MicroRNA-Based Machine Learning Classifier Accurately Predicts Molecular Subtypes in Endometrial Carcinoma

Keywords
prognosis, endometrial carcinoma, machine learning, molecular subtypes, microrna (mirna)

Abstract

Introduction

Molecular classification of endometrial carcinoma (EC) has revolutionized prognostic stratification. MicroRNAs (miRNAs) offer a potential cost-effective alternative for molecular subtyping. This study presents a proof-of-concept computational analysis to evaluate the feasibility of a miRNA-based machine learning classifier for discriminating the four distinct EC molecular subtypes.

Material and methods

We analyzed 232 EC samples from the TCGA-UCEC cohort with complete molecular annotations. To confirm the clinical relevance of the ground-truth labels used for modeling, survival analyses (Kaplan-Meier and cumulative incidence function) were first performed on the established TCGA subtypes. Feature selection was conducted exclusively on the training set, identifying a signature of 20 differentially expressed miRNAs. A benchmark of seven machine learning algorithms was conducted. A pilot external validation was performed on 15 independent clinical samples using qRT-PCR, with ground-truth subtypes confirmed via POLE sequencing and IHC (ProMisE criteria).

Results

Survival analysis confirmed that the ground-truth TCGA subtypes exhibited statistically significant prognostic differences (p=0.0011 for OS). In the computational benchmark, the LASSO logistic regression model demonstrated superior performance on the independent test set (multiclass AUC = 0.850). In the pilot external validation cohort, the classifier achieved an accuracy of 86.7% (95% CI: 59.5%–98.3%), correctly identifying all POLE and CN-high cases.

Conclusions

This study demonstrates the computational feasibility of using a 20-miRNA signature to classify EC molecular subtypes, particularly for the prognostically distinct POLE and CN-high groups. While the survival analysis reaffirms the prognostic value of the molecular subtypes themselves, our findings regarding the classifier represent a preliminary proof-of-concept.

MicroRNA-Based Machine Learning Classifier Accurately Predicts Molecular Subtypes in Endometrial Carcinoma

Gedan Liu[1], Sifeng Zhou[2], Yi Miao[1*]

Affiliations:

[1] Department of Gynecology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

[2] School of Clinical Medicine, Medical College of Yangzhou University, Yangzhou 225009, P. R. China

* Corresponding author: Yi Miao, MD, Email: miaoyi1980@yeah.net

Abstract

Background: Molecular classification of endometrial carcinoma (EC) has revolutionized prognostic stratification, yet the standard TCGA-based integrated genomic profiling remains resource-intensive and complex for routine clinical implementation. MicroRNAs (miRNAs) offer a potential cost-effective alternative for molecular subtyping. This study presents a proof-of-concept computational analysis to evaluate the feasibility of a miRNA-based machine learning classifier for discriminating the four distinct EC molecular subtypes: POLE ultramutated, microsatellite instability (MSI) hypermutated, copy-number low (CN-low), and copy-number high (CN-high).

Methods: We analyzed 232 EC samples from the TCGA-UCEC cohort with complete molecular annotations. To confirm the clinical relevance of the ground-truth labels used for modeling, survival analyses (Kaplan-Meier and cumulative incidence function) were first performed on the established TCGA subtypes. For classifier development, the dataset was strictly partitioned into training (80%) and independent test (20%) sets to prevent data leakage. Feature selection was conducted exclusively on the training set, identifying a signature of 20 differentially expressed miRNAs. A benchmark of seven machine learning algorithms—including Random Forest, SVM, XGBoost, and LASSO—was conducted. A pilot external validation was performed on 15 independent clinical samples using qRT-PCR, with ground-truth subtypes confirmed via POLE sequencing and IHC (ProMisE criteria).

Results: Survival analysis confirmed that the ground-truth TCGA subtypes exhibited statistically significant prognostic differences (p=0.0011 for OS), with POLE ultramutated and CN-high subtypes showing the most favorable and poorest outcomes, respectively. In the computational benchmark, the LASSO logistic regression model demonstrated superior performance on the independent test set (multiclass AUC = 0.850). The model showed high discriminative ability for POLE (AUC = 0.889) and CN-high (AUC = 0.875) subtypes, though MSI stratification was more challenging (AUC = 0.756). In the pilot external validation cohort, the classifier achieved an accuracy of 86.7% (95% CI: 59.5%–98.3%), correctly identifying all POLE and CN-high cases.

Conclusion: This study demonstrates the computational feasibility of using a 20-miRNA signature to classify EC molecular subtypes, particularly for the prognostically distinct POLE and CN-high groups. While the survival analysis reaffirms the prognostic value of the molecular subtypes themselves, our findings regarding the classifier represent a preliminary proof-of-concept. The results suggest that miRNA profiling warrants further investigation as a complementary tool,

though large-scale prospective studies are required to establish clinical utility.

Keywords: Endometrial Carcinoma; Molecular Subtypes; MicroRNA; Machine Learning; Pilot Feasibility Study.

1. Introduction

Endometrial carcinoma (EC), originating from the epithelial lining of the uterine corpus, represents the most common malignancy of the female reproductive tract in developed countries and the second most prevalent gynecological cancer globally after cervical cancer(Concin et al., 2025; Dou et al., 2020). Over the past several decades, the global burden of EC has exhibited a steady upward trend, closely linked to changes in demographic profiles, lifestyle factors, and increasing prevalence of metabolic disorders(Zhao, Hughes, Altaf-Haroon, Schiller, & Kadambi, 2021). According to the Global Cancer Observatory (GLOBOCAN) 2020 database, EC accounted for approximately 417,000 new cases and 97,000 cancer-related deaths worldwide, with the highest incidence rates reported in North America and Northern and Eastern Europe(Elmadani et al., 2025; Filho et al., 2025). In contrast, incidence rates remain relatively lower in developing regions such as sub-Saharan Africa and Southeast Asia. However, a concerning pattern of rising incidence and mortality has recently emerged in low- and middle-income countries, likely reflecting urbanization, increased life expectancy, and westernization of diet and lifestyle(Bray et al., 2024). The global disease burden of EC is not only reflected in its rising incidence but also in its clinical complexity and evolving mortality trends. While most cases are diagnosed at an early stage and are generally associated with favorable prognoses, a significant subset of patients—particularly those with high-grade tumors or non-endometrioid histologies—exhibit aggressive clinical behavior and poor survival outcomes(Hong & Ding, 2025; Whitaker, 2020). Epidemiologically, EC is often stratified into two major clinical types: type I, which includes low-grade endometrioid carcinomas and is typically estrogen-dependent, and type II, which encompasses serous, clear cell, and high-grade endometrioid carcinomas that tend to be estrogen-independent and more aggressive(Obon-Santacana et al., 2014; Uccella et al., 2013). Type II tumors account for a disproportionate share of EC-related deaths despite being less common, underscoring the limitations of traditional histopathological classifications in capturing the underlying molecular heterogeneity of this disease.

