



Integrative single-cell and bulk transcriptomics identify an *ALK*-associated three-gene signature predicting neuroblastoma outcomes

Aiguo Zhu^{1,2#}, Yuren Xia^{1,2#}, Xin Li^{1,2#}, Zhengchao Jin^{1,2}, Jian Wang^{1,2}, Yi Pan^{1,2}

¹Tianjin Cancer Hospital Airport Hospital, National Clinical Research Center for Cancer, Tianjin, China; ²Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China

Contributions: (I) Conception and design: Z Jin, J Wang, Y Pan; (II) Administrative support: Z Jin, J Wang, Y Pan; (III) Provision of study materials or patients: A Zhu, X Li, Y Xia; (IV) Collection and assembly of data: A Zhu, X Li, Y Xia; (V) Data analysis and interpretation: A Zhu, X Li, Y Xia; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Zhengchao Jin, MD; Jian Wang, MD; Yi Pan, MD. Tianjin Cancer Hospital Airport Hospital, National Clinical Research Center for Cancer, Tianjin, China; Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, East Five Road, Dongli District, Tianjin 300060, China. Email: tzkg_jinzc@foxmail.com; wang5862@163.com; tjphyLP@126.com.

Background: Activating anaplastic lymphoma kinase (*ALK*) mutations define a clinically relevant subset of neuroblastoma (NB), yet *ALK*-linked transcriptional programs with robust prognostic value remain insufficiently characterized. This study aimed to identify *ALK*-associated transcriptional features across single-cell and bulk datasets and to develop a reproducible prognostic signature for NB.

Methods: We integrated a public single-cell atlas of high-risk NB from CELLxGENE with bulk RNA sequencing (RNA-seq) cohorts and NB cell-line datasets. Differential expression between *ALK*-mutant and *ALK*-wild-type samples was analyzed in (I) pseudobulk malignant cells from the single-cell dataset, (II) the Kids First (Maris) cohort, and (III) GSE89413 NB cell lines, followed by direction-consistent intersection. A prognostic model was developed in GSE62564 using Cox regression and evaluated by Kaplan-Meier survival analysis and time-dependent receiver operating characteristic (ROC) curves. External validation was performed in GSE181559 and TARGET-NBL. Biological significance was further explored using preranked gene set enrichment analysis (GSEA).

Results: Three concordant *ALK*-associated genes, *DLK1*, *RTL1*, and *ISLR2*, were identified across all three analytical contexts and were consistently upregulated in *ALK*-mutant samples. A three-gene risk score significantly stratified overall survival and event-free survival in GSE62564 and remained predictive in GSE181559 and TARGET-NBL without model re-fitting. High-risk tumors were enriched for translational and ribosome biogenesis pathways, whereas low-risk tumors were enriched for adaptive immune activation programs.

Conclusions: Cross-context transcriptomic integration identified a minimal *ALK*-associated three-gene signature that reproducibly predicts NB outcomes and reflects distinct underlying biological states, supporting its potential value for risk stratification and hypothesis-driven therapeutic development.

Keywords: Neuroblastoma (NB); anaplastic lymphoma kinase mutation (*ALK* mutation); single-cell RNA sequencing (scRNA-seq); prognostic signature; bulk RNA sequencing integration (bulk RNA-seq integration)

Submitted Feb 23, 2026. Accepted for publication Mar 31, 2026. Published online Apr 24, 2026.

doi: 10.21037/tcr-2026-1-0396

View this article at: <https://dx.doi.org/10.21037/tcr-2026-1-0396>

Introduction

Neuroblastoma (NB) is one of the most common extracranial solid tumors in children and is characterized by marked biological and clinical heterogeneity (1,2). Among its recurrent genetic alterations, activating mutations in the anaplastic lymphoma kinase (*ALK*) receptor tyrosine kinase define a clinically important and therapeutically actionable subset (3). Although the development of next-generation *ALK* inhibitors has expanded treatment options for patients with *ALK*-aberrant NB, therapeutic responses remain heterogeneous, and acquired resistance frequently emerges. These challenges highlight the need for biologically informed biomarkers that can capture *ALK*-associated downstream transcriptional states and improve clinically relevant risk stratification (4).

Recent advances in transcriptomic profiling have greatly expanded our understanding of NB at both bulk and single-cell resolution. In particular, single-cell RNA sequencing

(scRNA-seq) has revealed extensive intratumoral heterogeneity, diverse malignant cell states, and a complex tumor microenvironment, while large public cohorts have enabled transcriptome-based discovery linked to clinical outcomes (5-7). However, few studies have systematically examined *ALK* mutation-associated transcriptional programs across single-cell malignant profiles, patient tumors, and NB cell lines in parallel, and even fewer have distilled such cross-context signals into compact signatures with reproducible prognostic value. Therefore, the identification of a minimal and biologically meaningful marker set that is both *ALK*-associated and transferable across cohorts may improve current risk stratification and facilitate the development of rational therapeutic strategies.

