Integrative Bioinformatics Analysis of microRNA Expression and Genomic Signatures for Early Detection and Survival Prediction in Prostate Cancer

Abstract

Prostate cancer remains the second most common malignancy among men worldwide and a leading cause of cancer-related mortality. MicroRNAs (miRNAs) and genomic signatures are critical regulators of oncogenic processes and represent promising biomarkers for cancer diagnosis and prognosis. In this study, we conducted an integrative analysis of miRNA expression profiles and genomic data from public bioinformatics resources, including The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Differential expression analysis, Cox proportional hazards modeling, and network-based functional enrichment were applied to identify key miRNAs and gene signatures associated with prostate cancer progression and patient survival. Our findings revealed a panel of dysregulated miRNAs and co-expressed gene modules with significant prognostic value. These biomarkers provide novel insights into the molecular mechanisms underlying prostate cancer and may support non-invasive early detection, survival prediction, and personalized therapeutic strategies in clinical practice.

**Keywords:** microRNA, genomic signatures, prostate cancer, early detection, survival prediction, bioinformatics, prognostic biomarkers, precision medicine

1. Introduction

Prostate cancer represents a significant global health burden, ranking as the second most frequently diagnosed cancer among men worldwide, with an estimated 1.4 million new cases annually (Sung et al., 2021). Despite advances in screening and therapeutic interventions, prostate cancer remains a leading cause of cancer-related mortality, particularly in cases where early detection is missed or treatment response is suboptimal (Siegel et al., 2023). Current diagnostic approaches, primarily relying on prostate-specific antigen (PSA) screening and histopathological examination, face limitations including false positives, overdiagnosis of indolent tumors, and insufficient prognostic accuracy for treatment stratification (Mottet et al., 2021).



Fig 1 : Prostate gland showing tumor causing urethral stricture

MicroRNAs (miRNAs) have emerged as critical post-transcriptional regulators of gene expression, playing pivotal roles in carcinogenesis, tumor progression, and metastasis (Bartel, 2018). These small non-coding RNAs, typically 18-25 nucleotides in length, regulate approximately 60% of human protein-coding genes through sequence-specific binding to target mRNAs, leading to translational repression or mRNA degradation (Friedman et al., 2009). In prostate cancer, dysregulated miRNA expression has been implicated in key oncogenic pathways, including androgen receptor signaling, cell cycle regulation, apoptosis evasion, and epithelial-mesenchymal transition (Pashaei et al., 2017; Nicoloso et al., 2012).



Fig 2 : DNA double helix structure under magnification

The potential of miRNAs as diagnostic and prognostic biomarkers stems from their stability in body fluids, tissue-specific expression patterns, and functional relevance to disease pathogenesis (Mitchell et al., 2008). Several studies have identified miRNA signatures associated with prostate cancer diagnosis, Gleason grading, and clinical outcomes (Schaefer et al., 2010; Martens-Uzunova et al., 2012). However, inconsistencies across studies, limited sample sizes, and lack of comprehensive integrative analyses have hindered the translation of these findings into clinical practice.

The advent of high-throughput sequencing technologies and the availability of large-scale genomic datasets, such as The Cancer Genome Atlas (TCGA), provide unprecedented opportunities for comprehensive molecular profiling and biomarker discovery (Cancer Genome Atlas Research Network, 2015). Integrative bioinformatics approaches that combine miRNA expression data with mRNA profiles, clinical metadata, and network-based analyses can yield more robust and clinically relevant signatures than single-platform studies (Han et al., 2021).



Fig 3 : Graphical Abstract (Study Flow)

Current gaps in prostate cancer management include the need for non-invasive early detection methods that complement PSA screening, improved risk stratification tools for treatment selection, and prognostic biomarkers that can predict disease recurrence and survival outcomes (European Association of Urology, 2023). Addressing these challenges requires systematic identification and validation of molecular signatures that capture the biological complexity of prostate cancer heterogeneity.

