



Identification of Two Subtypes and Prognostic Characteristics of Lung Adenocarcinoma Based on Pentose Phosphate Metabolic Pathway-Related Long Non-coding RNAs

Chuan Liu and Yongjie Wang*

Department of Thoracic Surgery, Affiliated Hospital of Qingdao University, Qingdao, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Kai Yang,
First Affiliated Hospital of Chongqing
Medical University, China
Min Wei,
Shanghai Jiao Tong University, China

*Correspondence:

Yongjie Wang
wangyongjie@qduhospital.cn

Specialty section:

This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 23 March 2022

Accepted: 25 May 2022

Published: 21 June 2022

Citation:

Liu C and Wang Y (2022) Identification of Two Subtypes and Prognostic Characteristics of Lung Adenocarcinoma Based on Pentose Phosphate Metabolic Pathway-Related Long Non-coding RNAs.
Front. Public Health 10:902445.
doi: 10.3389/fpubh.2022.902445

This study analyzed the differences in subtypes and characteristics of advanced lung adenocarcinoma (LUAD) patients based on the pentose phosphate metabolic pathway-related long non-coding RNAs (lncRNAs), along with their potential regulatory mechanisms. Using the expression profiling and corresponding clinical information of LUAD patients from Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA). Differential pathway scores between normal and tumor samples from TCGA were identified by rank-sum tests. Pearson correlation coefficients between pentose phosphate scores of the pentose phosphate samples and lncRNAs of the corresponding datasets were calculated. Next, the clusterProfiler software package was used for functional annotation. Clustering of pentose phosphate-related lncRNAs from LUAD samples categorized two molecular subtypes (C1, and C2). C1 was associated with a lower pentose phosphate score and a good prognosis; the C2 showed a higher pentose phosphate score and was related to poorer prognoses. The C2 was markedly associated with energy metabolic pathways. The expression of most immune cells were markedly higher in C1 subtype. Some crucial immune checkpoints, including CTLA4, CD274, and CD47, were also significantly upregulated in C1 subtype, leading to a higher score of clinical effect on the C1 subtype. Finally, one TF, BACH1, was found to be significantly upregulated in C1 subtypes; the pathways activated by this TF may be associated with tumor progression and poor prognoses. LUAD typing based on pentose phosphate metabolic pathway-related lncRNAs was confirmed. Differences in characteristics between C1 and C2 subtypes improved the current LUAD detection and treatment.

Keywords: lung adenocarcinoma, long non-coding RNAs, pentose phosphate metabolic pathway, transcription factors, immunotherapy

INTRODUCTION

Among the malignancies associated with the highest mortality, worldwide, lung cancer leads (1). The most frequent lung cancer histological subtype, accounting for ~50% of all the cases, is lung adenocarcinoma (LUAD). Patients with LUAD are at a high risk of distant metastases at each stage of the disease progression (2), along with high malignancy and poor prognoses (3, 4). Treatment of LUAD is based on grading and stage, mainly determined using the patients' characteristics and the pathological assessment of the tumor histology (5).

Especially, for metastatic LUAD patients, traditional therapy such as surgery and radiotherapy do a few efficacy for them. Targeted therapy or immunotherapy could be a superior choice for metastatic patients. Targeted therapeutic drugs such as anaplastic lymphoma kinase (ALK) inhibitors (6) and epidermal growth factor receptor (EGFR) inhibitors (7) have showed a promising performance in clinical trials for treating lung cancer patients with ALK or EGFR mutations. In the case of metastatic patients without these specific mutations, immunotherapy may be a wise choice. For example, programmed cell death protein 1/programmed cell death ligand 1 (PD-1/PD-L1) blockade therapy is considered as the most efficient in many cancer types (8). By the in combination with other treatments such as radiotherapy, chemotherapy or targeted therapy, the overall survival could be prolonged (9). However, not all patients can benefit from these therapies because of the individual differences and responses. Therefore, it is necessary to explore key molecular biomarkers for understanding further tumorigenesis mechanism and assisting personalized treatment in lung cancer.

Long non-coding RNAs (lncRNAs) are transcripts with a length of more than 200 nucleotides; they do not encode for proteins (10). Their functions are related to the subcellular localization of lncRNAs. Typically, lncRNAs in the nucleus can repress or activate the expression of their target genes by binding to them directly. In addition, they regulate gene expression through histone modification or recruitment of TFs. lncRNAs in the cytoplasm function as competitive endogenous RNAs through their interactions with corresponding miRNAs for the regulation of target gene expression (11–13). Several studies over the last decade show that lncRNAs exert different biological functions (14, 15). Aberrant expression of lncRNAs may result in a substantial effect on cancer progression, including cellular migration, proliferation, metastasis, and invasion (16).

Metabolism-related lncRNAs have been demonstrated to play an important role in regulating cancer metabolism and the crosstalk between cancer metabolism and tumor microenvironment (17). Previous studies have also discovered a series of lncRNA biomarkers or signatures related to metabolism for lung cancer based on bioinformatics analysis (18–20). However, limited research has explored molecular subtypes and prognostic biomarkers based on lncRNAs related to pentose phosphate pathway that is one of metabolism pathways involved in cancer development (21). PTTG3P, a lncRNA involved in the pentose phosphate pathway, is attenuated in non-small cell lung cancer, thereby leading to a poor prognosis (22). However, the

specific regulatory mechanisms of LUAD and pentose phosphate-related lncRNAs remain unclear.

Herein, we systematically investigated the correlation of pentose phosphate-related lncRNAs with the enrichment score of pentose phosphate pathway (pentose phosphate score) in patients with advanced stage (stages IV, III, and II) and examined their associated signaling pathways, along with their clinical significance.

MATERIALS AND METHODS

Expression Profile Data

The workflow of this study was shown in **Supplementary Figure 1**. The clinical information and corresponding RNA sequencing (RNA-seq) data of lung adenocarcinoma samples were obtained from The Cancer Genome Atlas (23) (TCGA, <https://portal.gdc.cancer.gov/>), named as TCGA dataset. In addition, expression data and clinical information of GSE30219 dataset (24) were extracted from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>).

Metabolism-Related Genes

Using the R package, Immuno-Oncology Biological Research (IOBR), 114 metabolism-related pathways with gene lists were extracted (**Supplementary Table 1**) (25).

Data Pre-processing

The RNA-seq data from TCGA was preprocessed in the following manner: (1) samples having incomplete survival information were excluded; samples of patients in stages II, III, or IV were retained; (2) ENSG was matched to GeneSymbol. The GSE30219 dataset was preprocessed as following: (1) the normalized dataset was downloaded; (2) samples with complete information on survival time and survival status were retained; (3) only the pentose phosphate samples of patients in stages IV, III, and II were retained; (4) GPL570 was re-annotated and the probes were then transformed to gene symbols.