In this study, we integrated a longitudinal high-risk NB single-cell atlas with bulk RNA sequencing (RNA-seq) data from clinical cohorts and NB cell lines to identify transcriptional features consistently associated with *ALK* mutation. By intersecting *ALK*-mutant *vs.* *ALK*-wild-type differential expression across malignant-cell pseudobulk single-cell profiles, the Kids First (Maris) cohort, and GSE89413 cell lines, we nominated a three-gene panel (*DLK1*, *RTL1*, and *ISLR2*) and constructed a parsimonious prognostic signature. The model was trained in GSE62564 and validated in independent datasets (GSE181559 and TARGET-NBL) without re-fitting. Finally, we linked the risk score to coherent biological programs, including enhanced translational activity in high-risk tumors and adaptive immune activation in low-risk tumors. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2026-1-0396/rc>).

Methods

Study design and overall workflow

We performed an integrative, multi-cohort analysis to investigate the association of *ALK* mutation status with transcriptomic programs and clinical outcomes in NB. The study workflow consisted of four sequential steps: identification of *ALK*-associated transcripts in a single-cell atlas, cross-context confirmation in bulk patient tumors and NB cell lines, construction of a parsimonious multigene prognostic model in a training cohort, and external validation with downstream pathway interpretation. Only de-identified, publicly available datasets were included, and all analyses were conducted using R and Python. The

Highlight box

Key findings

- By integrating single-cell and bulk transcriptomic datasets, we identified an anaplastic lymphoma kinase (*ALK*)-associated three-gene prognostic signature (*DLK1*, *RTL1*, and *ISLR2*) in neuroblastoma (NB). The risk score stratified survival in the training cohort and validated across independent cohorts without re-fitting, and was linked to distinct biological programs between risk groups.

What is known and what is new?

- ALK* mutations are recurrent, actionable drivers in NB, but transcriptome-based biomarkers that are both *ALK*-related and broadly reproducible across platforms remain limited. Single-cell atlases reveal heterogeneous malignant states and microenvironmental features, yet translation of these insights into compact, outcome-linked predictors is still challenging.
- By requiring directionally concordant differential expression across malignant-cell pseudobulk single-cell data, patient tumors, and cell lines, we distilled *ALK*-associated biology into a three-gene panel with cross-platform generalizability and coherent biological interpretation.

What is the implication, and what should change now?

- This three-gene signature is simple to implement from bulk transcriptomes and can complement established clinical risk factors for outcome stratification in NB. It also provides a rationale to test risk-adapted therapeutic strategies in *ALK*-mutant disease, including *ALK*-targeted regimens with biologically informed combinations. Prospective validation is warranted before clinical adoption.

study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

Data sources

Single-cell transcriptomic profiles were obtained from the longitudinal multi-omic atlas of high-risk NB, accessed through CELLxGENE (collection: cee845e3-ec04-4781-9e2a-28734bb4f7ba) (5). The authors' sample-level *ALK* annotations were adopted, and mutation-stratified analyses were restricted to malignant cells as defined below. Bulk RNA-seq expression and survival data for model training were derived from GSE62564. For *ALK* mutation-stratified differential expression analysis in patient tumors, we used the Kids First cBioPortal Maris cohort (x01_fy16_nbl_maris). *ALK*-stratified expression in NB cell lines was evaluated using GSE89413. External validation was performed in GSE181559 and the TARGET-NBL RNA-seq cohort with available clinical outcome data. The baseline demographic and clinical characteristics of the training and validation cohorts are summarized in Table S1.

Single-cell processing and malignant-cell subsetting

Single-cell preprocessing followed the protocol of Yu *et al.* (5). Raw snRNA-seq reads were aligned to GRCh38.p13 using Cell Ranger v3.1.0. High-quality cells were retained if they had 2,000–40,000 unique molecular identifiers, 1,000–10,000 detected genes, and <10% mitochondrial unique molecular identifiers. Filtered cells were log-normalized in Seurat v3, doublets were removed using DoubletFinder, and ambient RNA contamination was corrected using decontX; decontaminated matrices were rounded to integer counts and re-normalized. For integration, Seurat objects were split by patient, 2,000 integration features were selected, genes with a Gini index >0.8 or expressed in <0.2% of cells across patients were excluded, and integration was performed using FindIntegrationAnchors with k.anchor = 10. Uniform Manifold Approximation and Projection (UMAP) embedding and Louvain clustering were computed using the first 50 principal components with a resolution of 0.2. For malignant-cell reintegration, 2,500 highly variable genes were selected, initial UMAP and clustering were performed using the first 30 principal components at resolution 0.2, and Harmony-based reintegration was subsequently carried out using the first 20 Harmony components with clustering at resolution 0.2.

Bulk expression processing and mutation annotation

For the Kids First (Maris) cohort, *ALK* status was obtained from the mutation annotations provided by the cBioPortal database. Samples harboring non-silent pathogenic or likely pathogenic *ALK* variants, including the canonical hotspot mutations F1174, F1245, and R1275, were classified as *ALK*-mutant, whereas samples without such alterations were classified as *ALK*-wild-type. Expression values were analyzed in the format provided by the portal. For GSE89413, repository-provided RNA-seq quantification data were used, and *ALK* mutation status for each NB cell line was curated from the associated dataset annotations and relevant literature before differential expression analysis. For GSE62564 and the other bulk RNA-seq cohorts in which expression values were provided as fragments per kilobase of transcript per million mapped reads (FPKM), gene-level FPKM matrices were downloaded as deposited, transformed to $\log_2(\text{FPKM} + 1)$, and mapped to HGNC gene symbols. When multiple entries corresponded to the same gene symbol, the entry with the largest interquartile range across samples was retained.

