Jacobs *et al. Journal of Translational Medicine* (2026) 24:392
https://doi.org/10.1186/s12967-026-07917-5

Journal of Translational Medicine

RESEARCH Open Access

Unraveling lung cancer dynamics: a new metabolic signature improving the prediction of recurrence in resected lung adenocarcinoma

Francesca Jacobs[1,2], Lorenzo Manganaro[3], Lorenzo D'Ambrosio[1,4], Davide Corà[5,6], Martina Olivero[7], Francesca Napoli[1], Marco De Filippis[1], Valeria Cetoretta[1], Edoardo Garbo[1], Teresa Mele[1], Maddalena Arigoni[8], Eugenia R. Zanella[9], Sushant Parab[9,10], H. M. Waqas Munir[7], Francesca Picca[7], Riccardo Taulli[7,11], Francesca Bersani[7,11], Alessandra Merlini[4,7], Raffaele Calogero[8], Livio Trusolino[9,10], Luca Primo[9,10], Luisella Righi[1], Marco Volante[1], Francesco Leo[1,12], Enrico Ruffini[13,14], Mauro Papotti[15,16], Federico Bussolino[9,10], Silvia Novello[1,4,7], Giorgio V. Scagliotti[1], Paolo Bironzo[1,4†] and Gabriella Doronzo[7*†]

Abstract

Background Lung cancer is characterized by wide genetic, molecular, and phenotypic alterations that may challenge diagnosis and clinical decision-making. This heterogeneity often leads to variable responses to therapies, resulting in suboptimal outcomes for many patients. Recent advancements in *omics* technologies have enabled a deeper exploration of mechanisms driving tumor behavior and identification of specific molecular signatures. Tumor metabolic reprogramming, one of the hallmarks of cancer development, progression, and recurrence, represents a promising field of research.

Methods In this study, we developed a comprehensive metabolic signature using RNA-sequencing data from independent cohorts of patients diagnosed with stage I-III resectable lung adenocarcinoma (LUAD) to enhance patient stratification and prognostic accuracy.

Results We identified a novel prognostic signature "LMetSig" consisting of 10 metabolic genes that significantly stratified LUAD patients into high- and low-risk subgroups for disease-free survival (DFS). Cox regression analysis demonstrated that LMetSig is an independent prognostic biomarker for DFS. Among the LMetSig, TK1 gene emerged as a promising LUAD-specific biomarker. It was undetectable in normal tissue, showed variable expression in tumor samples and correlated with shorter DFS when expressed at high levels.

†Paolo Bironzo and Gabriella Doronzo contributed equally to this work.

*Correspondence:
Gabriella Doronzo
gabriella.doronzo@unito.it

Full list of author information is available at the end of the article

Conclusion Our findings suggest that LMetSig can significantly improve LUAD patients' stratification alongside conventional pathological and clinical parameters. By distinguishing high-risk patients from those with more favorable prognosis, this approach has the potential for informing personalized treatment strategies and improving clinical decision-making.

Keywords Lung adenocarcinoma (LUAD), Non-Small Cell Lung Cancer (NSCLC), Metabolism, Prognostic signature

Background

Lung cancer remains a leading cause of cancer-related death worldwide, with tobacco smoking being the main risk factor. The high mortality rate is mainly due to locally advanced or metastatic stage at diagnosis and high relapse rate following therapies with radical or palliative intent [1]. Therefore, efforts to improve prevention, early detection, and treatment strategies remain essential in addressing this major health challenge [2].

Non-Small Cell Lung Cancer (NSCLC) is the most common histological type of lung cancer and includes several distinct subtypes, the most common being adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC). These histotypes are characterized by different molecular and clinical features, even when classified into the same pathological stage [3–5]. The current approach for patients with upfront radically resected early stage LUAD includes the use of adjuvant systemic therapies based on pathological stage and several predictive biomarkers, including molecular alterations such as Epidermal Growth Factor Receptor (EGFR) activating mutations and Anaplastic Lymphoma Kinase (ALK) gene rearrangements, as well as Programmed Death Ligand protein 1 (PD-L1) protein expression [6]. Nonetheless, adjuvant treatment strategies are still far from being really tailored to the individual patient, with many patients still being over- or under-treated, experiencing relapse even after adjuvant therapies. Therefore, there is an urgent need for novel predictive and prognostic biomarkers that can be integrated with those already available [7, 8].

Metabolic reprogramming is necessary to support the relentless proliferation of tumor cells, their migration and growth in metastatic sites [9, 10]. Moreover, it influences the tumor microenvironment, immune milieu, extracellular matrix deposition/remodeling and prepares a permissive soil for colonization of metastatic cells or drug resistance [11]. High expression of key metabolic enzymes is coupled with the upregulation of genetic and epigenetic factors, oncogenes, and proliferative pathways [12].

Several investigators have developed and proposed different *metabolic signatures* for LUAD, detecting a significant correlation with patients' prognosis in some series [13–20]. Although these studies were encouraging, to date no metabolic biomarker for LUAD has been introduced in clinical practice, even if tumor metabolism is tightly connected with therapeutic interventions in lung cancer [13–20]. The only cancer metabolic drugs approved for LUAD treatment are several cytotoxic agents classified as "antimetabolites", including those targeting pyrimidine synthesis [21].

Considering the above background, this study was designed to identify a novel *metabolic gene signature* using clinical and transcriptomic data together with bioinformatics tools. The potential predictive role of the signature was then evaluated.

Methods

Study design and population

This is an observational, retrospective, multicenter study and male and female patients were considered. Enrolment was conducted at San Luigi Gonzaga and Città della Salute e della Scienza University Hospitals in Orbassano and Turin (Italy), respectively. DEFLeCT patients were part of the prospective observational clinical trial PROMOLE approved by the Ethical Committee of the San Luigi Gonzaga University Hospital (protocol n.14057 approved on September 28, 2018, n.73/2018, version-3 January 30, 2023). MAGA patients were included in the PROFILING protocol (No. 001-IRCC-00IIS-10) of Candiolo Cancer Institute FPO IRCCS, Candiolo, Turin Italy. The informed consent was collected for each patient at the time of the surgical resection. Demographic characteristics, smoking habit, clinical and radiological data were collected, together with data on pathologic staging (according to the VIII edition of the AJCC/UICC TNM staging system) and molecular analysis.

Some analyses were performed on LUAD integrated cohort (LUAD-IC), a dataset reported by Gyorffy et al. [22]. Briefly, the dataset gathered together 2852 lung tumour specimens from 17 independent patient cohorts obtained from NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) and the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). The cohorts included both LUAD (55%) and LUSC (40%) tumors. Most tumors were predominantly stage I (60%), with a significant proportion of stage II samples (32%), while stage III samples were fewer (8%) [22]. For more detailed and comprehensive information, the original study provides a full description of the dataset and its characteristics [22]. To perform our investigation, we restricted the analysis on LUAD histology, which consists of 1308 LUAD samples.

Transcriptomic analysis and exome sequencing

As part of DEFLeCT project, we reanalyzed mRNA-sequencing of the DEFLeCT cohort collected as previously described [23]. RNA samples and mRNA-sequencing data of MAGA cohort were instead both collected and analyzed ex novo. To ensure data consistency and uniformity across all results, we performed the analysis on the new samples using the same methodology as previously applied [23]. Briefly, tissues were isolated at the time of surgical resection and RNA extraction was performed by Maxwell-RSC simply RNA tissue kits (Promega Corporation). RNA samples were purified from DNA by treatment with RNase-Free DNase Set (QIAGEN). The quality of samples was verified by 2100 Bioanalyzer (Agilent Technologies) and Qubit assays (Thermo-Fisher Scientific).

Libraries for RNA sequencing were generated using TruSeq Stranded mRNA Library Prep (Illumina Inc.) following manufacturer's instructions, using 1 µg of total RNA as input material and 15 PCR cycles for DNA amplification. Each library was analyzed with the DNA High Sensitivity chip using Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified by Qubit 2.0 Fluorometer using ds DNA High Sensitivity Qubit Assay (ThermoFisher Scientific). Libraries were pooled together in equimolar amounts and run at the concentration of 1000 pM on the NextSeq1000 sequencer (Illumina Inc.) in 75 nts paired end sequencing mode following manufacturer instruction. Only genes with a value of Transcripts Per Million (TPM) > 2 in at least 80% of patients were considered for the analysis. TPM normalisation has been computed by Python script calculating the reads per kilobase (RPK) ratio for each gene and then scaling the results to the sum of the RPK per sample divided by 10^6. The sum of TPM for each sample, by construction, resulted equal to 10^6". As described by Gyorffy et al. [22] that performed the analysis, LUAD-IC samples were examined using the in-situ oligonucleotide array platforms GPL96 (Affymetrix Human Genome U133A Array), GPL3921 (Affymetrix HT Human Genome U133A Array) and GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array). Data obtained from these analyses were subjected to two rounds of normalization [22]. For further details, please refer to Gyorffy et al. [22].

Exome sequencing was carried out using the SureSelectXT Human All Exon Kit, following the manufacturer's guidelines (Agilent Technologies). The sequencing was conducted on the Illumina NovaSeq 6000 (Illumina Inc.) platform, with 35 million and 70 million paired-end reads for normal and tumor samples, respectively. To identify tumor-specific mutations, each tumor sample was compared to the corresponding normal DNA from the normal tissue. After the initial quality check, NGS sequencing of exomes was processed through a bioinformatics pipeline to assess the mutational landscape of LUAD patients. For each sample, TMB was determined by counting the total number of single nucleotide variants (SNVs) per megabase within the sequenced genomic region. In this study, we employed the IDEA pipeline workflow, which enables the identification of SNVs, insertions or deletions (INDELs), and gene copy-number alterations (CNAs). The results of somatic variant calling were compared with those from the established Genome Analysis Toolkit (GATK) Mutect2 and found to be consistent. SNV and INDEL annotations were based on the Catalogue of Somatic Mutations In Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic).

Metabolic genes screening and analysis

For metabolic gene enrichment analysis two databases were used: "Metabolomics Workbench tool" of Enrichr software and the Molecular signatures database (MSigDB) of the gene set enrichment analysis (GSEA) software [24–26]. The Gene Ontology biological processes (GO:BP) with a cut-off of $p < 0.05$ were considered in our analysis.

Methodology for sample stratification

To define the cutpoints for the different metabolic subsets, we applied an unsupervised hierarchical clustering approach using Morpheus software (https://software.broadinstitute.org/morpheus) [27], allowing the data to naturally segregate according to the expression levels of the genes composing the signature. This unbiased clustering procedure identified three stable and biologically meaningful groups that reflected low, intermediate, and high expression patterns. The resulting dendrogram and cluster structure provided clear separation points, which were then used to assign samples to the low-, medium-, and high-expression subsets. This data-driven strategy ensured that the cutpoints were determined objectively based on intrinsic transcriptional variability rather than arbitrary thresholds, thereby preserving the biological relevance of the metabolic subgroups.

The transcriptomic data were also analysed and visualized as Principal Component Analysis (PCA) plots via the web tool ClustVis [28].

To identify DEGs between the subset high and low of patients, we considered a log2 TPM ratio ≥ 1 and performed a 2-sample t test with a p value threshold of 0.05.

Survival analysis

For all cohorts considered, according to the definitions adopted in the AJCC Cancer Staging Manual (Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al., editors. AJCC Cancer Staging System, Version 9. Chicago, IL: American College of Surgeons; 2025) DFS represents the interval from curative-intent

treatment to the first occurrence of disease recurrence (either local or distant).There was no hierarchy among the events, the first that occurred (either local or distant) was considered as an event for the Kaplan-Meier estimation. Comparisons were made with log-rank test and hazard ratio (HR) estimates calculated by means of Cox regression. Multivariate analysis was performed using the Cox proportional hazards model including in the multivariate analysis only covariates with p-value ≤0.1 at the univariate analysis. Whenever indicated, Tests were two-sided and results were reported with 95% confidence intervals (95%CI) or interquartile ranges (IQR).

To perform our investigations on the LUAD-IC cohort, we used the Kaplan Meier plotter platform (https://www.kmplot.com) [22], restricting the analysis on LUAD histology. Depending on the selected clinical parameters, the software analysed only the samples for which the requested clinical information was available, excluding cases with missing data. Therefore, the reported data for the different analyses referred to the subset of samples with available information. The software indicated this number as"Using the selected parameters, the analysis will run on n patients".

