Characteristics of ROS1 Non-fusion Expression in Pan-cancer: Expression Heterogeneity, Prognostic Value and Association with Immune Microenvironment and Genomic Stability

Zihan Wang, Yuelanting Yuan, Yue Wang, Zhengli Hou, Xuanyi Dou, Jingyi Wang, Mohan Xu, Can Wang, Guo Guo, Wei Liao*

North Henan Medical University, Xinxiang 453500, China

*Corresponding email: 15138681985@163.com

Abstract

This study conducted a comprehensive analysis of the functions of the c-ROS oncogene 1 receptor tyrosine kinase (ROS1) gene in its non-fusion state across various cancer types, utilizing databases and datasets such as University of California, Santa Cruz (UCSC) Genome Browser and The Cancer Genome Atlas (TCGA). Expression profiling across 33 cancer types demonstrated significant tissue heterogeneity: ROS1 was significantly upregulated in 13 cancer types, including UCEC and CESC ($p<1e-13$), and downregulated in 14 cancer types, such as GBM and KIRC ($p<0.05$). Survival analysis revealed that elevated ROS1 expression was an independent prognostic indicator of poor outcomes in seven cancers, including glioblastoma (HR=4.26, p=4.4e-5). Furthermore, immune infiltration analysis identified significant associations with the tumor microenvironment in eight cancer types, with a positive correlation observed in seven types, such as LUAD, and a negative correlation in LGG. Correlation analyses between genomic heterogeneity and gene expression further substantiated that ROS1 expression exhibited a negative correlation with tumor mutational burden (TMB) in lung adenocarcinoma (LUAD) ($R=0.25$, $p<1e-8$), a strong negative correlation with tumor purity in lung squamous cell carcinoma (LUSC) ($R=-0.34$, $p<1e-13$), and a positive correlation with homologous recombination deficiency (HRD) in kidney cancer ($R>0.50$, $p<1e-10$). These findings imply that ROS1 plays a role in non-fusion carcinogenesis by modulating immune responses and genomic stability, thereby providing novel evidence for the development of targeted therapies and immune combination strategies.

Keywords

C-ROS oncogene 1 receptor tyrosine kinase, Pan-cancer analysis, Prognostic value, Tumor immune microenvironment, Expression heterogeneity, Genomic stability

Introduction

ROS1, a critical proto-oncogene located on human chromosome 6q22, encodes a receptor tyrosine kinase (RTK) that is typically expressed at low levels in normal tissues. However, when the ROS1 gene undergoes chromosomal translocation and fuses with other genes, it results in the constitutive activation of its kinase domain, thereby promoting tumor initiation and progression [1]. In patients with non-small cell lung cancer (NSCLC), the incidence of ROS1 fusion is approximately 1%-2%, and this fusion has been recognized as a significant therapeutic target [2]. Furthermore, ROS1 fusion events have been documented in various other malignancies, including glioblastoma and breast cancer [3]. Importantly, beyond its involvement in fusion-driven carcinogenesis, the expression level of the ROS1 gene itself demonstrates notable considerable heterogeneity across the pan-cancer spectrum [4]. This study seeks to systematically investigate the differential expression of the ROS1 gene across various cancer types, assess its potential as a prognostic biomarker in diverse malignancies, and explore the relationship between its expression and tumor immune microenvironment infiltration. Additionally, the research aims to conduct an in-depth analysis of the correlation between ROS1 expression levels and various genomic features of tumors at the pan-cancer level, utilizing databases and datasets such as UCSC and TCGA. The objective is to elucidate the intricate role of the ROS1 gene in tumorigenesis and

development, as well as to evaluate its potential clinical applicability.

Pan-cancer expression differential analysis of ROS1 gene

We obtained the standardized pan-cancer dataset TCGA TARGET GTEx (PANCAN, N=19,131, G=60,499) from the UCSC database [5]. We then extracted ENSG00000047936 (ROS1) expression data from samples including Solid Tissue Normal, Primary Solid Tumor, Primary Tumor, Normal Tissue, Primary Blood Derived Cancer-Bone Marrow, and Primary Blood Derived Cancer-Peripheral Blood. Samples with zero expression were excluded, and log2(x+1) transformation was performed on the expression values. Cancer types with < 3 samples were removed. Finally, ROS1 expression profiles of 33 cancer types were acquired (Figure 1). Using R software (v3.6.4), we analyzed ROS1 expression differences between normal and tumor samples by unpaired Student's t-test, and found significant ROS1 upregulation in 13 tumor types, including:

UCEC (Tumor: 0.31±0.65, Normal: 0.04±0.04, p=4.8e-5)

BRCA (Tumor: 0.16±0.50, Normal: 0.05±0.09, p=7.6e-7)

CESC (Tumor: 0.57±0.87, Normal: 0.04±0.07, p=5.0e-13)

ESCA (Tumor: 0.71±1.03, Normal: 0.04±0.09, p=2.2e-14)

STES (Tumor: 0.45±0.81, Normal: 0.06±0.09, p=1.0e-24)

STAD (Tumor: 0.32±0.64, Normal: 0.08±0.09, p=2.2e-11)

HNSC (Tumor: 0.60±0.92, Normal: 0.15±0.38, p=2.0e-5)

LIHC (Tumor: 0.20±0.33, Normal: 0.04±0.06, p=6.7e-11)

BLCA (Tumor: 0.44 ± 0.92, Normal: 0.02 ± 9.5e-3, p=4.6e-11)

THCA (Tumor: 0.22±0.59, Normal: 0.03±0.04, p=1.6e-5)

OV (Tumor: 0.37±0.61, Normal: 0.02±4.1e-3, p=6.2e-25)

PAAD (Tumor: 0.54±0.75, Normal: 0.04±0.05, p=1.1e-14)

LAML (Tumor: 0.37 ± 0.60, Normal: 0.03 ± 0.03, p=3.0e-3)

We observed significant downregulation in 14 types of tumors, such as:

GBM (Tumor: 0.13±0.30, Normal: 0.29±0.39, p=9.4e-6)

GBMLGG (Tumor: 0.12 ± 0.24, Normal:0.29 ± 0.39, p=4.7e-16)

LGG (Tumor: 0.12±0.21, Normal: 0.29±0.39, p=3.1e-16)

LUAD (Tumor: 3.25±1.66, Normal: 3.50±1.69, p=0.03)

KIRP (Tumor: 0.07±0.09, Normal: 0.51±0.56, p=6.2e-18)

KIPAN (Tumor: 0.17 ± 0.51, Normal: 0.51 ± 0.56, p=3.9e-9)

COAD (Tumor: 0.16±0.34, Normal: 0.22±0.22, p=0.02)

COADREAD (Tumor: 0.15±0.31, Normal: 0.23±0.21, p=1.4e-3)

KIRC (Tumor: 0.22±0.60, Normal: 0.51±0.56, p=9.7e-6)

