Association of LEMD1 Overexpression and Chromosome 1q Gain with Chromosomal Instability in Esophageal Squamous Cell Carcinoma: A TCGA Multi-omics Analysis

Xiaobo Wang[1, 2, 4], Xianfei Wang[1, 4], Sairah Abdul Karim[3, *], Chin Siang Kue[3, †]

[1]Department of Gastroenterology, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, China

[2]Post Graduate Centre, Management and Science University, Shah Alam, Selangor 40100, Malaysia

[3]Faculty of Health & Life Sciences, Management and Science University, Shah Alam, Selangor 40100, Malaysia

[4]Sichuan Branch of National Clinical Research Center for Digestive Diseases, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, China

*Corresponding email: sairah@msu.edu.my, †Corresponding email: cskue@msu.edu.my

Abstract

Background: Esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) are molecularly distinct histological subtypes. LAP2-Emerin-MAN1 (LEM) domain containing 1 (LEMD1), a cancer-testis antigen implicated in squamous malignancies, remains insufficiently characterized at the multi-omic level in ESCC. Objectives: To perform a reproducible multi-omics analysis of LEMD1 using The Cancer Genome Atlas (TCGA)-esophageal carcinoma (ESCA) data, integrating gene expression, copy-number alteration (CNA) status, chromosomal instability metrics, and ABSOLUTE-derived subclonal genome fraction, with a specific focus on ESCC. Methods: We integrated TCGA-ESCA ribonucleic acid (RNA)-seq expression, clinical annotations, and survival endpoints. Copy-number profiles were assessed using genomic identification of significant targets in cancer, version 2 (GISTIC2) gene-level calls, arm-level calls with aneuploidy score, Affymetrix genome-wide human snp array 6.0 (SNP6) segmentation, and ABSOLUTE-derived purity/ploidy metrics. Differential expressions, non-parametric tests, Spearman correlation, and Kaplan-Meier analyses were applied. Results: LEMD1 expression was significantly elevated in ESCC compared with EAC ($\log_2$FC = 0.805, false discovery rate (FDR) = 0.019). Gene-level CNA analysis revealed frequent LEMD1 gain/amplification in ESCC (51.0%), with lower deletion rates compared with EAC (p=0.023). LEMD1 CNA strongly correlated with 1q arm-level status (Spearman ρ=0.666, $p=7.6\times10^{-10}$), consistent with its chromosomal location at 1q32.1. Furthermore, LEMD1 CNA correlated with fraction genome gained (ρ=0.279, p=0.00596) and ABSOLUTE subclonal genome fraction (ρ=0.312, p=0.004104), indicating associations with chromosomal instability and intra-tumor heterogeneity. Conclusions: LEMD1 is a lineage-associated transcript elevated in ESCC and frequently gained as part of chromosome 1q alterations. LEMD1 CNA status tracks genome-wide gain burden and subclonal copy number variation (CNV) fraction, supporting its utility as a multi-omic indicator linked to chromosomal instability in ESCC.

Keywords

ABSOLUTE, Chromosomal instability, Copy-number alteration, Esophageal squamous cell carcinoma, LAP2-Emerin-MAN1 domain containing 1, 1q gain, The Cancer Genome Atlas

Introduction

Esophageal carcinoma is a highly lethal malignancy comprising two major histologic subtypes: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). These subtypes differ markedly in geographical distribution, risk factors, and molecular landscapes, and are increasingly regarded as distinct clinical entities. The TCGA Integrated Genomic Characterization of Oesophageal Carcinoma highlighted subtype-specific genomic programs and copy-number patterns that underscore this biological divergence [1,2]. LEMD1, located on chromosome 1q32.1, has been characterized as a cancer-testis antigen with tumor-promoting properties [3]. In oral squamous cell carcinoma, LEMD1 expression correlates with tumor

progression, nodal metastasis, and adverse prognosis, with functional evidence supporting roles in invasion and endothelial transmigration [4]. Recent studies have further implicated LEMD1 as an oncogenic factor and a potential therapeutic target across diverse multiple cancer types [5-7].

