ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Nucleolin overexpression correlates with poor prognosis and immune checkpoint regulation across various cancer types: Insights from The Cancer Genome Atlas and GTEx analyses

Kruthika Prakash[1], Janani Balaji[1], Surya Babu[1], Ramya Lakshmi Rajendran[2,3,4], Prakash Gangadaran[2,3,4*], ArulJothi Kandasamy Nagarajan[1*], and Byeong-Cheol Ahn[2,3,4,5*]

[1]Department of Genetic Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[2]BK21 FOUR KNU Convergence Educational Program of Biomedical Sciences for Creative Future Talents, Department of Biomedical Sciences, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[3]Department of Nuclear Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[4]Cardiovascular Research Institute, Kyungpook National University, Daegu, Republic of Korea

[5]Department of Nuclear Medicine, Kyungpook National University Hospital, Daegu, Republic of Korea

***Corresponding authors:**
Prakash Gangadaran (prakashg@knu.ac.kr);
ArulJothi Kandasamy Nagarajan (aruljotn@srmist.edu.in);
Byeong-Cheol Ahn (abc2000@knu.ac.kr)

Citation: Prakash K, Balaji J, Babu S, *et al*. Nucleolin overexpression correlates with poor prognosis and immune checkpoint regulation across various cancer types: Insights from The Cancer Genome Atlas and GTEx analyses. *Eurasian J Med Oncol.* 2026;10(5):026220241. doi: 10.36922/EJMO026220241

Received: May 27, 2026

Revised: July 29, 2026

Accepted: August 17, 2026

Published online: September 4, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Introduction: Nucleolin (NCL) is an established cancer target proposed to modulate the tumor immune microenvironment, but has not been systematically tested across cancers.

Objective: To systematically investigate the pan-cancer association of NCL expression with tumor progression, prognosis, and the tumor immune microenvironment, with emphasis on immune infiltration and immune checkpoint genes.

Methods: We analyzed 9,358 TCGA tumors across 33 cancer types against 7,262 GTEx and 727 adjacent normal samples using uniformly reprocessed transcriptomes, assessing differential expression with Wilcoxon tests and Cliff's delta, and stage association with the Jonckheere–Terpstra trend test. Three survival endpoints were modeled using multivariable Cox regression, adjusting for age, sex, stage, and grade, with proportional hazards checks. Immune infiltration was estimated by seven deconvolution algorithms as purity-adjusted partial correlations, with checkpoint correlations additionally proliferation-adjusted. All test families were Benjamini–Hochberg-corrected, and the findings were validated in nine independent Clinical Proteomic Tumor Analysis Consortium (CPTAC) cohorts.

Results: NCL was overexpressed in 24 of 29 evaluable cancers (Cliff's delta up to +0.93) and reduced in ovarian carcinoma; 7 of 9 CPTAC cohorts confirmed elevated NCL protein. A stage-ordered increase was present in only 2 of 17 cancers. After covariate adjustment and false discovery rate correction, NCL remained independently prognostic in only two: kidney renal papillary cell carcinoma (overall survival hazard ratio 2.12, 95% CI 1.37–3.29) and adrenocortical carcinoma (progression-free hazard ratio 2.41, 95% CI 1.40–4.16).

Conclusion: NCL correlated negatively with the microenvironment score in every

cancer where the association was significant, and positively with B7-H3 in 21 of 33 cancers, whereas CTLA-4, PD-1, and TIM-3 were largely uncorrelated. NCL, therefore, associates with tumor-intrinsic immunosuppressive ligands and an immune-excluded phenotype rather than with T-cell checkpoints, and its transcript-level prognostic value is limited, consistent with prior evidence that this effect is localization-dependent.

Keywords: Nucleolin; The Cancer Genome Atlas; Tumorigenesis; Immune regulation; Prognostic biomarker; Therapeutic target

1. Introduction

Cancer is among the most common causes of morbidity and mortality worldwide, representing an important public health issue.[1,2] Large public consortia, most notably The Cancer Genome Atlas (TCGA), the Genotype-Tissue Expression project (GTEx), and the Clinical Proteomic Tumor Analysis Consortium (CPTAC), now make it possible to evaluate a candidate molecule across the full spectrum of human malignancy rather than one tumor type at a time.[3-5] This has produced a large body of literature on pan-cancer biomarker reports; however, much of this literature is based on univariate association testing without correction for multiple comparisons or reporting of effect sizes. Thus, statistically significant but clinically negligible associations can be presented as meaningful.[6,7] This emphasizes the need for biomarkers that are both mechanistically informative and rigorously validated, and the identification of such biomarkers remains a central focus of cancer research.

Nucleolin (NCL) is a highly conserved, multifunctional phosphoprotein predominantly localized in the nucleolus, although it also shuttles to the nucleoplasm, cytoplasm, and cell surface under certain conditions.[8,9] It plays an essential role in ribosome biogenesis, ribosomal RNA processing, chromatin remodeling, messenger RNA (mRNA) stability, and cellular stress responses. It is required for the high rates of ribosome production that proliferating cells demand.[10,11] NCL has attracted increasing attention in cancer research for its dynamic localization and diverse functions. It is frequently overexpressed in tumors and implicated in oncogenic processes. In particular, cell-surface NCL is targeted by agents such as the aptamer AS1411, the pseudopeptide N6L, and the F3 peptide, several of which are in early-phase clinical evaluation.[12-14] NCL has been reported to be overexpressed and prognostically adverse in individual malignancies, including endometrial carcinoma, lung cancer, hepatocellular carcinoma, and pancreatic ductal adenocarcinoma (PDAC). Previous studies have reported that its subcellular localization determines the direction of its prognostic effect.[15-18] NCL is thus regarded as a promising biomarker for cancer diagnosis and prognosis, as well as a potential therapeutic target.

Given NCL's involvement in cancer-related pathways, a comprehensive understanding of its role across cancer types is essential. In particular, whether NCL carries prognostic information independent of clinical variables already used in practice, such as stage and grade, remains underexplored. NCL expression is closely linked to proliferation, which is strongly associated with tumor grade and stage; a correlation between NCL and survival may reflect established prognostic markers rather than contribute to them. The current NCL literature has largely failed to consider the multivariable modeling needed to discriminate among these alternatives. Here, we evaluated NCL expression in normal and cancer tissues across 33 cancer types, utilizing consistently reprocessed TCGA, GTEx, and nearby normal transcriptomes to directly compare tumor and normal tissues. We also assessed the association between NCL expression and cancer progression, including pathological stage. We modeled the relationship with patient survival using multivariable Cox regression, adjusting for age, sex, stage, and grade.

In addition, we investigated the relationship between NCL and the tumor microenvironment, focusing on its correlation with immune cell infiltration. NCL has been proposed to shape the tumor immune microenvironment; therefore, NCL targeting reduces immunosuppression in pancreatic cancer models, and the midkine (MDK)–NCL axis has been implicated in an immunosuppressive niche in endometrial carcinoma.[14,19] We assessed the association between NCL expression and immune cell populations, including cancer-associated fibroblasts (CAFs), macrophages, and $CD8^+$ T cells, all of which are critical modulators of tumor progression and immune evasion, across seven independent transcriptome deconvolution algorithms, adjusting for tumor purity because such algorithms differ in their reference signatures

and sensitivity to purity.[20-24] We further examined NCL's relationship with established immune checkpoint and immunomodulatory genes.

To gain further insight into the molecular mechanisms focusing on the role of NCL in cancer, we ranked genes by their within-cancer correlation with NCL. We analyzed the resulting profiles using gene set enrichment analysis (GSEA) against pathways involved in proliferation, DNA repair, apoptosis, and angiogenesis. We sought independent confirmation of our transcript-level findings at the protein level across nine CPTAC proteomic cohorts. Our findings provide a foundation for evaluating NCL as a pan-cancer biomarker and therapeutic target, with potential implications for identifying the settings in which NCL-directed treatment strategies are potentially relevant. This study adds to the growing body of evidence characterizing NCL's role in cancer biology and its relationship with the tumor immune microenvironment.

2. Materials and methods

All analyses were performed in Python 3.14 using NumPy,[25] SciPy,[26] pandas, statsmodels, *lifelines*[27] and *gseapy*.[28] Analysis code, intermediate result tables, and the exact software environment are openly available (see Data availability), and the workflow is documented step by step in the accompanying repository.

2.1. Expression data and cohort analyses

Transcriptome data were obtained from the University of California, Santa Cruz (UCSC) Xena Toil recomputation of TCGA, Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and GTEx (dataset TcgaTargetGtex_rsem_gene_tpm), comprising 60,498 genes across 19,131 samples quantified as $\log_2$(transcripts per million [TPM] + 0.001).[3,29,30] TCGA and GTEx data were originally processed using different pipelines, and a direct comparison between them can introduce batch effects larger than the biological differences of interest. The Toil recompute applies a single alignment and quantification pipeline to both datasets, which makes a tumor-versus-GTEx comparison valid.

Samples were assigned to 33 TCGA cohorts using the Xena phenotype annotation. Tumor samples were labeled as "Primary Tumor," and each TCGA–GTEx pairing was assigned a match quality of good, approximate, or none. Cancers with no acceptable GTEx counterpart (mesothelioma [MESO], thymoma [THYM], and uveal melanoma [UVM]) were analyzed against adjacent normal tissue only. Expression values were RNA-Seq by Expectation-Maximization (RSEM)-derived TPM estimates.[31] Gene identifiers were mapped from GENCODE v23[32] to gene symbols; where a symbol mapped to multiple identifiers, the row with the highest mean expression was retained for single-gene analyses, and the most variable row for correlation analyses (Table S1).

2.2. Clinical, genomic, and immune annotation

Clinical data, survival endpoints, tumor mutational burden, and microsatellite instability (MSI) scores for all 32 TCGA PanCancer Atlas studies were retrieved from cBioPortal.[33,34] Survival endpoints follow the curated definitions of the TCGA Clinical Data Resource, such as overall survival (OS), disease-specific survival (DSS), and progression-free survival (PFS), which harmonize endpoint definitions across studies and are recommended over raw TCGA follow-up fields.[35] Covariates extracted were age at diagnosis, sex, American Joint Committee on Cancer (AJCC) pathological stage, and histological grade. Tumor mutational burden (non-synonymous), Microsatellite Analysis for Normal-Tumor Instability (MANTIS) score, aneuploidy score, and fraction of genome altered were also retrieved.[36]

Immune infiltration estimates for TCGA samples were obtained from Tumor Immune Estimation Resource 2.0 (TIMER2.0), which provides pre-computed estimates from seven algorithms, including TIMER,[37] Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT), CIBERSORT Absolute (CIBERSORT-ABS), quantification of immune cell types from bulk RNA sequencing data (quanTIseq), Microenvironment Cell Populations-counter (MCP-counter), xCell, and Estimating the Proportion of Immune and Cancer cells (EPIC).[20-24,38] xCell immune, stromal, and microenvironment scores were taken from the same resource. Tumor purity was represented by the EPIC "uncharacterized cell" fraction, which is EPIC's estimate of the compartment neither immune nor stromal.

2.3. Differential expression and stage association

Nucleolin expression in tumors was assessed against each normal comparator for each cancer using two-sided Wilcoxon rank-sum testing. Effect sizes are reported as Cliff's delta[39] with 95% percentile bootstrap confidence intervals (CI; 2,000 resamples), and additionally as Hedges' *g* (Table S2).[40] A non-parametric effect size was preferred because expression distributions are skewed, and several cohorts have very small numbers of adjacent normals. Bootstrap intervals were used in preference to Cliff's asymptotic variance because that approximation is unreliable at the sample sizes some cohorts provide.

