International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 599-612

https://doi.org/10.70517/ijhsa46454

Time-series analysis of AKR1B10 gene expression pattern in hepatocellular carcinoma patients combined with LSTM neural network

Xiaoran Li[1] and Changlin Ma[2,*]
[1] Jining First People's Hospital Affiliated to Shandong First Medical University, Shandong First Medical University, Jinan, Shandong, 250117, China
[2] Jining First People's Hospital, Shandong University, Jinan, Shandong, 250100, China
Corresponding authors: (e-mail: machanglin80@sina.com).

Abstract Hepatocellular carcinoma (HCC) is a common and fatal malignant tumor worldwide, which is difficult to diagnose early and has a low survival rate. In this paper, we investigated the temporal expression pattern of AKR1B10 gene in patients with hepatocellular carcinoma (HCC) and constructed a GL-TGRN model using LSTM neural network combined with gene regulatory network. By analyzing multiple sets of student data (including mRNA expression data, miRNA data, and DNA methylation data), we conducted an in-depth exploration of the temporal expression of the AKR1B10 gene and its relationship with hepatocellular carcinoma development. The results showed that the GL-TGRN model performed excellently in inferring AKR1B10 gene expression, and the AUROC and AUPR values were increased by 26.23% and 35.69%, which were significantly higher than the comparison methods (e.g., GC-SIN and JUMP). In addition, through differential expression analysis, we screened 786 differential genes and 14 miRNAs related to hepatocellular carcinoma, and these molecules are closely related to hepatocarcinogenesis. Ablation experiments demonstrated that the fusion of multi-omics features in the GL-TGRN model significantly improved the accuracy of gene regulatory inference. This paper provides new data support for early diagnosis and personalized treatment of hepatocellular carcinoma.

Index Terms Hepatocellular carcinoma, AKR1B10 gene, gene regulatory network, LSTM neural network, temporal expression analysis, multi-group biomarker data

I. Introduction

Hepatocellular carcinoma is the most common primary liver malignancy and is the leading cause of cancer-related deaths worldwide [1]. As of 2024, 865,000 people have been diagnosed with liver cancer worldwide. Importantly, the main causes of hepatocellular carcinoma are hepatitis B and C infections, alcoholic liver disease, and the incidence of hepatocellular carcinoma will continue to rise as the incidence of clinical nonalcoholic steatohepatitis-associated cirrhosis tends to increase with dietary structure and standard of living [2]-[5]. Hepatocellular carcinoma has a very poor prognosis due to its advanced stage at initial diagnosis, high rate of postoperative recurrence and metastasis as well as multidrug resistance [6], [7]. Currently, the treatment of choice for hepatocellular carcinoma remains surgical, but only less than 20% have access to surgical resection [8]. Despite extensive research on the molecular mechanisms of hepatocellular carcinogenesis, knowledge of the genetic alterations that lead to the development and progression of hepatocellular carcinoma remains fragmented. The key drivers of its carcinogenesis remain unknown, which limits the development of targeted therapies for hepatocellular carcinoma [9]. In addition, the overlap of the most notable dysregulated genes in multiple studies is very low, and the inconsistency of the results is due to a variety of factors such as measurement errors, small sample sizes, and different statistical methods [10], [11].

The TCGA database discloses a large number of clinical and molecular phenotypes of several tumor patients, and also lists several genes with significant temporal expression patterns. one of these genes is the AKR1B10 gene, which is a novel protein for identifying human hepatocellular carcinoma, and whose dynamic changes are of value in the early diagnosis of hepatocellular carcinoma, and can inhibit the proliferation and metastasis of cancer cells by regulating related signaling pathways [12]-[14] The high expression of AKR1B10 in hepatocellular carcinoma tissues enhances cellular resistance to aldehyde-containing antitumor drugs, such as zorubicin and mitomycin, which may be due to the fact that these drugs are similar to the AKR1B10 substrates, which are cytotoxic compounds containing carbocytes, aromatic compounds containing aldehydes, and aliphatic compounds, to which AKR1B10 has a powerful reducing and detoxifying function, and thus resistance arises [15]-[18]. AKR is induced in rat hepatocellular carcinoma and it is thought that it may have a crucial detoxification role for harmful metabolites

produced by rapidly growing cancer cells [19], [20]. A protein called Spot17, which is highly homologous to rat AKR, is induced by rat hepatocellular carcinoma, and approximately 29% of hepatocellular carcinomas overexpress AKR, and about 54% of this is overexpressed AKR1B10 [21]. However, AKR1B10 gene expression inflection point detection has a continuous and intensive sampling requirement, and traditional qPCR technical means are costly [22]. Meanwhile, due to the small sample size of liver cancer studies, conventional machine learning models have low adaptability and face high bias in nonlinear time-series prediction results and dynamic detection [23]. And the LSTM neural network model has the advantages of small sample size, long period, and feature capture, which breaks through the above bottleneck [24].

Hepatocellular carcinoma (HCC) is a malignant tumor with high morbidity and mortality worldwide, and its etiology is diverse, including hepatitis virus infection, aflatoxin exposure, and chronic alcohol abuse. Although the diagnosis and treatment of hepatocellular carcinoma have advanced with the development of medical technology, early symptoms of hepatocellular carcinoma are not obvious and diagnosis is often lagging behind, resulting in the majority of patients being in the advanced stage at the time of diagnosis, which makes treatment more difficult. Therefore, how to realize early screening and diagnosis of liver cancer has become the key to improve the survival rate.

Currently, commonly used clinical screening methods for liver cancer include serum alpha-fetoprotein (AFP) test, imaging examination and liver function test. However, these methods have some limitations, especially in AFP-negative or low value patients, with low diagnostic accuracy. Therefore, finding new biomarkers for the early diagnosis of hepatocellular carcinoma is of great clinical value.The AKR1B10 gene is an enzyme in the aldo-keto reductase family, and studies have shown that it exhibits a high level of expression in a variety of tumors, and especially in hepatocellular carcinoma, the level of expression of AKR1B10 is significantly increased, suggesting that it may play an important role in hepatocarcinogenesis and development.

In this study, a gene regulatory network-based analysis model (GL-TGRN) was constructed to infer the temporal expression pattern of AKR1B10 gene in hepatocellular carcinoma patients by combining multi-omics data, including mRNA, miRNA and DNA methylation data, and LSTM neural network.LSTM neural network can effectively capture the time-dependence in gene expression, and combined with the graph neural network's topological features, provided more precise analysis results. Through the application of this model, we are able to further understand the complexity of gene regulation in hepatocellular carcinoma patients on the basis of multi-omics data, mine the gene features closely related to hepatocarcinogenesis, and provide new clues for early diagnosis.

II. Theoretical basis of AKR1B10 gene regulatory network

Hepatocellular carcinoma (HCC) is the most common primary liver cancer (referred to as hepatocellular carcinoma), which seriously jeopardizes human health. The main causative factors of primary liver cancer include chronic infection with hepatitis viruses (HBV, HCV), aflatoxin exposure, alcoholism, etc., among which liver cancer caused by non-alcoholic steatohepatitis is showing a significant rising trend. In recent years, despite advances in treatment, the overall survival rate of hepatocellular carcinoma is still far from satisfactory. In addition to the malignant characteristics of the tumor, liver cancer often loses the chance of treatment when it is detected due to its insidious onset, which is also an important factor contributing to the poor survival rate of liver cancer patients. Therefore, how to detect liver cancer at an early stage has been an important challenge for the medical community.

