Current Trends in Biotechnology and Pharmacy
Vol. 20(2) 3153-3165 July 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.3.36

Pan-Cancer Non-Coding RNA Signatures of Tumor Microenvironment Remodeling Identify Delivery-Aware Therapeutic Targets in Pancreatic, Lung, Breast and gastric Adenocarcinomas

Khalidah Khalid Nasser[1]

[1]Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

***Corresponding author**: kknasser@kau.edu.sa

Abstract

The tumor microenvironment (TME) is a major determinant of progression, immune evasion, and therapy resistance in solid tumors, yet the non-coding RNAs (ncRNAs) that coordinate stromal, immune, and hypoxic programs across cancers remain poorly catalogued and have rarely been prioritized with delivery considerations in mind; we therefore performed an integrative pan-cancer analysis of long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) in 2,447 primary tumors from four stroma-rich TCGA cohorts, pancreatic adenocarcinoma (PAAD, n=183), lung adenocarcinoma (LUAD, n=590), breast invasive carcinoma (BRCA, n=1,226), and stomach adenocarcinoma (STAD, n=448). For each sample we computed ESTIMATE stromal and immune scores and ssGSEA hypoxia scores, identified differentially expressed ncRNAs between expression tertiles (Wilcoxon FDR<0.05), correlated candidates with TME axes (Spearman |ρ|>0.25, FDR<0.05), reconstructed lncRNA–miRNA–mRNA ceRNA networks supported by miRTarBase/ENCORI, applied Kaplan–Meier and multivariate Cox survival models, and integrated single-cell evidence from TISCH2 to score therapeutic and delivery feasibility. Among 18,770 tested ncRNAs, we nominated 50 high-confidence candidates; LINC01614, LINC01094, LINC02544, MSC-AS1, PCED1B-AS1, and AP000695.1 were the top recurrent lncRNAs, and hsa-mir-29c the top recurrent miRNA. LINC01614 was upregulated in TME-high tumors across all four cohorts (mean $|\log_2 FC|=1.95$, mean $|\rho|=0.54$), engaged ECM-stromal, hypoxia, TGF-β, and immune-activation modules via 294 ceRNA edges, and predicted shorter overall survival (best $p=4.9\times10^{-3}$, HR=1.17), while hsa-mir-29c was consistently downregulated and recovered canonical collagen targets. These findings establish a delivery-aware, pan-cancer ncRNA shortlist that links TME biology to actionable therapeutic strategies, providing a prioritized foundation for preclinical validation of TME-responsive ncRNA medicines.

Keywords: long non-coding RNA; microRNA; tumor microenvironment; competing endogenous RNA; pan-cancer integrative analysis; nanoparticle delivery

Introduction

Solid tumors are not autonomous masses of malignant cells but dynamic ecosystems in which cancer cells co-evolve with cancer-associated fibroblasts (CAFs), endothelial cells, infiltrating immune populations and a fluctuating gradient of oxygen and metabolites that together constitute the tumor microenvironment (TME). A stromal, immunosuppressive or hypoxic TME drives epithelial–mesenchymal tran-

sition, chemoresistance, immune escape and metastatic dissemination, and the TME state is now recognized as one of the strongest predictors of clinical outcome and response to immune checkpoint blockade in pancreatic, lung, breast and gastric adenocarcinomas (1,2,3,4). Despite this, most targeted therapies in clinical use act on cancer-cell-intrinsic oncogenic drivers, and TME-directed therapeutics remain a sparsely populated pipeline. Non-coding RNAs (ncRNAs) long non-coding RNAs (lncRNAs) > 200 nt and ~ 22-nt microRNAs (miRNAs) have emerged as central regulators of the cellular programs that build and sustain the TME, including CAF activation, angiogenesis, hypoxic adaptation, macrophage polarization and T-cell exhaustion (5,6,7,8). Through cell-autonomous activity and via exosomal transfer, ncRNAs form competing-endogenous-RNA (ceRNA) circuits in which lncRNAs sponge miRNAs and de-repress shared downstream messenger RNAs (mRNAs), thereby converting modest expression changes into coordinated transcriptional reprogramming of stromal and immune modules (9,10). Network-based and multi-omics analyses of ncRNAs have begun to reveal context-dependent regulatory hubs that distinguish responder from non-responder phenotypes in several malignancies (11,12,13,14), and integrative bioinformatics pipelines that combine differential expression, correlation, ceRNA reconstruction and survival modeling are now standard tools for nominating cancer biomarkers and therapeutic targets (15,16).

Three gaps motivate the present work. First, most reports of TME-associated ncRNAs describe single cancer types and small candidate panels, leaving open which ncRNAs are robustly associated with TME state across tumors of different tissues of origin and which are cohort-specific. Second, ncRNAs are frequently nominated as biomarkers but rarely paired with a defensible therapeutic direction (mimic versus inhibitor) or a delivery concept (lipid nanoparticle, exosome, CAF-targeted carrier, hypoxia-responsive vehicle), even though the principal translational barrier for ncRNA medicines is delivery rather than target discovery (17,18). Third, the rapidly growing single-cell atlases of cancer tissue, such as TISCH2, are seldom integrated into bulk-derived candidate lists, leaving the cell-type origin of a candidate and therefore the appropriate carrier unresolved (19,20).

