Integrating AI and Novel Data Sources for miRNA Knowledge Graph Construction in Biomedical Discovery

Yanyan Ma[1], Dong Sha[2**], Jiong Ma[1], Xuejun Chen[1], Gaowei Mao[2], Long Qian[2], Yuehua Sheng[3*], Wei Zheng[1*], Tingyi Lu[4], Junbo Xu[5]**

[1]Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310000, China
[2]Ningbo Intellectual Property Center, Ningbo, Zhejiang 315048, China
[3]Zhejiang Chinese Medical University, Hangzhou, Zhejiang 310053, China
[4]Karazin Business School, V. N. Karazin Kharkiv National University, Kharkiv 61022, Ukraine
[5]Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China

Keywords: miRNA; Knowledge graph construction; Patent data; Knowledge discovery in biomedical research; Bioinformatics; Network analysis; Precision medicine

Citation: Ma Y.Y., S D., Ma J., et al.: Integrating AI and Novel Data Sources for miRNA Knowledge Graph Construction in Biomedical Discovery. Data Intelligence, 2026, Vol. 8, Art. No.: 20251169. DOI: https://doi.org/10.3724/2096-7004.di.2025.1169

ABSTRACT

MicroRNAs (miRNAs) are key regulators of gene expression and disease processes, making them essential targets in biomedical discovery. While existing miRNA knowledge graphs rely primarily on literature and structured databases, they may overlook emerging signals embedded in alternative repositories.

To enable systematic discovery, we employ a pipeline leveraging Microsoft's GraphRAG framework, combining domain-specific prompting with expert validation to extract reliable entity–relation pairs from long-form patent texts. Using this workflow, we construct a miRNA knowledge graph from 2,937 global invention patents (2021–2023), uncovering novel associations among miRNA subtypes, diseases, drugs, and biomarkers.

The graph reveals several discoveries: lung, breast, and brain cancers form prominent pathological cores; miR-21, miR-17-5p, and miR-26a exhibit strong multidisease connectivity; and underexplored signals such as preeclampsia and hair graying/loss emerge. Putative immuno-regulatory axes (e.g., PD-L1–miRNA links) further suggest translational potential.

*Corresponding author: Yuehua Sheng (E-mail: shengyuehua2024@outlook.com); Wei Zheng (E-mail: zhengwei@zju.edu.cn).
**Yanyan Ma and Sha Dong contributed equally to this work and are co-first authors.

© 2026 Chinese Academy of Sciences. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

HIGHLIGHTS

- By integrating GraphRAG with tailored prompting and expert review, we provide a reproducible framework that offers a new perspective on mining biomedical insights from long-form patent data.
- We present, to our knowledge, the first patent-derived miRNA knowledge graph, integrating 2,937 invention patents (2021–2023) and 10,560 entities.
- The graph reveals clinically relevant associations (e.g., miR-485 → NSCLC → anlotinib; the miR-21 family in breast/lung cancer) and emerging applications (hair health, ocular diseases) not visible in traditional databases.
- Quantitative validation on 100 randomly sampled patents shows micro-averaged Precision/Recall/F1 = 56.32/86.12/68.10.

1. INTRODUCTION

MicroRNAs (miRNAs) are a class of endogenous noncoding RNAs that regulate gene expression by binding to specific messenger RNAs (mRNAs), thereby modulating a wide range of biological processes, including immune responses [1-2], cell cycle progression [3-4], and tumor invasion [5-6]. The discovery of the first miRNA, *lin-4*, in *Caenorhabditis elegans* in 1993 [7-8] marked a seminal advancement, and a landmark study by Calin et al. in 2002 further revealed the aberrant expression of miRNAs in cancer, establishing their crucial role in human disease, particularly malignancies [9-10].

Owing to their high stability, tissue-specific expression patterns, and ease of detection, miRNAs have been recognized as promising molecular biomarkers with substantial potential for clinical translation [11-13]. They have demonstrated significant utility in early disease detection, therapeutic decision-making, and prognosis monitoring [14-15]. With the rapid development of computational biology, artificial intelligence (AI) and machine learning techniques have been increasingly integrated into miRNA research, enabling high-throughput analysis of complex biological datasets and facilitating the identification of disease-associated expression signatures. Representative studies in this area include the use of naive Bayes classifiers combined with TarBase for miRNA target prediction [16], the construction of multidimensional molecular interaction networks based on OMIM and HPRD for disease gene discovery [17], and the application of supervised learning algorithms to TCGA breast cancer datasets for molecular subtyping [18]. More recently, large language models (LLMs) have been utilized for the automated construction of knowledge graphs from unstructured biomedical text. Studies have shown that LLM-based approaches exhibit improved accuracy in entity and relationship extraction [19], highlighting the increasing convergence of advanced computational methods and miRNA research.

Despite these developments, existing miRNA knowledge graphs still draw predominantly on peer-reviewed literature, clinical narratives, and curated biomedical databases. While these sources provide high-quality evidence, they often exhibit delays in dissemination, uneven topical coverage, and limited visibility into translational or application-oriented innovations. As a result, current graphs may struggle to capture emerging biomedical insights in a timely and comprehensive manner.

Knowledge graphs have proven effective in structuring heterogeneous biomedical information and elucidating intricate relationships among biomedical entities [20]. However, their application in miRNA research has largely focused on traditional data domains, leaving other potentially valuable sources underutilized. In particular, patent documents constitute a valuable yet largely untapped resource. While patents are not conventionally employed as primary scientific evidence, they contain detailed technical disclosures, chronological innovation trajectories, and early indicators of research direction and enterprise strategy [21].

The absence of a patent-derived knowledge graph represents a notable gap in the current miRNA informatics landscape. Patents can provide a complementary perspective, capturing nascent applications, translational potential, and cross-disciplinary linkages that may elude traditional scientific databases—not by replacing these sources but rather by enriching the conceptual and technical breadth of the knowledge graph.

Moreover, large language models (LLMs) have exhibited remarkable capabilities in parsing and extracting information from unstructured text. However, without domain-specific grounding, LLMs are prone to hallucinations—generating grammatically correct but factually inaccurate outputs. Meanwhile, mining patent documents remains challenging because of their unstructured format, heterogeneous terminology, and lack of standardized annotation. To address these challenges, we adopt Microsoft's GraphRAG framework, a retrieval-augmented generation (RAG) approach that combines the linguistic capabilities of large language models (LLMs) with the structured guidance of knowledge graphs. This hybrid framework enhances the reliability and reproducibility of entity–relation extraction from long-form technical texts [22].

In this study, we construct a miRNA knowledge graph on the basis of 2,937 global invention patents published between 2021 and 2023. By systematically extracting and integrating key biomedical entities—such as miRNA subtypes, genes, proteins, diseases, drugs, and therapeutic strategies—this work highlights the potential of patents as valuable complementary data sources. These documents capture innovation trajectories and translational signals that are often missed by literature-based approaches. To our knowledge, this is the first systematic effort to construct a miRNA-focused knowledge graph using patent data. This study addresses a critical gap in current miRNA informatics and provides a new lens for uncovering emerging biomedical insights.

2. MATERIALS AND METHODS

The overall workflow for constructing the patent-based miRNA knowledge graph is shown in Figure 1.

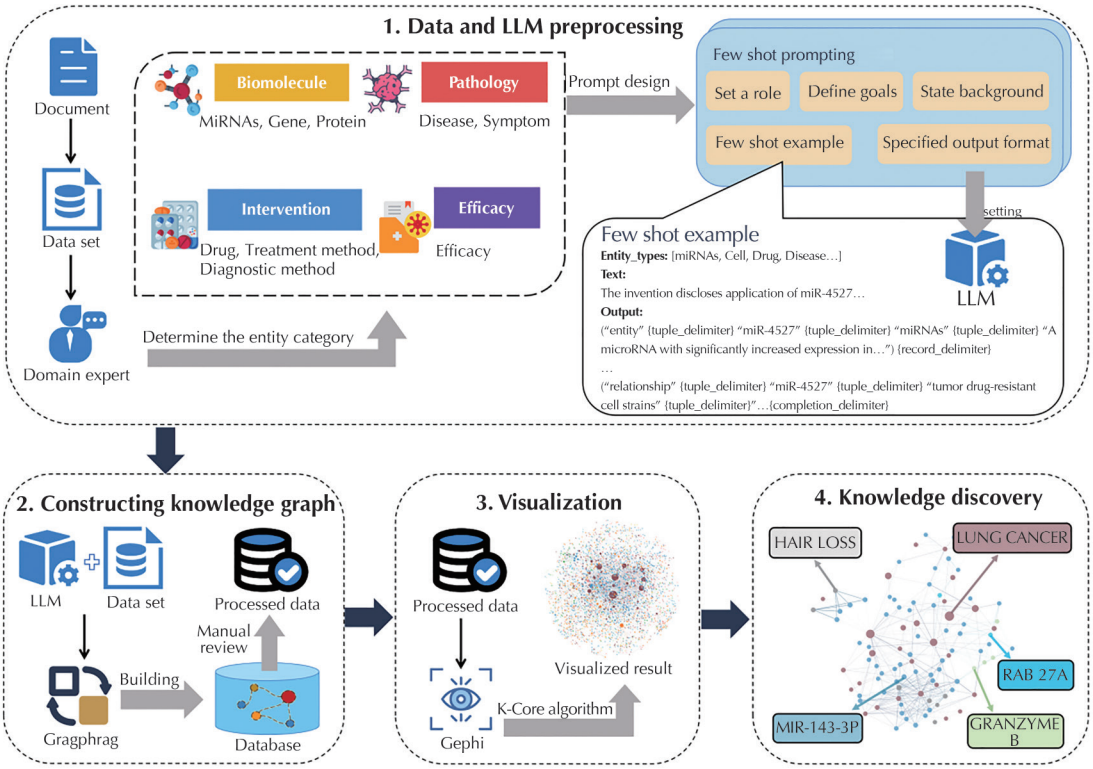


Figure 1. Workflow for constructing the patent-based miRNA knowledge graph via the GraphRAG framework.

