



# Complementary ROC-Derived Indices for Screening Improper Expression Profiles in RNA-Seq Differential Expression Analysis

● Merve Başol Göksülük<sup>1</sup>, ● Ebru Öztürk<sup>2</sup>, ● Ünal Erkorkmaz<sup>1</sup>, ● Asuman Devenci Özkan<sup>3</sup>, ● Hamdi Furkan Kepenek<sup>2</sup>, ● Dinçer Göksülük<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Sakarya University Faculty of Medicine, Sakarya, Türkiye

<sup>2</sup>Department of Biostatistics, Hacettepe University Faculty of Medicine, Ankara, Türkiye

<sup>3</sup>Department of Medical Biology, Sakarya University Faculty of Medicine, Sakarya, Türkiye

**Background:** Differential expression (DE) analysis of RNA sequencing (RNA-Seq) data are cornerstone of transcriptomic research. Widely used statistical frameworks are primarily optimized to detect monotonic mean shifts between conditions and may therefore overlook genes or microRNAs whose disease association arises at both low and high expression levels. Such non-monotonic patterns, referred to here as improper expression profiles, may reflect biologically relevant heterogeneity but remain difficult to identify using standard tools.

**Aims:** To evaluate whether receiver operating characteristic (ROC)-based indices, specifically the generalized area under the curve (gAUC) and the length of the ROC curve (LROC), can support exploratory screening and prioritization of improper expression profiles in RNA-Seq data, as a complement to conventional DE methods.

**Study Design:** Methodological study.

**Methods:** Using simulated negative binomial count data, we compared DESeq2, classical AUC (cAUC), gAUC, and LROC across varying sample sizes and dispersion levels, focusing on improper expression profiles. Performance was summarized using true positive rate and positive predictive value under ranking-based feature selection, including a

one-shot benchmark operating point (available only in simulations) and sensitivity analyses across selection sizes. The methods were also applied to a publicly available CC miRNA dataset using heuristic post-hoc screening rules informed by simulation diagnostics.

**Results:** cAUC was largely insensitive to improper expression patterns. DESeq2 performed robustly for conventionally differentially expressed features but recovered a smaller fraction of simulated improper profiles under ranking-based selection. Across simulation scenarios, gAUC showed the highest and most stable recovery of improper profiles, whereas LROC provided complementary signal under low-to-moderate dispersion but degraded under extreme overdispersion. In the CC dataset, ROC-derived indices identified candidate improper miRNAs that were not prioritized by DESeq2, and several top candidates had literature support consistent with biological plausibility.

**Conclusion:** gAUC, supported by LROC as an auxiliary index, provides a practical ROC-based screening extension to standard RNA-Seq workflows. Because these indices are applied using heuristic thresholds without controlled error rates, the resulting candidates should be interpreted as exploratory prioritization and require independent validation.

## INTRODUCTION

Gene expression reflects the transcription of DNA into RNA and, in many cases, its subsequent translation into proteins that regulate cellular function. In contrast, the collection of all expressed RNA molecules defines the transcriptome. Transcriptomic profiling provides a dynamic snapshot of cellular activity and is essential for understanding how gene regulation changes across biological

conditions and disease states.<sup>1</sup> Disruptions in transcriptomic patterns are central to disease mechanisms, including cancer development and therapeutic resistance, thereby positioning transcriptomic analysis as a foundation for precision medicine and biomarker discovery.<sup>2,3</sup> While early transcriptome studies relied on microarray technologies,<sup>4,6</sup> RNA sequencing (RNA-Seq) has become the dominant platform due to its higher sensitivity, broader dynamic range, and ability to identify novel transcripts.<sup>1,7</sup> A major application of RNA-Seq



**Corresponding author:** Dinçer Göksülük, Department of Biostatistics, Sakarya University Faculty of Medicine, Sakarya, Türkiye

**e-mail:** dincergoksuluk@sakarya.edu.tr

**Received:** January 30, 2026 **Accepted:** April 16, 2026 **Available Online Date:** June 1, 2026 • **DOI:** 10.4274/balkanmedj.galenos.2026.2026-1-309

Available at [www.balkanmedicaljournal.org](http://www.balkanmedicaljournal.org)

**ORCID iDs of the authors:** M.B.G. 0000-0002-2223-7856; E.Ö. 0000-0001-9206-6856; Ü.E. 0000-0002-8497-4704; A.D.Ö. 0000-0002-3248-4279; H.F.K. 0009-0003-6339-4203; D.G. 0000-0002-2752-7668.

**Cite this article as:** Başol Göksülük M, Öztürk E, Erkorkmaz Ü, Devenci Özkan A, Kepenek HF, Göksülük D. Complementary ROC-Derived Indices for Screening Improper Expression Profiles in RNA-Seq Differential Expression Analysis. *Balkan Med J.* 2026;43:330-342

Copyright© Author(s) - Available online at <http://balkanmedicaljournal.org/>

is differential expression (DE) analysis, which detects differences in gene expression across biological conditions and links transcriptional changes to disease mechanisms and treatment response.<sup>8,9</sup>

To perform DE analysis of RNA-Seq data, several statistical frameworks are widely used, including DESeq2, edgeR, and limma-voom. DESeq2 and edgeR model count data using negative binomial generalized linear models with dispersion estimation and shrinkage.<sup>8,10</sup> These methods perform reliably even with small numbers of replicates and are effective in resource-limited experimental designs. In contrast, limma-voom uses mean–variance modeling to derive precision weights for linear modeling of log-scale expression values. This approach offers computational efficiency and stability across large gene sets.<sup>11</sup> Alternative approaches include baySeq,<sup>12</sup> PoissonSeq,<sup>13</sup> and SAMseq,<sup>14</sup> and a broader overview of DE analysis tools is provided by Rosati et al.<sup>15</sup> These methods provide robust inferential testing for features whose association with condition is well captured by group-level mean differences (i.e., consistent up- or downregulation). However, because their primary hypotheses target mean shifts, they may be less sensitive to non-monotonic relationships in which both unusually low and unusually high expression levels are informative. In this study, we refer to such features as improper genes (IGs), also described as bimodal patterns in the literature.

The presence of genes or microRNAs (miRNAs) with improper expression profiles is now well recognized in biomedical research, as both very low and very high expression levels can contribute to disease. This “Goldilocks” phenomenon, in which either loss-of-function or gain-of-function of the same gene can lead to pathology, has been observed in diverse contexts.<sup>16–19</sup> Similarly, many miRNAs may play dual roles as tumor suppressors or oncogenes depending on their expression levels, underscoring the importance of balanced expression in normal physiology.<sup>20</sup> Detecting these patterns is important because they may indicate hidden biological variability, distinct disease subtypes, or multiple mechanisms operating within the same condition.

Several statistical and computational frameworks have been proposed to detect bimodal or multimodal expression patterns in microarray data. Early efforts, such as the Cancer Outlier Profile Analysis, aimed to identify oncogenes with extreme expression in a subset of tumor samples and successfully uncovered events such as the TMPRSS2–ERG fusion in prostate cancer.<sup>21,22</sup> Extensions include methods based on outlier detection and subgroup identification statistics,<sup>23–25</sup> clustering approaches,<sup>26</sup> and bimodality filtering techniques.<sup>27</sup> However, these methods were primarily developed for microarray platforms, which assume approximately Gaussian distributions, thereby limiting their direct applicability to RNA-Seq data characterized by discrete and overdispersed count distributions.

To address this limitation, two general strategies have emerged: (i) adapting existing workflows through RNA-Seq-specific extensions, such as SIBER<sup>28</sup> and GMMchi,<sup>29</sup> or enabling the use of microarray-based methods by applying variance-stabilizing transformations (VSTs) that render count data approximately normal; and (ii) employing distribution-free approaches, including receiver operating characteristic (ROC)-based frameworks, which evaluate

discriminatory performance independently of the underlying expression distribution. ROC-derived indices typically provide descriptive scores and rankings rather than formal hypothesis tests with  $p$  values or explicit error-rate control.

In this study, we evaluated ROC-derived indices alongside DESeq2, treating DESeq2 as the primary hypothesis-testing framework for DE and ROC-based indices as complementary exploratory screening tools. Specifically, ROC-based rankings were used to prioritize candidate genes and miRNAs with improper expression profiles that may be missed by mean-shift–focused testing. This evaluation was conducted through simulation studies and illustrated using a cervical cancer (CC) miRNA-Seq dataset, with literature-based plausibility assessments for selected candidates.

## MATERIALS AND METHODS

### DE analysis

Several frameworks exist for RNA-Seq DE analysis; see Rosati et al.<sup>15</sup> for an overview. In this study, DESeq2 was used as the primary hypothesis-testing benchmark for DE inference.<sup>8</sup> ROC-derived indices were used only as complementary screening scores, providing rankings without  $p$  values or formal error-rate control.