Association with pathological stage was assessed with the Jonckheere–Terpstra test for monotonic trend across

ordered stages I–IV, using the tie-corrected normal approximation, alongside a Kruskal–Wallis test for any difference among stages (Table S3).[41,42] The distinction matters when a Kruskal–Wallis result establishes only that stages differ, not that expression increases with stage, and comparisons of each stage against normal tissue, as shown by several web tools, do not test a stage-ordered trend at all. Cancers were included when at least three stage groups contained ≥5 patients, and the cohort totaled ≥40 staged patients.

2.4. Survival analysis

Nucleolin expression was *z*-scored within each cohort for each cancer and each outcome (OS, DSS, and PFS), so that hazard ratios (HRs) are reported per standard deviation and are comparable across cancers. Three models were fitted using *lifelines*,[27] a log-rank test comparing patients above and below the cohort median (the univariate analysis reported in most prior work), a univariate Cox model on continuous NCL, and a multivariable Cox model including age, sex, stage, and grade. Covariates were retained only when available for >80% of the cohort and showed more than one distinct value; the covariates actually used were recorded for every model. Cancers were modeled when ≥30 patients and ≥10 events were available.

The proportional hazards assumption was tested for each multivariable model using Schoenfeld residuals with rank-transformed time.[43] Models in which the NCL term violated the assumption ($p < 0.05$) were identified explicitly in Results and in Table S4, because an HR from such a model summarizes a time-varying effect and should not be read as a single constant risk multiplier.

2.5. Immune microenvironment analysis

For each NCL expression and estimated infiltration association in each cancer, the correlation coefficient was calculated as a Spearman correlation coefficient and as a partial Spearman correlation coefficient adjusting for tumor purity using a residual approach, implemented by regressing the ranks of both variables on the ranks of purity and correlating the residuals (*t*-test on $n-3$ degrees of freedom; Table S5).

To assess robustness to algorithm choice, 10 cell types resolvable by more than one algorithm were designated canonical. For each cancer and canonical cell type, we recorded how many algorithms yielded a significant purity-adjusted association after false discovery rate (FDR) correction and whether they agreed in sign. A combination was called concordant when at least two algorithms were significant, and none disagreed in direction, and conflicting when significant estimates of opposite sign were both present.

Associations between NCL and immune, stromal, and microenvironment scores, tumor mutational burden, MANTIS MSI score, aneuploidy score, and fraction of genome altered were computed as Spearman correlations, with CIs obtained by Fisher *z* transformation using the standard error of Bonett and Wright (Table S6).[44]

2.6. Immune checkpoint analysis

Fifteen immune checkpoint and immunomodulatory genes were examined including *CD274* (programmed cell death 1 ligand 1 [PD-L1]), *PDCD1LG2* (programmed cell death 1 ligand 2 [PD-L2]), *PDCD1* (programmed cell death 1 [PD-1]), *CTLA4*, *HAVCR2* (T-cell immunoglobulin and mucin-domain containing-3 [TIM-3]), *LAG3*, *TIGIT*, *IDO1*, *BTLA*, *SIGLEC15*, *CD276* (B7-H3), *IL10*, *TGFB1*, *ENTPD1* (CD39), and *NT5E* (CD73). For each cancer, the Spearman correlation between NCL and each gene was computed and reported in three forms: unadjusted, adjusted for tumor purity, and adjusted for proliferation.

The proliferation adjustment is essential to interpretation. NCL is a ribosome-biogenesis gene whose expression tracks proliferative rate, and proliferative tumors differ systematically in immune composition. A raw correlation between NCL and a checkpoint gene may therefore reflect nothing more specific than shared covariation with proliferation. The proliferation score was defined as the mean within-cohort *z*-score of 10 canonical markers (*MKI67*, *PCNA*, *TOP2A*, *CCNB1*, *BUB1*, *AURKA*, *CDK1*, *TYMS*, *RRM2*, and *TK1*). Associations described below as robust are those that remain significant after adjustments for both purity and proliferation.

2.7. Gene set enrichment analysis

All genes were ranked according to their within-cancer Spearman correlation with NCL, followed by pre-ranked GSEA against the Molecular Signatures Database (MSigDB) Hallmark and Reactome collections (v2024.1.Hs), using 1,000 permutations and gene sets restricted to 15–500 genes.[28,45-47] Correlations were computed within cancers rather than across pooled samples, where tissue-of-origin differences in composition dominate pooled pan-cancer correlations and do not describe any within-tumor relationship.

Ranked lists were restricted to genes expressed in the cancer of interest, defined as having TPM > 1 in at least 25% of tumors in that cohort. Without this restriction, approximately 10% of genes carried tied ranking statistics, because genes that are essentially undetected yield degenerate correlations that GSEA then orders arbitrarily; the restriction reduced ties to below 0.2%.

2.8. Independent proteomic validation

Protein-level validation used CPTAC cohorts accessed through the CPTAC Python package, including breast invasive carcinoma (BRCA), clear cell renal cell carcinoma (ccRCC), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), ovarian serous cystadenocarcinoma (OV), PDAC, and uterine corpus endometrial carcinoma (UCEC).[5] CPTAC provides an independent patient series, measured by mass spectrometry rather than sequencing, at the protein rather than the transcript level, with adjacent normal tissue from the same patients, making it the most stringent validation available without new experimental work. Tumor and normal samples were compared using the Wilcoxon rank-sum test with Cliff's delta and the Wilcoxon signed-rank test, with the same patient contributing to both.

2.9. Multiple testing and reporting

The Benjamini–Hochberg FDR correction was applied separately within each family of statistical tests. Corrections were performed within each cancer type for differential expression analyses; within each cancer type and survival endpoint for survival analyses; across all combinations of cancer type, immune cell type, and infiltration algorithm for immune infiltration analyses; across all combinations of cancer type and immune checkpoint gene; across all combinations of cancer type and genomic score; and across all GSEA. Adjusted p-values are reported as q values, and statistical significance was defined as $q < 0.05$ throughout. All associations are presented as effect estimates with corresponding 95% CIs, along with the number of samples included in each analysis.

2.10. Reproducibility

Analyses are deterministic given fixed random seeds (bootstrap resampling of Cliff's delta, seed 0, 2,000 resamples, GSEA permutation, seed 0, 1,000 permutations). The public resources used in this study are dynamically updated rather than archival, including UCSC Xena, TIMER2.0, MSigDB release files accessed via stable URLs, the live cBioPortal API, and the cptac package, which retrieves data at runtime. The manifest records the SHA-256 checksum, file size, and retrieval date of each input file, together with the dimensions and checksum of each output table, as well as the resolved versions of all software packages that could influence numerical results. The manifest is distributed alongside the source code to ensure computational reproducibility. Differences between a future run and the values reported here can therefore be localized to a specific input, package version, or code change. Statistical utility functions were verified against brute-force implementations and against *statsmodels* before use, including a null-calibration check confirming a 5.7% type-I error rate for the trend test under the null and confirmation that a non-monotonic pattern is not reported as a trend. The complete workflow, environment specification, verification suite, run manifest, and all intermediate result tables are openly available at https://github.com/ApexBlue11/NCL-PanCancer-Analysis.

3. Results

3.1. Nucleolin is overexpressed across most cancers, and the choice of normal comparator matters

Across 29 cancers with an acceptable GTEx comparator, NCL was significantly overexpressed in tumors in 24 (all $q < 0.05$), showing no significant difference in four (adrenocortical carcinoma [ACC], cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC], bladder urothelial carcinoma [BLCA], and kidney chromophobe [KICH]) and significantly reduced expression in one (OV; Cliff's delta −0.26, 95% CI −0.36 to −0.17, $q = 1.1\times10^{-4}$; Figure 1a, Table 1). Effect sizes were large in several cancers, brain lower grade glioma (LGG; delta +0.93, 95% CI +0.91 to +0.94, $q = 1.3 \times 10^{-199}$), cholangiocarcinoma (CHOL; +0.92, +0.85 to +0.97), testicular germ cell tumors (TGCT; +0.91, +0.84 to +0.96), rectum adenocarcinoma (READ; +0.90, +0.82 to +0.96), pancreatic adenocarcinoma (PAAD; +0.86, +0.80 to +0.91), and COAD (+0.86, +0.81 to +0.90).

An important qualification emerged when the same comparison was repeated against TCGA adjacent normal tissue (Figure 1b). Twelve cancers were significant against both comparators; 10 showed the same direction, but 2 did not. In kidney renal papillary cell carcinoma (KIRP), NCL was higher than in GTEx kidney (delta +0.63, $q = 5.1 \times 10^{-8}$), yet lower than in adjacent normal kidney (delta −0.29, $q = 1.1 \times 10^{-2}$). The same reversal occurred in thyroid carcinoma (THCA; +0.36 versus −0.27), and the discordance extended to KICH (+0.19, not significant, versus −0.80, $q = 1.6 \times 10^{-8}$) and kidney renal clear cell carcinoma (KIRC; +0.77 versus −0.06, not significant). The kidney and thyroid are therefore cancers in which the reported direction of NCL dysregulation is determined by the reference tissue chosen rather than by the tumor. The adjacent normal kidney is not histologically inert; it is subject to field effects, inflammation, and, in nephrectomy specimens, ischemic injury, whereas GTEx tissue is post-mortem, and ribosome-biogenesis genes such as *NCL* are sensitive to both. We therefore report both comparators throughout and restrict confident claims of overexpression to cancers where they agree.

