

Review

Intervention of machine learning in bladder cancer research using multi-omics datasets: systematic review on biomarker identification

Blessy Kiruba¹  · P. S. Athul Narayan¹ · Badhari Raj¹ · S. Rohieth Raj¹ · Sam George Mathew¹ · Sudhakaran Sajitha Lulu²  · Vino Sundararajan¹ 

Received: 10 January 2025 / Accepted: 16 May 2025

Published online: 05 June 2025

© The Author(s) 2025 

Abstract

Bladder cancer (BC) is one of the most prevalent types of cancer in developed countries. BC is characterized by its highly heterogeneous and dynamic nature, with significantly higher morbidity and mortality rates in men compared to women. Diagnosing BC requires traditional methods, such as cystoscopy, which can be invasive and costly. Recent research has heavily focused on multi-omics analysis, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics, for biomarker identification. However, challenges such as computational complexity and data integration prevent these methods from achieving robust diagnostic capabilities. Hence, machine learning (ML), with its ability to process high-dimensional data and identify complex patterns, offers a promising patient outcome. By exploiting genomics, epigenomics, transcriptomics, proteomics, and metabolomics data, these models facilitate the discovery of reliable biomarkers, which are critical for early detection, prognosis, and risk stratification of the disease. Integrated models combining computational techniques with large multi-omics datasets have gained significant attention, enabling the identification of significant BC biomarkers that include genes coding for diverse cellular functions, differentially expressed genes, proteins, and metabolites. A substantial amount of multi-omics data collected from clinics and laboratories are utilized to train powerful ML models such as Support Vector Machines (SVM), random forests (RF), decision trees (DT), and gradient boosting methods (e.g., XGBoost) to perform complex tasks, including biomarker discovery, classification of subtypes and feature selection. This comprehensive review highlights the essence of integrated multiomics-ML approaches for the improvement of prognosis and diagnosis of BC.

1 Introduction

Bladder cancer (BC) is the tenth most prevalent malignancy worldwide with a significant rate of morbidity and mortality [1]. Moreover, it is the fourth most prevalent cancer among men and the ninth most prevalent among women in Western countries [2]. The American Cancer Society predicts that in 2025 there will be approximately 84,870 new BC cases and 17,420 related deaths in the US [3]. Urothelial carcinoma is the most common subtype, further classified into muscle-invasive (MIBC) and non-muscle-invasive (NMIBC) forms based on the tumor penetration [4].

✉ Vino Sundararajan, svino@vit.ac.in; Blessy Kiruba, blessy.kiruba2024@vitstudent.ac.in; P. S. Athul Narayan, athulnarayan.ps2021@vitstudent.ac.in; Badhari Raj, badhari.raj2021@vitstudent.ac.in; S. Rohieth Raj, rohithraj.s2021@vitstudent.ac.in; Sam George Mathew, samgeorge.mathew2021@vitstudent.ac.in; Sudhakaran Sajitha Lulu, ssajithalulu@vit.ac.in | ¹Integrated Multiomics Laboratory, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India. ²Integrated Multiomics Laboratory, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India.



Despite advances in treatment approaches, including surgery, cisplatin-based chemotherapy, targeted therapy, and immunotherapy, the recurrence rate and early and accurate diagnosis remain a significant challenge [5]. The gold standard diagnostic tool, cystoscopy, is invasive, costly, and lacks sensitivity in detecting low-grade tumors, likewise, urinary cytology is limited by its low sensitivity for low-grade tumors and is prone to high rates of false positives. This highlights a critical need for non-invasive and cost-effective biomarkers with high sensitivity and specificity to improve early detection and disease monitoring [6]. Biomarkers are crucial for diagnosis, prognosis, cancer subtype differentiation, and recurrence risk assessment [7, 8]. In recent years, multi-omics exploration encompassing genomics, epigenomics, transcriptomics, proteomics, and metabolomics has gained significant attention in biomarker discovery. However, the complexity and heterogeneity of these large datasets often limit traditional statistical methods. Machine learning (ML) offers a promising alternative by applying mathematical models to learn from datasets, effectively identifying diagnostic, prognostic, and predictive biomarkers. Moreover, its application in healthcare offers benefits such as improved accuracy, early detection, efficiency, personalized medicine, and cost-effectiveness [9].

This review provides a comprehensive overview of the application of ML in BC biomarker discovery across multi-omics datasets, highlighting potential diagnostic, prognostic, and therapeutic advancements for improved patient outcomes.

2 Machine learning in multi-omics studies

2.1 Benefits of machine learning in multi-omics

The Tumor Microenvironment (TME) is extremely complex and heterogeneous, comprising various cell types, including tumor cells, immune cells, blood vessels, extracellular matrix, and fibroblasts. This complex network can make it difficult to identify biomarkers [10]. Therefore, ML could be used to identify potential biomarkers from the immense amount of multi-omics data associated with TME. Cost-effectiveness is another aspect, where a study depicted an average cystoscopy to be around \$430 on average for patients in the US [11]. Additionally, bladder cancer also requires lifetime cystoscopic surveillance, making it the most expensive cancer to treat per patient [12]. However, ML models can lower unnecessary cystoscopies by 30%–40% in NMIBC patients [13]. ML is particularly crucial in biomarker discovery due to the complexities of large, high-dimensional datasets. Traditional statistical methods like t-tests and ANOVA are limited because they assume specific data distributions (e.g., normality), often violated in genomic data. Moreover, deep sampling can create non-linear relationships, outliers, and distributions with kurtosis, complicating analysis. As dataset complexity grows, conventional methods become computationally infeasible [14].

There are various ML algorithms used in multi-omics data analysis. For instance, algorithms like Boruta, SVM-RFE, and pamr, are used to detect the most relevant key genes, proteins, or metabolites that can be utilized as biomarkers for prognosis or diagnosis and to differentiate between cancer subtypes [15]. By focusing on key features, these algorithms reduce computational burden while enhancing efficiency. Moreover, Classification algorithms, such as Random Forest, XGBoost, SVM, KNN, and decision trees, are used to classify patients into risk groups based on their molecular profiles, enabling personalized treatment decisions. Additionally, Regression methods, such as LASSO, Ridge, and Enet, are employed to identify biomarkers linearly associated with disease severity or treatment response, reducing dimensionality in omics datasets while preserving predictive power. Lastly, survival analysis algorithms like coxBoost, Stepwise COX, and plsRcox model the relationship between dataset parameters and survival time, allowing for the identification of biomarkers that predict patient prognosis. These survival analysis techniques are particularly valuable for uncovering biomarkers with complex, non-linear associations with outcomes, which might be overlooked by other methods [16].

Furthermore, the algorithm selection largely relies on the type of data being analyzed. For instance, LASSO and SVM are particularly effective for high-dimensional datasets, where feature selection and classification are critical. In contrast, Random Forest is better suited for handling relatively less complex datasets. Therefore, selecting the appropriate ML algorithm, applying effective feature selection methods, optimizing model parameters, and evaluating model performance are critical for ensuring robust predictions and meaningful biological interpretations. Following the selection of an algorithm, various systematic approaches are employed to optimize its hyperparameters, which are manually assigned before training and govern the model's learning process. This is particularly important to prevent overfitting and ensure consistent performance. The choice of hyperparameter optimization techniques depends on the user's preference and computational constraints. A more efficient approach involves Bayesian optimization, a probability-based technique, or

Grid Search, which systematically assesses all possible combinations of hyperparameters to determine the most effective setting. During training, the algorithm learns its parameters, such as coefficients and intercepts in the case of a linear regression model, based on the provided data to improve the model's predictive performance [17].

A schematic workflow representing the identification of biomarkers by ML algorithms using omics datasets is given in Fig. 1.

2.2 Common machine learning algorithms

LASSO is a linear regression model derived from traditional ordinary linear square (OLS) regression. Its objective is to minimize the cost of the regression model while improving feature selection. LASSO achieves this by including a penalty term to OLS model. As a result, the model automatically shrinks the coefficients of less important features to zero, effectively eliminating trivial features [19]. For a given set of observed values X_i , where $i = 1, 2, \dots, n$, the LASSO cost function is defined as:

$$\min_{\beta} \left(\sum (y_i - X_i \beta)^2 + \lambda \sum |\beta_j| \right)$$

The objective is to determine the optimal regression coefficients β that minimize the total cost function. The L1 regularization term $|\beta_j|$ induces sparsity by forcing some coefficients to zero. Upon solving for β , the resultant coefficient vector $\beta = (\beta_1, \beta_2, \dots, \beta_j)$ is obtained. The penalty term is strictly controlled by a regularization parameter λ that determines the degree of shrinkage applied to the coefficients. Ridge regression, in contrast, applies the L2 regularization penalty that penalizes large coefficient magnitudes rather than forcing them to zero [20]. Elastic Net (Enet) Regression combines both L1 (LASSO) and L2 (Ridge) penalties. In biomarker discovery using large and high-dimensional datasets, LASSO is applied to eliminate less important features, improving its performance while reducing noise that might overfit the data [21].

Support Vector Machine (SVM) is widely used for both regression and classification analysis. SVM finds and optimizes a hyperplane that acts as a $(d-1)$ -dimensional classification boundary in a d -dimensional dataset, ensuring better generalization [22]. Given a dataset with n data points:

$$D = \{ (x_i, y_i) \}_{i=1}^n, x_i \in \{R\}^d, y_i \in \{-1, 1\}$$

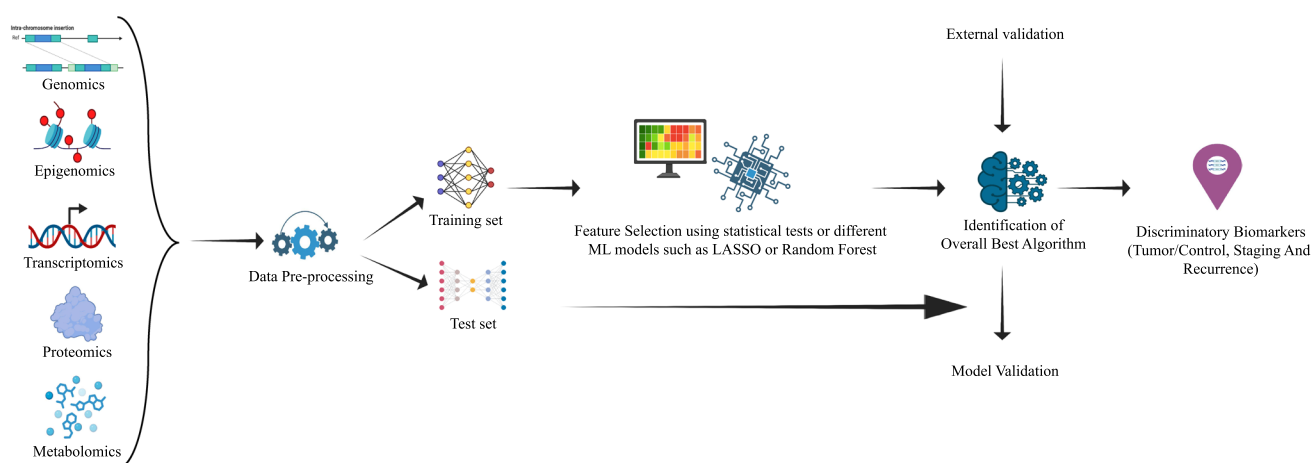


Fig. 1 Schematic representation of the workflow for biomarker identification using machine learning. Multi-omics data, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics, are collected from patient samples and pre-processed (e.g., normalization, batch correction). The data is divided into training and test sets, with the training set undergoing feature selection to identify key genes or molecules using statistical tests or machine learning models. An appropriate algorithm is selected for further filtering, and external datasets are used for validation. The test set is employed to evaluate the model using metrics such as accuracy, ROC-AUC, sensitivity, and specificity. Biomarkers are ultimately selected based on their contribution to model performance and biological relevance, providing insights into potential diagnostic, prognostic, or therapeutic applications [18]

where:

- x_i is the feature vectors in a d -dimensional space
- y_i is the class label (+1 or -1).

A hyperplane in d -dimensional space is given by:

$$w^T x + b = 0$$

where:

- w is the weight vector (normal to the hyperplane).
- b is the bias term

Classification is perfect when:

$$y_i(w^T x_i + b) \geq 1, \forall i$$

SVM can be solved using kernel function (kernel-SVM) or Lagrangian Dual Formulation to handle complex datasets. SVM is a widely used classification algorithm in bioinformatics recognized for its robustness and handling of complex biological data [23]. Similarly, Decision Trees (DT) have applications like those of SVM. DT is a recursive algorithm that splits features based on a threshold value, forming a hierarchical tree-like structure, using criteria such as Gini Index or Shannon Entropy [24]. However, DTs are prone to overfitting, which can be mitigated by Random Forest (RF), an ensemble learning method that constructs multiple decision trees using bootstrap aggregation (bagging). It follows a Bootstrap Aggregation approach to aggregate their outputs. RF improves classification performance and robustness, making it highly effective in analyzing multi-omics datasets [25]. K-Nearest Neighbors (K-NN) is another widely used algorithm for both regression and classification tasks, which determines the class or value of an input point by analyzing its *k-nearest neighbors*. The closeness of data points is generally measured using the Euclidean distance or other distance metrics, such as the Manhattan distance. The class label is assigned centered around the majority vote among the *k-nearest neighbors* [26]. K-NN is highly flexible since it does not assume any prior distribution about the data. The choice of k determines the performance of the model. A small k might cause overfitting due to sensitivity to noise, whereas a large k may lead to underfitting. K-NN struggles with high-dimensional data. Therefore, it is essential to apply dimensionality reduction techniques, such as Principal Component Analysis (PCA), to improve its performance and efficiency [27, 28].

Furthermore, the performance of the various ML models is evaluated with evaluation metrics. The model evaluation decides whether a model will generalize to unseen data and biases through various techniques and metrics according to the type of model. For classification algorithms, the confusion matrix divides the test samples based on the true and predicted values: true positive (TP), correctly predicts a positive instances, true negative (TN) correctly predicts a negative instances, a false positive (FP) is incorrectly predicting a positive instances, and a false negative (FN) is incorrectly predicting a negative instances. One of the basic metrics derived from these values is accuracy, which measures the proportion of correct predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

However, accuracy can be misleading in imbalanced datasets. In such cases, precision recall (sensitivity) and specificity offer more informative assessments. The formulas are given as:

$$Precision = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$\text{Sensitivity/Recall} = \frac{TP}{TP + FN}$$

Importantly in diagnostics, sensitivity/recall are considered as true positive, precision as positive predictive label, and specificity as true negative. Furthermore, the F1 score is the harmonic mean of precision and recall that provides a balanced evaluation:

$$F1\text{Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Another robust measure is the Mathew's correlation coefficient (MCC), which accounts for the correlation between the real and predicted values of the given instances.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

Additionally, the AUC-ROC curve evaluates a model's ability to distinguish between classes by plotting the true positive rate against the false positive rate across different thresholds. A higher area under the curve (AUC) indicates better classification performance.

For the regression algorithms, i.e., tasks like predicting a continuous variable based on one or more other variables, the evaluation measures how closely the model predicts the value of the dependent variable based on an independent variable to the actual value. One of the most common metrics include the Pearson's correlation coefficient, which is defined as:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Another key metric, R^2 (coefficient of determination), represents the proportion of variance in the target variable y that is accounted for by the predicted values $\hat{y}$, in comparison to the variance explained by the mean $\bar{y}$ in the test dataset. An $R^2 = 1$ indicates a perfect fit, while $R^2 = 0$ suggests no useful prediction. It is defined as:

$$R^2 = 1 - \frac{SS(y - \hat{y})}{SS(y - \bar{y})}$$

where SS denotes the sum of squares calculated on the test data. Moreover, in regression models, error calculations play a crucial role in quantifying how close the predicted values are to the actual values. Mean Absolute Error (MAE) is a simple metric that computes the average absolute difference between the predicted and actual values, indicating overall error. Additionally, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) take these differences a little further by squaring them, which penalizes large errors more compared to small ones. Further, these give the error in the outcome scale. An extended version of MSE is the Mean Absolute Percentage Error (MAPE), which is also a popular metric that presents the error in percentage terms, giving a more interpretable view of the respective error in terms of scales of actual values [29, 30].