The workflow comprises four stages:

(1) Patent data collection and preprocessing: Invention patents published between 2021 and 2023 were retrieved and curated into the "miRNA Invention Patent Dataset," with entities classified into four semantic domains—biomolecules (miRNAs, genes, proteins, cells), pathology (diseases, symptoms), interventions (drugs, treatment and diagnostic methods), and efficacy indicators.

(2) Entity and relationship extraction: Using Zhipu AI's GLM-Long with a few-shot prompting design (clear objectives, role definitions, background context, standardized output formats, and examples), entities and relationships were extracted from unstructured patent abstracts to construct an initial knowledge graph.

(3) Graph validation and analysis: Domain experts reviewed the extracted graph for accuracy, followed by K-core decomposition to identify core structures and highly connected entities.

(4) Knowledge discovery: Multilevel analyses were conducted across biomedical domains, revealing associations such as miR-143-3p with hair loss and RAB27A with lung cancer. These results provide interpretable and scalable insights into the translational potential of miRNA-related technologies.

2.1 Data acquisition and screening

The patent data utilized in this study were obtained from the Himmpat Patent Search System (https://www.himmpat.com). To retrieve miRNA-related invention patents, we employed the following Boolean query:

(miRNA/ta OR microRNA/ta) AND (apd = 2021–2023) AND (Invention Application OR Invention Grant)/pty

Here, "/ta" specifies that the term must appear in the title or abstract, "apd" restricts the application/publication date to the years 2021–2023, and "/pty" limits the results to invention patent applications or granted invention patents. This query initially returned 3,468 records. A rigorous cleaning procedure was then performed, including the removal of entries with missing abstracts, duplicate elimination, and manual expert review to confirm their technical relevance. After these filtering steps, a final dataset of 2,937 validated patents was compiled, forming the "miRNA Invention Patent Dataset" used in subsequent analyses (Supplementary Information S2).

2.2 Prompt Design

In this study, the GraphRAG framework was employed to construct a knowledge graph from patent texts. Since GraphRAG relies on prompt engineering to guide large language models (LLMs) in entity and relation extraction, manual prompt tuning—implemented in the form of few-shot prompting—was adopted as a core step of the workflow, which is consistent with Microsoft's official documentation on GraphRAG manual prompt tuning.

To optimize the extraction process, we applied few-shot prompting [23] to design, refine, and adjust the instruction templates. The design followed four principles: (1) setting clear task objectives; (2) specifying contextual conditions; (3) defining a standardized output format; and (4) providing illustrative few-shot examples. Through iterative adjustment of task instructions, role definitions, and domain-specific background information, the prompts were refined to maximize extraction accuracy from patent abstracts, which are typically lengthy and heterogeneous in style.

In the broader prompting literature, manual prompt design—especially few-shot prompting—has consistently been shown to outperform zero-shot prompting in tasks such as entity and relation extraction [24]. This is consistent with our finding that refined prompts yield more stable and domain-consistent extractions compared to zero-shot baselines.

A taxonomy of nine entity types relevant to the miRNA biomedical domain was defined with expert input: miRNA subtypes, genes, proteins, diseases, symptoms, drugs, treatment methods, diagnostic methods, and efficacy indicators. This taxonomy ensured domain-specific precision in extraction. Each prompt explicitly framed the task as being performed by a "knowledge graph expert" and enforced a structured JSON output format for downstream integration.

To validate the reliability of our prompt design, we quantitatively evaluated two independent random samples of 50 patent abstracts each, drawn via a Python script from the full corpus of 2,937 miRNA-related invention patents. For each sample, two domain experts independently annotated entities; disagreements were adjudicated to produce the final gold annotations. Extracted entity strings were compared against gold strings after deduplication by unique entity string and case-insensitive matching, additionally removing a single pair of wrapping quotes in predicted strings when present. Precision/Recall/F1 were 56.79/85.28/68.18 on the first sample and 55.63/87.37/67.98 on the second; micro-averaged across 100 patents, performance was 56.32/86.12/68.10. The complete prompting framework and the detailed evaluation protocol and results (including the reported Precision/Recall/F1, sampling procedure, and matching rules) are provided in Supplementary Information S3. To illustrate the few-shot design, we show the core instruction template and a representative extraction example below.

Role: You are an expert in the field of knowledge graph construction.

Task: Identify relevant entities and relationships from the patent abstract provided below.

Constraints:

Entity Extraction: Extract entities belonging to types: [miRNAs, Gene, Protein, Disease, Symptom, Drug, Treatment method, Diagnostic method, Efficacy].

Relationship Extraction: Identify pairs of (source_entity, target_entity) that are clearly related, providing a description and a strength score (1-10).

Format: Output as a delimited list using ("entity" <|> <name> <|> <type> <|> <description>) and ("relationship" <|> <source> <|> <target> <|> <description> <|> <strength>).

Few-shot Example:

Input Text:

"The invention discloses application of miR-4527 in preparation of a preparation for diagnosing or treating tumor drug resistance. The invention discovers that the expression level of miR-4527 in drug-resistant tumor cells is remarkably increased, and the inhibition of miR-4527 in drug-resistant tumor cells can remarkably reduce the drug resistance of the tumor cells to cis-platinum."

Model Output:

("entity" <|> "miR-4527" <|> "miRNAs" <|> "A microRNA with significantly increased expression in drug-resistant tumor cells, involved in diagnosing or treating tumor drug resistance.")##

("entity" <|> "cis-platinum" <|> "Drug" <|> "A chemotherapy drug whose resistance in tumor cells is reduced by inhibiting miR-4527.")##

("entity" <|> "tumor drug resistance" <|> "Disease" <|> "A condition involving reduced effectiveness of drugs in treating tumor cells.")##

("relationship" <|> "miR-4527" <|> "cis-platinum" <|> "Inhibition of miR-4527 reduces tumor cells' resistance to cis-platinum" <|> 9)##

("relationship" <|> "miR-4527" <|> "tumor drug resistance" <|> "miR-4527 is involved in the diagnosis and treatment of tumor drug resistance" <|> 8) <|COMPLETE|>

2.3 Knowledge graph construction

In this study, a knowledge graph was constructed via the GraphRAG framework, which consists of four primary stages: document segmentation, natural language processing (NLP), hierarchical community clustering, and data integration. Initially, input patent documents are segmented and embedded, breaking the texts into multiple granular units to facilitate detailed analysis and comprehensive information extraction. Subsequently, advanced large language model (LLM) techniques are employed to identify and extract key entities, their interrelationships, and core claims within these segmented units, resulting in the formation of a preliminary knowledge graph [25-26]. The extraction process is guided by the few-shot prompting strategy described in Section 2.2, which ensures alignment between the user intent and model output. The LLM component is implemented via Zhipu AI's GLM-Long, which supports ultralong context windows of up to 128K tokens. This extended context capacity significantly reduces semantic fragmentation and information loss when processing lengthy patent texts, thereby improving the accuracy and completeness of entity–relationship extraction. Following this step, the Leiden community detection algorithm is applied to perform hierarchical clustering, effectively uncovering intrinsic community structures and assigning entities to their corresponding clusters [27]. In the final stage, the clustered community data are systematically integrated into a database, where entities and relationships are stored as structured semantic triples. Moreover, the knowledge graph is further enriched by incorporating vector-based representations, substantially improving semantic depth and retrieval efficiency. The detailed workflow of the knowledge graph construction process is illustrated in Figure 2 (Supplementary Information S4), where key implementation parameters (e.g., chunk size, embedding/LLM models, parallelization settings) are documented for reproducibility.