Given the subtype-specific biology of ESCC and the established relevance of LEMD1 in squamous malignancies, we performed a reproducible multi-omics analysis of LEMD1 using data from the TCGA-ESCA cohort, with a specific focus on ESCC. This analysis integrated LEMD1 gene expression, copy-number alteration (CNA) status, chromosomal instability metrics derived from segmentation data, and ABSOLUTE-derived subclonal genome fraction, and evaluated their association with clinical characteristics.

Materials and methods

(1) Data acquisition: We analyzed TCGA-derived datasets including TCGA-ESCA RNA-seq expression (IlluminaHiSeq_RNASeqV2, $\log_2(x+1)$ RSEM), clinical annotations, and gene-level GISTIC2 thresholded copy-number calls. We also analyzed genome-wide SNP6 segmentation, arm-level CNV calls with aneuploidy score, ABSOLUTE mastercalls (purity, ploidy, and subclonal genome fraction), and standardized survival endpoints from TCGA-CDR [8,9].

(2) Cohort definition: ESCC and EAC were defined using histological_type annotations. Primary tumor samples (type 01) were prioritized. Patient-level TCGA barcodes were matched to TCGA-CDR endpoints for survival analyses.

(3) Expression analysis: LEMD1 RNA expression was extracted from the TCGA IlluminaHiSeq_RNASeqV2 matrix. Differential expressions between ESCC and EAC were performed using Welch's t-test with Benjamini-Hochberg FDR correction. LEMD1 tumor-versus-normal comparisons were exploration due to limited normal ESCC samples.

(4) Copy-number analysis: Gene-level LEMD1 CNA status was extracted from GISTIC2 thresholded calls (−2 deep deletion, −1 deletion, 0 diploid, +1 gain, +2 amplification). Histology-specific CNV distributions were compared using χ^2 tests. Arm-level 1q calls were obtained from the pan-cancer arm-level table and correlated with gene-level LEMD1 calls using Spearman correlation.

(5) Chromosomal instability metrics: Using SNP6 segmentation data, we computed fraction genome gained (FGG) and lost (FGL) as the proportion of the genome with segment mean >0.2 or <-0.2. Fraction genome altered (FGA) was defined as FGG + FGL. Genome length was approximated using hg19 chromosome lengths.

(6) ABSOLUTE-derived heterogeneity: Subclonal genome fraction was obtained from ABSOLUTE mastercalls and correlated with LEMD1 CNA using Spearman tests.

(7) Survival analysis: Overall survival (OS), disease-specific survival (DSS), progression-free interval (PFI), and disease-free interval (DFI) were obtained from TCGA-CDR. Kaplan-Meier and log-rank analyses were performed using a median split of LEMD1 expression within ESCC. Cox proportional hazards models were used for hazard ratio estimation. Two-sided $p<0.05$ was considered statistically significant.

Results

Cohort overview

The TCGA-ESCA clinical matrix comprised 204 samples, including 100 ESCC and 104 EAC cases, 185 were primary tumors and 19 were normal tissues. RNA-seq expression data were available for 196 samples. For ESCC-focused integrative analyses, we retained 96 ESCC primary tumors with complete clinical and survival information (Table 1).

Table 1. Clinical characteristics of the ESCC cohort (TCGA-CDR, n=96).

Characteristic	Value
Age (years), median (interquartile range)	57.0 (51.0-64.2)
Gender: male/female	81 (84.4%)/15 (15.6%)
AJCC stage: I/II/III/IV/NA	7/56/27/4/2
OS/DSS/PFI/DFI events	32/20/42/19

LEMD1 is overexpressed in ESCC and distinguishes histologic subtypes

LEMD1 expression differed significantly across ESCA histologies and sample types (Figure 1). Across tumors, LEMD1 expression was markedly higher in ESCC than in EAC (median 4.20 vs. 3.08, $P=0.009$), suggesting histology-specific regulation of this gene. Genome-wide

differential expression analysis confirmed that LEMD1 was significantly overexpressed in ESCC relative to EAC ($log_2FC = 0.805$, FDR = 0.019, Figure 2), further supporting its potential role as a histology-specific biomarker. An exploratory ESCC tumor-versus-normal comparison suggested elevated LEMD1 expression in tumors ($p = 0.038$). However, this finding should be interpreted with caution owing to the limited number of ESCC normal samples ($n = 3$). Collectively, these expression patterns point to a potentially distinct functional role for LEMD1 in ESCC compared with EAC.