3 Multi-omics data in bladder cancer

3.1 Genomics

Genomics represents the complete DNA sequence of an organism and plays a critical role in identifying cancer-specific biomarkers as a basis for diagnosis, prognosis, and personalized therapy [31]. It is identified as the most relevant field in the research of cancer as the disease progresses due to changes or variations the genetic material. The genomic alterations associated with cancer, includes SNPs, mutations in coding regions, genomic loss or amplification, chromosomal rearrangements, aberrant methylation, and changes in gene expression [32–34]. Machine learning algorithms, with

the help of bioinformatics tools are used to identify potential bladder cancer biomarkers, where it was trained on gene expression data to distinguish between the phenotypes of bladder cancer [35, 36].

A study conducted by Liosis et al. used the elastic-net regularized logistic regression, which combines LASSO and ridge regression to handle high-dimensional genomic data. With statistical modelling, it led to the identification of two novel gene signatures for bladder cancer, where 55 genes including ABCC2, OASL, PROM2 are correlated with Neoadjuvant chemotherapy therapy response, achieving an accuracy rate of 77.42% with an AUC of 0.7174, alongside 57 genes including ACBD7, DAPK2, CLDN6 are associated with BCa disease progression, with an impressive accuracy of 92.11% of AUC 1.0 [37]. This capability to identify specific genetic markers was crucial for tailoring personalized treatment strategies, which allowed clinicians to predict patient outcomes more effectively. Such advancements in biomarker discovery are vital, especially in a landscape where individualized therapy is becoming the norm in cancer management.

Similarly, Barbour et al., evaluated the association of ERCC2 mutations with bladder cancer prognosis through whole-genome sequencing data. They found that ERCC2 mutations were independently associated with favourable prognosis in bladder cancer and it was a helpful clinical marker for risk stratification and treatment decisions. To further elucidate the clinical relevance of ERCC2 mutations, a SVM model was developed by exploiting the altered patterns of somatic mutation distributions in tumours harbouring ERCC2 mutations. The SVM model was able to predict, with high accuracy, whether a given ERCC2 mutation was pathogenic or not and this would be useful to identify patients likely to respond to cisplatin therapy [38]. The clinical significance was further confirmed by another related study where the researchers analysed pretreatment transurethral resection material from patients with MIBC who received NAC before radical surgery. On analysis, they found out that deleterious mutations in ERCC2 strongly correlated with pathological response to NAC. Interestingly, ERCC2 mutations accounted for 13% in responders but only 2% in non-responders. This indicates that ERCC2 mutation status could be a very important biomarker for predicting the efficacy of NAC in bladder cancer patients and thus could direct the decision-making process regarding treatment, which may benefit the patient [39].

Furthermore, the study done by Yu et al. employed the use of the GenEpi ML model where it could identify interactions between genetic variants like within-gene and cross-gene interactions linked to bladder cancer phenotypes [40]. The study recruited Caucasian participants who are controls in the UK Biobank, matching cases to controls with age, sex, BMI, and smoking status. Four interactions including ST7L-ADSS2, FHIT-CHDH, LARP4B-LHPP, and RBFOX3-MPRIP were replicated in an independent set of 856 non-Caucasian participants who were associated with bladder cancer risk, where they developed an iPRS (Interaction-empowered polygenic risk score) that incorporates the interactions which showed a hazard ratio of 1.81 when comparing the highest tertile to the lowest tertile, and achieved an AUC of 0.91 which indicated an enhanced bladder cancer screening [41].

Additionally, Li et al. developed a novel diagnostic model to identify biomarkers in mitochondria-related genes as its dysfunction played a role in bladder cancer progression. LASSO and SVM-RFE algorithms were used and major biomarkers like GLRX2, NMT1, OXSM, TRAF3IP3 were identified, which had an AUC of 0.912 which meant mitochondria-related genes help in early detection of bladder cancer which may aid in patients' identification who would benefit from aggressive monitoring or intervention [42]. Similarly, Qian Yu et al., resorted to computational methods for studying the expression of genes involved in immuno-regulation. Using such study, the NR4A1 gene is shown to be an essential prognostic biomarker, which impacts on bladder cancer invasion through the involvement of immune cells like regulatory T cells and macrophages [43]. It establishes the incorporation of these immune cell interactions into bladder cancer, which would eventually characterize the disease progression or response to treatments.

Despite the advancement in machine learning for the exploration of biomarkers for bladder cancer, certain drawbacks exist. The biomarkers that are discovered using machine learning models should be verified in the laboratory. Moreover, there is a challenge posed by the limited availability of large well-annotated genomic databases for bladder cancer that can impede the construction of effective ML models. The genomic profiling of patients suffering from bladder cancer with African American, Asian American, and Tunisian origins revealed that AKT levels were substantially higher in these groups as compared to European American patients featured above underscored the importance of considering diverse populations in the study of genomes to identify the biomarkers that may hold relevance to clinical practice [44]. Table 1 presents key biomarkers identified from genomic data using various machine learning approaches, detailing their functions, associated pathways, regulators, and their roles as biomarkers in other cancers based on the GeneCards database and literature.

Table 1 Prominent biomarkers found by ML models using Genomic data

Biomaker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
ABCC2	Efflux transporter by actively pumping chemotherapeutic drugs out of cancer cells	Drug ADME and Statin Pathway	FOXO, FOXD3, FOXJ1, FOXJ2, FOXL1, GATA-2, GATA-3, HIF-3, Sox5	Gastric	Prognostic	[37, 45, 46]
OASL	A crucial prognostic and immunological biomarker and play an important role in the tumour microenvironment	Antiviral mechanism by IFN-stimulated genes	Elk-1, Irf-2, Nrf-2, Olf-1, PPARgamma1/2, RFX1, STAT1alpha/beta	Breast	Prognostic	[37, 47, 48]
PROM2	Its upregulation inhibits ferroptosis through facilitated iron export and promotes tumorigenesis	Ferritin pathway	c-Myc, CP2, GATA-6, Max, MyoD, NF-1, NF-1/L, p53	Lung squamous cell carcinoma	Prognostic and Diagnostic	[37, 49, 50]
ACBD7	Energy regulation via leptin-melanocortin pathway and involved in cell proliferation and differentiation	Insulin-like growth factor signaling pathway	C/EBPbeta, CUTL1, Evi-1, Lyf-1, NF-AT, NF-AT1, NF-AT2, NF-AT3, NF-AT4, Nrf-2	Thyroid	Diagnostic	[37, 51–53]
DAPK2	Upregulation causes the transition of autophagy to apoptosis	AKT1/CyclinD1 pathway	Ahr, Arnt, C/EBPalpha, CREB, MyoD, Pax-4a, POU2F1, POU2F1a, Roaz, ZID	Colorectal	Prognostic	[37, 54]
CLDN6	A tight junction protein that increases drug resistance by altering cell adhesion	PI3K/Akt signaling pathway	AP-1, ATF-2, c-Jun, COUP, COUP-TF	Ovarian Prognostic		[37, 55, 56]
GLRX2	Involved in protein folding, redox regulation, and is associated with mitochondrial biological activity that regulates its function and energy production processes such as ATP synthesis	PAK Pathway	aMEF-2, E47, GATA-1, ITF-2, MEF-2A	Oral squamous cell carcinoma	Prognostic	[42, 57]
NMT1	Involved in protein modification and signalling processes and promotes tumour	Apoptotic pathway and Programmed Cell Death	AP-1, c-Fos, c-Jun, Cart1, CBF(2), CBF-A, CBF-B, CBF-C, IRF-1, MAZR	Breast cancer	Prognostic	[42, 58]
TRAF3IP3	Codes proteins that play an important role in various cellular functions	c-Jun N-terminal kinase signal transduction pathway	AML1a, AP-1, ATF-2, c-Jun, NF-kappaB, NF-kappaB1	Breast	Prognostic	[42, 59]
ERCC2	Encodes a DNA helicase essential for nucleotide excision repair, a process critical for correcting DNA damage. Its mutations were an independent predictor for prognosis in bladder cancer patients	RNA Polymerase II Transcription Initiation and Promoter Clearance	AML1a, CBF(2), En-1, GR, Irf-2, LCR-F1, MZF-1, p53, ZID	Breast	Prognostic	[38, 60]

Table 1 (continued)

Biomaker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
ST7L-ADSS2	ST7L is part of the Wnt/GSK-3 β -catenin signalling pathway. Contact of this ADSS2 with ST7L might inhibit the tumour suppressive role of the latter for the development of BC	Denovo pathway and in the salvage pathway of purine nucleotide biosynthesis	CUTL1, FOXC1, IRF-7A, NR5F1, Pax-5, Roaz, STAT3, USF1, USF2	N/A	Diagnostic	[41]
FHIT-CHDH	FHIT is a known tumour suppressor, and aberrant expression in this gene leads to decreased apoptosis and increased resistance to genotoxic agents. The association of CHDH with arsenic metabolism may relate to arsenic exposure, an established risk factor for BC	Methionine de novo and salvage pathway, One-carbon metabolism and related pathways	c-Ets-1, GATA-2, ITF-2, NF-kB, Tal1B, YY1	N/A	Diagnostic	[41]
LARP4B-LHPP	Their interactions inhibit the proliferation of BC cells by inhibition of the AKT/p65 pathway	Pyrophosphate hydrolysis pathway	AP-1, AREB6, ARP-1, GCNF, ISGF-3, Pax-2	N/A	Diagnostic	[41]

*Taken from Genecards database [61]
N/A not available

3.2 Epigenomics

Genes that undergo epigenetic modification can lead to uncontrolled cell growth and division; in the case of the urothelial lining of the bladder, this may lead to urothelial cancer (UCa). Several epigenetic modifications such as methylation, acetylation, phosphorylation, SUMOylation have been identified in the progression of cancer. Considering the dynamic nature of biological systems, ML can be utilized to analyze large datasets of epigenetic data to identify biomarkers that have undergone significant epigenetic modifications. The necessity of these studies is to improve the prognostic and diagnostic significance in BCa.

Multiple independent studies utilized ML to analyze BCa prognostic markers and to develop models in order to achieve a reliable prognostic tool, also to improve BCa surveillance in patients. A study conducted by Zhang et al. [62] employed 101 combinations of 10 different algorithms to identify genes that exhibit aberrant methylation, which may serve as potential epigenetic biomarkers. From this analysis, nine genes were identified based on the C-indices (concordance indices), with StepCox and survivalSVM recognized as the top-performing algorithms. Following this, the identified genes underwent further evaluation of their prognostic capabilities. Ultimately, three genes-RNH1, TAP1, and AHNK were selected to construct a nomogram along with age and tumor stage, which were identified as independent factors through both univariate and multivariate Cox analyses of the nine genes. Similarly, Jiang et al. [63] developed a machine learning-based urinary DNA methylation diagnostic panel to distinguish BCa patients from non-BCa patients by using genes that have undergone aberrant methylation. The performance of the panel was evaluated, and three individual genes with significant prognostic potential were identified: ZNF671, OTX1, and IRF8, which achieved an impressive area under the curve (AUC) values of 0.86, 0.82, and 0.81, respectively, indicating effective classification performance. In contrast, Köhler et al. [64] applied three machine learning methods (random forest (RF), boosted trees (BT), and LASSO) in tandem to the urine samples, identifying 65 abnormally methylated CpG sites and arrived at a combination of biomarkers, which led to the construction of a decision matrix for an array of all possible combinations ranked for their prognostic ability. Notably, the combination of ALOX and TRPS1 manifested the highest sensitivity of 61%. In addition, van der Heijden et al. [65] developed a urinary methylation biomarker diagnostic classifier designed to enhance the monitoring of BCa and reduce the reliance on invasive cystoscopy procedures in patients undergoing surveillance for BCa. Urine samples were analyzed, and using logistic regression, a model was developed with a combination of three markers CFTR, SALL3, and TWIST1 that outperformed other combinations with a formidable AUC value of 0.87. Furthermore, the combined analysis of the patients' cytological data and the three-gene methylation classifier was superior to the analysis of cytological data alone. Additionally, the study indicated that this combined approach could have prevented cystoscopies in 36% of patients within the validation cohort that was used to assess the model.

In addition to the identification of biomarkers, recent studies have increasingly focused on evaluating potential machine learning-based BCa risk predicting models, to substantially enhance diagnostic and prognostic capabilities across diverse medical domains. For instance, Mohanad et al. [66] analysed aberrant methylation patterns in promoter sequences of DNA damage repair (DDR) genes and found 154 DDR genes that have undergone aberrant methylation in BCa patients. Following hierarchical clustering of 154 genes, 12 genes were selected. Notably, two genes that revealed significant methylation in their promoter regions: MutS homolog (MSH4) and Retinoblastoma binding protein 8 (RBBP8), commonly referred to as C-terminal binding protein (CtBP)-interacting protein (CtIP), were culled for model development. Hypermethylation of RBBP8 alters DNA damage repair pathways [67]. The human MutS homolog 5 (hMSH5) maintains genome stability by inhibiting double-strand breaks (DSBs) through Non-homologous end joining (NHEJ) pathway. hMSH5 lead to the formation of 53BP1 foci as DSB sites. They are protective in nature and hMSH5 deficiency leads to the slow formation of 53BP1 [68]. Several predictive machine learning algorithms (Decision Tree (DT), Kernel Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), and Random Forest (RF) were employed. Among these, the KNN algorithm achieved an outstanding accuracy of $90.05 \pm 4.5\%$, indicating excellent predictive performance.

A notable study conducted by Chatterjee et al. [69] investigated the potential of a BCa prognostic risk model (risk score) utilizing a signature gene panel by combining methylation, RNAseq and patients' clinical data analyses. Methylation data were collected and subjected to differential analysis to identify prognostic genes whose methylation status varies significantly. Subsequently, the most significant prognostic genes were subjected to network analysis, which revealed six hub genes (given in table) that were most suitable for developing the risk model. It was demonstrated that the model can independently predict the survival of BCa patients. A similar study by Peng et al. [70] utilized methylation data from The Cancer Genome Atlas Program (TCGA) to develop a multiclass prediction system and built a methylation-based model to identify significant biomarkers for predicting common urothelial carcinomas such as bladder urothelial

carcinoma (BLCA), renal clear cell carcinoma (KIRC) (also known as clear cell renal cell carcinoma (ccRCC)), a variant of renal carcinoma characterized by distinct clear cell histology and prostate-adenocarcinoma (PRAD), a neoplastic lesion of the prostate gland. The multiclass prediction system distinguished between normal tissue and cancer tissue and the origin to which it belongs. Remarkably, this model was also able to differentiate between metastatic lymph nodes and normal lymph nodes. Furthermore, the study also examined the association between their methylation model and tumor node metastasis (TNM) stage. The performance of the methylation model alone was superior and the combination of the two proved to be a promising prognostic tool.

One intriguing study by Guo et al. [71] intended to develop a novel methylation risk model for bladder cancer (MRSB) by selecting various candidate CpG sites using the support vector machine-recursive feature elimination (SVM-RFE) and least absolute shrinkage and selection operation (LASSO) methods. The eight candidate genes that were identified are subsequently used to build the model, that includes APC, SLCO4A1, ZC3H3, ADARB2, COL9A2, KRTDAP, POU3F3 and TNFAIP8L3. An impressive AUC value of 0.86 was obtained for the developed MRSB, which corroborates its predictive ability. Additionally, an effective nomogram was constructed to predict the individual progression risk, integrating factors such as the methylation risk score (MRSB), sex, age and tumor stage.

