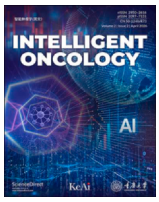
Intelligent Oncology 2 (2026) 100056

Contents lists available at ScienceDirect

Intelligent Oncology

journal homepage: www.keaipublishing.com/en/intonc/



Full Length Article

Integrative machine learning-driven prioritization of ceRNA networks in adrenocortical carcinoma

Javad Omidi

Department of Chemical Engineering, Columbia University, New York, NY 10027, USA

ARTICLE INFO

Keywords:
Adrenocortical carcinoma
ceRNA network
Machine learning
miRNA-mRNA interactions

ABSTRACT

Adrenocortical carcinoma (ACC) is a rare and highly aggressive endocrine malignancy in urgent need of robust biomarkers and novel therapeutic targets. In this study, a machine learning (ML)-driven framework is introduced to systematically model microRNA-messenger RNAmiRNA-mRNA regulatory associations and reconstruct context-specific ceRNA networks in ACC, leveraging harmonized RNA-Seq and miRNA-Seq data from The Cancer Genome Atlas-ACC and Genotype-Tissue Expression (2025). Multiple regression models were benchmarked, with random forest achieving superior predictive performance (R^2 = 0.9467 for normal and 0.9044 for the tumor datasets), enabling accurate identification of high-confidence miRNA-mRNA interactions. Integration of ML predictions with established reference datasets (TargetScan and miRTarBase) was demonstrated to have strong biological validity, while subsequent network analyses revealed extensive topological rewiring and a loss of regulatory hub connectivity in tumor tissue versus normal adrenal tissue. Moreover, a set of new, high-confidence miRNA-mRNA interactions in ACC was identified, implicating new candidates in oncogenic signaling and extracellular matrix remodeling. Together, a robust resource is provided for prioritizing functional interactions and regulatory hubs for future experimental validation, and the power of integrative ML approaches is underscored in advancing systems-level understanding and biomarker discovery in rare cancers such as ACC.

1. Introduction

Adrenocortical carcinoma (ACC) has been recognized as a rare endocrine malignancy of the adrenal cortex that exhibits aggressive clinical behavior and a dismal prognosis.[1,2] An annual incidence of only 0.7–2 cases per million has been reported, yet ACC has been estimated to account for most adrenal cancer deaths because of its high metastatic potential and resistance to conventional treatment.[2–4] Five-year survival below 40% in advanced disease has been documented despite modern surgery and adjuvant therapy, chiefly owing to late detection, frequent relapse, and limited systemic efficacy.[5,6] Consequently, an urgent requirement for robust molecular biomarkers suitable for early diagnosis, risk stratification, and personalized therapy has been articulated.[7,8]

Within competing endogenous RNA (ceRNA) networks, microRNAs (miRNAs) functioning as topological hubs have been described as master regulators of post-transcriptional programmes.[9] Such hubs have been shown to coordinate oncogene and tumor-suppressor expression by engaging numerous long non-coding RNAs, circular RNAs, and messenger RNAs through shared miRNA-response elements.[10,11] miR-21 and miR-155 have consistently been identified as high-degree nodes in breast, lung, and colorectal tumors, underscoring extensive roles in apoptosis, immune signaling, and proliferation.[12] Extensive network rewiring between normal and malignant tissue has been revealed, and the prognostic and therapeutic utility of hub targeting has been highlighted; supporting evidence has been provided that specific long non-coding RNAs and their miRNA partners, exemplified by LINC01929, act as oncogenic drivers of metastasis.[13–15]

Accurate delineation of functional miRNA-mRNA interactions has been impeded by the complexity of post-transcriptional regulation and tumor heterogeneity. Sequence-based prediction tools have been reported to generate thousands of putative targets, although only a small subset has been verified as active in tumors because of effects from RNA structure, cofactors, and contextual factors.[16] Linear-correlation measures have been noted to overlook non-linear or indirect associations, and integrative probabilistic frameworks, such as TaLasso and GenMiR++, have therefore been proposed.[17] Context-specific regulation has been highlighted by lung cancer studies[18,19] and breast cancer re-

E-mail address: jo2668@columbia.edu(J. Omidi)

https://doi.org/10.1016/j.intonc.2026.100056
Received 26 January 2026; Received in revised form 15 March 2026; Accepted 30 March 2026
Available online 2 April 2026
2950-2616/© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

search,[20] while a broad contribution of miRNA dysregulation to tumor initiation, progression and therapy resistance has been suggested by bibliometric surveys.[21]

Statistically significant yet weak associations have been produced by traditional expression-correlation approaches due to combinatorial miRNA action, dominance of translational repression over mRNA degradation, and ceRNA competition, whereas seed-sequence databases have exhibited false-positive rates of 20%-70% and poor concordance.[22,23] These shortcomings have been attributed to biases that favor canonical 3′-UTR sites, neglect tissue-specific 3′-UTR shortening, and rely on outdated conservation filters, thereby reinforcing the need for context-aware models that integrate isoform-resolved expression data.[23]

Machine learning (ML) methodologies have increasingly been adopted in ACC biomarker research. High-dimensional multi-omics datasets have been integrated to classify tumors into prognostically relevant subtypes and to identify novel marker genes.[24–26] Algorithms, such as random forests (RF), support vector machines, and neural networks, have been applied with high predictive accuracy for phenotype discrimination and survival prediction.[27–29] Latent regulatory modules and ceRNA circuits underpinning tumor progression have been revealed through advanced ML pipelines,[30,31] and early decision support systems that combined ML with adrenal gene-expression profiles have been demonstrated.[32] Together, these advances have underscored the potential of ML to reshape ACC biomarker discovery and to guide personalized management.

Previous systems-level studies by the author have highlighted the importance of non-coding RNA-mediated regulation and ceRNA network rewiring in ACC, including integrative ceRNA network analyses identifying miR-507 and miR-665 as regulatory hubs in ACC,[33] network-driven biomarker discovery revealing the synergistic oncogenic roles of CKS2 and ACAT2,[34] and systems-level characterization of miR-466 as a tumor-suppressive regulatory miRNA in ACC.[35] In addition, a recent review has synthesized the current molecular landscape and biomarker discovery efforts in ACC, highlighting emerging regulatory mechanisms and potential diagnostic targets.[36]

Although ML algorithms such as RF have been previously applied in bioinformatics studies, their use for context-aware reconstruction of ceRNA regulatory networks in ACC remains largely unexplored. In particular, no prior study has systematically integrated ensemble ML, normal-versus-tumor network analysis, and model interpretability to prioritize miRNA-mRNA regulatory interactions in ACC using harmonized Genotype-Tissue Expression (GTEx) in 2025[37] and The Cancer Genome Atlas-ACC (TCGA-ACC)[38] transcriptomic datasets. A disease-focused integrative framework is presented in this study, in which ensemble ML is combined with network-based analysis and model explainability to reconstruct and prioritize context-specific miRNA-mRNA interactions in ACC. By comparing tumor and normal adrenal regulatory landscapes and integrating predictions with curated interaction databases, a systematic strategy is provided for identifying candidate regulatory hubs and biologically meaningful interaction modules. For the first time, advanced ensemble learning and interpretable modeling techniques are leveraged in ACC to uncover key regulatory hubs and interaction modules, findings that were further benchmarked against curated interaction datasets. By integrating the most current GTEx and TCGA-ACC expression profiles via in silico evaluation against recognized interaction repositories, this work delivers an unprecedented set of candidate regulatory hubs and network modules, providing a new foundation for biomarker discovery and therapeutic development in ACC.

2. Materials and methods

2.1. Data sources and preprocessing

Transcriptomic data for ACC were retrieved from TCGA-ACC and GTEx databases to ensure comprehensive representation of tumor and normal adrenal tissue. Specifically, RNA-Seq and miRNA-Seq datasets from 78 ACC samples were downloaded from TCGA-ACC, encompassing both gene and miRNA expression profiles. All tumor data were collated and merged across samples to generate structured expression matrices for downstream analysis.

To represent the normal tissue counterpart, RNA-Seq and miRNA-Seq data from 250 normal adrenal gland samples were obtained from GTEx.[37] For miRNA data, additional filtering was required to remove redundant entries and retain only mature human (hsa) miRNAs. The miRNA data have been reported recently in GTEx. Preprocessing steps included data harmonization, format standardization, and the generation of unified case-by-feature matrices for both RNA and miRNA datasets. A summary of the datasets, sample types, and processing steps is provided in Table 1.

2.2. Quality control and filtering strategies

Quality control and expression filtering were applied uniformly across mRNA and miRNA datasets to improve data reliability while preserving relevant biological signals. To eliminate transcriptionally inactive or near-zero expression genes, a filtering step was applied whereby genes with transcripts per million (for RNA-Seq) or reads per million (for miRNA) values less than 1 in more than 80% of the samples were excluded. This threshold was intentionally set at a permissive level to retain weakly expressed yet potentially informative transcripts, given the relatively small sample size in the ACC dataset. The aim was to minimize background noise while preserving subtle expression patterns that might otherwise be lost under stricter filtering criteria. Gene expression distributions were assessed visually after log-transformation to ensure consistency between tumor and normal tissue and to verify the effectiveness of the normalization process. No major deviations or technical artifacts were observed that warranted further sample removal. Although principal component analysis and unsupervised clustering were considered for outlier detection, no samples showed patterns indicative of being technical outliers. Filtering steps were implemented using custom R scripts with consistent thresholds across mRNA and miRNA data. This filtering approach ensured the retention of biologically informative features while reducing technical noise, providing a solid foundation for subsequent integrative and network-based analyses.

Table 1
Summary of transcriptomic datasets, sample types, and processing steps.

Source	Dataset name	Sample type	No. of samples	Data type	Platform
TCGA	TCGA-ACC	Tumor	78	RNA-Seq	Illumina HiSeq
GTEx	Adrenal Gland	Normal	250	RNA-Seq	Illumina HiSeq
TCGA	TCGA-ACC	Tumor	78	miRNA-Seq	Illumina HiSeq
GTEx	Adrenal Gland	Normal	250	miRNA-Seq	Illumina HiSeq

Abbreviations: 1. TCGA: The Cancer Genome Atlas 2. GTEx: Genotype-Tissue Expression 3. ACC: adrenocortical carcinoma 4 miRNA: microRNA.

2.3. Feature matrix construction

Expression profiles of miRNAs and mRNAs were transformed into numerical vectors representing each miRNA-mRNA pair. Pre-processed expression matrices were filtered to retain only hsa miRNAs, while duplicated identifiers were removed. For each valid interaction, expression values across matched samples were extracted and converted into 1D arrays. Only those pairs with equal vector lengths and non-zero variance in both components were selected for further processing.

Feature vectors derived from individual miRNA and mRNA expression profiles were concatenated to form a joint representation of their interaction. Each pair's combined vector was included in the final feature matrix, and the corresponding target variable was defined as the Pearson correlation coefficient between the two expression profiles. In this way, a unified feature matrix (X) and a response vector (Y) were constructed, capturing the coordinated regulatory dynamics of miRNA-mRNA pairs across biological samples.

To clarify the construction of the feature matrix, the process was implemented through the following sequential steps. First, matched expression matrices for miRNAs and mRNAs were obtained after preprocessing and filtering steps described in Sections 2.1 and 2.2. Each matrix contained expression values for all retained features across the same set of biological samples. Second, candidate miRNA-mRNA pairs were generated based on experimentally supported interactions obtained from the miRTarBase database.[39] For each pair, the expression profile of the corresponding miRNA and mRNA was extracted across all matched samples, producing two vectors of equal length representing their expression levels. Third, these two vectors were concatenated to form a combined feature representation for each interaction pair. Specifically, the miRNA expression vector and the corresponding mRNA expression vector were joined into a single feature vector describing the joint expression behavior of the pair across samples. Finally, the response variable for each pair was computed as the Pearson correlation coefficient between the miRNA and mRNA expression vectors. In this way, the final dataset consisted of a feature matrix X representing concatenated expression profiles of miRNA-mRNA pairs and a response vector Y containing the corresponding correlation values. This representation enabled ML models to learn patterns linking joint expression profiles to the strength of regulatory association.