GraphPad*Prism* 8.0.1 and IBM SPSS v29.0 were used to identify the correlation between the expression of metabolic genes and the clinical outcomes of DEFLeCT and MAGA patients.

Immunohistochemistry (IHC)

All cases were analysed using IHC ($n = 11$ low and $n = 30$ high representative samples for LUAD metabolic "subtype"). Briefly, 5 µm-thick serial paraffin sections from representative paraffin blocks were processed using an automated platform (Ventana BenchMark Ultra, Roche) with a primary rabbit anti-TK1 (JF0970) antibody (Thermo-Fisher Scientific). TK1 tumor positivity was measured by two pathologists (L.R. and M.V.) using a semiquantitative histological score (H-score) and calculated as previously described [29].

Flow chart

To provide a clearer and more comprehensive illustration of the study design, analytical workflow and data integration steps, we have included a detailed flow chart summarising the entire methodology (see Fig. 1). Created in BioRender. https://BioRender.com/h37y798

Statistical analysis

The statistical analyses were performed using IBM SPSS v29.0 and GraphPad*Prism 8.0.1 software*. Descriptive

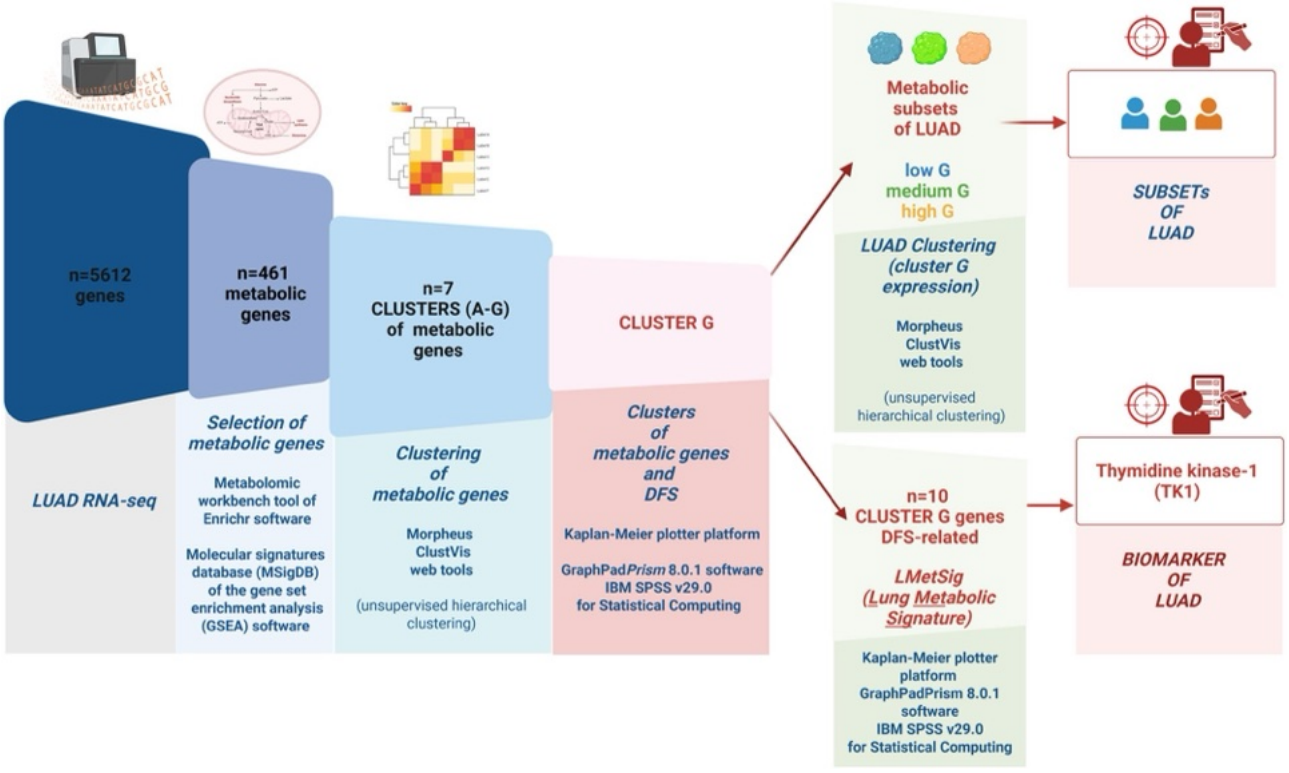


Fig. 1 Flow chart. The flow chart provides a detailed overview of the key points of the study, from the initial set of genes derived from RNA-sequencing analysis to the identification of metabolic clusters, with particular emphasis on Cluster G and the derivation of the LMetSig signature associated with poor prognosis. For each step, we have also specified the tools or software employed in the analysis, together with any relevant analytical indications, in order to enhance clarity and reproducibility

statistics were used for patients' characteristics. Qualitative variables were compared using the χ2 and Fisher's exact tests. Appropriate statistical tests were performed as indicated in the Results section, and $p < 0.05$ was considered to indicate statistical significance in all analysis.

Results

Study population

In the present study we included 87 patients from a larger cohort of early-stage resectable LUAD patients who were enrolled in a prospective observational clinical study, called PROMOLE, within the research project DEFLeCT (Digital Technology For Lung Cancer Treatment) [23, 30]. All patients were ≥18 years of age and had pathological stage I-III LUAD (according to the VIII edition of the AJCC/UICC TNM staging system) [31]. No patients received neo-adjuvant therapies. A summary of patient characteristics is provided in Fig. 2, panel A.

Thirty-one patients (35.6%) were females and the median age at diagnosis was 66.7 years (range 37–81). All patients underwent radical surgery: 78 (89.7%) received lobectomy, 3 (3.5%) bi-lobectomy, and 3 (3.5%) pneumonectomy, while for 3 (3.5%) patients details on the type of surgical procedure were missing. Twenty-two patients (25.3%) were active smokers, 47 (54%) former smokers, and 18 (20.7%) never smokers. Pathologic stage was I in 47 patients (54.1%), II in 20 patients (22.9%), and III in 17 patients (19.5%). Data were missing for 3 (3.5%) patients who were excluded from survival analysis. Maximum standardized uptake (SUVmax) values at the [(18)F] fluorodeoxyglucose positron emission tomography/computed tomography (PET/CT) scan performed at the time of diagnosis were also collected, with a median SUVmax of 7.7 (range 2.1–16.1). Seventeen patients (19.5%) received adjuvant chemotherapy; of these, 4 (23.5%) received cisplatin plus vinorelbine, 5 (29.4%) cisplatin plus gemcitabine, while the remaining 8 patients (47.1%) received different adjuvant treatments within clinical trials. A total of 7 patients (8.1%) received postoperative thoracic radiotherapy.

Identification of metabolism related gene clusters in LUAD

We performed a bulk mRNA-seq analysis of the 87 samples to identify metabolic genes that characterized LUAD cohort. After gene expression quantification and normalization to exclude genes with irrelevant expression and background noise, a total of 5612 genes were considered for the further analysis. To identify metabolic genes from this gene set we interrogated the specific "Metabolomics Workbench tool" of Enrichr software [24, 25] and the Gene Set Enrichment Analysis (GSEA) database [26]. We found 461 genes involved in the metabolism of lipids, carbohydrates, amino acid, nucleotides as highlighted by the analysis of the 100 most significantly enriched Gene Ontology categories related to "Biological Process" (GO:BP) (FDR-adjusted $p < 0.05$) (Supplementary figure 1, Supplementary data 1).

We performed a prompt investigation for the expression of the 461 metabolic genes in the 87 tumor samples. By using Morpheus clustering web application [27], genes were categorized in seven different metabolic clusters (A-G) through an unsupervised hierarchical clustering (Supplementary Figure 2). A subsequent functional analysis of these 7 clusters using the "Metabolomics Workbench" metadata [25] showed no overlap of metabolic function among the clusters (Fig. 2 panel B).

The most significant GO:BP for each cluster were Amino Sugar Catabolic Process (GO:0046348) for Cluster A, Branched-Chain Amino Acid catabolic Process (GO:0009083) for Cluster B, Steroid Metabolic Process (GO:0008202) for Cluster C, Purine-Containing Compound Biosynthetic Process (GO:0072522) for Cluster D, Phosphatidylinositol Biosynthetic Process (GO:0006661) for Cluster E, Glycerophospholipid Biosynthetic Process (GO:0046474) for Cluster F and Glycolytic Process (GO:0006096) for Cluster G (Fig. 2 panel B, Supplementary data 5).

Identification of a putative metabolic cluster of genes for LUAD stratification

We explored the potential association of the genes within the A-G clusters for Disease-Free Survival (DFS). Considering the mean expression of selected genes, we documented that only the high expression levels of "Cluster G" of genes was significantly associated with worse DFS of DEFLeCT LUAD patients (log-rank *p-value* = 0.04) (Fig. 3 panel A). We also analyzed a larger LUAD cohort, the "Integrated cohort" (LUAD-IC), encompassing 1308 tumor specimens derived from 12 independent LUAD cohorts as previously described [22]. In LUAD-IC cohort the high expression level of "Cluster G" of genes was significantly associated with worse DFS (log-rank *p-value* < 0.0001) while the high expression levels of "Cluster B" (log-rank *p-value* < 0.0001) or "Cluster F" of genes (log-rank *p-value* = 0.01) were associated with an improved DFS (Fig. 3 panel B).

"Cluster G" of genes identifies different metabolic subsets of LUAD

By means of an unsupervised hierarchical clustering using Morpheus software [27], we categorized the 87 DEFLeCT LUAD tumor samples according to expression levels of "Cluster G" of genes into 3 different subsets: low- (subset low, $n = 15$ patients), medium- (subset medium, $n = 36$ patients), and high-expression (subset high, $n = 36$ patients), as shown in Fig. 4 by the Heat map (panel A) and Principal Component Analysis (PCA) plot (panel B).

A)

Clinical Data of DEFLeCT cohort	No. (%)
No. of patients	87
Histology	Adenocarcinoma
Gender	
Male	56 (64.4)
Female	31 (35.6)
Surgery	
Lobectomy	78 (89.7)
Bilobectomy	3 (3.5)
Pneumectomy	3 (3.5)
N/A	3 (3.5)
Age of diagnosis average [range]	66.7 [37-81]
Smoking habit	
Current	22 (25.3)
Former	47 (54)
Never	18 (20.7)
Stage	
Stage I	47 (54.1)
Stage II	20 (22.9)
Stage III	17 (19.5)
N/A	3 (3.5)
Pathology grading	
Grade I	12 (13.9)
Grade II	43 (50)
Grade III	31 (36)
N/A	1 (1.2)
PET imaging of tumor SUVmax average [range]	7.7 [2.1-16.1]

B)