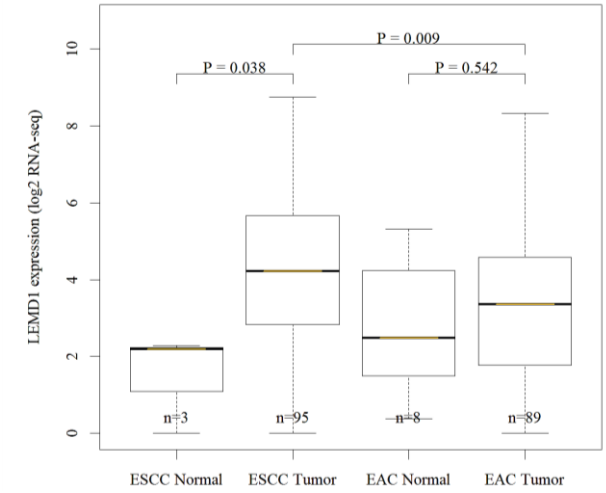

Figure 1. LEMD1 expression across ESCA histology's and sample types (log_2 RNA-seq).

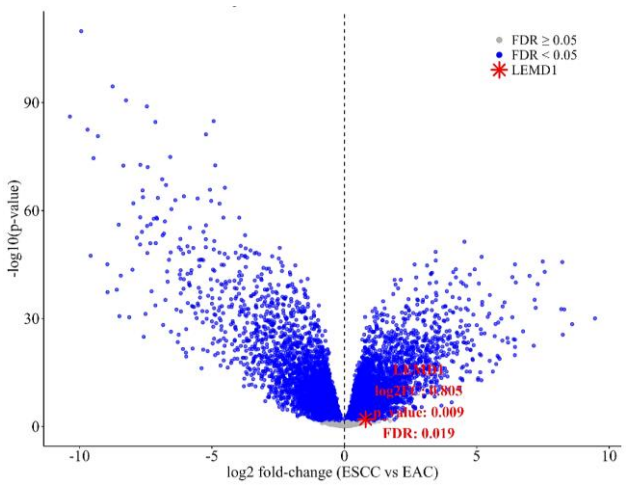

Figure 2. Volcano plot of differential expression between ESCC and EAC tumors.

LEMD1 expression correlates with squamous lineage markers in ESCC

To further characterize the lineage association of LEMD1, we examined correlations with established differentiation markers within ESCC tumors. LEMD1 expression correlated positively with canonical squamous markers, such as KRT14 and MKI67, and negatively with glandular or epithelial markers, including KRT4 and KRT8, strongly supporting that LEMD1 expression is linked to squamous differentiation programs (Figure 3).

