

Machine Learning Approach and Bioinformatics Analysis Discovered Key Genomic Signatures for Hepatitis B Virus-Associated Hepatocyte Remodeling and Hepatocellular Carcinoma

Adane Adugna¹ , Gashaw Azanaw Amare¹ 
and Mohammed Jemal² 

Abstract

Hepatitis B virus (HBV) causes liver cancer, which is the third most common cause of cancer-related death worldwide. Chronic inflammation via HBV in the host hepatocytes causes hepatocyte remodeling (hepatocyte transformation and immortalization) and hepatocellular carcinoma (HCC). Recognizing cancer stages accurately to optimize early screening and diagnosis is a primary concern in the outlook of HBV-induced hepatocyte remodeling and liver cancer. Genomic signatures play important roles in addressing this issue. Recently, machine learning (ML) models and bioinformatics analysis have become very important in discovering novel genomic signatures for the early diagnosis, treatment, and prognosis of HBV-induced hepatic cell remodeling and HCC. We discuss the recent literature on the ML approach and bioinformatics analysis revealed novel genomic signatures for diagnosing and forecasting HBV-associated hepatocyte remodeling and HCC. Various genomic signatures, including various microRNAs and their associated genes, long noncoding RNAs (lncRNAs), and small nucleolar RNAs (snoRNAs), have been discovered to be involved in the upregulation and downregulation of HBV-HCC. Moreover, these genetic biomarkers also affect different biological processes, such as proliferation, migration, circulation, assault, dissemination, antiapoptosis, mitogenesis, transformation, and angiogenesis in HBV-infected hepatocytes.

Keywords

HBV, HCC, machine learning approach, bioinformatics analysis, genomic signatures

Received: 5 November 2024; accepted: 24 March 2025

Introduction

Hepatitis B virus (HBV)-induced hepatocellular carcinoma (HCC) is responsible for approximately 90% of primary liver cancer cases worldwide.¹ Prospective cohort studies have shown that patients with chronic HBV infection have a nearly 100-fold increased risk of developing HCC.^{2–4}

There is evidence that the transformed protein products of the HBV accelerate the onset of cancer. When HBV infection occurs, hepatocyte remodeling and the development of HCC are closely linked to the transcriptional activation of differentiation-regulating genes by HBV-encoded proteins. Dysregulation of gene expression can result from the interaction of HBV-encoded proteins with the machinery of the host cell, especially those genes linked to cell differentiation.⁵ In other words, comprehending the

methods by which proteins encoded by HBV alter the transcription of genes that regulate differentiation is crucial for clarifying persistent HBV infection and hepatocyte transformation.^{6,7} Important transcription factors, epigenetic changes, and signaling pathways that affect hepatocyte gene expression patterns may interact throughout this process.⁸

¹Medical Laboratory Sciences, College of Health Sciences, Debre Markos University, Ethiopia

²Department of Biomedical Sciences, School of Medicine, Debre Markos University, Ethiopia

Corresponding author:

Adane Adugna, Medical Laboratory Sciences, College of Health Sciences, Debre Markos University, Debre Markos, P.O. Box: 269, Ethiopia.
Email: adaneadugna29@gmail.com



Moreover, the integration of HBV DNA from hepatocytes into their genomes is a significant step in the pathophysiology of HCC during persistent HBV infection. The incorporation of HBV DNA can result in several molecular and cellular alterations that support the development of HCC and hepatocyte remodeling.⁹ It is also imperative to understand how incorporated HBV DNA impacts hepatocyte remodeling and plays a role in the development of HCC to devise tailored treatments that attempt to impede these mechanisms, which may lead to the discovery of biomarkers for early detection in people with persistent HBV infection.^{10,11}

Furthermore, viral DNA sequences have the potential to disrupt genes in host cells during integration. This can result in changes in gene expression patterns, which can encourage cell division, prevent apoptosis, and cause genomic instability.¹² These alterations believed to play a critical role in the development of HCC from chronic HBV infection.¹³

Hepatocyte remodeling and the emergence of HBV-associated HCC can also be attributed to the dysregulation of many signaling pathways caused by the integration of HBV DNA into the hepatocyte genome.¹⁴ Hence, integrated HBV DNA has been found to impact multiple signaling pathways, including the PI3K/Akt/mTOR pathway, the Ras/Raf/MEK/ERK pathway, and the Wnt/ β -catenin pathway.^{15,16} These cascades are essential for controlling the survival, differentiation, and proliferation of hepatic cells. Through several different means, the incorporation of HBV DNA can activate these signaling pathways. Viral proteins generated by integrated HBV DNA, for instance, may interact directly with these pathways' constituent parts or interfere with their regular control.¹⁷ Within infected hepatocytes, this dysregulation may result in prolonged activation of pro-proliferative signals and the prevention of apoptosis.¹⁸

Moreover, integrated HBV DNA may also cause persistent hepatic inflammation, which can activate inflammatory signaling cascades such as the JAK/STAT and NF- κ B pathways. A microenvironment that fosters cellular changes and the growth of HCC can be produced by persistent inflammation. In addition, hepatic stellate cell signals are linked to extracellular matrix remodeling and the growth of myofibroblast-like cell types during HBV-associated hepatocyte immortalization.¹⁹

Precise estimation of HBV-induced hepatocyte remodeling and HCC biomarker prediction contributes to the early detection of HBV-related liver cancer and decreases mortality. The direct selection of predictive biomarkers, without the need for subjective preselection, is made possible by machine learning (ML) techniques.²⁰

As the fields of big data science and statistical approaches have advanced, artificial intelligence (AI) has become increasingly applicable to every facet of human existence. AI focuses on creating analogs of human intelligence, reasoning, and understanding to solve problems, make decisions, and perform tasks. ML, a subset of AI, is a process that involves processing large volumes of data with a specific objective, autonomously grouping the information, and enabling the machine to evaluate it.²¹ This is achieved by ingesting data and applying algorithms to develop a predictive model.²²

Currently, ML techniques are frequently employed to identify various medical signatures for patients with HBV-HCC.²³ By converting genetic and molecular data into a streamlined input mechanism, ML algorithms can utilize this information represented at the sequence level and analyze it through various methods. In other words, the ML approach has a complex role in HBV-induced hepatocyte remodeling and HCC in many facets of illness, and more advanced technologies have become available.²⁴

Because ML techniques can enhance diagnostic precision and predictive power, they have become increasingly popular in the analysis of HBV integration. A variety of ML models, such as support vector machines (SVMs), random forests (RFs), and eXtreme Gradient Boosting (XGBoost), have been utilized to forecast HBV infection and assess its integration into host genomes. Research has demonstrated that ensemble learning approaches, which integrate multiple algorithms, can improve the prediction accuracy and reduce the generalization errors associated with single classifiers.²⁵

Furthermore, studies have demonstrated the value of ML in identifying key clinical predictors for HBV integration, which is essential for early diagnosis and treatment strategies.^{25,26} Personalized treatment approaches are made possible by the incorporation of advanced algorithms that not only facilitate the stratification of individuals at risk but also contribute to the understanding of the molecular pathways underlying diseases associated with HBV.²⁷

Currently, the ML approach provides a thorough grasp of HBV progression through predictive analysis and identifies blood biomarkers that could be utilized as diagnostic, therapeutic, and prognostic signals for patients with HBV-caused hepatocyte remodeling and HCC.²⁰

Furthermore, a new field of study called bioinformatics has evolved that offers a great means of comprehending biological research, including the diagnostic, therapeutic, and prognostic value of miRNAs and their related genes in HBV-induced hepatic malignancy via the analysis of large-scale biological datasets.²⁸ Moreover, innovations in bioinformatics and science have recently revolutionized the generation and analysis of massive omics datasets. This has made it possible to integrate numerous datasets, such as the transcriptome, proteome, glycoproteome, metabolome, epigenome, and genome.¹⁷

Various types of genomic signatures have been discovered via machine learning approaches and bioinformatics analysis to indicate HBV-associated hepatocyte remodeling and HCC. This review summarizes recent advances in understanding the role of these signatures in HBV-induced hepatocyte remodeling and hepatic cancer.

Machine Learning (ML) Approaches Revealed Genomic Signatures for Hepatitis B Virus-Associated Hepatocyte Remodeling and HCC

The identification of genetic biomarkers for HBV-HCC has undergone dramatic transformation, shifting from

conventional statistical approaches to sophisticated ML strategies.²⁹ Traditional statistical analysis techniques, such as ANOVA, Pearson correlation, and *t*-tests, rely on specific assumptions about data distributions, typically normality, which genomic data often do not satisfy. While these methods are straightforward to interpret, making them suitable for smaller datasets, they struggle with high-dimensional data where the number of features exceeds the number of observations. This limitation can lead to overfitting and poor generalization of new data.³⁰

As a result, standard statistical tests are primarily used for exploratory data analysis and preliminary hypothesis testing. While they can identify genes with differential expression, their linear nature may lead to the oversight of complex associations.³¹ In contrast, ML techniques offer greater flexibility by effectively managing intricate, non-linear correlations and interactions between variables, making them better suited for analyzing high-dimensional genomic datasets.³²

Depending on the characteristics of the dataset, a variety of methods, including SVM, RF, and XGBoost, can be employed. Each algorithm has advantages; for example, SVM excel in high-dimensional environments, while RF are resistant to overfitting.³³

In addition, naïve Bayes performs classification by assuming the independence of features, while *K*-nearest neighbors (KNN) is another classification algorithm, and *K*-means is used for clustering.³⁴ Furthermore, dimensionality reduction algorithms, such as principal component analysis (PCA), simplify datasets by retaining only the most significant features.³⁵ Moreover, gradient boosting and AdaBoost are boosting algorithms designed to improve the accuracy of predictive models, particularly in the genetic and molecular biomarker analysis of HBV-HCC.³⁶

Moreover, to ensure the robustness of predictions, ML techniques often utilize cross-validation and a variety of performance metrics, including accuracy, sensitivity, and specificity. Recent research has demonstrated that ML methods can identify potential biomarkers with greater accuracy than traditional approaches. For example, LASSO regression combined with other ML techniques has successfully identified diagnostic biomarkers for liver diseases associated with HBV.³⁷

In other words, evaluating the performance of ML algorithms is essential for ensuring their effectiveness in making predictions, and this evaluation is conducted via various performance measures and validation techniques. The choice of measures depends on the specific problem being addressed, whether it involves classification or regression.³⁸

Among the key performance measures, accuracy is the simplest; it calculates the ratio of correctly predicted instances to the total instances in a dataset, making it particularly useful for balanced datasets where all classes hold equal importance. $\text{Accuracy} = \frac{\text{total number of predictions}}{\text{number of correct predictions}}$. However, when false positives carry significant consequences, as in medical diagnoses, precision becomes crucial; it measures the proportion of true positive predictions out of all positive predictions made by the model.³⁹

Similarly, recall (sensitivity) assesses how well the model identifies all relevant instances, focusing on true positives, which is critical in applications such as cancer detection, where failing to identify a positive instance (false negative) can have severe implications.⁴⁰ To balance precision and recall, the *F1* score, which is the harmonic mean of the 2, is commonly used, especially in cases involving imbalanced datasets. Additionally, the ROC AUC score indicates how effectively a model can separate classes across various thresholds; a higher AUC signifies better model performance.⁴¹ For more detailed insights into errors, a confusion matrix provides a breakdown of true positives, true negatives, false positives, and false negatives, while logarithmic loss (LogLoss) quantifies prediction uncertainty by penalizing incorrect classifications that are made with high confidence.⁴²

To ensure that performance metrics accurately reflect model performance, a variety of validation techniques are employed. Cross-validation, particularly *K*-fold cross-validation, divides the dataset into multiple training and validation subsets to prevent overfitting and provide reliable performance estimates.⁴³ Alternatively, the train-test split method partitions the data into separate training and test sets to evaluate how well the model generalizes to unseen data.⁴⁴ For more robust performance estimation, bootstrap sampling involves repeatedly drawing samples with replacement to create multiple training sets. Additionally, benchmarking against simpler baseline models helps determine whether a more complex model adds incremental value.⁴⁵

ML can analyze multi-genomic data to uncover genotype-phenotype correlations that conventional methods might overlook. Therefore, while traditional statistical methods provide foundational insights into genetic data, ML serves as a powerful alternative that enhances predictive accuracy and reveals complex patterns essential for biomarker discovery in HBV-HCC.⁴⁶ Model selection, training, evaluation, and data preprocessing are the core tenets of ML. Data preprocessing ensures high-quality input by cleaning and standardizing genetic data, which is essential for accurate model performance. Data preprocessing consists of several steps in HBV-HCC genomic biomarker discovery, starting with data exploration and analysis to understand the the genomic dataset, followed by handling missing values using techniques like imputation and performing data cleaning to ensure data integrity.³⁴ In addition, standardizing and normalizing bring uniformity to the data, while handling outliers addresses extreme values.⁴⁷ This is followed by data reduction, which simplifies the dataset, and feature selection, which is aimed at identifying significant attributes to enhance model efficiency. Feature Engineering involves experimenting with new features to optimize model performance.⁴⁸

Model selection involves choosing the appropriate algorithm, such as decision trees, SVMs, or neural networks, on the basis of the specific characteristics of the genetic data and the desired outcome.⁴⁹ During the training phase of deployment, the model is fed labeled genomic data to identify patterns associated with HBV-HCC. This process utilizes optimization techniques to minimize prediction errors.

