Spring 2026, Volume 11, Issue 41 (17- 25) DOI: 10.22034/ppmj.2026.738542

Personalized & Precision Medicine Journal

Homepage: www.pmjournal.ir

Exploratory Transcriptomic Profiling of Three RNA-Seq Samples Reveals Heterogeneity in KRAS-Associated Signaling and Epithelial Lineage Programs

Negar Fallah Azad [1,*] Mohammad Sadegh Mahmoudvandyam[2]

[1] Department of Pathogenic Microorganisms, Faculty of Basic Sciences and Advanced Technologies In Biology, University of science and culture , Tehran, Iran.

[2] Department of Microbiology (Pathogenic Microorganisms), Faculty of Biological Sciences, University of Science and Culture, Tehran, Iran.

ARTICLE INFO

Paper Type: Review Article

Submitted: 2026-01-28
Accepted: 2025-05-10

Keywords:
RNA-seq
KRAS signaling
Transcriptomic heterogeneity
Epithelial lineage
Lung adenocarcinoma

Corresponding author:
Negar Fallah Azad
Email: Ngarfallahazad@gmail.com.

ABSTRACT

KRAS-driven lung adenocarcinoma is molecularly heterogeneous; however, evaluation of mutation-associated programs requires matched mutation and clinical metadata. Here, three GDC STAR-count RNA-sequencing files annotated with GENCODE v36 were analyzed to determine which transcriptomic conclusions could be supported from the available data alone. Protein-coding TPM values were used for library-level quality assessment, transcriptome-wide correlation, principal component analysis, sample-specific expression contrasts, and exploratory scoring of curated KRAS/MAPK, PI3K–AKT–mTOR, cell-cycle, epithelial-lineage, epithelial–mesenchymal transition, hypoxia, inflammatory, interferon, apoptosis, and NRF2-related gene sets. The samples contained 24.3–49.5 million assigned reads and 12,808–14,574 protein-coding genes with TPM ≥1. Pairwise transcriptome correlations ranged from r=0.745 to 0.825. Marked biological heterogeneity was observed. Sample 1 showed a secretory epithelial profile characterized by PAEP, CEACAM5, CEACAM6, BPIFA1, and EPCAM. Sample 2 displayed a keratinizing/squamous-like program dominated by KRT5, KRT6A, KRT14, KRT17, SPRR family genes, and elevated NRF2-associated and glycolytic scores. Sample 3 exhibited a strong alveolar-lineage and immune-associated profile, with high SFTPA1, SFTPA2, SFTPB, SFTPC, HLA-DRA, CD74, and inflammatory pathway scores. KRAS expression was detectable in all samples (16.33–30.09 TPM), but RNA expression alone could not establish KRAS mutation status. These findings demonstrate substantial lineage and pathway heterogeneity among the three transcriptomes and provide a hypothesis-generating framework for subsequent mutation-informed analysis. Mutation-stratified differential expression, prognostic modeling, and survival analysis were not performed because mutation and clinical outcome data were unavailable.

How to Cite this Article:

F. Eghbalpour, M. Saremi. "Autoimmune Diseases and Their Relationship with Environmental Pollution" Personalized & Precision Medicine Journal, Vol. 11, no. 41, pp. 17- 25.

INTRODUCTION

Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small-cell lung cancer and is characterized by extensive genomic, transcriptomic, and phenotypic heterogeneity. Large-scale molecular profiling by The Cancer Genome Atlas demonstrated that LUAD contains recurrent alterations in receptor tyrosine kinase–RAS–RAF signaling, cell-cycle control, oxidative-stress regulation, chromatin remodeling, and lineage-determining transcriptional programs (1). This diversity has important biological and clinical consequences because tumors with

a similar histopathological diagnosis may differ markedly in cellular differentiation, metabolic state, immune contexture, and therapeutic vulnerability. Transcriptome profiling therefore provides a functional layer of information that complements genomic testing by revealing the cellular programs that are active in an individual tumor specimen.

KRAS is one of the most frequently altered oncogenic drivers in LUAD. The encoded small GTPase operates as a molecular switch that couples upstream growth-factor signals to several downstream effector networks, including RAF–MEK–ERK, PI3K–AKT–mTOR, RAL-GEF, and stress-adaptation pathways. Activating substitutions, most commonly affecting codon 12, impair normal GTP hydrolysis and favor persistent signaling that promotes proliferation, survival, metabolic reprogramming, invasion, and remodeling of the tumor microenvironment (2, 7). The clinical development of covalent KRAS G12C inhibitors has confirmed that mutant KRAS can be therapeutically targeted; however, incomplete responses and acquired resistance indicate that the biological state of a KRAS-driven tumor cannot be explained by the driver mutation alone (8).

A major source of heterogeneity in KRAS-associated LUAD is the presence of co-occurring genomic alterations. Mutations or losses involving TP53, STK11/LKB1, KEAP1, and CDKN2A/B define biologically distinct subsets with different inflammatory profiles, metabolic dependencies, differentiation states, and treatment responses (2, 9). In particular, STK11 loss can produce an immune-cold phenotype, whereas concurrent KEAP1–NRF2 pathway activation enhances antioxidant defense, redox homeostasis, and nutrient dependence in KRAS-mutant tumors (9, 10). These observations emphasize that pathway activity should be interpreted as the product of interacting genomic events and cellular context rather than as a direct readout of KRAS messenger RNA abundance.

