# RESEARCH



# Multi-omics analysis identifies different molecular subtypes with unique outcomes in early-stage poorly differentiated lung adenocarcinoma



Bing Liu<sup>1†</sup>, Wei Tao<sup>2†</sup>, Xuantong Zhou<sup>1†</sup>, Li-Di Xu<sup>2</sup>, Yanrui Luo<sup>2</sup>, Xin Yang<sup>3</sup>, Qingjie Min<sup>4</sup>, Miao Huang<sup>1</sup>, Yuge Zhu<sup>1</sup>, Xinrun Cui<sup>1</sup>, Yaqi Wang<sup>1</sup>, Tongyang Gong<sup>1</sup>, Enli Zhang<sup>2</sup>, Yu S. Huang<sup>2</sup>, Weizhi Chen<sup>2</sup>, Shi Yan<sup>1\*</sup> and Nan Wu<sup>5,6\*</sup>

# Abstract

**Introduction** Early-stage poorly differentiated lung adenocarcinoma (LUAD) is plagued by a high risk of postoperative recurrence, and its prognostic heterogeneity complicates treatment and surveillance planning. We conducted this integrative multi-omics study to identify those patients with a truly high risk of adverse outcomes.

**Methods** Whole-exome, RNA and whole methylome sequencing were carried out on 101 treatment-naïve earlystage poorly differentiated LUADs. Integrated analyses were conducted to disclose molecular characteristics and explore molecular subtyping. Functional validation of key molecules was carried out through in vitro and in vivo experiments.

**Results** Recurrent tumors exhibited significantly higher ploidy (p = 0.024), the fraction of the genome altered (FGA, p = 0.042), and aneuploidy (p < 0.05) compared to non-recurrent tumors, as well as a higher frequency of CNVs. Additionally, recurrent tumors showed hypomethylation at both the global level and in CpG island regions. Integrative transcriptomic and methylation analyses identified three molecular subtypes (C1, C2, and C3), with the C1 subtype presenting the worst prognosis (p = 0.024). Although frequently mutated genes showed similar mutation frequencies across the three subtypes, the C1 subtype exhibited the highest tumor mutation burden (TMB), mutant-allele tumor heterogeneity (MATH), aneuploidy, and HLA loss of heterozygosity (HLA-LOH), along with relatively lower immune cell infiltration. Furthermore, *GINS1* and *CPT1C* were found to promote LUAD progression, and their high expression correlated with a poor prognosis.

**Conclusions** This multi-omics study identified three integrative subtypes with distinct prognostic implications, paving the way for more precise management and postoperative monitoring of early-stage poorly differentiated LUAD.

Keywords Multi-omics, Poorly differentiated, Lung adenocarcinoma

<sup>†</sup>Bing Liu, Wei Tao and Xuantong Zhou contributed equally to this work.

\*Correspondence: Shi Yan yanshi@bjmu.edu.cn Nan Wu nanwu@bjmu.edu.cn Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

# Introduction

Lung adenocarcinoma (LUAD), the predominant pathological subtype of lung cancer, demonstrates significant histological heterogeneity. In 2020, the pathology committee of the International Association for the Study of Lung Cancer (IASLC) proposed a novel grading system for invasive LUAD, where any tumor with 20% or more high-grade patterns (solid, micropapillary, and complex glandular patterns) was classified as a poorly differentiated (Grade 3) invasive pulmonary adenocarcinoma [1], which accounted for 34-55% of all resected LUADs and predicted the worst survival outcome [2]. While this pattern-based grading system represents a significant advancement, it faces limitations in prognostic stratification across the heterogeneous spectrum of disease outcomes, particularly in the poorly differentiated group. Notably, only  $\sim 30\%$  of patients with early-stage poorly differentiated LUAD experience postoperative recurrence [3], underscoring the critical need for additional molecular or biological parameters to complement the current grading system. Such multi-dimensional refinement would enable a more precise prognostic evaluation and formulate proper plans for treatment and surveillance.

Omics studies are capable of providing multi-dimensional and high-resolution molecular information, thereby, assisting in identifying molecular characteristics and refining disease subtypes in solid tumors. Over the past decade, large-scale omics studies have significantly advanced our understanding of LUAD by delineating comprehensive mutational profiles, establishing molecular classification systems based on actionable driver mutations (e.g., EGFR and ALK), and uncovering critical roles of epigenetic regulation (e.g., chromatin modifications) and post-transcriptional processes (e.g., alternative splicing) in tumor pathogenesis [4]. Moreover, several transcriptomic stratifications have been also established, which are associated with specific genomic alterations for targeted therapy or immunotherapy, and describe different clinical outcomes [5, 6]. These studies have further refined the molecular classification for the personalized treatment of LUADs. However, since the clinicopathological phenotype still dominates the routine framework of clinical diagnosis, treatment and prognostic evaluation, incorporating an evolving understanding of molecular profiling with histopathological development is an urgent need for precise decision-making. Based on this requirement, adding a molecular classification could more accurately judge the prognosis within a certain pathological domain with a negative survival impact. Recent studies have shown that clustered molecular signatures can contribute to prognostic discrepancies in ovarian cancer and glioma presenting with poorly differentiated features [7-9]. Therefore, elucidating the molecular landscape of early-stage poorly differentiated LUAD could lay a foundation for comprehending the high-risk molecular characteristics of this heterogeneous entity and developing reliable prognostic biomarkers and precise strategies.

Pioneering studies have explored the mutational features of poorly differentiated (Grade 3) LUADs, which showed that they had a greater proportion of ALK rearrangements and KRAS mutations than those with Grades 1-2 [10]. Additionally, other studies have shown that micropapillary or solid predominant poorly differentiated LUADs have a high tumor mutational burden (TMB), fraction of genome altered (FGA) and copy number amplifications (CNV) [11]. These studies indicate that poorly differentiated LUADs have relatively specific genomic characteristics. Some recent studies have evaluated the correlation between histological grade and PD-L1 expression and immune cell infiltration, and found that poor differentiated tumors exhibited higher PD-L1 expression and more T lymphocyte infiltration [12]. These results suggest that patients with poor differentiated tumor are more likely to benefit from immunotherapy. Despite these studies accelerating our understanding of the genomic and immune microenvironmental features of poor differentiated LUAD to some extent, comprehensive and in-depth molecular characterization of this disease entity remains elusive, especially from a multi-omics perspective.

In this study, we conducted an integrative multi-omics analysis of genomic, epigenetic (methylation) and transcriptomic data from 101 early-stage poorly differentiated LUAD tumors and their paired normal tissues. Our study delineated the comprehensive characteristics of this aggressive disease entity and identified molecular subtypes with distinct prognoses, which could facilitate precise treatment and postoperative monitoring.

# **Materials and methods**

#### Collection of clinical specimens and public dataset

We enrolled 101 treatment-naïve patients with earlystage poorly differentiated LUAD who underwent radical resection between July 2012 and December 2017 at Peking University Cancer Hospital & Institute. All patients were confirmed as pathological T1 - 3 N0M0 stage (stage I-II) according to the 8th edition of the lung cancer staging system [13] and did not received neoadjuvant therapy, including chemotherapy, targeted therapy or immunotherapy. All hematoxylin and eosin (HE) stained slides were reviewed according to the 2015 WHO classification of lung cancer and the new grading system proposed by the IASLC pathology committee [1, 14]. All tumors were identified as grade 3 invasive LUAD. Detailed clinical information of the individual patients is listed in Table S1. As for specimen collection, primary tumor specimens and paired normal tissues were collected immediately after resection and then snap-frozen and stored at -80 °C at the Biobank of Peking University Cancer Hospital & Institute until further processing. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Peking University Cancer Hospital & Institute (Institutional Review Board No. 2024 KT65). Written informed consent was obtained from each patient before surgery.

RNA read count files for 566 patients from the TCGA LUAD project were obtained from the TCGA legacy archive (https://gdc.cancer.gov/about-data/publicatio ns/pancanatlas). Corresponding clinical data for these patients were retrieved from cBioPortal (https://www.cbioportal.org/). For comparative analysis with the HG cohort, only patients with early stage (TNM stage I and II) from the TCGA cohort were included, referred to as the TCGA cohort. A second independent cohort (GSE31210) [15] consisting of more than 200 early-stage LUAD cases was utilized to further validate the transcriptomic subtyping.

# Nucleic acid extraction

Genomic DNA and total RNA was extracted from pairs of tumor specimens and normal tissues in Genecast Biotechnology Co., Ltd. (Wuxi, China). Briefly, DNA and RNA was extracted from fresh frozen tissue blocks using the AllPrep DNA/RNA Mini Kit (80204, Qiagen) according to the manufacturer's protocol. A total of 101 patients had nucleic acid extracted. Seventy-nine patients had both RNA and DNA extracted, while 22 patients had only RNA or DNA.

# Whole-exome sequencing (WES) and Genomic data processing

Extracted DNA was quantified by Qubit dsDNA HS Assay kit (Life Technologies, California, USA). Then DNA was fragmented into 150–200 bp by using Covaris M220 Focused-ultrasonicator<sup>™</sup> Instrument (Covaris, Massachusetts, USA). Library construction and wholeexome capture of genomic DNA were performed using the KAPA Hyper Prep Kit (Illumina platforms) (KAPA Biosystems, Wilmington, MA) and Twist Human Core Exome kit (Twist Bioscience, San Francisco, USA) following the manufacturer's instruction. The captured DNA was then sequenced on an Illumina NovaSeq 6000 platform with 100-bp paired-end sequencing. The average sequencing depth was 229-fold for tumors and 200-fold for normal tissues.

