doi.org/10.1016/j.ridd.2011.09.022.

[2] Liu, Q., He, H., Yang, J., Feng, X., Zhao, F. and Lyu, J. Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. Journal of Psychiatric Research. 2020, 126, pp. 134-140. https://doi.org/10.1016/j.jpsychires.2019.08.002.

[3] Ustun, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C. and Murray, C. J. Global burden of depressive disorders in the year 2000. British Journal of Psychiatry. 2004, 184(5), pp. 386-92. https://doi.org/10.1192/bjp.184.5.386.

[4] Disease, G. B. D., Injury and Risk Factor, C. Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet. 2025, 406(10513), pp. 1873-1922. https://doi.org/10.1016/S0140-6736(25)01637-X.

[5] Zhou, X., Li, J., Gu, W., Wang, J., Zhu, Y., Zhang, G., Ding, Y. and Tang, Y. Prevalence and associated factors of anxiety and depression among patients with chronic respiratory diseases in eight general hospitals in Jiangsu Province of China: A cross-sectional study. Psychiatry Research. 2017, 251, pp. 48-53. https://doi.org/10.1016/j.psychres.2017.01.070.

[6] Petersen, M. C. and Shulman, G. I. Mechanisms of Insulin Action and Insulin Resistance. Physiological Reviews. 2018, 98(4), pp. 2133-2223. https://doi.org/10.1152/physrev.00063.2017.

[7] Zimmet, P., Alberti, K. G. and Shaw, J. Global and societal implications of the diabetes epidemic. Nature. 2001, 414(6865), pp. 782-7. https://doi.org/10.1038/414782a.

[8] Sharma, V. R., Matta, S. T., Haymond, M. W. and Chung, S. T. Measuring Insulin Resistance in Humans. Hormone Research in Paediatrics. 2020, 93(11-12), pp. 577-588. https://doi.org/10.1159/000515462.

[9] Zhou, X., Kang, C., Hu, Y. and Wang, X. Study on insulin resistance and ischemic cerebrovascular disease: A bibliometric analysis via CiteSpace. Front Public Health. 2023, 11, p. 1021378. https://doi.org/10.3389/fpubh.2023.1021378.

[10] Fernandes, B. S., Salagre, E., Enduru, N., Grande, I., Vieta, E. and Zhao, Z. Insulin resistance in depression: A large meta-analysis of metabolic parameters and variation. Neuroscience and Biobehavioral Reviews. 2022, 139, p. 104758. https://doi.org/10.1016/j.neubiorev.2022.104758.

[11] Brouwer, A., van Raalte, D. H., Lamers, F., Rutters, F., Elders, P. J. M., Van Someren, E. J. W., Snoek, F. J., Beekman, A. T. F. and Bremmer, M. A. Insulin resistance as a marker for the immune-metabolic subtype of depression. Journal of Affective Disorders. 2021, 295, pp. 1371-1376. https://doi.org/10.1016/j.jad.2021.08.151.

[12] Xia, M. F., Chen, Y., Lin, H. D., Ma, H., Li, X. M., Aleteng, Q., Li, Q., Wang, D., Hu, Y., Pan, B. S., et al. A indicator of visceral adipose dysfunction to evaluate metabolic health in adult Chinese. Scientific Reports. 2016, 6(1), p. 38214. https://doi.org/10.1038/srep38214.

[13] Amato, M. C., Giordano, C., Galia, M., Criscimanna, A., Vitabile, S., Midiri, M., Galluzzo, A. and AlkaMeSy Study, G. Visceral Adiposity Index: a reliable indicator of visceral fat function associated with cardiometabolic risk. Diabetes Care. 2010, 33(4), pp. 920-2. https://doi.org/10.2337/dc09-1825.

[14] Yang, Y., Li, S., Ren, Q., Qiu, Y., Pan, M., Liu, G., Zheng, R., An, Z. and Li, S. The interaction between triglyceride-glucose index and visceral adiposity in cardiovascular disease risk: findings from a nationwide Chinese cohort. Cardiovascular Diabetology. 2024, 23(1), p. 427. https://doi.org/10.1186/s12933-024-02518-2.

[15] Pan, L., Xu, Q., Liu, J., Gao, Y., Li, J., Peng, H., Chen, L., Wang, M., Mai, G. and Yang, S. Dose-response relationship between Chinese visceral adiposity index and type 2 diabetes mellitus among middle-aged and elderly Chinese. Frontiers in Endocrinology. 2022, 13, p. 959860. https://doi.org/10.3389/fendo.2022.959860.

[16] Li, H., Zhang, Y., Luo, H. and Lin, R. The lipid accumulation product is a powerful tool to diagnose metabolic dysfunction-associated fatty liver disease in the United States adults. Frontiers in Endocrinology. 2022, 13, p. 977625. https://doi.org/10.3389/fendo.2022.977625.

[17] Zhou, T., Chen, S., Mao, J., Zhu, P., Yu, X. and Lin, R. Association between obstructive sleep apnea and visceral adiposity index and lipid accumulation product: NHANES 2015-2018. Lipids in Health and Disease. 2024, 23(1), p. 100. https://doi.org/10.1186/s12944-024-02081-5.

[18] Simental-Mendia, L. E., Rodriguez-Moran, M. and Guerrero-Romero, F. The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. Metabolic Syndrome and Related Disorders. 2008, 6(4), pp. 299-304. https://doi.org/10.1089/met.2008.0034.

[19] Bello-Chavolla, O. Y., Almeda-Valdes, P., Gomez-Velasco, D., Viveros-Ruiz, T., Cruz-Bautista, I., Romo-Romo, A., Sanchez-Lazaro, D., Meza-Oviedo, D., Vargas-Vazquez, A., Campos, O. A., et al. METS-IR, a novel score to evaluate insulin sensitivity, is predictive of visceral adiposity and incident type 2 diabetes. European Journal of Endocrinology of the European Federation of Endocrine Societies. 2018, 178(5), pp. 533-544. https://doi.org/10.1530/EJE-17-0883.

[20] Duan, M., Zhao, X., Li, S., Miao, G., Bai, L., Zhang, Q., Yang, W. and Zhao, X. Metabolic score for insulin resistance (METS-IR) predicts all-cause and cardiovascular mortality in the general population:

evidence from NHANES 2001-2018. Cardiovascular Diabetology. 2024, 23(1), p. 243. https://doi.org/10.1186/s12933-024-02334-8.

[21] Su, X., Zhao, C. and Zhang, X. Association between METS-IR and heart failure: a cross-sectional study. Frontiers in Endocrinology. 2024, 15, p. 1416462. https://doi.org/10.3389/fendo.2024.1416462.

[22] Feng, L., Chen, T., Wang, X., Xiong, C., Chen, J., Wu, S., Ning, J. and Zou, H. Metabolism Score for Visceral Fat (METS-VF): A New Predictive Surrogate for CKD Risk. Diabetes, Metabolic Syndrome and Obesity. 2022, 15, pp. 2249-2258. https://doi.org/10.2147/DMSO.S370222.

[23] Tan, A., Yang, S., Pan, Y. and Lin, Q. Metabolism score for visceral fat (METS-VF): an innovative and powerful predictor of stroke. Archives of Medical Science. 2024, 20(5), pp. 1710-1714. https://doi.org/10.5114/aoms/193709.

[24] Bello-Chavolla, O. Y., Antonio-Villa, N. E., Vargas-Vazquez, A., Viveros-Ruiz, T. L., Almeda-Valdes, P., Gomez-Velasco, D., Mehta, R., Elias-Lopez, D., Cruz-Bautista, I., Roldan-Valadez, E., et al. Metabolic Score for Visceral Fat (METS-VF), a novel estimator of intra-abdominal fat content and cardio-metabolic health. Clinical Nutrition. 2020, 39(5), pp. 1613-1621. https://doi.org/10.1016/j.clnu.2019.07.012.

[25] Li, T., Ren, Y., Liu, M., Wang, Q., Sun, T., Cao, J. and Cui, H. Association between METS-VF and sarcopenia among middle-aged and older adults in China: The first longitudinal evidence from CHARLS. Experimental Gerontology. 2025, 206, p. 112778. https://doi.org/10.1016/j.exger.2025.112778.

[26] Xie, M., Zhang, Y., Wu, H., Wu, Z., Han, H., Xie, X., Zhang, R., Cheng, J. and Xu, J. Correlation of triglyceride-glucose index with the incidence and prognosis of hyperglycemic crises in critically ill patients with diabetes mellitus: a machine-learning-based multicenter retrospective cohort study. Frontiers in Nutrition. 2025, 12, p. 1649553. https://doi.org/10.3389/fnut.2025.1649553.

[27] Cheng, N., Chen, Y., Jin, L. and Chen, L. Nonlinear association between visceral fat metabolism score and heart failure: insights from LightGBM modeling and SHAP-Driven feature interpretation in NHANES. BMC Medical Informatics and Decision Making. 2025, 25(1), p. 223. https://doi.org/10.1186/s12911-025-03076-7.

[28] Chen, H. and Mui, A. C. Factorial validity of the Center for Epidemiologic Studies Depression Scale short form in older population in China. International Psychogeriatrics. 2014, 26(1), pp. 49-57. https://doi.org/10.1017/S1041610213001701.

[29] Boey, K. W. Cross-validation of a short form of the CES-D in Chinese elderly. International Journal of Geriatric Psychiatry. 1999, 14(8), pp. 608-17. https://doi.org/10.1002/(sici)1099-1166(199908)14:8<608::aid-gps991>3.0.co;2-z.

[30] Andresen, E. M., Malmgren, J. A., Carter, W. B. and Patrick, D. L. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). American Journal of Preventive Medicine. 1994, 10(2), pp. 77-84.

[31] Zhang, P., Wang, L., Zhou, Q., Dong, X., Guo, Y., Wang, P., He, W., Wang, R., Wu, T., Yao, Z., et al. A network analysis of anxiety and depression symptoms in Chinese disabled elderly. Journal of Affective Disorders. 2023, 333, pp. 535-542. https://doi.org/10.1016/j.jad.2023.04.065.

[32] Sheng, G., Lu, S., Xie, Q., Peng, N., Kuang, M. and Zou, Y. The usefulness of obesity and lipid-related indices to predict the presence of Non-alcoholic fatty liver disease. Lipids in Health and Disease. 2021, 20(1), p. 134. https://doi.org/10.1186/s12944-021-01561-2.

[33] Huo, R. R., Liao, Q., Zhai, L., You, X. M. and Zuo, Y. L. Interacting and joint effects of triglyceride-glucose index (TyG) and body mass index on stroke risk and the mediating role of TyG in middle-aged and older Chinese adults: a nationwide prospective cohort study. Cardiovascular Diabetology. 2024, 23(1), p. 30. https://doi.org/10.1186/s12933-024-02122-4.

[34] Bai, S., Wang, J., Liu, J., Miao, Y., Zhang, A. and Zhang, Z. Analysis of depression incidence and influence factors among middle-aged and elderly diabetic patients in China: based on CHARLS data. BMC Psychiatry. 2024, 24(1), p. 146. https://doi.org/10.1186/s12888-023-05473-6.

[35] Crisp, A. H. and McGuiness, B. Jolly fat: relation between obesity and psychoneurosis in general population. British Medical Journal. 1976, 1(6000), pp. 7-9. https://doi.org/10.1136/bmj.1.6000.7.

[36] Kundakovic, M. and Rocks, D. Sex hormone fluctuation and increased female risk for depression and anxiety disorders: From clinical evidence to molecular mechanisms. Frontiers in Neuroendocrinology. 2022, 66, p. 101010. https://doi.org/10.1016/j.yfrne.2022.101010.

[37] Huang, Y., Zhao, D., Yang, Z., Wei, C. and Qiu, X. The relationship between VAI, LAP, and depression and the mediation role of sleep duration-evidence from NHANES 2005-2020. BMC Psychiatry. 2025, 25(1), p. 228. https://doi.org/10.1186/s12888-025-06631-8.

[38] Huang, Q., Wang, D., Chen, S., Tang, L. and Ma, C. Association of METS-IR index with depressive symptoms in US adults: A cross-sectional study. Journal of Affective Disorders. 2024, 355, pp. 355-362. https://doi.org/10.1016/j.jad.2024.03.129.

[39] Liu, H., Dong, H., Zhou, Y., Jin, M., Hao, H., Yuan, Y. and Jia, H. The association between Metabolic Score for Visceral Fat and depression in overweight or obese individuals: evidence from NHANES. Frontiers in Endocrinology. 2024, 15, p. 1482003. https://doi.org/10.3389/fendo.2024.1482003.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260114

Noninvasive Brain Stimulation Techniques for the Recovery of Unilateral Visual Impairment After Stroke: A Systematic Review

Xinyi Zhang and Yuxia Ma *

School of Nursing, Lanzhou University, Lanzhou 730000, China

**Corresponding author: Yuxia Ma*

Abstract

According to data from the Global Stroke Report 2025, stroke accounted for 10.72% of all deaths worldwide in 2021. Monocular visual field defects represent one of the most prevalent and severe cognitive impairments following stroke, with a prevalence rate reaching up to 30%. Current conventional treatments for monocular visual field defects yield limited efficacy, whilst non-invasive brain stimulation techniques are advancing rapidly. With the continuous development of these non-invasive brain stimulation technologies, an increasing body of research indicates that they may offer certain restorative effects for post-stroke hemianopia. To synthesize the effects of noninvasive brain stimulation techniques in the recovery of unilateral visual impairment after stroke. A comprehensive search was conducted in six databases, including PubMed, Web of Science, Cochrane Library, CNKI, Wan fang, and VIP, as of January 3, 2026. The articles were screened based on the exclusion criteria. The included articles were analyzed based on their impact on visual function improvement, quality of life, and adverse effects. Sixteen studies were analyzed, demonstrating that noninvasive brain stimulation significantly improved visual function and quality of life for patients. Additionally, adverse effects were minimal. Noninvasive brain stimulation technology shows promise in the recovery of unilateral visual impairment after stroke. However, further multi-center, large-sample, high-quality randomized controlled trials are necessary to verify its efficacy and safety.

Keywords

noninvasive brain stimulation, stroke, unilateral visual impairment

1. Introduction

Stroke is a common neurological disorder. According to the Global Stroke Report 2025, there were 11.946 million new stroke cases worldwide in 2021, resulting in 7.253 million deaths, accounting for 10.7% of all deaths globally. Among these deaths, those caused by ischemic stroke (IS), intracerebral hemorrhage (ICH), and subarachnoid hemorrhage (SAH) accounted for 49.5% (3.591 million), 45.6% (3.308 million), and 4.9% (0.353 million), respectively. Stroke is the second leading cause of death worldwide and ranks third in the global burden of disease [1]. Stroke can lead to unilateral visual impairment, which is one of the most common and severe cognitive deficits following right-hemisphere stroke, with a prevalence of up to

30% [2]. Brain injury affecting the fronto-parietal subcortical cortical network involved in spatial representation and awareness may explain the mechanism underlying hemispatial neglect [3]. Unilateral neglect is characterized by a failure to report, respond to, or orient toward stimuli presented on the side opposite the brain lesion. The presence of unilateral neglect significantly delays recovery from mild hemiparesis, and patients frequently experience more difficulties in activities of daily living [4].

Among the therapeutic approaches for hemispatial neglect, the three most commonly used interventions are visual scanning training, limb activation therapy, and prism adaptation therapy [5]. However, conventional rehabilitation methods often show limited effectiveness, highlighting the need to explore new treatment strategies. In recent years, non-invasive brain stimulation (NIBS) has developed rapidly and has attracted considerable attention because of its non-invasive and safe characteristics. Repetitive transcranial magnetic stimulation (rTMS) can stimulate or inhibit cortical activity by modulating the frequency of the magnetic field, thereby inducing corresponding electrical currents. Theta burst stimulation (TBS) is a patterned form of rTMS with a shorter total stimulation duration [6] and lower stimulation intensity [7]. Its pulses are delivered at a faster rate and can be applied intermittently or continuously. TBS is believed to induce long-lasting changes in cortical excitability that persist beyond the stimulation period [8].

With the continuous development of non-invasive brain stimulation techniques, an increasing number of studies have demonstrated their potential value in the recovery of unilateral visual impairment after stroke [9]. However, systematic evaluations of the role of non-invasive brain stimulation in the recovery of unilateral visual impairment following stroke remain limited. Therefore, this systematic review aims to clarify the role of non-invasive brain stimulation techniques in the recovery of unilateral visual impairment after stroke.

2. Materials and Methods

2.1 Search Strategy

This systematic review was conducted in accordance with the PRISMA 2020 Statement. Relevant literature was retrieved from six databases, including PubMed, Web of Science, Cochrane Library, CNKI, Wanfang Data, and VIP Database. The search period extended from database inception to January 3, 2026. The following keywords were used for the literature search: “stroke”, “cerebrovascular accident”, “unilateral spatial neglect”, “spatial neglect”, “unilateral neglect”, “transcranial magnetic stimulation”, “TMS”, “repetitive transcranial magnetic stimulation”, “rTMS”, “theta burst stimulation”, and “TBS”.

2.2 Inclusion and Exclusion Criteria

Inclusion Criteria:

1) Original studies published in peer-reviewed journals.
2) Participants were adult patients diagnosed with visuospatial neglect after stroke (ischemic or hemorrhagic).
3) The included studies were randomized controlled trials (RCTs).
4) The control group received sham stimulation or conventional stroke rehabilitation, while the experimental group received non-invasive brain stimulation interventions in addition to the treatments given to the control group, including but not limited to tDCS, rTMS, TBS, iTBS, tACS, and PAS.

Exclusion Criteria:

1) Study types including meta-analyses, reviews, conference papers, or case reports.
2) Studies with insufficient data, and complete data could not be obtained even after contacting the authors.
3) Participants with visual impairment, abnormal vision, or corrected vision problems unrelated to stroke.
4) Patients with metal implants, cardiac pacemakers, or cochlear implants.