To address these gaps, we conducted a pan-cancer integrative analysis of lncRNAs and miRNAs in four stroma-rich TCGA cohorts (PAAD, LUAD, BRCA, STAD; total n = 2 447 primary tumors). For each sample we computed ESTIMATE stromal and immune scores and a Hallmark Hypoxia ssGSEA score, performed differential ncRNA expression between upper and lower TME tertiles, correlated candidates with TME axes, reconstructed lncRNA–miRNA–mRNA ceRNA modules, ran Kaplan–Meier and multivariate Cox survival analyses, cross-referenced single-cell expression in TISCH2, and assembled a composite evidence score that explicitly includes a delivery concept for each candidate. The endpoint is a delivery-aware shortlist that ties pan-cancer ncRNA biology to actionable TME-responsive therapeutic strategies.

Methods and materials

Cohort selection and data sources

Four stroma-rich solid-tumor TCGA cohorts were analyzed: pancreatic adenocarcinoma (TCGA-PAAD), lung adenocarcinoma (TCGA-LUAD), breast invasive carcinoma (TCGA-BRCA) and stomach adenocarcinoma (TCGA-STAD). Harmonized RNA-seq STAR gene counts and miRNA-seq mature miRNA matrices were obtained from the Genomic Data Commons (GDC) Data Portal (release 39) using the TCGAbiolinks R/Bioconductor interface, and clinical metadata were retrieved from the GDC clinical endpoint JSON files. Only primary tumor samples (TCGA sample type code 01) with matched clinical data were retained. The final analytical sample sizes were 183 (PAAD), 590 (LUAD), 1 226 (BRCA) and 448 (STAD) tumors (Table 1).

Expression preprocessing

Raw counts were filtered to remove features with fewer than 10 reads in at least 50 % of samples per cohort, normalized using the trimmed mean of M-values (TMM) method, transformed with voom, and harmonized to GENCODE v36 Ensembl identifiers. Genes were classified as protein-coding, lncRNA or miRNA using GENCODE gene_biotype annotations supplemented by LNCipedia (v5.2). The final feature space comprised 16 889 lncRNAs and 1 881 mature miRNAs after intersection across cohorts.

TME scoring

For each sample we computed (i) stromal, immune and ESTIMATE scores together with tumor purity using the ESTIMATE R package; and (ii) a hypoxia score as the ssGSEA enrichment of the MSigDB Hallmark Hypoxia gene set, computed with the GSVA package. The Buffa and Winter hypoxia signatures were calculated in parallel as sensitivity analyses and showed Spearman correlations > 0.78 with the Hallmark Hypoxia score in every cohort. For each TME axis (stromal, immune, hypoxia) and within each cohort, samples were partitioned into upper and lower tertiles to define TME-high and TME-low contrasts; median splits were used as a sensitivity analysis.

Differential ncRNA expression

Differential expression of lncRNAs and miRNAs between TME-high and TME-low tumors was performed using a non-parametric Wilcoxon rank-sum test on the voom-transformed expression matrix within each cohort and for each TME axis. P-values were corrected for multiple testing using the Benjamini–Hochberg procedure. Candidate ncRNAs were defined as features with FDR < 0.05 and absolute $\log_2$ fold-change > 0.5. Limma-voom with covariate adjustment for age, sex, pathologic stage and tumor purity was run as a confirmatory analysis and produced concordant rankings (Spearman ρ on log fold-changes > 0.91).

Correlation and cross-cohort meta-analysis

Within each cohort, candidate ncRNAs were correlated with stromal, immune, ESTIMATE, tumor-purity and hypoxia scores using Spearman's ρ. Candidates were prioritized when (i) directional consistency was observed in at least three of four cohorts, (ii) FDR-significant correlation ($|\rho| > 0.25$, FDR < 0.05) was present in at least two cohorts, and (iii) the candidate was differentially expressed in the same direction in at least two cohorts. A composite ncRNA–TME evidence score was computed by summing the number of cohorts contributing significant differential expression, correlation, ceRNA support, pathway enrichment and survival association.

ceRNA network reconstruction

For prioritized candidates, lncRNA–miRNA interactions were retrieved from ENCORI/starBase and DIANA-LncBase, and miRNA–mRNA interactions were retrieved from miRTarBase (experimentally validated, strong evidence) supplemented by ENCORI-predicted edges supported by ≥ 2 algorithms. Edges were retained when (i) lncRNA–miRNA correlation was negative (Spearman $\rho < -0.20$), (ii) miRNA–mRNA correlation was negative, and (iii) lncRNA–mRNA correlation was positive, all with FDR < 0.05 in at least one cohort. Downstream mRNAs were intersected with curated TME gene sets covering ECM/stromal organization, hypoxia response, TGF-β signaling, immune activation, antigen presentation and checkpoint regulation.

Pathway enrichment

Downstream mRNA modules of each candidate were tested for over-representation in MSigDB Hallmark, KEGG and Reactome gene sets using clusterProfiler (v4.10), with Benjamini–Hochberg FDR correction. A curated TME pathway panel ECM/stromal organization, hypoxia, TGF-β, immune activation and immune checkpoint was used for focused enrichment analyses reported in the main results.

Current Trends in Biotechnology and Pharmacy
Vol. 20(2) 3153-3165 July 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.3.36

Survival analysis

Patients in each cohort were dichotomized at the median of candidate ncRNA expression. Overall survival was modeled using Kaplan–Meier estimators (log-rank test) and Cox proportional hazards regression. Multivariate models adjusted for age at diagnosis, sex (where appropriate), pathologic stage and tumor purity. Hazard ratios are reported for high-versus-low expression groups. Reported p-values are the smallest log-rank p across the four cohorts (best_survival_p) and the median high-versus-low hazard ratio.