II. A.AKR1B10 gene in hepatocellular carcinoma patients

II. A. 1) Common liver cancer serum markers

(1) Protein markers

Alpha-fetoprotein (AFP) is located on chromosome 4, arm q (4q25), and is a member of the albumin-like gene superfamily.AFP is one of the most abundant proteins in fetal plasma and is synthesized in large quantities by the fetal yolk sac and liver, and begins to decline gradually in the second and third trimesters of pregnancy, but at the same time albumin begins to increase. This phenomenon often leads to AFP being viewed as the albumin of the fetus, and studies have also found a high degree of homology between the sequences encoding the two proteins. In normal individuals, AFP usually decreases to normal levels by 9 to 11 months of age, but it rises again in the presence of liver disease, especially in the presence of primary tumors of the liver, germ cell tumors, or benign liver disease [25].

De-gamma-carboxy prothrombinogen (DCP) is an abnormal thrombospondin protein that is highly expressed in the serum of HCC patients. Studies have shown that DCP is an effective diagnostic marker for hepatocellular carcinoma, and its level is related to the prognosis of the patients. The diagnosis of hepatocellular carcinoma is more accurate when AFP is used in combination with DCP than when it is used alone, and the abnormal elevation of DCP due to a number of factors may interfere with the diagnosis of the disease [26].

(2) Liquid biopsy

Liquid biopsy is the detection of circulating tumor cells, mainly including cell free DNA (cfDNA), extracellular vesicles and other markers. cfDNA refers to the DNA molecules that are released into the body fluids by the cells undergoing apoptosis, necrosis, or active secretion, which are presented as a double-stranded fragment with a length of about 160~240 bp. The cfDNA from tumor cells, i.e., circulating tumor DNA (ctDNA), has even demonstrated its potential and value in reflecting the tumor ontology from different dimensions, and has been widely used in various aspects of tumor diagnosis, treatment, and prognosis. Genetic and epigenetic alterations are important factors driving the occurrence, development and metastasis of HCC. As DNA fragments derived from tumor cells, ctDNA should share the same molecular alterations as the tumor cells from which it originates. The feasibility of detecting HCC-specific genetic or epigenetic alterations in plasma DNA has been confirmed, making it possible to use ctDNA to determine the extent of the patient's hoar tumors as well as genetic information about the primary tumor [27].

II. A. 2) AKR1B10 and liver cancer diagnosis

(1) Structure of AKR1B10

Aldo-keto reductase family 1 member B10 (AKR1B10), whose coding gene is localized in the region of chromosome 7q33, encodes a protein consisting of 316 amino acid residues, and AKR1B10 is a member of the aldo-keto reductase (AKR) superfamily [28]. AKR is an oxido-reductase that uses reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor and catalyzes the conversion of aldehydes and ketones to the corresponding alcohols, the specific process of which is shown in Figure 1.AKR may be involved in different biological processes in different organisms, including carbonyl detoxification, osmoregulation, hormone metabolism, lipid synthesis, diabetic complications, tumorigenesis, and therapy. As a member of AKR, AKR1B10's also has the physiological function of reducing aldehydes and ketones, which can reduce the damage to nucleic acids and proteins by carbonyl compounds in order to reduce cell mutation and tumorigenesis.

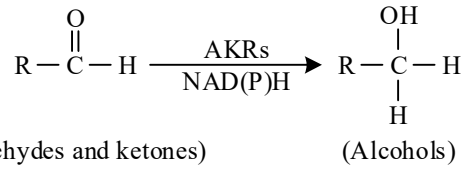


Figure 1: The working principle of AKR enzyme

All AKRs use NADPH as a coenzyme for their catalytic function, and the binding sites of AKR1B10 to coenzyme NADPH are Ser159, Asn160, Gln183 and Lys263, with hydrogen bonding provided by Tyr48 and His110, and Cys299 can serve as a binding site for AKR1B10 inhibitors and play an important role in the regulation of AKR1B10 activity.

(2) Relationship between AKR1B10 and hepatocellular carcinoma diagnosis

In normal humans, AKR1B10 is expressed in the small intestine and colon, with a small amount of expression in thymus, prostate and testis tissues, and negative expression in other human tissues.AKR1B10 is highly expressed in a variety of tumors such as hepatocellular carcinoma, lung carcinoma, breast carcinoma, and colorectal carcinoma, suggesting that it has a correlation with human cancers. Existing studies have found that the expression level of AKR1B10 protein increases in human normal liver, paraneoplastic liver and hepatocellular carcinoma tissues in a sequential order, and the expression rate is higher in hepatocellular carcinoma tissues, especially serum alpha-fetoprotein-negative hepatocellular carcinoma tissues, which suggests that it has a potential clinical application in early diagnosis of hepatocellular carcinoma and assessment of its prognosis. In addition, it has been reported that silencing of AKR1B10 in hepatocellular carcinoma cells by siRNA technology resulted in the down-regulation of proliferation-related genes and up-regulation of pro-apoptotic genes in hepatocellular carcinoma cells, suggesting that AKR1B10 may be involved in the regulation of hepatocellular carcinoma cell proliferation by altering the expression of apoptosis-related genes.

II. B.Multi-group Biological Data and Gene Regulatory Networks

II. B. 1) Multi-group student biology data

Histological data refers to data that are comprehensively measured and analyzed by high-throughput techniques on the composition and function of molecules at multiple levels within an organism. These data provide detailed information about biomolecules such as genes, transcription, proteins and metabolism, as well as their interactions and regulatory relationships. Histomics data include genomics, transcriptomics, proteomics, and metabolomics at

multiple levels.The TCGA database provides a wide range of cancer-related histomics data types. In this paper, we focus on mRNA expression data, miRNA expression data, and DNA methylation data.

(1) mRNA expression data. Gene expression refers to the process by which genes are transcribed into mRNA in cells and then converted into proteins and RNA molecules through the translation process. Genes are units of genetic information in the DNA molecule, and they carry the instructions necessary for the cell to synthesize proteins. The process of gene expression consists of three main steps, namely transcription, RNA processing and translation. The study of gene expression is important for understanding cellular functions, mechanisms of disease occurrence, and the development of drug therapy.

(2) miRNA expression data. miRNAs are a class of short non-coding RNA molecules, about 20 to 25 nucleotides in length. miRNAs regulate gene expression by interacting with the mRNA of target genes. They play important regulatory functions in cells and participate in the regulation of a variety of biological processes, including cell proliferation, differentiation, apoptosis and metabolism.

(3) DNA methylation data.DNA methylation is achieved by the addition of methyl groups to cytosine bases at the CpG site by DNA methyltransferases. This process is reversible and can be replicated by DNA replicase during cell division and replication. Under normal conditions DNA methylation plays an important role in the maintenance of genome stability, regulation of gene expression, gene silencing and cell differentiation etc. Abnormalities in DNA methylation can lead to abnormalities in gene expression and the development of diseases. Abnormal DNA methylation patterns are common in diseases such as cancer, where specific gene regions may be aberrantly methylated or unmethylated, leading to abnormal gene expression and cellular dysfunction.