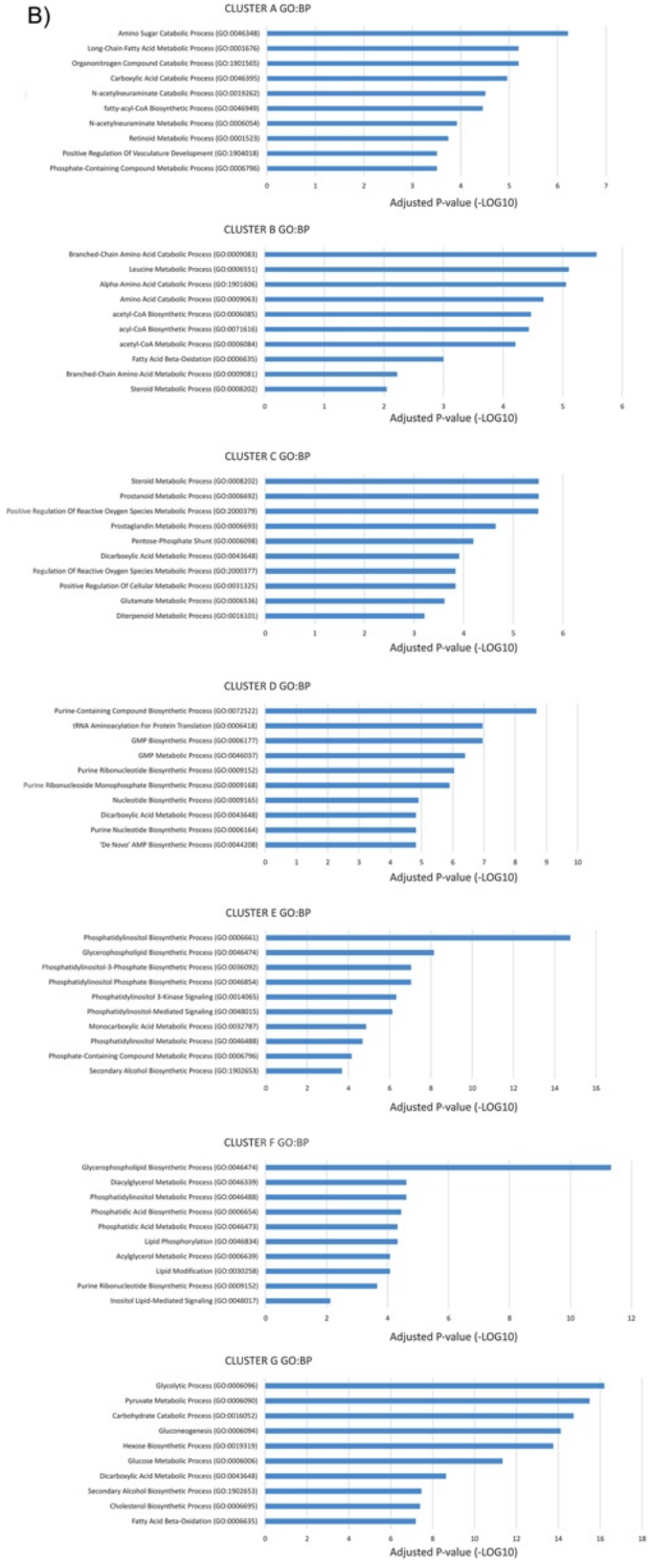


Fig. 2 Overview of DEFLeCT LUAD cohort and functional analysis of metabolic gene clusters. (**A**) Clinical characteristics of the patient population, reported as number of patients with percent or median with interquartile range. PET, positron emission tomography; SUV, standardized uptake value; (**B**) "Metabolomics Workbench" metadata analysis and relative top 10 enriched "Biological Process" (GO:BP) characterizing metabolic clusters of DEFLeCT LUAD cohort

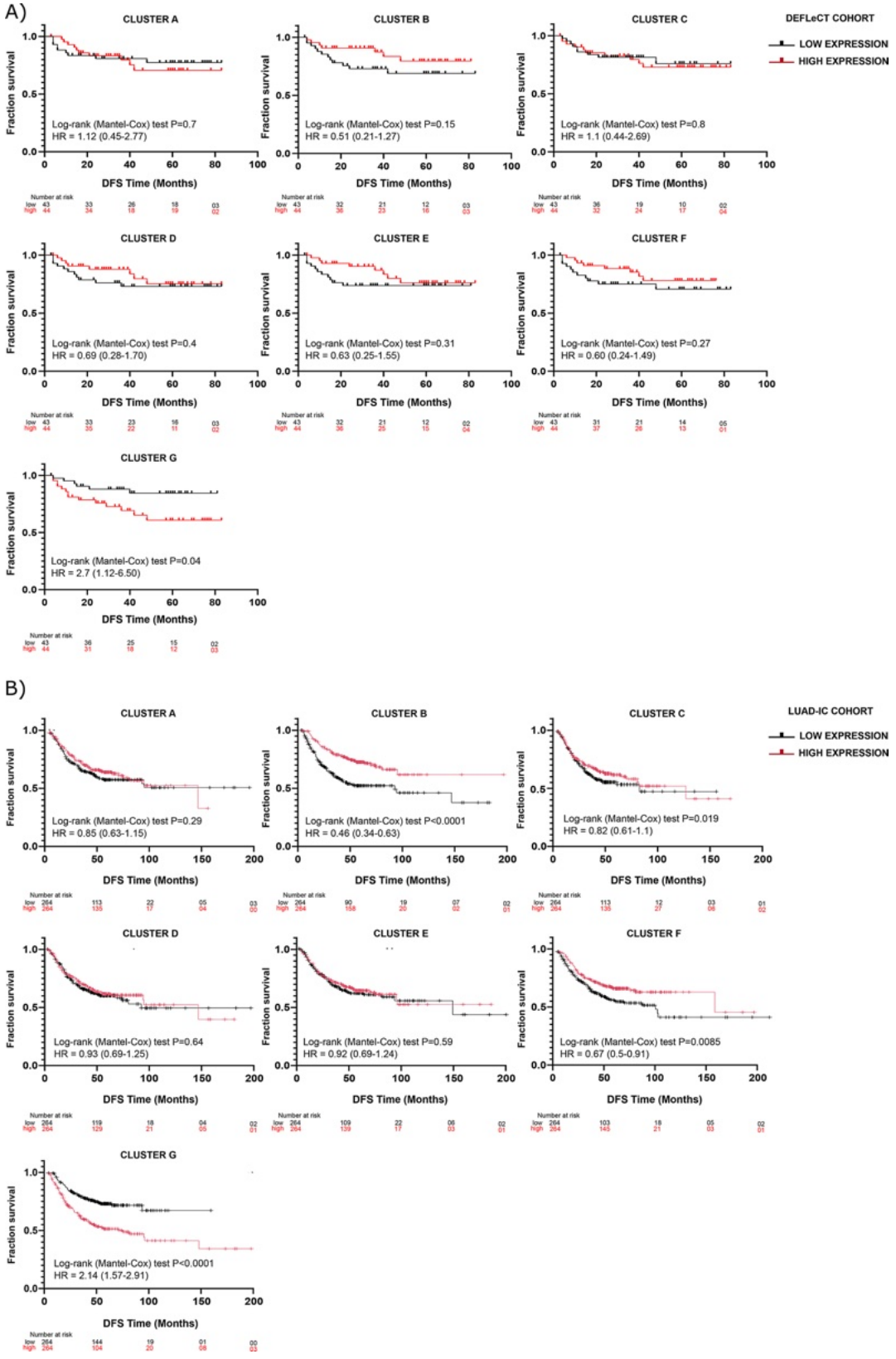


Fig. 3 Metabolic signature and LUAD DFS. Kaplan-Meier curves representing DFS of DEFLeCT (panel A) and LUAD-IC (panel B) patients related to the expression level of the genes of the distinct metabolic clusters identified

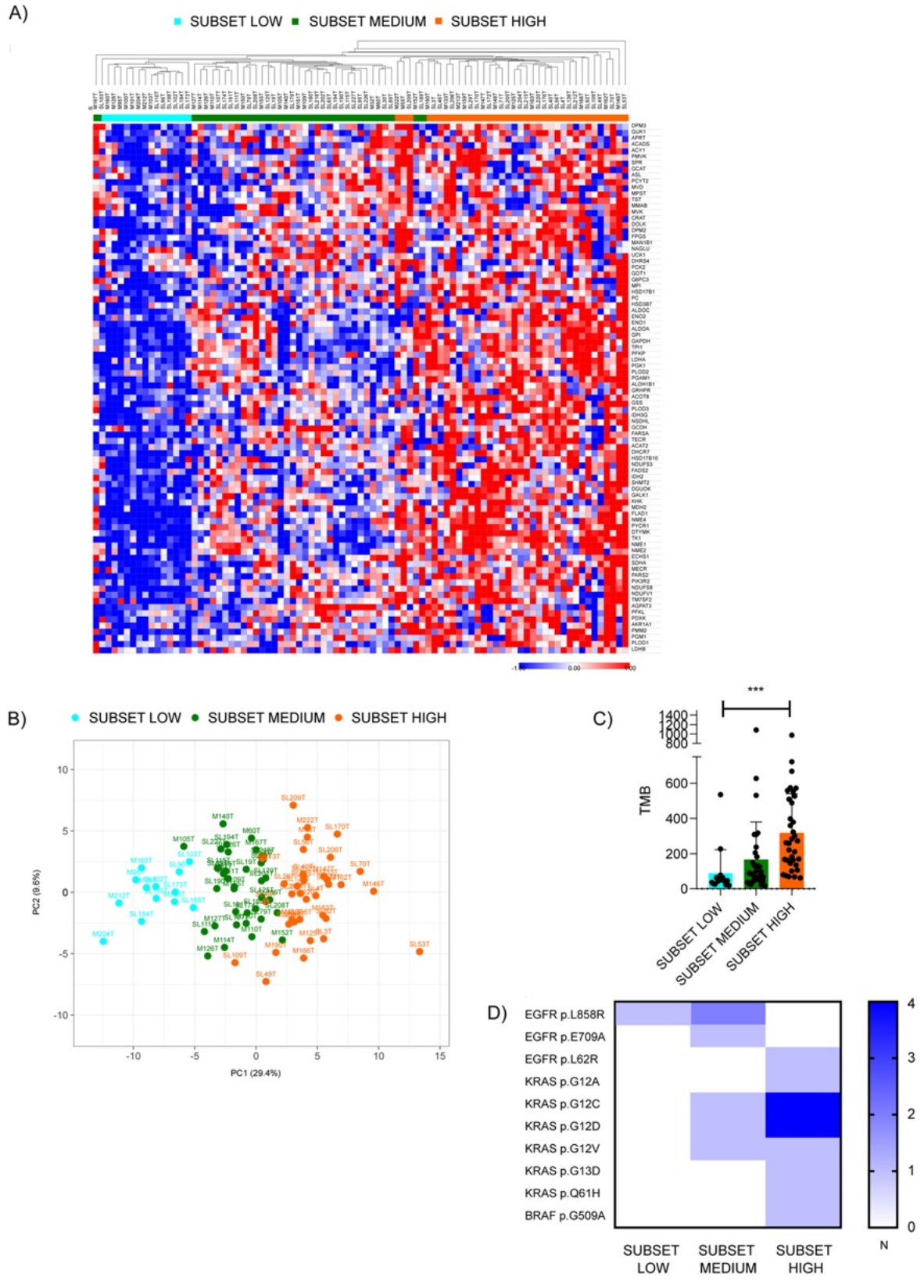


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Metabolic Cluster G and DEFLeCT LUAD cohort DFS. (**A**) Heatmap showing unsupervised hierarchical clustering of DEFLeCT LUAD samples (column) based on the expression of Cluster G genes (n=87, row). Red: up-regulated genes; blue: down-regulated genes. DEFLeCT LUAD subsets are flagged via color annotation: subset low (sky-blue), subset medium (green), subset high (orange); (**B**) Principal component Analysis (PCA) of DEFLeCT LUAD samples performed using as variable the mean expression values of Cluster G genes. Each dot represents a sample and each color represents a subset (subset low=sky-blue, subset medium=green, subset high=orange); (**C**) Bar graph about TMB distribution in the different metabolic subsets; (**D**) Heatmap showing the distribution as number of the mutations of the oncogenes in the different metabolic subsets

We then deepened into the clinical parameters of patients in each metabolic subset (Supplementary data 6). Although the "subset high" was characterized by a higher percentage of smokers, stage-III or high-grade tumors, the differences were not statistically significant among the three metabolic subsets for smoking habits (*p-value* = 0.283), gender (*p-value* = 0.554), or TNM stage (stage I vs. stages II-III, *p-value* = 0.227).

The SUVmax values were categorized into tertiles and we observed that the majority of patients in the metabolic subset-high showed SUVmax values in the higher tertile (SUVmax > 8) (Supplementary data 6). Nonetheless, SUVmax values were not significantly associated with the three metabolic subsets, regardless of whether they were categorized into tertiles (SUVmax < 4, 4–8, >8) or using the median value of 7.7 to differentiate the population into "high" and "low" SUVmax groups (*p-values* of 0.779 and 0.698, respectively). Furthermore, baseline SUVmax values did not show a significant correlation with the metabolic subsets, even when considering high and low categories only, yielding two-sided *p*-values of 0.423 and 0.541 for the SUVmax tertiles and median thresholds, respectively.

Mutational profile of LUAD

For 84 out of 87 tumors (96.5%) included in DEFLeCT LUAD cohort, the mutational status was available. We analyzed the distribution of Tumor Mutation Burden (TMB) in the different metabolic subsets and we evidenced that the high expression of "Cluster G" of genes was positively associated with TMB (*p-value* < 0.0001) (Fig. 4 panel C). Focusing on the distribution of the most important pathological mutated oncogenes characterizing LUAD (i.e. EGFR, KRAS, BRAF, MET) we observed an significant enrichment of KRAS mutant tumors in metabolic subset-high (% of KRAS mutant tumors vs all tumors in subset medium 5%, high 14%, *p-value* = 0.022) (Fig. 4 panel D).