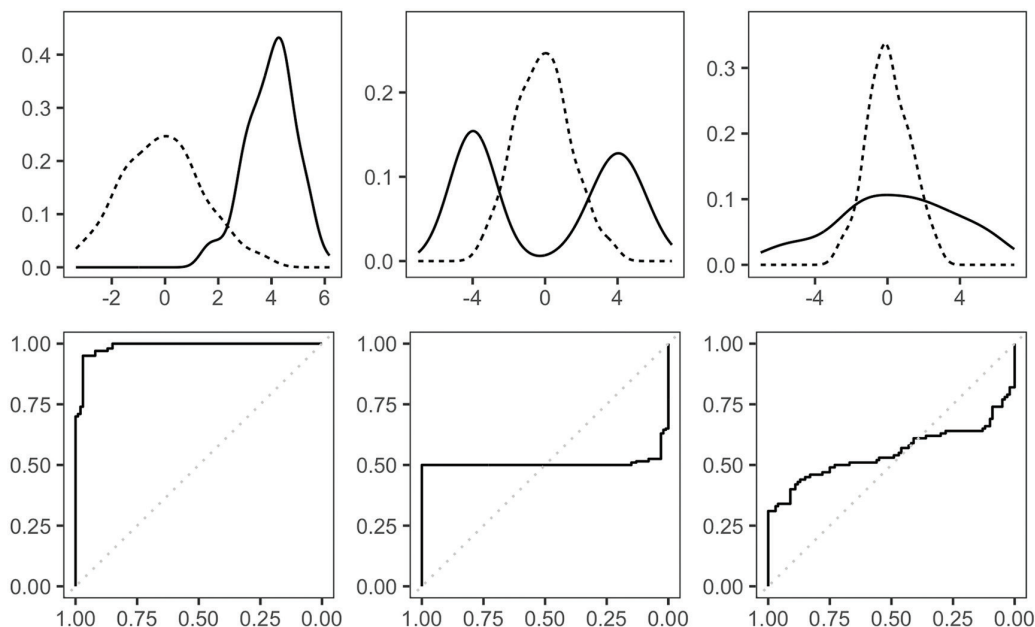
### IG expression profiles

IGs, introduced conceptually in the Introduction, exhibit non-monotonic relationships with phenotypes, where both unusually low and high expression levels may be associated with disease. In this study, these profiles were defined using the toy distributions shown in Figure 1. These simplified scenarios illustrate two generic mechanisms that can generate improper expression profiles: (i) a “high-low” mixture, in which diseased samples occupy both extremes of the expression distribution (Figure 1, middle panel), and (ii) group-specific variance inflation, in which one group shows markedly higher dispersion without a corresponding mean-shift (Figure 1, right panel).

In the simulation study, IGs were generated using the high-low mixture mechanism. A separate variance-inflation-only mechanism was not explicitly simulated, which limits the generalizability of conclusions to other forms of improper expression profiles. The variance-based scenario is retained here to highlight an alternative pathway through which non-monotonic signals may arise in practice and which ROC-based methods are also expected to capture.

### Data preprocessing

All count-based datasets, including simulated and real RNA-Seq data, were processed using a standard pipeline implemented in the DESeq2 framework. Raw counts were first filtered by removing low-quality genes (i.e., retaining genes with counts  $\geq 10$  in at least three samples), followed by near-zero-variance filtering using the nearZeroVar function from the caret R package.<sup>30</sup> The remaining counts were normalized using DESeq2’s median-of-ratios method to correct for library size and compositional effects, after which a VST was applied to obtain continuous-scale expression values with approximately stabilized variance across the dynamic range. Unless



**FIG. 1.** Distribution of proper and improper gene expression profiles and corresponding ROC curves—(left) proper, (middle/right) improper; dashed lines correspond healthy group. ROC, receiver operating characteristic.

otherwise stated, all ROC-based analyses and visualizations were performed using these variance-stabilized expression values.

### Detecting DEGs with improper expression profiles

DE analysis in this study was performed using the DESeq2 workflow. In addition to DESeq2's hypothesis-testing-based strategy for identifying differentially expressed genes (DEGs), we applied three ROC-based methods as supporting tools for gene selection and downstream performance evaluation, specifically targeting DEGs with improper expression profiles: (i) the classical area under the ROC curve (cAUC), (ii) the generalized area under the curve (gAUC), and (iii) the length of the ROC curve (LROC). Thus, cAUC, gAUC, and LROC were not used as standalone DE analysis methods; rather, they were integrated into the DESeq2-based workflow and applied to the preprocessed, variance-stabilized expression values described in the "Data preprocessing" subsection. The gAUC extends the conventional AUC by allowing dual thresholds, thereby capturing cases in which both high and low expression levels are informative.<sup>31</sup> In contrast, LROC quantifies discriminatory ability based on the LROC and may capture signals that conventional AUC underestimates.<sup>32</sup> Here, cAUC, gAUC, and LROC were used as descriptive screening scores for VST-transformed expression values. These methods provide rankings without  $p$  values or type I error control; therefore, all ROC-based results are interpreted as exploratory prioritization rather than formal DE inference.

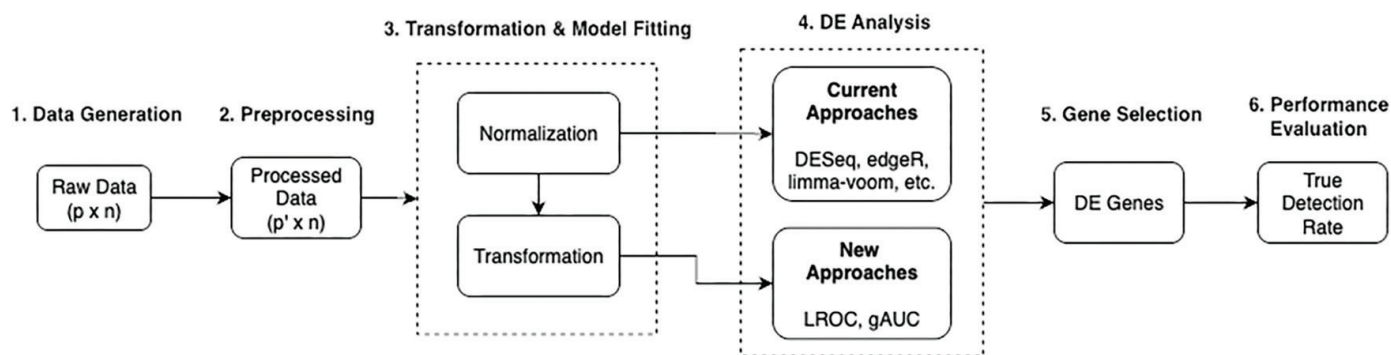
### Simulation study and real-data analysis

A comprehensive simulation study was designed to evaluate the performance of DESeq2, cAUC, gAUC, and LROC in detecting IGs under realistic RNA-Seq conditions. We evaluated each method

across 18 simulation scenarios defined by sample size ( $n$ ), number of features ( $p$ ), proportions of DEGs ( $\pi_{\text{DEGs}}$ ) and DEGs with improper profiles ( $\pi_{\text{IR}}$ ), overdispersion ( $\phi$ ), and the dispersion of log-fold-change values. For each scenario, 1,000 datasets were generated. The overall analysis pipeline is shown in Figure 2.

**Performance evaluation and sensitivity analysis:** Method performance was summarized using the true positive rate (TPR) and positive predictive value (PPV), computed for DEGs and IGs (a subset of true DEGs). For each dataset replicate, each method produced a ranked feature list. DESeq2 features were ranked by Benjamini-Hochberg-adjusted  $p$  values (log<sub>2</sub> fold change was not used for ranking), whereas cAUC, gAUC, and LROC features were ranked by their respective ROC-derived scores. Because the nominal number of DEGs is fixed by design at, we first report a reference "one-shot" benchmark at to summarize ranking performance at the design operating point. Due to prefiltering, the number of retained true DEGs may vary around across replicates. We then performed a sensitivity analysis by selecting the top- $K$  features for and evaluating how TPR and PPV changed with  $K$ , thereby reflecting the trade-off between true and false positives under ranking-based screening.

**CC dataset:** In addition to the simulation study, we analyzed a real RNA-Seq dataset originally published by Witten et al.<sup>33</sup> The dataset comprises small RNA-Seq measurements from 58 human cervical tissue samples, including 29 tumor and 29 matched non-tumor controls. Sequencing was performed using the Illumina (Solexa) platform, yielding expression profiles for 714 distinct miRNAs. The data are publicly available in the Gene Expression Omnibus under accession number GSE20592.



**FIG. 2.** Simulation workflow. Number of total features ( $p$ ) is fixed at 3,000. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; DE, differential expression.

In the CC analysis, candidate improper miRNAs were identified using heuristic cut-offs informed by simulation-based diagnostic plots: combined with either or . Because these thresholds are post-hoc screening rules rather than formal inferential tests, findings based on ROC-derived indices are considered exploratory and are interpreted alongside the DESeq2-based results.

For detailed information, including the mathematical background of the ROC-based approaches, simulation workflow, and dataset description, see [Supplementary Materials (Section A)].

## RESULTS

This section summarizes the comparative performance of DESeq2, cAUC, gAUC, and LROC based on simulation studies and presents an illustrative application to a CC miRNA dataset, accompanied by a literature-based plausibility assessment.

