Original Research Article

C1orf87, a predictive biomarker for lung adenocarcinoma

Theodore Hwang

Davis Senior High School, 315 W 14th St, Davis, CA 95616, United States

ABSTRACT

Uncharacterized genes represent a substantial fraction of the human genome and remain largely unexplored despite their potential importance in human physiology. C1orf87, also known as carcinoma-related EF-hand protein, has been annotated as a protein-coding gene, yet its molecular function and disease relevance have not been established. The analysis of transcriptomic data from The Cancer Genome Atlas (TCGA) comprising 576 lung cancer patients, encompassing more than 20,000 genes per case, revealed its association with patient survival and distinct pathway alterations in gene set enrichment analysis (GSEA). In addition, C1orf87 was found to be significantly upregulated in lung adenocarcinoma relative to adjacent normal tissue. However, elevated C1orf87 expression correlated with favorable overall survival ($p = 0.097$), suggesting a prognostic role. These findings support C1orf87 as a potential predictive biomarker in lung adenocarcinoma and highlight the value of investigating previously uncharacterized genes in uncovering novel cancer biology.

Keywords: C1orf87; Lung cancer; Tumor suppressor gene; Inflammation; Gene Set Enrichment Analysis; Airway epithelial cells; Prognostic marker

INTRODUCTION

Uncharacterized genes represent genomic regions predicted to encode a protein based on DNA sequence, but lacking definitive evidence of biological function. These genes remain poorly studied, yet their investigation can reveal novel mechanisms underlying human physiology and disease, including cancer. One example is C1orf87, an uncharacterized gene that appears to be expressed and translated also known as 'carcinoma-related EF-hand protein' (1). Despite its annotation, the molecular role of C1orf87 in cellular processes and disease states has not been established.

Lung cancer is the leading cause of cancer-related death in United States, accounting for 234,580 new lung cancer cases and 125,070 deaths (2). This represents nearly 20% of all cancer deaths in the United States. Lung cancers are broadly classified into non-small cell lung cancer (NSCLC), accounting for 80% of cases (3). Among NSCLC, lung adenocarcinoma is the most common histological subtype of lung cancer, followed by squamous cell carcinoma and then large cell carcinoma.

Lung adenocarcinoma arises due to complex factors such as genetics, environmental and inflammatory processes. While cigarette smoking is one of the common risk factors for lung cancers, a significant proportion of cases also occur in never smokers. Beyond genetic mutations, chronic inflammation induced by environmental pollutants or other disease conditions contributes to lung cancer development. Therefore, it is important to understand other contributing factors for lung cancer development. This study suggests that C1orf87 can be a potential tumor suppressor

Corresponding author: Theodore Hwang, E-mail: timtedhwang2@gmail.com.

Accepted November 7, 2025
https://doi.org/10.70251/HYJR2348.36425431

involved in inflammatory responses. In addition, down-regulation of C1ORF87 can be a biomarker to predict patient's outcomes for LUAD. In sum, this study aims to determine whether C1orf87 expression is associated with inflammatory signaling and clinical outcomes in lung adenocarcinoma.

METHODS AND MATERIALS

Gene expression analysis of C1orf87 across tissues

RNA and protein expression data for C1orf87 were obtained from the Human Protein Atlas and BioGPS databases. Expression patterns across tissues and cell types were analyzed to identify organ- and cell-specific expression profiles. Transcription factor binding site predictions for the upstream promoter region of C1orf87 were accessed through the Protein Atlas annotation resources.

Inflammatory response dataset analysis

To examine regulation of C1orf87 by inflammatory stimuli, the publicly available GDS4981 microarray dataset was downloaded from the NCBI Gene Expression Omnibus (GEO). This dataset profiles gene expression in human airway epithelial cells treated with IL-13 in vitro (n = 4 per group). Expression levels of C1orf87 between IL-13–treated and control groups were compared to assess cytokine-mediated regulation.