All raw Illumina sequence data were demultiplexed and converted to fastq files, with the subsequent trimming of adaptors, contamination, and low-quality nucleotides to obtain clean data by using Trimmomatic (version 0.36) [16]. Sentieon [17] (version 202112.04) was used to align the clean reads to the human reference genome (hg19) by the bwa mem algorithm with default parameters. The raw BAM files obtained were subjected to various processing steps including sorting, removal of duplicate reads, local realignment, and base quality score recalibration (BQSR) by using Sentieon tools. These steps were performed to generate final BAM files, which were used for subsequent analysis such as coverage and depth statistics, as well as mutation calling analysis.

### Somatic mutation calling

Somatic nucleotide variants (SNVs) and insertions/deletions (InDels) were detected for each paired sample using GATK Mutect2 [18] (version 4.1.9.0). The resulting variant calls in VCF format were subsequently annotated with ANNOVAR [19]. High-confidence somatic mutations were retained based on the following stringent criteria: total sequencing depth of  $\geq$  40X, at least 4 supporting reads, variant allele frequency (VAF)  $\geq$  0.05, classification as nonsynonymous variants (resulting in amino acid changes), and a maximum population frequency of < 0.02 in the 1000 Genomes Project, Exome Aggregation Consortium (ExAC), and Genome Aggregation Database (gnomAD). The top 30 frequently mutated genes in the HG cohort were subsequently analyzed for mutual exclusivity and co-occurrence using the maftools R package [20] (version 2.12.05).

# Tumor mutational burden (TMB) and Mutant-allele tumor heterogeneity (MATH) score calculation

Tumor mutational burden (TMB) for the HG cohort was quantified as the total number of somatic nonsynonymous variants within the entire covered exome region (sequencing depth >40X), expressed in mutations per megabase (Mb). The mutant-allele tumor heterogeneity (MATH) score was computed using all somatic variants with a variant allele frequency (VAF) ranging from 0.05 to 1, applying the formula: 100 × median absolute deviation (MAD)/median of the VAF [21].

# Somatic copy number alteration calling

FACETS [22] (version 0.6.2) was utilized to identify somatic copy number variant (CNV), as well as to determine tumor purity and ploidy. The total fraction of genome altered (FGA) was calculated as the percentage of a tumor genome showing a copy number different from the whole genome based on the CNV segment file for each tumor. Genomic Identification of Significant Targets in Cancer, version 2.0 (GISTIC2, version 2.0.23) [23] was employed to analyze focal genomic regions that exhibited significant amplification or deletion across all or Recurrence/Recurrence-free subgroup tumors. The aneuploidy score was also calculated based on the total number of altered arms for each tumor as previously suggested [24]. For the gene-wise result from GISTIC2, spearman correlation coefficients were computed to assess the correlation between gene level CNV and mRNA abundances, with an FDR threshold of less than 0.05.

# Chromosomal number instability (CNI) and microsatellite instability (MSI) score calculation

We utilized CNVkit (version 0.9.2) [25] to detect copy number variations (CNVs) in tumor samples from each patient. The"-reference"parameter was employed to specify a copy number baseline derived from the Genecast normal database, serving as a negative control. After correction for GC content and length of target region using proprietary algorithms for each region, the read counts were transformed into log2 ratios and converted into Z-score based on Gaussian transformations versus a normal control group. The target regions that satisfied the Z-score greater than the 95<sup>th</sup> percentile plus twicetimes absolute standard deviation of the normal control group were retained, and the Z-score was summed as the CNI score [26]. MSIsensor2 (version v0.1) (https:// github.com/niu-lab/msisensor2) was used with default parameters to detect microsatellite instability (MSI) score.

## HLA genotyping and HLA-LOH analysis

Reads from regions of HLA genes were extracted from normal BAM files using SAMtools (version 1.3) [27] and subsequently analyzed by HLA-HD software (version 1.2.0.1) [28] to identify Human Leukocyte Antigen class I (HLA-I) genotypes, employing the following parameters: minimum tag size set to 50 and cutting rate set to 0.95. Loss of heterozygosity (LOH) in HLA genes was determined using LOHHLA software (version 1.1.6) [29].

#### Neoantigen prediction

Neoantigens were predicted using the netMHC- 4.0 (version 4.0a) algorithm, incorporating somatic SNVs, InDels, and HLA genotypes [30]. Predicted results meeting the criteria of a binding affinity (Aff) < = 500 for the mutant (mut) and an Aff(mut)/Aff(wild) ratio < 1 were identified as neoantigens. The neoantigen burden (TNB) for each tumor sample was determined by summing the number of predicted binder mutations per Mb (whole covered exome region was same to TMB).

# Pathway alteration analysis

The somatic mutation genes were categorized into 10 canonical oncogenic signaling (COS) pathways and 8 DNA damage repair (DDR) pathways, based on previous research conducted separately [31]. The COS pathway

comprised 335 genes and included pathways such as cell cycle, Hippo, Myc, Notch, Nrf2, PI- 3-Kinase/Akt, RTK-RAS, TGF signaling, p53, and  $\beta$ -catenin/Wnt. The DDR pathway encompassed mismatch repair (MMR), base excision repair (BER), checkpoint factors (CPF), Fanconi anemia (FA), homologous recombination repair (HRR), nucleotide excision repair (NER), nonhomologous endjoining (NHEJ), and DNA translesion synthesis (TLS). It involved a total of 233 genes. If a mutated gene was found in a specific pathway, it was inferred that the patient had a mutation in that pathway. The mutation frequency within a cohort was calculated as the number of patients with mutations in that pathway divided by the total number of patients. Comparison of the mutation frequency of each pathway across different cohorts or groups was performed using Fisher's exact test.

### Mutational signatures analysis and comparison

We performed mutational signature analysis for the HG cohort using the DeconstructSigs R package (v1.8.0) [32]. Thirty COSMIC cancer signatures (https://cancer.sanger. ac.uk/signatures/signatures\_v2/) were considered, and the contributions (weights) of these signatures in each tumor were normalized to a range between 0 and 1. The weight values of signatures between the Re and Rf groups were compared using the Wilcoxon rank-sum test.

# RNA sequencing (RNA-seq) and data processing

The quantity of extracted RNA was measured using a Qubit 3.0 Fluorometer, while the quality was assessed using the Agilent 2100 Bioanalyzer system assay. After the rRNA removal from the total RNA, the cDNA library was constructed using the SMARTer Stranded Total RNA-Seq Kit v2 (634,412, Takara). Following PCR enrichment and purification of adapter-ligated fragments, the libraries were paired-end sequenced (PE150) using the Illumina NovaSeq 6000 Sequencing System. The average sequencing depth was 56 million reads for both tumor and normal tissue samples.

Cutadapt (version 4.4) (https://doi.org/https://doi.org/ 10.14806/ej.17.1.200) was utilized to remove the last 3 bases from each read. Trimmomatic (version 0.36) [16] was utilized to eliminate reads that contained adaptors, poly-N sequences, and low-quality reads using default parameters. The resulting trimmed reads were then aligned to the human hg19 reference transcriptome using hisat2 (version 2.1.0) [33]. The alignment data in BAM format was sorted and indexed using SAMtools (version 1.3) [27].

# Differential gene expression and pathway enrichment analysis

To estimate the expression level of each gene, feature-Counts (version 1.6.5) [34] were applied. Transcripts per million (TPM) values were calculated by normalizing the read counts, dividing them by the gene length and the total number of reads mapped to protein-coding genes. In total, 18,071 genes were initially profiled. DEGs (Differential Gene Expression) between different groups were identified using the DESseq2 package (version 1.38.3) [35] in the R software, with the criteria of |log2 (Fold Change)|> 1 and a Bonferroni-adjusted p-value < 0.05. Gene Set Enrichment Analysis (GSEA, version 4.3.3) was employed to perform pathway enrichment analysis among different groups [36]. Molecular Signatures Database (MSigDB, version 7.1) of hallmark gene sets (H), curated gene sets (C2), ontology gene sets (C5), oncogenic signature gene sets (C6), and immunologic signature gene sets (C7) were used in GSEA analysis. Differential gene enrichment analysis between groups was performed using the clusterProfiler R package using Reactome databases. The analysis included all differentially expressed genes, as well as those that were upregulated and downregulated. A significance threshold of p-value < 0.05 was applied.

# Immune cell infiltration and gene expression signatures analysis

To evaluate the tumor microenvironment, the ESTI-MATE R package (v1.0.13) [37] was employed with default settings to derive the ImmuneScore and StromalScore from gene expression data. A higher Immune-Score indicates a greater immune cell infiltration within the tumor. To predict responses to immune checkpoint blockade (ICB) therapy in our HG cohort, we utilized an 18-gene T cell-inflamed gene expression profile (GEP) [38]. The GEP score for each sample was calculated as a weighted sum of these 18 genes, normalized against 11 housekeeping genes. Signature enrichment scores for 28 immune cell subsets within the tumor microenvironment [39] and 14 functional states derived from the CancerSEA database [40] were calculated using the gene set variation analysis (GSVA) R package (version 1.42.0) [41].