The objective of this formulation is not to approximate the Pearson correlation itself, which is mathematically computable from the expression vectors, but rather to provide a supervised framework that allows ML models to learn complex expression patterns associated with strong regulatory coupling between miRNAs and their targets. In this context, the correlation coefficient serves as a continuous interaction-strength label that enables the model to rank candidate miRNA-mRNA pairs based on learned patterns from large numbers of training examples.

Unlike simple pairwise correlation analysis, which evaluates each pair independently, the ML model learns global relationships across the entire dataset and captures nonlinear dependencies between miRNA and mRNA expression patterns. This approach enables more robust prioritization of biologically meaningful regulatory interactions, particularly when applied to large-scale datasets containing thousands of candidate pairs.

Several quality control steps were applied to eliminate unreliable or uninformative features. Pairs with inconsistent vector shapes, zero standard deviation, or missing expression values were excluded. Additionally, to ensure the inclusion of experimentally validated interactions, only pairs confirmed by the miRTarBase dataset were selected and integrated with the ACC expression data. Additionally, only those sample columns without missing values in both miRNA and mRNA matrices were retained. Through this filtering process, a clean and statistically valid dataset was obtained, suitable for ML-based modeling of regulatory associations.

2.4. ML model training

A comprehensive evaluation of ML algorithms was conducted using five distinct models: RF, XGBoost regressor, LightGBM, gradient boosting regressor, and linear regression. These models were selected based on their established performance in high-dimensional biological datasets and their ability to capture complex nonlinear interactions. Both ensemble-based and linear approaches were considered, allowing performance comparisons across different algorithmic paradigms. Separate analyses were performed on normal and tumor-derived data from normal and tumor tissue to ensure context-specific optimization and robustness.

Transcriptomic profiles of miRNAs and mRNAs were independently retrieved from TCGA[38] (tumor samples, ACC) and GTEx[37] (normal adrenal gland tissue), ensuring biologically matched data from normal and tumor tissue for comparative modeling. Only human-specific pairs with adequate expression across all samples and no missing values were retained for training. The primary objective of model development was to perform regression analysis that could assign a quantitative score to each miRNA-mRNA pair, reflecting the predicted strength of their regulatory association. After filtering, each valid pair was encoded as a combined expression vector, and the Pearson correlation coefficient between their profiles was used as the response variable. The dataset was partitioned into training and test subsets using an 80:20 ratio. Model training was performed using default or heuristic-based hyperparameters. A five-fold cross-validation procedure was applied to each algorithm to evaluate generalization performance and reduce overfitting, particularly given the high dimensionality and biological variability inherent in transcriptomic data.

Model performance was quantitatively assessed using three key regression metrics: coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE). These metrics were calculated on the held-out test set for each model. Among all evaluated models, RF demonstrated superior predictive power, achieving the highest R^2 and the lowest MAE and MSE. The comparative performance of all models highlights the clear advantage of ensemble tree-based methods over linear regression in capturing the regulatory complexity of miRNA-mRNA interactions.

A limitation of this approach is that Pearson correlation captures statistical association rather than direct mechanistic regulation. Consequently, the resulting prediction score should be interpreted as a stabilized estimate of interaction strength derived from expression covariation, rather than as direct evidence of causal regulatory interaction. Furthermore, Pearson correlation may be influenced by confounding factors such as sample heterogeneity, indirect regulatory effects, or shared upstream drivers. The ML framework mitigates part of this limitation by integrating nonlinear modeling and ensemble aggregation, which reduces sensitivity to noise and outliers while enabling more robust prioritization of candidate regulatory interactions. The trained model is subsequently used to rank miRNA-mRNA interactions and to enable consistent comparison of interaction strength across biological conditions. In this context, the ML model acts as a denoising and regularization layer that integrates expression variability across samples, rather than as a deterministic calculator of correlation.

2.5. Model interpretation

Feature importance scores were computed using the built-in mechanism of the trained RF model. The importance of each feature, corresponding to individual miRNAs and mRNAs, was estimated based on the degree to which it contributed to reducing impurity across decision trees. Expression features were ranked according to these scores, and the top 10 most informative features were visualized using a horizontal bar plot. Gene and miRNA identifiers were mapped directly from the processed expression matrices, allowing biological interpretation of the model's prioritization. These rankings may assist in identifying key regulatory molecules within the ACC-specific transcriptomic landscape.

Model explainability was further explored using SHapley Additive exPlanations (SHAP) values, which provide consistent and locally accurate estimates of each feature's contribution to individual predictions. SHAP analysis was restricted to high-confidence samples with predicted correlation scores above 0.7, ensuring focus on biologically meaningful interactions. A TreeExplainer object was applied to a subset of the feature matrix, and SHAP values were calculated for the top 300 features. A beeswarm plot was generated to summarize the distribution and impact of features across all evaluated interactions, offering insights into the dominant transcriptomic signals driving model predictions. Unlike traditional importance scores, SHAP provides both global and local interpretability, offering a more nuanced understanding of regulatory dynamics in the tumor network.

2.6. Scoring and ceRNA network generation

Prediction scores representing the regulatory association strength between miRNAs and mRNAs were generated using RF regression models trained on matched expression data. For each miRNA-mRNA pair, the expression profiles were vectorized, combined, and used as input features, while the Pearson correlation between expression levels served as the target value. Separate models were trained and applied to tumor samples from TCGA-ACC and normal adrenal samples from GTEx, enabling the identification of context-specific regulatory patterns. These scores reflected the learned miRNA-mRNA relationships and were subsequently used for filtering and network construction.

Following score generation, a filtering step was employed to retain only high-confidence interactions. The primary network was constructed using a score threshold of 0.7, and additional sensitivity analyses were performed using thresholds of 0.8 and 0.9. This thresholding approach minimized the inclusion of weak or potentially spurious interactions, thereby improving the specificity and interpretability of the inferred regulatory network.

A ceRNA network was constructed by integrating high-confidence miRNA-mRNA interactions with additional mRNA-mRNA edges derived from co-expression analysis. For each miRNA, all mRNA targets jointly regulated by that miRNA were identified, and pairwise Pearson correlations were computed among them. mRNA-mRNA interactions were added if their correlation coefficient exceeded 0.4 with an associated P value below 0.05, indicating statistically significant co-regulation. The resulting network consisted of bipartite edges (miRNA-mRNA) and lateral, functionally linked mRNA-mRNA interactions, capturing both regulatory input and co-expressed competitive endogenous relationships within the tumor and normal transcriptomes. This comprehensive interaction map served as the structural foundation for subsequent topological and hub-based analyses.

Topological characteristics of the constructed networks were analyzed to identify key regulatory and structural components. Centrality measures, including node degree, closeness centrality, and betweenness centrality, were computed for each node using standard graph algorithms. Nodes with high degree were considered potential hubs due to their extensive connectivity, while those with high betweenness were interpreted as critical intermediaries in information flow. The results were stratified by node type (miRNA or mRNA) and stored for further comparison between tumor and normal ceRNA architectures. These metrics enabled the identification of candidate regulatory hubs with potential relevance to ACC pathophysiology.

2.7. Computational tools and implementation

Core preprocessing and statistical analyses were implemented in R (version 4.5.1). Transcriptomic count matrices were first imported and organized into expression matrices for downstream modeling. Lowly expressed genes were removed using the edgeR framework based on transcript abundance across samples. Specifically, genes with expression values below 1 TPM in more than 80% of samples were excluded using a filtering procedure implemented as:

```
filter <- rowSums(expr_matrix > 1) >= 0.2 * ncol(expr_matrix)
expr_filtered <- expr_matrix[filter, ]
```

This filtering step reduced background noise while retaining weak but potentially informative transcripts.Normalization and linear modeling were subsequently performed using the limma-voom pipeline to stabilize the mean-variance relationship in RNA-seq data. Raw count data were transformed using the voom function prior to empirical Bayes moderation:

```
library(limma)
library(edgeR)
dge <- DGEList(counts = expr_filtered)
dge <- calcNormFactors(dge)
v <- voom(dge, design)
fit <- lmFit(v, design)
fit <- eBayes(fit)
```

Normalized expression matrices were exported and subsequently used as input for downstream ML analyses implemented in Python (version 3.13.5). Transcriptomic expression matrices for miRNAs and mRNAs were first imported using the pandas library and organized into structured data frames. Raw expression tables were reshaped into expression matrices using pivot operations, allowing miRNA and gene identifiers to form matrix indices while sample identifiers were assigned to columns.

Similarly, mRNA expression matrices were filtered and harmonized by removing non-coding transcript identifiers and standardizing gene names prior to matrix construction. To ensure comparability between regulatory components, matched samples shared by both expression matrices were retained through an intersection of sample identifiers. Candidate miRNA-mRNA regulatory pairs were then filtered based on a curated interaction table. Only hsa miRNA entries and genes present in the expression matrices were retained, generating a validated interaction list for downstream modeling.

A feature matrix was subsequently constructed by concatenating the expression profiles of each miRNA-mRNA pair across matched samples. For a given interaction pair, the miRNA and mRNA expression vectors were merged into a single feature vector, while the Pearson correlation coefficient between the two profiles was used as the response variable:

```
X = np.concatenate([mirna_vec, mrna_vec])
Y = pearsonr(mirna_vec, mrna_vec)[0]
```

The resulting feature matrix X and response vector Y were used for supervised regression modeling. Model training and evaluation were performed using the scikit-learn framework. The dataset was partitioned into training and testing subsets using an 80/20 split:

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
```

A RF regression model was then trained to predict miRNA-mRNA association strength:

```
model = RandomForestRegressor(n_estimators=100, n_jobs=-1,
random_state=42)
model.fit(X_train, Y_train)
```

Five-fold cross-validation was performed using the `cross_val_score` function implemented in scikit-learn to reduce potential overfitting during the regression modeling stage:

```
cross_val_score(model, X, Y, cv=5)
```

Model performance was evaluated on the held-out test set using R^2, MAE, and MSE. The trained model was subsequently applied to all candidate miRNA–mRNA pairs to generate prediction scores representing regulatory association strength. These predicted scores were subsequently used to construct regulatory networks under multiple confidence thresholds (0.7, 0.8, and 0.9). For each threshold level, edges representing high-confidence miRNA-mRNA interactions were retained, and additional mRNA-mRNA associations were derived from pairwise expression correlations exceeding 0.4. Network size and connectivity metrics, including the number of nodes and edges, were then calculated to evaluate the robustness of the inferred regulatory architecture.

For network visualization and high-resolution figure generation, Python (version 3.13.5) was employed. Expression data matrices and interaction results were first structured and processed using pandas and numpy. Correlation heatmaps, volcano plots, and other expression-based visualizations were generated using matplotlib and seaborn. Graph-based visualizations of the miRNA-mRNA regulatory network were created with networkx, allowing for computation of topological metrics such as betweenness centrality and layout optimization. Where necessary, dimensionality reduction and clustering visualization were facilitated by scikit-learn. In specific instances involving ML-based prediction and model interpretation, SHAP and lime were optionally utilized to illustrate feature importance and interaction effects.

2.8. Computational and analytical workflow

In this study, the analytical pipeline for miRNA-mRNA interaction prediction in ACC followed a structured sequence of computational steps (Figure 1). First, harmonized RNA-Seq and miRNA-Seq data from TCGA-ACC and GTEx were acquired and preprocessed to generate matched tumor and normal expression matrices. Multiple regression-based ML models were then benchmarked, leading to the selection of RF as the top-performing predictor. The selected model's performance was assessed through metrics such as R^2, MAE, and MSE to ensure robustness. High-confidence miRNA-mRNA pairs were subsequently identified, forming the predicted interaction landscape. These predictions were cross-referenced with curated interaction databases (miRTarBase and TargetScan[40]) to confirm biological plausibility. The database-supported interactions were integrated into a ceRNA network, enabling topological analysis and biological interpretation of key regulatory hubs. This workflow ensured that only statistically and biologically supported interactions progressed to downstream network-level investigation, thereby enhancing the reliability and translational potential of the findings.

3. Results

3.1. Model performance comparison

The performance of five regression models was systematically evaluated on tumor-derived data from ACC samples. As shown in Figure 2, the RF model outperformed all others across all evaluation metrics, achieving the lowest MSE (MSE = 0.0023) and MAE (MAE = 0.0242), while yielding the highest R^2 (R^2 = 0.9044). XGBoost regressor and LightGBM, both gradient boosting-based approaches, demonstrated moderate performance, with R^2 values of 0.5635 and 0.4308, respectively. Linear regression and gradient boosting regressor yielded substantially lower predictive accuracy, with R^2 values below 0.3 and markedly higher error rates. These findings suggest that the RF model captured the underlying nonlinear relationships between miRNA and mRNA expression most effectively and was therefore selected as the optimal predictor for downstream analyses on tumor samples.

3.2. ML model performance

Scatter plots of predicted values in comparison with true Pearson correlation values were generated for both normal and tumor tissue using the RF regression model. In Figure 3A, corresponding to normal adrenal tissue, a strong alignment along the identity line was observed, reflecting high predictive performance (R^2 = 0.9467, MAE = 0.0116, MSE = 0.0006). In contrast, Figure 3B, based on tumor data from ACC samples, demonstrated a slightly broader dispersion, though still indicating strong agreement between observed and predicted values (R^2 = 0.9045, MAE = 0.0242, MSE = 0.0023). The performance stability was further supported by 5-fold cross-validation, yielding a mean R^2 of 0.9346 (± 0.0151) in the normal model and 0.8811 (± 0.0366) in the tumor model.

Prediction residuals and error patterns were evaluated for both normal and tumor datasets to assess model accuracy and bias. In Figure 3C (normal) and 3D (tumor), the distribution of true versus predicted

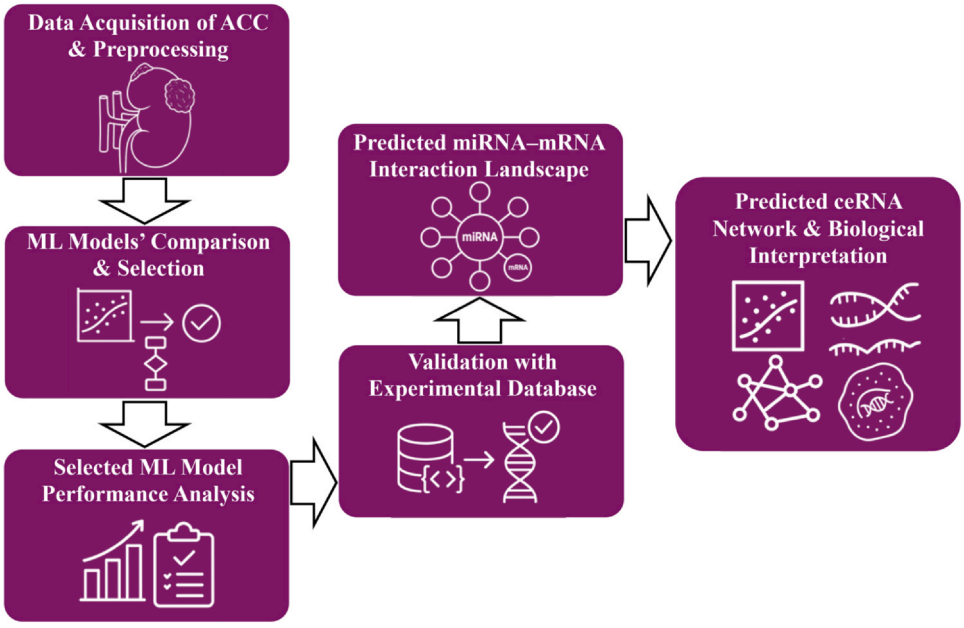


Figure 1. Workflow for predicting miRNA-mRNA interactions in normal adrenal tissue and ACC tissue. Abbreviations: 1. ML: machine learning 2. ACC: adrenocortical carcinoma 3. miRNA: microRNA 4. mRNA: messenger RNA 5. ceRNA: competing endogenous RNA.

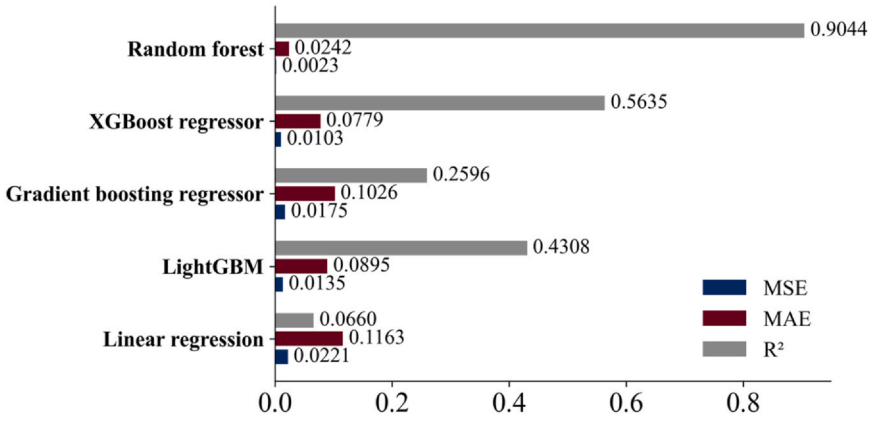


Figure 2. Performance comparison of five regression models on tumor-derived ACC transcriptomic data. Abbreviations: 1. MSE: mean squared error 2. MAE: mean absolute error 3. R^2: coefficient of determination.

correlation scores showed close alignment, with nearly symmetric and centered histograms in both conditions. The corresponding residual distributions in Figure 3E (normal) and F (tumor) revealed sharp peaks around zero, indicating minimal deviation between predicted and observed values across the majority of interactions. Additionally, Figure 3G (normal) and H (tumor) display absolute and signed prediction errors across low, medium, and high gene expression groups. In both datasets, error magnitudes remained low across expression levels, though slightly higher variability was observed in the tumor condition, particularly within the high expression group. These analyses confirmed the robustness and generalization capability of the model across different expression contexts.

3.3. Feature importance and explainability

Feature importance was evaluated based on impurity reduction in RF models trained separately on normal and tumor datasets. As shown in Figure 4A, the most influential miRNAs in the normal condition included hsa-mir-15a, hsa-mir-3677, and hsa-mir-3609, which exhibited the highest contribution to model prediction. These features were primarily associated with strong regulatory signals across the normal adrenal transcriptome. In contrast, Figure 4B revealed a distinct set of miRNAs as top contributors in the tumor context, with hsa-mir-1537, hsa-mir-138-1, and hsa-mir-10a ranking highest in importance. The divergence in dominant features between normal and tumor models suggested substantial rewiring of regulatory networks in ACC. These findings highlighted condition-specific miRNA drivers potentially involved in tumorigenesis.

Leveraging SHAP to interrogate the RF regressors trained on high-confidence interaction pairs (prediction score >0.70) revealed a sharp divergence in the miRNA drivers that modulate correlation strength in normal versus ACC tissue. In the normal model, a compact subset of miRNAs, most prominently hsa-mir-329-1, hsa-mir-155, and hsa-mir-3675, exerted uniformly positive SHAP contributions, indicating that higher expression of these regulators consistently elevates the predicted miRNA-mRNA correlation (Figure 4C). By contrast, the tumor model was dominated by a different cadre of miRNAs, headed by hsa-mir-126, hsa-mir-153-2, and hsa-mir-1255a, whose SHAP distributions spanned a wider positive range (up to approximately 0.30) and displayed more heterogeneous density patterns (Figure 4D). This shift underscores a substantial rewiring of post-transcriptional control in ACC, whereby tumor-specific miRNAs gain influence while homeostatic regulators lose prominence.