Acquisition of the lncRNA Expression Profiles

The gene transfer format (GTF) file (version: V32) was obtained from the GENCODE website (<https://www.genecodegenes.org/>). TCGA expression profiles and GSE30219 data were divided into mRNAs or lncRNAs according to the annotations in the file.

Identification of Hubs in the Pentose Phosphate Pathway

Scores for metabolic pathways were calculated separately for TCGA and GSE30219 samples by single sample gene set enrichment analysis (ssGSEA) (26). Differences in scores between normal and tumor samples were identified using Wilcoxon test. Subsequently, prognosis-related metabolic pathways in TCGA and GSE30219 datasets were identified by univariate Cox analyses. Pentose phosphate pathway was identified in both two datasets and was considered as an important metabolic pathway for further analysis.

Identification of Pentose Phosphate-Related lncRNAs

Pearson correlation coefficients between pentose phosphate scores of the pentose phosphate samples and the lncRNAs of the corresponding datasets, along with their corresponding *P*-values were calculated. Filtering for significance was performed based on the set thresholds of $|\text{correlation}(\text{cor})| > 0.25$ and $P < 0.05$.

Identification of Pentose Phosphate-Related lncRNA Subtypes

The samples were clustered and typed by constructing a consistency matrix using ConsensusClusterPlus (27). The pentose phosphate-related lncRNA subtypes of the samples were obtained based on the pentose phosphate score-related lncRNAs. In addition, the Kaplan-Meier (KM) algorithm was executed and euclidean was set as the distance metric to perform 500 bootstraps, with each bootstrap process incorporating 80% of the patients belonging to the training set. The best classification was evaluated by computing the consistency matrix and the consistency cumulative distribution function (CDF), after setting the number of clusters between two and 10.

Gene Set Enrichment Analysis and Functional Annotation

The “GSEA” (26) package was employed to study the pathways of different molecular subtypes in biological processes. All candidate gene sets of KEGG pathways from the Molecular Signatures Database (MSigDB) database (28) were used to perform GSEA. Significantly enriched pathways were screened under the criteria of $P < 0.05$ and $Q < 0.25$. In addition, the clusterProfiler package (29) was used for functional annotation. Limma R package was used to identify differentially expressed genes (DEGs) between two subtypes (30). DEGs with $|\log_2(\text{fold change})| > 1$ and adjust $P < 0.05$ were screened.

Analysis of TF Activity

TF activity was scored according to the method developed by Garcia-Alonso (31) and TF activation levels were compared between clusters by analysis of variance (ANOVA). The screening criterion for significantly differentially expressed TFs was set at $P < 0.05$.

First-Order Partial Correlation Analysis

Analysis using first-order partial correlation was performed to evaluate the correlation of glycolysis-related genes and glycolysis scores with lncRNAs. Assuming that glycolysis score is x and glycolysis-related gene expression is y , the first-order partial correlation between x and y based on lncRNAs is as follows:

$$r_{xy|lncRNA} = \frac{r_{xy} - r_{xlncRNA} * r_{ylncRNA}}{\sqrt{(1 - r_{xlncRNA}^2) * (1 - r_{ylncRNA}^2)}}$$

Statistical Analysis

All statistical analysis and R packages were implemented in R software (v4.0). Log-rank test was conducted in KM survival

analysis. Wilcoxon test was conducted in testing the difference between two groups. $P < 0.05$ was considered as significant.

RESULTS

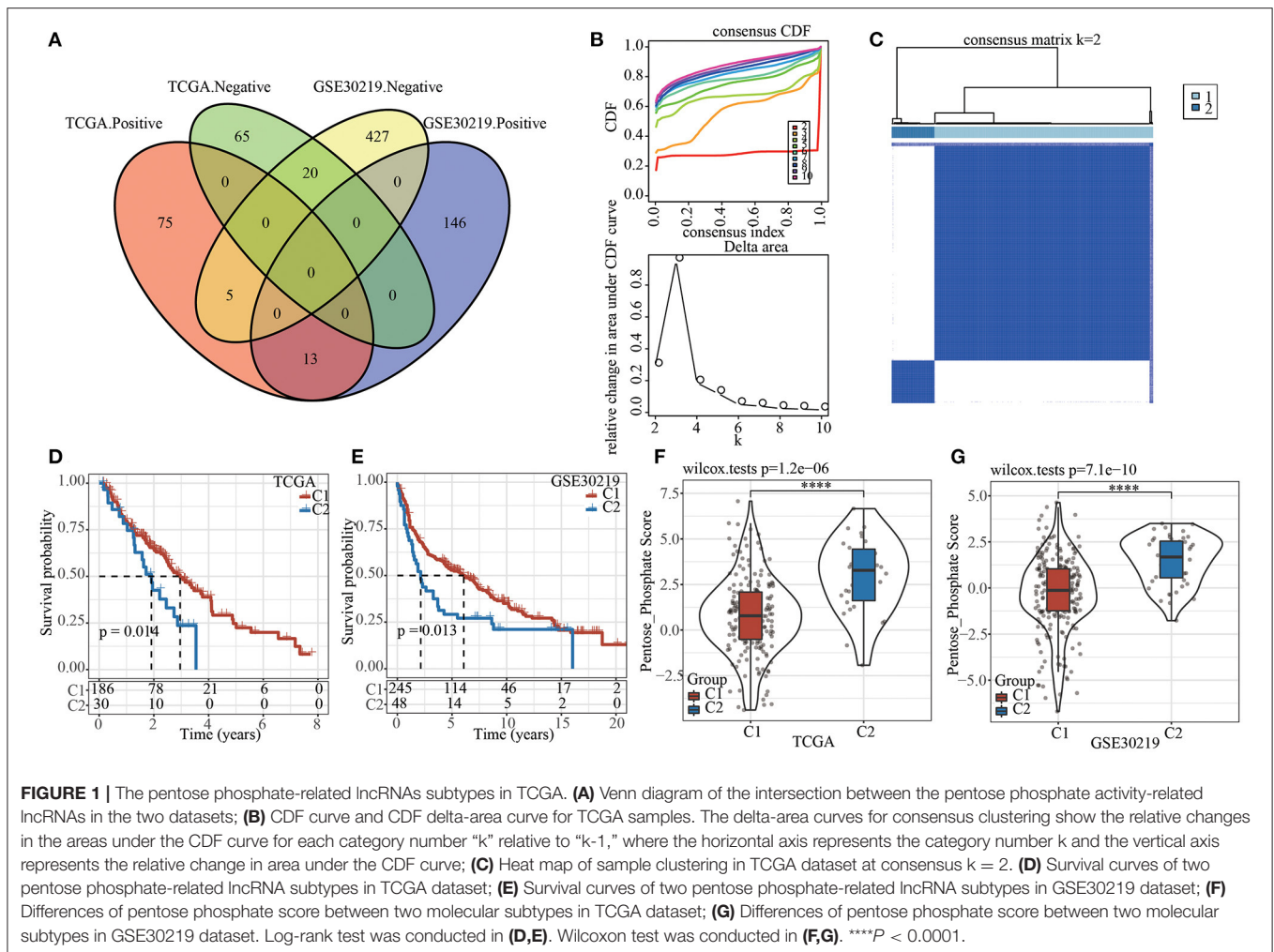
Molecular Typing Based on Pentose Phosphate Score-Related lncRNAs

