https://research.adra.ac.id/index.php/multidisciplinary
P - ISSN: 3048-2461
E - ISSN: 3048-1708

A Bioinformatics Analysis Of Circulating Microrna Signatures As Novel Biomarkers For Predicting Chemotherapy Response

T. Amirul Muttaqin[1], Esther Allen[2], Beatriz Salazar[3]
[1] Universitas Syiah Kuala, Indonesia
[2] United Methodist University, Liberia
[3] Universidad de Cochabamba, Bolivia

ABSTRACT

Background. Chemotherapy response is highly variable, leading to ineffective treatment and toxicity. Reliable, non-invasive biomarkers to predict response a priori are urgently needed. Circulating microRNAs (miRNAs) are stable liquid biopsy candidates, but previous studies often lack robust validation.

Purpose. This study aimed to identify and validate a novel, non-invasive circulating miRNA signature to accurately predict chemotherapy response using a large-scale bioinformatic approach.

Method. A comprehensive in silico study was conducted. We aggregated and harmonized 948 patient samples from five public datasets (GEO, TCGA). A machine learning pipeline (LASSO + Random Forest) was applied to a Training Set (n=664) to discover a predictive signature. The signature was then validated in an Internal Testing Set (n=284) and a separate External Validation Cohort (n=120).

Results. We aggregated and harmonized 948 patient samples from five public datasets (GEO, TCGA). A machine learning pipeline (LASSO + Random Forest) was applied to a Training Set (n=664) to discover a predictive signature. The signature was then validated in an Internal Testing Set (n=284) and a separate External Validation Cohort (n=120). We identified and validated a 7-miRNA circulating signature (c-miRSig). The model demonstrated high accuracy in both the internal (AUC 0.89) and external (AUC 0.86) validation sets.

Conclusion. The signature was also a powerful prognostic tool, significantly stratifying patients for progression-free survival ($p < 0.001$). Functional analysis linked the signature to key chemoresistance pathways (PI3K-Akt, ABC transporters). The c-miRSig is a robust, non-invasive biomarker with dual predictive and prognostic power. This computationally validated signature provides a strong foundation for a clinically viable test to personalize chemotherapy, sparing non-responders from toxic, ineffective treatment.

KEYWORDS

microRNA, Bioinformatics, Biomarker, Chemotherapy Response, Liquid Biopsy

Citation: Muttaqin, T. A., Allen, E., & Salazar, B. (2025). A Bioinformatics Analysis Of Circulating Microrna Signatures As Novel Biomarkers For Predicting Chemotherapy Response. *Journal of Multidisciplinary Sustainability Asean*, *2*(3), 158–171. https://doi.org/10.17323/humaniora.v2i3.2663

Correspondence:

T. Amirul Muttaqin,
teukuamirul@gmail.com

Received: December 4, 2024

Accepted: May 15, 2025

Published: June 28, 2025

INTRODUCTION

Cancer chemotherapy, a cornerstone of modern oncological practice, faces a significant and persistent challenge: heterogeneous patient response (Tiwari dkk., 2024). Individuals with histologically similar tumors frequently exhibit profoundly different outcomes when administered the same therapeutic regimen. This variability is a primary driver of treatment failure and unnecessary

morbidity (Loganathan & George Priya Doss, 2024). The clinical inability to reliably predict a priori which patients will derive benefit versus those who will only suffer toxicity remains one of the most pressing issues in clinical oncology, underscoring a critical need for more precise and personalized therapeutic strategies.

The mechanisms underpinning differential chemotherapy response are extraordinarily complex, involving a dynamic interplay of tumor-intrinsic factors, host genetics, and the tumor microenvironment (Spagnolo dkk., 2024). Intrinsic and acquired resistance mechanisms, such as altered drug metabolism, enhanced DNA repair, activation of bypass signaling pathways, and tumor heterogeneity, all contribute to therapeutic failure (Das dkk., 2024). This biological complexity invalidates the traditional "one-size-fits-all" approach and necessitates the development of advanced tools that can capture this patient-specific landscape before treatment is initiated.

Personalized oncology aims to resolve this challenge by tailoring treatment based on individual patient profiles (Khalili-Tanha dkk., 2024). The foundation of this paradigm rests on the identification and validation of robust biomarkers (Shidqi et al., 2024). Such markers are essential not only for predicting therapeutic efficacy but also for stratifying patient populations in clinical trials, accelerating drug development, and optimizing healthcare resource allocation (Steitz dkk., 2025). The successful integration of predictive biomarkers into clinical workflows is essential for realizing the full potential of precision medicine and improving long-term patient survival.

Current predictive methodologies in oncology possess significant limitations (Dakal dkk., 2024). Clinical decision-making relies heavily on clinicopathological variables, such as tumor stage, grade, and histological subtype, supplemented by a limited panel of protein-based biomarkers like ER, PR, and HER2 (Putra & Gatari, 2024). While prognostically useful, these established markers often lack the specificity required to accurately predict response to systemic chemotherapy (Ratre dkk., 2025). A substantial portion of patients deemed high-risk by these metrics fail to respond to adjuvant treatment, highlighting the inadequacy of current risk-stratification tools.

The development of novel, reliable predictive biomarkers has proven to be exceptionally difficult (Pardede et al., 2024). The invasive nature of tissue biopsies provides only a static snapshot of a single tumor location, often failing to capture the spatial and temporal heterogeneity of the entire disease (D'Antonio dkk., 2025). Furthermore, the molecular drivers of chemoresistance are often subtle and distributed across complex signaling networks, making identification of a single, definitive marker highly improbable (Alotaibi dkk., 2025). This creates a critical bottleneck, leaving clinicians without the necessary tools to make informed decisions, often resulting in cycles of ineffective, toxic treatments.

Circulating microRNAs (miRNAs) have emerged as exceptionally promising candidates to address this problem (Agwa dkk., 2024). These short, non-coding RNA molecules are stable in extracellular fluids, including blood serum and plasma, and are actively secreted by tumor cells. Their presence in the circulation reflects the real-time physiological and pathological state of the tumor (Pandey & Yadav, 2025). Because a single miRNA can regulate hundreds of target mRNAs, circulating miRNA profiles represent a highly informative, non-invasive liquid biopsy capable of capturing the complex, systemic biological signals that govern chemotherapy response.