Differential expression and cross-context intersection

Mutation-stratified differential expression results from three independent comparisons were consolidated and filtered to generate a harmonized, direction-consistent candidate set prior to intersection. Specifically, we imported the differential expression results from the single-cell pseudobulk tumor-cell comparison (*ALK*-mutant *vs.* *ALK*-wild-type), the NB cell-line comparison from GSE89413, and the patient-tumor comparison from the Kids First Maris cohort. For the single-cell pseudobulk analysis, genes were filtered using Benjamini-Hochberg-adjusted P values (adjusted P value <0.01) and effect-size criteria ($|\log_2\text{fold change (FC)}| \geq 1$). For the GSE89413 and Maris datasets, genes were filtered using nominal P values (P value <0.05) and effect-size criteria ($|\log\text{FC}| \geq 0.5$). Genes were further required to show concordant directionality across all three comparisons, with positive log fold-change indicating upregulation in *ALK*-mutant relative to *ALK*-wild-type samples.

Prognostic model training

Prognostic model development was conducted in the GSE62564 cohort. Expression values for *DLK1*, *RTL1*, and

ISLR2 were standardized as z-scores within the cohort, and a multivariable Cox proportional hazards model including these three genes was fitted using overall survival (OS) as the endpoint. Clinical covariates such as age, disease stage, and *MYCN* status were not incorporated into model construction. The estimated regression coefficients from the training cohort were used to construct a linear risk score, defined as: $\text{RiskScore} = 0.07354 \times \text{DLK1} + 0.13085 \times \text{RTL1} + 0.34775 \times \text{ISLR2}$. Patients were then dichotomized into high- and low-risk groups according to the cohort median RiskScore. Survival differences between groups were assessed using Kaplan-Meier analysis with the log-rank test, and predictive performance was evaluated using time-dependent receiver operating characteristic (ROC) curves at prespecified time points (2, 3, 5, and 8 years). The proportional hazards assumption was assessed using Schoenfeld residuals.

To assess potential collinearity among the three model genes, pairwise Pearson correlation analysis and variance inflation factor (VIF) calculations were performed in the GSE62564 training cohort.

External validation and biological interpretation

The regression coefficients derived from the training cohort were directly applied to the external validation cohorts, GSE181559 and TARGET-NBL, without model re-fitting. Within each validation cohort, gene expression values were standardized as z-scores, and the RiskScore was calculated using the same formula as that used in the training set. Patients were stratified into high- and low-risk groups according to the median RiskScore within each cohort, and survival analyses together with time-dependent ROC evaluation were performed using the same procedures as in the training cohort. To explore the biological characteristics associated with the RiskScore, we compared gene expression profiles between the high- and low-risk groups in the training cohort and generated a ranked gene list for preranked gene set enrichment analysis (GSEA). Enrichment analyses were conducted using the MSigDB Hallmark, GO biological process, and KEGG gene set collections, and pathways with a false discovery rate (FDR) <0.05 were considered significantly enriched.

Visualization, statistics, and software

UMAP embeddings, dot plots, and heatmaps for the single-cell analyses were generated in Python using scanpy,

anndata, and matplotlib, following the recommended settings of the source atlas. Survival curves, time-dependent ROC analyses, and Cox regression models were performed in R using the survival, survminer, and timeROC packages. Multiple testing was controlled using the Benjamini-Hochberg method where applicable, and all statistical tests were two-sided unless otherwise specified.

Statistical analysis

Statistical analyses were performed using R and Python. Continuous variables are summarized as medians with ranges, and categorical variables are presented as counts and percentages. Survival differences between groups were evaluated using Kaplan-Meier analysis and the log-rank test. Time-dependent ROC curves were used to assess predictive performance at prespecified time points. Pairwise gene correlations were assessed using Pearson correlation analysis, and collinearity among model variables was evaluated using VIF calculations. Multiple testing was controlled using the Benjamini-Hochberg method where applicable. All statistical tests were two-sided unless otherwise specified, and $P < 0.05$ was considered statistically significant.

Results

Single-cell landscape and cross-cohort nomination of ALK-linked genes

We first delineated the cellular landscape of the published high-risk NB single-cell atlas. Unsupervised clustering and cell-type annotation resolved malignant NB cells together with multiple stromal and immune compartments (Figure 1A). Mapping *ALK* status onto the same embedding showed that *ALK*-mutant and *ALK*-wild-type tumor cells were broadly intermingled across malignant states (Figure 1B), suggesting that *ALK*-associated transcriptional differences occur within shared cellular programs rather than through a major shift in overall cell-type composition. To identify robust *ALK* mutation-associated transcripts, we performed mutation-stratified differential expression analyses in three independent contexts: pseudobulk malignant cells from the single-cell dataset, NB cell lines (GSE89413), and patient tumors from the Kids First (Maris) cohort. After applying context-specific significance and effect-size thresholds and requiring concordant directionality, a three-way intersection yielded a minimal set of three concordant genes—*DLK1*, *RTL1*, and *ISLR2* (Figure 1C). At single-