The integration of machine learning and DNA methylation analysis is proving to be a transformative approach in the early detection and prognosis of bladder cancer. Studies have successfully identified multiple genes with aberrant methylation that serve as potential biomarkers, demonstrating high predictive accuracy through various models and also significantly enhancing the quality of care and treatment results for patients with bladder cancer. The prominent biomarkers found using epigenomic studies by ML models along with their functions associated pathways, regulators, and their roles as biomarkers in other cancers based on the GeneCards database and literature are given in Table 2.

3.3 Transcriptomics

Transcriptomics is the study of the total RNA transcripts produced from a genome. It aims at analysing the gene expression, regulation, and, most importantly, cellular functioning patterns [111]. These RNA molecules can be measured and analysed using methods like RNA sequencing, which would grant scientists the knowledge about how genes get activated, interact, and respond to other biological or environmental stimuli [112]. This is critical in establishing comprehensive analysis of gene expression and regulation therapeutic targets in domains like developmental biology, personalized medicine, and disease research [113]. Transcriptomic studies represent a key method of identifying biomarkers for bladder cancer by yielding insights into the molecular mechanisms underlying the disease. From analysing the total set of RNA transcripts within cancer cells, gene expression patterns can be found to distinguish bladder cancer from healthy tissue [114]. These studies enable the discovery of potential biomarkers that can serve as indicators for cancer diagnosis, prognosis, or treatment response [115]. Following the advancement of high-throughput sequencing and bioinformatics, with transcriptomic data, classifications of bladder cancer subtypes are more precise and assist in personalized treatment strategies [116].

A study by Wang et al. analysed gene expression datasets using ML models such as LASSO logistics regression and identified the overexpressed gene PTHLH as a biomarker. Moreover, it was observed to be a prominent biomarker for BC with excellent diagnostic and prognostic capability with a satisfactory prediction efficiency of AUC = 0.775. In vitro validation done by knocking out the PTHLH gene from BC cells and using the co-culture method and immunofluorescence demonstrated supporting evidence of the involvement of PTHLH in BC proliferation [117]. In another study, the role of SMAD6 in BC proliferation was studied by Chen et al., with mRNAseq data using LASSO-Logistic Regression, Boruta, random-forest, svmREF, and XGboost ML models. In this transcriptome study, SMAD6 was found to be a evident prognostic biomarker with the downregulation of SMAD6 resulting in the inhibition of BC proliferation. Wound healing assays and colony forming assays were performed for validation, and sufficient evidence was obtained [118]. Zhang et al., conducted a transcriptome study using bulk RNAseq and single cell RNA seq data of BC to find a novel biomarker. The analysis of the data was done using CoxBoost ML model and LDLRAD3 was found as a prognostic biomarker with great predictive efficiency. The upregulation of LDLRAD3 expression was found to be a leading cause for poor prognosis of BC patients. The results were further validated using invivo mouse models [119].

Another study done by Song et al., analysed the gene expression data and single cell transcriptome data using ML models such as LASSO and Logistic regression model, and found LIG1 to be upregulated in BC tissues, leading to poor prognosis. LIG1 was also able to show good prognostic prediction of BCa with a prediction efficiency of AUC = 0.793 and identified as a prognostic biomarker of BCa [120]. Another transcriptomic study by Shen et al., analyzed gene expression

Table 2 Prominent biomarkers found by ML models using epigenomic data

Biomaker	Function	Associated pathways*	Regulating transcription fac-tors*	Reported in other cancers	Biomarker type	References
TAP1	Functions in antigen presen-tation, facilitating immune response through MHC class I molecules. Downregulation of the protein causes the cancer cells to evade immu-nosurveillance	Antigen processing-Cross presentation and Class I MHC mediated antigen processing and presentation	AML1a, STAT3	Ovarian	Prognostic and Diagnostic	[18, 62, 72, 73]
AHNAK	Serves as a scaffold pro-tein, participating in cell signalling and cytoskeletal organization. The upregula-tion of AHNAK is observed in BC tissues	Phospholipase-C Pathway	Bach1, Egr-2, Elk-1, FOXD3, HTF, LUN-1, Pax-5, RelA, Spz1	Breast	Prognostic and Diagnostic	[62, 74–76]
RPP21	An essential subunit of the RNase P complex is involved in the maturation of tRNA by cleaving its precursor forms, thereby promoting proper protein synthesis, and also alongside RPP29, is involved in the DNA damage response	rRNA processing in the nucleus and cytosol and tRNA processing	E2F, Nkx2-5, p53, PPAR-gamma1/2	Liver	Prognostic and Diagnostic	[62, 77–79]
ZNF671	Functions as a tumor suppres-sor, inhibiting cell prolifera-tion and metastasis	RNA Polymerase II Transcrip-tion pathway	AREB6	Cervical	Prognostic and Diagnostic	[63, 80, 81]
OTX1	Acts as a transcription factor that promotes oncogenesis	Beta-2 adrenergic-dependent CFTR expression pathway	AREB6, E2F, HOXA5, POU2F1, SRY	Ovarian	Prognostic and Diagnostic	[63, 82, 83]
IRF8	Involved in regulating immune responses and myeloid cell differentiation	NF-kappaB Signaling pathway, Type II interferon signaling pathway, Toll-Like Receptors Pathway	C/EBPbet, COMP1, E4BP4, Evi-1, FOXO3b,HNF-1	Rectal	Prognostic and Diagnostic	[63, 84, 85]
ALOX	Encodes arachidonate 5-lipoxygenase, crucial for leukotriene synthesis and inflammation	Biosynthesis of DPA-derived SPMs, Biosynthesis of spe-cialized proresolving media-tors (SPMs) and NF-kappaB Signaling pathways	AP-1, ATF-2, c-Jun, Egr-1, GR, p53, Sp1	Breast	To construct a decision matrix	[64, 86, 87]
TRPS1	A marker that is crucial in metastasis and identifying triple negative breast cancer	N/A	aMEF-2, E2F, HFH-1, MEF-2A, STAT3	Breast	To construct a decision matrix	[64, 88]

Table 2 (continued)

Biomarker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
CFTR	Encodes the cystic fibrosis transmembrane conductance regulator, crucial for chloride ion transport in epithelial cells	AMPK Enzyme Complex Pathway	AP-1, ATF-2, c-Jun, Sp1, SRF, YY1	Colorectal	Monitoring Bladder cancer	[65, 89, 90]
SALL3	A transcription factor involved in embryonic development and tissue patterning	N/A	E2F, Evi-1, FOXL1, POU3F2	Liver	Monitoring Bladder cancer	[65, 91, 92]
TWIST1	A transcription factor that regulates transition from epithelial cells to mesenchymal cells (EMT), facilitating cell migration and invasion during development	Notch-mediated HES/HEY network pathway, Cell migration and invasion through p75NTR and CKAP4 signaling pathway	AP-1, AREB6, ATF-2, c-Jun, MyoD, p53, Pax-3, STAT3	Breast	Monitoring Bladder cancer	[65, 93, 94]
RBBP8	An endonuclease involved in DNA damage repair and cell cycle checkpoint control	G2/M DNA damage checkpoint pathway	Egr-3, FOXJ2, Nkx2-2, Nkx2-5, SRF	Gastric	Prognostic and Diagnostic	[66, 95, 96]
MSH4	Repair of DNA mismatches during homologous recombination	Cell cycle and meiotic pathway	AML1a, AP-4, C/EBPalpha, E4BP4, Evi-1, FOXO3, POU2F1	Colorectal	Prognostic and Diagnostic	[66, 97, 98]
EGFR	Encodes the epidermal growth factor receptor, which is crucial for cell growth, differentiation, and proliferation and driver of cancer progression	Cellular Apoptosis Pathway, Signaling by EGFR in Cancer, GPCR Pathway	c-Myc, Max1, p53, STAT3	Non-Small Cell Lung cancer	Prognostic	[69, 99, 100]
NFE2	A transcription factor playing an important role in erythropoiesis and antioxidant defence	RNA Polymerase II Transcription pathway, Generic Transcription Pathway	AML1a	Liver	Prognostic	[69, 101, 102]
ARL4D	Involved in cell signaling and cytoskeletal organization, it influences processes like cell migration through its effects on cytoskeletal dynamics	N/A	AML1a, ATF-2, c-Jun, FOXJ2, NF-kappaB, Nkx5-1, Pax-5, USF1	Gastric	Prognostic	[69, 103, 104]
SH3RF2	An E3 ubiquitin ligase that regulates protein degradation and cellular signaling pathways	Anti-apoptotic regulator of the JNK pathway	AML1a, Arnt, Cdc5, E47, Gfi-1, HSF2, Lmo2, POU6F1, Tal-1beta	Ovarian	Prognostic	[69, 105, 106]

Table 2 (continued)

Biomaker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
CDH3	Encodes cadherin 3 (P-cadherin), which regulates aspects of cell adhesion and maintaining tissue architecture	ERK signalling pathway	AP-1, AREB6, ATF-2, c-Jun, COUP, HNF-4alpha1, Sp1	Colorectal	Prognostic	[69, 107, 108]
TNFAIP8L3	Involved in regulating inflammation and apoptosis	Enhance the activity of the PI3K-AKT and MEK-ERK pathways	AREB6, ATF6, FOXC1, HFH-3, NF-AT, TBP	Head and Neck	To build MRSB	[71, 109, 110]

*Taken from Genecards database [38]
N/A not available

and single-cell RNASeq datasets using the ML models SVM-RFE and LASSO. Further, PPP2R2B was found to be a prognostic biomarker for BC, with PPP2R2B being downregulated in BC tissues [121]. Li et al., conducted a similar transcriptomic study with single cell and bulk RNAseq data using the ML models CoxBoost, Enet, Lasso, RSF, plsRcox, GBM, Stepwise Cox, Boruta, pamr, and Ridge. After the analysis, the study generated 3 novel prognostic biomarkers FN1, CRYAB, and COL6A1 with good prediction accuracy [122]. Another transcriptomic study utilized 10 different ML models and found that SVM model was the best one among them in predicting the prognosis of BC. It also identified FOXP3 and TOX as prominent prognostic biomarkers for BC. The upregulation of FOXP3 and downregulation of TOX was found to be the leading cause for poor prognosis among BC patients [123].

By reviewing these different transcriptomic studies, SVM AND LASSO logistic regression model was found to be used the most for predicting the biomarkers from the transcriptomic data, although the biomarkers found by each of them seemed to vary. Most of the biomarkers found were prognostic in nature and were genes with their upregulation or downregulation having an impact on the proliferation of BC cells. Most of the biomarkers found were oncogenic in nature, ie, their enhanced expression in the bladder tissues aid in the proliferation of BC.

The common limitations that have been encountered in these studies were of the input data; the data collected were old, retrospective in nature, less sample size, or confined to a particular geographical location. So, obtaining data from a large population with different ethnicities can aid in more accurate predictions by ML models. Most of the biomarkers found were not clinically validated and this poses a concern over their clinical usage. Hence, further studies can concentrate on clinically validating these biomarkers for them to be used in the diagnosis of BC. Table 3 denotes the different biomarkers found in these studies, along with their functions, associated pathways, regulators, and roles as biomarkers in other cancers based on the GeneCards database and literature.

3.4 Proteomics

Proteomes from urinary exosomes are considered excellent sources for bladder cancer biomarkers mainly due to two reasons: one, being a non-invasive detection method, and second, its direct contact with the tumor [137, 138]. However, early studies utilizing proteomic biomarkers for classification were hindered by high intra-tumor heterogeneity and difficulties in comparing biomarkers across diverse studies [138]. This would change as the need for biomarker-based classification, a painless, non-invasive method compared to traditional cystoscopy, is imperative. Traditional urine cytology shows an AUC of only 0.718, coupled with low sensitivity [139].

Most proteomic studies directed to find proteomic biomarkers follow the usual discovery-internal-external validation phases. Efforts have been made to make these phases more accurate and efficient by introducing new screening technologies and, most importantly, using ML models to classify biomarkers that can effectively discriminate. Discovery phases mostly use Capillary electrophoresis coupled to Mass Spectrometry (CE-MS), Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS) by Data-Dependent Acquisition (DDA) or Liquid Chromatography coupled with Tandem Mass Spectrometry (LC-MS/MS) to effectively profile proteins for urine sources. In one study, an aptamer-based screening (called SOMASCAN) of urinary samples was done; This makes it possible for thousands of proteins to be screened at once, improving sensitivity [140]. These proteins, upon identification, usually undergo quantification and sequencing. Validation, yet another important phase, employs ELISA by assessing the fold changes and AUC [141] for internal validation by splitting the dataset into training and testing sets, a well-practiced process. External validation is almost always done on larger cohorts of independent sets through ELISA. In most proteomic studies, many biomarkers are identified as robust.

The advent of Machine learning rises from the need for both higher accuracy and efficiency of classification, as repeated tests are necessary for disease diagnosis and recurrence surveillance [140]. A rather narrow types of algorithms are implemented in classification due to the biomarkers being high dimensional [138], and being a better option compared to unsupervised algorithms. Support Vector Machines (SVM) (also Msvm-RFE), with its inherent capability to classify high dimensional parameters, is an excellent resource. Another well-used algorithm is Random Forest Analysis (RFA). Vanarsa, Castillo and Wang et al. constructed an RFA model that rates the importance of the biomarkers that are found, based on their ability to distinguish BC from non-cancer tumors [140]. Principal Component Analysis (PCA) was used to constrict the data to visualize how well the Biomarkers discriminated between different cancer stages, particularly NMIBC, which is usually regarded as an early stage, from MIBC, which refers to the advanced stages.