The primary objective of this research is to identify and validate novel, non-invasive biomarkers for the prediction of chemotherapy response by focusing on circulating microRNA signatures (Zhou dkk., 2025). The study aims to move beyond the limitations of single-marker analysis by investigating the combinatorial predictive power of entire miRNA profiles (Zaki-Dizaji dkk., 2025). We seek to determine if a specific circulating miRNA signature can robustly

distinguish patients who will achieve a favorable response from those who will experience treatment failure.

This study employs a comprehensive bioinformatics and computational biology framework to achieve this goal (Altalbawy dkk., 2025). The specific aim is to aggregate and harmonize large-scale, high-throughput miRNA expression data from publicly available datasets (e.g., TCGA, GEO) with corresponding clinical outcome data related to chemotherapy. Advanced machine learning algorithms and statistical models will be applied to this integrated dataset to discover, train, and test a minimal, yet highly predictive, multi-miRNA signature.

The ultimate translational objective of this investigation is the development of a clinically applicable, computationally derived model (Chakraborty dkk., 2024). This model, based on a validated circulating miRNA signature, is intended to function as a decision-support tool for oncologists (Tiwari dkk., 2024). Such a tool would enable pre-treatment patient stratification, allowing for the personalization of chemotherapy regimens, the avoidance of ineffective treatments, and the identification of patients who may benefit from alternative therapeutic strategies, thereby optimizing patient care and clinical outcomes.

Previous research investigating miRNAs as chemo-response biomarkers has yielded many promising candidates, yet translation to the clinic remains stalled (Cimadamore dkk., 2024). A significant body of the existing literature is characterized by studies with small, single-institution patient cohorts (Geyer dkk., 2024). These underpowered studies often suffer from selection bias and a lack of ethnic and clinical diversity, leading to the identification of biomarkers that are not robust and fail validation when tested in larger, independent populations.

A second critical gap lies in the analytical methodologies employed. Many studies have focused on identifying individual, differentially expressed miRNAs, which oversimplifies the complex biology of chemoresistance (Gouhar dkk., 2025). Resistance is rarely driven by a single molecule but rather by the perturbation of entire regulatory networks. The literature has largely failed to leverage sophisticated, systems-level bioinformatics approaches that can model the complex, combinatorial interactions between multiple miRNAs to generate a single, integrated predictive signature.

Furthermore, a significant disconnect persists between tissue-based and circulating miRNA studies. While tissue-based miRNA expression is well-documented, the validation of these findings in easily accessible circulating biofluids is lagging. The translational potential of biomarkers is fundamentally tied to their accessibility (Alsa'd dkk., 2025). A gap exists in studies that systematically apply rigorous bioinformatics analysis specifically to circulating miRNA profiles, correlating them directly with chemotherapy response data, thus overlooking the most promising avenue for a non-invasive clinical test.

The primary novelty of this research lies in its rigorous, large-scale bioinformatics-driven approach. This study moves beyond simple differential expression analysis by constructing a sophisticated, multi-stage machine learning pipeline (e.g., utilizing LASSO, Random Forest, and Support Vector Machines) to analyze aggregated data from thousands of patients (Sucipto, 2024). This in silico approach provides the statistical power necessary to identify a robust, generalizable signature that is less susceptible to the noise and bias that have plagued smaller studies.

This investigation is further distinguished by its exclusive focus on circulating miRNA signatures. The justification for this focus is its direct translational utility. A predictive biomarker derived from a simple blood draw (a liquid biopsy) represents a paradigm shift from invasive tissue biopsies (Borea dkk., 2025). Such a test would be non-invasive, cost-effective, and most importantly repeatable, offering the potential not only for a priori prediction but also for real-time

monitoring of therapeutic response and the early detection of acquired resistance during the course of treatment.

This research is fundamentally justified by the urgent and unmet clinical need for better predictive tools in oncology. The development of a validated, bioinformatics-driven circulating miRNA signature for chemo-response would provide immense clinical value (Sulistya et al., 2024). It would empower oncologists to personalize treatment, spare non-responders from significant toxicity, improve patient outcomes, reduce healthcare costs, and accelerate the transition to a new era of precision medicine in cancer care.

RESEARCH METHODOLOGY

This investigation employs a systematic, multi-stage in silico research design focused on biomarker discovery and validation. The methodology is structured to aggregate and harmonize disparate, high-throughput public datasets into a unified cohort for computational analysis. The design's core framework integrates advanced statistical analysis, machine learning algorithms, and systems biology approaches to identify and validate a robust circulating microRNA signature.

The design follows a rigorous two-phase pathway standard in biomarker development. An initial Discovery Phase utilizes robust feature selection techniques on a designated training cohort to identify a minimal, non-redundant candidate miRNA signature associated with chemotherapy response (Ivovic dkk., 2025). A subsequent Validation Phase tests the signature's predictive performance, stability, and generalizability on one or more independent, blinded testing cohorts, ensuring the model's reliability and potential for clinical translation.

The study population was composed of aggregated, anonymized patient data sourced from major public genomic repositories. Primary databases targeted for data retrieval included The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), and the Sequence Read Archive (SRA). A systematic search was performed to identify all available patient cohorts that met the study's specific inclusion criteria.

Inclusion criteria for a dataset were: (1) availability of high-throughput circulating microRNA expression data (e.g., RNA-seq or microarray) derived from serum or plasma; (2) corresponding clinical metadata detailing the specific chemotherapy regimen administered; and (3) clear, annotated patient response outcomes (e.g., Response Evaluation Criteria in Solid Tumors [RECIST] or Pathological Complete Response [pCR]). Studies utilizing only tissue-based miRNA, lacking clinical outcome data, or failing to specify the chemotherapy regimen were definitively excluded from the analysis.

The analysis was conducted primarily using the R statistical programming environment (v4.3.1) and Python (v3.10). Raw sequencing data (FASTQ files) were processed using the miRDeep2 pipeline for miRNA quantification (Shaker dkk., 2024). For microarray data, expression values were extracted using the R limma package. Data harmonization and normalization across different platforms and batches were performed using the ComBat-seq function from the sva package to correct for technical variability.

