GMR
The Original

INTEGRATED COMPUTATIONAL CHARACTERIZATION OF CIRCULAR RNA BIOMARKERS, REGULATORY NETWORKS, AND DERIVED PEPTIDES IN LUNG ADENOCARCINOMA

Ayushi Malviya[1], Rajabrata Bhuyan[2*]

[1]Department of Bioscience and Biotechnology, Banasthali Vidyapith, Rajasthan, India, Email: malviya.ayushi29@gmail.com; ORCID: 0000-0001-5399-027X
[2]Assistant Professor, Department of Bioscience and Biotechnology, Banasthali Vidyapith, Niwai-Tonk, Rajasthan, India. 304022,
E-mail: rajabrata001@gmail.com, ORCID: 0000-0001-7777-8230
*Corresponding Author: Rajabrata Bhuyan: E-mail: rajabrata001@gmail.com,

ABSTRACT
Lung adenocarcinoma (LUAD) represents a molecularly heterogeneous malignancy characterised by a paucity of validated biomarkers for early-stage detection and precision therapeutic targeting. Although circular RNAs (circRNAs) have emerged as critical regulators of tumour biology, their mechanistic contributions to LUAD progression at both the post-transcriptional and translational levels remain insufficiently characterised. In this study, we performed an integrated computational analysis combining circRNA expression profiling, regulatory network reconstruction, multi-criteria computational biomarker prioritization, RNA-binding protein (RBP) interaction, and characterization of dysregulated circRNA-derived peptides in LUAD. We analysed publicly available RNA-Seq datasets and identified 52,744 circRNAs, among which 1,480 were significantly dysregulated, including 855 downregulated and 625 upregulated circRNAs. A multi-algorithm consensus framework shortlisted 34 high-confidence biomarker candidates from which, hsa_circ_0058622, hsa_circ_0085740, hsa_circ_0012248, and hsa_circ_0024109 showed the strongest consensus across different models. Integrated circRNA-miRNA-mRNA and RBP interaction analyses predicted EIF4A3, AGO2, and HuR as important regulatory hubs linked with epithelial-mesenchymal transition, metabolic adaptation, and oncogenic signalling pathways. Additionally, we predicted four novel circRNA-derived peptides and characterized them using physicochemical analysis and 500 ns molecular dynamics simulations. Nonetheless, these results are computational and depend on experimental validation to establish their biological significance.

KEYWORDS: LUAD; circRNA; circRNA-encoded peptides; machine learning; cancer biomarkers; RNA-binding proteins

1. INTRODUCTION

Lung cancer represents one of the most formidable challenges in modern oncology, accounting for more cancer-related deaths annually than any other malignancy worldwide, with lung adenocarcinoma (LUAD) emerging as the predominant histological subtype among non-small cell lung cancers (NSCLC), responsible for nearly 85% of all diagnosed cases (Sausville et al., 2017; Thandra et al., 2021). Despite major improvements in targeted therapy and precision oncology, the overall survival rate of LUAD patients remains poor due to late diagnosis, extensive molecular heterogeneity, and frequent therapeutic resistance (Chen et al., 2014; Seguin et al., 2022). This clearly shows the urgent need for reliable molecular biomarkers and mechanistic targets that can improve early detection and therapeutic stratification.

Recent studies have highlighted the importance of non-coding RNAs in LUAD progression, particularly circular RNAs (circRNAs) (Alimohammadi et al., 2023; Chen et al., 2018; Fan et al., 2020; Gilyazova et al., 2024; Ju et al., 2021). CircRNAs are covalently closed RNA molecules generated through back-splicing events, making them highly stable compared with linear RNAs (Chen, 2016; Cocquerelle et al., 1993). Their remarkable stability, tissue-specific expression, and disease-associated dysregulation make them attractive candidates for biomarker discovery (Jeck et al., 2013; Salzman et al., 2013). Initially considered transcriptional artefacts, circRNAs are now recognised as active regulators of gene expression, with many reported to function through circRNA-miRNA-mRNA regulatory axes or interact directly with RNA-binding proteins (RBPs), thereby influencing transcriptional and post-transcriptional cellular processes (Malviya and Bhuyan, 2025) .

A growing body of evidence also suggests that certain circRNAs possess coding potential. Through internal ribosome entry sites (IRESs) or m6A-mediated translation, some circRNAs can produce biologically active peptides from small open reading frames (smORFs)(Chen et al., 2021; Wen et al., 2022). These circRNA-derived peptides have been linked with tumour proliferation, apoptosis, metabolic reprogramming, and oncogenic signalling in several cancers (Lin et al., 2025). Also, the circular structure of circRNAs provides greater resistance to degradation, potentially supporting sustained peptide translation. This makes circRNA-derived peptides particularly interesting from both mechanistic and therapeutic perspectives.

However, important gaps still remain. Most LUAD studies have examined individual circRNAs or isolated regulatory pathways, while the broader regulatory landscape, biomarker relevance, and translational potential of circRNAs remain poorly explored within a unified framework. And very few studies have systematically integrated network biology, AI/ML-based biomarker prioritisation, and peptide prediction approaches in LUAD.

To address this gap, the present study performed an integrated computational investigation of the LUAD circRNA landscape. We integrated differential expression analysis, circRNA-miRNA-mRNA network reconstruction, RBP interaction mapping, AI/ML-driven biomarker prioritisation, and structural characterization of circRNA-derived peptides using molecular dynamics simulations. This study aims to improve the current understanding of circRNAs in LUAD and to identify functionally relevant candidates with possible diagnostic and therapeutic significance.

2. MATERIALS AND METHODS

2.1. Dataset collection and study design

The NCBI Sequence Read Archive (SRA) was systematically explored to retrieve publicly available RNA-Seq datasets optimised for circRNA detection (Leinonen et al., 2011). Only transcriptomic paired-end reads from LUAD samples enriched for non-coding RNAs were considered. The selected datasets were categorised based on their intended application: ncRNA-enriched libraries were retained for circRNA identification (Accession IDs: SRP335547; SRP089923; SRP048484; SRP047399), while libraries containing both cancerous and non-cancerous samples were utilised for differential expression analyses. All RNA-Seq libraries were obtained from the European Nucleotide Archive (Harrison et al., 2021). The human reference genome (GRCh37.fa) along with its corresponding annotation file (ucsc_GRCh37.gtf) was downloaded from the UCSC Genome Browser (Kent et al., 2002). For annotation of identified circRNAs, reference files from the CircBase database (hsa_hg19_circRNA.bed; hsa_hg19_circRNA.gtf) were employed (Glažar et al., 2014) . In addition, mature and annotated miRNA sequences were retrieved from the miRBase database (homo_sapiens.fa) (Miranda et al., 2006).

2.2. CircRNA detection

Pre-processing and quality control of each dataset were conducted using FastQC (http:// www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess read quality. Reads with a base quality score below Q20 were dropped and the sequences were trimmed using Trimmomatic (Bolger et al., 2014). The quality-filtered reads were then subjected to circRNA detection using CIRI2, the updated version of the CIRI (circRNA identifier) algorithm, which relies on chiastic clipping signals and multiple filtration steps to reliably and unbiasedly detect circRNAs (Gao et al., 2015). It utilises BWA for spliced alignment against the human reference genome (Li and Durbin, 2009). CIRI2 was chosen for this study because it offers a strong balance of accuracy, computational efficiency, and low resource demand. CircRNA detection was performed individually for each sample, and the resulting output files were subsequently merged into a consolidated BED file for downstream analyses.

2.3. Differential gene expression studies

StringTie (Pertea et al., 2015) was used for transcript normalisation, abundance estimation, and generation of count data for downstream analysis using the Bioconductor package DESeq2 (Love et al., 2017, 2014). Differential expression analysis was performed using TPM-normalised count values. CircRNA transcripts from cancer and normal samples were further filtered based on an adjusted p-value threshold of 0.05, followed by refinement using a $\log_2$ fold change cut-off of ±2.

2.4. Prediction of target miRNAs and RBPs

The interactions between significantly differentially expressed circRNAs and their target miRNAs were predicted using the miRanda tool (Miranda et al., 2006), applying a threshold score of 150 and a minimum free energy cutoff of -25 kcal/mol to ensure the inclusion of a sufficiently comprehensive pool for feature extraction required in the next steps. An interaction network comprising the top-ranking circRNAs and their high-confidence miRNA targets was subsequently constructed and visualised using Cytoscape (Shannon et al., 2003). Simultaneously, target RNA-binding proteins (RBPs) of the significantly dysregulated circRNAs were retrieved from the circInteractome database (Dudekula et al., 2016), and circRNA-RBP interaction networks were generated in Cytoscape for both upregulated and downregulated circRNAs.

2.5. Machine Learning-mediated biomarker potential prediction

A circRNA feature matrix was constructed comprising six biologically relevant attributes: Exclusivity Score, Consistency Score, Differential Expression Score (DE_Score), Binding Affinity Score, Regulatory Potential Score, and Sponge Capacity Score (Supplementary Table S1). Collectively, these features capture complementary aspects of circRNA biology, including disease specificity, expression dysregulation, molecular interaction strength, regulatory influence, and miRNA sponge activity, thereby providing a multidimensional representation of biomarker potential. This feature matrix (Supplementary Table S2) served as the input to three complementary computational ranking strategies. These included (i) a feature-weighted statistical scoring framework, (ii) unsupervised clustering approaches (PCA, K-means, Isolation Forest), and (iii) a multilayer machine learning (ML) algorithm used as a nonlinear feature integration model to generate relative prioritization scores (Sarker, 2021). The approach was not intended as a diagnostic classifier but rather as a nonlinear feature integration model capable of capturing complex relationships among multiple biologically derived scoring attributes. The resulting output was used solely for candidate prioritization and consensus ranking. These complementary approaches were selected because they capture distinct aspects of candidate prioritization, including feature-based ranking, unsupervised pattern discovery, anomaly detection, and nonlinear feature interactions.

2.6. External validation of candidate biomarker genes

To further assess the clinical relevance of the prioritised circRNA candidates, the expression patterns and prognostic significance of their corresponding host genes (PTK2, MMP1, KRT17, CYP24A1, PPP2R5A, SP100, NASP, SMCHD1, SHC2, and TELO2) were evaluated using independent publicly available datasets. Differential gene expression between lung adenocarcinoma (LUAD) and normal lung tissues was validated using the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) platform, which integrates RNA sequencing data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) project. Overall survival (OS) analysis was performed in GEPIA2 using the median expression value as the cut-off to stratify patients into high- and low-expression groups, and statistical significance was determined using the log-rank test. Additionally, the diagnostic performance of the selected host genes was evaluated using TCGA-LUAD RNA sequencing count data downloaded through the TCGAbiolinks R package (v2.4.0). Raw gene counts were normalised to $\log_2$ counts per million (logCPM) using the edgeR package (v4.10.1). Receiver operating characteristic (ROC) curve analysis was performed in R (v4.6) using the pROC package to distinguish tumour from normal samples, and the area under the curve (AUC), 95% confidence intervals, sensitivity, and specificity were calculated to assess the diagnostic accuracy of each candidate gene.

2.7. Gene Ontology (GO) and pathway enrichment studies

To identify the miRNA-targeted genes for subsequent GO analysis, the multimiR (Ru et al., 2014) package in R was employed. Three target prediction databases (miRanda, miRBase and TargetScan) were queried (Kozomara et al., 2019; Miranda et al., 2006; Riffo-Campos et al., 2016), and only those genes that were predicted by at least two of the three databases and/or were experimentally validated were included in the final list. Functional annotation of significantly dysregulated circRNA target genes, predicted through their interacting miRNAs, was performed using Enrichr server (Chen et al., 2013). Gene functions were categorised into three domains: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). These categories were separately analysed for the upregulated and downregulated gene sets. Additionally, pathway enrichment analysis was conducted using the KEGG database (Kanehisa et al., 2017).

2.8. Identification of ORFs and IRES Elements

The protein-coding potential of detected circRNAs was assessed using CPC2 in combination with circPrimer and the TransCirc database (Huang et al., 2021; Kang et al., 2017; Zhong and Feng, 2022). These tools were employed to predict open reading frames (ORFs) and internal ribosome entry sites (IRES) within the circRNA sequences. CircRNAs that consistently exhibited coding potential across all three tools were subjected to further analyses, including evaluation of back-splice junction (BSJ) spanning ORF length, assessment for the presence of Kozak consensus sequences, and confirmation of IRES elements. To strengthen confidence in the predictions, data were integrated from public circRNA repositories, including circBank, circBase, IRESsite, and TransCirc (Glažar et al., 2014; Huang et al., 2021; Liu et al., 2019; Mokrejs et al., 2006), enabling cross-prioritization and enrichment of the dataset with experimentally supported translational circRNAs.