Epithelial lineage identity is another central determinant of LUAD biology. Alveolar type II cells are an established cell of origin for experimental KRAS-driven lung adenocarcinoma, and surfactant genes such as SFTPA1, SFTPA2, SFTPB, and SFTPC provide transcriptional evidence of alveolar differentiation (11). The lineage transcription factor NKX2-1/TTF-1 regulates pulmonary identity and may function as a lineage-survival factor in differentiated LUAD, while loss or redistribution of its transcriptional program can accompany progression toward poorly differentiated, mucinous, gastric-like, or metastatic states (6, 12-14). Conversely, increased expression of basal keratins and small proline-rich proteins may indicate a basal or squamous-like differentiation program. Secretory epithelial markers, including CEACAM5, CEACAM6, EPCAM, and AGR2, represent a further state associated with epithelial differentiation, adhesion, secretory-protein processing, and potentially invasive behavior (15). Thus, differences in lineage composition can substantially modify the apparent output of oncogenic signaling pathways in bulk RNA-sequencing data.

The interpretation of bulk tumor RNA sequencing is additionally influenced by non-neoplastic stromal and immune cells. Expression of HLA class II genes, CD74, interferon-response genes, and leukocyte-associated transcripts may reflect immune infiltration, tumor-cell-intrinsic inflammatory signaling, or both. Without matched pathology, tumor-purity estimates, somatic-variant calls, and clinical metadata, such signals should be treated as biologically informative but not assigned to a specific cell population or prognostic category. Nevertheless, carefully constrained exploratory analysis can identify dominant epithelial states, stress-response programs, and immune-associated features that guide subsequent hypothesis-driven investigation.

In the present study, three GDC STAR-count RNA-sequencing files were analyzed to characterize their principal transcriptomic features and to determine which conclusions could be supported from gene-expression data alone. The analysis focused on library-level quality metrics, global transcriptional similarity, sample-enriched genes, epithelial and alveolar lineage markers, KRAS-associated signaling components, oxidative-stress programs, and immune-related expression. Because matched mutation and survival information was unavailable, the study was intentionally framed as an exploratory three-sample transcriptomic analysis rather than a mutation-stratified or prognostic investigation. This conservative design allowed biologically meaningful heterogeneity to be described while avoiding unsupported claims regarding KRAS genotype, differential expression, or clinical outcome.

MATERIALS AND METHODS

RNA-sequencing data and preprocessing

Three GDC gene-expression files generated by the STAR-count workflow were analyzed. Each file contained gene identifiers, gene symbols, gene biotypes, unstranded and stranded read counts, transcripts per million (TPM), fragments per kilobase per million, and upper-quartile-normalized FPKM values. The files were annotated against GENCODE v36, consistent with the GDC harmonized mRNA pipeline (3). The unstranded count column was used for library summaries, whereas TPM values were used for cross-sample descriptive analyses. Rows representing alignment summary categories were separated from gene-level records. Analyses were restricted to protein-coding genes, and duplicated gene symbols, where present, were aggregated.

Quality assessment and exploratory transcriptomic analysis

For each sample, assigned reads, unmapped reads, multimapping reads, reads without an annotated feature, ambiguous reads, the number of detected protein-coding genes, mitochondrial read fraction, and ribosomal read fraction were calculated. Expression values were transformed as log2(TPM+1). Pairwise Pearson and Spearman correlations were calculated across protein-coding genes. Principal component analysis was performed on the 1,000 most variable protein-coding genes after feature standardization.

Because the dataset contained only three unlabelled samples and no biological replicates per group, formal differential-expression testing and false-discovery-rate estimation were not performed.

Sample-specific genes and pathway scoring

For descriptive identification of sample-enriched transcripts, each sample was compared with the mean TPM of the other two samples using log2((TPM+0.1)/(mean TPM of the other samples+0.1)). Genes with TPM ≥5 were ranked by this descriptive ratio; these values were not treated as inferential fold changes. Exploratory pathway scores were calculated from curated gene sets representing KRAS/MAPK, PI3K–AKT–mTOR, cell cycle/E2F, epithelial–mesenchymal transition (EMT), epithelial and alveolar lineage, hypoxia/glycolysis, inflammatory/NF-κB, interferon/antigen presentation, apoptosis, and KEAP1–NRF2 oxidative-stress programs. For each gene, log2(TPM+1) values were standardized across the three samples, and pathway scores were defined as the mean standardized expression of genes in the corresponding set. These scores indicate relative activity only within this three-sample dataset and are not equivalent to formal GSEA or GSVA.

RESULTS

Library characteristics

The three libraries contained 49.5, 35.5, and 24.3 million assigned gene-level reads, respectively. The proportion of assigned reads among all STAR accounting categories ranged from 38.6% to 41.6%. Protein-coding genes with TPM ≥1 numbered 12,808, 14,516, and 14,574. Mitochondrial reads represented 1.77%, 1.52%, and 0.70% of assigned gene counts, indicating no marked mitochondrial-read excess (Table 1 and Figure 1).

Global transcriptomic similarity and separation

Pairwise Pearson correlations of log2-transformed protein-coding expression were 0.825 between Samples 1 and 2, 0.745 between Samples 1 and 3, and 0.822 between Samples 2 and 3. Spearman correlations showed the same pattern, indicating that Sample 1 and Sample 3 were the most transcriptionally divergent pair. Principal component analysis of the 1,000 most variable genes separated all three samples, with Sample 2 and Sample 3 occupying distinct expression states and Sample 1 forming a third independent profile (Figures 2 and 3).