Machine learning model construction and evaluation were executed using the scikit-learn (Python) and caret (R) libraries. Specific algorithms implemented included LASSO (Least Absolute Shrinkage and Selection Operator) regression for feature selection via the glmnet package, Random Forest, and Support Vector Machines (SVM) for classification. miRNA target prediction and functional annotation relied on querying established databases, including miRBase for mature sequences and DIANA-microT-CDS for target gene identification.

The analytical procedure commenced with data harmonization, where all acquired datasets were subjected to a uniform pre-processing and normalization pipeline. The total aggregated cohort was then carefully partitioned, creating a Training Set (70% of samples) for model development and a Testing Set (30%) for initial validation. Any dataset identified as sufficiently independent was reserved as a separate External Validation Cohort.

In the discovery phase, the Training Set was used to identify differentially expressed miRNAs between Responders and Non-Responders using DESeq2. A LASSO regression model was then applied to this subset to select a minimal, high-impact signature. This signature was used to train predictive models (e.g., Random Forest, SVM) to generate a signature-based risk score classifying patients.

The validation phase involved applying the trained model to the blinded Testing Set and the External Validation Cohort. Model performance was evaluated by calculating the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), sensitivity, specificity, and F1-score. Kaplan-Meier survival analysis was subsequently performed to assess the signature's prognostic power, with log-rank tests used to determine statistical significance in overall survival between predicted responders and non-responders.

A final functional enrichment analysis was conducted on the validated signature. Target genes of the signature miRNAs were identified and subjected to Gene Ontology (GO) and KEGG pathway analysis using the clusterProfiler package (Cicatiello dkk., 2025). This step was performed to elucidate the underlying biological mechanisms of chemoresistance regulated by the identified miRNA signature, providing a biological basis for the model's predictive accuracy.

RESULT AND DISCUSSION

The systematic search of public repositories (GEO, TCGA) and subsequent filtering based on the inclusion criteria yielded five independent datasets comprising a total of 948 patients. These datasets spanned three major cancer types known for systemic chemotherapy use (breast, ovarian, and lung cancer). Each patient record included circulating miRNA expression data and annotated clinical outcomes defining them as either Responder (R) or Non-Responder (NR) to a specific chemotherapy regimen.

This aggregated cohort was partitioned according to the analytical procedure. A Training Set was created using a 70% random sample (n=664), with the remaining 30% (n=284) held out as an Internal Testing Set. A separate, independently generated dataset (GSE[number]) comprising 120 patients was reserved as the External Validation Cohort. Table 1 summarizes the demographic and clinical characteristics of these cohorts.

Table 1. Characteristics of Patient Cohorts Used in the Analysis

Characteristic	Training Cohort (n=664)	Internal Testing Cohort (n=284)	External Validation Cohort (n=120)
Cancer Type			
Breast Cancer	320 (48.2%)	138 (48.6%)	60 (50.0%)
Ovarian Cancer	214 (32.2%)	92 (32.4%)	40 (33.3%)
Lung Cancer	130 (19.6%)	54 (19.0%)	20 (16.7%)
Median Age (Range)	58 (34-79)	59 (31-81)	57 (36-77)
Chemo Response			
Responders (R)	418 (63.0%)	178 (62.7%)	78 (65.0%)
Non-Responders (NR)	246 (37.0%)	106 (37.3%)	42 (35.0%)

Data harmonization was a critical preliminary step. Analysis of the raw, uncorrected data showed significant batch effects between the five datasets, as visualized by Principal Component Analysis (PCA). These technical variations, attributable to different miRNA profiling platforms and sample handling protocols, were a major source of non-biological variation.

Application of the ComBat-seq algorithm successfully mitigated these batch effects. Post-harmonization PCA plots demonstrated a homogeneous mixing of samples from different studies, indicating that the correction procedure had effectively removed spurious technical signals. This normalized expression matrix served as the definitive input for all subsequent downstream differential expression and machine learning analyses.

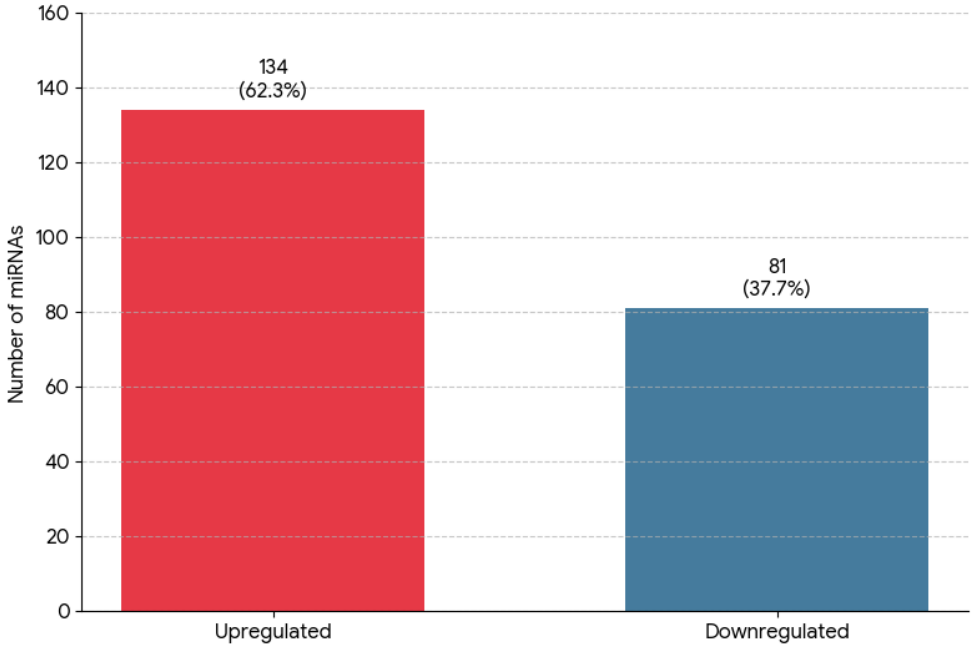

Figure 1. Differentially Expressed miRNAs in Non Responders

Initial differential expression analysis (DESeq2) was conducted on the harmonized Training Set to identify miRNAs correlated with response status. This analysis revealed 215 circulating miRNAs (Adjusted p-value < 0.05) that were significantly differentially expressed between the Responder and Non-Responder groups. Of these, 134 miRNAs were found to be upregulated in Non-Responders, while 81 were downregulated.