Expression analysis in lung cancer

C1orf87 expression in tumor versus normal tissue was analyzed using RNA-seq data from The Cancer Genome Atlas (TCGA) for lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC). Normalized gene expression values were downloaded from the cbioportal.

Patient survival analysis

TCGA LUAD and LUSC patient cohorts were stratified into C1orf87-high and C1orf87-low expression groups based on median expression. Kaplan–Meier survival curves were generated in the Graphpad Prism v10.0, and log-rank tests were used to evaluate statistical significance of survival differences between groups.

Gene Set Enrichment Analysis (GSEA)

RNA-seq data from TCGA LUAD samples (n = 497) were analyzed using Gene Set Enrichment Analysis (GSEA). Patients were divided into C1orf87-high and C1orf87-low groups, consistent with survival analysis. Enrichment was tested against the Molecular Signatures Database (MSigDB Hallmark gene sets). Pathways with nominal p-value < 0.05 and false discovery rate (FDR) q-value < 0.25 were considered significant.

RESULTS

C1orf87 expression is highly restricted in respiratory system, specifically in lung airway and bronchial epithelial cells

The analysis of the C1orf87 structure (available in the Protein atlas) revealed that it contains an EF hand domain, suggesting it is a calcium-binding protein with a potential transporter function (4). Two single nucleotide polymorphisms (SNPs) have been shown to be associated with migraine disorder and alcoholic chronic pancreatitis, although the mechanisms remain elusive. The sequence analysis of the upstream promoter region of C1orf87 gene revealed a number of potential transcription factor binding sites such as AML1a, ATF-2, Brachyury, C/EBPbeta, GR-alpha, HFH-1, LCR-F1, NF-1, Pax-4a (ref: protein atlas). Notably, C/EBPbeta and Glucocorticoid receptor alpha are known to be involved in various inflammatory responses in the body.

RNA expression of C1orf87 across human body suggests its expression is restricted to brain and respiratory system as well as reproductive organs (Figure 1). Interestingly, protein expression of C1orf87 is highly expressed in the respiratory system compared to RNA expression level except for male/female reproductive organs. BioGPS analysis further confirmed the high expression of C1orf87 in the respiratory system, specifically in airway and bronchial epithelial cells (Figure 2).

C1orf87 can be regulated by inflammatory signals and implicated in lung cancer

Since lung airway epithelial cells express high level of C1orf87 expression, publicly available gene expression datasets were examined to explore potential regulatory contexts. Analysis of the GDS4981 microarray dataset, in which human airway epithelial cells were treated with IL-13 *in vitro* (5), revealed a decrease in C1orf87 expression. These findings suggest that inflammatory responses downregulate C1orf87 and that its function may play an important role in modulating airway inflammation (Figure 3).

Since inflammation is a well-known risk factor across many different cancer types[3], it is highly possible that C1orf87 is implicated in cancer, particularly in