Beyond its epidemiological and clinical significance, EC presents notable challenges in precision medicine(Chen, Chen, Chen, Xiao, & Huang, 2025). Current diagnostic and prognostic stratification largely rely on histological subtype, tumor grade, and FIGO stage, which offer limited insight into tumor biology and frequently fail to predict clinical behavior or therapeutic response. In recent years, advances in high-throughput sequencing technologies and multi-omics profiling have catalyzed a paradigm shift in cancer classification. The landmark study by The Cancer Genome Atlas (TCGA) has redefined EC as a molecularly heterogeneous disease comprising four distinct subtypes—POLE ultramutated, microsatellite instability hypermutated (MSI), copy-number low (CN-low), and copy-number high (CN-high)—each with unique genomic, transcriptomic, and clinical features(Alexa, Hasenburg, & Battista, 2021; Zhang et al., 2024). This refined classification offers new opportunities for prognostication and personalized treatment, yet its implementation in clinical practice remains limited due to the cost and complexity of multi-omics assays.

There is a growing interest in identifying accessible and robust biomarkers that can

recapitulate the TCGA molecular subtypes using more practical platforms. Among various biomolecular candidates, microRNAs (miRNAs) have gained considerable attention due to their regulatory roles in gene expression, their tissue- and cancer-specific expression patterns, and their stability in formalin-fixed paraffin-embedded (FFPE) tissues and bodily fluids(Nalbant & Akkaya-Ulum, 2024; Seida, Ogami, Yoshino, & Suzuki, 2025). Accumulating evidence suggests that miRNAs not only participate in the pathogenesis of EC but also reflect its molecular subtypes and aggressiveness(He, Liao, Peng, Zou, & Guo, 2025; Xie, Huang, Xie, Hu, & Jin, 2025). Therefore, developing a miRNA-based classifier to predict molecular subtypes of EC may provide a clinically feasible alternative to comprehensive sequencing, enabling more precise risk stratification and guiding adjuvant therapy in routine practice. However, the practical application of full-scale multi-omics analyses remains limited in most clinical settings due to logistical, financial, and technical constraints. MicroRNAs, as small non-coding RNAs that regulate post-transcriptional gene expression, have shown great promise as potential diagnostic and prognostic biomarkers across various cancers, including EC. They are not only stable and quantifiable in diverse biological samples but also reflect key molecular processes such as DNA damage repair, mismatch repair deficiency, chromatin remodeling, and oncogenic pathway activation—all of which are hallmarks of the TCGA EC subtypes.

This study presents a proof-of-concept computational analysis to evaluate the feasibility of a miRNA-based molecular signature for endometrial carcinoma. Our survival analyses reconfirmed the established prognostic and biological relevance of the ground-truth TCGA subtypes, serving to validate the clinical significance of the target classification system rather than the predictive performance of the machine learning model itself. By exploring a potential alternative to resource-intensive genomic sequencing, this pilot feasibility validation provides preliminary evidence that miRNA profiling could complement existing molecular subtyping methods. While currently exploratory, these findings offer a computational foundation for future large-scale prospective studies to investigate the potential utility of miRNA signatures in precision oncology frameworks.

2. Methods

2.1 Data Source and Sample Selection

In this study, we utilized publicly available data from The Cancer Genome Atlas (TCGA) Uterine Corpus Endometrial Carcinoma (TCGA-UCEC) cohort. A total of 232 EC samples with clearly annotated molecular TCGA subtypes were included in the analysis. The subtype classification—comprising POLE ultramutated, microsatellite instability (MSI) hypermutated, copy-number low (CN-low), and copy-number high (CN-high)—was derived based on the integrated genomic analysis conducted by Douglas A. Levine and colleagues, as part of the TCGA Research Network initiative(Akhtar, Al Hyassat, Elaiwy, Rashid, & Al-Nabet, 2019; Gargiulo et al., 2016; Winterhoff, Thomaier, Mullany, & Powell, 2020; Yoshimura et al., 2025). These molecular subtypes were established through comprehensive multi-omics profiling, including whole-exome sequencing, copy number variation, DNA methylation, and transcriptomic analyses.

Corresponding microRNA expression data were also retrieved from the TCGA-UCEC dataset via the Genomic Data Commons (GDC) data portal. All miRNA profiling was performed using the Illumina HiSeq platform, and raw count data were normalized using standard TCGA workflows. Only primary tumor samples with complete miRNA expression profiles and subtype annotations were retained for downstream analyses. Sample identifiers were cross-matched to

ensure consistency between molecular subtype metadata and miRNA expression matrices. All data processing was conducted in accordance with TCGA data usage policies and guidelines.

2.2 Prognostic Assessment of Molecular Subtypes

To evaluate the prognostic significance of the four molecular subtypes of endometrial carcinoma, we conducted survival analyses using multiple clinical endpoints. Patients with missing subtype annotations were excluded, resulting in a final analytical cohort of 232 cases. Overall survival (OS), progression-free interval (PFI), disease-specific survival (DSS), and disease-free interval (DFI) were selected as the primary clinical outcomes for evaluating survival differences among the POLE ultramutated, MSI hypermutated, copy-number low, and copy-number high groups. Kaplan–Meier survival curves were constructed for each endpoint to visualize survival probability across the four molecular subtypes. Log-rank tests were employed to assess statistical differences between survival distributions. For all time-to-event outcomes, survival times were defined based on standardized criteria provided in the TCGA-CDR dataset, ensuring comparability across endpoints. To account for the presence of competing risks—particularly death unrelated to tumor recurrence—we further employed cumulative incidence function (CIF) analysis for the PFI endpoint. Subtype-stratified CIF curves were generated, and differences among groups were tested using Gray's test.

2.3 Identification of Subtype-Associated miRNAs

To identify robust miRNA signatures specific to the four molecular subtypes while strictly preventing data leakage, the dataset was first randomly partitioned into a training set (80%) and an independent test set (20%) using the caret package with a fixed random seed for reproducibility. Feature selection was conducted exclusively within the training cohort to define the classifier inputs. We applied the Kruskal-Wallis rank-sum test to screen for miRNAs showing statistically significant differential expression across the POLE, MSI, CN-low, and CN-high subtypes. To control for false positives inherent in high-dimensional genomic data, p-values were adjusted using the False Discovery Rate (FDR) method. To construct a concise and clinically translatable signature, we ranked the significant features (FDR < 0.05) by statistical significance and selected the top 20 miRNAs as the core feature set for downstream modeling.

2.4 Machine Learning Model Development and Evaluation

Prior to model training, missing expression values were imputed using the median value calculated from the training set; this imputation logic was subsequently applied to the test set to maintain data independence. We implemented a comprehensive benchmark analysis using seven distinct machine learning algorithms to determine the optimal classifier: Random Forest (RF), Support Vector Machine (SVM) with a linear kernel, Extreme Gradient Boosting (XGBoost), Least Absolute Shrinkage and Selection Operator (LASSO) multinomial logistic regression, k-Nearest Neighbors (KNN), Naive Bayes (NB), and a baseline non-regularized multinomial logistic regression. All models were trained on the processed training set using the identified 20-miRNA signature. Performance was rigorously evaluated on the unseen test set to assess generalizability. We utilized a multi-metric approach, reporting overall Accuracy, Macro-averaged F1 score, Weighted F1 score, and the multiclass Area Under the Curve (AUC). Based on the comparative results, the LASSO model demonstrated superior performance and was selected for detailed characterization, including One-vs-Rest (OvR) Receiver Operating Characteristic (ROC) curves and confusion matrix analysis.