Firstly, pentose phosphate pathway was identified as an important metabolic pathway related to prognosis in both TCGA and GSE30219 datasets (see Materials and methods 2.3). Correlation analysis between pentose phosphate score and lncRNA expression yielded 178 and 611 lncRNAs that were significantly associated with pentose phosphate activity in TCGA and GSE30219 datasets, respectively. A total of 38 lncRNAs were found in both two datasets (Figure 1A), but the correlations (negative and positive correlations) of five genes were found to be inconsistent in TCGA and GSE30219 datasets. Therefore, the remained 33 lncRNAs were included for identifying prognostic lncRNAs through univariate Cox analysis. Of them, 26 prognosis-related lncRNAs were screened, and their expression was used as an input for consensus clustering. Samples in TCGA dataset were clustered using ConsensusClusterPlus and the optimal number of clusters was determined based on the CDF. The CDF delta area curve shows that the clustering results were stable when the number of clusters was 2 (Figure 1B). Hence, we chose $k = 2$ and effectively obtained two molecular subtypes (Figure 1C). Subsequently, the prognostic characteristics for the two subtypes showed that the samples of the C1 subtype exhibited a longer survival time (Figure 1D). A similar trend was observed in GSE30219 (Figure 1E). Moreover, differential analysis of the pentose phosphate scores between the two molecular subtypes yielded lower scores in the C1 subtype relative to the C2 subtype in both TCGA and GSE30219 datasets (Figures 1E,G).

Differential Pathways Between Two Subtypes

The TCGA dataset's distributions of key clinical features for the two molecular subtypes were examined to see if there were any changes. The TCGA dataset revealed no significant variations in clinical features between the subtypes in ages, genders, and stages (Supplementary Figure 2). We used GSEA to look into the KEGG pathways connected with distinct molecular subtypes in biological processes (Figure 2). The results showed that metabolism-related pathways, including cytokine-cytokine receptor interaction, focal adhesion, and chemokine signaling pathway were significantly enriched in the C2 subtype showing a better prognosis, which indicated metabolic pathways were more activated in C2 subtype.

In addition, we performed differential expression analysis between two subtypes, and identified 266 DEGs ($|\log_2(\text{fold change})| > 1$ and adjust $P < 0.05$), where 126 and 140 DEGs were up-regulated in C1 and C2, respectively (Supplementary Figure 3A). GO analysis revealed that these DEGs were significantly enriched in metabolic terms such as quinone metabolic process, secondary metabolic process, and prostaglandin metabolic process (Supplementary Figure 3B).



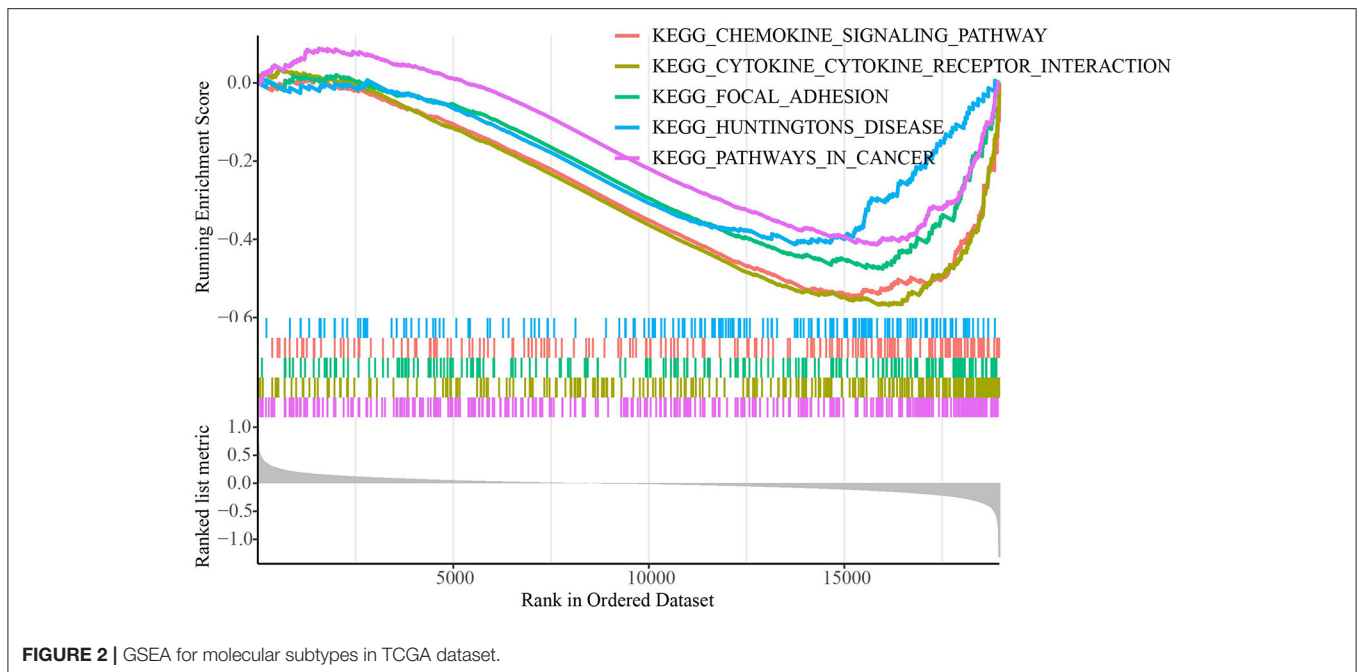
KEGG analysis showed that disease-related pathways such as autoimmune thyroid disease, type I diabetes mellitus, and graft-vs.-host disease were significantly enriched (**Supplementary Figure 3C**).

Mutational Characteristics of Pentose Phosphate-Related lncRNAs Subtypes

The differences in genomic alterations between these two molecular subtypes in TCGA were further examined. Correlation analysis indicated that pentose phosphate activity was not significantly associated with mutational characteristics, including the aneuploidy score, homologous recombination defects, number of segments, tumor mutation burden, and fraction altered (**Supplementary Figure 4A**). In addition, results of the mutation analysis suggested a significant correlation between subtypes and mutation in genes, including SPTA1, TNR, KEAP1, STK11, and FAT4, all of which had a significantly higher proportion of mutations in the C2 subtype relative to the C1 subtype (**Supplementary Figure 4B**).