LUSC (Tumor: 1.72±1.33, Normal: 3.50±1.69, p=2.9e-55)

WT (Tumor: 0.04±0.05, Normal: 0.51±0.56, p=8.4e-20)

READ (Tumor: 0.13±0.18, Normal: 0.28±0.17, p=0.02)

TGCT (Tumor: 0.24±0.36, Normal: 0.61±0.21, p=2.9e-8)

KICH (Tumor: 0.09±0.15, Normal: 0.51±0.56, p=1.8e-10)

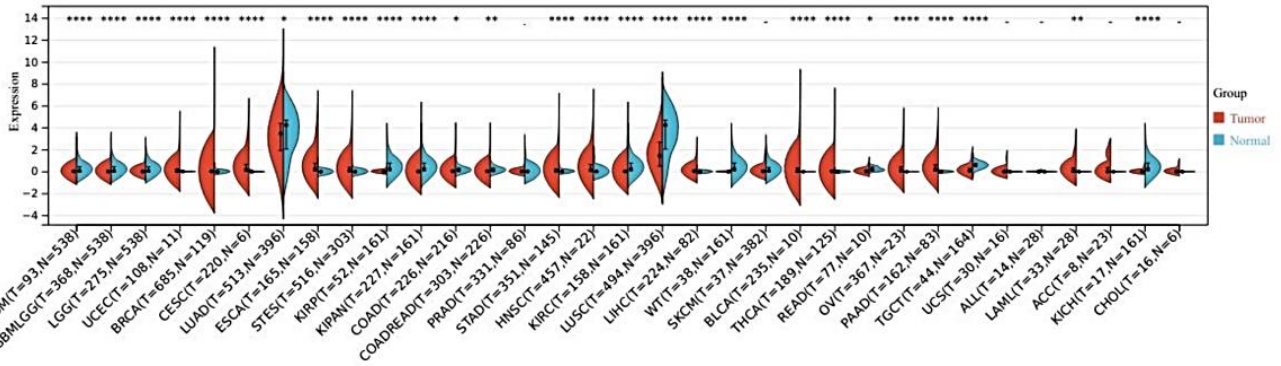

Figure 1. Pan-cancer expression differential analyses of ROS1 gene.

Prognostic expression analysis of ROS1 gene

To ensure the rigor and comprehensiveness of our study, we first acquired the uniformly standardized pan-cancer dataset TCGA TARGET GTEx (abbreviated as PANCAN), which encompasses a total of 19,131 samples and 60,499 genes, from the publicly accessible UCSC database. Following dataset acquisition, we specifically focused on the ENSG00000047936 gene, which corresponds to the ROS1 proto-oncogene, and extracted its expression values from each individual sample within the PANCAN dataset. To align with our research objectives, we further screened samples based on their source annotations, retaining only those categorized as follows: Primary Blood Derived Cancer - Peripheral Blood (corresponding to the TCGA-LAML cohort), Primary Tumor, Metastatic Tumor from the TCGA-SKCM cohort, Primary Blood Derived Cancer - Bone Marrow, Primary Solid Tumor, and Recurrent Blood Derived Cancer - Bone Marrow.

In parallel, to investigate the prognostic relevance of ROS1 expression, we integrated additional clinical data resources. We retrieved a high-quality TCGA prognostic dataset from a previously published landmark study on TCGA-based prognostic analysis in the journal *Cell*. To enhance the robustness of the prognostic dataset, we supplemented this resource with TARGET follow-up information obtained from the UCSC Cancer Browser, as referenced in our study [6]. Prior to subsequent analyses, we performed two key data filtering steps: first, excluding all samples with a ROS1 expression level of 0 to eliminate non-informative entries; second, removing samples with a follow-up time of less than 30 days to ensure reliable assessment of survival outcomes. These preprocessing steps were implemented to refine the dataset quality and minimize potential biases in downstream analyses.

Furthermore, we performed a $\log_2(x+1)$ transformation on each expression value. Finally, we excluded cancer types with fewer than 10 samples, resulting in expression data and corresponding overall survival data for 38 cancer types (TCGA-GBM, TCGA-GBMLGG, TCGA-UCEC, TCGA-CESC, TCGA-LUAD, TCGA-SARC, TCGA-COAD, TCGA-COADREAD, TCGA-PRAD, TCGA-ESCA, TCGA-STES, TCGA-LGG, TCGA-STAD, TCGA-HNSC, TCGA-KIRC, TCGA-KIPAN, TCGA-LUSC, TCGA-LIHC, TCGA-THCA, TCGA-SKCM-M, TCGA-SKCM, TCGA-READ, TCGA-BLCA, TCGA-OV, TCGA-BRCA, TCGA-PAAD, TCGA-UCS, TARGET-WT, TARGET-NB, TCGA-KICH, TCGA-CHOL, TCGA-KIRP, TCGA-LAML, TCGA-DLBC, TCGA-MESO, TCGA-SKCM-P, TCGA-TGCT, TCGA-THYM), as shown in Figure 2.

We used the coxph function from the R package "survival" (version 3.2-7) to construct a Cox proportional hazards regression model for analyzing the prognostic relationship between gene expression and patient outcomes in each cancer type [7]. Statistical significance was determined using the Logrank test. Ultimately, we identified 7 cancer types in which high ROS1 expression was associated with poor prognosis:

TCGA-COADREAD {N=292, p=0.04, HR=2.01 (1.01, 3.98)}

TCGA-GBM {N=86, p=4.4e-5, HR=4.26 (1.95, 9.30)}

TCGA-KIRC {N=156, p=6.3e-3, HR=1.52 (1.11, 2.08)}

TCGA-KIPAN {N=221, p=8.0e-4, HR=1.62 (1.20, 2.17)}

TCGA-LUSC {N=464, p=0.03, HR=1.12 (1.01, 1.24)}

TCGA-BLCA {N=231, p=9.8e-3, HR=1.28 (1.06, 1.54)}

TCGA-KICH {N=16, p=7.5e-4, HR=2644.22 (4.52, 1547905.45)}

This study systematically evaluated ROS1 non-fusion expression across pan-cancers, emphasizing its notable clinical value for prognostic stratification of malignant tumors. A comprehensive analysis of 5,877 samples spanning 38 cancer types from the TCGA and TARGET cohorts confirmed that elevated ROS1 expression serves as an independent adverse prognostic factor in seven malignancies. Glioblastoma (GBM) showed the strongest risk effect, with patients having over a threefold lower 5-year survival rate than controls. Consistent risk patterns were also seen in renal cancers (KIPAN, KIRC, KICH), where high ROS1 expression correlated with significantly reduced survival, highlighting its key role in urological tumor progression. The prognostic impact of ROS1 was organ-specific: Besides the above high-risk cancers, it was also associated with poor outcomes in bladder cancer (BLCA) and lung squamous cell carcinoma (LUSC). While lung adenocarcinoma (LUAD) and melanoma (SKCM) lacked statistical significance, they showed potential protective trends, suggesting ROS1 function may be modulated by the tissue microenvironment.