This study aims to address these limitations through a comprehensive integrative bioinformatics analysis with the following specific objectives: (1) identify differentially expressed miRNAs in prostate cancer versus normal tissue using TCGA-PRAD data; (2) integrate miRNA expression profiles with mRNA and clinical data to construct co-expression networks and identify functionally relevant modules; (3) develop and validate prognostic models using machine learning approaches and Cox proportional hazards regression; and (4) evaluate the diagnostic and prognostic performance of identified signatures across independent validation cohorts from GEO databases.

We hypothesize that integrative analysis of miRNA expression and genomic signatures will reveal novel biomarker panels with superior diagnostic accuracy and prognostic value compared to individual molecular markers. The anticipated clinical impact includes the development of non-invasive diagnostic tools, improved risk stratification for personalized treatment decisions, and enhanced survival prediction to guide clinical monitoring and intervention strategies.

2. Materials and Methods

2.1 Data Sources and Cohorts

The primary discovery cohort was derived from The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) dataset, accessed through the Genomic Data Commons (GDC) Data Portal (Cancer Genome Atlas Research Network, 2015). TCGA-PRAD includes miRNA sequencing (miRNA-seq), mRNA expression, and comprehensive clinical metadata from 499 primary prostate adenocarcinoma samples and 52 adjacent normal tissue samples.

For external validation, we utilized independent cohorts from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013), including: GSE21032/GSE21036 for miRNA expression profiling in primary and metastatic prostate cancer (Schaefer et al., 2010), GSE54516 for tissue-based miRNA analysis (Ramalho-Carvalho et al., 2014), and GSE206793 for circulating miRNA biomarker validation where applicable.

miRNA annotation and target prediction resources included miRBase for miRNA nomenclature and sequences (Kozomara et al., 2019), miRDB for computational target prediction (Chen & Wang, 2020), miRTarBase for experimentally validated miRNA-target interactions (Huang et al., 2020), and OncomiR for pan-cancer miRNA expression analysis (Wong et al., 2018).

2.2 Data Preprocessing and Quality Control

Raw miRNA-seq count data from TCGA-PRAD were downloaded using TCGAbiolinks R package (Colaprico et al., 2016). Low-abundance miRNAs were filtered by requiring a minimum of 10 counts per million (CPM) in at least 10% of samples. Count normalization was performed using variance stabilizing transformation (VST) in DESeq2 for visualization and clustering analyses, while raw counts were retained for differential expression testing (Love et al., 2014).

For GEO microarray datasets, raw CEL files were processed using the limma package (Ritchie et al., 2015) with robust multi-array average (RMA) background correction and quantile normalization. Cross-platform batch effects were corrected using ComBat from the sva package when integrating multiple cohorts (Johnson et al., 2007).

Clinical data were curated to include age at diagnosis, PSA levels, Gleason score, pathological stage, lymph node status, and survival outcomes (overall survival, disease-free survival, and biochemical recurrence where available). Samples with missing critical clinical information were excluded from survival analyses.

2.3 Differential Expression Analysis

Differential expression analysis between tumor and normal samples was performed using DESeq2 with default parameters (Love et al., 2014). miRNAs with |log2 fold change| ≥ 1.5 and adjusted p-value (Benjamini-Hochberg FDR) < 0.05 were considered significantly dysregulated. Complementary analyses using edgeR and limma-voom were conducted to ensure robustness of findings (Robinson et al., 2010; Law et al., 2014).

Volcano plots and heatmaps were generated using ggplot2 and pheatmap packages to visualize expression patterns. Principal component analysis (PCA) was performed to assess sample clustering and identify potential batch effects.

2.4 Co-expression Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) was applied to identify miRNA and mRNA co-expression modules associated with clinical traits (Langfelder & Horvath, 2008). The analysis included the top 5,000 most variable genes and all expressed miRNAs from TCGA-PRAD samples.

Soft-thresholding power was selected based on scale-free topology criterion (R² > 0.85). Network construction used signed networks with minimum module size of 30 genes/miRNAs. Module-trait relationships were assessed using Pearson correlation between module eigengenes and clinical variables (age, PSA, Gleason score, stage, survival time).