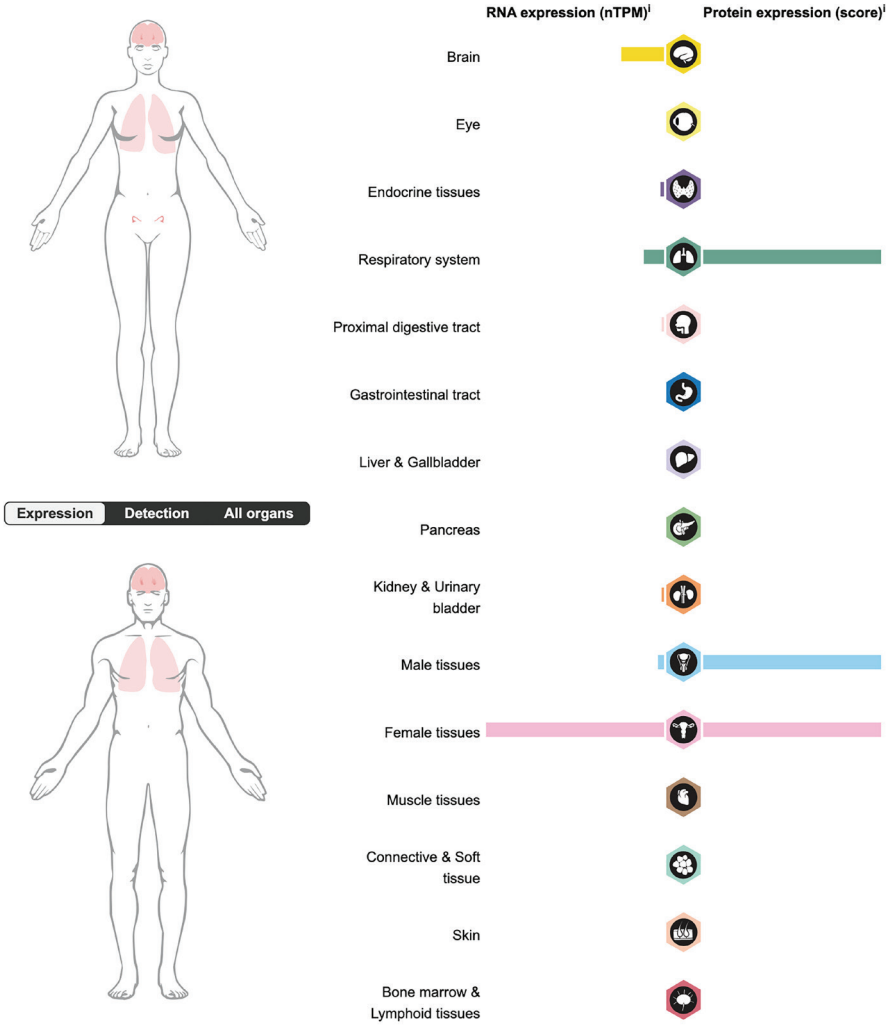

Figure 1. C1orf87 RNA and Protein expression summary across different organs. Data was obtained from the Human Protein Atlas.

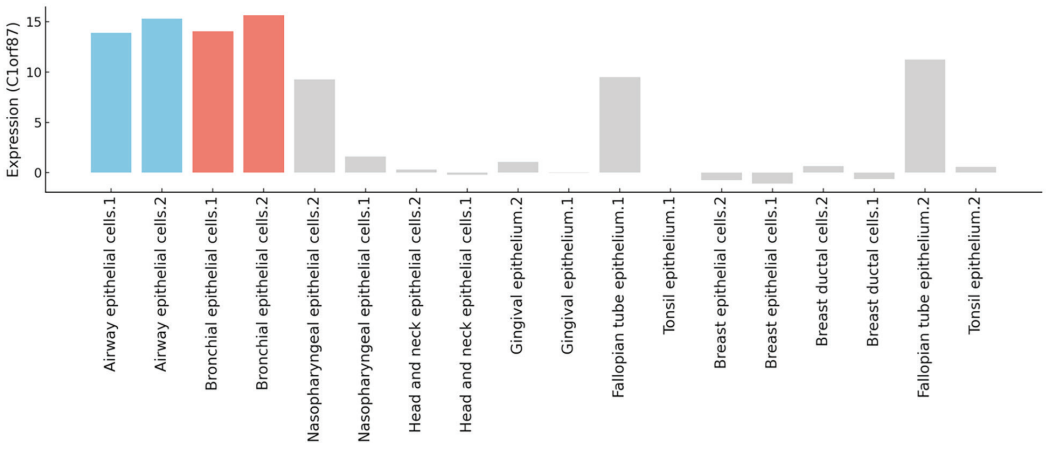

Figure 2. RNA expression of C1orf87 across various cell types. Data was obtained from the Human Protein Atlas. Note the highest expression of C1orf87 in Airway epithelial (blue) and bronchial epithelial (red) cells.

lung cancer given the expression pattern of C1Orf87. Comparing expression between normal and tumor tissues, both LUAD (lung adenocarcinoma) and LUSC (lung squamous carcinoma) showed that tumor tissues have decreased expression of C1orf87 (Figure 4), suggesting its potential role as a tumor suppressor. Patient survival analysis revealed that C1orf87 expression showed a trend for its association with favorable prognosis in LUAD, but no association in LUSC (Figure 5 and Figure 6).