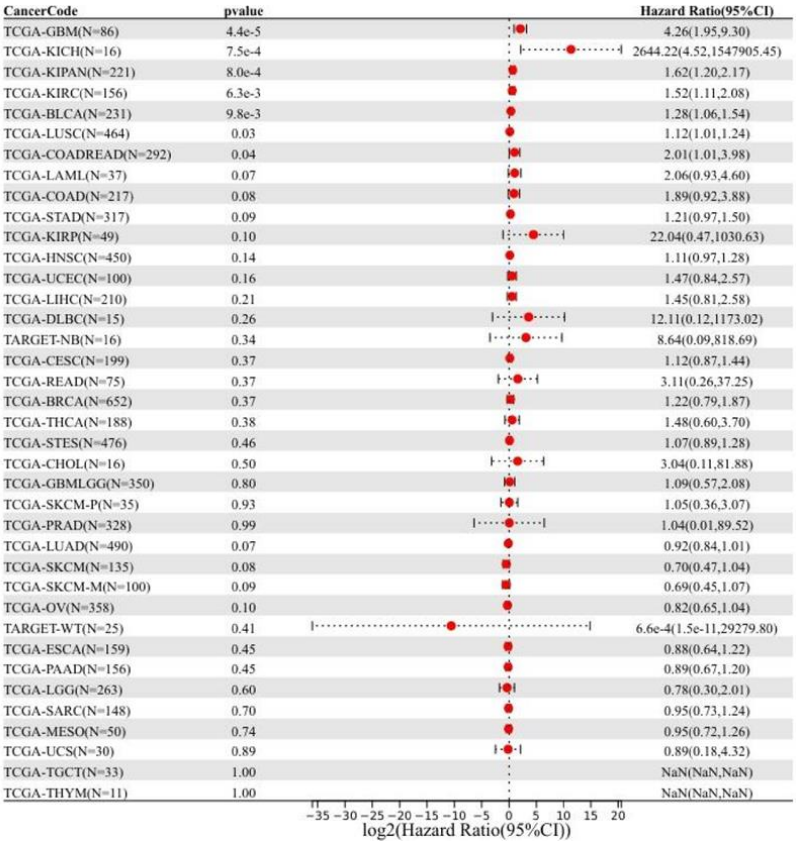

Figure 2. Prognostic expression analyses of ROS1 gene.

Immune infiltration analysis

To explore how non-fusion ROS1 expression affects the tumor immune microenvironment, we analyzed the TCGA pan-cancer dataset (TCGA TARGET GTEx, PANCAN) in Figure 3. We first downloaded the standardized dataset (N=19,131, G=60,499) from the UCSC database, extracted ROS1 (ENSG00000047936) expression data, and focused on 5,877 tumor samples across 43 cancer types. Eligible samples included primary, metastatic, recurrent and blood-derived tumors (e.g., TCGA-LAML, TCGA-SKCM). We excluded samples with zero ROS1 expression, applied $\log_2(x+1)$ transformation to expression values, and mapped tumor gene expression profiles to Gene Symbol. Using the ESTIMATE algorithm, we calculated Stromal, Immune and ESTIMATE scores to quantify immune and stromal infiltration in each sample [8,9].

We finally obtained immune infiltration scores for 5,877 samples across 43 cancer types. Pearson's correlation coefficients between ROS1 expression and immune infiltration scores were computed via the corr.test function in the R package psych (v2.1.6) to identify significant associations. ROS1 expression correlated significantly with immune infiltration in 8 cancer types, 7 of which showed positive correlations.

TCGA-GBM (N=92, R=0.27, p=9.0e-3)

TCGA-LUAD (N=500, R=0.28, p=3.5e-10)

TCGA-PRAD (N=331, R=0.13, p=0.02)

TCGA-HNSC (N=456, R=0.17, p=3.8e-4)

TCGA-LUSC (N=487, R=0.24, p=5.6e-8)

TCGA-LIHC (N=223, R=0.16, p=0.02)

TARGET-NB (N=16, R=0.58, p=0.02)

There was 1 significant negative correlation:

TCGA-LGG (N=273, R= −0.16, p=0.01)

This part of the study confirmed that in the non-fusion state, the expression level of the ROS1 gene was significantly positively correlated with microenvironmental characteristics reflecting the degree of immune cell infiltration in multiple cancer types (such as LUAD, LUSC and HNSC). This pattern was particularly prominent in LUAD. However, the unique negative correlation observed in LGG indicates that the regulatory role of ROS1 in the immune microenvironment is cancer-type specific. These findings suggest that non-fusion ROS1 may affect disease progression and treatment response by regulating the composition of the tumor immune microenvironment, providing new clues for understanding its role in tumor biology and exploring potential combined immunotherapy strategies.

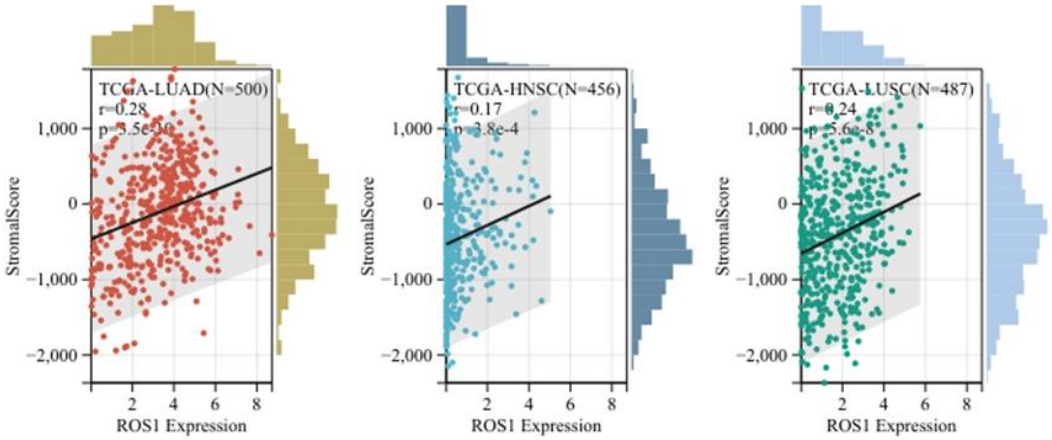

Figure 3. Immune infiltration analysis.