Pathways Related to Pentose Phosphate-Related lncRNAs Subtypes

Next, we analyzed the differentially activated pathways between the two pentose phosphate-related lncRNAs subtypes. To identify these pathways, we performed a GSEA using all candidate gene sets from the KEGG tool. We defined false discovery rate (FDR) < 0.05 as the criterion for significant enrichment of the corresponding pathway. The C2-activated pathways included VEGF signaling pathway, ECM receptor interaction, intestinal immune network for IgA production, and Toll-like receptor signaling pathway (**Figure 3A**). In addition, we also compared the pathways that were consistently upregulated between C1 and C2 subtypes consisting of different LUAD patients (NES > 0 in C1 vs. C2, **Figures 3B,C**). GSEA for different subtypes showed that patients with the C2 subtype showed significantly upregulated glutathione metabolism, porphyrin and chlorophyll metabolism, steroid hormone biosynthesis, and arachidonic acid metabolism. Thus, the lncRNAs used for molecular typing may be significantly associated with energy metabolism.



Immune Characteristics of Pentose Phosphate-Related LncRNAs Subtypes

To further elucidate the differences in the immune microenvironment of patients belonging to two pentose phosphate-related lncRNAs subtypes, we determined the degree of infiltration of different immune cells in patients of TCGA dataset using the expressions of gene markers of immune cells from previously published literature (32). Based on the results of the differential analysis for the expressions of markers of immune cells, the two defined molecular subtypes were found to be significantly different for most immune cells as well as some pathways. Additionally, the expressions of most immune cells in the C1 subtype were markedly greater than those in the C2 subtype, including eosinophils, effector memory T, and T helper follicular cells (Figures 4A,C). Moreover, ESTIMATE was used to assess immune cell infiltration. The findings suggested that the “ImmuneScore,” “Stromal Score,” and “ESTIMATE Score” of the C1 subtype in TCGA were substantially greater than those of the C2 subtype, thereby indicating that, overall, the C1 subtype had a high degree of immune cell infiltration (Figures 4B,D).

Differences in the Responses to Immunotherapy Between Different Pentose Phosphate-Related LncRNAs Subtypes

Furthermore, differences in the responses to immunotherapy of different pentose phosphate-related lncRNAs subtypes were analyzed. First, we investigated whether there were any differences in the expressions of immune checkpoints between the two subtypes, wherein the immune checkpoints were derived from the HiscAtlas database (33). In the TCGA dataset, the expressions of the hub immune checkpoints, CTLA4, CD274, and CD47, were found to be markedly upregulated in the

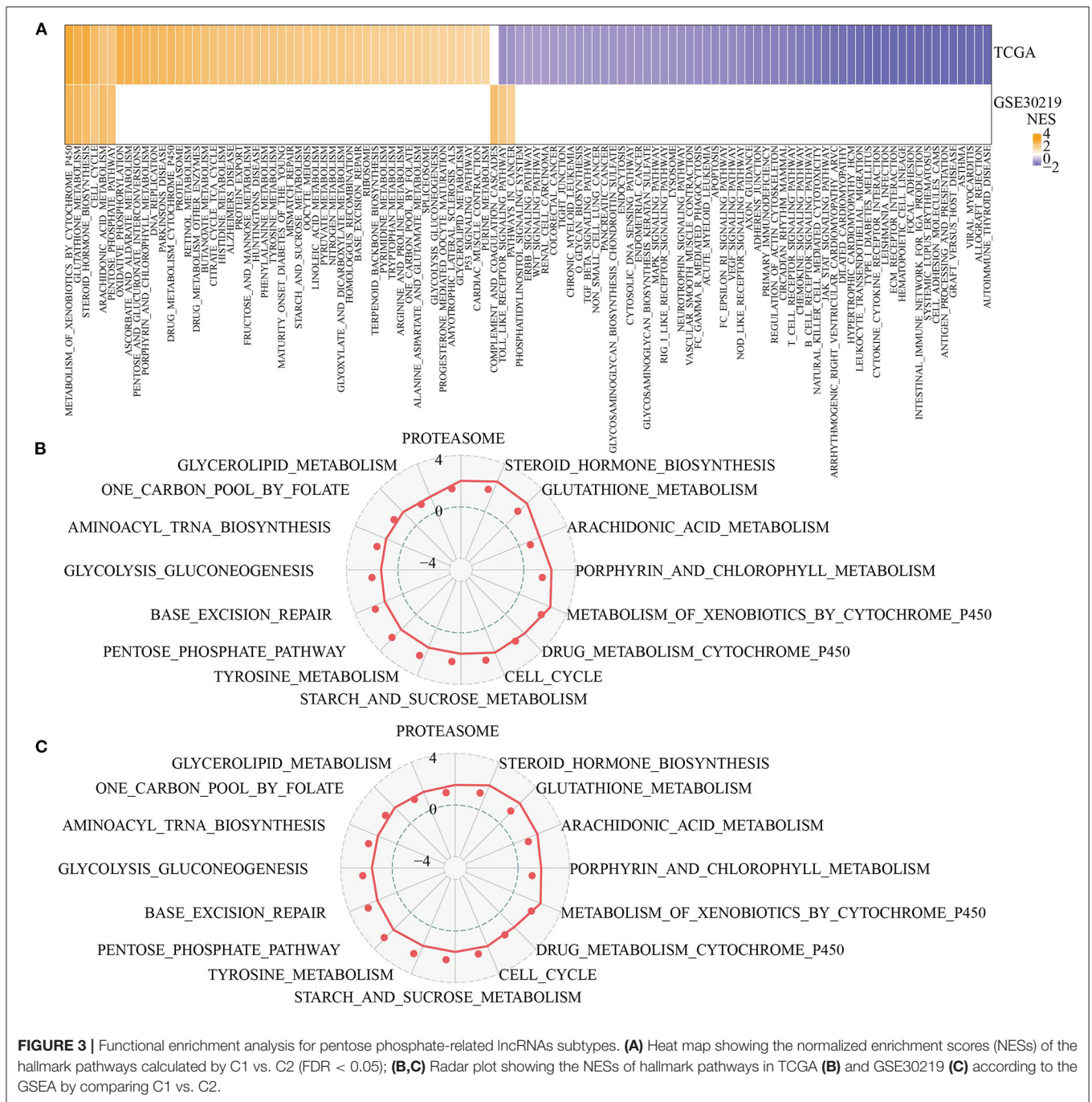
C1 subtype relative to those in the C2 subtype. In addition, expressions of CTLA4 and VISTA were found to be substantially high in the C1 subtype (Figure 5A); CD47 and CEACAM1 expressions were high in the C1 subtype in the GSE30219 dataset (Figure 5B). Moreover, using the TIDE (<http://tide.dfci.harvard.edu/>) software, the potential clinical effects of immunotherapy on the two molecular subtypes were assessed. The higher TIDE prediction score indicated a greater likelihood for immune escape, which suggested that patients were less likely to benefit from immunotherapy. We found a significantly higher dysfunction score for the C1 subtype as compared to the C2 subtype in TCGA dataset (Figure 5C). However, no significant differences were observed in the GSE30219 dataset (Figure 5D).