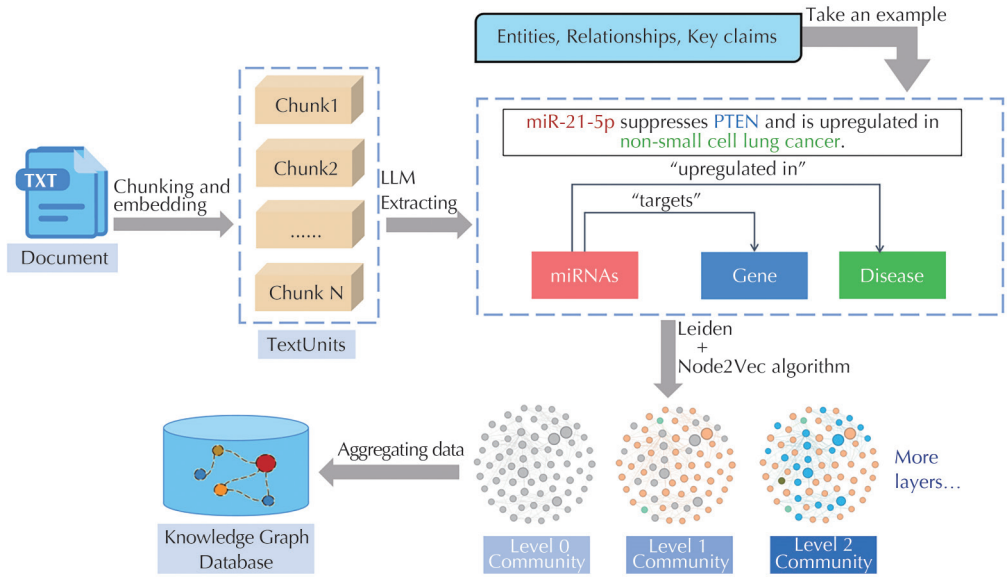


Figure 2. Flowchart of knowledge graph construction.

2.4 Entity normalization and synonym consolidation

Entity names were harmonized to standard vocabularies—miRBase for miRNAs (miR-XX-5p/3p; optional hsa-), HGNC for genes (preferred symbols), and MeSH/ICD for diseases (acronyms expanded; spelling unified). Other categories were standardized for casing/terminology (e.g., ELISA, qRT-PCR, extracellular vesicles; siRNA/shRNA/CRISPR; IFN-γ). Obvious synonyms were merged within each category under a canonical normalized label, with weights summed across merged entries. The weighted degree of a node is defined as the sum of weights of all edges incident to it. Accordingly, the reported per-entity values in the tables denote the node weighted degree, whereas the average weighted degree used by Gephi is a global network metric calculated as the mean of node weighted degrees across all nodes.

2.5 Knowledge graph visualization

In this study, Gephi software [28] was used to visualize the knowledge graph. Gephi is a tool widely used in network science for processing and analyzing relational data. It supports various layout methods for graph visualization, efficiently handles large-scale network datasets, and statistically analyzes network attributes at the node level [29]. During the visualization process, the node size was mapped to each node's weighted degree (sum of incident edge weights); the node color encodes the entity type.

2.6 K-Core algorithm

In the miRNA knowledge discovery process, this study employed the K-Core algorithm to extract structurally cohesive subgraphs from the constructed knowledge graph for visualization and further

analysis. The K-Core algorithm identifies subgraphs in which each vertex is connected to at least k other vertices [30-31]. By iteratively removing vertices with degrees lower than k, the procedure highlights the most interconnected and structurally important nodes within the graph.

In practice, we tested multiple k values to evaluate their impact on subgraph characteristics (the k-core principle is illustrated in Figure 3). Smaller k values (e.g., k = 1) produced dense graphs that lacked interpretability, whereas larger k values fragmented the network into overly sparse components. To balance graph density and clarity, we selected the most suitable k values for visualization. This choice allowed us to effectively highlight representative clusters of miRNA-related entities and to identify core nodes that reveal the internal organization of the network. The resulting core subgraphs provided a clearer and more interpretable basis for analyzing key associations in the patent-derived miRNA knowledge graph. We selected k = 4, 3, and 2 in the subsequent result visualizations to balance readability and structural connectivity based on iterative screening of values from 1–6. Specifically, the 4-core subgraph (k = 4) corresponds to Discovery 1 , highlighting tightly connected biomolecular–pathological hubs, while the 2-core and 3-core subgraphs (k = 2, 3) are used for Discoveries 2–4 , capturing broader intervention and efficacy relationships.

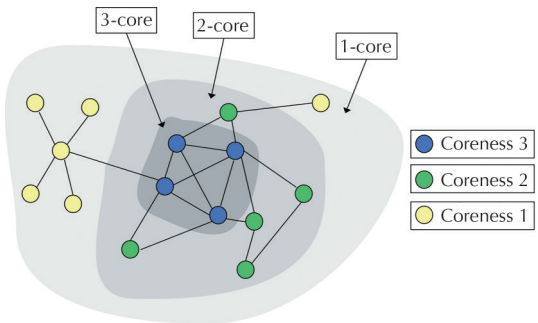


Figure 3. Diagram of k-core decomposition.

3. Results

In this study, manual inspection and correction were performed on the preliminary knowledge graph to ensure the accuracy and consistency of the extracted entities and relationships. The finalized graph comprises 10,560 nodes and 11,754 edges. Visualization was conducted via Gephi, as shown in Figure 4, to provide a global view of the graph's structural characteristics.

In the visualized graph, several large nodes are prominently located near the center. The size of each node is proportional to its weighted degree (the sum of incident edge weights), indicating that nodes with higher connectivity strength appear larger in the layout. Node color is determined by the "type" attribute, enabling visual partitioning by entity category—such as miRNAs, genes, diseases, drugs, and so on. This visual encoding facilitates intuitive interpretation of the distribution, importance, and interconnectivity of different types of entities within the knowledge graph. The complete graph structure and attribute annotations are detailed in Figure 4 and Supplementary Information S5.

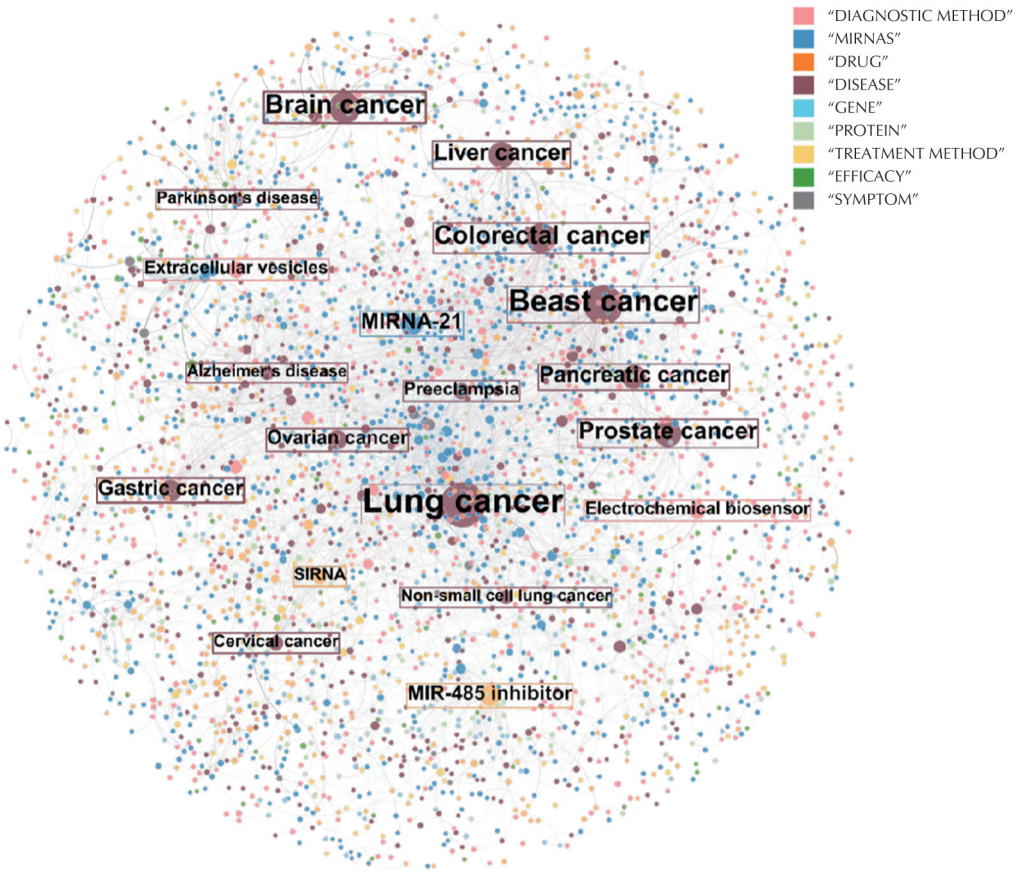


Figure 4. Overall visualization of the miRNA patent knowledge graph.

Figure 5 illustrates the data table interface after importing the constructed knowledge graph into Gephi software. Using non-small cell lung cancer as an example, researchers can thoroughly explore the relevant data through careful review of this interface. In the table, the "Source" field clearly indicates that non-small cell lung cancer is the starting node. The "Target" field lists the associated target nodes, whereas the "Description" field provides a concise and accurate explanation of the relationship between the starting node and the target entity. Additionally, the "weight" field quantifies the strength of the association; the higher the weight value is, the closer the association between the two. This structured table design not only clarifies the complex connections between non-small cell lung cancer and other entities but also offers an intuitive and efficient way to display the data. This interface also enables task-oriented retrieval, where a user can input a query node (e.g., a miRNA or disease) and quickly obtain the top-ranked connected entities and short paths based on edge weights for hypothesis screening. It provides strong support for further research in this field.