2.9. Preprocessing of amino acid sequences and functional characterization

The predicted short peptide sequences were manually curated before downstream structural and functional studies to remove duplication and standardize formatting for cross-platform analytical compatibility. The ProtParam program was then used to characterize the physicochemical parameters of the selected peptides, including molecular weight, isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY).(Wilkins et al., 1999) When taken as a whole, these factors offered a quantitative foundation for determining peptide stability, aqueous solubility, and initial biological activity profiles.

2.10. Protein Structure Prediction and Molecular Dynamics Simulations

The three-dimensional structures of the predicted small peptides were modelled using AlphaFold2 (Abramson et al., 2024).All-atom MD simulations of 500 ns duration were performed in GROMACS, following the standard lysozyme-in-water tutorial protocol with parameters adapted for peptide systems (Lemkul, 2024; Van Der Spoel et al., 2005). Each peptide was placed in a cubic simulation box with a minimum distance of 1.0 nm from the box edge, explicitly solvated using the TIP3P water model, and neutralised with Na^+/Cl^- counter ions. CHARMM36 force field was applied, and energy minimisation was performed using the steepest descent algorithm until the maximum force convergance was achieved. Equilibration was carried out sequentially under NVT and NPT ensembles. Production MD simulations of 500 ns duration were performed in GROMACS. Structural stability and conformational dynamics were assessed throughout the trajectory using RMSD, RMSF, radius of gyration (Rg), and SASA. Free Energy Landscape (FEL) analysis was performed by projecting the trajectory onto RMSD and Rg as collective variables and computing the Gibbs free energy ln, where conformational minima identify thermodynamically stable states sampled by each peptide over the simulation period (Riccardi et al., 2012)

3. RESULTS

3.1. Characterization of circRNA expression and host gene associations

We predicted the circRNA landscape in LUAD using the CIRIv2 pipeline and later annotated the detected circRNAs against the circBase database. CIRI2 uses a maximum likelihood-based multiple seed matching strategy to separate genuine back-splice junction (BSJ) reads from incorrectly mapped non-BSJ reads. This helps reduce false discoveries while still maintaining good sensitivity. The tool also improves read-classification accuracy and supports multithreading, so the analysis runs faster and consumes less RAM. Simple, but effective.

This workflow helped us accurately identify and classify circRNAs into two broad groups: Novel circRNAs and Known circRNAs. Novel circRNAs referred to circular RNA transcripts detected in LUAD samples that did not match any previously annotated entries in circBase. In other words, these circRNAs remain uncharacterised and may represent newly predicted transcripts. Known circRNAs, on the other hand, showed genomic coordinates and sequence features consistent with entries already reported in circBase.

In the LUAD dataset, we identified 18,922 Novel circRNAs, whereas 33,822 circRNAs belonged to the Known category. Most circRNAs mapped to exonic regions (32,452), which aligns well with existing reports showing that exonic circRNAs dominate cancer-associated circRNA populations. A smaller fraction originated from intronic regions (1,194); this suggests that circRNA biogenesis is not restricted only to exons. We also identified 2,656 overlapping circRNAs spanning multiple genomic regions, indicating complex transcriptional hotspots that may support alternative circularisation events or transcript overlap.

Interestingly, 1,199 circRNAs belonged to non-coding regions, highlighting their likely regulatory significance beyond protein-coding potential. And only 86 antisense circRNAs were detected. Though small in number, these antisense circRNAs may still play specific roles in transcriptional regulation, gene silencing, or expression interference mechanisms.

3.2. Differential expression of circRNAs in LUAD vs normal lung cells

Differential expression analysis revealed major transcriptomic alterations in lung adenocarcinoma (LUAD) when compared with normal lung tissue. The heatmap first showed noticeable variation in circRNA expression patterns across tumour and normal samples (Figure 1a). Distinct clustering appeared among both samples and circRNAs, suggesting strong expression similarities within particular groups. Some circRNAs remained consistently overexpressed in tumour tissues, while others showed strong suppression. And this uneven expression pattern points towards major regulatory disturbances during LUAD progression.

Moving on to the volcano plot analysis, we identified 1,480 significantly dysregulated circRNAs with an adjusted p-value below 0.05 (Figure 1b). Among them, 625 circRNAs were upregulated, whereas 855 circRNAs showed downregulation in tumour samples. The larger fraction of downregulated circRNAs (57.8%) compared with upregulated circRNAs (42.2%) suggests that LUAD progression involves widespread transcriptional repression along with activation of tumour-associated pathways. Quite expected in aggressive tumours; still, the scale of suppression was notable.

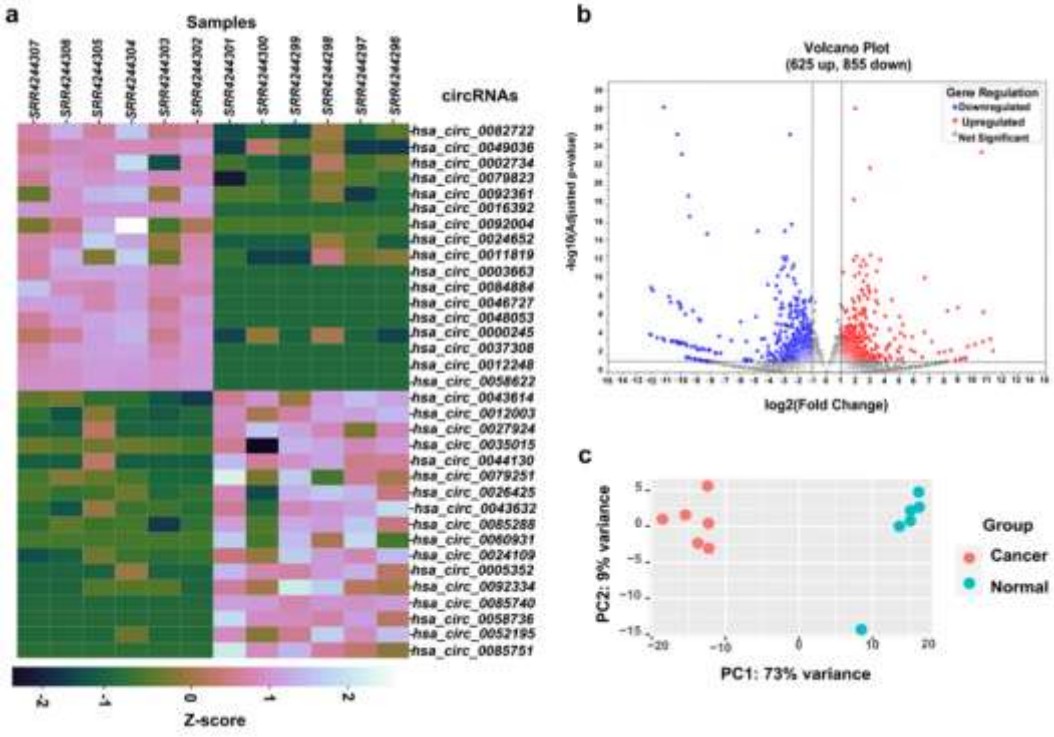


Figure 1 Differential expression of circRNAs. (a) Heatmap displaying normalised circRNA expression (Z-scores) of the top 34 prioritised circRNAs. Rows represent selected circRNAs, and columns denote individual samples. Expression patterns reveal clustering among circRNAs and samples based on expression similarity, (b) Volcano plot depicting differential circRNA expression in LUAD samples. The plot displays $\log_2$ fold changes versus -$\log_{10}$ adjusted p-values for detected circRNAs, highlighting 625 upregulated and 855 downregulated circRNAs. Circles represent individual circRNAs classified as upregulated, downregulated, or not significant based on statistical thresholds; (c) Principal component analysis (PCA) plot showing separation between LUAD cancer and normal samples based on circRNA expression. PC1 and PC2 together explain major variance, with distinct clustering observed for each group.

The PCA analysis also showed a clear separation between LUAD and normal samples based on circRNA expression profiles (Figure 1c). PC1 and PC2 together explained a major proportion of the total variance, allowing distinct clustering of the two sample groups. The separation was visually evident, with very little overlap between tumour and normal tissues.

This clearly shows why circRNA expression signatures can effectively distinguish cancerous samples from non-cancerous tissues. The predominance of downregulated circRNAs may reflect the gradual loss of normal epithelial identity and differentiation pathways during tumour progression. But the upregulated circRNAs likely include transcripts linked with oncogenic signalling, uncontrolled proliferation, metabolic rewiring, and immune escape mechanisms. So, the simultaneous presence of both suppressed and activated circRNA populations indicates extensive reorganisation of transcriptional networks during LUAD tumourigenesis.

3.3. CircRNA-miRNA-mRNA regulatory network and RBP interactions

Analysis of the upregulated circRNAs revealed a distinct interaction network characterised by a smaller but relatively focused set of circRNAs that engage with specific miRNAs (Figure 2a). Prominent circRNAs included hsa_circ_0012003, hsa_circ_0026425, hsa_circ_0027924, hsa_circ_0035015, hsa_circ_0043614, hsa_circ_0043632, hsa_circ_0044130, hsa_circ_0052195, hsa_circ_0085288, and hsa_circ_0085740. Several of these circRNAs acted as hub molecules by interacting with multiple miRNAs; for example, hsa_circ_0035015 (targeting hsa-miR-665, hsa-miR-770-5p, and hsa-miR-4739), hsa_circ_0043614 (targeting hsa-miR-7158-5p and hsa-miR-6894-5p), and hsa_circ_0043632 (targeting hsa-miR-4776-3p, hsa-miR-7106-5p, hsa-miR-1293, and hsa-miR-6791-3p). This pattern suggests that a few highly expressed circRNAs may exert substantial influence over a diverse repertoire of miRNAs.

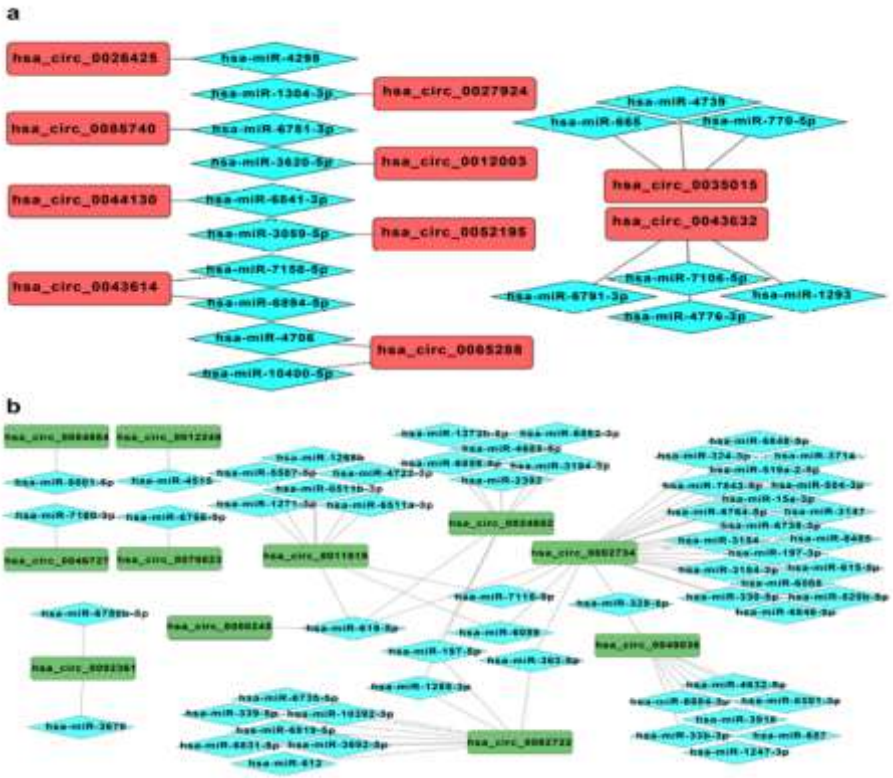


Figure 2 Network visualisation of highly dysregulated circRNAs and their predicted interacting miRNAs in LUAD. Rectangular nodes represent individual circRNAs, and diamond-shaped nodes represent miRNAs, and edges show predicted regulatory relationships, highlighting possible regulatory roles in LUAD progression. (a) Depiction of upregulated circRNAs (red, rectangular nodes) and their miRNA associations (teal, diamond-shaped nodes) and (b) downregulated circRNAs (green, rectangular nodes) and their miRNA (teal, diamond -shaped nodes) associations.