5) Patients with comorbidities that are contraindications to non-invasive brain stimulation, such as malignant tumors, open wounds, a history of epilepsy, or a family history of epilepsy.

2.3 Study Selection

All retrieved records were imported into EndNote 9.0 reference management software. After removing duplicates, the studies were screened. The titles, abstracts, and full texts were independently reviewed according to the predefined inclusion and exclusion criteria, and the reasons for exclusion were documented for each excluded article.

2.4 Quality Assessment

Since the included studies were randomized controlled trials, the Cochrane Risk of Bias (ROB) assessment tool was used to evaluate methodological quality. Evidence quality was assessed from several aspects, including research objectives, study design, study population, observation or measurement methods, outcome analysis, quality control, and result reporting. After completing the quality assessment, Review Manager 5.3 software was used to generate the risk-of-bias summary figures and individual quality assessment results.

2.5 Data Extraction and Analysis

Information extraction and analysis were conducted for studies that passed the quality assessment. The extracted study characteristics included: Clinical characteristics, including the first author and year of publication, sample size (experimental group/control group), and treatment conclusions; Spatial neglect-related outcome measures, including the Star Cancellation Test, Line Bisection Test, Line Cancellation Test, and Albert Test, among others. Due to the substantial heterogeneity among the included studies, performing a meta-analysis might introduce bias. Therefore, a systematic review approach was adopted to comprehensively analyze the role of non-invasive brain stimulation techniques in the functional recovery of patients with unilateral visuospatial impairment after stroke.

3. Results

3.1 Study Selection Results and Characteristics of Included Studies

A total of 16 relevant studies were included, comprising 15 English-language articles and 1 Chinese-language article. The studies involved a total sample size of 375 patients, among whom 209 received non-invasive brain stimulation treatment, while 160 received conventional therapy or sham stimulation as placebo treatment. The literature screening process is shown in Figure 1.

3.2 Quality Assessment Results

Among the 16 included studies [10–25], the participants were randomly assigned using methods such as random number tables or computer-generated randomization. In 11 studies, the method of allocation concealment was not clearly described, and one study did not employ allocation concealment, which may have introduced a certain degree of selection bias. Six studies explicitly reported a double-blind experimental design, and most studies did not present other sources of bias. Overall, the methodological quality of the included studies was moderate. The results of the methodological quality assessment are presented in Figure 2 and Figure 3.

3.3 Improvement in Visual Function

Among the included studies, 14 used the Line Bisection Test, and 9 used the Star Cancellation Test as outcome measures for evaluating unilateral neglect function. The results of multiple studies indicated that visual spatial bias improved after intervention with non-invasive brain stimulation techniques. One study reported that after 10 sessions of rTMS treatment administered daily, the Line Bisection Test showed that the mean improvement in leftward deviation of line bisection in the rTMS experimental group was greater than that in the control group. Studies conducted by Kim Yong-Kyun et al. [11] and Lim et al. [23] both

confirmed and recommended the use of low-frequency rTMS at a frequency of 10 sessions per day for non-invasive brain stimulation rehabilitation training. The basic characteristics of the included studies are summarized in Table 1.

Figure 1: Flow diagram of the literature selection process

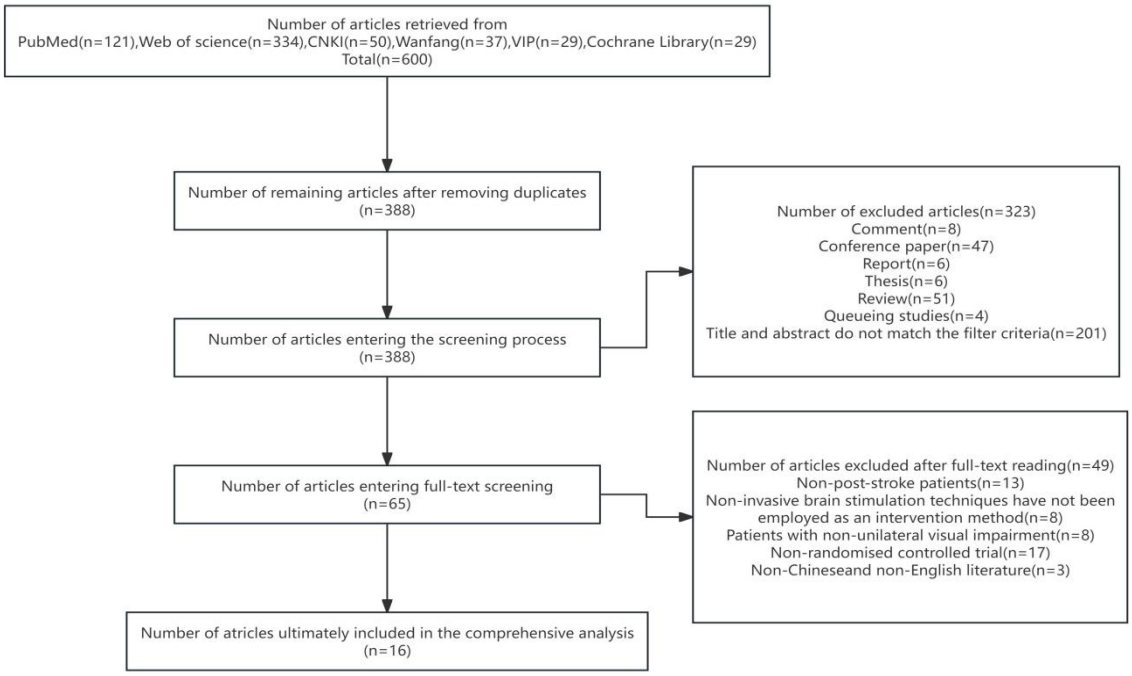


Figure 2: Risk-of-bias graph

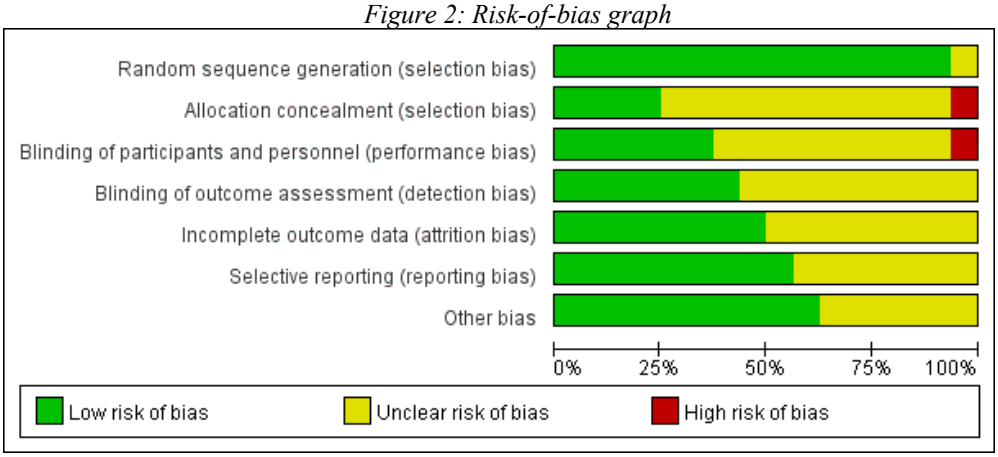


Figure 3: Risk-of-bias graph

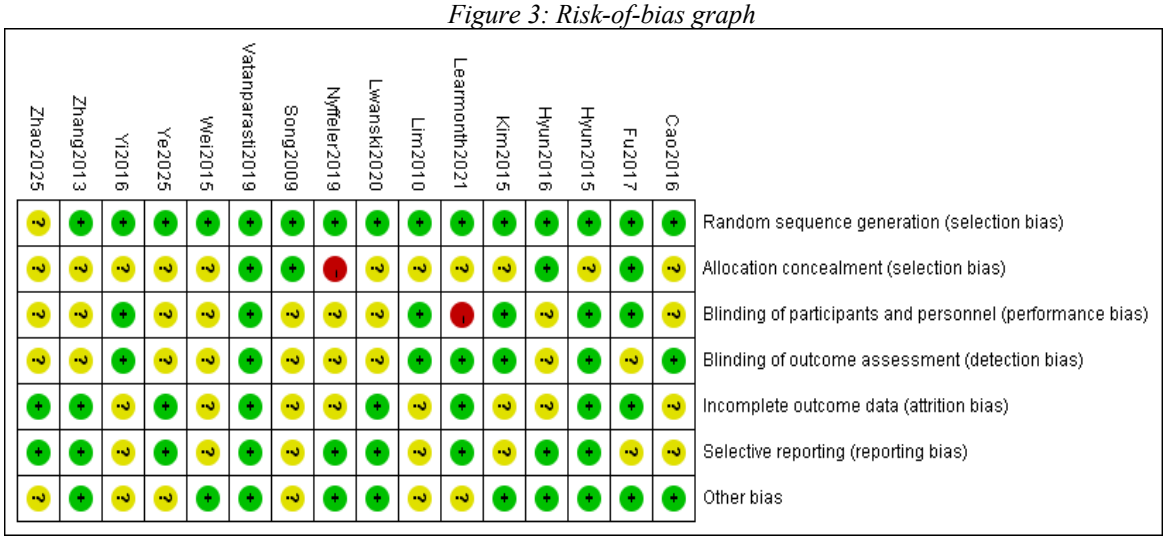


Table 1: Characteristics of the Included Studies

First Author / Year	Sample Size (Experimental / Control)	Study Type	Outcome Measures	Study Conclusions
Kim Yong·Kyun/2015	19/15	RCT	Letter cancelation test, Line bisection test, Ota's task	Compared with a single daily session of rTMS, ten sessions of low-frequency rTMS per day applied to the left parietal cortex had a more positive effect on spatial neglect. Therefore, low-frequency rTMS administered 10 times per day may be used for the treatment of left hemispatial neglect after stroke.
Wei Yang/2015	28/10	RCT	Star cancelation test, Line bisection test	The study provided strong evidence that rTMS significantly improved the neurocognitive function of VSN, with continuous TBS showing the best therapeutic effect. According to DTI assessment, enhanced connectivity within white-matter tract networks related to visual attention may represent a potential mechanism underlying the recovery of VSN observed with continuous TBS.
Fu Wei/2017	6/6	RCT	Star cancelation test, Line bisection test	The present study indicated that continuous theta stimulation combined with visual scanning training and motor function training can regulate RSFC within the attention network, thereby promoting recovery from behavioral deficits in patients with post-stroke VSN. The stimulation parameters used in this study may provide valuable references for clinical treatment strategies for VSN and functional recovery after stroke.
Hyun Gyu Cha/2015	12/12	RCT	Motor Free Visual Perception Test, Star cancelation test, Line bisection test, Albert test	The findings suggest that repetitive transcranial magnetic stimulation may help reduce unilateral neglect in patients with stroke.
Hyun Gyu·Cha/2016	15/15	RCT	Line bisection test, Albert test	The study suggests that rTMS may be effective in improving unilateral neglect and motor function.
Cao Lei/2016	7/6	RCT	Star cancelation test, Line	The results support the critical role of the left dorsolateral prefrontal cortex (LDLPFC) in regulating the attention network involved in neglect,

First Author / Year	Sample Size (Experimental / Control)	Study Type	Outcome Measures	Study Conclusions
			bisection test	and enhancing LDLPFC activity through intermittent theta burst stimulation may facilitate recovery of VSN in stroke patients.
Yi You Gyoung/2016	20/10	RCT	Motor Free Visual Perception Test, Star cancelation test, Line bisection test, Catherine Bergego Scale	The results showed that anodal tDCS facilitated the right cortex while cathodal tDCS inhibited the left cortex, thereby improving symptoms of visuospatial neglect. Therefore, tDCS may serve as an effective adjunctive therapy for the recovery of neglect symptoms.
Vatanparasti, S./2019	7/7	RCT	Star cancelation test, Line bisection test, Figure Copying Test, Clock Drawing, Modified Rankin Scale	The study indicated that transcranial magnetic stimulation did not enhance the effect of prism adaptation training on neglect symptoms in stroke patients.
lwarski, S./2020	14/14	RCT	BIT-c, BIT-b, Visuospatial Scale	No statistically significant differences were found in outcome measures between the rTMS group and the sham stimulation group. The magnitude of stimulation effects was not related to lesion volume, lesion location, or baseline motor threshold. The study did not confirm the effectiveness of 1 Hz rTMS as an adjunct to visual scanning training in patients with subacute stroke and VSN.
Nyffeler T/2019	20/10	RCT	Catherine Bergego Scale, Fluff test, Two-Part-Picture Test, bird cancellation task	The study showed that cTBS applied to the contralesional posterior parietal cortex significantly improved and accelerated recovery from neglect, as well as related overall functional outcomes, in patients with intact interhemispheric connections.
Song W./2009	7/7	RCT	Line bisection test, Line cancelation test	This study suggested that low-frequency rTMS applied to the unaffected hemisphere may improve visuospatial neglect after stroke, supporting the interhemispheric competition theory of the attention network, and highlighting the need for further research.
Lim, J. Y./2010	7/7	RCT	Line bisection test, Albert test	The study demonstrated that low-frequency rTMS applied to the unaffected left parietal region was safe and improved performance on the Line Bisection Test, suggesting that non-invasive cortical stimulation may serve as an adjunctive strategy for cognitive rehabilitation in patients with hemispatial neglect. Further prospective randomized sham-controlled trials are required to confirm the beneficial effects.
Zhao, W. Y./2025	10/10	RCT	Line bisection test, Line cancelation test, Star	The study showed that iTBS not only alleviated symptoms of unilateral neglect (UN) but also significantly enhanced cortical excitability at the stimulation site, contributing to reconstruction of

First Author / Year	Sample Size (Experimental / Control)	Study Type	Outcome Measures	Study Conclusions
			cancelation test	brain networks, particularly functional connectivity between the hemispheres.
Ye, L. L./2025	10/10	RCT	Line bisection test, Line cancelation test, Star cancelation test	The findings indicated that iTBS significantly improved symptoms of UN and enhanced excitability at the stimulation site. Moreover, iTBS promoted early enhancement of connectivity within the fronto-parietal network, demonstrating potential therapeutic value for restoring neural network balance.
Learmonth, G./2021	18/6	RCT	BIT, Line bisection test, The Balloons test, The Broken Hearts test, Visual field testing	The results showed that patients receiving tDCS alone or tDCS combined with behavioral training exhibited varying degrees of improvement in core unilateral neglect assessment indicators such as the Line Bisection Test and Balloon Test. The findings also suggest that future studies may explore the feasibility of home-based tDCS interventions for post-stroke unilateral spatial neglect.
Zhang/2013	15/15	RCT	Line bisection test, Line cancelation test, Clock Drawing	The study showed that rTMS inhibited excessive cortical excitability in the unaffected hemisphere, promoted restoration of excitatory balance between the bilateral hemispheres, improved neglect symptoms, and also contributed to recovery of motor function.

Notes:
RCT = randomized controlled trial
TBS / rTMS / tDCS = transcranial magnetic stimulation techniques
DTI = diffusion tensor imaging
VSN / UN = visuospatial neglect / unilateral neglect
iTBS = intermittent theta burst stimulation
BIT = Behavioural Inattention Test
BIT-C = conventional subtests of the BIT
BIT-B = selected behavioral subtests of the BIT

3.4 Quality of Life and Adverse Effects

The results indicated that patients receiving non-invasive brain stimulation (NIBS) showed significant improvements in visual function and quality of life. Specifically, patients exhibited improvements in visual spatial bias after treatment, and their quality of life was also significantly enhanced. In addition, most patients did not experience obvious adverse reactions. Only a small number of patients reported mild symptoms such as headache and localized skin itching.

4. Discussion

This study included 16 randomized controlled trials and comprehensively evaluated the effectiveness of non-invasive brain stimulation techniques in the recovery of unilateral visual impairment after stroke. The results suggest that non-invasive brain stimulation has a certain therapeutic effect on the recovery of unilateral visual impairment after stroke. The underlying mechanisms may be related to regulation of neurotransmitter secretion, improvement of neural conduction pathways, and promotion of neural regeneration.

Non-invasive brain stimulation techniques directly act on damaged brain regions through external stimulation, particularly for behavioral and cognitive disorders such as unilateral neglect, enabling targeted improvement of related symptoms. This approach can better meet patients' rehabilitation needs and achieve optimal therapeutic outcomes. Moreover, NIBS may not only improve symptoms in the short term, but also promote recovery and reorganization of brain function through modulation of neuroplasticity, thereby providing sustained long-term therapeutic effects [26].

With the development of neuromodulation techniques and multimodal neuroimaging, the mechanisms underlying the effects of NIBS on post-stroke unilateral neglect have been increasingly explored. Mechanistic research has gradually shifted from focusing on local changes in cortical excitability to investigating reorganization of brain network topology. Previous studies have suggested that relative hyperexcitability of the contralesional hemisphere and functional suppression of the lesioned hemisphere after stroke, together with interhemispheric inhibition imbalance via the corpus callosum, constitute an important mechanism underlying unilateral spatial neglect (USN). In recent years, evidence from TMS-EEG, functional connectivity analysis, and structural connectivity studies has provided more refined neurophysiological support for this theory.