The model's performance on unseen data is then evaluated using metrics such as accuracy and precision.⁵⁰

Additionally, feature engineering is crucial, as it involves selecting and transforming genetic features that significantly impact the model's predictive capability. Hyperparameter tuning further enhances model performance by adjusting parameters that influence learning dynamics. By following these guidelines, researchers can effectively leverage ML to identify potential biomarkers that may aid in the early detection and management of HBV-related HCC.³⁴

High-throughput sequencing technologies and well-established databases provide the majority of input data used to derive genomic biomarkers associated with HBV-HCC.⁵¹ Important resources include the Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), which offer comprehensive gene expression profiles. For instance, RNA-seq data from patients with HBV infection, encompassing both normal and HCC tissue samples, can be found in datasets such as GSE25599, GSE77509, and GSE94660 from GEO.⁵² These databases facilitate the identification of differentially expressed genes (DEGs) by comparing gene expression levels across various conditions. Furthermore, multi-omics techniques that integrate data from transcriptomics, proteomics, and genomics enhance our understanding of the molecular pathways underlying HBV-HCC, aiding in the discovery of potential biomarkers for diagnosis and treatment.^{53,54}

Different ML techniques possess distinct characteristics and applications for identifying genetic biomarkers associated with HCC linked to HBV. Supervised learning algorithms, such as SVM and RF, are commonly employed for classification tasks, while unsupervised methods like clustering assist in uncovering patterns within data without predefined labels. Additionally, deep learning techniques, including Convolutional Neural Networks (CNNs), excel at processing complex information, particularly high-dimensional genomic data and image analysis.⁵⁵

The range of applications for these techniques includes patient stratification, disease prediction, and biomarker identification. RNA-seq and microarray studies are common sources of high-quality, well-annotated data that are essential for successful implementation.⁵⁶ ML models to accurately identify relevant biomarkers and DEGs. Furthermore, feature selection methods are frequently employed to focus on the most informative genetic features, reducing dimensionality and enhancing model performance.⁵⁷

Based on genomic data, ML methods have been created to predict the likelihood of developing HCC and hepatocyte remodeling caused by HBV.⁵⁸ These models can precisely determine a person's risk of acquiring HBV-HCC by combining genetic and epigenetic data, which enables early intervention and individualized treatment plans. In other words, patients with HBV-induced HCC and hepatocyte remodeling have new genetic biomarkers that are predictive of prognosis through the use of ML techniques.⁵⁹

The ML methods can find genetic signatures for hepatocyte remodeling and HCC associated with HBV from gene

expression datasets.⁶⁰ ML models can uncover unique biomarkers linked with disease progression and survival outcomes in HBV-HCC patients by examining gene expression profiles and other genomic data.⁶¹ Moreover, ML techniques have completely transformed our knowledge of the genomes of HBV-induced HCC and hepatocyte remodeling, providing the way for the identification of novel molecular targets for the diagnosis, prognosis, and treatment of HBV-HCC.⁶²

ML techniques have also been applied to the identification of particular genomic alterations linked to HBV-induced HCC and hepatocyte remodeling.⁶³ Through the examination of extensive genomic data, ML algorithms can detect patterns and signatures specific to HBV-HCC and hepatocyte remodeling, offering a valuable understanding of the molecular processes that underlie the illness.⁶⁴

To estimate possible genomic and molecular signatures for patients with HBV-HCC and hepatocyte remodeling from large genomic datasets, the following steps should be followed; first, samples of HBV-infected liver tissue that have developed HCC or undergone hepatocyte remodeling can be used to gather gene expression data. The expression levels of thousands of genes in each sample are usually included in this data.⁶⁵

Subsets of genes that are particularly important for differentiating between distinct phenotypes, including normal hepatocytes, modified hepatocytes, and HCC cells, can then be found using feature selection algorithms.⁶⁶ With the least amount of duplication among the features that are chosen, these algorithms seek to identify the genes that are most strongly associated with the desired phenotype.⁶⁷ In this technique, several feature selection algorithms can be used. In other words, using gene expression data, feature selection algorithms are 1 potential machine learning method to find genetic signatures for hepatocyte remodeling and HCC associated with HBV.⁶⁸

Recursive feature elimination (RFE), for instance, is a technique that repeatedly eliminates less significant features until the target number is retained. As an alternative, sparsity in the chosen features can be enforced using L1 regularization techniques like LASSO (Least Absolute Shrinkage and Selection Operator).⁶⁹ Following the identification of a subset of relevant genes through feature selection techniques, machine learning models, such as neural networks, SVM, random forests, gradient boosting machines, and k-means clustering, can be trained on this reduced feature set to categorize new samples into distinct phenotypes according to their gene expression profiles.⁷⁰

A variety of performance indicators, including accuracy, precision, recall, and area under the receiver operating characteristic curve (AUROC), can then be used to assess the trained model. Additionally, using enrichment analysis techniques like gene ontology or KEGG pathway analysis may involve finding important, additional pathways or biological processes linked to the chosen genomic signatures.^{53,71}

It is conceivable to find genomic signatures predictive of remodeling and HCC development by applying this ML approach to genomic data from HBV-associated liver tissue

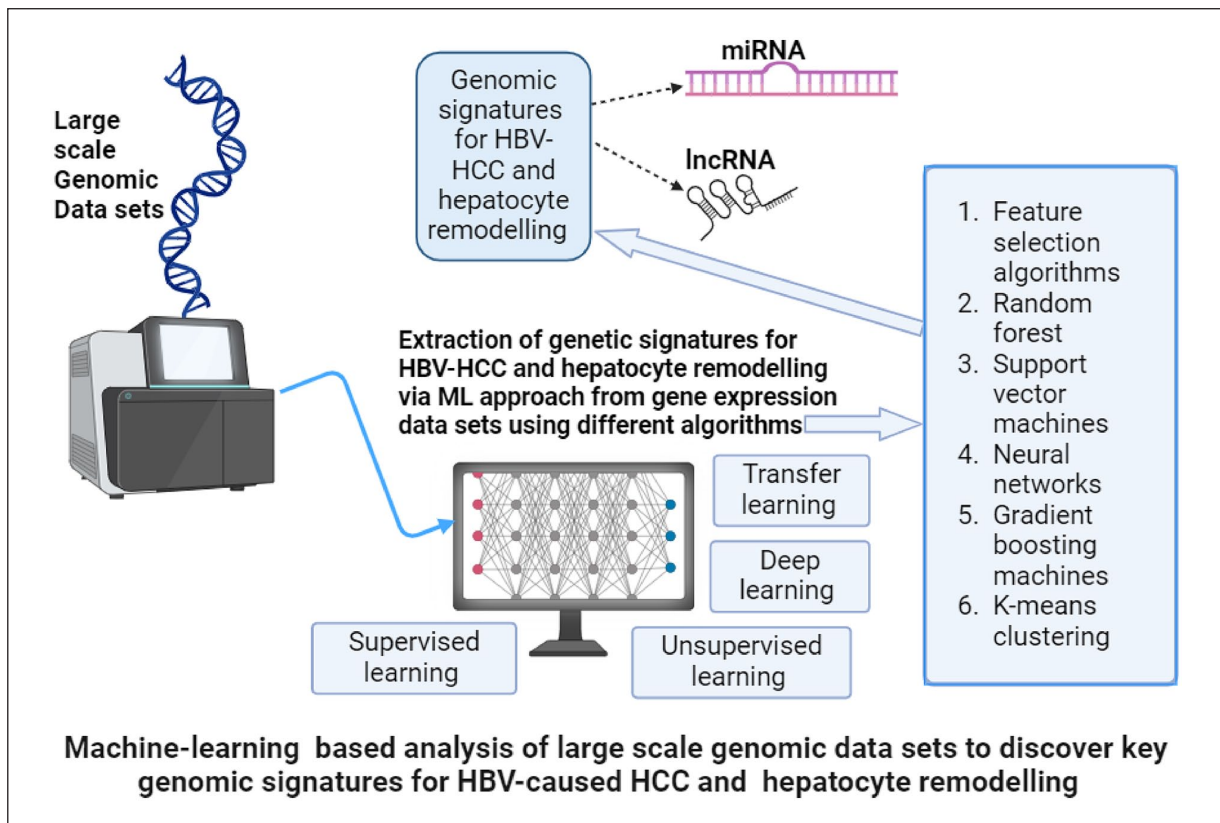


Figure 1. The analysis of large-scale genomic datasets using a machine learning approach to discover key genomic signatures for HBV-HCC and hepatocyte remodeling.

samples with known phenotypes relevant to these diseases.³⁴ These signs may function as biomarkers for the advancement of the disease or as targets for therapeutic interventions, and they may also shed light on the underlying molecular mechanisms underlying HBV pathogenesis. To identify genetic signatures associated with the onset of HCC and HBV-associated hepatocyte remodeling, several machine-learning approaches might be investigated.⁷² Among the techniques that are commonly used is supervised learning, which involves analyzing labeled data to identify molecular markers associated with the progression of HBV-related disease and training a model to predict particular outcomes.⁷³ Another type of ML approach, unsupervised learning, can identify distinct molecular signatures or subtypes linked to HBV-caused liver disorders without the need for predefined labels.⁷⁴

In addition, deep learning uses techniques such as deep neural networks to analyze enormous amounts of genomic data and find complex relationships and patterns that may shed light on the pathophysiology of HBV.^{64,75} Transfer learning, when working with sparsely labeled data for diseases associated with HBV, makes use of information gleaned from related activities or datasets to enhance predictive performance,⁷⁶ and feature selection and dimensionality reduction methods can be used to find important genomic features related to HBV-associated hepatocyte remodeling and HCC development^{77,78} (Figure 1).

ML Approaches Revealed microRNAs and Their Related Genes as Genomic Signatures for Hepatitis B Virus-Associated Hepatocyte Remodeling and HCC

The identification of HBV-related HCC is one possible clinical use for the discovery of miRNAs and their target genes.⁷⁹ Through the evaluation of particular miRNA levels in patient samples, medical professionals could be better equipped to identify HCC and distinguish it from other liver illnesses. This might result in an early diagnosis and course of treatment, which would benefit the patient.⁸⁰ Furthermore, these biomarkers may be useful in prognosticating the course and fate of HBV-related HCC in patients. Clinicians may be better equipped to gauge a patient's response to treatment and the chance of disease recurrence by tracking changes in miRNA levels over time.⁸¹

Moreover, developing focused therapeutic approaches may result from knowledge of the function of particular miRNAs and the genes that they target in HBV-related HCC.⁸² For instance, targeted medicines that attempt to alter the expression or activity of specific miRNAs may be investigated as possible therapeutic options if these miRNAs promote tumor growth or metastasis.⁸³ These miRNAs can be used as targets for innovative therapeutic strategies and can be integrated into the diagnostic and prognostic systems for HBV-related HCC.⁸⁴ A large amount of genomic data,

including microRNA expression profiles, can be analyzed by ML algorithms to find trends and correlations with the course of disease. Through training these algorithms on datasets comprising data from patients with HBV-HCC and healthy controls, researchers may be able to identify particular microRNAs linked to the onset or course of the disease.⁸⁵

A previous study illustrated the use of an ML approach to determine potential miRNA signals from miRNA-mRNA regulatory network modules in HBV-HCC and hepatocyte remodeling via comprehensive analysis. In this study, a regulatory network was built to integrate the regulatory network interactions between miRNAs and mRNAs via the random forest method to rank possible signals.⁸⁶ In other words, through the application of computational approaches to analyze intricate gene-microRNA interactions, scientists can obtain a deeper understanding of the molecular mechanisms that underlie these activities, and this information may result in better HBV-HCC patient diagnosis, prognosis, and treatment regimens.⁸⁷

About 1 ml technique called eXtreme gradient boosting (XGB) can identify significant genes that could lead to HCC.⁸⁸ In accordance with previous studies, model performance was assessed using balanced measures for accuracy, sensitivity, selectivity, positive-predictive value, and negative-predictive value. For the XGBoost model, the accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and *F1* rating were 98.1%, 98.6%, 100%, 97.2%, 94.4%, 100%, and 97.1%, respectively. Genes such as RNF26, FLJ10233, ACBD6, RBM12, PFAS, H3C11, and GKP5 are possible signals for HBV-related hepatocyte remodeling and HCC.³⁴

Moreover, SVM-HCC identified a 23-miRNA signature that was associated with cancer stages in patients with HCC and achieved a 10-fold cross-validation accuracy, sensitivity, specificity, Matthews correlation coefficient, and area under the receiver operating characteristic curve (AUC) of 92.59%, .98, .74, .80, and .86, respectively,⁸⁹ and an accuracy and test AUC of 74.28 and 73%, respectively. A hepatitis B pipeline involving of hsa-let-7i and its 13 integrated miRNAs was identified through a correlation evaluation of the miRNA profile and its jointly expressed miRNAs.⁹⁰

Additionally, possible differentially expressed miRNAs from the Cancer Genome Atlas database were identified via ML model.⁹¹ The 5 of these miRNAs include NK3 homeobox 1 (NKX3-1), forkhead box O1 (FOXO1), nuclear receptor subfamily 4 groups A member 3 (NR4A3), secreted frizzled-related protein 1 (SFRP1), interleukin 6 signal transducer (IL6ST), and NK3 homeobox 1 (NKX3-1).⁹² In the context of HBV-associated HCC, the genes NKX3-1, FOXO1, NR4A3, SFRP1, and IL6ST are important regulators of cellular processes including proliferation, differentiation, apoptosis, inflammation, and tumorigenesis. NKX3-1 is involved in cell proliferation, and its downregulation is linked to tumor progression.⁹³ In HCC, FOXO1 controls apoptosis and prevents tumor growth. NR4A3, which regulates cellular proliferation, is downregulated in HCC.⁹⁴

As an antagonist of the Wnt signaling pathway, SFRP1 downregulation is associated with dysregulated Wnt signaling during the development of HCC.⁹⁵ The interleukin-6

signaling system, in which IL6ST is involved, can stimulate cancer and inflammation in addition to, when dysregulated, liver regeneration.⁹⁶

ML Approaches Revealed Long Noncoding microRNAs (lncRNAs) and Their Related Genes as Genomic Signatures for HBV-Associated Hepatocyte Remodeling and HCC

The identification of lncRNAs from large genomic datasets as genetic markers for hepatitis B virus-associated hepatocyte remodeling and HCC can be accomplished via a variety of ML techniques, including SVM, random forests, and deep learning neural networks.^{97,98} LncRNAs can be grouped according to their expression patterns by clustering algorithms, including k-means clustering, hierarchical clustering, and principal component analysis (PCA).⁹⁹ These algorithms can also be used to find any correlations between lncRNAs, HBV-associated hepatocyte remodeling, and HCC.¹⁰⁰

The most significant lncRNAs that are predictive of HBV-associated hepatocyte remodeling and HCC can be identified via feature selection algorithms.¹⁰¹ Additionally, network-based techniques like gene co-expression networks, protein-protein interaction networks, or regulatory networks can be used to identify putative relationships between lncRNAs and other genes or proteins implicated in HCC and hepatocyte remodeling linked to the HBV.¹⁰²

These ML algorithms may discriminate between known cases of HCC and healthy controls by training them on the expression levels of particular lncRNAs and other biological cues.¹⁰³ After being trained, these models can be used to predict, on the basis of lncRNA profiles, whether newly diagnosed HBV patients are more likely to develop HCC. This method shows promise for enhancing HCC linked to early identification of HBV and tailored therapy.⁴⁸

In a previous study, lncRNAs such as AL356056.2, AL445524.1, TRIM52-AS1, AC093642.1, EHMT2-AS1, AC003991.1, AC008040.1, LINC00844, and LINC01018 were filtered out as possible diagnostic lncRNA biological signals for HCC associated with HBV via ML.¹⁰⁴ Since particular indicators or biological pathways relevant to HBV-HCC development for these genes are still being intensively explored, the functions of these genes in HBV-HCC still need to be precisely described in the literature.¹⁷ In contrast, given their known characteristics and connections with other genes or proteins involved in cellular processes relevant to HCC and HBV, probable roles can be deduced from them. Since some long noncoding RNAs (lncRNAs) have been linked to chromatin remodeling and epigenetic control, lncRNAs like EHMT2-AS1 may be engaged in mechanisms that are comparable to those that lead to the development of HCC.^{105,106}

TRIM52-AS1 may function as an antisense RNA for the TRIM52 gene, which is connected to immunological control and the antiviral response. As a result, it may affect how the host reacts to HBV infection.^{106,107} Each of these

factors might have altered the gene expression linked to chronic inflammation inside the immune milieu of tumors. New information about the specific functions of lncRNAs in HCC associated with HBV was obtained through a functional annotation of the target DEmRNAs.¹⁰⁸

The specificity and sensitivity of the random forest and support vector machine models were 94.3% and 86.5% and 95.7% and 90.4%, respectively, as indicated by their respective areas under the curve (AUCs). The integrated co-expressed DEmRNAs matched similar pathways in the retinol metabolism, PI3K-Akt signaling cycle, biological cancer development, and p53 signaling processes, according to the findings of the functional enrichment investigation.¹⁰⁹

Non-coding RNA molecules known as “snoRNAs” are mostly involved in the processing and alteration of other RNA molecules. They play a role in several biological functions, including splicing, gene control, and ribosomal RNA processing.¹¹⁰ The specific functions of different snoRNA species might, however, differ greatly from one another.