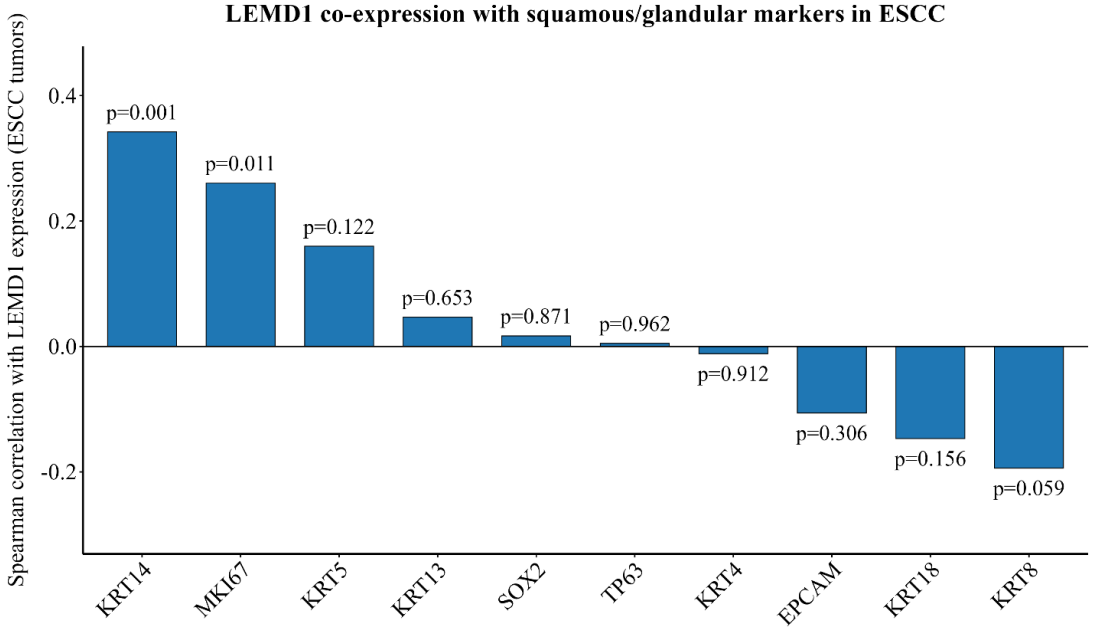

Figure 3. Spearman correlations between LEMD1 and selected lineage markers within ESCC tumors.

LEMD1 is frequently gained in ESCC and aligns with 1q arm-level status

Using GISTIC2 thresholded CNV calls, LEMD1 gain or amplification occurred in 49 (51.0%) ESCC tumors, whereas deletions were rare (6, 6.2%). In contrast, among EAC cases, LEMD1 gain or amplification was observed in 35 (39.8%), while deletions occurred more frequently (17, 19.3%). The distribution of LEMD1 CNA categories differed significantly between the two histological subtypes (χ^2 test, $p<0.001$, Figure 4A). Consistent with

the chromosomal location of LEMD1 on 1q32.1, gene-level LEMD1 CNV status showed a strong positive correlation with arm-level chromosome 1q copy-number status in ESCC (Spearman $\rho = 0.666$, $p = 7.6\times10^{-10}$, Figure 4B). This finding suggests that LEMD1 copy-number alterations in ESCC are predominantly driven by large-scale chromosomal events affecting the entire 1q arm, further supporting its role as a potential oncogenic driver in this subtype.

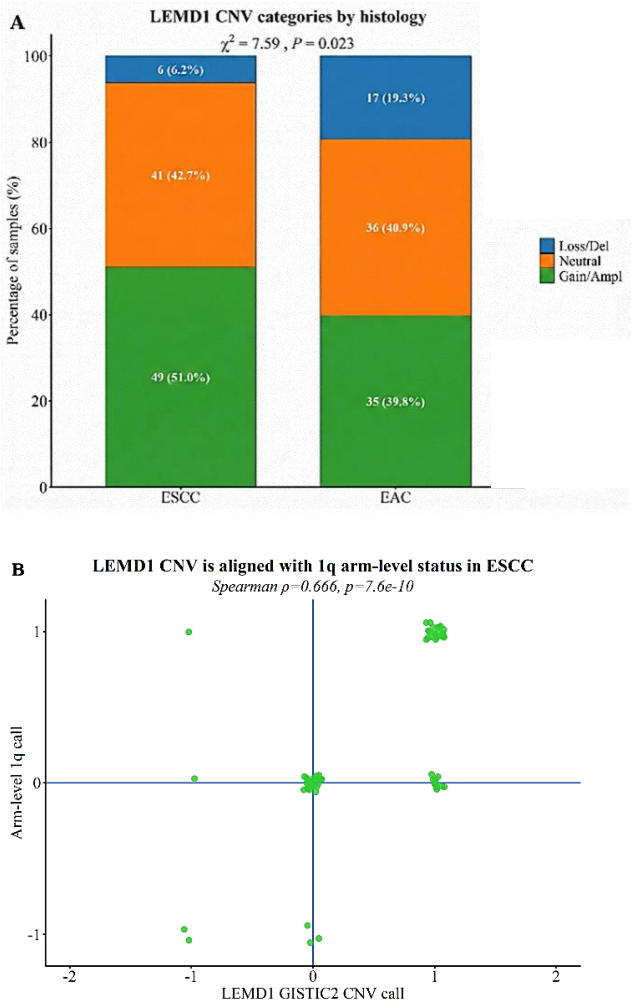

Figure 4. (A) Distribution of LEMD1 GISTIC2 CNV categories in ESCC versus EAC tumors. (B) Relationship between gene-level LEMD1 CNV and arm-level 1q call in ESCC.