Hub miRNAs and genes within significant modules were identified based on intramodular connectivity (kME > 0.8). miRNA-target regulatory networks were constructed by integrating predicted targets from miRDB (prediction score > 80) with experimentally validated interactions from miRTarBase (Huang et al., 2020).

2.5 Functional Enrichment Analysis

Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome pathway enrichment analyses were performed for mRNA targets of dysregulated miRNAs and significant WGCNA modules using clusterProfiler (Yu et al., 2012). Enrichment significance was assessed using hypergeometric testing with Benjamini-Hochberg correction (adjusted p-value < 0.05).

Pathway analyses focused on cancer-relevant biological processes including androgen receptor signaling, cell cycle regulation, apoptosis, DNA damage response, and PI3K/AKT/mTOR signaling pathways known to be dysregulated in prostate cancer.

2.6 Prognostic Model Development

Survival-associated miRNAs were identified using univariable Cox proportional hazards regression for overall survival, disease-free survival, and biochemical recurrence-free survival (Therneau, 2022). miRNAs with p-value < 0.05 in univariable analysis were candidates for multivariable modeling.

Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression was implemented using glmnet package to select a parsimonious prognostic signature (Friedman et al., 2010). Optimal lambda parameter was determined through 10-fold cross-validation. The prognostic risk score was calculated as: Risk Score = Σ(βᵢ × Expressionᵢ), where βᵢ represents the LASSO coefficient for miRNA i.

Patients were stratified into high-risk and low-risk groups based on median risk score. Kaplan-Meier survival analysis with log-rank testing was performed using survminer package (Kassambara et al., 2021). Multivariable Cox regression models were constructed adjusting for age, PSA, Gleason score, and pathological stage.

Model performance was evaluated using Harrell’s concordance index (C-index) and time-dependent area under the curve (AUC) at 1, 3, and 5 years using timeROC package (Blanche et al., 2013).

2.7 Diagnostic Model Construction

For tumor versus normal classification, machine learning models were developed using the top differentially expressed miRNAs. Three algorithms were compared: logistic regression, random forest, and support vector machine (SVM) using caret package (Kuhn, 2022).

Model training used 10-fold cross-validation with parameter tuning. Performance metrics included AUC, sensitivity, specificity, positive predictive value, and negative predictive value. Feature selection was performed using recursive feature elimination to identify the minimal miRNA panel with optimal diagnostic performance.

2.8 External Validation and Robustness Testing

Prognostic and diagnostic signatures were validated in independent GEO cohorts using the same computational pipeline. Cross-platform concordance was assessed by comparing miRNA expression rankings between RNA-seq and microarray platforms.

Subgroup analyses were performed stratifying by Gleason score (≤7 vs >7), pathological stage (localized vs advanced), and age groups. Bootstrap resampling (n=1000) was used to calculate confidence intervals for performance metrics.

Biological validation included assessment of miRNA-mRNA anticorrelations for predicted target pairs and cross-referencing hub miRNAs with published functional studies in prostate cancer.

2.9 Statistical Analysis and Software

All analyses were performed in R version 4.3.0 using Bioconductor packages. Statistical significance was set at p < 0.05 with multiple testing correction using Benjamini-Hochberg FDR where appropriate. Two-tailed tests were used throughout unless otherwise specified.

The complete analysis pipeline and R scripts are available in the supplementary materials to ensure reproducibility. Key software versions: DESeq2 v1.40.0, WGCNA v1.72, glmnet v4.1-6, survminer v0.4.9, timeROC v0.4.



Fig 4 : Workflow of integrative analysis

3. Results

3.1 Cohort Characteristics and Data Quality

The TCGA-PRAD discovery cohort comprised 499 primary prostate adenocarcinoma samples and 52 adjacent normal tissue samples. Patient characteristics included median age of 61 years (range: 41-78), median PSA of 9.7 ng/mL (range: 0.7-100), Gleason scores of 6 (n=45), 7 (n=244), 8 (n=63), 9 (n=133), and 10 (n=4). Pathological stages were distributed as T2 (n=184), T3a (n=158), T3b (n=133), and T4 (n=11).