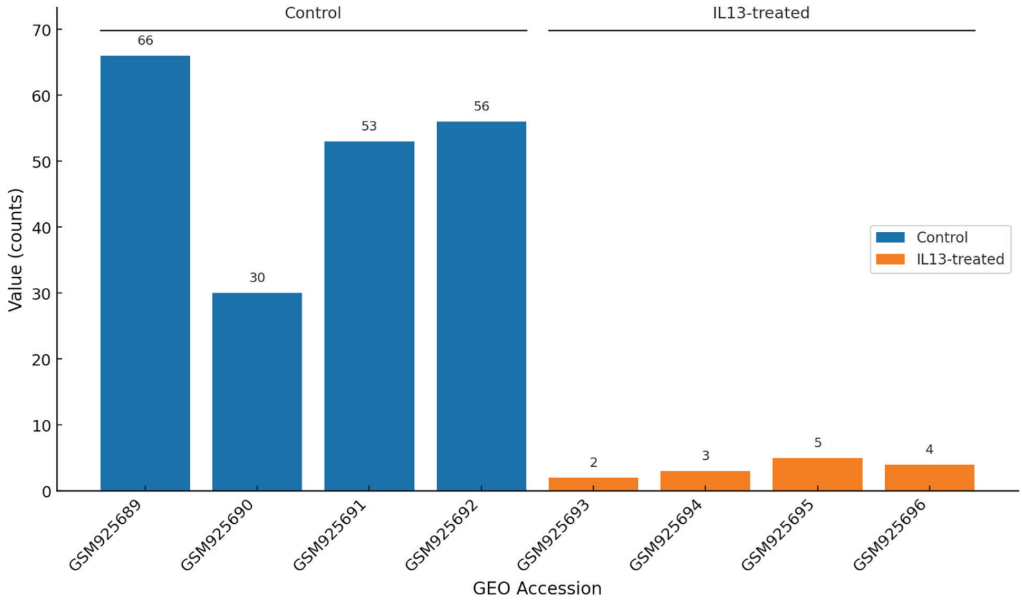

Figure 3. Analysis of the publicly available microarray dataset (GDS4981). IL13 treated human epithelial cells were profiled for the gene expression. RNA expression levels for C1orf87 for the control (n = 4) and IL13-treated (n = 4) groups were plotted. Note a significant down-regulation of C1orf87 in the IL13-treated group compared to the control.