Whole methylome sequencing (WMS) and data processing

After the DNA extraction, WMS (Whole-methylome sequencing) libraries were generated using the NEBNext Enzymatic Methyl-seq Kit from New England Biolabs, according to the manufacturer's instructions. The quantification of the libraries was carried out using the Qubit dsDNA HS Assay Kit from Thermo Fisher Scientific. The libraries were then subjected to paired-end sequencing with a read length of 100 base pairs on the NovaSeq 6000

platform from Illumina. The average sequencing depth was nine fold for both tumor and normal tissue samples.

The raw methylation sequencing reads were processed using Trimmomatic (version 0.36) [16] to remove adaptors and eliminate low-quality reads. The clean reads were then aligned to the human reference genome (hg19) and deduplicated using BisMark (version 0.23.0) [42]. SAMtools (version 1.3) [27] and BamUtil (https://github. com/statgen/bamUtil) were used for sorting and overlapclipping of mapped reads. Reads with mapping quality below 20 were filtered out by SAMtools. The methylation status of each CpG site was extracted from a sorted bam file using the bismark methylation extractor function from BisMark. The beta value for each CpG site was calculated as the ratio of methylated CpGs to the sum of methylated and unmethylated CpGs in each sample. To examine genome-wide methylation patterns, the genome was divided into 1,846 non-overlapping 1-Mb segments after excluding regions that overlapped with Duke blacklisted regions or the hg19 gap track38 [43]. Regions in 1-Mb segment and CpG island (https://genome.ucsc. edu/cgi-bin/hgTables) were then used to calculate the mean methylation. Tumor fraction was also extracted from the whole-methylome sequencing data using ichorCNA (version 0.2.0) [44], with normal copy number variation (CNV) files as a reference. Furthermore, an updated plasma aneuploidy score (PAscore) was calculated to summarize chromosome arm-level copy number alternation by adapting a previously described approach, using 30 PBMC profiles as reference baseline [45].

# Differentially methylation region (DMRs) analysis

To identify differentially methylated regions (DMRs) between recurrence and non-recurrence tumors, we utilized the DSS package (version 2.47.1) [46], with the results obtained from bismark methylation extractor as input. The following parameters were configured for DSS analysis: smoothing set to TRUE, smoothing span set to 500, minimum number of CpGs (minCG) set to 3, minimal length (minlen) set to 50, delta set to 0.1, distance threshold for merging (dis.merge) set to 100, percentage of significant probes (pct sig) set to 0.5, and p value <0.05 was considered as statistical significant. DMRs were categorized as hyperDMRs or hypoDMRs based on the direction of methylation change. The average methylation levels in both directions were calculated using the smoothened CpG methylation levels. ChIPseeker (version 1.32.0) [47] was used for DMR annotation.

# Association analysis between gene expression and promoter methylation

To investigate the relationship between gene expression and DNA methylation at the gene level, we categorized

the gene expression levels into deciles based on the ascending distribution of log2-transformed TPM values [48]. The methylation level of a gene was determined by calculating the average methylation level of the CpGs within the nearest gene interval to the transcription start site (TSS). The criteria for selecting the interval were as follows: 1) If more than three CpGs were found within a 200 bp range of the TSS, we assigned the average DNA methylation value of the adjacent CpGs in the TSS200 region to the corresponding gene; 2) If there were fewer than three CpGs within the TSS200 region, we calculated the average of the CpGs located in both TSS200 and the first exon of the gene; 3) If there were fewer than three CpGs in TSS200 or first exon, we calculated the average of the CpGs located in the upstream 1500 bp region extending towards the first exon of the gene. Genes with fewer than three CpGs in the interval from TSS1500 to the first exon were excluded from the analysis.

### **Consensus clustering for RNA datasets**

Prior to clustering analysis, we selected the top 500, 1,000, 1,500, 2,000, 2,500, 3,000, and 3,500 most variable coding genes based on their median absolute deviation (MAD) from the differentially expressed genes (DEGs) between tumor and normal samples, using the Cancer-Subtypes R package (version 1.20.0) [49]. Consensus clustering was then performed on these gene subsets using the ConsensusClusterPlus R package (version 1.62.0) [50]. The partitioning around medoids (PAM) algorithm with Spearman distance was employed for clustering. Due to the limited number of samples (n = 82), the maximum number of clusters was set to six. To identify the most representative samples within each cluster, silhouette scores were computed for all samples using the CancerSubtypes R package. A two-cluster solution based on the top 3,500 most variably expressed genes (MADranked) was selected due to its relatively higher average silhouette value, distinct separation between clusters in the consistent heatmap, and significant association with patient recurrence-free survival (RFS). Tumors exhibiting a silhouette value less than 0, as well as those from patients lacking recurrence-free survival (RFS) data (n = 2), were excluded from both the survival analysis and subsequent single-sample transcriptomic classifier analysis.

# Construction and validation of the single-sample transcriptomic classifier

Building on previously published studies [51], we developed a Spearman nearest-centroid classifier specifically tailored for lung adenocarcinoma (LUAD) patients, utilizing RNA sequencing-derived gene expression data. The construction of the classifier involved first calculating the mean expression levels of samples within each predefined subtype (as described in the aforementioned section). Subsequently, for each individual sample, we determined the Spearman correlation between its gene expression profile and the centroid expression profiles of the subtypes. Each sample was then assigned to the subtype with the highest Spearman correlation coefficient. Samples exhibiting a maximal correlation value below 0.2 were deemed to have a weak association with any subtype and were excluded from the subsequent survival analysis.

To validate the robustness of this RNA-based classification method, we applied it to transcriptomic data from the TCGA cohort and a second independent cohort (GSE31210). For TCGA cohort, the transcripts per million (TPM) values were calculated from raw read counts using the same method as described for the HG cohort.

#### **Consensus Clustering for DNA methylome datasets**

To profile the DNA methylation patterns across 98 tumor samples, we first calculated the average DNA methylation value for CpG sites located within a region of  $\pm 2000$  base pairs of the transcription start site (TSS). This approach allows us to capture the DNA methylation status in the promoter regions of genes, which often play a critical role in regulating gene expression [52]. The subtyping method was similar to that used for RNA datasets. For this analysis, the distance metric applied was'euclidean,'and the clustering algorithm employed was'k-means'.A twocluster solution based on the top 1,000 most variably expressed genes (ranked by median absolute deviation) was selected. Tumors with a silhouette value less than 0, as well as those from patients without recurrence-free survival (RFS) data, were excluded from the survival analysis.

#### Integrate analysis of RNA and WMS subtyping

We utilized the NetworkD3 package (version 0.4) in R to generate a Sankey diagram that illustrates the relationships among RNA subtypes, WMS subtypes, and patient recurrence status. Our analysis revealed that the WMS subgroup effectively partitioned the RNA R1 subgroup into two distinct clusters. In contrast, nearly all samples from the RNA R2 subgroup remained within the same category in the WMS clustering analysis and exhibited fewer recurrences. To integrate the clustering results from both RNA and WMS omics data, we reclassified the samples as follows: samples belonging to both the RNA R1 and WMS M2 subgroups were combined into subtype C1; those belonging to the RNA R1 and WMS M1 subgroups were designated as subtype C2; and samples within the RNA R2 subgroup were retained as subtype C3.

# Identification of stably expressed differentially expressed genes (DEGs) across three integrative subtypes and their prognostic evaluation

Differentially expressed gene (DEG) lists were obtained for each pairwise comparison of the three subtypes (FDR < 0.05 and |log2 FC|> 1). For each subtype, candidate gene lists for both high expression and low expression were generated by taking the intersection of relevant comparisons. For example, in the C1 subtype, genes highly expressed in C1 were identified by intersecting C1vsC2\_up and C1vsC3\_up, while genes with low expression in C1 were identified by intersecting C1vsC2\_down and C1vsC3\_down. To obtain a list of stably expressed DEGs in each subtype, gene filtering was performed. Initially, genes were filtered based on their Transcripts Per Million (TPM) values, retaining those with TPM >1 in more than 50% of samples within the relevant subtype. Subsequently, genes were further filtered based on their coefficient of variation (CV). The CV of TPM values was calculated within each specific subtype, and genes were ranked in ascending order according to their CV values. The top 20 genes with the lowest CV, indicating stable expression within the subtype, were selected. If fewer than 20 genes met this criterion, all available genes were included. To assess the prognostic impact of highly and stably expressed genes in the C1 subtype, we calculated *p*-values and hazard ratios (HRs) for each gene across percentiles from 0.1 to 0.9 in both the HG and TCGA cohorts. Bubble plots were generated to visualize the HRs and *p*-values for recurrence-free survival (RFS) in the HG cohort and disease-free survival (DFS) in the TCGA cohort. Additionally, we utilized the surv\_cutpoint function from the survminer R package to estimate the optimal cutoff for each gene.