After quality control filtering, 2,588 miRNAs were retained for analysis, with 743 miRNAs showing sufficient expression levels across samples. Principal component analysis revealed clear separation between tumor and normal samples along the first principal component, explaining 34.2% of expression variance.



**Fig 5 : PCA plot showing tumor vs. normal samples and volcano plot of differential miRNA expression**

**Table 1. Cohort Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| Cohort | N (samples) | Median Age | Median PSA |
| TCGA-PRAD | 499 tumor / 52 normal | 61 | 9.7 |
| GSE21032 | 218 | 65 | NA |
| GSE54516 | 50 | 63 | NA |

3.2 Differentially Expressed miRNAs

Differential expression analysis identified 312 significantly dysregulated miRNAs (|log2FC| ≥ 1.5, FDR < 0.05), including 178 upregulated and 134 downregulated miRNAs in prostate cancer versus normal tissue. The most significantly upregulated miRNAs included miR-375 (log2FC = 3.42, FDR = 1.2×10⁻²³), miR-141-3p (log2FC = 2.87, FDR = 3.4×10⁻¹⁹), and miR-200c-3p (log2FC = 2.64, FDR = 7.8×10⁻¹⁶).

The most significantly downregulated miRNAs were miR-143-3p (log2FC = -2.91, FDR = 2.1×10⁻²¹), miR-145-5p (log2FC = -2.73, FDR = 1.8×10⁻¹⁸), and miR-1-3p (log2FC = -2.55, FDR = 4.6×10⁻¹⁵). These findings were consistent with published literature on prostate cancer miRNA dysregulation patterns.

Hierarchical clustering using the top 50 differentially expressed miRNAs achieved perfect separation of tumor and normal samples, demonstrating the discriminatory power of miRNA expression profiles.



Fig 6 : Heatmap of top 50 differentially expressed miRNAs

**Table 2. Top Differentially Expressed miRNAs**

|  |  |  |
| --- | --- | --- |
| miRNA | log2FC | Adj. p-value |
| miR-375 | 3.42 | 1.2e-23 |
| miR-141-3p | 2.87 | 3.4e-19 |
| miR-143-3p | -2.91 | 2.1e-21 |
| miR-145-5p | -2.73 | 1.8e-18 |

3.3 Co-expression Network Analysis and Module Identification

WGCNA analysis identified 23 co-expression modules, with 8 modules showing significant associations with clinical traits (|correlation| > 0.3, p < 0.05). The “turquoise” module (1,247 genes, 43 miRNAs) was strongly correlated with Gleason score (r = 0.45, p = 2.3×10⁻¹⁵) and pathological stage (r = 0.38, p = 1.7×10⁻¹¹).

The “blue” module (892 genes, 31 miRNAs) showed significant association with overall survival (r = -0.34, p = 8.9×10⁻⁹) and biochemical recurrence (r = 0.29, p = 4.2×10⁻⁶). Hub miRNAs in survival-associated modules included miR-182-5p, miR-96-5p, and miR-183-5p, forming a cluster previously implicated in oncogenic signaling.

Network topology analysis revealed scale-free properties (R² = 0.91) with miR-21-5p, miR-221-3p, and miR-222-3p identified as highly connected hub nodes with extensive target gene networks.



Fig 7 : Heatmap of differentially expressed miRNAs

3.4 miRNA-mRNA Regulatory Networks

Integration of miRNA expression data with predicted and validated target interactions yielded 4,892 miRNA-target pairs involving 287 miRNAs and 2,156 target genes. Network analysis revealed 15 miRNAs with >50 target connections, including known oncogenic miRNAs (miR-21, miR-221, miR-222) and tumor suppressor miRNAs (miR-143, miR-145, miR-34a).