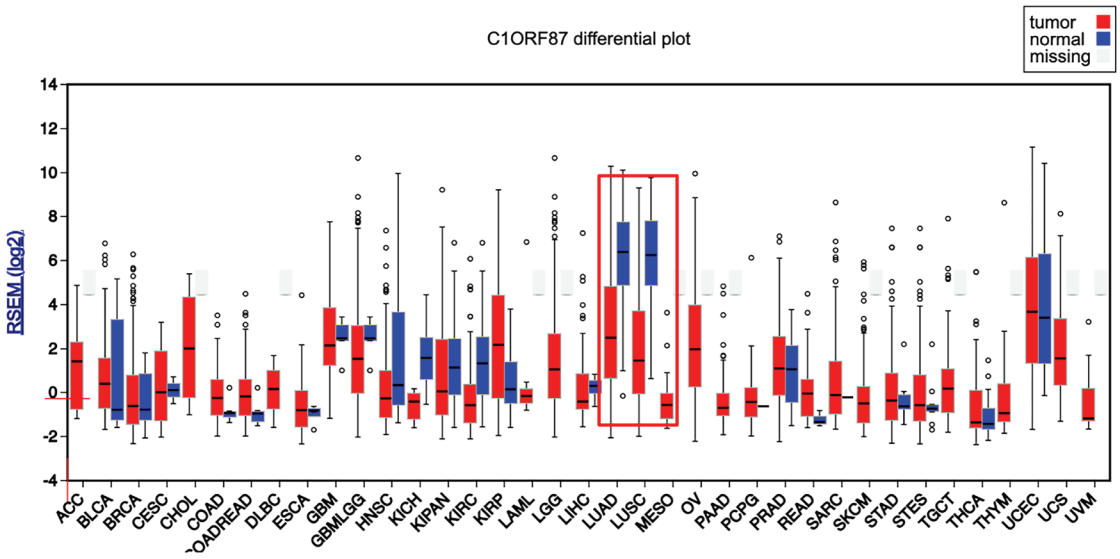

Figure 4. C1ORF87 RNA expression comparison between normal and tumor across cancer types. Note that C1ORF87 expression is down-regulated in both LUAD (lung adenocarcinoma) and LUSC (lung squamous carcinoma).

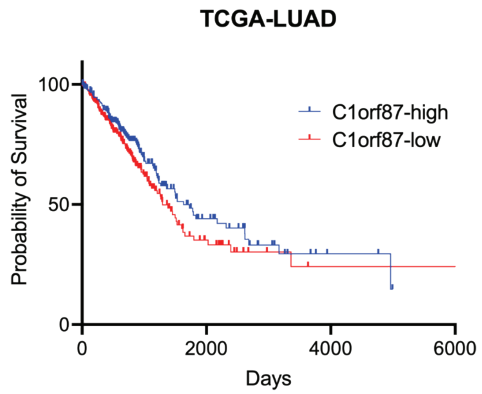

Figure 5. Patient survival analysis of C1orf87 in lung adenocarcinoma (LUAD, n=576). The patient group with high expression of C1orf87 (n = 288) has a favorable prognosis (p-val = 0.097, log-rank test).

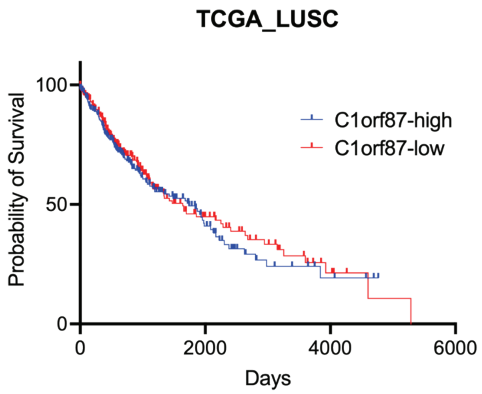

Figure 6. Patient survival analysis of C1orf87 in lung squamous carcinoma (LUSC, n=489). There is no statistically significant difference between two groups(C1orf87-high, n=245 and C1orf87-low, n=244, p-val = 0.56, long-rank test).

Gene Set Enrichment Analysis (GSEA)

Given my observation that C1orf87 was downregulated in lung cancer and it is associated with lung cancer patient survival, it is likely that this gene is involved in key signaling pathways that shape distinct transcriptional programs. To address this, RNA-seq data from TCGA LUAD patient samples (n = 576) were further analyzed for the gene set enrichment analysis (GSEA) by stratifying patients into C1orf87-high and C1orf87-low groups, consistent with the patient survival analysis described above. This analysis identified a number of genes that are differentially expressed between two groups shown in the heatmap (Figure 7). This analysis revealed that low C1orf87 expression was associated with enrichment of MYC, E2F, and G2M checkpoint pathways (Figure 8). MYC is known to induce E2F transcription factor family which plays an important role in cell cycle progression, implicating enhanced cell-cycle progression and proliferative activity in tumors with low C1orf87 expression (6). In contrast, high C1orf87 expression correlated with TNF-alpha signaling via NF-kB, inflammatory response, and IL6-JAK-STAT3 signaling pathways (Figure 9). NF-κB is a key transcription factor that drives the expression of IL-6. The IL-6/JAK/STAT3 signaling pathway is shown to be aberrantly hyperactivated in patients with chronic inflammatory conditions and in those with hematopoietic malignancies or solid tumors (7). Together, these findings suggest that C1orf87 levels may differentially modulate cancer cell proliferation and inflammatory signaling programs in lung adenocarcinoma.