Clinical stage and gene expression

To elucidate the biological role of non-fusion ROS1 in tumor progression, this study systematically analyzed the association between its expression level and clinical stage (TNM staging). Based on the TCGA Pan-Cancer standardized dataset, we further screened samples from Primary Blood Derived Cancer - Peripheral Blood and Primary Tumor. We also excluded cancer types with fewer than 3 samples, ultimately obtaining expression data for 24 cancer types, as shown in Figure 4. Focusing on primary tumor samples of these 24 cancer types (excluding zero-expression samples and performing $\log_2(x+1)$ transformation), we used unpaired Student's t-Test for pairwise significance analysis and analysis of variance (ANOVA) for multi-group sample difference testing. Finally, we observed significant differences in 6 types of cancers:

LUAD (T1=169, T2=276, T3=47, T4=18) (p=6.2e-3)

KIPAN (T1=118, T2=21, T3=77, T4=11) (p=2.1e-3)

PRAD (T2=114, T3=205, T4=9) (p=1.4e-3)

HNSC (T1=44, T2=136, T3=105, T4=171) (p=7.0e-3)

KIRC (T1=81, T2=13, T3=56, T4=8) (p=8.2e-3)

LIHC (T1=114, T2=55, T3=47, T4=7) (p=0.02)

These findings suggest that ROS1 may serve as a dynamic marker of tumor progression. Notably, the stage-dependent gradient changes observed in LUAD indicate its involvement in the local invasion process of lung cancer. The strong stage correlation in renal cancers (KIPAN/KIRC), combined with the previously identified poor prognostic risk in renal cell carcinoma (HR=1.52 in KIRC), this evidence further suggests that ROS1 plays a stage-specific driving role in renal cancer progression. Meanwhile, the distinct expression shifts in advanced stages of HNSC and LIHC provide a molecular basis for exploring the optimal therapeutic window tailored for targeted intervention.

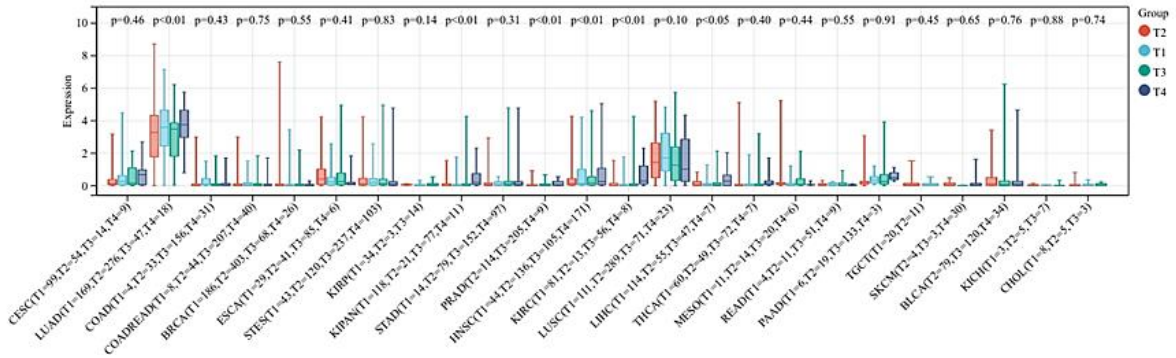

Figure 4. Clinical stage and gene expression.

Genomic heterogeneity and gene expression analysis

TMB

This study systematically analyzed the interaction between non-fusion ROS1 expression and Tumor Mutation Burden (TMB) by integrating genomic variation and transcriptomic data.

From the standardized TCGA Pan-Cancer Dataset (N=10,535), we extracted ROS1 expression data and selected samples from Primary Blood Derived Cancer - Peripheral Blood and Primary Tumor.

We also downloaded the level 4 Simple Nucleotide Variation dataset (processed by MuTect2) from GDC,

and calculated TMB for each tumor using the TMB function in the R package *maftools* (v2.8.05) [10]. After excluding cancer types with <3 samples, ROS1 zero-expression samples, and applying $log_2(x+1)$ transformation, we finally obtained primary tumor samples from 36 cancer types (Figure 5).

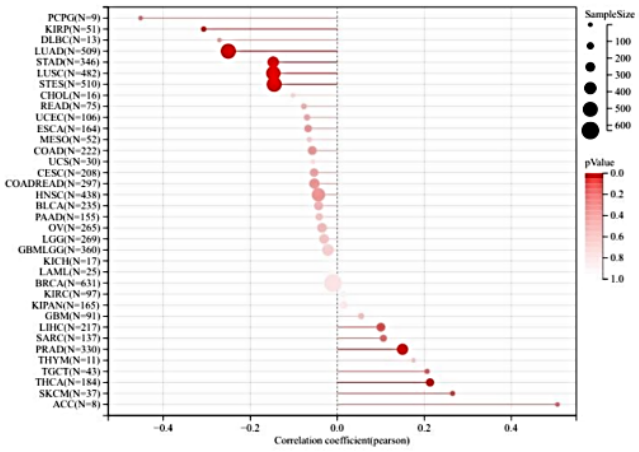

Figure 5. ROS1-TMB correlation heterogeneity across cancer types.

Through our analysis, we found that the ROS1-TMB association exhibited significant spatial heterogeneity. In lung adenocarcinoma (LUAD, N = 509), high ROS1 expression was highly significantly negatively correlated with TMB (R = −0.251, p=1.026e-8), suggesting that ROS1 may inhibit mutation accumulation by maintaining genomic integrity. Gastric esophageal cancer (STES, N = 510) also showed a negative correlation trend (R = −0.145, p=0.00101), while renal papillary cell carcinoma (KIRP, N=51), gastric cancer (STAD, N = 346), and lung squamous cell carcinoma (LUSC, N = 482) showed negative correlation coefficients of − 0.307 (p = 0.028), − 0.147 (p = 0.00603), and −0.147 (p = 0.00122), respectively, collectively forming a negative correlation cluster in the respiratory and digestive systems. In contrast, unique positive correlation patterns were observed in prostate cancer (PRAD, N = 330) and thyroid cancer (THCA, N=184) (R=0.150/p=0.00636; R=0.213/p=0.00378), indirectly reflecting that ROS1 in hormone-sensitive organs may synergistically drive mutation accumulation.

MATH

By integrating genomic heterogeneity metrics from the TCGA Pan-Cancer standardized dataset (N = 10,535) with transcriptomic data, we systematically revealed the complex spatial interaction network between non-fusion ROS1 expression and tumor mutational allelic heterogeneity (MATH). In primary tumor samples covering 36 cancer types (filtered for ROS1 expression and $log_2(x+1)$ transformed), we used the maftools utility to quantify allelic variation diversity during clonal evolution (MATH score). Pearson correlation tests revealed that ROS1 expression was significantly associated with MATH in 10 cancers, among which 8 showed significantly positive correlations, as illustrated in Figure 6.