Network analysis of the downregulated circRNAs (Figure 2b) revealed a complex regulatory catalogue comprising several circRNAs, notably hsa_circ_0049036, hsa_circ_0046727, hsa_circ_0024652, hsa_circ_0092361, hsa_circ_0002734, hsa_circ_0012248, hsa_circ_0084884, hsa_circ_0000245, hsa_circ_0011819, hsa_circ_0082722, and hsa_circ_0079823, that interact with a wide spectrum of miRNAs, including hsa-miR-4632-5p, hsa-miR-330-5p, hsa-miR-3714, hsa-miR-519a-2-5p, hsa-miR-6088, hsa-miR-615-5p, hsa-miR-197-5p, hsa-miR-612, hsa-miR-15a-3p, and hsa-miR-324-3p, among others. Several circRNAs exhibited hub-like properties by targeting multiple miRNAs, suggesting that they may exert broad regulatory influence. Conversely, certain miRNAs such as hsa-miR-197-5p, hsa-miR-1285-3p, hsa-miR-619-5p, hsa-miR-363-5p, hsa-miR-7110-5p, and hsa-miR-328-5p were recurrently targeted by multiple circRNAs, indicating their central roles in the post-transcriptional network.

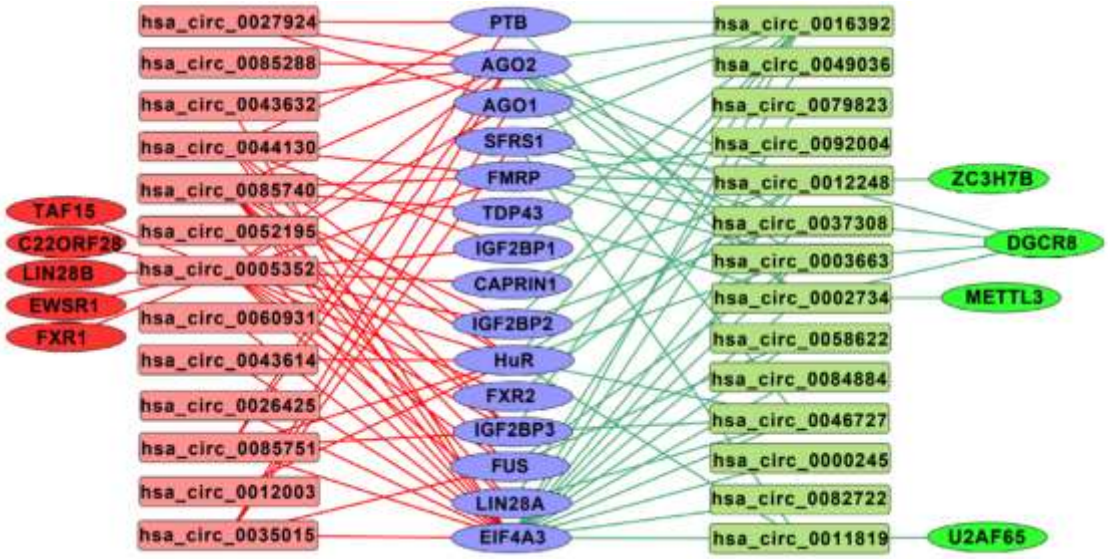


Figure 3. Network visualisation of circRNA interactions with RNA-binding proteins (RBPs) in LUAD samples. Nodes represent circRNAs (green: downregulated, red: upregulated) and RBPs (purple: targeted by both circRNA categories, light green: targeted only by downregulated circRNAs, yellow: targeted only by upregulated circRNAs). Edges indicate predicted binding interactions, with colour coordination reflecting circRNA regulation status (green edges for downregulated circRNAs, red edges for upregulated circRNAs). This network highlights the complex regulatory relationships between circRNAs and RBPs involved in LUAD.

The circRNA-RBP interaction network visualised in Figure 3 indicates a complex regulatory architecture involving 27 differentially expressed circRNAs (13 upregulated and 14 downregulated) and 24 RNA-binding proteins exhibiting distinct interaction patterns. Network topology analysis demonstrated that upregulated circRNAs displayed significantly higher RBP connectivity than their downregulated counterparts, with hsa_circ_0005352 emerging as a central hub interacting with 17 RBPs, followed by hsa_circ_0085740 (10 RBPs) and hsa_circ_0044130 (6 RBPs). Among the RBPs predicted, EIF4A3 functioned as the most promiscuous binding partner, interacting with 22 circRNAs from both expression cohorts, while AGO2 targeted 15 circRNAs and HuR bound 11 circRNAs, establishing these proteins as convergent regulatory hubs bridging upregulated and downregulated circRNA populations. The IGF2BP protein family (IGF2BP1/2/3) demonstrated preferential binding to upregulated circRNAs, particularly hsa_circ_0005352, hsa_circ_0044130, and hsa_circ_0085740, suggesting their role in stabilising oncogenic circRNA transcripts.
Notably, METTL3 and DGCR8 exclusively interacted with downregulated circRNAs, while C22ORF28, EWSR1, TAF15, and specific members of the LIN28 and FXR families showed selective binding to upregulated circRNAs. Splicing factors U2AF65 and ZC3H7B uniquely associated with downregulated circRNAs, whereas stress granule components CAPRIN1 and FXR1/2 exhibited differential targeting patterns. The extensive interconnectivity of upregulated circRNAs with multiple RBPs suggests enhanced post-transcriptional stability mechanisms driving their accumulation in lung cancer cells.

3.4. Computational prioritization of candidate circRNA biomarkers

We prioritised 34 circRNA candidates through an integrated computational ranking framework combining three complementary analytical strategies: a MLneural-derived nonlinear scoring model (DL score), unsupervised clustering (Cluster score), and a feature-weighted statistical scoring approach (Stats score). Candidate circRNAs that consistently ranked highly across all three methods were considered high-confidence biomarkers (Figure 4). All prioritised candidates received consistently high ML-derived scores, indicating strong agreement within the feature integration framework, and were derived from genes with diverse biological functions. Each analytical approach highlighted different aspects of biological significance: the statistical model emphasised disease specificity, consistency, and regulatory or sponge capacity; the unsupervised methods predicted circRNAs central to biologically meaningful clusters or exhibiting high-impact outlier profiles; and the ML prioritization score captured nonlinear feature interactions predictive of disease relevance (Supplementary Tables 1,2). The convergence of these orthogonal approaches on the same set of circRNAs highlights their multifaceted strengths, including structural importance within regulatory networks, reproducibility across datasets, and biological distinctiveness, while simultaneously reducing false positives and increasing confidence in their potential therapeutic relevance. CircRNAs that were consistently prioritized across all three analytical frameworks were designated as consensus biomarker candidates. To ensure biological relevance and avoid overreliance on computational predictions, the outputs generated at each stage were further subjected to manual evaluation and biological interpretation. This integrative multi-tiered strategy improved prioritization robustness, reduced the likelihood of false-positive candidate selection, enhanced cross-method reproducibility, and provided a more biologically meaningful framework for circRNA biomarker discovery than conventional single-parameter screening approaches.
Among the upregulated candidates, hsa_circ_0085740 derived from PTK2 (focal adhesion kinase, LFC: 10.6) demonstrated exceptional performance, ranking within the top four in both clustering and statistical analyses. hsa_circ_0060931 from CYP24A1, a key enzyme involved in vitamin D metabolism, exhibited the highest fold change, while hsa_circ_0052195 from PTPRH (protein tyrosine phosphatase receptor, LFC: 8.97) consistently maintained top-10 rankings across all analytical methods. Other notable upregulated candidates included hsa_circ_0058736 from C2orf82,

hsa_circ_0024109 from MMP1 (a matrix metalloproteinase involved in extracellular matrix remodelling), and hsa_circ_0043632 from KRT17 (keratin 17, a cytoskeletal protein). Among the downregulated circRNAs, hsa_circ_0058622 derived from the SP100 gene showed exceptional consensus, ranking first across all three methods, followed by hsa_circ_0012248 from NASP and hsa_circ_0037308 from TELO2, both of which consistently ranked within the top five. Notably, hsa_circ_0016392 derived from PPP2R5A, a regulatory subunit of protein phosphatase 2A, displayed the strongest downregulation (LFC: -13.21) with robust statistical support. Additional downregulated candidates of interest included hsa_circ_0048053 from SHC2 (an adaptor protein involved in signalling pathways), hsa_circ_0046727 from SMCHD1 (a chromatin modifier), and hsa_circ_0003663 from TOP3A (DNA topoisomerase), all showing coordinated top-10 rankings.

The top 34 dysregulated circRNAs (17 upregulated and 17 downregulated) (Table 1) were subsequently selected as the final high-confidence biomarker panel for downstream functional evaluation, including GO analysis and pathway enrichment studies, to investigate their roles in lung cancer. Collectively, this integrative workflow suggests that combining multi-algorithm consensus with biological pathway screening can strengthen biomarker discovery and prioritisation.

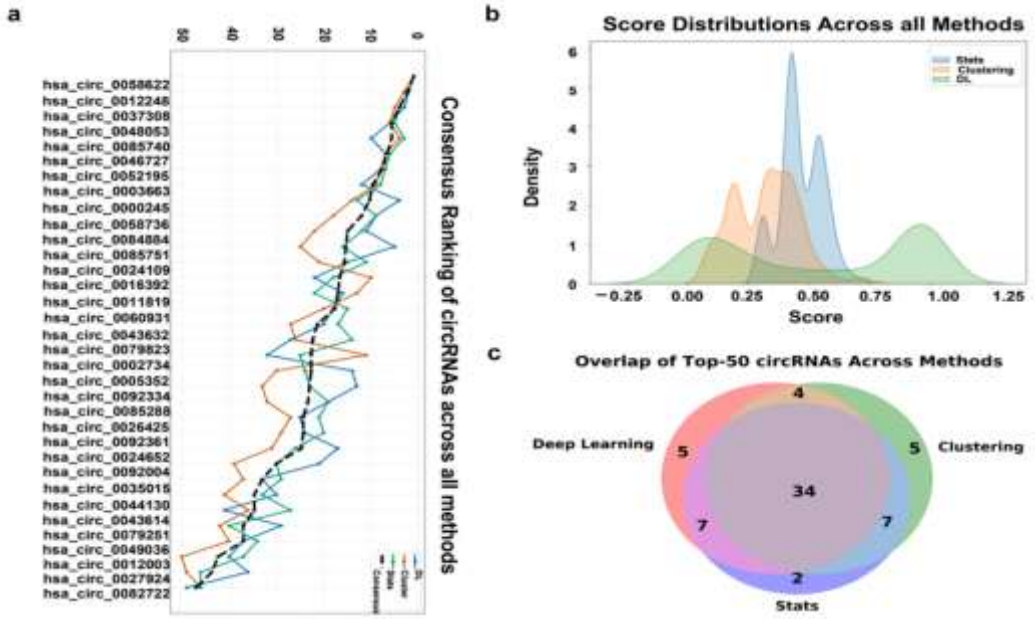


Figure 4 Overview of circRNA prioritisation results across multiple analytical methods. (a) Consensus ranking of circRNAs across analytical methods with the plot depicting top circRNAs with similar consensus rankings derived from MLP, cluster, and statistical (stats) algorithms, indicating strong agreement among methods in identifying consistently significant circRNA candidates; (b) Venn diagram illustrating the overlapping top circRNAs predicted via all three algorithms; (c) Density plots showing the distribution of circRNA scores obtained from each algorithm, illustrating comparable scoring trends and overlapping regions that reflect methodological consistency in circRNA prioritisation.