LEMD1 CNV tracks genome-wide gain burden and subclonal CNV fraction

Based on SNP6 segmentation analysis, ESCC exhibited a significantly higher genome-wide gain burden than EAC, as measured by the fraction of genome gained (FGG, p=0.013, Figure 5). This histological difference suggests that ESCC tumors may rely more heavily on copy-number gain-driven oncogenic mechanisms compared to EAC. Within the ESCC cohort, LEMD1 copy-number alteration status was positively correlated with FGG (Spearman ρ=0.279, p=0.00596, Figure 6A), indicating that tumors with LEMD1 gain or amplification tend to harbor a greater overall burden of genomic gains. Furthermore, among ESCC cases with available ABSOLUTE estimates (n=83), LEMD1 copy-number status showed a significant positive correlation with subclonal genome fraction (Spearman $\rho = 0.312$, $p = 0.004104$, Figure 6B), a metric reflecting the proportion of tumor cells harboring subclonal alterations. Collectively, these findings suggest that LEMD1 copy-number states are associated with elevated subclonal copy-number burden and may reflect enhanced intra-tumor heterogeneity in ESCC (Table 2). Notably, the correlations with both FGG and subclonal genome fraction were independent of the strong association between LEMD1 CNV and arm-level 1q status, supporting the potential functional relevance of LEMD1 copy-number gains in ESCC progression.

Table 2. Correlations between LEMD1 CNV and chromosomal instability/heterogeneity metrics in ESCC.

Association	Spearman ρ	p-value	n
LEMD1 CNV vs. 1q arm-level call	0.666	7.56×10^{-10}	67
LEMD1 CNV vs. FGG	0.268	8.55×10^{-3}	95
LEMD1 CNV vs. subclonal genome fraction	0.312	4.10×10^{-3}	83

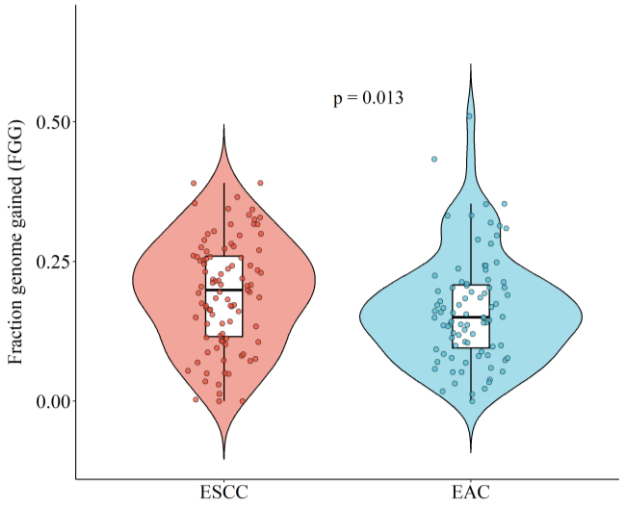

Figure 5. Fraction genome gained (FGG) differs by histology (ESCC vs. EAC).