A striking feature of both beeswarm plots is the scarcity of negative SHAP contributions. The primary explanation is methodological: SHAP values were calculated only for interaction pairs whose predicted Pearson correlation exceeded 0.70, a narrow, high-confidence slice of the data, rather than the full correlation spectrum was examined. A secondary factor is compounded by this effect: within that filtered subset, strong positive correlations dominate, so the model output is naturally nudged further in the positive direction by most perturbations of the input features, leaving little room for sizable negative impacts. This interpretation is supported by the global patterns in Figure 3A-B, where truly negative correlations are comparatively rare once the same score filter is applied. Taken together, Figures 3A-B and 4C-D show that a tumor-biased cadre of miRNAs controls the strongest miRNA-mRNA couplings in ACC, and that our explainability analysis faithfully reflects this focused, positively skewed landscape.

3.4. Predicted miRNA-mRNA interaction landscape

High-confidence miRNA-mRNA interactions were identified by selecting the top 50 predicted pairs with the highest correlation scores in both normal and tumor tissue. In the normal dataset, top scores ranged from 0.73 to 0.82, while in the tumor dataset, the range extended from 0.84 to 0.98, indicating a stronger regulatory signal captured in the cancer context. The corresponding Sankey plots in Figure 5A-B visualize the regulatory architecture formed by these interactions. In the network in normal tissue, miRNAs such as miR-485, miR-7, and miR-329 were linked to a moderate number of diverse mRNA targets, suggesting balanced and distributed regulatory roles (Figure 5A). In contrast, the network in tumor tissue exhibited a denser and more centralized interaction pattern, with miRNAs such as miR-205 and miR-4772 connecting to a broader set of transcripts, potentially reflecting enhanced miRNA activity or compensatory regulatory expansion under malignant transformation (Figure 5).

To further investigate the model's predicted regulatory landscape, heatmaps of predicted correlation scores for the top 50 miRNA-mRNA pairs were generated (Figure 5C-D). In the normal condition, the heatmap exhibited a compact and relatively uniform structure, with most scores clustered within a narrow range, reflecting consistent yet moderate predicted interaction strengths (Figure 5C). In contrast, the tumor heatmap demonstrated greater score variability and wider distribution, with several high-scoring regions, indicating a stronger and more heterogeneous regulatory influence predicted by the model (Figure 5D). These differences may reflect a reprogrammed post-transcriptional environment in ACC, characterized by more pronounced or diversified miRNA-driven modulation of target transcripts.

3.5. Comparison with known interaction databases

Two widely used reference datasets, miRTarBase and TargetScan, were employed to evaluate the biological relevance of top-scoring predicted interactions. The datasets were first filtered to retain only human-specific entries, and in parallel, the top 100 predicted miRNA-mRNA pairs with the highest model scores were selected separately from normal and tumor data. These high-confidence predictions were then cross-referenced against both databases to assess their overlap and assess potential biological support.

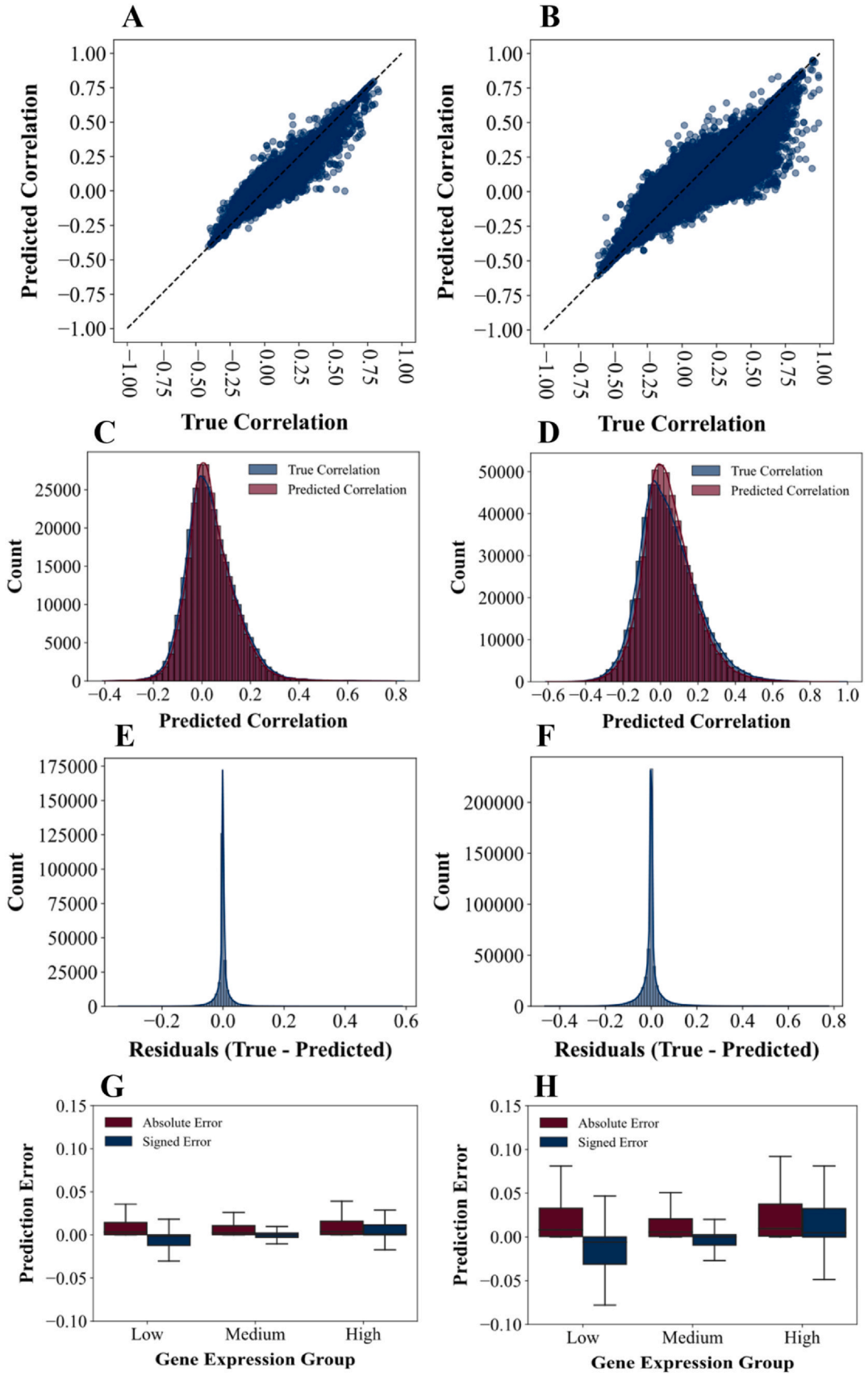


Figure 3. Random forest regression modeling of miRNA-mRNA interactions in normal adrenal and ACC tissue. A & B, Scatter plots comparing predicted versus true Pearson correlation values for normal (A) and tumor (B) tissue; C–H, Comparison of prediction distributions (C, D), residuals (E, F), and error stratification (G, H) for normal tissue (C, E, G) and tumor tissue (D, F, H).

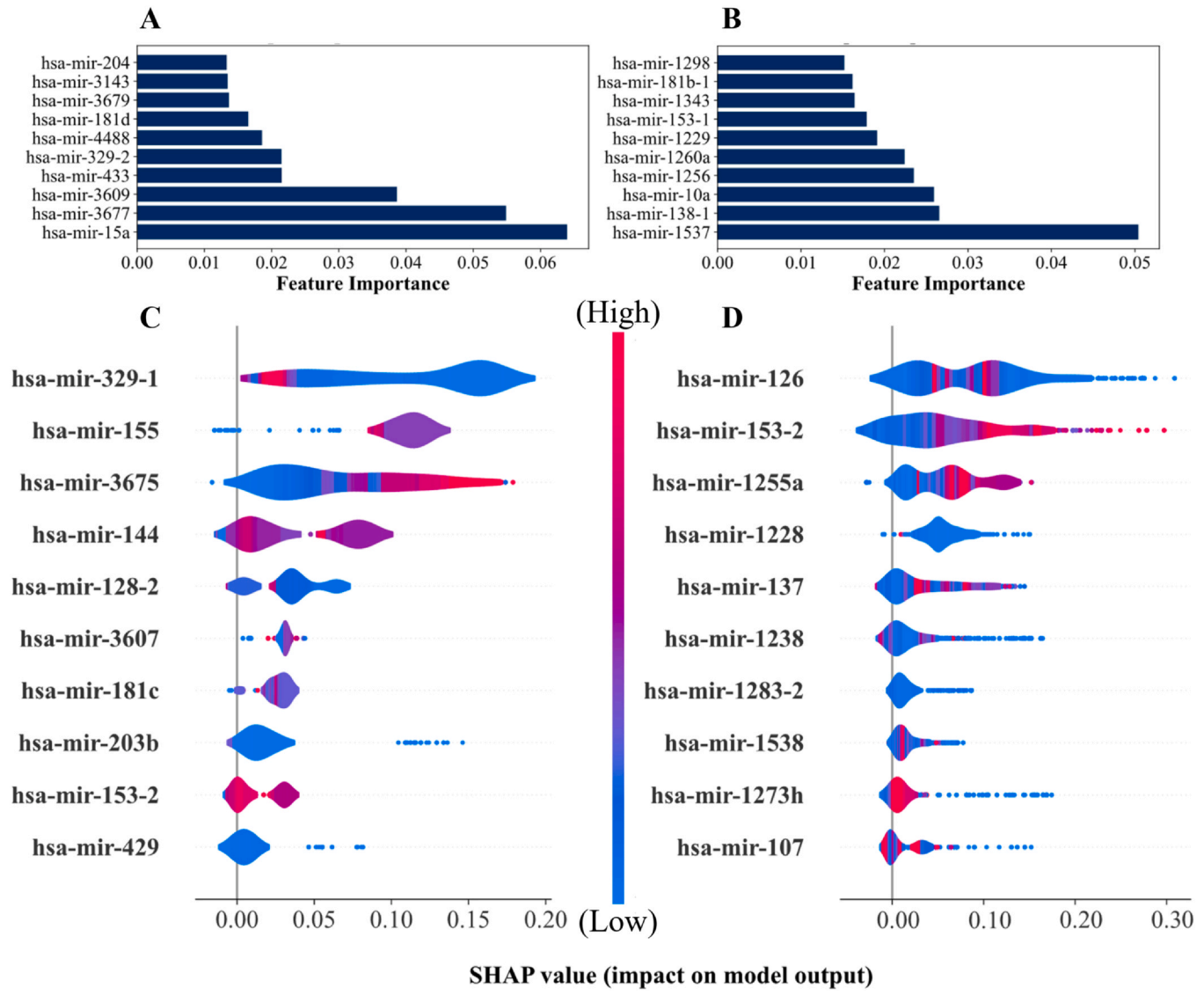


Figure 4. Top miRNA features and SHAP analysis of model predictions. A & B, Top 10 most important miRNA features in normal (A) and tumor (B) tissue based on random forest feature importance scores; C & D, SHAP beeswarm plots for the top 10 miRNA features driving predicted miRNA-mRNA correlations (prediction score >0.70) in normal (C) and tumor (D) tissue. Color denotes relative miRNA expression (blue = low, pink = high); the x-axis shows feature impact on model output. Abbreviation: SHAP: SHapley Additive exPlanations.

The results of this comparison are summarized in Figure 6A-B. In the normal tissue, 44 of the top 100 predicted pairs were found to overlap with both miRTarBase and TargetScan, indicating strong support from both computational and experimental evidence (Figure 6A). In the tumor tissue, 36 overlapping pairs were identified (Figure 6B). These intersecting interactions were considered the most reliable subset for further biological interpretation.

It is important to note that the ML model predicted interaction strength based on expression-derived patterns and therefore reflected statistical association rather than direct regulatory causality. The overlap with curated databases such as miRTarBase and TargetScan was used solely as an external source of biological support for a subset of predicted pairs. Consequently, interactions identified in this comparison should be interpreted as computationally prioritized candidates supported by existing evidence rather than experimentally validated regulatory relationships.

Sankey plots in Figure 6C-D illustrate these database-supported interactions, with Figure 6C representing the 44 consensus pairs in the normal tissue and Figure 6D representing the 36 pairs in the tumor tissue. In the normal network, miRNAs such as miR-7, miR-329, miR-149, and miR-592 were shown to regulate a diverse set of targets with corroborating evidence from both databases. The tumor-specific plot revealed prominent regulators including miR-205, miR-223, and miR-214, each associated with multiple database-supported targets. These database-supported interactions not only reinforce the credibility of the ML-based predictions but also highlight condition-specific miRNA regulators with potential functional relevance in ACC.

A subset of highly confident miRNA-mRNA interactions was identified among the tumor-specific predictions, characterized by model scores exceeding 0.90. These interactions were not present in either miRTarBase or TargetScan, suggesting the possibility of novel or context-specific regulatory relationships in ACC. Notably, eight tumor-specific pairs with particularly high scores were hsa-mir-205-IL24, hsa-mir-6499-SECTM1, hsa-mir-4772-IFNE, hsa-mir-548az-IGF2BP1, hsa-mir-199b-MMP2, hsa-mir-214-COL6A3, hsa-mir-4772-FFAR2, and hsa-mir-7847-BMP3. These pairs exhibited predicted regulatory strength surpassing many of the previously reported interactions in reference databases and may represent tumor-enriched regulatory events undetectable under normal physiological conditions.

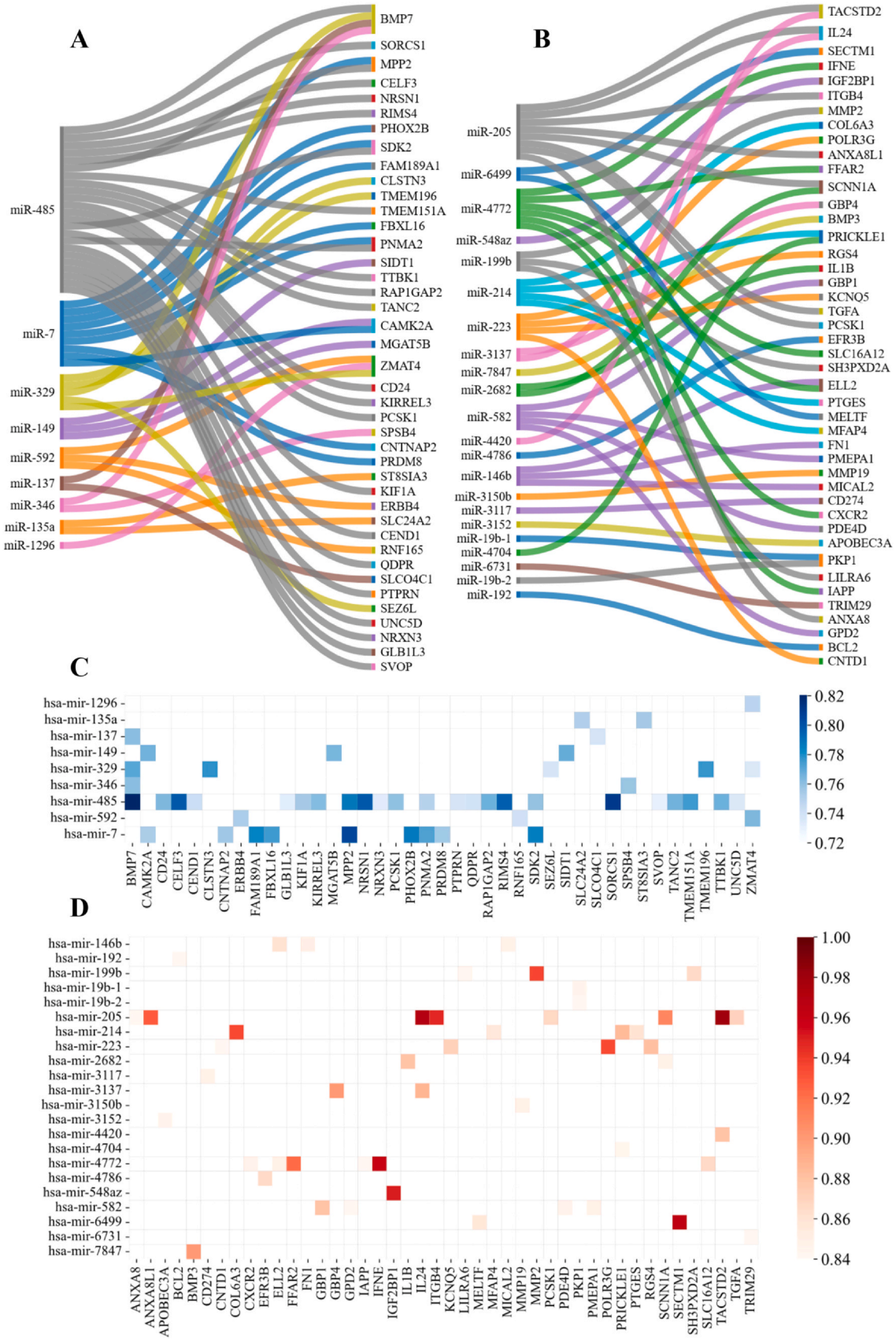


Figure 5. Predicted top 50 miRNA-mRNA interactions and correlation scores. A & B, Sankey plots showing the top 50 predicted miRNA-mRNA interactions in normal (A) and tumor (B) tissue based on the highest predicted correlation scores from random forest models; C & D, Heatmaps representing the prediction scores for the top 50 predicted miRNA-mRNA pairs in normal (C) and tumor (D) tissue.

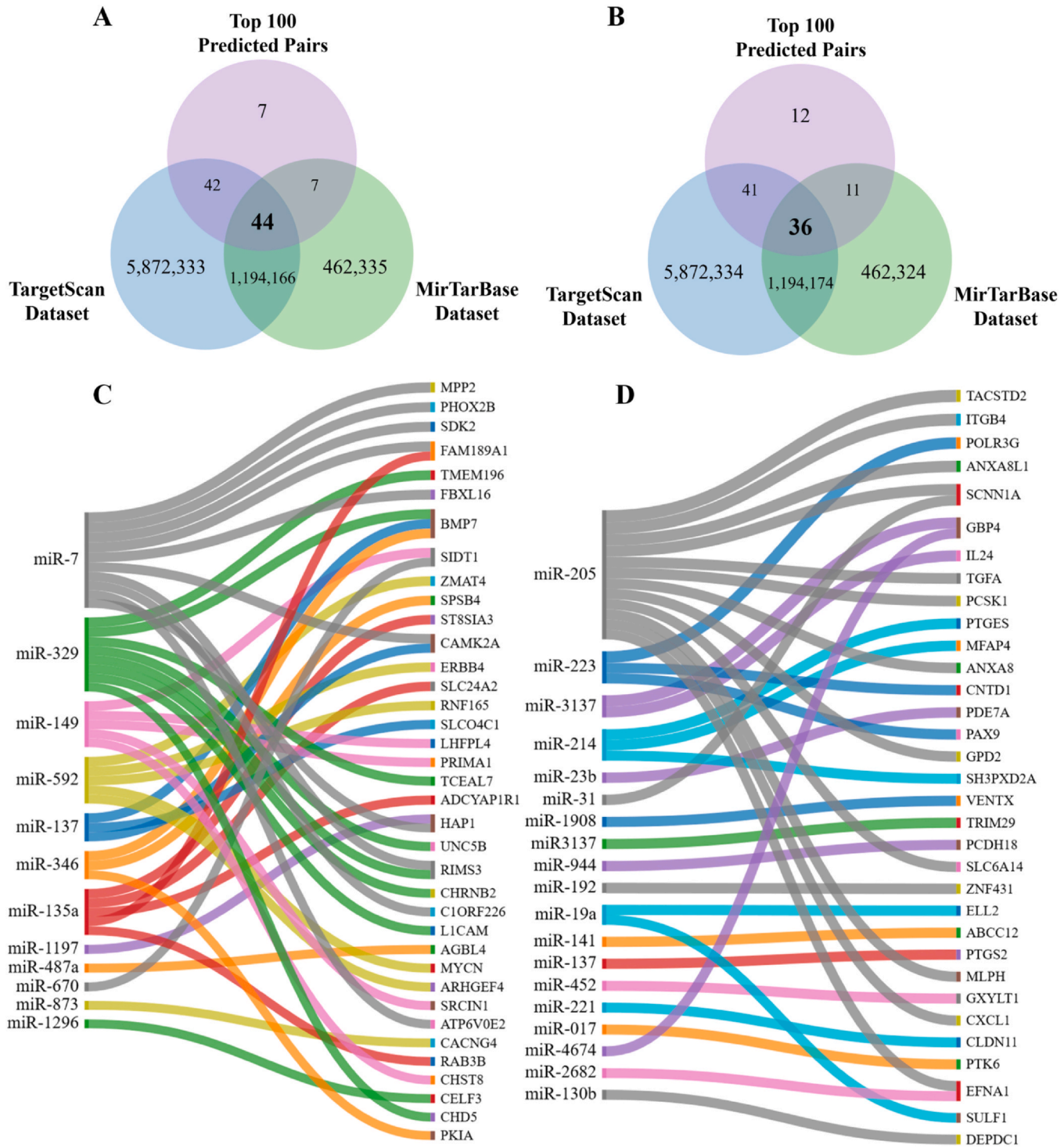


Figure 6. Overlap of predicted miRNA-mRNA interactions with validated databases. A & B, Venn diagrams showing the overlap between the top 100 predicted miRNA-mRNA interactions and entries in miRTarBase and TargetScan databases for normal (A) and tumor (B) tissue; C & D, Sankey plots illustrating the overlapping miRNA-mRNA interactions supported by both miRTarBase and TargetScan databases among the top 100 predicted pairs in normal (C) and tumor (D) tissue.

Several of these targets are involved in cancer-associated pathways. For instance, MMP2 and COL6A3 are known to modulate extracellular matrix remodeling and tumor invasion, while IL24, BMP3, and IGF2BP1 have been implicated in immune signaling and oncogenic processes across various malignancies. Their appearance as high-confidence miRNA targets in the tumor context may reflect tumor-specific activation of compensatory or oncogenic pathways governed post-transcriptionally by miRNAs. The full list of new high-scoring miRNA-mRNA interactions is provided in Table 2, comprising 64 unique pairs in the tumor dataset and 56 in the normal dataset. These predictions, though not yet supported by existing interaction databases, offer a rich resource for future experimental investigation and may contribute to the discovery of previously unrecognized layers of gene regulation in ACC.