The models have been shown to have higher levels of AUC and specificity. The study by Frantzi et al, resulted in two biomarker panels for classifying into primary or recurrent bladder cancer. The primary panel, after validation, exhibited an AUC of 0.87 and the Recurrent panel, 0.75, respectively [138]. Interestingly, when incorporating traditional cytology,

Table 3 Prominent biomarkers found by ML models using transcriptomics data

Biomaker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
PTHLH	Regulate aspects of cell proliferation, migration and survival	GPCR downstream signalling pathway	AP-1, deltaCREB	Non-small cell lung cancer	Diagnostic and prognostic	[117, 124, 125]
SMAD6	Regulates the TGF-β and BMP signalling pathway	TGF-beta Pathway, BMP receptor signalling pathway	AML1a	Colorectal	Diagnostic	[118, 126, 127]
LDLRAD3	Regulates formation of High endothelial venules	N/A	ATF-2, CUTL1, Evi1, FOXO1, FOXO1a, GATA-1, HNF-1, HNF-1A, IRF-2, Pax-6	Colorectal	Prognostic	[119, 128]
LIG1	Involvement in senescence, DNA replication, cell cycle and p53 pathway	DNA repair pathways	E2F, Elk-1, GATA-2, MAZR	Breast	Prognostic	[120, 129]
PPP2R2B	Regulates the Wnt pathway	p70S6K Signaling, Beta-Adrenergic Signaling, Wnt signaling pathway	MyoD, NF-E2, p45, Nkx2-5 YY1	Breast	Prognostic	[121, 130]
FN1	Functions in regulating apoptosis	Epithelial to mesenchymal transition pathways, ERK Signaling pathway	AP-1, STAT3	Thyroid	Prognostic	[122, 131]
CRYAB	Acts in the process of regulation of cell motility	Cellular heat stress pathway	c-Myc, Max1, NF-kappaB, Sp1	Colorectal	Prognostic	[122, 132]
COL6A1	Converts normal fibroblast cells to cancer-associated fibroblast cells and can activate the TGF-β pathway	Collagen chain Trimerization and Integrin Pathway	AML1a, AP-1, ATF-2, c-Jun, c-Myc, Ik-1, Max1, NF-kappaB, RelA	Osteosarcoma	Prognostic	[122, 133]
FOXP3	Functions as a TF for regulatory T cells and also regulates the HIF-1α gene expression	NF-kappaB Signaling, Wnt/Hedgehog/Notch signalling pathways	AML1a, CREB, deltaCREB	Pancreatic Adenocarcinoma	Prognostic	[123, 134]
TOX	Regulates the tumour microenvironment and is involved in the pathways of immune regulation and T-cell exhaustion	Mesodermal commitment pathway	GATA-1, NF-AT, Pax-4a/5	Lung Adenocarcinoma	Prognostic	[123, 135, 136]

*Taken from Genecards database [38]
N/A not available

Table 4 Prominent biomarkers found by ML models using proteomic data

Biomarker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
Collagen Fragments	Extracellular matrix remodelling, tissue repairing and cell signaling	TGF- β signaling pathway, PI3K/AKT pathway	LEF-1, B-MYB, GLI1	Breast, Pancreas, Gastric, And Colorectal	Diagnostic	[138, 142, 143]
Apolipoprotein A-I	Liquid transport, anti-inflammation, (potential) role in tumor cell proliferation and angiogenesis	Plasma lipoprotein assembly, remodeling, and clearance and Response to elevated platelet cytosolic Ca ²⁺	C/EBPalpha, COUP HNF-4alpha1/2, PPAR-gamma1/2	Ovarian, Breast, Gallbladder, Liver, Colon, Cervix, Uterus, Prostrate, Kidney, Lung, Head & Neck	Diagnostic and recurrence	[138, 140, 144, 145]
Hemoglobin chains	Oxygen transport, (potential) indicator of hematuria	ERK1/2, P38/MAPK, and JNK pathways	C/EBPbeta, FOXD1/3b, GATA1, IRF1, STAT3, TBP, USF1	Lung, Breast,	Primary and recurrent. More prominent in primary	[138, 146]
Fibrinogen A (FGA)	Blood clotting, indicator of disrupted coagulation pathways	RAS, FAK-ERK, integrin-AKT, and TLR7/8 pathway	E47, Evi-1, FOXO3, ISGF-3, Pbx1a, STAT3, Tal-1	Lung, Gastric, Colorectal	Diagnostic	[141, 144, 147]
Matrix Metalloproteinases (MMPs)	ECM degradation, associated with disease progression and aggressiveness	EGF receptor pathway	AP-1, ATF-2, c-Jun, CREB, SRF	Breast, Pancreas, Lung, Bladder, Colorectal, Ovarian, Prostrate, Brain	Diagnostic and Prognostic	[138, 140, 144, 148]
α -2-macroglobulin (A2M)	protease inhibition and transport of cytokines	Classical pathway of the complement system	AREB6, CREB, FAC1, GR, Lmo2, STAT5A	Leukemia, Prostrate	Diagnosis and recurrence	[141, 144, 149]
Uromodulin	function not fully understood, but it may play a role in preventing kidney stones and urinary tract infections	Metabolism of proteins and transport to the Golgi and subsequent modification	AhR, Brachyury, HFH-1, Nkx3-1, XBP-1	Prostrate	Muscle invasive (advanced) vs Non-Muscle invasive (early)	[150, 151]
Inter-Alpha-Trypsin Inhibitor Heavy Chain 2 (ITIH2)	Regulate inflammation and extracellular matrix remodeling	PI3K/AKT signaling pathway, Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) and Metabolism of proteins	FOXJ2, HFH-3, MyoD, MZF-1, Olf-1, PPAR-gamma1, RSRFC4, STAT5A	Glioma, Ovarian, Liver	Diagnostic	[143, 144, 152]
Complement components	Inflammation and defense against pathogens. Dysregulation has been implicated in cancer development and progression	classical and complement pathway	AP-1, ATF-2, c-Jun, PPAR-gamma1/2, STAT3	Lung, Colorectal, Ovarian, Gastric, Oropharyngeal	Diagnosis and recurrence	[138, 143, 144, 153]

Table 4 (continued)

Biomaker	Function	Associated pathways*	Regulating transcription factors*	Reported in other cancers	Biomarker type	References
Fascin-1 (FSCN1)	Embryogenesis, tissue repair, fibrogenesis, cell motility invasion, and immune response factor	Cytokine signalling, innate immune system	CREB, p53	Breast, Ovarian, Pancreatic, Skin, Lung, Gastric, Esophageal, Kidney, Colon And Prostrate	Recurrence	[138, 144, 154]

*Taken from Genecards database [61]

the AUC of the recurrent panel increased to 0.87, signifying high accuracy. The study also mentioned that biomarkers in both the panels overlap with each other and with previous studies that reported the same set of proteins, further validating the panel. This panel is also great for identifying low-risk recurrent BC biomarkers. Suh et al. noted that a panel of 8 proteins after validation, from DEPs by Data-Independent Acquisition (DIA) methods performed well, with AUC of both their models being 0.845 and 0.842 respectively [141]. Table 4 depicts the key biomarkers found using proteomic studies by ML models, along with their functions, associated pathways, regulators, and their roles as biomarkers in other cancers based on the GeneCards database and literature.

As much as the identification of proteomic biomarkers is of great importance for efficiency and diagnostic milestones for urothelial cancer therapy, several limitations and shortcomings loom. The most important one is the limited transition to clinical settings. These identified biomarkers are not yet commonly employed in clinical practice for diagnosis, staging, and recurrence monitoring. More focus should be placed on clinical trials and longitudinal studies to validate these biomarkers in real-world scenarios [155]. This can bridge the gap between research findings and clinical application. The narrow set of ML algorithms being used, mainly just SVM and Random Forest, albeit reliable, can have limitations in handling very complex data structures. The high amount of false-positives in recurrent detection remains another problem [144]. Balancing specificity and sensitivity are critical factors for improved models. Additionally, scarce amounts of biomarkers identified for early-stage cancer is another concern. Many biomarkers have been identified for diagnosing and monitoring advanced urothelial carcinoma, however, very few studies focus on early-stage prediction of urothelial carcinoma, particularly NMIBC.

3.5 Metabolomics

Metabolomics, the study of small molecules involved in biological processes, is another great approach for bladder cancer diagnosis, staging, and recurrence prediction, using biomarkers, here called 'metabolomes'. Urine and serum are commonly used biofluids for metabolomic profiling as they contain metabolites that are reflective of changes in bladder tissue.

Several studies have used various analytical platforms, including Ultra-performance liquid chromatography time-of-flight mass spectrometry (UPLC-TOF-MS) and Gas chromatography-mass spectrometry (GC-MS), to analyze urine metabolites and to differentiate bladder cancer from controls. Shao et.al, used UPLC-TOF-MS and a 4-step marker discovery pipeline to identify 6 potential bladder cancer markers. This pipeline includes selection by detection count, three-parameter Gaussian fitting, statistical test, and AUC. Building a decision tree model based on these 6 markers, the model achieved an accuracy of 76.60% and specificity of 86.67% [156].

Another study by Wang.et.al, utilized UPLC-MS to profile urine metabolites from bladder cancer patients and healthy controls. Through Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) analysis and univariate analysis, 19 metabolites were identified as potential biomarkers. Further analysis done using SVM, Random Forest and Boruta algorithms narrowed down the metabolites to 11. A logistic regression model with these 11 biomarkers showed an AUC of 0.983 and a specificity of 100%, demonstrating exceptional diagnostic capability [157].

GUDCA and Palmitoylcarnitine are two of the most prominent biomarkers among metabolomes. GUDCA is a bile acid that has roles in fat digestion and absorption, solubilizing cholesterol, and facilitating its elimination from the body. It is found to be significantly upregulated in bladder cancer patients [157], perhaps indicating altered bile acid metabolism or a response to cellular stress induced by the tumor. Palmitoylcarnitine is involved in the transport of fatty acids into mitochondria. It is also observed to be highly upregulated in cells (almost tenfold) in bladder cancer patients [157, 158], possibly from the Warburg effect.

Serum Metabolomics, although less common, also holds great promise in diagnosis. Troisi et al. constructed ensemble ML models using serum metabolomics profiles to differentiate between cancer patients and controls, low-grade (G1-2) and high-grade (G3) cancers, and different degrees of muscle invasiveness. These ensemble models demonstrated an accuracy of at least 80% across classification tasks. The study also identified Waskey metabolic pathways contributing to bladder cancer, including perturbations in amino acid, glutathione, glyoxylate, and dicarboxylase metabolism [159].

Even with astounding progress in using metabolomic biomarkers for diagnosis, challenges still loom. Variability in study design, analytical platforms, and data analysis methods across different studies makes it quite difficult to compare and validate results. Using larger cohorts that are well-defined, implementing standardized protocols for sample collection, preparation, and analysis, or focusing on validating identified biomarker panels through independent cohorts

Table 5 Prominent biomarkers found by ML models using metabolomic data

Biomarker	Function	Associated pathways*	Reported in other cancers	Biomarker type	References
ADENOSINE MONOPHOSPHATE (AMP)	Activates AMPK, which may give tumor cells a growth advantage by controlling the metabolic adaptability of cells	ATP and adenosine signaling pathways	N/A	Diagnosis	[157]
GLYCOURSODEOXYCHOLIC ACID (GUDCA)	Bile acid, overexpression in urine, leads to worse bladder conditions	Cellular proliferation, mTOR/S6K1 pathway	Colon	Diagnosis	[157, 160]
CHENODEOXYCHOLIC ACID (CDCA)	Bile acid, overexpression in urine leads to worse bladder conditions and promotes epithelial-mesenchymal transition	Lipid peroxidation, ROS/p38 MAPK/DGAT1 pathway	Acute Myeloid Leukemia	Diagnosis	[157, 161]
GLYCOCHENODEOXYCHOLIC ACID (GCDCA)	Bile acid, overexpression in urine leads to worse bladder conditions	Cellular proliferation	Hepatocellular	Diagnosis	[157, 162]
Palmitoylecarnitine	Induces inflammatory cytokines, drives bladder tumor development	Metabolism of fatty acids, Ca ²⁺ influx	Prostrate	Diagnosis	[157, 163]
Imidazoleacetic acid	Derived from oxidizing histamine, which is released by mast cells during inflammation	Histidine metabolism	Rectal	Diagnosis	[156, 164]
Taurine	Involved in osmoregulation, antioxidant defence and cell signaling	Alanine-3-sulfinate pathway, NF-κB pathway	Lung, Breast	Diagnosis	[156, 165, 166]
Histidine	Derives histamine and therefore imidazole acid, which was found enriched in bladder cancer metabolomes	Cell signaling, Protein synthesis	Prostrate, Ovarian	Diagnosis	[156, 167]
Lactate	High levels associated with Warburg effect in cancer cells	PD-1/PD-L1 pathway, cellular metabolism	Esophageal, Colorectal, Hepatocellular, Pancreatic, Gastric, Breast, Prostrate, Ovarian, Glioma	Diagnosis	[159, 168]
Glucose	Primary energy source	Glycolysis, TCA	Hepatocellular, Pancreatic, Ovarian, Breast, And Colorectal	Diagnosis	[159, 169]

*Taken from GeneCards database [61]
N/A not available

and clinical trials can hopefully facilitate the translation of metabolomic biomarker panels into clinical tools for bladder cancer diagnosis, staging, and recurrence. Table 5 contains the different biomarkers found in these studies, along with their functions, associated pathways and their roles as biomarkers in other cancers based on the GeneCards database and literature.

4 Multi-omics data integration in bladder cancer research

Attaining a comprehensive understanding of the intricate biological processes in cancer requires the integration of variations across multiple levels, including genomics, epigenomics, transcriptomics, metabolomics, and proteomics. This synergy provides a detailed catalogue of these processes, enabling classification, disease subtyping, prediction of biomarkers, and obtaining insights into the pathophysiology of the disease more evident [170]. For instance, identifying an SNP associated with the disease outcome, Copy number variations and its influence on gene expression, epigenetic markers, proteins, and metabolites all give valuable diagnostic and prognostic markers [171]. These synergic effects can be observed, identified, and interlinked by utilizing novel AI and machine learning algorithms. For instance, MOGONET is a supervised multi-omics integration algorithm to analyze and classify data. It can learn omic-specifics using graph convolutional networks (GCN) and efficiently perform cross-omics correlations using view correlation discovery network (VCDN) [172]. Furthermore, Zang et al. introduced a novel approach by combining machine learning and deep learning for multi-omics analysis of MIBC to enhance the prognosis of bladder cancer. Autoencoders were used to integrate gene expression, CNVs, miRNA expression, and DNA methylation from TCGA-MIBC data. Additionally, Machine learning algorithms, including random forest, Naïve Bayes, k-NN, and Adaboost, classified patients into high-risk and low-risk subtypes. Significant genomic, immune component, and pathway differences were observed between groups, and importantly, KRT7 was identified as a key biomarker for MIBC from the RNA and protein levels [173].

Moreover, the multi-omics integration methods can be categorized into two types: horizontal and vertical. Horizontal integration is studying the same omics (e.g., transcriptomics) across different sample groups, allowing comparative analysis and identification of condition-specific variations. In contrast, vertical integration is used for multi-omics analysis, where different omics layers (e.g., genomics, proteomics, transcriptomics) are integrated for the same set of samples but contain different features [174]. The vertical integration methods can be classified into network, fusion, correlation-based, and similarity-based methods. Several computational tools have been developed to facilitate multi-omics integration. iClusterPlus is a statistical tool used to systematically cluster multi-omics data and identify shared patterns, such as identifying tumor subtypes. MIXOmics, an R package that enables integration and analysis of multi-omics data. MoFA (Multi-omics factor analysis) employs Bayesian tool used to integrate multi-omics data, focusing on factor models while Multiple factor analysis (MFA) integrates omics datasets by projecting them in a low-dimensional space. Similarity network fusion (SNF), which uses a network approach to first create individual networks and later using a non-linear network fusion approach, fuses them into a single similarity network [175].

Several BC studies have successfully integrated multi-omics data to identify key biomarkers, classify tumor subtypes, and improve prognostic predictions. A study conducted by Chu et al. [176] on muscle-invasive urothelial cancer utilized 10 clustering algorithms, including SNF, COCA, IntNMF, PINSPlus, iClusterBayes, NEMO, moCluster, and LRA to integrate genomics, transcriptomics, and epigenomics datasets. The candidate genes obtained through these approaches were further analyzed using prognostic ML models for patient prognosis and screening potential therapeutic targets [176]. Similarly, another study with MIBC with mutation data, DNA copy number, RNA-Seq, and methylation data were analyzed with the iClusterBayes method. With the integrative analysis, it was found that patients with low expression of *MTAP/CDKN2A/2B* had a low response rate to immunotherapy and worse survival, revealing intrinsic MIBC subtypes and prognostic biomarkers [177]. Another study conducted by Shi et al [178], with BC employing whole exome sequencing, RNA sequencing, reduced-representation bisulfite sequencing (RRBS), and oxidative reduced-representation bisulfite sequencing (oxRRBS). The integrative analysis depicted the increased involvement of epigenetic alterations compared to genetic mutations. Similarly, Transcriptomic, proteomic, and protein acetylation sequences were integrated to identify key prognostic genes in BC. A risk model and nomogram incorporating clinical factors were developed, aiding in patient stratification and highlighting the potential of multi-omics and ML in prognosis and therapeutic decision-making [178].