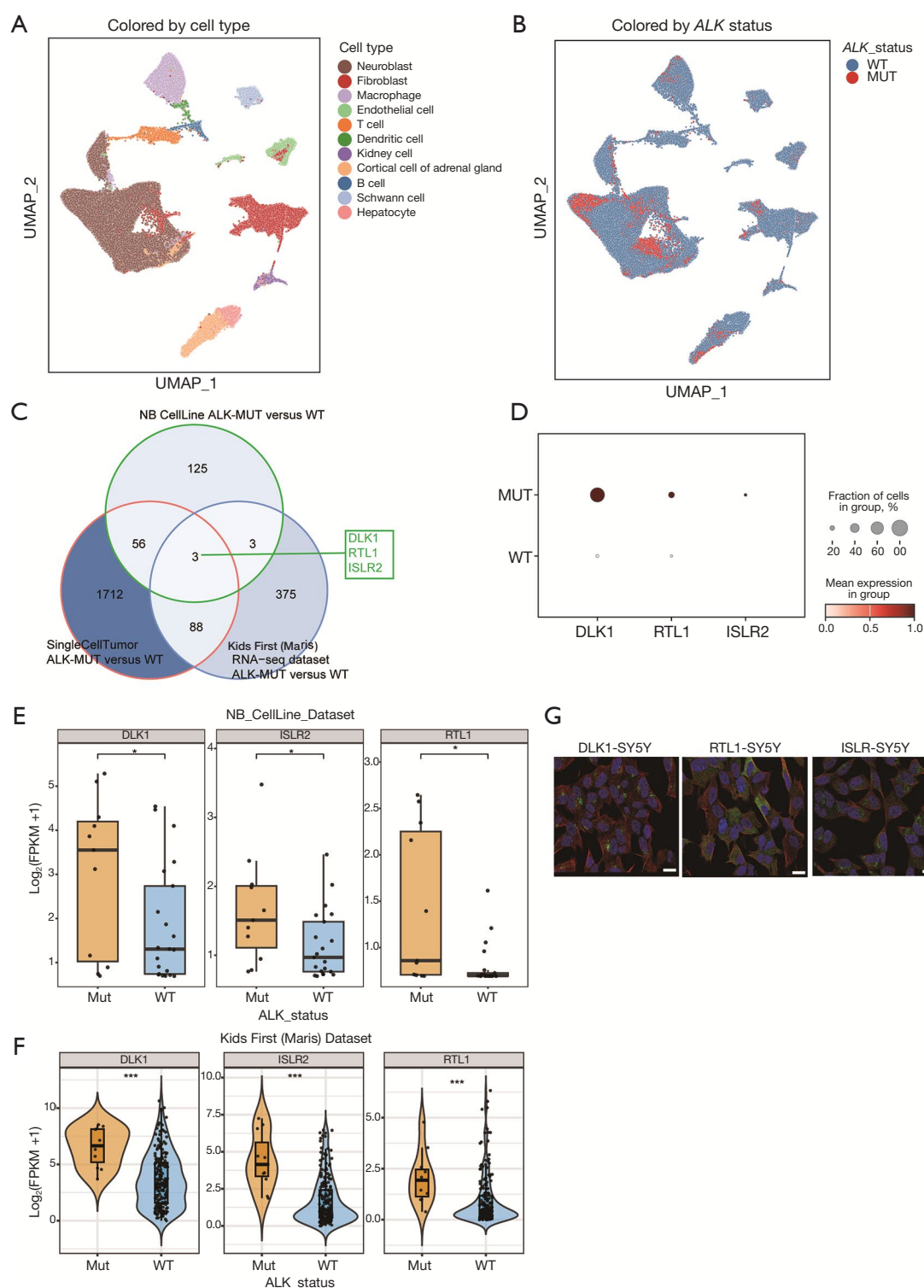


Figure 1 Single-cell landscape, *ALK*-status mapping, cross-context DEG intersection, and candidate-gene expression. (A) UMAP embedding of the published high-risk NB single-cell atlas, annotated into major malignant and stromal/immune compartments. (B) The same UMAP colored by sample-level *ALK* status in malignant cells. (C) Three-way intersection of *ALK*-MUT *vs.* *ALK*-WT DEGs derived from (I) single-cell malignant-cell pseudobulk profiles, (II) NB cell lines (GSE89413), and (III) patient tumors from the Kids First (Maris)

cohort. Dataset-specific filtering criteria were applied (single-cell: adjusted P value <0.01 and $|\log_2FC| \geq 1$; cell lines/Maris: P value <0.05 and $|\log_2FC| \geq 0.5$), and only directionally concordant genes were retained, yielding *DLK1*, *RTL1*, and *ISLR2*. (D) Single-cell expression of *DLK1*, *RTL1*, and *ISLR2* in malignant cells stratified by *ALK* status. (E) Expression of the three genes in NB cell lines stratified by *ALK* status. (F) Expression of the three genes in the Kids First (Maris) cohort stratified by *ALK* status. (G) Representative protein-level images showing predominantly cytoplasmic/perinuclear staining patterns for the three candidates, consistent with the transcriptomic findings. Scale bar = 20 μm . *, $P < 0.05$; ***, $P < 0.001$. *ALK*, anaplastic lymphoma kinase; DEG, differentially expressed gene; FC, fold change; FPKM, fragments per kilobase of transcript per million mapped reads; MUT, mutant; NB, neuroblastoma; UMAP, Uniform Manifold Approximation and Projection; WT, wild-type.

cell resolution, all three genes showed higher expression in *ALK*-mutant malignant cells (Figure 1D) and displayed consistent upregulation in *ALK*-mutant cell lines and *ALK*-mutant patient tumors in the Maris cohort (Figure 1E,1F). Representative protein-level images further supported the predominately cytoplasmic/perinuclear staining patterns of these candidates and were consistent with their transcript-level differences (Figure 1G).