### Performance evaluation

Using the simulation framework described in the Materials and Methods section, we evaluated the performance of DESeq2 and three ROC-derived screening indices (cAUC, gAUC, and LROC) under varying RNA-Seq conditions. Performance metrics were computed from 1000 simulated datasets and summarized across levels of overdispersion (very low, moderate, and very high) and sample sizes (100, 300, and 500) for both DEGs and IGs. Figure 3 presents results for the low-DE scenario ( $\pi_{\text{DEGs}}=0.05$ ), in which differences among methods are most pronounced and comparative interpretation is most informative. Corresponding high-DE results ( $\pi_{\text{DEGs}}=0.30$ ) are provided in the [Supplementary Materials (Section B)]. We focus on the low-DE setting because it yields conclusions consistent with the high-DE case while providing clearer contrasts among methods.

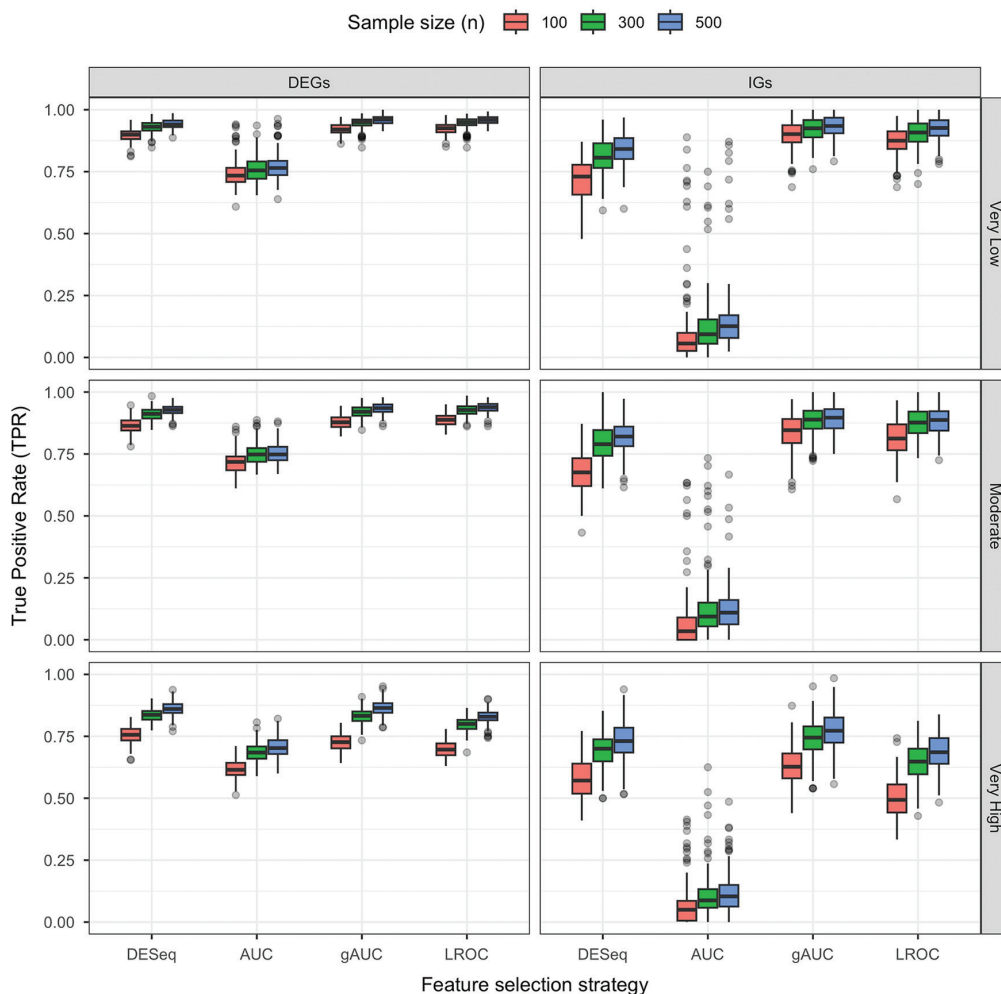
Across all scenarios, increasing sample size improved detection performance for all methods, although the magnitude of improvement varied. The cAUC performed worst overall, particularly for IGs, with TPR values often below 0.25, indicating weak discriminatory ability. DESeq2 showed strong and stable performance for DEGs but lower sensitivity for IGs. At the benchmark level, gAUC and LROC achieved higher recovery of IGs than DESeq2 under very low and moderate overdispersion, reflecting their ability to capture non-monotonic expression patterns.

Under very high overdispersion, particularly for IGs, LROC performance declined substantially and, in some cases, fell below that of DESeq2, whereas gAUC remained robust, with median detection rates exceeding 0.75 except at small sample sizes. Increasing overdispersion also increased variability in detection performance across all methods, while low-dispersion settings produced more stable results. Overall, the one-shot results summarize ranking behavior at the design operating point, whereas practical trade-offs are further examined in the sensitivity analysis below.

### Sensitivity analysis

To evaluate the sensitivity of the methods to the choice of DEG set size, we examined both TPR and PPV for DEGs and IGs across a range of selection sizes. Although the simulation design specifies a nominal DEG count of 150 per dataset, prefiltering can result in varying numbers of retained true DEGs around this value across replicates. This empirical range, which may vary slightly across simulation scenarios, is indicated by the gray-shaded band in Figures 4-7. Because PPV directly reflects the proportion of false positives among selected features, these curves summarize the sensitivity analysis, that is, the trade-off between true detection and false positives under ranking-based screening. We did not calibrate a nominal false-positive rate (FPR) for the ROC-derived indices using an inferential procedure; therefore, false positives are summarized empirically via PPV rather than through a controlled error rate, while empirical FPR curves are reported in the [Supplementary Materials (Section B)] as a specificity benchmark.

For DEGs, the sensitivity analysis indicated that all methods achieved broadly similar TPR values when only a small number of features were selected. TPR increased as the selection size grew; however, once the selected set size exceeded the retained number of true DEGs (i.e., the gray-shaded band), marginal gains diminished and method-specific differences became more apparent. Selecting additional features beyond this range primarily increased false positives rather than improving true detection (Figures 4 and 5). In this setting, DESeq2 showed a modest advantage under small sample sizes and very high overdispersion.



**FIG. 3.** Simulation results at the one-shot benchmark. The figure reports TPRs across simulated datasets when the selected set size is fixed at the nominal design value  $K=150$  (i.e.,  $p \times \pi_{\text{DEGs}}$ ) under the low-DE scenario  $\pi_{\text{DEGs}} = 0.05$ . This operating point is available only in simulations and is used here as a descriptive ranking benchmark. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; AUC, area under the curve; TPRs, true positive rates; DE, differential expression.

For IGs, gAUC consistently achieved the highest TPR and PPV. LROC generally ranked second, except under very high overdispersion (Figure 6). Notably, gAUC began recovering IGs even at small selection sizes, whereas other methods showed limited sensitivity in this range. Under very high overdispersion, DESeq2 also began recovering IGs at smaller selection sizes than in low-to-moderate overdispersion settings and generally ranked second after gAUC (Figure 7).

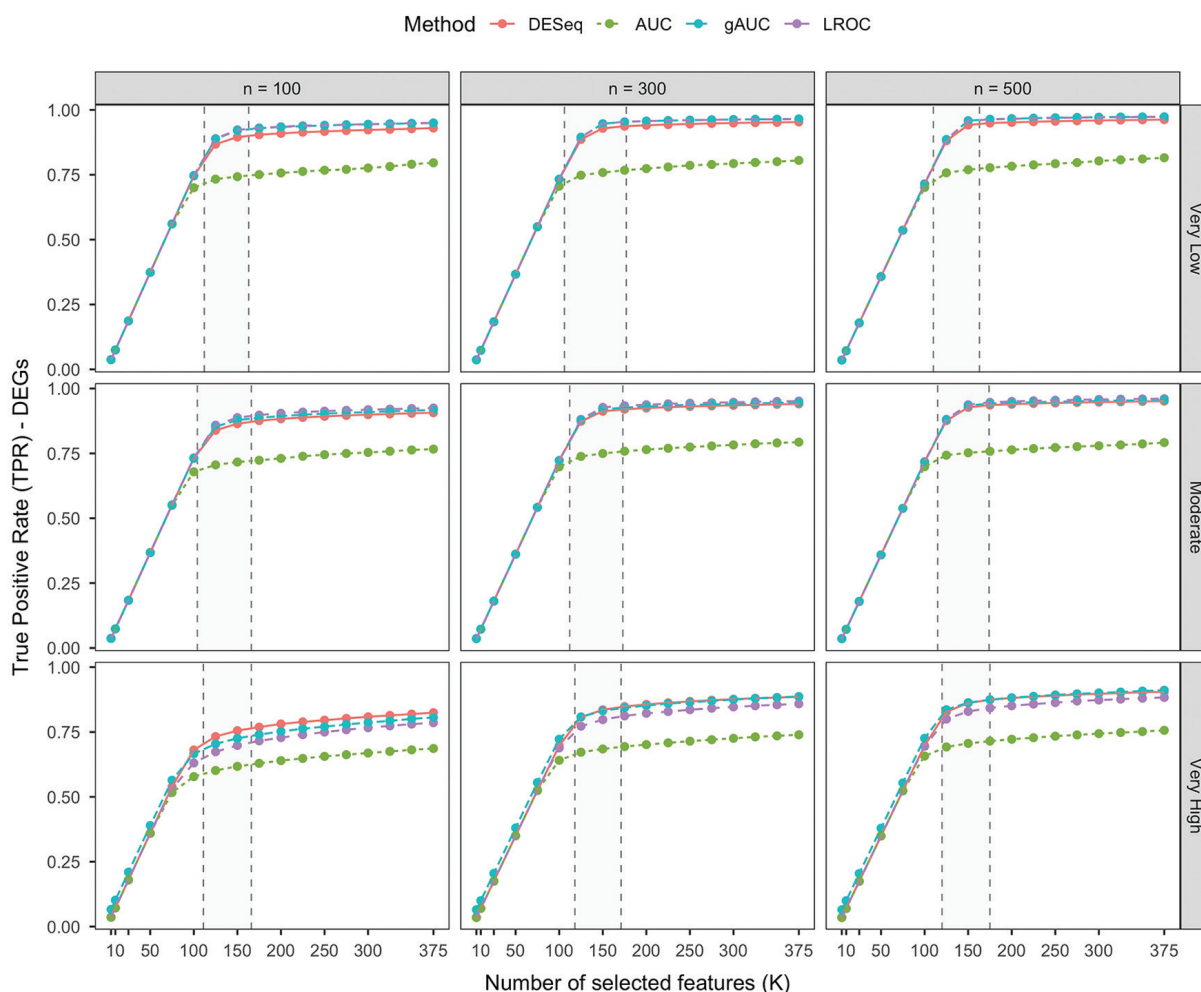
### Visual inspection of improper expression profiles