Characteristics of Pentose Phosphate-Related LncRNAs

The majority of phosphate-related lncRNAs showed a negative association with pentose phosphate activity relative to the protein-coding genes (Figure 6A).

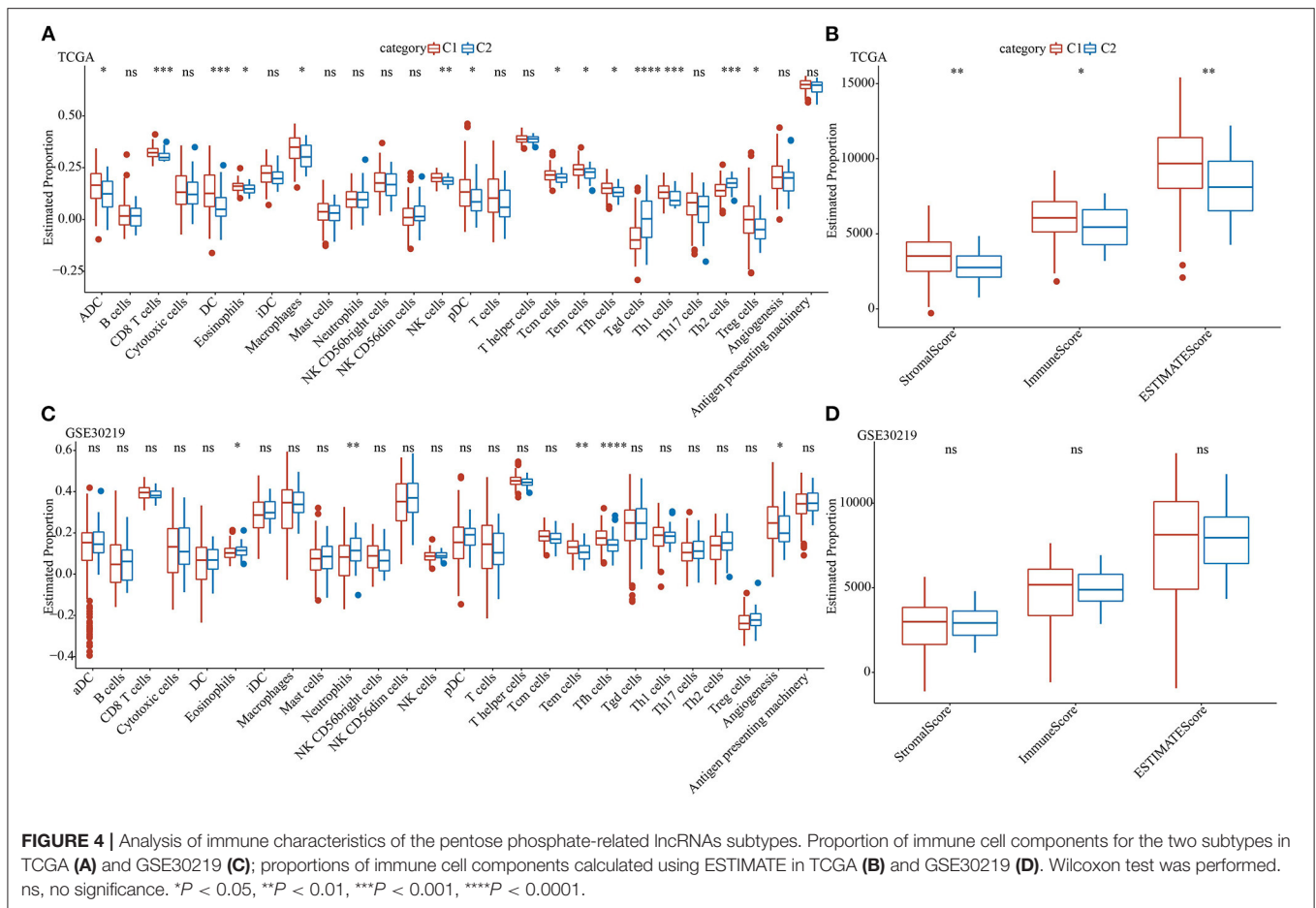
Next, using the LncAtlas database, we determined the cellular localization of these pentose phosphate-related lncRNAs. The majority of these lncRNAs were localized to the nucleus (Figure 6B); 59.22% of RCIs were negative ($RCI < 0$) in TCGA and 60.36% of RCIs were negative ($RCI < 0$) in GSE30219.

To evaluate the differences in TF activities between the two subtypes, the TF activity scores for each sample in TCGA and GSE30219 were calculated, according to the method described by Garcia-Alonso et al. (31). Further, an ANOVA test was performed to compare the TF activation levels between the two clusters in an effort to identify significantly differentially expressed TFs. A total of 91 and 51 significantly differentially expressed TFs in TCGA and GSE30219 datasets, respectively, were obtained.



Further, we investigated the association between pentose phosphate-related lncRNAs and the dysregulated TFs. Specifically, we performed a correlation analysis for the lncRNAs located in the nucleus with differentially expressed TFs; a series of TFs showing a significant negative correlation with pentose phosphate-related lncRNAs were identified (Figure 6C). In our previous study, we have reported a total of 33 pentose phosphate-related lncRNAs, of which 13 show a positive correlation with the pentose phosphate score and the remaining 20 show a negative correlation. Some of the critically significant lncRNAs

were selected based on their correlation with the activity of TFs (Figure 6D). Further, the distributions of activated and repressed TFs between the C1 subtype and C2 subtype were determined. A total of 11 repressed TFs and five activated TFs were found in LUAD-TCGA, whereas one activated TF was present in the GSE30219 dataset (Figure 6E). To examine whether the activity of TFs differed between the two molecular subtypes, one TF, BACH1, that was significantly upregulated in the C2 subtype (Figure 6F) was identified. We hypothesized that these TFs, which were consistently upregulated in the C2 subtype, may