II. B. 2) Gene regulatory networks

The intensive development of molecular biology has revealed that complex life phenomena are the result of the interaction of gene activities. Gene regulatory networks are the manifestation of life functions at the level of gene expression. The completion of whole genome sequencing in many species and the development of high-throughput experimental techniques have yielded functional genomic data such as genome expression profiles, protein interactions, RNAi analysis of genome expression, and protein-DNA binding, which have made it possible to study gene regulatory networks at the system level. The biggest challenge encountered during the study is to find the location of genes and their products on functional pathways, circuits and networks.

A successful biochemical model should be able to characterize the relationships among the components of a cellular interaction network. These biochemical networks can be constructed from different perspectives to demonstrate the corresponding type of action. Usually one considers three types of biochemical networks:

(1) Metabolic networks, which mainly demonstrate the chemical reaction chains between various metabolic substrates and products in the cell.

(2) Protein networks, which mainly demonstrate the interactions between various proteins in the cell, e.g., the formation of protein complexes and protein modification during signaling.

(3) Gene transcriptional regulatory networks, which demonstrate an abstract interaction relationship among all genes, i.e., the effect of the expression level of one gene on the expression of other genes. Each of these types of networks is only a simplification of the cellular life system.

II. C.Theories related to neural networks

II. C. 1) Graph Neural Network Foundations

A graph neural network is a type of neural network specialized for processing graph-structured data.Many real-world problems can be naturally represented by graphs, including social networks, molecular structures, transportation networks, biological networks, and so on. Graph neural networks can capture the relationships and interactions between entities in these networks. A graph neural network consists of two main parts, information aggregation and feature update, which are included in every graph neural network layer. In the aggregation step, each node collects information from its neighboring nodes, a process often referred to as “message passing”, and the aggregated information from its neighbors can be used to obtain structural features of the graph, which are generally aggregated by summing, taking the mean, or taking the most. The update step consists of stitching together the node's own features and neighborhood messages to update the feature representation of the current node, which is generally accomplished using a neural network or transformation function.

The graph filtering based on the null domain is similar to the traditional convolution, and the node features are updated by aggregating the features of the neighboring nodes and the center node to pass the information. Among them, GraphSAGE and GAT belong to the null-domain based graph filtering methods.GAT introduces the attention mechanism into the process of aggregating the neighbor information of the graph.This method allows the model to measure the importance of each neighboring feature.Compared to taking the average and the maximum, the attention mechanism can better access the information provided by different neighbors through weighted

aggregation. The attention mechanism generally measures the importance of information originating from neighboring nodes by measuring the relevance of neighboring nodes relative to the central node as an attention coefficient between neighboring nodes and the central node. The attention coefficient is usually computed using an inner product, and the nodes are transformed into key features required for downstream tasks by linear mapping before computation to improve the expressiveness of the attention mechanism. Specifically, it is computed as:

$$a_{ij} = \frac{\exp\left(Leaky\,\mathrm{Re}\,LU\left(\vec{a}^{T}\left[W\vec{h}_i \,\|\, W\vec{h}_j\right]\right)\right)}{\sum_{k\in \mathrm{N}_i}\exp\left(Leaky\,\mathrm{Re}\,LU\left(\vec{a}^{T}\left[W\vec{h}_i \,\|\, W\vec{h}_k\right]\right)\right)} \tag{1}$$

$$\vec{h}_{i'} = \sigma\left(\sum_{j\in \mathrm{N}_i} a_{ij} W\vec{h}_j\right) \tag{2}$$

where $\left[W\vec{h}_i \,\|\, W\vec{h}_j\right]$ denotes that node i and node j are improved in expressiveness by the shared parameter matrices, and $\vec{a}^{T}$ is a single-layer feed-forward neural network used to extract the two node vectors of the Key information. a_{ij} denotes the attention weight of node i for node j, which is normalized according to the neighborhood of node i, and the obtained weight is weighted with the corresponding node vectors to obtain the updated node i vectors $\vec{h}_{i'}$.

II. C. 2) Long and short-term memory neural networks

Long Short-Term Memory (LSTM) is a temporal recurrent neural network for processing sequential data.LSTM is a special variant of Recurrent Neural Network (RNN), which has a “gate” structure and is designed to overcome the limitations of RNN in long-term memory.LSTM achieves the filtering of memories by dividing memories into two types: long-term and short-term, which store information important to the model output and the current time-step output, respectively. LSTM can filter memories by dividing them into long-term and short-term, storing information that has an important effect on the output of the model and the output of the current time step respectively.The basic unit of the LSTM neural network is the memory cell, which consists of the forgetting gate, the input gate, and the output gate, and the LSTM is composed of countless memory cells connected in a series.

The role of the forgetting gate is to screen long-term memories. The incoming long-term memory C_{t-1} from the previous memory cell is multiplied by a memory factor f_t with a value domain of $[0,1]$, so that it retains a portion of useful long-term information; and through learning, the memory factor is optimized so that the model can better handle long-time dependent data. The formula for f_t is:

$$f_t = sigmoid\left(W_f \cdot [h_{t-1}, X_t] + b_f\right) \tag{3}$$

where h_{t-1} is the short-term information of the previous memory cell, X_t is the input of the current time step, W_f is the weight matrix of the forgetting gate, and b_f is the bias term of the forgetting gate.

The role of the input gate is to decide how much information at the current time step needs to be added to long-term memory. The information of the current time step consists of h_{t-1} and X_t together, which is multiplied by i_t with a value domain of $[0,1]$ through the nonlinear transformation of $\tanh$. In this way, the information that is most useful for the prediction of the current time step is filtered out and added to the long-term memory to obtain a new long-term memory C_t. The formulae for i_t and $\tilde{C}_t$ are:

$$i_t = sigmoid\left(W_i \cdot [h_{t-1}, X_t] + b_i\right) \tag{4}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, X_t] + b_C\right) \tag{5}$$

where W_i is the weight matrix of i_t, W_C is the weight matrix of $\tilde{C}_t$, b_i is the bias term of i_t, and b_C is the bias term of $\tilde{C}_t$.

The role of the output gate is to filter the most effective information for the current time step from the long-term memory for the prediction of the current time step. The long-term memory C_t of the current time step is nonlinearly

transformed and multiplied by o_t with value domain $[0,1]$ to get h_t as the output of the current time step. The formulae for C_t, o_t, and h_t are:

$$C_t = C_{t-1} \cdot f_t + \tilde{C}_t \cdot i_t \tag{6}$$

$$o_t = sigmoid\left(W_o \cdot [h_{t-1}, X_t] + b_o\right) \tag{7}$$

$$h_t = \tanh(C_t) \cdot o_t \tag{8}$$

where W_o is the weight matrix of o_t and b_o is the bias term of o_t.

The LSTM with the introduction of gating mechanism has better nonlinear fitting ability and is able to handle long-term dependencies. However, its computational resource consumption is large, and gradient explosion is still possible if the sequence is too long.