2.5 Model Validation of Clinical Applicability

To assess the translational feasibility of the identified miRNA signature, we performed a preliminary external validation using a pilot cohort of 15 independent endometrial carcinoma tumor samples. The expression levels of the signature miRNAs were quantified using quantitative real-time PCR (qRT-PCR). The ground-truth molecular subtypes of these clinical samples were confirmed via standard clinical workflows equivalent to TCGA criteria, including POLE exonuclease domain sequencing and immunohistochemistry (IHC) for mismatch repair proteins (MSH2, MSH6, MLH1, PMS2) and p53, adhering to the ProMisE classification guidelines. The trained computational model was applied to the PCR-derived expression data to predict subtypes. Given the pilot nature and limited sample size of this validation cohort, model performance was quantified using classification accuracy, and statistical uncertainty was estimated using 95% Confidence Intervals (CI) calculated via the exact binomial test.

3. Results

3.1 Prognostic Assessment of Molecular Subtypes

A total of 232 patients with complete molecular subtype annotations were included in this analysis (Table S1). Kaplan-Meier survival curves were generated to assess differences in overall survival, progression-free interval), disease-specific survival, and disease-free interval across the subtypes. Our findings demonstrated significant differences in prognosis among the four subtypes. As shown in Figure 1a, the KM analysis for OS showed a clear separation of survival curves, with a log-rank p-value of 0.0011, indicating a strong association between molecular subtype and long-term survival. Patients with the POLE ultra-mutated subtype exhibited the most favorable overall survival, while those with the copy-number high (serous-like) subtype had the worst prognosis. As shown in Figure 1b, for PFI, the difference among the subtypes was even more pronounced. The POLE and MSI subtypes demonstrated significantly longer progression-free intervals, whereas the serous-like group experienced more rapid disease progression. This distinction was statistically robust, with a p-value of < 0.0001. As shown in Figure 1c and 1d, the DSS and DFI curves also revealed subtype-specific patterns, with POLE and MSI subtypes generally associated with better outcomes. Although the log-rank p-value for DFI was comparatively modest ($p = 0.03$), the cumulative incidence analysis (as shown in Figure 1e) accounting for competing risks further substantiated the survival differences across subtypes. Collectively, these survival analyses reinforce the clinical relevance of the four molecular subtypes in EC and validate their prognostic stratification power. The POLE subtype consistently demonstrates superior prognosis, while the copy-number high (serous-like) subtype is associated with the poorest outcomes.

3.2 Machine Learning Model Development and Evaluation

To construct a robust molecular classifier, we first performed feature selection on the training set, identifying a signature of 20 miRNAs significantly associated with the four EC subtypes. The top-ranked features included hsa-miR-497-5p, hsa-miR-195-5p, hsa-miR-18a-3p, hsa-miR-34a-5p, and hsa-miR-9-3p, among others (as shown in Figure 2 and Table S2). Using this 20-miRNA signature, we conducted a comprehensive benchmark analysis of seven machine learning algorithms on the independent test set. As illustrated in the model comparison analysis, the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model emerged as the top-performing algorithm, achieving the highest multiclass Area Under the Curve (AUC) of 0.850, followed closely by Random Forest (AUC = 0.837) and Naive Bayes (AUC = 0.833) (as shown in Figure 3 and Table S3). The LASSO model notably outperformed the baseline non-regularized

logistic regression (AUC = 0.786) and the k-Nearest Neighbors classifier, which showed the poorest performance (AUC = 0.540). Consequently, the LASSO model was selected as the optimal classifier for further characterization.

Detailed evaluation of the LASSO model revealed strong predictive capabilities across the specific molecular subtypes (as shown in Figure 4). The One-vs-Rest ROC analysis demonstrated high sensitivity and specificity for the prognostically distinct subtypes, with AUC values of 0.889 for the POLE ultra-mutated subtype and 0.890 for the copy-number low (Endometrioid) subtype. The aggressive copy-number high (Serous-like) subtype was also classified with high accuracy (AUC = 0.875). The MSI hyper-mutated subtype proved the most challenging to distinguish, yielding a comparatively lower AUC of 0.756. Examination of the confusion matrix for the test set corroborated these findings; while the model effectively separated POLE and copy-number high tumors, there was a degree of overlap between the MSI and copy-number low subtypes, reflecting the shared biological characteristics often observed between these groups (as shown in Figure 5).

3.3 Model Validation of Clinical Applicability

To assess the translational feasibility and robustness of our identified miRNA signature, we performed an external validation using a pilot cohort of 15 independent endometrial carcinoma tissue samples (as shown in Figure 6). The ground-truth molecular subtypes were confirmed via POLE sequencing and IHC markers (p53, MMR proteins), comprising 4 POLE, 4 MSI, 4 CN-low, and 3 CN-high cases. When applied to the qRT-PCR expression data from these clinical samples, the LASSO classifier achieved an overall accuracy of 86.7% (13/15), with a 95% confidence interval of 59.5% to 98.3%.

The confusion matrix for this validation cohort highlighted the model's reliability in identifying the most clinically critical subtypes. The classifier achieved 100% sensitivity for the POLE ultra-mutated subtype (4/4 correctly identified) and the copy-number high (Serous-like) subtype (3/3 correctly identified), which are associated with the best and worst prognoses, respectively. The misclassifications were confined to the intermediate groups: one MSI sample was misclassified as copy-number low, and one copy-number low sample was misclassified as copy-number high. These results indicates that despite the limited sample size of the pilot cohort, the 20-miRNA signature maintains high diagnostic accuracy and clinical relevance, particularly in distinguishing the subtypes with the most divergent clinical outcomes.

3. Discussion

In this proof-of-concept computational study, we explored the feasibility of using a miRNA-based molecular classifier to distinguish between the four biologically and prognostically distinct subtypes of endometrial carcinoma (EC) established by The Cancer Genome Atlas (TCGA). While integrated genomic profiling (combining sequencing, copy number variation, and methylation) remains the gold standard for classification, its implementation in routine clinical practice is often limited by cost and technical complexity. By leveraging a rigorously constructed 20-miRNA signature and a penalized machine learning approach, our LASSO model achieved a multiclass AUC of 0.850 on an independent test set and demonstrated an overall accuracy of 86.7% in a pilot external validation cohort. These findings suggest that miRNA profiling may represent a viable, cost-effective surrogate for molecular subtyping, though further validation is required to establish clinical utility.