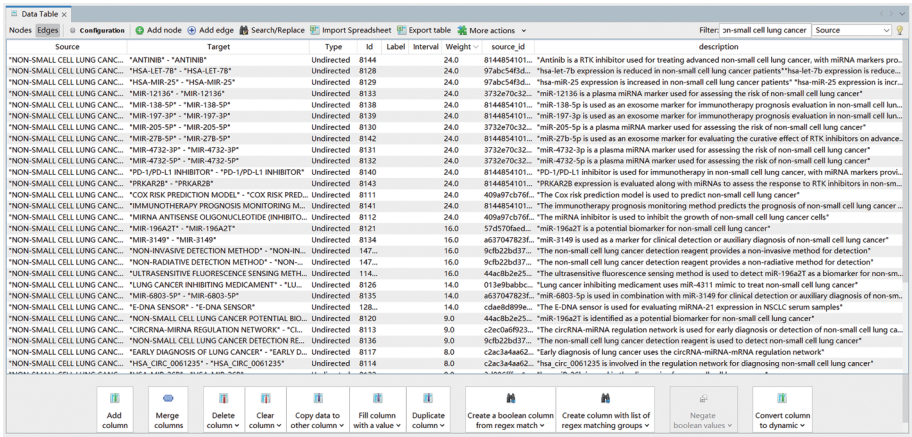


Figure 5. Edge-table view of non-small cell lung cancer (NSCLC).

3.1 Entity extraction results

In this study, named entity recognition was categorized into nine specific types. The proportional distribution of each category is shown in Figure 6, and the detailed results of entity extraction are provided in Table 1 (Supplementary Information S6).

The entity distribution—after harmonizing names to standard vocabularies (miRBase for miRNAs, HGNC for genes, MeSH/ICD for diseases) and merging clear synonyms—shows the five largest categories (Figure 6): Diagnostic Method (N = 2676), MiRNAs (N = 2241), Drug (N = 1369), Disease (N = 1102),

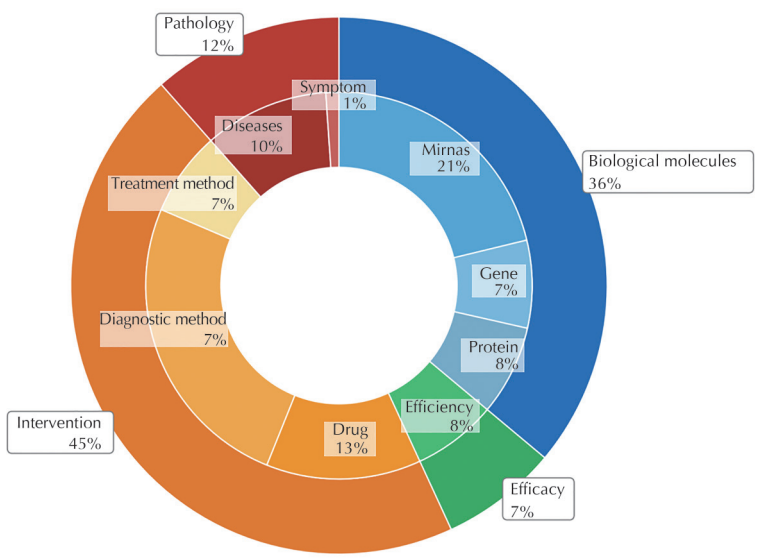


Figure 6. Proportion of entity categories.

Table 1. Entity categories and representative leaders (node weighted degree).

Category (count)	Entity leaders (NodeWeighted Degree)
Diagnostic Method (N = 2676)	Extracellular Vesicles (970); Electrochemical Biosensor (937); miRNA-Disease Association Prediction Method (707); miRNA Biomarkers (637); Disease/Constitution Analysis Method (538); miRNA Detection Method (485); Predicting Pregnancy Complications (380); miRNA Detection Kit (337); Antitubercular Curative Effect Evaluation (336); 4-miRNA Biomarker Model (331);
MiRNAs (N = 2241)	miR-21 (2000); miR-17-5p (659); miR-26a (628); miR-143-3p (547); miR-133 (536); miR-155 (534); miR-8485 (519); miR-449a (494); miR-21-5p (484); miR-485 (483);
Drug (N = 1369)	miR-485 Inhibitor (1291); siRNA (810); shRNA (470); Antisense Oligonucleotides (402); Clozapine (360); PAX6 Inhibitor (332); Modified RT-let-7 (325); 5-Fluorouracil (292); siRNA Molecules (292); Neural Stem Cell Microparticles (288);
Disease (N = 1102)	Lung Cancer (3197); Breast Cancer (2744); Brain Cancer (2140); Colorectal Cancer (2027); Prostate Cancer (1787); Liver Cancer (1713); Pancreatic Cancer (1553); Gastric Cancer (1391); Ovarian Cancer (1175); Preeclampsia (960);
Protein (N = 796)	SMN1 Protein (355); Polypeptide (208); Engineered Human Mitofusin-2 (207); RBFOX2 Protein (192); Therapeutic Protein (180); Serine/Arginine-Rich Splicing Factor 1 (SRSF1) (177); PD-L1 (171); Viral Envelope Glycoprotein (166); Regulatory Peptides (miPEPs) (146); Silk Fibroin Affinity Peptide (135);
Gene (N = 773)	CTNNB1 (390); MBNL1 (262); MBNL2 (232); SIRT1 (159); ICP34.5 (146); Brown Planthopper LNCINR2 (144); Caspase-3 (143); PHD2 (141); Open Reading Frame (ORF) (131); RAB27A (128);
Treatment Method (N = 744)	mRNA Construct (405); Gene Therapy (388); Nucleic Acid Delivery System (299); Lipid-Based Topical Injection Formulation (292); Recombinant Oncolytic Viral Vectors (289); RNA Interference (284); AAV9-Based Recombinant Virus (266); Titanium Dental Implant (207); Non-Integrating Viral Delivery System (198); Adoptive Cell Therapy (156);
Efficacy (N = 744)	Hair growth (430); Protein expression (249); Immune enhancement (192); Payload gene expression (176); Melanocytes activation (142); Immune function (136); Proliferation of sheep dermis fibroblasts (136); Apoptosis (135); Cardiomyocyte proliferation (133); Cognitive function (131)
Symptom (N = 115)	Fetal growth restriction (627); Fibrosis (485); Hair graying (456); Hair loss (420); Acute stress (363); Spontaneous abortion (360); Stillbirth (360); Cachexia (185); Tumor drug resistance (144); Sensorineural hearing loss (118)

Note: Values in parentheses denote the node weighted degree (sum of weights of edges incident to the node). In Gephi, the average weighted degree is a global network metric computed as the sum of all node weighted degrees divided by the number of nodes. The names reflect normalized forms (miRBase for miRNAs, HGNC for genes, and MeSH/ICD for diseases), and obvious synonyms were merged. The detailed normalization rules and synonym mappings are available in Methods (§2.4) and Supplementary Information S6. The values are not directly comparable across categories due to differing degree distributions.

and Protein (N = 796). The representative leaders are Diagnostic Method—Extracellular Vesicles, Electrochemical Biosensor, miRNA–Disease Association Prediction Method, miRNA Biomarkers, miRNA Detection Method/Kit; MiRNAs—miR-21, miR-17-5p, and miR-26a; Drugs—miR-485 inhibitor, siRNA, shRNA, and antisense oligonucleotides; Disease—Lung Cancer, Breast Cancer, Brain Cancer, and Colorectal Cancer; and Proteins—SMN1 protein, PD-L1, and SRSF1.

It is important to observe that the entity distribution in our knowledge graph follows a characteristic long-tail pattern. While Table 1 highlights the high-degree "hub" nodes that dominate the network structure (e.g., miR-21, Lung Cancer), a significant portion of the graph consists of low-frequency entities situated at the periphery. These long-tail entities often represent specific chemical modifications, rare disease applications, or nascent innovations that have not yet accumulated high patent citations or broad industry attention. Although the current analysis prioritizes core subgraphs to validate the framework against established biomedical knowledge, these sparse entities hold substantial value for detecting early-stage technological trends. Mining these "hidden gems" in the long tail via link prediction algorithms will be a key focus of our future work to enhance the comprehensiveness of knowledge discovery.

3.2 MiRNA knowledge discovery

To enhance the knowledge discovery process, this study utilized the K-Core algorithm to filter the knowledge graph, with the goal of identifying and extracting closely related subgraphs that meet specific core degree requirements. Owing to the powerful visualization capabilities of Gephi software, these core subgraphs can be intuitively displayed.

Through in-depth analysis of these core subgraphs, researchers can more accurately and efficiently extract key information from the miRNA knowledge graph, gaining a deeper understanding of miRNA-related biomolecules, pathological mechanisms, intervention strategies, and their therapeutic effects. Additionally, this analysis helps reveal the complex interactions and connections between different fields of knowledge, providing robust data support and a theoretical foundation for advancing miRNA research.

3.2.1 Discovery 1: Pathology hubs and miRNA leaders

Figure 7 shows the 4-core subgraph screened out by the K-Core algorithm, which reveals the connections between biomolecule categories (miRNA subtypes, proteins, genes) and pathological categories (diseases and symptoms). Among the disease-related clusters, lung cancer appeared most frequently, followed by breast cancer and brain cancer. Specific miRNA subtypes, such as miR-21, miR-143-3p, and miR-17-5p, are strongly linked to cancers, including breast, gastric, and lung cancer. In addition, androgen receptor (AR) H875Y was identified as a diagnostic gene marker closely correlated with bladder and lung cancer.