Correlation between metabolic subsets and DFS in LUAD patients

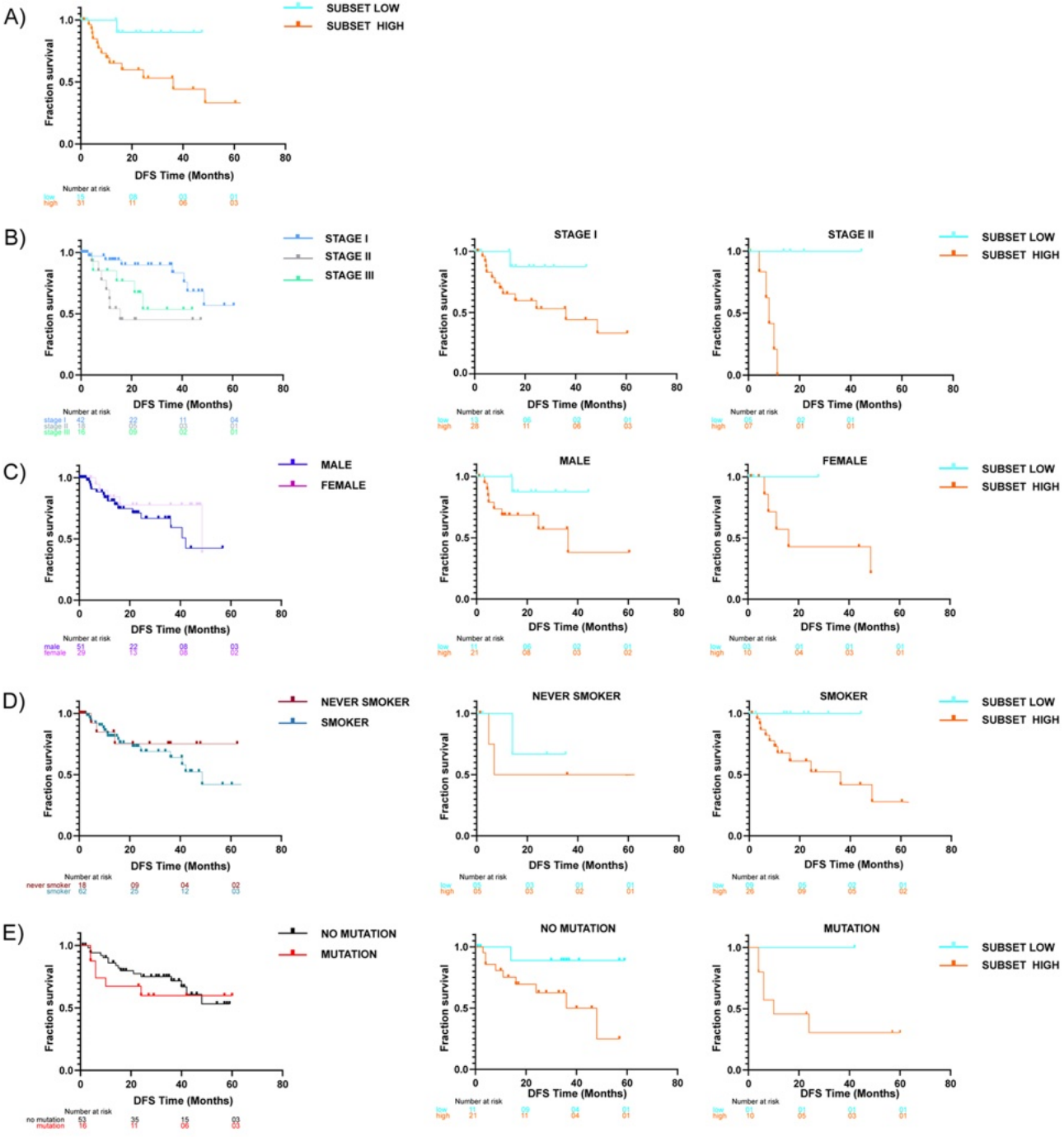
We investigated the putative correlation between the metabolic subsets and the DFS of DEFLeCT LUAD cohort. After exclusion of 7 patients due to missing relevant data or inadequate follow-up, the survival analysis was performed on the remaining 80 patients. Overall, with a median follow-up of 56.5 months (95% CI: 56.6 -62.3), the median DFS for the entire cohort was 49 months (95% CI: not estimable), with an estimated DFS at 3 years of 70.6%. Among these 80 patients, DFS showed a statistically significant difference according to metabolic subset (*p-value* = 0.025) (Fig. 5 panel A).

Namely, patients in "subset low" exhibited a significantly better DFS compared to those in "subset high", with a median DFS that was not reached (NR) versus 32 months (95% CI: 5–67, log-rank test *p-value* = 0.040) with a HR of 0.156 (95% CI: 0.020–1.201, *p-value* = 0.074) (Fig. 5 panel A).

As expected, TNM staging demonstrated a significant association with DFS (*p-value* = 0.017). It is noteworthy that, although with the limitations due to the small sample size, the metabolic signature had a greater impact in stages I-II (Fig. 5 panel B). In contrast, the correlation between metabolic signature and outcomes in stage III patients was not statistically significant (*p-value* = 0.799). Specifically, in the early stages (stage I-II), none of the patients in "subset low" experienced relapse, resulting in a significantly better DFS compared to patients in "subset high" (median DFS NR vs 36.1 months; *p-value* = 0.022) (Fig. 5 panel B). The correlation with metabolic subset was particularly significant in stage II patients, where all but two patients with metabolic "subset high" experienced an early relapse, while none of the patients in metabolic "subset low" relapsed. The median DFS for "subset high" was only 8 months vs NR for "subset low" (*p-value* = 0.002) (Fig. 5 panel B). Patients with grade 3 tumors exhibited worse DFS compared to those with grade 1–2 tumors (*p-value* = 0.005). Notably, a higher tumor grade was not associated with "subset high" (*p-value* = 0.224).

In line with these findings across the cohort, gender did not correlate with DFS or metabolic subsets (*p-value* = 0.293). However, metabolic subsets were significantly associated with DFS in both males and females (*p-value* = 0.049) (Fig. 5 panel C).

Smoking history was not associated with DFS (*p-value* = 0.471) (Fig. 5 panel D). Nevertheless, patients with a history of smoking had a higher likelihood of belonging to "subset high", although this difference did not reach statistical significance (OR 3.100, 95% CI: 0.742–12.953, *p-value* = 0.135). Notably, high metabolic subset was significantly associated with DFS also in patients with a smoking history (*p-value* = 0.035) (Fig.



F)

Variable	HR (95% CI)	P value	HR (95% CI)	P value
Staging (1 vs ≥2)	0.286 (0.105-0773)	0.014	0.155 (0.027-0.882)	0.036
Grading (≤2 vs 3)	0.270 (0.103-0.709)	0.008	0.546 (0.116-2.579)	0.445
Metabolic subset (low vs high)	0.156 (0.020-1.201)	0.074	0.081 (0.008-0.816)	0.033

Fig. 5 DFS analysis of DEFLeCT LUAD cohort. Kaplan-Meier curves representing DFS of DEFLeCT patients considering metabolic subset and all cohort (**A**) or after a stratification of the patients according to stage, gender, smoking habit or the oncogene mutational profile (**B-E**); (**F**) Table summarizes the results of multivariate analysis relative to the correlation between DFS and pathological staging, grading, and metabolic subset

5 panel D). Finally, SUVmax values were not associated with DFS (*p-value* = 0.956).

We investigated the putative correlation between the oncogene mutational profile and DFS of DEFLeCT cohort (Fig. 5 panel E). The comparison between non-mutated and mutated tumors indicated a not significant difference in terms of DFS (*p-value* = 0.97). Probably due to the small number of data, the significant difference in DFS between "subset low" and "subset high" is lost by separating patients into non-mutated and mutated ones (*p-value* = 0.056 and 0.29 respectively). It is noteworthy that DFS in non-mutated and mutated patients within both the low subset and the high subset was not statistically significant (*p-value* = 0.73 and 0.08 respectively).

In the multivariate analysis, we included variables that had a *p*-value < 0.1 at the univariate analysis: pathological staging, grading, and metabolic subset. Pathological staging and metabolic subset maintained a statistically significant association with DFS (*p-value* = 0.036 and 0.033 respectively), whereas grading did not maintain a significant correlation with the survival outcome (*p-value* = 0.44)(Fig. 5 panel F).

Identification of Lung Metabolic Signature "LMetSig" related to poor prognosis in LUAD

Based on exploratory analyses in DEFLeCT LUAD patients, we compared the gene expression level between "subset high" vs "subset low" (5612 genes) and we identified 429 differentially expressed genes (DEGs) (Supplementary figure 3, Supplementary data 7). The intersection of DEGs with "Cluster G" of genes revealed a list of 23 differentially expressed genes (gDEGs) up-regulated in tumors of "subset high" versus "subset low" (Fig. 6 panel A and B, Supplementary data 7). The high expression of 10 out of the 23 gDEGs was significantly associated with a reduced DFS (*p-value* = 0.002) (Fig. 6 panel C). This set of 10 genes, that was named "Lung Metabolic Signature" (LMetSig) is composed by Aldolase-A (ALDOA), 7-Dehydrocholesterol Reductase (DHCR7), Deoxythymidylate Kinase (DTYMK), Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH), NAD (P) Dependent Steroid Dehydrogenase-Like Protein (NSDHL), 6-Phosphofructokinase Type C-F (PFKP), Phosphoglycerate Kinase 1 (PGK1), Glucose Phosphate Isomerase (GPI), Triosephosphate Isomerase (TPI1) and Thymidine Kinase (TK1) (Fig. 6 panel C).

We analyzed the expression level of LMetSig genes in tumor and normal tissues of the same patient, if available. We did not detect any significant difference between the expression of LMetSig genes in the normal tissues of the "subset high" and "subset low" (Fig. 7) while, as expected, the transcription of these genes increased significantly in tumor tissues compared to normal tissues in both subsets (except for DHCR7 and NSDHL in "subset low") (Fig. 7). Notably, this increase was higher in tumors of patients belonging to "subset high" when compared to "subset low" (Fig. 7).

The predictive ability of LMetSig was verified in the LUAD-IC cohort (*n* = 528 patients): considering the mean expression of LMetSig genes patients with low expression of LMetSig showed a significantly better DFS compared to high expressors (median DFS of 20.07 vs 50.56 months, log-rank *p-value* < 0.0001)(Supplementary figure 4 panel A). Analyzing LUAD-IC cohort according to TNM stage (*n* = 274 patients), LMetSig showed a significant impact on DFS in stage-I patients (log-rank *p-value* = 0.0017) but not in stage-II patients (*n* = 98 patients)(log-rank *p-value* = 0.76)(Supplementary figure 4 panel B). Sample sizes of other stages were too low for meaningful analysis (*n* = 8 patients for stage-III and *n* = 0 for stage-IV).

LMetSig expression level was significantly associated with DFS in both sexes (*n* = 240 female patients, log-rank *p-value* = 0.004; *n* = 288 male patients, log-rank *p-value* = 0.0021) (Supplementary figure 4 panel B). The high metabolic subset was associated with a worse DFS for both never-smoker and smoker (former and current) patients although not significantly for the latter ones (*n* = 140 never-smoker, log-rank *p-value* = 0.024); *n* = 229 smoker, log-rank *p-value* = 0.135) (Supplementary figure 4 panel B).

We performed a multivariate analysis, including stage and LMetSig as covariates showing that both variables maintained a statistically significant correlation with DFS also in the LUAD-IC cohort (*n* = 380 patients, stage log-rank *p-value* < 0.0001, LMetSig log-rank *p-value* = 0.014) (Supplementary figure 4 panel C).

Exploration of LMetSig expression in a new early-stage radically resected LUAD cohort

LMetSig expression distribution was investigated in a different independent prospective cohort of early-stage radically resected LUAD (MAGA cohort *n* = 48 patients, Supplementary data 6). A summary of patient characteristics data is provided in Figure 8 panel A.