Table 1. RNA-sequencing library summary

Sample	Assigned reads	Assigned (%)	PC genes TPM>0	PC genes TPM≥1	Mitochondrial (%)	KRAS TPM
Sample_1	49,496,117	41.6	17,503	12,808	1.77	16.33
Sample_2	35,507,173	38.8	17,977	14,516	1.52	30.09
Sample_3	24,300,711	38.6	17,847	14,574	0.70	23.16

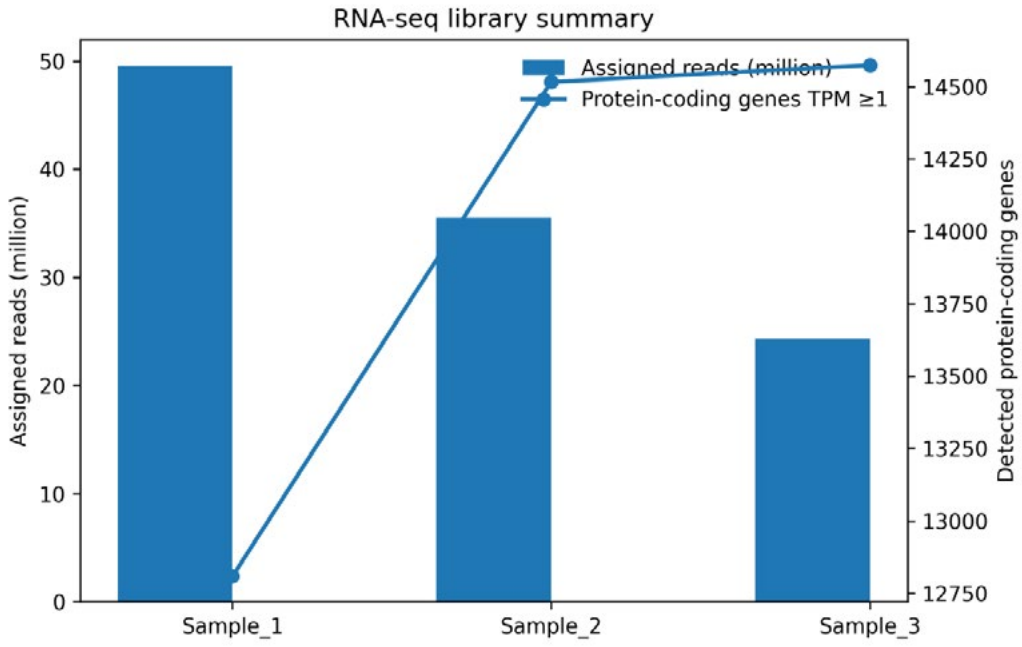

Fig 1. RNA-seq library summary. Bars show assigned gene-level reads; the line shows the number of protein-coding genes with TPM ≥1.

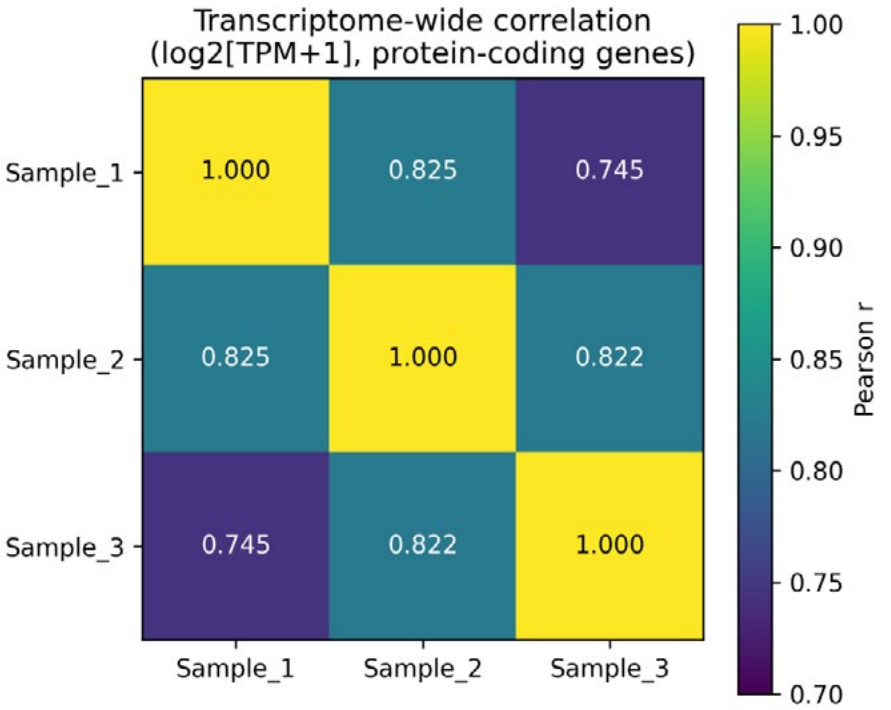

Fig2. Transcriptome-wide Pearson correlation based on log2(TPM+1) values for protein-coding genes.