To further characterize the behavior of ROC-based metrics, we visually examined their joint distributions in a simulated dataset under the very high overdispersion scenario (i.e., and ). Figure 8 provides a diagnostic view of how the ROC-derived indices behave across different simulated feature types.

IGs tend to cluster at low cAUC values (approximately 0.5-0.65) but exhibit higher LROC values and often elevated gAUC despite low cAUC, reflecting non-monotonic expression patterns. In this high-dimensional setting, visual separation is descriptive and does not quantify statistical uncertainty; therefore, these plots are interpreted as exploratory diagnostics that may guide screening and threshold tuning rather than as formal evidence of signal.

### CC results

As a real-data counterpart to the simulation analyses, we applied the proposed methods to a CC miRNA dataset. After filtering low-quality reads and near-zero-variance features, 416 miRNAs were retained for downstream analysis. DE analysis was first performed using DESeq2, with significance defined by an absolute  $\log_2$  fold-change greater than 0.6 and a Benjamini-Hochberg-adjusted  $p$  value



**FIG. 4.** Sensitivity analysis of the methods across simulations under the low-differential expression scenario—true positive rate for DEGs. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; AUC, area under the curve; DEGs, differentially expressed genes.

below 0.05. Subsequently, gAUC and LROC were applied as screening scores to identify candidate improper miRNA profiles that may not be prioritized by DESeq2. Visual inspection of the ROC-based indices (Figure 9) revealed patterns similar to those observed in the simulation diagnostics (Figure 8).

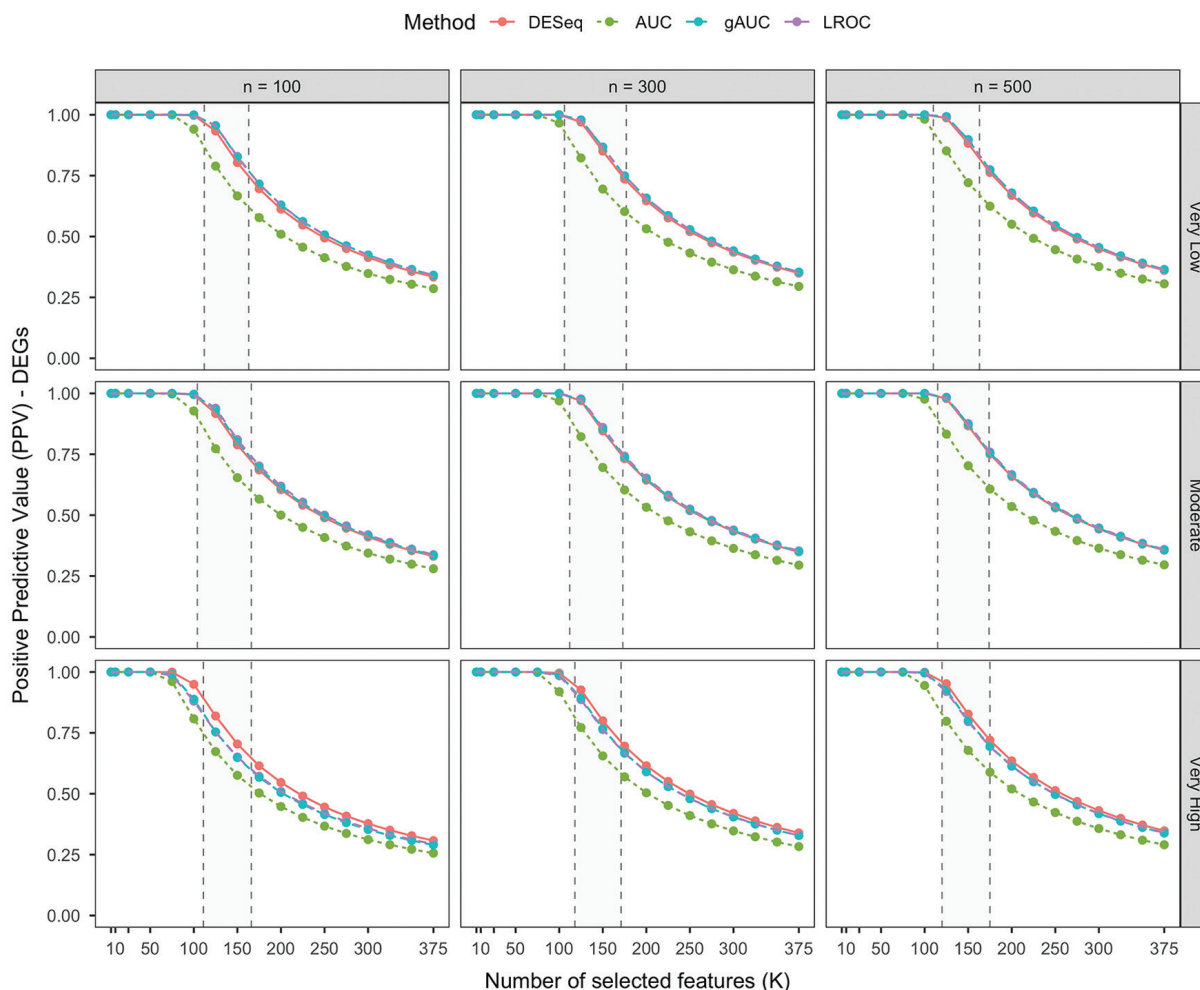
The volcano plot (Figure 10) summarizes the DESeq2 results, with slight vertical jitter applied to improve visual clarity near the significance boundary. miRNAs flagged by the heuristic screening rule but not detected by DESeq2 were considered candidate improper miRNAs and were primarily located in the non-significant region of the volcano plot. Using this rule, 56 miRNAs were identified as candidate improper miRNAs, of which 34 were not detected by DESeq2 (Figure 11).<sup>34</sup> ROC curves for the top candidates (Figure 12) exhibit non-monotonic patterns, suggesting potential associations between both low and high expression levels and tumor status. Collectively, these results illustrate how ROC-derived screening indices can complement DESeq2 by prioritizing candidate

non-monotonic miRNA profiles for downstream follow-up, such as literature-based plausibility assessment.

## DISCUSSION

In this study, we evaluated ROC-derived indices (cAUC, gAUC, and LROC) as complementary screening tools alongside DESeq2. Because DESeq2 primarily targets mean-shift signals, it may deprioritize genes and miRNAs with IG profiles. Our results suggest that ROC-derived rankings can help flag candidate improper features for follow-up analysis; however, these indices do not provide  $p$  values or formal error control and should therefore be interpreted as exploratory diagnostics.

**Importance of IG expression profiles:** IG expression profiles have important biological relevance beyond their statistical characterization. For example, in breast cancer, the estrogen receptor gene exhibits a bimodal distribution that distinguishes ER-



**FIG. 5.** Sensitivity analysis of the methods across simulations under the low-differential expression scenario—positive predictive value for DEGs. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; AUC, area under the curve; DEGs, differentially expressed genes.