Ting et al. reported that rTMS improves unilateral neglect primarily by suppressing attentional bias toward the contralesional side and restoring orienting balance between the two hemispheres [27]. With further research, studies by Zhao et al. and Ye et al., using TMS-EEG techniques, demonstrated that NIBS improves post-stroke unilateral neglect by enhancing cortical excitability at the stimulation target and reshaping functional connectivity within brain networks [17,22]. These findings suggest that the effects of NIBS are not limited to simply "activating the lesioned hemisphere" or "inhibiting the contralesional hemisphere," but rather involve reconstruction of cross-regional network synchronization to restore the dynamic balance of the attention system. From the perspective of structural connectivity, Yang et al. provided complementary evidence, indicating that enhanced connectivity of white matter tract networks associated with visual attention may represent a potential mechanism underlying the therapeutic effects of continuous TBS (cTBS) in USN [10]. This finding extends the effects of NIBS from functional connectivity modulation to structural connectivity remodeling, suggesting that stimulation may improve spatial attention deficits by promoting compensatory strengthening of residual white matter pathways or enhancing network integration efficiency. Therefore, the therapeutic mechanism of NIBS in post-stroke unilateral neglect has evolved from the early "single-region excitation/inhibition hypothesis" to a "brain network plasticity-based systemic regulation model." This theoretical advancement not only provides neurophysiological evidence for optimizing stimulation parameters, but also lays the foundation for developing individualized neurorehabilitation strategies.

However, there is considerable heterogeneity in stimulation parameters used in NIBS for the treatment of post-stroke unilateral neglect, mainly due to the lack of uniform standards and consensus. Taking transcranial direct current stimulation (tDCS) as an example, the effects of parameters such as stimulation intensity, duration, and frequency remain unclear. Some studies have suggested that high-intensity tDCS may more effectively improve symptoms of unilateral neglect, whereas other studies indicate that low-intensity tDCS may be safer and equally effective [23]. In addition, there are differing views regarding stimulation duration and frequency [28,29].

Furthermore, repetitive transcranial magnetic stimulation (rTMS) and other non-invasive brain stimulation techniques, such as transcranial ultrasound stimulation, also exhibit heterogeneity in stimulation parameters when used to treat post-stroke unilateral neglect [7]. The selection of stimulation parameters in these techniques is often based on researchers' experience and experimental design, rather than standardized guidelines.

Therefore, to facilitate the clinical application of NIBS in the treatment of post-stroke unilateral neglect, further research is required to determine optimal stimulation parameters and establish standardized protocols. In addition, greater attention should be paid to individual differences among patients and heterogeneity in treatment outcomes.

5. Conclusion

This study updates the current evidence regarding the role of non-invasive brain stimulation in post-stroke unilateral visual impairment. The results indicate that non-invasive brain stimulation has certain benefits in promoting recovery from unilateral visual impairment after stroke and can be used as an adjunctive technique in routine rehabilitation training. Low-frequency rTMS administered 10 sessions per day is recommended as a training protocol. Moreover, in recent years, randomization, blinding, and allocation concealment in randomized controlled trials have been implemented more rigorously, resulting in higher methodological quality compared with earlier studies.

However, several limitations should be acknowledged. First, the number of included studies was relatively small, and the sample sizes were limited, which may affect the stability and reliability of the results. Second, treatment protocols and outcome measures varied across studies, which may have influenced the results of the overall analysis. In addition, some studies had relatively low methodological quality or small sample sizes, which may have introduced bias. Therefore, further high-quality randomized controlled trials are needed to verify the effectiveness and mechanisms of non-invasive brain stimulation in the recovery of unilateral visual impairment after stroke.

Future research should focus on the mechanisms of action of non-invasive brain stimulation and the upstream and downstream regulatory targets involved. Large-scale, multicenter randomized controlled trials are also needed to validate the efficacy and safety of non-invasive brain stimulation techniques. Moreover, exploring the combined application of non-invasive brain stimulation with other therapeutic approaches may be more suitable for clinical implementation. In addition, the current stimulation parameter settings for non-invasive brain stimulation lack standardized protocols, which warrants further investigation in future studies.

References

[1] MARTIN S S, ADAY A W, ALLEN N B, et al. 2025 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association [J]. Circulation, 2025, 151(8): e41-e660.

[2] ZHANG J, WANG X, HUANG F, et al. Effect of Eye-Tracking Technology-Based Visual Scanning Training on Unilateral Spatial Neglect after Stroke [J]. Journal of visualized experiments : JoVE, 2025, (223).

[3] HEILMAN K M, VALENSTEIN E, WATSON R T. Neglect and related disorders [J]. Seminars in neurology, 2000, 20(4): 463-70.

[4] BOWEN A, HAZELTON C, POLLOCK A, et al. Cognitive rehabilitation for spatial neglect following stroke [J]. The Cochrane database of systematic reviews, 2013, 2013(7): Cd003586.

[5] SUPPA A, HUANG Y Z, FUNKE K, et al. Ten Years of Theta Burst Stimulation in Humans: Established Knowledge, Unknowns and Prospects [J]. Brain stimulation, 2016, 9(3): 323-35.

[6] RINGMAN J M, SAVER J L, WOOLSON R F, et al. Frequency, risk factors, anatomy, and course of unilateral neglect in an acute stroke cohort [J]. Neurology, 2004, 63(3): 468-74.

[7] HAO Z, WANG D, ZENG Y, et al. Repetitive transcranial magnetic stimulation for improving function after stroke [J]. The Cochrane database of systematic reviews, 2013, 2013(5): Cd008862.

[8] PASCUAL-LEONE A, VALLS-SOLé J, WASSERMANN E M, et al. Responses to rapid-rate transcranial magnetic stimulation of the human motor cortex [J]. Brain : a journal of neurology, 1994, 117 (Pt 4): 847-58.

[9] OBERMAN L, EDWARDS D, ELDAIEF M, et al. Safety of theta burst transcranial magnetic stimulation: a systematic review of the literature [J]. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society, 2011, 28(1): 67-74.

[10] YANG W, LIU T T, SONG X B, et al. Comparison of different stimulation parameters of repetitive transcranial magnetic stimulation for unilateral spatial neglect in stroke patients [J]. J Neurol Sci, 2015, 359(1-2): 219-25.

[11] KIM Y K, JUNG J H, SHIN S H. A comparison of the effects of repetitive transcranial magnetic stimulation (rTMS) by number of stimulation sessions on hemispatial neglect in chronic stroke patients [J]. Exp Brain Res, 2015, 233(1): 283-9.

[12] FU W, CAO L, ZHANG Y, et al. Continuous theta-burst stimulation may improve visuospatial neglect via modulating the attention network: a randomized controlled study [J]. Topics in stroke rehabilitation, 2017, 24(4): 236-41.

[13] VATANPARASTI S, KAZEMNEJAD A, YOONESSI A, et al. The Effect of Continuous Theta-Burst Transcranial Magnetic Stimulation Combined with Prism Adaptation on the Neglect Recovery in Stroke

Patients [J]. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association, 2019, 28(11): 104296.

[14] YI Y G, CHUN M H, DO K H, et al. The Effect of Transcranial Direct Current Stimulation on Neglect Syndrome in Stroke Patients [J]. Annals of rehabilitation medicine, 2016, 40(2): 223-9.

[15] CHA H G, KIM M K. Effects of repetitive transcranial magnetic stimulation on arm function and decreasing unilateral spatial neglect in subacute stroke: a randomized controlled trial [J]. Clinical rehabilitation, 2016, 30(7): 649-56.

[16] CHA H G, KIM M K. The effects of repetitive transcranial magnetic stimulation on unilateral neglect of acute stroke patients: A randomised controlled trial [J]. Hong Kong physiotherapy journal : official publication of the Hong Kong Physiotherapy Association Limited = Wu li chih liao, 2015, 33(2): 53-8.

[17] YE L, ZHAO W, ZHANG Y, et al. Exploring Neurobiological Effects of Intermittent Theta-Burst Stimulation on the Left Cerebellum for Post-stroke Unilateral Neglect: A Preliminary Transcranial Magnetic Stimulation-Electroencephalography Investigation [J]. Cerebellum, 2025, 24(4): 103.

[18] CAO L, FU W, ZHANG Y, et al. Intermittent θ burst stimulation modulates resting-state functional connectivity in the attention network and promotes behavioral recovery in patients with visual spatial neglect [J]. Neuroreport, 2016, 27(17): 1261-5.

[19] SONG W, DU B, XU Q, et al. Low-frequency transcranial magnetic stimulation for visual spatial neglect: a pilot study [J]. Journal of rehabilitation medicine, 2009, 41(3): 162-5.

[20] IWAŃSKI S, LEŚNIAK M, POLANOWSKA K, et al. Neuronavigated 1 Hz rTMS of the left angular gyrus combined with visuospatial therapy in post-stroke neglect [J]. NeuroRehabilitation, 2020, 46(1): 83-93.

[21] LEARMONTH G, BENWELL C S Y, MäRKER G, et al. Non-invasive brain stimulation in Stroke patients (NIBS): a prospective randomized open blinded end-point (PROBE) feasibility trial using transcranial direct current stimulation (tDCS) in post-stroke hemispatial neglect [J]. Neuropsychological rehabilitation, 2021, 31(8): 1163-89.

[22] ZHAO W, YE L, ZHANG Y, et al. Potential neurophysiological mechanisms of intermittent theta-burst stimulation to the left prefrontal cortex for post-stroke unilateral neglect: An exploratory transcranial magnetic stimulation - electroencephalography study [J]. IBRO Neurosci Rep, 2025, 19: 604-13.

[23] LIM J Y, KANG E K, PAIK N J. Repetitive transcranial magnetic stimulation to hemispatial neglect in patients after stroke: an open-label pilot study [J]. Journal of rehabilitation medicine, 2010, 42(5): 447-52.

[24] NYFFELER T, VANBELLINGEN T, KAUFMANN B C, et al. Theta burst stimulation in neglect after stroke: functional outcome and response variability origins [J]. Brain : a journal of neurology, 2019, 142(4): 992-1008.

[25] ZHANG Y, M., FU W, HU J, et al. The effect of transcranial magnetic stimulation on unilateral spatial neglect and motor function rehabilitation in stroke patients [J]. Chinese journal of cerebrovascular diseases, 2013, 10(02): 74-8.

[26] HUMMEL F C, COHEN L G. Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? [J]. The Lancet Neurology, 2006, 5(8): 708-12.

[27] TING D S, POLLOCK A, DUTTON G N, et al. Visual neglect following stroke: current concepts and future focus [J]. Survey of ophthalmology, 2011, 56(2): 114-34.

[28] KIM Y H, MIN S J, KO M H, et al. Facilitating visuospatial attention for the contralateral hemifield by repetitive TMS on the posterior parietal cortex [J]. Neurosci Lett, 2005, 382(3): 280-5.

[29] SALAZAR A P S, VAZ P G, MARCHESE R R, et al. Noninvasive Brain Stimulation Improves Hemispatial Neglect After Stroke: A Systematic Review and Meta-Analysis [J]. Archives of physical medicine and rehabilitation, 2018, 99(2): 355-66.e1.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260115

Chemical Biology Approaches in Drug Target Discovery: From Affinity Probes to Thermal Stability Analysis

Xinning Kang*

College of Chemistry and Life Sciences, Beijing University of Technology, Beijing, China

**Corresponding author: Xinning Kang*

Abstract

Drug target discovery is a critical step in new drug research and development, and chemical biology techniques and theories provide powerful tools to advance this process. This review summarizes two major categories of drug target discovery methods—affinity probes and thermal stability analysis—and their recent research progress, highlighting the role of chemical biology approaches. It also aims to deepen the author's understanding of future directions for advanced study in chemical biology. The review focuses on activity-based protein profiling (ABPP) and affinity-based protein profiling (AfBPP) as representative affinity probe technologies, as well as label-free thermal stability techniques including DARTS, CETSA, and TPP, with extensions to related developments. Both categories of drug target discovery technologies have distinct advantages, and their combined use enables a complete workflow from theory to practice and from discovery to validation. Chemical biology methods play a positive role in drug target identification.

Keywords

chemical biology, drug target discovery, affinity probe, thermal stability analysis, ABPP, AfBPP, DARTS, CETSA, TPP

1. Introduction

Drug target discovery aims to elucidate the mechanisms by which drugs exert their effects, evaluate potential side effects, predict drug resistance, and constitutes a key component in the pipeline of novel drug research and development. Clearly identifying the direct protein targets of drug molecules not only helps define the drug's mechanism of action but also provides precise scientific guidance for clinical application and drug optimization. However, traditional target discovery approaches are largely hypothesis-driven and validated one by one, resulting in low efficiency and difficulty in addressing the diverse drug–protein interactions in complex biological systems. Statistics show that among the human proteome, approximately 4,000 proteins are disease-related, yet only about 500 are targeted by approved drugs. Many disease-associated proteins remain "undruggable" due to the lack of well-defined active sites, low ligandability, etc. These "undruggable targets" represent a critical bottleneck that traditional drug discovery urgently needs to overcome. Therefore, developing efficient and precise target discovery technology platforms has become a key pathway to break through multiple obstacles in drug development.

Chemical biology, as an interdisciplinary field, uses chemical tools to dissect biological processes and brings new methodological innovations to drug target discovery. Current technical systems in this field mainly advance along two routes: one is affinity/activity-based probe technologies, exemplified by activity-based protein profiling (ABPP), which offers high specificity and quantifiability but requires chemical modification of the drug; the other is label-free thermal stability-based techniques, represented by drug affinity responsive target stability (DARTS) and thermal proteome profiling (TPP), which require no drug modification and are particularly suitable for natural products that are difficult to chemically derivatize. In recent years, these two categories have increasingly been used in combination to build integrated research workflows. Against this background, this review systematically summarizes the principles, development progress, and application examples of these two major classes of chemical biology techniques in drug target discovery under the theme "From Affinity Probes to Thermal Stability Analysis." It compares their advantages and limitations, and looks forward to the potential value of emerging technologies—such as peptide-centric local stability assay (PELSA) and artificial intelligence-assisted prediction—in overcoming undruggable targets, with the hope of providing methodological references for researchers in chemical biology and drug discovery.

2. Target Discovery Technologies: Affinity Probes

2.1 ABPP Technology: Principles and Applications

2.1.1 Technical Principles

ABPP, or activity-based protein profiling [1,2], is a core method in chemical proteomics that uses small-molecule probes to profile active proteins. It excels at identifying direct targets of small-molecule drugs or toxicants. By precisely localizing these direct targets, researchers can gain deeper and more comprehensive insights into the pharmacological and toxicological mechanisms of small molecules. As a key chemical proteomics tool, ABPP focuses on direct interactions between small molecules and proteins. Its core mechanism involves specially designed small-molecule probes for targeted labeling, enrichment, and identification. These probes consist of a bioactive functional group, a linker connecting different modules, and a reporter tag that generates a detectable signal, enabling specific capture of proteins that directly interact with the drug molecule. Compared with traditional omics approaches that reflect macroscopic changes at the gene transcription or protein expression level, ABPP more sensitively and directly reveals the initial molecular targets of drug action, providing crucial clues for understanding drug mechanisms. To avoid potential interference of reporter groups with compound bioactivity, ABPP often incorporates bioorthogonal handles (e.g., alkyne or azide groups), facilitating efficient identification of covalently binding molecular entities.

2.1.2 Applications

Two representative examples illustrate the application of ABPP:

First, in studies of off-target effects of the anticancer drug mitocan family [1], previous understanding held that mitocans primarily act on DNA. However, ABPP revealed that they also directly bind to aldehyde dehydrogenase (ALDH). This discovery uncovered a novel mechanism of action and provided a theoretical basis for subsequent structural modifications to reduce toxicity.

Second, in off-target effect research on the marketed anti-obesity drug orlistat [1], orlistat inhibits lipase to reduce calorie intake. Researchers used ABPP to systematically "fish" for potential protein targets of orlistat. Three alkyne-bearing probes were designed and incubated with live cells, followed by enrichment and identification, yielding 8 proteins that interact with the drug. In addition to the known fatty acid synthase (FAS), newly identified targets included ribosomal proteins L7a, 14, and S9, which are highly expressed in many cancer cells. These findings provide important clues for further exploring orlistat's antitumor effects and off-target mechanisms.

2.2 AfBPP Technology: Principles

AfBPP (affinity-based protein profiling) shares similarities and differences with ABPP [1]. Key distinctions and advantages/disadvantages include:

(1) Probe preparation: ABPP requires only the attachment of bioorthogonal groups (e.g., alkyne or azide) to the target molecule; AfBPP additionally requires incorporation of photoaffinity labeling groups such as benzophenone, aryl azide, diazirine, or aryl tetrazole. Among these, diazirine is currently the most widely used due to its small size and high photocrosslinking efficiency.

(2) Protein capture workflow: In ABPP, after incubation with cells or tissue lysates, direct enrichment and purification of labeled proteins can proceed; in AfBPP, UV irradiation is required after incubation to trigger covalent crosslinking between the photoaffinity group and nearby proteins before subsequent enrichment.

(3) Target identification: ABPP captures only proteins that form covalent bonds with the probe; AfBPP can identify both covalently and non-covalently interacting proteins (via hydrogen bonds, van der Waals forces, etc.), offering broader coverage and a more comprehensive reflection of the compound's targets. However, this comprehensiveness also increases the risk of nonspecific background interference, complicating result interpretation.

2.3 Emerging Developments in Probe Technologies

Beyond ABPP and AfBPP, photoaffinity labeling represents another frontier in probe-based methods [3]. Photoaffinity probes, as core tools in chemical proteomics, rely on exquisitely designed molecular architectures to capture transient protein interactions in cells. These probes typically comprise four functional modules: (1) a target recognition module that specifically binds to the active or affinity site of the target protein, ensuring selectivity; (2) a photoreactive trigger unit (e.g., diazirine, benzophenone, aryl azide) that, upon specific light excitation, generates highly reactive intermediates and forms covalent bonds with nearby amino acid residues via radical-mediated mechanisms (see Table 1 for detailed comparison of photoreactive groups); (3) a linker module that modulates overall probe conformation and flexibility (e.g., simple alkane chains, PEG, or peptide chains) to optimize binding efficiency and labeling yield—for instance, Winter's team screened 407 candidates from the Enamine database and selected a six-carbon alkane linker after considering steric hindrance and hydrophobicity to effectively capture diverse small-molecule fragment–protein interactions; and (4) a reporter tag module, usually containing a bioorthogonal handle that enables highly specific enrichment and sensitive detection of labeled products via click chemistry. Using carefully designed photoaffinity probes in cell lysates or live cells, specific wavelengths trigger the photoreactive group, covalently anchoring the probe to interacting target proteins. The reporter tag then undergoes efficient click chemistry with biotin-azide, and streptavidin-coated magnetic beads exploit the high-affinity biotin–streptavidin interaction for specific capture and enrichment of labeled proteins. Enriched products are identified by mass spectrometry, enabling real-time tracking and precise identification of dynamic cellular molecular interactions. This approach provides innovative solutions and new perspectives to overcome limitations of traditional methods.