III. Gene regulatory networks based on integrated neural networks

Hepatocellular carcinoma is one of the common malignant tumors, characterized by high degree of malignancy, rapid progression, poor prognosis and high mortality rate, and early detection, early diagnosis and early treatment are the keys to improve the therapeutic effect of hepatocellular carcinoma. At present, the diagnosis of this disease mainly relies on serum AFP, AFU indexes and imaging examination, but some scholars pointed out that some hepatocellular carcinoma patients have negative or low detection values of serum AFP and AFU. In order to explore the significance of AKR1B10 protein in the early diagnosis of hepatocellular carcinoma, this paper analyzed and compared the expression of AKR1B10 protein in hepatocellular carcinoma tissues and tissues of benign hepatic lesions, and analyzed the relationship between the positive expression rate of AKR1B10 protein in hepatocellular carcinoma tissues and patients' preoperative serum AFP and AFU levels.

III. A. Hepatocellular carcinoma AKR1B10 gene data processing and characterization

III. A. 1) AKR1B10 gene data processing

In this paper, the time-sequence expression analysis of AKR1B10 gene in hepatocellular carcinoma was mainly carried out on the basis of multi-omics student data. However, before reusing the multi-omics data to do gene time-sequence expression analysis, it is necessary to preprocess a large amount of raw data generated by the experimental group, the specific steps of which are as follows:

(1) Quality control is the first step in analyzing multi-omics biology data, and is used to exclude barcodes that are unlikely to represent intact cells, including the number of counts (count depth), the number of genes, and the proportion of mitochondrial gene counts for each barcode, as well as to filter out anomalous peaks that may correspond to dead cells, cells with ruptured cell membranes, or two cells through a threshold value.

(2) Normalization is performed to ameliorate differences due to uneven sequencing depths to obtain the correct relative gene expression abundance between cells. This effect is more pronounced in multiple sets of biomass data, as the RNA abundance of each cell may be significantly different due to the cell cycle or other biological factors.

(3) The purpose of data correction is to further remove differences caused by nonbiological factors such as experiment time, experimenter or reagent differences. Batch effects are a common problem in biology and are technical confounders that must be considered for downstream single-cell data analysis tasks in order to discover true biological signals.

(4) Feature selection is designed to reduce the dimensionality of multi-group biomarker data and reduce the computational burden of downstream analysis tasks. Also, feature selection reduces noise in the data and facilitates visualization of the data.

(5) Dimensionality reduction is designed to embed the expression matrix into a low-dimensional space to capture the underlying structure in the data with the lowest possible dimensionality, which works on the principle that multi-group student biology data is inherently low-dimensional. In other words, the biofluidic shape in which the cellular expression profile resides can be efficiently described in much fewer dimensions than the number of genes.

III. A. 2) Multi-omics data feature extraction

(1) Gene network reconstruction and feature selection

Assume that there are n genes in the gene regulatory network, i.e., $G = \{G_1, G_2, \cdots, G_n\}$, and G_i denotes gene i. G_i is regulated by other genes and can be expressed as:

$$G_i(t) = f(G_1(t), G_2(t), \cdots, G_n(t)) \tag{9}$$

where $G_i(t)$ denotes the expression level of gene i at time t. If the f function is assumed to be a linear model, then:

$$G_i(t) = j_{1i}G_1(t) + j_{2i}G_2(t) + \cdots + j_{ni}G_n(t) + b(t) \tag{10}$$

where j_{mi} denotes the strength of regulation of gene m on gene i and $b(t)$ denotes the external perturbation or input.

Reconstruction of the network in the framework of linear modeling is actually to find, one by one, the set of transcription factors that are regulatory for $G_i\left(i = 1, 2, \cdots, n\right)$, i.e., to find out, one by one, the strength of regulation of the gene by the transcription factors. This can be regarded as a feature selection problem, according to the calculated regulatory strength j_{mi} to sort, selected intensity greater than the threshold of the regulatory edge is regarded as the existence of regulatory relationship, the intensity below the threshold is regarded as the absence of regulatory relationship. To screen out transcription factors for a target gene that are important for its influence, if transcription factors are considered as features, is to reconstruct the gene regulatory network using the feature selection method.

The problem is redefined as a learning data (LS) set $LS = \{x_1, x_2, \cdots, x_N\}$, $x_i = \{x_i^1, x_i^2, \cdots, x_i^n\}^T$ denoting the expression data of n genes in the i rd sample. In the learning samples, n learning is performed to compute all j_{mi} and the entire gene regulatory network is reconstructed after weight ordering filtering.

(2) Feature extraction of multi-group student biology data

The multi-group student data used in this paper are mainly the time series expression data of genes, protein sequence data, promoter sequence data and CDS data. Similarly, due to the different types of data corresponding to different genetic characteristics and data forms, they need to be processed separately in different ways. For protein sequence data, we used the combined triad (CT) coding method for coding processing. In this paper, the data filling method is to fill the previous value or the next value in the sequence data. Because there is a dependency between the previous and next values of the time series data, it is better to use this filling method than mean value filling. The time series expression data of each gene is represented as:

$$V_{t\exp} = [e_{t1}, e_{t2}, \cdots, e_{tn}]^T \tag{11}$$

Here $e_{tp}(p = 1, \cdots, n)$ denotes the expression value of the gene at the p th time point.

By merging the time series data feature vector, protein data feature vector, promoter data feature vector and CDS feature vector of a transcription factor we can get the overall biological data feature representation of a transcription factor as:

$$A_{tf} = [V_{t\exp-tf}, D_{pep-tf}, D_{prom-tf}, D_{cds-tf}]^T \tag{12}$$

Such an overall biological data feature representation vector can also be obtained for each target gene as:

$$A_{tg} = [V_{t\exp-tg}, D_{pep-tg}, D_{prom-tg}, D_{cds-tg}]^T \tag{13}$$

For each transcription factor-target gene pair (TF-TG), by splicing the time series expression feature vectors of the transcription factors and target genes therein, protein sequence feature vectors, promoter sequence feature vectors, CDS feature vectors, and overall data representation feature vectors, respectively, the following five feature representations can finally be obtained for each pair of transcription factor-target gene pairs, viz:

$$Exp_{(T,G)} = [V_{t\exp-tf}, V_{t\exp-tg}]^T \tag{14}$$

$$Pep_{(T,G)} = [D_{pep-tf}, D_{pep-tg}]^T \tag{15}$$

$$\Pr om_{(T,G)} = [D_{prom-tf}, D_{prom-tg}]^T \tag{16}$$

$$CDS_{(T,G)} = [D_{cds-tf}, D_{cds-tg}]^T \tag{17}$$

$$A_{(T,G)} = [A_{tf}, A_{tg}]^T \tag{18}$$

where $Exp_{(T,G)}$ represents the TF-TG temporal expression value feature vector, $Pep_{(T,G)}$ represents the TF-TG protein sequence feature vector, $\mathrm{Pr}om_{(T,G)}$ represents the TF-TG promoter sequence feature vector , $CDS_{(T,G)}$ denotes the CDS feature vector of TF-TG, and $A_{(T,G)}$ denotes the overall data feature of TF-TG.