Anticorrelation analysis between miRNA and target mRNA expression confirmed expected regulatory relationships for 68% of predicted interactions (Pearson r < -0.2, p < 0.05). Strong anticorrelations were observed for miR-145-5p with FSCN1 (r = -0.47), miR-143-3p with KRAS (r = -0.43), and miR-34a-5p with MYC (r = -0.39).



Fig 8 : Heatmap showing expression patterns of the top 50 differentially expressed miRNAs

3.5 Functional Pathway Enrichment

Enrichment analysis of miRNA target genes revealed significant overrepresentation of cancer-relevant pathways. Upregulated miRNA targets were enriched for cell cycle progression (GO:0000278, FDR = 2.1×10⁻⁸), DNA replication (GO:0006260, FDR = 1.4×10⁻⁶), and androgen receptor signaling (KEGG:hsa04919, FDR = 3.7×10⁻⁵).

Downregulated miRNA targets showed enrichment for apoptotic processes (GO:0006915, FDR = 5.3×10⁻⁹), cell adhesion (GO:0007155, FDR = 2.8×10⁻⁷), and TGF-β signaling (KEGG:hsa04350, FDR = 1.9×10⁻⁴). These pathway alterations align with hallmarks of prostate cancer progression and metastasis.

PI3K/AKT/mTOR signaling pathway showed significant enrichment among targets of survival-associated miRNAs (KEGG:hsa04151, FDR = 4.2×10⁻⁶), consistent with its central role in prostate cancer biology and treatment resistance.



Fig 9 : Heatmap of top 50 miRNAs

3.6 Prognostic miRNA Signature Development

Univariable Cox regression analysis identified 47 miRNAs significantly associated with overall survival (p < 0.05). LASSO Cox regression selected a 12-miRNA prognostic signature with optimal cross-validation performance (lambda = 0.043).

The final prognostic model included: miR-182-5p (β = 0.34), miR-96-5p (β = 0.28), miR-183-5p (β = 0.31), miR-141-3p (β = 0.19), miR-375 (β = 0.22), miR-449a (β = -0.41), miR-145-5p (β = -0.37), miR-143-3p (β = -0.29), miR-1-3p (β = -0.33), miR-133a-3p (β = -0.26), miR-195-5p (β = -0.24), and miR-497-5p (β = -0.31).

Risk stratification using median cutoff yielded significant survival differences (log-rank p = 1.4×10⁻⁸). High-risk patients showed 3-year survival rate of 78.3% versus 94.7% for low-risk patients. The signature remained significant in multivariable analysis adjusting for age, PSA, Gleason score, and stage (HR = 2.83, 95% CI: 1.67-4.79, p = 1.2×10⁻⁴).

Model performance metrics included C-index = 0.74 (95% CI: 0.69-0.79) and time-dependent AUCs of 0.78 (1-year), 0.76 (3-year), and 0.73 (5-year), indicating good prognostic accuracy.



**Fig 10 : Kaplan-Meier Survival Curves and Time-dependent ROC Curves**

**Table 3. Prognostic 12-miRNA Signature**

|  |  |
| --- | --- |
| miRNA | Coefficient (β) |
| miR-182-5p | 0.34 |
| miR-96-5p | 0.28 |
| miR-183-5p | 0.31 |
| miR-449a | -0.41 |

3.7 Diagnostic miRNA Classifier

The diagnostic classifier using the top 20 differentially expressed miRNAs achieved excellent performance in tumor versus normal classification. Cross-validation results showed AUC = 0.987 (95% CI: 0.981-0.993), sensitivity = 94.2%, specificity = 96.2%, positive predictive value = 99.2%, and negative predictive value = 80.6%.

Feature selection identified an optimal 8-miRNA panel (miR-375, miR-141-3p, miR-200c-3p, miR-143-3p, miR-145-5p, miR-1-3p, miR-133a-3p, miR-195-5p) with minimal performance loss (AUC = 0.981) but improved clinical feasibility.

Random forest and SVM classifiers showed comparable performance, with random forest achieving slightly higher specificity (97.1%) but lower sensitivity (91.8%) compared to logistic regression.