Construction and performance of a three-gene prognostic model in the training cohort

We next developed a prognostic model in the GSE62564 training cohort using the three *ALK*-associated genes. After standardizing the expression values of *DLK1*, *RTL1*, and *ISLR2* within the cohort, a multivariable Cox proportional hazards model was fitted to generate a linear risk score, defined as $\text{RiskScore} = 0.07354 \times \text{DLK1} + 0.13085 \times \text{RTL1} + 0.34775 \times \text{ISLR2}$. Patients were stratified into high- and low-risk groups according to the cohort median RiskScore. Kaplan-Meier analysis demonstrated that patients in the high-risk group had significantly worse OS and event-free survival (EFS) than those in the low-risk group (Figure 2A,2B). Time-dependent ROC analysis further showed that the three-gene signature maintained stable predictive performance across multiple follow-up time points (Figure 2C). To explore the biological features associated with the model-defined risk groups, we compared transcriptomic profiles between the high- and low-risk groups in the training cohort and performed preranked GSEA. High-risk tumors were significantly enriched for pathways related to translation, ribosome biogenesis, and ribosomal RNA (rRNA) processing, whereas low-risk tumors were enriched for adaptive immune-related programs, including T-cell activation and antigen receptor signaling (Figure 2D). These findings indicate that the three-gene signature not only stratifies clinical outcome but also captures distinct underlying biological states.

Because *DLK1* and *RTL1* are both located within the *DLK1-DIO3* imprinted domain, we further assessed collinearity among *DLK1*, *RTL1*, and *ISLR2* in the training cohort. *DLK1* and *RTL1* showed a moderate positive correlation ($r=0.606$, $P < 2.2 \times 10^{-16}$), whereas correlations involving *ISLR2* were weaker (*DLK1* vs. *ISLR2*: $r=0.216$, $P=1.161 \times 10^{-6}$; *RTL1* vs. *ISLR2*: $r=0.228$, $P=2.575 \times 10^{-7}$). VIF analysis showed low VIF values for all three genes (all VIFs <1.6), supporting the stability of the three-gene model.

External validation across independent cohorts

We further evaluated the robustness of the three-gene signature in two independent external cohorts. In GSE181559, the RiskScore calculated using the locked training coefficients significantly stratified OS, with patients in the high-risk group showing markedly poorer outcomes than those in the low-risk group (Figure 3A). Time-dependent ROC analysis demonstrated good predictive performance across multiple follow-up time points in this validation cohort (Figure 3B). Consistent findings were observed in the TARGET-NBL cohort, where the high-risk group also exhibited significantly inferior OS compared with the low-risk group (Figure 3C), and time-dependent ROC analysis confirmed the predictive value of the model (Figure 3D). Together, these results demonstrate that the three-gene signature retains prognostic utility across independent datasets and transcriptomic platforms.

Discussion

In the present study, we integrated single-cell and bulk transcriptomic datasets to identify *ALK* mutation-associated expression patterns in NB and to develop a compact three-gene prognostic signature composed of *DLK1*, *RTL1*, and *ISLR2*. By requiring directionally concordant differential expression across malignant-cell pseudobulk profiles, patient tumors, and NB cell lines, we minimized dataset-