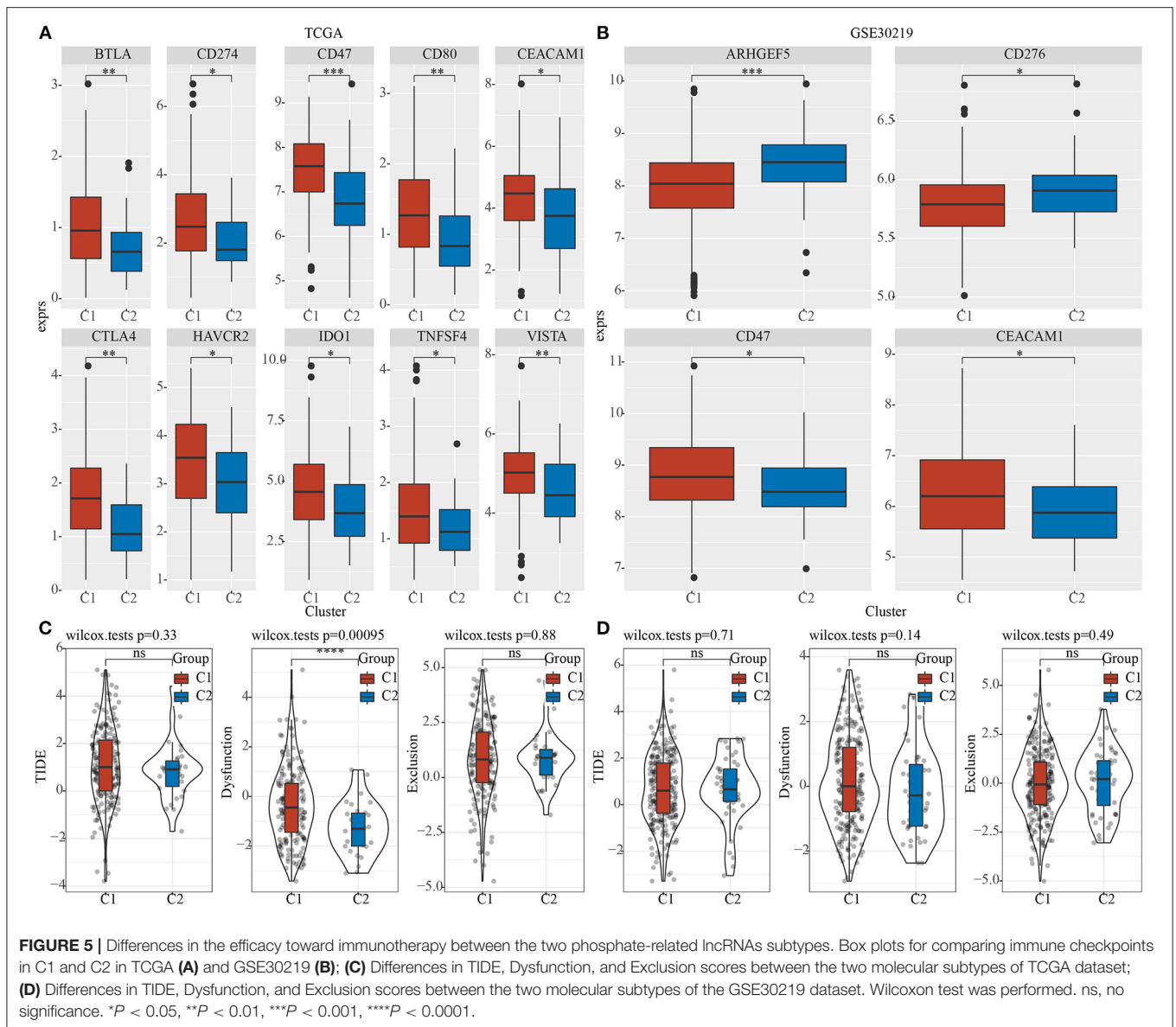
be associated with the poorer prognoses of the C2 subtype. To verify our hypothesis, a functional enrichment analysis for the TFs was performed (Figure 6G), which showed significant enrichment of genes targeted by TFs in transduction cascades involved in cellular senescence, atherosclerosis, fluid shear stress, HIF-1 signaling pathway, hepatocellular carcinoma, and ferroptosis. The activation of these pathways was related to tumor progression and poor prognosis, which suggested that pentose phosphate-related lncRNAs may exert important effects in the activation and progression of some cancer-related pathways.

DISCUSSION

Pentose phosphate has an important function in the metabolic reprogramming of tumor cells and is usually mediated through glucose-6-phosphate dehydrogenase (G6PD) (34). However, the role of pentose phosphate in LUAD remains unclear. Thus, in this study, two molecular subtypes of LUAD, C1, and C2, were obtained based on pentose phosphate-related lncRNAs by clustering analysis of samples from TCGA and GSE datasets. The C1 subtype showed a lower pentose phosphate score and was associated with good prognoses, whereas the C2 subtype yielded a higher pentose phosphate score and was related to poor prognoses. GSEA indicated that the C2 subtype was substantially

associated with pathways of energy metabolism. Unlike protein-coding genes, most pentose phosphate-related lncRNAs are negatively associated with pentose phosphate activity (13, 22). As a majority of them are localized to the nucleus, they are likely to suppress the expression of pentose phosphate by directly binding to the target genes (10, 11, 16, 35). For lncRNA-mediated regulation, we identified significantly differentially expressed TFs. These TFs may be involved in the activation of pathways related to cellular senescence, which is in turn related to tumor progression and poor prognoses in patients with the C2 subtype.

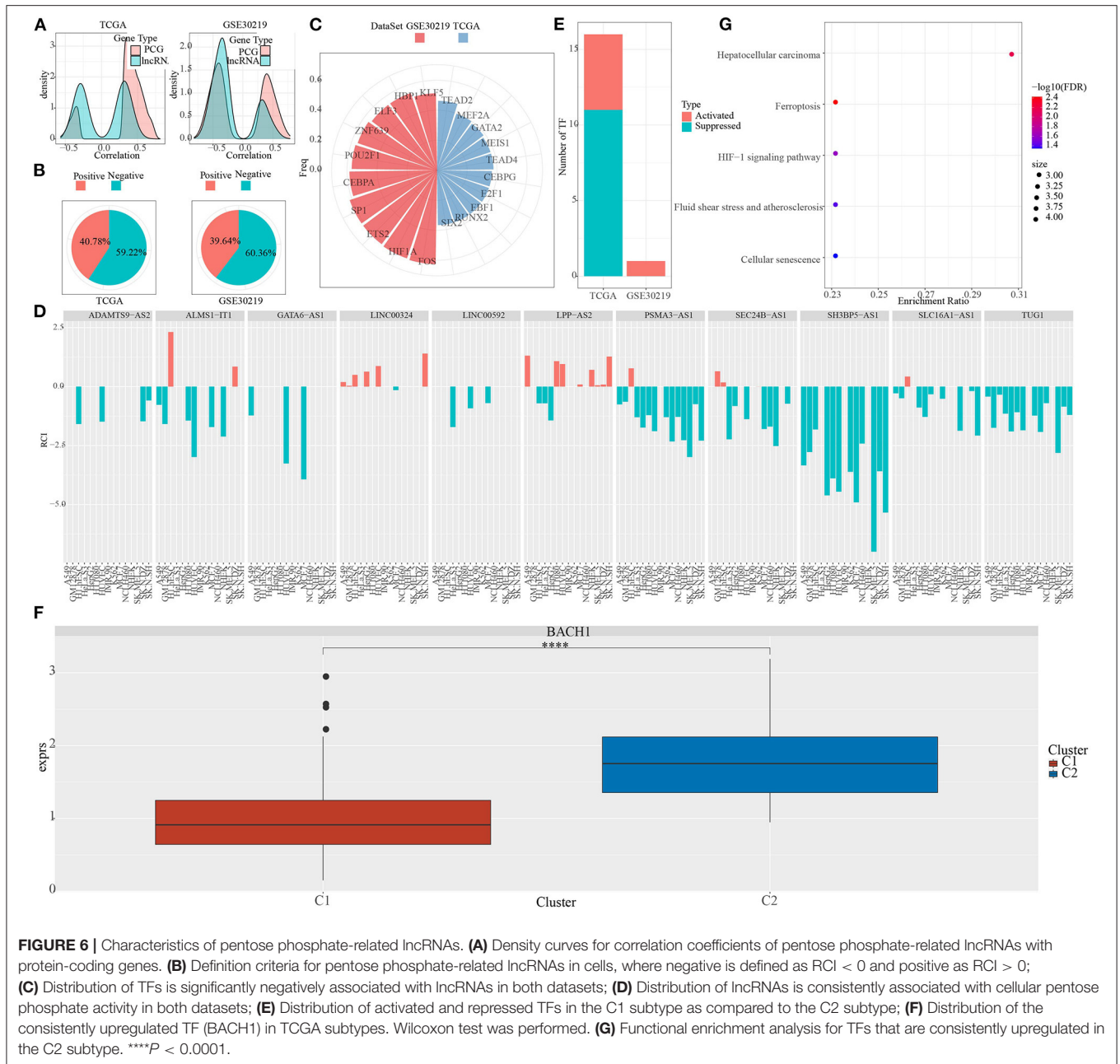
The characteristics of LUAD tumor cells are determined by the stromal and immune cells recruited to and activated in the tumor microenvironment (TME). Additionally, the TME in which tumor cells proliferate, develop, and metastasize, is also infiltrated by immune-related molecules and immune cells (36–38). Thus, we examined the differences in the TME between different pentose phosphate-related lncRNAs-based subtypes. A higher degree of immune cell infiltration for most immune cells was found in the C1 subtype relative to the C2 subtype, suggesting that the C1 subtype had a relatively high degree of immune cell infiltration, consistent with the findings of our previous study. The higher immune cell abundance and immune scores for C1 may be a reason for favorable prognoses among these patients. Moreover, some hub immune checkpoints, such as CTLA4,