3. Target Discovery Technologies: Label-Free Thermal Stability Analysis

3.1 DARTS Technology and Applications

3.1.1 Technical Principles

Drug affinity responsive target stability (DARTS) is widely used for identifying and screening drug targets [2,4]. Its core principle is as follows: the target drug (or equivalent volume of vehicle control) is incubated with protein samples, followed by limited proteolysis with a protease. After digestion, samples are separated by SDS-PAGE. In the gel, proteins specifically bound and protected by the small-molecule drug are more resistant to protease degradation, resulting in stronger band intensity compared with the vehicle control. DARTS can be applied to purified proteins or whole-cell lysates. After SDS-PAGE separation, depending on research goals, Western blotting or mass spectrometry can be used for qualitative and quantitative detection of differential protein expression between drug-treated and control groups. A major advantage is that no drug modification is required, making it particularly suitable for natural products.

3.1.2 Applications

One example is the identification of tanshinol sodium target proteins using DARTS [5]. Tanshinol sodium may exert cardiovascular protective effects by directly binding to the target protein CD44. This study

provides a scientific basis for deeper mechanistic investigation and new drug development. In this work, DARTS acted as a "label-free fishing rod," successfully identifying direct drug targets from complex cell lysates without altering the natural structure of the compound, followed by confirmation through subsequent experiments.

Another example is the discovery of the brain protein target AK1 for ginsenoside AK1 using DARTS, with validation by BLI [6]. DARTS served as a label-free target screening tool, successfully "fishing" potential targets of ginsenoside from the complex brain tissue proteome, followed by BLI and molecular docking to complete the full target discovery workflow from "screening" to "validation," highlighting DARTS's practical value in studying mechanisms of complex traditional Chinese medicine components.

3.2 CETSA and TPP Technologies and Applications

3.2.1 Technical Principles

Cellular thermal shift assay (CETSA) [4,7] is based on the principle that specific binding of a small-molecule drug to its target protein induces conformational changes that enhance the protein's thermodynamic stability, making it more resistant to thermal denaturation. In experiments, intact cells or cell lysates are exposed to a temperature gradient. Unbound proteins, with lower structural stability, readily denature and form insoluble aggregates upon heating, precipitating out; in contrast, drug-bound target proteins retain solubility due to higher thermal stability. A major advantage is the ability to monitor drug–target interactions in live cells or near-physiological conditions in real time and directly measure binding affinity. However, CETSA relies heavily on the availability of specific antibodies, limiting its use to validation of known targets or a small number of candidates and making high-throughput screening of unknown proteins difficult.

Thermal proteome profiling (TPP) [4,7] shares the same core mechanism as CETSA—enhanced thermal stability upon drug–target binding—but represents a major breakthrough by expanding from validation of known targets to discovery of novel targets. By performing unbiased, system-wide analysis of the entire proteome's thermal stability, TPP not only precisely identifies direct drug targets but also reveals downstream regulatory proteins or complexes affected by target functional changes.

Both techniques rely on the principle that drug binding enhances target protein thermal stability. CETSA, as a foundational method, focuses on antibody-based precise validation of specific known proteins. TPP, by replacing antibodies with mass spectrometry, extends the application from a few targeted proteins to comprehensive proteome-wide coverage, achieving unbiased discovery of novel targets in a high-throughput manner.

3.2.2 Applications

In a study identifying TH588 target proteins [7], CETSA was first used to optimize experimental conditions, followed by TPP for proteome-wide scanning, successfully mapping the action profile of TH588 and demonstrating the powerful complementary nature of these two techniques in discovering novel drug targets and elucidating complex mechanisms without modification.

In research on the therapeutic mechanism of Erigeron breviscapus extract for encephalopathy [8], TPP was successfully applied to a multi-component traditional Chinese medicine system. Without isolating individual compounds, TPP captured overall action targets, highlighting its unique advantage in natural product studies. The study went beyond target identification to integrate "target–pathway–downstream molecule–disease" layered analysis, linking TPP-discovered targets to clinical diseases. This scientifically explained traditional efficacy (e.g., treatment of Alzheimer's and Parkinson's) and prospectively predicted new indications (anti-glioma and neuroregeneration).

3.3 Brief Introduction to Other Label-Free Techniques

3.3.1 SPROX

Stability of proteins from rates of oxidation (SPROX) [3,4] is based on the principle that drug binding stabilizes target protein structure, increasing resistance to chemical denaturants (e.g., urea or guanidine hydrochloride). This stability change is reflected in methionine (Met) residues: folded-state Met is resistant

to hydrogen peroxide oxidation, while denatured-state Met is readily oxidized. By detecting the extent of Met oxidation, one can infer whether a protein binds the drug.

3.3.2 SIP

Solvent-induced protein precipitation (SIP) [4] relies on the principle that drug binding increases target protein stability, conferring greater resistance to organic solvent-induced denaturation and precipitation (e.g., acetone–ethanol–acetic acid mixtures). Cell lysates are treated with varying solvent concentrations, and soluble proteins in the supernatant are quantified by mass spectrometry. Shifts in precipitation curves between drug-treated and control groups identify target proteins. Advantages include no chemical modification, simple operation, use of common reagents, and complementarity with CETSA and SPROX. Limitations include applicable only to cell lysates, not live cells.

3.3.3 LiP

Limited proteolysis (LiP) [4] is based on the principle that drug binding stabilizes target protein conformation, protecting certain flexible regions from short-term hydrolysis by nonspecific proteases (e.g., proteinase K). After initial limited digestion, samples are fully digested with trypsin, and differential peptides are quantified by mass spectrometry to infer drug-binding proteins and potential binding regions. Advantages: provides domain-level structural information, clearly delineating binding regions; good applicability to complex proteomes. Limitations: requires exogenous protease, limited to lysates, multiple steps, and complex sample composition.

4. Technical Comparison and Integration Strategies

4.1 Comparison of Advantages and Limitations

The two major categories—affinity probe and thermal stability analysis—each have strengths and weaknesses, as summarized in Table 1.

Table 1: Comparison of the Two Categories of Technologies

Technology Type	Representative Techniques	Advantages	Limitations	LimitationsApplicable Scenarios
Affinity Probes	ABPP, AfBPP	High specificity, quantitative	Requires chemical modification, complex probe design	Known active sites, covalent drugs
Thermal Stability	DARTS, TPP	No modification needed, proteome-wide screening	Insensitive to weak binding, limited for membrane proteins	Natural products, difficult-to-modify drugs

4.2 Combined Applications

The study identifying methylenetetrahydrofolate dehydrogenase 1-like protein (MTHFD1L) as the binding target of natural product pseudolaric acid A (PAA) [9] exemplifies the integrated use of affinity probe and thermal stability technologies:

Phase 1: Preliminary screening with DARTS

PAA was incubated with HeLa cell lysates, followed by limited proteolysis. Drug-bound proteins are protected from protease degradation. MS identified 42 potential binding proteins (overlapping from two replicates).

Phase 2: Cross-validation with ABPP

Alkyne-bearing PAA activity probes were used for ABPP, with a competition group pre-incubated with excess free PAA. Probe-labeled proteins are specifically enriched if free PAA competitively inhibits labeling. MS identified 18 potential binding proteins (overlapping from two replicates).

Phase 3: Intersection to lock in the most likely target

Intersection of 42 DARTS proteins and 18 ABPP proteins yielded only one overlapping protein: MTHFD1L. Conclusion: MTHFD1L is the most likely target of PAA.

Phase 4: In vitro biophysical validation

NMR saturation transfer difference (STD) confirmed direct binding between PAA and MTHFD1L, identifying key binding groups (e.g., C4-acetoxy) consistent with structure–activity relationship studies. Surface plasmon resonance (SPR) measured binding constants, further confirming direct interaction.

Phase 5: Functional validation

Gene silencing of MTHFD1L in HeLa cells significantly reduced cell viability, phenocopying PAA treatment, confirming MTHFD1L mediates PAA's antitumor activity. Mechanistic studies showed PAA treatment induces reactive oxygen species (ROS) accumulation, with transcriptomic analysis revealing related pathway changes.

5. Challenges and Prospects

Although chemical biology-based target discovery has advanced significantly and multi-technology integration is maturing, methodological systems still face challenges, primarily stemming from technical limitations such as insufficient sensitivity for low-abundance targets and poor applicability to insoluble systems like membrane proteins.

Emerging technologies offer hope for overcoming these hurdles:

First, upgraded applications of peptide-centric local stability assay (PELSA) [10]; second, in situ live-cell applications of photoaffinity labeling [2]; third, deep integration of artificial intelligence with chemical proteomics.

In the future, chemical biology target discovery is expected to trend toward multidimensional integration, combining precise chemical probe design, biophysical in situ detection, high-sensitivity mass spectrometry, and AI-driven high-throughput prediction. This will build a comprehensive technical framework from in vitro screening to live-cell validation, and from single-target analysis to full interaction network mapping. With breakthroughs in these emerging directions, many previously "undruggable" targets are likely to gain new therapeutic opportunities, and chemical biology will play an increasingly pivotal role in the era-defining challenge of precise drug discovery.

6. Conclusion

Drug target discovery technologies are advancing toward multidimensional integration. Affinity probe-based methods (e.g., ABPP) enable specific enrichment and quantification of target proteins, while label-free thermal stability techniques (e.g., DARTS, TPP) allow unbiased proteome-wide screening. These two categories are complementary in principle. Single technologies struggle to balance sensitivity and accuracy, making multi-technology integration the mainstream strategy. A typical workflow involves preliminary screening with label-free methods, cross-validation with probe-based techniques, in vitro biophysical confirmation, and cellular functional validation—forming a closed "discovery–validation" loop that effectively reduces false positives. The target identification of pseudolaric acid A exemplifies successful integration of DARTS and ABPP.

Current technologies still face challenges such as insufficient sensitivity for low-abundance targets, limited applicability to membrane proteins, and difficulty capturing transient/weak interactions. Emerging directions including PELSA, live-cell photoaffinity labeling, and AI-assisted prediction continue to develop and are likely to enable chemical biology to make greater contributions to overcoming undruggable targets.

References

[1] Chen, J. S., Zhang, X. Y., Liu, J. H., Meng, W. Q., & Zhang, L. (2023). Progress in the application of activity-based and affinity-based protein expression profiling in pharmacological and toxicological research. *Chinese Journal of Pharmacology and Toxicology*, 37(12), 951-958.

[2] Zhang, M. T., & Ding, M. (2023). Application of chemical proteomics technology in drug target identification. *Life Sciences*, 35(06), 816-823. https://doi.org/10.13376/j.cbls/20230006

[3] Li, Y. Z., Wan, N., & Ye, H. (2025). Breakthroughs in target discovery and application of photoaffinity probes in chemical proteomics. *Chinese Journal of Modern Applied Pharmacy*, 42(23), 4184-4190. https://doi.org/10.13748/j.cnki.issn1007-7693.20250955

[4] Lü, J. W., & Ye, M. L. (2021). Research progress on label-free drug target identification proteomics methods. *Journal of Mass Spectrometry*, 42(05), 845-861.

[5] Chen, Y. J., Jiang, W. W., Li, L. H., Zhang, C. Z., Zhang, C., & Jia, D. (2022). Identification of direct binding target proteins of sodium danshensu based on DARTS technology. *Pharmaceutical and Clinical Research*, 30(01), 58-61+65. https://doi.org/10.13664/j.cnki.pcr.2022.01.007

[6] Chen, F. Y., Qin, W., Li, R. M., Cheng, Y., Zhu, Z., Chen, L., & Zhao, Y. N. (2022). Discovery and confirmation of the brain protein action site AK1 of ginsenoside based on DARTS technology. *China Journal of Chinese Materia Medica*, 47(05), 1336-1342. https://doi.org/10.19540/j.cnki.cjcmm.20211117.703

[7] Liu, Z. Y., Zhu, S. B., Deng, H. T., & Chen, Y. L. (2021). Identification of target proteins of TH588 by thermal proteomics analysis. *Journal of Mass Spectrometry*, 42(05), 951-961.

[8] Rao, Q. L., Xu, Y. T., Chen, H., Huang, Z. X., Zhang, G., Lü, J. R., ... & Deng, X. Q. (2025). Study on the mechanism of action and potential new effects of Erigeron breviscapus extract in treating encephalopathy based on thermal proteomics technology. *Modern Chinese Medicine*, 27(03), 460-471. https://doi.org/10.13313/j.issn.1673-4890.20241111001

[9] Dong, H., Yang, X., Wang, P., et al. (2025). Identification and verification of methylenetetrahydrofolate dehydrogenase 1-like protein as the binding target of natural product pseudolaric acid A. *Natural Products and Bioprospecting*, 15, 21. https://doi.org/10.1007/s13659-025-00502-1

[10] Tang, Z. H., & Liu, X. Y. (2025). A new method for studying small molecule-protein interactions—PELSA. *Journal of Mass Spectrometry*, 46(01), 1-2.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260116

Parkinson's Disease: Calcium Dysregulation as a Multilevel Mechanism of Selective Neuronal Vulnerability, with Implications for Therapeutic Strategy

Weijie Lai*

ZJU-UoE Institute, Zhejiang University, Haining, China

**Corresponding author: Weijie Lai*

Abstract

Parkinson's disease (PD) is a prevalent neurodegenerative disorder characterised by the selective degeneration of dopaminergic neurons in the substantia nigra pars compacta. Despite extensive investigation into genetic and environmental risk factors, the mechanisms underlying this selective neuronal vulnerability remain incompletely understood. Increasing evidence suggests that calcium dysregulation represents a unifying pathogenic mechanism linking molecular, cellular, and systems-level processes in PD. This review synthesises evidence demonstrating how the intrinsic reliance of substantia nigra dopaminergic neurons on calcium-dependent autonomous pacemaking imposes a sustained metabolic burden that predisposes these neurons to degeneration. It examines how genetically diverse PD-associated mutations, including those affecting PINK1/Parkin, LRRK2, and GBA, converge on disrupted calcium homeostasis through distinct cellular pathways involving mitochondrial buffering, lysosomal calcium signalling, and organelle quality control. At the cellular level, chronic calcium dysregulation drives oxidative stress, α-synuclein aggregation, and impaired autophagy, forming a self-reinforcing cycle of neuronal dysfunction. These processes extend beyond individual neurons, promoting neuroinflammation and basal ganglia circuit reorganisation that ultimately underlie motor symptom expression. The review further discusses the therapeutic implications of this multilevel framework, highlighting why broad calcium channel blockade has shown limited clinical efficacy. It argues that future disease-modifying strategies will require greater molecular and temporal precision, as well as integration with systems-level neuromodulation approaches that reduce activity-dependent calcium and metabolic burden. Overall, calcium dysregulation emerges as a central integrator of vulnerability across biological scales in Parkinson's disease, offering a coherent framework for understanding pathogenesis and guiding therapeutic innovation.

Keywords

Parkinson's disease, calcium dysregulation, selective neuronal vulnerability, dopaminergic neurons, α-synuclein pathology

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterised by both motor dysfunction and selective neuronal loss. Clinically, it presents with bradykinesia, rigidity, resting tremor, and postural instability, while pathologically it is defined by degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNc) [1]. As one of the most prevalent neurodegenerative diseases worldwide, PD represents a growing biomedical and societal challenge in ageing populations [2]. Despite extensive research into genetic, molecular, and environmental contributors, the mechanisms underlying the selective vulnerability of SNc dopaminergic neurons relative to neighbouring neuronal populations remain poorly understood [3-5].

Recent advances suggest that this selective vulnerability cannot be explained by a single genetic mutation or toxic insult [3, 5]. Instead, accumulating evidence points towards calcium dysregulation as a unifying mechanistic framework linking genetic susceptibility, cellular stress, and systems-level dysfunction in PD [6-8]. SNc dopaminergic neurons exhibit a distinctive form of autonomous pacemaking that is tightly coupled to sustained calcium entry via L-type calcium channels, particularly those containing the Cav1.3 subunit [9, 10]. Although physiologically essential for tonic dopamine release, this calcium-dependent firing strategy places these neurons under chronic metabolic stress, increasing their vulnerability to mitochondrial dysfunction, oxidative stress, and protein aggregation [3, 11]. Importantly, this calcium burden coincides with the intrinsic metabolic liability of dopamine itself: its cytosolic metabolism and oxidation generate reactive oxygen species and quinone intermediates, thereby imposing an additional oxidative load [12, 13]. The convergence of sustained calcium influx and dopamine-related oxidative stress therefore creates a uniquely hostile intracellular environment in SNc neurons, helping to explain their disproportionate susceptibility to degeneration [11, 14, 15].

Under physiological conditions, intracellular calcium levels in dopaminergic neurons are tightly regulated through coordinated buffering, active transport, and sequestration by intracellular organelles, particularly the endoplasmic reticulum and mitochondria [16, 17]. Activity-dependent calcium entry couples neuronal firing to neurotransmitter release and mitochondrial ATP production, aligning metabolic output with energetic demand [10]. However, the reliance of SNc dopaminergic neurons on sustained calcium influx renders this homeostatic balance especially fragile. Even modest perturbations in calcium handling, mitochondrial efficiency, or buffering capacity can shift calcium signalling from a physiological regulator to a chronic source of cellular stress, providing a critical baseline for understanding pathological calcium dysregulation in PD [11, 17].