The activities of TEP SNORD12B, SNORD14E, and SNORA63 in the context of HBV-HCC are not well understood.¹¹¹ These snoRNAs may play a role in controlling the production of genes or altering particular target RNAs related to HBV infection or the growth of HCC. TEP from HBV-associated liver cancer up-regulated SNORA63 and down-regulated SNORD12B and SNORD14E, which may serve as diagnostic biomarkers for both the early stages of the disease and HBV-induced HCC.¹¹²

Additionally, when paired with platelet characteristics, TEP SNORD12B, SNORD14E, and SNORA63 improve the testing efficacy of AFP and obtain good diagnostic effectiveness for HBV-induced HCC. SNORA63, SNORD14E, and SNORD12B, which are abnormally expressed TEPs may be new, non-invasive indicators for identifying HBV-related HCC.¹¹³

One of the snoRNAs that is most obviously elevated in HBV-caused HCC was SNORD12B, and there is a strong correlation between elevated SNORD12B expression and unfavorable patient prognosis.¹¹² Significantly, SNORD12B aids in the development, circulation, assault, and dissemination of tumor hepatocytes both in vivo and in vitro during chronic hepatitis B virus infection, demonstrating its oncogenic character.¹¹³ SNORA18L5 enhances the likelihood of HBV-related hepatocellular carcinoma by changing where ribosomal proteins are located and lowering p53 expression. Furthermore, through the miR-30a-5p/REEP3 pathway, CircFAT1 accelerates the expansion of cancer in hepatocytes. Circulating miRNA-483-5p, 21, and 155 may be novel prognostic and early diagnostic biomarkers for HCC, according to machine learning algorithms.¹¹⁴

Among patients with reoccurring HCC, the genes CETN2, HMGA1, RACGAP1, and SNRPB were elevated, and they can serve as prognostic biomarkers to predict the recurrence of HCC following surgical removal.¹¹⁵ HBV-HCC has also been linked to 3 genes, such as TOP3B, SSBP3, and COX7A2L. Previous studies emphasized the significance of biomarkers including COX7A2L, SSBP3, and TOP3B, which were continuously significant in a variety of models, indicating their potential to enhance the prediction accuracy for HCC prognosis.^{52,116}

A component of cytochrome c oxidase, which is vital to both oxidative phosphorylation and the electron transport chain, is COX7A2L.¹¹⁷ It is essential to the creation of cellular energy because it makes it easier for electrons to move from cytochrome c to molecular oxygen, which in turn drives the synthesis of ATP. Changes in COX7A2L expression or function may affect the pathways involved in cellular metabolism and energy production in HBV-HCC.¹¹⁸

Comparably, SSBP3 genes encode a single-stranded DNA-binding protein that is involved in several metabolic functions related to DNA, including transcriptional control, repair, recombination, and replication. Its precise function in HBV-HCC is unknown; however, it might be related to controlling the processing or stability of viral or host genetic material.¹¹⁹

Moreover, TOP3B encodes the beta enzyme DNA topoisomerase III, which catalyzes modifications to DNA molecules' topological states during cellular functions like replication, recombination, and repair.¹²⁰ Changes in TOP3B activity or expression in HBV-HCC may influence genomic stability by affecting the resolution or processing of complex DNA structures.¹²¹

On the other hand, applying a multivariate Cox regression model, the predictive biomarkers YWHAB, PPAT, and NOL10 were ultimately found. YWHAB is involved in several biological functions, including signal transduction, cell cycle regulation, and apoptosis.¹²² Through its interactions with viral proteins or cellular components linked to HBV replication and the formation of HCC, YWHAB facilitates the growth and progression of tumors in the context of HBV-HCC.¹²³ Moreover, YWHAB controls the activity of important carcinogenic pathways like the PI3K/AKT/mTOR pathway, which is commonly dysregulated in HCC.¹²⁴

An enzyme called phosphoribosyl pyrophosphate amidotransferase (PPAT) is involved in the metabolism of nucleotides and purines. In HBV-HCC, PPAT plays a role in the elevated need for nucleotides necessary for the fast growth of tumor cells brought on by HBV infection.¹²⁵ Infected hepatocytes with overexpressed PPAT have increased de novo purine production, which supports these needs and encourages tumor growth. Additionally, through its possible role in altering signaling pathways associated with cell survival or proliferation, PPAT may have non-canonical effects on the course of cancer.¹²⁶

Nucleolar Protein 10 (NOL10) is mostly found in the nucleolus, where it plays a role in ribosome synthesis, which is essential for regulating cellular development and responding to stress brought on by long-term infections such as HBV.¹²⁷ NOL10 influences ribosome biogenesis, which affects translation control within cells and results in the uncontrolled growth characteristic of cancer cells observed at advanced stages following long-term chronic hepatitis B infection. This may be a factor contributing to malignant transformation in HBV-HCC pathogenesis.¹²⁸

The control over the methylation of the YWHAB, PPAT, and NOL10 promoters differed in malignant hepatocytes and cancers. The examination of the immune invasion indicated that B cells and dendritic cells were probably responsible for the aberrant modulation of the aforementioned

signals.¹²⁹ These indicators' important role in the growth of HCC was demonstrated by their gene set variation analysis (GSVA). Three hub genes and our findings imply that the circTMC03/miR-577/RHOA, circHMGCS1/miR-892a/KIF5B, and circHMGCS1/miR-581/AURKA pathways may be essential for the development of HCC. Both in vitro and in vivo carcinogenesis were significantly reduced when the expression of these genes was specifically inhibited.¹³⁰

A transcriptomics study revealed that GABRR1, SOX11, COL24A1, and MYLK2 were expressed in the Cancer Genome Atlas (TCGA) dataset. These genes were markedly up-regulated in cancer cells. In particular, Kaplan-Meier analysis revealed an unfavorable relationship between the mRNA levels of these genes and the overall survival (OS) and progression-free survival (PFS) of HCC patients.¹³¹⁻¹³⁴ These genes can predict the longevity of HCC patients. Additional research revealed the significance of the GABRP, HBG1, and DAK (TKFC) genes in HCC.^{135,136} BIRC5, CDC20, and UBE2C demonstrated substantial associations with the OS. As a result, 3 hub genes, BIRC5, CDC20, and UBE2C, were discovered and confirmed to be associated with the course and outcome of HCC.^{137,138} These factors might be possible HCC treatment targets. According to a survival study, increased GLMP, SLC38A6, and WDR76 mRNA expression levels were linked to poor prognosis, and the combined expression of these 3 genes was found to be a standalone predictor of outcome in patients with HBV-HCC.¹³⁹

Five cancer suppressors that were confirmed to be hsa-miR-183-5p target genes (AKAP12, DYRK2, FOXN3, FOXO1, and LATS2) had their expression restored by LNC-HC.¹⁴⁰ All things considered, human LNC-HC was found to be a new tumor suppressor that, through competitively binding hsa-miR-183-5p as a competing endogenous RNA (ceRNA), could decrease HCC cell expansion in vitro and limit tumor development in vivo. These results imply that LNC-HC may function as an indicator for HCC and offer an intriguing therapeutic target for the disease's management.⁸⁵

Several clinical variables in the emergence of HCC may be associated with higher levels of CASC7 and the levels of systemic CASC7 and FOXA1 are anticipated to grow in significance as a diagnostic tool for HCC and a way to track the course of the illness.¹⁴¹ Especially in male individuals with hepatocellular carcinoma, FOXA1 suppresses PIK3R1 activation to slow the progression of the cancer. A tumor suppressor that is inhibited in HCC, miR-212-3p, may be able to block the expression of FOXA1. FOXA1 controlled the expression of AGR2, which is how it carried out its biological role. Low levels of miR-212-3p might be attributed to FOXA1's aberrant overexpression.¹⁴²

Furthermore, the production of CASC7 had an adverse association with the generation of miR-30a-5p in HCC tumor hepatocytes. Additionally, CASC7 and miR-30a-5p modulated KLF10, a target of miR-30a-5p. By coupling to miR-30a-5p, CASC7 controlled the KLF10/TGF- β /SMAD3 route, which enhanced the growth of HCC cells.¹⁴²

To identify diagnostic biomarkers for HBV-related hepatocyte remodeling, other algorithms such as LASSO

regression, SVM-RFE, RF, and WGCNA are employed.¹⁰⁸ The genes encoding versican (VCAN) and phosphatidic acid phosphatase type 2C (PPAP2C) were shown to be potentially useful diagnostic indicators for HBV-hepatocyte remodeling based on these investigations, along with their performance characteristic curves.²⁷ Moreover, SPINK1 was also shown to be a standalone predictor of overall survival for patients with HCC by multivariate analysis.¹⁴³

By utilizing ML, it is now possible to examine novel forms of clinical data to identify HCC. Genes including FCN3, CLEC1B, and PRC1 were linked with overall survival, progression-free survival, and disease-free survival, according to the large-scale transcriptome profiling data analysis.¹⁴⁴ Using ML techniques, important therapeutic targets associated with HCC in its early and late stages were also discovered, and vitronectin, lactate dehydrogenase D (LDHD), thrombin-activable fibrinolysis inhibitor, and miR-590 were among the key players in the early stage of HBV-HCC, while regucalcin, the SPRY domain containing 4, miR-3199-1, miR-194-2, and miR-4999 were involved in the late stage of HBV-HCC.¹⁴³

Unique transcriptome signatures can be discovered with the use of ML techniques in conjunction with RNA sequencing data. When it came to distinguishing between the HCC and normal cell models, the 3 signals that were found, PARP2-202, SPON2-203, and CYREN-211, showed the greatest accuracy compared to all additional transcriptomes.¹⁴⁵ The results of this investigation's approach can be applied to any RNA sequencing dataset to identify new transcriptional indicators.¹⁴⁶

PARP2-202 may contribute to the survival and proliferation of HCC cells by repairing DNA and maintaining genomic stability. It also interacts with biological components linked to HBV replication or viral proteins.¹⁴⁷ Conversely, it has been proposed that SPON2-203 contributes to tumor growth and dissemination and that elevated levels of this protein are associated with a worse prognosis.¹⁴⁸ It has been suggested that CYREN-211 may have a role in the development of tumors by modifying the course of the cell cycle through interactions with important regulatory proteins that govern the G1/S transition.¹⁴⁹

Research has shown that human cartilage contains a cellular matrix glycoprotein called cartilage oligomeric matrix protein (COMP), which controls cell phenotype throughout tissue development and transformation.¹⁵⁰ It has been observed that COMP does not circulate in normal hepatic tissues and that its amount is markedly elevated in HCC cells. COMP worked with CD36 and, afterward, was critical for the advancement of MEK/ERK and PI3K/AKT-mediated HCC. Increased COMP is a useful indicator of the likelihood of developing HCC as well as a diagnostic marker for the disease.¹⁵¹

Hepatic stellate cells generate a glycoprotein known as Mac-2 attaching protein (M2BP), which is extensively glycosylated to form the Mac-2-bound protein glycosylation isomer (M2BPGi). As a mediator linking HSCs and Kupffer cells, M2BPGi may collaborate with extracellular matrix protein molecules such as galectin-3 (Mac-2) and has been demonstrated to help determine the extent of cirrhosis in

long-term liver disease. According to recent research, M2BPGi may be able to predict the progression of HCC.¹⁵²

Furthermore, according to the prediction model, combining the data from serum Gal-3BP and alpha-fetoprotein (AFP) increased the accuracy, specificity, and sensitivity of the initial HCC diagnosis from 90.63% to 95%, 93.75% to 95%, and 87.5% to 95%, respectively. According to these findings, serum Gal-3BP level is a possible indicator for identifying HBV-induced HCC, and when serum Gal-3BP and AFP are used together, as opposed to AFP alone, in current clinical practice, the ability to diagnose HBV-HCC is improved.^{86,153}

In HBV-HCC and hepatocyte remodeling, sex-specific gene biomarkers can also be found using machine learning techniques. It is well recognized that there are gender variations in HBV-caused liver cancer. Researchers can learn more about the underlying molecular pathways that underlie the onset, development, and progression of HBV-HCC as well as hepatocyte remodeling by finding sex-specific genetic signatures.¹⁵³

A recent study on sex-specific transcriptome analysis reveals gender-related dysregulated pathways in human HCC, highlighting the usefulness of machine learning in the identification of sex-specific gene biomarkers. To find gender-related dysregulated pathways in HCC patients, the authors of this work employed a cross-validation strategy in conjunction with a machine-learning technique known as Random Forests.⁸⁶

In addition, examining the sex-related variations in the expression and operation of predicted genes implicated in hepatocyte remodeling and HCC linked to HBV is a fascinating area of study.¹⁵⁴ It is becoming better acknowledged that sex has a biological role in liver illnesses, such as HCC and HBV infection. In several liver illnesses, sex-specific variations in immune responses, gene expression, and disease progression have been noted.¹⁵⁵

Compared to women, men are more likely to be diagnosed with and experience aggressive HCC. A growing body of research indicates that the protective effects of estrogen and the stimulatory effects of androgens in the development and progression of HCC appear to mediate the sex discrepancy.¹⁵⁶ The effect of sex hormones on the transactivation of the hepatitis B virus X protein and the release of inflammatory cytokines has been the primary focus of studies on the sex difference in HCC over the past few decades, and these studies have recently become even more intense.¹³⁹ Sex hormones connect to certain cellular receptors in hepatocytes and alter the related signaling pathways, which in turn affect genetic modifications and DNA damage repair.¹⁵⁷

Certain genes may be expressed or regulated differently in males and females in the setting of HBV-associated hepatocyte remodeling and HCC. Comprehending these variations can offer valuable perspectives on the molecular processes that underlie gender differences in illness susceptibility, progression, and response to treatment.¹⁵⁸

Examining extensive omics data (transcriptomics or proteomics) via an ML approach from patients with HBV-related liver disorders, both male and female, may identify

putative sex-specific gene signatures linked to carcinogenesis and hepatocyte remodeling.¹⁵⁹ Further understanding of the underlying mechanisms causing sex inequalities in these illnesses may come from examining the effects of sex hormones on host immunological responses, HBV replication, and tumor formation.¹⁶⁰

Using a sex-specific differential expression analysis of the tumor and tissues close to the tumor, males showed PI3K, PI3K/AKT, FGFR, EGFR, NGF, GF1R, Rap1, DAP12, and IL-2 signaling pathway enrichment, while both sexes showed activation of pathways related to apoptosis and cell cycle identified genes and pathways that are relevant to the etiology and that differentiate male from female HCC.¹⁶¹⁻¹⁶³

Moreover, the incidence of HCC is greater in men than in women. According to a recent study published in the gut, the development of HCC is significantly inhibited by CYP39A1, an autosomal gene specific to the liver that expressed preferentially in females.¹⁶⁴ According to the expression quantitative trait loci (eQTL) analysis, males, and females differ in the activation of several signaling pathways.¹⁴⁰ Sex has a significant role in modulating the effects of eQTLs in HCC, as seen by the 24.3% of identified eQTLs that show differential effects between the sexes. The genes exhibiting sex-specific dysregulation in tumors and those carrying a sex-specific eQTL converge in clinically significant pathways, implying that differential genetic effects on gene expression play a role in the molecular etiologies of male and female HCC.^{165,166}

Sex-specific molecular etiologies of HCC are identified via a sex-stratified approach. These findings provide a fresh understanding of how inherited genetic transcription control influences sex differences in the etiology of HCC and establish a foundation for further research on sex-biased malignancies.¹⁶¹ In general, investigating the sex-specific features of gene expression patterns in HCC and HBV-associated hepatocyte remodeling may provide significant biological insights with ramifications for patient-specific personalized medicine strategies for both male and female patients.^{161,167}

In the ML and bioinformatics-based search for biomarkers for HBV-HCC, multi-center longitudinal studies are essential for several reasons. First, longitudinal studies enable the monitoring of biomarker performance over time, which is crucial for understanding biomarker predictive capabilities at various stages of disease progression.¹⁶⁸ This approach enhances the reliability of results by establishing temporal relationships between biomarker levels and the development of HCC.¹⁶⁹ Furthermore, these investigations are essential for validating new biomarkers identified by ML algorithms. Although many biomarkers have shown promise in early research, they have not been sufficiently validated across diverse cohorts.²⁹ For instance, the Early Detection Research Network (EDRN) of the National Cancer Institute emphasizes the importance of extensive biorepositories and longitudinal data collection to facilitate the discovery and validation of novel biomarkers in high-risk populations¹⁷⁰ (Table 1).