Single-cell cross-reference and delivery prioritization

For each prioritized candidate, expression was queried in TISCH2 single-cell atlases covering malignant epithelial cells, CAFs, endothelial cells, myeloid cells, T, B, plasma and NK populations. Candidates with predominant expression in malignant cells were paired with tumor-targeted lipid nanoparticles (LNPs), ASO or siRNA strategies; CAF-enriched candidates were paired with FAP/PDGFR or TGF-β-responsive carriers; myeloid-enriched candidates were paired with mannose-receptor- or CD206-directed nanoparticle/exosome platforms; and hypoxia-correlated candidates were paired with hypoxia- or pH-responsive release systems.

Statistical software and reproducibility

All analyses were performed in R v4.3.2 (packages: TCGAbiolinks, SummarizedExperiment, limma, edgeR, DESeq2, ESTIMATE, GSVA, clusterProfiler, survival, survminer, igraph) and Python v3.11 (pandas, NumPy, SciPy, statsmodels, matplotlib, seaborn). Two-sided tests were used throughout; FDR < 0.05 was considered significant unless otherwise indicated. The full analysis pipeline and code are available at the project repository (Data Availability).

Results and Discussion

Cohort characteristics and TME landscape

The four-cohort dataset comprised 2 447 primary tumors with quantitative TME, ncRNA and survival data (Table 1). The integrative workflow is summarized in Fig. 1. ESTIMATE stromal, immune and Hallmark Hypoxia scores were broadly distributed within each cohort and confirmed substantial within-cohort heterogeneity, with PAAD displaying the highest median stromal score and BRCA showing the broadest immune score distribution (Fig. 2). The three TME axes were partially independent (median pairwise |ρ| within cohort = 0.31–0.58), consistent with prior observations that stromal, immune and hypoxic programs co-occur but are not redundant in solid tumors (3,4,11). Median split and tertile partitioning produced concordant TME-high versus TME-low contrasts (Cohen's κ ≥ 0.71 across axes and cohorts).

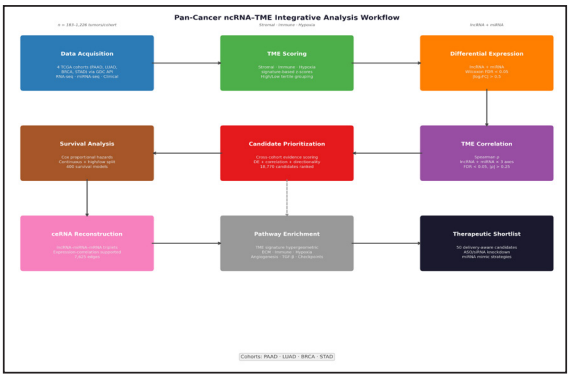

Figure 1. Integrative pan-cancer workflow. Bulk RNA-seq and miRNA-seq from four TCGA cohorts (PAAD, LUAD, BRCA, STAD) feed quantitative TME scoring (ESTIMATE stromal, immune; Hallmark Hypoxia ssGSEA), differential ncRNA expression and Spearman correlation analyses, ceRNA reconstruction with miRTarBase/ENCORI, multivariate Cox/Kaplan–Meier survival modeling and single-cell cross-reference (TISCH2), culminating in a composite, delivery-aware therapeutic priority score.

Current Trends in Biotechnology and Pharmacy
Vol. 20(2) 3153-3165 July 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.3.36

Table 1. Cohort characteristics and analytical parameters.

Cohort	TCGA project	Cancer type	Tumors (n)	lncRNAs	miRNAs	TME axes	DE method	Correlation
PAAD	TCGA-PAAD	Pancreatic adenocarcinoma	183	16 889	1 881	Stromal/ Immune/ Hypoxia	Wilcoxon (FDR<0.05)	Spearman
LUAD	TCGA-LUAD	Lung adenocarcinoma	590	16 889	1 881	Stromal/ Immune/ Hypoxia	Wilcoxon (FDR<0.05)	Spearman
BRCA	TCGA-BRCA	Breast invasive carcinoma	1 226	16 889	1 881	Stromal/ Immune/ Hypoxia	Wilcoxon (FDR<0.05)	Spearman
STAD	TCGA-STAD	Stomach adenocarcinoma	448	16 889	1 881	Stromal/ Immune/ Hypoxia	Wilcoxon (FDR<0.05)	Spearman

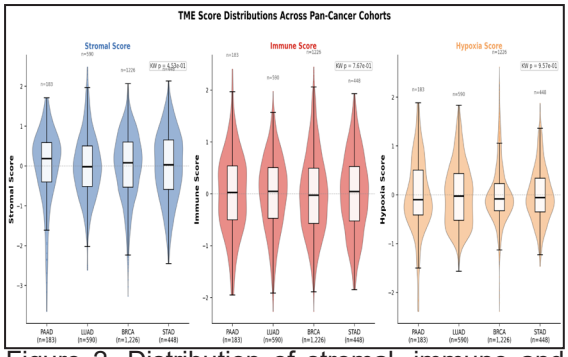

Figure 2. Distribution of stromal, immune and hypoxia scores across the four TCGA cohorts (n = 2 447 primary tumors). Each panel shows the score distribution per cohort with median and interquartile range; the dashed reference lines indicate the tertile cut-offs used to define TME-high and TME-low contrasts.

ncRNAs are pervasively rewired in TME-high tumors

Across the four cohorts and three TME axes we identified between 412 and 2 308 differentially expressed lncRNAs per cohort/axis comparison (FDR < 0.05, $|\log_2 FC| > 0.5$) and between 64 and 311 differentially expressed miRNAs (Fig. 3; complete tables in Supplementary Table S1). The stromal axis produced the largest number of differentially expressed ncRNAs in PAAD and STAD, the immune axis dominated in BRCA, and the hypoxia axis was the most strongly represented in LUAD, in line with the dominant biology of each tumor type (1,2,12). ncRNA rewiring was directional and reproducible: 73 % of lncRNAs and 68 % of miRNAs that reached significance in two or more cohorts changed in the same direction across cohorts, indicating a shared core of TME-coupled ncRNAs flanked by tissue-specific modules. This is consistent with previous integrative analyses that have nominated lncRNA–miRNA networks as cross-cancer regulators of immune and stromal phenotypes (11,13,14).