Table 1. Consensus prioritization of high-confidence circRNA candidates using three complementary computational scoring frameworks.

circRNA	Gene name	Expression	LFC value	ML score	ML rank	Cluster score	Cluster rank	Stats score	Stats rank
hsa_circ_0058622	SP100	Downregulated	-12.17	0.99663204	1	0.757	1	0.740	1
hsa_circ_0012248	NASP	Downregulated	-11.18	0.99612737	2	0.744	3	0.727	2
hsa_circ_0037308	TELO2	Downregulated	-9.94	0.9952877	3	0.728	5	0.71	4
hsa_circ_0000245	None	Downregulated	-2.56	0.9949515	4	0.608	14	0.650	13
hsa_circ_0085751	PTK2	Upregulated	4.94	0.99449205	5	0.558	25	0.643	16
hsa_circ_0048053	SHC2	Downregulated	-9.48	0.99448025	6	0.716	6	0.700	5
hsa_circ_0046727	SMCHD1	Downregulated	-9.41	0.99440217	7	0.713	7	0.696	6
hsa_circ_0052195	PTPRH	Upregulated	8.97	0.9940776	8	0.707	8	0.695	7
hsa_circ_0058736	C2orf82	Upregulated	8.27	0.99389577	9	0.587	18	0.671	9
hsa_circ_0085740	PTK2	Upregulated	10.6	0.99388254	10	0.733	4	0.716	3
hsa_circ_0084884	TMEM67	Downregulated	-5.89	0.99361026	11	0.572	22	0.655	12
hsa_circ_0003663	TOP3A	Downregulated	-8.2	0.9934958	12	0.696	9	0.679	8
hsa_circ_0092334	PRELID2	Upregulated	5.03	0.9927882	13	0.537	33	0.630	23

hsa_circ_0005352	CDCA2	Upregulated	5.34	0.9924915	14	0.539	30	0.629	24
hsa_circ_0024109	MMP1	Upregulated	6.72	0.99242914	15	0.575	21	0.659	11
hsa_circ_0011819	MACF1	Downregulated	-3.45	0.9921859	16	0.636	13	0.631	22
hsa_circ_0024652	GRIK4	Downregulated	-4	0.9918901	17	0.538	31	0.624	26
hsa_circ_0060931	CYP24A1	Upregulated	28.43	0.9917231	18	0.579	20	0.645	15
hsa_circ_0085288	CTHRC1	Upregulated	4.46	0.9914436	19	0.538	32	0.636	19
hsa_circ_0043632	KRT17	Upregulated	5.21	0.99133587	20	0.553	27	0.643	17
hsa_circ_0092004	FLNA	Downregulated	-3.98	0.99129635	21	0.517	39	0.612	30
hsa_circ_0016392	PPP2R5A	Downregulated	-13.21	0.991119	22	0.690	10	0.639	18
hsa_circ_0092361	VIPR1	Downregulated	-4.84	0.99085486	24	0.545	29	0.635	20
hsa_circ_0026425	KRT6A	Upregulated	4.97	0.99053043	25	0.548	27	0.634	21
hsa_circ_0079823	BMPER	Downregulated	-4.75	0.9904053	27	0.553	26	0.647	14
hsa_circ_0079251	SLC29A4	Upregulated	4.12	0.9901022	29	0.503	42	0.602	40
hsa_circ_0044130	KIF18B	Upregulated	4.07	0.9897237	30	0.508	41	0.606	33
hsa_circ_0002734	SEMA5A	Downregulated	-2.94	0.98954356	32	0.669	11	0.627	25
hsa_circ_0035015	CKMT1B	Upregulated	4.33	0.9894872	33	0.525	37	0.618	29
hsa_circ_0027924	C12orf48	Upregulated	3.66	0.98910236	36	0.499	49	0.599	49
hsa_circ_0049036	FBN3	Downregulated	-3.65	0.98865277	37	0.5117	40	0.605	34
hsa_circ_0012003	SLC2A1	Upregulated	3.35	0.9880233	40	0.497	50	0.603	37
hsa_circ_0043614	KRT14	Upregulated	4.01	0.98796755	41	0.526	36	0.619	27
hsa_circ_0082722	DENND2A	Downregulated	-2.81	0.9872055	49	0.501	46	0.598	46
ML score: Machine learning-based score; Cluster score: clustering-based score; Stats scores: statistically calculated scores. All three represent independent ranking methods; LFC shows expression fold change.									

Several prioritised circRNAs originated from genes previously implicated in lung cancer or other malignancies. For example, circRNAs derived from PTK2, MMP1, KRT17, CYP24A1, and PPP2R5A are associated with tumour invasion, extracellular matrix remodelling, epithelial plasticity, or oncogenic signalling in previous studies. However, to the best of our knowledge, these circRNAs have not previously been prioritised within an integrated framework combining differential expression analysis, regulatory network reconstruction, RNA-binding protein interactions, multi-criteria computational prioritisation, and circRNA-derived peptide prediction.

3.5. Validation of prioritised host genes

To independently validate the computationally prioritised circRNA candidates, the expression profiles and prognostic significance of their corresponding host genes (PTK2, MMP1, KRT17, CYP24A1, PPP2R5A, SP100, NASP, SMCHD1, SHC2, and TELO2) were assessed using TCGA-LUAD datasets through GEPIA2, while their diagnostic performance was evaluated by ROC curve analysis. Consistent with the discovery analysis, several host genes exhibited differential expression between LUAD and normal lung tissues in GEPIA2 (Supplementary Figure 1a). Kaplan–Meier survival analysis revealed that high KRT17 expression was significantly associated with poorer overall survival in LUAD patients (log-rank $p = 0.0031$; HR = 1.6), while no significant survival differences were observed for the remaining host genes ($p > 0.05$) (Supplementary Figure 2b). ROC analysis further demonstrated varying diagnostic performance among the selected genes, with CYP24A1 (AUC = 0.913), MMP1 (AUC = 0.893), PPP2R5A (AUC = 0.863), and TELO2 (AUC = 0.842) exhibiting excellent to very good discriminatory ability, whereas NASP (AUC = 0.796) and SP100 (AUC = 0.772) showed moderate diagnostic accuracy. In contrast, PTK2 (AUC = 0.667), KRT17 (AUC = 0.609), SHC2 (AUC = 0.504), and SMCHD1 (AUC = 0.485) displayed comparatively lower diagnostic performance (Supplementary Figure 2). Collectively, these independent analyses support the diagnostic potential of several prioritised host genes while identifying KRT17 as the only candidate with significant prognostic relevance.

3.6. Target gene set enrichment analysis

The target genes of the circRNA-miRNA regulatory axis were subjected to GO analysis to characterise the biological processes underlying the mechanisms of lung cancer progression and to provide insights into disease pathogenesis. Gene functions were systematically categorised into the three major GO domains: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). To capture directional biological changes, GO enrichment analyses were performed separately for the upregulated and downregulated gene sets, enabling the identification of distinct functional regulations associated with each expression pattern. In parallel, pathway enrichment analysis was conducted using the KEGG database. This integrated GO-KEGG framework enabled a comprehensive investigation of the functional landscape, facilitating the identification of key pathways, cellular modules, and molecular mechanisms potentially driving disease progression.

GO enrichment analysis of genes associated with upregulated circRNAs revealed strong activation of pathways linked to transcriptional activity, epithelial remodelling, and membrane-associated processes (Supplementary Figure 3). Elevated transcriptional activity was reflected by the enrichment of key BP terms, including DNA-templated transcription, RNA biosynthetic processes, and RNA polymerase II-mediated transcription. Enrichment of epithelium development, positive regulation of epithelial cell proliferation, and cell-cell adhesion suggested enhanced proliferative and structural dynamics within epithelial cells. Additionally, terms such as neuron projection development and projection morphogenesis suggested cytoskeletal reorganisation and increased cellular plasticity. The CC terms associated with upregulated genes mapped strongly to the recycling endosome, exocytic vesicle membrane, synaptic vesicle membrane, and cortical endoplasmic reticulum, suggesting increased vesicle trafficking and membrane turnover. The presence of motile cilium and peroxisome further highlighted the involvement of signalling and metabolic microdomains. MF analysis revealed enrichment in transporter activities, including amino acid transporters and metal cation antiporters, along with regulatory functions such as kinase activator activity, PDZ domain binding, and actin binding. The simultaneous activation of proliferative and transport-related pathways, together with cytoskeletal reorganisation, suggests that upregulated circRNA networks may drive growth-oriented and invasive cellular programmes in LUAD.

GO analysis of downregulated circRNA-associated genes showed a contrasting enrichment pattern dominated by processes such as regulation of DNA-templated transcription, gene expression regulation, RNA polymerase II regulation, regulation of apoptotic processes, and intracellular signalling modulation, indicating attenuation of multiple regulatory checkpoints governing transcription and cell death (Supplementary Figure 4). Suppression of pathways controlling cell population proliferation and intracellular signal transduction suggested reduced regulatory constraints on growth and signalling fidelity. The major enriched CC terms included focal adhesion, cell-substrate junction, ER membrane, mitochondrial membrane, and nuclear lumen, highlighting downregulation within adhesion structures, energy-regulating organelles, and nuclear regulatory compartments. In the MF domain, key downregulated terms included DNA-binding transcription activator activity, cis-regulatory region binding, kinase binding, serine/threonine kinase activity, and GTPase regulator activity, suggesting decreased regulation of kinase pathways and transcriptional control mechanisms.

3.7. Enrichment of significant pathways

A comparative KEGG pathway analysis of target genes regulated by dysregulated circRNA-miRNA interactions revealed a clear functional division between upregulated and downregulated circRNA-miRNA networks, visualised in Figure 5, suggesting that they coordinate distinct yet complementary activities contributing to cancer progression. Upregulated circRNAs showed predominant associations with several disease- and metabolism-related pathways. Enriched pathways included Herpes simplex virus 1 infection, pathogenic E. coli infection, and Shigellosis, indicating activation of infection-responsive signalling modules. Additional enrichment was observed in pathways related to fluid shear stress and atherosclerosis, cellular senescence, gastric cancer, breast cancer, and the TGF-β signalling pathway, suggesting involvement in stress responses and cancer-associated signalling cascades. Metabolic pathways such as glycolysis/gluconeogenesis, RNA degradation, and amino acid metabolism, including glycine-serine-threonine metabolism and phenylalanine metabolism, were also significantly overrepresented. Pathways related to cytoskeletal regulation, particularly regulation of the actin cytoskeleton, were similarly enriched. Collectively, these findings suggest that genes associated with upregulated circRNAs may support LUAD progression through enhanced metabolic activity, stress adaptation, and oncogenic signalling programmes.

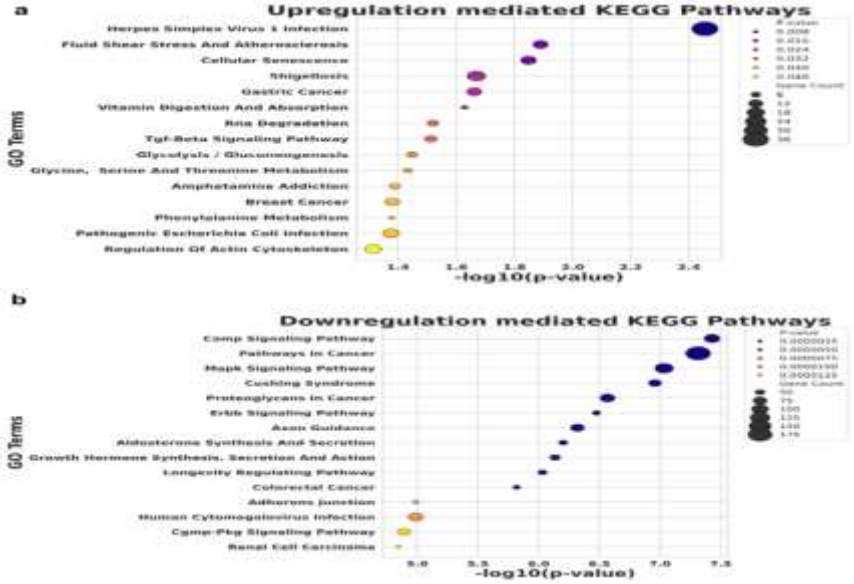


Figure 5 Pathway enrichment of target genes of dysregulated circRNAs. (a) Bubble plot illustrating the enriched pathways targeted by upregulated circRNAs; (b) Bubble plot illustrating the enriched pathways targeted by downregulated circRNAs. The plot provides insights into the potential signalling pathways influenced by dysregulated circRNAs in each case, contributing to our understanding of their roles in lung cancer development and progression.

In contrast, downregulated circRNAs showed enrichment in several classical oncogenic signalling cascades encompassing signalling, cancer-associated, and hormonal regulation pathways. Enriched pathways included the cAMP signalling pathway, cGMP-PKG signalling pathway, and MAPK signalling pathway, indicating reduced activity across major

intracellular signalling networks. Several cancer-related pathways were also predicted, including pathways in cancer, proteoglycans in cancer, colorectal cancer, and renal cell carcinoma. Additional downregulated pathways included ErbB signalling, adherens junction, and axon guidance, reflecting impaired signalling coordination and structural regulation. Hormonal and endocrine-associated pathways, such as aldosterone synthesis and secretion, growth hormone synthesis, secretion and action, and the longevity regulating pathway, were also enriched. Viral response pathways, including human cytomegalovirus infection, were similarly represented. Collectively, the downregulated pathways suggest diminished signalling fidelity, compromised adhesion and structural integrity, and suppression of regulatory cancer-associated modules, which may together facilitate LUAD progression by weakening cellular control systems that normally restrict malignant behaviour.