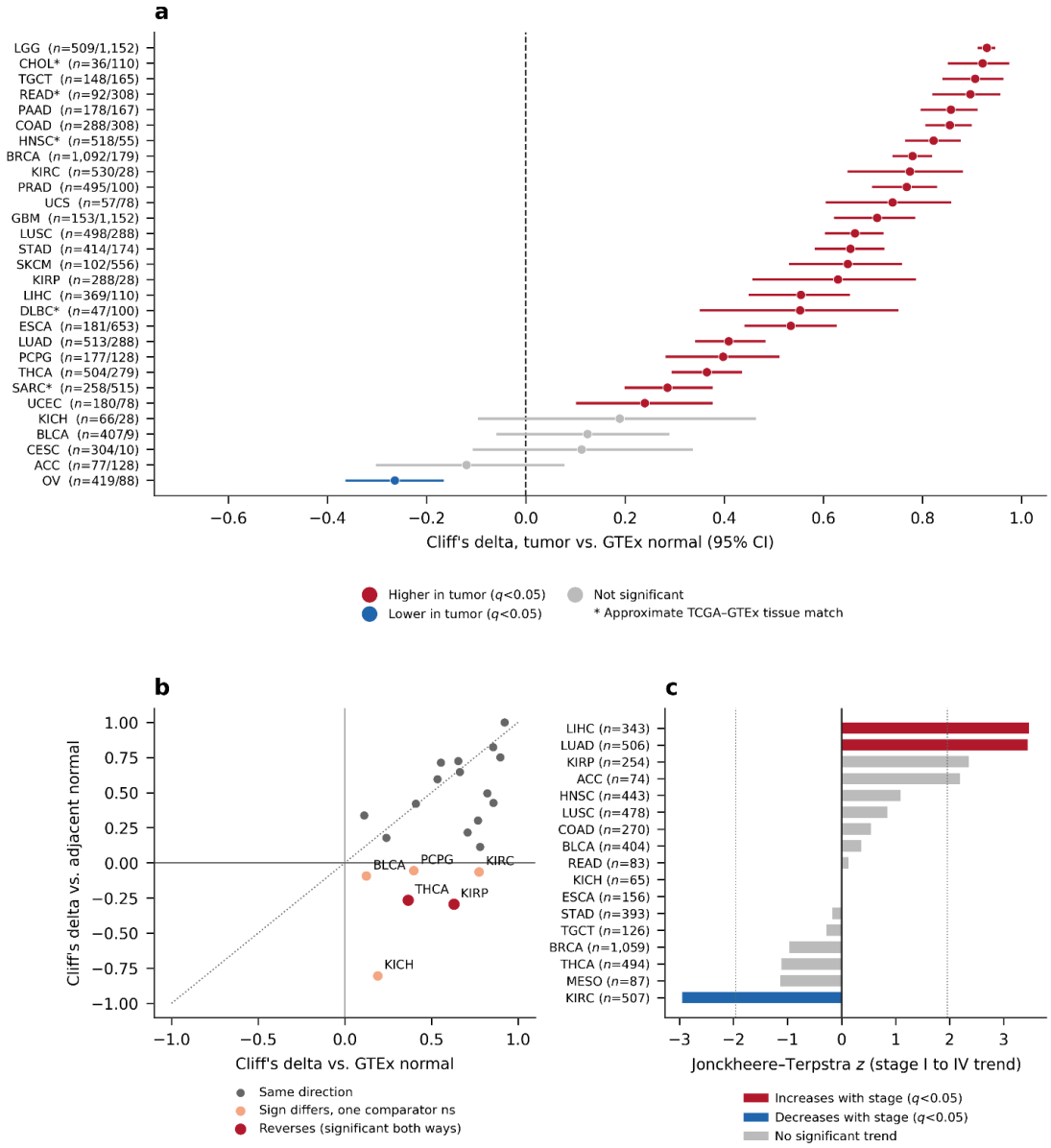


Figure 1. Nucleolin (NCL) expression across cancers and pathological stage. (a) Cliff's delta with 95% bootstrap confidence intervals (CIs) for NCL expression in tumors versus Genotype-Tissue Expression project (GTEx) normal tissues, ordered by effect size. Point size indicates cohort size. Filled markers indicate $q < 0.05$ after Benjamini–Hochberg correction. (b) Comparison of effect sizes obtained using GTEx normal tissues versus The Cancer Genome Atlas (TCGA) adjacent normal tissues. Points deviating from the diagonal indicate comparator-dependent results; KIRP, KICH, KIRC, and THCA are labeled. (c) Standardized Jonckheere–Terpstra trend statistic for NCL expression across pathological stages I–IV; positive values indicate increasing expression with stage. Only cancers with ≥3 stage groups, each containing ≥5 patients, are shown. Sample sizes for each stage group are indicated.

This has practical consequences for how NCL is described in individual cancers. Reports that NCL is reduced in renal or thyroid tumors are reconcilable with reports that it is elevated; both can be correct, depending on the reference tissue used. In cancers where the two comparators agree, including READ (delta +0.90, $q = 2.2 \times 10^{-38}$) and skin cutaneous melanoma (SKCM; +0.65, $q = 4.3 \times 10^{-25}$), elevation is unambiguous.

3.2. A stage-ordered increase in nucleolin is the exception, not the rule

Across 17 cancers with sufficient staging data, a significant

Table 1. Nucleolin differential expression by cancer type

Cancer	Tumor (*n*)	GTEx (*n*)	Cliff's delta (95% CI)	*q*	Tissue match	Adjacent-normal delta	Adjacent *q*
LGG	509	1,152	+0.93 (+0.91, +0.94)	1.3e−199	Good	–	–
CHOL	36	110	+0.92 (+0.85, +0.97)	2.1e−16	Approximate	+1.00	1.2e−05
TGCT	148	165	+0.91 (+0.84, +0.96)	6.9e−43	Good	–	–
READ	92	308	+0.90 (+0.82, +0.96)	2.2e−38	Approximate	+0.75	0.0002
PAAD	178	167	+0.86 (+0.80, +0.91)	1.7e−42	Good	+0.43	0.2044
COAD	288	308	+0.86 (+0.81, +0.90)	9.4e−72	Good	+0.82	3.2e−16
HNSC	518	55	+0.82 (+0.77, +0.87)	2.3e−23	Approximate	+0.50	1.7e−07
BRCA	1092	179	+0.78 (+0.74, +0.82)	6.0e−62	Good	+0.11	0.0670
KIRC	530	28	+0.77 (+0.65, +0.88)	7.3e−12	Good	−0.06	0.4332
PRAD	495	100	+0.77 (+0.70, +0.83)	2.3e−33	Good	+0.30	0.0007
UCS	57	78	+0.74 (+0.61, +0.86)	3.9e−13	Good	–	–
GBM	153	1152	+0.71 (+0.62, +0.78)	2.3e−45	Good	+0.22	0.4717
LUSC	498	288	+0.66 (+0.61, +0.72)	1.8e−53	Good	+0.65	3.3e−13
STAD	414	174	+0.65 (+0.59, +0.72)	1.5e−35	Good	+0.73	2.6e−12
SKCM	102	556	+0.65 (+0.53, +0.76)	4.3e−25	Good	–	–
KIRP	288	28	+0.63 (+0.46, +0.78)	5.1e−08	Good	−0.29	0.0106
LIHC	369	110	+0.55 (+0.45, +0.65)	1.9e−18	Good	+0.71	2.9e−15
DLBC	47	100	+0.55 (+0.35, +0.75)	8.5e−08	Approximate	–	–
ESCA	181	653	+0.53 (+0.44, +0.62)	9.4e−28	Good	+0.60	0.0007
LUAD	513	288	+0.41 (+0.34, +0.48)	1.3e−21	Good	+0.42	3.4e−07
PCPG	177	128	+0.40 (+0.28, +0.51)	4.3e−09	Good	−0.05	0.8830
THCA	504	279	+0.36 (+0.30, +0.43)	4.8e−17	Good	−0.27	0.0014
SARC	258	515	+0.29 (+0.20, +0.37)	1.4e−10	Approximate	–	–
UCEC	180	78	+0.24 (+0.10, +0.37)	0.0026	Good	+0.18	0.2133
KICH	66	28	+0.19 (−0.09, +0.46)	0.1619	Good	−0.80	1.6e−08
BLCA	407	9	+0.12 (−0.06, +0.29)	0.5433	Good	−0.09	0.5146
CESC	304	10	+0.11 (−0.11, +0.33)	0.5473	Good	+0.34	0.3881
ACC	77	128	−0.12 (−0.30, +0.08)	0.1619	Good	–	–
OV	419	88	−0.26 (−0.36, −0.17)	0.0001	Good	–	–

Notes: Differential expression of nucleolin in tumors versus Genotype-Tissue Expression project (GTEx) and The Cancer Genome Atlas (TCGA) adjacent normal tissues, with sample sizes, median expression, Cliff's delta with 95% confidence intervals (CIs), Hedges' *g*, and false discovery rate (FDR)-adjusted *p*-values for each comparison.

monotonic trend across stages I–IV was present in only three, and in one of these, the direction was negative: liver hepatocellular carcinoma (LIHC; Jonckheere–Terpstra z = +3.47, q = 0.005, n = 343), LUAD (z = +3.45, q = 0.005, n = 506), and, decreasing, KIRC (z = −2.95, q = 0.018, n = 507) (Figure 1c, Table 2). Fourteen cancers showed no stage-ordered trend. The corresponding stage IV versus stage I effect sizes were modest, even when the trend was significant: LUAD Cliff's delta +0.26 (95% CI +0.05 to +0.45) and KIRC −0.13 (−0.28 to +0.00).

These results warrant methodological caution. Stage-stratified expression plots produced by several widely used web tools annotate each stage against normal tissue rather than against the preceding stage. Such annotations can be uniformly significant even when expression is flat or falling across stages, and reading them as evidence of stage-wise progression will systematically overstate how often expression tracks the stage. A trend test on the ordered groups is required, and when applied to NCL, it supports a stage-ordered increase in a small minority of cancers. Even in LIHC, where the trend is significant, stage IV includes only six patients, so the trend is driven by stages I–III, and the late-versus-early effect size is not distinguishable from zero (delta −0.01, 95% CI −0.53 to +0.53).

3.3. Nucleolin is rarely an independent prognostic factor

Univariate analyses reproduced the pattern that has led NCL to be described as a prognostic biomarker; NCL was associated with OS in 6 of 24 cancers, with DSS in 7 of 21, and with PFS in 8 of 29 ($q < 0.05$).

After adjustment for age, sex, stage, and grade, and FDR correction, these patterns largely did not survive (Figure 2, Table 3). NCL remained independently associated with outcome in a single cancer per endpoint, KIRP for OS

Table 2. Association between nucleolin expression and pathological stage

Cancer	*n*	Stages I/II/III/IV	Kruskal–Wallis *q*	Trend *z*	Trend *q*	Stage IV vs. I delta (95% CI)
LIHC	343	167/86/84/6	0.0376	+3.47	0.0048	−0.01 (−0.53, +0.53)
LUAD	506	274/123/82/27	0.0376	+3.45	0.0048	+0.26 (+0.05, +0.45)
KIRC	507	248/54/124/81	0.0376	−2.95	0.0182	−0.13 (−0.28, +0.00)
KIRP	254	167/22/50/15	0.1642	+2.35	0.0790	+0.40 (+0.09, +0.67)
ACC	74	9/36/15/14	0.2717	+2.19	0.0961	+0.59 (+0.19, +0.90)
MESO	87	10/16/45/16	0.1365	−1.14	0.5858	−0.49 (−0.81, −0.06)
THCA	494	280/51/109/54	0.7103	−1.12	0.5858	−0.06 (−0.21, +0.09)
HNSC	443	27/73/78/265	0.2570	+1.09	0.5858	+0.09 (−0.11, +0.30)
BRCA	1,059	180/612/248/19	0.3340	−0.97	0.6291	−0.28 (−0.48, −0.05)
LUSC	478	233/156/82/7	0.4806	+0.85	0.6762	+0.02 (−0.50, +0.50)
COAD	270	44/108/81/37	0.2570	+0.54	0.9054	−0.10 (−0.34, +0.15)
BLCA	404	2/129/140/133	0.1994	+0.36	0.9933	+0.03 (−0.11, +0.18)
TGCT	126	100/12/14/0	0.7103	−0.28	0.9933	+0.05 (−0.20, +0.30)
STAD	393	57/122/173/41	0.8081	−0.17	0.9933	−0.00 (−0.24, +0.24)
READ	83	12/25/31/15	0.7103	+0.13	0.9933	−0.11 (−0.57, +0.36)
KICH	65	20/25/14/6	0.4806	+0.01	0.9933	+0.27 (−0.25, +0.72)
ESCA	156	18/75/54/9	0.7103	−0.01	0.9933	+0.20 (−0.30, +0.67)

Note: Per-stage sample sizes, Kruskal–Wallis and Jonckheere–Terpstra test statistics, and late-versus-early stage effect sizes for nucleolin expression across cancer types.