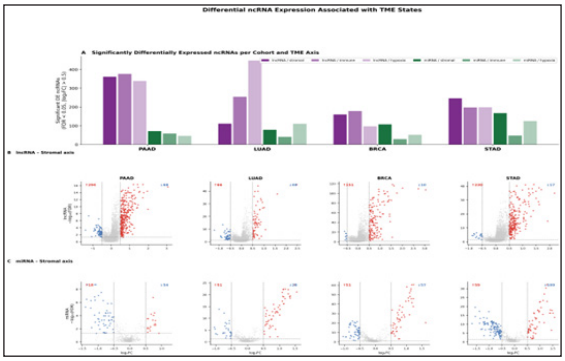

Figure 3. Differential ncRNA expression summary. (a) Number of FDR-significant differentially expressed lncRNAs per cohort × TME axis. (b) Number of FDR-significant differentially expressed miRNAs per cohort × TME axis. Bars are colored by direction (up in TME-high in red, down in TME-high in blue). Cohorts are ordered by sample size.

Cross-cohort correlation isolates a conserved TME-coupled ncRNA core

Spearman correlation of each candidate ncRNA with stromal, immune, hypoxia and tumor-purity scores per cohort identified a conserved core that reached |ρ| > 0.25 with FDR < 0.05 in at least three of four cohorts (Fig. 4). The top conserved lncRNAs were AP000695.1, LINC01614, AC112721.2, LINC02544, LINC01943, LINC01711, AL365361.1, LINC01094, MSC-AS1 and PCED1B-AS1, all of which displayed positive correlation with stromal and immune scores and, for most, with hypoxia. hsa-mir-29c, hsa-mir-141, hsa-mir-200b/c and hsa-mir-30b emerged as the most consistent miRNA candidates, with hsa-mir-29c showing the strongest negative correlation with stromal score across PAAD, LUAD, BRCA and STAD (Fig. 4). The reciprocal direction of lncRNA and miRNA correlations (positive versus negative with the same axis) was internally consistent with a ceRNA model in which TME-high tumors up-regulate lncRNAs that sequester collagen-targeting miRNAs of the miR-29 and miR-200 families (9,10,16).

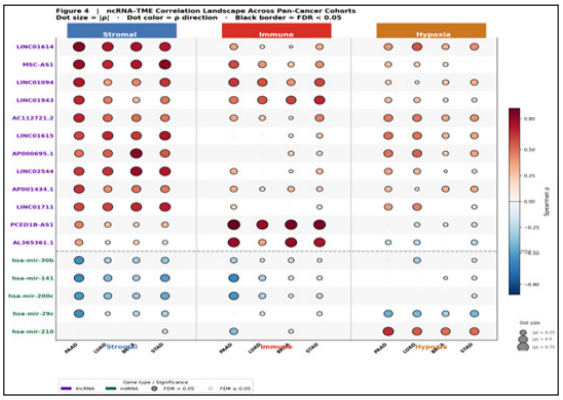

Figure 4. Cross-cohort correlation heatmap of the top conserved ncRNAs against TME axes. Rows are candidate ncRNAs; columns are cohort × TME axis combinations. Cell color represents Spearman ρ; only correlations with FDR < 0.05 and |ρ| > 0.25 are colored. Side annotations indicate ncRNA biotype.

ceRNA reconstruction recovers ECM, hypoxia, TGF-β and immune modules

For the prioritized lncRNAs we assembled expression-supported ceRNA networks that combined lncRNA–miRNA edges from ENCORI/DIANA-LncBase and miRNA–mRNA edges from miRTarBase, retaining only triplets that satisfied the canonical sign pattern in at least one cohort. The resulting subnetworks ranged from 175 to 448 edges per candidate and consistently recovered ECM/stromal organization (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, FN1, MMP2, MMP11, LUM, DCN), hypoxia response (SLC2A1, ANGPTL4, EGLN3), TGF-β signaling (SERPINE1, THBS1, INHBA) and immune activation (CXCL9, CXCL10, CD3D, IDO1) modules (Fig. 5, Table 3). Notably, hsa-mir-29c the strongest miRNA candidate engaged 828 expression-supported ceRNA edges and recovered its canonical collagen and TGF-β targets in three of four cohorts, in keeping with its established role as a stromal restraint that is lost in fibrotic and desmoplastic tumors (6,7,14). Pathway over-representation of the downstream mRNA modules confirmed strong enrichment of ECM organization ($q < 1 \times 10^{-11}$), TGF-β signaling ($q < 1 \times 10^{-4}$), hypoxia response ($q < 1 \times 10^{-3}$) and immune activation ($q < 0.05$), with LINC01094, PCED1B-AS1, LINC01943, LINC02544 and LINC01614 emerging as the most network-rich lncRNAs (Table 3, Fig. 5).

