

Materials and methods

The Cancer Genome Atlas/Genomic Data Commons (TCGA/GDC) data acquisition. Mature microRNA (miRNA or miR) isoform expression data for hsa-miR-324-5p (accession no. MIMAT0000761) were obtained from TCGA (cancer.gov/ccg/research/genome-sequencing/tcga) through the National Cancer Institute GDC (gdc.cancer.gov/; accessed on July 21, 2026). Data were queried using the TCGAbiolinks R package (version 2.34.1; bioconductor.org/packages/TCGAbiolinks/), with the data category set to ‘Transcriptome Profiling’, data type to ‘Isoform Expression Quantification’, workflow type to ‘BCGSC miRNA Profiling’ and access type to ‘open’. TCGA analysis was performed using R version 4.4.1 (r-project.org/) with TCGAbiolinks version 2.34.1 (bioconductor.org/packages/TCGAbiolinks/), data.table version 1.18.2.1 (cran.r-project.org/package=data.table), dplyr version 1.2.1 (cran.r-project.org/package=dplyr), and ggplot2 version 4.0.3 (cran.r-project.org/package=ggplot2). A total of 33 TCGA cancer cohorts were queried.

Sample selection and classification. Samples were classified according to the sample-type annotations provided by GDC. Samples annotated as Primary Tumor, Metastatic, Recurrent Tumor, Additional-New Primary, Additional Metastatic, or Primary Blood Derived Cancer-Peripheral Blood were assigned to the tumor group, whereas samples annotated as Solid Tissue Normal were assigned to the normal group. When sample-type annotation alone did not permit classification, TCGA sample codes derived from the sample barcode were used as a secondary criterion: Sample codes 01-09 were classified as tumor and codes 10-19 as normal. Samples without a valid tumor/normal classification or finite expression values were excluded.

To avoid duplication at the sample level, one record was retained for each unique TCGA sample identifier when multiple technical aliquots corresponded to the same sample. Distinct biological samples from the same participant were retained. The final dataset comprised 11,331 samples from 33 TCGA cohorts, including 10,656 tumor and 675 normal samples. Cohort-specific tumor and normal sample numbers are provided in Table SI.

hsa-miR-324-5p quantification and normalization. For each sample, all miRNA isoform records corresponding to hsa-miR-324-5p were identified, and their read counts were summed to obtain the total hsa-miR-324-5p target count. hsa-miR-324-5p expression was normalized to counts per million mature miRNA reads (CPM), calculated as: $CPM = \text{target count} / \text{total mature miRNA count} \times 1 \times 10^6$.

The normalized expression values were transformed as $\log_2(CPM + 1)$ for statistical analysis and visualization.

Statistical analysis. Within each TCGA cancer cohort, tumor and normal samples were compared using a two-sided Wilcoxon rank-sum test with unpaired observations. Statistical testing was performed only when both the tumor and normal groups contained ≥ 2 samples. Cohorts without sufficient normal samples were displayed descriptively but were not subjected to tumor-normal statistical comparison.

To account for multiple cohort-level comparisons, Benjamini-Hochberg adjustment was applied to the raw Wilcoxon P values. Statistical significance was based on the unadjusted Wilcoxon P-values, consistent with the prespecified plotting procedure. $P < 0.05$ was considered to indicate a statistically significant difference. Both raw and Benjamini-Hochberg-adjusted P-values are reported in Table SI.

Table SI. Sample counts of hsa-microRNA-324-5p expression across The Cancer Genome Atlas cancer cohorts.

Cohort	Tumor	Normal	Total	Tumor median log ₂ (CPM + 1)	Normal median log ₂ (CPM + 1)	Raw P-value	BH-adjusted P-value
Adrenocortical carcinoma	80	0	80	5.535	NA	NA	NA
Bladder urothelial carcinoma	413	19	432	4.886	2.749	6.15x10 ⁻¹⁰	2.83x10 ⁻⁹
Breast invasive carcinoma	1,098	104	1,202	4.425	3.923	4.19x 10 ⁻⁴	8.04x10 ⁻⁴
Cervical squamous cell carcinoma	309	3	312	4.353	3.037	0.0265	0.0381
Cholangiocarcinoma	36	9	45	4.386	4.052	0.0631	0.0806
Colon adenocarcinoma	453	8	461	3.903	5.764	1.15x10 ⁻⁵	3.78x10 ⁻⁵
Lymphoid neoplasm diffuse large B cell lymphoma	47	0	47	5.126	NA	NA	NA
Esophageal carcinoma	185	13	198	4.443	3.686	4.20 x 10 ⁻⁵	1.07 x 10 ⁻⁴
Glioblastoma multiforme	268	5	273	10.657	5.547	2.14 x 10 ⁻⁴	4.48 x 10 ⁻⁴
Head and Neck Squamous Cell Carcinoma	525	44	569	4.507	2.806	<1.00 x10 ⁻³⁰⁰	<1.00 x10 ⁻³⁰⁰
Kidney Chromophobe	66	25	91	3.110	3.943	1.76 x 10 ⁻⁷	6.73 x 10 ⁻⁷
Kidney Renal Clear Cell Carcinoma	521	71	592	4.129	4.563	2.12 x 10 ⁻⁵	6.09 x 10 ⁻⁵
Kidney renal papillary cell carcinoma	292	34	326	4.281	4.216	0.8273	0.8273
Acute myeloid leukemia	188	0	188	7.250	NA	NA	NA
Brain lower grade glioma	530	0	530	4.123	NA	NA	NA
Liver hepatocellular carcinoma	375	50	425	4.360	4.031	0.0021	0.0034
Lung adenocarcinoma	518	46	564	4.609	2.698	3.64 x 10 ⁻¹³	2.09 x 10 ⁻¹²
Lung squamous cell carcinoma	488	45	533	5.121	3.880	<1.00 x10 ⁻³⁰⁰	<1.00 x10 ⁻³⁰⁰
Mesothelioma	87	0	87	4.591	NA	NA	NA
Ovarian serous cystadenocarcinoma	531	0	531	5.558	NA	NA	NA
Pancreatic adenocarcinoma	179	4	183	4.121	5.233	0.0144	0.0220
Pheochromocytoma and paraganglioma	184	3	187	5.438	5.086	0.2994	0.3346
Prostate adenocarcinoma	499	52	551	3.644	3.003	5.50 x10 ⁻⁵	1.26 x 10 ⁻⁴
Rectum adenocarcinoma	162	3	165	4.114	4.693	0.0448	0.0607
Sarcoma	263	0	263	4.331	NA	NA	NA
Skin cutaneous melanoma	450	2	452	5.873	4.485	0.2467	0.2987
Stomach adenocarcinoma	436	41	477	4.462	3.799	0.0012	0.0021
Testicular germ cell tumors	156	0	156	4.462	NA	NA	NA
Thyroid carcinoma	514	59	573	5.037	4.962	0.6899	0.7213

Thymoma	124	2	126	5.176	4.794	0.3055	0.3346
Uterine corpus endometrial carcinoma	542	33	575	5.276	3.176	$<1.00 \times 10^{-300}$	$<1.00 \times 10^{-300}$
Uterine carcinosarcoma	57	0	57	6.070	NA	NA	NA
Uveal melanoma	80	0	80	5.654	NA	NA	NA

CPM, counts per million; BH, Benjamini-Hochberg; NA, not applicable.