CD274, and CD47, were found to be significantly upregulated in the C1 subtype as compared to the C2 subtype as evidenced by the differences in immune checkpoint expressions. Furthermore, the putative clinical effects of immunotherapy on both molecular subtypes were assessed. The findings suggested that the C1 subtype had a higher score and a greater likelihood for immune escape, which indicated that the patients belonging to the C2 subtype had poorer prognoses, along with a downregulated expression of the immune checkpoints. This further suggested that these patients may be better suited to and benefit more from immunotherapy as compared to those with the C1 subtype. TFs play key regulatory roles in several aspects of tumor progression in multiple cancers, including LUAD. For example, the expression of SOX12 is significantly high in LUAD tissues and is closely associated with the survival and prognoses of patients. It is suggested that SOX12 expression may be an auxiliary

predictor for LUAD progression (39). Thus, it is crucial to investigate the activity of TFs in LUAD. Through our analysis, we identified one TF, BACH1, that was consistently and significantly upregulated in the C1 subtype. Moreover, enrichment analysis indicated that BACH1 was involved in the activation of certain pathways and was tightly related to tumor progression and poor prognoses.

In this study, the pentose phosphate correlation-based typing could provide a new perspective for further studies on LUAD, particularly at the level of lncRNAs. A previous study (40) reports the two existing typing criteria for LUAD, namely TCGA subtype and Immune subtype. Comparisons with the existing molecular subtypes revealed no significant differences in the distributions among the subtypes. However, no significant differences in survival curves for these distributions were also observed. Thus, the subtypes proposed in this study were meaningful. We



have identified the possible regulatory involvement of pentose phosphate-related lncRNAs.

Nevertheless, further experimental evidence is needed to confirm the findings of this study. For instance, experimental validation of the differences in the expressions of pentose phosphate-related lncRNAs between the C1 and C2 subtypes is needed. Further investigations on the hub immune checkpoints identified herein and if they are significantly different at the transcript and protein levels between the two different subtypes are also warranted. The impact of BACH1 on tumor progression and prognoses needs to be confirmed experimentally. Moreover,

the possible mechanism of regulatory interplay also needs to be investigated in the future.

In conclusion, we typed LUAD into two subtypes based on pentose phosphate-related lncRNAs, as C1 and C2 subtypes. The C1 type showed a lower pentose phosphate score and was related to good prognoses, whereas C2 had a higher pentose phosphate score and was associated with poor prognoses. The potential clinical effects differed significantly between the two molecular subtypes, with higher scores associated with clinical effects and a higher likelihood of immune escape in the C1 subtype. In addition, the C1 subtype showed a relatively high

degree of immune cell infiltration. There was a significant correlation between subtypes and mutations, as also with genes, including SPTA1, TNFR, KEAP1, STK11, and FAT4, of which, the proportion of mutations were significantly higher in the C2 subtype as compared to the C1 type. Finally, the C2 subtype was found to be significantly associated with certain energy metabolic pathways.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

CL and YW designed the study and conducted literature review. CL collected the data, analyzed the data, and wrote the first

draft of the manuscript. YW interpreted the data and edited the manuscript. Both authors read and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.902445/full#supplementary-material>

Supplementary Table 1 | 114 metabolism-related pathways with gene lists.

Supplementary Figure 1 | The workflow of this study.

Supplementary Figure 2 | The distribution of different clinical features in C1 and C2 subtypes.

Supplementary Figure 3 | Differential analysis between C1 and C2 subtypes in TCGA dataset. **(A)** A heatmap of screened DEGs. **(B)** GO analysis of DEGs. **(C)** KEGG analysis of DEGs.

Supplementary Figure 4 | Genomic alterations in molecular subtypes in TCGA dataset. **(A)** Differences in Homologous Recombination Defects, Aneuploidy Score, Fraction Altered, Number of Segments, and Tumor Mutation Burden between the two molecular subtypes in TCGA dataset and correlation between them. $n-C1 = 186$, $n-C2 = 30$. **(B)** Somatic mutation analysis (Fisher test) of the top 16 mutated genes in both molecular subtypes.