Among the seven machine learning algorithms evaluated, the LASSO logistic regression model emerged as the top performer, outperforming complex non-linear models like SVM and

Random Forest, as well as the baseline non-regularized logistic regression. The success of the LASSO algorithm—which enforces sparsity by shrinking less informative coefficients to zero—suggests that the molecular distinctions between EC subtypes are driven by a specific subset of high-impact miRNAs rather than diffuse, genome-wide variations. This finding is particularly advantageous for translational research, as it implies that a simplified PCR-based panel targeting these specific miRNAs could potentially replicate the classification performance of high-throughput sequencing platforms.

Crucially, the selected 20-miRNA signature is not merely a statistical artifact but reflects the underlying pathophysiology of endometrial tumorigenesis. Our biological enrichment analysis highlights that these miRNAs serve as key post-transcriptional regulators of pathways defining the specific subtypes. For instance, hsa-miR-34a-5p, a prominent feature in our signature, is a well-characterized downstream effector of TP53. Its dysregulation is mechanically linked to the TP53-mutated status defining the Copy-Number High (Serous-like) subtype, consistent with its role in cell cycle arrest and apoptosis. Similarly, members of the miR-15/16 family (e.g., hsa-miR-497-5p, hsa-miR-195-5p) were found to be significantly downregulated in high-grade subtypes; these miRNAs functionally target IGF1R and mTOR signaling pathways, which are frequently altered in Endometrioid (Copy-Number Low) and MSI-hypermutated tumors. The ability of the classifier to distinguish the MSI subtype, albeit with slightly lower sensitivity, may be attributed to miRNAs regulating mismatch repair genes (e.g., MLH1, MSH2), suggesting a direct functional link between the miRNA profile and the genomic instability characterizing this group.

It is important to clarify the role of the survival analyses presented in this study. Our Kaplan-Meier and cumulative incidence function analyses confirmed that the ground-truth TCGA subtypes (POLE, MSI, CN-low, and CN-high) exhibit statistically distinct prognostic trajectories, with POLE-mutated patients showing superior survival and CN-high patients experiencing the poorest outcomes. These results validate the clinical relevance of the classification system itself, but they do not directly validate the prognostic power of our machine learning predictions. The strong association between the ground-truth subtypes and clinical outcome serves as the rationale for developing this classifier, emphasizing that accurate molecular subtyping—whether by genomic sequencing or our proposed miRNA surrogate—is essential for risk stratification.

Despite these promising preliminary results, this study has several limitations inherent to its retrospective design and scope. First, while the LASSO model performed well on the internal test set, the pilot external validation was conducted on a small cohort. While this demonstrated technical feasibility using qRT-PCR, the wide confidence intervals indicate that estimates of diagnostic accuracy must be interpreted with caution. Second, the current model relies solely on miRNA expression; future iterations could potentially enhance performance by integrating clinical variables or histopathological features. Finally, while we identified biologically plausible miRNA targets, the specific regulatory mechanisms in the context of POLE ultramutation remain to be fully elucidated experimentally.

In conclusion, this work serves as a pilot feasibility validation for a miRNA-based EC classifier. We demonstrate that a compact signature of 20 miRNAs can effectively discriminate clinically relevant molecular subtypes, offering a potential low-cost alternative for stratification in resource-limited settings. However, to transition from this proof-of-concept study to clinical application, further rigorous evaluation is needed. It is also necessary to do a prospective study to

know the potential application of the identified microRNAs in EC. Such future work should focus on large-scale, multi-center cohorts to confirm the robustness of the signature and evaluate its efficacy in guiding real-time therapeutic decision-making.

4. Conclusion

In conclusion, this study presents a proof-of-concept computational analysis demonstrating the feasibility of using a miRNA-based machine learning classifier for the molecular subtyping of endometrial carcinoma. Among the seven evaluated algorithms, the LASSO logistic regression model achieved superior performance, suggesting that a compact panel of 20 miRNAs can effectively capture the molecular distinctions between subtypes, particularly for the prognostically critical POLE and copy-number high groups. Our pilot external validation using qRT-PCR on independent clinical samples provides preliminary evidence of the technical transferability of this signature. Crucially, while the survival analyses conducted in this study reconfirmed the strong prognostic relevance of the established TCGA molecular subtypes, the developed classifier serves as a surrogate tool for subtype assignment rather than a direct prognostic predictor. These findings indicate that miRNA profiling holds promise as a cost-effective, accessible alternative to comprehensive genomic sequencing in resource-limited settings. However, large-scale prospective studies are necessary to validate these retrospective findings and fully establish the clinical utility of this signature in routine precision oncology.

Funding

Not applicable.

Author Contributions

Gedan Liu (First author): Data curation, formal analysis, methodology, visualization, writing—original draft.

Sifeng Zhou (Second author): Methodology, software, validation, writing—review & editing.

Yi Miao (Corresponding author): Conceptualization, project administration, supervision, funding acquisition, writing—review & editing.

All authors read and approved the final manuscript.

Conflict of Interest

The authors declare that they have no competing interests.

Consent for Publication

Not applicable.

Data Availability

The datasets analyzed during the current study are publicly available in the TCGA repository (https://portal.gdc.cancer.gov/) under the TCGA-UCEC project. Additional data generated during the clinical validation are available from the corresponding author upon reasonable request.

Acknowledgments

The authors thank the TCGA Research Network for generating and providing the public datasets. We also thank the clinical staff of the Department of Gynecology, Shanghai General Hospital, for assistance in collecting tissue samples and performing laboratory experiments.

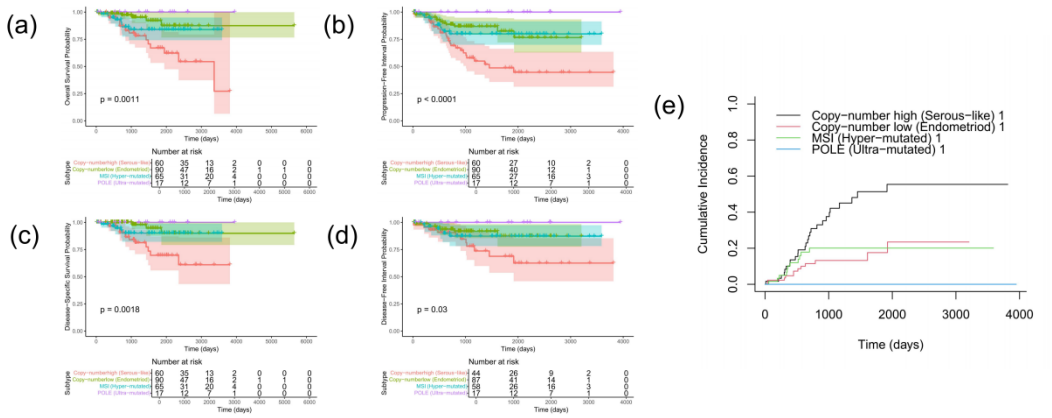

Figure 1. Survival analysis and disease progression of endometrial carcinoma molecular subtypes. Panels show (a) Overall Survival, (b) Progression-Free Interval, (c) Disease-Specific Survival, and (d) Disease-Free Interval probability. Panel (e) shows the Cumulative Incidence Function for disease progression across the four subtypes.