To further investigate the biological relevance of database-supported interactions, protein-protein interaction (PPI) networks were constructed for the database-supported miRNA-mRNA (Figure 6A-B). As illustrated in Figure 7A (normal) and 7B (tumor), networks were generated using the STRING database[41] with medium confidence (score

Table 2
List of new high-scoring miRNA-mRNA interactions identified in the tumor and normal tissue.

Normal tissue			Tumor tissue		
miRNA	mRNA	Score	miRNA	mRNA	Score
hsa-mir-485	BMP7	0.822678	hsa-mir-205	IL24	0.96939
hsa-mir-485	SORCS1	0.810574	hsa-mir-6499	SECTM1	0.964284
hsa-mir-485	CELF3	0.797858	hsa-mir-4772	IFNE	0.959806
hsa-mir-485	NRSN1	0.797622	hsa-mir-548az	IGF2BP1	0.950735
hsa-mir-485	RIMS4	0.795677	hsa-mir-199b	MMP2	0.935874
hsa-mir-485	MPP2	0.787412	hsa-mir-214	COL6A3	0.935303
hsa-mir-329	CLSTN3	0.778377	hsa-mir-4772	FFAR2	0.92143
hsa-mir-485	TMEM151A	0.775236	hsa-mir-7847	BMP3	0.900262
hsa-mir-7	PNMA2	0.772844	hsa-mir-214	PRICKLE1	0.884073
hsa-mir-485	TTBK1	0.766177	hsa-mir-223	RGS4	0.880115
hsa-mir-485	RAP1GAP2	0.766138	hsa-mir-2682	IL1B	0.879338
hsa-mir-485	TANC2	0.766049	hsa-mir-582	GBP1	0.879076
hsa-mir-149	CAMK2A	0.765818	hsa-mir-4420	TACSTD2	0.878834
hsa-mir-149	MGAT5B	0.76452	hsa-mir-223	KCNQ5	0.871829
hsa-mir-485	CD24	0.76303	hsa-mir-4786	EFR3B	0.865378
hsa-mir-485	KIRREL3	0.761461	hsa-mir-4772	SLC16A12	0.865143
hsa-mir-485	PCSK1	0.759088	hsa-mir-199b	SH3PXD2A	0.864624
hsa-mir-485	SDK2	0.759035	hsa-mir-146b	ELL2	0.861416
hsa-mir-7	CNTNAP2	0.755275	hsa-mir-6499	MELTF	0.855652
hsa-mir-7	PRDM8	0.755024	hsa-mir-146b	FN1	0.849613
hsa-mir-485	KIF1A	0.754503	hsa-mir-582	PMEPA1	0.848944
hsa-mir-485	PNMA2	0.74879	hsa-mir-4772	ELL2	0.848876
hsa-mir-1296	ZMAT4	0.747821	hsa-mir-3150b	MMP19	0.848404
hsa-mir-485	CEND1	0.742449	hsa-mir-146b	MICAL2	0.847267
hsa-mir-485	QDPR	0.738649	hsa-mir-3117	CD274	0.847023
hsa-mir-485	PTPRN	0.737363	hsa-mir-2682	SCNN1A	0.846987
hsa-mir-329	SEZ6L	0.736669	hsa-mir-4772	CXCR2	0.846777
hsa-mir-329	ZMAT4	0.735235	hsa-mir-582	PDE4D	0.846773
hsa-mir-485	UNC5D	0.734661	hsa-mir-3152	APOBEC3A	0.843983
hsa-mir-485	NRXN3	0.732447	hsa-mir-19b-1	PKP1	0.843817
hsa-mir-485	GLB1L3	0.731484	hsa-mir-4704	PRICKLE1	0.843312
hsa-mir-485	SVOP	0.729224	hsa-mir-199b	LILRA6	0.841902
hsa-mir-592	CD24	0.729146	hsa-mir-4772	IAPP	0.841232
hsa-mir-149	TM6SF2	0.727832	hsa-mir-6731	TRIM29	0.841098
hsa-mir-329	SLC6A17	0.727543	hsa-mir-19b-2	PKP1	0.840651
hsa-mir-485	RHPN2	0.727372	hsa-mir-582	GPD2	0.839473
hsa-mir-485	SYT11	0.726824	hsa-mir-192	BCL2	0.839139
hsa-mir-485	GNG4	0.721761	hsa-mir-3117	GREM1	0.836347
hsa-mir-329	PPP2R2C	0.720872	hsa-mir-4420	SRGN	0.834783
hsa-mir-329	NAT16	0.719137	hsa-mir-4433b	EMP1	0.827083
hsa-mir-149	STMN4	0.718741	hsa-mir-208a	SP8	0.826231
hsa-mir-149	RAP1GAP2	0.714978	hsa-mir-6499	SLC7A5	0.82604
hsa-mir-149	KIF1A	0.711112	hsa-mir-582	IGFBP4	0.822331
hsa-mir-485	KCNQ5	0.71093	hsa-mir-205	SLC7A5	0.820583
hsa-mir-149	PNMA2	0.709827	hsa-mir-6731	SAA1	0.816992
hsa-mir-487a	ZDHHC22	0.708606	hsa-mir-221	ADCYAP1	0.815002
hsa-mir-211	PTPRT	0.708056	hsa-mir-4786	DIO2	0.812931
hsa-mir-485	AMOTL1	0.707522	hsa-mir-4772	PLEKHG2	0.812545
hsa-mir-485	SYT13	0.706918	hsa-mir-199b	CERCAM	0.812126
hsa-mir-485	SCN3B	0.706405	hsa-mir-582	NET1	0.811492
hsa-mir-296	RIMS4	0.705516	hsa-mir-146b	CXCR1	0.808679
hsa-mir-485	VGF	0.704701	hsa-mir-222	CLDN11	0.808407
hsa-mir-137	CERS1	0.704185	hsa-mir-6731	RUNX3	0.806699
hsa-mir-7	KCNH2	0.702447	hsa-mir-146b	GBP5	0.806657
hsa-mir-485	FAIM2	0.702447	hsa-mir-4420	IRX3	0.806146
hsa-mir-296	DRD2	0.701249	hsa-mir-2682	SLC6A14	0.805577
-	-	-	hsa-mir-4772	ITPRIPL1	0.805274
-	-	-	hsa-mir-2682	SHISA2	0.805006
-	-	-	hsa-mir-130b	SLC7A5	0.803278
-	-	-	hsa-mir-20a	FOXL1	0.802825
-	-	-	hsa-mir-146b	NAV2	0.801249
-	-	-	hsa-mir-137	IL6	0.800462
-	-	-	hsa-mir-146b	EGR3	0.799996
-	-	-	hsa-mir-582	PCSK1	0.796783

threshold: 0.4), incorporating a maximum of five interactors per node due to the limited number of database-supported target genes. The full STRING network was selected, integrating evidence from text mining, experimental data, co-expression, and curated databases.

The PPI network corresponding to the normal condition revealed a highly clustered topology centered around synaptic signaling components such as GRIN1, GRIN2A, GRIN2D, and CAMK2A, suggesting tight regulatory modules in non-tumor samples (Figure 7A). In contrast, the network in tumor tissue featured key mediators including PTGS2, ERBB2, CXCL1, and IL24, pointing toward involvement in inflammation and growth factor pathways, potentially indicative of tumor-specific reprogramming (Figure 7B).

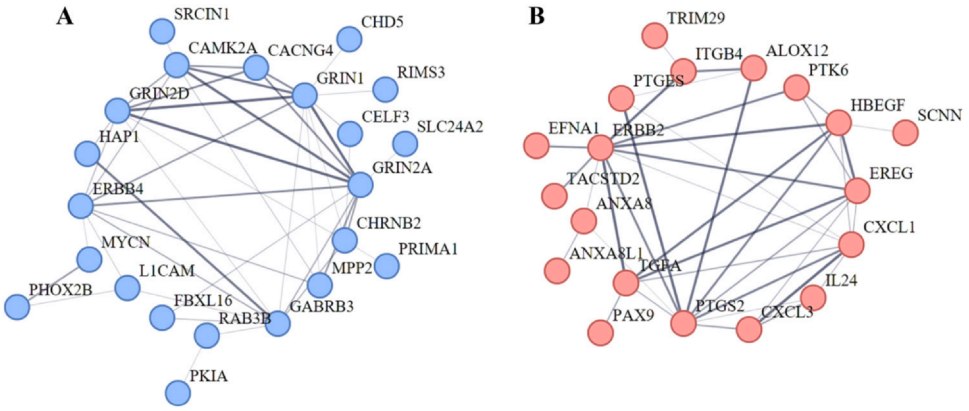


Figure 7. Protein-protein interaction networks of target mRNAs in predicted miRNA-mRNA pairs in normal (A) and tumor (B) tissue. Blue and red nodes represent genes in normal and tumor tissue, respectively.

Table 3
Summary of topological parameters for PPI networks in normal and tumor tissue using STRING.

Network type	No. of node	No. of degree	Average node degree	Average local clustering coefficient	PPI enrichment *P*-value
Normal	42	46	2.19	0.441	2.89×10^{-15}
Tumor	38	41	2.16	0.389	2.01×10^{-8}

Abbreviation: PPI: protein-protein interaction.

A comparison of network metrics is presented in Table 3. The normal PPI network comprised 42 nodes and 46 edges, with an average node degree of 2.19 and a clustering coefficient of 0.441. In contrast, the network in tumor tissue included 38 nodes and 41 edges, with a slightly lower average node degree (2.16) and clustering coefficient (0.389). Notably, the PPI enrichment *P*-values were significant in both conditions (normal: 2.89×10^{-15}; tumor: 2.01×10^{-8}), confirming that the observed interactions were unlikely to occur by chance and supporting the functional coherence of the predicted targets.

3.6. Predicted ceRNA network

High-confidence ceRNA networks were constructed by selecting the 44 database-supported miRNA-mRNA pairs (supported by both miRTarBase and TargetScan) in the normal tissue and 36 database-supported pairs in the tumor tissue, as identified in Section 3.5 (Figure 6A-B). The resulting networks are visualized in Figure 8A (normal) and 8B (tumor), where blue nodes represent miRNAs and red nodes indicate mRNAs.

The network in normal tissue exhibits a denser and more interconnected structure, with a higher number of edges and clustered regions, reflecting a more robust regulatory landscape under physiological conditions (Figure 8A). In contrast, the network in normal tissue displays reduced connectivity and more dispersed nodes, suggesting a loss or rewiring of regulatory interactions in the tumor tissue (Figure 8B). Notably, the number of database-supported interactions is lower in the network in tumor tissue (36 pairs) compared to the network in normal tissue (44 pairs), consistent with a potential disruption of ceRNA crosstalk during tumorigenesis. Betweenness centrality was calculated for each node to further investigate topological features; the network in normal tissue contained several high-centrality miRNAs and mRNAs acting as key hubs, while in the network in tumor tissue, hub nodes appeared less prominent, with a tendency toward fragmentation.

A comparative analysis of network centrality metrics revealed distinct patterns of hub node organization between tumor and normal ceRNA networks. These centrality-based comparisons are primarily descriptive and are intended to highlight potential regulatory hubs rather than establish statistically significant differences between conditions. Formal statistical testing of network topology differences represents an important direction for future work as larger matched cohorts become available. In the tumor network, the genes SCNN1A, EFNA1, hsa-mir-205, TACSTD2, and ITGB4 consistently ranked among the top nodes by closeness and degree centrality, highlighting their central role in network communication and connectivity. Notably, SCNN1A and EFNA1 emerged as dominant hubs across all three metrics, while GBP4 was uniquely prominent for betweenness centrality, suggesting its critical function as a bottleneck for information flow in the tumor context. The prominence of hsa-mir-205 among the top hubs underscores its putative regulatory importance in tumor-specific ceRNA interactions. In contrast, the network in normal tissue exhibited a broader and more diverse set of central nodes. RIMS3 demonstrated the highest centrality across all metrics, indicating its central integrative role under physiological conditions. Other notable hubs included CAMK2A, FAM189A1, BMP7, and hsa-mir-7, each of which appeared repeatedly among the top-ranked nodes for closeness, degree, and betweenness centrality. The presence of both mRNAs and miRNAs as key hubs in the normal tissue reflects a more distributed and redundant regulatory architecture, potentially conferring robustness against perturbations. These findings illustrate a marked shift in network topology from a robust, highly interconnected normal tissue to a more fragmented and hierarchically constrained architecture in tumors, with key hubs emerging that may represent potential drivers or vulnerabilities in ACC.