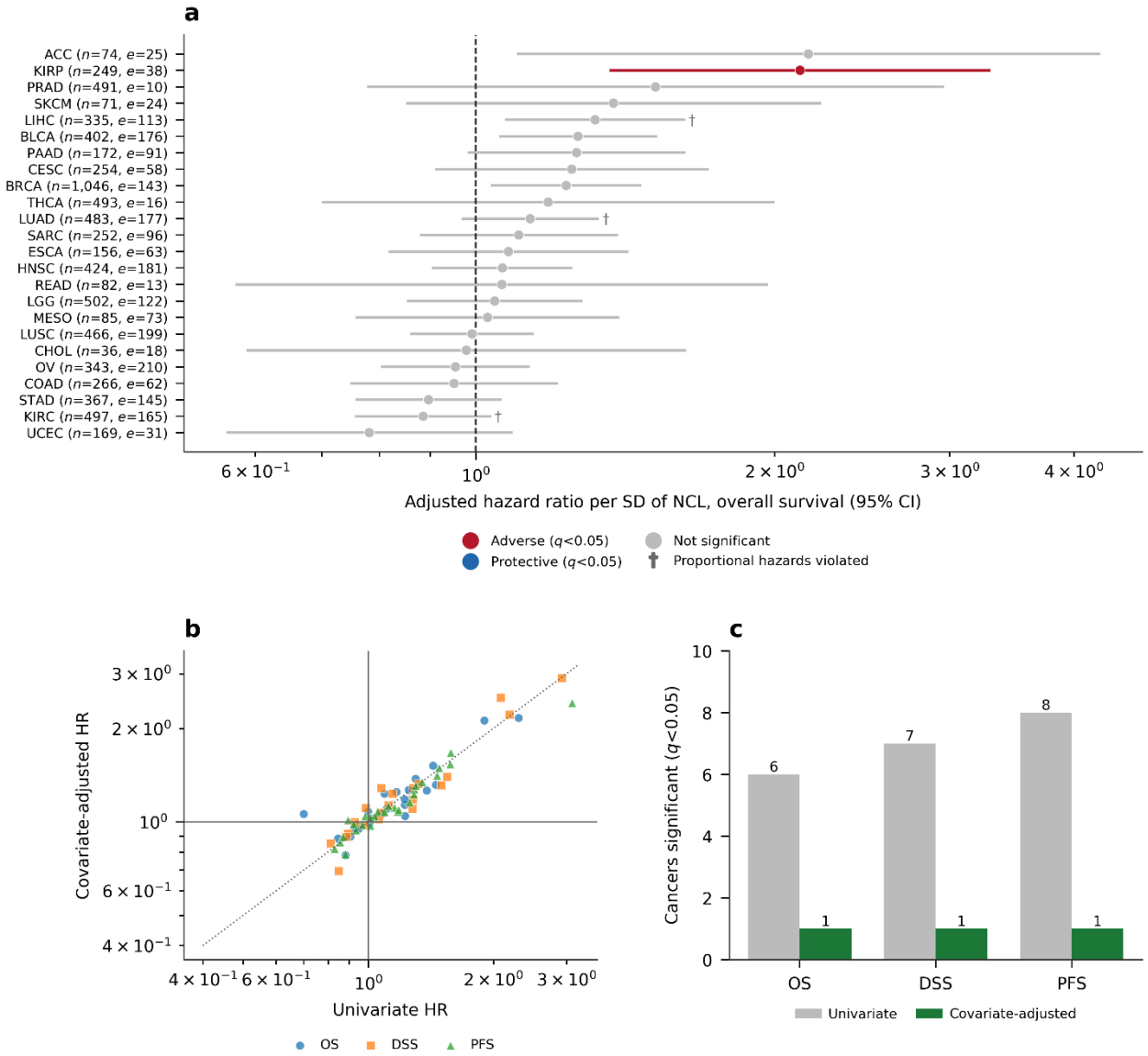


Figure 2. Survival analysis showing the association between nucleolin (NCL) expression and patient survival across cancers. (a) Hazard ratios (HRs) per standard deviation (SD) increase in NCL expression for overall survival (OS) from multivariable Cox models adjusted for age, sex, stage, and grade, with 95% confidence intervals (CIs). Each row is annotated with the number of patients and events included in the model. Red indicates an adverse association significant at $q < 0.05$, blue indicates a protective association, and gray indicates no significant association. A dagger marks models in which the proportional-hazards assumption was violated for the NCL term (Schoenfeld residuals, $p < 0.05$); these HRs reflect a time-varying effect and should not be interpreted as constant risk multipliers. (b) Univariate versus covariate-adjusted HRs for each cancer and endpoint. Points below the diagonal indicate cancers in which the apparent prognostic association is attenuated after adjustment. (c) Number of cancers reaching $q < 0.05$ before and after adjustment, by endpoint. Full per-model results, including the covariates available in each cohort, are provided in Table 3.
Abbreviations: DSS: Disease-specific survival; PFS: Progression-free survival.

(HR 2.12 per SD, 95% CI 1.37–3.29, $q = 0.019$, 38 events) and DSS (HR 2.90, 1.60–5.27, $q = 0.0099$, 25 events), and ACC for PFS (HR 2.41, 1.40–4.16, $q = 0.044$, 38 events). The proportional-hazards assumption held for the NCL term in both, supporting the interpretation of these HRs as constant effects. Across all endpoints, the assumption was violated in 10 of 74 models, consistently in KIRC and LUAD; the corresponding HRs are shown in Table 3 and should not be interpreted as time-invariant.

The attenuation between the univariate and adjusted models is substantial and warrants explicit mention, as transcript-level NCL has been reported to be prognostic

Table 3. Multivariable Cox regression of nucleolin and survival outcomes

Cancer	*n*	Events	Adjusted HR (95% CI)	Adjusted *q*	Univariate *q*	Covariates	PH violated
ACC	74	25	2.16 (1.10, 4.24)	0.1184	0.0441	Age, sex, stage	No
KIRP	249	38	2.12 (1.37, 3.29)	0.0187	0.0151	Age, sex, stage	No
PRAD	491	10	1.52 (0.78, 2.95)	0.4058	0.4819	Age	No
SKCM	71	24	1.38 (0.85, 2.22)	0.3946	0.4819	Age, sex, stage	No
LIHC	335	113	1.32 (1.07, 1.62)	0.0789	0.0061	Age, sex, stage, grade	Yes
BLCA	402	176	1.27 (1.06, 1.52)	0.0789	0.0487	Age, sex, stage	No
PAAD	172	91	1.26 (0.98, 1.62)	0.2668	0.0441	Age, sex, stage, grade	No
CESC	254	58	1.25 (0.91, 1.71)	0.3946	0.4819	Age, grade	No
BRCA	1,046	143	1.23 (1.04, 1.46)	0.0995	0.4819	Age, sex, stage	No
THCA	493	16	1.18 (0.70, 1.99)	0.7840	0.6460	Age, sex, stage	No
LUAD	483	177	1.13 (0.97, 1.33)	0.3650	0.0456	Age, sex, stage	Yes
SARC	252	96	1.10 (0.88, 1.39)	0.6652	0.5690	Age, sex	No
ESCA	156	63	1.08 (0.82, 1.42)	0.7840	0.9989	Age, sex, stage	No
HNSC	424	181	1.06 (0.91, 1.25)	0.7197	0.4819	Age, sex, stage, grade	No
READ	82	13	1.06 (0.57, 1.96)	0.9320	0.3817	Age, sex, stage	No
LGG	502	122	1.04 (0.85, 1.28)	0.8127	0.1746	Age, sex, grade	No
MESO	85	73	1.03 (0.76, 1.39)	0.9320	0.8915	Age, sex, stage	No
LUSC	466	199	0.99 (0.86, 1.14)	0.9320	0.9513	Age, sex, stage	No
CHOL	36	18	0.98 (0.59, 1.62)	0.9320	0.8915	Age, sex, stage, grade	No
OV	343	210	0.95 (0.81, 1.13)	0.7840	0.4819	Age, grade	No
COAD	266	62	0.95 (0.75, 1.21)	0.8127	0.8465	Age, sex, stage	No
STAD	367	145	0.90 (0.76, 1.06)	0.3946	0.4819	Age, sex, stage, grade	No
KIRC	497	165	0.89 (0.76, 1.03)	0.3650	0.0501	Age, sex, stage, grade	Yes
UCEC	169	31	0.78 (0.56, 1.09)	0.3781	0.5690	Age, grade	No

Note: Associations between nucleolin (NCL) expression and overall, disease-specific, and progression-free survival, reported as hazard ratios (HRs) per standard deviation increase in NCL expression with 95% confidence intervals (CIs), covariates included, numbers of patients and events, false discovery rate (FDR)-adjusted *p*-values, and proportional-hazards (PH) test results.

in several of these cancers. In LIHC, the univariate association is strong (q = 0.006) but does not remain significant after adjustment (HR 1.32, 95% CI 1.07–1.62, q = 0.079). In HNSC, the adjusted estimate is null (HR 1.06, 95% CI 0.91–1.25, q = 0.72). KICH, frequently cited in this context, has too few events in TCGA to support a multivariable model. The general pattern is that transcript-level NCL's apparent prognostic value is largely accounted for by variables already used clinically.

3.4. Nucleolin marks an immune-excluded tumor phenotype

Across 3,910 cancer × cell type × algorithm tests, 1,229 purity-adjusted associations were significant at $q < 0.05$. The dominant pattern was negative. NCL correlated inversely with the xCell microenvironment score across all

20 cancers, with the association significant for the immune score in 18 of 20 (positive only in prostate adenocarcinoma [PRAD] and UVM) and for the stromal score in 14 of 15 (positive only in KIRC) (Figure 3b, Table 4). The strongest inverse immune-score associations were in GBM (ρ = −0.38, 95% CI −0.52 to −0.22, n = 151), SKCM (−0.36,

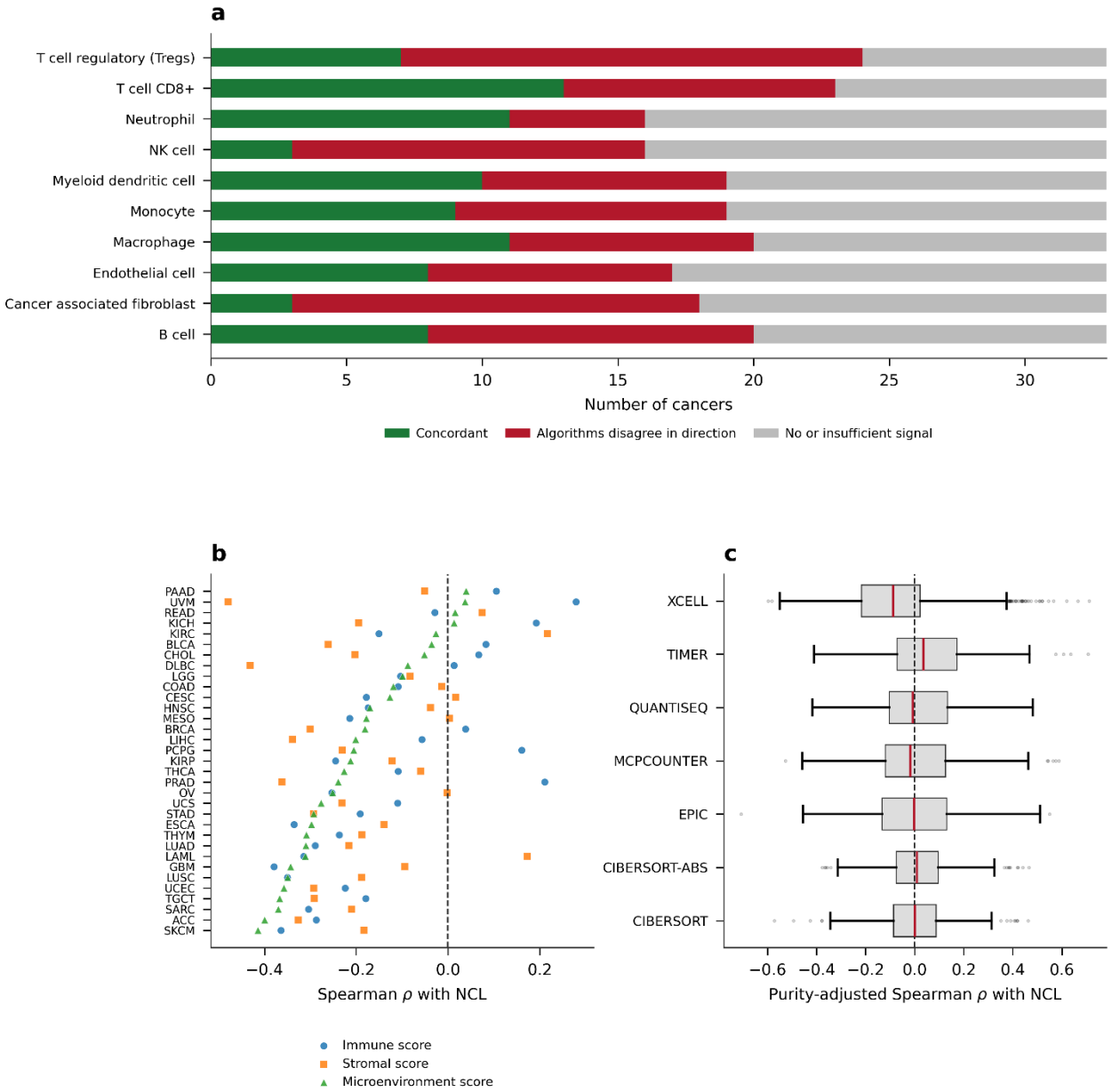


Figure 3. Immune infiltration associations with nucleolin (NCL) expression across cancers and deconvolution algorithms. (a) For each of 10 canonical cell types, the number of cancers in which algorithms concordantly detect an association with NCL (≥2 algorithms significant at $q < 0.05$ after purity adjustment, with no disagreement in sign), the number in which significant associations of opposite sign are returned for the same cell type, and the number with no or insufficient signal. Each bar represents the 33 cancers analyzed. The number of algorithms available varied by cell type, ranging from two to seven; cancer-associated fibroblasts, endothelial cells, and macrophages had the fewest (up to three), whereas $CD8^+$ T cells had the most (up to seven). Algorithm counts for each cancer and cell type are provided in Table S6. (b) Correlation between NCL expression and the xCell immune, stromal, and microenvironment scores in each cancer, with cancers ordered by microenvironment score. Points to the left of zero indicate that NCL-high tumors are associated with lower immune or stromal scores. (c) Distribution of purity-adjusted partial Spearman correlations between NCL and infiltration estimates for each of the seven deconvolution algorithms, pooled across cancers and cell types. The spread illustrates the extent to which estimated associations depend on the algorithm used.