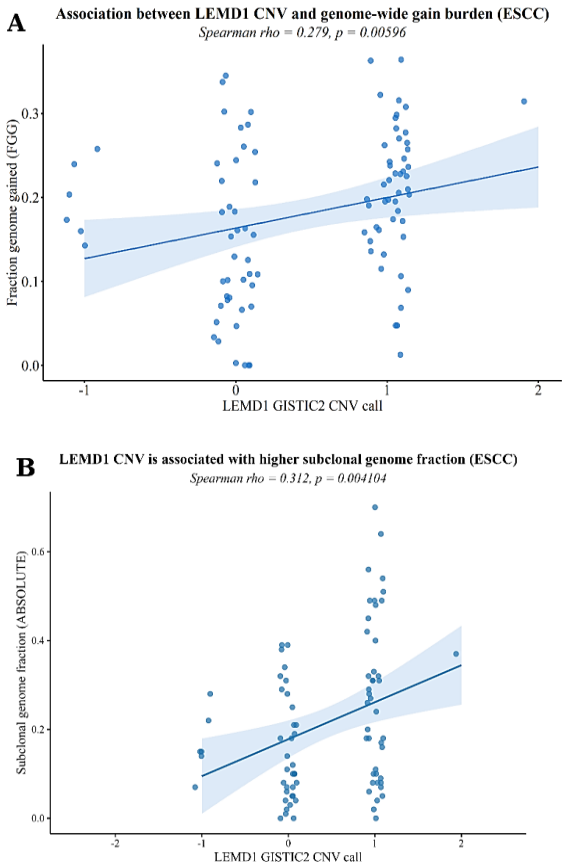

Figure 6. (A) LEMD1 CNV is associated with fraction genome gained (FGG) in ESCC. (B) LEMD1 CNV is associated with ABSOLUTE subclonal genome fraction in ESCC.

Clinical outcome associations

Using TCGA-CDR endpoints within the ESCC cohort, no statistically significant associations were observed between LEMD1 expression and overall survival (OS), progression-free interval (PFI), or disease-free interval (DFI). A non-significant trend toward poorer disease-specific survival (DSS) was noted (Figure 7). These analyses were exploratory in nature and may be limited by sample size and clinical heterogeneity.

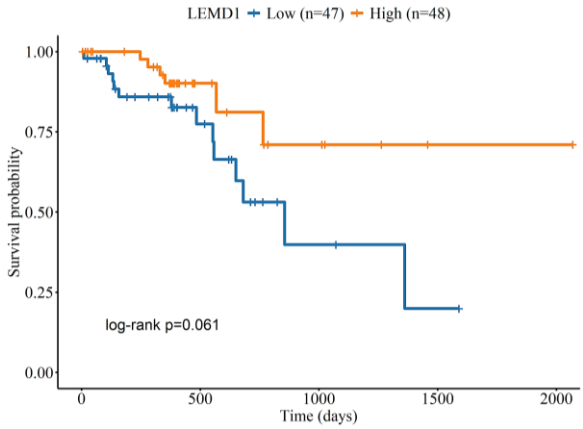

Figure 7. Kaplan-Meier curve for DSS stratified by LEMD1 expression.

Discussion

This TCGA-based multi-omics analysis provides a comprehensive characterization of LEMD1 in esophageal squamous cell carcinoma. We demonstrate that LEMD1 is transcriptionally upregulated in ESCC relative to EAC, consistent with prior reports implicating LEMD1 as a tumor-progressive factor in squamous malignancies. The observed positive correlation between LEMD1 expression and canonical squamous lineage markers further supports its association with squamous differentiation.

We further show that LEMD1 is frequently gained or amplified in ESCC and that its gene-level copy-number alteration (CNA) status closely reflects chromosome 1q arm-level copy-number state. However, gene-level copy-number did not fully account for inter-sample variability in LEMD1 mRNA expression within ESCC, suggesting that additional regulatory mechanisms may contribute to its transcriptional dysregulation. These may include epigenetic activation characteristics of cancer-testis antigens, transcriptional control, or post-transcriptional regulation mechanisms [10].

By integrating SNP6 segmentation data and ABSOLUTE-derived estimates, we linked LEMD1 CNA status to chromosomal instability-related features. The positive association between LEMD1 copy-number state and genome-wide gain burden (FGG) suggests that LEMD1 gain likely arises within the context of broad copy-number gains rather than as an isolated focal event. Moreover, the correlation between LEMD1 CNA status and subclonal genome fraction suggests an association with elevated subclonal copy-number burden, a surrogate measure of intra-tumor heterogeneity.