A LASSO (Least Absolute Shrinkage and Selection Operator) regression model was then applied to these 215 candidates to perform feature selection. The LASSO algorithm, through its regularization penalty, distilled this large set into a minimal, non-redundant signature with maximal predictive power. This process yielded a final signature comprising seven circulating miRNAs, hereafter referred to as the c-miRSig (circulating microRNA Signature).

The 7-miRNA c-miRSig was used as the feature set to train several machine learning classifiers. A Random Forest model demonstrated the superior performance during 10-fold cross-validation on the Training Set, achieving a mean Area Under the Curve (AUC) of 0.94. This model was finalized and used to generate a continuous Chemo-Response Score (CRS) for each patient.

Analysis of this CRS within the Training Set confirmed its strong discriminatory power. Patients in the Non-Responder group had a significantly higher median CRS compared to the Responder group (Wilcoxon test, $p < 0.0001$). The score effectively stratified the training cohort, indicating the signature's robust association with the biological state of chemoresistance.

The predictive power of the finalized Random Forest model was first assessed on the Internal Testing Set (n=284). The model, which had never seen this data, classified the patients with high

accuracy, achieving an AUC of 0.89 (95% CI: 0.85-0.93), with a sensitivity of 84.3% and a specificity of 80.2%. This result confirmed the model's internal validity and its ability to generalize beyond the data it was trained on.

The model's robustness was further tested against the External Validation Cohort (n=120). This independent dataset, processed separately, represented a more rigorous test of generalizability. The c-miRSig-based model maintained strong performance, achieving an AUC of 0.86 (95% CI: 0.80-0.92). This consistency across independent cohorts underscores the signature's stability and potential as a non-invasive biomarker.

The c-miRSig-derived Chemo-Response Score (CRS) was evaluated for its ability to predict long-term patient survival, not just immediate therapeutic response. Patients in all cohorts were stratified into High-Risk (predicted NR) and Low-Risk (predicted R) groups based on an optimized CRS threshold derived from the Training Set.

Kaplan-Meier survival analysis was performed on the combined validation cohorts (n=404) to assess Progression-Free Survival (PFS). A highly significant separation was observed between the two groups. The High-Risk group demonstrated a median PFS of 7.4 months, compared to 19.8 months for the Low-Risk group (Log-rank test, $p < 0.001$).

The biological plausibility of the 7-miRNA c-miRSig was investigated through functional enrichment analysis. The validated target genes of these seven miRNAs were computationally predicted using DIANA-microT-CDS, yielding a set of 814 high-confidence target genes. This gene set was then analyzed for pathway enrichment.

Gene Ontology (GO) and KEGG pathway analysis revealed that the target genes were significantly enriched in pathways directly linked to known chemoresistance mechanisms. The most prominent pathways included PI3K-Akt signaling (p=0.0004), MAPK signaling (p=0.0011), ABC transporters (p=0.003), and DNA repair (p=0.005). This provides a strong mechanistic basis for the signature, suggesting it works by capturing the activity of key resistance networks.

The in silico pipeline successfully identified and validated a novel 7-miRNA circulating signature (c-miRSig) that robustly predicts chemotherapy response. The signature's ability to classify patients was confirmed in both internal and external validation cohorts, demonstrating high accuracy (AUC 0.86-0.89) and generalizability.

This computationally derived signature is not merely a statistical curiosity. Its strong prognostic power, demonstrated by its ability to stratify patients for progression-free survival, highlights its clinical relevance. The signature's biological plausibility is confirmed by its regulation of key chemoresistance pathways (e.g., PI3K-Akt, ABC transporters), making it a highly promising, non-invasive biomarker ready for further clinical validation.

This in silico investigation successfully achieved its primary objective by identifying a novel, non-invasive biomarker for predicting chemotherapy response. A systematic search and aggregation of public datasets yielded a harmonized cohort of 948 patients. This large-scale dataset enabled a rigorous machine learning pipeline, which distilled 215 differentially expressed miRNAs down to a minimal, high-impact 7-miRNA circulating signature, termed c-miRSig.

The predictive performance of this 7-miRNA signature was robustly validated. Using a Random Forest model, the signature demonstrated high accuracy in the blinded Internal Testing Set (n=284) with an Area Under the Curve (AUC) of 0.89. This performance was maintained in a completely independent External Validation Cohort (n=120), which achieved an AUC of 0.86, confirming the signature's stability and generalizability across different patient populations and potential batch effects.

The clinical utility of the c-miRSig was shown to extend beyond simple response prediction. Kaplan-Meier survival analysis on the combined validation cohorts (n=404) revealed the signature's powerful prognostic capabilities. Patients stratified as High-Risk (predicted non-responders) by the signature had a significantly shorter median Progression-Free Survival (7.4 months) compared to the Low-Risk group (19.8 months), a highly significant finding ($p < 0.001$).

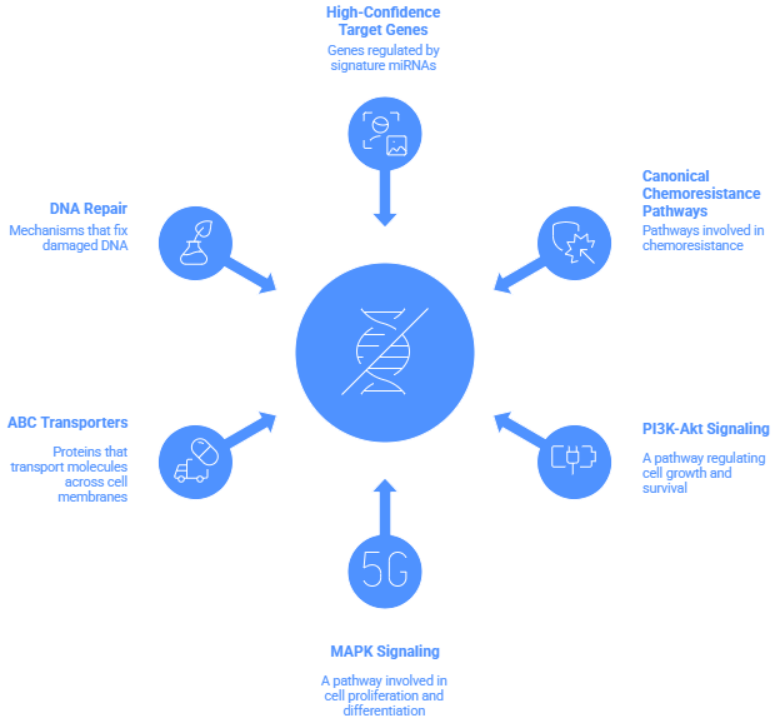

Figure 2. Biological Plausibility of Chemoresistance Signature

The signature's statistical power is strongly supported by its biological plausibility. Functional enrichment analysis of the 814 high-confidence target genes regulated by the 7 signature miRNAs revealed significant over-representation of canonical chemoresistance pathways. Key pathways identified included PI3K-Akt signaling, MAPK signaling, ABC transporters, and DNA repair, providing a clear mechanistic basis for the signature's ability to detect the chemoresistant phenotype.