In terms of symptom-related associations, miR-3199, miR-8485, miR-3189, and miR-3139 were connected to hair-related conditions such as graying and hair loss, whereas miR-16-5p, miR-17-5p, and miR-20a-5p were associated with pregnancy complications.

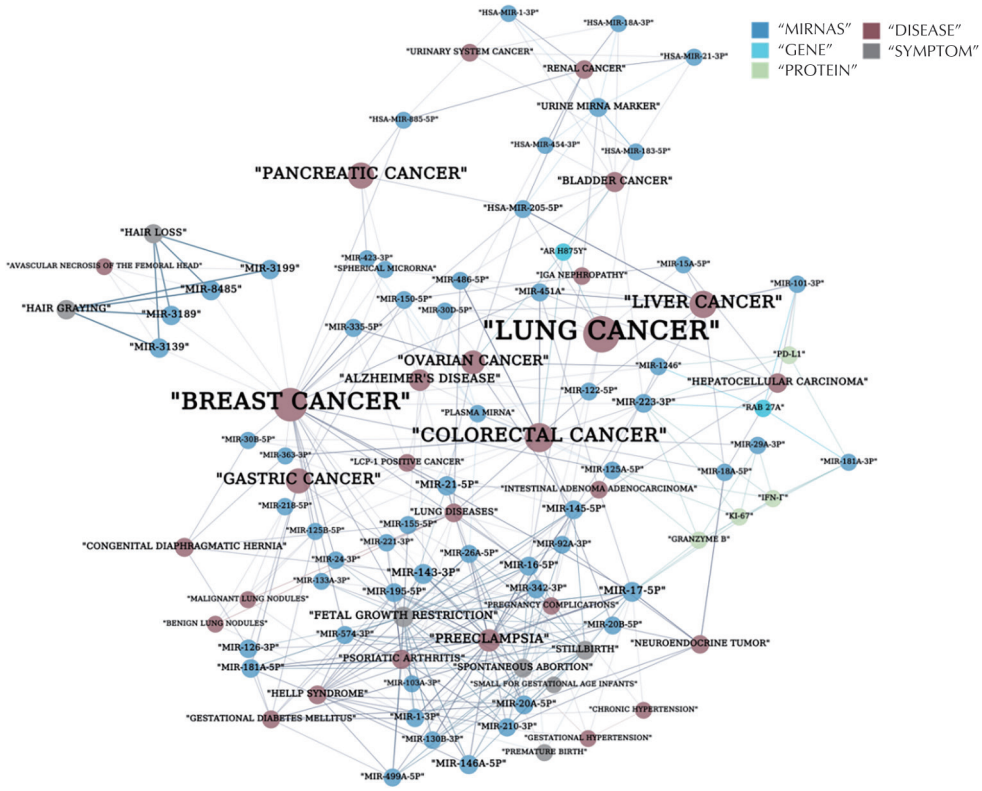


Figure 7. Biomolecular-pathological core subgraph.

Regarding biomolecular interactions, the subgraph highlights links between RAB27A and PD-L1 and several miRNAs (e.g., miR-101-3p, miR-181a-3p, miR-223-3p, and miR-1246) that are related to anticancer activity. Similarly, GRANZYME B, KI-67, and interferon-gamma (IFN-γ) are closely connected with miRNAs such as miR-17-5p, miR-223-3p, and miR-29a-3p, which are associated with immunomodulatory functions. These associations indicate that the subgraph effectively captures clinically relevant miRNA–disease–gene relationships. Further details are provided in Supplementary Information S7.

3.2.2 Discovery 2: Intervention patterns linked to pathology

Figure 8 shows the 2-core subgraph screened out by the K-Core algorithm, which reveals the connections between intervention categories (drugs, diagnostics, and therapies) and pathological categories (diseases and symptoms). Among drug interventions, miR-485 inhibitors are the most frequently mentioned, showing associations with both neurodegenerative diseases and hair-related conditions. For example, miR-485 inhibition was associated with the regulation of SIRT1, PGC-1α, and CD36 expression in tau protein–related disorders and with symptoms such as alopecia areata, abnormal hair shafts, cicatricial alopecia, and telogen effluvium.

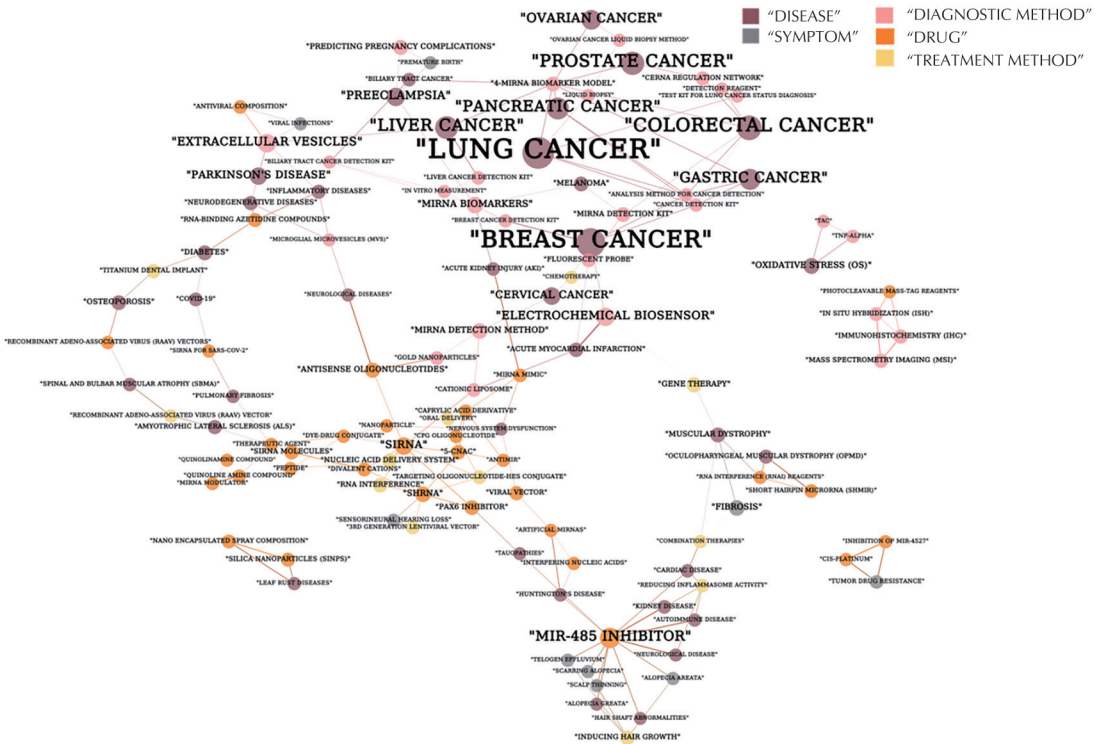


Figure 8. Pathology-intervention core subgraph.

With respect to diagnostic interventions, extracellular vesicles (EVs) are prominent and are associated with neurodegenerative, inflammatory, and autoimmune diseases. EVs carry specific miRNA subtypes that serve as biomarkers or therapeutic agents. Electrochemical biosensors have also been broadly applied in clinical analysis and disease detection, particularly for acute myocardial infarction.

In terms of therapies, gene therapy is one of the most common, together with nucleic acid delivery systems and RNA interference technologies, which are frequently represented in the subgraph. Further details are provided in Supplementary Information S8.

3.2.3 Discovery 3: From interventions to functional endpoints

Figure 9 shows the 2-core subgraph screened via the K-Core algorithm, which reveals the connections between intervention categories (drugs, diagnostics, therapies) and the efficacy category. In diagnostics, electrochemical biosensors are widely used, reducing molecular diffusion and enabling shorter detection times, whereas lateral flow chromatographic test strips employ colloidal gold or fluorescence immunochromatography for rapid miRNA detection.

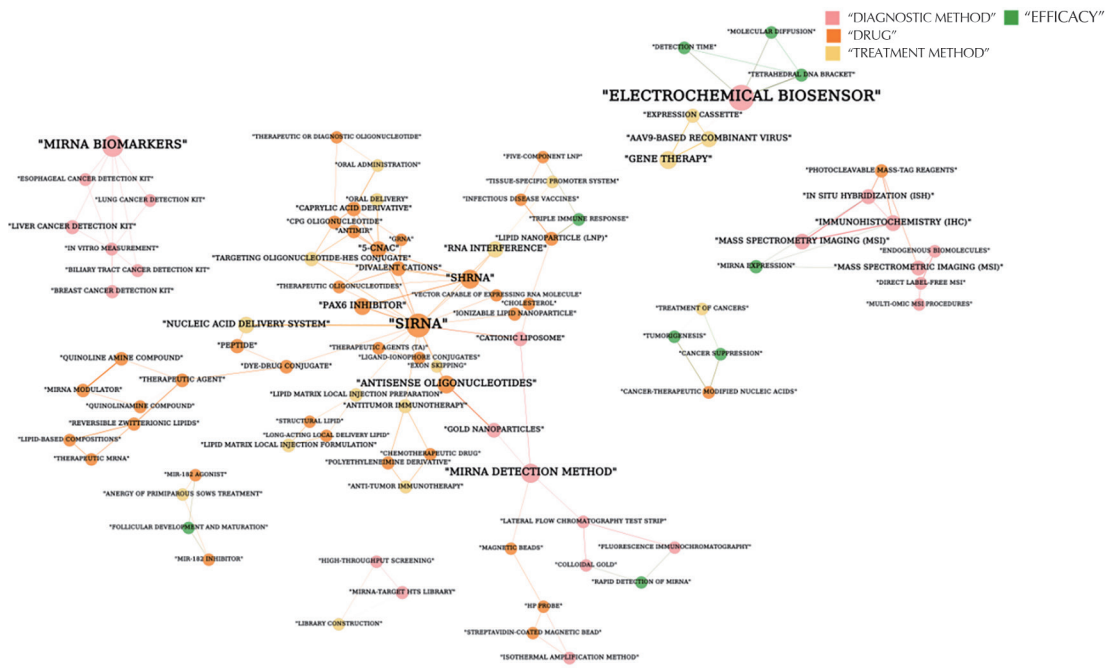


Figure 9. Intervention-Efficacy core subgraph.