Consistent with this framework, genetic, iPSC-based, and imaging studies have demonstrated how disruptions in calcium homeostasis interact with PD-associated pathways, including those involving PINK1, Parkin, LRRK2, and α-synuclein, to drive neurodegeneration [7, 18, 19]. These findings indicate that calcium dysregulation does not act as an isolated molecular defect, but rather as a central integrator that amplifies diverse genetic and cellular stressors.

Building on this perspective, this review argues that calcium dysregulation operates across genetic, cellular, and systems levels to shape Parkinson's disease pathogenesis, providing a coherent framework for selective neuronal vulnerability. Unlike previous reviews that have primarily focused on intracellular calcium channels or single-cell calcium signalling, this article adopts a multiscale approach to examine how calcium-dependent stress propagates from subcellular organelles to neural circuits and ultimately contributes to motor dysfunction. It further examines how insights into calcium-dependent mechanisms have both informed and constrained therapeutic development, highlighting the limitations of calcium-targeted pharmacological interventions. Finally, the review considers how, at later disease stages marked by substantial neuronal loss, therapeutic focus increasingly shifts toward systems-level neuromodulation aimed at reducing activity-dependent calcium and metabolic burden rather than correcting underlying cellular pathology.

2. Calcium Dysregulation as a Mechanism of Selective Neuronal Vulnerability

Increasing evidence suggests that calcium dysregulation represents a critical mechanistic convergence point through which diverse genetic and cellular risk factors drive selective neuronal vulnerability in Parkinson's disease [3, 6, 7]. Rather than acting as isolated causes, mutations associated with both familial and sporadic PD appear to compromise dopaminergic neurons' capacity to maintain calcium homeostasis, thereby amplifying metabolic stress and lowering the threshold for neurodegeneration.

2.1 Genetic risk factors converging on calcium homeostasis

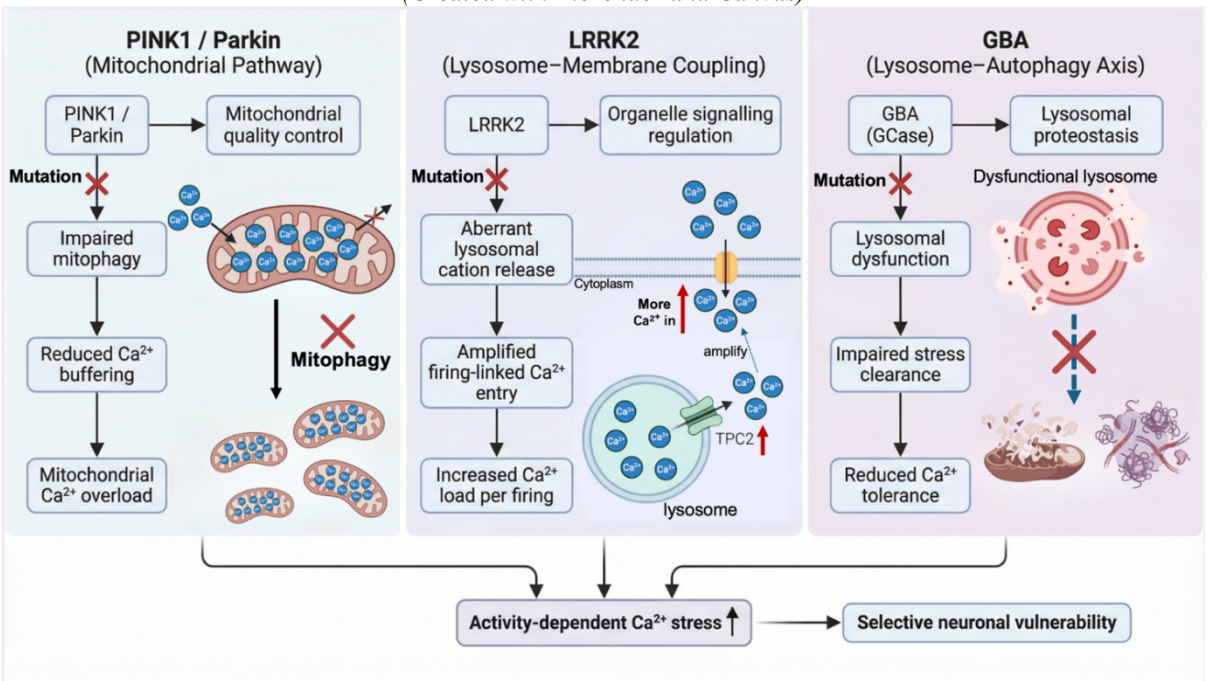
To illustrate how genetically diverse forms of Parkinson's disease converge on calcium dysregulation, this section focuses on three representative genetic risk factors: PINK1/Parkin, LRRK2, and GBA. Pathogenic mutations in these genes can disrupt calcium homeostasis at distinct cellular control points, including mitochondrial calcium buffering, lysosomal calcium signalling, and organelle quality control (Figure 1), thereby linking genetic heterogeneity to selective neuronal vulnerability [20-22].

Mutations in PINK1 and Parkin, key regulators of mitochondrial quality control, provide direct genetic evidence that impaired mitochondrial calcium handling can drive neurodegeneration [3]. Under physiological conditions, mitochondria transiently sequester cytosolic calcium during neuronal activity to support ATP production while limiting toxic calcium accumulation [17]. Calcium imaging studies demonstrate that PINK1-deficient neurons exhibit defective mitochondrial calcium efflux, leading to sustained intramitochondrial calcium accumulation and reduced buffering capacity (Figure 1) [18]. Notably, neuronal death can be triggered by physiological calcium loads rather than excessive calcium entry, reflecting a reduced tolerance to activity-dependent calcium stress. Parkin dysfunction further exacerbates this vulnerability by impairing mitophagy, allowing calcium-intolerant mitochondria to persist [21].

In contrast, LRRK2 mutations disrupt calcium homeostasis primarily through lysosome-associated mechanisms [23]. A recent study combining human neuronal models, in vivo Drosophila assays, and calcium imaging demonstrates that pathogenic LRRK2 exaggerates depolarisation-induced calcium entry via aberrant activation of the lysosomal cation channel TPC2 [22]. This maladaptive coupling between lysosomal cation release and plasma membrane calcium influx increases calcium load during neuronal firing, thereby lowering the threshold for cellular stress (Figure 1) [23]. Importantly, pharmacological modulation of TPC2 ion selectivity was sufficient to normalise calcium entry and rescue behavioural deficits, highlighting dysregulated calcium signalling dynamics rather than acute cytotoxicity as the key pathogenic mechanism [24].

GBA mutations, which impair the lysosomal enzyme glucocerebrosidase and represent the most common genetic risk factor for Parkinson's disease, further reinforce this convergence on calcium stress [20]. Although GBA does not directly regulate calcium flux, loss of glucocerebrosidase activity disrupts lysosomal function and TFEB-mediated autophagy, secondarily reducing the cell's capacity to cope with sustained calcium stress (Figure 1) [25-27]. In dopaminergic neurons operating near the limits of calcium tolerance, such secondary destabilisation may further lower the threshold for calcium-induced toxicity.

Figure 1. Genetically distinct Parkinson's disease risk factors converge on activity-dependent calcium stress. (Created with Biorender and Canvas)

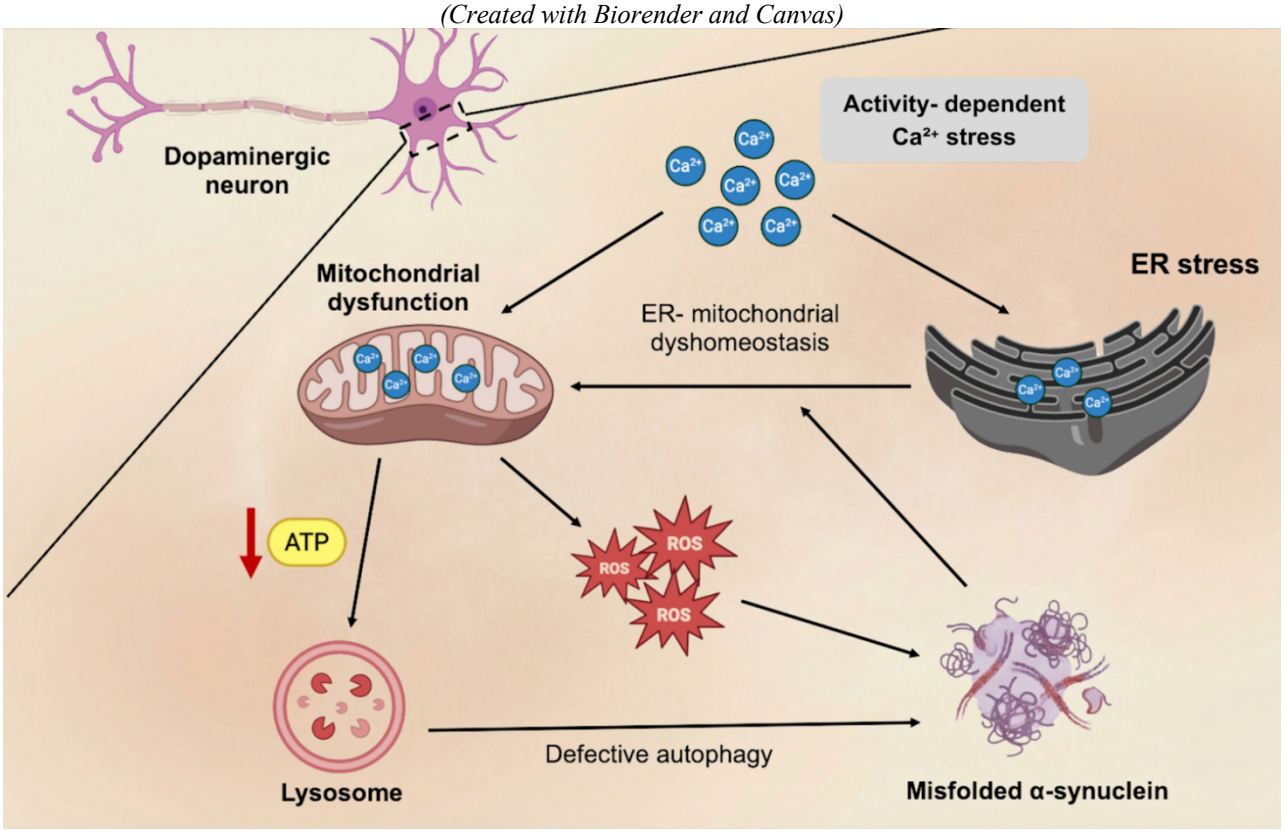


2.2 Cellular Pathways Linking Calcium Dysregulation to Neurodegeneration

At the cellular level, sustained calcium dysregulation initiates a cascade of pathological processes that amplify neuronal vulnerability [28]. The endoplasmic reticulum (ER), the major intracellular calcium store, normally ensures precise calcium release for physiological signalling. Under sustained calcium dysregulation, chronic calcium overload drives excessive ER-to-mitochondria calcium transfer, leading to mitochondrial calcium accumulation and metabolic overstimulation [29]. This metabolic strain increases the production of reactive oxygen species (ROS), which progressively compromises mitochondrial function and reduces the capacity of mitochondria to buffer subsequent calcium influx [28].

Apart from directly damaging mitochondria, excessive ROS production promotes the misfolding and aggregation of α-synuclein, a presynaptic protein involved in synaptic vesicle dynamics and a central pathological hallmark of Parkinson's disease [28]. Aggregated α-synuclein, in turn, interferes with the recovery of mitochondrial calcium dynamics following calcium release from the endoplasmic reticulum, prolonging mitochondrial calcium burden and thereby increasing oxidative stress and mitochondrial dysfunction [30]. In parallel, misfolded and aggregated α-synuclein is inefficiently degraded by lysosomes, placing a substantial burden on the autophagic system and contributing to defective autophagic flux [31]. Because lysosomal function is energetically demanding and depends on ATP supplied by mitochondria, mitochondrial dysfunction further impairs lysosomal homeostasis [28]. Together, these processes establish a self-reinforcing vicious cycle linking calcium stress, ER stress, mitochondrial dysfunction, lysosomal impairment, and α-synuclein pathology, ultimately rendering dopaminergic neurons functionally compromised and progressively vulnerable to degeneration (Figure 2).

Figure 2. Cellular network linking activity-dependent calcium stress to dopaminergic neuronal vulnerability. (Created with Biorender and Canvas)



Although substantial progress has been made in describing these pathways, the precise causal relationships among calcium dysregulation, mitochondrial failure, α-synuclein aggregation, and lysosomal dysfunction remain incompletely resolved [32]. Whether calcium overload acts as an initiating trigger or a downstream amplifier likely depends on disease stage, genetic background, and neuronal context, underscoring that selective neuronal vulnerability in Parkinson's disease emerges from the dynamic convergence of multiple calcium-dependent stressors.

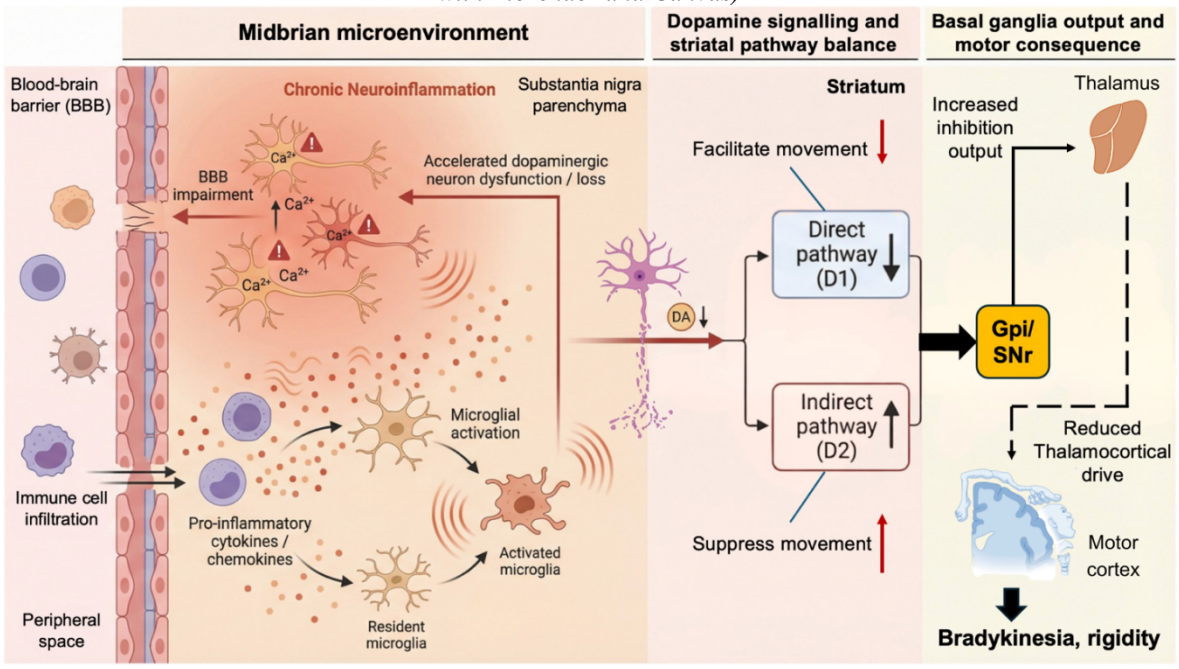
3. From Cellular Pathology to Network Dysfunction: A Systems-Level Perspective

Although calcium-driven stress originates within individual substantia nigra dopaminergic neurons, it propagates across tissue and circuit levels (Figure 3). Excessive calcium signalling also contributes to the disruption of the blood–brain barrier (BBB) [33]. BBB impairment permits increased infiltration of peripheral immune cells and facilitates sustained activation of resident microglia within the midbrain [34, 35]. Activated microglia and infiltrating immune cells release pro-inflammatory cytokines and chemokines, thereby promoting a chronic neuroinflammatory environment that accelerates dopaminergic neuronal dysfunction and degeneration [33].

As dopaminergic neurons progressively lose function or degenerate, dopamine levels in their projection targets, such as the striatum, decrease. This dopamine deficit has direct consequences for basal ganglia circuit organisation. Dopamine normally maintains a balance between the direct pathway, which facilitates movement, and the indirect pathway, which suppresses movement [36]. Dopamine depletion weakens direct pathway activity while relatively enhancing indirect pathway influence [37]. The net effect is excessive inhibitory output from basal ganglia output nuclei to thalamocortical motor circuits, leading to reduced cortical motor drive and the emergence of cardinal motor features such as bradykinesia and rigidity [38].

Evidence for this multilevel transition from cellular pathology to circuit dysfunction comes from complementary experimental approaches. In animal models, genetic or toxin-induced disruption of dopaminergic neurons reproduces both dopamine loss and characteristic motor deficits [37]. Electrophysiological studies demonstrate widespread reorganisation of firing patterns across basal ganglia circuits following dopamine depletion, indicating network-level adaptation rather than isolated neuronal failure [39, 40]. In humans, PET imaging consistently links reduced striatal dopaminergic signalling with motor symptom severity [41]. Together, these findings illustrate how calcium-driven cellular stress is amplified through neuroinflammatory and circuit-level mechanisms, ultimately transforming local neuronal vulnerability into systems-level dysfunction in Parkinson's disease [37].