Table 4. Nucleolin correlations with tumor microenvironment and genomic features

Measure	Cancers tested	Significant ($q < 0.05$)	Positive	Negative	Median ρ	Range of significant ρ
Immune score	33	20	2	18	−0.174	−0.38 to +0.28
Stromal score	33	15	1	14	−0.188	−0.48 to +0.22
Microenvironment score	33	20	0	20	−0.212	−0.41 to −0.17
TMB	33	7	6	1	+0.064	−0.12 to +0.34
MSI (MANTIS)	32	10	5	5	+0.035	−0.45 to +0.30
Aneuploidy	33	7	5	2	+0.044	−0.18 to +0.24
FGA	33	11	9	2	+0.118	−0.14 to +0.26

Note: Correlations between nucleolin expression and immune, stromal, and microenvironment scores, tumor mutational burden (TMB), Microsatellite Analysis for Normal-Tumor Instability (MANTIS) microsatellite instability score, aneuploidy score, and fraction of genome altered (FGA).

−0.53 to −0.17, n = 102), LUSC (−0.35, −0.43 to −0.27, n = 498), and esophageal carcinoma (ESCA; −0.34, −0.47 to −0.19, n = 181). The microenvironment score showed the same pattern (SKCM −0.41, −0.57 to −0.23; ACC −0.40, −0.58 to −0.18; sarcoma [SARC] −0.37, −0.48 to -0.25), as did the stromal score (UVM −0.48, −0.64 to −0.28; PRAD −0.36, −0.44 to −0.28; LIHC −0.34, −0.43 to −0.24). Only PRAD and UVM showed positive immune-score associations.

Cell-type-level results were consistent with this. The values below are the median purity-adjusted correlation across all algorithms that resolve the cell type, with the range of estimates that reached significance in parentheses. Among the strongest concordant associations were reduced $CD8^+$ T cells in SKCM (median ρ = −0.32, significant estimates −0.43 to −0.25 across 5 of 7 algorithms, n = 102) and ACC (−0.31, −0.34 to −0.31, 4 of 7, n = 77), reduced macrophages in GBM (−0.46, −0.46 to −0.46, n = 151), KIRP (−0.39, −0.44 to −0.39, n = 288), LUSC (−0.33, −0.36 to −0.33, n = 498) and OV (−0.32, −0.39 to −0.31, n = 300), and reduced CAFs in PRAD (−0.46, −0.55 to −0.46, n = 495). NCL-high tumors are thus, in general, immune- and stroma-poor, a phenotype consistent with a proliferative, tumor-cell-rich compartment.

The strongest positive cell-type association we observed illustrates why both purity adjustment and multi-algorithm reporting matter. NCL correlates with neutrophil infiltration in THCA (ρ = 0.54) when estimated using the MCP-counter alone, without purity adjustment. Across all six algorithms that resolve neutrophils, and after purity adjustment, the median estimate fell to +0.30. Restricting to the four algorithms that reach significance yielded a median of +0.39, with estimates ranging from +0.25 to +0.54 (n = 500). The association is real and directionally consistent, but its magnitude varies roughly two-fold with these choices.

3.5. Deconvolution algorithms frequently disagree

Of 330 cancer × canonical cell type combinations, only 83 (25%) were concordant, meaning at least two algorithms gave a significant purity-adjusted association with no disagreement in sign. In 109 combinations (33%), algorithms produced significant estimates in the opposite direction for nominally the same cell type in the same cancer (Figure 3a).

This has a direct methodological implication. A pan-cancer immune association supported by a single deconvolution algorithm has roughly a one-in-three chance of being contradicted by another algorithm applied to the same data. Results in this section are therefore reported only where multiple algorithms agree. We regard single-algorithm immune associations, and particularly analyses in which the algorithm differs between comparisons, as insufficiently robust to support biological interpretation.

3.6. Nucleolin correlates with tumor-intrinsic immunosuppressive ligands, not T-cell checkpoints

Testing 495 cancer × gene associations produced a clearly structured result (Figure 4, Tables 5 and S7). After adjustment for both tumor purity and proliferation, the most consistent association, by a wide margin, was with CD276 (B7-H3), which was positive and significant in

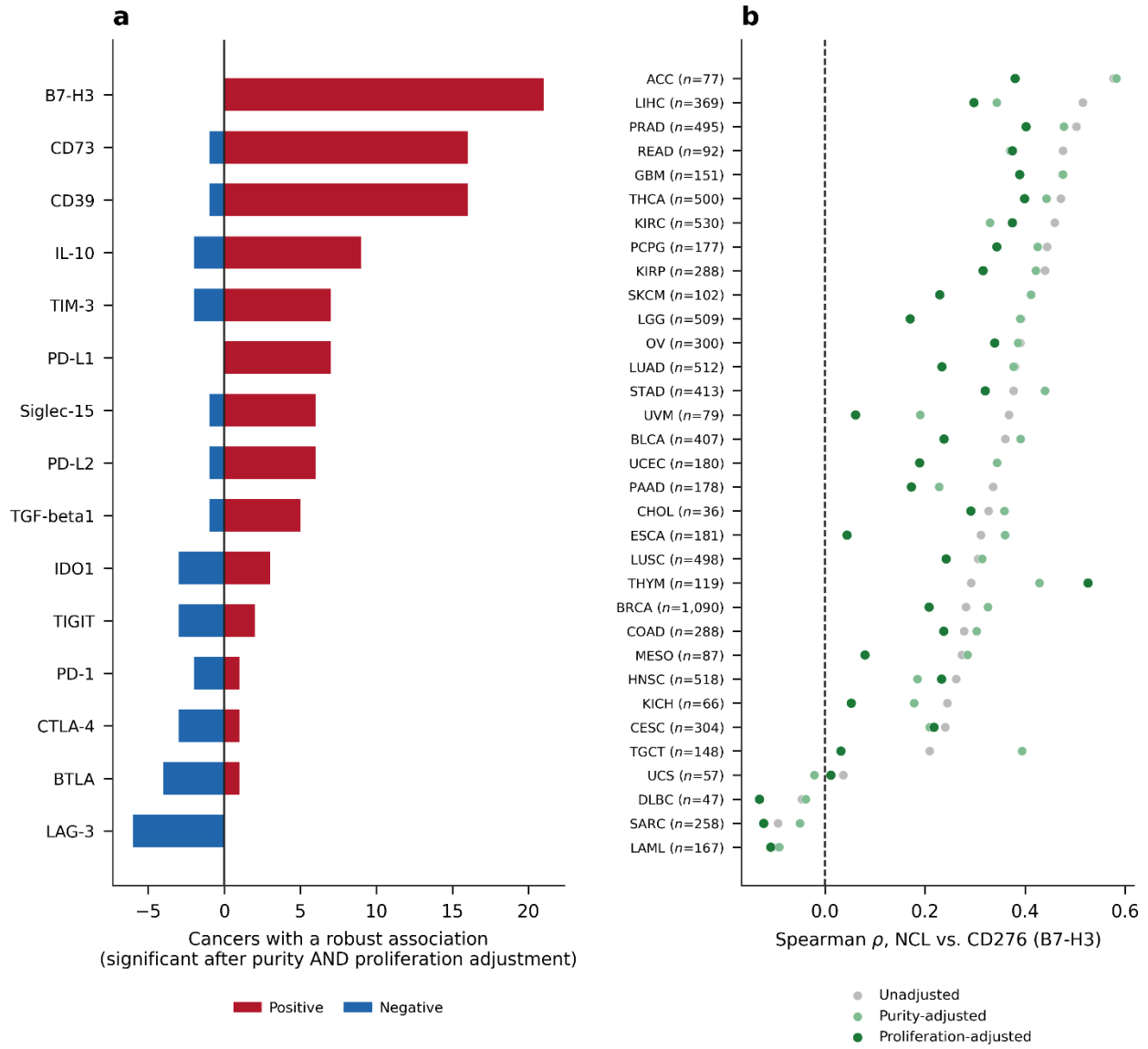


Figure 4. Immune checkpoint associations across cancers, highlighting nucleolin (NCL) correlations with tumor purity and proliferation adjustment. (a) Number of the 33 cancers in which each molecule shows a robust association with NCL, defined as significant at $q < 0.05$ after adjustment for both tumor purity and proliferation. Bars extend to the right for positive associations and to the left for negative associations; molecules with bars on both sides, therefore, show inconsistent directions of association across cancers. Cancers not reaching significance are not counted. (b) Per-cancer associations for CD276 (B7-H3), showing unadjusted, purity-adjusted, and proliferation-adjusted correlations, with cohort size indicated for each cancer. The ordering indicates that adjustment attenuates, but does not abolish, the association. Cancers in which the B7-H3 association did not meet the robustness criterion are described in Section 3.6.

21 of 33 cancers, with no cancer showing a significant negative association (median ρ = +0.34). The strongest were ACC (ρ = +0.58, 95% CI +0.39 to +0.72, n = 77), LIHC (+0.52, +0.43 to +0.59, n = 369), PRAD (+0.50, +0.43 to +0.57, n = 495), READ (+0.48, +0.29 to +0.63, n = 92), GBM (+0.47, +0.33 to +0.60, n = 151), THCA (+0.47, +0.40 to +0.54, n = 500), and KIRC (+0.46, +0.38 to +0.53, n = 530). After adjustment for both purity and proliferation, the association was attenuated but retained, with partial correlations across the 21 robust cancers spanning +0.17 to +0.53. The 12 cancers in which the association was not robust were CHOL, lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), ESCA, KICH, acute myeloid leukemia (LAML), MESO, PAAD, SARC, SKCM, TGCT, uterine carcinosarcoma (UCS), and UVM, a group weighted toward the smallest cohorts and

the non-carcinoma histologies. Associations with the two principal components of the adenosine axis followed the same pattern: *NT5E* (CD73) was robust in 17 of 33 cancers, 16 of them positive (proliferation-adjusted ρ up to +0.55; the single negative was COAD), and *ENTPD1* (CD39) in 17 of 33, again 16 positives (up to +0.56; the single negative was LAML).