5 Challenges and research directions

While machine learning (ML) has significantly enhanced biomarker identification by improving efficiency, accuracy, and scalability, several challenges remain. Overfitting is one of the most common. Multi-omics datasets, despite being highly dimensional, are often limited to less sample size, leading to the same. In such situations, it is prudent to implement k-fold cross-validation (fivefold or tenfold) during model training to evaluate robustness. For regularization, LASSO (L1 regularization) and Ridge (L2 regularization) are used in penalizing over-complex models [179]. Further, the other major hurdle is the lack of interpretability in ML models, particularly black-box models like deep learning and gradient boosting, which can hinder clinical adoptions. A promising solution involves explainable AI methods such as Shapley Additive exPlanations (SHAP) and Local interpretable Model-agnostic explanation (LIME) that can bridge the interpretability gap. For example, applying SHAP to proteomic ML models revealed that APOE and CD44 drive predictions by regulating EMT pathways, allowing biologically credible validation [180, 181]. Additionally, reproducibility and data quality continue to be a tough hurdle for robust ML models. A 2022 study demonstrated that only 22 of the 74 AI-derived biomarkers in bladder cancer studies were validated across independent cohorts [182]. ML models often overlook pre-analytical variables (such as urine sample handling) and platform-specific biases (eg. SWATH-MS vs. LC-MS/MS) [183]. Similarly, the barrier to clinical integration remains, with less than 5% of ML models being integrated into clinical workflows due to interoperability challenges with electronic health records and regulatory hurdles [184]. To improve generalizability, advanced architectures such as Deep MOIS-MC have been developed to enhance cross-omics correlation, identifying consensus biomarkers across genomics, proteomics, and metabolomics [185]. Additionally, deploying ML models as an interoperable docker container within hospital systems can streamline adoption. A pilot study integrating the DeepCysto AI tool into cystoscopy suites reduced false-positive rates by 22% through real-time contour analysis [180].

Most importantly, the biomarkers obtained from ML applications have to undergo clinical validation to ensure clinical translation. Initially, in the discovery phase, potential biomarkers are identified using ML-based analysis of genomic, proteomic, or metabolomic datasets. Following this, analytical validation is conducted to assess the precision, reproducibility, and specificity of biomarkers in laboratory settings using techniques like Immunohistochemistry, Western Blotting, Quantitative Polymerase Chain Reaction (qPCR), RT-PCR, Enzyme-Linked Immunoassay (ELISA), or Florescence In Situ Hybridization (FISH). Subsequently, clinical validation establishes the biomarker's correlation with clinical outcomes and disease phenotypes. Biomarkers then undergo clinical utility assessment, evaluating their impact on patient outcomes, diagnosis, or prognosis. Lastly, regulatory bodies such as the Food and Drug Administration (FDA) or the European Medicines Agency (EMA) approve these biomarkers with adequate evidence. However, challenges exist, including non-standardized isolation techniques and lack of reproducibility across studies [186, 187]. These challenges can be overcome by establishing standardized protocols, consistency in sample processing, and enhanced data interpretation, and the patient data should be encrypted. Additionally, conducting multi-center studies with diverse cohorts and optimizing workflows will enhance reproducibility and clinical applicability. Future advancements, including deep learning for unveiling complex biological interactions and integration of real-time multi-omics profiling with ML models, may enable dynamic monitoring of treatment response through biopsy platforms, hopefully replacing surveillance cystoscopies within the near future.

6 Conclusion

Bladder cancer poses a significant health burden, particularly in developed countries, which highlights the urgent need for improved diagnostic and prognostic strategies. Utilizing ML algorithms on the multi-omics data, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics, offers a wide range of applications in biomarker identification. ML enables the identification of novel biomarkers that enhance early detection, define risk stratification, predict recurrence, classify cancer subtype, and improve prognostic accuracy. However, challenges such as data integration complexities, model interpretability, and clinical validation must be addressed to ensure the reliability and applicability of these approaches. Future models need to leverage learning frameworks to harmonize datasets while preserving patient privacy, addressing critical gaps in ethnic/geographic diversity that currently limit biomarker generalizability. Advances in explainable AI tools like SHAP and LIME will bridge the interpretability gap, enabling more clinicians to

trust ML-derived biomarkers. As AI-driven methodologies continue to evolve, their seamless integration into clinical workflows could significantly improve patient outcomes, reduce reliance on invasive diagnostics, and advance the era of precision oncology in bladder cancer.

Author contributions B.K., Conceptualization, Writing—original draft, Data curation, Investigation, Visualization. A.N.P.S., Writing—original draft and design, Data curation, Investigation. B.R., Writing—original draft and design, Data curation, Investigation. R.R.S., Writing—original draft and design, Data curation, Investigation. S.M.G., Writing—original draft and design, Data curation, Investigation. S.L.S., Conceptualization, Writing—review & editing, Supervision. V.S., Conceptualization, Writing—review & editing, Supervision, Project administration. All authors reviewed the manuscript.

Funding Open access funding provided by Vellore Institute of Technology.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Nasrallah AA, Williams SB. Epidemiology of bladder cancer. In: Knowles MA, Dyrskjot L, editors. *Biology of bladder cancer: from molecular insights to clinical strategies*. Cham: Springer Nature Switzerland; 2024. p. 3–22.
2. Falabella R, De Simone V, Crocetto F, Del Giudice F, Porreca A, Foschi N, Barone B, Di Gianfrancesco L, Di Pasquale V, Caputo VF. New biomarkers for diagnosis of bladder cancer: a bibliometric analysis. *Arch Ital Urol Androl*. 2025. <https://doi.org/10.4081/aiua.2025.13396>.
3. Key Statistics for Bladder Cancer. <https://www.cancer.org/cancer/types/bladder-cancer/about/key-statistics.html>. Accessed 27 Mar 2025
4. Kumbham S, Md Mahabubur Rahman K, Foster BA, You Y. A comprehensive review of current approaches in bladder cancer treatment. *ACS Pharmacol Transl Sci*. 2025. <https://doi.org/10.1021/acsptsci.4c00663>.
5. Wang G, Jin W, Xu Z, Ju L, Shan D, Li S, Yu M, Cao X, Liu N, Qian K, Zhang Y, Xiao Y, Wang X. Urine-based liquid biopsy in bladder cancer: Opportunities and challenges. *Clin Transl Discov*. 2023;3: e176. <https://doi.org/10.1002/ctd2.176>.
6. Lopez-Beltran A, Cookson MS, Guercio BJ, Cheng L. Advances in diagnosis and treatment of bladder cancer. *BMJ*. 2024;384: e076743. <https://doi.org/10.1136/bmj-2023-076743>.
7. Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol*. 2012;6:140–6. <https://doi.org/10.1016/j.molonc.2012.01.010>.
8. Zhang X, Jonassen I, Goksøyr A. Machine learning approaches for biomarker discovery using gene expression data. In: Helder IN, editor. *Bioinformatics*. Brisbane: Exon Publications; 2021.
9. Al-Tashi Q, Saad MB, Muneer A, Qureshi R, Mirjalili S, Sheshadri A, Le X, Vokes NI, Zhang J, Wu J. Machine learning models for the identification of prognostic and predictive cancer biomarkers: a systematic review. *Int J Mol Sci*. 2023;24:7781. <https://doi.org/10.3390/ijms24097781>.
10. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, Wang L, Lu T, Zhang Y, Sun Z, Han X. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13:816. <https://doi.org/10.1038/s41467-022-28421-6>.
11. Berookhim BM, Sethi AS, Wen CC, Cui J, Liou LS. A Cost comparison of the diagnostic modalities used in the detection of urothelial carcinoma in patients undergoing evaluation for hematuria. *UIJ*. 2008. <https://doi.org/10.3834/uij.1939-4810.2008.12.01>.
12. Sievert KD, Amend B, Nagele U, Schilling D, Bedke J, Horstmann M, Hennenlotter J, Kruck S, Stenzl A. Economic aspects of bladder cancer: what are the benefits and costs? *World J Urol*. 2009;27:295–300. <https://doi.org/10.1007/s00345-009-0395-z>.
13. Abbas S, Shafik R, Soomro N, Heer R, Adhikari K. Reviewing AI's Role in Non-Muscle-Invasive Bladder Cancer Recurrence Prediction. 2024.
14. Ng S, Masarone S, Watson D, Barnes MR. The benefits and pitfalls of machine learning for biomarker discovery. *Cell Tissue Res*. 2023;394:17–31. <https://doi.org/10.1007/s00441-023-03816-z>.
15. Kumar RMR. Exosome- machine learning integration in biomedicine: advancing diagnosis and biomarker discovery. *Curr Med Chem*. 2024;31:1–12. <https://doi.org/10.2174/0109298673319827240812052102>.
16. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Briefings Bioinform*. 2022;23:bbab454. <https://doi.org/10.1093/bib/bbab454>.

17. Schratz P, Muenchow J, Iturrutxa E, Richter J, Brenning A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol Model*. 2019;406:109–20. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>.
18. Wang Z-D, Tian X, Wang Y, Wang J-J, Ye S-Q, Huang Y-Q, Qu Y-Y, Chang K, Shi G-H, Ye D-W, Gu C-Y. The expression and prognostic value of transporter 1, ATP binding cassette subfamily B member in clear cell renal cell cancer with experimental validation. *Front Oncol*. 2022;12:1013790. <https://doi.org/10.3389/fonc.2022.1013790>.
19. Ransam J, Cook JA. LASSO regression. *Br J Surg*. 2018;105:1348. <https://doi.org/10.1002/bjs.10895>.
20. de Vlaming R, Groenen PJ. The current and future use of ridge regression for prediction in quantitative genetics. *BioMed Res Int*. 2015. <https://doi.org/10.1155/2015/143712>.
21. Galal A, Talal M, Moustafa A. Applications of machine learning in metabolomics: disease modeling and classification. *Front Genet*. 2022. <https://doi.org/10.3389/fgene.2022.1017340>.
22. Suthaharan S. Support Vector Machine. In: Suthaharan S, editor. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*. Boston: Springer; 2016. p. 207–35.
23. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom Proteom*. 2018;15:41–51.
24. Somvanshi M, Chavan P, Tambade S, Shinde SV. A review of machine learning techniques using decision tree and support vector machine. In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA). 2016; pp 1–7.
25. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV. A random forest based biomarker discovery and power analysis framework for diagnostics research. *BMC Med Genom*. 2020;13:178. <https://doi.org/10.1186/s12920-020-00826-6>.
26. Zhang S, Li X, Zong M, Zhu X, Cheng D. Learning k for kNN Classification. *ACM Trans Intell Syst Technol*. 2017. <https://doi.org/10.1145/2990508>.
27. Zhou X, Wang H, Xu C, Peng L, Xu F, Lian L, Deng G, Ji S, Hu M, Zhu H, Xu Y. Application of kNN and SVM to predict the prognosis of advanced schistosomiasis. *Parasitol Res*. 2022. <https://doi.org/10.1007/s00436-022-07583-8>.
28. Jagga Z, Gupta D. Machine learning for biomarker identification in cancer research—developments toward its clinical application. *Pers Med*. 2015;12:371–87. <https://doi.org/10.2217/pme.15.5>.
29. Rainio O, Teuhio J, Klén R. Evaluation metrics and statistical tests for machine learning. *Sci Rep*. 2024;14:6086. <https://doi.org/10.1038/s41598-024-56706-x>.
30. Varoquaux G, Colliot O. Evaluating machine learning models and their diagnostic value. In: Colliot O, editor. *machine learning for brain disorders*. New York: Springer; 2023. p. 601–30.
31. Jagga Z, Gupta D. Machine learning for biomarker identification in cancer research—developments toward its clinical application. *Per Med*. 2015;12:371–87. <https://doi.org/10.2217/pme.15.5>.
32. Bouras E, Karhunen V, Gill D, Huang J, Haycock PC, Gunter MJ, Johansson M, Brennan P, Key T, Lewis SJ, Martin RM, Murphy N, Platz EA, Travis R, Yarmolinsky J, Zuber V, Martin P, Katsoulis M, Freisling H, Nøst TH, Schulze MB, Dossus L, Hung RJ, Amos CI, Ahola-Olli A, Palaniswamy S, Männikkö M, Auvinen J, Herzig K-H, Keinänen-Kiukkaanniemi S, Lehtimäki T, Salomaa V, Raitakari O, Salmi M, Jalkanen S, Jarvelin M-R, Dehghan A, Tsilidis KK, The PRACTICAL consortium, Crick, BPC3, Caps, Pegasus. Circulating inflammatory cytokines and risk of five cancers: a Mendelian randomization analysis. *BMC Med*. 2022;20:3. <https://doi.org/10.1186/s12916-021-02193-0>.
33. Li Y, Sundquist K, Zhang N, Wang X, Sundquist J, Memon AA. Mitochondrial related genome-wide Mendelian randomization identifies putatively causal genes for multiple cancer types. *EBioMedicine*. 2023;88:104432. <https://doi.org/10.1016/j.ebiom.2022.104432>.
34. Mateo J, Seed G, Bertan C, Rescigno P, Dolling D, Figueiredo I, Miranda S, Nava Rodrigues D, Gurel B, Clarke M, Atkin M, Chandler R, Messina C, Sumanasuriya S, Bianchini D, Barrero M, Petermolo A, Zafeiriou Z, Fontes M, Perez-Lopez R, Tunariu N, Fulton B, Jones R, McGovern U, Ralph C, Varughese M, Parikh O, Jain S, Elliott T, Sandhu S, Porta N, Hall E, Yuan W, Carreira S, De Bono JS. Genomics of lethal prostate cancer at diagnosis and castration resistance. *J Clin Invest*. 2020;130:1743–51. <https://doi.org/10.1172/JCI132031>.
35. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16:321–32. <https://doi.org/10.1038/nrg3920>.
36. Zhang C, Berndt-Paetz M, Neuhaus J. Bioinformatics analysis identifying key biomarkers in bladder cancer. *Data*. 2020;5:38. <https://doi.org/10.3390/data5020038>.
37. Liosis KC, Marouf AA, Rokne JG, Ghosh S, Bismar TA, Alhaji R. Genomic biomarker discovery in disease progression and therapy response in bladder cancer utilizing machine learning. *Cancers*. 2023;15:4801. <https://doi.org/10.3390/cancers15194801>.
38. Barbour JA, Ou T, Yang H, Fang H, Yue NC, Zhu X, Wong-Brown MW, Wong YT, Bowden NA, Wu S, Wong JWH. ERCC2 mutations alter the genomic distribution pattern of somatic mutations and are independently prognostic in bladder cancer. *Cell Genom*. 2024;4:100627. <https://doi.org/10.1016/j.xgen.2024.100627>.
39. Gil-Jimenez A, Van Dorp J, Contreras-Sanz A, Van Der Vos K, Vis DJ, Braaf L, Broeks A, Kerkhoven R, Van Kessel KEM, Ribal MJ, Alcaraz A, Wessels LFA, Seiler R, Wright JL, Mengual L, Boormans J, Van Rhijn BWG, Black PC, Van Der Heijden MS. Assessment of predictive genomic biomarkers for response to cisplatin-based neoadjuvant chemotherapy in bladder cancer. *Eur Urol*. 2023;83:313–7. <https://doi.org/10.1016/j.eururo.2022.07.023>.
40. Chang Y-C, Wu J-T, Hong M-Y, Tung Y-A, Hsieh P-H, Yee SW, Giacomini KM, Oyang Y-J, Chen C-Y, For the Alzheimer's Disease Neuroimaging Initiative. GenEpi: gene-based epistasis discovery using machine learning. *BMC Bioinform*. 2020;21:68. <https://doi.org/10.1186/s12859-020-3368-2>.
41. Yu EY, Tang Q, Chen Y, Zhang Y, Dai Y, Wu Y, Li W, Mehrkanon S, Wang S, Zeegers MP, Wesselius A. Genome-wide exploration of genetic interactions for bladder cancer risk. *Int J Cancer*. 2024;154:81–93. <https://doi.org/10.1002/ijc.34690>.
42. Li J, Wang Z, Wang T. Machine-learning prediction of a novel diagnostic model using mitochondria-related genes for patients with bladder cancer. *Sci Rep*. 2024;14:9282. <https://doi.org/10.1038/s41598-024-60068-9>.
43. Yu Q, Dong H, Liu S, Li Y, Luo J, Wu X. Identification potential biomarker for bladder cancer using feature selection. 2024; 5: 628
44. Mitchell KA, Williams H. Emerging genomic biomarkers for improving kidney, prostate, and bladder cancer health disparities outcomes. *Urol Oncol Semin Orig Invest*. 2022;40:126–32. <https://doi.org/10.1016/j.urolonc.2019.04.024>.