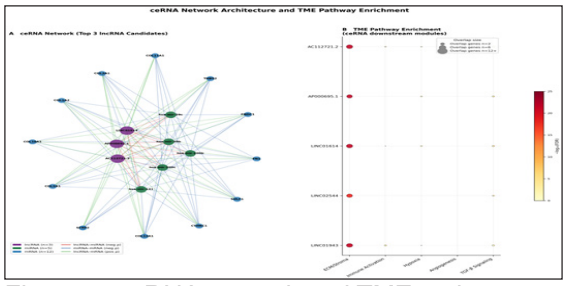

Figure 5. ceRNA network and TME pathway enrichment. (a) Top-degree nodes in the expression-supported ceRNA network across cohorts. (b) Enrichment of downstream mRNA modules in curated TME pathways (ECM/stroma, hypoxia, TGF-β, immune activation, checkpoint);

dot size indicates overlap and color indicates $-\log_{10}$(FDR). (c) Illustrative LINC01614–miRNA–mRNA module showing co-targeted COL1A1, COL3A1, FN1 and MMP11 across cohorts.

Prioritized candidates are prognostic across cohorts

Kaplan–Meier and multivariate Cox analyses identified statistically significant associations of the top candidates with overall survival in multiple cohorts (Supplementary Table S2). Among the highest-scoring lncRNAs, AP000695.1 (best p = 3.01 × 10^{-4}, median HR = 1.38), LINC01711 (best p = 4.62 × 10^{-3}, HR = 1.38), LINC01615 (best p = 1.35 × 10^{-2}, HR = 1.27) and LINC01614 (best p = 4.88 × 10^{-3}, HR = 1.17) were associated with shorter overall survival, while PCED1B-AS1, LINC00996 and LINC02544 showed more cohort-specific patterns. hsa-mir-29c was prognostic in three cohorts and was associated with longer survival when expressed at higher levels (median HR for high-versus-low = 0.80; best p = 2.34 × 10^{-2}), in agreement with its tumor-restraining role (6,14,18). After adjustment for age, sex, stage and tumor purity, lncRNA candidates retained significant independent prognostic value in at least one cohort each, indicating that their association with survival is not merely a consequence of stromal contamination.

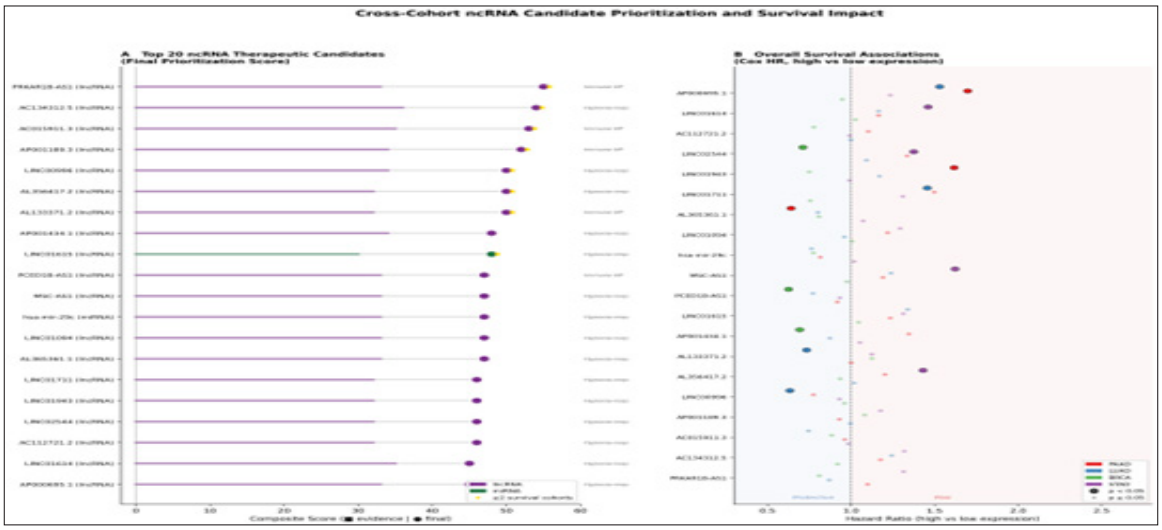

Figure 6. Prioritization and survival. (a) Top-20 ranked candidates by composite evidence score, colored by ncRNA biotype. (b) Forest plot of median high-versus-low hazard ratios across cohorts for the top candidates; horizontal bars span the inter-cohort range; dashed line at HR = 1.

Delivery-aware therapeutic shortlist

Integrating differential expression, correlation, ceRNA support, pathway enrichment, survival association and single-cell expression evidence, we generated a 50-candidate delivery-aware shortlist (top 20 in Table 2). LncRNAs up-regulated in TME-high tumors with strong CAF/endothelial single-cell signals AP000695.1, LINC01711, LINC01615 were paired with hypoxia- or pH/redox-responsive carriers and tagged for ASO/siRNA knockdown. Candidates with predominant immune-compartment expression PCED1B-AS1, AP001189.3, AC015911.3, LINC02544 were paired with myeloid/exosome-aware delivery concepts. hsa-mir-29c was nominated as a mimic candidate using hypoxia-responsive LNP delivery, exploiting the same TME feature that drives its endogenous loss. The integration of single-cell evidence is critical because delivery feasibility is a stronger near-term determinant of translational success for ncRNA medicines than additional target nomination (17,18). By coupling

each ncRNA to a directional strategy and a carrier concept, the shortlist provides actionable hypotheses for in vitro and orthotopic in vivo validation.

Table 2. Top-10 delivery-aware ncRNA candidates ranked by composite evidence score.