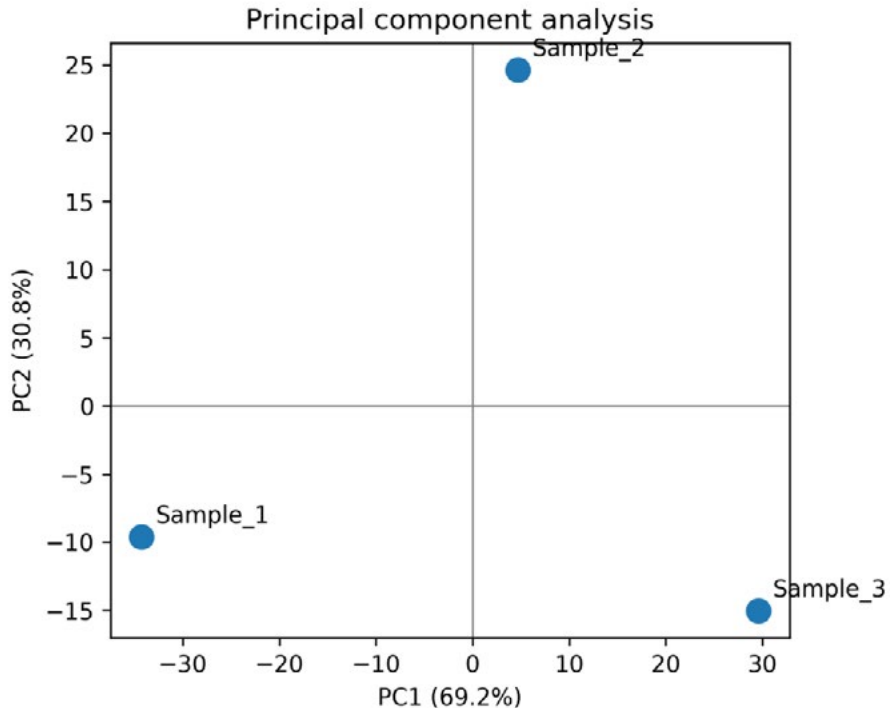

Fig 3. Principal component analysis of the 1,000 most variable protein-coding genes. The analysis is descriptive because only three samples were available.

Distinct epithelial and lineage-associated expression programs

Sample 1 was characterized by high expression of CEACAM6, CEACAM5, KRT18, AGR2, BPIFA1, EPCAM, and KRT19, together with marked sample-specific expression of PAEP, COL2A1, SEPTIN14, CYP1A1, and several secretory epithelial transcripts. This profile was consistent with a differentiated secretory epithelial state. Sample 2 showed a striking keratinizing program, with KRT5, KRT6A, KRT14, KRT17, KRT16, DSG3, SFN, and multiple small proline-rich proteins among its dominant or most sample-enriched genes. Sample 3 showed a strong alveolar-associated program, with SFTPA1, SFTPA2, SFTPB, and SFTPC among the most abundant transcripts, accompanied by HLA-DRA, HLA-DRB1,

CD74, B2M, and JCHAIN. These profiles indicate pronounced differences in epithelial differentiation and immune-cell or inflammatory contributions among the samples.

Exploratory KRAS-associated pathway patterns

KRAS transcript expression was detectable in all samples and was highest in Sample 2 (30.09 TPM), followed by Sample 3 (23.16 TPM) and Sample 1 (16.33 TPM). However, transcript abundance is not a surrogate for mutation status. Relative pathway scoring showed that Sample 2 had the highest PI3K–AKT–mTOR, hypoxia/glycolysis, interferon/antigen-presentation, apoptosis, and KEAP1–NRF2 oxidative-stress scores. Sample 3 had the highest KRAS/MAPK, EMT, epithelial/alveolar-lineage, and inflammatory/NF-κB scores, whereas its cell-cycle score was substantially lower than those of Samples 1 and 2. Sample 1 showed comparatively higher proliferation than Sample 3 but lower inflammatory, interferon, EMT, and oxidative-stress scores (Table 3 and Figure 5).

Scores are mean gene-wise z scores and indicate only relative activity within this dataset.

DISCUSSION

The present exploratory analysis shows that the three RNA-sequencing profiles represent biologically distinct transcriptional states despite retaining moderate-to-high transcriptome-wide correlations. The dominant source of variation was not the abundance of KRAS transcripts themselves, but the combination of epithelial lineage, differentiation status, stress-response activity, and immune-associated expression. This observation is consistent with the broader molecular heterogeneity of lung adenocarcinoma, in which tumors sharing a histological diagnosis may differ substantially in cell of origin, differentiation program, co-occurring molecular alterations, and microenvironmental composition (1). These factors are particularly relevant to studies of RAS signaling because the transcriptional consequences of pathway activation depend strongly on cellular context and cannot be interpreted independently of lineage identity (1, 2, 9, 10).

Sample 1 was characterized by a secretory epithelial program marked by CEACAM5, CEACAM6, AGR2, EPCAM, KRT18, and KRT19, together with pronounced PAEP expression. CEACAM5 and CEACAM6 are frequently associated with epithelial tumor differentiation, cell adhesion, and invasive behavior (15), whereas AGR2 is an endoplasmic-reticulum protein linked to secretory activity and maintenance of epithelial tumor phenotypes. The coordinated expression of these genes suggests a differentiated, secretory carcinoma-like state rather than a highly mesenchymal phenotype. The presence of EPCAM and keratin 18/19 further supports preservation of epithelial identity. However, because no histopathological or genomic metadata were available, this profile should be interpreted as a transcriptional phenotype and not as evidence for a specific LUAD subtype or molecular driver.