3.8. Characterization of predicted circRNA-encoded short peptides

The circRNAs have conventionally been classified as a functionally distinct subclass of non-coding RNAs. However, several evidences from multiple independent studies have demonstrated their ability to harbour small open reading frames (smORFs) capable of ribosomal translation, thereby generating short but biologically active peptides. These findings place circRNAs at a previously underexplored interface between non-coding RNA biology and the expanding landscape of the cancer proteome. In the present study, a curated set of four candidate peptides was computationally predicted from the dysregulated circRNA repertoire predicted in LUAD using circPrimer and CPC2 as the primary analytical frameworks for systematic evaluation of coding potential within circular RNA transcripts. The parental circRNA loci from which these peptide-coding sequences were derived showed significant and directionally distinct transcriptional dysregulation in LUAD, with two circRNAs markedly upregulated and the remaining two consistently downregulated. The physicochemical properties of all four computationally predicted peptides, including ORF lengths, molecular weights, isoelectric points, instability indices, aliphatic indices, GRAVY scores, and primary amino acid sequences, are systematically summarised in Table 2 for comparative evaluation.

The predicted physicochemical properties of all four circRNA-encoded peptides are presented in Table 2; the following discussion focuses on the structural and functional implications of these parameters in the context of LUAD-associated dysregulation.

Table 2. The predicted peptide-coding potential of selected dysregulated circRNAs predicted in LUAD.

CircRNA name	Peptide name	ORF Length	Peptide Length	Mol. Wt. (Da)	Kozak Score	pI	Instability Index	Aliphatic Index	GRAVY score	Peptide Sequence
hsa_circ_0054839 (Upregulated)	Pep_0054839_47	144	47	5366.19	0.92	8.27	26.61	109.79	0.274	MNYLIVDNLLVHSITTTTTTITTTTMGIYASSRIRSYDRVQLVTWLT
hsa_circ_0079873 (Upregulated)	Pep_0079873_37	114	37	4234.23	0.88	5.94	34.72	157.84	1.06	MLVVLVPGIEDGVFFLETVYLIGLIQMMRNARKLLLQ
hsa_circ_0002734 (Downregulated)	Pep_0002734_80	147	80	8711.19	0.88	10.04	18.14	91.38	0.096	MKGTCVIAWLFSSLGLWRLAHPEAQGTTQCQRTEHPVISYKALVNHLSRKPFGTSLVGKQVVGSQPSASTVALSLNFKIM
hsa_circ_0024805 (Downregulated)	Pep_0024805_18	57	18	2010.34	1	6.52	26.57	97.78	-0.53	MAKTLEGEKEHGRLALNI

a) Pep_0054839_47 encoded from hsa_circ_0054839 (Upregulated)

b) Physico-chemical profile: Pep_0054839_47 comprises 47 residues (5,366.19 Da) with a near-neutral pI. The peptide is predicted to be highly stable, and its aliphatic index is the highest among downregulated and comparably high overall

reflects substantial thermostability attributable to a high proportion of hydrophobic aliphatic side chains (Leu, Ile, Val, Ala). The positive GRAVY score corroborates an overall hydrophobic character, and inspection of the sequence indicates a striking central threonine-rich motif, which may confer O-glycosylation sites or serve as a flexible polar linker between hydrophobic segments.

c) Structural stability: Pep_0054839_47 displays the most rigid and compact dynamics of all peptides studied. Its backbone RMSD is remarkably low and the RMSF is similarly minimal, indicating highly constrained, near-crystalline backbone dynamics during the simulation (Figure 6a and Figure 6b). Paradoxically, its Rg is the highest of the cohort, implying that while the peptide is internally rigid, it adopts an extended, non-globular conformation (Figure 6c). The SASA of 41.99 nm² is intermediate, suggesting a partial hydrophobic core with extensive peripheral solvent exposure through the threonine-rich segment (Figure 6d).

d) FEL analysis: The FEL of Pep_0054839_47 is the most energetically constrained in this study, characterized by a single, narrow, deep minimum (Figure 7a). This topology reflects exceptional conformational rigidity and a well-defined preferred structure, consistent with the lowest MD fluctuation metrics. Structurally, this peptide behaves as a pre-organized binder: its narrow energy well implies that it incurs minimal entropic cost upon engaging a binding partner, a thermodynamic property associated with high-affinity interactions. The upregulation of hsa_circ_0054839 in LUAD, combined with this peptide's rigid hydrophobic scaffold and translational robustness (Kozak 0.92), suggests that Pep_0054839_47 may function as a constitutively active oncogenic effector capable of stable, high-affinity engagement of pro-tumorigenic targets.

3.8.1. Pep_0079873_37 encoded from hsa_circ_0079873 (Upregulated)

a) Physico-chemical profile: Pep_0079873_37 is a 37-residue peptide (4,234.23 Da) with a slightly acidic pI of 5.94. It has a moderately elevated instability index which is still below the threshold for instability (40), classifying it as a stable molecule with moderate susceptibility to post-translational modification or protease cleavage. Its aliphatic index of is by far the highest in the cohort, indicating exceptional thermostability driven by a high density of aliphatic hydrophobic residues, a property particularly characteristic of membrane-interacting or transmembrane-spanning peptides. This is powerfully corroborated by the strongly positive GRAVY score, the highest among all, which marks Pep_0079873_37 as the most hydrophobic peptide in the peptide set. Indeed, its sequence inspection indicates a canonical hydrophobic core flanked by polar termini, structurally reminiscent of a signal anchor or single-pass transmembrane domain.

b) Structural stability: Despite its markedly hydrophobic nature, Pep_0079873_37 exhibits moderate backbone dynamics. The RMSD nm and RMSF are intermediate values that suggest a peptide whose hydrophobic core constrains internal flexibility while the termini retain mobility (Figure 6a and Figure 6b). The Rg of is consistent with a compact, globular-to-cylindrical conformation expected of a hydrophobic helix (Figure 6c). The SASA of is also intermediate, which (given the highly positive GRAVY score) may indicate that the peptide partially buries its hydrophobic surface in simulations conducted in aqueous medium, as would be expected for a membrane-targeting sequence simulated outside its native lipid environment (Figure 6d).

c) FEL analysis: The FEL of Pep_0079873_37 indicates a bistable conformational landscape two discrete thermodynamic basins (Figure 7b). This bistability is mechanistically significant: it implies that the peptide transitions between two structurally distinct states, a property consistent with conformational switching behavior. For a hydrophobic, membrane-targeting peptide, this dual stability may reflect alternation between an aqueous-exposed disordered state and a membrane-inserted helical state; a characteristic topology of cell-penetrating and fusogenic peptides. The upregulation of hsa_circ_0079873 in LUAD, encoding a thermostable, highly hydrophobic bistable peptide, raises the compelling hypothesis that Pep_0079873_37 may participate in membrane remodeling, mitochondrial outer membrane interaction, or organelle-targeting activities that favor tumor cell survival.

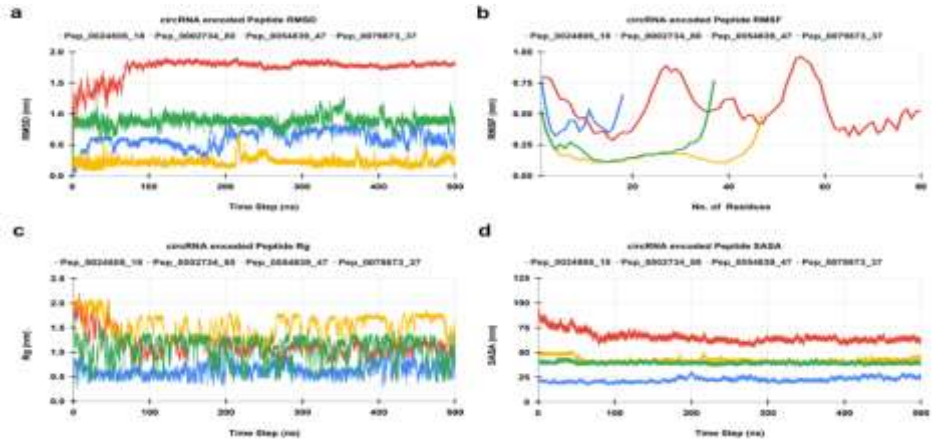


Figure 6 Graphical illustration of the MD simulation analysis of all four circRNA-encoded peptides (Pep_0054839_47, Pep_0079873_37, Pep_0002734_80, and Pep_0024805_18 ; all distinguished by colour scheme) over a 500 ns simulation, including (a) RMSD graph, (b) RMSF per residue, (c) Rg, (d) and SASA

3.8.2. Pep_0002734_80 encoded from hsa_circ_0002734 (Downregulated)

a) Physico-chemical profile: Pep_0002734_80 is the largest peptide in this cohort (80 residues, 8,711.19 Da) and the only candidate with a basic isoelectric point (pI = 10.04), suggesting a net positive charge under physiological conditions. This charge state may facilitate electrostatic interactions with negatively charged cellular components, including nucleic

acids or membrane phospholipids. It shows the lowest instability index of all peptides studied, which classifies Pep_0002734_80 as highly stable in solution, while its aliphatic index predicts reasonable thermostability. The near-neutral GRAVY score denotes a mildly hydrophobic character that may support partial membrane association or hydrophobic core formation.

b) Structural stability: The dynamics of Pep_0002734_80 reflect its larger sequence length and structural complexity. The mean RMSD of 1.739 nm is the highest observed in this study, indicative of substantial conformational transitions during the simulation trajectory. However, this should be interpreted in context: for an 80-residue peptide with a highly stable predicted physico-chemical profile (instability index 18.14), large RMSD values may reflect slow, large-scale domain motions rather than disordered sampling. The mean RMSF of 0.554, while the highest in absolute terms, is relatively modest on a per-residue basis, implying that individual residues are comparatively constrained even as the global conformation evolves mobility (Figure 6a and Figure 6b). The Rg and SASA are consistent with an extended, partially structured peptide that maintains significant solvent exposure (Figure 6c and Figure 6d).

c) FEL analysis: The FEL of Pep_0002734_80 is characterized by a single, broad conformational basin with the widest conformational spread of the series (Figure 7c). This profile is consistent with a folding funnel topology, wherein the peptide samples a large configurational space but is directed toward a single thermodynamic attractor. Broad funnels of this type are frequently observed for peptides undergoing coupled folding-and-binding, a mechanism whereby the peptide adopts its final structure upon encountering its physiological partner. The downregulation of hsa_circ_0002734 in LUAD therefore raises the possibility that a structurally adaptive peptide effector potentially involved in chromatin or RNA-binding is lost in the tumorigenic context.

3.8.3. Pep_0024805_18 encoded from hsa_circ_0024805 (Downregulated)

a) Physico-chemical profile: Pep_0024805_18 is the smallest peptide characterized in this study, comprising 18 residues with a molecular weight of 2,010.34 Da. Its predicted isoelectric point (pI = 6.52) situates it near physiological neutrality, suggesting compatibility with intracellular environments without strong electrostatic precipitation. The low instability index classifies the peptide as stable in vitro, while the aliphatic index implies moderate thermostability relative to its diminutive size. Critically, the negative GRAVY score denotes an overall hydrophilic character, consistent with a surface-accessible or solvent-exposed functional conformation. The Kozak score of absolute 1.0, the highest among all four candidates provides the strongest translational initiation signal, indicating robust ribosomal recognition of the upstream ORF within hsa_circ_0024805.

b) Structural stability: Despite its small size, Pep_0024805_18 exhibits notable structural fluctuation characteristics. The mean RMSD suggests a moderately dynamic peptide that does not converge to a single rigid conformation during the simulation. The per-residue RMSF indicates considerable local flexibility, which is proportionally elevated relative to its sequence length and consistent with the absence of extensive hydrophobic burial mobility (Figure 6a and Figure 6b). The radius of gyration is the lowest across all peptides, as expected given its 18-residue span, and the SASA reflects broad solvent exposure that supports its hydrophilic GRAVY profile (Figure 6c and Figure 6d).

c) FEL analysis: The free energy landscape of Pep_0024805_18 indicates a multi-stable conformational ensemble, with three thermodynamically accessible basins (Figure 7d). This multi-basin topology is a hallmark of intrinsically disordered or partially disordered peptides and suggests that the downregulation of hsa_circ_0024805 in LUAD may compromise the availability of a dynamically versatile signalling effector. The peptide's capacity to interconvert between multiple metastable states could underpin promiscuous binding interactions.