Another high-confidence ceRNA network was constructed by selecting the top 100 miRNA-mRNA interactions with prediction scores exceeding 0.7 for both normal and tumor tissue. A noticeably denser structure was observed in the normal network, with a higher number of interactions compared to its tumor counterpart. A comparative boxplot of log-transformed betweenness scores is shown in Figure 8C. The distribution highlights a broader spread and higher centrality values in the normal samples, suggesting a more interconnected and potentially robust regulatory network. In contrast, the network in tumor tissue displayed a sparser architecture with fewer connections, consistent with the reduced betweenness observed in its corresponding boxplot. These findings indicate that while high-confidence ceRNA interactions are present in both tissue states, their organizational complexity and regulatory dynamics differ markedly between normal and tumor tissue.

To evaluate the robustness of the reconstructed ceRNA networks with respect to the selected interaction score threshold (>0.7), a sensitivity analysis was performed by reconstructing the networks using

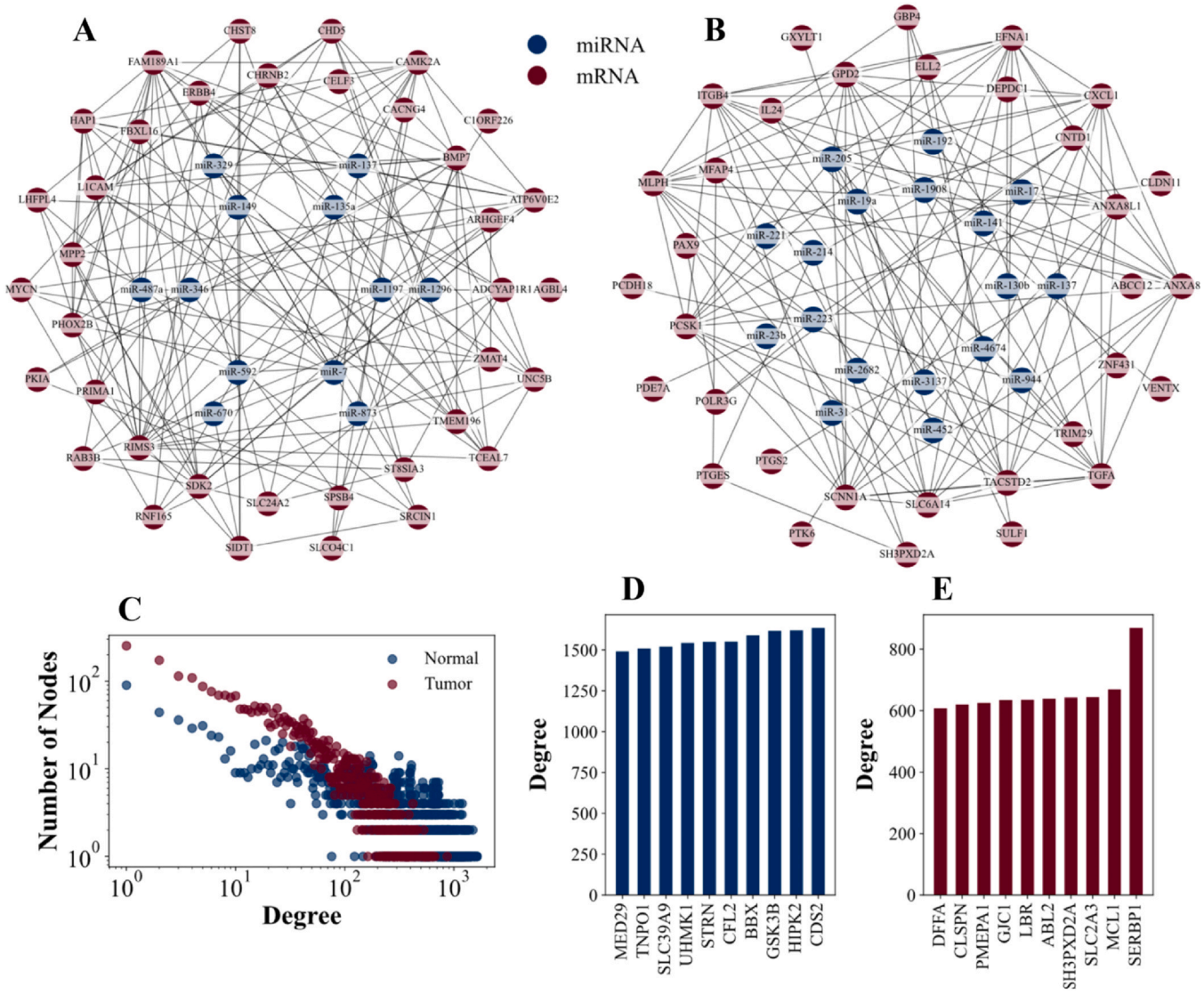


Figure 8. Topological comparison of predicted ceRNA interactions and degree distributions in normal and tumor tissue. A & B, Visualization and topological comparison of the top database-supported high-confidence predicted ceRNA interactions in normal (A) and tumor (B) tissue; C, Top 10 genes with the highest degree values in the normal ceRNA network constructed from 10k predicted interactions; D, Top 10 genes with the highest degree values in the tumor ceRNA network based on the same number of predicted interactions; E, Log-log scatter plot of degree distributions for all nodes in the normal (blue) and tumor (red) networks.

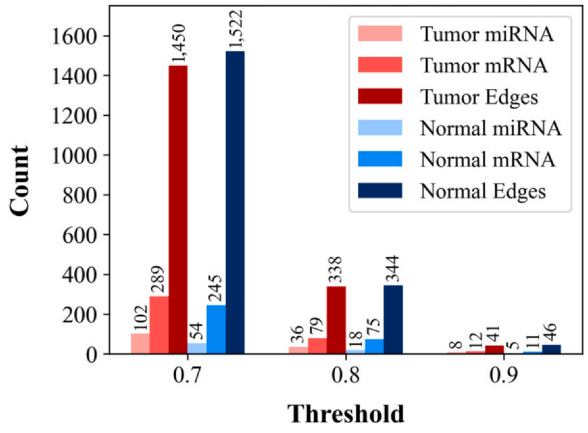


Figure 9. Threshold sensitivity analysis of the reconstructed ceRNA networks across three interaction score cutoffs (0.7, 0.8, and 0.9) in tumor and normal tissue.

three different score cutoffs (0.7, 0.8, and 0.9). As shown in Figure 9, increasing the threshold progressively reduced the number of retained miRNA nodes, mRNA nodes, and regulatory edges in both tumor and normal networks. At a threshold of 0.7, the network in tumor tissue contained 102 miRNAs, 289 mRNAs, and 1,450 edges, while the network in normal tissue included 54 miRNAs, 245 mRNAs, and 1,522 edges. Increasing the threshold to 0.8 reduced the networks to 36 miRNAs, 79 mRNAs, and 338 edges in tumor samples, and 18 miRNAs, 75 mRNAs, and 344 edges in normal samples. At the most stringent threshold of 0.9, only the highest-confidence interactions were retained, resulting in 8 miRNAs, 12 mRNAs, and 41 edges in tumor samples and 5 miRNAs, 11 mRNAs, and 46 edges in normal samples. Importantly, although the absolute number of interactions decreased as the threshold increased, the overall structural patterns of network sparsity and tumor-normal differences remained consistent. These results indicate that the key network-level observations reported in this study are not driven by a single arbitrary threshold selection but are stable across a range of interaction score cutoffs.

To capture the broader connectivity architecture of the predicted ceRNA networks, an extended analysis was conducted using the top

Table 4
Summary of the top 10 genes ranked by closeness, betweenness centrality, and degree across different interaction score thresholds in ceRNA networks in tumor and normal tissue.

ceRNA network type	Score range	Top 10 closeness	Top 10 betweenness centrality	Top 10 degree
Tumor	>0.8	IL24, GPD2, SCNN1A, SLC7A5, TACSTD2, SLC6A14, EFNA1, hsa-mir-205, ITGB4, ANXA8L1	IL24, GPD2, SLC7A5, TACSTD2, TRIM29, ELL2, SCNN1A, GBP4, SLC6A14, EFNA1	GPD2, SCNN1A, SLC7A5, TACSTD2, IL24, SLC6A14, EFNA1, ELL2, hsa-mir-205, ITGB4
Tumor	>0.7	PMEPA1, CXCL2, RAB27B, TGFA, SCNN1A, RNF144B, ELL2, SLC6A14, CXCL1, EFNA1	CXCL2, IL20RB, PMEPA1, RAB27B, PRICKLE1, ELL2, RNF144B, POLR3G, PRR11, HMGA1	ELL2, PMEPA1, PCSK1, SLC7A5, CXCL2, TGFA, XPR1, RAB27B, POLR3G, CDCP1
Tumor	>0.6	PMEPA1, RAB27B, ELL2, SLC2A3, SLC7A5, CEP170, FOSL1, GPD2, MCL1, MYC	EHF, SH3PXD2A, SLC7A5, POLD3, RAB27B, WWC1, PMEPA1, SLC2A3, BIRC5, SERBP1	ELL2, RAB27B, PMEPA1, SH3PXD2A, SLC7A5, SLC2A3, GJC1, GPD2, MYC, FOSL1
Normal	>0.7	PNMA2, BMP7, MPP2, SDK2, RAP1GAP2, KIF1A, CD24, RIMS4, hsa-mir-485, SORCS1	BMP7, PNMA2, CD24, FAM189A1, RIMS3, CAMK2A, ZMAT4, MPP2, SDK2, RIMS4	BMP7, PNMA2, MPP2, SDK2, RAP1GAP2, KIF1A, CD24, RIMS4, hsa-mir-485, SORCS1
Normal	>0.6	BMP7, TANC2, KIF1A, PNMA2, SLC7A14, RAP1GAP2, RIMS4, FAIM2, AMOTL1, GNG4	BMP7, TANC2, KIF1A, SLC7A14, RIMS3, UNC5B, PRIMA1, CNTNAP2, CPLX2, B3GALT5	BMP7, TANC2, KIF1A, PNMA2, SLC7A14, RAP1GAP2, RIMS4, FAIM2, CNTNAP2, GNG4
Normal	>0.5	RAB3B, BMP7, TANC2, KIF1A, SLC7A14, PNMA2, KSR2, GRIK3, SORL1, FAIM2	RAB3B, TANC2, RIMS3, BMP7, KIF1A, RIMS4, SLC7A14, KSR2, GRIK3, SORL1	RAB3B, BMP7, TANC2, KIF1A, SLC7A14, PNMA2, KSR2, GRIK3, SORL1, FAIM2

10,000 predicted interactions in both normal and tumor samples. The degree distribution of all nodes within these expanded networks is visualized in Figure 8D-E. As expected, a power-law-like behavior was observed in both conditions, although the network in tumor tissue exhibited a sharper drop-off at higher degrees, suggesting a sparser yet more modular topology.

The identification of high-degree nodes (hubs) was carried out to highlight central regulators within the networks. The top 10 hub genes with the highest degree values in the normal and tumor networks are presented in Figure 8D and E, respectively. In the normal context, nodes such as CDS2, HIPK2, and GSK3B exhibited notably high connectivity, suggesting widespread regulatory influence (Figure 8D). Conversely, in the network in tumor tissue, nodes such as SERBP1, MCL1, and SLC2A3 emerged as prominent hubs, potentially reflecting tumor-specific rewiring of ceRNA interactions (Figure 8E).

A comparative summary of the top 10 nodes based on key network metrics, closeness, betweenness centrality, and degree across varying score thresholds is presented in Table 4. In the tumor-derived ceRNA networks, these metrics were evaluated at confidence thresholds of >0.8, >0.7, and >0.6, while for the network in normal tissue, thresholds of >0.7, >0.6, and >0.5 were applied. This multi-level assessment was conducted to identify consistently central genes that may represent potential regulatory hubs and biomarker candidates for further experimental validation.

4. Discussion

4.1. Biological interpretation of predicted regulatory networks

The reliability and biological plausibility of the predicted miRNA-mRNA interactions in this study have been substantiated through several layers of validation and are broadly consistent with the growing literature on systems-level biomarker discovery in ACC. Notably, the integration of ML-derived regulatory networks with established experimental databases such as miRTarBase has yielded a set of high-confidence interactions, underscoring the robustness of the analytical approach. This multilayered validation strategy parallels previous studies, where hub genes and prognostic markers were confirmed through the convergence of multiple computational and experimental modalities.[24,25]