Sample 2 displayed the most distinctive basal and keratinizing program, with high expression of KRT5, KRT6A, KRT14, KRT17, and multiple small proline-rich proteins. This combination is more typical of basal or squamous differentiation than of a classical surfactant-producing LUAD phenotype. Several explanations are possible, including a poorly differentiated or transdifferentiated adenocarcinoma, mixed histology, sampling of a squamous component, or inaccurate histological assignment. Bulk RNA-sequencing alone cannot distinguish among these possibilities. Nevertheless, the finding is important because inclusion of such a sample in a small mutation-stratified analysis could generate apparently strong differential-expression signals that actually reflect lineage imbalance rather than KRAS mutation status. Histological review and expression of lineage markers such as NKX2-1, NAPSA, TP63, and KRT5 should therefore be considered when constructing mutation-

Table 2. Top sample-enriched protein-coding transcripts

Rank	Sample 1	Sample 2	Sample 3
1	PAEP (2879.4 TPM)	SPRR1A (2432.5 TPM)	SFTPC (12371.1 TPM)
2	SEPTIN14 (513.5 TPM)	KRT6A (3312.0 TPM)	PGC (937.4 TPM)
3	COL2A1 (610.2 TPM)	KRT14 (3085.7 TPM)	SFTPA1 (9524.9 TPM)
4	FGB (191.4 TPM)	KRT5 (3307.0 TPM)	SFTPB (4274.9 TPM)
5	MAGEA12 (94.5 TPM)	SPRR1B (1241.9 TPM)	SFTA2 (133.8 TPM)
6	OBP2A (127.9 TPM)	SPRR3 (638.6 TPM)	MT1M (537.7 TPM)
7	BPIFA2 (45.8 TPM)	SPRR2A (613.9 TPM)	HLA-DQA2 (236.5 TPM)
8	SPAG11B (27.3 TPM)	DSG3 (717.6 TPM)	PEBP4 (142.6 TPM)
9	CSAG1 (59.4 TPM)	SPRR2E (439.9 TPM)	PLA2G1B (55.3 TPM)
10	CYP1A1 (237.0 TPM)	KRT16 (617.8 TPM)	HAS1 (76.7 TPM)

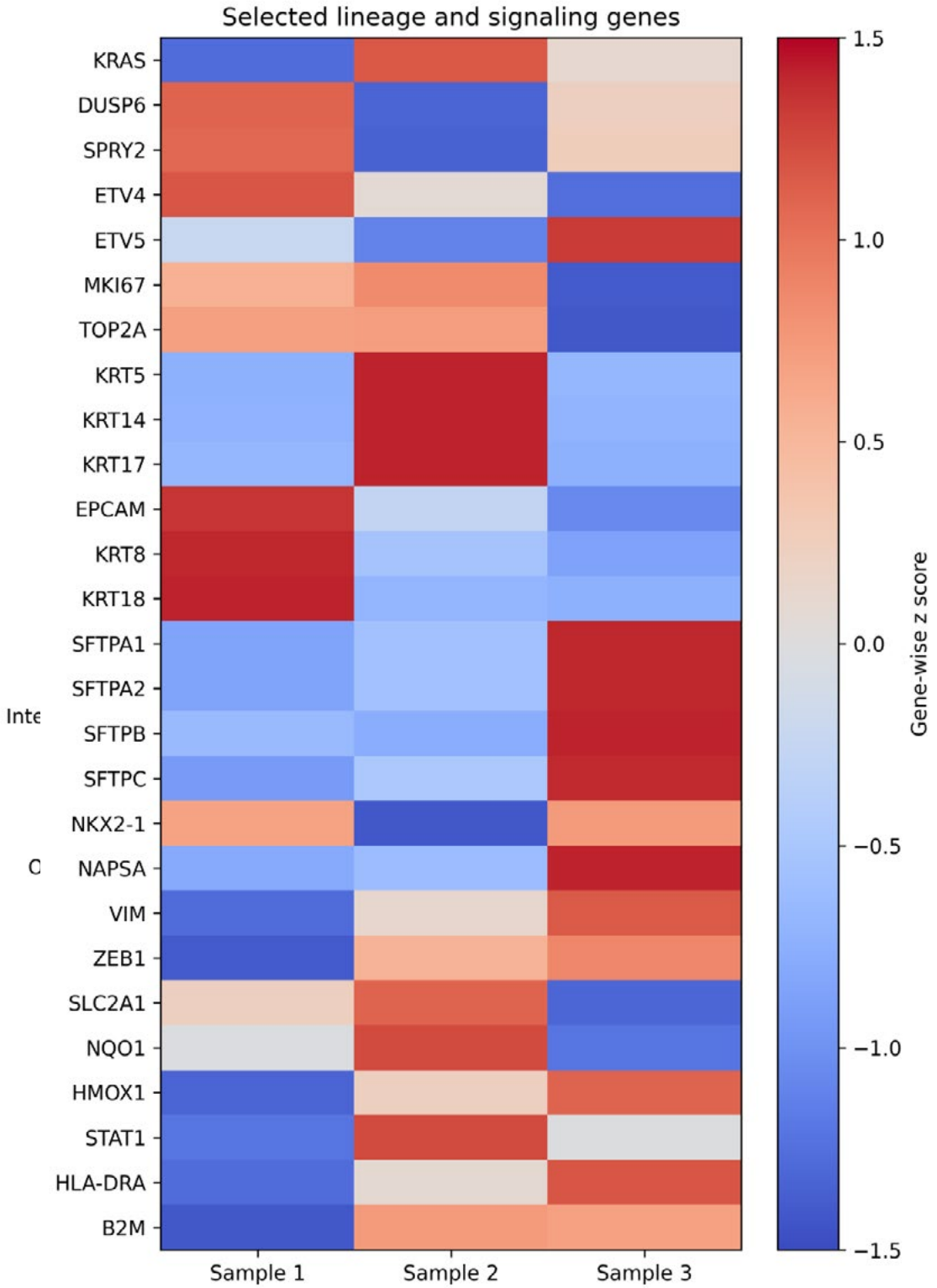

Fig4. Relative expression of selected lineage, KRAS-associated signaling, proliferation, stress-response, and immune genes. Values are gene-wise z scores across the three samples.

comparison cohorts.