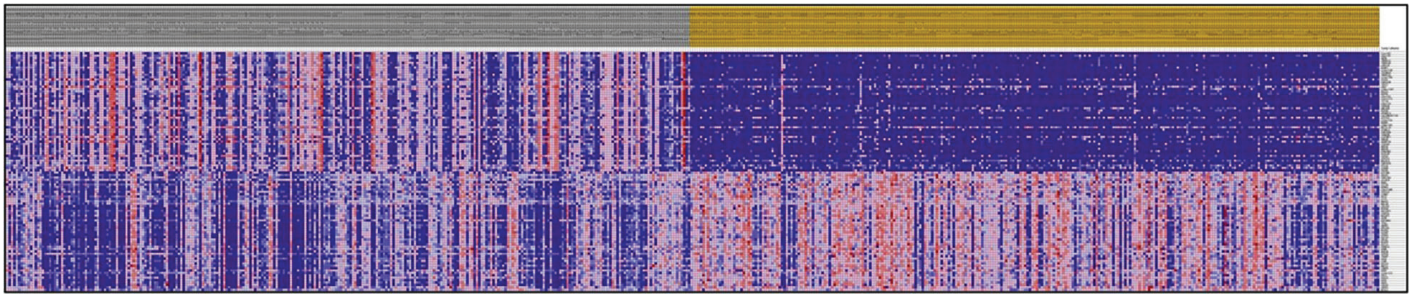

Figure 7. A heatmap of differentially expressed genes between C1orf87-high and C1orf87-low. Patients were divided into C1orf87-high (left) and C1orf87-low (right) groups. Each column represents an individual patient, and each row represents a gene. Gene expression levels were displayed as a color scale, with red indicating upregulation and blue indicating downregulation.

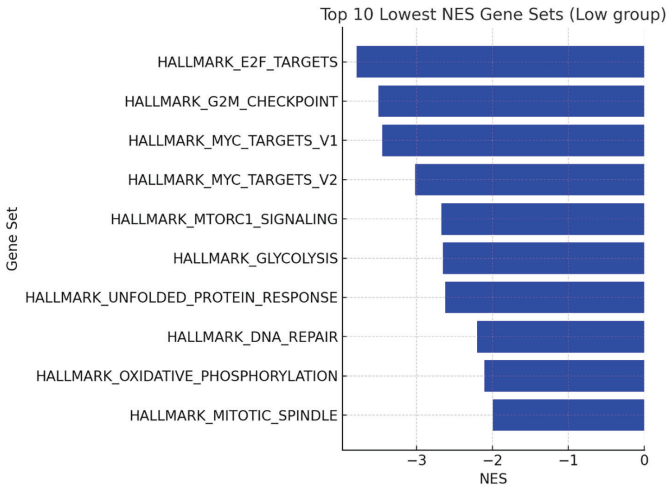

Figure 8. Gene set enrichment analysis of TCGA LUAD samples stratified by C1orf87 expression. Bar graph showing normalized enrichment score (NES) for selected pathways. Negative NES values indicate relative enrichment in the C1orf87-low group. All pathways were statistically significant (p-value < 0.05).

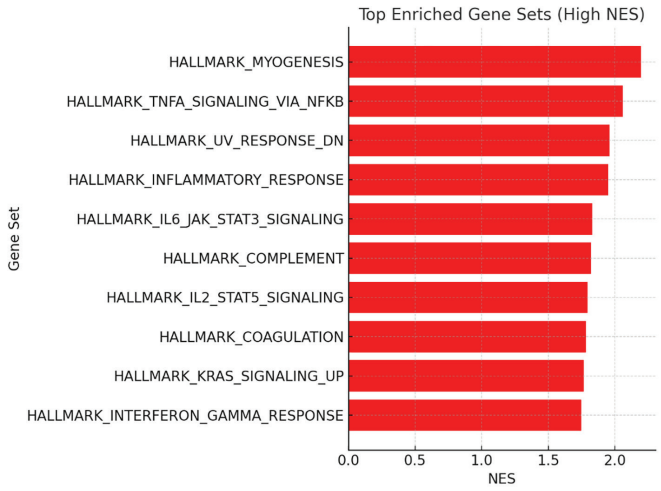

Figure 9. Gene set enrichment analysis of TCGA LUAD samples stratified by C1orf87 expression. Bar graph showing normalized enrichment score (NES) for selected pathways. Positive NES values indicate relative enrichment in the C1orf87-high group. All pathways were statistically significant (p-value < 0.05).