45. Selim MS, Kassem AB, El-Bassiouny NA, Salahuddin A, Abu El-Ela RY, Hamza MS. Polymorphic renal transporters and cisplatin's toxicity in urinary bladder cancer patients: current perspectives and future directions. *Med Oncol*. 2023. <https://doi.org/10.1007/s12032-022-01928-0>.
46. Wang Y, Gan X, Cheng X, Jia Y, Wang G, Tang X, Du H, Li X, Liu X, Xing X, Ji J, Li Z. ABCC2 induces metabolic vulnerability and cellular ferroptosis via enhanced glutathione efflux in gastric cancer. *Clin Transl Med*. 2024. <https://doi.org/10.1002/ctm2.1754>.
47. Liu Y, Yang R, Zhang M, Yang B, Du Y, Feng H, Wang W, Xue B, Niu F, He P. Multi-omics landscape of Interferon-stimulated gene OASL reveals a potential biomarker in pan-cancer: from prognosis to tumor microenvironment. *Front Immunol*. 2024. <https://doi.org/10.3389/fimmu.2024.1402951>.
48. Zhang Y, Yu C. Prognostic characterization of OAS1/OAS2/OAS3/OASL in breast cancer. *BMC Cancer*. 2020. <https://doi.org/10.1186/s12885-020-07034-6>.
49. Luo W, Wang J, Xu W, Ma C, Wan F, Huang Y, Yao M, Zhang H, Qu Y, Ye D, Zhu Y. LncRNA RP11-89 facilitates tumorigenesis and ferroptosis resistance through PROM2-activated iron export by sponging miR-129-5p in bladder cancer. *Cell Death Dis*. 2021. <https://doi.org/10.1038/s41419-021-04296-1>.
50. Wu K-L. PROM2 upregulation promotes cancer cell migration and confers a poor prognosis in lung squamous cell carcinoma. *Am J Cancer Res*. 2024;14:1561–76. <https://doi.org/10.62347/eqfy1219>.
51. Liu J, Miao X, Yao J, Wan Z, Yang X, Tian W. Investigating the clinical role and prognostic value of genes related to insulin-like growth factor signaling pathway in thyroid cancer. *Aging*. 2024. <https://doi.org/10.18632/aging.205524>.
52. Wang X, Song D, Zhu B, Jin Y, Cai C, Wang Z. Urinary exosomal mRNA as a biomarker for the diagnosis of bladder cancer. *Anticancer Drugs*. 2024. <https://doi.org/10.1097/cad.0000000000001571>.
53. Lanfray D, Caron A, Roy M-C, Laplante M, Morin F, Leprince J, Tonon M-C, Richard D. Involvement of the Acyl-CoA binding domain containing 7 in the control of food intake and energy expenditure in mice. *Elife*. 2016. <https://doi.org/10.7554/elife.11742>.
54. Li C, Weng J, Yang L, Gong H, Liu Z. Development of an anoikis-related gene signature and prognostic model for predicting the tumor microenvironment and response to immunotherapy in colorectal cancer. *Front Immunol*. 2024. <https://doi.org/10.3389/fimmu.2024.1378305>.
55. Maesaka F, Kuwada M, Horii S, Kishi S, Fujiwara-Tani R, Mori S, Fujii K, Mori T, Ohmori H, Owari T, Miyake M, Nakai Y, Tanaka N, Bhawal UK, Luo Y, Kondoh M, Fujimoto K, Kuniyasu H. Hypomethylation of CLDN4 gene promoter is associated with malignant phenotype in urinary bladder cancer. *IJMS*. 2022;23:6516. <https://doi.org/10.3390/ijms23126516>.
56. Hu P, Lei L, Wang Y, Tian X, Wei X, Jiang N, Liu L. CLDN4 as a novel diagnostic and prognostic biomarker and its association with immune infiltrates in ovarian cancer. *Mediators Inflamm*. 2023;2023:1–17. <https://doi.org/10.1155/2023/1075265>.
57. Banerjee S, Mukherjee S, Mitra S, Singhal P. Altered expression of mitochondrial antioxidants in oral squamous cell carcinoma. *J Oral Sci*. 2017;59:439–46. <https://doi.org/10.2334/josnusd.16-0655>.
58. Deng L, Gao X, Liu B, He X, Xu J, Qiang J, Wu Q, Liu S. NMT1 inhibition modulates breast cancer progression through stress-triggered JNK pathway. *Cell Death Dis*. 2018;9:1143. <https://doi.org/10.1038/s41419-018-1201-x>.
59. Zhang Y, Zheng A, Shi Y, Lu H. SNORA5A regulates tumor-associated macrophage M1/M2 phenotypes via TRAF3IP3 in breast cancer. *Braz J Med Biol Res*. 2024;57: e13809. <https://doi.org/10.1590/1414-431x2024e13809>.
60. Adico MDW, Zouré AA, Sombié HK, Kiendrebeogo TI, Dabré S, Amegnona LJ, Bakyono BS, Traoré L, Ouedraogo TC, Ouedraogo RA, Zohoncon TM, Yonli AT, Bayala B, Bambara HA, Djigma FW, Simpore J. Involvement of ERCC1 (rs3212986) and ERCC2 (rs1799793, rs13181) polymorphisms of DNA repair genes in breast cancer occurrence in Burkina Faso. *Molec Gen Gen Med*. 2023;11: e2134. <https://doi.org/10.1002/mggg.2134>.
61. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The genecards suite: from gene data mining to disease genome sequence analyses. *CP Bioinform*. 2016. <https://doi.org/10.1002/cpbi.5>.
62. Zhang J, Chen J, Xu M, Zhu T. Exploring prognostic DNA methylation genes in bladder cancer: a comprehensive analysis. *Discov Onc*. 2024;15:331. <https://doi.org/10.1007/s12672-024-01206-7>.
63. Jiang Y-H, Liu Y-S, Wei Y-C, Jhang J-F, Kuo H-C, Huang H-H, Chan MWY, Lin G-L, Cheng W-C, Lin S-C, Wang H-J. Hypermethylation Loci of ZNF671, IRF8, and OTX1 as potential urine-based predictive biomarkers for bladder cancer. *Diagnostics*. 2024;14:468. <https://doi.org/10.3390/diagnostics14050468>.
64. Köhler CU, Schork K, Turewicz M, Eisenacher M, Roghmann F, Noldus J, Marcus K, Brüning T, Kärferlein HU. Use of Multiple machine learning approaches for selecting urothelial cancer-specific DNA methylation biomarkers in urine. *Int J Mol Sci*. 2024;25:738. <https://doi.org/10.3390/ijms25020738>.
65. van der Heijden AG, Mengual L, Ingelmo-Torres M, Lozano JJ, van de Rijst-van Westerlo CCM, Baixauli M, Geavlete B, Moldoveanu C, Ene C, Dinney CP, Czerniak B, Schalken JA, Kiemeny LALM, Ribal MJ, Witjes JA, Alcaraz A. Urine cell-based DNA methylation classifier for monitoring bladder cancer. *Clin Epigenet*. 2018;10:71. <https://doi.org/10.1186/s13148-018-0496-x>.
66. Mohanad M, Yousef HF, Bahnassy AA. Epigenetic inactivation of DNA repair genes as promising prognostic and predictive biomarkers in urothelial bladder carcinoma patients. *Mol Genet Genomics*. 2022;297:1671–87. <https://doi.org/10.1007/s00438-022-01950-x>.
67. Mijnes J, Veeck J, Gaisa NT, Burghardt E, de Ruijter TC, Gostek S, Dahl E, Pfister D, Schmid SC, Knüchel R, Rose M. Promoter methylation of DNA damage repair (DDR) genes in human tumor entities: RBBP8/CtIP is almost exclusively methylated in bladder cancer. *Clin Epigenetics*. 2018;10:15. <https://doi.org/10.1186/s13148-018-0447-6>.
68. Al-Soodani AT, Wu X, Kelp NC, Brown AJ, Roberts SA, Her C. hMSH5 regulates NHEJ and averts excessive nucleotide alterations at repair joints. *Genes*. 2022;13:673. <https://doi.org/10.3390/genes13040673>.
69. Chatterjee D, Mou SI, Sultana T, Hosen MI, Faruk MO. Identification and validation of prognostic signature genes of bladder cancer by integrating methylation and transcriptomic analysis. *Sci Rep*. 2024;14:368. <https://doi.org/10.1038/s41598-023-50740-x>.
70. Peng D, Ge G, Xu Z, Ma Q, Shi Y, Zhou Y, Gong Y, Xiong G, Zhang C, He S, He Z, Li X, Ci W, Zhou L. Diagnostic and prognostic biomarkers of common urological cancers based on aberrant DNA methylation. *Epigenomics*. 2018;10:1189–99. <https://doi.org/10.2217/epi-2018-0017>.

71. Guo Y, Yin J, Dai Y, Guan Y, Chen P, Chen Y, Huang C, Lu Y-J, Zhang L, Song D. A novel CpG methylation risk indicator for predicting prognosis in bladder cancer. *Front Cell Dev Biol*. 2021. <https://doi.org/10.3389/fcell.2021.642650>.
72. Ling A, Löfgren-Burström A, Larsson P, Li X, Wikberg ML, Öberg Å, Stenling R, Edin S, Palmqvist R. TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncol Immunology*. 2017;6: e1356143. <https://doi.org/10.1080/2162402X.2017.1356143>.
73. Mantel I, Sadiq BA, Blander JM. Spotlight on TAP and its vital role in antigen presentation and cross-presentation. *Mol Immunol*. 2022;142:105–119. <https://doi.org/10.1016/j.molimm.2021.12.013>.
74. Lee H, Kim K, Woo J, Park J, Kim H, Lee KE, Kim H, Kim Y, Moon KC, Kim JY, Park IA, Shim BB, Moon JH, Han D, Ryu HS. Quantitative proteomic analysis identifies AHNAK (Neuroblast Differentiation-associated Protein AHNAK) as a novel candidate biomarker for bladder urothelial carcinoma diagnosis by liquid-based cytology. *Mol Cell Proteom*. 2018;17:1788–802. <https://doi.org/10.1074/mcp.RA118.000562>.
75. Yang I, Son Y, Shin JH, Kim IY, Seong JK. Ahnak depletion accelerates liver regeneration by modulating the TGF- β /Smad signaling pathway. *BMB Rep*. 2022;55:401–6. <https://doi.org/10.5483/BMBRep.2022.55.8.071>.
76. Ss. Cyril and Methodious University, Faculty of Medicine, Institute of Pathology, Skopje, North Macedonia, Komina S, Petrusevska G, Ss. Cyril and Methodious University, Faculty of Medicine, Institute of Pathology, Skopje, North Macedonia, Jovanovic R, Ss. Cyril and Methodious University, Faculty of Medicine, Institute of Pathology, Skopje, North Macedonia, Kostadinova Kunovska S, Ss. Cyril and Methodious University, Faculty of Medicine, Institute of Pathology, Skopje, North Macedonia, Stavridis S, Ss. Cyril and Methodious University, Faculty of Medicine, University Urology Clinic, Skopje, North Macedonia, Dohcev S, Ss. Cyril and Methodious University, Faculty of Medicine, University Urology Clinic, Skopje, North Macedonia, Saidi S, Ss. Cyril and Methodious University, Faculty of Medicine, University Urology Clinic, Skopje, North Macedonia, Topuzovska S, Ss. Cyril and Methodious University, Faculty of Medicine, Institute of Medical and Experimental Biochemistry, Skopje, North Macedonia. AHNAK2 urinary protein expression as potential biomarker for bladder cancer detection: a pilot study. *Turk J Urol*. 2022;48:423–30. <https://doi.org/10.5152/tud.2022.22132>.
77. Abu-Zhayia ER, Khoury-Haddad H, Guttman-Raviv N, Serruya R, Jarrous N, Ayoub N. A role of human RNase P subunits, Rpp29 and Rpp21, in homology directed-repair of double-strand breaks. *Sci Rep*. 2017;7:1002. <https://doi.org/10.1038/s41598-017-01185-6>.
78. Shaukat A-N, Kaliatsi EG, Skeparnias I, Stathopoulos C. The dynamic network of RNP RNase P subunits. *IJMS*. 2021;22:10307. <https://doi.org/10.3390/ijms221910307>.
79. Tsai W-L, Cheng J-S, Liu P-F, Chang T-H, Sun W-C, Chen W-C, Shu C-W. Sofosbuvir induces gene expression for promoting cell proliferation and migration of hepatocellular carcinoma cells. *Aging*. 2022;14:5710–26. <https://doi.org/10.18631/aging.204170>.
80. Zhang J, Luo J, Jiang H, Xie T, Zheng J, Tian Y, Li R, Wang B, Lin J, Xu A, Huang X, Yuan Y. The tumor suppressor role of zinc finger protein 671 (ZNF671) in multiple tumors based on cancer single-cell sequencing. *Front Oncol*. 2019;9:1214. <https://doi.org/10.3389/fonc.2019.01214>.
81. Zhu P, Xiong J, Yuan D, Li X, Luo L, Huang J, Wang B, Nie Q, Wang S, Dang L, Chen Y, Li S, An Y, Yang L, Liu R, He Y, Li X, Chen P, Zhou H, Li Q. ZNF671 methylation test in cervical scrapings for cervical intraepithelial neoplasia grade 3 and cervical cancer detection. *Cell Rep Med*. 2023;4:101143. <https://doi.org/10.1016/j.xcrm.2023.101143>.
82. Tu X-P, Li H, Chen L-S, Luo X-N, Lu Z-M, Zhang S-Y, Chen S-H. OTX1 exerts an oncogenic role and is negatively regulated by miR129-5p in laryngeal squamous cell carcinoma. *BMC Cancer*. 2020;20:794. <https://doi.org/10.1186/s12885-020-07279-1>.
83. Zhang Q, Li X, Ren L, Gu X, Xiao N, Li N. OTX1 silencing suppresses ovarian cancer progression through inhibiting the JAK/STAT signaling. *Tissue Cell*. 2023;82:102082. <https://doi.org/10.1016/j.tice.2023.102082>.
84. Moorman HR, Reategui Y, Poschel DB, Liu K. IRF8: mechanism of action and health implications. *Cells*. 2022;11:2630. <https://doi.org/10.3390/cells11172630>.
85. Tse BCY, Bergamin S, Steffen P, Hruby G, Pavlakis N, Clarke SJ, Evans J, Engel A, Kneebone A, Molloy MP. CD11c+ and IRF8+ cell densities in rectal cancer biopsies predict outcomes of neoadjuvant chemoradiotherapy. *Oncol Immunology*. 2023;12:2238506. <https://doi.org/10.1080/2162402X.2023.2238506>.
86. Halade GV, Kain V, Hossain S, Parcha V, Limdi NA, Arora P. Arachidonate 5-lipoxygenase is essential for biosynthesis of specialized pro-resolving mediators and cardiac repair in heart failure. *Am J Physiol Heart Circulatory Physiol*. 2022;323:H721–37. <https://doi.org/10.1152/ajpheart.00115.2022>.
87. Mongioli JM, Hong C-C, Zircoli GR, Khoury T, Omilian AR, Qin B, Bandera EV, Yao S, Ambrosone CB, Gong Z. Genetic variants in COX2 and ALOX genes and breast cancer risk in white and black women. *Front Oncol*. 2021;11:679998. <https://doi.org/10.3389/fonc.2021.679998>.
88. Ai D, Yao J, Yang F, Huo L, Chen H, Lu W, Soto LMS, Jiang M, Raso MG, Wang S, Bell D, Liu J, Wang H, Tan D, Torres-Cabala C, Gan Q, Wu Y, Albarracín C, Hung M-C, Meric-Bernstam F, Wistuba II, Prieto VG, Sahin AA, Ding Q. TRPS1: a highly sensitive and specific marker for breast carcinoma, especially for triple-negative breast cancer. *Mod Pathol*. 2021;34:710–9. <https://doi.org/10.1038/s41379-020-00692-8>.
89. Liou TG. The clinical biology of cystic fibrosis transmembrane regulator protein: its role and function in extrapulmonary disease. *Chest*. 2018;155:605. <https://doi.org/10.1016/j.chest.2018.10.006>.
90. Scott P, Anderson K, Singhanian M, Cormier R. Cystic fibrosis, CFTR, and colorectal cancer. *IJMS*. 2020;21:2891. <https://doi.org/10.3390/ijms21082891>.
91. Kuroda T, Yasuda S, Tachi S, Matsuyama S, Kusakawa S, Tano K, Miura T, Matsuyama A, Sato Y. SALL3 expression balance underlies lineage biases in human induced pluripotent stem cell differentiation. *Nat Commun*. 2019;10:2175. <https://doi.org/10.1038/s41467-019-09511-4>.
92. Zhu H, Xie Z. Therapeutic potential of tLyp-1-EV-shCTCF in inhibiting liver cancer stem cell self-renewal and immune escape via SALL3 modulation in hepatocellular carcinoma. *Transl Oncol*. 2024;49:102048. <https://doi.org/10.1016/j.tranon.2024.102048>.
93. Zhu Q-Q, Ma C, Wang Q, Song Y, Lv T. The role of TWIST1 in epithelial-mesenchymal transition and cancers. *Tumor Biol*. 2016;37:185–97. <https://doi.org/10.1007/s13277-015-4450-7>.
94. Xu J, Guo R, Wen N, Li L, Yi Y, Chen J, He Z, Yang J, Xiao Z-XJ, Niu M. FBXO3 stabilizes USP4 and Twist1 to promote PI3K-mediated breast cancer metastasis. *PLoS Biol*. 2023;21: e3002446. <https://doi.org/10.1371/journal.pbio.3002446>.
95. Zhou Z, Liu Y, Gao S, Zhou M, Qi F, Ding N, Zhang J, Li R, Wang J, Shi J, Yu R, Wang Y, Li Y, Pan J, Du J, Wang D. Excessive DNA damage mediates ECM degradation via the RBBP8/NOTCH1 pathway in sporadic aortic dissection. *Biochimica et Biophysica Acta (BBA)—Molecular Basis of Disease*. 2022;1868:166303. <https://doi.org/10.1016/j.bbadis.2021.166303>.