Sample 2 also showed the highest relative NRF2-associated and glycolytic scores. This pattern was supported by increased expression of oxidative-stress and detoxification genes, including NQO1, members of the AKR1C family, and SLC7A11. Activation of this transcriptional program can promote adaptation to reactive oxygen species, maintenance of redox balance, and metabolic survival under stressful conditions. In LUAD, similar expression patterns may occur in tumors with KEAP1 or NFE2L2 pathway alterations, but they can also be induced by non-genetic oxidative

Table 3. Relative pathway scores across the three samples

Pathway	Sample 1	Sample 2	Sample 3	Genes
KRAS/MAPK signaling	-0.34	-0.12	0.45	20
PI3K–AKT–mTOR	-0.63	0.55	0.08	18
Cell cycle/E2F	0.61	0.76	-1.37	19
EMT	-0.84	0.18	0.67	16
Epithelial/LUAD markers	0.14	-0.64	0.51	13
Hypoxia/glycolysis	-0.33	0.69	-0.36	13
Inflammatory/NF-kB	-0.84	-0.44	1.29	13
Interferon/antigen presentation	-1.32	0.90	0.42	17
Apoptosis	-0.84	0.55	0.30	14
Oxidative stress/KEAP1-NRF2	-0.48	1.13	-0.64	13

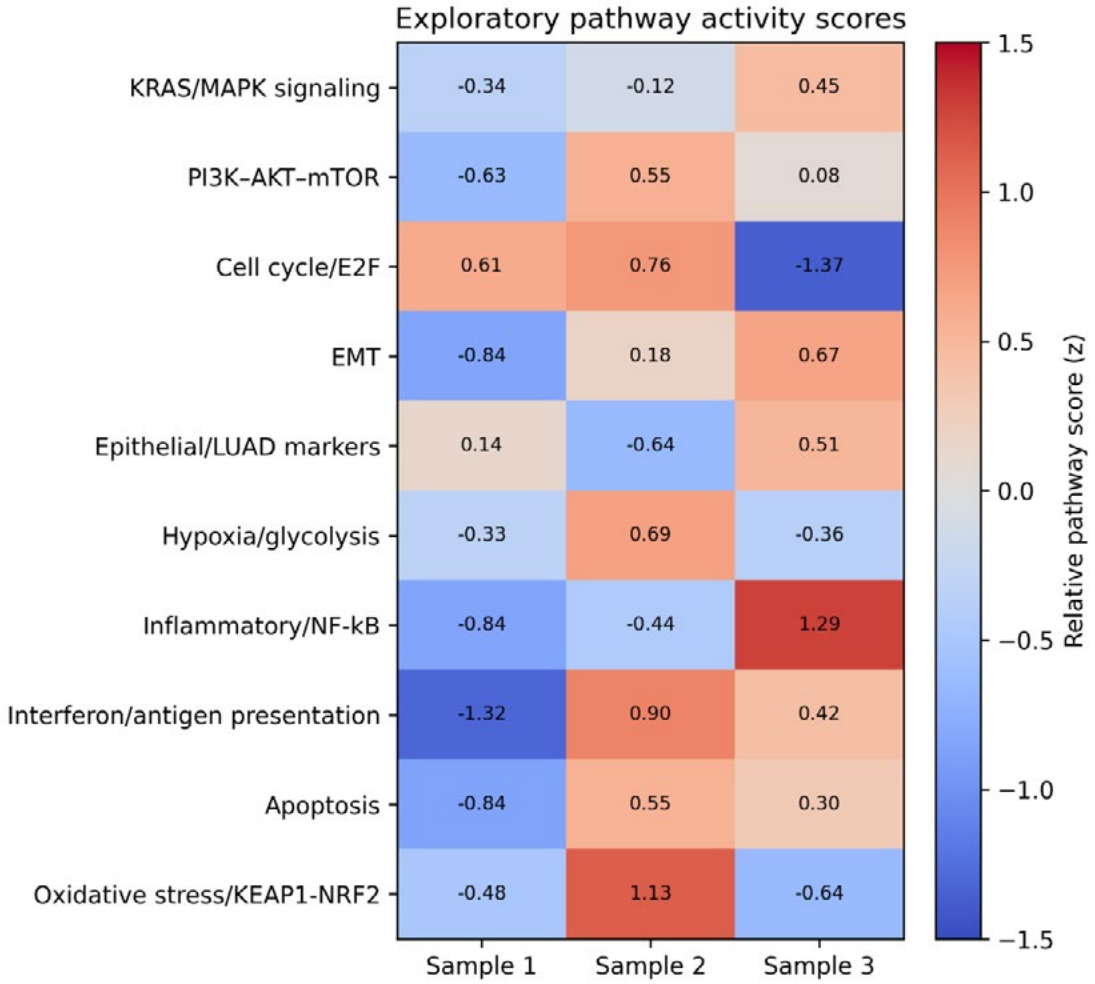

Fig5. Exploratory pathway activity scores. Red indicates relatively higher and blue relatively lower standardized expression within the three-sample dataset.

stress and metabolic remodeling. Consequently, the current findings indicate an NRF2-like transcriptional state but do not establish the presence of a KEAP1/NFE2L2 mutation. This distinction is critical because pathway activity inferred from RNA expression is a functional readout, whereas mutation status is a genomic classification, and the two are related but not interchangeable.