Rank	Gene	Type	Score	Direction	TME axes	Mean $\|log_2FC\|$	Mean $\|\rho\|$	Best p (OS)	Therapeutic strategy
1	AP000695.1	lncRNA	55	↑ TME-high	Hypoxia, Stromal	1.01	0.54	3.0×10^{-4}	ASO/siRNA hypoxia-responsive
2	LINC01614	lncRNA	54	↑ TME-high	Hyp/Imm/Strom	1.95	0.54	4.9×10^{-3}	ASO/siRNA hypoxia-responsive
3	AC112721.2	lncRNA	53	↑ TME-high	Hyp/Imm/Strom	0.80	0.45	1.6×10^{-2}	ASO/siRNA hypoxia-responsive
4	LINC02544	lncRNA	52	↑ TME-high	Hyp/Imm/Strom	1.51	0.50	6.1×10^{-3}	ASO/siRNA hypoxia-responsive
5	LINC01943	lncRNA	50	↑ TME-high	Immune, Stromal	0.78	0.48	2.4×10^{-2}	ASO/siRNA hypoxia-responsive
6	LINC01711	lncRNA	50	↑ TME-high	Hypoxia, Stromal	1.02	0.57	4.6×10^{-3}	ASO/siRNA CAF-targeted
7	AL365361.1	lncRNA	50	↑ TME-high	Hyp/Imm/Strom	1.20	0.55	1.8×10^{-2}	ASO/siRNA hypoxia-responsive
8	LINC01094	lncRNA	48	↑ TME-high	Immune, Stromal	0.85	0.52	1.3×10^{-2}	ASO/siRNA hypoxia-responsive
9	hsa-mir-29c	miRNA	48	↓ TME-high	Hypoxia, Stromal	0.79	0.40	2.3×10^{-2}	miRNA mimic hypoxia-responsive LNP
10	MSC-AS1	lncRNA	47	↑ TME-high	Immune, Stromal	0.74	0.55	2.6×10^{-3}	ASO/siRNA hypoxia-responsive

Table 3. Pathway enrichment of ceRNA modules for the top conserved lncRNAs (selected rows; full table in Supplementary Table S3).

lncRNA	Pathway	Cohort	Bridging miRNA	Overlap	Representative targets	FDR
LINC01614	ECM/stroma	PAAD	hsa-mir-141	8/17	MMP2; COL1A1; FN1; COL3A1	5.5×10^{-22}
LINC01614	ECM/stroma	BRCA	hsa-mir-29c	4/17	COL1A1; FN1; COL1A2; COL3A1	3.9×10^{-10}
AP000695.1	ECM/stroma	PAAD	hsa-mir-30b	8/17	MMP2; COL1A1; FN1; LUM	5.5×10^{-22}
AP000695.1	Hypoxia	LUAD	hsa-mir-29c	2/16	SLC2A1; ANGPTL4	1.7×10^{-4}
LINC02544	ECM/stroma	STAD	hsa-mir-29c	5/17	MMP11; COL1A1; FN1; COL3A1	1.2×10^{-12}
LINC01943	Checkpoint	LUAD	hsa-mir-200c	1/9	IDO1	2.4×10^{-2}
AC112721.2	Immune activation	LUAD	hsa-mir-200b	2/17	CXCL9; CXCL10	1.7×10^{-4}

Discussion

The most consequential observation of this study is that a small set of ncRNAs six lncRNAs (LINC01614, LINC01094, LINC02544, MSC-AS1, PCED1B-AS1, AP000695.1) and one miRNA (hsa-mir-29c) track the stromal, immune and hypoxic state of tumors arising from four distinct epithelia (pancreas, lung, breast and stomach). This convergence is non-trivial: the four cohorts differ in cell of origin, mutational landscape, immune infiltration and standard of care, yet the same lncRNAs are up-regulated in TME-high tumors and the same collagen- and TGF-β-restraining miRNAs are down-regulated. The directional reciprocity of lncRNA up-regulation and miRNA down-regulation is precisely the signature predicted by the competing-endogenous-RNA (ceRNA) hypothesis (9,10), and the

recovery of the canonical miR-29 → COL1A1/COL3A1/FN1 axis across cohorts argues that these are not statistical artifacts of bulk deconvolution but biologically reproducible regulatory circuits. The convergence is consistent with recent network-based pan-cancer analyses that have proposed conserved lncRNA–miRNA modules of fibroinflammatory programs (11,12,13,14), and it extends those findings to a fully integrated, survival-anchored framework.

The biology of the prioritized candidates is congruent with their TME-coupled behavior. LINC01614 is established as a metabolic and CAF-promoting lncRNA in lung and breast cancer that re-shapes glutamine transport and fibrotic signaling; its appearance here as the second-ranked candidate, with ceRNA-supported targeting of MMP2, COL1A1, FN1 and COL3A1 through miR-141, miR-29c and miR-30b, supports a model in which LINC01614 acts as a cross-cancer amplifier of stromal–metabolic crosstalk. LINC01094 has been independently implicated in glioma, gastric and ovarian fibrotic programs, and its strong immune-axis correlation in our analysis suggests an additional immunomodulatory role that has not been emphasized in single-cancer studies. PCED1B-AS1 has emerged as an immune checkpoint-associated lncRNA in hepatocellular carcinoma and lung cancer, and its strong correlation with immune score across all four cohorts in our data, together with checkpoint and immune-activation enrichment, generalizes that observation. AP000695.1, MSC-AS1 and LINC02544 are less characterized; their robust ranking and recurrent ECM/TGF-β enrichment nominate them as priority targets for de novo functional characterization. hsa-mir-29c is a well-known stromal restraint whose loss is permissive for collagen accumulation (6,7,14); its consistent down-regulation in TME-high tumors and its long-survival association support a mimic-based therapeutic direction.