Table 1. Role of major gene signatures in HBV-induced hepatocyte remodeling and HCC.

Group of gene signatures	The activities of gene signatures in HBV-induced hepatocyte remodeling and HCC	Ref.
SNORD12B, SNORA63, SNORD14E, CircFAT, and SNORA18L5	They improve the effectiveness of the AFP test for HBV-induced HCC and are indicators for patient prognosis and poor treatment outcomes. Furthermore, their elevated expressions help the progression, migration, invasion, and dissemination of malignant liver cells.	Zhao et al, ¹¹³ Lux and Bullinger ¹⁷¹
CETN2, HMGA1, RACGAP1, and SNRPB	They serve as prognostic biomarkers to predict the re-appearance of HCC following surgical removal and anti-viral therapy.	Li et al ¹⁷²
COX7A2L, SSBP3, and TOP3B	They are potential enhancers to predict the responsiveness of HBV-HCC medications.	Lobo-Jarne et al ¹⁷³
circTMCO3/miR-577/RHOA, circHMGCS1/miR-892a/KIF5B, and circHMGCS1/miR-581/AURKA	They enhance the carcinogenesis of HBV-induced HCC.	Li et al ¹⁷⁴
YWHAB, PPAT, NOL10, S100A9, CHEK1, KIF11, PBK, and MCM3	They provoke the advancement of the chronically infected hepatocytes to HCC because of the ongoing inflammatory responses from immune cells and are important for the prompt identification of HBV-HCC.	Li et al, ¹⁷⁴ Quinn and Chang ¹⁷⁵
GABRR1, SOX11, COL24A1, and MYLK2)	They can forecast and predict the longevity of HCC patients. Thus, they indicate the poor overall survival of patients with HBV-caused HCC.	Liang et al ¹⁷⁶
GABRP, HBG1, and DAK (TKFC)	They are involved in the encroachment, re-occurrence, and tumorigenesis of HCC.	Huang et al ¹⁷⁷
BIRC5, CDC20, and UBE2C	They are associated with the course and outcome of HCC. These might end up being possible HCC treatment candidates.	Ren and Feng ¹⁷⁸
GLMP, SLC38A6, and WDR76	They show a bad prognosis and are the standalone predictors of treatment efficacy in HCC patients.	Yu et al ¹⁷⁹
AKAP12, DYRK2, FOXN3, FOXO1, and LATS2	They suppress the progression of hepatocellular carcinoma cells by decoding hsa-miR-183-5p. Hence, they decrease the expansion of malignant hepatocytes and have an anti-cancer effect.	Takahashi et al ¹⁸⁰
CASC7, FOXA1	They suppress the proliferation of hepatocellular carcinoma by focusing on microRNA-30a-5p. They also suppress PIK3R1 activation to slow the progression of the cancer.	Verma et al, ¹⁸¹ Moyo et al ¹⁸²

Abbreviations: HBV: hepatitis B virus; HCC: hepatocellular carcinoma.

Table 1 summarizes the major genomic signatures with their respective functions, which were identified by the machine learning approach.

Bioinformatics Analysis Revealed Genomic Signatures for Hepatitis B Virus-Associated Hepatocyte Remodeling and HCC

By combining several omics datasets (genomics, transcriptomics, proteomics, and metabolomics), using a variety of statistical techniques, and utilizing ML algorithms, bioinformatics analysis facilitates the identification of important genetic signatures for HBV-HCC and hepatocyte remodeling.¹⁷⁸ Using a comprehensive approach, it is possible to find potential biomarkers for diagnostic or treatment targets in addition to learning more about the molecular pathways behind the development and course of the disease¹¹³ (Figure 2).

Bioinformatics tools are crucial for identifying genomic biomarkers associated with HBV-HCC by leveraging extensive data from genomic databases such as the National

Center for Biotechnology Information (NCBI), GEO, and TCGA.⁵² These tools facilitate the identification of DEGs through differential RNA-Seq analysis using libraries like the limma and edgeR packages in R.¹⁸³

Open-source software and repositories, including cBioPortal, ArrayExpress, and GEO, enable comprehensive cancer omics data analysis. Additionally, web servers such as gene expression profiling interactive analysis (GEPIA), tumor-infiltrating immune cell (TIMER), and OncoPrint support the analysis of associations between clinical factors and gene expression in tumors, aiding in the identification of prognostic biomarkers.¹⁸⁴ Furthermore, pathway analysis, network analysis, and multi-omics data integration enhance biomarker discovery and patient prognosis prediction. Algorithms for sequence alignment and variant calling are essential for identifying genetic variations and mutations while examining gene functionalities.¹⁸⁵

Furthermore, numerous bioinformatic investigations have led to the identification of prospective miRNAs with varying expression patterns and target genes.¹⁸⁰ A possible miRNA-mRNA regulatory loop including miR-93-5p-JUN/STAT3 (Jun proto-oncogene, AP-1 transcription factor subunit/signal transducer and activator of

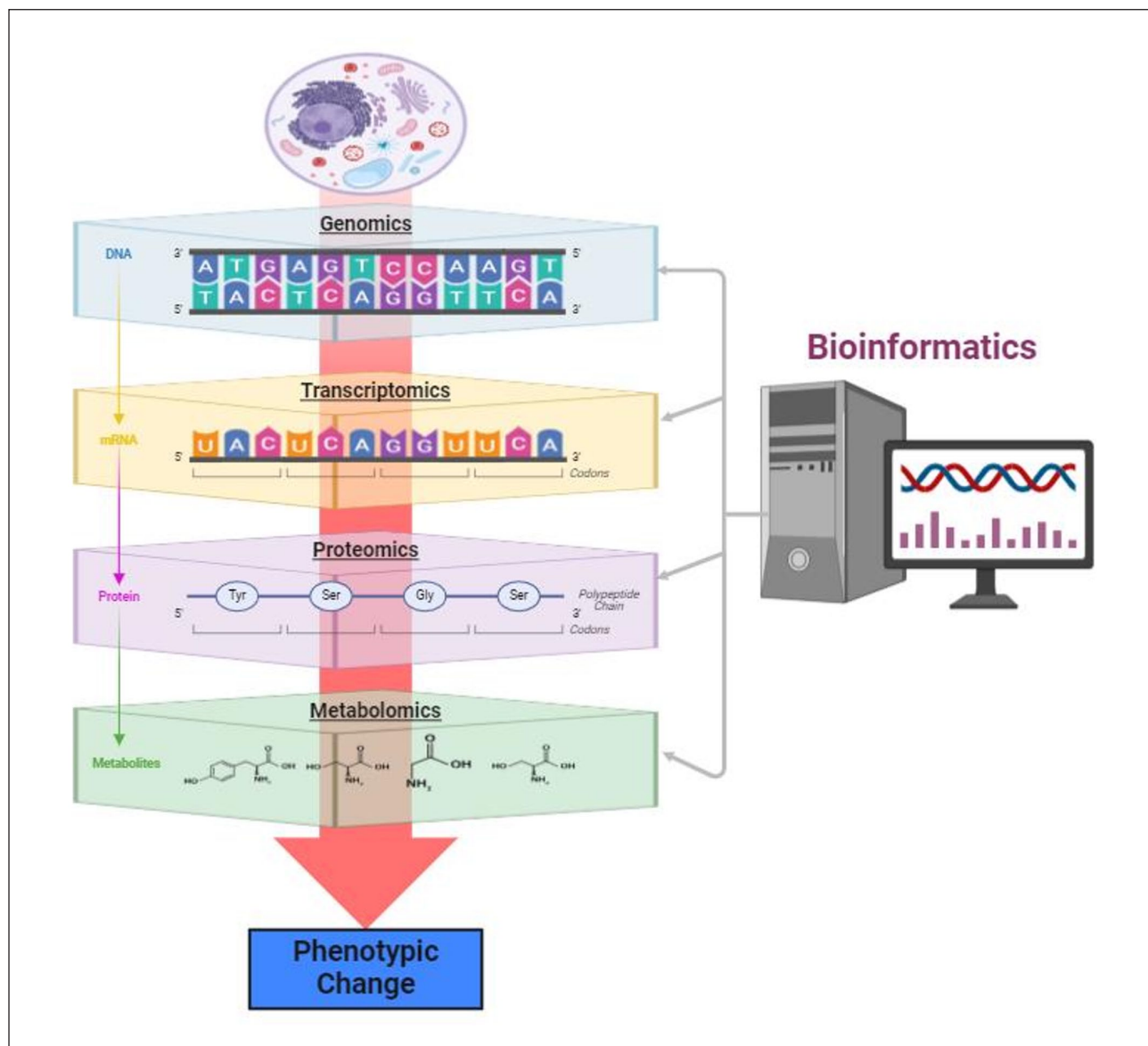


Figure 2. The analysis of large-scale multi-omics datasets via a bioinformatics approach to discover key genomic signatures for HBV-HCC and hepatocyte remodeling.

transcription 3) pathway, miR-106b-5p-STAT3 pathway, miR-21-5p-STAT3/PIK3R1 (phosphoinositide-3-kinase regulatory subunit 1) pathway, miR-125b-5p-E2F2/E2F3 pathway, and miR-let7c-5p-NRAS (neuroblastoma ras viral oncogene homolog) pathway has been developed by matching miRNA-mRNA pairs in the miRNet dataset.^{112,186}

Furthermore, elevated expression of the miRNAs miR-130b-5p, miR-320d, miR-483-3p, miR-1246, miR-320b, miR-192-5p, miR-4532, miR-320c, miR-483-5p, and miR-122-5p may affect expression across various HCC pathways, according to KEGG bioinformatic analysis (mirPATHv4), which distinguishes verified targets and pathways in HCC. Furthermore, the overexpression of miR-483-5p regulates the expression of PPAR α /TIMP2 and CDK15 (peroxisome proliferator-activated receptor alpha/tissue inhibitor of metalloproteinases 2 and cyclin-dependent kinases) target genes.¹⁸⁷

During HBV-HCC and hepatocyte remodeling, the miRNA-mRNA regulatory network is performed by RNA polymerase enzymes and controlled by the chromatin environment, nucleosome occupancy, histone modifications, transcription factor availability, and regulatory elements.¹¹³ Moreover, researchers can find genomic markers for HBV-HCC and hepatocyte remodeling by using bioinformatics analysis of the miRNA-mRNA regulatory network, which also sheds light on the underlying molecular mechanisms behind the progression of the disease. The ramifications of these results extend beyond diagnosis, prognosis, and possible treatment therapies aimed at certain dysregulated pathways.⁵² During miRNA-miRNA interaction, Bowtie, Burrows-Wheeler Aligner (BWA), SAMtools, and Genome Analysis Toolkit (GATK) are frequently utilized algorithms for bioinformatics analysis.¹⁸⁸ BWA is a tool for mapping low-divergent sequences to a large reference genome, while Bowtie is a short-read aligner used to align sequencing

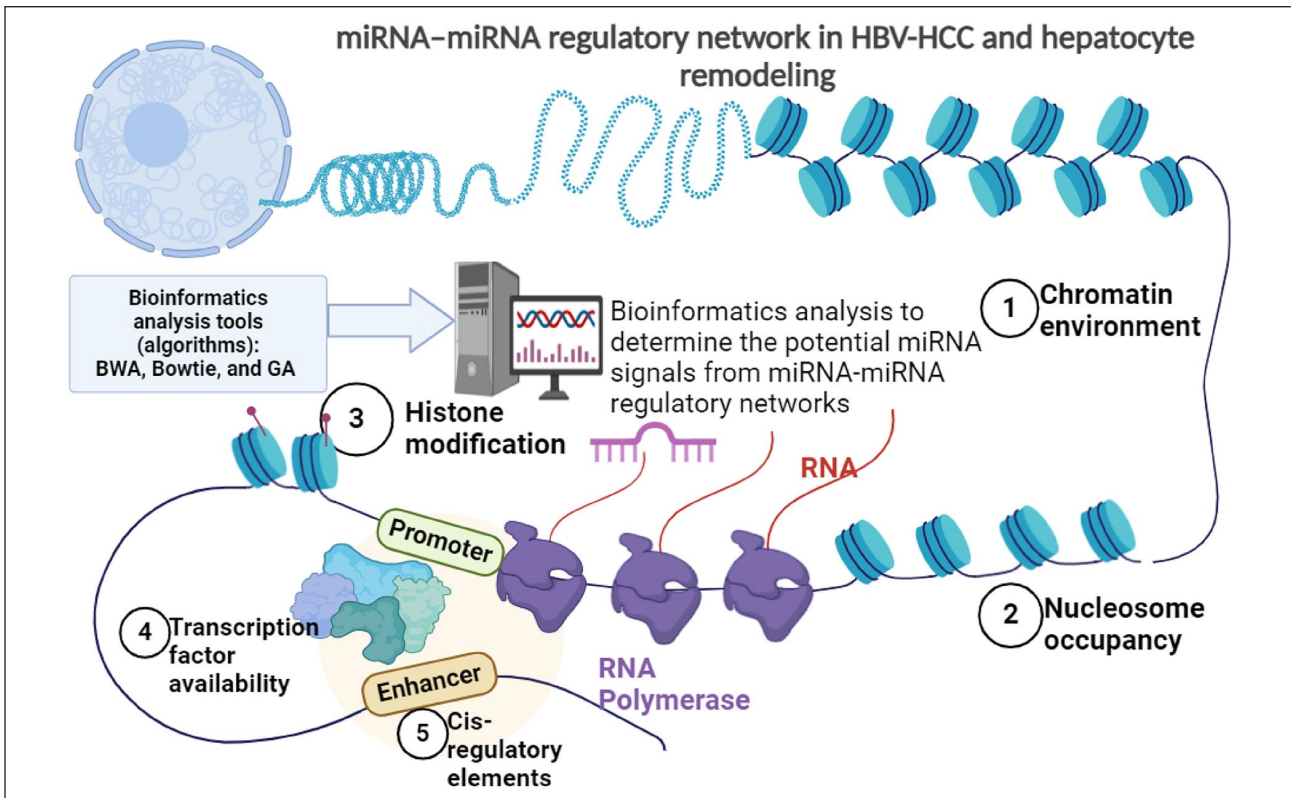


Figure 3. The application of bioinformatics analysis to determine the potential miRNA signals for HBV-HCC and hepatocyte remodeling from miRNA-miRNA interaction networks.

reads to a reference genome.¹⁸⁹ In addition, the software package GA was created at the Broad Institute to examine next-generation sequencing data, and SAMtools is a suite of programs for working with high-throughput sequencing data¹⁹⁰ (Figure 3).