After mRNA-seq analysis and unsupervised clustering according on LMetSig expression level, we profiled tumors in "subset low", "subset medium", and "subset high" (Fig. 8 panels B and C).

As previously defined for the DEFLeCT LUAD cohort, the MAGA cohort did not exhibit a statistically significant association between metabolic subsets and smoking habit (*p-value* = 0.409), gender (*p-value* = 0.305), or TNM stage (stage I vs. stages II-III, *p-value* = 0.173). A borderline significant association was observed with tumor grading (*p-value* = 0.061). Indeed, in this cohort, high-grade tumors accounted for 64, 53, and 21% of the samples classified as metabolic subsets high, intermediate, and low, respectively (Supplementary data 6).

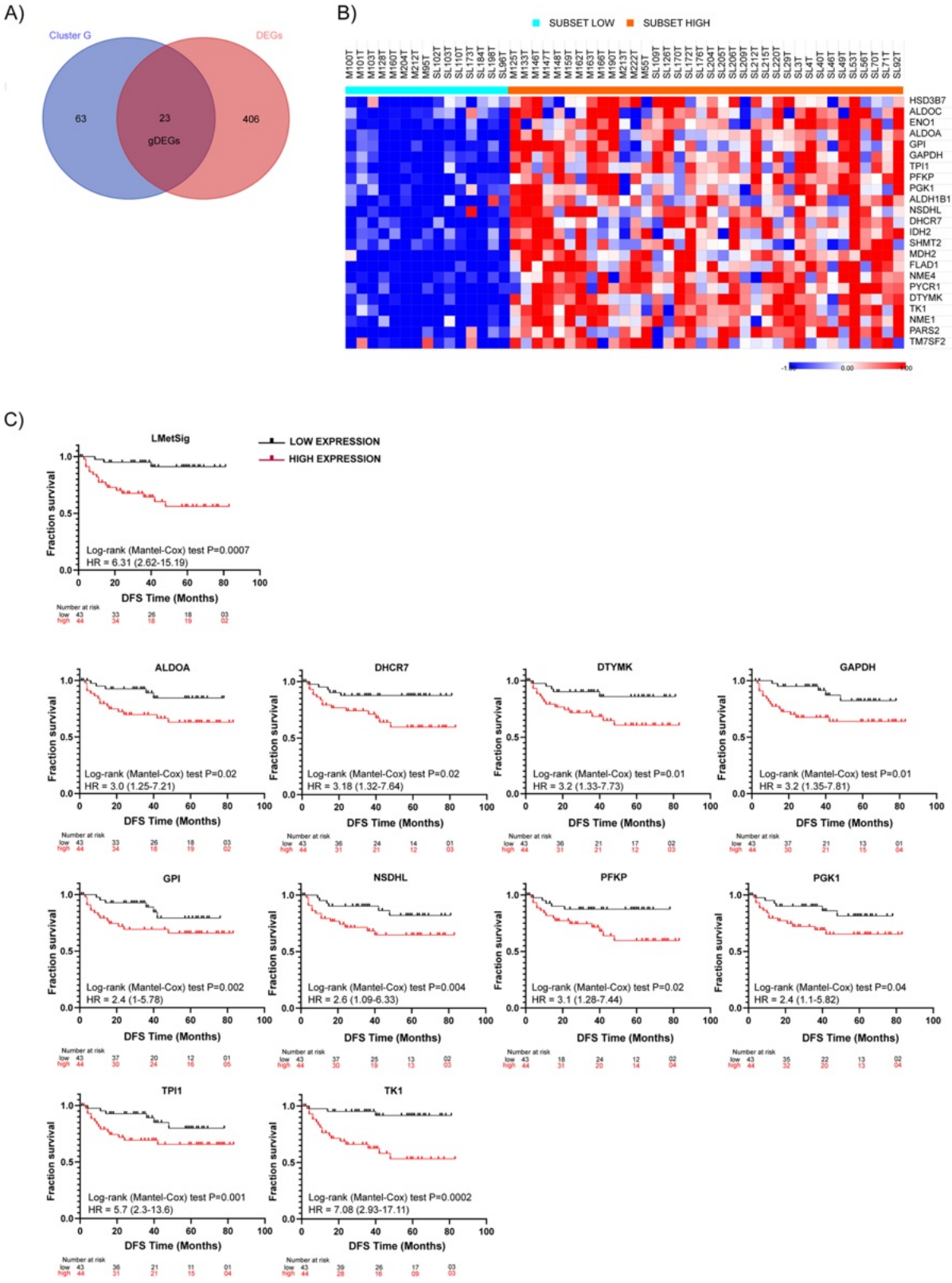


Fig. 6 Identification of LMetSig. (**A**) Venn diagram showing the number of gDEGs shared between genes of Cluster G and differential expressed genes (DEGs) characterizing subset low and high of DEFLeCT LUAD samples; (**B**) Heatmap showing the expression of gDEGs in DEFLeCT LUAD samples. Red: up-regulated genes; blue: down-regulated genes. Sample subsets are flagged via color annotation: low (sky-blue) and high (orange); (**C**) Kaplan-Meier curves representing DFS of DEFLeCT LUAD patients considering the mean expression of LMetSig genes or the punctual expression of its components ALDOA, DHCR7, DTYMK, GAPDH, NSDHL, PFKP, PGK1, GPI, TPI1 and TK1

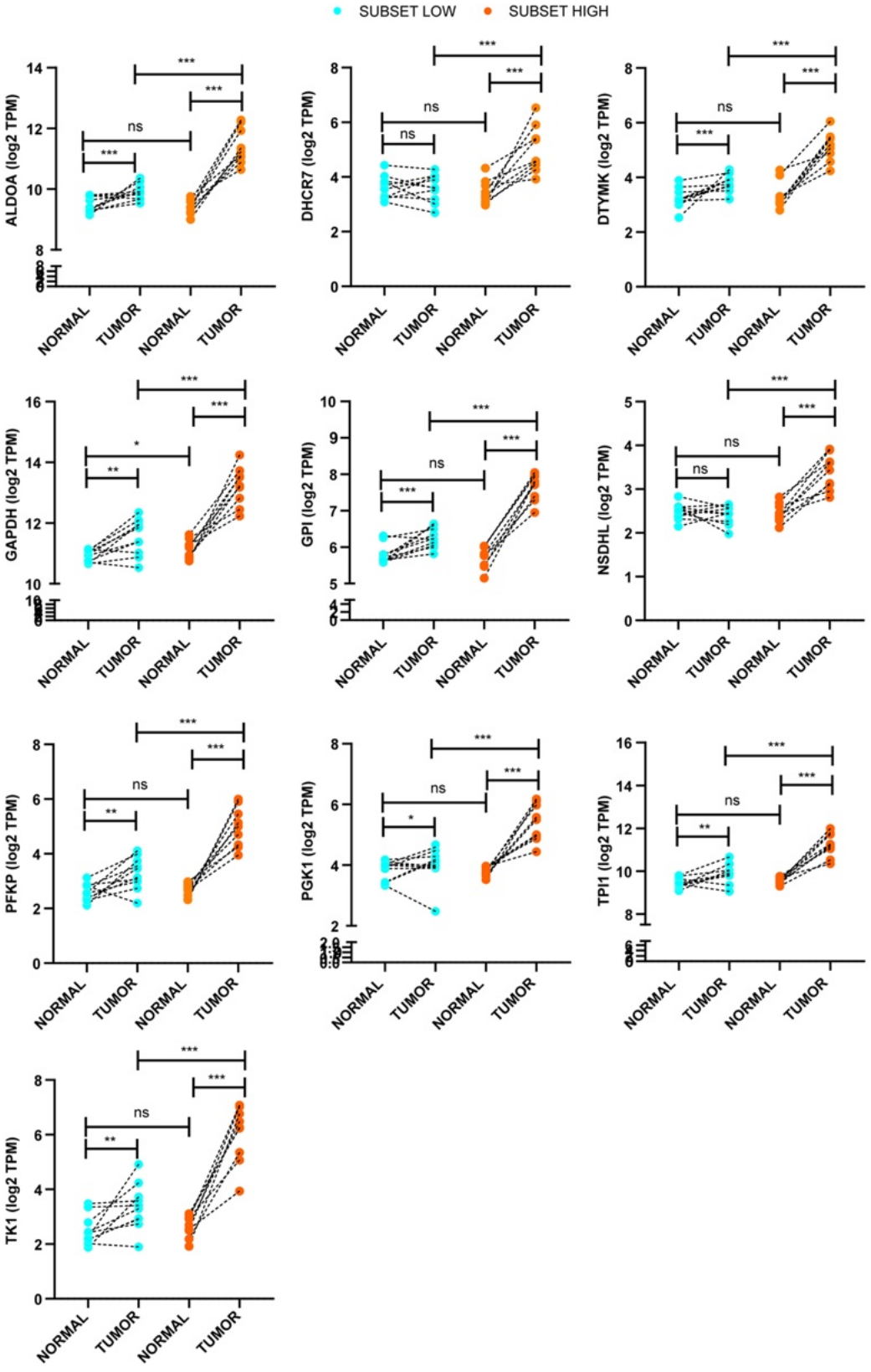
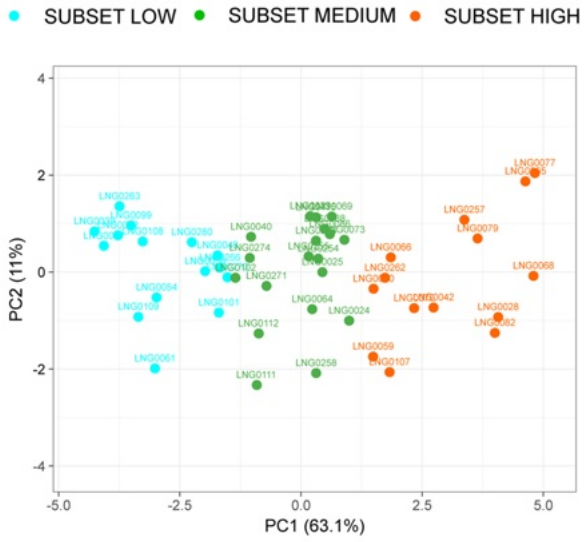


Fig. 7 LMetSig expression in LUAD and normal lung tissues. Comparison of the LMetSig mRNA expression level (log2TPM) in paired adjacent normal tissues and LUAD tissues from DEFIeCT cohort considering the metabolic subsets. Sample subsets are flagged via color annotation: low (sky-blue) and high (orange). (n=10 normal lung tissue; n=9 LUAD samples; ***p<0.0001, **p<0.001, *p<0.01 in all comparisons by Student's t-test)

A)

Clinical Data of MAGA cohort	No. (%)
No. of patients	48
Histology	Adenocarcinoma
Gender	
Male	26 (54.2)
Female	22 (45.8)
Surgery	
Lobectomy	42 (87.5)
Bilobectomy	2 (4.2)
Pneumectomy	3 (6.3)
Atipical Surgery	1 (2)
Age of diagnosis average [range]	70 [35-82]
Smoking habit	
Current	9 (18.8)
Former	22 (45.8)
Never	11 (22.9)
N/A	6 (12.5)
Stage	
Stage I	19 (39.5)
Stage II	14 (29.2)
Stage III	15 (31.3)
Pathology grading	
Grade I	5 (10.5)
Grade II	20 (42.5)
Grade III	22 (46.7)
N/A	1 (2)



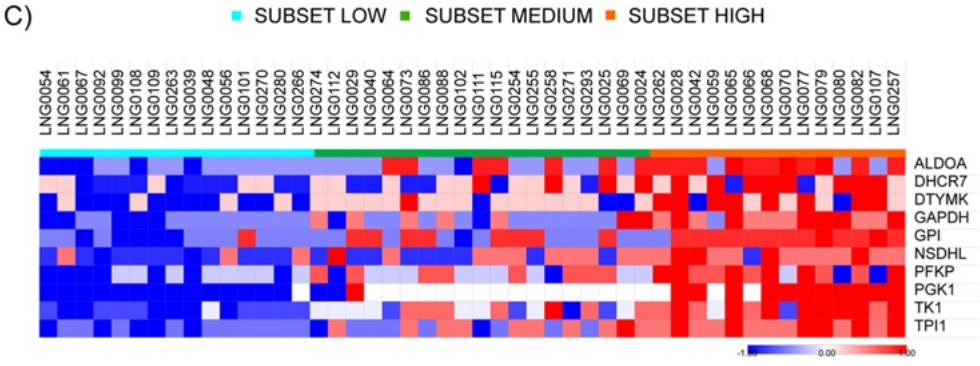


Fig. 8 MAGA LUAD cohort stratification. (**A**) Clinical characteristics of the patient population, reported as number of patients with percent or median with interquartile range; (**B**) Principal component Analysis (PCA) of MAGA LUAD samples performed using as variable the mean expression values of LMetSig genes. Each dot represents a sample and each color represents a subset (subset low=sky-blue, subset medium=green, subset high=orange); (**C**) Heatmap showing the expression of LMetSig genes in MAGA LUAD samples. Red: up-regulated genes; blue: down-regulated genes. Sample subsets are flagged via color annotation: low (sky-blue), medium (green) and high (orange)

TK1 as a putative new prognostic biomarker for LUAD

Comparing subsets "low" and "high" in DEFLeCT and MAGA LUAD cohorts, TK1 emerged as the most significant up-regulated DEG both among the metabolic genes considered and among LMetSig genes (Fig. 9 panel A, Supplementary data 7)

The high expression of TK1 was associated with a shorter DFS in both DEFLeCT (median DFS 42 vs NR months, log-rank *p-value* = 0.001)(Fig. 6 panel C) and LUAD-IC patients (median DFS 11.2 vs 32 months, log-rank *p-value* < 0.0001) (Supplementary figure 4 panel D).