III. B. Methods for constructing time-sequenced gene regulatory networks

III. B. 1) GL-TGRN network model construction

In order to achieve the effect of analyzing the temporal expression of AKR1B10 gene in hepatocellular carcinoma patients, this paper combines LSTM with GraphSAGE and establishes GL-TGRN for analyzing the temporal expression of AKR1B10 gene in hepatocellular carcinoma patients. Figure 2 shows the specific architecture of the GL-TGRN model.The main structure of GL-TGRN can be divided into three main parts.The first part is the part of the null domain graph convolution method using GraphSAGE as an application example, which is used to learn useful network topology features from the temporal gene regulatory networks from the first T moments and generate network node feature representations for gene nodes at each moment. The second part is the recurrent neural network part with LSTM as an application example, this part of the algorithm content is based on the fact that recurrent neural networks can learn the relevant features of the data in the presence of a time dependency, and generate and update the hidden state for the node feature representation of the gene nodes at each moment. The third part is a feature fusion task to make predictions about the structure of the gene regulatory network at future moments.

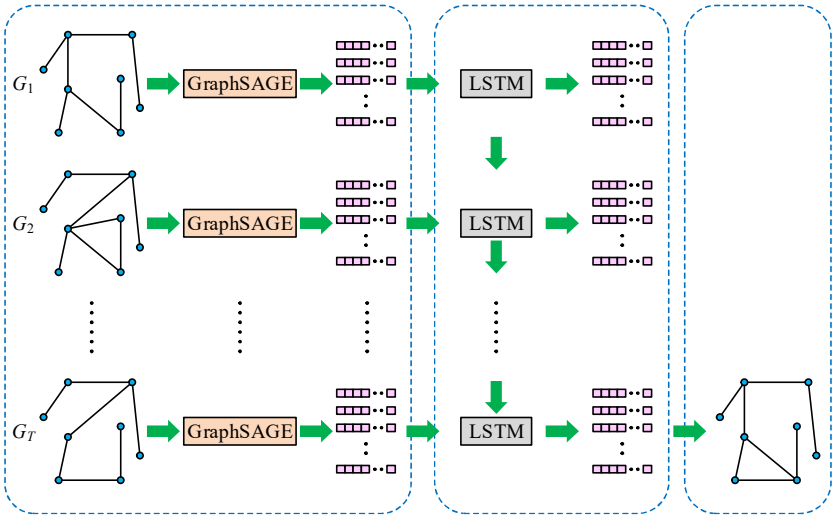


Figure 2: GL-TGRN procedure

The first part of the graph, Graph SAGE, receives inputs from multiple momentary gene regulatory networks, which are represented here in the mean aggregator as an aggregator formalism with its computational formula as:

$$x_{v_t}^{l+1} = \sigma\left(W^{l+1} \cdot MEAN\left(\{x_{v_t}^{l}\} \cup \{x_{u_t}^{l}, \forall u_t \in N_t(v_t)\}\right)\right)$$
$$v_t \in V, t \in [1,T] \tag{19}$$

In this paper, it is assumed that the number of genes is the same at all moments, then $v_t\left(t \in [1,T]\right)$ refers to the same set, and the subscript t is a feature representation of gene v_t that is convenient to distinguish genes at different moments, that is, $x_{v_t}^{l+1}$ is the feature representation of gene v after the processing of the layered convolutional network of $l+1$. u_t is the set of all the genes in the regulatory network at the t th moment that have a regulatory relationship with gene v_t. W^{l+1} denotes the weight matrix of the $l+1$ th graph convolution layer, and σ is the activation function. This formula indicates that for each graph convolutional layer, all the gene regulatory networks from 1 to T moments use the same weight matrix to extract the topological information of the regulatory network and generate a corresponding node feature representation for each gene. When the algorithm is initialized, it first generates a unit matrix of $|V| \times |V|$, where each row of the unit matrix represents the initialized features of a gene, i.e., all genes contain only their own information when they have not been processed by the graph

convolutional network. Finally the first part outputs the node feature representations $x_{v_t}, v_t \in V, t \in [1,T]$ of all genes at the corresponding moments after $l(l \geq 1)$ layers of graph convolutional network processing.

The second part of the recurrent neural network part of the figure can directly receive the processing results of the first part of the graph convolutional neural network for further learning, generating feature representations with temporal information for all genes at different moments, which is mathematically processed by the following formula:

$$F_t = \sigma\left(W_f\left[h_{t-1}, x_{v_t}\right] + b_f\right) \tag{20}$$

$$I_t = \sigma\left(W_i\left[h_{t-1}, x_{v_t}\right] + b_i\right) \tag{21}$$

$$\tilde{C}_t = \tanh\left(W_c\left[h_{t-1}, x_{v_t}\right] + b_c\right) \tag{22}$$

$$C_t = F_t * C_{t-1} + I_t * \tilde{C}_t \tag{23}$$

$$O_t = \sigma\left(W_o\left[h_{t-1}, x_{v_t}\right] + b_o\right) \tag{24}$$

$$h_t = O_t * \tanh(C_t) \tag{25}$$

where $v_t \in V$, $t \in [1,T]$, the LSTM will process the features that are organized into different time snapshots of the same gene with temporal structure. In the formula, F_t, I_t, $\tilde{C}_t$, C_t, and h_t represent the states of the LSTM units when gene v_t is processed by LSTM, and the specific meanings can be directly referred to the interpretation of the LSTM formula. Where O_{v_t} is the feature representation of gene v_t containing temporal information after LSTM processing at t.

After the third part of the figure obtains the feature representation of all genes at different moments, the algorithm can use a variety of measures for different node pairs to generate the probability of generating a concatenation of edges between two and two or other metrics, this algorithm ultimately uses the vector inner product to generate the probability of having a regulatory relationship between genes and genes, which is computed by the formula:

$$p(v_{t+1}, u_{t+1}) = \sigma(O_{v_t} \cdot O_{u_t}) \tag{26}$$

where v_{t+1} and u_{t+1} are the two genes that need to be predicted whether there is a regulatory relationship or not at the moment $t+1$, and O_{v_t} and O_{u_t} are the node feature vectors of the two genes that have been processed by LSTM at the moment t, respectively. σ is the Sigmoid activation function and $(\cdot)$ is the inner product operation of the vectors.

III. B. 2) Time-Sequenced Gene Regulatory Network Training Strategy

Probabilistic augmented program evolution algorithm is an algorithm used to train the structure of the proposed model. The probabilistic augmented program evolution algorithm consists of three parts, i.e., incorporating program instructions encoded in probabilistic vectors, population-based incremental learning, and tree-structured programming. The probabilistic augmented program evolution algorithm consists of two learning processes, population-based learning (GBL)) and elite learning (EL). Among them, population-based learning is the main learning process and elite learning is used to get the best structure.

A probabilistic prototype tree is a complete n-forked tree with each node containing a random constant P_{el}, a vector of probabilities of variables $\overrightarrow{P_j}$, and $l+k$ sets of instructions (where l sets of terminal instructions and k sets of function instructions). The probabilistic prototype tree is generated by the following two steps: (1) generating the random constant $P_{el} = \mathrm{U}[0,1]$; (2) for all sets of terminal instructions $P_j(I) = P_T / l$ and for all sets of functional instructions $P_j(I) = (1 - P_T) / k$, where P_T denotes the probability of using the set of terminals.