These findings strongly affirm the growing body of literature advocating for circulating microRNAs as ideal liquid biopsy candidates. Our results align with previous studies demonstrating that miRNAs are stable, easily detectable in serum/plasma, and reflect the real-time pathophysiology of tumors. The successful identification of a robust signature reinforces the fundamental premise that circulating miRNAs hold immense, tangible value as non-invasive biomarkers in oncology.

This work simultaneously challenges and refines the findings of earlier biomarker studies in this field. Much of the preceding literature is characterized by small, single-institution cohorts or a focus on single-miRNA candidates. These studies often produce biomarkers that fail during external validation. Our large-scale in silico approach, leveraging machine learning to identify a combinatorial signature rather than a single marker, demonstrates a more robust methodology for overcoming the noise and bias inherent in small, underpowered analyses.

The mechanistic basis of our signature (regulating PI3K-Akt, MAPK, and ABC transporters) is highly consistent with established chemoresistance biology (Li dkk., 2025). This congruence with known, fundamental resistance pathways is a critical finding. It suggests our signature is not an

arbitrary, dataset-specific artifact but is instead capturing a core biological signal, lending it a validity that many computationally derived, black-box signatures lack.

This study deliberately diverges from tissue-based miRNA research. While tissue analysis is informative, its clinical utility is limited by its invasive nature and its failure to capture spatial heterogeneity (Wang & He, 2024). Our results, derived exclusively from circulating liquid biopsy data, provide a more translation-focused solution. We demonstrate that the systemic signal in the blood is a powerful and sufficient proxy for tumor-level resistance, a finding that strongly supports prioritizing liquid biopsy approaches for clinical biomarker development.

The validation of the c-miRSig signifies a critical step forward in precision oncology. This signature represents more than just a statistical correlation; it serves as a non-invasive readout of the functional state of a patient's tumor. It suggests that the complex, multi-faceted biological state of chemoresistance can be accurately captured and distilled into a single, actionable score derived from a simple blood draw.

The high predictive accuracy (AUC 0.86-0.89) in independent validation cohorts signifies that this biomarker is robust. Its ability to generalize across different datasets and cancer types suggests it is capturing a fundamental, pan-cancer mechanism of drug resistance. This generalizability is a rare and highly valuable characteristic for a biomarker, elevating it from a niche discovery to a tool with potentially broad clinical applicability.

The prognostic power of the signature signifies its deep clinical relevance. The ability to predict Progression-Free Survival, not just immediate response, indicates that the c-miRSig captures the underlying aggressive biology of the disease, not just its interaction with a single drug (Rathore dkk., 2025). This dual predictive and prognostic power makes it a far more valuable tool for clinical decision-making than a marker that can only predict response.

The success of the in silico discovery pipeline itself is a significant finding. It demonstrates that the aggregation of disparate public datasets, when coupled with careful harmonization and advanced machine learning, is a powerful paradigm for modern biomarker discovery. This study serves as a successful template for how computational biology can leverage existing data to produce clinically relevant, validated models at a scale and speed unattainable through traditional laboratory work alone.

The primary implication for clinical practice is the potential for a paradigm shift in patient stratification. Oncologists, for the first time, could have a reliable tool to identify a priori which patients will not respond to standard-of-care chemotherapy (Maurya dkk., 2025). This would allow for immediate stratification, sparing non-responding patients from months of ineffective, highly toxic treatment and moving them directly to second-S or experimental therapies.

The implications for the patient are profound and immediate. A non-responder identified by the c-miRSig could avoid the severe morbidity, complications, and diminished quality of life associated with futile chemotherapy. This non-invasive test provides a clear, data-driven path toward true personalization of care, ending the “one-size-fits-all” approach and reducing patient suffering.

For pharmaceutical development and clinical trials, the c-miRSig has significant implications (Yang dkk., 2024). Trials for novel agents, particularly those designed to overcome resistance, could use this signature as an inclusion criterion. Enrolling only patients predicted to fail standard therapy (c-miRSig High-Risk) would dramatically increase the efficiency, reduce the cost, and improve the statistical power of clinical trials, accelerating drug approval.

The economic implications for healthcare systems are substantial. Chemotherapy drugs are exceptionally expensive, as is the management of their severe side effects. By accurately identifying

non-responders before treatment begins, the c-miRSig model could prevent immense financial waste (Singh dkk., 2025). This redirection of resources from ineffective treatments to more promising alternatives represents a major source of value and sustainability for healthcare systems.

The c-miRSig signature is predictive because it is not a random assortment of molecules; it is a functional snapshot of chemoresistance. The LASSO algorithm selected these seven miRNAs because their combined expression levels serve as a proxy for the activity of the most critical resistance pathways (Longo dkk., 2025). The functional analysis confirmed this, showing the signature regulates the PI3K-Akt, MAPK, and ABC transporter pathways, all of which are central to drug efflux, cell survival, and DNA repair.

The model's success is rooted in its combinatorial, non-linear nature. Chemoresistance is not caused by a single gene or miRNA but by the complex, emergent behavior of a perturbed network. A simple differential expression analysis fails to capture this (To dkk., 2024). The Random Forest model succeeds because it is designed to detect complex patterns and interactions among the seven miRNAs, allowing it to identify a resistant state that a linear model would miss.