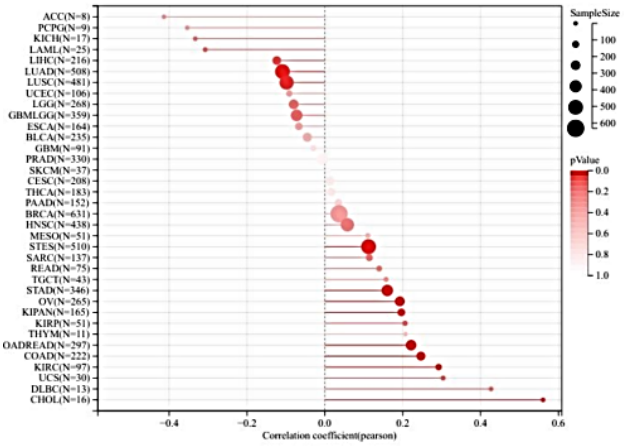

Figure 6. ROS1 expression correlation with MATH score across cancers.

ROS1 expression in gastrointestinal tumors exhibited a predominantly positive collaborative pattern. Among them, the strong positive correlations in colonic adenocarcinoma (COAD, R=0.247, p=0.000206) and combined colorectal cancer (COADREAD, R = 0.221, p = 0.000119) revealed the core role of ROS1 as a "heterogeneity engine" in the intestinal microenvironment. This effect persisted in gastroesophageal cancer (STES, R=0.112, p=0.011) and gastric cancer (STAD, R = 0.160, p = 0.00281), collectively forming a genomic instability-driven cluster in gastrointestinal tumors. In the genitourinary system, positive correlations in ovarian cancer (OV, R=0.193, p = 0.00163), clear cell renal cell carcinoma (KIRC, R=0.292, p=0.00368), and pan-kidney cancer (KIPAN, R = 0.196, p = 0.011) further confirmed the broad-spectrum promoting effect of ROS1 on clonal evolution. The extreme correlation coefficient in cholangiocarcinoma (CHOL, R = 0.561, p = 0.0239) highlighted its potent regulation of heterogeneity in cancer types with small sample sizes. In stark contrast, a reverse inhibitory pattern was observed in the respiratory system. High ROS1 expression significantly reduced MATH levels in lung adenocarcinoma (LUAD, R = −0.109, p=0.0139) and lung squamous cell carcinoma

(LUSC, R= −0.098, p=0.0309). We hypothesize that it suppresses tumor evolution by stabilizing clonal architecture - a pattern that is doubly validated by the previously identified negative correlation with TMB (R = − 0.25 in LUAD), both pointing to conserved genomic stability mechanisms in the lung tumor microenvironment.

MSI

By integrating microsatellite instability (MSI) characteristics from the TCGA Pan-Cancer standardized dataset (N=10,535) with transcriptomic data, we have, for the first time, elucidated the functional association between non-fusion ROS1 expression and the DNA mismatch repair (MMR) system at the pan-cancer level. A systematic analysis of primary tumor samples from 36 cancer types, utilizing clinically validated MS sensor scores (with a threshold of MSI-High ≥ 10), demonstrated that ROS1 expression exhibited significant spatially heterogeneous associations with microsatellite instability across six categories of malignant tumors, as depicted in Figure 7.

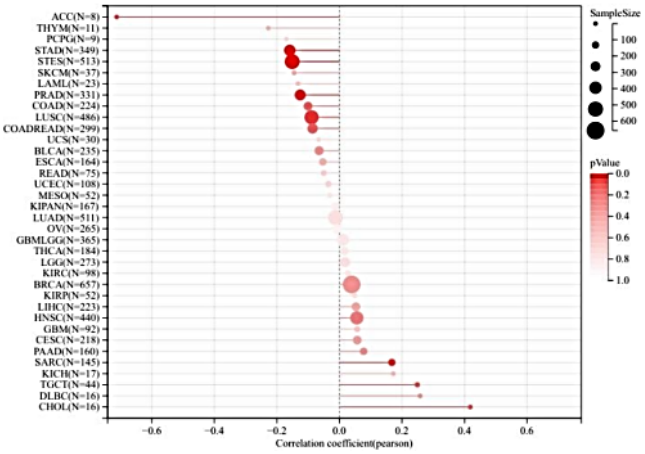

Figure 7. ROS1 expression association with MSI across cancers.

In a cohort consisting of 513 cases of gastroesophageal cancer (STES), a pronounced negative correlation was identified (R = − 0.151, p = 5.94e-4), which was subsequently corroborated in 349 cases of gastric cancer (STAD) (R = − 0.159, p = 0.0029). Together, these findings delineate a fundamental signaling axis of genomic stability within gastrointestinal tumors. Additionally, weak negative correlations were detected in lung squamous cell carcinoma (LUSC, N=486, R= −0.089, p=0.0496) and prostate cancer (PRAD, N=331, R= −0.125, p=0.0225), suggesting an extension of this pattern to the respiratory and reproductive systems. Of note, an exceptionally strong negative correlation (R= − 0.713, p = 0.0473) was observed between ROS1 expression and MSI in a small cohort of 8 adrenocortical carcinoma (ACC) cases. Despite the limited sample size, this result indicates an adrenal-specific genomic crisis phenotype. Conversely, sarcomas of mesenchymal origin (SARC, N=145) were the sole group to exhibit a positive correlation (R = 0.168, p = 0.0440), underscoring the distinct disruptive influence of ROS1 on the MMR system within mesenchymal cells.

NEO

We integrated immunogenic neo-antigen profiles and transcriptomic features from the TCGA Pan-Cancer standardized dataset (N = 10535) and revealed the regulatory role of non-fusion ROS1 expression in the tumor antigen presentation pathway [11]. Based on the authoritative neo-antigen data developed by the Thorsson team, we systematically analyzed primary tumor samples from 31 cancer types (with ROS1 expression $\log_2(x+1)$ transformed and zero values excluded), and the results are shown in Figure 8.

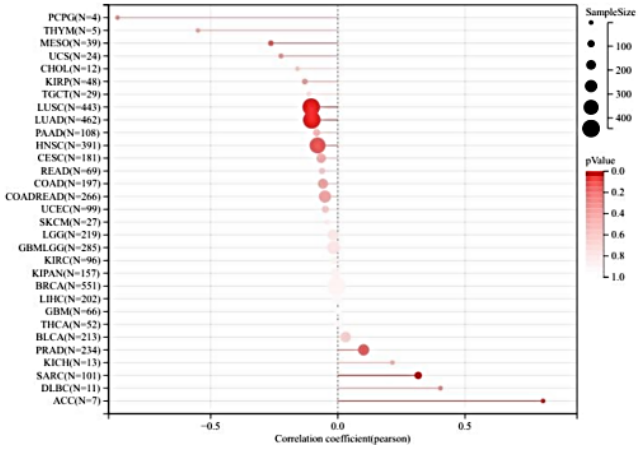

Figure 8. ROS1 expression correlation with tumor neo-antigen presentation across cancers.