Figure 2. Heatmap of the top 20 selected miRNA signatures. miRNA expression levels are shown as Z-scores across samples grouped by the four molecular subtypes: POLE (Ultra-mutated), MSI (Hyper-mutated), Copy-number low, and Copy-number high.

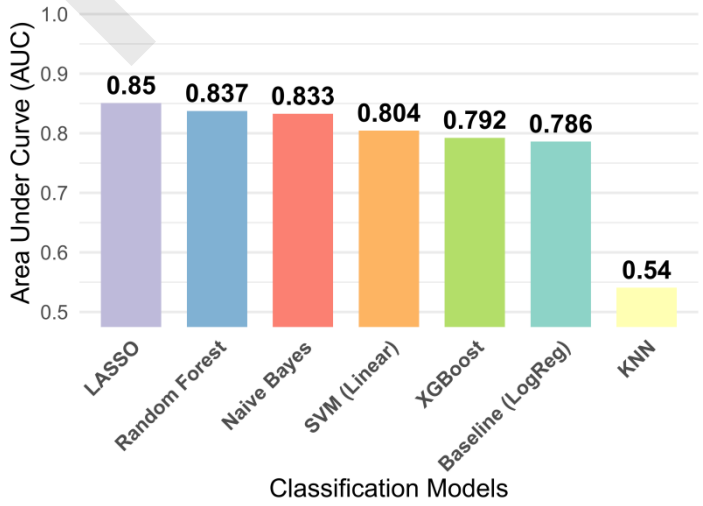

Figure 3. Performance comparison of seven machine learning models. The bar chart displays the

multiclass Area Under the Curve (AUC) for LASSO, Random Forest, Naive Bayes, SVM, XGBoost, Baseline (LogReg), and KNN.

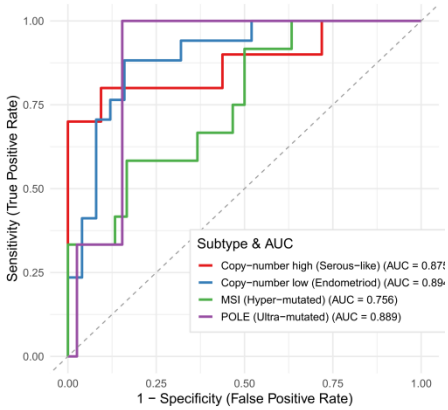

Figure 4. One-vs-Rest ROC curves for the LASSO classifier. Each curve represents the classification performance for a specific subtype against all others, with corresponding AUC values indicated in the legend.

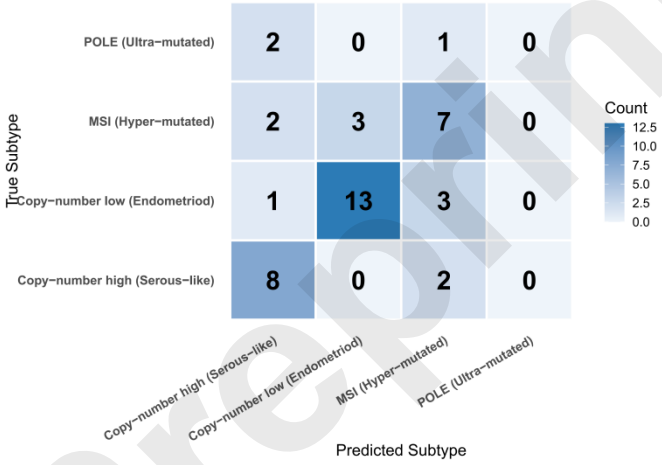

Figure 5. Confusion matrix of the LASSO model on the independent test set. The plot illustrates the distribution of true versus predicted molecular subtypes with sample counts indicated in each cell.

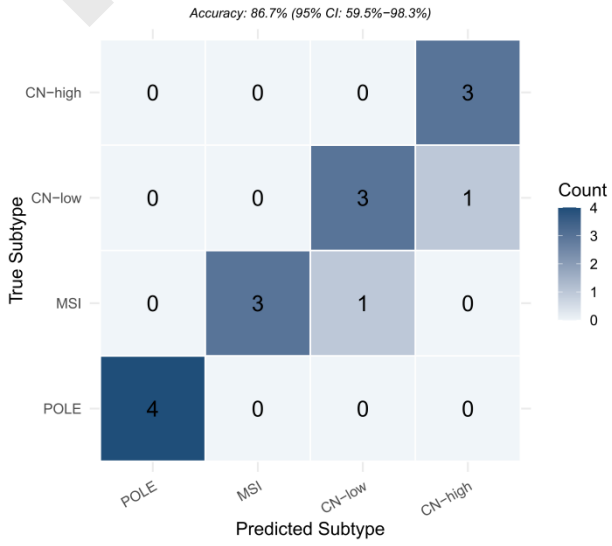

Figure 6. Confusion matrix of the pilot external validation cohort. The model achieved an overall accuracy of 86.7% (95% CI: 59.5%-98.3%) in classifying independent clinical samples confirmed via POLE sequencing and IHC.

Reference

Akhtar, M., Al Hyassat, S., Elaiwy, O., Rashid, S., & Al-Nabet, A. (2019). Classification of Endometrial Carcinoma: New Perspectives Beyond Morphology. *Adv Anat Pathol, 26*(6), 421-427. doi:10.1097/PAP.0000000000000251

Alexa, M., Hasenburg, A., & Battista, M. J. (2021). The TCGA Molecular Classification of Endometrial Cancer and Its Possible Impact on Adjuvant Treatment Decisions. *Cancers (Basel), 13*(6). doi:10.3390/cancers13061478

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, 74*(3), 229-263. doi:10.3322/caac.21834

Chen, F., Chen, M., Chen, R., Xiao, L., & Huang, J. (2025). An assisted diagnostic and prognostic model for endometrial cancer using 36 serological markers and clinical variables from 562 patients. *Sci Rep, 15*(1), 40441. doi:10.1038/s41598-025-23853-8

Concin, N., Matias-Guiu, X., Cibula, D., Colombo, N., Creutzberg, C. L., Ledermann, J., . . . Nout, R. A. (2025). ESGO-ESTRO-ESP guidelines for the management of patients with endometrial carcinoma: update 2025. *Lancet Oncol, 26*(8), e423-e435. doi:10.1016/S1470-2045(25)00167-6

Dou, Y., Kawaler, E. A., Cui Zhou, D., Gritsenko, M. A., Huang, C., Blumenberg, L., . . . Clinical Proteomic Tumor Analysis, C. (2020). Proteogenomic Characterization of Endometrial Carcinoma. *Cell, 180*(4), 729-748 e726. doi:10.1016/j.cell.2020.01.026

Elmadani, M., Mokaya, P. O., Omer, A. A. A., Kiptulon, E. K., Klara, S., & Orsolya, M. (2025). Cancer burden in Europe: a systematic analysis of the GLOBOCAN database (2022). *BMC Cancer, 25*(1), 447. doi:10.1186/s12885-025-13862-1