Sample 3 exhibited a markedly different phenotype characterized by strong expression of the surfactant genes SFTPA1, SFTPA2, SFTPB, and SFTPC. These genes are hallmarks of alveolar type II cell differentiation and support an alveolar lineage state. Maintenance of alveolar identity can influence oncogenic signaling,

tumor differentiation, and therapeutic dependencies. Experimental and clinical studies have shown that lineage-defining transcription factors, including NKX2-1 and FOXA-family factors, can modulate the biological output of oncogenic pathways and preserve distinct LUAD cell states (6, 11-14, 16). The surfactant-rich profile of Sample 3 therefore provides a plausible biological explanation for its separation from the keratinizing Sample 2, even before consideration of any underlying mutation differences.

In addition to alveolar markers, Sample 3 showed increased expression of HLA class II-related genes, CD74, B2M, and other immune-associated transcripts. This pattern may reflect greater infiltration by antigen-presenting immune cells, interferon-responsive tumor cells, or a combination of both. Because the analyzed files were derived from bulk tissue, expression originating from malignant epithelial cells cannot be separated from stromal or immune-cell expression without deconvolution or single-cell data. Nevertheless, the coordinated elevation of antigen-presentation genes and inflammatory pathway scores suggests that Sample 3 had a more immune-active transcriptional environment than the other two samples. In a larger cohort, this phenotype would warrant evaluation using immune-deconvolution methods and comparison with tumor purity, interferon signaling, and immune-checkpoint expression.

KRAS transcripts were detected in all three samples, with the highest TPM value in Sample 2. This variation should not be interpreted as evidence of KRAS mutation or oncogenic activity. Activating KRAS mutations alter the biochemical state and downstream signaling of the protein and do not necessarily cause a proportional increase in KRAS mRNA abundance. Conversely, elevated KRAS expression may occur in wild-type tumors. Similarly, the relative KRAS/MAPK pathway score observed in Sample 3 is not specific to mutant KRAS, because MAPK-responsive genes can be activated by EGFR, ERBB2, MET, BRAF, NF1 loss, inflammatory signaling, and other upstream mechanisms. The present analysis therefore distinguishes between KRAS expression, RAS/MAPK-related transcriptional output, and confirmed KRAS mutation status; only the first two can be evaluated from the available files.

The data also illustrate why pathway-level interpretation may be more informative than inspection of a single oncogene transcript. The three samples differed simultaneously in epithelial differentiation, oxidative-stress adaptation, glycolysis, inflammatory signaling, and antigen-presentation programs. Such coordinated changes can alter tumor behavior and may modify response to targeted or immune-based therapies even when the initiating genomic alteration is shared. In larger mutation-annotated LUAD cohorts, combining somatic-variant data with transcriptomic pathway scores could help distinguish biologically different KRAS-mutant subgroups, including tumors dominated by epithelial lineage programs, oxidative-stress adaptation, or immune activation. This strategy would be more robust than defining tumors solely by KRAS mRNA abundance.

From a methodological perspective, the moderate-to-high global correlations observed among the samples indicate that all three retained a common core human epithelial transcriptome, whereas PCA and marker-gene analysis captured biologically meaningful differences in a smaller subset of highly variable genes. This pattern is expected in bulk tumor transcriptomics: most housekeeping and essential cellular genes remain broadly correlated, while lineage markers, immune genes, metabolic programs, and stress-response pathways drive sample separation. The concordance between PCA, sample-enriched genes, and curated pathway scores increases confidence that the observed distinctions are not attributable to a single isolated transcript. However, because pathway scores were calculated as relative standardized means across only three samples, they should be viewed as descriptive indicators rather than estimates of absolute pathway activation.

Collectively, these findings support a model in which transcriptional heterogeneity among the three specimens is shaped primarily by epithelial lineage and microenvironmental context. Sample 1 represents a secretory epithelial state, Sample 2 a basal/keratinizing and oxidative-stress-adapted state, and Sample 3 an alveolar and immune-associated state. These profiles generate testable hypotheses for subsequent work. Mutation sequencing could determine whether the samples carry KRAS or other RAS-pathway alterations; histopathological reassessment could clarify the basal/squamous-like phenotype of Sample 2; and immune deconvolution or immunohistochemistry could evaluate the antigen-presentation signature of Sample 3. A larger cohort with matched mutation, clinical, and survival data would then be required to identify statistically supported KRAS-associated hub genes and determine whether any of the observed programs have prognostic significance.

Limitations

The study included only three samples without known biological grouping, mutation status, histopathological annotation, treatment information, or survival data. Accordingly, formal differential-expression testing, multiple-testing correction, mutation-stratified comparisons, protein–protein interaction network inference, hub-gene selection, Kaplan–Meier analysis, and Cox regression were not statistically justified and were not performed. The curated pathway scores were exploratory and were based on relative standardized expression rather than formal enrichment against a reference cohort. Potential differences in tumor purity and stromal or immune-cell content could also contribute to the observed patterns.