For drug and therapeutic interventions, there are fewer functional connections, but they are still present. Anticancer-modified nucleic acids are linked to cancer suppression and tumorigenesis, whereas lipid nanoparticles are associated with immune modulation, including a triple immune response. In animal health research, a miR-182 inhibitor was connected to follicular development in primiparous sows, highlighting links between intervention strategies and reproductive outcomes. Further details are provided in Supplementary Information S9.

3.2.4 Discovery 4: Molecular links to efficacy endpoints

Figure 10 shows the 3-core subgraphs screened via the K-Core algorithm, demonstrating the connections between biological categories (miRNA subtypes, proteins, genes) and the efficacy category. Among all efficacy endpoints, those related to hair health, including hair growth, melanogenesis, outer root sheath development, and keratinocyte proliferation, appeared most frequently.

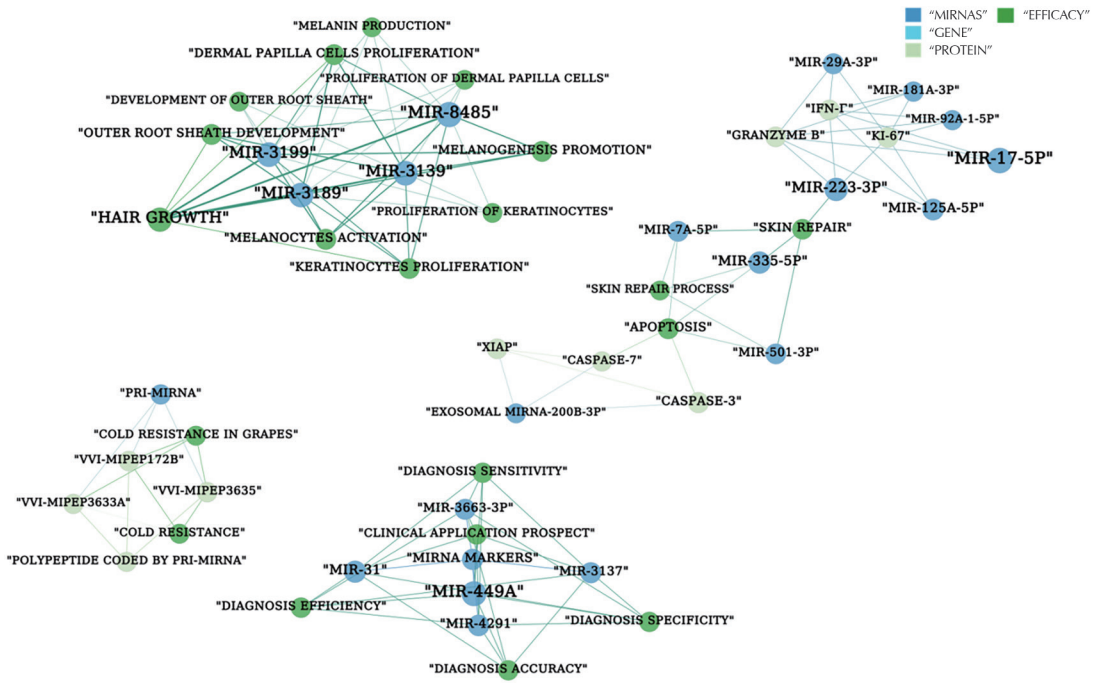


Figure 10. Efficacy-biomolecular core subgraph.

Specific miRNA subtypes, such as miR-223-3p, miR-7a-5p, miR-335-5p, and miR-501-3p, are associated with skin repair functions, whereas CASPASE-7, CASPASE-3, miR-501-3p, and miR-335-5p are linked to apoptosis-related processes. These associations highlight the molecular connections among miRNA subtypes, proteins, and cell life cycle regulation.

In the plant research subgraph, polypeptides such as VVI-MIPEP172B, VVI-MIPEP3633A, and VVI-MIPEP3635 were connected with enhanced cold resistance in grapes, reflecting cross-domain applications of biomolecular efficacy. Further details are provided in Supplementary Information S10.

4. DISCUSSION AND KEY FINDINGS

This study integrated global invention patent data related to miRNAs from 2021–2023 and constructed a miRNA knowledge graph, providing an innovative perspective and analytical framework for biomedical research and drug development. The graph captures multiple knowledge dimensions, including biological entities (e.g., miRNA subtypes, proteins, genes), pathological elements (e.g., diseases, symptoms), intervention strategies (e.g., drugs, diagnostic and therapeutic methods), and efficacy indicators. By

systematically analyzing the internal relationships across these domains, intricate interconnections between miRNA function, disease mechanisms, and clinical applications can be revealed. These findings have significant implications for precision diagnosis and personalized treatment, and offer a hypothesis-generation resource for drug repurposing while providing multidimensional knowledge support for biomedical researchers and clinicians.

The findings suggest that, in the constructed patent knowledge graph from 2021–2023, lung cancer appears most frequently in disease treatment applications. This finding is highly consistent with global disease burden data [32], underscoring the central importance of lung cancer—a malignancy with high incidence and mortality rates—in medical research [33]. Breast cancer ranks second in frequency, aligning with its status as the most prevalent malignant tumor among women [34-35], which is consistent with recent epidemiological evidence. Notably, non-small cell lung cancer (NSCLC), the primary type of lung cancer [36], is frequently represented in the patent-derived graph. A comparison of the NSCLC edge table (Figure 5) with recent studies from 2024–2025 revealed good consistency with current biomedical evidence. In terms of therapeutics, anlotinib, a receptor tyrosine kinase inhibitor, is commonly used for advanced NSCLC treatment [37-38]. Diagnostic advances, including changes in the expression of hsa-let-7b and hsa-miR-25, which have been investigated as potential biomarkers for NSCLC, have also been highlighted [39-40]. Additionally, miR-196a2T, a single-nucleotide polymorphism (SNP) variant of miR-196a2, was identified as a potential biomarker [41]. For immunotherapy, several miRNAs—miR-138-5p, miR-17-5p, miR-197-3p, and miR-21-5p—are frequently mentioned as exosomal markers that are predictive of treatment response [42-45]. These findings align with those of subsequent studies, supporting the relevance of the patent-derived graph for the discovery of NSCLC-related knowledge.

In terms of the functional analysis of miRNA subtypes, the miR-21 family is particularly notable. Among them, miR-21 was the most frequently occurring subtype and showed high expression levels in this study. miR-21 is widely involved in various biological processes and disease diagnoses, including serving as a biomarker for breast cancer, promoting corneal epithelial repair, and playing a crucial role in the diagnosis of multiple diseases. There are diverse detection techniques, such as the use of gold nanocluster carriers and quantum dot sensors, which enable highly sensitive and specific detection. miR-21-5p, the mature form of miR-21, has demonstrated extensive regulatory capabilities in multiple pathological processes and has played an important role in predicting pregnancy complications, the prognosis of ALS and FTD, cancer detection, the risk assessment of heart failure, the identification of a marker for lung diseases, and the diagnosis of breast cancer.

Among the pathological intervention methods, drug intervention has received relatively high attention because of its direct effects. This study revealed that miRNA-485 inhibitors are frequently mentioned in patents for diverse disease contexts and are also associated with hair health applications, suggesting exploratory potential in this nondisease area. By regulating the expression of hair follicle-related genes, it promotes melanin production and the proliferation of hair follicle cells, which not only helps prevent hair loss but also promotes hair regeneration and significantly reduces hair graying [46]. Similarly, the

miRNA-328 antagonist is used to treat ocular surface injuries, infections, dry eye diseases, and other eye diseases [47]. This information, which is difficult to obtain from traditional electronic medical records and static databases, can be systematically uncovered using the knowledge graph constructed in this study, and the related knowledge discoveries have been verified and supported by the latest literature.

In the field of diagnosis, miRNAs have great potential as biomarkers. The knowledge graph reveals that specific miRNA subtypes, such as miR-144-3p, miR-451a, and miR-363-3P, can be used as biomarkers for the diagnosis of colorectal cancer [48-51]; subtypes, such as miR-122, miR-29c-3p, and miR-8485, can be used as biomarkers for the diagnosis of diabetic neuropathy [52-53].