Filho, A. M., Laversanne, M., Ferlay, J., Colombet, M., Pineros, M., Znaor, A., . . . Bray, F. (2025). The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. *Int J Cancer, 156*(7), 1336-1346. doi:10.1002/ijc.35278

Gargiulo, P., Della Pepa, C., Berardi, S., Califano, D., Scala, S., Buonaguro, L., . . . Pignata, S. (2016). Tumor genotype and immune microenvironment in POLE-ultramutated and MSI-hypermutated Endometrial Cancers: New candidates for checkpoint blockade immunotherapy? *Cancer Treat Rev, 48*, 61-68. doi:10.1016/j.ctrv.2016.06.008

He, Y., Liao, K., Peng, H., Zou, X., & Guo, Z. (2025). Advances in MiRNAs Involved in Endometrial Carcinoma. *Comb Chem High Throughput Screen, 28*(1), 3-11. doi:10.2174/0113862073299444240308145725

Hong, M. K., & Ding, D. C. (2025). Early Diagnosis of Ovarian Cancer: A Comprehensive Review of the Advances, Challenges, and Future Directions. *Diagnostics (Basel), 15*(4). doi:10.3390/diagnostics15040406

Nalbant, E., & Akkaya-Ulum, Y. Z. (2024). Exploring regulatory mechanisms on miRNAs and their implications in inflammation-related diseases. *Clin Exp Med, 24*(1), 142. doi:10.1007/s10238-024-01334-y

Obon-Santacana, M., Kaaks, R., Slimani, N., Lujan-Barroso, L., Freisling, H., Ferrari, P., . . . Duell, E. J. (2014). Dietary intake of acrylamide and endometrial cancer risk in the European Prospective Investigation into Cancer and Nutrition cohort. *Br J Cancer, 111*(5), 987-997. doi:10.1038/bjc.2014.328

Seida, M., Ogami, K., Yoshino, S., & Suzuki, H. I. (2025). Fine Regulation of MicroRNAs in Gene Regulatory Networks and Pathophysiology. *Int J Mol Sci, 26*(7). doi:10.3390/ijms26072861

Uccella, S., Mariani, A., Wang, A. H., Vierkant, R. A., Cliby, W. A., Robien, K., . . . Cerhan, J. R. (2013). Intake of coffee, caffeine and other methylxanthines and risk of Type I vs Type II endometrial cancer. *Br J Cancer, 109*(7), 1908-1913. doi:10.1038/bjc.2013.540

Whitaker, K. (2020). Earlier diagnosis: the importance of cancer symptoms. *Lancet Oncol, 21*(1), 6-8. doi:10.1016/S1470-2045(19)30658-8

Winterhoff, B., Thomaier, L., Mullany, S., & Powell, M. A. (2020). Molecular characterization of endometrial cancer and therapeutic implications. *Curr Opin Obstet Gynecol, 32*(1), 76-83. doi:10.1097/GCO.0000000000000602

Xie, Q., Huang, J., Xie, Y., Hu, J., & Jin, L. (2025). Identification of prognostic biomarkers for endometrioid endometrial carcinoma based on the miRNA and mRNA co-expression network regulated by estradiol. *Clinics (Sao Paulo), 80*, 100672. doi:10.1016/j.clinsp.2025.100672

Yoshimura, T., Nakamura, K., Chiyoda, T., Takamatsu, R., Kawano, R., Aimono, E., . . . Yamagami, W. (2025). Next-generation sequencing outperforms Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) in endometrial cancer molecular classification. *BJC Rep, 3*(1), 37. doi:10.1038/s44276-025-00145-2

Zhang, B., Zhou, D., Zhang, S., Yan, J., Meng, Q., & Lv, Q. (2024). Clinical analysis of molecular typing of 146 cases of endometrial carcinoma. *Front Oncol, 14*, 1482817. doi:10.3389/fonc.2024.1482817

Zhao, Q., Hughes, R., Altaf-Haroon, I., Schiller, E., & Kadambi, A. (2021). Systematic literature review of the real-world burden and use of chemotherapies for treatment of advanced or recurrent endometrial carcinoma. *Journal of Clinical Oncology, 39*(15_suppl), e17571-e17571. doi:10.1200/JCO.2021.39.15_suppl.e17571

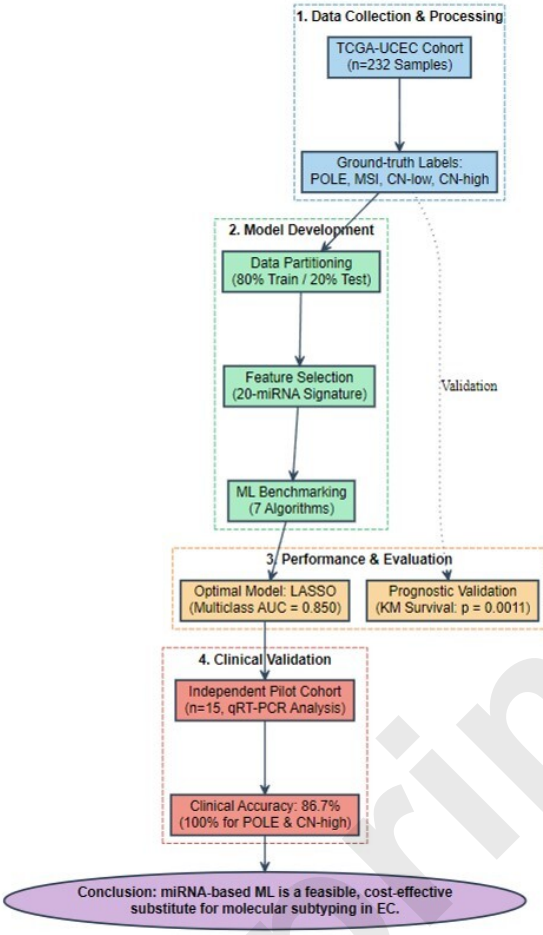
1. Data Collection & Processing
TCGA-UCEC Cohort
(n=232 Samples)
Ground-truth Labels:
POLE, MSI, CN-low, CN-high
2. Model Development
Data Partitioning
(80% Train / 20% Test)
Feature Selection
(20-miRNA Signature)
ML Benchmarking
(7 Algorithms)
Validation
3. Performance & Evaluation
Optimal Model: LASSO
(Multiclass AUC = 0.850)
Prognostic Validation
(KM Survival: p = 0.0011)
4. Clinical Validation
Independent Pilot Cohort
(n=15, qRT-PCR Analysis)
Clinical Accuracy: 86.7%
(100% for POLE & CN-high)
Conclusion: miRNA-based ML is a feasible, cost-effective
substitute for molecular subtyping in EC.

Table Legends

Table S1. Clinical and molecular characteristics of 232 endometrial carcinoma patients in the TCGA-UCEC c

Table S2. miRNAs significantly associated with endometrial carcinoma molecular subtypes.