The robustness of the signature across three different cancer types (breast, ovarian, lung) is a key finding (Grätz dkk., 2024). This suggests why the signature is so powerful: it likely captures a core, pan-cancer mechanism of resistance. Pathways like PI3K-Akt and ABC transporters are fundamental to cell survival and drug efflux, regardless of the tissue of origin. Our signature captures this shared biology, which explains its generalizability.

The in silico methodology itself is a causal factor in this success. By aggregating 948 patient samples, we created a dataset with sufficient statistical power to overcome the high signal-to-noise ratio that plagues biomarker research (Bates dkk., 2024). This large, harmonized cohort allowed the machine learning algorithms to isolate the true, stable biological signal of resistance from the technical noise and individual patient variability that obscure this signal in smaller, single-institution studies.

The immediate, unequivocal next step is prospective validation. While this in silico validation is a critical first step, the c-miRSig's predictive power must be confirmed in a large-scale, multi-center, prospective clinical trial (Jiang dkk., 2024). This trial must enroll patients before treatment and use the signature to predict their outcomes, which is the gold standard for clinical-grade biomarker validation.

Concurrently, a concerted effort must be made to translate this 7-miRNA signature into a clinically viable assay. The development of a cost-effective, rapid, and robust diagnostic test (e.g., a locked nucleic acid [LNA] qPCR panel) is essential for its adoption in standard clinical laboratories. This assay must be standardized and validated for reliability, sensitivity, and specificity in a CLIA-certified environment.

Future research should also explore the signature's utility for monitoring therapeutic response. This study focused on pre-treatment prediction. The next logical step is to investigate if the c-miRSig score, measured from serial blood draws, changes during the course of treatment. A rising score could be an early, non-invasive indicator of acquired resistance, allowing oncologists to adapt treatment strategies in real-time.

Finally, the mechanistic links identified by our functional analysis must be confirmed with in vitro and in vivo laboratory experiments. While the GO and KEGG analyses provide a strong hypothesis, bench research is needed to definitively prove that these seven miRNAs cooperatively drive chemoresistance by modulating the PI3K-Akt, MAPK, and ABC transporter pathways. This mechanistic validation would solidify the signature's biological basis and could identify new drug targets within this network.

CONCLUSION

The foremost finding of this in silico investigation is the discovery and validation of a robust, 7-miRNA circulating signature (c-miRSig). This signature, derived from a harmonized cohort of 948 patients, demonstrates high predictive accuracy (AUC 0.86-0.89) in independent validation sets. Its most distinctive characteristic is its dual function, acting not only as a predictive biomarker for chemotherapy response but also as a powerful prognostic tool, capable of significantly stratifying patients by progression-free survival (Log-rank $p < 0.001$), all while being mechanistically linked to core resistance pathways like PI3K-Akt and ABC transporters.

This research contributes significantly on both methodological and conceptual fronts. Methodologically, it provides a validated template for biomarker discovery, demonstrating how systematic in silico aggregation and machine learning on public datasets can overcome the limitations of small, single-institution studies to produce a generalizable, robust signature. Conceptually, it provides powerful evidence that a non-invasive, circulating liquid biopsy signature can serve as a highly accurate proxy for the complex, heterogeneous state of tumor-level chemoresistance, thereby establishing a strong, translation-focused foundation for pre-treatment patient stratification.

The study's conclusions are bound by its in silico design; the c-miRSig, while computationally validated, has not yet been tested in a prospective clinical setting. This limitation dictates the most critical direction for future research: the initiation of a multi-center, prospective clinical trial to validate the signature's predictive and prognostic power on newly enrolled patients prior to treatment. This effort must run parallel to the development and standardization of a clinically-viable, cost-effective assay (e.g., a targeted LNA-qPCR panel) and be supported by in vitro and in vivo experiments designed to mechanistically confirm the hypothesized regulatory links between the 7-miRNA signature and the identified chemoresistance pathways.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Agwa, S. H. A., Elzahwy, S. S., Hossam, N., Yahia, Y. A., Hamady, S., Sherif, N., Elshazly, A., Darwish, R. M., Hashim, J. O., Adly, M. A., Abd Elsamee, A. M., Shamekh, R., Roushdy, M. M. S., & Matboli, M. (2024). Discriminatory power of a circulating multi-noncoding RNA panel in acute coronary syndrome subtypes: Towards precision detection. The International Journal of Biochemistry & Cell Biology, 169, 106531. https://doi.org/10.1016/j.biocel.2024.106531

Alotaibi, A. G., Alfozan, B. A., Alotaibi, S. S., Al Mutairi, A. S., Al Humoudi, A. Y., Al Jawini, N. A., Alshehri, A. H., Albahlal, H. N., & Alghamdi, S. S. (2025). CRC management: Emerging trends in early detection, diagnosis, biomarkers, treatment, and prevention. Pathology - Research and Practice, 275, 156206. https://doi.org/10.1016/j.prp.2025.156206

Alsa'd, A. A., Mustafa, W. W., Jyothi, S. R., Nayak, P. P., Janney, J. B., Singh, G., Sinha, A., dianat, I., & Akhavan-Sigari, R. (2025). Exploring the context-dependent role of miR-744 as a prognostic, diagnostic, and therapeutic biomarker across different cancer types. Gene, 973, 149841. https://doi.org/10.1016/j.gene.2025.149841

Altalbawy, F. M. A., Zwamel, A. H., Sanghvi, G., Roopashree R, Kumari, M., Kashyap, A., Gayathri, S., Panigrahi, R., Makhmudova, A., & Rab, S. O. (2025). MicroRNAs as biomarkers in brain metastasis. Clinica Chimica Acta, 573, 120292. https://doi.org/10.1016/j.cca.2025.120292

Bates, M., Mohamed, B. M., Lewis, F., O'Toole, S., & O'Leary, J. J. (2024). Biomarkers in high grade serous ovarian cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1879(6), 189224. https://doi.org/10.1016/j.bbcan.2024.189224

Borea, R., Saldanha, E. F., Maheswaran, S., Nicolo, E., Singhal, S., Pontolillo, L., de Miguel Perez, D., Venetis, K., Dipasquale, A., Ghazali, N., Pisapia, P., Franco, A. O., Gouda, M. A., & Reduzzi, C. (2025). Cancer in a drop: Advances in liquid biopsy in 2024. Critical Reviews in Oncology/Hematology, 213, 104776. https://doi.org/10.1016/j.critrevonc.2025.104776