Immunohistochemistry analyses on DEFLeCT samples highlighted the specific TK1 protein expression in tumor cells only and also validated the sample classification in "subset low" and "subset high" according to gene expression level (Fig. 9 panel B).

Discussion

Cancer cell metabolism deregulation leads to uncontrolled proliferation of tumor cells and it is also associated with other malignant "phenotypes", such as migration and invasion [11]. This deregulation is generated through a complex set of genetic alterations, aberrant signaling pathways, crosstalk with stromal cells, and the tumor microenvironment [10–12]. Because of the relevant role of metabolic networks and the superior performance of multi-gene models compared to single gene tests, we could reasonably expect that metabolism-related prognostic models might play a quite significant role in this setting. Some attempts have been made to propose genomic signatures with prognostic roles based on metabolic genes in LUAD, although the results have been highly variable and poorly standardized [13–20].

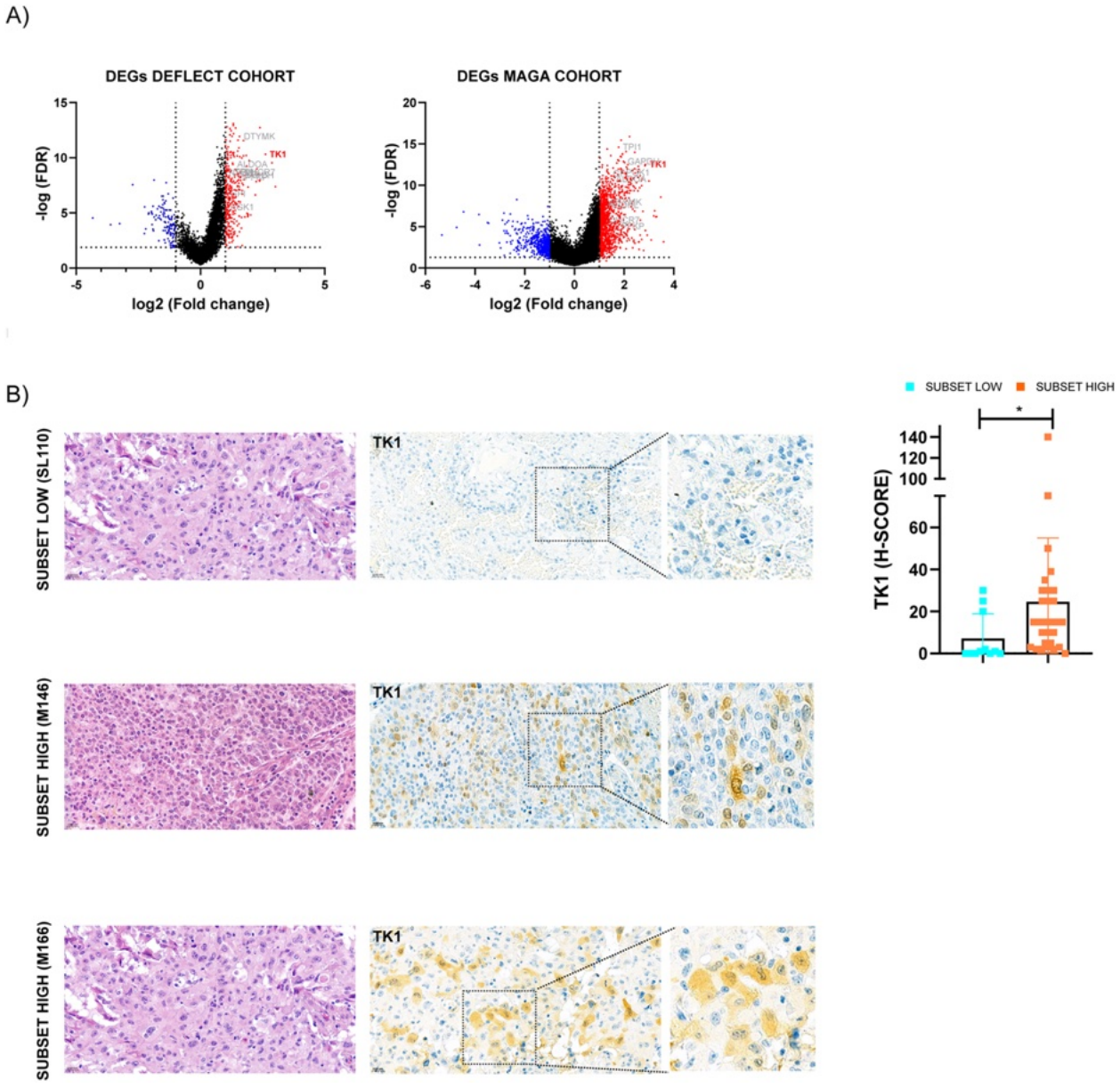


Fig. 9 Expression of TK1 in DEFLeCT LUAD samples. (**A**) Volcano plot of DEGs expression comparing samples of subset high and low of DEFLeCT and MAGA cohort. The LMetSig were labeled (grey). Red: up-regulated genes; blue: down-regulated genes; (**B**) Representative image and relative quantification (as H-score) of TK1 IHC staining of DEFLeCT LUAD samples belonging to low or high metabolic subsets (n=11 low; n=30 high LUAD samples; *p<0.01 in comparisons by Student's t-test)

In this study we identified *"LMetSig",* a *new metabolic signature,* able to stratify early-stage LUAD patients and to predict prognosis. Due to the limited number of events at the time of analysis, OS data were not mature. Furthermore, OS might be greatly influenced by the type of treatment received in case of relapse that was driven by the tumor profile (e.g., oncogene addiction, PDL1 expression, etc). Therefore, our main focus was on the correlation between the metabolic signature and DFS. The high expression of LMetSig was associated with shorter DFS in two different independent early-stage LUAD cohorts: we detected a statistically significant correlation of LMetSig with DFS in DEFLeCT LUAD cohort composed by 87 patients but also in the LUAD-IC composed by more than 1300 patients, further highlighting its potential prognostic value.

Notably, metabolic subsets had no significant association with clinical parameters, but had an independently significant correlation with DFS. With the limitation of the sample size, LMetSig up-regulation showed a greater impact on DFS in patients with stage I-II than in those with stage III disease. Our data suggest that the metabolic

signature might play a pivotal role in identifying high-risk patients within stage I-II. The lack of a significant correlation between metabolic signature and outcomes in stage III patients deserves further investigation. Indeed, this result may be due to the small sample size in our study or it might suggest that, in more advanced stages, other tumor hallmarks overshadow the predictive value of metabolic signatures. Indeed, at a more advanced stage of the disease, the substantial tumour burden, increased genomic instability, and multiple concurrent pathophysiological processes can accelerate tumour progression. These factors may introduce confounding variables that interfere with the reliability of analytical measurements.

The absence of significant associations between LMetSig and standard clinical parameters such as smoking status, gender, TNM stage or tumor grade suggests that the signature may act as an independent biological determinant rather than reflecting conventional clinical features. This independence might reinforce its potential value as a complementary prognostic marker, capable of capturing punctual molecular aspects of tumor behavior not explained by routinely assessed clinical variables.

"Subtype high" tumors exhibited a trend towards higher uptake values (quantified as SUVmax) of the primary tumor at PET/CT scan. To date, several studies have indicated that SUVmax may represent an independent risk factor for recurrence but available evidence is not conclusive [32, 33].

The study also suggested a putative relationship between the metabolic profile and the mutational landscape of tumors. The significant correlation between high expression of LMetSig and TMB is worthy of further investigation as TMB is a potential biomarker for response to immune checkpoint inhibitors, overall tumor aggressiveness and often associated with the occurrence of neoantigens. In addition, the analysis of the distribution of the key LUAD mutated oncogenes (such as EGFR, KRAS, BRAF, and MET) showed that KRAS mutations are enriched in the metabolic "subset high" tumors. Only 30% of all patients in DEFLeCT LUAD cohort who relapsed harbored KRAS mutations, and all had a medium or high expression of LMetSig. The lack of a significant difference in DFS between mutated and non-mutated tumours, both overall and within metabolic subsets, emphasizes the complexity of the genetic and epigenetic networks in tumor progression.

LMetSig is composed by ten genes: ALDOA, DHCR7, DTYMK, GAPDH, NSDHL, PFKP, PGK1, GPI, TPI1 and TK1. We examined the degree of overlap between LMetSig and previously published metabolic signatures. Indeed, GPI, ALDOA, TPI1 and PFKP were indicated as metabolic markers related to the poor prognosis of LUAD patients in the studies of Xue C. et al. [15] and Zhao Z. et al. [16], as well as in the study of Wang Z. et al. [17]. In particular, these studies found that these genes were associated with a reduction in OS of LUAD patients. Although the overall overlap was limited, these data emphasised the prognostic value of these metabolic genes.

ALDOA gene, encodes the enzyme that facilitates the reversible transformation of fructose-1,6-bisphosphate into glyceraldehyde-3-phosphate and dihydroxyacetone phosphate and has a clear role in the progression of several cancer types, including lung cancer [34]. Recently, ALDOA has also emerged as a potential therapeutic target [35]. The PFKP gene is responsible for encoding the enzyme that facilitates the conversion of fructose-6-phosphate into fructose-1,6-bisphosphate. PFKP may be involved in metabolic reprogramming of lung cancer and it is often found to be overexpressed [36]. Specifically, Shen et al. have shown that higher PFKP expression in lung tumors correlates with worse OS while the reduction of PFKP level leads to decreased glucose uptake, which has potential antitumor effects [36]. PGK1 is a key gene in glycolysis, known for its overexpression in LUAD, which has been associated with worse prognosis [37]. Its function extends beyond metabolism, influencing hypoxia mechanisms, DNA synthesis, the recruitment of immune cells, including M2 macrophages and exhausted T cells, contributing to the immunosuppressive tumor microenvironment [37]. TPI1 has shown a significant correlation between overexpression and unfavorable prognosis in LUAD [38]. Studies in mouse models indicate that inhibiting TPI1 can significantly reduce cell migration, colony formation, and tumor growth. Finally, two genes of the LMetSig are involved in nucleotide synthesis: DTYMK and TK1. Regarding DTYMK, its prognostic role has already been demonstrated in patients with NSCLC [39]. In LUAD, high expression levels of DTYMK had been significantly associated with reduced DFS and OS, with an increase of immune-infiltration, tumor cell resistance to chemotherapy, and cell proliferation and invasion [40, 41].

Among LMetSig, TK1 gene was identified as a candidate prognostic biomarker for LUAD due to three peculiar features: i) TK1 was undetectable in normal tissue; ii) TK1 emerged as one of the most significant DEGs comparing different tumor tissues; iii) TK1 showed a significant correlation with worse DFS.