DISCUSSION AND CONCLUSION

Chronic inflammation is a well-established driver of tumor initiation and progression. Prolonged exposure to inflammatory cytokines, such as IL-13, can reshape the airway microenvironment and promote carcinogenesis. In this study, inflammatory signaling downregulates C1orf87 in airway epithelial cells, suggesting a link between inflammation and loss of this gene's activity in lung adenocarcinoma. Given that low C1orf87 expression is associated with enhanced proliferative pathways and poor prognosis, its suppression by inflammatory cues may represent a critical step in lung cancer development.

This work also highlights how computational analysis of uncharacterized genes can provide new insights into cancer biology. While up-regulation of C1orf87 in lung cancer can suggest that C1orf87 may be a potential oncogene in lung cancer, it does not necessarily establish causative roles in lung cancer. By computationally analyzing publicly available patient datasets, C1orf87 is shown to be a candidate tumor suppressor and predictive biomarker for favorable prognosis. Such approaches can uncover overlooked genes with significant biological and clinical relevance. Moving forward, experimental studies are needed to validate these computational findings. For example, artificially overexpressing C1orf87 in lung adenocarcinoma cells and testing its effects on tumor growth could directly reveal whether it exerts tumor suppressor activity, offering a foundation for translational applications.

In conclusion, C1orf87 is a previously uncharacterized gene with restricted expression in airway epithelial cells that is downregulated in lung cancer. Its reduced expression correlates with poor patient survival and activation of proliferative signaling programs, whereas higher expression is associated with inflammatory pathways. These findings suggest C1orf87 functions at the intersection of inflammation and cancer progression, with potential utility as a prognostic biomarker in lung adenocarcinoma.

ACKNOWLEDGEMENT

The author would like to thank Dr. Micaela Vargas in the Rosetta Institute of Biomedical Research who provided an opportunity to investigate the functions of C1orf87 using bioinformatic tools. The interest in this gene began during the two-week Bioinformatic workshop held in University of California San Diego, which motivated to pursue this project further. The author subsequently carried out an independent bioinformatic analysis of lung cancer patient datasets, leading to the findings presented here.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Delgado AP, Brandao P, Chapado MJ, Hamid S & Narayanan R. Open reading frames associated with cancer in the dark matter of the human genome. *Cancer Genomics Proteomics.* 2014; 11: 201-213 .
2. Siegel RL, Kratzer TB, Giaquinto AN, Sung H & Jemal A. Cancer statistics, 2025. *CA Cancer J Clin.* 2025; 75: 10-45. https://doi.org:10.3322/caac.21871
3. Chen Z, Fillmore CM, Hammerman PS, Kim CF & Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer.* 2014; 14: 535-546 . https://doi.org:10.1038/nrc3775
4. Uhlen M, *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics.* 2005; 4: 1920-1932. https://doi.org:10.1074/mcp.M500279-MCP200
5. Alevy YG, *et al.* IL-13-induced airway mucus production is attenuated by MAPK13 inhibition. *J Clin Invest.* 2012; 122: 4555-4568. https://doi.org:10.1172/JCI64896
6. Ahmadi SE, Rahimi S, Zarandi B, Chegeni R & Safa M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. *J Hematol Oncol.* 2021; 14: 121 (2021). https://doi.org:10.1186/s13045-021-01111-4
7. Johnson DE, O'Keefe RA & Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat Rev Clin Oncol.* 2018; 15: 234-248. https://doi.org:10.1038/nrclinonc.2018.8