Taken together, the four circRNA-encoded peptides described above present a remarkably heterogeneous structural repertoire, which maps non-randomly onto their expression patterns in LUAD. The two downregulated peptides: Pep_0024805_18 and Pep_0002734_80; exhibit multi-basin and broad-funnel FEL topologies respectively, suggesting conformationally adaptive molecules whose loss may disrupt flexible signalling or coupled folding-and-binding processes that are protective against tumorigenic transformation. Conversely, the two upregulated peptides: Pep_0054839_47 and Pep_0079873_37, display energetically constrained and bistable FEL profiles respectively, implying high-affinity constitutive effectors or membrane-active switches whose overproduction may actively drive oncogenic phenotypes. The divergence in GRAVY scores (−0.53 to +1.06) and aliphatic indices (91 to 158) further underscores that these circRNA-derived microproteins are not a homogeneous functional class, but rather a diverse set of regulatory entities acting across distinct cellular compartments. This analysis provides a structural mechanistic rationale for functional characterization and warrants experimental validation through pull-down proteomics, liposome association assays, and cancer cell line perturbation studies.

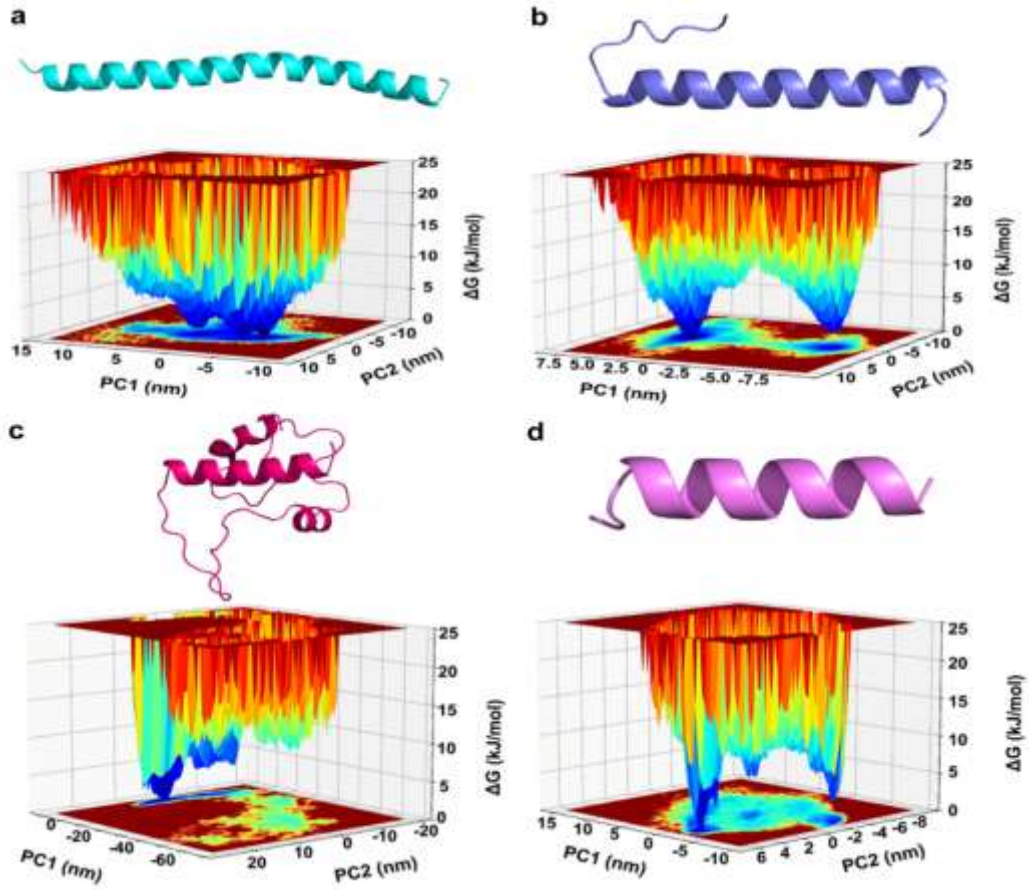


Figure 7 Stable structures and free energy landscapes of circRNA-encoded peptides: (a) Pep_0054839_47, (b) Pep_0079873_37, (c) Pep_0002734_80, and (d) Pep_0024805_18

4. DISCUSSION

This study maps the circRNA regulatory landscape in LUAD and shows how deeply these molecules are linked with tumour development and progression. We combined differential expression profiling, circRNA-miRNA-mRNA network reconstruction, RBP interaction prediction, functional enrichment analysis, and a three-layer computational prioritization workflow involving statistical scoring, unsupervised clustering, and machine learning. Using this integrated approach, we predicted 34 high-confidence circRNA candidates that were consistently prioritised across all analytical frameworks (Table 1). And this is significant because no single computational approach can fully capture the complex biological characteristics of circRNAs. Statistical scoring provides an interpretable measure of feature importance, unsupervised clustering identifies intrinsic similarities among candidates without requiring labelled data, and ML
-based feature integration captures nonlinear relationships among multiple biological attributes. Agreement across these independent approaches therefore increases confidence in the prioritization of candidate biomarkers while reducing dependence on any single analytical strategy. Most previous studies investigated circRNAs using differential expression and ceRNA analyses alone. In contrast, our workflow combines transcriptomic profiling with RNA-binding protein interaction analysis, multi-criteria computational prioritisation, prediction of circRNA-derived peptides, and molecular dynamics simulations, enabling functional prioritisation of candidate circRNAs beyond expression differences alone.

To strengthen the biological and clinical relevance of the computationally prioritised circRNA candidates, their corresponding host genes were independently evaluated using TCGA-derived datasets. The concordance between the discovery analysis, GEPIA2 expression validation, and ROC analysis provides additional support for several candidate genes, particularly CYP24A1, MMP1, PPP2R5A, and TELO2, which demonstrated good to excellent diagnostic performance. Although most host genes were not significantly associated with overall survival, KRT17 emerged as an exception, where elevated expression correlated with poorer patient survival, further supporting its potential role in LUAD progression. The absence of significant prognostic associations for the remaining genes does not diminish their potential diagnostic utility, as biomarkers involved in disease detection and those associated with patient prognosis often represent distinct biological processes. Overall, these independent validation analyses reinforce the robustness of the computational prioritisation framework and provide additional evidence that several of the identified circRNA-associated host genes warrant further functional and experimental investigation.

Global circRNA profiling revealed a structurally diverse circular transcriptome dominated by exon-derived circRNAs, which matches the general biogenesis patterns reported in several cancers. Most dysregulated circRNAs were downregulated. This points towards widespread loss of epithelial identity and transcriptional suppression within tumour tissues. But the upregulated subset appeared more closely linked with proliferative and oncogenic functions. Several major candidates originated from genes associated with focal adhesion signalling, extracellular matrix remodelling, cytoskeletal organisation, metabolism, and chromatin regulation. Our findings are also consistent with several previously reported cancer-associated circRNAs, while simultaneously extending their potential functional significance through an integrated

computational framework. CircRNAs originating from PTK2, MMP1, KRT17, and CYP24A1 have each been implicated in tumour progression, invasion, epithelial plasticity, or dysregulated signalling across lung and other malignancies (Dai et al., 2021; Ji et al., 2022; Wang et al., 2025; Wu et al., 2022; Yang et al., 2020). The recurrence of these cancer-associated loci in our analysis provides independent support for the robustness of the computational workflow, suggesting that the prioritisation strategy successfully recovers biologically relevant candidates rather than producing arbitrary predictions. However, previous investigations have largely focused on differential expression or individual circRNA-mediated ceRNA mechanisms. In contrast, the present study integrates transcriptomic profiling with ceRNA network reconstruction, RNA-binding protein interaction mapping, consensus-based computational prioritisation, and evaluation of circRNA coding potential, thereby providing a more comprehensive assessment of the possible functional roles of these molecules in LUAD. Furthermore, several highly prioritised circRNAs identified in this study have received little or no previous attention in lung adenocarcinoma, indicating that the proposed framework is capable of highlighting both well-established cancer-associated candidates and relatively unexplored circRNAs that warrant future experimental investigation.

The circRNA-miRNA interaction networks showed two very different regulatory patterns. Upregulated circRNAs formed a compact but highly focussed network where a small number of molecules, particularly hsa_circ_0035015, hsa_circ_0043614, and hsa_circ_0043632, exerted strong regulatory influence over multiple miRNAs. In contrast, downregulated circRNAs formed a broader and more complex network centred on candidates such as hsa_circ_0049036 and hsa_circ_0046727. This pattern hints at widespread post-transcriptional disruption during LUAD progression.

We also analysed circRNA-RBP interactions. Here, upregulated circRNAs displayed much stronger RBP connectivity than the downregulated group. EIF4A3 emerged as the most extensively interacting RBP and bound 22 circRNAs across both expression classes. AGO2 and HuR also appeared repeatedly within the network. Interestingly, IGF2BP family proteins preferentially interacted with upregulated circRNAs, which fits well with their known roles in stabilising oncogenic transcripts. But METTL3 and DGCR8 selectively associated with downregulated circRNAs. These expression-dependent interaction patterns suggest that altered circRNA-RBP binding may actively rewire post-transcriptional regulatory systems controlling proliferation, survival, and therapy resistance.

Functional enrichment analysis supported these observations quite clearly. Genes linked with upregulated circRNAs showed enrichment for transcriptional activation, epithelial remodelling, vesicle trafficking, cytoskeletal reorganisation, and intracellular transport. Together, these pathways reflect metabolically adaptive and invasive tumour behaviour. On the other hand, genes associated with downregulated circRNAs showed reduced transcriptional control, weakened apoptotic regulation, disrupted adhesion structures, and impaired signalling fidelity. KEGG analysis showed the same directional trend. Upregulated networks were enriched for TGF-β signalling, glycolysis, cytoskeletal regulation, and cancer-associated pathways, whereas downregulated networks displayed suppression of MAPK, ErbB, cAMP, and cGMP-PKG signalling pathways along with reduced adherens junction integrity.

Moving on to the translational aspect, we explored whether selected dysregulated circRNAs could encode functional peptides. This area is still relatively underexplored in lung cancer biology. Using circPrimer and CPC2, we predicted four candidate peptides: Pep_0054839_47, Pep_0079873_37, Pep_0002734_80, and Pep_0024805_18. We restricted the screening workflow to circRNAs containing intact BSJ-spanning ORFs with canonical AUG start codons and valid in-frame stop codons. All four candidates showed strong Kozak consensus scores, indicating good translational potential.

Their physicochemical properties varied considerably. Pep_0002734_80 displayed strong basicity, suggesting possible nucleic acid or membrane interaction capacity. The remaining peptides showed near-neutral or mildly acidic profiles, more consistent with soluble regulatory functions. All four peptides remained stable according to instability index calculations. And the hydrophobicity analysis was particularly interesting. Pep_0079873_37 showed strong hydrophobic character and a high aliphatic index, indicating possible membrane-associated behaviour. In contrast, Pep_0024805_18 appeared strongly hydrophilic and resembled soluble microregulatory peptides. When viewed alongside parental circRNA expression patterns, peptides derived from upregulated circRNAs appeared more compatible with membrane-proximal oncogenic signalling, whereas peptides from downregulated circRNAs showed features more consistent with tumour-suppressive or cytoplasmic regulatory roles.

We then tested the structural stability of all four peptides using 500 ns all-atom molecular dynamics simulations based on AlphaFold-predicted structures. All peptides retained their predicted folds throughout the simulation trajectories. RMSD values remained low across the entire run, indicating stable structural behaviour. Pep_0024805_18 emerged as the most conformationally rigid peptide and showed the lowest RMSD, RMSF, Rg, and SASA values. Its PCA-derived free-energy landscape also displayed a sharp and highly localised energy minimum: quite remarkable for an 18-residue peptide.

Pep_0054839_47 also maintained strong structural conservation with moderate solvent exposure, which usually supports stable intermolecular interaction capacity. And Pep_0079873_37 retained a coherent hydrophobic core despite slightly increased backbone flexibility. This observation fits nicely with its predicted membrane-associated nature. In contrast, Pep_0002734_80 sampled a broader range of conformational states because of its larger size and mixed helical-loop topology. Such structural flexibility may actually favour interactions with multiple signalling proteins or transcription-associated regulatory partners.