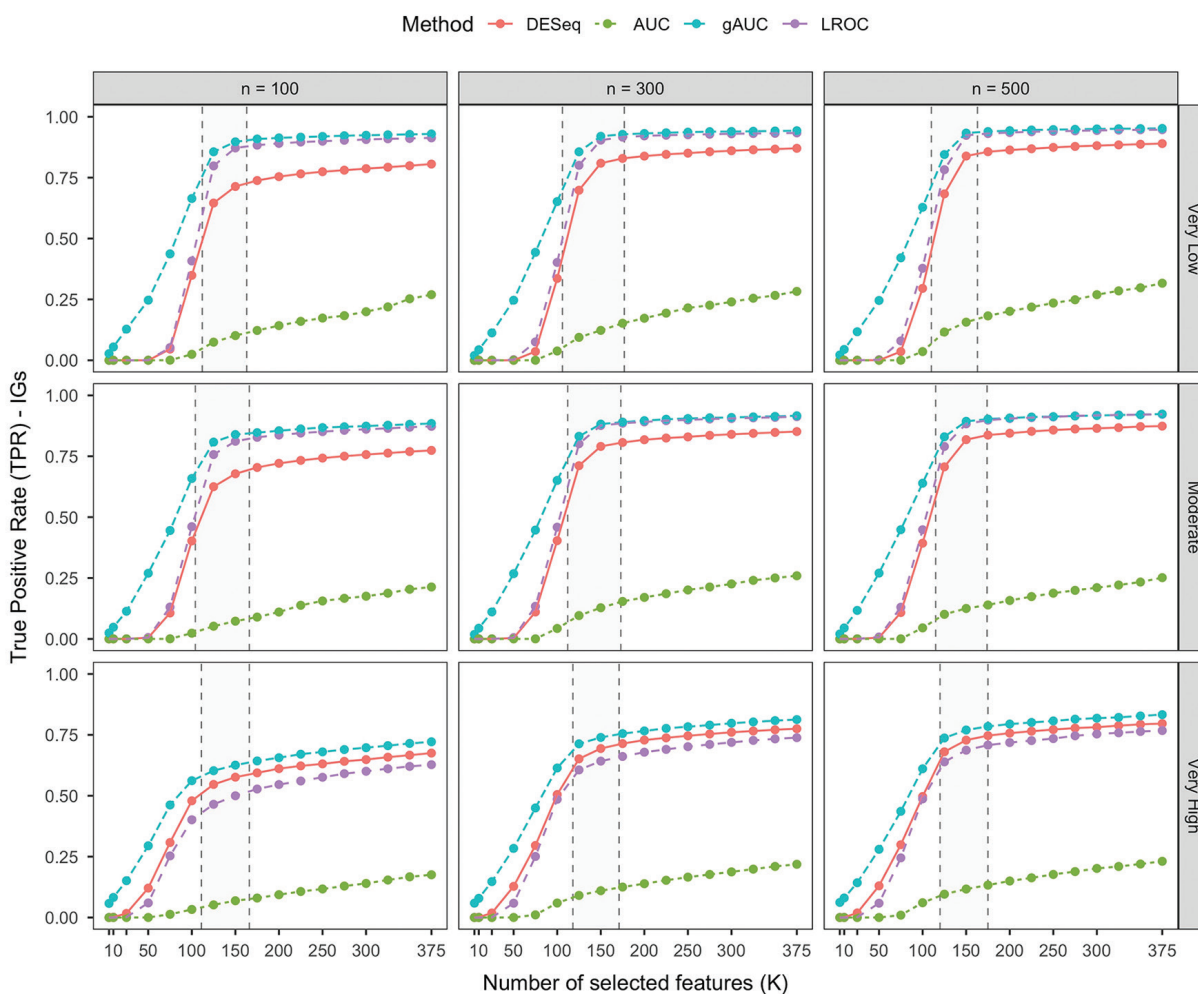
positive and ER-negative subtypes. In contrast, tumor-suppressor pathways involving genes such as PTEN may be disrupted by both loss-of-function and aberrant overexpression.<sup>35,36</sup> The activation marker TNFRSF9 also shows a bimodal expression pattern in tumor-infiltrating regulatory T-cells in non-small-cell lung cancer.<sup>17</sup> Together, these examples illustrate that transcriptomic heterogeneity can manifest as bimodality in both mRNA and miRNA expression, underscoring the need for approaches that can flag non-monotonic expression–disease relationships as candidates for further investigation.

Conventional DE tools primarily detect mean-shift signals and may miss improper expression patterns. Earlier microarray-based studies sought to address this limitation. Parodi et al.<sup>37</sup> introduced “not proper ROC curves” and the ABCR statistic to detect genes whose ROC curves cross the random expectation line, indicating subgroup effects. Silva-Fortes et al.<sup>38</sup> proposed the “arrow plot,” which combines ROC AUC and overlap measures to identify bimodal or subgroup-specific expression patterns missed by standard

approaches. Building on these conceptually related ROC-based ideas, we use gAUC and LROC as screening indices and illustrate their behavior through simulations and a real dataset application.

**Insights from simulation study:** In our simulations, ranking behavior was primarily influenced by sample size and overdispersion. At the one-shot benchmark (available only in simulations), gAUC and LROC generally recovered more IGs than DESeq2 under low-to-moderate overdispersion, whereas LROC performance deteriorated under extreme overdispersion. The sensitivity analysis summarizes the trade-off between true positives and false positives under ranking-based screening. This limitation likely reflects LROC’s reliance on a parametric binormal assumption, which may be violated in RNA-Seq data. Future work may address this limitation by developing non-parametric or regularized variants of LROC to improve robustness while preserving its geometric interpretability.

Although we did not explicitly simulate an IG-generating mechanism driven solely by variance inflation, ROC-derived screening may,



**FIG. 6.** Sensitivity analysis of the methods across simulations under the low-differential expression scenario — true positive rate for improper genes. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; AUC, area under the curve; IGs, improper genes.

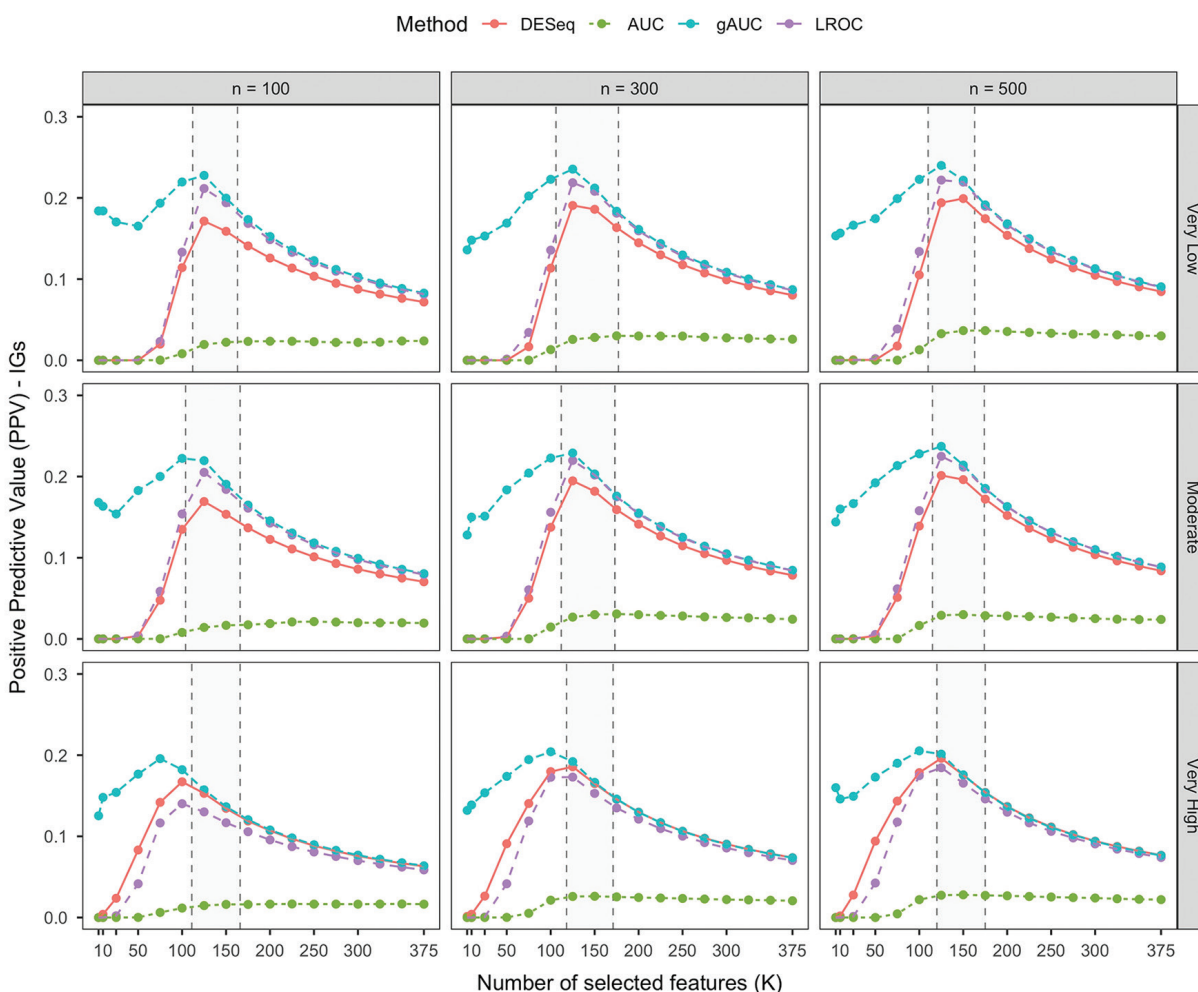
in principle, detect IGs arising from variance inflation–driven differences; however, this scenario was not evaluated in the present study. We also observed that DESeq2 recovered a larger fraction of simulated IGs under very high overdispersion. One plausible explanation is that, as dispersion increases, some mixture-generated IGs may resemble variance-inflation–like profiles or exhibit stronger mean-shift components, thereby becoming more detectable by DESeq2. Consequently, under a dedicated variance-inflation simulation design, the unimodal structure of a variance-inflation-only mechanism may lead DESeq2 and ROC-based screening to produce more similar IG recovery patterns. This hypothesis should be evaluated in future work.