Figure 3. System-level amplification of calcium-driven dopaminergic vulnerability in Parkinson's disease. (Created with Biorender and Canvas)



4. Therapeutic Implications: Calcium-Targeted Interventions and Neuromodulation

Given growing evidence that calcium dysregulation contributes to selective dopaminergic vulnerability in Parkinson's disease, targeting calcium signalling has been considered a logical therapeutic approach [42]. Cav1.3-containing L-type calcium channels were therefore targeted as a major source of calcium entry in substantia nigra dopaminergic neurons. Epidemiological and preclinical studies suggested that blocking these

channels could reduce cumulative calcium stress, leading to clinical testing of the calcium channel blocker isradipine [43, 44]

Despite encouraging epidemiological and preclinical data, the phase III STEADY-PD trial failed to demonstrate disease-modifying efficacy [45]. Rather than refuting the role of calcium dysregulation, this outcome can be hypothesised to reflect key limitations of current pharmacological strategies. First, isradipine lacks cell-type and subcellular specificity, resulting in broad suppression of calcium influx across neuronal populations [46]. Given the indispensable role of calcium signalling in neuronal metabolism, synaptic transmission, and gene regulation, such global inhibition likely disrupts physiological function and severely constrains the therapeutic window. Second, although intervention occurred at an early clinical stage [42], key pathological processes, such as dopaminergic axonal loss and organellar dysfunction, may already have been established, limiting the capacity of calcium modulation to meaningfully alter disease progression.

These clinical and mechanistic outcomes highlight the need for therapeutic precision rather than broad calcium suppression. Future strategies aim to preserve physiological signalling while mitigating pathogenic overload through Cav1.3-selective inhibitors, mitochondria-targeted antioxidants, and enhancement of PINK1/Parkin-mediated mitochondrial quality control [19, 45]. Although still at early translational stages, these approaches represent a shift toward mechanism-informed interventions that balance calcium homeostasis without compromising neuronal function [46].

As Parkinson's disease progresses, the therapeutic impact of pharmacological calcium modulation becomes increasingly limited, as substantial dopaminergic neuronal loss and circuit reorganisation have already occurred [47]. Under these conditions, restoring function requires strategies that address established network-level dysfunction rather than solely correcting cellular calcium dysregulation [48]. Deep brain stimulation (DBS) exemplifies this systems-level approach by modulating pathological basal ganglia output and alleviating motor symptoms independently of ongoing neurodegeneration [49]. Clinically, DBS has demonstrated significant improvements in motor symptoms, with meta-analyses reporting reductions in UPDRS-III motor scores comparable to medical therapy [50]. Studies show an average improvement of 60.7% to 66% in UPDRS-motor scores after DBS implantation, sustained for several years [51]. This includes substantial decreases in tremor and rigidity, with some patients experiencing as high as 85% improvement in tremor and 66% improvement in rigidity at 5 years post-DBS [52]. However, conventional DBS relies on continuous open-loop stimulation, which can impose additional metabolic demand and activity-dependent calcium influx, potentially aggravating calcium and energy stress in vulnerable neural circuits [47, 48].

These limitations have driven the development of adaptive DBS and brain–computer interface (BCI)–based neuromodulation. By adjusting stimulation in real time and delivering it only when pathological activity emerges, adaptive systems reduce unnecessary neural activation, lower cumulative energy expenditure, and minimise stimulation-induced calcium influx [53]. Adaptive DBS (aDBS), also known as closed-loop DBS, has been shown to be equally or even more effective than conventional continuous DBS in alleviating motor symptoms such as rigidity, while significantly reducing stimulation time and energy consumption by up to 56% [54, 55]. Some reports indicate that aDBS can lead to motor score improvements that are 27% to 29% higher than open-loop DBS and provides better control of symptoms and dyskinesias, particularly improving bradykinesia without additional side effects [55, 56]. Rather than directly suppressing calcium signalling, this approach limits activity-driven calcium and metabolic burden, representing a mechanism-informed evolution of system-level intervention for Parkinson's disease [49].

5. Conclusion

Parkinson's disease is best understood not as the consequence of a single genetic defect or isolated molecular pathway, but as the outcome of multilevel processes converging on calcium dysregulation. The intrinsic reliance of substantia nigra dopaminergic neurons on sustained calcium-dependent pacemaking creates a physiological vulnerability that, when combined with genetic and cellular stressors, predisposes these neurons to degeneration.

Genetically diverse risk factors, including PINK1/Parkin, LRRK2, and GBA, disrupt calcium handling at distinct cellular sites yet converge functionally to impair mitochondrial buffering, lysosomal clearance, and organelle quality control. These disturbances reinforce one another through calcium-driven oxidative stress,

α-synuclein aggregation, and defective autophagy, forming a self-sustaining cycle of neuronal dysfunction. Importantly, calcium-associated pathology extends beyond individual neurons, promoting neuroinflammation and network-level reorganisation that ultimately manifests as basal ganglia circuit dysfunction and motor impairment.

This integrated framework helps explain the limited clinical success of calcium-targeted pharmacological interventions, as global calcium inhibition risks disrupting essential physiological signalling and is unlikely to reverse established pathology. Future disease-modifying strategies will therefore require greater molecular and temporal precision and will likely need to be integrated with systems-level approaches. In this context, adaptive neuromodulation strategies that reduce activity-dependent calcium and metabolic burden offer a promising, mechanism-informed direction for therapeutic development.

References

[1] Bloem, B. R., Okun, M. S. and Klein, C. Parkinson's disease. Lancet. 2021, 397(10291), pp. 2284-2303. https://doi.org/10.1016/s0140-6736(21)00218-x.

[2] Ben-Shlomo, Y., Darweesh, S., Llibre-Guerra, J., Marras, C., San Luciano, M. and Tanner, C. The epidemiology of Parkinson's disease. Lancet. 2024, 403(10423), pp. 283-292. https://doi.org/10.1016/s0140-6736(23)01419-8.

[3] Michel, P. P., Hirsch, E. C. and Hunot, S. Understanding Dopaminergic Cell Death Pathways in Parkinson Disease. Neuron. 2016, 90(4), pp. 675-91. https://doi.org/10.1016/j.neuron.2016.03.038.

[4] Verma, A. and Ravindranath, V. Ca(V)1.3 L-Type Calcium Channels Increase the Vulnerability of Substantia Nigra Dopaminergic Neurons in MPTP Mouse Model of Parkinson's Disease. Frontiers in Aging Neuroscience. 2019, 11, p. 382. https://doi.org/10.3389/fnagi.2019.00382.

[5] Filimontseva, A., Fu, Y., Vila, M. and Halliday, G. M. Neuromelanin and selective neuronal vulnerability to Parkinson's disease. Trends in Neurosciences. 2025, 48(6), pp. 445-459. https://doi.org/10.1016/j.tins.2025.04.005.

[6] Benkert, J., Hess, S., Roy, S., Beccano-Kelly, D., Wiederspohn, N., Duda, J., Simons, C., Patil, K., Gaifullina, A., Mannal, N., et al. Cav2.3 channels contribute to dopaminergic neuron loss in a model of Parkinson's disease. Nature Communications. 2019, 10(1), p. 5094. https://doi.org/10.1038/s41467-019-12834-x.

[7] Beccano-Kelly, D. A., Cherubini, M., Mousba, Y., Cramb, K. M. L., Giussani, S., Caiazza, M. C., Rai, P., Vingill, S., Bengoa-Vergniory, N., Ng, B., et al. Calcium dysregulation combined with mitochondrial failure and electrophysiological maturity converge in Parkinson's iPSC-dopamine neurons. iScience. 2023, 26(7), p. 107044. https://doi.org/10.1016/j.isci.2023.107044.

[8] Lin, J., Pang, D., Li, C., Ou, R., Yu, Y., Cui, Y., Huang, J. and Shang, H. Calcium channel blockers and Parkinson's disease: a systematic review and meta-analysis. Therapeutic Advances in Neurological Disorders. 2024, 17, p. 17562864241252713. https://doi.org/10.1177/17562864241252713.

[9] Guzman, J. N., Sánchez-Padilla, J., Chan, C. S. and Surmeier, D. J. Robust pacemaking in substantia nigra dopaminergic neurons. Journal of Neuroscience. 2009, 29(35), pp. 11011-9. https://doi.org/10.1523/jneurosci.2519-09.2009.

[10] Guzman, J. N., Sanchez-Padilla, J., Wokosin, D., Kondapalli, J., Ilijic, E., Schumacker, P. T. and Surmeier, D. J. Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1. Nature. 2010, 468(7324), pp. 696-700. https://doi.org/10.1038/nature09536.

[11] Surmeier, D. J., Guzman, J. N., Sanchez-Padilla, J. and Schumacker, P. T. The role of calcium and mitochondrial oxidant stress in the loss of substantia nigra pars compacta dopaminergic neurons in

Parkinson's disease. Neuroscience. 2011, 198, pp. 221-31. https://doi.org/10.1016/j.neuroscience.2011.08.045.

[12] Zucca, F. A., Segura-Aguilar, J., Ferrari, E., Muñoz, P., Paris, I., Sulzer, D., Sarna, T., Casella, L. and Zecca, L. Interactions of iron, dopamine and neuromelanin pathways in brain aging and Parkinson's disease. Progress in Neurobiology. 2017, 155, pp. 96-119. https://doi.org/10.1016/j.pneurobio.2015.09.012.

[13] Zhou, Z. D., Yi, L. X., Wang, D. Q., Lim, T. M. and Tan, E. K. Role of dopamine in the pathophysiology of Parkinson's disease. Translational Neurodegeneration. 2023, 12(1), p. 44. https://doi.org/10.1186/s40035-023-00378-6.

[14] Dryanovski, D. I., Guzman, J. N., Xie, Z., Galteri, D. J., Volpicelli-Daley, L. A., Lee, V. M., Miller, R. J., Schumacker, P. T. and Surmeier, D. J. Calcium entry and α-synuclein inclusions elevate dendritic mitochondrial oxidant stress in dopaminergic neurons. Journal of Neuroscience. 2013, 33(24), pp. 10154-64. https://doi.org/10.1523/jneurosci.5311-12.2013.

[15] Ricke, K. M., Paß, T., Kimoloi, S., Fährmann, K., Jüngst, C., Schauss, A., Baris, O. R., Aradjanski, M., Trifunovic, A., Eriksson Faelker, T. M., et al. Mitochondrial Dysfunction Combined with High Calcium Load Leads to Impaired Antioxidant Defense Underlying the Selective Loss of Nigral Dopaminergic Neurons. Journal of Neuroscience. 2020, 40(9), pp. 1975-1986. https://doi.org/10.1523/jneurosci.1345-19.2019.

[16] Verkhratsky, A. Physiology and pathophysiology of the calcium store in the endoplasmic reticulum of neurons. Physiological Reviews. 2005, 85(1), pp. 201-79. https://doi.org/10.1152/physrev.00004.2004.

[17] Matuz-Mares, D., González-Andrade, M., Araiza-Villanueva, M. G., Vilchis-Landeros, M. M. and Vázquez-Meza, H. Mitochondrial calcium: Effects of its imbalance in disease. Antioxidants (Basel). 2022, 11(5), p. 801. https://doi.org/10.3390/antiox11050801.

[18] Gandhi, S., Wood-Kaczmar, A., Yao, Z., Plun-Favreau, H., Deas, E., Klupsch, K., Downward, J., Latchman, D. S., Tabrizi, S. J., Wood, N. W., et al. PINK1-associated Parkinson's disease is caused by neuronal vulnerability to calcium-induced cell death. Molecular Cell. 2009, 33(5), pp. 627-38. https://doi.org/10.1016/j.molcel.2009.02.013.

[19] Dong-Chen, X., Yong, C., Yang, X., Chen-Yu, S. and Li-Hua, P. Signaling pathways in Parkinson's disease: molecular mechanisms and therapeutic interventions. Signal Transduction and Targeted Therapy. 2023, 8(1), p. 73. https://doi.org/10.1038/s41392-023-01353-3.

[20] Chiasserini, D., Paciotti, S., Eusebi, P., Persichetti, E., Tasegian, A., Kurzawa-Akanbi, M., Chinnery, P. F., Morris, C. M., Calabresi, P., Parnetti, L., et al. Selective loss of glucocerebrosidase activity in sporadic Parkinson's disease and dementia with Lewy bodies. Molecular Neurodegeneration. 2015, 10, p. 15. https://doi.org/10.1186/s13024-015-0010-2.

[21] Barazzuol, L., Giamogante, F., Brini, M. and Calì, T. PINK1/Parkin mediated mitophagy, Ca^{2+} signalling, and ER-mitochondria contacts in Parkinson's disease. International Journal of Molecular Sciences. 2020, 21(5), p. 1772. https://doi.org/10.3390/ijms21051772.

[22] Gregori, M., Pereira, G. J. S., Allen, R., West, N., Chau, K. Y., Cai, X., Bostock, M. P., Bolsover, S. R., Keller, M., Lee, C. Y., et al. Lysosomal TPC2 channels disrupt Ca^{2+} entry and dopaminergic function in models of LRRK2-Parkinson's disease. Journal of Cell Biology. 2025, 224(6), p. e202412055. https://doi.org/10.1083/jcb.202412055.

[23] Xiong, Y. and Yu, J. LRRK2 in Parkinson's disease: upstream regulation and therapeutic targeting. Trends in Molecular Medicine. 2024, 30(10), pp. 982-996. https://doi.org/10.1016/j.molmed.2024.07.003.

[24] Alharbi, A. F. and Parrington, J. TPC2 in drug development: Emerging target for cancer, viral infections, cardiovascular diseases, and neurological disorders. Pharmacological Research. 2025, 213, p. 107655. https://doi.org/10.1016/j.phrs.2025.107655.

[25] Medina, D. L., Di Paola, S., Peluso, I., Armani, A., De Stefani, D., Venditti, R., Montefusco, S., Scotto-Rosato, A., Prezioso, C., Forrester, A., et al. Lysosomal calcium signalling regulates autophagy through calcineurin and TFEB. Nature Cell Biology. 2015, 17(3), pp. 288-99. https://doi.org/10.1038/ncb3114.

[26] Burbulla, L. F. and Krainc, D. The role of dopamine in the pathogenesis of GBA1-linked Parkinson's disease. Neurobiology of Disease. 2019, 132, p. 104545. https://doi.org/10.1016/j.nbd.2019.104545.

[27] Giamogante, F., Barazzuol, L., Maiorca, F., Poggio, E., Esposito, A., Masato, A., Napolitano, G., Vagnoni, A., Calì, T. and Brini, M. A SPLICS reporter reveals [Formula: see text]-synuclein regulation of lysosome-mitochondria contacts which affects TFEB nuclear translocation. Nature Communications. 2024, 15(1), p. 1516. https://doi.org/10.1038/s41467-024-46007-2.

[28] Kaur, S., Sehrawat, A., Mastana, S. S., Kandimalla, R., Sharma, P. K., Bhatti, G. K. and Bhatti, J. S. Targeting calcium homeostasis and impaired inter-organelle crosstalk as a potential therapeutic approach in Parkinson's disease. Life Sciences. 2023, 330, p. 121995. https://doi.org/10.1016/j.lfs.2023.121995.

[29] Filadi, R., Theurey, P. and Pizzo, P. The endoplasmic reticulum-mitochondria coupling in health and disease: Molecules, functions and significance. Cell Calcium. 2017, 62, pp. 1-15. https://doi.org/10.1016/j.ceca.2017.01.003.

[30] Ramezani, M., Wagenknecht-Wiesner, A., Wang, T., Holowka, D. A., Eliezer, D. and Baird, B. A. Alpha synuclein modulates mitochondrial Ca^{2+} uptake from ER during cell stimulation and under stress conditions. NPJ Parkinsons Dis. 2023, 9(1), p. 137. https://doi.org/10.1038/s41531-023-00578-x.

[31] Bellomo, G., Paciotti, S., Gatticchi, L. and Parnetti, L. The vicious cycle between α-synuclein aggregation and autophagic-lysosomal dysfunction. Movement Disorders. 2020, 35(1), pp. 34-44. https://doi.org/10.1002/mds.27895.

[32] Lang, M., Pramstaller, P. P. and Pichler, I. Crosstalk of organelles in Parkinson's disease - MiT family transcription factors as central players in signaling pathways connecting mitochondria and lysosomes. Molecular Neurodegeneration. 2022, 17(1), p. 50. https://doi.org/10.1186/s13024-022-00555-7.

[33] Bastioli, G., Piccirillo, S., Graciotti, L., Carone, M., Sprega, G., Taoussi, O., Preziuso, A. and Castaldo, P. Calcium deregulation in neurodegeneration and neuroinflammation in Parkinson's disease: Role of calcium-storing organelles and sodium-calcium exchanger. Cells. 2024, 13(15), p. 1301. https://doi.org/10.3390/cells13151301.

[34] Chung, Y. C., Shin, W. H., Baek, J. Y., Cho, E. J., Baik, H. H., Kim, S. R., Won, S. Y. and Jin, B. K. CB2 receptor activation prevents glial-derived neurotoxic mediator production, BBB leakage and peripheral immune cell infiltration and rescues dopamine neurons in the MPTP model of Parkinson's disease. Experimental and Molecular Medicine. 2016, 48(1), p. e205. https://doi.org/10.1038/emm.2015.100.

[35] Ruan, Z., Zhang, D., Huang, R., Sun, W., Hou, L., Zhao, J. and Wang, Q. Microglial activation damages dopaminergic neurons through mmp-2/-9-mediated increase of blood-brain barrier permeability in a Parkinson's disease mouse model. International Journal of Molecular Sciences. 2022, 23(5), p. 2793. https://doi.org/10.3390/ijms23052793.

[36] Rommelfanger, K. S. and Wichmann, T. Extrastriatal dopaminergic circuits of the Basal Ganglia. Frontiers in Neuroanatomy. 2010, 4, p. 139. https://doi.org/10.3389/fnana.2010.00139.

[37] McGregor, M. M. and Nelson, A. B. Circuit mechanisms of Parkinson's disease. Neuron. 2019, 101(6), pp. 1042-1056. https://doi.org/10.1016/j.neuron.2019.03.004.