In addition, the study also revealed the relevant efficacies of biological categories and intervention categories in different fields. During the regulation of pathological processes, specific miRNA subtypes play key regulatory roles in functions such as hair growth, immune enhancement, melanocyte activation, apoptosis, and skin repair. In the field of animals, they can be used for the rapid detection of early pregnancy in sows; in the field of plants, they can be used for the identification of cold resistance and heat resistance, providing a theoretical basis for related research.

In conclusion, this study constructed a miRNA knowledge graph from patent data, highlighting the regulatory roles of specific miRNAs in multiple diseases and their potential as biomarkers. The graph highlights key miRNA subtypes in prevalent diseases such as lung and breast cancer, offering insights into their applications in diagnosis and treatment. By integrating drug interventions and diagnostic methods, the graph demonstrates the added value of patent data for capturing emerging biomedical knowledge, providing a real-time and multidimensional resource that complements conventional databases and supports ongoing research in precision medicine. Technically, the extracted graph data is exported in standard GraphML format (as detailed in Supplementary Information S4), which supports modular data integration. This allows researchers to periodically process newly published patents and merge the incremental outputs into the existing network using visualization tools like Gephi (Supplementary Information S5). This interoperability ensures that the knowledge graph remains a dynamic, evolving repository capable of continuously tracking the latest technological trajectories.

Beyond the biological findings, this work also makes methodological contributions: (1) it demonstrates the feasibility of applying the GraphRAG framework to long and complex patent documents, showing that large-context prompts combined with expert validation can yield reliable biomedical knowledge graphs; (2) it shows that patent-derived graphs can reveal earlier translational signals and enterprise-driven innovations often absent in literature- or database-based KGs; and (3) it proposes a cross-layer analytical framework (biomolecules–pathology–interventions–efficacy) that is transferable to other biomedical domains. These contributions highlight not only the biological insights but also the broader methodological and practical value of patent-driven GraphRAG approaches.

5. CONCLUSION

This study demonstrates the feasibility of constructing a miRNA knowledge graph from patent data, revealing clinically meaningful miRNA–disease–drug associations and emerging biomarkers. The findings complement conventional literature-based graphs, highlighting patents as a dynamic and temporally sensitive data source.

Beyond the miRNA domain, the approach provides a transferable framework for other biomedical areas where innovation signals are embedded in patent disclosures. By integrating 2,937 patents into a graph of 10,560 nodes and 11,754 edges, the study captured entities spanning biological, pathological, interventional, and efficacy domains. The knowledge graph revealed subtype-specific biomarkers (e.g., miR-21-5p in breast cancer; miR-138-5p, miR-17-5p, and miR-197-3p in immunotherapy prediction) and clinically relevant associations (e.g., miR-485 $\rightarrow$ NSCLC $\rightarrow$ anlotinib), many of which have been supported by recent biomedical literature, underscoring the graph's real-world relevance.

The methodology provides a practical and extensible framework for biomedical informatics. By enabling periodically updatable and semantically enriched knowledge modeling, it provides a foundation for advancing precision medicine and fostering interdisciplinary innovation while also highlighting the potential of patent-derived knowledge graphs in biomedical discovery.

Nevertheless, several limitations should be acknowledged. First, the dataset was restricted to invention patents published between 2021 and 2023, which may omit earlier patents or the latest developments. Second, patent documents differ from scientific literature by employing specific legal jargon and broad defensive descriptions, which can introduce semantic ambiguity. Although this study mitigated such risks by specifically targeting patent abstracts—which contain denser technical facts—and using refined prompting to filter non-technical phrasing, some degree of semantic noise in entity recognition may persist. This is consistent with our quantitative evaluation (micro-averaged Precision/Recall/F1 $\approx$ 0.56/0.86/0.68): recall is relatively high because the prompts favor exhaustive coverage across nine entity types, whereas precision is lower due to boundary ambiguity, synonym/variant surface forms, and occasional inclusion of broadly phrased method-related expressions common in patent-style descriptions; consequently, F1 reflects a trade-off between coverage and specificity. Third, the knowledge graph has not yet been directly integrated with clinical datasets or validated experimentally, which constrains its translational applicability. Future work will explore integrating patent-derived graphs with literature-based resources, incorporating multimodal evidence, and extending the temporal scope of patent analysis to better capture emerging developments.

DATA AVAILABILITY STATEMENT

The data resources and supplementary materials supporting this study are openly available in ScienceDB at https://doi.org/10.57760/sciencedb.32205 under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

AUTHOR CONTRIBUTIONS

Yanyan Ma contributed to conceptualization, methodology, investigation, and writing—review & editing. Sha Dong contributed to investigation, writing—original draft, review & editing, resources, project administration, and funding acquisition. Jiong Ma was involved in the investigation, data curation, and writing of the original draft. Xuejun Chen contributed resources; curated the data; and wrote, reviewed, and edited the manuscript. Gaowei Mao contributed to data curation, software support, and writing—review & editing. Long Qian contributed to the software, investigation, data curation, and formal analysis. Yuehua Sheng contributed to conceptualization, methodology, and writing—review & editing. Wei Zheng contributed to the methodology, data curation, and formal analysis. Tingyi Lu contributed to the validation, formal analysis, and visualization. Junbo Xu contributed to the software, original draft preparation, and visualization. All the authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by the China National Intellectual Property Administration (CNIPA) through a research project entitled "Study on Intellectual Property Protection of Pharmaceuticals" (Project Number: SS24-A-21); by the Key Project of Ningbo Municipal Soft Science Research Program "Research on Innovation Paths and Countermeasures in Key Areas of People's Livelihood Science and Technology in Ningbo" (Project Number: 2025R14); by the Zhejiang Higher Education Society Project "Knowledge Graph Construction and Large Language Model Applications for Women's Health Care Education in Zhejiang Universities" (Project Number: KT2025354); and by the Zhejiang Provincial Center for Disease Control and Prevention (CDC) Project "Construction and Application of a Large Language Model–Based Knowledge System for Human Papillomavirus (HPV) Prevention and Control in Universities" (Project Number: 2025JK217).

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

REFERENCES

[1] C. L. Holley and V. K. Topkara, "An introduction to small noncoding RNAs: miRNA and snoRNA," Cardiovasc. Drugs Ther., vol. 25, pp. 151–159, 2011.

[2] X. Jia, X. Wang, X. Guo, et al., "MicroRNA-124: an emerging therapeutic target in cancer," Cancer Med., vol. 8, no. 12, pp. 5638–5650, 2019.

[3] M. Carleton, M. A. Cleary, and P. S. Linsley, "MicroRNAs and cell cycle regulation," Cell Cycle, vol. 6, no. 17, pp. 2127–2132, 2007.

[4] M. M. J. Mens and M. Ghanbari, "Cell cycle regulation of stem cells by microRNAs," Stem Cell Rev. Rep., vol. 14, pp. 309–322, 2018.

[5] J. Tang, A. Ahmad, and F. H. Sarkar, "The role of microRNAs in breast cancer migration, invasion and metastasis," Int. J. Mol. Sci., vol. 13, no. 10, pp. 13414–13437, 2012.

[6] Q. Huang, K. Gumireddy, M. Schrier, et al., "The microRNAs miR-373 and miR-520c promote tumor invasion and metastasis," Nat. Cell Biol., vol. 10, no. 2, pp. 202–210, 2008.
[7] R. Conrad, M. Barrier, and L. P. Ford, "Role of miRNA and miRNA processing factors in development and disease," Birth Defects Res. C Embryo Today Rev., vol. 78, no. 2, pp. 107–117, 2006.
[8] T. D. Resnick, K. A. McCulloch, and A. E. Rougvie, "miRNAs give worms the time of their lives: small RNAs and temporal control in Caenorhabditis elegans," Dev. Dyn., vol. 239, no. 5, pp. 1489, 2010.
[9] G. A. Calin and C. M. Croce, "MicroRNA signatures in human cancers," Nat. Rev. Cancer, vol. 6, no. 11, pp. 857–866, 2006.
[10] E. Tili, J. J. Michaille, V. Gandhi, et al., "miRNAs and their potential for use against cancer and other diseases," Future Oncol., vol. 3, no. 5, pp. 521–537, 2007.
[11] H. H. Wu, W. Lin, and K. W. Tsai, "Advances in molecular biomarkers for gastric cancer: miRNAs as emerging novel cancer markers," Expert Rev. Mol. Med., vol. 16, Art. no. e1, 2014.
[12] H. Wei, Y. Xu, W. Xu, et al., "Serum exosomal miR-223 serves as a potential diagnostic and prognostic biomarker for dementia," Neuroscience, vol. 379, pp. 167–176, 2018, doi: 10.1016/j.neuroscience.2018.03.016.
[13] P. T. B. Ho, I. M. Clark, and T. T. L. Le, "MicroRNA-based diagnosis and therapy," Int. J. Mol. Sci., vol. 23, no. 13, Art. no. 7167, 2022.
[14] A. Rodríguez-Martínez, D. de Miguel-Pérez, F. G. Ortega, et al., "Exosomal miRNA profile as complementary tool in the diagnostic and prediction of treatment response in localized breast cancer under neoadjuvant chemotherapy," Breast Cancer Res., vol. 21, pp. 1–9, 2019.
[15] H. Ueda, H. Takahashi, R. Sakaniwa, et al., "Preoperative treatment response prediction for pancreatic cancer by multiple microRNAs in plasma exosomes: Optimization using machine learning and network analysis," Pancreatology, vol. 24, no. 7, pp. 1097–1106, 2024.
[16] H. Chi, H. Chen, R. Wang, et al., "Proposing new early detection indicators for pancreatic cancer: Combining machine learning and neural networks for serum miRNA-based diagnostic model," Front. Oncol., vol. 13, Art. no. 1244578, 2023.
[17] H. Jamialahmadi, A. Asadnia, G. Khalili-Tanha, et al., "Identification of miR-20a as a diagnostic and prognostic biomarker in colorectal cancer: MicroRNA sequencing and machine learning analysis," MicroRNA, vol. 14, no. 1, pp. 73–91, 2025.
[18] M. Sherafatian, "Tree-based machine learning algorithms identified minimal set of miRNA biomarkers for breast cancer diagnosis and molecular subtyping," Gene, vol. 677, pp. 111–118, 2018.
[19] X. Bai, S. He, Y. Li, et al., "Construction of a knowledge graph for framework material enabled by large language models and its application," npj Comput. Mater., vol. 11, no. 1, Art. no. 51, 2025.
[20] D. N. Nicholson and C. S. Greene, "Constructing knowledge graphs and their biomedical applications," Comput. Struct. Biotechnol. J., vol. 18, pp. 1414–1428, 2020.
[21] M. Fankhauser, C. Moser, and T. Nyfeler, "Patents as early indicators of technology and investment trends: analyzing the microbiome space as a case study," Front. Bioeng. Biotechnol., vol. 6, Art. no. 84, 2018.
[22] D. Edge et al., "From local to global: A Graph RAG approach to query-focused summarization," arXiv:2404.16130, 2024.
[23] O. Sabo, Y. Elazar, Y. Goldberg, et al., "Revisiting few-shot relation classification: Evaluation data and classification schemes," Trans. Assoc. Comput. Linguistics, vol. 9, pp. 691–706, 2021.
[24] Y. Li, "A practical survey on zero-shot prompt design for in-context learning," 2023, arXiv:2309.13205.
[25] J. D'souza and S. Auer, "Sentence, phrase, and triple annotations to build a knowledge graph of natural language processing contributions—a trial dataset," J. Data Inf. Sci., vol. 6, no. 3, pp. 6–34, 2021.