96. Yu Y, Wang S, Yin Y, Wang G. Mechanism of RBBP8-mediated homologous recombination repair in gastric cancer synthetic lethal. *Chron Dis Transl Med*. 2023;9:250–7. <https://doi.org/10.1002/cdt3.75>.
97. Gonzalo A, Lucas M-O, Charpentier C, Sandmann G, Lloyd A, Jenczewski E. Reducing MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. *Nat Commun*. 2019;10:2354. <https://doi.org/10.1038/s41467-019-10010-9>.
98. Chang Y, Lee C, Ke T, Chang C, Chao D, Huang H, Chang J. Molecular characterization of colorectal cancer using whole-exome sequencing in a Taiwanese population. *Cancer Med*. 2019;8:3738–47. <https://doi.org/10.1002/cam4.2282>.
99. Qi Y. Receptor tyrosine kinases in breast cancer treatment: unraveling the potential. *Am J Cancer Res*. 2024;14:4172–96. <https://doi.org/10.62347/KIVS3169>.
100. Zhang X, Zhang Y, Tang H, He J. EGFR gene copy number as a predictive/biomarker for patients with non-small-cell lung cancer receiving tyrosine kinase inhibitor treatment: a systematic review and meta-analysis. *J Investig Med*. 2017;65:72–81. <https://doi.org/10.1136/jim-2016-000252>.
101. He F, Ru X, Wen T. NRF2, a transcription factor for stress response and beyond. *Int J Mol Sci*. 2020;21:4777. <https://doi.org/10.3390/ijms21134777>.
102. Xu W, Liu J, Liu Q, Xu J, Zhou L, Liang Z, Huang H, Huang B, Xiao GG, Guo J. NFE2-driven neutrophil polarization promotes pancreatic cancer liver metastasis progression. *Cell Rep*. 2025;44:115226. <https://doi.org/10.1016/j.celrep.2024.115226>.
103. Lin S-J, Huang C-F, Wu T-S, Li C-C, Lee F-JS. Arl4D-EB1 interaction promotes centrosomal recruitment of EB1 and microtubule growth. *Mol Biol Cell*. 2020;31:2348. <https://doi.org/10.1091/mbc.E18-10-0611>.
104. Peng Y, Wu Q, Wang L, Wang H, Yin F. A DNA methylation signature to improve survival prediction of gastric cancer. *Clin Epigenet*. 2020;12:15. <https://doi.org/10.1186/s13148-020-0807-x>.
105. Gong T-T, Liu F-H, Xiao Q, Li Y-Z, Wei Y-F, Xu H-L, Cao F, Sun M-L, Jiang F-L, Tao T, Ma Q-P, Qin X, Song Y, Gao S, Wu L, Zhao Y-H, Huang D-H, Wu Q-J. SH3RF2 contributes to cisplatin resistance in ovarian cancer cells by promoting RBPMS degradation. *Commun Biol*. 2024;7:67. <https://doi.org/10.1038/s42003-023-05721-1>.
106. Boncheva V, Linnebacher M, Kdimati S, Draper H, Orchard L, Mills K, O'Sullivan G, Tangney M, Guinn B. Identification of the antigens recognised by colorectal cancer patients using sera from patients who exhibit a Crohn's-like lymphoid reaction. *Biomolecules*. 2022;12:1058. <https://doi.org/10.3390/biom12081058>.
107. Vieira AF, Paredes J. P-cadherin and the journey to cancer metastasis. *Mol Cancer*. 2015;14:178. <https://doi.org/10.1186/s12943-015-0448-4>.
108. Song J, Jin Y, Fan S, Wei Y, Dong D, Jia L, Fan S, Zhang A, Zhou W, Jiang W, Ren L. CDH3 is an effective serum biomarker of colorectal cancer distant metastasis patients. *J Cancer*. 2024;15:5218–29. <https://doi.org/10.7150/jca.98337>.
109. Yuan Z, Zhang B, Zhuang J, Yang X, Yu J, Xu Q, Azam N, Ur Rahman K, Feng B. Chlorotoxin fusion protein regulates miR-374a and TNFAIP8 expression and inhibits glioma cell proliferation and promotes apoptosis. *Cytotechnology*. 2020;72:685–94. <https://doi.org/10.1007/s10616-020-00411-w>.
110. Lan G, Yu X, Sun X, Li W, Zhao Y, Lan J, Wu X, Gao R. Comprehensive analysis of the expression and prognosis for TNFAIPs in head and neck cancer. *Sci Rep*. 2021;11:15696. <https://doi.org/10.1038/s41598-021-95160-x>.
111. Breschi A, Gingeras TR, Guigó R. Comparative transcriptomics in human and mouse. *Nat Rev Genet*. 2017;18:425–40. <https://doi.org/10.1038/nrg.2017.19>.
112. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLOS Comput Biol*. 2017;13: e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>.
113. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019;20:631–56. <https://doi.org/10.1038/s41576-019-0150-2>.
114. Pepe P, Salemi M, Marchese G, Salluzzo MG, Lanza G, Marino S, Schillaci F, Truda A, Pepe L, Pennisi M. Transcriptome analysis in patients with muscle-invasive bladder cancer. *In Vivo*. 2024. <https://doi.org/10.21873/invivo.13615>.
115. Giulietti M, Occhipinti G, Righetti A, Bracci M, Conti A, Ruzzo A, Cerigioni E, Cacciamani T, Principato G, Piva F. Emerging biomarkers in bladder cancer identified by network analysis of transcriptomic data. *Front Oncol*. 2018. <https://doi.org/10.3389/fonc.2018.00450>.
116. Goel A, Ward DG, Noyvert B, Yu M, Gordon NS, Abbotts B, Colbourne JK, Kissane S, James ND, Zeegers MP, Cheng KK, Cazier J-B, Whalley CM, Beggs AD, Palles C, Arnold R, Bryan RT. Combined exome and transcriptome sequencing of non-muscle-invasive bladder cancer: associations between genomic changes, expression subtypes, and clinical outcomes. *Genome Med*. 2022;14:1–16. <https://doi.org/10.1186/s13073-022-01056-4>.
117. Wang J, He X, Bai Y, Du G, Cai M. Identification and validation of novel biomarkers affecting bladder cancer immunotherapy via machine learning and its association with M2 macrophages. *Front Immunol*. 2022. <https://doi.org/10.3389/fimmu.2022.1051063>.
118. Chen Z, Ou Y, Ye F, Li W, Jiang H, Liu S. Machine learning identifies the role of SMAD6 in the prognosis and drug susceptibility in bladder cancer. *J Cancer Res Clin Oncol*. 2024;150:1–17. <https://doi.org/10.1007/s00432-024-05798-z>.
119. Zhang J, Huang Y, Tan X, Wang Z, Cheng R, Zhang S, Chen Y, Jiang F, Tan W, Deng X, Li F. Integrated analysis of multiple transcriptomic approaches and machine learning integration algorithms reveals high endothelial venules as a prognostic immune-related biomarker in bladder cancer. *Int Immunopharmacol*. 2024;136:112184. <https://doi.org/10.1016/j.intimp.2024.112184>.
120. Song D, Shen T, Feng K, He Y, Chen S, Zhang Y, Luo W, Han L, Tong M, Jin Y. LIG1 is a novel marker for bladder cancer prognosis: evidence based on experimental studies, machine learning and single-cell sequencing. *Front Immunol*. 2024. <https://doi.org/10.3389/fimmu.2024.1419126>.
121. Shen D, Kang S. Comprehensive analysis of mitochondria-related genes indicates that PPP2R2B is a novel biomarker and promotes the progression of bladder cancer via Wnt signaling pathway. *Biol Direct*. 2024;19:17. <https://doi.org/10.1186/s13062-024-00461-6>.
122. Li J, Kong Z, Qi Y, Wang W, Su Q, Huang W, Zhang Z, Li S, Du E. Single-cell and bulk RNA-sequence identified fibroblasts. *Int J Surg*. 2024;110:4911.
123. Cho H, Tong F, You S, Jung S, Kim WH, Kim J. Prediction of the immune phenotypes of bladder cancer patients for precision oncology. *IEEE Open J Eng Med Biol*. 2022;3:47–57. <https://doi.org/10.1109/OJEMB.2022.3163533>.
124. Edwards CM, Johnson RW. From good to bad: the opposing effects of PTHrP on tumor growth, dormancy, and metastasis throughout cancer progression. *Front Oncol*. 2021. <https://doi.org/10.3389/fonc.2021.644303>.

125. Lyu G, Dai L, Deng X, Liu X, Guo Y, Zhang Y, Wang X, Huang Y, Wu S, Guo J, Liu Y. Integrative analysis of cuproptosis-related mitochondrial depolarisation genes for prognostic prediction in non-small cell lung cancer. *J Cell Mol Med*. 2025;29: e70438. <https://doi.org/10.1111/jcmm.70438>.
126. Wang Q, Xiong F, Wu G, Wang D, Liu W, Chen J, Qi Y, Wang B, Chen Y. SMAD proteins in TGF- β signalling pathway in cancer: regulatory mechanisms and clinical applications. *Diagnostics*. 2023;13:2769. <https://doi.org/10.3390/diagnostics13172769>.
127. Bayat Z, Ghaemi Z, Behmanesh M, Soltani BM. Hsa-miR-186-5p regulates TGF β signaling pathway through expression suppression of SMAD6 and SMAD7 genes in colorectal cancer. *Biol Chem*. 2021;402:469–80. <https://doi.org/10.1515/hsz-2019-0407>.
128. Hussein NA, El Sewedy SM, Zakareya MM, Youssef EA, Ibrahim FAR. Expression status of circ-SMARCA5, circ-NOL10, circ-LDLRAD3, and circ-RHOT1 in patients with colorectal cancer. *Sci Rep*. 2023;13:13308. <https://doi.org/10.1038/s41598-023-40358-4>.
129. Sun L, Liu X, Song S, Feng L, Shi C, Sun Z, Chen B, Hou H. Identification of LIG1 and LIG3 as prognostic biomarkers in breast cancer. *Open Medicine*. 2021;16:1705–17. <https://doi.org/10.1515/med-2021-0388>.
130. Bao Y, Oguz G, Lee WC, Lee PL, Ghosh K, Li J, Wang P, Lobie PE, Ehmsen S, Ditzel HJ, Wong A, Tan EY, Lee SC, Yu Q. EZH2-mediated PP2A inactivation confers resistance to HER2-targeted breast cancer therapy. *Nat Commun*. 2020;11:5878. <https://doi.org/10.1038/s41467-020-19704-x>.
131. Pan H, Luo Z, Lin F, Zhang J, Xiong T, Hong Y, Sun B, Yang Y. FN1, a reliable prognostic biomarker for thyroid cancer, is associated with tumor immunity and an unfavorable prognosis. *Oncol Lett*. 2024;28:510. <https://doi.org/10.3892/ol.2024.14643>.
132. Deng J, Chen X, Zhan T, Chen M, Yan X, Huang X. CRYAB predicts clinical prognosis and is associated with immunocyte infiltration in colorectal cancer. *PeerJ*. 2021;9: e12578. <https://doi.org/10.7717/peerj.12578>.
133. Yao F, Zhu ZF, Wen J, Zhang FY, Zhang Z, Zhu LQ, Su GH, Yuan QW, Zhen YF, Wang XD. PODN is a prognostic biomarker and correlated with immune infiltrates in osteosarcoma. *Cancer Cell Int*. 2021;21:381. <https://doi.org/10.1186/s12935-021-02086-5>.
134. Gong R, Wang J, Xing Y, Wang J, Chen X, Lei K, Yu Q, Zhao C, Li S, Zhang Y, Wang H, Ren H. Expression landscape of cancer-FOXP3 and its prognostic value in pancreatic adenocarcinoma. *Cancer Lett*. 2024;590:216838. <https://doi.org/10.1016/j.canlet.2024.216838>.
135. Qiu S, Lin W, Zhou Z, Hong Q, Chen S, Li J, Zhong F, Zhou Q, Cui D. TOX: a potential new immune checkpoint in cancers by pancancer analysis. *Discov Onc*. 2024;15:354. <https://doi.org/10.1007/s12672-024-01236-1>.
136. Guo L, Li X, Liu R, Chen Y, Ren C, Du S. TOX correlates with prognosis, immune infiltration, and T cells exhaustion in lung adenocarcinoma. *Cancer Med*. 2020;9:6694–709. <https://doi.org/10.1002/cam4.3324>.
137. Lee N, Canagasingham A, Bajaj M, Shanmugasundaram R, Hutton A, Buccì J, Graham P, Thompson J, Ni J. Urine exosomes as biomarkers in bladder cancer diagnosis and prognosis: From functional roles to clinical significance. *Front Oncol*. 2022;12:1019391. <https://doi.org/10.3389/fonc.2022.1019391>.
138. Frantzi M, van Kessel KE, Zwarthoff EC, Marquez M, Rava M, Malats N, Merseburger AS, Katafigiotis I, Stravodimos K, Mullen W, Zoidakis J, Makridakis M, Pejchinovski M, Critselis E, Lichtinghagen R, Brand K, Dakna M, Roubelakis MG, Theodorescu D, Vlahou A, Mischak H, Anagnou NP. Development and validation of urine-based peptide biomarker panels for detecting bladder cancer in a multi-center study. *Clin Cancer Res*. 2016;22:4077–86. <https://doi.org/10.1158/1078-0432.CCR-15-2715>.
139. Liu Y-R, Ortiz-Bonilla CJ, Lee Y-F. Extracellular vesicles in bladder cancer: biomarkers and beyond. *Int J Mol Sci*. 2018;19:2822. <https://doi.org/10.3390/ijms19092822>.
140. Vanarsa K, Castillo J, Wang L, Lee KH, Pedroza C, Lotan Y, Mohan C. Comprehensive proteomics and platform validation of urinary biomarkers for bladder cancer diagnosis and staging. *BMC Med*. 2023;21:133. <https://doi.org/10.1186/s12916-023-02813-x>.
141. Suh J, Han D, Ku JH, Kim HH, Kwak C, Jeong CW. Next-generation proteomics-based discovery, verification, and validation of urine biomarkers for bladder cancer diagnosis. *Cancer Res Treat*. 2022;54:882–93. <https://doi.org/10.4143/crt.2021.642>.
142. Necula L, Matei L, Dragu D, Pitica I, Neagu A, Bleotu C, Diaconu CC, Chivu-Economescu M. Collagen family as promising biomarkers and therapeutic targets in cancer. *Int J Mol Sci*. 2022;23:12415. <https://doi.org/10.3390/ijms232012415>.
143. Schiffer E, Vlahou A, Petrolekas A, Stravodimos K, Tauber R, Geschwend JE, Neuhaus J, Stolzenburg J-U, Conway MR, Mischak H, Theodorescu D. Prediction of muscle-invasive bladder cancer using urinary proteomics. *Clin Cancer Res*. 2009;15:4935–43. <https://doi.org/10.1158/1078-0432.CCR-09-0226>.
144. Chang Q, Chen Y, Yin J, Wang T, Dai Y, Wu Z, Guo Y, Wang L, Zhao Y, Yuan H, Song D, Zhang L. Comprehensive urinary proteome profiling analysis identifies diagnosis and relapse surveillance biomarkers for bladder cancer. *J Proteome Res*. 2024;23:2241–52. <https://doi.org/10.1021/acs.jproteome.4c00199>.
145. Georgila K, Vyrla D, Drakos E. Apolipoprotein A-I (ApoA-I), Immunity, inflammation and cancer. *Cancers*. 2019;11:1097. <https://doi.org/10.3390/cancers11081097>.
146. Xu X, Cai H, Peng J, Liu H, Chu F. HBB as a novel biomarker for the diagnosis and monitoring of lung cancer regulates cell proliferation via ERK1/2 pathway. *Technol Cancer Res Treat*. 2024;23:15330338241249032. <https://doi.org/10.1177/15330338241249032>.
147. Raungrut P, Jirapongsak J, Tanyapattapong S, Bunsong T, Ruklert T, Kueakool K, Thongsuksai P, Nakwan N. Fibrinogen alpha chain as a potential serum biomarker for predicting response to cisplatin and gemcitabine doublet chemotherapy in lung adenocarcinoma: integrative transcriptome and proteome analyses. *Int J Mol Sci*. 2025;26:1010. <https://doi.org/10.3390/ijms26031010>.
148. Roy R, Yang J, Moses MA. Matrix metalloproteinases as novel biomarkers and potential therapeutic targets in human cancer. *J Clin Oncol*. 2009;27:5287–97. <https://doi.org/10.1200/JCO.2009.23.5556>.
149. Michelis L, Milhem L, Galouk E, Stemer G, Aviv A, Tadmor T, Shehadeh M, Shvidel L, Barhoum M, Braester A. Increased serum level of alpha-2 macroglobulin and its production by B-lymphocytes in chronic lymphocytic leukemia. *Front Immunol*. 2022;13:953644. <https://doi.org/10.3389/fimmu.2022.953644>.
150. De Souza Dutra C, Da Cruz Schafhauser D, Hentz M, Mayer NR, Pinheiro RM, Baierle G, Kist DR, Bullé DJ, Donaduzzi RC, Bohmgahren MF, Zaha A. Urinary endogenous peptides as biomarkers for prostate cancer. *Oncol Lett*. 2023. <https://doi.org/10.3892/ol.2023.13759>.
151. Schaeffer C, Devuyst O, Rampoldi L. Uromodulin: roles in health and disease. *Annu Rev Physiol*. 2021;83:477–501. <https://doi.org/10.1146/annurev-physiol-031620-092817>.
152. Zhang XF, Zhang XL, Guo L, Bai YP, Tian Y, Luo HY. The function of the inter-alpha-trypsin inhibitors in the development of disease. *Front Med*. 2024. <https://doi.org/10.3389/fmed.2024.1432224>.