# Cell culture

SPCA- 1 was purchased from Chinese Academy of Sciences Cell Bank (Shanghai, China) and cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (HyClone, South Logan, UT, USA), penicillin (100 U/mL), and streptomycin (100  $\mu$ g/mL). HEK293 T was purchased from the American Type Culture Collection (ATCC) (Manassas, VA) and cultured in DMEM supplemented with 10% fetal bovine serum (HyClone, South Logan, UT, USA), penicillin (100 U/mL), and streptomycin (100  $\mu$ g/mL). All cells were maintained at 37°C in a humidified cell incubator with 5% CO<sub>2</sub>.

# Lentivirus infections

shRNA oligos targeting GINS1 or CPT1C and a nontargeting oligo control were engineered into pSIH-puro plasmid. The target sequences for short hairpin RNA were as followed: shGINS1 #1: 5'-CAAGTTCTGGAG GAGATGAAA- 3'; shGINS1 #2: 5'- CTTGCCAAATGC ATTACGATT- 3'; shCPT1C #1: 5'-CTCACGTTTCTG GAATGACTT- 3'; shCPT1C #2: 5'- CCTGCTGATGAC CATGGTTAT- 3'. For pSIH-puro lentivirus production, the packaging plasmids vSVG, pLP1 and pLP2 were used. The indicated packaging plasmids and lentiviral vectors were co-transfected into HEK293 T cells. After 48 h transfection, the supernatant containing lentivirus particles was collected and stored in aliquots at - 80 °C. For lentivirus infection, cells were first treated with polybrene  $(5 \,\mu g/mL)$  (TR- 1003, Sigma), then infected with the indicated lentivirus. Stable cell populations were established by selecting with puromycin  $(2 \mu g/mL)$  (540222, Sigma).

### **RNA extraction and qRT-PCR**

Total RNA was extracted with TRIzol reagent (Thermo Fisher Scientific). The cDNAs were obtained using Quantscript RT kit (Tiangen, Beijing, China) according to the manufacture's protocol. Real-time RT-PCR was performed by using SYBR Premix Ex TaqTM II (TaKaRa, Japan) on Step-one plus real-time PCR system (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. The sequences of primers for qRT-PCR were as followed: GINS1 Forward primer-ACGAGGATGGACTCAGACAAG; GINS1 Reverse primer-TGCAGCGTCGATTTCTTAACA; CPT1C Forward primer-GGATGGCACTGAAGAGGA AA; CPT1C Reverse primer- TCCTGGAAAAGGCATCTCTC; GAP DH Forward primer-CCGGGAAACTGTGGCGTGATGG; GAPDH Reverse primer-AGGTGGAGGAGTGGGTGT CGCTGTT.

# Antibodies and reagents

Antibodies used in this study were as follows: Anti-GINS1 (PA562341, Invitrogen), Anti-CPT1C (66072, Proteintech), Anti- $\beta$ -actin (A5316, Sigma). Secondary antibodies included HRP Goat Anti-Mouse (926–80010, LI-COR) and HRP Goat Anti-Rabbit (926–80011, LI-COR).

# Transwell assay

SPCA- 1 cells  $(8 \times 10^4$  per insert) were suspended in FBS-free RPMI1640 and seeded into the upper chambers with or without pre-coated matrigel (BD, Franklin Lakes, NJ, USA). The bottom chambers were added with RPMI 1640 medium supplemented with 10% FBS. After 24 h

incubation, the migratory or invasive cells were methanol-fixed and stained with crystal violet. Cells in three randomly selected fields were photographed and statistically analyzed.

# Cell viability and colony formation assays

Cell viability was quantified by CCK- 8 assays. SPCA- 1 cells with indicated treatment were seeded into 96-well plates (4 × 10<sup>4</sup> cells/mL; 100 µL/well). Cell Proliferation Reagent CCK- 8 (#CK04, Dojindo Molecular Technologies, Japan) was added to the cell culture medium at a ratio of 1:10. After 1 h of incubation at 37°C, absorbance at 450 nm was measured using a microplate reader (BioTek). For colony formation assays, SPCA- 1 cells with indicated treatment were seeded into 6-well plates (4 × 10<sup>3</sup> cells/well), followed by incubation at 37 °C for 9–12 days until the development of visible colonies. Colonies were stained with crystal violet staining solution and counted.

# Xenograft tumor model

Week-old, female BALB/c nude mice were purchased from Vital River (Beijing, China).  $2 \times 10^6$  SPCA- 1 cells with stable GINS1/CPT1C depletion or control cells were subcutaneously injected into the flanks of mice in separate groups. Each group was composed of 5 mice, randomly chosen. Following 3 weeks, the mice were sacrificed. Tumors were weighed and analyzed by a twotailed, unpaired Student's t test. The Institutional Animal Care and Use Committee of Peking University Cancer Hospital and Institute approved all animal experiments.

# Quantification and statistical analysis

All data statistics and figure generation were conducted using R version 4.1.2. Various statistical tests were employed based on the type of variables being compared. Continuous variables were compared using either the two-sided Wilcoxon rank-sum test or the Kruskal-Wallis rank-sum test. Categorical variables were compared using either the chi-square test or the two-sided Fisher's exact test. Survival analyses were performed using the Kaplan-Meier method and two-sided log-rank tests were used to compare survival curves (R packages survival and survminer). Univariate and multivariate Cox regression analyses were conducted using the survival and survminer packages. To evaluate the predictive ability of each Cox module at specific time points (first year, third year, and fifth year), recurrence prediction analyses were performed using the survival and timeROC packages. Chi-square proportion calculated using R package rms was used to assess the relative contribution of each variable to survival risk. Statistical significance was defined as *p* < 0.05.

# Results

# Mutational landscape of early-stage poorly differentiated LUAD

Totally, 101 patients were enrolled in this study (the HG cohort). The median age of the cohort was 60 years (range: 34–80 years). Sixty-three patients (62.4%) were male, and thirty-eight patients were female (37.6%). At the end of the follow-up period (median follow-up: 43.2  $\pm$  20.1 months), 33 (32.7%) patients experienced recurrence. Detailed individual patient information is listed in Table S1. Totally, WES, WMS and RNA-seq was successfully conducted on matched tumor-normal tissue pairs from 97, 98 and 82 patients in this cohort, respectively (Fig. 1A). Among them, 78 patients had WES, RNA-seq and WMS data, simultaneously (Fig. 1B).

We first investigated the mutational landscape of the HG cohort (Table S2), and the top 30 mutated genes are shown in Fig. 1C. The most frequently mutated genes

Fig. 1 Genomic features of early-stage poorly differentiated LUAD. A Next-generation sequencing platforms applied for the paired normal and tumor tissue in the HG cohort including whole exome sequencing (WES), whole methylome sequencing (WMS) and RNA-seq. Patient numbers were indicated for each platform. B Venn plot to show the overlapping amounts of patients analyzed with three platforms and available clinical parameters. C Mutational landscape of the HG cohort, displaying the top 30 mutated genes. Clinical parameters including age, gender, smoking status, pTNM stage, visceral pleural invasion (VPI), lymphovascular invasion (LVI) and spread through air spaces (STAS), and molecular biomarkers including TNB, TMB, MATH, aneuploidy score, purity, ploidy, FGA and HLA-LOH status are indicated. D Mutational signature analysis of the HG cohort. Signature analysis was performed using the deconstructSigs R package. The x-axis displays 30 signatures from the COSMIC database, and the y-axis representing the weight score of each sample for the corresponding signature. Signature presented in more than 25% samples (Age, BRCA, Smoking, MMR deficiency, APOBEC, Aflatoxin and Ultraviolet) are marked in red. E The co-occurrence and mutually exclusive analysis for the top 30 mutated genes. Negative log transformed p values were indicated with red represents mutually exclusive and blue represents co-occurrence. The determination of an event is based on the odds ratio: an odds ratio > 1 indicates \*\*Co-occurrence\*\*, while an odds ratio < 1 indicates \*\*Mutual exclusivity\*\*. The intensity of the color reflects the significance of the p-value, with darker colors indicating smaller p-values. For both co-occurrence and mutually exclusive events, the smaller the p-value, the higher the corresponding -log10(p-value) in the label, indicating a stronger confidence level for each event (F, G) GISTIC amplification (F) and deletion (G) plot to visualize focal-level copy number variations (CNV)s frequencies. The chromosome is oriented vertically from top to bottom, and GISTIC q-values at each region are plotted from left to right on a log scale. The green line represents the significance threshold (q-value = 0.25).

<sup>(</sup>See figure on next page.)



Fig. 1 (See legend on previous page.)

were EGFR (51%), TP53 (45%), MUC16 (28%), RYR2 (28%), TTN (26%) and KRAS (19%). These findings differed from a recent proteogenomic study of unselected Chinese LUAD patients [53]. In that cohort, the mutational frequencies of MUC16, RYR2 and TTN were lower, and were not among the top 10 mutated genes. Somatic mutational signatures were also investigated in the HG cohort (Table S3), and our analysis identified Signatures 1, 3, 4, 6, 7, 13, and 24 as significant contributors (Fig. 1D). Some of these signatures have been reported to be associated with lung cancer. For instance, Signature 4 is associated with tobacco smoking and is likely due to DNA damage caused by tobacco smoke-derived mutagens [54]. The co-occurrence and mutual exclusivity patterns of the most mutated genes were further investigated (Fig. 1E). Notably, EGFR mutations were mutually exclusive with mutations in many other genes, including TTN, KRAS, RYR1 and MUC16, which was consistent with previous studies [55, 56]. As revealed by FACETS, significant CNVs were observed in the HG cohort (Figure S1). Using GISTIC 2.0, remarkable focal level copy number gains were observed at 1q21.1, 7p11.2, 10q11.21, 14q13.3, 16q11.2, 19q12 and 20q13.33 (Fig. 1F; Table S4). Additionally, losses were detected at 1p11.2, 2q21.2, 9p21.3, 16q11.2, and 16p13.2 (Fig. 1G; Table S4).