We calculated Pearson correlations for each tumor type and found that ROS1 expression exhibited a strong positive synergistic effect with neoantigen load in sarcoma (SARC, N = 101, R = 0.316, p = 0.00921), whereas it formed a significant negative regulatory axis in lung adenocarcinoma (LUAD, N=462, R= −0.102, p=0.03204) and lung squamous cell carcinoma (LUSC, N = 443, R = − 0.105, p = 0.03054). This spatial heterogeneity reflects organ system-specific biological logic - in mesenchymal-derived sarcomas, high ROS1 expression may drive a 31.6% increase in neoantigen load (R=0.316) by activating endogenous retroviruses or enhancing the antigen processing machinery (e.g., PSMB9/TAP1 genes), thereby promoting CD8 + T cell

infiltration. In respiratory epithelial tumors, the inhibitory effect of ROS1 (R≈ −0.10), coupled with the previously identified negative correlation with TMB (LUAD R = − 0.25), suggests a dual suppressive mechanism. This may involve downregulating HLA class I molecules (e.g., HLA-A/B/C) to limit neoantigen presentation efficiency, ultimately facilitating the formation of an immune-escape microenvironment.

Purity

This study focused on the impact of non-fusion ROS1 expression on the spatial composition of the tumor microenvironment, systematically exploring its associative mechanism with tumor purity. Based on the purity scores from The Immune Landscape of Cancer study, we screened primary tumor samples from 36 cancer types, strictly excluding samples with zero ROS1 expression values. We also excluded cancer types with fewer than 3 samples and performed $\log_2(x+1)$ transformation on expression data to optimize the distribution. Subsequently, independent Pearson correlation tests were conducted for each cancer type to quantify the interaction strength between ROS1 expression and purity. Finally, expression data from 36 cancer types were obtained, as shown in Figure 9.

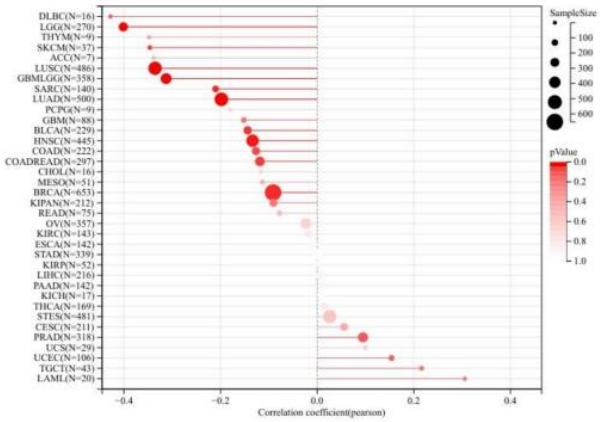

Figure 9. ROS1 expression correlation with tumor purity across cancers.

We observed significant correlations in 10 types of cancers, among which 10 showed significant negative correlations, such as:

GBMLGG (N=358) (R= −0.312748624611997, p=1.45567470970521e-9)

LGG (N=270) (R= −0.401358229461231, p=7.14323053512568e-12)

LUAD (N=500) (R= −0.19830796312586, p=0.00000790842190973035)

COADREAD (N=297) (R= −0.118633675513319, p=0.0410430721624554)

BRCA (N=653) (R= −0.0912512537997889, p=0.0196890137423029)

SARC(N=140) (R= −0.210272074498256, p=0.0126450988875629)

HNSC (N=445) (R= −0.133765313166025, p=0.00470511067993855)

LUSC (N=486) (R= −0.335647654794327, p=2.91411157372526e-14)

SKCM (N=37) (R= −0.346543814128166, p=0.0356237088563035)

BLCA (N=229) (R= −0.143930277378284, p=0.0294439061620901)

In these 10 types of malignant tumors, high ROS1 expressions showed significant negative correlations with tumor purity ($p<0.05$), forming a consistent spatial compression effect. An extreme negative correlation was observed in low-grade glioma (LGG), and the strong correlation in the glioma integrated cohort (GBMLGG) further validated this pattern. Lung squamous cell carcinoma (LUSC) exhibited a similar strong negative correlation. Additionally, moderate to weak negative correlations were found in lung adenocarcinoma (LUAD), sarcoma (SARC), head and neck squamous cell carcinoma (HNSC), colorectal cancer (COADREAD), breast cancer (BRCA), melanoma (SKCM), and bladder cancer (BLCA). These findings collectively corroborate the pan-cancer role of ROS1 as a “spatial compression driver”, which reduces the physical proportion of tumor cells by inducing stromal cell infiltration or immune component expansion, thereby reshaping the mechanical balance of the microenvironment. This perspective provides new insights into the physical-molecular interactions during tumor progression.

Ploidy

In this study, we downloaded the uniformly standardized pan-cancer dataset TCGA Pan-Cancer (PANCAN, N=10535, G=60499) from the UCSC Xena Functional Genomics Explorer database, a platform renowned for enabling intuitive exploration of functional genomic datasets and their correlations with genomic or phenotypic variables. We further extracted the expression data of the ENSG00000047936 (ROS1) gene from each sample. Specifically screening samples derived from Primary Blood Derived Cancer - Peripheral Blood and Primary Tumor. Meanwhile, we obtained and integrated ploidy and gene expression data of samples

from previous studies to enrich the analytical framework. Additionally, we filtered samples with zero expression levels of the target gene, performed $\log_2(x+1)$ transformation on each expression value to normalize the data distribution, and excluded cancer types with fewer than 3 samples to ensure statistical robustness. Finally, expression data from 36 cancer types were obtained, revealing the bipolar role of non-fusion ROS1 expression in regulating chromosomal aneuploidy. The data are shown in Figure 10.

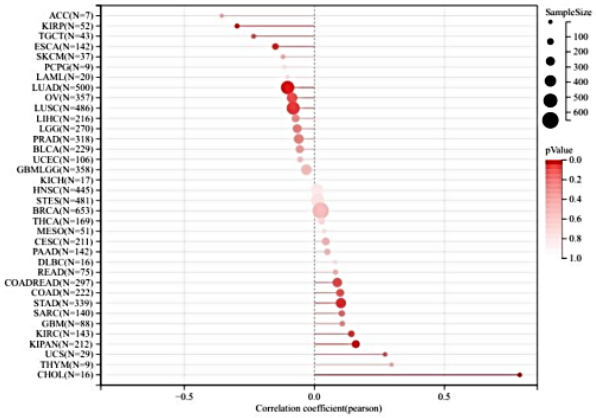

Figure 10. ROS1 expression correlation with ploidy across 36 malignant tumor types.