Table S3. Performance metrics of six machine learning classifiers for predicting EC molecular subtypes.

ohort.

Patient.ID	Study.ID	Cancer.Type	Subtype
TCGA-A5-A0G5	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-A5-A0G9	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GA	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GB	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GE	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0GH	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GI	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GJ	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-A5-A0GM	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0GN	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-A5-A0GP	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-A5-A0GQ	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0GU	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0GV	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0GW	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0GX	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0R6	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-A5-A0R7	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-A5-A0R8	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0R9	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0RA	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-A5-A0VP	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-A5-A0VQ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AJ-A23M	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A051	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-AP-A052	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A053	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A054	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A056	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-AP-A059	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-AP-A05A	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A05D	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A05H	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A05J	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A05N	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A05P	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A0L8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A0L9	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A0LD	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A0LE	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A0LF	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A0LG	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A0LH	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A0LI	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AP-A0LJ	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A0LL	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A0LM	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)

TCGA-AP-A0LN	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A0LO	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A0LP	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A0LT	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AP-A0LV	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AP-A1DQ	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A05S	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A05T	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AX-A05U	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AX-A05Y	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A05Z	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-AX-A060	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A062	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AX-A063	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A064	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A06B	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A06H	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A0IS	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-AX-A0IU	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A0IW	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A0J0	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-AX-A0J1	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-AX-A1C7	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A1C8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A1CP	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A2H5	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-AX-A2HF	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-B5-A0JS	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0JT	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0JV	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A0JY	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-B5-A0JZ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A0K0	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0K1	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0K2	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A0K3	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-B5-A0K4	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0K6	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A0K7	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A0K8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-B5-A11E	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-B5-A11F	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11G	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A11H	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A11I	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-B5-A11J	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A11N	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-B5-A11O	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)

TCGA-B5-A11Q	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11S	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11U	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-B5-A11V	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11W	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11Y	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A11Z	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A121	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-B5-A1MU	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-B5-A1MY	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BG-A0LW	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0LX	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0M0	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0M2	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0M3	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0M4	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0M6	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BG-A0M7	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0M8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BG-A0M9	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MC	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MG	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MI	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MO	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MQ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0MS	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MT	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0MU	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0RY	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0VT	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0VV	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0VW	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0VX	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0VZ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A0W1	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0W2	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0YU	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A0YV	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BG-A186	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A187	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A18A	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BG-A18B	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BG-A18C	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BG-A2AE	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BK-A0C9	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BK-A0CA	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BK-A0CB	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BK-A0CC	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)

TCGA-BK-A139	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BK-A13C	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0T9	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0TA	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0TC	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-BS-A0TD	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0TE	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0TG	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0TI	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BS-A0TJ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0U5	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0U7	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0U8	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0U9	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BS-A0UA	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0UF	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-BS-A0UJ	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0UL	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-BS-A0UT	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0UV	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-BS-A0V6	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-BS-A0V8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-BS-A0WQ	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZN	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZO	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A0ZP	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A0ZQ	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZR	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZS	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A0ZU	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZV	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A0ZZ	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A101	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A102	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A103	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-D1-A15V	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A15W	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A15X	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A15Z	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A160	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A161	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A163	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A165	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A167	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A168	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A169	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A16B	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A16D	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)

TCGA-D1-A16E	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A16G	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A16I	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A16J	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A16N	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A16O	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A16Q	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A16S	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A16X	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-D1-A16Y	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-D1-A174	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A176	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A177	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17A	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17B	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17C	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A17D	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17F	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17H	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17K	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A17L	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A17M	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17N	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A17Q	ucec_tcga_pub	Endometrial Cancer	POLE (Ultra-mutated)
TCGA-D1-A17R	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A17S	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A17T	ucec_tcga_pub	Endometrial Cancer	Copy-number low (Endometriod)
TCGA-D1-A17U	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-D1-A1NU	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-D1-A1NX	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-DI-A0WH	ucec_tcga_pub	Endometrial Cancer	MSI (Hyper-mutated)
TCGA-DI-A1NN	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-E6-A1LZ	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-EO-A1Y5	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-EO-A1Y8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-EY-A1GS	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-EY-A212	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-FI-A2D2	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-FI-A2EW	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-FI-A2EX	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)
TCGA-FI-A2F8	ucec_tcga_pub	Endometrial Cancer	Copy-number high (Serous-like)

miRNA	p.value	q.value
hsa.miR.497.5p	1,63E-10	3,08E-07
hsa.miR.195.5p	1,82E-09	1,72E-06
hsa.miR.18a.3p	4,24E-09	2,06E-06
hsa.miR.34a.5p	4,71E-09	2,06E-06
hsa.miR.9.3p	5,44E-09	2,06E-06
hsa.miR.195.3p	1,41E-08	3,84E-06
hsa.miR.99b.3p	1,47E-08	3,84E-06
hsa.miR.375	1,65E-08	3,84E-06
hsa.miR.196a.5p	1,83E-08	3,84E-06
hsa.miR.452.5p	2,57E-08	4,87E-06
hsa.miR.9.5p	9,27E-08	1,59E-05
hsa.miR.452.3p	1,54E-07	2,42E-05
hsa.miR.34c.5p	1,85E-07	2,69E-05
hsa.miR.652.3p	2,51E-07	3,39E-05
hsa.miR.224.5p	3,20E-07	4,03E-05
hsa.miR.4787.3p	4,69E-07	5,54E-05
hsa.miR.449a	5,27E-07	5,86E-05
hsa.miR.16.2.3p	6,73E-07	7,07E-05
hsa.miR.34c.3p	8,24E-07	8,20E-05
hsa.miR.4728.3p	1,13E-06	1,06E-04
hsa.let.7b.5p	1,19E-06	1,07E-04
hsa.miR.10a.5p	1,40E-06	1,19E-04
hsa.miR.15b.3p	1,44E-06	1,19E-04
hsa.miR.146b.3p	2,04E-06	1,60E-04
hsa.miR.34b.3p	2,25E-06	1,70E-04
hsa.miR.190b	2,85E-06	2,07E-04
hsa.miR.1468.5p	2,96E-06	2,07E-04
hsa.miR.205.5p	3,32E-06	2,20E-04
hsa.miR.146b.5p	3,37E-06	2,20E-04
hsa.miR.1224.5p	3,87E-06	2,42E-04
hsa.miR.449c.5p	3,96E-06	2,42E-04
hsa.miR.3074.5p	8,35E-06	4,92E-04
hsa.miR.18a.5p	8,60E-06	4,92E-04
hsa.miR.497.3p	1,13E-05	6,30E-04
hsa.miR.31.5p	1,23E-05	6,59E-04
hsa.let.7a.5p	1,28E-05	6,59E-04
hsa.miR.330.5p	1,29E-05	6,59E-04
hsa.let.7g.5p	1,62E-05	8,03E-04
hsa.miR.1266.5p	1,66E-05	8,03E-04
hsa.miR.132.3p	2,04E-05	9,63E-04
hsa.miR.449b.3p	2,10E-05	9,69E-04
hsa.miR.34b.5p	2,28E-05	1,03E-03
hsa.miR.222.5p	2,80E-05	1,23E-03
hsa.miR.31.3p	3,07E-05	1,32E-03
hsa.miR.1226.3p	3,14E-05	1,32E-03
hsa.miR.301a.5p	3,30E-05	1,36E-03
hsa.miR.222.3p	4,30E-05	1,73E-03