# Recurrence-associated molecular events at genomic, epigenomic and transcriptomic levels

Through multi-dimensional omics data, we explored the interactions between the genome, epigenome, and transcriptome (Tables S5 -7). We initially assessed the impact of CNVs on mRNA expression levels. As illustrated in Fig. 2A, CNVs have the potential to influence gene expression either positively or negatively, in both cis and trans modes. A comprehensive analysis of 16,455 CNV-mRNA pairs revealed significant correlations: 588 pairs exhibited significant cis effects, whereas 6,569 CNVs influenced the expression of 4,075 mRNAs in trans (Spearman correlation analysis, FDR < 0.05). We estimated tumor purity in 97 patients using the FACETS algorithm for WES data and the ichorCNA R package for DNA methylation data (Table S8). As depicted in Fig. 2B, the tumor purity estimates derived from these two methods exhibited a strong correlation (Pearson's correlation coefficient; R = 0.69, p < 0.001), indicating the robustness and reliability of these computational approaches for assessing tumor purity. We subsequently investigated the correlation between the epigenome and transcriptome using whole methylome sequencing (WMS) and RNA-seq data. As illustrated in Fig. 2C, DNA methylation exhibited an inverse relationship with gene expression in the first three to four deciles, after which it remained low across higher expression deciles within the specific regions (detailed in the Methods) of both tumor

#### (See figure on next page.)

Fig. 2 Analysis of recurrence-associated molecular features in the HG cohort. A Correlation plot illustrating the relationship between WES-based gene-level CNV and RNA-seq derived RNA expression across the entire genome. Significant (FDR < 0.05) positive and negative correlations are indicated in red and blue, respectively. Genes are ordered by chromosomal location on both the x and y axis. Diagonal line indicates potential cis-effects of CNA on mRNA. B Correlation analysis between tumor purity as measured by WES and WMS. Tumor purity for WMS was calculated using ichorCNA, while Facets software was employed for WES-based estimates. The Pearson correlation line is plotted in blue and its 95% confidence interval (CI) is indicated in the grey shadow. Pearson's correlation coefficient (R) and significant p-value are shown. C Correlation analysis between methylation levels and RNA expression levels for both normal samples (blue) and tumor samples (red). The x-axis displays ten groups ranked by increasing expression levels. Gene expression levels were determined by calculating the median expression for each gene across all samples, log2-transforming the values, sorting the genes by increasing median expression, and dividing them into deciles. Methylation levels, representing the average methylation of CpG sites within specific regions (detailed in the Methods), as shown on the y-axis. Spearman's correlation coefficient (R) and p-value are indicated. D Comparison of clinical and molecular features between recurrence (Re) and recurrence-free (Rf) patients. Clinical parameters including age, smoking status, VPI, LVI, pT stage and receipt of adjuvant therapy were analyzed. Molecular features such as ploidy and fraction of genome altered (FGA) were also assessed. E Mutational signature comparison between Re and Rf groups. Signature analysis was performed using the deconstructSigs R package. The x-axis displays 30 signatures from the COSMIC database, and the y-axis representing the weight score of each sample for the corresponding signature. F CNV frequencies comparison between Re and Rf groups. CNV were calculated using GISTIC2 software. The x-axis represents genomic chromosome positions, while the y-axis depicts CNV mutation frequencies for amplifications (AMP) (upper panel) and deletions (DEL) (down panel). G Comparison of immune profiles between Re and Rf groups. The enrichment scores of 28 immune cell subsets within the tumor microenvironment was guantitatively assessed using the GSVA R package and the ssGSEA method. RNA data in TPM format served as input. The x-axis represents samples, while the y-axis represents immune cell subsets. Each square indicates an enrichment score. H Methylation comparison between Re and Rf groups. Average methylation level within a 1 Mb interval were depicted as blue rings for Rf and Re groups, respectively. Differential methylation regions (DMR)s representing the absolute methylation level difference between Re vs Rf are highlighted in orange (hypermethylation) and green (hypomethylation). DMRs located within CpG islands are denoted as CGI DMRs. Differential methylation analysis was conducted using DSS software, excluding CpG sites with depth < 5. Note: For D, E and G, statistical comparison between RFS groups was performed using the fisher's test for categorical variables, and Wilcoxon test for continuous variables. ns: p > 0.05 (not marked on the figures **E** and **G**), \*: 0.01 , \*\*: <math>0.001 , \*\*\*: <math>p < = 0.001.



Fig. 2 (See legend on previous page.)

(Spearman correlation; R = -0.51, p < 0.001) and normal tissues (Spearman correlation; R = -0.46, p < 0.001).

To identify the determinants of recurrence in this specific disease entity, we analyzed the clinicopathological and molecular differences between recurrent and nonrecurrent cases in the HG cohort. Among the clinicopathological variables examined, visceral pleural invasion (VPI, p = 0.001) and pathological T stage (pT stage, p =0.022) were significantly correlated with recurrence. Other factors such as age, smoking status, lymphovascular invasion (LVI), tumor size, receipt of adjuvant therapy, tumor mutational burden (TMB), and mutant-allele tumor heterogeneity (MATH) were not significantly associated with recurrence (Fig. 2D; Figure S2A; Table S8). We also compared the frequently mutated genes between recurrent and non-recurrent cases and found mutations of some frequently mutated genes, including EGFR, TP53, RYR2 and MUC16, were not correlated with recurrence (Figure S2B). Conversely, mutations in STK11, XIRP2, MXRA5, and ZNF536 were more frequent in the recurrent cases. Previous studies have also indicated that mutations in EGFR could not determine the prognosis of LUAD [57], whereas mutations in STK11 and MXRA5 have been associated with poor prognosis [58, 59]. Additionally, analysis of the 10 canonical oncogene signaling pathways and DNA damage response (DDR) pathways in recurrent and non-recurrent cases revealed no significant differences between the two groups (Figure S2C; Table S9). Next, we investigated the potential association between mutational signatures and recurrence, and found no significant differences in mutational signatures between recurrent and non-recurrent cases (Fig. 2E). Next, we investigated the differences in copy number variations (CNVs) and observed that recurrent tumors exhibited a higher frequency of CNVs, including both amplifications and deletions (Fig. 2F; Table S10). According to our analyses, several molecular features related to chromosomal instability, such as abnormal ploidy, FGA and aneuploidy, were enriched in recurrent tumors, suggesting that chromosomal instability indicates a worse prognosis.

In addition to the intrinsic properties of tumor cells, immune cell populations are crucial prognostic determinants of LUAD. Using single-sample Gene Set Enrichment Analysis (ssGSEA), we compared the enrichment scores of 28 immune cell subsets between recurrent and non-recurrent cases (Table S11). As shown in Fig. 2G, recurrent tumors showed significantly lower enrichment scores for central memory CD4 + T cells and immature dendritic cells (p < 0.05). Previous studies have demonstrated that activated CD8 + T cells, effector memory CD8 + T cells and macrophages are associated with LUAD prognosis, which was not observed in the present

study. Based on the MSigDB database(C2), gene set enrichment analysis (GSEA) of recurrent and non-recurrent cases demonstrated that recurrent cases were significantly enriched in pathways, including the adenylate cyclase activating pathway, tumorigenesis by ret c634r, aml methylation cluster 7 dn, DNA methylation, peptide hormone biosynthesis, and inflammatory response tgfb1 (nominal *p*-value <0.05, Figure S2D; Table S12). We investigated methylation differences between recurrent and non-recurrent cases using WMS data. Our analysis revealed a global pattern of hypomethylation in 1 M-bp regions in recurrent cases compared to that in non-recurrent cases. Specifically, among the 10,569 differentially methylated regions (DMRs) identified, 95.7% exhibited global hypomethylation, and 77.8% of the CpG island (CGI) regions displayed similar trends (Fig. 2H; Tables S13, 14).