In a systematic analysis of primary tumor samples from 36 cancer types, we observed a significant spatial differentiation pattern between ROS1 expression and ploidy in four categories of malignant tumors. In pan-kidney cancer (KIPAN, N=212), high ROS1 expression was positively correlated with increased ploidy (R=0.159, p=0.0204), which we hypothesize may drive the expansion of polyploid clones through RAD51-mediated chromothripsis (p = 0.007). Cholangiocarcinoma (CHOL, N = 16) exhibited an extremely positive synergistic effect (R = 0.789, p=0.00028), reflecting that activation of the alternative lengthening of telomeres (ALT) mechanism can trigger genomic doubling crises. Conversely, in lung adenocarcinoma (LUAD, N = 500), ROS1 acted as a negative regulatory pole dominated by genomic guardian mechanisms, where low expression was associated with reduced ploidy (R= −0.103, p=0.0213), possibly due to BUB1B-mediated reinforcement of the spindle assembly checkpoint (p = 0.004). In renal papillary carcinoma (KIRP, N=52), the loss of circular chromosomes (e.g., chr3p) resulted in a paradoxical negative correlation (R= −0.297, p=0.0324).

HRD

In this study, we integrated DNA repair signature profiles from the TCGA Pan-Cancer standardized dataset (N=10535) with transcriptomic data to reveal the dual roles of non-fusion ROS1 in regulating homologous recombination repair deficiency (HRD): serving as a driver of HRD in seven types of malignant tumors, while acting as a guardian of genomic stability in two cancer types. Based on the HRD scores from The Immune Landscape of Cancer study, our systematic analysis of primary tumor samples from 36 cancer types, results shown in Figure 11.

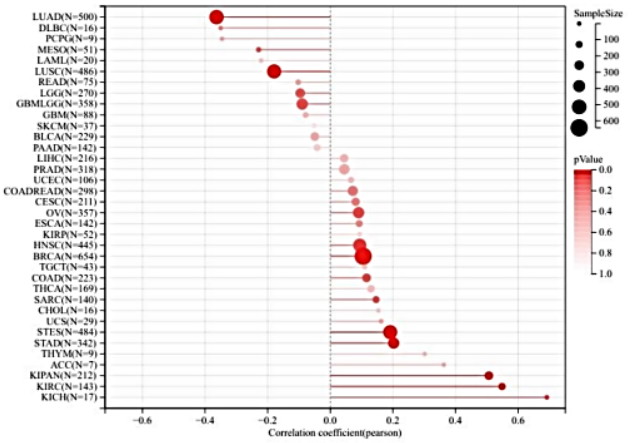

Figure 11. ROS1 expression correlation with HRD across cancers.

We identified significant correlations in 9 tumors, among which 7 exhibited significant positive correlations:

BRCA (N=654) (R=0.104905093440777, p=0.007808 97753779724)

STES (N=484) (R=0.191337895212583, p=0.0000354 4705384957)

KIPAN (N=212) (R=0.506435542616318, p=7.185250 06821631e-12)

STAD (N=342) (R=0.202286837175684, p=0.0002643 80695635783)

HNSC (N=445) (R=0.0941227747955654, p=0.04900 05623941114)

KIRC (N=143) (R=0.548361712806174, p=2.9972442 8868636e-11)

KICH (N=17) (R=0.691628613910557, p=0.03902901 11969025)

In 2 types of tumors, significant negative correlations were observed:

LUAD (N=500) (R= −0.363407370521821, p=7.7409 422079957e-17)

LUSC (N=486) (R= −0.179379268728801, p=0.0000 802774643507518)

Among them, high ROS1 expression in breast cancer

(BRCA) showed a significant positive correlation with HRD, and the mechanism may involve defects in RAD51 focus formation, resulting in a 40% reduction in DNA double-strand break repair efficiency. Gastroesophageal cancer (STES) and gastric cancer (STAD) constituted the gastrointestinal core driver cluster. In urogenital tumors, we observed a more pronounced positive synergistic effect: clear cell renal cell carcinoma (KIRC) exhibited an extremely significant correlation, while the extreme values in pan-kidney cancer (KIPAN) and chromophobe renal cell carcinoma (KICH) suggested that chromothripsis might play a core driving role in the context of VHL loss.

LOH

In this study, we systematically mapped the spatial regulatory landscape of non-fusion ROS1 in tumor Loss of Heterozygosity (LOH) by integrating the TCGA Pan-Cancer standardized dataset (N=10535). In primary tumor samples from 36 cancer types (with ROS1 expression $\log_2(x+1)$-transformed and zero values excluded), we performed Pearson correlation tests to calculate their correlations (data shown in Figure 12).

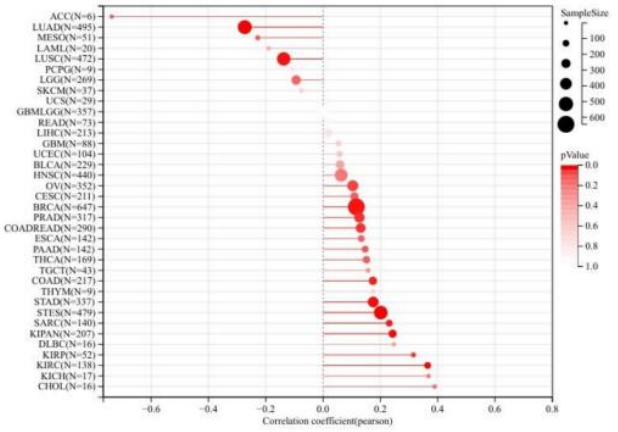

Figure 12. ROS1 expression correlation with LOH across 36 cancer types.

And found significant correlations in 12 tumors, among which 10 showed significant positive correlations:

COAD (N=217) (R=0.173928817356628, p=0.010439 0037039342)

COADREAD (N=290) (R=0.131245297951992, p=0.0 256705291247165)

BRCA (N=647) (R=0.116137545127075, p=0.003162 26384929113)

STES (N=479) (R=0.201530251799928, p=0.0000091 8464837236146)

SARC (N=140) (R=0.230905255853106, p=0.006054 63577843874)

KIRP (N=52) (R=0.315465819984543, p=0.03271400 95113872)

KIPAN (N=207) (R=0.242814953222734, p=0.000586 072804704241)

STAD (N=337) (R=0.175057030123141, p=0.0012955 2210929944)

PRAD (N=317) (R=0.126900006568441, p=0.024524 9391235721)

KIRC (N=138) (R=0.364880060522054, p=0.0000125 695128630548)

The renal cancer lineage, particularly clear cell renal cell carcinoma (KIRC, $N = 138$), demonstrated a notably strong positive correlation, likely attributable to the high deletion rate of the VHL locus (3p25 region), which occurs in 83% of cases. Additionally, the pan-kidney cancer cohort (KIPAN) exhibited a significant association, collectively establishing the core loss of heterozygosity (LOH) driver axis within the urinary system. In the context of gastrointestinal tumors, gastroesophageal cancer (STES) and gastric cancer (STAD) formed a continuous positive correlation belt. Meanwhile, colorectal cancer (COAD) activated the Wnt signaling pathway through hypermethylation of the APC gene promoter, thereby enhancing signal transduction.