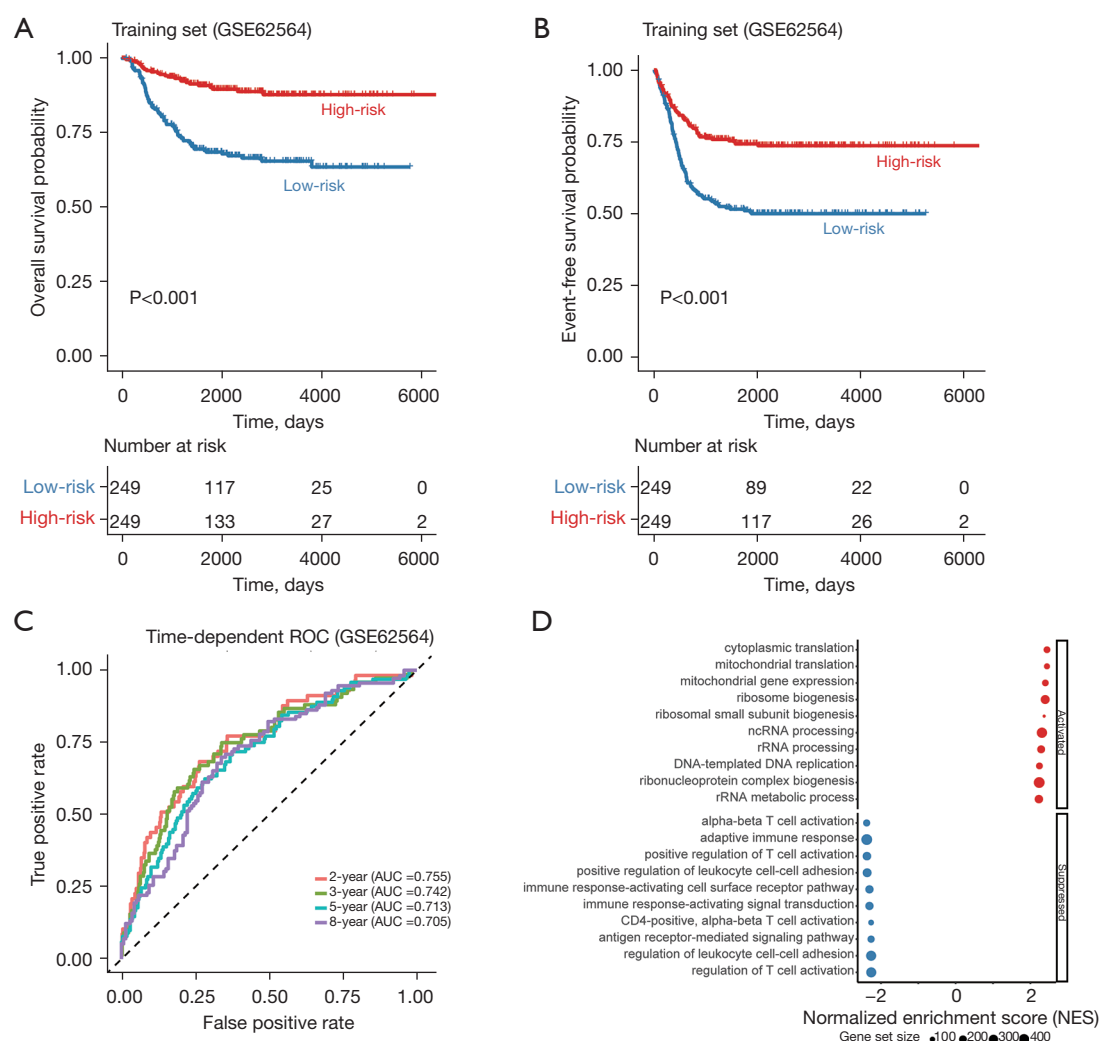


Figure 2 Construction of the three-gene prognostic signature in GSE62564 and associated biological programs. (A,B) Kaplan-Meier curves for OS and EFS in the GSE62564 training cohort. RiskScore was calculated using the three-gene model, and patients were dichotomized into high- and low-risk groups according to the cohort median. Log-rank P values are shown. (C) Time-dependent ROC curves evaluating predictive performance of the signature at prespecified time points. (D) Preranked GSEA comparing high- vs. low-risk groups in the training cohort. High-risk tumors were enriched for translation- and ribosome-related pathways, whereas low-risk tumors were enriched for adaptive immune-related pathways. AUC, area under the curve; EFS, event-free survival; GSEA, gene set enrichment analysis; ncRNA, non-coding RNA; OS, overall survival; ROC, receiver operating characteristic; rRNA, ribosomal RNA.

specific bias and obtained a parsimonious marker set with robust cross-cohort reproducibility. The resulting signature stratified survival in the training cohort and retained predictive value in two independent validation datasets without model re-fitting. In addition to its prognostic performance, the signature was associated with biologically coherent pathway differences, with high-risk tumors enriched for translational and biosynthetic programs and low-risk tumors enriched for adaptive immune-related

pathways. Collectively, these findings support the potential utility of this *ALK*-associated three-gene signature as a biologically informed prognostic tool in NB.

A key observation in this study is that *DLK1* and *RTL1* (*PEG11*) are both located within the *DLK1-DIO3* imprinted domain on chromosome 14q32, a developmentally regulated locus that contains paternally expressed protein-coding genes (including *DLK1*, *RTL1*, and *DIO3*) as well as maternally expressed long non-coding RNAs and a