These peptide findings are important for another reason. The candidates originated from both upregulated and downregulated circRNAs, suggesting that peptide-mediated regulation may occur across distinct oncogenic and tumour-suppressive layers within LUAD. Also, circRNA-derived peptides themselves remain poorly characterised in cancer systems. Stable micropeptides like these may eventually serve as useful biomarkers because circRNAs naturally resist exonuclease-mediated degradation. Some may even participate in protein-protein or protein-DNA interactions linked with transcriptional regulation, stress adaptation, or cancer signalling pathways (Fan et al., 2020; Gilyazova et al., 2024; Ju et al., 2021; Li et al., 2025)

But there is an obvious limitation. We relied entirely on computational prediction and in silico analyses. We still need experimental confirmation of peptide expression, localisation, intermolecular interactions, and biological activity in cellular and animal systems.

5. CONCLUSION

This study systematically characterised the circRNA regulatory landscape in LUAD using integrated expression profiling, ceRNA network analysis, RBP interaction mapping, functional enrichment, and multi-layer biomarker prioritisation. We prioritized 34 high-confidence circRNA candidates that consistently remained dysregulated across statistical analysis, unsupervised clustering, and machine learning-assisted prediction models. The associated regulatory networks pointed towards major cancer-related processes, including transcriptional reprogramming, epithelial remodelling, metabolic adaptation, and disruption of signalling homeostasis. Upregulated circRNAs showed stronger connectivity with RBPs and oncogenic pathways, whereas downregulated circRNAs were associated with weakened regulatory and adhesion mechanisms. And this clear directional separation suggests that circRNA dysregulation in LUAD is not random but functionally organised. The study also predicted four circRNA-derived smORF peptides with favourable translational potential and stable structural behaviour during molecular dynamics simulations. These peptides may represent an additional regulatory layer in lung cancer biology and could have future value as stable biomarkers or regulatory microproteins. Finally, the experimental validation of circRNA expression, peptide translation, molecular interactions, and biological function will be necessary before clinical relevance can be firmly established.

6. Acknowledgements

The study was supported and funded by ICMR, New Delhi (grant number ISRM/12(98)/2020, ID. No. 2020-4064). The authors also acknowledge the Bioinformatics Centre at Department of Bioscience and Biotechnology and "Workshop in Bioinformatics" project at Banasthali Vidyapith supported by DBT, Government of India, and the DST-FIST program of Government of India

7. Authors' contributions:

AM: Data Curation, Formal Analysis, Visualization, Writing – Original Draft, Writing – Review & Editing.
RB: Conceptualization, Supervision, Resources, Writing – Review & Editing.

8. Data Availability Statement

All datasets analyzed in this study are publicly available in the repositories and databases cited within the Methodology section. While the technical context is provided via accession numbers in the text, direct URLs to the datasets are provided below to facilitate immediate access
https://www.ebi.ac.uk/ena/browser/view/SRP335547
https://www.ebi.ac.uk/ena/browser/view/SRP089923
https://www.ebi.ac.uk/ena/browser/view/SRP048484
https://www.ebi.ac.uk/ena/browser/view/SRP047399
https://www.circbase.org/
https://www.mirbase.org/

9. Ethics Statement

This study was purely computational in nature and did not involve the collection of new data from human participants or animals. All data used in this study were obtained from publicly available, pre-existing datasets that had already been anonymized prior to access, such that no identifiable personal information was used at any stage of this research. Ethical approval for the original collection of these data had already been obtained by the respective original investigators/data providers, in accordance with the ethical standards applicable at the time and institution of collection. As this study involved secondary analysis of de-identified, publicly accessible data only, and did not entail any new data collection, direct participant contact, or intervention, separate ethical approval was not required for the present work.

REFERENCES

1. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630. https://doi.org/10.1038/s41586-024-07487-w
2. Alimohammadi, M., Gholinezhad, Y., Mousavi, V., Kahkesh, S., Rezaee, M., Yaghoobi, A., Mafi, A., Araghi, M., 2023. CIRCULAR RNAS: NOVEL ACTORS OF WNT SIGNALING PATHWAY IN LUNG CANCER PROGRESSION. EXCLI J. https://doi.org/10.17179/excli2023-6209
3. Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30. https://doi.org/10.1093/bioinformatics/btu170
4. Chen, C.K., Cheng, R., Demeter, J., Chen, J., Weingarten-Gabbay, S., Jiang, L., Snyder, M.P., Weissman, J.S., Segal, E., Jackson, P.K., Chang, H.Y., 2021. Structured elements drive extensive circular RNA translation. Mol. Cell 81. https://doi.org/10.1016/j.molcel.2021.07.042

5. Chen, D., Ma, W., Ke, Z., Xie, F., 2018. CircRNA hsa_circ_100395 regulates miR-1228/TCF21 pathway to inhibit lung cancer progression. Cell Cycle 17. https://doi.org/10.1080/15384101.2018.1515553
6. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14. https://doi.org/10.1186/1471-2105-14-128
7. Chen, L.L., 2016. The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm.2015.32
8. Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., Wong, K.K., 2014. Non-small-cell lung cancers: A heterogeneous set of diseases. Nat. Rev. Cancer. https://doi.org/10.1038/nrc3775
9. Cocquerelle, C., Mascrez, B., Hétuin, D., Bailleul, B., 1993. Mis-splicing yields circular RNA molecules. The FASEB Journal 7. https://doi.org/10.1096/fasebj.7.1.7678559
10. Dai, W., Zhai, X., Chen, Y., Bai, Y., Deng, H., Zhu, R., Fan, W., Cai, S., 2021. CircMMP1 promotes colorectal cancer growth and metastasis by sponging miR-1238 and upregulating MMP family expression. Ann. Transl. Med. 9, 1341–1341. https://doi.org/10.21037/atm-21-3930
11. Dudekula, D.B., Panda, A.C., Grammatikakis, I., De, S., Abdelmohsen, K., Gorospe, M., 2016. Circinteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol. 13. https://doi.org/10.1080/15476286.2015.1128065
12. Fan, Z., Bai, Y., Zhang, Q., Qian, P., 2020. CircRNA circ_POLA2 promotes lung cancer cell stemness via regulating the miR-326/GNB1 axis. Environ. Toxicol. 35. https://doi.org/10.1002/tox.22980
13. Gao, Y., Wang, J., Zhao, F., 2015. CIRI: An efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol. 16. https://doi.org/10.1186/s13059-014-0571-3
14. Gilyazova, I., Gimalova, G., Nizamova, A., Galimova, E., Ishbulatova, E., Pavlov, V., Khusnutdinova, E., 2024. Non-Coding RNAs as Key Regulators in Lung Cancer. Int. J. Mol. Sci. https://doi.org/10.3390/ijms25010560
15. Glažar, P., Papavasileiou, P., Rajewsky, N., 2014. CircBase: A database for circular RNAs. RNA 20. https://doi.org/10.1261/rna.043687.113
16. Harrison, P.W., Ahamed, A., Aslam, R., Alako, B.T.F., Burgin, J., Buso, N., Courtot, M., Fan, J., Gupta, D., Haseeb, M., Holt, S., Ibrahim, T., Ivanov, E., Jayathilaka, S., Kadhirvelu, V.B., Kumar, M., Lopez, R., Kay, S., Leinonen, R., Liu, X., O'Cathail, C., Pakseresht, A., Park, Y., Pesant, S., Rahman, N., Rajan, J., Sokolov, A., Vijayaraja, S., Waheed, Z., Zyoud, A., Burdett, T., Cochrane, G., 2021. The European Nucleotide Archive in 2020. Nucleic Acids Res. 49. https://doi.org/10.1093/nar/gkaa1028
17. Huang, W., Ling, Y., Zhang, S., Xia, Q., Cao, R., Fan, X., Fang, Z., Wang, Z., Zhang, G., 2021. TransCirc: An interactive database for translatable circular RNAs based on multi-omics evidence. Nucleic Acids Res. 49. https://doi.org/10.1093/nar/gkaa823
18. Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., Sharpless, N.E., 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19. https://doi.org/10.1261/rna.035667.112
19. Ji, Y., Zhao, Q., Feng, W., Peng, Y., Hu, B., Chen, Q., 2022. N6-Methyladenosine Modification of CIRCKRT17 Initiated by METTL3 Promotes Osimertinib Resistance of Lung Adenocarcinoma by EIF4A3 to Enhance YAP1 Stability. Cancers (Basel). 14. https://doi.org/10.3390/cancers14225582
20. Ju, Y., Yuan, B., Wu, W., Zhao, J., Shi, X., 2021. CircRNA ANXA2 Promotes Lung Cancer Proliferation and Metastasis by Upregulating PDPK1 Expression. J. Oncol. 2021. https://doi.org/10.1155/2021/4526609
21. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45. https://doi.org/10.1093/nar/gkw1092
22. Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., Gao, G., 2017. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 45. https://doi.org/10.1093/nar/gkx428
23. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, and D., 2002. The Human Genome Browser at UCSC. Genome Res. 12. https://doi.org/10.1101/gr.229102
24. Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019. MiRBase: From microRNA sequences to function. Nucleic Acids Res. 47. https://doi.org/10.1093/nar/gky1141
25. Leinonen, R., Sugawara, H., Shumway, M., 2011. The sequence read archive. Nucleic Acids Res. 39. https://doi.org/10.1093/nar/gkq1019
26. Lemkul, J.A., 2024. Introductory Tutorials for Simulating Protein Dynamics with GROMACS. Journal of Physical Chemistry B. https://doi.org/10.1021/acs.jpcb.4c04901
27. Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25. https://doi.org/10.1093/bioinformatics/btp324
28. Li, H., wu, F., Han, Y., Guo, Z., Chen, T., Ma, Z., 2025. CircRNA regulates lung cancer metastasis. Gene. https://doi.org/10.1016/j.gene.2024.149060
29. Lin, Y., Wang, Y., Li, L., Zhang, K., 2025. Coding circular RNA in human cancer. Genes Dis. https://doi.org/10.1016/j.gendis.2024.101347
30. Liu, M., Wang, Q., Shen, J., Yang, B.B., Ding, X., 2019. Circbank: a comprehensive database for circRNA with standard nomenclature. RNA Biol. 16. https://doi.org/10.1080/15476286.2019.1600395
31. Love, M.I., Huber, W., Anders, S., 2017. Analyzing RNA-seq data with DESeq2. Bioconductor.
32. Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15. https://doi.org/10.1186/s13059-014-0550-8
33. Malviya, A., Bhuyan, R., 2025. Circular RNAs in cancer: roles, mechanisms, and therapeutic potential across colorectal, gastric, liver, and lung carcinomas. Discover Oncology. https://doi.org/10.1007/s12672-025-01743-9

34. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B., Rigoutsos, I., 2006. A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. Cell 126. https://doi.org/10.1016/j.cell.2006.07.031
35. Mokrejs, M., Vopálenský, V., Kolenaty, O., Masek, T., Feketová, Z., Sekyrová, P., Skaloudová, B., Kríz, V., Pospísek, M., 2006. IRESite: the database of experimentally verified IRES structures (www.iresite.org). Nucleic Acids Res. 34. https://doi.org/10.1093/nar/gkj081
36. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33. https://doi.org/10.1038/nbt.3122
37. Riccardi, L., Nguyen, P.H., Stock, G., 2012. Construction of the free energy landscape of peptide aggregation from molecular dynamics simulations. J. Chem. Theory Comput. 8. https://doi.org/10.1021/ct200911w
38. Riffo-Campos, Á.L., Riquelme, I., Brebi-Mieville, P., 2016. Tools for sequence-based miRNA target prediction: What to choose? Int. J. Mol. Sci. https://doi.org/10.3390/ijms17121987
39. Ru, Y., Kechris, K.J., Tabakoff, B., Hoffman, P., Radcliffe, R.A., Bowler, R., Mahaffey, S., Rossi, S., Calin, G.A., Bemis, L., Theodorescu, D., 2014. The multiMiR R package and database: Integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res. 42. https://doi.org/10.1093/nar/gku631
40. Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., Brown, P.O., 2013. Cell-Type Specific Features of Circular RNA Expression. PLoS Genet. 9. https://doi.org/10.1371/journal.pgen.1003777
41. Sarker, I.H., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput. Sci. https://doi.org/10.1007/s42979-021-00592-x
42. Sausville, L.N., Jones, C.C., Aldrich, M.C., Blot, W.J., Pozzi, A., Williams, S.M., 2017. Genetic variation in the eicosanoid pathway is associated with non-small-cell lung cancer (NSCLC) survival. PLoS One 12. https://doi.org/10.1371/journal.pone.0180471
43. Seguin, L., Durandy, M., Feral, C.C., 2022. Lung Adenocarcinoma Tumor Origin: A Guide for Personalized Medicine. Cancers (Basel). https://doi.org/10.3390/cancers14071759
44. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. 13. https://doi.org/10.1101/gr.1239303
45. Thandra, K.C., Barsouk, Adam, Saginala, K., Aluru, J.S., Barsouk, Alexander, 2021. Epidemiology of lung cancer. Wspolczesna Onkologia. https://doi.org/10.5114/wo.2021.103829
46. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: Fast, flexible, and free. J. Comput. Chem. https://doi.org/10.1002/jcc.20291
10. Wang, L., Tong, X., Zhou, Z., Wang, S., Lei, Z., Zhang, T., Liu, Z., Zeng, Y., Li, C., Zhao, J., Su, Z., Zhang, C., Liu, X., Xu, G., Zhang, H.T., 2025. Correction: Circular RNA hsa_circ_0008305 (circPTK2) inhibits TGF-β-induced epithelial-mesenchymal transition and metastasis by controlling TIF1γ in non-small cell lung cancer (Molecular Cancer, (2018), 17, 1, (140), 10.1186/s12943-018-0889-7). Mol. Cancer. https://doi.org/10.1186/s12943-025-02509-1
11. Wen, S. yang, Qadir, J., Yang, B.B., 2022. Circular RNA translation: novel protein isoforms and clinical significance. Trends Mol. Med. https://doi.org/10.1016/j.molmed.2022.03.003
12. Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., Hochstrasser, D.F., 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. https://doi.org/10.1385/1-59259-584-7:531
13. Wu, X., Zhou, J., Zhao, L., Yang, Z., Yang, C., Chen, Y., Xue, W., 2022. CircCYP24A1 hampered malignant phenotype of renal cancer carcinoma through modulating CMTM-4 expression via sponging miR-421. Cell Death Dis. 13. https://doi.org/10.1038/s41419-022-04623-0
14. Yang, H., Li, Xiaobo, Meng, Q., Sun, H., Wu, S., Hu, W., Liu, G., Li, Xianjing, Yang, Y., Chen, R., 2020. CircPTK2 (hsa_circ_0005273) as a novel therapeutic target for metastatic colorectal cancer. Mol. Cancer 19. https://doi.org/10.1186/s12943-020-1139-3
15. Zhong, S., Feng, J., 2022. CircPrimer 2.0: a software for annotating circRNAs and predicting translation potential of circRNAs. BMC Bioinformatics 23. https://doi.org/10.1186/s12859-022-04705-y

Supplementary Table 1. Description of features used for circRNA biomarker prioritization

Feature	Biological Significance	Data Source	Interpretation
Exclusivity Score	Measures the degree to which a circRNA is specific to the cancer samples CircRNAs exhibiting greater tissue specificity are considered more suitable diagnostic biomarkers	Transcript count values, dataset associated studies	Higher values indicate greater cancer specificity
Consistency Score	Evaluates the reproducibility of differential expression patterns across all samples and studies.	GEO datasets, published studies, and integrated expression analyses	Higher values indicate more consistent differential expression across samples
Differential Expression Score (DE_Score)	Represents the magnitude and direction of differential expression between cancer and control samples. This feature captures the extent of transcriptional dysregulation associated with disease	Differential expression analysis using DESeq2	Larger absolute values indicate stronger dysregulation, while the sign denotes

			upregulation (+) or downregulation (−).
Binding Affinity Score	Estimates the strength of predicted circRNA–miRNA interactions, reflecting the likelihood of stable molecular binding.	circRNA–miRNA interaction prediction using miRanda tool	Higher values indicate stronger predicted interaction potential
Regulatory Potential Score	Quantifies the potential biological influence of a circRNA based on its associated targets	circInteractome database	Higher values indicate greater involvement in biologically relevant regulation
Sponge Capacity Score	Quantifies the potential biological influence of a circRNA based on its associated targets, pathways, and network connectivity.	miRNA response element (MRE) prediction and interaction analyses using miRanda	Higher values indicate greater miRNA sponge potential.
The six features explained above collectively capture complementary aspects of circRNA biomarker potential, including disease specificity, reproducibility, expression dysregulation, interaction strength, regulatory influence, and miRNA sponge activity. These features were subsequently integrated into statistical, machine-learning, and deep-learning frameworks for biomarker prioritization			

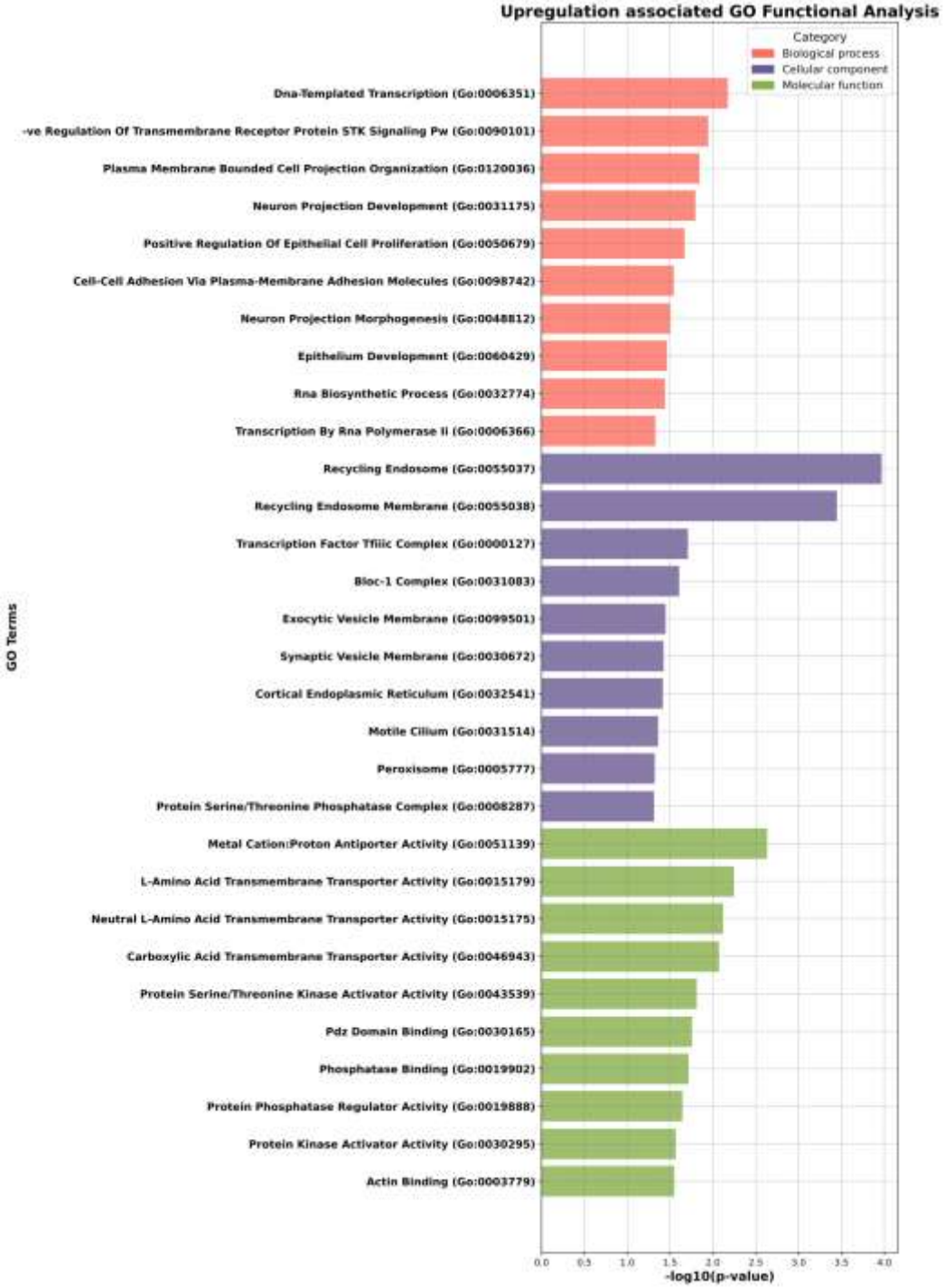
It should be noted that the computational pipeline was not employed as a fully automated decision-making system. At each stage of the analysis, including statistical scoring, clustering-based prioritization, and deep-learning-based ranking, the computational outputs were subjected to manual evaluation and biological interpretation. Candidate circRNAs were examined in the context of existing literature, expression patterns, functional relevance, and biological plausibility to ensure that the final shortlist reflected both computational evidence and domain-specific expertise. This combined strategy minimized the risk of overreliance on any individual algorithm and enhanced the robustness and interpretability of the final biomarker candidates.

The complete feature matrix generated during the biomarker prioritization workflow contained a large number of candidate circRNAs and intermediate feature values. Inclusion of the entire dataset would result in a highly extensive supplementary file that may reduce readability and limit the practical utility of the presented information. Therefore, for clarity and ease of interpretation, Supplementary Table S2 presents a representative subset of the highest-priority circRNA candidates together with their corresponding feature values. These candidates adequately illustrate the feature profiles and scoring characteristics used throughout the prioritization framework.

Supplementary Table 2. Representative entries from the circRNA feature matrix used for biomarker prioritization

circRNA	Exclusivity Score	Consistency Score	DE_Score	Binding Affinity	Regulatory Potential	Sponge Capacity Score
hsa_circ_0016392	0.999999999	1	-0.582351661	0	0.407646163	0
hsa_circ_0058622	0.999999999	1	-0.509434472	0.562791659	0.525441628	0.004770798
hsa_circ_0012248	0.999999998	1	-0.443005448	0.575988118	0.482900249	0.009541596
hsa_circ_0085740	0.999999997	1	0.39227102	0.578060516	0.448007869	0.014312394
hsa_circ_0011179	0.999999996	1	-0.397198312	0	0.278038818	0
hsa_circ_0037308	0.999999995	1	-0.373347056	0.581629099	0.435831669	0.009541596
hsa_circ_0046727	0.999999992	1	-0.311085563	0.57970936	0.391672702	0.009541596
hsa_circ_0003663	0.999999982	1	-0.265987572	0.548239268	0.350663081	0.004770798
hsa_circ_0058736	0.999999981	1	0.201578105	0.582226688	0.31577268	0.009541596
hsa_circ_0084884	0.999999902	1	-0.132770513	0.57997082	0.266930605	0.028624788
chr1:23878112|23878797	0.999999647	1	-0.057950281	0	0.040565197	0
hsa_circ_0052195	0.996037639	1	0.221890135	0.674833188	0.357773051	0.030237691
hsa_circ_0024109	0.97962375	1	0.189109013	0.563591357	0.301453716	0.019083192
hsa_circ_0051240	0.977381614	1	0.119809302	0	0.083866512	0
hsa_circ_0060944	0.971291859	1	0.105876189	0	0.074113332	0
hsa_circ_0043632	0.950054883	1	0.113747645	0.603880366	0.260787462	0.033395586
hsa_circ_0045522	0.946360132	1	0.102942956	0	0.072060069	0
hsa_circ_0005352	0.942857102	1	0.107576047	0.562752636	0.244129024	0.004770798
hsa_circ_0026425	0.940775557	1	0.114040618	0.569420351	0.250654538	0.052478778
hsa_circ_0092361	0.939271249	1	-0.110513431	0.58422628	0.252627286	0.028624788

CircRNAs that were consistently prioritized across all three analytical frameworks were designated as consensus biomarker candidates. To ensure biological relevance and avoid overreliance on computational predictions, the outputs generated at each stage were further subjected to manual evaluation and biological interpretation. This integrative multi-tiered strategy improved prioritization robustness, reduced the likelihood of false-positive candidate selection, enhanced cross-method reproducibility, and provided a more biologically meaningful framework for circRNA biomarker discovery than conventional single-parameter screening approaches.



Supplementary Fig1: *Gene Ontology functional enrichment analysis of upregulated circRNA target genes. GO terms are categorized into biological processes, cellular components, and molecular functions, as depicted in the graoh above. Bar length represents statistical significance (-log10 p-value). Top enriched terms include transcription regulation, neuron projection development, recycling endosome components, and transporter activities.*

Supplementary Fig2: *Gene Ontology functional enrichment analysis of downregulated circRNA target genes. GO terms are categorized into biological processes, cellular components, and molecular functions, as depicted in the graoh above. Bar length represents statistical significance (-log10 p-value). Top enriched terms include transcription regulation (both positive and negative), apoptosis regulation, focal adhesion components, and DNA-binding transcription activator activity.*