One of the procedures for including the very important update and mutation probability prototype trees in training is as follows:

$$P_{TARGET} = P(P_{ROGb}) + (1 - P(P_{ROGb})) \cdot l_r \cdot \frac{\varepsilon + FIT(P_{ROG}^{el})}{\varepsilon + FIT(P_{ROG})} \tag{27}$$

The process of updating the probabilistic prototype tree is denoted as:

$$P_j(I_j(P_{ROGb})) = P_j(I_j(P_{ROGb})) + c^{l_r} \cdot l_r \cdot \left(1 - P_j(I_j(P_{ROGb}))\right) \tag{28}$$

The process of variational probability prototype tree is denoted as:

$$P_j(I) = P_j(I) + m_r \cdot (1 - P_j(I)) \tag{29}$$

where P_{ROGb} represents an individual, P_{ROG}^{el} represents the best individual, $I_j(P_{ROGb})$ represents the instruction of individual P_{ROGb}, and l_r represents the learning rate, which is a well-defined constant. c^{l_r} is a constant that affects the number of iterations, and m_r represents the mutation rate. Repeat until $P(P_{ROGb}) \geq P_{TARGET}$. When $P_j(I) > T_p$, the pruning probability prototype tree, generally taken as $T_p = 0.85$.

III. C. Results and analysis of gene regulation experiments

III. C. 1) AKR1B10 gene inference

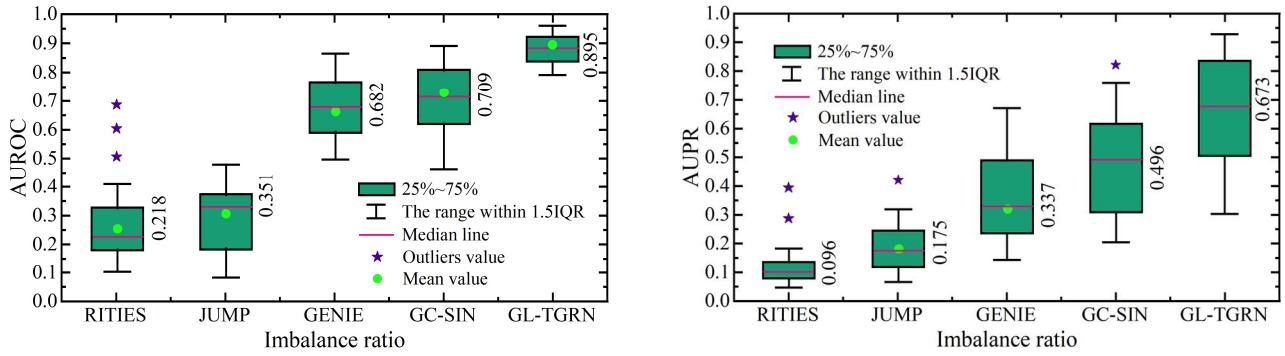
Based on the previously selected multi-group biomarker dataset of hepatocellular carcinoma patients, single-cell transcriptional data and knockdown data of AKR1B10 gene were selected as inputs from them, so as to infer the regulatory relationship between AKR1B10 genes in hepatocellular carcinoma patients.AURO and AUPR are the two most commonly used assessment indexes for evaluating the inference method of the gene regulatory network.The values of AUROC and AUPR are able to show the inference AUROC and AUPR values can show the accuracy of the inference methods, and the higher the value, the more accurate the inference results. Therefore, AUROC and AUPR are used in this paper to quantitatively evaluate the performance of this method.

Comparison experiments were conducted using RITIES, JUMP, GENIE, and GC-SIN. where GENIE utilizes a holistic tree-based strategy with multifactor perturbation data, and JUMP uses a hybrid strategy combining a nonparametric decision tree approach with a dynamics on/off model with time-series expression data. The data used in both methods are averages of gene expression, and RITIES, GC-SIN, and GL-TGRN designed in this paper used single-cell transcriptional data to infer AKR1B10 gene expression in hepatocellular carcinoma patients.

Figure 3 shows the results of AKR1B10 gene regulation inference by different methods, in which Figure 3(a)~(b) shows the comparison results of AUROC and AUPR, respectively. The results showed that the accuracy of GL-TGRN, GC-SIN and RITIES inference results were better than that of JUMP and GENIE, indicating that the single-cell transcriptional data were more effective for inference of AKR1B10 gene expression compared to the mean gene expression values. This is due to the fact that single-cell transcriptional data contains the gene expression levels of multiple cells, and analyzing this data can capture the cellular differential information to infer the gene expression of AKR1B10 more accurately.GL-TGRN improved the AUROC and AUPR by 26.23% and 35.69%, respectively, compared with the comparative method GC-SIN. This indicates that the method proposed in this paper to classify genes using knockout data before analyzing single-cell transcriptional data is effective. The gene knockout data contains steady-state gene expression information, which makes up for the deficiencies in the single-cell data, and the results of gene classification directly identify many non-existent regulatory relationships, which provides a great contribution to the improvement of the inference accuracy. In summary, the GL-TGRN model designed in this paper showed better performance results than the comparative methods on the multi-omics data of AKR1B10 gene in hepatocellular carcinoma patients, either AUROC or AUPR performance results, which indicated that the inferred results of the temporal sequencing of AKR1B10 gene in hepatocellular carcinoma patients had a higher accuracy rate, i.e., the inference was closer to the temporal expression of the AKR1B10 gene in the liver cancer serum proteins. The higher accuracy of the extrapolation indicates the more accurate the originally unknown intergenic regulatory roles reduced by analyzing the expression levels of genes that can be observed in the cells.

III. C. 2) Gene-modulated ablation experiments

In the GL-TGRN model established in this paper, three modules, GraphSAGE, LSTM and feature fusion (FF), are mainly included as a way to ensure the model's ability to optimize the prediction of AKR1B10 gene temporal expression in liver cancer patients. In order to further analyze the effect of different structures on enhancing the regulatory network of AKR1B10 gene, the ablation experiments were designed based on the GraphSAGE network and the introduction of the LSTM and feature fusion modules, respectively, on the basis of the data of multiple groups of student organisms from hepatocellular carcinoma patients obtained in the previous paper. The results of the above two ablation experiments were compared with the performance of the GL-TGRN model for the purpose of ablation experiments. AUROC and AUPR were chosen as evaluation indexes to obtain the effect of AKR1B10 gene regulation under different features, and Figure 4 shows the ablation experiments of the model under different evaluation indexes.



(a) AUROC (b) AUPR

Figure 3: Inference results of AKR1B10 gene regulation

From the comparison of the experimental results in the figure, it can be seen that the AUROC and AUPR values obtained from the GL-TGRN model are significantly higher than those of a single model or two network structures, then it indicates that the GL-TGRN model designed in this paper can realize the precise analysis of the time-sequence expression of the AKR1B10 gene in hepatocellular carcinoma patients. The ablation experiments verified that the addition of the fusion multi-omics feature module to the GL-TGRN model is useful for improving the accuracy of the model in inferring the regulatory network of the AKR1B10 gene in hepatocellular carcinoma patients. This suggests that fusion multi-omics gene features not only enriched the input information of the model, but also enhanced the ability of the model to predict the regulatory relationships among genes.