In other words, the bioinformatics study aligned with the expression verification of the 5 miRNAs. In particular, the prognostic value of hsa miR-10b-5p and hsa miR-10b-3p in patients with HCC is high. Therefore, the 5 differentially expressed miRNAs could be used as medical indicators for patients with HBV-HCC. Furthermore, the carcinogenesis of HCC may be influenced by the varying degrees of expression of the targets of each of these 5 mRNAs, which include SFRP1, EDNRB, NR4A3, FHL2, NKX3 1, IL6ST, and FOXO 1.¹⁹¹

Furthermore, bioinformatic analysis-based simultaneous evaluation of erratically methylated DEGs may offer novel insights into the epigenetic process underlying HCC. Hub genes such as S100A9 (S100 calcium-binding protein A9), CHEK1 (checkpoint kinase 1), KIF11 (kinesin family member 11), PBK (PDZ binding kinase), and MCM3 (minichromosome maintenance complex component 3) may function as sensors to enable an accurate diagnosis of HCC.¹⁹²

Moreover, according to earlier research, the ubiquitin-conjugating enzyme E2C (UBE2C) is implicated in HBV-HCC as a carcinogen.¹⁴⁴

A thorough bioinformatic analysis was initially used to examine UBE2C activation in HCC. The diagnostic and predictive functions of UBE2C in the hepatocarcinogenesis of HBV-caused HCC, as well as the UBE2C promoter

methylation rate and upstream regulatory miRNAs of UBE2C in HCC, were evaluated via receiver operating characteristic (ROC) curve modeling and survival analysis.¹⁹³

UBE2C was substantially more abundant in patients HCC than in normal controls. In HCC, there was also a clear decrease in UBE2C promoter methylation, which was inversely linked with UBE2C mRNA expression. The combination of an in vitro experiment and a bioinformatic correlation analysis suggested that hsa-miR-193b-3p could be a significant additional upstream regulatory pathway of UBE2C in HCC. These findings suggests that UBE2C is elevated in HCC and could be a crucial biomarker for prognosis and diagnosis in individuals with HCC.¹⁹⁴

A previous study also screened for miRNA gene variants that indirectly target mRNA transcripts via the regulation of hepatitis B infection by characterizing the functional effects of single nucleotide polymorphisms (SNPs) on mRNA-miRNA and miRNA-binding region interactions and most SNPs were found in Has-miR-139-3p.¹⁹⁵

Specific changes were observed in miRNAs associated with HBV-related HCC, including miR-150, miR-342-3p, miR-663, miR-20b, miR-92a-3p, miR-376c-3p, and miR-92b. On the basis of GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis, these HBV-related HCC miRNAs are predicted to regulate transcription, the promoter region of RNA polymerase II, focal binding, activation of proteins via the MAPK (mitogen-activated protein kinase) signaling pathway, and the actin superstructure. Additionally, according to IPA (ingenuity pathway analysis), these miRNAs have a substantial

effect on HCC incidence and HBV infection through their actions on AGO2 (argonaute RISC catalytic component 2), TP53 (tumor protein53 gene), and CCND1 (cyclin D1) genes.¹⁹⁶

In other words, miRNA-mRNA axes were found in HCC caused by HBV. For HBV-related HCC, hsa-let-7c-3p/CKS2 (cyclin-dependent kinase regulatory subunit 2), hsa-mir-195-5p/CDK1, and hsa-mir-5589-3p/CCNB1 may be used as predictive markers and treatment strategies.¹⁹⁷ Previous investigation has shown that let-7a, miR-34a, and miR-199 a/b function as cancer silencers and have unique interactions in several immunological pathways. Hence, they serve as significant HBV-associated signatures.¹⁷⁴

Furthermore, in HBV-induced HCC, miRNA-99a-5p is erratically generated and may alter the metabolic processes of hepatocellular carcinoma cells. Compared with healthy tissue, cancerous tissue expresses this gene at a substantially reduced level. According to the KEGG and GO analysis, miRNA-99a-5p operates through several pathways. ROC analysis also revealed that it has much promise for prognostic prediction in HCC.¹⁹⁸

The findings of the upregulation and knockdown experiments verified that miRNA-99a-5p could prevent malignant hepatocytes from proliferating, suggesting that it might play a significant role as a tumor suppressor in HCC.¹⁹⁹ These findings suggest that miRNA-99a-5p is adversely associated with the advancement of HCC in HBV-infected individuals and may serve as an innovative treatment option for HCC.²⁰⁰

According to the clusterProfiler R package, KEGG, and GO enrichment studies of DEGs, the aberrant expression of DEGs, including CDK1 (cyclin-dependent kinase 1), CDC20 (cell division cycle protein 20), CCNB1 (cyclin B1), CENPF (centromere protein F), and MAD2L1 (mitotic arrest deficient 2 like 1) is instructive for early detection, tumor phase classification, and prediction of the poor prognosis of HBV-caused hepatic malignancy.²⁰¹

The results of the GO enrichment study demonstrated that these DEGs had substantial increases in molecular functions, such as binding to calcium ions, protein kinases, DNA, and heme, and in physiological processes such as division, growth, redox processes, immune response, and proteolysis. These DEGs were shown to be considerably enriched in the framework of KEGG and GO investigations for the development of cells, oocyte meiosis, metabolic routes, antibiotic biosynthesis, and the p53 pathway.²⁰²

On the other hand, the development and progression of HBV-related HCC may also be significantly influenced by CDK1, KIF11 (kinesin family member 11), KIF20A (kinesin family member 20A), CCNB1, NDC80 (NDC80 kinetochore complex component), MCM2 (minichromosome maintenance complex component 2), RFC4 (replication factor C subunit 4), and RRM2 (ribonucleotide reductase regulatory subunit M2).²⁰³

KIF11 and KIF20A are involved in cell division and mitosis, and CDK1 is involved in the regulation of the cell cycle. Furthermore, CCNB1 is a crucial cell cycle regulator. Furthermore, chromosomal segregation during mitosis is linked to NDC80.^{204,205} Moreover, MCM2 plays a role in the start of DNA replication. Furthermore, RRM2 is

involved in nucleotide synthesis, whereas RFC4 is involved in DNA replication and repair activities. All of these proteins work together to cause cellular processes to become dysregulated, which may result in uncontrolled development traits that are subsequently observed on in the course of a long-term, chronic hepatitis B infection. Thus, these proteins play a significant role in the initiation and advancement of HBV-associated HCC.²⁰⁶ Furthermore, they may also serve as possible targets for therapy and screening biomarkers for HBV-associated HCC, according to PPI (protein-protein interaction) network modeling.²⁰⁷

In addition, on the basis of integrated bioinformatics and survival analysis, the poor overall survival and pathogenesis of HCC patients were linked to elevated gene expression of CDC20, MAD2L1, BUB1 (mitotic checkpoint serine/threonine kinase), BUB1B, AURKA (aurora kinase A), TOP2A (topoisomerase II alpha), CCNB2, TPX2 (TPX2 microtubule nucleation factor), CDC7, CDC20, and MCM3 (mini-chromosome maintenance3).²⁰⁸⁻²¹⁰

CDC20, MAD2L1, BUB1, BUB1B, AURKA, TOP2A, CCNB2, TPX2, CDC7, and MCM3 are all proteins with significant biological roles in HBV-HCC. Cell cycle progression is regulated in part by CDC20.²¹¹ During cell division, MAD2L1 facilitates appropriate chromosomal segregation by participating in the spindle assembly checkpoint.²¹² Additionally, BUB1 and BUB1B are crucial for appropriate chromosomal segregation during mitosis.²¹³

AURKA controls centrosome the development and the production of mitotic spindles. In both chromosomal condensation and DNA replication, TOP2A is essential. Another important cell cycle regulator that manages the start of mitosis is CCNB2.^{214,215} TPX2 plays a role in spindle assembly by nucleating microtubules.²¹⁶ CDC7 plays a crucial role in the initiation of DNA replication.²¹⁷ Moreover, MCM3 plays a crucial role in the pre-replicative complex, which controls the replication of DNA.²¹⁸

Collectively, these proteins are involved in dysregulated cellular processes that could result in the uncontrollable growth characteristics observed at advanced stages after a protracted, chronic infection with hepatitis B. Therefore, these proteins play a significant role in the onset and advancement of HBV-associated HCC.²¹¹

A high Tripartite Motif-Containing 28 (TRIM28) expression level was linked to T categorization, clinical stage, histopathological grade, and circulating AFP levels, according to genomics and empirical analysis, and it was found to be a distinct risk indicator for an unfavorable outcome in patients with HBV-caused HCC.²¹⁹ TRIM28 expression closely linked to the mechanism of ligand-receptor interaction, and its upregulation may inhibit DC activation. Therefore, in HBV-associated hepatic malignancies, increased expression of TRIM28 may be a possible prognostic biomarker associated with immunological infiltration.²²⁰ However, the expression of CYP2C8 (cytochrome P450 family 2 subfamily C member 8) is downregulated in HCC and malignant hepatocytes.²²¹

Identifying biomarkers for HBV-HCC requires the integration of computational findings with experimental validation. Bioinformatics analyses enable the processing of large datasets and the identification of significant patterns

related to gene expression and disease progression.²²² For example, researchers have utilized bioinformatics tools to develop prognostic models based on gene signatures, which are subsequently validated through experimental techniques such as Western blotting and quantitative PCR to ensure their clinical applicability. This combined strategy not only enhances the reliability of biomarker candidates but also improves their utility in predicting patient outcomes and guiding treatment plans. Ultimately, the combination of computational insights and laboratory validation provides a stronger foundation for biomarker development in HCC, paving the way for improved diagnostic and therapeutic interventions.²²³

Data Standardization

In the realm of bioinformatics and ML techniques, data standardization is vital, particularly when analyzing genetic markers associated with HCC and HBV-induced hepatocyte remodeling.¹³³ Variability in data processing and experimental protocols can lead to discrepancies that complicate the replication and comparison of findings across different studies. By standardizing these procedures, significant genetic signatures can be identified, ensuring that data collected from various sources can be effectively integrated and analyzed.³⁴ This consistency not only fosters collaborative research efforts but also enhances the reliability of results, allowing scientists to confidently build upon one another's work. Ultimately, addressing variability through data standardization is essential for improving clinical outcomes and deepening our understanding of hepatitis B virus pathophysiology.²²⁴

Clinical Validation of Discovered Biomarkers for Patients with HBV-HCC

For predictive models to be both accurate and useful in actual clinical situations, clinical validation is essential in ML.²²⁵ To establish their efficacy and reliability in HCC surveillance among patients with chronic HBV infection, further validation through extensive clinical trials is necessary. This stage is crucial for enhancing early detection and ultimately improving patient outcomes.²²⁶

To ensure their dependability and effectiveness in a range of patient populations, however, the transfer from theoretical models to clinical practice requires thorough validation against external cohorts.²²⁷ For instance, a 2-stage ML model has demonstrated strong performance across multi-center datasets, providing recommendations for initial treatments and forecasting post-treatment survival for HCC patients. Additionally, models utilizing readily accessible clinical characteristics, such as tumor size and α -fetoprotein levels, have proven effective in accurately predicting the likelihood of microvascular invasion before surgery, which is crucial for tailoring individualized treatment plans for patients.²²⁸

Both internal and external datasets are typically utilized in the validation process to ensure the generalizability and robustness of the models. For example, an ensemble ML model was employed to predict early mortality among HCC patients with bone metastases, yielding promising results that could enhance clinical decision-making.²²⁹ Additionally, a variety of ML techniques have been applied to analyze genetic data and imaging features, improving the diagnostic accuracy in distinguishing between benign and malignant liver lesions.¹⁰³ These advancements underscore the potential of ML to enhance clinical management strategies for HBV-associated HCC, paving the way for precision medicine approaches based on validated genomic biomarkers.²³⁰

Strategies for Avoiding Overfitting in ML Models

One major issue is the risk of overfitting in ML models, particularly when searching for biomarkers associated with HBV-HCC. Overfitting occurs when a model learns the noise and fluctuations in the training data rather than the underlying patterns, resulting in poor generalization of unseen data.²³¹ Overly complex models may perform well during training but poorly during the validation or testing phases of research focused on gene expression data for potential biomarkers.¹⁸⁰ The implementation of strategies such as cross-validation and regularization is essential to mitigate overfitting, especially when dealing with small or non-diverse datasets. By carefully balancing model complexity and performance, we can increase the accuracy of biomarker discovery for HBV-HCC, ultimately improving clinical outcomes.²⁹

For reliable outcomes, strategies to prevent overfitting in ML models are essential during the genomic biomarker discovery process for HBV-HCC. Feature selection is an effective strategy that reduces noise and enhances model performance by lowering the dimensionality of the data and retaining only the most relevant features.²³² This goal can be achieved through techniques such as regularization and recursive feature elimination. Additionally, cross-validation is crucial for assessing the robustness of the model; by repeatedly splitting the data into training and validation sets, we can ensure that the model performs well when applied to new data. A diverse and sufficiently large dataset can help reduce overfitting and enhance the reliability of biomarker discovery efforts by providing a more representative sample of the underlying population.¹⁴⁶

Conclusions

Despite the growing amount of knowledge about how HBV-HCC develops, early identification and treatment of HBV-induced hepatocyte remodeling and HCC are still difficult. Emerging genomic signatures (indicators) unique to liver cancers that guarantee greater identification and longevity for patients still require evaluation. Our review summarized


the combined use of machine learning algorithms and bioinformatics analysis to extract novel genomic signatures from genomic, transcriptomic, epigenomic, proteomic, metabolomics, and multi-omics datasets. These genomic signatures serve as diagnostic, therapeutic, and prognostic targets for HBV-induced hepatocyte remodeling and HCC.