Patient subtyping based on multiple molecular dimensions Next, we explored the molecular subtyping of early-stage poorly differentiated LUADs using different omics datasets. Initially, bulk RNA-seq data from 82 patients were analyzed to explore the transcriptomic subtypes of earlystage poorly differentiated LUADs. A total of 3,725 genes were differentially expressed between tumor and normal tissue ( $|(\log 2 \text{ FC})| > 1$ , FDR < 0.05). Of these, the expression of 2,197 genes was upregulated and the expression of 1,528 genes was downregulated in tumor tissues (Fig. 3A; Table S15). Through unsupervised consensus clustering of the top 3,500 differentially expressed genes (DEGs), we identified two distinct transcriptomic groups in the HG cohort: R1 (n = 38) and R2 (n = 44) (Fig. 3B). To enable the application of these transcriptomic classes in future research and clinical settings, we developed a single-sample classifier based on the clustering results. This classifier was constructed by using an approach adopted in a recent study of non-muscle invasive bladder cancer [51], wherein a group label was assigned to the transcriptomic profile of a tumor based on its correlation with group-specific mean expression profiles (detailed in the Methods section). We applied this classifier to both the HG (n = 82) and TCGA cohorts (n = 394), and reclassified the tumor samples into R1 and R2 subtypes (Table S8, 16). The reclassification resulted in R1 (n = 56) and R2 (n = 26) in the HG cohort. The R1 subtype exhibited significantly poorer recurrence-free survival (RFS) in the HG cohort (p = 0.018; log-rank test, Fig. 3C) and worse disease-free survival (DFS) in TCGA cohort (p =0.005; log-rank test, Fig. 3I) than the R2 subtype. The same trend was also observed in another independent validation cohort comprising more than 200 early-stage LUAD cases (Figure S3 A). We subsequently analyzed the WMS data from 98 patients to explore the methylation



Fig. 3 Molecular subtyping of the HG cohort based on transcriptomic and methylation data. A Volcano plot of differentially expressed genes (DEGs) between tumor (T) and normal samples (N). DEGs were identified using DESeq2 with an adjusted *p*-value (FDR) < 0.05 and absolute log2 fold change > 1. A total of 3725 DEGs were identified, with 2197 upregulated and 1528 downregulated in tumors.14345 genes non-significant.
B Heatmap of clustering consistency in 82 tumor samples based on the top 3500 MAD DEGs. Tumor samples were clustered into two groups (R1 and R2) based on their gene expression patterns (detailed in method). C Kaplan–Meier survival curves of recurrence-free survival (RFS) for the single-sample transcriptomic classifier predicted two subtypes R1 (red) and R2 (blue). Hazard ratio, 95% CI, as well as *p* value were displayed. D Heatmap comparing methylation levels in the promotor region of 55,638 gene between tumor (T) and normal samples (N). Average methylation levels of all CpG sites within ± 2 kb of the transcript start site for each gene are visualized, ranging from full methylation (depicted in red) to no methylation (in blue). E Heatmap of methylation clustering consistency in 98 tumor samples. Tumor samples were clustered into two groups (M1 and M2) based on their methylation profiles. F Kaplan–Meier survival curves of RFS for the methylation-clustering based subtypes, and relapse status. H Kaplan–Meier survival curves of RFS for combined RNA and methylation subtypes. Based on the results in G, the R1 group was further divided into C1 (R1 & M2) and C2 (R1 & M1), and the original R2 group was labeled as C3. The C3 group had the best RFS, followed by C2 and C1. I Kaplan–Meier survival curves of DFS for the single-sample transcriptomic classifier predicted subtypes in the TCGA early-stage cohort.

subtyping of early-stage poorly differentiated LUAD. Figure 3D shows methylation levels of all gene promoter regions (TSS  $\pm 2$  kb) between tumor and normal tissues (Table S17). Unsupervised consensus clustering based on the top 1000 gene promoter regions was used to identify methylation subtypes, and two subtypes were identified: M1 (n= 67) and M2 (n= 31) (Fig. 3E). The M2 subtype exhibited a significantly inferior RFS compared to the M1 subtype (p= 0.029; log-rank test; Fig. 3F).

The emergence of cancer necessitates molecular alterations in both the transcriptome and epigenome. We further leveraged data from 78 patients who underwent both RNA-seq and WMS to develop integrative subtyping. As shown in Fig. 3G, tumors classified under the R1 subtype were further divided into two subsets based on methylation profiling, termed C1 (26/53, R1&M2) and C2 (27/53, R1&M1) in integrative subtyping. Most tumors predicted to be of the R2 subtype were enriched in the M1 subtype (23/25, 92.00%), with only four (16%) recurring after surgery; thus, the R2 subtype was designated as the C3 subtype in integrative subtyping. Consequently, the 78 patients included in integrative subtyping were classified into three groups: C1 (n = 26), C2 (n = 27), and C3 (n = 25). Survival analysis revealed significant differences in prognosis among the three subtypes, with C1 subtype patients exhibiting the shortest recurrence-free survival (RFS) and C3 subtype patients demonstrating the longest RFS (p = 0.024; log-rank test; Fig. 3H, Table S8).

Univariate Cox regression analysis of the selected clinical features, as shown in Figure S3B, identified age, tumor size, STAS, VPI, and pT stage as significant prognostic factors (all p < 0.05). In multivariate Cox regression analysis, integrative subtyping remained an independent prognostic indicator for recurrence-free survival (RFS) (p <

(See figure on next page.)

0.05, Figure S3 C). We then performed a time-dependent receiver operating characteristic (ROC) analysis to assess the predictive accuracy for recurrence at 1, 3 and 5 years using a Cox model that incorporated integrative subtyping along with other prognostic clinical features. The predictive accuracy, indicated by the area under the curve (AUC), was approximately 0.7 when using integrative subtyping alone or in combination with another clinical feature (Figure S3D). Further analysis combining integrative subtyping with the five aforementioned clinical factors yielded AUC values of 0.74, 0.80, and 0.74 for the first, third, and fifth years, respectively (Figure S3E). These results suggest that the integrative model can more effectively predict the progression of early-stage poorly differentiated lung adenocarcinomas (LUADs), particularly at the 3-year mark. Notably, integrative subtyping emerged as the most significant predictor of RFS, accounting for 38.9% of the variance based on ANOVA of the Cox proportional hazards model (Figure S3 F), which represents the largest Chi-square value relative to the total Chi-square values of all variables included in the model.

# Clinicopathological and molecular characteristics of the three integrative subtypes

The three integrative subtypes identified using RNA-seq and WMS data exhibited different prognoses, suggesting that they are different disease subsets. Therefore, we further investigated their characteristics in depth. We initially compared the clinicopathological characteristics of the three integrative subtypes. As shown in Fig. 4A, the C3 subtype included more males (p = 0.005, Fisher's exact test) and smokers (p = 0.014, Fisher's exact test) than the C2 subtype, but there was no significant difference

Fig. 4 Comparison of the clinicopathological and molecular characteristics of the three integrative subtypes. A Three subtypes C1 (red), C2 (yellow) and C3 (blue) were compared based on clinicopathological factors. B Comparison of molecular features, including TMB, MATH, aneuploidy, ploidy (WES), PAscore (WMS) and HLA-LOH status among the three subtypes C1, C2 and C3. C, D Comparison of top 50 mutated genes (SNV) and copy number variations (CNV). The left panel (C) shows the mutation frequency of the top 50 mutated genes in each subgroup (C1, C2 and C3). The y-axis represents the genes, and the x-axis represents the frequency (red: mutation, blue: wild type (WT)). Genes are sorted by mutation frequency from highest to lowest. The right panel (D) shows the comparison of gene CNV status between subgroups. The y-axis represents the genes, and the x-axis represents the frequency. Genes with a total copy number greater than the gene-level median ploidy were considered gains (light red). Genes with more than twice the median ploidy were considered amplifications (red). Genes with less than the median ploidy were considered losses (light blue). Genes with a total copy number of 0 were considered deletions (blue). Genes with no copy number variants were set to none (white)."Group comparison "denotes comparisons between subgroups:"All" represents an overall comparison to assess whether the presence of mutations in genes differs across the three subgroups, while"C1 vs C2,"etc., represents whether mutation frequencies differ between two specific subgroups. E The box plot showing statistical differences in immune cell enrichment score among three subtypes C1, C2 and C3. The enrichment scores of 28 immune cell subsets within the tumor microenvironment was quantitatively assessed using the GSVA R package and the ssGSEA method. The x-axis represents immune cell subtypes, and the y-axis represents the enrichment score of each subtype. F-I The box plot showing a statistical difference in GEP score (detailed in the Methods), PD-L1 (CD274) gene expression, total tumor immune score and stromal score (estimated by ESTIMATE, detailed in the Methods) among three subtypes C1, C2 and C3. Note: To calculate p-values between molecular subtypes, fisher's test was utilized for categorical variables, and Wilcoxon test for continuous variables. ns: p > 0.05 (not marked on the figures A, B and E), \*: 0.01 < p < = 0.05, \*\*: 0.001 < p < = 0.01, \*\*\*: p < = 0.001



Fig. 4 (See legend on previous page.)

between subtypes C1 and C2 and subtypes C1 and C3. In terms of pT stage, the C1 subtype enriched more patients with T2 or T3 stage disease than the C3 subtype (p = 0.013, Fisher's exact test). Additionally, tumor size decreased in subtypes C1, C2, and C3, but the difference was not statistically significant. Other clinical characteristics, including age, STAS, VPI, LVI, pTNM stage, and adjuvant therapy, showed no significant differences among the three subtypes (Fig. 4A, Figure S4 A). We further explored whether the three integrative subtypes have significantly different molecular characteristics. As shown in Fig. 4B, we observed increased TMB and MATH in the C1 subtype (all p < 0.05, Wilcoxon test), indicating that patients exhibiting the C1 subtype could harbor more mutations and intra-tumor heterogeneity (ITH). Parameters related to chromosome stability, such as aneuploidy, ploidy, FGA, MSI, CNI, and PA score, were also higher in the C1 subtype (all p < 0.05, Wilcoxon test; Fig. 4B and Figure S4B). Additionally, we observed that the C1 subtype was significantly more likely to exhibit HLA loss of heterozygosity (HLA-LOH) and a higher tumor neoantigen burden (TNB) than subtypes C2 and C3 (all p < 0.05, Fisher's exact test). Increased HLA-LOH levels are usually accompanied by tumor immune evasion. Together, these results indicate that the C1 subtype has significant genomic instability, mutation burden, and ITH.