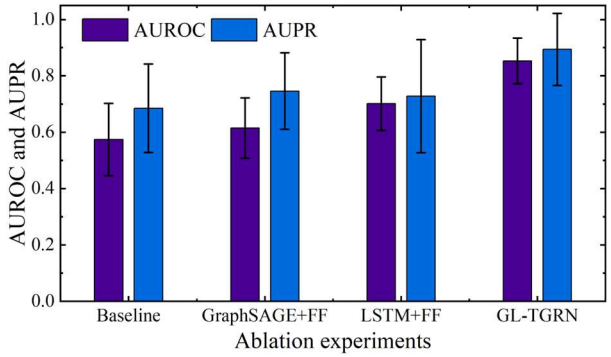


Figure 4: Ablation experiments of the model

III. C. 3) AKR1B10 gene differential expression analysis

After obtaining gene and microRNA expression data for hepatocellular carcinoma, the data were subjected to differential expression analysis to screen for differential genes and microRNAs that contribute to the development of hepatocellular carcinoma.In practice, the observed data for genes and microRNAs are discrete values containing random errors, and comparisons of the differences in expression levels between conditions are generally considered as comparisons of two distributions, which, in addition to considering the mean value, also Consider the effect of variance on the analysis of differences. In this paper, we choose the R language and call the function of edgeR package of Bioconductor to help carry out variance analysis.

The exactTest() function of edgeR is called to accurately test the differential genes and microRNAs in normal and diseased samples, where the null hypothesis of the exactTest is that the relative abundance of microRNAs or genes g in normal and diseased samples are equal, and those that are contrary to the null hypothesis are considered to be unequal. exactTest() uses the FDR-corrected q-values to analyze the variance of the samples. FDR-corrected q-value as the criterion for selecting differential genes and microRNAs associated with hepatocellular carcinoma, and the FDR-corrected q-value was chosen to have a significant level of 0.01.

In order to reduce the false positives of the data, this topic simultaneously used the logarithmic fold change (logFC) to determine the differential genes and differential microRNAs in the two samples and averaged the logFC values of the genes or microRNAs in each group of comparisons after removing the very large and very small values. In

this case, the absolute value of logFC was chosen to have a significant level of 1, i.e., a difference of two times was considered a significant difference, and when the ratio value was greater than 2, it was considered to be up-regulated in normal samples, and when the ratio value was less than -2, it was considered to be down-regulated in diseased samples. The expression levels of genes and microRNAs were considered significantly different when they met the criteria for differences in both logFC and FDR values under both samples. Tables 1 and 2 show the results of the differences of the first 20 differential genes and 14 microRNAs, respectively. After processing and analysis, 786 significantly different genes and 14 microRNAs were obtained in this study, and these biomolecules were considered to be significantly associated with liver cancer development.

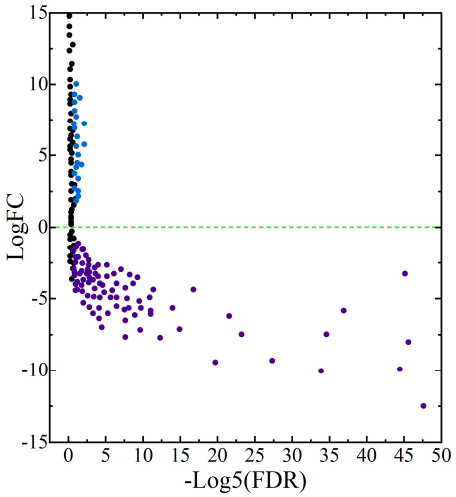
Table 1: Top 20 differential genes

Name	Name	Name	Name
CD5LS	CTD-2573B13	RP12-532J6	LMAN12L2
FAM8C9A	SPICT63	KCNT1Y7	RP11-878D25
STAB78B2	CD16BET2	NKX3-1H	SLCF9E5
GPTR189	CXCR3P21	CTB-135C14	MMP13E2
RP11-1213A24	IDI2-ASI1A	ITG1AD8	FEF1A2P35

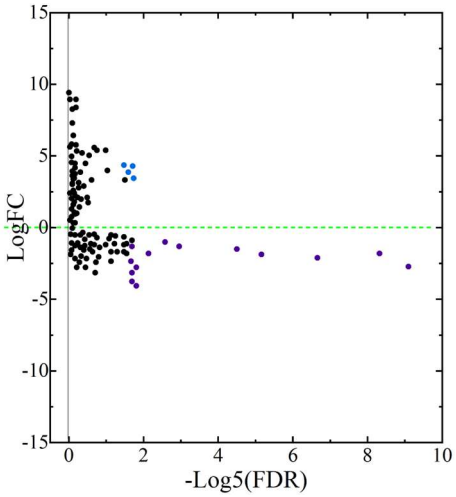
Table 2: Differential microRNA

Name	Name	Name	Name
HCC-miR-443b	HCC-miR-452a	HCC-miR-138	HCC-miR-141
HCC-miR-145	HCC-miR-428	HCC-miR-129	HCC-miR-485-2
HCC-miR-12b	HCC-miR-486-7	HCC-miR-207	HCC-miR-4612
HCC-miR-892b	HCC-miR-195-2	-	-

Figure 5 shows the volcano plots of AKR1B10 gene expression, where Figure 5(a)~(b) shows the volcano plots of differential genes and differential microRNAs, respectively. The horizontal coordinate of the volcano plot is taken as the negative logarithm of FDR with 5 as the base (-log5), and the larger the horizontal coordinate is, the smaller the FDR value is, and the more different the molecule is in the two samples. The vertical coordinate is taken as logFC and when the value is greater than 2, it is the point of up-regulation in the normal sample, labeled as blue, and the point less than -2 is the point of up-regulation in the diseased sample, labeled as purple. Based on the volcano plot, it can be seen that the differential expression of AKR1B10 gene in hepatocellular carcinoma patients can be clearly understood, which provides reliable data support for clarifying the prevalence of hepatocellular carcinoma patients.



(a) Differential genes



(b) Differential microRNA

Figure 5: Volcano map of AKRIB10 gene expression

IV. Conclusion

The study demonstrated that the GL-TGRN model showed significant advantages in the temporal expression inference of AKR1B10 gene in hepatocellular carcinoma patients. Compared with the traditional method, GL-TGRN improved the AUROC and AUPR values by 26.23% and 35.69%, respectively, indicating that the model can more accurately capture the temporal features of gene expression in hepatocellular carcinoma patients. The comparison with the contrasting methods indicated that the fusion of multi-omics data plays an important role in improving the accuracy of gene regulatory network inference. In addition, the results of differential expression analysis screened 786 differential genes and 14 miRNAs, molecules that were significantly associated with hepatocarcinogenesis. It was further verified by ablation experiments that the fusion of multi-omics features effectively enhanced the model performance. The study in this paper provides strong data support and new research perspectives for the early diagnosis and precise treatment of hepatocellular carcinoma.

References

[1] Rumgay, H., Arnold, M., Ferlay, J., Lesi, O., Cabasag, C. J., Vignat, J., ... & Soerjomataram, I. (2022). Global burden of primary liver cancer in 2020 and predictions to 2040. Journal of hepatology, 77(6), 1598-1606.