ML methods have demonstrated considerable advantages over traditional correlation-based approaches, as reflected by the superior predictive performance (R^2 up to 0.95) and the capacity to capture nonlinear, context-dependent regulatory associations. These capabilities align with recent efforts to employ advanced algorithms such as RF and support vector machine for prognostic stratification and regulatory module identification in ACC and other cancers.[24,25,27] In particular, the use of ensemble models has enabled the identification of new regulatory axes and biomarker candidates not previously detected by correlation or mutual information methods, thereby extending the frontier of post-transcriptional regulatory biology.[30]

Several miRNAs and mRNAs with high regulatory potential have been highlighted by the current framework, many of which have not been previously implicated in ACC. For example, the identification of distinct tumor-specific hubs and rewired network architectures in malignant samples echoes findings from other multi-omics studies, where central network nodes were shown to be tightly associated with disease prognosis, immune infiltration, and drug response.[26,28] Moreover, the predicted enrichment of new miRNA-mRNA pairs in pathways such as extracellular matrix remodeling and immune signaling is consistent with the biological mechanisms identified in multi-cohort integrative analyses.[24,26]

4.2. Methodological assumptions and limitations

Despite these strengths, several limitations must be acknowledged. First, the integration of transcriptomic data from TCGA-ACC and GTEx introduces potential sources of technical variability, including cohort-specific batch effects and differences in sample processing pipelines.[28] In addition, the imbalance between ACC samples ($n = 78$) and normal adrenal samples ($n = 250$) may influence the statistical properties of correlation estimates and model training. Although the larger number of normal samples provides improved statistical stability for correlation estimates, separate models were trained for tumor and normal tissue to prevent direct bias arising from unequal sample sizes. Cross-validation procedures were applied to reduce overfitting; however, these factors may still affect model generalizability and should be considered when interpreting the predicted interaction scores. Additional harmonization steps were applied to mitigate technical discrepancies between cohorts, including consistent preprocessing, expression filtering, and feature construction procedures across datasets. Nevertheless, residual cross-cohort variability cannot be completely excluded and should be considered when interpreting network differences between tumor and normal tissue.

Additionally, as the predictions are derived entirely from in silico models, further experimental validation, such as luciferase assays or in vivo studies, is required to confirm the functional relevance of new regulatory interactions. Similar concerns have been noted in the literature, emphasizing the need for orthogonal validation and integration with additional omics layers such as proteomics or epigenomics.[25,32]

From a methodological perspective, it is important to clarify the relationship between Pearson correlation and the ML objective used in this study. While Pearson correlation is a deterministic function of the input vectors, its empirical estimate is highly sensitive to sample size, expression noise, and outliers, particularly in small and heterogeneous cohorts such as TCGA-ACC. While the Pearson correlation coefficient was used as a target variable during model training, the purpose of the ML framework was not to reproduce this statistic directly but to learn global expression patterns that characterize strong miRNA-mRNA regulatory coupling across the dataset. This strategy enables systematic prioritization of candidate interactions beyond simple pairwise correlation screening. The ensemble learning framework aggregates information across multiple decision paths and subsamples, producing a stabilized interaction score that is less sensitive to sampling noise than raw correlation estimates. In addition, ceRNA network inference inherently assumes that correlated miRNA-mRNA expression patterns reflect regulatory competition mediated through shared miRNA response elements. While this assumption is widely adopted in ceRNA studies, indirect regulatory effects and shared upstream drivers may also generate correlated expression patterns. Consequently, a subset of predicted interactions may represent false-positive associations rather than direct regulatory relationships.

Importantly, separate ML models were trained for tumor and normal tissue to capture condition-specific regulatory structures rather than to enable direct quantitative comparison of prediction scores across cohorts. Consequently, the inferred networks should be interpreted as condition-specific regulatory landscapes rather than outputs of a single unified predictive model.

Furthermore, the interaction scores produced by the ML framework reflect statistical co-expression patterns rather than direct biochemical evidence of miRNA-mediated regulation. Although overlap with curated databases such as miRTarBase and TargetScan provides supporting biological context, definitive confirmation of regulatory relationships ultimately requires targeted experimental validation.

The comparative analysis with reference databases such as miRTarBase and TargetScan has showed both substantial overlap and meaningful divergence. Many high-scoring predicted interactions have been independently supported by previous experimental or computational evidence,[24,31] thereby supporting the reliability of the modeling pipeline. Conversely, a subset of top-ranked tumor-specific interactions appears to be novel, and these findings echo prior ML studies that identified previously unreported biomarkers with significant prognostic potential.[25,27] The context specificity of such regulatory events, as well as their potential for functional impact, merit further mechanistic investigation.

4.3. Future directions and experimental prioritization

Among the predicted interactions, those with high machine-learning scores (>0.9) and strong network centrality may represent the most promising candidates for experimental validation. Future directions for this line of research are clear. Validation with expanded normal tissue datasets and additional independent cohorts will be essential to mitigate sample-specific biases.[28] Experimental confirmation of predicted interactions, ideally through high-throughput reporter assays and integration with multi-omics datasets, will facilitate translation from computational prediction to clinical application.[27,30] Furthermore, the generalizability of this ML pipeline to other cancers and disease contexts may enable the rapid discovery of novel biomarkers and therapeutic targets, as demonstrated in recent pan-cancer analyses.[26,32]

In summary, the present study has extended the boundaries of miRNA-mRNA regulatory analysis in ACC, leveraging state-of-the-art ML methodologies and integrating findings with a broad spectrum of published multi-omics and computational studies.[24–32] These results not only reinforce the biological plausibility and reliability of predicted interactions but also highlight the translational potential of advanced ML for biomarker discovery and systems-level understanding in rare endocrine malignancies such as ACC.

5. Conclusions