[26] K. R. Chowdhary, "Natural language processing," in Fundamentals of Artificial Intelligence, pp. 603–649, 2020.
[27] J. H. Patil, P. Potikas, W. B. Andreopoulos, et al., "Community detection using deep learning: Combining variational graph autoencoders with Leiden and K-Truss techniques," Information, vol. 15, no. 9, Art. no. 568, 2024.
[28] M. A. M. Faysal and S. Arifuzzaman, "A comparative analysis of large-scale network visualization tools," in Proc. 2018 IEEE Int. Conf. Big Data (Big Data), 2018, pp. 4837–4843.
[29] A. Sapountzi and K. E. Psannis, "Social networking data analysis tools & challenges," Future Gener. Comput. Syst., vol. 86, pp. 893–913, 2018.
[30] Y. X. Kong, G. Y. Shi, R. J. Wu, et al., "k-core: Theories and applications," Phys. Rep., vol. 832, pp. 1–32, 2019.
[31] W. Khaouid, M. Barsky, V. Srinivasan, et al., "K-core decomposition of large networks on a single PC," Proc. VLDB Endow., vol. 9, no. 1, pp. 13–23, 2015.
[32] A. Leiter, R. R. Veluswamy, and J. P. Wisnivesky, "The global burden of lung cancer: current status and future trends," Nat. Rev. Clin. Oncol., vol. 20, no. 9, pp. 624–639, 2023.
[33] A. Lahiri, A. Maji, P. D. Potdar, et al., "Lung cancer immunotherapy: progress, pitfalls, and promises," Mol. Cancer, vol. 22, no. 1, Art. no. 40, 2023.
[34] K. M. Cuthrell and N. Tzenios, "Breast cancer: Updated and deep insights," Int. Res. J. Oncol., vol. 6, no. 1, pp. 104–118, 2023.
[35] N. Florez, L. Kiel, I. Riano, et al., "Lung cancer in women: the past, present, and future," Clin. Lung Cancer, vol. 25, no. 1, pp. 1–8, 2024.
[36] Y. Alduais, H. Zhang, F. Fan, et al., "Non-small cell lung cancer (NSCLC): a review of risk factors, diagnosis, and treatment," Medicine, vol. 102, no. 8, Art. no. e32899, 2023.
[37] J. Wei, Y. Zhang, Y. Zheng, et al., "Efficacy and safety of anlotinib monotherapy or combination therapy in the treatment of patients with advanced non-small cell lung cancer: a retrospective real-world study conducted in East China," BMC Pulm. Med., vol. 25, Art. no. 170, 2025.
[38] Z. Jiang, X. Sun, Y. Li, et al., "Anlotinib induced ferroptosis through the p53/xCT/GPX4 pathway in non-small cell lung cancer," Transl. Oncol., vol. 53, Art. no. 102289, 2025.
[39] J. Lu, H. Zhong, T. Chu, et al., "MEK inhibitor combined with RTK inhibitor (trametinib plus anlotinib) in non-G12C KRAS-mutant non-small cell lung cancer: A new strategy," J. Clin. Oncol., vol. 42, no. 16, pp. 8542–8542, 2024.
[40] A. A. I. Hassanin and K. S. Ramos, "Circulating exosomal miRNA profiles in non-small cell lung cancers," Cells, vol. 13, no. 18, Art. no. 1562, 2024, doi: 10.3390/cells13181562.
[41] F. Su, W. Zhao, F. Zhao, et al., "Pyrococcus furiosus Argonaute-based fluorometric biosensor for one-tube detection of cancer-associated single nucleotide polymorphisms in microRNAs," Anal. Chem., vol. 97, no. 8, pp. 4678–4686, 2025.
[42] F. Rostami, Z. Tavakol Hamedani, A. Sadoughi, et al., "PDL1 targeting by miR-138-5p amplifies anti-tumor immunity and Jurkat cells survival in non-small cell lung cancer," Sci. Rep., vol. 14, no. 1, Art. no. 13542, 2024.
[43] D. Mukhopadhyay, P. Cocco, S. Orrù, et al., "The role of microRNAs as early biomarkers of asbestos-related lung cancer: A systematic review and meta-analysis," Pulmonology, vol. 31, no. 1, Art. no. 2416792, 2025.
[44] G. Zheng, H. Ye, J. Bai, et al., "Downregulation of lncRNA MIR17HG reduced tumorigenicity and Treg-mediated immune escape of non-small-cell lung cancer cells through targeting the miR-17-5p/RUNX3 axis," J. Biochem. Mol. Toxicol., vol. 38, no. 5, Art. no. e23715, 2024.

[45] R. Q. da Silva Lima, C. F. M. Vasconcelos, J. P. A. Gomes, et al., "miRNA-21, an oncomiR that regulates cell proliferation, migration, invasion and therapy response in lung cancer," Pathology–Research and Practice, vol. 263, Art. no. 155601, 2024.

[46] S. Begum, M. J. Hossain, I. Kim, et al., "Modulating immune responses in alopecia: therapeutic insights and potential targets of antisense oligonucleotides," BMC Immunol., vol. 26, no. 1, Art. no. 26, 2025.

[47] J. L. Chen, W. Y. Lai, R. T. Lin, et al., "Safety and tolerability of anti-microRNA-328 ophthalmic solution, SHJ002, in pediatric subjects: First-in-human clinical study," Clin. Ther., vol. 46, no. 10, pp. 768–772, 2024.

[48] M. Kalimutho, G. Del Vecchio Blanco, S. Di Cecilia, et al., "Differential expression of miR-144* as a novel fecal-based diagnostic marker for colorectal cancer," J. Gastroenterol., vol. 46, pp. 1391–1402, 2011.

[49] Z. Zhang, D. Zhang, Y. Cui, et al., "Identification of microRNA-451a as a novel circulating biomarker for colorectal cancer diagnosis," Biomed Res. Int., vol. 2020, no. 1, Art. no. 5236236, 2020.

[50] O. Pös, J. Styk, G. Buglyó, et al., "Cross-kingdom interaction of miRNAs and gut microbiota with noninvasive diagnostic and therapeutic implications in colorectal cancer," Int. J. Mol. Sci., vol. 24, no. 13, Art. no. 10520, 2023.

[51] X. Zhou, Y. Li, L. Wu, et al., "Upregulated lncRNA LINC01128 in colorectal cancer accelerates cell growth and predicts malignant prognosis through sponging miR-363-3p," J. Cancer Res. Clin. Oncol., vol. 150, no. 5, Art. no. 276, 2024.

[52] N. Pastukh, A. Meerson, D. Kalish, et al., "Serum miR-122 levels correlate with diabetic retinopathy," Clin. Exp. Med., vol. 19, pp. 255–260, 2019.

[53] S. Wang, P. Yi, N. Wang, et al., "LncRNA TUG1/miR-29c-3p/SIRT1 axis regulates endoplasmic reticulum stress-mediated renal epithelial cells injury in diabetic nephropathy model in vitro," PLoS One, vol. 16, no. 6, Art. no. e0252761, 2021.