Acknowledgements

Not applicable.

ORCID iDs

Adane Adugna  <https://orcid.org/0000-0003-1467-2724>

Gashaw Azanaw Amare  <https://orcid.org/0009-0001-4803-1514>

Mohammed Jemal  <https://orcid.org/0009-0000-6700-9972>

Statements and Declarations

Ethical Considerations

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Author Contributions/CRedit

AA: Involved in the study conception, formal analysis, investigation, manuscript draft, design, manuscript writing-up, manuscript editing, and validation. GA: Involved in the study conception, design, and manuscript writing-up. MJ: Involved in manuscript writing-up, design, conception, language editing, manuscript editing, and validation.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The data presented in this manuscript are available from the corresponding author upon reasonable request.

References

1. Ringehan M, McKeating JA, Protzer U. Viral hepatitis and liver cancer. *Phil Trans R Soc A*. 2017;372(1732):20160274.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30.
3. Sukowati CH, El-Khobar KE, Ie SI, et al. Significance of hepatitis virus infection in the oncogenic initiation of hepatocellular carcinoma. *World J Gastroenterol*. 2016;22(4):1497-1512.
4. Higgs MR, Chouteau P, Lerat H. 'Liver let die': oxidative DNA damage and hepatotropic viruses. *J Gen Virol*. 2014;95(5):991-1004.
5. Prange R. Hepatitis B virus movement through the hepatocyte: an update. *Biol Cell*. 2022;114(12):325-348.
6. Yang L, Zou T, Chen Y, et al. Hepatitis B virus X protein mediated epigenetic alterations in the pathogenesis of hepatocellular carcinoma. *Hepatol Int*. 2022;16(4):741-754.
7. Zhao F, Xie X, Tan X, et al. The functions of hepatitis B virus encoding proteins: viral persistence and liver pathogenesis. *Front Immunol*. 2021;12:691766.
8. Villanueva RA, Loyola A. Pre- and post-transcriptional control of HBV gene expression: the road traveled towards the new paradigm of HBx, its isoforms, and their diverse functions. *Biomedicines*. 2023;11(6):1674.
9. Salpini R, D'Anna S, Benedetti L, et al. Hepatitis B virus DNA integration as a novel biomarker of hepatitis B virus-mediated pathogenetic properties and a barrier to the current strategies for hepatitis B virus cure. *Front Microbiol*. 2022;13:972687.
10. Tu T, Zhang H, Urban S. Hepatitis B virus DNA integration: in vitro models for investigating viral pathogenesis and persistence. *Viruses*. 2021;13(2).
11. Zhao K, Liu A, Xia Y. Insights into Hepatitis B virus DNA integration-55 years after virus discovery. *Innovation*. 2020;1(2):100034.
12. Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA integration: molecular mechanisms and clinical implications. *Viruses*. 2017;9(4).
13. Pollicino T, Caminiti G. HBV-integration studies in the clinic: role in the natural history of infection. *Viruses*. 2021;13(3).
14. Yang L, Ye S, Zhao X, et al. Molecular characterization of HBV DNA integration in patients with hepatitis and hepatocellular carcinoma. *J Cancer*. 2018;9(18):3225-3235.
15. Zoulim F, Chen PJ, Dandri M, Kennedy PT, Seeger C. Hepatitis B virus DNA integration: implications for diagnostics, therapy, and outcome. *J Hepatol*. 2024;81(6):1087-1099.
16. Shoraka S, Hosseinian SM, Hasibi A, Ghaemi A, Mohebbi SR. The role of hepatitis B virus genome variations in HBV-related HCC: effects on host signaling pathways. *Front Microbiol*. 2023;14:1213145.
17. Torresi J, Tran BM, Christiansen D, et al. HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer*. 2019;19(1):707.
18. Jiang Y, Han Q, Zhao H, Zhang J. The mechanisms of HBV-Induced hepatocellular carcinoma. *J Hepatocell Carcinoma*. 2021;8:435-450.
19. Yang G, Wan P, Zhang Y, et al. Innate immunity, inflammation, and intervention in HBV infection. *Viruses*. 2022;14(10):2275.
20. Bara AW, Braszewska A, Kwasniewska J. DNA methylation-an epigenetic mark in mutagen-treated Brachypodium distachyon cells. *Plants*. 2021;10(7):1408.
21. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol*. 2019;16(7):391-403.
22. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc*. 2020;92(4):807-812.
23. Mamoshina P, Volosnikova M, Ozerov IV, et al. Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Front Genet*. 2018;9:242.

24. Yakimovich A. Machine learning and Artificial Intelligence for the prediction of host–pathogen interactions: a viral case. *Infect Drug Resist.* 2021;14:3319–3326.
25. Harabor V, Mogos R, Nechita A, et al. Machine learning approaches for the prediction of Hepatitis B and C seropositivity. *Int J Environ Res Public Health.* 2023;20(3):2380.
26. Rui F, Yeo YH, Xu L, et al. Development of a machine learning-based model to predict hepatic inflammation in chronic hepatitis B patients with concurrent hepatic steatosis: a cohort study. *EClinicalMedicine.* 2024;68:102419.
27. Yuan M, Hu X, Yao L, et al. Comprehensive bioinformatics and machine learning analysis identify VCAN as a novel biomarker of hepatitis B virus-related liver fibrosis. *Front Mol Biosci.* 2022;9:1010160.
28. Unfried JP, Sangro P, Prats-Mari L, Sangro B, Fortes P. The landscape of lncRNAs in hepatocellular carcinoma: a translational perspective. *Cancers.* 2021;13(11):2651.
29. Ng S, Masarone S, Watson D, Barnes MR. The benefits and pitfalls of machine learning for biomarker discovery. *Cell Tissue Res.* 2023;394(1):17–31.
30. DeGroat W, Mendhe D, Bhusari A, et al. IntelliGenes: a novel machine learning pipeline for biomarker discovery and predictive analysis using multi-genomic profiles. *Bioinformatics.* 2023;39(12).
31. Brar A, Zhu A, Baciuc C, et al. Development of diagnostic and prognostic molecular biomarkers in hepatocellular carcinoma using machine learning: A systematic review. *Liver Cancer International.* 2022;3(4):141–161.
32. Liu X, Lu J, Zhang G, et al. A machine learning approach yields a multiparameter prognostic marker in liver cancer. *Cancer Immunol Res.* 2021;9(3):337–347.
33. Perakakis N, Yazdani A, Karniadakis GE, Mantzoros C. Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism.* 2018;87:A1–A9.
34. Kucukakcali Z, Akbulut S. Machine learning-based prediction of HBV-related hepatocellular carcinoma and detection of key candidate biomarkers. *Medeni Med J.* 2022;37(3):255–263.
35. Kim HY, Lampertico P, Nam JY, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B. *J Hepatol.* 2022;76(2):311–318.
36. Wong GL, Hui VW, Tan Q, et al. Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis. *JHEP Rep.* 2022;4(3):100441.
37. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* 2018;15(4):233–234.
38. TechTarget. Training performance evaluation, prediction performance evaluation and baseline modeling can refine machine learning models, 2021. Accessed February 19, 2025. <https://www.techtarget.com/searchenterpriseai/post/3-ways-to-evaluate-and-improve-machine-learning-models>
39. Labelle. Performance Metrics in Machine Learning: A Comprehensive Guide, 2024. Accessed February 19, 2025. <https://www.labellerr.com/blog/performance-metrics-in-machine-learning/>
40. Deepai. Accessed February 19, 2025. <https://deepai.org/machine-learning-glossary-and-terms/precision-and-recall>
41. Simplilearn. What is a ROC Curve and How to Use it in Performance Modeling? Accessed February 19, 2025. <https://www.simplilearn.com/what-is-a-roc-curve-and-how-to-use-it-in-performance-modeling>
42. Iguazio. Mastering ML Model Performance: Best Practices for Optimal Results, 2023. Accessed February 19, 2025. <https://www.iguazio.com/blog/mastering-ml-model-performance-best-practices-for-optimal-results/>
43. Neptune. Cross-Validation in Machine Learning: How to Do It Right, 2024. Accessed February 19, 2025. <https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right>
44. Machinelearningmastery. Train-Test Split for Evaluating Machine Learning Algorithms, 2020. Accessed February 19, 2025. <https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/>
45. V7. Top Performance Metrics in Machine Learning: A Comprehensive Guide, 2023. Accessed February 19, 2025. <https://www.v7labs.com/blog/performance-metrics-in-machine-learning>
46. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicinar.* 2020;56(9):455.
47. Hur MH, Yip TCF, Kim SU, et al. A machine learning model to predict liver-related outcomes after the functional cure of chronic hepatitis B. *J Hepatol.* 2025;82(2):235–244.
48. Zhang Z-M, Huang Y, Liu G, et al. Development of machine learning-based predictors for early diagnosis of hepatocellular carcinoma. *Sci Rep.* 2024;14(1):5274.
49. Leng S, Xu Z, Ma H. Reconstructing directional causal networks with random forest: causality meeting machine learning. *Chaos.* 2019;29(9):093130.
50. Tao K, Bian Z, Zhang Q, et al. Machine learning-based genome-wide interrogation of somatic copy number aberrations in circulating tumor DNA for early detection of hepatocellular carcinoma. *EBioMedicine.* 2020;56:102811.
51. Zeng H, Hui Y, Qin W, et al. High-throughput sequencing-based analysis of gene expression of hepatitis B virus infection-associated human hepatocellular carcinoma. *Oncol Lett.* 2020;20(4):18.
52. Zeng XC, Zhang L, Liao WJ, et al. Screening and identification of potential biomarkers in hepatitis B virus-related hepatocellular carcinoma by bioinformatics analysis. *Front Genet.* 2020;11:555537.
53. Hasan MAM, Maniruzzaman M, Shin J. Differentially expressed discriminative genes and significant meta-hub genes based key genes identification for hepatocellular carcinoma using statistical machine learning. *Sci Rep.* 2023;13(1):3771.
54. Qiang R, Zhao Z, Tang L, et al. Identification of 5 hub genes related to the early diagnosis, tumour stage, and poor outcomes of Hepatitis B virus-related hepatocellular carcinoma by bioinformatics analysis. *Comput Math Methods Med.* 2021;2021:1–20.
55. Foltz SM, Greene CS, Taroni JN. Cross-platform normalization enables machine learning model training on microarray and RNA-seq data simultaneously. *Commun Biol.* 2023;6(1):222.
56. Akond Z, Alam M, Mollah MNH. Biomarker identification from RNA-Seq data using a robust statistical approach. *Bioinformation.* 2018;14(4):153–163.
57. Agraz M, Goksuluk D, Zhang P, et al. ML-GAP: machine learning-enhanced genomic analysis pipeline using autoencoders and data augmentation. *Front Genet.* 2024;15:1442759.

58. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. *Med Oncol.* 2022;39(8):120.
59. Shen C, Jiang X, Li M, Luo Y. Hepatitis virus and hepatocellular carcinoma: recent advances. *Cancers.* 2023;15(2):533.
60. Angelis I, Exarchos T. Hepatocellular carcinoma detection using machine learning techniques. *Adv Exp Med Biol.* 2021;1338:21-29.
61. Zhang X, Jonassen I, Goksøyr A. Machine learning approaches for biomarker discovery using gene expression data. In: Helder IN, ed. *Bioinformatics.* Exon Publications; 2021.
62. Monaco A, Pantaleo E, Amoroso N, et al. A primer on machine learning techniques for genomic applications. *Comput Struct Biotechnol J.* 2021;19:4345-4359.
63. Schreiber J, Singh R. Machine learning for profile prediction in genomics. *Curr Opin Chem Biol.* 2021;65:35-41.
64. Tran KA, Kondrashova O, Bradley A, et al. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 2021;13(1):152.
65. Melis M, Diaz G, Kleiner DE, et al. Viral expression and molecular profiling in liver tissue versus microdissected hepatocytes in hepatitis B virus-associated hepatocellular carcinoma. *J Transl Med.* 2014;12(1):230.
66. Acharya S, Cui L, Pan Y. Multi-view feature selection for identifying gene markers: a diversified biological data driven approach. *BMC Bioinformatics.* 2020;21(18):483.
67. Feng G. Feature selection algorithm based on optimized genetic algorithm and the application in high-dimensional data processing. *PLoS One.* 2024;19(5):e0303088.
68. Montesinos-López OA, Montesinos-López A, Pérez-Rodríguez P, et al. A review of deep learning applications for genomic selection. *BMC Genomics.* 2021;22(1):19.
69. Xiong Y, Ling QH, Han F, Liu QH. An efficient gene selection method for microarray data based on LASSO and BPSO. *BMC Bioinformatics.* 2019;20(22):715.
70. Siemers FM, Bajorath J. Differences in learning characteristics between support vector machine and random forest models for compound classification revealed by Shapley value analysis. *Sci Rep.* 2023;13(1):5983.
71. Ajuwon BI, Richardson A, Roper K, et al. The development of a machine learning algorithm for early detection of viral hepatitis B infection in Nigerian patients. *Sci Rep.* 2023;13(1):3244.
72. Al-Bzour NN, Al-Bzour AN, Qasaymeh A, et al. Machine learning approach identifies inflammatory gene signature for predicting survival outcomes in hepatocellular carcinoma. *Sci Rep.* 2024;14(1):30328.
73. Pu X, Deng D, Chu C, Zhou T, Liu J. High-dimensional hepatopath data analysis by machine learning for predicting HBV-related fibrosis. *Sci Rep.* 2021;11(1):5081.
74. Fujita M, Chen MJM, Siwak DR, et al. Proteo-genomic characterization of virus-associated liver cancers reveals potential subtypes and therapeutic targets. *Nat Commun.* 2022;13(1):6481.
75. Coudray N, Tsirigos A. Deep learning links histology, molecular signatures and prognosis in cancer. *Nat Cancer.* 2020;1(8):755-757.
76. Abdulsahib FI, Al-Khateeb B, Kóczy LT, Nagy S. A transfer learning approach for the classification of liver cancer. *J Intell Syst.* 2023;32(1).
77. Mehmood M, Alshammari N, Alanazi SA, Ahmad F. Ahamd: systematic framework to predict early-stage liver carcinoma using hybrid of feature selection techniques and regression techniques. *Complexity.* 2022;2022(1):1-11.
78. Mostafa G, Mahmoud H, Abd El-Hafeez T, ElAraby ME. Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms. *J Big Data.* 2024;11(1):88.
79. Rizzo GEM, Cabibbo G, Craxi A. Hepatitis B virus-associated hepatocellular carcinoma. *Viruses.* 2022;14(5).
80. Tian J, Meng J, Yang Z, et al. Hepatitis B-related hepatocellular carcinoma: classification and prognostic model based on programmed cell death genes. *Front Immunol.* 2024;15:1411161.
81. Loureiro D, Tout I, Narguet S, et al. Mansouri A: miRNAs as potential biomarkers for viral hepatitis B and C. *Viruses.* 2020;12(12):1440.
82. Diener C, Keller A, Meese E. Emerging concepts of miRNA therapeutics: from cells to clinic. *Trends Genet.* 2022;38(6):613-626.
83. Menon A, Abd-Aziz N, Khalid K, Poh CL, Naidu R. Poh CL: miRNA: a promising therapeutic target in cancer. *Int J Mol Sci.* 2022;23(19):11502.
84. Shi Y, Liu Z, Lin Q, et al. MiRNAs and cancer: key link in diagnosis and therapy. *Genes.* 2021;12(8).
85. Yan R, Li K, Yuan D, et al. miR-183-5p promotes proliferation and migration in hepatocellular carcinoma by targeting IRS1 and its association with patient survival. *Int J Biol Markers.* 2020;35(3):83-89.
86. Liu T, Liu D, Liu R, et al. Discovering potential serological biomarker for chronic hepatitis B virus-related hepatocellular carcinoma in Chinese population by MAL-associated serum glycoproteomics analysis. *Sci Rep.* 2017;7:38918.
87. Sternby Eilard M, Åberg F. Combination of biomarkers for the detection of hepatocellular carcinoma. *Ann Transl Med.* 2020;8(20):1283.
88. Chen X, Huang L, Xie D, Zhao Q. EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. *Cell Death Dis.* 2018;9(1):3.
89. Yerukala Sathipati S, Ho SY. Novel miRNA signature for predicting the stage of hepatocellular carcinoma. *Sci Rep.* 2020;10(1):14452.
90. Letafati A, Najafi S, Mottahedi M, et al. MicroRNA let-7 and viral infections: focus on mechanisms of action. *Cell Mol Biol Lett.* 2022;27(1):14.
91. Zhao X, Dou J, Cao J, et al. Uncovering the potential differentially expressed miRNAs as diagnostic biomarkers for hepatocellular carcinoma based on machine learning in the cancer genome atlas database. *Oncol Rep.* 2020;43(6):1771-1784.
92. Jiang J, Liu Z, Ge C, et al. NK3 homeobox 1 (NKX3.1) up-regulates forkhead box O1 expression in hepatocellular carcinoma and thereby suppresses tumor proliferation and invasion. *J Biol Chem.* 2017;292(47):19146-19159.
93. Antao AM, Ramakrishna S, Kim KS. The role of Nkx3.1 in cancers and stemness. *Int J Stem Cells.* 2021;14(2):168-179.
94. Cui X, Zhao H, Wei S, et al. Hepatocellular carcinoma-derived FOXO1 inhibits tumor progression by suppressing IL-6 secretion from macrophages. *Neoplasia.* 2023;40:100900.
95. Xie Q, Chen L, Shan X, et al. Epigenetic silencing of SFRP1 and SFRP5 by hepatitis B virus X protein enhances hepatoma cell tumorigenicity through Wnt signaling pathway. *Int J Cancer.* 2014;135(3):635-646.

