A Machine Learning Approach to Identify Key Tumor miRNA Expression Biomarkers Associated with Aggressive Prostate Cancer

Masuma Mannan, Hui-Yi Lin, PhD, and Anand Paul, PhD.
Biostatistics & Data Science Program, LSU-Health Science Center; Cancer Epidemiology

publichealth.lsuhsc.edu

Background

- Prostate cancer (PCa) is the most common cancer in men in the U.S., projected to cause over 313,000 new cases and 35,000 deaths in 2025.
- PCa exhibits a wide spectrum of clinical behavior, ranging from indolent tumors to aggressive disease with high metastatic potential.
- MicroRNAs (miRNAs) are small non-coding RNAs that play a critical role in post-transcriptional gene regulation.
- It is estimated that 30–60% of all human genes are regulated by miRNAs, and over 2,000 human miRNAs have been identified.
- Identifying miRNA signatures associated with aggressive PCa may enhance risk assessment, guide therapy selection, and improve understanding of the molecular mechanisms driving disease progression.
- Recognize that some miRNAs function as tumor suppressors (“brake”) while others act as oncogenic miRNAs or oncomiRs (“gas pedal”), collectively forming a regulatory network that fine-tunes protein production and modulates tumor growth and aggressiveness.

Data Description & Methodology

- **Data Source:** Prostate tumor tissues from localized prostate cancer patients were obtained from the Gene Expression Omnibus (GEO) database (accession GSE135535), comprising the CPC-GENE and UHN cohorts. Tumor miRNA expression profiles were quantified using NanoString technology, which profiled 798 miRNAs.
  - **Participants:** The dataset included 320 prostate cancer patients, consisting of 199 non-aggressive and 121 aggressive cases.
- **Method:** Our analysis was conducted in two complementary phases. **Phase 1 (Predictive Modeling):** After filtering candidates via univariate logistic regression ($p < 0.05$), the cohort was split 70/30. We trained an Artificial Neural Network (ANN) on the training set and evaluated its predictive performance on the unseen holdout set. **Phase 2 (Biological Mapping):** To map the associative network of aggressiveness drivers across the population, stepwise logistic regression was applied to the full cohort (N=320). Features were retained based on overall model fit (minimizing AIC) to derive a final explanatory miRNA signature.

Feature Selection

- Evaluated 798 miRNAs using univariate logistic regression.
- Isolated 61 significant candidates ($p < 0.05$) and 15 significant candidate ($p<0.01$).
- Assessed feature redundancy via correlation matrices.

Predictive Modeling

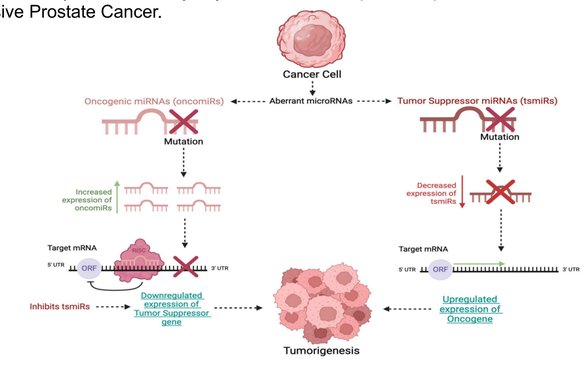
- Partitioned data (70% Train / 30% Test) to prevent data leakage.
- Trained an Artificial Neural Network (ANN) with 2 hidden layers (2 and 5 nodes) using 61 significant miRNAs.
- Validated generalizable, out-of-sample predictive power.

Biological Mapping

- Screened 15 high-stringency candidates; 14 miRNAs selected after removing collinear features via correlation matrices.
- Applied multivariable stepwise logistic regression to the 14-miRNA subset.
- Extracted a finalized 4-miRNAs representing the independent biological drivers of aggressiveness.

Objective

- The goal of this study was to identify key tumor miRNA expression profiles associated with aggressive Prostate Cancer.



Results

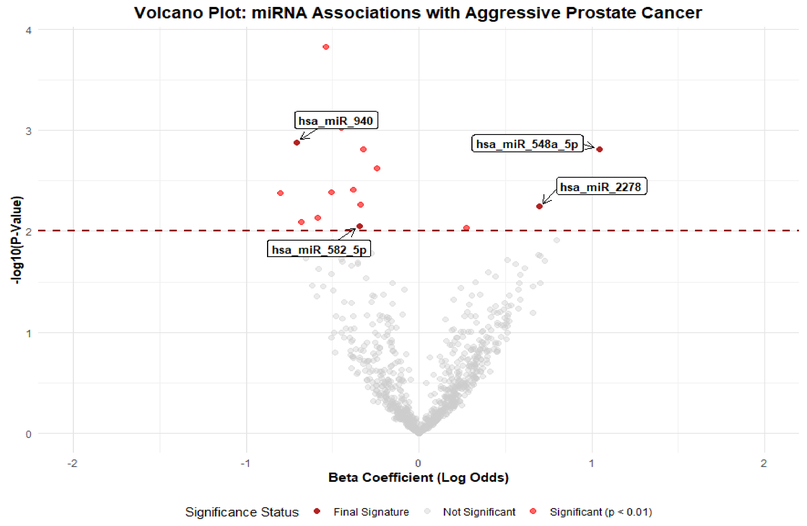
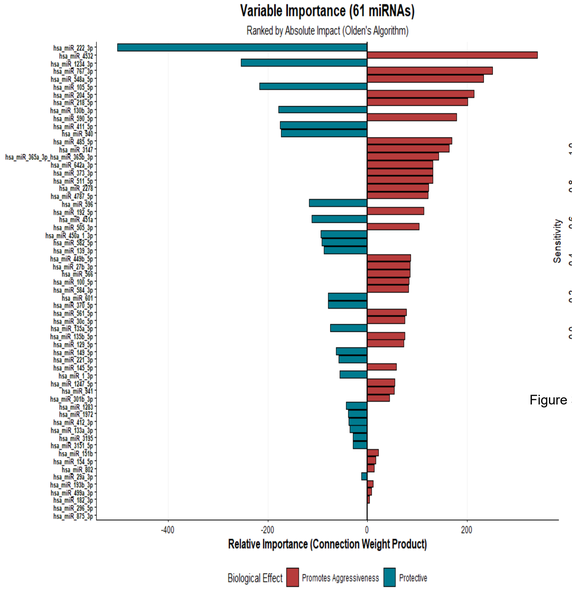