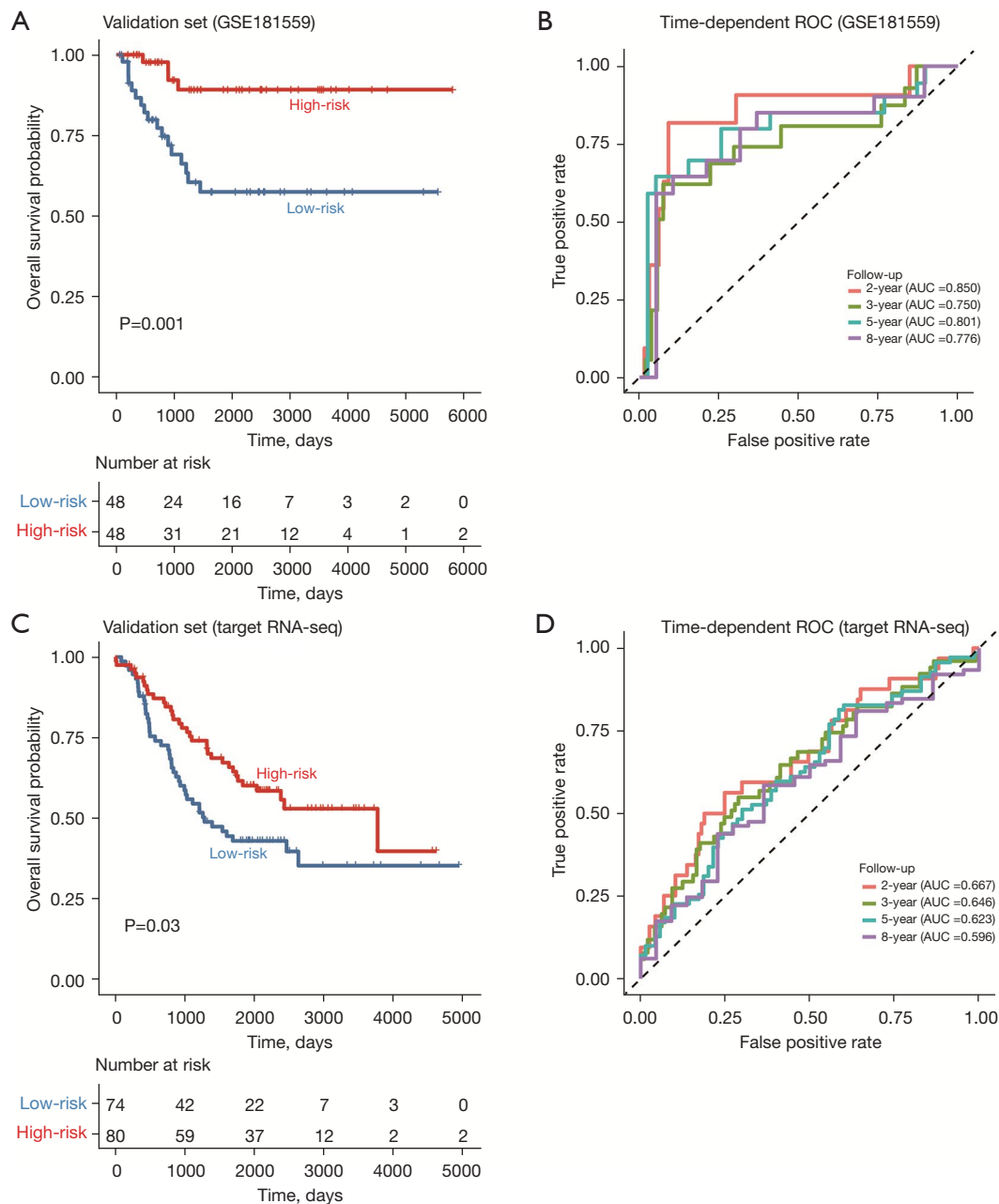


Figure 3 External validation of the three-gene signature across independent cohorts. (A,B) Validation in GSE181559: (A) Kaplan-Meier OS analysis for high- and low-risk groups stratified by the median RiskScore within the cohort, and (B) time-dependent ROC curves. (C,D) Validation in TARGET-NBL: (C) Kaplan-Meier OS analysis and (D) time-dependent ROC curves using the same scoring procedure. In all validation analyses, the RiskScore was computed using the locked coefficients derived from the training cohort. AUC, area under the curve; OS, overall survival; RNA-seq, RNA sequencing; ROC, receiver operating characteristic.

large microRNA cluster (8,9). Although *DLK1* and *RTL1* showed coordinated expression consistent with their shared imprinting-domain context, collinearity analysis indicated that this did not introduce problematic instability into the

three-gene prognostic model. Epigenetic regulation of this region can lead to coordinated expression changes among genes within the locus, providing a plausible explanation for the consistent co-occurrence of *DLK1* and *RTL1* in

our multi-context analyses. Importantly, this domain-level coordination does not imply a direct upstream–downstream regulatory relationship between *DLK1* and *RTL1*; rather, it supports the interpretation that these genes may jointly reflect an *ALK*-associated transcriptional context linked to developmental or differentiation-related programs. This genomic and biological framework strengthens the plausibility of the identified three-gene signature while avoiding overinterpretation of causality.

In contrast to *DLK1* and *RTL1*, *ISLR2* is located on chromosome 15q24 and is not genomically linked to the *DLK1-DIO3* imprinted domain (10). *ISLR2* belongs to the leucine-rich repeat and immunoglobulin superfamily and has been implicated in neuronal connectivity, axon guidance, and neural development. Its inclusion in the present signature, therefore, likely reflects an independent, lineage-related dimension of *ALK*-associated transcriptional states rather than a shared imprinted regulatory mechanism. From this perspective, the three-gene signature may be understood as a set of parallel transcriptional readouts, with *DLK1* and *RTL1* representing developmentally regulated, imprinting-associated signals and *ISLR2* reflecting a distinct neural adhesion or differentiation-related feature. This interpretation provides a biologically coherent framework for the signature while remaining appropriately cautious with respect to mechanistic inference.

The biological differences observed between the model-defined risk groups are consistent with the known signaling consequences of *ALK* activation in NB. Activating *ALK* mutations, particularly the recurrent hotspot variants F1174, F1245, and R1275, can drive downstream PI3K-AKT-mTOR and RAS-MAPK signaling, thereby promoting anabolic activity, protein synthesis, and cellular proliferation. In line with this, high-risk tumors in our study were enriched for pathways related to translation, ribosome biogenesis, and rRNA processing, suggesting a more biosynthetically active and proliferative transcriptional state. By contrast, low-risk tumors were enriched for adaptive immune-related pathways, including T-cell activation and antigen receptor signaling, indicating a relatively more immune-active tumor context. These findings suggest that the three-gene signature captures not only prognostic differences but also biologically meaningful variation in tumor-intrinsic and tumor microenvironment-associated programs.