96. Schmidt-Arras D, Rose-John S. IL-6 pathway in the liver: from physiopathology to therapy. *J Hepatol.* 2016;64(6):1403-1415.
97. Tao S, Hou Y, Diao L, et al. Long noncoding RNA study: genome-wide approaches. *Genes Dis.* 2023;10(6):2491-2510.
98. GeeksforGeeks: Random Forest vs Support Vector Machine vs Neural Network, 2024. Accessed July 8, 2024. <https://www.geeksforgeeks.org/random-forest-vs-support-vector-machine-vs-neural-network/>
99. Azlan A, Obeidat SM, Yunus MA, Azzam G. Systematic identification and characterization of Aedes aegypti long noncoding RNAs (lncRNAs). *Sci Rep.* 2019;9(1):12147.
100. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA.* 2019;5(1):17.
101. Bonidia RP, Machida JS, Negri TC, et al. A novel decomposing model with evolutionary algorithms for feature selection in long non-coding RNAs. *IEEE Access.* 2020;8:181683-181697.
102. Ovens K, Eames BF, McQuillan I. Comparative analyses of gene co-expression networks: implementations and applications in the study of evolution. *Front Genet.* 2021;12:695399.
103. Feng S, Wang J, Wang L, et al. Current status and analysis of machine learning in hepatocellular carcinoma. *J Clin Transl Hepatol.* 2023;11(5):1184-1191.
104. Nong S, Chen X, Wang Z, et al. Potential lncRNA biomarkers for HBV-related hepatocellular carcinoma diagnosis revealed by analysis on coexpression network. *Biomed Res Int.* 2021;2021(1):9972011.
105. Ferrer J, Dimitrova N. Transcription regulation by long non-coding RNAs: mechanisms and disease relevance. *Nat Rev Mol Cell Biol.* 2024;25(5):396-415.
106. Jan S, Dar MI, Wani R, et al. Targeting ehmt2/g9a for cancer therapy: progress and perspective. *Eur J Pharmacol.* 2021;893:173827.
107. Liu Y, Wu Y, Liu S, Dai Y. Long non-coding RNA TRIM52-AS1 promotes growth and metastasis via miR-218-5p/ROBO1 in hepatocellular carcinoma. *Cancer Manag Res.* 2021;13:547-558.
108. Sun T-H, Wang C-C, Wu YL, Hsu K-C, Lee TH. Machine learning approaches for biomarker discovery to predict large-artery atherosclerosis. *Sci Rep.* 2023;13(1):15139.
109. Diaz-Uriarte R, Gómez de Lope E, Giugno R, et al. Ten quick tips for biomarker discovery and validation analyses using machine learning. *PLoS Comput Biol.* 2022;18(8):e1010357.
110. Huang ZH, Du YP, Wen JT, Lu BF, Zhao Y. snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. *Cell Death Discov.* 2022;8(1):259.
111. Wajahat M, Bracken CP, Orang A. Emerging functions for snoRNAs and snoRNA-derived fragments. *Int J Mol Sci.* 2021;22(19):10193.
112. Zhang X, Wang C, Xia S, et al. The emerging role of snoRNAs in human disease. *Genes Dis.* 2023;10(5):2064-2081.
113. Zhao X, Chen G, Wu Y, et al. TEP SNORD12B, SNORA63, and SNORD14E as novel biomarkers for hepatitis B virus-related hepatocellular carcinoma (HBV-related HCC). *Cancer Cell Int.* 2024;24(1):3.
114. Sayed GI, Solymán M, El Gedawy G, et al. Circulating miRNA's biomarkers for early detection of hepatocellular carcinoma in Egyptian patients based on machine learning algorithms. *Sci Rep.* 2024;14(1):4989.
115. Son JA, Ahn HR, You D, et al. Novel gene signatures as prognostic biomarkers for predicting the recurrence of hepatocellular carcinoma. *Cancers.* 2022;14(4):865.
116. De Stefano F, Chacon E, Turcios L, Marti F, Gedaly R. Novel biomarkers in hepatocellular carcinoma. *Dig Liver Dis.* 2018;50(11):1115-1123.
117. Yu J, Park R, Kim R. Promising novel biomarkers for hepatocellular carcinoma: diagnostic and prognostic insights. *J Hepatocell Carcinoma.* 2023;10:1105-1127.
118. Watson SA, McStay GP. Functions of cytochrome c oxidase assembly factors. *Int J Mol Sci.* 2020;21(19):7254.
119. Kur J, Olszewski M, Długołęcka A, Filipkowski P. Single-stranded DNA-binding proteins (SSBs) — sources and applications in molecular biology. *Acta Biochim Pol.* 2005;52(3):569-574.
120. Moreira F, Arenas M, Videira A, Pereira F. Molecular evolution of DNA Topoisomerase III beta (TOP3B) in Metazoa. *J Mol Evol.* 2021;89(6):384-395.
121. Pommier Y, Nussenzweig A, Takeda S, Austin C. Human topoisomerases and their roles in genome stability and organization. *Nat Rev Mol Cell Biol.* 2022;23(6):407-427.
122. Hu X, Bao M, Huang J, Zhou L, Zheng S. Identification and validation of novel biomarkers for diagnosis and prognosis of hepatocellular carcinoma. *Front Oncol.* 2020;10:541479.
123. Nathan KG, Lal SK. The multifarious role of 14-3-3 family of proteins in viral replication. *Viruses.* 2020;12(4):436.
124. Tian LY, Smit DJ, Jücker M. The role of PI3K/AKT/mTOR signaling in hepatocellular carcinoma metabolism. *Int J Mol Sci.* 2023;24(3):2652.
125. Chu Q, Gu X, Zheng Q, Wang J, Zhu H. Phosphoribosyl pyrophosphate amido transferase: a new prognostic biomarker for hepatocellular carcinoma. *Int J Gen Med.* 2022;15:353-358.
126. Tran DH, Kim D, Kesavan R, et al. De novo and salvage purine synthesis pathways across tissues and tumors. *Cell.* 2024;187(14):3602-3618.e20.
127. Jin X, Tanaka H, Jin M, et al. PQBP5/NOL10 maintains and anchors the nucleolus under physiological and osmotic stress conditions. *Nat Commun.* 2023;14(1):9.
128. Jiao L, Liu Y, Yu XY, et al. Ribosome biogenesis in disease: new players and therapeutic targets. *Signal Transduct Target Ther.* 2023;8(1):15.
129. Liu X, Xiao C, Yue K, et al. Identification of multi-omics biomarkers and construction of the novel prognostic model for hepatocellular carcinoma. *Sci Rep.* 2022;12(1):12084.
130. Wang F, Xu X, Zhang N, Chen Z. Identification and integrated analysis of hepatocellular carcinoma-related circular RNA signature. *Ann Transl Med.* 2020;8(6):294.
131. Chen H, Ao Q, Wang Y, et al. SOX11 as a potential prognostic biomarker in hepatocellular carcinoma linked to immune infiltration and ferroptosis. *Chin J Cancer Res.* 2024;36(4):378-397.
132. Yan L, Xu F, Dai C. Overexpression of COL24A1 in hepatocellular carcinoma predicts poor prognosis: a study based on multiple databases, clinical samples and cell lines. *Oncotargets Ther.* 2020;13:2819-2832.
133. Liu Y, Veeraraghavan V, Pinkerton M, et al. Viral biomarkers for hepatitis B virus-related hepatocellular carcinoma occurrence and recurrence. *Front Microbiol.* 2021;12:665201.

134. Hayashi S, Nagaoka K, Tanaka Y. Blood-based biomarkers in Hepatitis B virus-related hepatocellular carcinoma, including the viral genome and glycosylated proteins. *Int J Mol Sci.* 2021;22(20).
135. Walakira A, Skubic C, Nadižar N, et al. Integrative computational modeling to unravel novel potential biomarkers in hepatocellular carcinoma. *Comput Biol Med.* 2023;159:106957.
136. Jiang N, Zhang X, Qin D, et al. Identification of core genes related to progression and prognosis of hepatocellular carcinoma and small-molecule drug predication. *Front Genet.* 2021;12:608017.
137. Li J, Zhang D, Liu H, et al. Identification of potential prognostic biomarkers among gene models for coiled-coil domain-containing family members in hepatocellular carcinoma elucidates their influence on the hypoxia pathway and immune microenvironment. *J Gastrointest Oncol.* 2023;14(6):2559-2573.
138. Xu R, Lin L, Zhang B, et al. Identification of prognostic markers for hepatocellular carcinoma based on the epithelial-mesenchymal transition-related gene BIRC5. *BMC Cancer.* 2021;21(1):687.
139. Wu S, Liu M, Xiao S, et al. Identification and verification of novel ferroptosis biomarkers predicts the prognosis of hepatocellular carcinoma. *Genomics.* 2023;115(6):110733.
140. Duan X, Li W, Hu P, et al. MicroRNA-183-5p contributes to malignant progression through targeting PDCD4 in human hepatocellular carcinoma. *Biosci Rep.* 2020;40(10).
141. Yao S, Cao B, Li T, et al. Prediction-oriented prognostic biomarker discovery with survival machine learning methods. *NAR Genom Bioinform.* 2023;5(2):lqad055.
142. Rojas Á, Gil-Gómez A, de la Cruz-Ojeda P, et al. Long non-coding RNA H19 as a biomarker for hepatocellular carcinoma. *Liver Int.* 2022;42(6):1410-1422.
143. Keshtkar S, Soleimanian S, Kaviani M, et al. Immune cell-derived extracellular vesicles in the face of pathogenic infections. *Front Immunol.* 2022;13:906078.
144. Sun J, Zhang Z, Cai J, Li X, Xu X. Identification of hub genes in liver hepatocellular carcinoma based on weighted gene co-expression network analysis. *Biochem Genet.* 2024.
145. models ML. Accessed January 14, 2025. <https://machine-learningmodels.org/solving-overfitting-in-deep-learning-models>
146. Bejani MM, Ghatee M. A systematic review on overfitting control in shallow and deep neural networks. *Artif Intell Rev.* 2021;54(8):6391-6438.
147. Gupta R, Kleinjans J, Caiment F. Identifying novel transcript biomarkers for hepatocellular carcinoma (HCC) using RNA-Seq datasets and machine learning. *BMC Cancer.* 2021;21(1):962.
148. Zhang J, Liu G, Liu Y, et al. The biological functions and related signaling pathways of SPON2. *Front Oncol.* 2023;13:1323744.
149. Toyoda H, Kumada T, Tada T, et al. Tumor markers for hepatocellular carcinoma: simple and significant predictors of outcome in patients with HCC. *Liver Cancer.* 2015;4(2):126-136.
150. Demšar J, Zupan B. Hands-on training about overfitting. *PLoS Comput Biol.* 2021;17(3):e1008671.
151. Ghosh S, Zhao X, Alim M, Brudno M, Bhat M. Artificial intelligence applied to 'omics data in liver disease': towards a personalised approach for diagnosis, prognosis and treatment. *Gut.* 2024;74(2):295-311.
152. Attia MF, Anton N, Wallyn J, Omran Z, Vandamme TF. An overview of active and passive targeting strategies to improve the nanocarriers efficiency to tumour sites. *J Pharm Pharmacol.* 2019;71(8):1185-1198.
153. Luo P, Yin P, Hua R, et al. A large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology.* 2018;67(2):662-675.
154. Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, et al. The impact of sex on gene expression across human tissues. *Science.* 2020;369(6509):eaba3066.
155. Xu C, Cheng S, Chen K, et al. Sex differences in genomic features of hepatitis B-associated hepatocellular carcinoma with distinct antitumor immunity. *Cell Mol Gastroenterol Hepatol.* 2023;15(2):327-354.
156. Ajayi F, Jan J, Singal AG, Rich NE. Racial and sex disparities in hepatocellular carcinoma in the USA. *Curr Hepatol Rep.* 2020;19(4):462-469.
157. Li Y, Xu A, Jia S, Huang J. Recent advances in the molecular mechanism of sex disparity in hepatocellular carcinoma. *Oncol Lett.* 2019;17(5):4222-4228.
158. Yang D, Hanna DL, Usher J, et al. Impact of sex on the survival of patients with hepatocellular carcinoma: a surveillance, epidemiology, and end results analysis. *Cancer.* 2014;120(23):3707-3716.
159. Brown R, Goulder P, Matthews PC. Sexual dimorphism in chronic hepatitis B virus (HBV) infection: evidence to inform elimination efforts. *Wellcome Open Res.* 2022;7:32.
160. Ye W, Siwko S, Tsai RYL. Sex and race-related DNA methylation changes in hepatocellular carcinoma. *Int J Mol Sci.* 2021;22(8):3820.
161. Natri HM, Wilson MA, Buetow KH. Distinct molecular etiologies of male and female hepatocellular carcinoma. *BMC Cancer.* 2019;19(1):951.
162. Shamsan E, Almezgagi M, Gamah M, et al. The role of PI3k/AKT signaling pathway in attenuating liver fibrosis: a comprehensive review. *Front Med.* 2024;11:1389329.
163. Huang CY, Tan KT, Huang SF, et al. Study of sex-biased differences in genomic profiles in East Asian hepatocellular carcinoma. *Oncol Discov.* 2024;15(1):276.
164. Kotsiliti E. Understanding HCC sex disparities. *Nat Rev Gastroenterol Hepatol.* 2022;19(3):147.
165. Ma J, Malladi S, Beck AH. Systematic analysis of sex-linked molecular alterations and therapies in cancer. *Sci Rep.* 2016;6:19119.
166. Yao C, Joehanes R, Johnson AD, et al. Sex- and age-interacting eQTLs in human complex diseases. *Hum Mol Genet.* 2014;23(7):1947-1956.
167. Melia T, Waxman DJ. Genetic factors contributing to extensive variability of sex-specific hepatic gene expression in diversity outbred mice. *PLoS One.* 2020;15(12):e0242665.
168. Choi J, Kim G, Han S, et al. Longitudinal assessment of three serum biomarkers to detect very early-stage hepatocellular carcinoma. *Hepatology.* 2019;69(5):1983-1994.
169. Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. *Comput Struct Biotechnol J.* 2020;18:2300-2311.
170. Borges KA, Dai J, Parikh ND, et al. Rationale and design of the hepatocellular carcinoma early detection strategy study: a multi-center longitudinal initiative of the National Cancer Institute's Early Detection Research Network. *Contemp Clin Trials.* 2019;76:49-54.