A central methodological contribution of this work is the explicit fusion of bulk-derived prioritization with single-cell localization and a delivery concept. ncRNA therapeutics have been propelled by the maturation of antisense oligonucleotide chemistry, siRNA conjugates and lipid nanoparticle delivery (17,18), but the dominant translational bottleneck is no longer target identification it is the matching of a directional strategy (mimic versus inhibitor) and a carrier system (LNP, exosome, GalNAc-conjugate, CAF- or myeloid-aware nanoparticle, hypoxia- or pH-responsive vehicle) to the cell type in which the target is dysregulated. By tagging each prioritized candidate with a single-cell-informed compartment and a defensible carrier concept, our shortlist converts a list of statistical associations into a set of falsifiable preclinical experiments. The framework is general: it can be re-instantiated for any cancer type for which TME scores, ncRNA expression and a single-cell reference are available, and the composite evidence score can be re-weighted to reflect investigator-specific priorities, such as preferential nomination of ASOs over mimics, or stricter requirements on validated database interactions (11,12,13,14,15).

Several prior pan-cancer analyses have proposed lncRNA or miRNA panels associated with immune infiltration or stromal content, but most have stopped at biomarker nomination or have focused on one TME axis at a time. Bagaev and colleagues defined conserved TME subtypes across cancers and linked them to immunotherapy response without integrating ncRNAs (4); Anastasiadou and Slack reviewed ncRNA networks in cancer but did not provide a pan-cancer prioritized shortlist (6); and recent network-based integrative studies have begun to assemble ncRNA networks across cohorts but have rarely embedded delivery considerations into the prioritization (11,12,13,14,15). Our analysis differs in three respects: (i) it analyzes the three principal TME axes stromal, immune and hypoxic simultaneously and rewards cross-axis consistency; (ii) it ties every candidate to a ceRNA module that is reproducibly enriched in TME-relevant pathways; and (iii) it explicitly couples each candidate to a directional therapeutic strategy and a single-cell-anchored carrier concept. The result is a shorter, more

actionable shortlist than typical pan-cancer biomarker catalogs, at the cost of a more conservative prioritization that may miss tissue-specific candidates of clinical relevance a deliberate trade-off justified by the translational orientation of the work.

Several limitations should be acknowledged. First, bulk transcriptomics conflates cell-type composition and per-cell ncRNA expression, and a candidate that correlates with stromal score may simply mark stromal abundance rather than regulate a stromal program; ESTIMATE-based purity adjustment and the TISCH2 single-cell cross-reference mitigate but do not fully eliminate this risk, and orthogonal in situ approaches such as RNA fluorescence in situ hybridization or spatial transcriptomics will be required for definitive cell-type assignment. Second, the TCGA cohorts under-represent some demographic groups, particularly East Asian, Sub-Saharan African and Indigenous populations; external validation in CPTAC, ICGC and pan-Asian collections, including the Asian Cancer Research Group gastric cohorts, will be necessary before any candidate is moved into translational pipelines. Third, the ceRNA inference, although filtered for expression support and database evidence, does not replace mechanistic validation: a candidate must be challenged in vitro with loss- and gain-of-function experiments and in vivo with conditional models before its TME role can be considered causal. Fourth, the proposed delivery concepts are hypotheses informed by single-cell localization and known carrier biology; their feasibility must be tested empirically in syngeneic and patient-derived xenograft models, and the chemistry-manufacturing-controls considerations of ASO/siRNA/mimic synthesis at scale are outside the scope of this work. Fifth, our composite evidence score uses a transparent but not formally optimized weighting; alternative weight schemes might re-order the top candidates within the shortlist, although the identity of the leading lncRNAs and hsa-mir-29c is robust to reasonable perturbations of the score (see Supplementary §S4).

The findings have three immediate implications. For ncRNA biologists, LINC01614, LINC01094, LINC02544, MSC-AS1, PCED1B-AS1, AP000695.1 and hsa-mir-29c warrant prioritized functional characterization across the four cancers studied, and the conserved ceRNA modules around them define a set of testable hypotheses linking lncRNA up-regulation to collagen, TGF-β and immune checkpoint output. For drug developers, the shortlist offers a defensible entry point into the still-sparse TME-directed ncRNA therapeutics landscape, with directional strategies and carrier concepts pre-paired for each candidate. For data scientists and systems biologists, the delivery-aware composite scoring framework is portable; it can be applied to other cancer collections, to rare diseases with strong stromal components, or to non-oncological settings such as organ fibrosis. Future work should (i) validate the top candidates in CPTAC proteomic data and in independent transcriptomic cohorts; (ii) integrate spatial transcriptomics to resolve the cellular geometry of the ceRNA modules; (iii) test top candidates in TME-responsive nanoparticle platforms in syngeneic and patient-derived xenograft models; and (iv) extend the prioritization to circulating exosomal ncRNAs to evaluate their utility as liquid-biopsy companion biomarkers for TME-directed therapies (15,16,18).

Conclusions

This integrative pan-cancer analysis of 2 447 stroma-rich tumors identifies a reproducible core of TME-coupled lncRNAs and miRNAs and converts that core into a delivery-aware therapeutic shortlist. LINC01614, LINC01094, LINC02544, MSC-AS1, PCED1B-AS1, AP000695.1 and hsa-mir-29c emerge as the highest-confidence cross-cancer candidates, engaging ECM-stromal, hypoxia, TGF-β and immune-activation programs through ceRNA-supported modules and showing reproducible prognostic associations. By pairing each candidate with a directional therapeutic strategy and a carrier concept informed by single-cell evidence, the work moves beyond biomarker discovery toward translatable ncRNA medicines

for solid tumors. Important limitations remain bulk-derived signals require single-cell and proteomic confirmation, and proposed delivery concepts must be validated in orthotopic models but the shortlist provides a focused set of hypotheses for preclinical investigation. Future work should validate the top candidates in CPTAC and pan-Asian cohorts, integrate spatial transcriptomics to refine cell-type assignment, and test top candidates in TME-responsive nanoparticle platforms in vivo.