In this study, a robust ML-based framework was established for the systematic modeling and interpretation of miRNA-mRNA regulatory associations within the ceRNA landscape of ACC. By leveraging harmonized RNA-Seq and miRNA-Seq data from both TCGA-ACC and GTEx, and applying rigorous feature engineering and multi-model regression, highly accurate prediction of regulatory interactions was achieved, with RF emerging as the top-performing algorithm. An integrative approach, which combined ML predictions with established interaction databases, revealed extensive network rewiring and a pronounced loss of regulatory hub connectivity in the tumor tissue compared to normal tissue, highlighting the complexity of post-transcriptional control in ACC. Notably, new, high-confidence tumor-specific miRNA-mRNA interactions, many of which are linked to oncogenic signaling and extracellular matrix remodeling, were identified, expanding the current repertoire of potential biomarkers and therapeutic targets for this challenging malignancy. These findings underscore not only the biological relevance of ML-inferred regulatory networks but also the capacity of this analytical pipeline to prioritize functionally meaningful molecular interactions for future experimental validation. Furthermore, the generalizability and scalability of the framework are positioned as a powerful tool for dissecting miRNA-mediated regulatory architecture across a range of cancers and transcriptomic datasets, with significant implications for systems-level biomarker discovery and precision oncology.

CRediT authorship contribution statement

Javad Omidi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethics statement

Not applicable.

Funding

No specific grant was received for this research.

Data availability

Data is available from the corresponding author upon reasonable request.

Declaration of Competing Interest

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

References

1. Allolio B, Fassnacht M. Clinical review: adrenocortical carcinoma: clinical update. *J Clin Endocrinol Metab.* 2006;91(6):2027–2037.
2. Kerkhofs TM, Verhoeven RH, Van der Zwan JM, et al. Adrenocortical carcinoma: a population-based study on incidence and survival in the Netherlands since 1993. *Eur J Cancer.* 2013;49(11):2579–2586.
3. Else T, Kim AC, Sabolch A, et al. Adrenocortical carcinoma. *Endocr Rev.* 2014;35(2):282–326.
4. Terzolo M, Angeli A, Fassnacht M, et al. Adjuvant mitotane treatment for adrenocortical carcinoma. *N Engl J Med.* 2007;356(23):2372–2380.
5. Allolio B, Hahner S, Weismann D, Fassnacht M. Management of adrenocortical carcinoma. *Clin Endocrinol (Oxf).* 2004;60(3):273–287.
6. Fassnacht M, Libé R, Kroiss M, Allolio B. Adrenocortical carcinoma: a clinician's update. *Nat Rev Endocrinol.* 2011;7(6):323–335.
7. Chukkalore D, MacDougall K, Master V, Bilen MA, Nazha B. Adrenocortical carcinomas: molecular pathogenesis, treatment options, and emerging immunotherapy and targeted therapy approaches. *Oncologist.* 2024;29(9):738–746.
8. Decmann A, Perge P, Turai PI, Patócs A, Igaz P. Non-coding RNAs in adrenocortical cancer: from pathogenesis to diagnosis. *Cancers.* 2020;12(2):461.
9. Chiu HS, Somvanshi S, Patel E, et al. Pan-cancer analysis of lncRNA regulation supports their targeting of cancer genes in each tumor context. *Cell Rep.* 2018;23(1):297–312.e12.
10. Liang WC, Fu WM, Wang YB, et al. H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. *Sci Rep.* 2016;6:20121.
11. Hashemi M, Gholamrezaie H, Ziyaei F, et al. Role of lncRNA PVT1 in the progression of urological cancers: Novel insights into signaling pathways and clinical opportunities. *Cell Signal.* 2025;131:111736.
12. Yin J, Zeng X, Ai Z, Yu M, Wu Y, Li S. Construction and analysis of a lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in oral cancer. *BMC Med Genom.* 2020;13(1):84.
13. Salimian N, Peymani M, Ghaedi K, Mirzaei S, Hashemi M. Diagnostic and therapeutic potential of LINC01929 as an oncogenic LncRNA in human cancers. *Pathol Res Pract.* 2023;244:154409.
14. Norouzinasab F, Salimian N, Mokhtari K, et al. Discovery of LINC01614 associated with the SPP1 gene in colorectal cancer. *Pathol Res Pr.* 2025;266:155761.
15. Sayaf H, Salimian N, Mohammadi M, et al. Botox-A induced apoptosis and suppressed cell proliferation in fibroblasts pre-treated with breast cancer exosomes. *Mol Cell Probes.* 2025;79:102007.
16. Sedaghat N, Fathy M, Modarressi MH, Shojaie A. Identifying functional cancer-specific miRNA-mRNA interactions in testicular germ cell tumor. *J Theor Biol.* 2016;404:82–96.
17. Andrés-León E, Cases I, Alonso S, Rojas AM. Novel miRNA-mRNA interactions conserved in essential cancer pathways. *Sci Rep.* 2017;7:46101.
18. Jia S, Yu L, Wang L, Peng L. The functional significance of circRNA/miRNA/mRNA interactions as a regulatory network in lung cancer biology. *Int J Biochem Cell Biol.* 2024;169:106548.
19. Wang J, Yu XF, Ouyang N, et al. MicroRNA and mRNA interaction network regulates the malignant transformation of human bronchial epithelial cells induced by cigarette smoke. *Front Oncol.* 2019;9:1029.
20. Balasundaram A, Mitra TS, Tayubi IA, Zayed H, Doss GPC. Deciphering the miRNA-mRNA Interaction Landscape between Breast Cancer and Triple-Negative Breast Cancer: An Integrated Bioinformatics Approach. *ACS Omega.* 2024;9(23):24379–24395.
21. Zhang Z, Song W, Chen W, et al. Unveiling hotspots of emerging research in the miRNA-related mechanism underlying cancer through comprehensive bibliometric analysis with implications for precision medicine and non-invasive diagnostics. *Front Oncol.* 2025;14:1521251.
22. Giles CB, Girija-Devi R, Dozmorov MG, Wren JD. mirCoX: a database of miRNA-mRNA expression correlations derived from RNA-seq meta-analysis. *BMC Bioinforma.* 2013;14(14):S17.
23. Afonso-Grunz F, Müller S. Principles of miRNA-mRNA interactions: beyond sequence complementarity. *Cell Mol Life Sci.* 2015;72(16):3127–3141.
24. Marquardt A, Landwehr LS, Ronchi CL, et al. Identifying new potential biomarkers in adrenocortical tumors based on mRNA expression data using machine learning. *Cancers.* 2021;13(18):4671.
25. Yi X, Wan Y, Cao W, Peng K, Li X, Liao W. Identification of four novel prognostic biomarkers and construction of two nomograms in adrenocortical carcinoma: a multi-omics data study via bioinformatics and machine learning methods. *Front Mol Biosci.* 2022;9:878073.
26. Martin-Hernandez R, Espeso-Gil S, Domingo C, et al. Machine learning combining multi-omics data and network algorithms identifies adrenocortical carcinoma prognostic biomarkers. *Front Mol Biosci.* 2023;10:1258902.
27. Eisenhofer G, Durán C, Chavakis T, Cannistraci CV. Steroid metabolomics: machine learning and multidimensional diagnostics for adrenal cortical tumors, hyperplasias, and related disorders. *Curr Opin Endocr Metab Res.* 2019;8:40–49.
28. Tang J, Fang Y, Xu Z. Establishment of prognostic models of adrenocortical carcinoma using machine learning and big data. *Front Surg.* 2023;9:966307.
29. Yang Y, Wang X, Wu L, Zhao S, Chen R, Yu G. Identification and validation of susceptibility modules and hub genes of adrenocortical carcinoma through WGCNA and machine learning. *Discov Oncol.* 2025;16(1):663.
30. Yan X, Guo ZX, Yu DH, et al. Identification and validation of a novel prognosis prediction model in adrenocortical carcinoma by integrative bioinformatics analysis, statistics, and machine learning. *Front Cell Dev Biol.* 2021;9:671359.
31. Saygili ES, Elhassan YS, Prete A, Lippert J, Altieri B, Ronchi CL. Machine learning-based survival prediction tool for adrenocortical carcinoma. *J Clin Endocrinol Metab.* 2025;110(10):e3185–e3192.
32. Yang JY, Yang MQ, Luo Z, et al. A hybrid machine learning-based method for classifying the Cushing's Syndrome with comorbid adrenocortical lesions. *BMC Genom.* 2008;9(1(1):S23.
33. Omidi J. miR-507 and miR-665 as central MicroRNA regulators in the ceRNA network of adrenocortical carcinoma: A systems biology approach. *Hum Gene.* 2025;46:201498.
34. Omidi J. Synergistic dual oncogenic role of CKS2 and ACAT2: enhanced regulatory coherence and biomarker potential in adrenocortical carcinoma. *Cancer Genet.* 2025;298-299:221–225.
35. Omidi J. miR-466 as a central miRNA with tumor-suppressive potential in the regulatory network of adrenocortical carcinoma. *Silico Res Biomed.* 2026:100214.
36. Omidi J. Molecular landscape and biomarker discovery in adrenocortical carcinoma: An integrative review of bioinformatics and translational insights. *Pathol Res Pr.* 2026;277:156295.
37. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648–660.
38. National Cancer Institute. TCGA Program. Updated 2024. Accessed January 2, 2026. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.
39. Huang HY, Lin YC, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020;48(D1):D148–D154.
40. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* 2015;4:e05005.
41. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–D613.