Several limitations should be acknowledged. The modest ESCC sample size and limited availability of matched normal esophageal tissues restrict statistical power and preclude definitive tumor-versus-normal comparisons. Additionally, the lack of orthogonal validation using independent cohorts, DNA methylation profiling, or protein-level analyses limits the generalizability of the findings. Future studies should validate LEMD1 expression and CNA patterns in independent ESCC cohorts (e.g., Gene Expression Omnibus/ArrayExpress datasets), investigate regulatory mechanisms such as DNA methylation, and assess the potential of LEMD1 as

a clinically actionable biomarker or immunotherapy target in ESCC.

Conclusion

LEMD1 is transcriptionally elevated in ESCC relative to EAC and is frequently gained as part of chromosome 1q copy-number alterations. LEMD1 gene-level CNA status correlates with genome-wide gain burden and ABSOLUTE subclonal genome fraction, linking LEMD1 CNA to chromosomal instability and increased subclonal copy-number burden in ESCC. These findings provide a reproducible genomic framework for further investigation and support the need for functional studies and external-cohort validation.

Funding

This work was not supported by any funds.

Author Contributions

Xiaobo Wang and Xianfei Wang contributed to the study conception and design, data acquisition, and analysis. Sairah Abdul Karim and Chin Siang Kue supervised the study, provided critical intellectual input, and revised the manuscript. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Data Availability

All analyses were performed using publicly available TCGA datasets distributed via UCSC Xena and the TCGA Pan-Cancer Atlas resources. The exact input files analyzed are listed in the materials and methods section. Analysis scripts and processed outputs will be provided as supplementary materials upon publication.

Acknowledgments

The authors acknowledge the TCGA Research Network for generating and providing the datasets used in this study.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Cui, Y., Chen, H., Xi, R., Cui, H., Zhao, Y., Xu, E., Liu, Z. (2020) Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Research*, 30(10), 902-913.

[2] Yang, H., Wang, F., Hallemeier, C. L., Lerut, T., Fu, J. (2024) Oesophageal cancer. *The Lancet*, 404(10466), 1991-2005.

[3] Xu, M., Lin, B., Zheng, D., Wen, J., Hu, W., Li, C., Qu, J. (2021) LEM domain containing 1 promotes thyroid cancer cell proliferation and migration by activating the Wnt/β-catenin signaling pathway and epithelial-mesenchymal transition. *Oncology Letters*, 21(6), 442.

[4] Sasahira, T., Kurihara, M., Nakashima, C., Kirita, T., Kuniyasu, H. (2016) LEM domain containing 1 promotes oral squamous cell carcinoma invasion and endothelial transmigration. *British Journal of Cancer*, 115(1), 52-58.

[5] Li, X., Jiang, S., Jiang, T., Sun, X., Guan, Y., Fan, S., Cheng, Y. (2023) LEM domain containing 1 acts as a novel oncogene and therapeutic target for triple-negative breast cancer. *Cancers*, 15(11), 2924.

[6] Li, L., Zhang, P. (2024) Elevation of LEM domain containing 1 predicts poor prognosis of NSCLC Patients and triggers malignant stemness and invasion of NSCLC cells by stimulating PI3K/AKT pathway. *Current Molecular Medicine*, 24(3), 366-378.

[7] Li, Z., Wang, J., Wu, J., Li, N., Jiang, C. (2022) Long noncoding RNA LEMD1-AS1 increases LEMD1 expression and activates PI3K-AKT pathway to promote metastasis in oral squamous cell carcinoma. *BioMed Research International*, (1), 3543948.

[8] Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Haussler, D. (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38(6), 675-678.

[9] Liao, C., Wang, X. (2023) TCGAplot: an R package for integrative pan-cancer analysis and visualization of TCGA multi-omics data. *BMC Bioinformatics*, 24(1), 483.

[10] Zhao, Z., Chen, Y., Cheng, X., Huang, L., Wen, H., Xu, Q., Ni, T. (2023) The landscape of cryptic antisense transcription in human cancers reveals an oncogenic noncoding RNA in lung cancer. *Science Advances*, 9(14), eadf3264.