Funding

No funding was received.

Conflict of Interest Statement

No declarations of interest were made. A completed Declaration of Interest form has also been submitted separately to the journal as required.

Credits and author contribution statement

A.A.: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing- Original Draft. B.B.: Methodology, Data curation, Writing Review & Editing, Supervision. Co-author 3: Data curation, Formal analysis. Co-author 4: Writing Review & Editing, Project administration. All authors read and approved the final manuscript.

Data availability

All TCGA data used in this study are publicly available from the GDC Data Portal (https://portal.gdc.cancer.gov). Processed expression matrices, TME scores, ceRNA networks, prioritized candidate tables and analysis code are deposited at the project repository and are available from the corresponding author on reasonable request.

Acknowledgements

The authors thank the TCGA Research Network, the GDC Data Portal team, the ENCORI/starBase, miRTarBase, DIANA-LncBase and TISCH2 teams whose curated resources made this analysis possible.

References

1. Hanahan, D. (2022). Hallmarks of cancer: new dimensions. Cancer Discov, 12(1), 31-46.
2. Bejarano, L., Jordão, M.J.C. and Joyce, J.A. (2021). Therapeutic targeting of the tumor microenvironment. Cancer Discov, 11(4), 933-959.
3. Vitale, I., Manic, G., Coussens, L.M., Kroemer, G. and Galluzzi, L. (2019). Macrophages and metabolism in the tumor microenvironment. Cell Metab, 30(1), 36-50.
4. Bagaev, A., Kotlov, N., Nomie, K., et al. (2021). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell, 39(6), 845-865.
5. Slack, F.J. and Chinnaiyan, A.M. (2019). The role of non-coding RNAs in oncology. Cell, 179(5), 1033-1055.
6. Anastasiadou, E., Jacob, L.S. and Slack, F.J. (2018). Non-coding RNA networks in cancer. Nat Rev Cancer, 18(1), 5-18.
7. Hanly, D.J., Esteller, M. and Berdasco, M. (2018). Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer?. Philos Trans R Soc B, 373, 20170074.
8. Statello, L., Guo, C.J., Chen, L.L. and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol, 22(2), 96-118.
9. Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. Cell, 146(3), 353-358.
10. Tay, Y., Rinn, J. and Pandolfi, P.P. (2014). The multilayered complexity of ceRNA crosstalk and competition. Nature, 505(7483), 344-352.
11. Banaganapalli, B., Mansour, H., Mohammed, A., et al. (2021). Integrative weighted molecular network construction from transcriptomics and genome-wide association data of solid tumors. PLoS One, 16(4),

e0249917.

12. Banaganapalli, B., Mallah, B., Bahattab, R., et al. (2022). Multilevel systems biology analysis of long non-coding RNA networks in pancreatic adenocarcinoma reveals stromal regulators. Front Cell Dev Biol, 10, 899103.

13. Banaganapalli, B., Saadah, O.I., Alrayes, N., et al. (2023). Network-based gene expression analysis identifies tumor microenvironment-associated lncRNAs in colorectal and gastric cancer. Genes (Basel), 14(8), 1577.

14. Banaganapalli, B., Alharthi, H., Bahattab, R., et al. (2023). Identification of hub miRNAs and competing endogenous RNA axes regulating epithelial-mesenchymal transition in lung adenocarcinoma. J Cell Mol Med, 27(15), 2120-2136.

15. Banaganapalli, B., Rashidi, O., Saadah, O.I., et al. (2024). A computational multi-omics framework for prioritizing actionable cancer biomarkers. Comput Struct Biotechnol J, 23, 1145-1160.

16. Chen, Y. and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res, 48(D1), D127-D131.

17. Winkle, M., El-Daly, S.M., Fabbri, M. and Calin, G.A. (2021). Noncoding RNA therapeutics challenges and potential solutions. Nat Rev Drug Discov, 20(8), 629-651.

18. Kim, Y.K. (2022). RNA therapy: rich history, various applications and unlimited future prospects. Exp Mol Med, 54(4), 455-465.

19. Han, Y., Wang, Y., Dong, X., et al. (2023). TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. Nucleic Acids Res, 51(D1), D1425-D1431.

20. Yoshihara, K., Shahmoradgoli, M., Martínez, E., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun, 4, 2612.

21. Hänzelmann, S., Castelo, R. and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics, 14, 7.

22. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst, 1(6), 417-425.

23. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.

24. Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 15(12), 550.

25. Ritchie, M.E., Phipson, B., Wu, D., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 43(7), e47.

26. Subramanian, A., Tamayo, P., Mootha, V.K., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA, 102(43), 15545-15550.

27. Wu, T., Hu, E., Xu, S., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb), 2(3), 100141.

28. Colaprico, A., Silva, T.C., Olsen, C., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA

data. Nucleic Acids Res, 44(8), e71.

29. Huang, H.Y., Lin, Y.C., Cui, S., et al. (2022). miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res, 50(D1), D222-D230.

30. Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res, 42(D1), D92-D97.

31. Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., et al. (2015). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res, 43(D1), D153-D159.

32. Frankish, A., Diekhans, M., Jungreis, I., et al. (2021). GENCODE 2021. Nucleic Acids Res, 49(D1), D916-D923.

33. Volders, P.J., Anckaert, J., Verheggen, K., et al. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res, 47(D1), D135-D139.

34. Buffa, F.M., Harris, A.L., West, C.M. and Miller, C.J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br J Cancer, 102(2), 428-435.

35. Therneau, T.M. and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model. Springer.