By contrast, the T-cell checkpoint receptors were weakly and inconsistently associated, and when associations reached significance, they were as often negative as positive. *PDCD1* was robust in 3 of 33 cancers (positive in BLCA, negative in SKCM and THCA) and *CTLA4* in 4 of 33 (positive in BLCA, negative in LUSC, OV, and THCA; median ρ = +0.02). *TIGIT* was robust in 5 (proliferation-adjusted ρ = −0.21 to +0.21), *LAG3* in 6 (all negative: KIRC, OV, PRAD, SARC, THCA, UCS; −0.55 to −0.17), and *TIM3* in 9 of 33 (median ρ = +0.05, −0.25 to +0.35). For *CTLA4* and *PDCD1*, the robust estimates ranged from −0.25 to +0.15 and from −0.27 to +0.12, respectively, so even when significant, they are small and of inconsistent sign. *CD274* was robust in 7 of 33 (all positive: BLCA, BRCA, KICH, KIRC, LIHC, PAAD, pheochromocytoma and paraganglioma [PCPG]; median ρ = +0.17), and *TGFB1* in 6 of 33.

This distinction is biologically coherent. B7-H3, CD73, and CD39 are expressed by tumor cells themselves, whereas PD-1, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), lymphocyte activation gene 3 (LAG-3), and T-cell immunoreceptor with Ig and ITIM domains (TIGIT) are expressed principally by infiltrating lymphocytes. A tumor-cell-expressed gene, such as *NCL*, would be expected to covary with the former rather than the latter. Given the inverse relationship between NCL and immune content described above, it correlates negatively rather than positively with lymphocyte-restricted transcripts. That the observed pattern matches this expectation and

Table 5. Nucleolin correlates with immune checkpoint and immunomodulatory genes

Molecule	Gene	Median ρ	Significant unadjusted	Robust after both adjustments	Positive/Negative	Robust ρ range
B7-H3	*CD276*	+0.344	27	21	21/0	+0.17 to +0.53
CD39	ENTPD1	+0.168	21	17	16/1	−0.24 to +0.56
CD73	*NT5E*	+0.232	19	17	16/1	−0.14 to +0.55
IL-10	*IL10*	+0.136	17	11	9/2	−0.25 to +0.26
TIM-3	*HAVCR2*	+0.048	13	9	7/2	−0.25 to +0.35
PD-L2	*PDCD1LG2*	+0.131	14	7	6/1	−0.11 to +0.34
Siglec-15	*SIGLEC15*	+0.096	10	7	6/1	−0.16 to +0.48
PD-L1	*CD274*	+0.169	14	7	7/0	+0.17 to +0.39
TGF-beta1	*TGFB1*	+0.105	9	6	5/1	−0.23 to +0.26
IDO1	*IDO1*	+0.040	12	6	3/3	−0.20 to +0.20
LAG-3	*LAG3*	−0.010	8	6	0/6	−0.55 to −0.17
TIGIT	*TIGIT*	+0.050	10	5	2/3	−0.21 to +0.21
BTLA	*BTLA*	+0.056	10	5	1/4	−0.32 to +0.19
CTLA-4	*CTLA4*	+0.017	10	4	1/3	−0.25 to +0.15
PD-1	*PDCD1*	+0.033	9	3	1/2	−0.27 to +0.12

Note**:** Unadjusted, tumor-purity-adjusted, and proliferation-adjusted correlations between nucleolin expression and immune checkpoint and immunomodulatory genes across cancer types.

persists after adjustment for proliferation argues against it being merely an artifact of tumor cellularity.

It is worth stating the negative result explicitly because NCL has been discussed in the context of classical checkpoint biology. Across 33 cancers, we found no consistent association between NCL and PD-1, CTLA-4, LAG-3, or TIGIT, and only a weak one with PD-L1 and transforming growth factor beta 1 (TGF-β1). The robust association is with B7-H3, which, to our knowledge, has not previously been linked to NCL.

3.7. Nucleolin is largely unrelated to tumor mutational burden and microsatellite instability

Nucleolin showed little relationship with genomic instability measures. Tumor mutational burden was significantly associated in 7 of 33 cancers (ACC, BRCA, LGG, LUAD, PRAD, stomach adenocarcinoma [STAD], THCA; median ρ = +0.06 overall, significant estimates −0.12 to +0.34), MANTIS MSI scores in 10 of 32 (BRCA, CESC, DLBC, KIRC, LGG, LUSC, OV, PRAD, STAD, THCA; median ρ = +0.04, significant estimates −0.45 to +0.30, with equal numbers positive and negative), and aneuploidy scores in 7 of 33 (median ρ = +0.04, −0.18 to +0.24). The fraction of the genome altered was more often associated with higher ρ values (11 of 33, median ρ = +0.12). NCL, therefore, does not track the mutational processes that predict immune checkpoint blockade response, thereby further distinguishing its immune association from a conventional immunogenicity signal (Table S8).

3.8. Independent proteomic validation

In nine CPTAC cohorts with adjacent normal tissue, NCL protein was significantly more abundant in the tumor in seven (Figure 5, Table 6). LUAD (Cliff's delta +0.99, $q = 5.1 \times 10^{-35}$, 102 matched pairs), LUSC (+0.97), GBM (+0.99), OV (+0.91), COAD (+0.89), ccRCC (+0.65), and HNSC (+0.52). Paired analyses within patients yielded the same conclusion wherever available (Table S9).

Two cohorts diverged from the transcript-level results and are reported as such. In UCEC, NCL protein did not

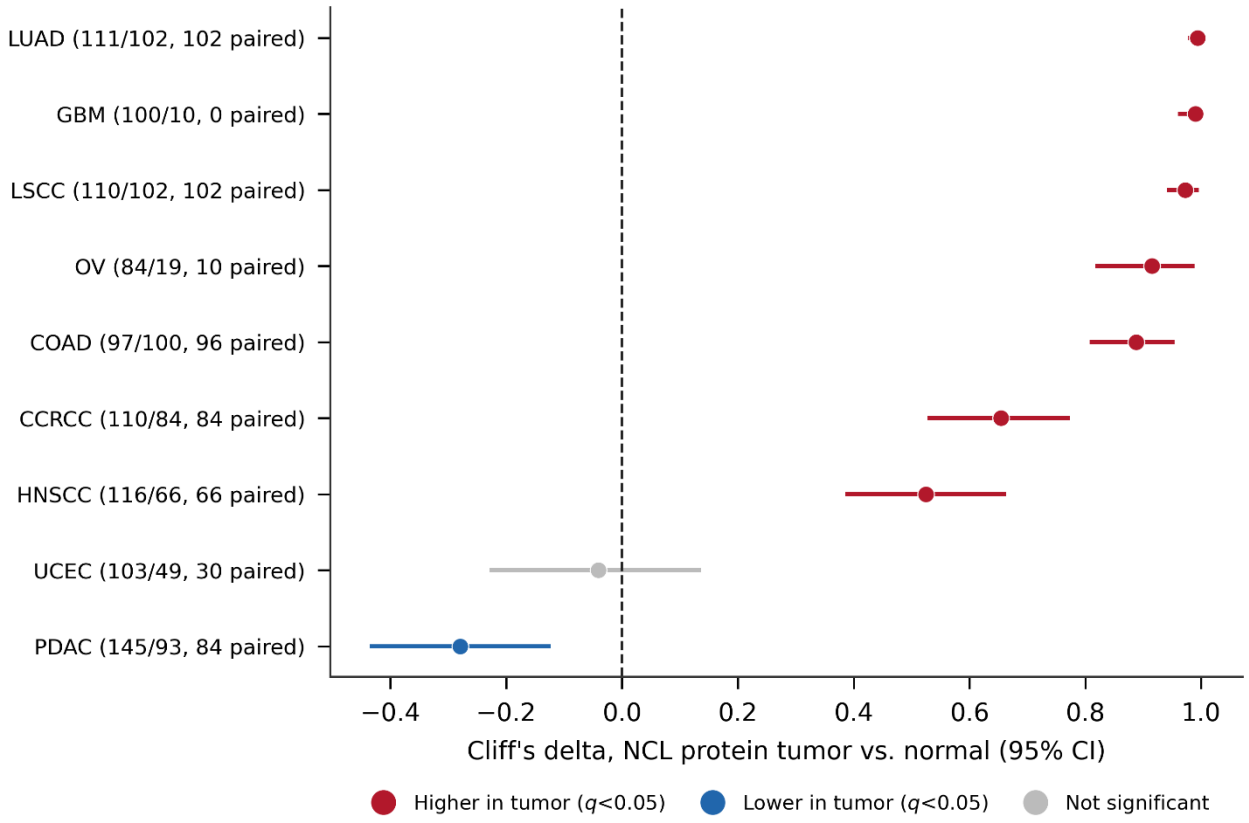


Figure 5. Cliff's delta with 95% bootstrap confidence intervals (CIs) for nucleolin (NCL) protein abundance in tumor versus adjacent normal tissues across nine Clinical Proteomic Tumor Analysis Consortium (CPTAC) cohorts. Numbers of tumors, normal samples, and matched pairs are indicated for each cohort. Cohorts discordant with the corresponding transcript-level result (PDAC and UCEC) are indicated.

Table 6. Independent validation of nucleolin protein abundance in nine CPTAC cohorts

CPTAC cohort	Tumor (*n*)	Normal (*n*)	Paired (*n*)	Cliff's delta (95% CI)	*q* (rank-sum)	*q* (paired)
LUAD	111	102	102	+0.99 (+0.98, +1.00)	5.1e−35	7.3e−18
GBM	100	10	0	+0.99 (+0.96, +1.00)	3.5e−7	–
LUSC	110	102	102	+0.97 (+0.94, +0.99)	1.0e−33	7.3e−18
OV	84	19	10	+0.91 (+0.82, +0.98)	1.0e−9	0.0078
COAD	97	100	96	+0.89 (+0.81, +0.95)	1.5e−26	1.2e−15
CCRCC	110	84	84	+0.65 (+0.53, +0.77)	1.4e−14	1.0e−11
HNSC	116	66	66	+0.52 (+0.39, +0.66)	6.3e−9	1.3e−7
UCEC	103	49	30	−0.04 (−0.22, +0.13)	0.6876	0.9515
PDAC	145	93	84	−0.28 (−0.43, −0.13)	0.0003	0.0108

Note: Nucleolin protein abundance in tumor and adjacent normal tissues across nine Clinical Proteomic Tumor Analysis Consortium (CPTAC) cohorts, including sample sizes, matched pairs, effect sizes, and false discovery rate (FDR)-adjusted *p*-values.

differ between tumor and normal samples (delta = −0.04, q = 0.69) despite modest transcript elevation. In PDAC, NCL protein was significantly lower in tumor (delta −0.28, 95% CI −0.43 to −0.13, $q = 3.2 \times 10^{-4}$), whereas the transcript was strongly elevated in TCGA PAAD (delta +0.86). Pancreatic tumors have a high stromal content and low epithelial cellularity, and bulk transcript and bulk protein measurements weigh these compartments differently; such discordance is a known feature of pancreatic proteogenomic data. We therefore do not claim protein-level overexpression of NCL in pancreatic cancer.

Notably, the ccRCC proteomic result (+0.65) supports the GTEx-based direction in kidney cancer rather than the adjacent-normal direction, providing independent evidence relevant to the comparator discordance described in Section 3.1.

3.9. Nucleolin is coupled to proliferation and RNA metabolism in every cancer

Pre-ranked GSEA of genes ranked by their within-cancer correlation with NCL produced 32,491 tests across 32 cancers, of which 9,349 were significant at $q < 0.05$ (Figure 6, Table S10). The result was strikingly uniform.