**Visual diagnostics of improper profiles:** Beyond numerical performance metrics, we propose visual inspection of the AUC space (e.g., plotting cAUC against gAUC and/or LROC) as a diagnostic tool. In our simulated data, DEGs, IGs, and non-significant genes occupied distinct regions: DEGs showed high cAUC and gAUC, IGs showed moderate cAUC but high gAUC, and non-DE genes clustered near

baseline. Such plots are descriptive and do not quantify statistical uncertainty; therefore, they are intended as exploratory diagnostics to guide screening and threshold selection.

In real-data applications, we used heuristic cut-offs, which should be interpreted as post-hoc screening rules. Because these cut-offs depend on data structure and analyst objectives, they are inherently subjective: more permissive thresholds increase sensitivity but also raise FPRs, whereas more stringent thresholds reduce false positives at the cost of increased false negatives. Accordingly, threshold tuning should reflect the intended trade-off between sensitivity and the burden of false discoveries.

**Application to CC miRNA data:** CC remains a significant global health burden. In the CC miRNA dataset, ROC-derived screening indices identified candidate improper miRNAs that were not prioritized by DESeq2 (Figures 10-13). Using the heuristic rule described in the materials and methods section, we identified candidate improper miRNAs for downstream analysis and summarized their patterns



**FIG. 7.** Sensitivity analysis of the methods across simulations under the low-differential expression scenario—positive predictive value for improper genes. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve; AUC, area under the curve; IGs, improper genes.

using diagnostic plots. These findings are exploratory and intended for prioritization rather than confirmation.

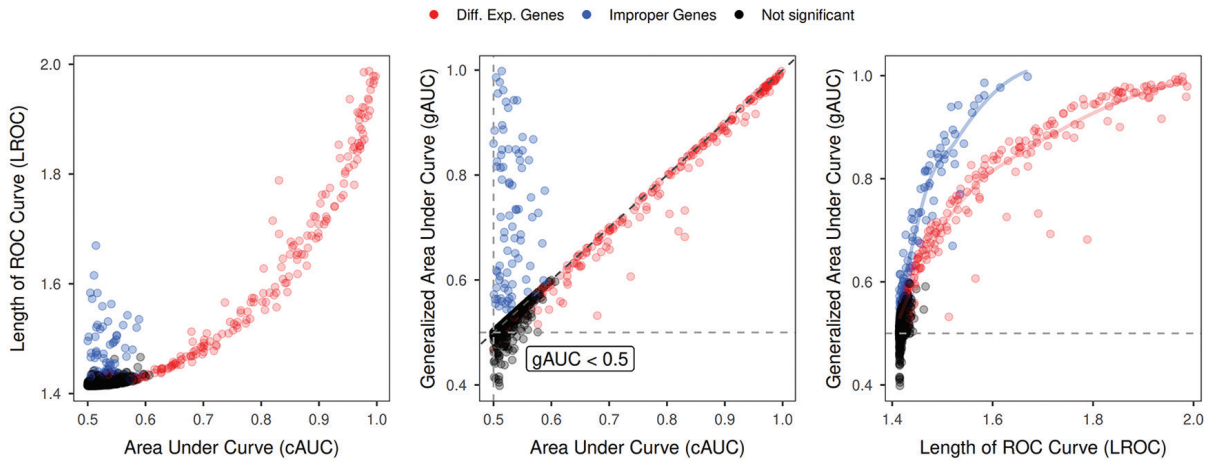
Improper or bimodal miRNA expression patterns may reflect underlying biological heterogeneity. ROC-derived indices can therefore help prioritize such candidates for downstream evaluation, including literature-based support and independent validation. For example, miR-15a-5p has been associated with tumor-suppressive activity when underexpressed, whereas miR-93-5p has been linked to oncogenic activity when overexpressed; their appearance among candidate improper miRNAs is consistent with these dual functional roles.

#### **Literature-based plausibility assessment of miRNAs in CC data**

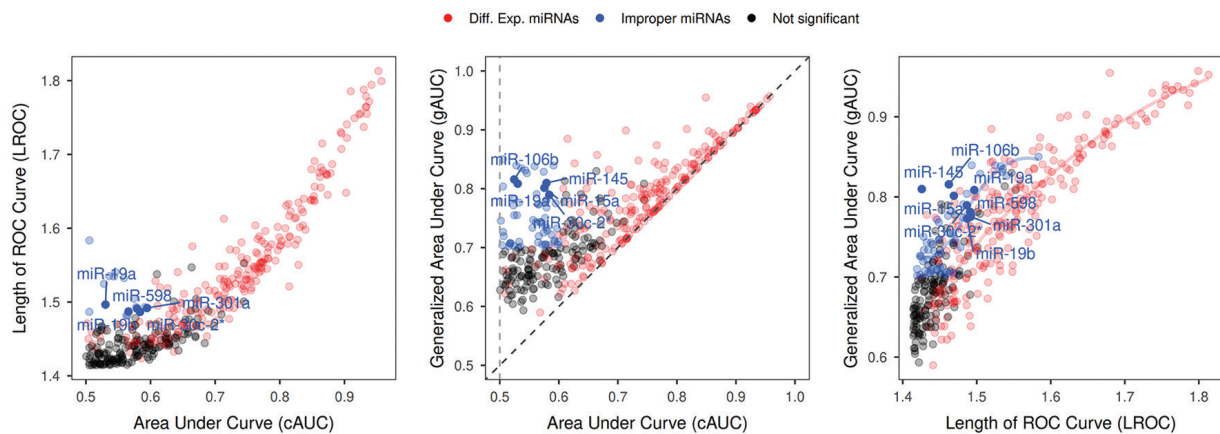
To evaluate the biological relevance of improper miRNAs identified in the CC dataset, we conducted a literature-based plausibility assessment of the top candidates. Several highly ranked miRNAs identified by our approach have well-established roles in CC

progression, supporting the biological plausibility of the ROC-based findings. Notably, tumor-suppressor miRNAs such as let-7d, miR-30c-5p, miR-382-5p, and miR-145 are consistently reported to be downregulated in cervical tumors.<sup>39-42</sup> These miRNAs regulate key oncogenic pathways involved in cell-cycle control, survival signaling, angiogenesis, and metastasis; thus, their reduced expression is consistent with loss of tumor-suppressive function.

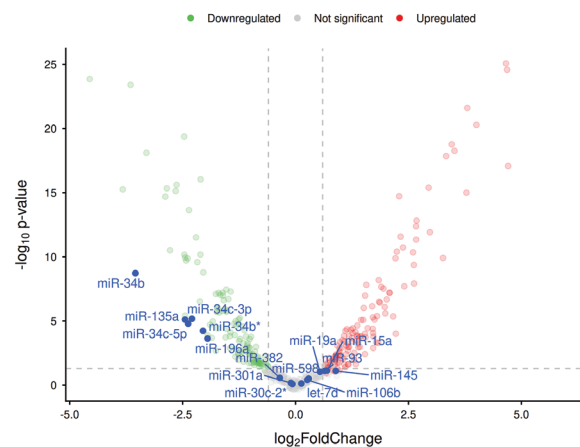
In contrast, several oncogenic miRNAs highlighted by our analysis, including miR-301a, miR-19a/b, miR-93-5p, miR-106b, and miR-15a-5p, have been repeatedly implicated in CC as drivers of proliferation, invasion, and resistance to apoptosis.<sup>43-47</sup> These miRNAs typically target key tumor-suppressor pathways, including PTEN/PI3K, TGF- $\beta$  signaling, and p53-related stress responses, thereby promoting epithelial-mesenchymal transition, angiogenesis, and metastatic potential. Being classified as improper miRNAs suggests that their dysregulation—particularly deviations toward both low and high expression levels—may contribute to malignant behavior, consistent



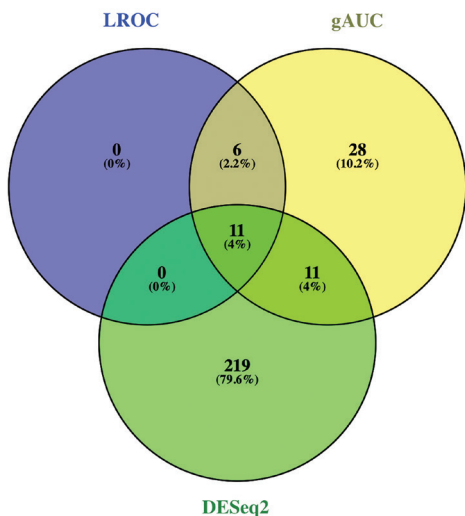
**FIG. 8.** Scatterplot matrix for the relationships among three receiver operating characteristic-derived diagnostic indices of simulated gene expression data. Points are colored according to the true status of the genes in the simulated dataset. ROC, receiver operating characteristic.