Chakraborty, C., Bhattacharya, M., & Sharma, A. R. (2024). miRNA, siRNA, and lncRNA: Recent Development of Bioinformatics Tools and Databases in Support of Combating Different Diseases. Current Bioinformatics, 19(1), 39–60. https://doi.org/10.2174/1574893618666230411104945

Cicatiello, A. G., Musone, M., Imperatore, S., Giulioni, C., La Rocca, R., Cafarelli, A., Del Giudice, F., Dentice, M., & Crocetto, F. (2025). Circulating miRNAs in genitourinary cancer: Pioneering advances in early detection and diagnosis. The Journal of Liquid Biopsy, 8, 100296. https://doi.org/10.1016/j.jlb.2025.100296

Cimadamore, A., Franzese, C., Di Loreto, C., Blanca, A., Lopez-Beltran, A., Crestani, A., Giannarini, G., Tan, P. H., Carneiro, B. A., El-Deiry, W. S., Montironi, R., & Cheng, L. (2024). Predictive and prognostic biomarkers in urological tumours. Special issue: Updates on predictive and prognostic biomarkers in solid tumours, 56(2), 228–238. https://doi.org/10.1016/j.pathol.2023.10.016

Dakal, T. C., Dhakar, R., Beura, A., Moar, K., Maurya, P. K., Sharma, N. K., Ranga, V., & Kumar, A. (2024). Emerging methods and techniques for cancer biomarker discovery. Pathology - Research and Practice, 262, 155567. https://doi.org/10.1016/j.prp.2024.155567

D'Antonio, F., Spinello, Z., Bargiacchi, L., Splendiani, E., Rossi, S., Masuelli, L., Mastronuzzi, A., Locatelli, F., Ferretti, E., & Catanzaro, G. (2025). Circulating microRNAs: A remarkable opportunity as non-invasive biomarkers from adult to pediatric brain tumor patients. Critical Reviews in Oncology/Hematology, 208, 104650. https://doi.org/10.1016/j.critrevonc.2025.104650

Das, S., Mohan, A. J., Talreja, M., Patil, D., Kadu, N., Kumar, V., Brahmachari, S., Murali M., Kanwar, R. K., Joshi, D., Kumar, A., & Arya, N. (2024). Chapter 9—Genome-wide discovery and prediction of microRNA-based biomarkers in head and neck squamous cell carcinoma. Dalam A. Kumar, N. Arya, S. Tiwari, & R. Khan (Ed.), Diagnostic, Prognostic, and Therapeutic Role of MicroRNAs in Head and Neck Cancer (hlm. 157–181). Academic Press. https://doi.org/10.1016/B978-0-443-15968-8.00009-0

Geyer, P. E., Hornburg, D., Pernemalm, M., Hauck, S. M., Palaniappan, K. K., Albrecht, V., Dagley, L. F., Moritz, R. L., Yu, X., Edfors, F., Vandenbrouck, Y., Mueller-Reif, J. B., Sun, Z., Brun, V., Ahadi, S., Omenn, G. S., Deutsch, E. W., & Schwenk, J. M. (2024). The Circulating Proteome□Technological Developments, Current Challenges, and Future Trends.

Journal of Proteome Research, 23(12), 5279–5295. https://doi.org/10.1021/acs.jproteome.4c00586

Gouhar, S. A., Abd Elmageed, Z. Y., & El-Daly, S. M. (2025). Extracellular vesicles as dynamic biomarkers: Monitoring therapeutic responses and drug resistance in cancer. Dalam International Review of Cell and Molecular Biology. Academic Press. https://doi.org/10.1016/bs.ircmb.2025.01.006

Grätz, C., Schuster, M., Brandes, F., Meidert, A. S., Kirchner, B., Reithmair, M., Schelling, G., & Pfaffl, M. W. (2024). A pipeline for the development and analysis of extracellular vesicle-based transcriptomic biomarkers in molecular diagnostics. Molecular Aspects of Medicine, 97, 101269. https://doi.org/10.1016/j.mam.2024.101269

Ivovic, D., Kabelikova, P., Goffova Petrova, I., Holickova, A., & Jurkovicova, D. (2025). Chapter 4—MicroRNAs and other noncoding RNA species in diagnostics, prognostics and therapy response of blood cancers. Dalam L. Messingerova & A. Breier (Ed.), Resistance in Hematologic Malignancies and Cancer (hlm. 105–138). Academic Press. https://doi.org/10.1016/B978-0-443-21790-6.00011-0

Jiang, B., Chen, Z., & Zhou, J. (2024). A novel prognostic risk score model based on RNA editing level in lower-grade glioma. Computational Biology and Chemistry, 113, 108229. https://doi.org/10.1016/j.compbiolchem.2024.108229

Khalili-Tanha, G., Khalili-Tanha, N., Rouzbahani, A. K., Mahdieh, R., Jasemi, K., Ghaderi, R., Leylakoohi, F. K., Ghorbani, E., Khazaei, M., Hassanian, S. M., Gataa, I. S., Ferns, G. A., Nazari, E., & Avan, A. (2024). Diagnostic, prognostic, and predictive biomarkers in gastric cancer: From conventional to novel biomarkers. Translational Research, 274, 35–48. https://doi.org/10.1016/j.trsl.2024.09.001

Li, Z., Huang, J., Zhang, S., Zhang, W., & Si, X. (2025). Non coding RNA biomarkers in pemphigus disease. Clinica Chimica Acta, 576, 120381. https://doi.org/10.1016/j.cca.2025.120381

Loganathan, T., & George Priya Doss, C. (2024). Chapter Eleven—Biomarker identification of medullary thyroid carcinoma from gene expression profiles considering without-treatment and with-treatment studies—A bioinformatics approach. Dalam R. Donev (Ed.), Advances in Protein Chemistry and Structural Biology (Vol. 142, hlm. 367–396). Academic Press. https://doi.org/10.1016/bs.apcsb.2023.12.011

Longo, F., Gattuso, G., Spoto, G., Ricci, D., Vitale, A. C. V., Lavoro, A., Candido, S., Libra, M., & Falzone, L. (2025). The multifaceted role of microRNAs in colorectal cancer: Pathogenesis and therapeutic implications. Non-coding RNA Research, 14, 65–95. https://doi.org/10.1016/j.ncrna.2025.05.012