From a clinical perspective, this three-gene signature offers several practical advantages. It is based on only three transcriptomic features, can be readily derived from routine bulk RNA-seq data, and shows reproducible prognostic

performance across independent cohorts without re-fitting. These characteristics make it a potentially useful complement to established clinical risk factors. However, because age, disease stage, and *MYCN* status were not incorporated into model construction, the independent prognostic value of the three-gene signature remains to be determined in future integrated clinicogenomic analyses. In addition to its value for prognostic stratification, the biological patterns associated with the signature suggest testable therapeutic hypotheses. For example, patients with higher scores may represent a subgroup with increased translational activity and could therefore be considered for combination strategies integrating *ALK* inhibition with agents targeting translational or mTOR-related pathways. Conversely, lower-risk tumors enriched for adaptive immune programs may provide a rationale for exploring immune-modulating approaches in selected settings. Although these possibilities require further validation, the present findings provide a biologically grounded framework for future translational investigation.

Several limitations of this study should be acknowledged. First, *ALK* mutation status in the single-cell dataset was assigned at the sample level rather than determined directly at single-cell resolution, and thus, intratumoral mutational heterogeneity could not be fully resolved. Second, the three differential expression datasets were filtered using context-specific statistical thresholds, which may introduce some degree of selection bias despite the requirement for concordant directionality across all comparisons. Third, because the bulk RNA-seq analyses were primarily based on FPKM-formatted data, the study did not use count-based normalization frameworks, which may limit direct comparability with other analytical pipelines. Finally, although the protein-level images were consistent with the transcriptomic findings, systematic experimental validation at the protein and spatial levels was not performed. These limitations should be considered when interpreting the present findings and highlight the need for further prospective and functional validation.

Conclusions

In summary, by integrating single-cell and bulk transcriptomic datasets, we identified a three-gene *ALK*-associated signature composed of *DLK1*, *RTL1*, and *ISLR2* that reproducibly stratified NB outcomes across independent cohorts. The signature was associated with biologically distinct states characterized by enhanced

translational activity in high-risk tumors and adaptive immune activation in low-risk tumors. As a compact and interpretable model, it may complement established risk factors and provide a useful framework for future translational studies in *ALK*-related NB.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2026-1-0396/rc>

Peer Review File: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2026-1-0396/prf>

Funding: This work was supported by the National Natural Science Foundation of China (No. 82403771 to Y.X.), the China Postdoctoral Science Foundation (No. 2024M752393 to Y.X.), the Tianjin Health Technology Project (No. 2022QN106 to X.L.), and the Hebei Provincial Traditional Chinese Medicine Scientific Research Project (No. T2026078 to X.L.).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2026-1-0396/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license).

See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Sharma R, Yadav J, Bhat SA, et al. Emerging Trends in Neuroblastoma Diagnosis, Therapeutics, and Research. *Mol Neurobiol* 2025;62:6423-66.
- Bhoopathi P, Mannangatti P, Emdad L, et al. The quest to develop an effective therapy for neuroblastoma. *J Cell Physiol* 2021;236:7775-91.
- Trigg RM, Turner SD. ALK in Neuroblastoma: Biological and Therapeutic Implications. *Cancers (Basel)* 2018;10:113.
- Carpenter EL, Mossé YP. Targeting ALK in neuroblastoma--preclinical and clinical advancements. *Nat Rev Clin Oncol* 2012;9:391-9.
- Yu W, Biyik-Sit R, Uzun Y, et al. Longitudinal single-cell multiomic atlas of high-risk neuroblastoma reveals chemotherapy-induced tumor microenvironment rewiring. *Nat Genet* 2025;57:1142-54.
- Jansky S, Sharma AK, Körber V, et al. Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. *Nat Genet* 2021;53:683-93.
- Hartlieb SA, Sieverling L, Nadler-Holly M, et al. Alternative lengthening of telomeres in childhood neuroblastoma from genome to proteome. *Nat Commun* 2021;12:1269.
- Wallace C, Smyth DJ, Maisuria-Armer M, et al. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet* 2010;42:68-71.
- Kagami M, Mizuno S, Matsubara K, et al. Epimutations of the IG-DMR and the MEG3-DMR at the 14q32.2 imprinted region in two patients with Silver-Russell Syndrome-compatible phenotype. *Eur J Hum Genet* 2015;23:1062-7.
- Abudureyimu S, Asai N, Enomoto A, et al. Essential Role of Linx/Islr2 in the Development of the Forebrain Anterior Commissure. *Sci Rep* 2018;8:7292.

Cite this article as: Zhu A, Xia Y, Li X, Jin Z, Wang J, Pan Y. Integrative single-cell and bulk transcriptomics identify an *ALK*-associated three-gene signature predicting neuroblastoma outcomes. *Transl Cancer Res* 2026;15(5):421. doi: 10.21037/tcr-2026-1-0396