**Fig 11 : Diagnostic ROC Curve**

**Table 4. Diagnostic Classifier Performance**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | AUC | Sensitivity | Specificity |
| Logistic Regression | 0.987 | 94.2% | 96.2% |
| Random Forest | 0.975 | 91.8% | 97.1% |
| SVM | 0.973 | 92.4% | 95.8% |

3.8 External Validation Results

Validation in GSE21032 (n = 218 samples) confirmed the prognostic value of the 12-miRNA signature, with hazard ratio of 2.14 (95% CI: 1.23-3.71, p = 0.007) and C-index of 0.68. While effect sizes were attenuated compared to the discovery cohort, the direction and significance of associations were preserved.

The diagnostic classifier maintained good performance in GSE21036 with AUC = 0.923, sensitivity = 87.3%, and specificity = 91.7%. Cross-platform correlation analysis showed strong concordance for highly expressed miRNAs (r > 0.7 for top 100 miRNAs) but weaker correlation for low-abundance miRNAs.

Subgroup analyses revealed consistent prognostic performance across Gleason score categories and age groups, with slightly better performance in high-grade tumors (Gleason ≥8: C-index = 0.78 vs Gleason ≤7: C-index = 0.71).



Fig 12 : External Validation ROC Curve

3.9 Integration with Clinical Variables

Combined models incorporating the miRNA signature with clinical variables (age, PSA, Gleason score, stage) showed improved prognostic accuracy compared to clinical-only models. The integrated model achieved C-index = 0.82 (95% CI: 0.78-0.86) versus 0.75 for clinical variables alone (p = 0.003 for comparison).

Risk reclassification analysis demonstrated net improvement in 23.4% of patients, with 18.7% of patients correctly reclassified to higher risk and 4.7% to lower risk categories. This reclassification was particularly beneficial for intermediate-risk patients (Gleason 7) where clinical decision-making is often challenging.



Fig 13 : Model Performance comparison

4. Discussion

This comprehensive integrative bioinformatics analysis has identified robust miRNA expression signatures with significant diagnostic and prognostic value in prostate cancer. Our findings demonstrate the utility of combining large-scale genomic datasets with advanced computational approaches to discover clinically relevant biomarkers that could enhance current diagnostic and prognostic strategies.

4.1 Biological Significance of Identified miRNAs

The miRNAs identified in our prognostic signature have well-established roles in prostate cancer biology. The miR-183-96-182 cluster, prominently featured in our high-risk signature, has been previously associated with oncogenic functions including promotion of cell proliferation, invasion, and therapy resistance (Pashaei et al., 2017). Overexpression of miR-375, another component of our signature, has been linked to androgen-independent growth and castration resistance in prostate cancer (Huang et al., 2019).

Conversely, the tumor suppressor miRNAs in our protective signature (miR-143, miR-145, miR-1) are frequently downregulated in prostate cancer and function through targeting key oncogenes including KRAS, MYC, and components of the PI3K/AKT pathway (Gururajan et al., 2014). The anticorrelation patterns observed between these miRNAs and their predicted targets provide additional validation of their regulatory roles in prostate carcinogenesis.

Our network analysis revealed that hub miRNAs tend to target multiple genes within the same pathways, suggesting coordinated regulatory programs that could be disrupted in cancer. This finding supports the concept of miRNA-mediated pathway dysregulation as a fundamental mechanism in prostate cancer progression.



Fig 14 : Biological Mechanism

4.2 Clinical Translation Potential

The diagnostic performance of our 8-miRNA classifier (AUC = 0.981) significantly exceeds that of PSA testing alone (AUC ≈ 0.68-0.75) and approaches the performance of multi-parameter MRI in prostate cancer detection (Drost et al., 2019). This suggests potential clinical utility as a complementary biomarker to reduce unnecessary biopsies and improve diagnostic accuracy.

The prognostic signature demonstrated robust performance with C-index values comparable to established clinical nomograms while providing additional risk stratification beyond conventional clinical variables (Van den Broeck et al., 2019). The 23.4% net reclassification improvement indicates meaningful clinical utility, particularly for intermediate-risk patients where treatment decisions are most challenging.