**FIG. 9.** Scatterplot relationships among ROC-based metrics for the cervical cancer dataset. Points are colored according to the differential expression results of DESeq2 and heuristic thresholds for ROC-derived metrics. ROC, receiver operating characteristic. cAUC, classical area under the curve.



**FIG. 10.** Volcano plot of DESeq2 results for the cervical cancer dataset. The top-ranked candidate miRNAs exhibiting improper expression profiles, as identified by the gAUC and LROC metrics, are annotated in the figure. Some of these improper miRNA profiles fall within the upregulated or downregulated regions, which are also detected by the DESeq2 framework. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve.

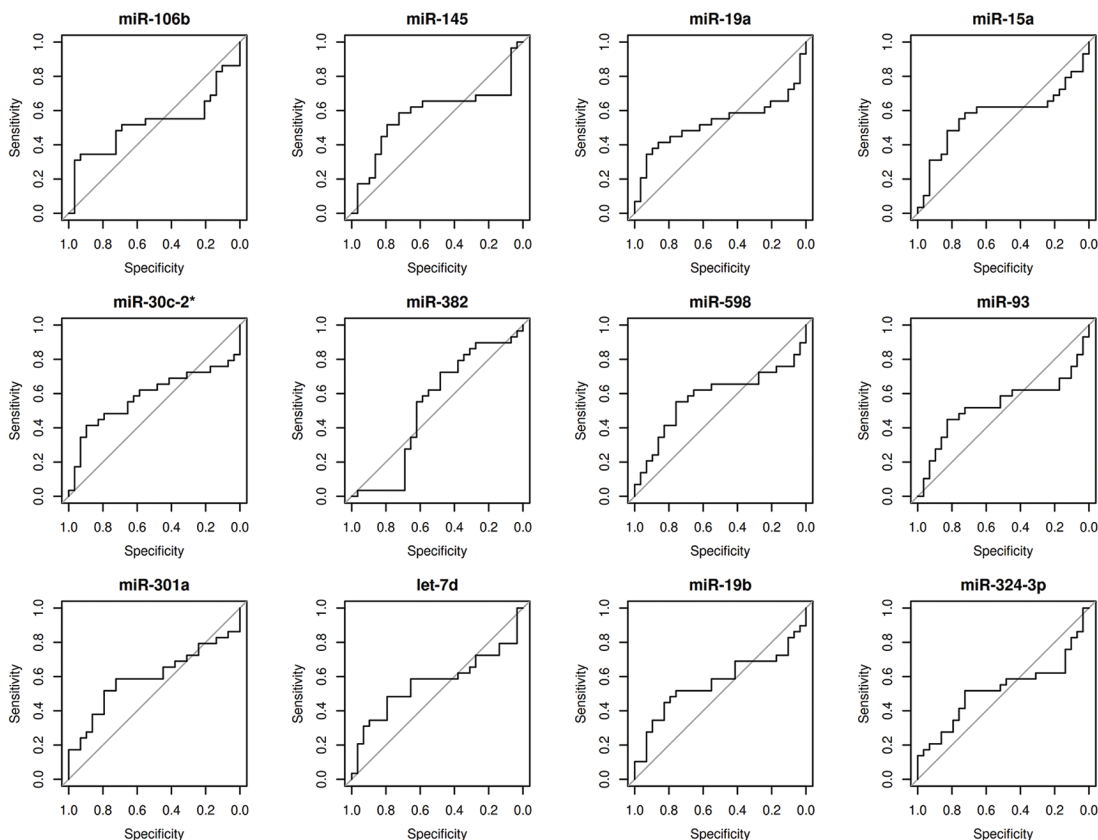


**FIG. 11.** Venn diagram showing the overlap between miRNAs identified by DESeq2, LROC, and gAUC in the cervical cancer dataset. Number of selected features by DESeq2 is defined using the criteria, e.g., BH-adjusted  $p$  value < 0.05 and absolute  $\log_2$  fold-change > 0.6. LROC, length of the receiver operating characteristic; gAUC, generalized area under the curve.

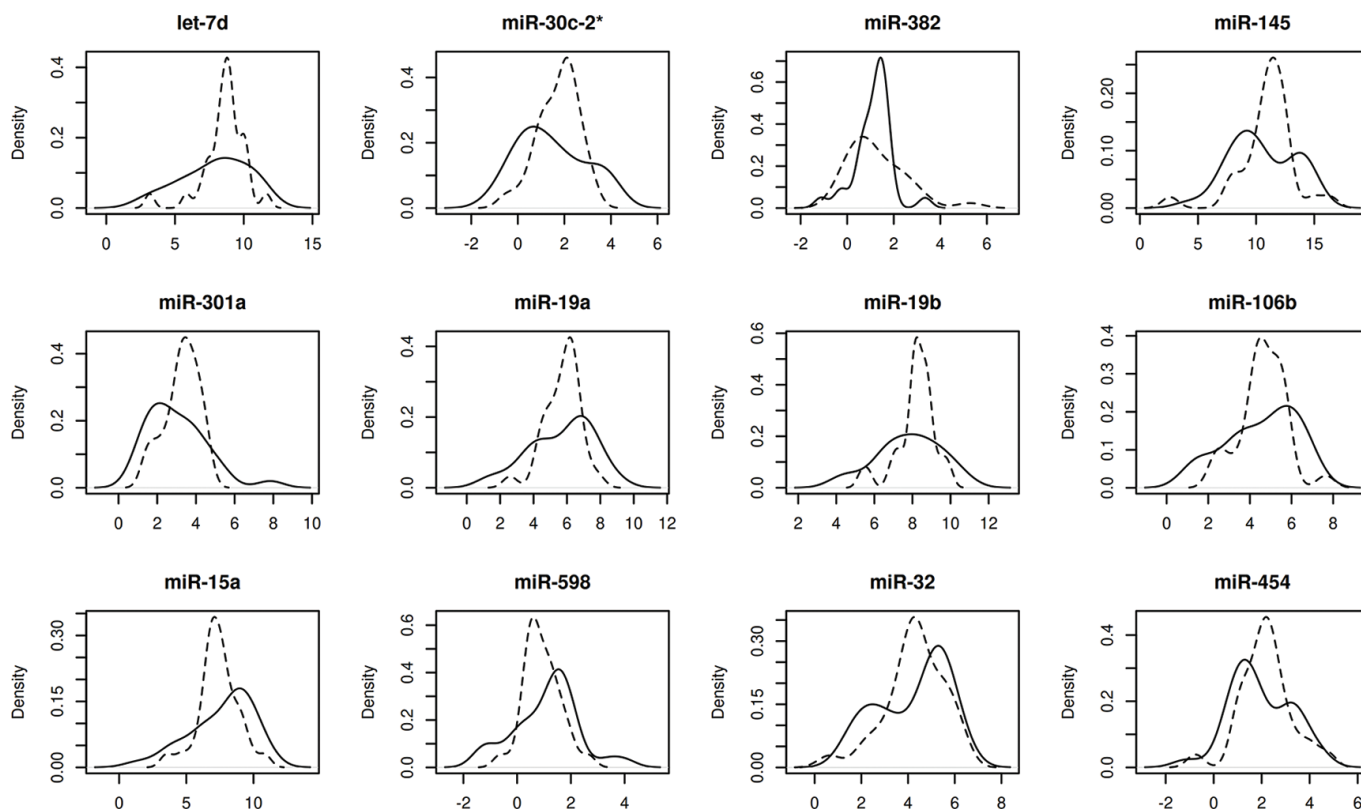
with experimentally observed context-dependent or dual functional roles. Together with additional less-characterized candidates (e.g., miR-598), these findings suggest that improper miRNA profiles may reflect biologically meaningful heterogeneity in CC and warrant further functional investigation.

We acknowledge several limitations of this work. The primary limitation is that gAUC and LROC are used as heuristic screening scores without  $p$  values or formal type I error control; therefore, results depend on post-hoc threshold selection and do not support strict inferential claims. Additional limitations include the restriction to binary outcomes (two-group comparisons) and the lack of an inferential calibration framework for the ROC-derived indices (e.g., null-only behavior, specificity, and nominal type I error or FPR control), beyond the empirical FPR sensitivity benchmarks provided in the [Supplementary Materials (Section B)]. These issues should be addressed in future methodological work, for example, by developing data-driven thresholding strategies or formal test statistics with known sampling distributions.

In conclusion, improper, non-monotonic expression profiles in RNA-Seq data can be difficult to prioritize using mean-shift-focused DE testing alone. In both the simulation studies and the CC case



**FIG. 12.** Receiver operating characteristic (ROC) curves for the top improper miRNAs identified in the cervical cancer dataset. Each panel represents an individual miRNA exhibiting a non-monotonic expression profile, with both low and high expression levels contributing to discrimination between tumor and control samples.



**FIG. 13.** Density plot of distributions for improper miRNA expression profiles detected by ROC-derived indices. These miRNAs are a subset of the top candidates supported by the literature.

ROC, receiver operating characteristic.

study, gAUC (and, in some settings, LROC) provided complementary screening rankings that may help flag candidate non-monotonic features for follow-up analysis. Because these indices rely on heuristic thresholds and lack error control, the findings should be interpreted as exploratory and require independent validation.