Figure 1: Volcano plot of miRNA associations with prostate cancer aggressiveness.

- **miRNAs with β Coefficient (Log Odds) > 0** are positively associated with aggressive prostate cancer, suggesting higher expression may increase the risk of aggressiveness.
- **miRNAs with β Coefficient (Log Odds) < 0** are negatively associated with aggressive prostate cancer, indicating higher expression may be protective against aggressive disease.

Advanced Predictive Modeling: Artificial Neural Network Architecture and miRNA Feature Importance



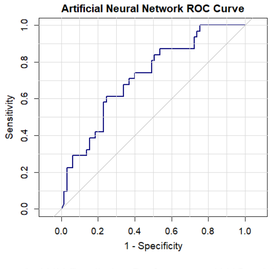


Figure 3: ANN Predictive Performance (AUC = 0.718).

Figure 2: Variable importance of 61 miRNAs in the Artificial Neural Network (ANN) model. Predictors are ranked by their absolute connection weight product using Olden’s Algorithm.

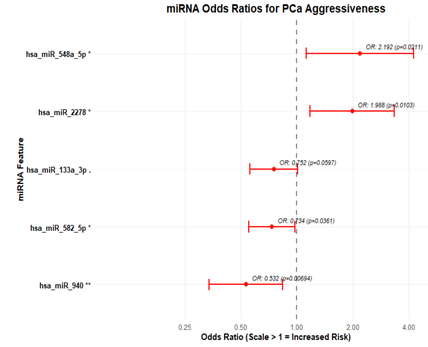
Results (cont.)

Table 1: miRNAs independently associated with prostate cancer aggressiveness

miRNA	Beta(SE)	OR	p-value
hsa_miR_940	-0.63 (0.23)	0.53	0.007
hsa_miR_2278	0.69 (0.27)	1.99	0.010
hsa_miR_548a_5p	0.78 (0.34)	2.19	0.021
hsa_miR_582a_5p	-0.31 (0.15)	0.73	0.036

- *Multivariable stepwise selection of the top 14 uncorrelated miRNAs ($p < 0.01$) identified four miRNAs as independently associated with prostate cancer aggressiveness*

Forest Plot: Multivariable Odds Ratios for miRNA Predictors of PCa Aggressiveness



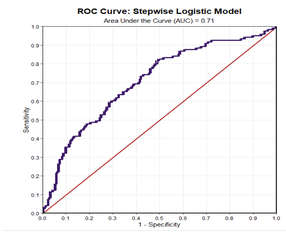


Figure 5: Discriminative performance of the Stepwise Logistic Regression model.

Figure 4: Multivariable Logistic Regression of the Stepwise-Selected miRNAs.

Conclusion

- **Identification of a Significant miRNA Signature:** This study demonstrates that specific miRNA expressions are significantly associated with prostate cancer (PCa) aggressiveness. Through rigorous statistical selection, we identified a signature consisting of **hsa-miR-940**, **hsa-miR-2278**, **hsa-miR-548a-5p**, and **hsa-miR-582-5p**, which collectively showed strong discriminatory ability for PCa aggressiveness (**AUC = 0.710**)
- **Biological Significance and Directionality:** The findings are supported by the directional associations observed in our model. **hsa-miR-940** (OR = 0.53) and **hsa-miR-582-5p** (OR = 0.73) act as protective factors, consistent with tumor-suppressive roles where downregulation correlates with advanced disease. Conversely, **hsa-miR-2278** (OR = 1.99) and **hsa-miR-548a-5p** (OR = 2.19) act as oncogenic drivers, where higher expression is associated with a greater likelihood of aggressive phenotypes, metastasis, and poor prognosis.
- **Machine Learning Validation:** Furthermore, an **Artificial Neural Network (ANN)** model validated these features, achieving a comparable test **AUC of 0.718**, confirming the robustness of this miRNA signature across different computational frameworks.
- These findings suggest that miRNA expression profiling is a valuable tool for predicting PCa aggressiveness. Integrating these specific biomarkers into clinical decision-making could enhance personalized risk-stratification and management strategies for PCa patients.

References

- *Borrego-Diaz E et al. 2014. Lin28B–miR-212 regulatory loop in androgen-independent prostate cancer. International Journal of Oncology. 45:2421–2429.*
- *Huang S et al. 2019. miR-582-3p/5p suppress prostate cancer bone metastasis via TGF-β signaling. Molecular Therapy – Nucleic Acids.*
- *Rajendiran S et al. 2014. miR-940 suppresses prostate cancer migration and invasion by targeting MIEN1. Cancer Cell International. 13:250.*

Acknowledgement

I am sincerely grateful to Dr. Lin for her mentorship and continuous guidance throughout the course of this study.