In the Hallmark collection, the G2/M checkpoint signature was positively enriched in all 32 cancers (median normalized enrichment score [NES] +1.91). MYC targets V1 were enriched in 31 of 32 (+1.83; the exception was OV), the mitotic spindle signature in 31 of 32 (+1.60; the exception was KICH), E2F targets in 30 of 32 (+1.91; the exceptions were THCA and THYM), and MYC targets V2 in 26 of 32 (+1.78; the exceptions were KICH, KIRP, PCPG, THCA, THYM, and UVM). Every one of these enrichments was positive; none was significantly negative in any cancer. Further positively enriched sets included the unfolded protein response (22 of 32), mechanistic target of rapamycin complex 1 (mTORC1) signaling (21 of 32), phosphoinositide 3-kinase (PI3K)–protein kinase B (AKT)–mechanistic target of rapamycin (mTOR) signaling (20 of 32), and DNA repair (14 of 32). Only one signature was consistently depleted, and in only three cancers (coagulation; DLBC, SARC, and TGCT).

Reactome results were concordant and dominated by nuclear RNA metabolism, transport of mature transcripts to the cytoplasm, processing of capped intron-containing pre-mRNA, small nuclear ribonucleoprotein (snRNP) assembly, nuclear transfer RNA (tRNA) processing, gene silencing by RNA, small ubiquitin-like modifier (SUMO) ylation of RNA-binding and DNA-repair proteins, chromatin-modifying enzymes, and histone acetylation, each positively enriched in 31 of 32 cancers (median NES +1.7 to +2.0).

Two features of this result matter for interpretation. First, it recovers NCL's established biology (ribosome biogenesis, RNA processing, and chromatin function) from an unsupervised, correlation-ranked analysis, supporting the validity of the ranking statistic. Second, and more importantly, no immune or inflammatory signature appears among the consistently enriched pathways in either collection. The pathways associated with NCL across cancers are those of a proliferating, biosynthetically active cell. This is the strongest single piece of evidence that NCL's immune associations (Sections 3.4–3.6) follow from the

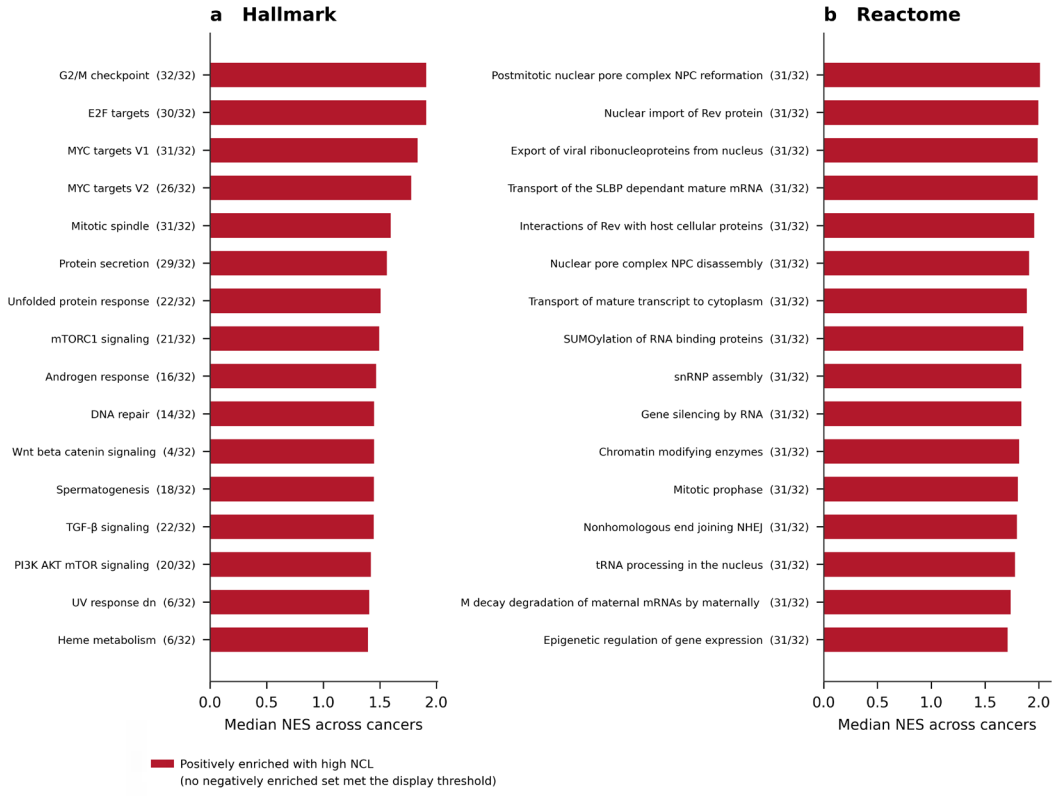


Figure 6. Pathway associations showing consistent nucleolin (NCL)-linked enrichment of (a) Hallmark and (b) Reactome gene sets across cancers. Normalized enrichment scores (NESs) from pre-ranked gene set enrichment analysis, by their within-cancer correlation with NCL, for Hallmark and Reactome gene sets, showing consistent enrichment directions across cancers. The 16 most frequently significant direction-consistent gene sets are shown for each collection. The fraction beside each pathway (e.g., 32/32, 26/32) indicates the number of the 32 analyzable cancers in which the gene set reached $q < 0.05$, out of all tested cancers. Bar length represents the median NES across cancers in which the gene set was significant. Cancers not counted are those in which the gene set did not reach significance; no gene set shown here showed significant enrichment in opposite directions across cancers. Exceptions among the near-complete sets are described in Section 3.9, and full per-cancer results are provided in Table S10.

proliferative state of the tumor compartment rather than from a dedicated immunoregulatory program, and it is why the checkpoint analysis in Section 3.6 was adjusted for proliferation before any association was described as robust (Table S10a).

The TGF-β signaling was positively enriched in 22 of 32 cancers, whereas, at the level of individual genes, *TGFB1* was robustly correlated with NCL in only 6 of 33 cancers (Section 3.6). Pathway-level and gene-level results should not be conflated; enrichment of a signaling signature reflects coordinated movement of many downstream targets and does not imply that the ligand transcript itself tracks NCL (Table S11).

4. Discussion

4.1. Nucleolin overexpression, prognostic value, and immune exclusion

Three conclusions follow. First, NCL overexpression in cancer is real, widespread, and reproducible: it is elevated at the transcript level in 24 of 29 evaluable cancers, with large effect sizes, and confirmed at the protein level in an independent patient series across 7 of 9 cohorts. This is the best-supported claim about NCL in cancer, and it survives rigorous testing.

Second, total NCL transcript abundance carries little independent prognostic information; this is informative rather than merely negative. In 74 multivariable models across three endpoints, NCL remained significant after covariate adjustment and FDR correction in two cancers.

This result should be interpreted alongside, not against, the existing prognostic literature. The most systematic prior assessment is a meta-analysis of 12 immunohistochemistry studies in 1,221 patients across 8 cancer types, demonstrating that cytoplasmic NCL was strongly associated with poor OS (HR 4.32, 95% CI 3.01–6.19) while nuclear NCL was associated with improved

OS (HR 0.42, 95% CI 0.20–0.86), and describing the role of total NCL as controversial.[8] Bulk transcriptome measurement cannot distinguish these compartments; it sums a fraction that is adverse and a protective fraction. On that reading, the near-absence of an independent transcript-level prognostic effect is consistent with the localization hypothesis, and our result supports rather than contradicts it.

The practical implication is specific. If NCL is to be used prognostically, the measurement must resolve subcellular localization, which means immunohistochemistry or an equivalent spatial assay rather than bulk expression. Reports of transcript-level NCL prognostic value in individual cancers, including our own earlier analysis, should be interpreted with that constraint in mind, particularly where clinical covariates were not modeled; the attenuation we observed between univariate and adjusted models is exactly what one expects of a proliferation-coupled gene, since proliferation is already embedded in grade and stage.

Third, the relationship between NCL and the tumor immune microenvironment is consistent across cancers, but not as assumed. NCL-high tumors are immune- and stroma-poor. Where NCL does correlate positively with an immune-relevant molecule, that molecule is characteristically expressed by tumor cells rather than by lymphocytes.

4.2. B7-H3 as the principal immune association

The most robust immune finding is the association between NCL and B7-H3 (CD276), which is positive in 21 of 33 cancers after adjustment for both purity and proliferation, and negative in none. B7-H3 is a compelling comparison to NCL; both are broadly overexpressed across carcinomas, reach the tumor-cell surface, and are being pursued as surface-targeting opportunities. B7-H3 is currently the subject of extensive clinical development, including antibody–drug conjugates and chimeric antigen receptor T-cell approaches. It is notable for being expressed in tumors that are not inflamed, the immune-excluded setting in which conventional PD-1/PD-L1 blockade performs poorly.[48]

The co-occurrence of NCL and B7-H3 in an immune-excluded context, together with the parallel associations with CD73 and CD39, suggests that NCL-high tumors may be characterized less by classical T-cell exhaustion than by an adenosine-mediated, tumor-cell-intrinsic mode of immune evasion.[49] We emphasize that this is an association-level observation and a hypothesis for testing, not a demonstrated mechanism; nothing in these data establishes that NCL regulates B7-H3 or that either is causal for the immune phenotype. The relevant experiment, which perturbs NCL and measures B7-H3 and immune composition, has not been performed here.

4.3. Nucleolin is primarily linked to proliferative and RNA-processing programs

The enrichment results are unusually consistent and provide context for the immune findings. Across 32 cancers, the programs associated with NCL are those of proliferation and nuclear RNA metabolism, the G2/M checkpoint signature in every cancer without exception, MYC and E2F target sets in almost all, and Reactome sets covering mRNA processing and export, snRNP assembly, tRNA processing, and chromatin modification in 31 of 32.

That this analysis recovers NCL's canonical biology from an unsupervised, correlation-ranked procedure is reassuring about the method. What it adds is the absence: no immune or inflammatory program is consistently enriched in either collection. Taken together with the negative correlations with immune, stromal, and microenvironmental scores, this supports interpreting NCL as a marker of the biosynthetic and proliferative state of the tumor compartment, whose immune associations follow from that state rather than from a dedicated immunoregulatory function.

This is also why we regard the B7-H3 result as the most interesting finding rather than a corollary of proliferation. Proliferation is the dominant axis of NCL covariation, so an immune association that persists after explicit adjustment for a proliferation score, such as the B7-H3, CD73, and CD39 associations, while most T-cell checkpoint associations do not, is not simply a restatement of that axis.

4.4. Methodological considerations for pan-cancer nucleolin analyses

Two methodological observations from this work generalize beyond NCL.

The first concerns deconvolution. One-third of the cancer × cell type combinations we examined produced significant estimates with opposite signs depending on the algorithm used. Since a typical pan-cancer immune analysis reports one algorithm, and since algorithms are often selected after inspecting results, the field's immune-association literature is likely to contain a substantial number of associations that would reverse under a different but equally defensible choice. Reporting concordance across algorithms costs little and is considerably more informative than a single point estimate.

The second concerns the choice of a normal comparator. In kidney and thyroid cancers, the direction of NCL

dysregulation depends on whether GTEx or adjacent normal tissue is used. Neither is simply correct; adjacent normal tissue shares the patient's genetic background and exposures but is subject to field effects and surgical ischemia, whereas GTEx tissue is unaffected by the tumor and is post-mortem and from different individuals. For genes sensitive to metabolic and ischemic states, which include most of the ribosome-biogenesis machinery, this choice is consequential and should be reported rather than kept silent.

4.5. Limitations

The central limitation is that this study generates no new experimental data. We have, as far as possible, computationally validated, in an independent patient series, measurements at the protein level on an orthogonal platform using paired adjacent normal tissue (CPTAC). This is a genuine and reasonably stringent form of external validation that substantiates the claim of overexpression. It is not, however, a substitute for functional experimentation, and none of the mechanistic hypotheses advanced in Section 4.2 is tested by it. The association between NCL and B7-H3 in particular requires perturbation experiments, namely *NCL* knockdown or overexpression with measurement of B7-H3 and of immune composition, before any regulatory relationship can be claimed. We state plainly that this work does not provide that evidence.