TK1 could serve as a putative therapeutic target, a predictive biomarker, and a marker of treatment response in LUAD. TK1 encodes an enzyme involved in DNA synthesis, particularly of pyrimidines, during the cell replication phase. Its expression significantly increases in active proliferating tumor cells where the "*salvage* pathway" of synthesis of pyrimidines supports the normal synthesis by "*de novo* pathway" [42, 43]. TK1 is particularly overexpressed during the S phase of the cell cycle, making it an

ideal biomarker for tumor proliferation. In vitro evidence suggests that TK1 is not only expressed in the cytoplasm but may also be present on the cell membrane of lung cancer cells [44]. Membrane TK1 is rarely detected in normal tissue, making it a potential tumor-specific target for antibody-based or cell-based therapies [44], particularly in patients with immune checkpoint inhibitors-resistant LUAD [45].

Inhibition of pyrimidine synthesis, particularly the *de novo* pathway, is an established therapeutic strategy in oncology. The most important group of antitumor agents acting at this level includes nucleoside analogous and antimetabolites, such as 5-fluorouracil, gemcitabine, and pemetrexed. Unfortunately, the use of these drugs is associated with the rapid onset of resistance, which is also due to the parallel activation of the "*salvage* pathway" [42, 43]. Currently, there are no specific TK1 inhibitors approved or under clinical investigation. Inhibition of pyrimidine recovery mediated by TK1 is currently possible only through siRNA. Interestingly, silencing the cytosolic TK1 via siRNA enhances the effect of thymidylate synthase inhibition involved in "*de novo*" synthesis, conferring greater sensitivity to pemetrexed in cell lines [46]. Silencing of TK1 gene reduces tumour cell proliferation, invasion and immune evasion. Concurrent knockdown of both KPNA2 and TK1 restores STAT3 regulation and enhances antigen presentation [47]. After pemetrexed administration, the transient increase in TK1 (as shown in salvage-pathway "flare" PET studies) suggests that TK1 activity is dynamically modulated [48].

In the context of identifying new predictors or tracers of lung cancer, molecules that reflect the course of the disease and can be detected in a patient's bloodstream are of fundamental importance. From this perspective, the levels of serum TK1 activity and serum TK1 protein concentration are related to tumor progression and might be used for monitoring outcomes both in early and advanced disease [49–56].

Epigenetic variables, including those linked to tumor cell metabolism, are increasingly recognized as key determinants of cancer progression. Alterations in metabolic pathways can reshape the epigenetic landscape by modulating the availability of essential cofactors, thereby influencing gene expression programs that promote tumor growth, adaptation, and therapeutic resistance. As a result, metabolic interactions are emerging as critical drivers of malignant evolution and represent a growing area of interest for both biomarker development and targeted therapeutic strategies [57].

Despite their limited numbers, the DEFLeCT, MAGA and LUAD-IC cohorts still include patients with a variety of demographic characteristics, smoking habits, clinical and radiological parameters, pathological staging, and molecular features. These characteristics collectively reflect the spectrum observed in most LUAD patients. Interestingly, LMetSig consistently clustered early-stage resected LUAD of DEFLeCT and MAGA cohorts: the comparison of the "subset low" and of "subset high" of the two cohorts showed that the two populations are similar in terms of smoking habit, stage and pathological grading. As previously described, lung tumors, especially LUAD, are heterogeneous, and prognosis varies even within the same TNM stage. The current staging system does not fully capture the molecular complexity of the disease, indicating the need for additional prognostic/predictive factors. A better understanding of tumor metabolism could improve predictions of recurrence and guide more tailored therapeutic strategies, enabling better patient selection for personalized treatments and optimizing follow-up care.

Putative clinical implementation strategy

In summary, LMetSig is a promising tool for improving the prognostic assessment of early-stage lung adenocarcinoma (LUAD) beyond traditional clinical and pathological parameters. It could help to guide risk-adapted management by identifying individuals who could benefit from intensified surveillance or adjuvant treatment despite having early-stage disease.

Although the present study is exploratory, it aims to generate preliminary evidence and lay the groundwork for future investigations based on larger, prospective cohorts.

LMetSig ability to predict disease-free survival independently and to stratify patients within the same TNM stage suggests that metabolic profiling could address a major unmet clinical need: it might distinguish between early-stage patients who are truly at low risk and those who might potentially benefit from more intensive surveillance or adjuvant therapy. Notably, the lack of association between LMetSig and most clinical variables, including tumor grade, highlights its potential as a complementary, biology-driven biomarker.

If prospectively validated, LMetSig could be incorporated as an additional molecular test performed on surgical or biopsy specimens, alongside established clinicopathological parameters. In particular, TK1 assessment might represent a surrogate and reliable biomarker of LMetSig.

Baseline TK1 expression (or its dynamic induction) could influence responsiveness to pemetrexed chemotherapy. Moreover, concurrent targeting or inhibition of TK1 may enhance the anti-tumor efficacy of pemetrexed. Thus, TK1 warrants investigation as both a predictive biomarker and a therapeutic target in combination with pemetrexed-based regimens in LUAD. Since TK1 levels reflect proliferation and salvage-pathway activity and given the fact that TK1 is overexpressed in many LUAD

tumors, measuring TK1 expression prior to chemotherapy might represent a novel predictive biomarker for pemetrexed activity (alone or in combination). For example, tumors with high TK1 might show different degrees of sensitivity than tumors with low TK1, depending on how salvage pathways compensate for Thymidine Synthase inhibition. As shown in salvage-pathway "flare" PET studies, the transient increase in TK1 after pemetrexed administration suggested that TK1 activity is dynamically modulated [48]. Given the presence of TK1 on the cell membrane in lung cancer but not in healthy lung tissue, targeted therapies (e.g., antibodies, ADCs, CAR-T cells) against TK1 might selectively kill LUAD tumor cells either alone, or in synergy with pemetrexed chemotherapy. This dual approach could both reduce tumor burden and overcome adaptive resistance. Finally, combining TK1 evaluation (as DFS predictor marker) with pemetrexed might limit metastatic potential, since TK1 appears to support not only proliferation but also invasion and metastasis in LUAD.

From a health economic perspective, the signature relies on a limited number of genes and could be investigated using widely available proteomic (e.g. immunohistochemistry), or transcriptomic platforms (e.g. RT-qPCR), which could support its feasible and cost-effective integration into existing testing pipelines. TK1 is particularly amenable to investigation by IHC, as its expression can easily be assessed using routinely archived FFPE.

Furthermore, given the evidence that serum levels of TK1 can serve as a non-invasive biomarker in different tumors [49–56], integrating serum TK1 assessment alongside LMetSig may represent an additionally, minimally invasive method for monitoring tumour dynamics, improving risk stratification and potentially informing longitudinal clinical decision-making. A prospective evaluation of serum TK1 levels may be carried out using a high-sensitivity commercial Enzyme-Linked Immunosorbent Assay (ELISA).

Overall, when combined with a serum-based evaluation of TK1 where appropriate, LMetSig has the potential to be incorporated into clinical workflows as a practical, scalable and cost-effective molecular classifier that strengthens evidence-based decision-making in LUAD, particularly where current prognostic tools are insufficient. This is particularly relevant given the heterogeneity of LUAD. Therefore, incorporating metabolic signatures such as LMetSig into clinical workflows could help personalise post-surgical management, reduce overtreatment, and ultimately improve patient outcomes.

Abbreviations

NSCLC	Non-small Cell Lung Cancer
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
EGFR	Epidermal Growth Factor Receptor
ALK	Anaplastic Lymphoma Kinase
PD-L1	Programmed Death Ligand protein 1
DEFLeCT	Digital Technology For Lung Cancer Treatment
SUVmax	Maximum standardized uptake
PET/CT scan	[(18)F] fluorodeoxyglucose positron emission tomography/computed tomography
GSEA	Gene Set Enrichment Analysis
GO	BP Gene Ontology Biological Process
DFS	Disease-Free Survival
LUAD-IC	LUAD integrated cohort
PCA	Principal Component Analysis
TMB	Tumor Mutation Burden
NR	Not reached
LmetSig	Lung Metabolic Signature
ALDOA	Aldolase-A
DHCR7	Dehydrocholesterol Reductase
DTYMK	Deoxythymidylate Kinase
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
NSDHL	NAD(P) Dependent Steroid Dehydrogenase-Like Protein
PFKP	6-Phosphofructokinase Type C-F
PGK1	Phosphoglycerate Kinase 1
GPI	Glucose Phosphate Isomerase
TPI1	Triosephosphate Isomerase
TK1	Thymidine Kinase

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-026-07917-5.

Supplementary material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Acknowledgements

Not applicable.

Author contributions

Conceptualization: G.D., P.B., M.O., L.D., F.J., F.B., GV.S., S.N.; methodology: G.D., M.O., M.A., E.Z., W.MHM., F.N.; bioinformatic and clinical data analysis: L.D., D.C., L.M, G.D., M.O., R.C., S.P., P.B., F.J. M.D., V.C., E.G.; pathological samples management: G.D., M.O., M.A., E.Z., F.N., L.R., M.V., L.P., M.P., L.T., F.L., E.R; writing—original draft preparation: G.D., F.J., P.B., L.D., M.O., GV.S., S.N.; writing—review and editing: G.D., F.J., P.B., L.D., M.O., P.F., R.T., F.BE., A.M., GV.S., S.N; supervision: F.B., GV.S, N.S.; funding acquisition: G.D., F.B., GV.S., S.N. All authors have read and agreed to the published version of the manuscript.

Funding

The study was supported by DEFLeCT project (Digital Technology For Lung Cancer Treatment) funded by Regione Piemonte (P.O.R. F.E.S.R. 2014/2020 technological platform "Health and Wellness") and by RILO project (Ricerca Locale 2023, University of Turin).

Data availability

As part of the DEFLeCT project, a portion of some genetic and clinical data were originally published by some of the authors in [23] and has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001007219 (metadata are available under accession number EGAD00001010838). New RNA- and exome-sequencing data about DEFLeCT and MAGA cohorts were produced for this study and requests for raw sequencing data access can be addressed to the corresponding authors.

Declarations

Ethics approval and consent to participate

Enrolment was conducted at San Luigi Gonzaga and Città della Salute e della Scienza University Hospitals in Orbassano and Turin (Italy), respectively. DEFLeCT patients were part of the prospective observational clinical trial PROMOLE approved by the Ethical Committee of the San Luigi Gonzaga University Hospital (protocol n0.14057 approved on September 28, 2018, n0.73/2018, version-3 January 30, 2023). MAGA patients were included in the PROFILING protocol (No. 001-IRCC-00IIS-10) of Candiolo Cancer Institute FPO IRCCS, Candiolo, Turin Italy. The informed consent was collected for each patient at the time of the surgical resection. This study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

No patient data has been published so the consent for publication of the current study was not required. All co-authors agreed for publication.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Oncology, University of Turin, San Luigi Gonzaga University Hospital, 10043 Orbassano, Italy
[2]Medical Oncology Unit, Cardinal Massaia Hospital, 14100 Asti, Italy
[3]aizoOn Technology & Consulting, 10146 Turin, Italy
[4]Oncology Unit, Department of Oncology, University of Turin, San Luigi Gonzaga University Hospital, , 10043 Orbassano, Italy
[5]Department of Translational Medicine (DIMET), University of Piemonte Orientale, I-28100 Novara, Italy
[6]Center for Translational Research on Autoimmune and Allergic Disease (CAAD), I-28100 Novara, Italy
[7]Department of Oncology, University of Turin, Translational Oncology Laboratory "Paola Gilardi", San Luigi Gonzaga University Hospital, 10043 Orbassano, Italy
[8]Department of Molecular Biotechnology and Health Sciences, University of Turin, 10125 Turin, Italy
[9]Candiolo Cancer Institute, FPO-IRCCS, 10060 Candiolo, Italy
[10]Department of Oncology, University of Turin, 10060 Candiolo, Italy
[11]Molecular Biotechnology Center "Guido Tarone", University of Turin, 10126 Turin, Italy
[12]Thoracic Surgery Division, Department of Oncology, University of Turin, San Luigi Gonzaga University Hospital, , 10043 Orbassano, Italy
[13]Department of Surgical Sciences, University of Turin, Città Della Salute e Della Scienza Hospital, , Turin, Italy
[14]Thoracic Surgery Department, Città Della Salute e Della Scienza Hospital, Turin, Italy
[15]Department of Oncology, University of Turin, Città Della Salute e Della Scienza Hospital, Turin, Italy
[16]Pathology Unit, Città Della Salute e Della Scienza Hospital, Turin, Italy