To further investigate molecular heterogeneity among the three integrative subtypes, we compared the SNV and CNV profiles, focusing on the top 50 most frequently mutated genes and CNVs (Table S2, Table S6). While frequently mutated genes, such as *EGFR*, *TP53*, *KRAS*, *RYR2*, and *TTN*, showed similar mutation frequencies across subtypes, several other genes showed significant differences (Fig. 4C). For example, *SEC16B* and *FSIP2* mutations were more common in the C1 subtype, whereas *AHNAK2* mutations were enriched in the C3 subtype (p < 0.05, Fisher's exact test). Additionally, *STK11* mutations were more common in the C1 subtype than in the C3 subtype (p < 0.05, Fisher's exact test). Next, we compared the CNV among these three subtypes and found that deletions of many pivotal cancer-associated genes were more common in the C1 subtype, including *CDKN2 A*, *MLLT3*, *CD274*, *JAK2*, *NOTCH1* and *NFIB* (Fig. 4D). We further compared 10 canonical oncogene signaling pathways and the DDR pathway between these three subtypes (Table S18). We found that the C1 subtype showed a higher mutational rate for the Notch and BER pathways than the C2 subtype (p < 0.05), and the C2 subtype showed a higher mutational rate for the PI3 K pathway than the C3 (p < 0.05, Figure S4 C).

Based on the RNA-seq data, analysis of the enrichment scores for 28 distinct immune cell subsets among the three integrative subtypes (Table S11) revealed significant disparities in the majority of immune cell types between these subtypes (Fig. 4E). Notably, there was a progressive increase in the proportions of nearly all these cell types from subtype C1 to C2 and further to C3. For instance, the C3 subtype tumors exhibited the highest abundance of various immune effector cells, including activated CD8 + T cells, natural killer cells, and mast cells (all p < 0.05, Wilcoxon test), compared to the C1 and C2 subtype tumors. Furthermore, our analysis revealed that the C3 subtype demonstrated elevated levels of CD274 (PD-L1) expression and a higher Gene Expression Profiling (GEP) score (Fig. 4F, G; Table S19). Notably, both the immune and stromal scores, as estimated by the ESTIMATE algorithm, were significantly higher in subtype C3 than subtypes C1 and C2 (Fig. 4H, I). These findings suggest that the C3 subtype is characterized by an immunologically active tumor microenvironment (TME). Additionally, an analysis of functional states across 25 cancer types revealed significant enrichment of cell cycle and DNA repair functions in subtype C1, whereas subtype C3 was significantly enriched for guiescence and stemness-related functions (Figure S4D; Table S20).

# The exploration of representative differentially expressed genes in the three integrative subtypes

Next, we identified DEGs that were stably expressed across the three integrative subtypes (Figure S5A; Table S21). To evaluate the prognostic significance of

<sup>(</sup>See figure on next page.)

**Fig. 5** Exploration the biological functions of GINS1 and CPT1C by in vitro and in vivo experiments. **A**, **B** Box plots showing statistical differences in GINS1 (**A**) and CPT1C (**B**) gene expression among normal and three subtypes (C1, C2 and C3) (**C**, **D**) Kaplan–Meier survival analysis of RFS comparing high and low expression levels of GINS1 (**C**) and CPT1C (**D**) gene (cutoff for each gene is detailed in the Methods). Hazard ratio, number of patients in each group, as well as *p* value are displayed. **E**, **G** The GINS1 (**E**) and CPT1C (**G**) knockdown efficiency was analyzed by western blotting in SPCA-1 cells. **F**, **H** Transwell assay was performed to evaluate the effect of GINS1 (**F**) and CPT1C (H) depletion on migration and invasion ability of SPCA-1 cells (*n* = 3). Representative images (left) and quantitative results (right). Scale bar, 200 µm. \*\*\*: *p* < =0.001. **I**, **K** Colony formation assay was performed to evaluate the effect of GINS1 (J) and CPT1C (L) depletion on cell viability of SPCA-1 cells (*n* = 3). \*\*\*: 0.001 < *p* < = 0.01, \*\*\*: *p* < =0.001. **J**, **L** CCK-8 assay was performed to evaluate the effect of GINS1 (J) and CPT1C (L) depletion on cell viability of SPCA-1 cells (*n* = 3). \*\*\*: *p* < =0.001. **M**, **N** Images (left) and weight (right) of xenograft tumors derived from control SPCA-1 cells and SPCA-1 cells with GINS1 (M) or CPT1C (N) depletion (*n* = 5). \*\*\*: *p* < =0.001.

highly and stably expressed genes in the C1 subtype, we calculated the p-values and hazard ratios (HRs) for each gene across percentiles from 0.1 to 0.9 in both the HG and TCGA cohorts (Figure S5B, C; Table S22). Among these genes, GINS complex subunit 1 (GINS1) and Carnitine palmitoyltransferase 1C (CPT1C) were of particular interest, because their high expression was significantly associated with poor prognosis in both cohorts (Fig. 5A, B). Previous studies have highlighted their pivotal roles in promoting tumor progression [60, 61]. We utilized the surv\_cutpoint function from the survminer R package to determine the optimal cutoff for each gene and found that higher expression levels of both genes were associated with shorter recurrence-free survival (RFS) in the HG cohort (GINS1: HR = 2.33, p = 0.050; CPT1C: HR =4.68, *p* = 0.002, Fig. 5C, D).

To further explore the biological functions of GINS1 and CPT1C in LUAD progression, we knocked down the expression of endogenous GINS1 and CPT1C in SPCA-1 cells using lentiviruses encoding shRNAs. We found that the downregulation of both GINS1 and CPT1C markedly attenuated cell migration and invasion in vitro (Fig. 5E-H and Figure S5D, E). In addition to inhibiting pro-metastatic function, the knockdown of GINS1 and CPT1C strikingly restrained cell proliferation, as demonstrated by the reduced proliferative ability in CCK- 8 and colony formation assays (Fig. 5I-L and Figure S5 F, G). We further performed the xenograft tumor model to consolidate the biological functions of GINS1 and CPT1C in vivo. As shown in Fig. 5M and N, compared to the control group, the downregulation of GINS1 and CPT1C markedly attenuated tumor progression in vivo, accompanied by a marked reduction in tumor size and weight.

# Potential prognostic and predictive significance of the integrative subtyping

In current clinical practice, adjuvant chemotherapy remains a cornerstone in the treatment of LUAD; however, its use in early-stage LUAD remains controversial. In the HG cohort, 37 patients received adjuvant chemotherapy after surgery. These patients tended to exhibit a worse prognosis than those who did not receive postoperative treatment, although the difference showed only a marginal statistical significance (p = 0.059, Fig. 6A). We explored the potential of integrative subtyping as a possible indicator for assessing adjuvant chemotherapy efficacy. As shown in Fig. 6B-D, the survival analysis suggested limited benefit from adjuvant chemotherapy in both C1 and C3 subtypes, with the C3 subtype exhibiting a particularly unfavorable outcome (p = 0.038). In contrast, a trend toward improved prognosis was observed in patients classified as the C2 subtype who received adjuvant chemotherapy, although this difference did not reach statistical significance (p = 0.301). These preliminary findings indicated that integrative subtyping might offer some clinical utility in evaluating heterogeneous responses to adjuvant chemotherapy in early-stage poorly differentiated LUAD, warranting further investigation to assess its potential value in therapeutic decision-making.

In summary, each integrative subtype exhibits unique clinical features, molecular characteristics, and TME, as shown in Fig. 6E. The C1 subtype is characterized by high genomic instability, low immune infiltration and high tumor heterogeneity. The C2 subtype displays low tumor heterogeneity and HLA-LOH, fewer neoantigens and moderate immune activation, and it may benefit from adjuvant chemotherapy. The C3 subtype exhibits substantial immune infiltration, high GEP score, high TNB, and low frequency of HLA-LOH, suggesting a potential benefit from immunotherapy.

# Discussion

Early-stage poorly differentiated LUADs exhibit a poor prognosis, with approximately 30% of them experiencing recurrence [3]. Precise management of this population is a challenge in clinical practice. Adding molecular dimensional features and stratification to the new grading system could render the prognostic evaluation more precise. In this study, we performed integrative multi-omics analysis, including genomics, epigenomics and transcriptomics, in an early-stage poorly differentiated lung adenocarcinoma cohort (n = 101). Our analysis of the multi-platform dataset revealed comprehensive molecular characteristics of this special disease entity. Furthermore, three molecular subtypes were identified based on the transcriptomic and methylation data. These integrative subtypes demonstrated distinct clinicopathological and molecular features with prognostic significance, enabling further precise stratification of early-stage poorly differentiated LUADs.