[38] Cazorla, M., Kang, U. J. and Kellendonk, C. Balancing the basal ganglia circuitry: a possible new role for dopamine D2 receptors in health and disease. Movement Disorders. 2015, 30(7), pp. 895-903. https://doi.org/10.1002/mds.26282.

[39] Ellens, D. J. and Leventhal, D. K. Review: electrophysiology of basal ganglia and cortex in models of Parkinson disease. Journal of Parkinson's Disease. 2013, 3(3), pp. 241-54. https://doi.org/10.3233/jpd-130204.

[40] Galvan, A., Devergnas, A. and Wichmann, T. Alterations in neuronal activity in basal ganglia-thalamocortical circuits in the parkinsonian state. Frontiers in Neuroanatomy. 2015, 9, p. 5. https://doi.org/10.3389/fnana.2015.00005.

[41] Stoessl, A. J., Lehericy, S. and Strafella, A. P. Imaging insights into basal ganglia function, Parkinson's disease, and dystonia. Lancet. 2014, 384(9942), pp. 532-44. https://doi.org/10.1016/s0140-6736(14)60041-6.

[42] Surmeier, D. J., Halliday, G. M. and Simuni, T. Calcium, mitochondrial dysfunction and slowing the progression of Parkinson's disease. Experimental Neurology. 2017, 298(Pt B), pp. 202-209. https://doi.org/10.1016/j.expneurol.2017.08.001.

[43] Swart, T. and Hurley, M. J. Calcium channel antagonists as disease-modifying therapy for Parkinson's disease: Therapeutic rationale and current status. CNS Drugs. 2016, 30(12), pp. 1127-1135. https://doi.org/10.1007/s40263-016-0393-9.

[44] Caulfield, M. E., Manfredsson, F. P. and Steece-Collier, K. The role of striatal Cav1.3 calcium channels in therapeutics for Parkinson's disease. Handbook of Experimental Pharmacology. 2023, 279, pp. 107-137. https://doi.org/10.1007/164_2022_629.

[45] Vijiaratnam, N., Simuni, T., Bandmann, O., Morris, H. R. and Foltynie, T. Progress towards therapies for disease modification in Parkinson's disease. Lancet Neurology. 2021, 20(7), pp. 559-572. https://doi.org/10.1016/s1474-4422(21)00061-2.

[46] Liss, B. and Striessnig, J. The potential of L-type calcium channels as a drug target for neuroprotective therapy in Parkinson's disease. Annual Review of Pharmacology and Toxicology. 2019, 59, pp. 263-289. https://doi.org/10.1146/annurev-pharmtox-010818-021214.

[47] Duda, J., Pötschke, C. and Liss, B. Converging roles of ion channels, calcium, metabolic stress, and activity pattern of Substantia nigra dopaminergic neurons in health and Parkinson's disease. Journal of Neurochemistry. 2016, 139 Suppl 1(Suppl Suppl 1), pp. 156-178. https://doi.org/10.1111/jnc.13572.

[48] Habets, J. G. V., Heijmans, M., Kuijf, M. L., Janssen, M. L. F., Temel, Y. and Kubben, P. L. An update on adaptive deep brain stimulation in Parkinson's disease. Movement Disorders. 2018, 33(12), pp. 1834-1843. https://doi.org/10.1002/mds.115.

[49] Chen, D., Zhao, Z., Shi, J., Li, S., Xu, X., Wu, Z., Tang, Y., Liu, N., Zhou, W., Ni, C., et al. Harnessing the sensing and stimulation function of deep brain-machine interfaces: a new dawn for overcoming substance use disorders. Transl Psychiatry. 2024, 14(1), p. 440. https://doi.org/10.1038/s41398-024-03156-8.

[50] Mao, Z., Ling, Z., Pan, L., Xu, X., Cui, Z., Liang, S. and Yu, X. Comparison of efficacy of deep brain stimulation of different targets in Parkinson's disease: A network meta-analysis. Frontiers in Aging Neuroscience. 2019, 11, p. 23. https://doi.org/10.3389/fnagi.2019.00023.

[51]Porta, M., Servello, D., Zekaj, E., Gonzalez-Escamilla, G. and Groppa, S. Pre-dopa deep brain stimulation: Is early deep brain stimulation able to modify the natural course of Parkinson's disease? Frontiers in Neuroscience. 2020, 14, p. 492. https://doi.org/10.3389/fnins.2020.00492.

[52] Limousin, P. and Foltynie, T. Long-term outcomes of deep brain stimulation in Parkinson disease. Nature Reviews: Neurology. 2019, 15(4), pp. 234-242. https://doi.org/10.1038/s41582-019-0145-9.

[53] Wang, S., Zhu, G., Shi, L., Zhang, C., Wu, B., Yang, A., Meng, F., Jiang, Y. and Zhang, J. Closed-loop adaptive deep brain stimulation in Parkinson's disease: Procedures to achieve it and future perspectives. Journal of Parkinson's Disease. 2023, 13(4), pp. 453-471. https://doi.org/10.3233/jpd-225053.

[54] Johnson, L. A., Nebeck, S. D., Muralidharan, A., Johnson, M. D., Baker, K. B. and Vitek, J. L. Closed-Loop Deep Brain Stimulation Effects on Parkinsonian Motor Symptoms in a Non-Human Primate - Is Beta Enough? Brain Stimulation. 2016, 9(6), pp. 892-896. https://doi.org/10.1016/j.brs.2016.06.051.

[55] Parastarfeizabadi, M. and Kouzani, A. Z. Advances in closed-loop deep brain stimulation devices. Journal of Neuroengineering and Rehabilitation. 2017, 14(1), p. 79. https://doi.org/10.1186/s12984-017-0295-1.

[56] Rosa, M., Arlotti, M., Ardolino, G., Cogiamanian, F., Marceglia, S., Di Fonzo, A., Cortese, F., Rampini, P. M. and Priori, A. Adaptive deep brain stimulation in a freely moving Parkinsonian patient. Movement Disorders. 2015, 30(7), pp. 1003-5. https://doi.org/10.1002/mds.26241.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Vol. 12 (2026): Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/lshe.20260117

LTβR Agonist and IFN-I Cooperatively Induce Antitumor Tertiary Lymphoid Structure Formation: Mechanisms and Therapeutic Prospects

Xiaoyin Wang*

Clinical Medicine, Harbin Medical University, Harbin 150023, China

**Corresponding author: Xiaoyin Wang*

Abstract

This review systematically explores the antitumor strategy of combining lymphoxin beta receptor agonists with type I interferons to induce tertiary lymphoid structure formation, thereby transforming “cold” tumors into “hot” tumors. The article pointed out that this combination therapy can synergistically overcome the microenvironmental barriers of “cold” tumors: LTβR agonists provide structural scaffolds for the formation of tertiary lymphoid structures by activating tumor stromal cells, especially inducing high endothelial vein generation to promote lymphocyte homing; Type I interferon, on the other hand, provides powerful immune activation signals through its downstream signaling pathway, including promoting the recruitment and activation of B cells and T cells, enhancing the function of antigen-presenting cells, and driving B cells to differentiate into follicular-like B cells. Preclinical studies have confirmed that this strategy can effectively enhance the efficacy of existing therapies such as immune checkpoint inhibitors and chimeric antigen receptor T cells, and induce durable anti-tumor immune memory in neoadjuvant therapy, showing broad clinical translation prospects.

Keywords

tertiary lymphoid structures, LTβR agonist, type I interferon, tumor immunotherapy

1. Introduction: TLS – The Central Hub of Tumor Immunotherapy

1.1 Definition and Structural Features of Tertiary Lymphoid Structures (TLS)

Tertiary lymphoid structures are highly organized ectopic aggregates of immune cells that form in non-lymphoid tissues under conditions of chronic inflammation, including infection, autoimmune disease, and cancer. They structurally and functionally mimic secondary lymphoid organs (SLOs) [1]. TLS formation is a dynamic, continuous process, initiated by the release of chemokines induced by inflammatory signals and culminating in the establishment of functional germinal centers (GCs). Their structural core is precisely composed of the following components. High endothelial venules (HEVs) are specialized vascular “entry points” for the lymphocyte-specific homing into TLS. They characteristically express addressins like PNAd, providing a directed migration pathway for CCR7+ lymphocytes, particularly naive T cells and central

memory T cells. Within the TLS, immune cells exhibit distinct spatial compartmentalization. The T-cell zone is formed by a network scaffold of stromal cells expressing CCL19 and CCL21. This chemokine environment recruits and clusters dendritic cells and naive T cells via their receptor CCR7, serving as the critical site for antigen presentation and CD8+ T cell activation. The adjacent B-cell zone is primarily orchestrated by CXCL13. This chemokine recruits CXCR5+ B cells to form follicles and induces the differentiation and maintenance of the follicular dendritic cell (FDC) network, providing survival and proliferation signals for B cells. Under persistent antigen stimulation and with help from follicular helper T (Tfh) cells, a germinal center can form within the B-cell zone. Here, B cells undergo clonal expansion, somatic hypermutation (SHM), antibody class switch recombination (CSR), and affinity maturation, ultimately differentiating into long-lived plasma cells and memory B cells capable of secreting high-affinity antibodies, thereby establishing the foundation for long-term humoral immunity [1, 2].

The maturity of TLS is directly correlated with its immune function. Mature TLS, as the subtype with the highest functional capacity, is strongly associated with favorable patient prognosis, potent anti-tumor immunity, and superior response to immune checkpoint inhibitors (ICIs).

1.2 TLS and Clinical Prognosis: One of the Strongest Immune Biomarkers

The presence and maturity of TLS correlate with better prognosis in multiple cancer types and can serve as a biomarker for immunotherapy efficacy. In cancers such as melanoma, non-small cell lung cancer, breast cancer, head and neck squamous cell carcinoma (HNSCC), and pancreatic cancer, mature (GC-containing) TLS is strongly associated with longer overall survival and enhanced response to ICIs [1, 3, 4]. Among these components, B cells are essential for coordinating the anti-tumor immune response and achieving long-term protective immunity [5].

The presence of mature TLS containing CD23+ follicular dendritic cells is associated with improved objective response rates to ICIs in various human cancers, and this association is independent of CD8+ T cell density and PD-L1 expression [6].

1.3 Deficiencies of Single-agent TLS Induction

Endogenous lymphotoxin-α1β2 (LTα1β2) signals are low in intensity, unstable, and cell-contact dependent, insufficient to drive high-intensity, rapid tissue restructuring. To address these limitations, exogenous administration of an LTβR agonistic antibody (e.g., 4H8) can directly, at high concentration, and independently of cell contact, activate the Lymphotoxin-β receptor (LTβR) on tumor stromal cells, inducing TLS and HEV formation.

LTβR signaling primarily drives the activation of stromal cells like fibroblastic reticular cells (FRCs) and the construction of tissue architecture. It provides relatively weak direct stimulation for the activation, clonal expansion, and functional differentiation of immune cells themselves, particularly B and T cells. TLS induced by LTβR agonist monotherapy exhibits defects in B cell activation, expansion, and functional maturation [5]. This deficiency suggests that while LTβR activation provides the core framework for TLS, the formation and optimization of a functional germinal center (GC) reaction within the TLS requires the synergy of a potent immune-activating signal.

1.4 TLS Neogenesis: LTβR & IFN-I Synergy Drives Inducibility, Maturity, and Anti-tumor Immunity

The neogenesis of TLS recapitulates key steps in lymphoid organ development. It begins with an inflammatory recruitment phase mediated by cytokines like type I interferon (IFN-I), followed by a tissue organization phase dependent on the LTα1β2-LTβR signaling pathway triggering stromal formation, and finally a functional phase characterized by germinal center formation and local humoral immune responses [7]. LTβR signaling provides the necessary organizational scaffold for TLS formation, while IFN-I activation confers critical immune activity to the TLS. The synergistic action of both achieves the core properties of TLS: inducibility, functional maturity, and the ability to mediate anti-tumor immune responses.

2. LTβR Signaling and IFN-I Signaling: Independent Biological Functions and Core Pathways

2.1 The LTβR Signaling Axis: The Skeletal Organizer of TLS Genesis

The LTβR (lymphotoxin-β receptor) signaling axis is a key molecular pathway regulating lymphoid organ development and TLS formation. Its activation depends on the membrane-bound ligand LTα1β2 (composed of 1 LTα and 2 LTβ subunits) or the soluble ligand LIGHT. Upon ligand binding, LTβR primarily activates the intracellular NIK/IKKα signaling module, initiating the non-canonical NF-κB (p52/RelB) transcriptional program and moderately regulating the canonical NF-κB pathway [8]. This axis primarily acts on lymphoid tissue organizer (LTo) cells, vascular endothelial cells, and stromal cells within the tumor microenvironment, inducing lymphoid stromal cells to highly express adhesion molecules like VCAM-1 and ICAM-1, and to secrete key chemokines such as CCL19, CCL21, and CXCL13. This establishes the initial framework for immune cell recruitment and retention [9]. Concurrently, this signaling directly acts on vascular endothelial cells, inducing phenotypic remodeling, upregulating marker molecules like PNAd and MAdCAM-1, thereby driving the transformation of ordinary blood vessels into high endothelial venules (HEVs) with lymphocyte-homing functionality [1].

Through the induced chemokine gradients, LTβR signaling mediates the regionalized arrangement of lymphocytes within the TLS. CXCL13 recruits CXCR5+ B cells and mature Tfh cells to form the follicular area, while CCL19/21 recruits CCR7+ CD4+ T cells, particularly naive/central memory T cells and dendritic cells, homing them to the T-cell zone. This process ultimately results in spatial segregation, forming B-cell and T-cell areas [9]. Furthermore, the CXCL13-CXCR5 axis can induce B cells and LTi cells to express LTα1β2, which binds to LTβR on stromal cells/HEVs, establishing a positive feedback loop that sustains and strengthens the expression of chemokines and adhesion molecules, accelerating TLS maturation [1, 9]. Sustained LTβR non-canonical NF-κB signaling is crucial for maintaining the follicular dendritic cell (FDC) network and stromal cell survival. This mechanism prevents the disintegration of formed TLS structures, ensuring their long-term function [9].

2.2 The IFN-I Signaling Axis: Activating TLS Function

2.2.1 JAK1/TYK2-STAT1/STAT2-IRF9-ISGF3

Type I interferons (IFN-I), primarily including IFN-α and IFN-β, are key cytokines that initiate anti-tumor innate and adaptive immunity. Their signaling is mediated by IFNAR (IFN-α/β receptor, composed of IFNAR1 and IFNAR2 subunits). Upon IFN-I binding, IFNAR activates intracellular JAK1/TYK2 kinases, leading to phosphorylation of STAT1 and STAT2, which then bind with IRF9 to form the ISGF3 transcription complex. This complex translocates to the nucleus and binds to ISRE cis-elements, initiating the expression of hundreds of interferon-stimulated genes (ISGs), including various chemokines (e.g., CXCL9, CXCL10, CXCL11) and their receptors [10, 11].

2.2.2 Tolerance Breakage, APC Activation, and Chemokine - Driven Immune Cell Recruitment

IFN-I signaling is a key driver in breaking tumor immune tolerance and reshaping the immune microenvironment of cold tumors. Its core mechanism lies in potently activating the antigen cross-presentation capacity of cDC1s through its downstream transcription factors (particularly ISGF3 and IRF family members), causing them to upregulate MHC-I and more efficiently present tumor antigens to CD8+ T cells, initiating tumor-specific cytotoxic T lymphocyte (CTL) responses [11]. IFN-I broadly enhances the maturation and function of antigen-presenting cells, not only upregulating MHC-II and co-stimulatory molecule expression but also maintaining MHC-II synthesis and antigen processing functions in intracellular vacuolar compartments, thereby enabling DCs to continuously sample and present antigens to T cells [12]. Concurrently, STAT1 homodimers strongly induce the production of chemokines CXCL9, CXCL10, and CXCL11, forming a chemotactic gradient that efficiently recruits immune cells expressing CXCR3, such as CD8+ effector T cells, into the tumor microenvironment [13]. Additionally, IFN-I acts directly on immunosuppressive cells like MDSCs and Tregs, limiting their suppressive functions, thereby promoting CD8+ T cell anti-tumor activity [11].

3. Mechanism of TLS Induction by LTβR and IFN-I Synergy

LTβR agonists and type I interferon (IFN-I) synergistically induce the formation of anti-tumor tertiary lymphoid structures (TLS) through spatiotemporal coupling, pathway crosstalk, and cellular network interactions.

3.1 Complementary Chemokine Networks and Positive Feedback Cascades in Cellular Networks

The LTβR signal is primarily responsible for promoting the structural foundation of TLSs. This includes driving the formation of the structural scaffold, inducing the differentiation of high endothelial venules (HEVs), and orchestrating the organized recruitment of T cells and B cells to their respective functional zones via chemokines like CXCL13, CCL19, and CCL21, thereby achieving initial tissue organization and compartmentalization [9]. In a lung inflammation model, IFN-I can initiate this process through two synergistic pathways. On one hand, IFN-I induces lymphocytes (particularly early-recruited T cells) to express lymphotoxin alpha (LTα), which subsequently activates LTβR on stromal cells, driving their production of the key B-cell chemoattractant CXCL13 [14]. Blocking LTβR signaling completely suppresses TLS development and CXCL13 expression, confirming the pivotal role of this pathway [14]. On the other hand, IFN-I can be directly sensed by stromal cells, independently of LTβR signaling, inducing their expression of T-cell chemokines CXCL9, CXCL10, CCL19, and CCL21. Notably, although CCL19 is expressed, its absence in this model does not affect TLS formation, suggesting functional redundancy within the chemokine network [14].