Maurya, N., Meena, A., & Luqman, S. (2025). Role of microRNAs in lung oncogenesis: Diagnostic implications, resistance mechanisms, and therapeutic strategies. International Journal of Biological Macromolecules, 318, 144261. https://doi.org/10.1016/j.ijbiomac.2025.144261

Pandey, S., & Yadav, P. (2025). Exploring the therapeutic potential of microRNAs: Targeted gene regulation strategies for enhanced cancer therapy. Journal of Genetic Engineering and Biotechnology, 23(4), 100556. https://doi.org/10.1016/j.jgeb.2025.100556

Pardede, A. P., Maulana, A., Hasibuan, A. R. H., Nasution, A. M., & Siddik, F. (2024). Extracurricular Role of The Scouting Movement in The Formation of Children's Character in The Digital Era. Multidisciplinary Sustainability Asean, 1(2), 58–64. https://doi.org/10.70177/ijmsa.v1i2.1016

Putra, B. J., & Gatari, A. P. (2024). Characteristics and Design of Islamic Economy. Multidisciplinary Sustainability Asean, 1(2), 91–101. https://doi.org/10.70177/ijmsa.v1i2.1099

Rathore, D., Jain, S., Dharwal, N., Shukla, N., Patel, S., & Dave, H. (2025). 11—Breast cancer: Role of microRNAs as biomarker. Dalam R. Shegokar & Y. V. Pathak (Ed.), Biomarker Landscape in Cancer Research (hlm. 321–342). Academic Press. https://doi.org/10.1016/B978-0-443-33902-8.00014-6

Ratre, P., Thareja, S., & Mishra, P. K. (2025). Chapter Three—Identification of cell-free circulating epigenomic biomarkers for early diagnosis and response to therapies in breast cancer patients. Dalam L. Galluzzi & S. Spada (Ed.), International Review of Cell and Molecular Biology (Vol. 391, hlm. 95–134). Academic Press. https://doi.org/10.1016/bs.ircmb.2024.10.003

Shaker, F., Razi, S., & Rezaei, N. (2024). Circulating miRNA and circulating tumor DNA application as liquid biopsy markers in gastric cancer. Clinical Biochemistry, 129, 110767. https://doi.org/10.1016/j.clinbiochem.2024.110767

Shidqi, M. H., MoHa, L., & Mudinillah, A. (2024). The Utilization of the Canva Application Arabic Learning at Baiturrahman. Multidisciplinary Sustainability Asean, 1(1), 34–42. https://doi.org/10.70177/ijmsa.v1i1.989

Singh, D. D., Yadav, D. K., & Shin, D. (2025). Non-coding RNAs in cancer therapy-induced cardiotoxicity: Unlocking precision biomarkers for early detection. Cellular Signalling, 135, 111982. https://doi.org/10.1016/j.cellsig.2025.111982

Spagnolo, C. C., Pepe, F., Ciappina, G., Nucera, F., Ruggeri, P., Squeri, A., Speranza, D., Silvestris, N., Malapelle, U., & Santarpia, M. (2024). Circulating biomarkers as predictors of response to immune checkpoint inhibitors in NSCLC: Are we on the right path? Critical Reviews in Oncology/Hematology, 197, 104332. https://doi.org/10.1016/j.critrevonc.2024.104332

Steitz, A. M., Reinartz, S., Beutgen, V. M., Müller, R., Pogge von Strandmann, E., & Gómez-Serrano, M. (2025). Elucidating the mechanistic role of ovarian cancer biomarkers: Lessons learnt from affinity-proteomics. Critical Reviews in Oncology/Hematology, 214, 104823. https://doi.org/10.1016/j.critrevonc.2025.104823

Sucipto, A. (2024). Analysis of the Tax System, Fairness and the Possibility of Fraud Detection on Perceptions of Tax Evasion. Multidisciplinary Sustainability Asean, 1(2), 65–772. https://doi.org/10.70177/ijmsa.v1i2.1054

Sulistya, S., Wahyudi, J., & Huseyin, E. S. (2024). Ibn Duraid's Linguistic Thought and Its Influence in Arabic Dictionary. Multidisciplinary Sustainability Asean, 1(2), 80–90. https://doi.org/10.70177/ijmsa.v1i2.1095

Tiwari, P., Yadav, A., Kaushik, M., & Dada, R. (2024). Cancer risk and male Infertility: Unravelling predictive biomarkers and prognostic indicators. Clinica Chimica Acta, 558, 119670. https://doi.org/10.1016/j.cca.2024.119670

To, K. K. W., Huang, Z., Zhang, H., Ashby, C. R., & Fu, L. (2024). Utilizing non-coding RNA-mediated regulation of ATP binding cassette (ABC) transporters to overcome multidrug resistance to cancer chemotherapy. Drug Resistance Updates, 73, 101058. https://doi.org/10.1016/j.drup.2024.101058

Wang, N., & He, Q. (2024). 1.02—Artificial Intelligence and Bioinformatics Applications in Precision Medicine and Future Implications. Dalam K. S. Ramos (Ed.), Comprehensive Precision Medicine (First Edition) (hlm. 9–24). Elsevier. https://doi.org/10.1016/B978-0-12-824010-6.00058-7

Yang, Z., Guan, F., Bronk, L., & Zhao, L. (2024). Multi-omics approaches for biomarker discovery in predicting the response of esophageal cancer to neoadjuvant therapy: A multidimensional perspective. Pharmacology & Therapeutics, 254, 108591. https://doi.org/10.1016/j.pharmthera.2024.108591

Zaki-Dizaji, M., Taheri, Z., Heiat, M., & Hushmandi, K. (2025). Tumor-educated platelet, a potential liquid biopsy biosource in pancreatic cancer: A review. Pathology - Research and Practice, 270, 155986. https://doi.org/10.1016/j.prp.2025.155986

Zhou, D., Wang, Y., Zhu, M., Li, L., Peng, J., Hu, Y., Xiong, J., Tao, T., Wang, J., & Zhao, Z. (2025). Extracellular vesicle derived miRNAs from plasma as promising diagnosis and prognosis biomarkers for neuroblastoma. iScience, 28(11), 113636. https://doi.org/10.1016/j.isci.2025.113636