Further limitations follow from the data sources. Transcriptome deconvolution infers cell composition from bulk expression and is not a measurement of infiltration; the poor cross-algorithm concordance documented in Section 3.5 is itself evidence of this limitation, and single-cell or spatial data would be required to resolve the cell-type associations convincingly. Five TCGA–GTEx tissue pairings are approximate and are flagged as such; three cancers have no acceptable GTEx counterpart, and LAML could not be compared against GTEx because no matched normal samples were available. Several cohorts are small, including CHOL (36 tumors), KICH (65), and UCS (57), so the absence of a significant association in those cancers reflects limited power as much as the absence of effect, and KICH in particular could not be modeled for OS. Stage and grade are unavailable for some cancers, so the covariate sets differ across models; the covariates used in each model are recorded. The proportional-hazards assumption was violated for the NCL term in 10 of 74 models, and those HRs summarize time-varying effects. Finally, TCGA is predominantly composed of treatment-naive primary resections from a limited set of populations, and prognostic associations derived from it need not transfer to treated, metastatic, or differently ascertained populations.

5. Conclusion

In conclusion, our comprehensive pan-cancer analysis of NCL revealed that its overexpression is broad and reproducible across cancer types at both the transcript and protein level, but that it is only rarely an independent prognostic factor once established clinical variables are accounted for. Our findings also clarify the role of NCL within the tumor microenvironment, showing that it marks an immune-excluded phenotype and co-expresses with tumor-intrinsic immunosuppressive ligands rather than with conventional immune checkpoint receptors, refining our understanding of its immunomodulatory role in cancer. Future research should focus on the mechanisms underlying this relationship between NCL and immunosuppressive signaling within the tumor microenvironment. Developing NCL-targeted therapeutic strategies, particularly those directed at its associated immunosuppressive pathway rather than conventional checkpoint pathways, could significantly improve treatment outcomes. Moreover, integrating protein-level and localization-based approaches to explore NCL's functional role could help translate these associations into viable therapeutic strategies.

Acknowledgments

None.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C2005057).

Conflict of interest

Prakash Gangadaran is an Editorial Board Member of this journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Authors contributions

Conceptualization: Kruthika Prakash, ArulJothi Kandasamy Nagarajan, Prakash Gangadaran, Byeong-Cheol Ahn

Formal analysis: Kruthika Prakash, ArulJothi Kandasamy Nagarajan

Investigation: Kruthika Prakash, Janani Balaji, Surya Babu, ArulJothi Kandasamy Nagarajan, Prakash Gangadaran, Byeong-Cheol Ahn

Methodology: Kruthika Prakash, Janani Balaji, Surya Babu, ArulJothi Kandasamy Nagarajan, Prakash Gangadaran, Byeong-Cheol Ahn
Writing–original draft: Kruthika Prakash, Janani Balaji, Surya Babu, Ramya Lakshmi Rajendran
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All primary data are publicly available: transcriptome data (UCSC Xena Toil recompute, https://xenabrowser.net/datapages/; dataset: *TcgaTargetGtex_rsem_gene_tpm*); clinical, survival, TMB, and MSI data (cBioPortal, https://www.cbioportal.org; and TCGA PanCancer Atlas studies); immune infiltration estimates (TIMER2.0, http://timer.cistrome.org); gene sets (MSigDB v2024.1.Hs, https://www.gsea-msigdb.org); and proteomic data (CPTAC via the *cptac* Python package). No new data were generated in this study. All analysis code, the pinned software environment, the run manifest recording input checksums and package versions, every intermediate result table, and a step-by-step methodology document are openly available at https://github.com/ApexBlue11/NCL-PanCancer-Analysis. The pipeline reproduces every number and figure reported here from public data.

Further disclosure

An AI assistant (Anthropic Claude) was used during the preparation of this revision for analysis code development, statistical implementation, and the execution of the computational workflow. All analyses were specified, inspected, and verified by the authors; all statistical implementations were validated against reference implementations prior to use; and the authors take full responsibility for the integrity of the data, the accuracy of the analyses, and the content of the manuscript. No AI system is an author, and no AI system was used to generate or fabricate data.

References

1. Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263.

   doi: 10.3322/caac.21834

2. Piña-Sánchez P, Chávez-González A, Ruiz-Tachiquín M, *et al*. Cancer Biology, Epidemiology, and Treatment in the 21st Century: Current Status and Future Challenges From a Biomedical Perspective. *Cancer Control*. 2021;28:10732748211038735.

   doi: 10.1177/10732748211038735

3. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, *et al*. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-1120.

   doi: 10.1038/ng.2764

4. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660.

   doi: 10.1126/science.1262110

5. Ellis MJ, Gillette M, Carr SA, *et al*. Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov*. 2013;3(10):1108-1112.

   doi: 10.1158/2159-8290.CD-13-0219

6. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

   doi: 10.1371/journal.pmed.0020124

7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300.

   doi: 10.1111/j.2517-6161.1995.tb02031.x

8. Yangngam S, Prasopsiri J, Hatthakarnkul P, *et al*. Cellular localization of nucleolin determines the prognosis in cancers: a meta-analysis. *J Mol Med (Berl)*. 2022;100(8):1145-1157.

   doi: 10.1007/s00109-022-02228-w

9. Berger CM, Gaume X, Bouvet P. The roles of nucleolin subcellular localization in cancer. *Biochimie*. 2015;113:78-85.

   doi: 10.1016/j.biochi.2015.03.023

10. Mongelard F, Bouvet P. Nucleolin: a multiFACeTed protein. *Trends Cell Biol*. 2007;17(2):80-86.

    doi: 10.1016/j.tcb.2006.11.010

11. Jia W, Yao Z, Zhao J, Guan Q, Gao L. New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sci*. 2017;186:1-10.

    doi: 10.1016/j.lfs.2017.07.025

12. Bates PJ, Reyes-Reyes EM, Malik MT, Murphy EM, O'Toole MG, Trent JO. G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochim Biophys Acta Gen Subj*. 2017;1861(5 Pt B):1414-1428.

    doi: 10.1016/j.bbagen.2016.12.015

13. Gilles ME, Maione F, Cossutta M, *et al*. Nucleolin Targeting Impairs the Progression of Pancreatic Cancer and Promotes

the Normalization of Tumor Vasculature. *Cancer Res*. 2016;76(24):7181-7193.

doi: 10.1158/0008-5472.CAN-16-0300

14. Ponzo M, Debesset A, Cossutta M, *et al*. Nucleolin Therapeutic Targeting Decreases Pancreatic Cancer Immunosuppression. *Cancers (Basel)*. 2022;14(17):4265.

doi: 10.3390/cancers14174265

15. Lin Q, Ma X, Hu S, *et al*. Overexpression of Nucleolin is a Potential Prognostic Marker in Endometrial Carcinoma. *Cancer Manag Res*. 2021;13:1955-1965.

doi: 10.2147/CMAR.S294035

16. Ezzatifar F, Rafiei A, Jeddi-Tehrani M. Nucleolin; A tumor associated antigen as a potential lung cancer biomarker. *Pathol Res Pract*. 2022;240:154160.

doi: 10.1016/j.prp.2022.154160

17. Ying J, Pan R, Tang Z, *et al*. Downregulation of NCL attenuates tumor formation and growth in HeLa cells by targeting the PI3K/AKT pathway. *Cancer Med*. 2022;11(6):1454-1464.

doi: 10.1002/cam4.4569

18. Ugrinova I, Petrova M, Chalabi-Dchar M, Bouvet P. Multifaceted Nucleolin Protein and Its Molecular Partners in Oncogenesis. *Adv Protein Chem Struct Biol*. 2018;111:133-164.

doi: 10.1016/bs.apcsb.2017.08.001

19. Yu X, Xie L, Ge J, Li H, Zhong S, Liu X. Integrating single-cell RNA-seq and spatial transcriptomics reveals MDK-NCL dependent immunosuppressive environment in endometrial carcinoma. *Front Immunol*. 2023;14:1145300.

doi: 10.3389/fimmu.2023.1145300

20. Newman AM, Liu CL, Green MR, *et al*. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.

doi: 10.1038/nmeth.3337

21. Finotello F, Mayer C, Plattner C, *et al*. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. 2019;11(1):34.

doi: 10.1186/s13073-019-0638-6

22. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220.

doi: 10.1186/s13059-017-1349-1

23. Becht E, Giraldo NA, Lacroix L, *et al*. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218.

doi: 10.1186/s13059-016-1070-5

24. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. 2017;6:e26476.

doi: 10.7554/eLife.26476

25. Harris CR, Millman KJ, van der Walt SJ, *et al*. Array programming with NumPy. *Nature*. 2020;585(7825):357-362.

doi: 10.1038/s41586-020-2649-2

26. Virtanen P, Gommers R, Oliphant TE, *et al*. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272.

doi: 10.1038/s41592-019-0686-2

27. Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw*. 2019;4(40):1317.

doi: 10.21105/joss.01317

28. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2023;39(1):btac757.

doi: 10.1093/bioinformatics/btac757

29. Vivian J, Rao AA, Nothaft FA, *et al*. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*. 2017;35(4):314-316.

doi: 10.1038/nbt.3772

30. Goldman MJ, Craft B, Hastie M, *et al*. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-678.

doi: 10.1038/s41587-020-0546-8

31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.

doi: 10.1186/1471-2105-12-323

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

doi: 10.1093/nar/gkaa1087

33. Cerami E, Gao J, Dogrusoz U, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.

doi: 10.1158/2159-8290.CD-12-0095

34. Gao J, Aksoy BA, Dogrusoz U, *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.

doi: 10.1126/scisignal.2004088

35. Liu J, Lichtenberg T, Hoadley KA, *et al*. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*. 2018;173(2):400-416.e11.

doi: 10.1016/j.cell.2018.02.052

36. Kautto EA, Bonneville R, Miya J, *et al*. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget*. 2017;8(5):7452-7463.

doi: 10.18632/oncotarget.13918

37. Li B, Severson E, Pignon JC, *et al*. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17(1):174.

doi: 10.1186/s13059-016-1028-7

38. Li T, Fu J, Zeng Z, *et al*. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48(W1):W509-W514.

doi: 10.1093/nar/gkaa407

39. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol Bull*. 1993;114(3):494-509.

doi: 10.1037/0033-2909.114.3.494

40. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Stat*. 1981;6(2):107-128.

doi: 10.2307/1164588

41. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika*. 1954;41(1-2):133-145.

doi: 10.1093/biomet/41.1-2.133

42. Terpstra TJ. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indag Math*. 1952;55:327-333.

doi: 10.1016/S1385-7258(52)50043-X

43. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526.

doi: 10.2307/2337547

44. Bonett DG, Wright TA. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*. 2000;65(1):23-28.

doi: 10.1007/BF02294183

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

doi: 10.1073/pnas.0506580102

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

doi: 10.1016/j.cels.2015.12.004

47. Gillespie M, Jassal B, Stephan R, *et al*. The reactome pathway knowledgebase 2022. *Nucleic Acids Res*. 2022;50(D1):D687-D692.

doi: 10.1093/nar/gkab1028

48. Koumprentziotis IA, Theocharopoulos C, Foteinou D, *et al*. New Emerging Targets in Cancer Immunotherapy: The Role of B7-H3. *Vaccines (Basel)*. 2024;12(1):54.

doi: 10.3390/vaccines12010054

49. Allard B, Allard D, Buisseret L, Stagg J. The adenosine pathway in immuno-oncology. *Nat Rev Clin Oncol*. 2020;17(10):611-629.

doi: 10.1038/s41571-020-0382-2