CONCLUSION

Analysis of the three available STAR-count RNA-sequencing files identified substantial transcriptomic heterogeneity. Sample 1 displayed a secretory epithelial program, Sample 2 a keratinizing/squamous-like and oxidative-stress-associated program, and Sample 3 an alveolar and inflammatory program. KRAS and multiple downstream signaling genes were expressed in all samples, but neither KRAS mutation status nor prognostic relevance could be determined. These results

provide a defensible exploratory manuscript based solely on the available data and identify the metadata required for a subsequent mutation-informed LUAD study.

Data Availability

The analysis was based on three GDC STAR-count gene-expression files supplied for the study. Derived summary tables and figures were generated from the unstranded count and TPM columns.

Conflict of Interest

The authors declare that they have no conflicts of interest related to this study.

Consent for Publication

Not applicable. This study did not involve the collection of identifiable personal information, and no individual-level identifiable data are presented in this manuscript.

Funding Support

This study received no specific funding from any public, commercial, or not-for-profit funding agency.

Ethical Considerations

Ethical approval was not required for this study because the analysis was based exclusively on previously generated RNA-sequencing data obtained from the Genomic Data Commons (GDC). No human participants were prospectively recruited, no intervention was performed, and no identifiable personal information was collected or analyze.

REFERENCES

1.Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 9;511(7511):543.

2. Skoulidis F, Byers LA, Diao L, Papadimitrakopoulou VA, Tong P, Izzo J, Behrens C, Kadara H, Parra ER, Canales JR, Zhang J. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. Cancer discovery. 2015 Aug 1;5(8):860-77.

3. Skoulidis F, Li BT, Dy GK, Price TJ, Falchook GS, Wolf J, Italiano A, Schuler M, Borghaei H, Barlesi F, Kato T. Sotorasib for lung cancers with KRAS p. G12C mutation. New England Journal of Medicine. 2021 Jun 24;384(25):2371-81.

4. Gu Y, Zhang Z, Camps MG, Ossendorp F, Wijdeven RH, Ten Dijke P. Genome-wide CRISPR screens define determinants of epithelial-mesenchymal transition mediated immune evasion by pancreatic cancer cells. Science Advances. 2023 Jul 14;9(28):eadf9915.

5. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15-21.

6.Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell systems. 2015 Dec 23;1(6):417-25.

7.Orstad G, Fort G, Parnell TJ, Jones A, Stubben C, Lohman B, Gillis KL, Orellana W, Tariq R, Klingbeil O, Kaestner K. FoxA1 and FoxA2 control growth and cellular identity in NKX2-1-positive lung adenocarcinoma. Developmental cell. 2022 Aug 8;57(15):1866-82.

8.Huang L, Guo Z, Wang F, Fu L. KRAS mutation: from undruggable to druggable in cancer. Signal transduction and targeted therapy. 2021 Nov 15;6(1):386.

9.Skoulidis F, Goldberg ME, Greenawalt DM, Hellmann MD, Awad MM, Gainor JF, Schrock AB, Hartmaier RJ, Trabucco SE, Gay L, Ali SM. STK11/LKB1 mutations and PD-1 inhibitor resistance in KRAS-mutant lung adenocarcinoma. Cancer discovery. 2018 Jul 1;8(7):822-35.

10.Galan-Cobo A, Sitthideatphaiboon P, Qu X, Poteete A, Pisegna MA, Tong P, Chen PH, Boroughs LK, Rodriguez ML, Zhang W, Parlati F. LKB1 and KEAP1/NRF2 pathways cooperatively promote metabolic reprogramming with enhanced glutamine dependence in KRAS-mutant lung adenocarcinoma. Cancer research. 2019 Jul 1;79(13):3251-67.

11. Lin C, Song H, Huang C, Yao E, Gacayan R, Xu SM, Chuang PT. Alveolar type II cells possess the capability of initiating lung tumor development. PloS one. 2012 Dec 20;7(12):e53817.

12. Winslow MM, Dayton TL, Verhaak RG, Kim-Kiselak C, Snyder EL, Feldser DM, Hubbard DD, DuPage MJ, Whittaker CA, Hoersch S, Yoon S. Suppression of lung adenocarcinoma progression by Nkx2-1. Nature. 2011 May 5;473(7345):101-4.

13. Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen TA, Wang DG, Crowley D, Whittaker CA, Meyerson M, Kimura S, Jacks T. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. Molecular cell. 2013 Apr 25;50(2):185-99.

14. Little DR, Lynch AM, Yan Y, Akiyama H, Kimura S, Chen J. Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves opposing murine alveolar cell fates in vivo. Nature communications. 2021 May 4;12(1):2509.

15. Kim EY, Cha YJ, Jeong S, Chang YS. Overexpression of CEACAM6 activates Src-FAK signaling and inhibits anoikis, through homophilic interactions in lung adenocarcinomas. Translational oncology. 2022 Jun 1;20:101402.

16.Tanaka H, Yanagisawa K, Shinjo K, Taguchi A, Maeno K, Tomida S, Shimada Y, Osada H, Kosaka T, Matsubara H, Mitsudomi T. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. Cancer research. 2007 Jul 1;67(13):6007-11.