171. Lux S, Bullinger L. Circular RNAs in cancer. *Adv Exp Med Biol.* 2018;1087:215-230.
172. Li HC, Yang CH, Lo SY. Long noncoding RNAs in hepatitis B virus replication and oncogenesis. *World J Gastroenterol.* 2022;28(25):2823-2842.
173. Lobo-Jarne T, Nývltová E, Pérez-Pérez R, et al. Human COX7A2L regulates complex III biogenesis and promotes supercomplex organization remodeling without affecting mitochondrial bioenergetics. *Cell Rep.* 2018;25(7):1786-1799.e1784.
174. Li X, Lei Y, Wu M, Li N. Regulation of macrophage activation and polarization by HCC-derived exosomal lncRNA TUC339. *Int J Mol Sci.* 2018;19(10).
175. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* 2016;17(1):47-62.
176. Liang W, Zhao Y, Meng Q, et al. The role of long non-coding RNA in hepatocellular carcinoma. *Aging.* 2024;16(4):4052-4073.
177. Huang Z, Zhou JK, Peng Y, He W, Huang C. The role of long noncoding RNAs in hepatocellular carcinoma. *Mol Cancer.* 2020;19(1):77.
178. Ren X, Feng N. Unveiling novel prognostic biomarkers and therapeutic targets for HBV-associated hepatocellular carcinoma through integrated bioinformatic analysis. *Medicine.* 2024;103(43):e40134.
179. Yu H, Samuels DC, Zhao YY, Guo Y. Architectures and accuracy of artificial neural network for disease classification from omics data. *BMC Genomics.* 2019;20(1):167.
180. Takahashi Y, Ueki M, Tamiya G, et al. Machine learning for effectively avoiding overfitting is a crucial strategy for the genetic prediction of polygenic psychiatric phenotypes. *Transl Psychiatry.* 2020;10(1):294.
181. Verma S, Sahu BD, Mugale MN. Role of lncRNAs in hepatocellular carcinoma. *Life Sci.* 2023;325:121751.
182. Moyo B, Nicholson SA, Arbuthnot PB. The role of long non-coding RNAs in hepatitis B virus-related hepatocellular carcinoma. *Virus Res.* 2016;212:103-113.
183. Clark AJ, Lillard JW, Jr. A comprehensive review of bioinformatics tools for genomic biomarker discovery driving precision oncology. *Genes.* 2024;15(8).
184. Xie L, Wang L, Zhu W, Zhao J, Guo X. Editorial: bioinformatics tools (and web server) for cancer biomarker development, volume II. *Front Genet.* 2022;13:959159.
185. Zeng L, Fan X, Wang X, et al. Bioinformatics analysis based on multiple databases identifies hub genes associated with hepatocellular carcinoma. *Curr Genomics.* 2019;20(5):349-361.
186. Medium. Accessed January 14, 2025. <https://neer975.medium.com/strategies-to-prevent-overfitting-in-machine-learning-models>
187. Morlando M, Fatica A. Alteration of epigenetic regulation by long noncoding RNAs in cancer. *Int J Mol Sci.* 2018;19(2):570.
188. Zheng X, Chen L, Li X, et al. Prediction of miRNA targets by learning from interaction sequences. *PLoS One.* 2020;15(5):e0232578.
189. Thankaswamy-Kosalai S, Sen P, Nookaew I. Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics. *Genomics.* 2017;109(3-4):186-191.
190. Chen L, Heikkinen L, Wang C, et al. Trends in the development of miRNA bioinformatics tools. *Brief Bioinform.* 2019;20(5):1836-1852.
191. Sivasudhan E, Blake N, Lu Z, Meng J, Rong R. Hepatitis B viral protein HBx and the molecular mechanisms modulating the hallmarks of hepatocellular carcinoma: a comprehensive review. *Cells.* 2022;11(4).
192. Yi L, Luo P, Zhang J. Identification of aberrantly methylated differentially expressed genes in breast cancer by integrated bioinformatics analysis. *J Cell Biochem.* 2019;120(9):16229-16243.
193. Li X, Ma Z, Mei L. Comprehensive analysis of UBE2C expression and its potential roles and mechanisms in hepatocellular carcinoma. *Aging.* 2023;15(15):7397-7407.
194. Chu Q, Li J, Chen J, Yuan Z. HBV induced the discharge of intrinsic antiviral miRNAs in HBV-replicating hepatocytes via extracellular vesicles to facilitate its replication. *J Gen Virol.* 2022;103(5).
195. Lee Y, Kim JH. The emerging roles of extracellular vesicles as intercellular messengers in liver physiology and pathology. *Clin Mol Hepatol.* 2022;28(4):706-724.
196. Wang X, Shen H, Zhangyuan G, et al. 14-3-3 ζ delivered by hepatocellular carcinoma-derived exosomes impaired anti-tumor function of tumor-infiltrating T lymphocytes. *Cell Death Dis.* 2018;9(2):159.
197. Cao LQ, Yang XW, Chen YB, et al. Exosomal miR-21 regulates the TETs/PTENp1/PTEN pathway to promote hepatocellular carcinoma growth. *Mol Cancer.* 2019;18(1):148.
198. Chen L, Guo P, He Y, et al. HCC-derived exosomes elicit HCC progression and recurrence by epithelial-mesenchymal transition through MAPK/ERK signalling pathway. *Cell Death Dis.* 2018;9(5):513.
199. Zhang JN, Wei F, Zheng BB, Tang L, Chen FY. Role of miRNA-99a-5p in modulating the function of hepatocellular carcinoma cells: bioinformatics analysis and in vitro assay. *Curr Cancer Drug Targets.* 2023;23(6):461-470.
200. Farzaneh Z, Farzaneh M. Prevention and treatment of hepatocellular carcinoma using miRNAs. *Arch Iran Med.* 2022;25(2):133-138.
201. Zhang W, Liu Y, Jiang J, et al. Extracellular vesicle long non-coding RNA-mediated crosstalk in the tumor microenvironment: tiny molecules, huge roles. *Cancer Sci.* 2020;111(8):2726-2735.
202. Gondaliya P, Sayyed AA, Driscoll J, Patel K, Patel T. Extracellular vesicle RNA signaling in the liver tumor microenvironment. *Cancer Lett.* 2023;558:216089.
203. Yang WX, Pan YY, You CG. CDK1, CCNB1, CDC20, BUB1, MAD2L1, MCM3, BUB1B, MCM2, and RFC4 may be potential therapeutic targets for hepatocellular carcinoma using integrated bioinformatic analysis. *Biomed Res Int.* 2019;2019:1-16.
204. Jin Z, Peng F, Zhang C, et al. Expression, regulating mechanism and therapeutic target of KIF20A in multiple cancer. *Heliyon.* 2023;9(2):e13195.
205. Zou Y, Ruan S, Jin L, et al. CDK1, CCNB1, and CCNB2 are prognostic biomarkers and correlated with immune infiltration in hepatocellular carcinoma. *Med Sci Monit.* 2020;26:e925289.
206. Jin X, Yu W, Wang A, Qiu Y. Serum ribonucleotide reductase subunit M2 in patients with chronic liver diseases and hepatocellular carcinoma. *Lab Med.* 2023;54(6):626-632.
207. Zhou X, Park B, Choi D, Han K. A generalized approach to predicting protein-protein interactions between virus and host. *BMC Genomics.* 2018;19(6):568.
208. Ng CKY, Di Costanzo GG, Terracciano LM, Piscuoglio S. Circulating cell-free DNA in hepatocellular carcinoma: current insights and Outlook. *Front Med.* 2018;5:78.

209. Zhao Y, Xue F, Sun J, et al. Genome-wide methylation profiling of the different stages of hepatitis B virus-related hepatocellular carcinoma development in plasma cell-free DNA reveals potential biomarkers for early detection and high-risk monitoring of hepatocellular carcinoma. *Clin Epigenetics*. 2014;6(1):30.
210. Tran NH, Kisiel J, Roberts LR. Using cell-free DNA for HCC surveillance and prognosis. *JHEP Rep*. 2021;3(4):100304.
211. Zhuang L, Yang Z, Meng Z. Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients. *Biomed Res Int*. 2018;2018:1-8.
212. Chen Q, Yang S, Zhang Y, et al. Identification of MAD2L1 as a potential biomarker in hepatocellular carcinoma via comprehensive bioinformatics analysis. *Biomed Res Int*. 2022;2022(1):9868022.
213. Fu J, Zhang X, Yan L, et al. Identification of the hub gene BUB1B in hepatocellular carcinoma via bioinformatic analysis and in vitro experiments. *PeerJ*. 2021;9:e10943.
214. Li S, Hao L, Hu X, Li L. A systematic study on the treatment of hepatitis B-related hepatocellular carcinoma with drugs based on bioinformatics and key target reverse network pharmacology and experimental verification. *Infect Agent Cancer*. 2023;18(1):41.
215. Panvichian R, Tantiwetrueangdet A, Angkathunyakul N, Leelaudomlapi S. TOP2A amplification and overexpression in hepatocellular carcinoma tissues. *Biomed Res Int*. 2015;2015:1-8.
216. Li G, Wang Z, Chen D, et al. Comprehensive analysis of a TPX2-related TRHDE-AS1/PKIA ceRNA network involving prognostic signatures in Hepatitis B virus-infected hepatocellular carcinoma. *Front Cell Infect Microbiol*. 2022;12:1025900.
217. Liu R, Huang Y. CDC7 as a novel biomarker and druggable target in cancer. *Clin Transl Oncol*. 2022;24(10):1856-1864.
218. Li HT, Wei B, Li ZQ, et al. Diagnostic and prognostic value of MCM3 and its interacting proteins in hepatocellular carcinoma. *Oncol Lett*. 2020;20(6):308-311.
219. Yang Y, Tan S, Han Y, et al. The role of tripartite motif-containing 28 in cancer progression and its therapeutic potentials. *Front Oncol*. 2023;13:1100134.
220. Coto-Llerena M, Benjak A, Gallon J, et al. Circulating cell-free DNA captures the intratumor heterogeneity in multinodular hepatocellular carcinoma. *JCO Precis Oncol*. 2022;6(6):e2100335.
221. Li K, Chen Y. CYP2C8 regulated by GAS5/miR-382-3p exerts anti-cancerous properties in liver cancer. *Cancer Biol Ther*. 2020;21(12):1145-1153.
222. Zhu XR, Zhu JQ, Chen YF, et al. Bioinformatics analysis and experimental verification of the prognostic and biological significance mediated by fatty acid metabolism related genes for hepatocellular carcinoma. *Front Oncol*. 2022;12:972744.
223. Han J, Wang Y, Zhou H, Ai S, Wan D. Integrated bioinformatics and experimental analysis identified TRIM28 a potential prognostic biomarker and correlated with immune infiltrates in liver hepatocellular carcinoma. *Comput Math Methods Med*. 2022;2022:6267851-6267917.
224. Zhang J, Liu X, Zhou W, et al. Identification of key genes associated with the process of hepatitis B inflammation and cancer transformation by integrated bioinformatics analysis. *Front Genet*. 2021;12:654517.
225. Shen Y, Huang J, Jia L, Zhang C, Xu J. Bioinformatics and machine learning driven key genes screening for hepatocellular carcinoma. *Biochem Biophys Rep*. 2024;37:101587.
226. Nio K, Shimakami T, Terashima T, et al. Serum laminin γ 2 monomer as a predictive biomarker for hepatocellular carcinoma in patients with chronic hepatitis B virus infection: a retrospective cohort study. *Sci Rep*. 2024;14(1):25395.
227. Xu F, Li Z, Guan H, Ma J. Multiple machine learning algorithms, validation of external clinical cohort and assessments of model gain effects will better serve cancer research on bioinformatic models. *Cancer Cell Int*. 2024;24(1):427.
228. Long Z, Yi M, Qin Y, et al. Development and validation of an ensemble machine-learning model for predicting early mortality among patients with bone metastases of hepatocellular carcinoma. *Front Oncol*. 2023;13:1144039.
229. Chen Z, Zuo X, Zhang Y, et al. Derivation and validation of machine learning models for preoperative estimation of microvascular invasion risk in hepatocellular carcinoma. *Ann Transl Med*. 2023;11(6):249.
230. Lee KH, Choi GH, Yun J, et al. Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: a multi-center study. *Digit Med*. 2024;7(1):2.
231. Huang S, Cai N, Pacheco PP, et al. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics*. 2018;15(1):41-51.
232. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res*. 2018;24(6):1248-1259.