153. Ajona D, Cragg MS, Pio R. The complement system in clinical oncology: applications, limitations and challenges. *Semin Immunol.* 2025;77:101921. <https://doi.org/10.1016/j.smim.2024.101921>.
154. Li C-H, Chan M-H, Liang S-M, Chang Y-C, Hsiao M. Fascin-1: updated biological functions and therapeutic implications in cancer biology. *BBA Adv.* 2022;2:100052. <https://doi.org/10.1016/j.bbadv.2022.100052>.
155. Lee S, Jue M, Lee K, Paulson B, Oh J, Cho M, Kim JK. Early-stage diagnosis of bladder cancer using surface-enhanced Raman spectroscopy combined with machine learning algorithms in a rat model. *Biosens Bioelectron.* 2024;246:115915. <https://doi.org/10.1016/j.bios.2023.115915>.
156. Shao C-H, Chen C-L, Lin J-Y, Chen C-J, Fu S-H, Chen Y-T, Chang Y-S, Yu J-S, Tsui K-H, Juo C-G, Wu K-P. Metabolite marker discovery for the detection of bladder cancer by comparative metabolomics. *Oncotarget.* 2017;8:38802–10. <https://doi.org/10.18632/oncotarget.16393>.
157. Wang R, Kang H, Zhang X, Nie Q, Wang H, Wang C, Zhou S. Urinary metabolomics for discovering metabolic biomarkers of bladder cancer by UPLC-MS. *BMC Cancer.* 2022;22:214. <https://doi.org/10.1186/s12885-022-09318-5>.
158. Feldner-Busztin D, Firas Nisantzis P, Edmunds SJ, Boza G, Racimo F, Gopalakrishnan S, Limborg MT, Lahti L, De Polavieja GG. Dealing with dimensionality: the application of machine learning to multi-omics data. *Bioinformatics.* 2023;39:btad021. <https://doi.org/10.1093/bioinformatics/btad021>.
159. Troisi J, Colucci A, Cavallo P, Richards S, Symes S, Landolfi A, Scala G, Maiorino F, Califano A, Fabiano M, Silvestre G, Mastella F, Caputo A, D'Antonio A, Altieri V. A serum metabolomic signature for the detection and grading of bladder cancer. *Appl Sci.* 2021;11:2835. <https://doi.org/10.3390/app11062835>.
160. Kühn T, Stepien M, López-Noguerol M, Damms-Machado A, Sookthai D, Johnson T, Roca M, Hüsing A, Maldonado SG, Cross AJ, Murphy N, Freisling H, Rinaldi S, Scalbert A, Fedirko V, Severi G, Boutron-Ruault M-C, Mancini FR, Sowah SA, Boeing H, Jakszyn P, Sánchez MJ, Merino S, Colorado-Yohar S, Barricarte A, Khaw KT, Schmidt JA, Perez-Cornago A, Trichopoulou A, Karakatsani A, Thriskos P, Palli D, Agnoli C, Tumino R, Sacerdote C, Panico S, Bueno-de-Mesquita B, van Gils CH, Heath AK, Gunter MJ, Riboli E, Lahoz A, Jenab M, Kaaks R. Prediagnostic plasma bile acid levels and colon cancer risk: a prospective study. *J Natl Cancer Inst.* 2020;112:516–24. <https://doi.org/10.1093/jnci/djz166>.
161. Liu J, Wei Y, Jia W, Can C, Wang R, Yang X, Gu C, Liu F, Ji C, Ma D. Chenodeoxycholic acid suppresses AML progression through promoting lipid peroxidation via ROS/p38 MAPK/DGAT1 pathway and inhibiting M2 macrophage polarization. *Redox Biol.* 2022;56:102452. <https://doi.org/10.1016/j.redox.2022.102452>.
162. Colosimo S, Tomlinson JW. Bile acids as drivers and biomarkers of hepatocellular carcinoma. *World J Hepatol.* 2022;14:1730–8. <https://doi.org/10.4254/wjh.v14.i9.1730>.
163. Al-Bakheit A, Traka M, Saha S, Mithen R, Melchini A. Accumulation of palmitoylcarnitine and its effect on pro-inflammatory pathways and calcium influx in prostate cancer. *Prostate.* 2016;76:1326–37. <https://doi.org/10.1002/pros.23222>.
164. Jia H, Shen X, Guan Y, Xu M, Tu J, Mo M, Xie L, Yuan J, Zhang Z, Cai S, Zhu J, Zhu Z. Predicting the pathological response to neoadjuvant chemoradiation using untargeted metabolomics in locally advanced rectal cancer. *Radiother Oncol.* 2018;128:548–56. <https://doi.org/10.1016/j.radonc.2018.06.022>.
165. El Agouza IM, Eissa SS, El Houseini MM, El-Nashar DE, Abd El Hameed OM. Taurine: a novel tumor marker for enhanced detection of breast cancer among female patients. *Angiogenesis.* 2011;14:321–30. <https://doi.org/10.1007/s10456-011-9215-3>.
166. Liang T-L, Pan H-D, Yan P-Y, Mi J-N, Liu X-C, Bao W-Q, Lian L-R, Zhang C-F, Chen Y, Wang J-R, Xie Y, Zhou H, Yao X-J, Graham P, Leung EL-H, Liu L, Li R-Z. Serum taurine affects lung cancer progression by regulating tumor immune escape mediated by the immune microenvironment. *J Adv Res.* 2024;S2090–1232(24):00389–98. <https://doi.org/10.1016/j.jare.2024.09.005>.
167. Mitraka M, Gore CR, Kumar A, Sarode SC, Sharma NK. Undetectable free aromatic amino acids in nails of breast carcinoma: biomarker discovery by a novel metabolite purification VTGE system. *Front Oncol.* 2020;10:908. <https://doi.org/10.3389/fonc.2020.00908>.
168. Wang W, Wang H, Wang Q, Yu X, Ouyang L. Lactate-induced protein lactylation in cancer: functions, biomarkers and immunotherapy strategies. *Front Immunol.* 2025. <https://doi.org/10.3389/fimmu.2024.1513047>.
169. Tang Q, Wu S, Zhao B, Li Z, Zhou Q, Yu Y, Yang X, Wang R, Wang X, Wu W, Wang S. Reprogramming of glucose metabolism: the hallmark of malignant transformation and target for advanced diagnostics and treatments. *Biomed Pharmacother.* 2024;178:117257. <https://doi.org/10.1016/j.biopha.2024.117257>.
170. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights.* 2020;14:1177932219899051. <https://doi.org/10.1177/1177932219899051>.
171. Molla G, Bitew M. Revolutionizing personalized medicine: synergy with multi-omics data generation, main hurdles, and future perspectives. *Biomedicines.* 2024. <https://doi.org/10.3390/biomedicines12122750>.
172. Arjmand B, Hamidpour SK, Tayanloo-Beik A, Goodarzi P, Aghayan HR, Adibi H, Larijani B. Machine learning: a new prospect in multi-omics data analysis of cancer. *Front Genet.* 2022. <https://doi.org/10.3389/fgene.2022.824451>.
173. Zhang X, Wang J, Lu J, Su L, Wang C, Huang Y, Zhang X, Zhu X. Robust prognostic subtyping of muscle-invasive bladder cancer revealed by deep learning-based multi-omics data integration. *Front Oncol.* 2021. <https://doi.org/10.3389/fonc.2021.689626>.
174. Picard M, Scott-Boyer M-P, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J.* 2021;19:3735–46. <https://doi.org/10.1016/j.csbj.2021.06.030>.
175. (2025) Multi-Omics Technology in Human Health and Diseases
176. Chu G, Ji X, Wang Y, Niu H. Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. *Mol Ther Nucleic Acids.* 2023;33:110–26. <https://doi.org/10.1016/j.omtn.2023.06.001>.
177. Mo Q, Li R, Adeegbe DO, Peng G, Chan KS. Integrative multi-omics analysis of muscle-invasive bladder cancer identifies prognostic biomarkers for frontline chemotherapy and immunotherapy. *Commun Biol.* 2020;3:784.
178. Shi Z-D, Han X-X, Song Z-J, Dong Y, Pang K, Wang X-L, Liu X-Y, Lu H, Xu G-Z, Hao L, Dong B-Z, Liang Q, Wu X-K, Han C-H. Integrative multi-omics analysis depicts the methylome and hydroxymethylome in recurrent bladder cancers and identifies biomarkers for predicting PD-L1 expression. *Biomark Res.* 2023;11:47. <https://doi.org/10.1186/s40364-023-00488-3>.
179. Ali N, Bolenz C, Todenhöfer T, Stenzel A, Deetmar P, Kriegmair M, Knoll T, Porubsky S, Hartmann A, Popp J, Kriegmair MC, Bocklitz T. Deep learning-based classification of blue light cystoscopy imaging during transurethral resection of bladder tumors. *Sci Rep.* 2021;11:11629. <https://doi.org/10.1038/s41598-021-91081-x>.

180. Yoo JW, Koo KC, Chung BH, Baek SY, Lee SJ, Park KH, Lee KS. Deep learning diagnostics for bladder tumor identification and grade prediction using RGB method. *Sci Rep*. 2022;12:17699. <https://doi.org/10.1038/s41598-022-22797-7>.
181. Khoraminia F, Fuster S, Kanwal N, Olislagers M, Engan K, Van Leenders GJLH, Stubbs AP, Akram F, Zuiverloon TCM. Artificial intelligence in digital pathology for bladder cancer: hype or hope? A systematic review. *Cancers*. 2023;15:4518. <https://doi.org/10.3390/cancers15184518>.
182. Roper K, Abdel-Rehim A, Hubbard S, Carpenter M, Rzhetsky A, Soldatova L, King RD. Testing the reproducibility and robustness of the cancer biology literature by robot. *J R Soc Interface*. 2022;19:20210821. <https://doi.org/10.1098/rsif.2021.0821>.
183. Mattsson-Carlsson N, Palmqvist S, Blennow K, Hansson O. Increasing the reproducibility of fluid biomarker studies in neurodegenerative studies. *Nat Commun*. 2020;11:6252. <https://doi.org/10.1038/s41467-020-19957-6>.
184. Abbas S, Shafik R, Soomro N, Heer R, Adhikari K. AI predicting recurrence in non-muscle-invasive bladder cancer: systematic review with study strengths and weaknesses. *Front Oncol*. 2025;14:1509362. <https://doi.org/10.3389/fonc.2024.1509362>.
185. Ji Y, Dutta P, Davuluri R. Deep multi-omics integration by learning correlation-maximizing representation identifies prognostically stratified cancer subtypes. *Bioinform Adv*. 2023;3:vbad075. <https://doi.org/10.1093/bioadv/vbad075>.
186. Ahangar M, Mahjoubi F, Mowla SJ. Bladder cancer biomarkers: current approaches and future directions. *Front Oncol*. 2024. <https://doi.org/10.3389/fonc.2024.1453278>.
187. Giordano A, Soria F. Role and efficacy of current biomarkers in bladder cancer. *AME Med J*. 2020;5:6–6. <https://doi.org/10.21037/amj.2020.02.06>.
188. Scientific Image and Illustration Software | BioRender. n.d. <https://www.biorender.com/>. Accessed 29 March 2025.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.