Complementing the LTβR signal, the IFN-I pathway focuses on activating and amplifying cellular immunity. It can significantly enhance the activation of antigen-presenting cells (particularly cDC1s). By upregulating chemokines like CXCL9 and CXCL10, IFN-I actively recruits effector CD8+ T cells, Th1 cells, and NK cells from the peripheral circulation or draining lymph nodes into the TLS-containing tumor microenvironment [11,15]. The IFN-I signal-promoted activation of cDC1s facilitates antigen presentation to CD4+ T cells in the T-cell zone, inducing their differentiation into CXCR5+ T follicular helper (Tfh) cells. These Tfh cells migrate along the CXCL13 gradient towards the T-B border and into the B-cell follicles [3].

In the early stages of TLS formation, IFN-I not only creates the necessary inflammatory conditions within an immunosuppressive microenvironment, but the effector T cells it recruits can also upregulate the expression of LTα1β2 upon antigen stimulation [16]. This LTα1β2 can feedback onto LTβR on stromal cells, thereby further consolidating and amplifying the local chemokine gradient. Concurrently, the mature HEVs constructed by LTβR signaling provide an efficient portal for the various immune cells recruited by IFN-I to enter the tissue site [2].

These two signaling pathways exhibit a synergistic effect, jointly amplifying the expression of chemokines and adhesion molecules. The ultimate outcome of this multi-layered collaboration is the sustained recruitment of lymphocytes, the orderly arrangement of T-cell and B-cell zones, the full maturation of functional HEVs, and the refinement of the germinal center (GC) reaction mediated by Tfh cells and B cells. Consequently, scattered immune cell aggregates are transformed into mature TLSs possessing complete lymphoid organ structure and potent anti-tumor activity, achieving a high degree of coupling between the core processes of structural construction and immune activation.

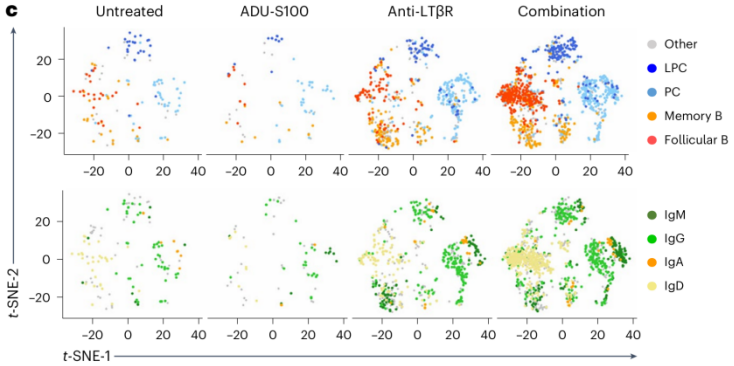
Furthermore, low concentrations of IFN produced in response to TNF are sufficient to increase STAT1 expression, thereby priming immune cells like macrophages to exhibit enhanced responsiveness to subsequent IFN-I signals during early infection [10]. This priming mechanism may also be applicable to the initial phase of TLS formation, where IFN-I in the microenvironment can increase the sensitivity of resident immune cells to subsequent activation signals.

3.2 STING/LTβR Combination Drives B-cell Maturation in TLS

STING (stimulator of interferon genes) agonists bind to the STING protein and activate downstream signaling pathways, ultimately promoting the nuclear translocation of interferon regulatory factor 3 (IRF3), thereby inducing the robust production of type I interferons. As shown in Figure 1, mouse experiments with a STING agonist (ADU-S100) confirmed that the combination of an LTβR activator and IFN-I can promote

the differentiation of memory B cells towards a follicular B-cell-like phenotype. In Figure 1 (top panel), t-SNE visualization shows the abundance of intratumoral B cell subsets across four treatment groups (untreated, ADU-S100, Anti-LTβR, combination therapy). The combination therapy group exhibited a significant increase in follicular B cells, memory B cells, and plasma cells, forming dense, well-defined cell clusters—a hallmark of TLS. In contrast, B cell populations in the monotherapy groups were sparse and dispersed, indicating incomplete TLS induction. Figure 1 (bottom panel) further demonstrates enrichment of IgG produced by plasma cells and IgD involved in maintaining follicular B cell homeostasis in the combination therapy group [2]. This shift towards class-switched, high-affinity antibodies is a key functional output of the germinal center reaction, strongly suggesting that combined activation drives the functional maturation of B cells within the TLS.

Figure 1: Synergistic induction of B cell subsets and immunoglobulin class switch by combined STING and LTβR activation [2].



4. Anti-tumor Immune Effects of Synergistically Induced TLS: From Local Activation to Systemic Response

4.1 In Situ Immune Activation: Tumor-specific T/B Cell Clonal Expansion and Killing

As an immune fortress within the tumor, the core function of a TLS is to achieve in situ antigen presentation and efficient lymphocyte activation [2]. In mature TLSs induced by the synergistic action of an LTβR agonist and IFN-I, CXCL13+ follicular dendritic cells (FDCs) and CD11c+ dendritic cells (DCs) capture tumor antigens. With the assistance of cytokines like IL-6 and IL-12, they activate CD4+ T cells to differentiate into CXCR5+Bcl6+ T follicular helper (Tfh) cells [17]. Tfh cells interact with B cells through CD40L-CD40 signaling, driving the germinal center (GC) reaction, manifested by high Ki-67 expression in B cells, activation-induced cytidine deaminase (AID)-mediated somatic hypermutation (SHM), and antibody class switching (from IgM to IgG) [18]. Simultaneously, CD8+ T cells activated in the T-cell zone of the TLS receive co-stimulatory signals, differentiating into effector cells (Teff) that highly express granzyme B (GZMB) and perforin, which directly infiltrate the tumor parenchyma to execute killing [18]. Single-cell sequencing has confirmed that the clonal diversity of T cells within TLSs is significantly higher than in the periphery, and TCR sequences show a high match with tumor neoantigens, proving tumor-specific clonal expansion [19].

4.2 Immune Memory Formation: Long-lasting Anti-tumor Immunity and Prevention of Recurrence

In mouse models of pancreatic cancer (KPC) and breast cancer (Py230), an experimental design involving initial combination therapy (to induce TLS formation), followed by surgical resection of the primary tumor, and finally tumor re-challenge in mice, simulated the clinical scenario of post-operative prevention of recurrence [2, 4]. Mice receiving the combination neoadjuvant therapy exhibited 100% long-term survival, with re-inoculated tumor growth completely inhibited or regressed, indicating the mice developed robust, durable anti-tumor immune memory [4]. Tumor growth was also significantly inhibited in naive recipient mice receiving serum from donors that underwent combination therapy, demonstrating that humoral immune memory can provide protection via passive antibody transfer [4].

4.3 Tumor Microenvironment Remodeling

The synergistic strategy achieves comprehensive remodeling of the TME through TLS formation. The LTβR signal acts directly on endothelial cells, inducing the differentiation of high endothelial venules (HEVs, MECA-79+), which replace the original leaky tumor vasculature, significantly improving the infiltration efficiency of lymphocytes (especially naive lymphocytes) [20]. The IFN-γ and TNF-α enriched within TLSs inhibit the expansion of myeloid-derived suppressor cells (MDSCs) and impair the suppressive function of regulatory T cells (Tregs) [15]. Tumors that were originally "immune-desert" or "immune-excluded" and lacked T-cell infiltration are transformed into "immune-hot" tumors rich in CD8+ T cells, B cells, and DCs following TLS formation. This transformation occurs concurrently with the upregulation of PD-L1 expression, creating conditions for subsequent immune checkpoint blockade (ICB) therapy [21].

5. Preclinical Models and Translational Evidence

In models of pancreatic ductal adenocarcinoma (PDAC), melanoma (B16-OVA), colorectal cancer (MC38), and breast cancer, combination therapy with an LTβR agonist (e.g., C91) and IFN-I (or a STING agonist as an IFN-I inducer) has demonstrated potent anti-tumor activity superior to monotherapies. In the KPC pancreatic cancer model, monotherapy with either LTβR agonist or IFN-I only slightly delayed tumor growth, whereas the combination therapy induced widespread TLS formation and led to tumor regression [2].

6. Future Perspectives

6.1 Golden Combinations for Combination Therapy

Based on the function of TLSs as niches for maintaining T cell stemness, the combination of an LTβR agonist + IFN-I + PD-1/PD-L1 inhibitor is considered a powerful triple therapy. The induced TLS provides a supportive microenvironment for T cells, particularly TCF1-expressing precursor exhausted T cells (Texprog) with stem-like properties. This helps maintain their stemness and prevents their differentiation into terminally exhausted T cells (Texterm) with irreversible functional impairment. This population serves as a key reservoir capable of proliferating and differentiating into effector cells following PD-1 blockade [2, 22]. Additionally, LTβR + IFN-I + cancer vaccine/oncolytic virus can strengthen antigen supply [23]; LTβR + IFN-I + CAR-T leverages the homing properties of TLSs: HEVs can significantly promote CAR-T cell infiltration into the tumor parenchyma, and concentration gradients of chemokines like CXCL13 facilitate efficient infiltration of CAR-T cells engineered to overexpress corresponding chemokine receptors (e.g., CXCR5) [2, 24]. The supportive microenvironment within the TLS, composed of activated APCs, Tfh-type cytokines (e.g., IL-21), and antibodies produced by B cells, provides co-stimulation for CAR-T cells, prevents exhaustion, and may generate synergistic anti-tumor effects through mechanisms like antibody-dependent cellular cytotoxicity (ADCC) and epitope spreading [1, 5].

6.2 Innovations in Delivery Technology

To overcome the challenges of the short half-life and systemic toxicity of IFN-I, novel delivery systems such as thermosensitive hydrogels, lipid nanoparticles (LNPs), and biodegradable scaffolds are being developed for intratumoral, localized release [25]. These systems can confine IFN-α/β or STING agonists to the tumor site, enabling sustained, low-dose release to mimic chronic inflammatory signals (beneficial for TLS initiation) while avoiding the immunosuppression or toxicity associated with acute inflammation, effectively preventing systemic cytokine storms [25]. Combined with agonistic antibodies targeting LTβR, these approaches hold promise for achieving safe and effective TLS-inducing therapies in the clinic.

7. Conclusions

LTβR agonist with IFN-I can synergistically induce the formation of functional tertiary lymphoid structures (TLS). The LTβR signaling pathway mainly constructs the physical framework of TLS, for example, driving the formation of high endothelial venules; while the IFN-I signaling pathway provides critical immune activation signals. Together, they can effectively convert cold tumors into hot tumors,

showing better anti-tumor effects than single therapy in preclinical models and enhancing the efficacy of existing immunotherapies.

This study clarifies the synergistic mechanism of two core pathways in TLS generation: structural construction and functional activation. The LTβR pathway is responsible for the initial formation and maintenance of tissue structure, while the IFN-I pathway drives the activation and recruitment of immune cells. They enhance each other through positive feedback loops, providing an integrated theoretical framework for understanding the dynamic generation of in-situ tumor immune structures. This strategy offers a new direction for improving the response to solid tumor immunotherapy. It has two main application pathways: first, as a sensitizer for therapies such as immune checkpoint inhibitors, to improve treatment response rates by remodeling the tumor immune microenvironment; second, for neoadjuvant/adjuvant therapy, aiming to induce long-term immune memory and reduce the risk of recurrence. In addition, the maturity of TLS can serve as a potential biomarker for predicting the efficacy of immunotherapy.

Future research will focus on exploring its optimized combinations with existing therapies, such as combining with PD-1 inhibitors, cancer vaccines, or CAR-T cell therapies, to achieve synergistic effects. At the same time, developing novel technologies that enable local and sustained drug delivery is key to overcoming the systemic toxicity of cytokines and promoting their clinical translation.

References

[1] Deng, S., Chen, Y., Song, B., Wang, H., Huang, S., Wu, K., & Chu, Q. (2025). Tertiary lymphoid structures in cancer: spatiotemporal heterogeneity, immune orchestration, and translational opportunities. Journal of Hematology & Oncology, 18(1), 1-29

[2] An, D., Chen, G., Cheng, W. Y., Mohrs, K., Adler, C., Gupta, N. T., ... & Kuhnert, F. (2024). LTβR agonism promotes antitumor immune responses via modulation of the tumor microenvironment. Cancer Research, 84(23), 3984-4001.

[3] Li, H., Zhang, M. J., Zhang, B., Lin, W. P., Li, S. J., Xiong, D., ... & Sun, Z. J. (2025). Mature tertiary lymphoid structures evoke intra-tumoral T and B cell responses via progenitor exhausted CD4+ T cells in head and neck cancer. Nature Communications, 16(1), 4228.

[4] Maxwell, D., Yasuhiro, K., Fumiaki, K., Tomoko, S., Shahidul, I., Brown, G., & Masanobu, K. (2025). 882 Intratumoral B cells are essential for anti-tumor immune responses of tertiary lymphoid structures: impact of STING and LTβR activation in pancreatic cancer. Journal for Immunotherapy of Cancer, 13(Suppl 2), A999-A1000.

[5] Sawada, J., Kikuchi, Y., Duah, M., Herrera, J. L., Kanamori, F., Csomos, K., ... & Komatsu, M. (2025). Simultaneous STING and lymphotoxin-β receptor activation induces B cell responses in tertiary lymphoid structures to potentiate antitumor immunity. Nature Immunology, 26(10), 1766-1780.

[6] Vanhersecke, L., Brunet, M., Guégan, J. P., Rey, C., Bougouin, A., Cousin, S., ... & Italiano, A. (2021). Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. Nature Cancer, 2(8), 794-802.

[7] Sautès-Fridman, C., Petitprez, F., Calderaro, J., & Fridman, W. H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. Nature Reviews Cancer, 19(6), 307-325.

[8] Dejardin, E., Droin, N. M., Delhase, M., Haas, E., Cao, Y., Makris, C., ... & Green, D. R. (2002). The lymphotoxin-β receptor induces different patterns of gene expression via two NF-κB pathways. Immunity, 17(4), 525-535.

[9] Cui, X., Gu, X., Li, D., Wu, P., Sun, N., Zhang, C., & He, J. (2025). Tertiary lymphoid structures as a biomarker in immunotherapy and beyond: advancing towards clinical application. Cancer Letters, 613, 217491.

[10] Ivashkiv, L. B., & Donlin, L. T. (2014). Regulation of type I interferon responses. Nature Reviews Immunology, 14(1), 36-49.

[11] Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M. J., & Kroemer, G. (2015). Type I interferons in anticancer immunity. Nature Reviews Immunology, 15(7), 405-414.

[12] Simmons, D. P., Wearsch, P. A., Canaday, D. H., Meyerson, H. J., Liu, Y. C., Wang, Y., ... & Harding, C. V. (2012). Type I IFN drives a distinctive dendritic cell maturation phenotype that allows continued class II MHC synthesis and antigen processing. The Journal of Immunology, 188(7), 3116-3126.

[13] Wang, Y. Y., Lin, J. F., Wu, W. W., Fu, Z., Cao, F., Chen, Y. X., ... & Xu, R. H. (2025). Inhibition of MBTPS1 enhances antitumor immunity and potentiates anti-PD-1 immunotherapy. Nature Communications, 16(1), 4047.

[14] Calvanese, A. L., Cecconi, V., Stäheli, S., Schnepf, D., Nater, M., Pereira, P., ... & van den Broek, M. (2024). Sustained innate interferon is an essential inducer of tertiary lymphoid structures. European Journal of Immunology, 54(10), 2451207.

[15] Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., ... & Jönsson, G. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature, 577(7791), 561-565.

[16] Schumacher, T. N., & Thommen, D. S. (2022). Tertiary lymphoid structures in cancer. Science, 375(6576), eabf9419.

[17] Crotty, S. (2014). T follicular helper cell differentiation, function, and roles in disease. Immunity, 41(4), 529-542.

[18] Noël, G., Fontsa, M. L., Garaud, S., De Silva, P., De Wind, A., Van den Eynden, G. G., ... & Willard-Gallo, K. (2021). Functional Th1-oriented T follicular helper cells that infiltrate human breast cancer promote effective adaptive immunity. The Journal of clinical investigation, 131(19).

[19] Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., ... & Wargo, J. A. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. Nature, 577(7791), 549-555.

[20] An, D., Chen, G., Cheng, W. Y., Mohrs, K., Adler, C., Gupta, N. T., ... & Kuhnert, F. (2024). LTβR agonism promotes antitumor immune responses via modulation of the tumor microenvironment. Cancer Research, 84(23), 3984-4001.

[21] Colbeck, E. J., Ager, A., Gallimore, A., & Jones, G. W. (2017). Tertiary lymphoid structures in cancer: drivers of antitumor immunity, immunosuppression, or bystander sentinels in disease?. Frontiers in immunology, 8, 1830.

[22] Siddiqui, I., Schaeuble, K., Chennupati, V., Marraco, S. A. F., Calderon-Copete, S., Ferreira, D. P., ... & Held, W. (2019). Intratumoral Tcf1+ PD-1+ CD8+ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. Immunity, 50(1), 195-211.

[23] Gujar, S. A., & Lee, P. W. (2014). Oncolytic virus-mediated reversal of impaired tumor antigen presentation. Frontiers in oncology, 4, 77.

[24] Li, G., Guo, J., Zheng, Y., Ding, W., Han, Z., Qin, L., ... & Luo, M. (2021). CXCR5 guides migration and tumor eradication of anti-EGFR chimeric antigen receptor T cells. Molecular Therapy-Oncolytics, 22, 507-517.

[25] Dosta, P., Cryer, A. M., Prado, M., & Artzi, N. (2025). Bioengineering strategies to optimize STING agonist therapy. Nature Reviews Bioengineering, 3(8), 660-680.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

ESA

EXPLORING SCIENCE ACADEMIC CONFERENCE SERIES

Life Sciences and Health Engineering

2026 International Conference on Life Sciences and Health Engineering · IC-LSHE 2026

EDITOR-IN-CHIEF

Prof. Svetlana V. Androsova

VOLUME

Vol. 12 (2026)

CONFERENCE

IC-LSHE 2026

E-ISSN 3105-0522 **P-ISSN** 3105-0514

ISSN 3105-0514

12>