hsa.miR.3127.5p	4,91E-05	1,93E-03
hsa.miR.30c.2.3p	5,29E-05	2,02E-03
hsa.miR.516a.5p	5,36E-05	2,02E-03
hsa.miR.221.3p	5,80E-05	2,15E-03
hsa.miR.942.5p	7,31E-05	2,66E-03
hsa.let.7i.3p	9,40E-05	3,35E-03
hsa.miR.499a.5p	1,01E-04	3,52E-03
hsa.miR.130b.3p	1,06E-04	3,60E-03
hsa.miR.135a.3p	1,07E-04	3,60E-03
hsa.miR.301b.3p	1,09E-04	3,62E-03
hsa.miR.455.5p	1,30E-04	4,23E-03
hsa.miR.942.3p	1,38E-04	4,42E-03
hsa.miR.130a.3p	1,92E-04	6,06E-03
hsa.miR.30a.5p	2,03E-04	6,16E-03
hsa.miR.7.5p	2,05E-04	6,16E-03
hsa.miR.17.3p	2,05E-04	6,16E-03
hsa.miR.744.5p	2,27E-04	6,72E-03
hsa.miR.135a.5p	2,34E-04	6,80E-03
hsa.miR.937.3p	2,41E-04	6,89E-03
hsa.miR.20a.3p	2,45E-04	6,90E-03
hsa.miR.196b.5p	3,27E-04	9,10E-03
hsa.miR.431.5p	3,43E-04	9,39E-03
hsa.miR.296.3p	3,60E-04	9,72E-03
hsa.miR.326	4,00E-04	1,06E-02
hsa.miR.499a.3p	4,10E-04	1,08E-02
hsa.miR.200b.3p	4,29E-04	1,11E-02
hsa.miR.3677.3p	4,78E-04	1,22E-02
hsa.miR.664a.3p	4,83E-04	1,22E-02
hsa.miR.676.3p	5,12E-04	1,27E-02
hsa.let.7b.3p	5,19E-04	1,27E-02
hsa.miR.6511b.3p	5,38E-04	1,30E-02
hsa.miR.135b.5p	5,57E-04	1,33E-02
hsa.miR.185.3p	5,71E-04	1,35E-02
hsa.miR.760	5,82E-04	1,35E-02
hsa.miR.642a.5p	5,84E-04	1,35E-02
hsa.miR.301a.3p	6,19E-04	1,41E-02
hsa.miR.125a.3p	6,40E-04	1,44E-02
hsa.miR.935	6,58E-04	1,45E-02
hsa.miR.182.5p	6,62E-04	1,45E-02
hsa.miR.4661.5p	7,33E-04	1,59E-02
hsa.miR.874.3p	7,61E-04	1,63E-02
hsa.miR.4674	7,75E-04	1,65E-02
hsa.miR.4746.5p	7,84E-04	1,65E-02
hsa.let.7i.5p	7,94E-04	1,65E-02
hsa.miR.429	8,41E-04	1,73E-02
hsa.miR.3131	8,52E-04	1,73E-02
hsa.miR.10b.3p	8,65E-04	1,74E-02
hsa.miR.99b.5p	9,14E-04	1,82E-02

hsa.miR.22.3p	9,62E-04	1,88E-02
hsa.miR.664a.5p	9,65E-04	1,88E-02
hsa.miR.2110	9,92E-04	1,91E-02
hsa.miR.301b.5p	1,00E-03	1,92E-02
hsa.miR.187.3p	1,04E-03	1,96E-02
hsa.miR.106b.3p	1,09E-03	2,03E-02
hsa.miR.335.3p	1,09E-03	2,03E-02
hsa.miR.4473	1,14E-03	2,09E-02
hsa.let.7a.3p	1,15E-03	2,09E-02
hsa.miR.500b.3p	1,22E-03	2,18E-02
hsa.miR.548v	1,23E-03	2,18E-02
hsa.miR.30a.3p	1,24E-03	2,18E-02
hsa.miR.125a.5p	1,25E-03	2,18E-02
hsa.miR.615.3p	1,27E-03	2,20E-02
hsa.miR.449b.5p	1,32E-03	2,27E-02
hsa.miR.3614.3p	1,38E-03	2,35E-02
hsa.miR.30c.5p	1,41E-03	2,38E-02
hsa.miR.324.3p	1,42E-03	2,38E-02
hsa.miR.625.3p	1,53E-03	2,53E-02
hsa.miR.501.3p	1,54E-03	2,53E-02
hsa.miR.148a.3p	1,58E-03	2,58E-02
hsa.miR.21.3p	1,67E-03	2,70E-02
hsa.miR.296.5p	1,77E-03	2,84E-02
hsa.miR.877.5p	2,01E-03	3,19E-02
hsa.miR.184	2,06E-03	3,24E-02
hsa.miR.181a.3p	2,11E-03	3,27E-02
hsa.miR.1229.3p	2,11E-03	3,27E-02
hsa.miR.27a.3p	2,26E-03	3,47E-02
hsa.miR.7702	2,55E-03	3,89E-02
hsa.miR.4668.3p	2,60E-03	3,93E-02
hsa.miR.561.5p	2,62E-03	3,93E-02
hsa.miR.185.5p	2,72E-03	4,04E-02
hsa.miR.224.3p	2,74E-03	4,04E-02
hsa.miR.92a.1.5p	2,79E-03	4,06E-02
hsa.miR.92a.3p	2,80E-03	4,06E-02
hsa.miR.455.3p	2,85E-03	4,11E-02
hsa.miR.190a.5p	2,89E-03	4,14E-02
hsa.miR.519a.3p	2,92E-03	4,15E-02
hsa.miR.4638.3p	2,98E-03	4,19E-02
hsa.miR.1269a	2,99E-03	4,19E-02
hsa.miR.934	3,09E-03	4,29E-02
hsa.miR.1306.5p	3,12E-03	4,30E-02
hsa.miR.95.3p	3,29E-03	4,51E-02
hsa.miR.605.5p	3,57E-03	4,85E-02
hsa.miR.4782.5p	3,69E-03	4,98E-02

Model	Macro_F1	Weighted_F1	AUC
Random Forest	0,494	0,6222	0,8373
SVM (Linear)	0,5134	0,6459	0,8042
XGBoost	0,5429	0,606	0,7922
LASSO	0,5109	0,6445	0,8504
KNN	0,3062	0,3857	0,5405
Naive Bayes	0,5993	0,6319	0,8326
Baseline (LogReg)	0,5417	0,6796	0,7858