Received: 10 September 2025 / Accepted: 17 February 2026
Published online: 10 March 2026

References

1. American Cancer Society. Key Stats Lung Ca. Available from: https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html.
2. Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, Cantarutti A, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. Ann Oncol. 2019;30:1162–69.
3. WHO Classification of Tumours Editorial Board. Publication of the WHO classification of tumours, 5th Edition ed. Vol.
4. Yatabe Y, Dacic S, Borczuk AC, Warth A, Russell PA, Lantuejoul S, et al. Best practices recommendations for diagnostic Immunohistochemistry in lung cancer. J Thorac Oncol. 2019;14:377–407.
5. Pisano C, Witel G, De Filippis M, Listì A, Napoli F, Righi L, et al. Moving through rare lung cancer histologies: a narrative review on diagnosis and treatment of selected infrequent entities. Precis Cancer Med, AME Publishing Co. 2022;5:27
6. Kris MG, Gaspar LE, Chaft JE, Kennedy EB, Azzoli CG, Ellis PM, et al. Adjuvant systemic therapy and adjuvant radiation therapy for stage I to IIIA completely resected non-small-cell lung cancers: American Society of clinical Oncology/Cancer Care Ontario clinical practice guideline update. J Clin Oncol. 2017;35:2960–74.
7. Passaro A, Al Bakir M, Hamilton EG, Diehn M, André F, Roy-Chowdhuri S, et al. Cancer biomarkers: emerging trends and clinical implications for personalized treatment. Cell. 2024;187:1617–35.
8. Rosell R, Jain A, Codony-Servat J, Jantus-Lewintre E, Morrison B, Ginesta JB, et al. Biological insights in non-small cell lung cancer. Cancer Biol Med. 2023;20:500–18.
9. Mariën H, Derveaux E, Vanhove K, Adriaensens P, Thomeer M, Mesotten L. Changes in metabolism as a diagnostic tool for lung cancer: systematic review. Metabolites. 2022;12:545.
10. Tufail M, Jiang CH, Li N. Altered metabolism in cancer: insights into energy pathways and therapeutic targets. Mol Cancer. 2024;23:203.
11. Elia I, Haigis MC. Metabolites and the tumour microenvironment: from cellular mechanisms to systemic metabolism. Nat Metab. 2021;3:21–32.
12. Zhou Z, Li J, Ousmane D, Peng L, Yuan X, Wang J. Metabolic reprogramming directed by super-enhancers in tumors: an emerging landscape. Mol Ther. 2024;32:572–79.
13. Qin S, Sun S, Wang Y, Li C, Fu L, Wu M, et al. Immune, metabolic landscapes of prognostic signatures for lung adenocarcinoma based on a novel deep learning framework. Sci Rep. 2024;14:527.
14. He L, Chen J, Xu F, Li J. Prognostic implication of a metabolism-associated gene signature in lung adenocarcinoma. Mol Ther Oncolytics. 2020;19:265–77.
15. Xue C, Dai YZ, Li GL, Zhang Y. Prediction of prognosis, efficacy of lung adenocarcinoma by machine learning model based on immune and metabolic related genes. Discov Oncol. 2024;15:778.
16. Zhao Z, He B, Cai Q, Zhang P, Peng X, Zhang Y, et al. A model of twenty-three metabolic-related genes predicting overall survival for lung adenocarcinoma. PeerJ. 2020;8:e10008.
17. Wang Z, Embaye KS, Yang Q, Qin L, Zhang C, Liu L, et al. Establishment and validation of a prognostic signature for lung adenocarcinoma based on metabolism related genes. Cancer Cell Int. 2021;21:219.
18. Tang X, Qi C, Zhou H. Liu Y.A novel metabolic-immune related signature predicts prognosis and immunotherapy response in lung adenocarcinoma. Heliyon. 2022;8:e10164.
19. Zhou Y, Wang Y. Prognostic implication of an energy metabolism-related 11-gene signature in lung cancer. J Biochem Mol Toxicol. 2022;36:e23171.
20. Hu J, Yu H, Sun L, Yan Y, Zhang L, Jiang G, et al. Identification of an individualized metabolism prognostic signature and related therapy regimens in early stage lung adenocarcinoma. Front Oncol. 2021;11:650853.
21. Stine ZE, Schug ZT, Salvino JM, Dang CV. Targeting cancer metabolism in the era of precision oncology. Nat Rev Drug Discov. 2022;21:141–62.
22. Gyorffy B. Transcriptome-level discovery of survival-associated biomarkers and therapy targets in non-small-cell lung cancer. Br J Pharmacol. 2024;181:362–74.
23. Manganaro L, Bianco S, Bironzo P, Cipollini F, Colombi D, Corà D, et al. Consensus clustering methodology to improve molecular stratification of non-small cell lung cancer. Sci Rep. 2023;13:7759.
24. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set KnowledgeDiscovery with Enrichr. Curr Protoc. 2021;1:e90.
25. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, et al. Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res. 2016;44:D463–470.
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.
27. Morpheus. https://software.broadinstitute.org/morpheus.
28. Metsalu T, Vilo J. Clustvis: a web tool for visualizing clustering of multivariate data using Principal Component analysis and heatmap. Nucleic Acids Res. 2015;43:W566–70.
29. Righi L, Volante M, Rapa I, Tavaglione V, Inzani F, Pelosi G, et al. Mammalian target of rapamycin signaling activation patterns in neuroendocrine tumors of the lung. Endocr Relat Cancer. 2010;17:977–87.
30. Bironzo P, Primo L, Novello S, Righi L, Candeloro S, Manganaro L, et al. Clinical-molecular prospective cohort study in non-small cell lung cancer (PROMOLE study): a comprehensive approach to identify new predictive markers of pharmacological response. Clin Lung Cancer. 2022;23:e347–52.

31. Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification of malignant Tumours. Wiley Blackwell. 8th Edition. 2016 December. ISBN: 978-1-119-26357-9.
32. Sujit SJ, Aminu M, Karpinets TV, Chen P, Saad MB, Salehjahromi M, et al. Enhancing NSCLC recurrence prediction with PET/CT habitat imaging, ctDNA, and integrative radiogenomics-blood insights. Nat Commun. 2024;15:3152.
33. Blumenthaler AN, Hofstetter WL, Mehran RJ, Rajaram R, Rice DC, Roth JA, et al. Preoperative maximum standardized uptake value associated with recurrence risk in early lung cancer. Ann Thorac Surg. 2022; 113:1835–44.
34. Fu H, Gao H, Qi X, Zhao L, Wu D, Bai Y, et al. Aldolase a promotes proliferation and G1/S transition via the EGFR/MAPK pathway in non-small cell lung cancer. Cancer Commun. 2018;38:18.
35. Chang YC, Chiou J, Yang YF, Su CY, Lin YF, Yang CN, et al. Therapeutic targeting of aldolase a interactions inhibits lung cancer metastasis and prolongs survival. Cancer Res. 2019;79:4754–66.
36. Shen J, Jin Z, Lv H, Jin K, Jonas K, Zhu C, et al. PFKP is highly expressed in lung cancer and regulates glucose metabolism. Cellular Oncol. 2020;43:617–29.
37. Yang Y, Cui H, Li D, Gao Y, Chen L, Zhou C, et al. Prognosis and immunological characteristics of PGK1 in lung adenocarcinoma: a systematic analysis. Cancers (Basel). 2022;14:5228.
38. Liu P, Sun SJ, Ai YJ, Feng X, Zheng YM, Gao Y, et al. Elevated nuclear localization of glycolytic enzyme TPI1 promotes lung adenocarcinoma and enhances chemoresistance. Cell Death Dis. 2022;13:205.
39. Heydari N, Mahdizadeh M, Jafari SM. The evolving landscape of involvement of DTYMK enzymes in cancer. Med Oncol. 2023;40:213.
40. Zhang Y, Wang H, Liu Y, Yang J, Zuo X, Dong M, et al. Comprehensive analysis of DTYMK in pan-cancer and verification in lung adenocarcinoma. Biosci Rep. 2022;42: BSR20221170.
41. Chen X, Yuan Y, Zhou F, Huang X, Pu J, Niu X, et al. Comprehensive analysis of DTYMK for estimating the immune microenvironment, diagnosis, prognosis effect in patients with lung adenocarcinoma. Aging. 2022;14:7866–76.
42. Mullen NJ, Singh PK. Nucleotide metabolism: a pan-cancer metabolic dependency. Nat Rev Cancer. 2023;23:275–94.
43. Walter M, Herr P. Re-discovery of pyrimidine salvage as target in cancer therapy. Cells. 2022;11:739.
44. Weagel EG, Burrup W, Kovtun R, Velazquez EJ, Felsted AM, Townsend MH, et al. Membrane expression of thymidine kinase 1 and potential clinical relevance in lung, breast, and colorectal malignancies. Cancer Cell Int. 2018;18:135.
45. Malvi P, Janostiak R, Nagarajan A, Cai G, Wajapeyee N. Loss of thymidine kinase 1 inhibits lung cancer growth and metastatic attributes by reducing GDF15 expression.Plos. Genet. 2019;15:e1008439.
46. Di Cresce C, Figueredo R, Ferguson PJ, Vincent MD, Koropatnick J. Combining small interfering RNAs targeting thymidylate synthase and thymidine kinase 1 or 2 sensitizes human tumor cells to 5-fluorodeoxyuridine and pemetrexed. J Pharmacol Exp Ther. 2011;338:952–63.
47. Liu W, Zou H, Guo L, Zhou Z, Xie Y, Guo H, et al. Coordinating oncogenesis and immune evasion: KPNA2, GOLM1, and TK1 as novel CAR T-cell targets in lung adenocarcinoma. Eur J Med Res. 2025;30:765.
48. Aravind P, Popat S, Barwick TD, Soneji N, Lythgoe M, Sreter KB, et al. A subset of non-small cell lung cancer patients treated with pemetrexed show 18F-Fluorothymidine "flare" on positron emission tomography. Cancers (Basel). 2023;15:3718.
49. Chen Y, Ying M, Chen Y, Hu M, Lin Y, Chen D, et al. Serum thymidine kinase 1 correlates to clinical stages and clinical reactions and monitors the outcome of therapy of 1, 247 cancer patients in routine clinical settings. Int J Clin Oncol. 2010;15:359–68.
50. Xu Y, Shi QL, Ma H, Zhou H, Lu Z, Yu B, et al. High thymidine kinase 1 (TK1) expression is a predictor of poor survival in patients with pT1 of lung adenocarcinoma. Tumor Biol. 2012;33:475–83.
51. Wang H, Wang X, Xu L, Zhang J, Cao H. High expression levels of pyrimidine metabolic rate-limiting enzymes are adverse prognostic factors in lung adenocarcinoma: a study based on the cancer Genome atlas and gene expression Omnibus datasets. Purinergic Signal. 2020;16:347–66.
52. He X, Wang M. Application value of serum TK1 and PCDGF, CYFRA21-1, NSE, and CEA plus enhanced CT scan in the diagnosis of non small cell lung cancer and chemotherapy monitoring. J Oncol. 2022;2022:8800787.
53. Nisman B, Nechushtan H, Biran H, Gantz-Sorotsky H, Peled N, Gronowitz S, et al. Serum thymidine kinase 1 activity in the prognosis and monitoring of chemotherapy in lung cancer patients: a brief report. J Thorac Oncol. 2014;9:1568–72.
54. Wang Z, Zhang G, Li Z, Li J, Ma H, Hei A, et al. STK1p as a prognostic biomarker for overall survival in non-small-cell lung carcinoma, based on real-world data. Future Sci OA. 2020;7:FSO661.
55. Jiang ZF, Wang M, Xu JL. Thymidine kinase 1 combined with CEA, CYFRA21-1 and NSE improved its diagnostic value for lung cancer. Life Sci. 2018;194:1–6.
56. Alegre MM, Weyant MJ, Bennett DT, Yu JA, Ramsden MK, Elnaggar A, et al. Serum Detection of thymidine kinase 1 as a means of early Detection of lung cancer. AnticancerResearch. 2014;34:2145–51.
57. Verma A, Lindroth AM. The emerging intertwined activities of metabolism and epigenetics unveils culprits and prospects in cancer. Exp Mol Med. 2025;57:1928–39.

Publisher's Note