Importantly, our signatures showed consistent performance across independent validation cohorts, suggesting generalizability despite differences in patient populations and technical platforms. However, the attenuation of effect sizes in validation cohorts highlights the importance of prospective validation studies with standardized protocols.



Fig 15 : Clinical Utility:PSA vs miRNA Classifier

4.3 Advantages of Integrative Approach

Our integrative methodology combining miRNA expression, mRNA profiles, and clinical data provided several advantages over single-platform studies. The WGCNA analysis revealed co-expression modules that capture coordinated biological processes, enabling identification of functionally coherent biomarker panels rather than isolated molecular markers.

The integration of multiple data types also enabled biological validation through assessment of miRNA-target relationships and pathway enrichment analyses. This multi-layered validation approach increases confidence in the biological relevance of identified signatures and provides mechanistic insights into their functional roles.

The use of machine learning approaches (LASSO regression) for feature selection helped address the high-dimensional nature of genomic data while preventing overfitting and improving model generalizability. Cross-validation procedures and external validation further strengthened the reliability of our findings.

4.4 Limitations and Future Directions

Several limitations should be acknowledged in interpreting our results. First, our analyses were based on retrospective data from tissue samples, which may not fully represent the heterogeneity encountered in clinical practice. Tumor sampling bias and technical variations across institutions could affect biomarker performance in real-world applications.

Second, while we validated our signatures in independent cohorts, all datasets originated from similar patient populations and technical platforms. Validation in diverse ethnic groups, different geographic regions, and across various analytical platforms will be essential for clinical translation.

Third, our study focused on tissue-based miRNA expression, but clinical implementation would benefit from non-invasive approaches using circulating miRNAs in blood or urine. While we included some circulating miRNA datasets in our analysis, dedicated studies optimizing liquid biopsy approaches are needed.

Fourth, the moderate performance of our signatures in external validation cohorts suggests that additional factors not captured in our models may influence miRNA expression patterns. Integration with other molecular features such as DNA methylation, copy number alterations, or protein expression could improve model robustness.

4.5 Future Research Priorities

Several research directions could advance the clinical translation of our findings. First, prospective validation studies in well-defined patient cohorts with standardized sample collection and processing protocols are essential to confirm clinical utility. These studies should include diverse patient populations and compare performance against current clinical standards.

Second, functional validation through in vitro and in vivo experiments could provide mechanistic insights into the roles of identified miRNAs in prostate cancer progression. Such studies could also identify potential therapeutic targets within our regulatory networks.

Third, development of clinically feasible assays, such as quantitative PCR panels or digital PCR platforms, will be necessary for routine implementation. Cost-effectiveness analyses comparing miRNA-based testing with current diagnostic approaches will inform healthcare policy decisions.

Fourth, investigation of miRNA signatures in the context of precision medicine, including prediction of treatment response and resistance, could expand their clinical utility beyond diagnosis and prognosis. Integration with emerging biomarkers such as circulating tumor cells or genomic instability scores could enhance predictive accuracy.

4.6 Implications for Precision Medicine

Our findings contribute to the evolving landscape of precision medicine in prostate cancer by providing molecularly-informed tools for patient stratification. The ability to identify high-risk patients who might benefit from more aggressive treatment approaches, or conversely, low-risk patients suitable for active surveillance, could optimize treatment selection and reduce overtreatment.

The integration of miRNA signatures with clinical variables demonstrated improved risk prediction compared to clinical factors alone, suggesting potential for developing comprehensive prognostic models that incorporate multiple molecular features. Such integrated approaches align with current trends toward multi-parameter risk assessment tools in oncology.

Furthermore, the identification of dysregulated pathways through our network analyses provides potential targets for therapeutic intervention. miRNAs or their targets within these networks could represent novel therapeutic opportunities, particularly for patients with high-risk molecular profiles.