[2] Petruzziello, A. (2018). Epidemiology of hepatitis B virus (HBV) and hepatitis C virus (HCV) related hepatocellular carcinoma. The open virology journal, 12, 26.

[3] Ganne-Carrié, N., & Nahon, P. (2019). Hepatocellular carcinoma in the setting of alcohol-related liver disease. Journal of hepatology, 70(2), 284-293.

[4] Kanwal, F., Kramer, J. R., Mapakshi, S., Natarajan, Y., Chayanupatkul, M., Richardson, P. A., ... & El-Serag, H. B. (2018). Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology, 155(6), 1828-1837.

[5] Wong, M. C., Jiang, J. Y., Goggins, W. B., Liang, M., Fang, Y., Fung, F. D., ... & Chan, H. L. (2017). International incidence and mortality trends of liver cancer: a global profile. Scientific reports, 7(1), 45846.

[6] Nault, J. C., Cheng, A. L., Sangro, B., & Llovet, J. M. (2020). Milestones in the pathogenesis and management of primary liver cancer. Journal of Hepatology, 72(2), 209-214.

[7] Niu, Z. S., Wang, W. H., & Niu, X. J. (2022). Recent progress in molecular mechanisms of postoperative recurrence and metastasis of hepatocellular carcinoma. World journal of gastroenterology, 28(46), 6433.

[8] Orcutt, S. T., & Anaya, D. A. (2018). Liver resection and surgical strategies for management of primary liver cancer. Cancer control, 25(1), 1073274817744621.

[9] Meier, T., Timm, M., Montani, M., & Wilkens, L. (2021). Gene networks and transcriptional regulators associated with liver cancer development and progression. BMC Medical Genomics, 14, 1-23.

[10] Huo, X., Han, S., Wu, G., Latchoumanin, O., Zhou, G., Hebbard, L., ... & Qiao, L. (2017). Dysregulated long noncoding RNAs (lncRNAs) in hepatocellular carcinoma: implications for tumorigenesis, disease progression, and liver cancer stem cells. Molecular cancer, 16, 1-10.

[11] Llovet, J. M., Montal, R., Sia, D., & Finn, R. S. (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. Nature reviews Clinical oncology, 15(10), 599-616.

[12] DiStefano, J. K., & Davis, B. (2019). Diagnostic and prognostic potential of AKR1B10 in human hepatocellular carcinoma. Cancers, 11(4), 486.

[13] Ma, L. N., Ma, Y., Luo, X., Ma, Z. M., Ma, L. N., & Ding, X. C. (2024). AKR1B10 expression characteristics in hepatocellular carcinoma and its correlation with clinicopathological features and immune microenvironment. Scientific Reports, 14(1), 12149.

[14] Tian, K., Deng, Y., Li, Z., Zhou, H., & Yao, H. (2023). AKR1B10 inhibits the proliferation and metastasis of hepatocellular carcinoma cells by regulating the PI3K/AKT pathway. Oncology Letters, 27(1), 18.

[15] Wang, Y. Y., Qi, L. N., Zhong, J. H., Qin, H. G., Ye, J. Z., Lu, S. D., ... & You, X. M. (2017). High expression of AKR1B10 predicts low risk of early tumor recurrence in patients with hepatitis B virus-related hepatocellular carcinoma. Scientific reports, 7(1), 42199.

[16] Maccari, R., & Ottanà, R. (2025). In Search for Inhibitors of Human Aldo-Keto Reductase 1B10 (AKR1B10) as Novel Agents to Fight Cancer and Chemoresistance: Current State-of-the-Art and Prospects. Journal of Medicinal Chemistry.

[17] Zhang, T., Guan, G., Zhang, J., Zheng, H., Li, D., Wang, W., ... & Chen, X. (2022). E2F1‐mediated AUF1 upregulation promotes HCC development and enhances drug resistance via stabilization of AKR1B10. Cancer science, 113(4), 1154-1167.

[18] Endo, S., Matsunaga, T., & Nishinaka, T. (2021). The role of AKR1B10 in physiology and pathophysiology. Metabolites, 11(6), 332.

[19] Rajak, S., Gupta, P., Anjum, B., Raza, S., Tewari, A., Ghosh, S., ... & Sinha, R. A. (2022). Role of AKR1B10 and AKR1B8 in the pathogenesis of non-alcoholic steatohepatitis (NASH) in mouse. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1868(4), 166319.

[20] Dai, T., Ye, L., Yu, H., Li, K., Li, J., Liu, R., ... & Wang, G. (2021). Regulation network and prognostic significance of aldo-keto reductase (AKR) superfamily genes in hepatocellular carcinoma. Journal of hepatocellular carcinoma, 997-1021.

[21] Shi, J., Chen, L., Chen, Y., Lu, Y., Chen, X., & Yang, Z. (2019). Aldo-Keto Reductase Family 1 Member B10 (AKR1B10) overexpression in tumors predicts worse overall survival in hepatocellular carcinoma. Journal of Cancer, 10(20), 4892.

[22] Banerjee, S. (2021). Aldo keto reductases AKR1B1 and AKR1B10 in cancer: molecular mechanisms and signaling networks. Cell Biology and Translational Medicine, Volume 14: Stem Cells in Lineage Specific Differentiation and Disease, 65-82.

[23] Shen, Y., Huang, J., Jia, L., Zhang, C., & Xu, J. (2024). Bioinformatics and machine learning driven key genes screening for hepatocellular carcinoma. Biochemistry and Biophysics Reports, 37, 101587.

[24] Huang, R., Wei, C., Wang, B., Yang, J., Xu, X., Wu, S., & Huang, S. (2022). Well performance prediction based on Long Short-Term Memory (LSTM) neural network. Journal of Petroleum Science and Engineering, 208, 109686.

[25] Blessing B Ekpenyong,Godwin M Ubi,M E Kooffreh,Anthony J Umoyen,Cecilia S James,Ivon A Ettah... & Ogar A Ambo. (2025). Tumor protein 53 gene polymorphism, demographic attributes and associated risk factors among liver cancer patients in Calabar, Nigeria. BMC cancer,25(1),430.

[26] M. Liu,X. Zhang,D. Han & J.P. Thiery. (2024). 1034P Preclinical development of genetically modified tumor-infiltrating lymphocytes using biopsy samples from liver cancer patients. Annals of Oncology,35(S2),S695-S695.

[27] Yongyan Chen,Wendi Zhang,Min Cheng,Xiaolei Hao,Haiming Wei,Rui Sun & Zhigang Tian. (2024). Galectin-3-ITGB1 Signaling Mediates Interleukin 10 Production of Hepatic Conventional Natural Killer Cells in Hepatitis B Virus Transgenic Mice and Correlates with Hepatocellular Carcinoma Progression in Patients. Viruses,16(5),737-.

[28] Yao Shen,Ailin Qiu,Xin Huang,Xiaosha Wen,Sundar Shehzadi,Yan He... & Shenghui Yang. (2024). AKR1B10 and digestive tumors development: a review. Frontiers in Immunology,15,1462174-1462174.