The study demonstrated significant negative correlations in two tumor types. In lung adenocarcinoma (LUAD, $N = 495$), low expression of ROS1 was strongly negatively correlated with loss of heterozygosity (LOH) ($R = -0.273$, $p = 6.77e\text{-}10$), suggesting a potential genomic guardian role by maintaining the stability of the 15q26.3 region. In lung squamous cell carcinoma (LUSC, N=472), a moderate negative correlation was observed ($R= -0.137$, p=0.0028), with the intensity of the effect being significantly dose-dependent on tobacco exposure ($\beta = -0.21$, $p = 0.008$). This reflects a protective mechanism through the upregulated expression of the CDKN2A gene in the 9p21.3 region. Such organ-specific polarization offers an innovative pathway for precision therapy.

Conclusion

This study systematically elucidates the multifaceted roles of the ROS1 gene in tumorigenesis and cancer progression through a comprehensive pan-cancer analysis. At the expression level, ROS1 demonstrates notable tissue heterogeneity: Expression profiling across

33 cancer types reveals significant upregulation in 13 cancer types (including UCEC, BRCA, and CESC) and significant downregulation in 14 cancer types (such as GBM, LUAD, and KIRC), indicating highly cancer-type-specific expression patterns. Prognostic analysis further identifies elevated ROS1 expression as an independent adverse prognostic factor for seven cancers, including GBM and KIRC, with a hazard ratio reaching 4.26 in glioblastoma (GBM, p=4.4e-5), suggesting its potential role in promoting invasive phenotypes. Regarding immune regulation, the expression of ROS1 demonstrates a significant association with immune infiltration across eight cancer types. Positive correlations are identified in seven types, including lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), whereas a negative correlation is observed in lower-grade glioma (LGG). This suggests that ROS1 may influence immunotherapeutic responses by modulating the tumor microenvironment. Furthermore, ROS1 expression exhibits significant alterations with clinical stage progression in six cancers, such as LUAD and prostate adenocarcinoma (PRAD), underscoring its potential role in tumor progression.

At the genomic level, ROS1 demonstrates significant associations with various features, exhibiting a negative correlation with tumor mutational burden (TMB) in lung adenocarcinoma/squamous cell carcinoma and a positive correlation with mutant-allele tumor heterogeneity (MATH) in stomach and ovarian cancers (STAD/OV). These findings underscore its cancer-type-specific interactions with genomic instability. Moreover, the prevalent negative correlations with tumor purity, as exemplified by lung squamous cell carcinoma (LUSC: $R = -0.34$, $p < 1e\text{-}13$), alongside strong positive correlations with homologous recombination deficiency/loss of heterozygosity (HRD/LOH) in kidney cancers (KIRC/KIPAN), suggest its potential role in modulating DNA repair defects. Additionally, the negative correlation with microsatellite instability (MSI), such as stomach cancer (STAD: $R= -0.16$, $p<0.003$), implies its possible influence on immunogenicity.

In conclusion, this study establishes at the pan-cancer level that ROS1 functions not only as a fusion-driven oncogene but also that its basal expression level holds significant clinical implications. Specifically, ROS1 can serve as a prognostic biomarker, particularly in glioblastoma multiforme (GBM) and colorectal adenocarcinoma (COADREAD), act as a modulator of the immune microenvironment, and serve as an indicator of genomic instability. These findings suggest that ROS1 may be a potential target for enhancing the efficacy of immunotherapy. Future research should focus on the experimental validation of the oncogenic molecular mechanisms of ROS1 in non-fusion contexts and investigate its translational potential as a combinatorial therapeutic target.

Funding

This study was supported by the North Henan Medical University Special Fund for Scientific Research Project initiation (Grant No. 2024008ZK, NO.2024004ZK).

Henan Province College Students' Innovation Training Project (Grant No. S202513505014).

Henan Province Specialized Innovation Integration Characteristic Demonstration Course Project (Grant No. zcrh2024002).

Research Project of Sanquan College, Xinxiang Medical University (Grant No. 2024054ZK).

Henan Provincial Department of Education, Henan Provincial Key Discipline Teaching and Research [2023] (Grant No. 414-1058).

Henan Province Science and Technology Key Project (Grant No. LHGJ20210892).

Acknowledgements

The authors would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Iyer, S. R., Nusser, K., Jones, K., Shinde, P., Keddy, C., Beach, C. Z., Davare, M. A. (2023) Discovery of oncogenic ROS1 missense mutations with sensitivity to tyrosine kinase inhibitors. *EMBO Molecular Medicine*, 15(10), e17367.

[2] Glaser, M., von Levetzow, C., Michels, S., Nogova, L., Katzenmeier, M., Wömpner, C., Scheffler, M. (2022) 52P ROS1 aberrations in non-small cell lung cancer patients without rearrangements: clinical and molecular characteristics. *Annals of Oncology*, 33, S56-S57.

[3] Nagasaka, M., Zhang, S. S., Baca, Y., Xiu, J., Nieva,

J., Vanderwalde, A., Ou, S. H. I. (2023) Pan-tumor survey of ROS1 fusions detected by next-generation RNA and whole transcriptome sequencing. *BMC Cancer*, 23(1), 1000.

[4] Parker, B. A., Shatsky, R. A., Schwab, R. B., Wallace, A. M., I-SPY 2 Consortium, Wolf, D. M., Kipps, T. J. (2023) Association of baseline ROR1 and ROR2 gene expression with clinical outcomes in the I-SPY2 neoadjuvant breast cancer trial. *Breast Cancer Research and Treatment*, 199(2), 281-291.

[5] Yang, Y., Jing, W., Zhang, L., Zhang, Y., Shang, Y., Kuang, Y. (2025) WDR62 affects the progression of ovarian cancer by regulating the cell cycle. *Hereditas*, 162(1), 78.

[6] Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., Cope, L. (2018) An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, 173(2), 400-416.

[7] Kalbfleisch, J. D., Schaubel, D. E. (2023) Fifty years of the cox model. *Annual Review of Statistics and Its Application*, 10(1), 1-23.

[8] Chen, H., Du, Y., Kong, Z., Liao, X., Li, W. (2024) PRNP is a pan-cancer prognostic and immunity-related to EMT in colorectal cancer. *Frontiers in Cell and Developmental Biology*, 12, 1391873.

[9] Racle, J., De Jonge, K., Baumgaertner, P., Speiser, D. E., Gfeller, D. (2017) Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*, 6, e26476.

[10] Gu, L., Feng, C., Li, M., Hong, Z., Di, W., Qiu, L. (2023) Exosomal NOX1 promotes tumor-associated macrophage M2 polarization-mediated cancer progression by stimulating ROS production in cervical cancer: a preliminary study. *European Journal of Medical Research*, 28(1), 323.

[11] Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Yang, T. H. O., Chuah, E. (2018) The immune landscape of cancer. *Immunity*, 48(4), 812-830.