**Ethics Committee Approval:** Not applicable.

**Informed Consent:** Not applicable.

**Data Sharing Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Authorship Contributions:** Concept- D.G.; Design- M.B.G., E.Ö., D.G.; Supervision- D.G.; Funding- M.B.G., Data Collection or Processing- M.B.G., A.D.Ö., H.F.K.; Analysis and/or Interpretation- M.B.G., E.Ö., Ü.E., H.F.K., D.G.; Literature Review- M.B.G., E.Ö., Ü.E., A.D.Ö., H.F.K., D.G.; Writing- M.B.G., E.Ö., Ü.E., A.D.Ö., H.F.K., D.G.; Critical Review- M.B.G., E.Ö., A.D.Ö., D.G.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**Funding:** This work was supported by the Scientific Research Projects Unit of Sakarya University, Türkiye (grant number: 2025-27-73-57). The funder had no role in the study design, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to submit the article for publication.

**Supplementary Materials Section A and B:** <https://balkanmedicaljournal.org/img/files/balkan-2026-1-309-supplementary.pdf>

## REFERENCES

1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10:57-63. [CrossRef]
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372:793-795. [CrossRef]
3. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol.* 2017;18:83. [CrossRef]
4. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning.* 2003;52:91-118. [CrossRef]
5. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc.* 2002;97:77-87. [CrossRef]
6. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA.* 2002;99:6567-6572. [CrossRef]
7. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13. Erratum in: *Genome Biol.* 2016;17:181. [CrossRef]
8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. [CrossRef]
9. Wang T, Li B, Nelson CE, Nabavi S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinformatics.* 2019;20:40. [CrossRef]
10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-140. [CrossRef]

11. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. [\[CrossRef\]](#)
12. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics.* 2010;11:422. [\[CrossRef\]](#)
13. Li J, Witten DM, Johnstone IM, Tibshirani R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics.* 2012;13:523-538. [\[CrossRef\]](#)
14. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res.* 2013;22:519-536. [\[CrossRef\]](#)
15. Rosati D, Palmieri M, Brunelli G, et al. Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: a review. *Comput Struct Biotechnol J.* 2024;23:1154-1168. [\[CrossRef\]](#)
16. Ma CS, Tangye SG. Inborn errors of immunity: the Goldilocks effect-susceptibility to disease due to a little too much or a little too little. *Clin Exp Immunol.* 2023;212:93-95. [\[CrossRef\]](#)
17. Guo X, Zhang Y, Zheng L, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med.* 2018;24:978-985. Erratum in: *Nat Med.* 2018;24:1628. [\[CrossRef\]](#)
18. Granno S, Nixon-Abell J, Berwick DC, et al. Downregulated Wnt/ $\beta$ -catenin signalling in the Down syndrome hippocampus. *Sci Rep.* 2019;9:7322. [\[CrossRef\]](#)
19. Carilli M, Gorin G, Choi Y, Chari T, Pachter L. Biophysical modeling with variational autoencoders for bimodal, single-cell RNA sequencing data. *Nat Methods.* 2024;21:1466-1469. [\[CrossRef\]](#)
20. Kuehn HS, Boast B, Rosenzweig SD. Inborn errors of human IKAROS: LOF and GOF variants associated with primary immunodeficiency. *Clin Exp Immunol.* 2022;212:129-136. [\[CrossRef\]](#)
21. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005;310:644-648. [\[CrossRef\]](#)
22. MacDonald JW, Ghosh D. COPA--cancer outlier profile analysis. *Bioinformatics.* 2006;22:2950-2951. [\[CrossRef\]](#)
23. Tibshirani R, Hastie T. Outlier sums for differential gene expression analysis. *Biostatistics.* 2007;8:2-8. [\[CrossRef\]](#)
24. Hartigan JA, Hartigan PM. The dip test of unimodality. *Ann Statist.* 1985;13:70-84. [\[CrossRef\]](#)
25. Rousseeuw PJ. Least median of squares regression. *J Am Stat Assoc.* 1984;79:871-880. [\[CrossRef\]](#)
26. Moody L, Mantha S, Chen H, Pan YX. Computational methods to identify bimodal gene expression and facilitate personalized treatment in cancer patients. *J Biomed Inform.* 2019;100:100001. [\[CrossRef\]](#)
27. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform.* 2009;7:199-216. [\[CrossRef\]](#)
28. Tong P, Chen Y, Su X, Coombes KR. SIBER: systematic identification of bimodally expressed genes using RNAseq data. *Bioinformatics.* 2013;29:605-613. [\[CrossRef\]](#)
29. Liu TC, Kalugin PN, Wilding JL, Bodmer WF. GMMchi: gene expression clustering using Gaussian mixture modeling. *BMC Bioinformatics.* 2022;23:457. [\[CrossRef\]](#)
30. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;28:1-26. [\[CrossRef\]](#)
31. Martínez-Cambor P, Corral N, Rey C, Pascual J, Cernuda-Morollón E. Receiver operating characteristic curve generalization for non-monotone relationships. *Stat Methods Med Res.* 2017;26:113-123. [\[CrossRef\]](#)
32. Franco-Pereira AM, Nakas CT, Pardo MC. Biomarker assessment in ROC curve analysis using the length of the curve as an index of diagnostic accuracy: the binormal model framework. *ASIA Advances in Statistical Analysis.* 2020;104(4):625-647. [\[CrossRef\]](#)
33. Witten D, Tibshirani R, Gu SG, Fire A, Lui WO. Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. *BMC Biol.* 2010;8:58. [\[CrossRef\]](#)
34. Oliveros JC. Venny: An interactive tool for comparing lists with Venn's diagrams; 2007-2015. Accessed: 2025-10-15. [\[CrossRef\]](#)
35. Moody L, Xu GB, Pan YX, Chen H. Genome-wide cross-cancer analysis illustrates the critical role of bimodal miRNA in patient survival and drug responses to PI3K inhibitors. *PLoS Comput Biol.* 2022;18:e1010109. [\[CrossRef\]](#)
36. Oleksiewicz U, Liloglou T, Tasopoulou KM, et al. Cytochrome has bimodal: tumour suppressor and oncogene functions in lung cancer cell lines. *Hum Mol Genet.* 2013;22:3207-3217. [\[CrossRef\]](#)
37. Parodi S, Pistoia V, Muselli M. Not proper ROC curves as new tool for the analysis of differentially expressed genes in microarray experiments. *BMC Bioinformatics.* 2008;9:410. [\[CrossRef\]](#)
38. Silva-Fortes C, Amaral Turkman MA, Sousa L. Arrow plot: a new graphical tool for selecting up and down regulated genes and genes differentially expressed on sample subgroups. *BMC Bioinformatics.* 2012;13:147. [\[CrossRef\]](#)
39. De Santis C, Götte M. The Role of microRNA Let-7d in female malignancies and diseases of the female reproductive tract. *Int J Mol Sci.* 2021;22:7359. [\[CrossRef\]](#)
40. Gong Y, Luo G, Zhang S, Chen Y, Hu Y. Transcriptome sequencing analysis reveals miR-30c-5p promotes ferroptosis in cervical cancer and inhibits growth and metastasis of cervical cancer xenografts by targeting the METTL3/KRAS axis. *Cell Signal.* 2024;117:111068. [\[CrossRef\]](#)
41. Guo H, Li J, Fan F, Zhou P. LINC00707 Regulates miR-382-5p/VEGFA pathway to enhance cervical cancer progression. *J Immunol Res.* 2021;2021:5524632. [\[CrossRef\]](#)
42. Li Q, Yu X, Yang L. MiR-145 inhibits cervical cancer progression and metastasis by targeting WNT2B by Wnt/ $\beta$ -catenin pathway. *Int J Clin Exp Pathol.* 2019;12:3740-3751. [\[CrossRef\]](#)
43. Peng LN, Shi WT, Feng HR, Wei CY, Yin QN. Effect of miR-301a/PTEN pathway on the proliferation and apoptosis of cervical cancer. *Innate Immun.* 2019;25:217-223. [\[CrossRef\]](#)
44. Wang W, Liu L, Tian Y. miR-19-3p targets PTEN to regulate cervical cancer cell proliferation, invasion, and autophagy. *Genet Res (Camb).* 2023;2023:4784500. [\[CrossRef\]](#)
45. Sun XY, Han XM, Zhao XL, Cheng XM, Zhang Y. MiR-93-5p promotes cervical cancer progression by targeting THBS2/MMPS signal pathway. *Eur Rev Med Pharmacol Sci.* 2019;23:5113-5121. [\[CrossRef\]](#)
46. Cheng Y, Guo Y, Zhang Y, You K, Li Z, Geng L. MicroRNA-106b is involved in transforming growth factor  $\beta$ 1-induced cell migration by targeting disabled homolog 2 in cervical carcinoma. *J Exp Clin Cancer Res.* 2016;35:11. [\[CrossRef\]](#)
47. Zhao XQ, Tang H, Yang J, Gu XY, Wang SM, Ding Y. MicroRNA-15a-5p down-regulation inhibits cervical cancer by targeting TP53INP1 in vitro. *Eur Rev Med Pharmacol Sci.* 2019;23:8219-8229. [\[CrossRef\]](#)