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
- Shi J, Hua X, Zhu B, Ravichandran S, Wang M, Nguyen C, et al. Somatic genomics and clinical features of lung adenocarcinoma: a retrospective study. *PLoS Med.* (2016) 13:e1002162. doi: 10.1371/journal.pmed.1002162
- Zhou X, Hua D, Gao C, Zhang Y, Qiu L, Wang L. Icotinib and pemetrexed in treatment of lung adenocarcinoma and the effects on prognostic survival rate of patients. *Oncol Lett.* (2019) 18:4153–9. doi: 10.3892/ol.2019.10763
- Gong S, Qu X, Yang S, Zhou S, Li P, Zhang Q. RFC3 induces epithelial-mesenchymal transition in lung adenocarcinoma cells through the Wnt/ β -catenin pathway and possesses prognostic value in lung adenocarcinoma. *Int J Mol Med.* (2019) 44:2276–88. doi: 10.3892/ijmm.2019.4386
- Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep.* (2019) 9:3358. doi: 10.1038/s41598-019-40041-7
- Wu J, Savooji J, Liu D. Second- and third-generation ALK inhibitors for non-small cell lung cancer. *J Hematol Oncol.* (2016) 9:19. doi: 10.1186/s13045-016-0251-8
- Khandekar MJ, Piotrowska Z, Willers H, Sequist LV. Role of epidermal growth factor receptor (EGFR) inhibitors and radiation in the management of brain metastases from EGFR mutant lung cancers. *Oncologist.* (2018) 23:1054–62. doi: 10.1634/theoncologist.2017-0557
- Wu X, Gu Z, Chen Y, Chen B, Chen W, Weng L, et al. Application of PD-1 blockade in cancer immunotherapy. *Comput Struct Biotechnol J.* (2019) 17:661–74. doi: 10.1016/j.csbj.2019.03.006
- Chowdhury PS, Chamoto K, Honjo T. Combination therapy strategies for improving PD-1 blockade efficacy: a new era in cancer immunotherapy. *J Intern Med.* (2018) 283:110–20. doi: 10.1111/joim.12708
- Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol.* (2015) 22:5–7. doi: 10.1038/nsmb.2942
- Si J, Ma Y, Lv C, Hong Y, Tan H, Yang Y. HIF1A-AS2 induces osimertinib resistance in lung adenocarcinoma patients by regulating the miR-146b-5p/IL-6/STAT3 axis. *Mol Ther Nucl Acids.* (2021) 26:613–24. doi: 10.1016/j.omtn.2021.09.003
- Ren P, Hong X, Chang L, Xing L, Zhang H. USF1-induced overexpression of long noncoding RNA WDFY3-AS2 promotes lung adenocarcinoma progression via targeting miR-491-5p/ZNF703 axis. *Mol Carcinog.* (2020) 59:875–85. doi: 10.1002/mc.23181
- Song J, Sun Y, Cao H, Liu Z, Xi L, Dong C, et al. A novel pyroptosis-related lncRNA signature for prognostic prediction in patients with lung adenocarcinoma. *Bioengineered.* (2021) 12:5932–49. doi: 10.1080/21655979.2021.1972078
- Koziol MJ, Rinn JL. RNA traffic control of chromatin complexes. *Curr Opin Genet Dev.* (2010) 20:142–8. doi: 10.1016/j.gde.2010.03.003
- Kondo Y, Shinjo K, Katsushima K. Long non-coding RNAs as an epigenetic regulator in human cancers. *Cancer Sci.* (2017) 108:1927–33. doi: 10.1111/cas.13342
- Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F, Ochiya T. Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* (2018) 109:2093–100. doi: 10.1111/cas.13642
- Lin YH. Crosstalk of lncRNA and cellular metabolism and their regulatory mechanism in cancer. *Int J Mol Sci.* (2020) 21:82947. doi: 10.3390/ijms21082947
- Gong W, Yang L, Wang Y, Xian J, Qiu F, Liu L, et al. Analysis of survival-related lncRNA landscape identifies a role for LINC01537 in energy metabolism and lung cancer progression. *Int J Mol Sci.* (2019) 20:153713. doi: 10.3390/ijms20153713
- Mai S, Liang L, Mai G, Liu X, Diao D, Cai R, et al. Development and validation of lactate metabolism-related lncRNA signature as a prognostic model for lung adenocarcinoma. *Front Endocrinol.* (2022) 13:829175. doi: 10.3389/fendo.2022.829175
- Yu X, Zhang X, Zhang Y. Identification of a 5-gene metabolic signature for predicting prognosis based on an integrated analysis of tumor microenvironment in lung adenocarcinoma. *J Oncol.* (2020) 2020:5310793. doi: 10.1155/2020/5310793
- Patra KC, Hay N. The pentose phosphate pathway and cancer. *Trends Biochem Sci.* (2014) 39:347–54. doi: 10.1016/j.tibs.2014.06.005
- Huang HT, Xu YM, Ding SG, Yu XQ, Wang F, Wang HF, et al. The novel lncRNA PTTG3P is downregulated and predicts poor prognosis in non-small cell lung cancer. *Arch Med Sci.* (2020) 16:931–40. doi: 10.5114/aoms.2020.93535
- Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol.* (2015) 19:A68–77. doi: 10.5114/wo.2014.47136
- Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic activation of germline and placental genes

- identifies aggressive metastasis-prone lung cancers. *Sci Transl Med.* (2013) 5:186ra66. doi: 10.1126/scitranslmed.3005723
25. Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front Immunol.* (2021) 12:687975. doi: 10.3389/fimmu.2021.687975
 26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
 27. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170
 28. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
 29. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
 30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
 31. Garcia-Alonso L, Iorio F, Matchan A, Fonseca N, Jaaks P, Peat G, et al. Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.* (2018) 78:769–80. doi: 10.1158/0008-5472.CAN-17-1679
 32. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* (2016) 17:231. doi: 10.1186/s13059-016-1092-z
 33. Liu Y, He M, Wang D, Diao L, Liu J, Tang L, et al. HisgAtlas 1.0: a human immunosuppression gene database. *Database.* (2017) 2017:bax094. doi: 10.1093/database/bax094
 34. Cossu V, Bonanomi M, Bauckneht M, Ravera S, Righi N, Miceli A, et al. Two high-rate pentose-phosphate pathways in cancer cells. *Sci Rep.* (2020) 10:22111. doi: 10.1038/s41598-020-79185-2
 35. Lin Z, Huang W, Yi Y, Li D, Xie Z, Li Z, et al. LncRNA ADAMTS9-AS2 is a prognostic biomarker and correlated with immune infiltrates in lung adenocarcinoma. *Int J Gen Med.* (2021) 14:8541–55. doi: 10.2147/IJGM.S340683
 36. Jin X, Zheng Y, Chen Z, Wang F, Bi G, Li M, et al. Integrated analysis of patients with KEAP1/NFE2L2/CUL3 mutations in lung adenocarcinomas. *Cancer Med.* (2021) 10:8673–92. doi: 10.1002/cam4.4338
 37. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell.* (2017) 169:750–65.e17. doi: 10.1016/j.cell.2017.04.014
 38. Li Y, Shen R, Wang A, Zhao J, Zhou J, Zhang W, et al. Construction of a prognostic immune-related LncRNA risk model for lung adenocarcinoma. *Front Cell Dev Biol.* (2021) 9:648806. doi: 10.3389/fcell.2021.648806
 39. Li L, Zhang T, Chen Y, Song J, Meng Y, Liu S, et al. Expression of transcription factor SOX12 in lung adenocarcinoma and its clinical significance. *Nan Fang Yi Ke Da Xue Xue Bao.* (2019) 39:186–91. doi: 10.12122/j.issn.1673-4254.2019.09.10
 40. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity.* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.