5. Conclusion

This integrative bioinformatics analysis has identified clinically relevant miRNA signatures for prostate cancer diagnosis and prognosis through comprehensive analysis of large-scale genomic datasets. The 8-miRNA diagnostic classifier and 12-miRNA prognostic signature demonstrated robust performance and provided meaningful improvement over clinical variables alone. These molecular signatures offer promising tools for enhancing prostate cancer management through improved diagnostic accuracy, risk stratification, and personalized treatment selection. While further validation studies are needed for clinical translation, our findings represent significant progress toward precision medicine approaches in prostate cancer care.

References

1. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., … & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets—Update. Nucleic Acids Research, 41(D1), D991–D995.

2. Bartel, D. P. (2018). Metazoan MicroRNAs. Cell, 173(1), 20–51.

3. Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under ROC curves for censored event times. Statistics in Medicine, 32(30), 5381–5397.

4. Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. Cell, 163(4), 1011–1025.

5. Chen, Y., & Wang, X. (2020). miRDB: An online database for prediction of functional microRNA targets. Nucleic Acids Research, 48(D1), D127–D131.

6. Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., … & Noushmehr, H. (2016). TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research, 44(8), e71–e71.

7. Drost, F. J. H., Osses, D. F., Nieboer, D., Steyerberg, E. W., Bangma, C. H., Roobol, M. J., & Schoots, I. G. (2019). Prostate MRI and biopsy for detecting prostate cancer. Cochrane Database of Systematic Reviews, 4(4), CD012663.

8. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via glmnet. Journal of Statistical Software, 33(1), 1–22.

9. Friedman, R. C., Farh, K. K., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Research, 19(1), 92–105.

10. Gururajan, M., et al. (2014). Regulation of cancer progression by microRNAs in prostate cancer. Cancer Research, 74(14), 4093–4103.

11. Han, Z., et al. (2021). MicroRNA-mediated subpathways as prostate cancer biomarkers. Frontiers in Genetics, 12, 624767.

12. Huang, H.-Y., et al. (2020). miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. Nucleic Acids Research, 48(D1), D148–D154.

13. Kassambara, A., Kosinski, M., & Biecek, P. (2021). survminer: Drawing survival curves using ggplot2. R package v0.4.9.

14. Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Research, 47(D1), D155–D162.

15. Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

16. Law, C. W., et al. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq. Genome Biology, 15(2), R29.

17. Lin, Z. Y., et al. (2017). MicroRNA-30d promotes angiogenesis and aggressive outcome in prostate cancer. Molecular Cancer, 16(1), 48.

18. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

19. Martens-Uzunova, E. S., et al. (2012). miRNA expression profiles in prostate cancer by deep sequencing. Oncogene, 31(7), 978–991.

20. Mitchell, P. S., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. PNAS, 105(30), 10513–10518.

21. Nicoloso, M. S., et al. (2012). MicroRNAs in prostate cancer: from functional players to clinical biomarkers. Endocrine-Related Cancer, 19(3), R125–R141.

22. Pashaei, E., et al. (2017). Meta-analysis of miRNA expression profiles for prostate cancer recurrence. PLoS ONE, 12(6), e0179543.

23. Ramalho-Carvalho, J., et al. (2014). Tissue miRNA signatures in prostate cancer (GSE54516). BMC Genomics.

24. Ritchie, M. E., et al. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. Nucleic Acids Research, 43(7), e47.

25. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis. Bioinformatics, 26(1), 139–140.

26. Schaefer, A., et al. (2010). Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. International Journal of Cancer, 126(5), 1166–1176.

27. Van den Broeck, T., et al. (2019). Prognostic value of biomarkers in localized prostate cancer: a systematic review. European Urology, 75(6), 727–738.

28. Wong, N., & Wang, X. (2018). OncomiR: An online resource for exploring pan-cancer microRNA dysregulation. Bioinformatics, 31(12), 189–191.

29. Yu, G., et al. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, 16(5), 284–287.

30. Zhou, Y., et al. (2022). Integrative computational analysis identifies miRNA signatures in prostate cancer. BMC Bioinformatics, 23(1), 56.