Consistent with previous genomic studies in the Chinese population [53, 62], *EGFR* (51%) and *TP53* (45%) were also the predominant mutated genes in early-stage poorly differentiated LUAD. However, the mutational frequencies of *MUC*16 (28%), *RYR2* (28%) and *TTN* (26%) were significantly higher in early-stage poorly differentiated LUAD than in the unselected Chinese LUAD patients. *MUC16* mutations have been found in several solid tumors, including melanoma and breast cancer [63–65]. Several recent studies have revealed that *MUC16* mutations are associated with TMB in solid tumors [66]. A recent study reported that *RYR2* was one of the most mutational genes which were almost always shared by primary lung cancers and brain metastasis lesions [67]. Previous studies have revealed that the



Fig. 5 (See legend on previous page.)



Fig. 6 Potential prognostic and predictive significance of the integrative subtyping. A Kaplan–Meier analysis of RFS in patients who received adjuvant therapy compared to those who did not. B-D Evaluation of the Impact of adjuvant therapy on RFS within different subgroups defined by the combined subtype (C1, C2 and C3). E Summary of clinicopathological and molecular features for each subtype, with potential therapeutic implications

mutation of *RYR2* was a significant biomarker associated with high TMB in LUAD [68]. As another gene with high-frequency mutations, *TTN* mutations were associated with the therapeutic efficacy of immune checkpoint blockade in advanced non-small cell lung cancer [69, 70]. Moreover, *TTN* mutations may function as crucial intratumoral drivers for micropapillary/solid components, as shown in another LUAD cohort [71]. In addition to oncogenic gene mutations, remarkable CNVs were also found in early-stage poorly differentiated LUADs. These results indicated that poorly differentiated LUADs had relatively special mutational landscape and chromosome structure variations, which may lead to the transition of growth patterns during the development of LUAD.

Accurate risk classification plays a pivotal role in guiding treatment strategies and enhancing patient outcomes. While the current grading system categorizes poorly differentiated LUADs as a homogeneous group, our study revealed three distinct molecular subtypes through transcriptomic and methylation profiling. These subtypes exhibit unique molecular signatures that correlate with specific clinical, pathological, and prognostic characteristics. The C1 subtype displayed the highest levels of TMB, MATH, aneuploidy and HLA-LOH levels, coupled with relatively reduced immune cell infiltration. These factors likely contribute to increased genomic instability and impaired anti-tumor immune responses and lead to its poor prognosis. Multiple studies have demonstrated that high TMB is associated with a poor prognosis of LUADs [72]. Aneuploidy is a hall mark of cancer, and patients with high aneuploidy often show a poor prognosis [73]. Lower PD-L1 expression further implies that these tumors might rely less on the PD-L1/PD- 1 pathway for immune escape, instead of utilizing alternative mechanisms such as HLA-LOH. In contrast to the C1 subtype, the C3 subtype exhibited lower genomic instability, as evidenced by reduced TMB, MATH and FGA. Notably, this subtype also demonstrated the highest level of immune cell infiltration, which may contribute to its more favorable prognosis. Furthermore, the elevated immune infiltration and higher GEP scores suggest that this subtype may be particularly responsive to immunotherapy, making it a promising candidate for such treatment strategies. For the C2 subtype, its molecular characteristics was intermediate and featured with low heterogeneity, low LOH and less neo antigens. In this study, we found the C1 and C3 subtypes did not benefit from adjuvant chemotherapy, especially for the later one. However, patients with C2 subtype receiving adjuvant chemotherapy tended to exhibit better prognosis, implying adjuvant chemotherapy was a potential treatment modality for the C2 subtype. However, these are preliminary findings that require further validation in an Page 20 of 24

additional large-scale cohort. We will continue to accumulate more data on this aspect and validate the results through both retrospective and prospective cohort studies. Currently, several molecular classification systems exist for unselected LUAD patients, yet they remain largely disconnected from pathological diagnostics. The integration of classical morphological information and molecular data remains a severe challenge for clinical utility. The integrative subtypes identified in this study offer a practicable approach to further enhance prognostic evaluation in poorly differentiated LUAD, building upon the current pathological diagnostic framework. To identify early-stage LUAD with a high recurrence risk, we recommend first using the histological grading system to select poorly differentiated cases, then performing molecular subtyping exclusively on these tumors. This stratified approach will facilitate more precise prognostic assessment and guide adjuvant therapy decisions.

Given the pressing demand for innovative treatment methods, the DEGs identified among the three integrative subtypes were selected for further studies to illustrate their potential as new therapeutic targets. Our study found that GINS1 and CPTIC exhibited high expression in the C1 subtype and were associated with a higher hazard ratio of recurrence in both the HG and TCGA cohorts, acting as potential therapeutic targets. GINS1, part of helicase at DNA replication forks, is involved in DNA replication initiation and elongation [74, 75]. Previous studies have shown that GINS1 plays essential roles in tumorigenesis and progression, including NSCLC [60]. In addition to DNA replication, GINS1 is associated with tumor metastasis, sorafenib resistance, doxorubicin resistance, and B-cell proliferation, suggesting essential role of GINS1 in tumor progression, drug resistance and immune microenvironment [75, 76]. CPT1C, a rate-limiting enzyme in fatty acid oxidation, fuels tumor growth under metabolic stress and acts as a prognostic marker in many tumors [77, 78]. Several studies have reported that CPT1C, a regulator of lipid metabolic reprogramming, is pivotal for the proliferation and metastasis of various tumors [61, 79]. Moreover, dysregulation of CPT1C can lead to plasma membrane remodeling and anthracycline resistance in breast cancer [80]. Collectively, all the evidence showed that both GINS1 and CPT1C are pivotal for tumor progression and are associated with drug resistance, which is consistent with the pro-proliferative and pro-metastatic functions of GINS1/CPT1C in our studies and indicates potential drug targets to enhance adjuvant chemotherapy in LUADs.

Moreover, our studies found that downregulated genes in both GINS1-high expression and C1 subtype samples were significantly enriched in pathways including"Neutrophil degranulation", "Signaling by Interleukins", "Arachidonic acid metabolism" and "PD-1 signaling"(Figure S6 A, B; Table S23). These findings suggest that GINS1 may promote tumor progression by suppressing T cell and neutrophil functions, thereby inducing an immunosuppressive microenvironment and contributing to the aggressive phenotype of the C1 subtype [81-83]. Notably, arachidonic acid metabolism, a critical metabolic pathway in tumor progression, has been shown to enhance antitumor immune responses by activating CD8 +T cells in colorectal cancer [84]. Therefore, subsequent studies on elucidating how GINS1 modulates arachidonic acid metabolism to influence C1 subtype transformation, may provide a stronger theoretical foundation for developing therapeutic strategies targeting arachidonic acid metabolism inhibitors. Additionally, we observed that downregulated genes in both CPT1C-high and C1 subtype samples were significantly enriched in pathways such as"Chemokine receptors bind chemokines","Neutrophil degranulation","Phosphorylation of CD3 and TCR zeta chains","PD- 1 signaling"and"Biosynthesis of specialized proresolving mediators (SPMs)"(Figure S6 A, C; Table S23). These results suggest that CPT1C may similarly promote tumor progression by creating an immunosuppressive microenvironment that contributes to the aggressive phenotype of the C1 subtype [82, 83, 85, 86]. Of particular interest, SPMs, a class of bioactive lipid mediators derived from polyunsaturated fatty acids (PUFAs), play crucial roles in actively resolving inflammation and inducing surrounding immune-infiltrating cells such as tumor-associated macrophages, representing a promising direction for anticancer therapy [87, 88]. While CPT1C is known to participate in fatty acid oxidation [77], its specific impact on SPMs and the function of SPMs in early-stage poorly differentiated lung adenocarcinoma remain unclear. Therefore, future studies may focus on elucidating the mechanism by which CPT1C regulates SPMs, with the goal of developing novel SPMbased therapeutic strategies.

While this study provides some valuable insights, several limitations should be acknowledged. First, we actually made substantial efforts to screen early-stage LUAD from our institution between 2012 and 2017, and screened out 101 poorly differentiated cases. Although this is the largest multi-omics cohort of early-stage poorly differentiated LUAD with long-term follow-up to date, the sample size remains limited. This constrained statistical power may explain why certain clinically apparent trends—particularly regarding differential responses to adjuvant chemotherapy across the three subtypes—failed to reach statistical significance. Consequently, these preliminary findings warrant validation through largerscale prospective studies. Second, owing to the scarcity of public multi-omics data encompassing transcriptomic and methylation profiles for poorly differentiated LUAD, we could not externally validate our integrative molecular subtyping system or its prognostic value. Future multi-center studies focusing on early-stage poorly differentiated LUAD cohorts will be essential to verify the clinical relevance and generalizability of our classification. Third, we performed comprehensive and in-depth analyses of the molecular characteristics of poorly differentiated LUAD. However, we did not profile the proteome or metabolome, which could reflect the biological characteristics of this special disease entity more directly. Therefore, we will perform proteomic and metabolomic analyses on poorly differentiated LUAD in the future to establish a more comprehensive and reliable subtyping.

In summary, through integrated analyses, we delineate a genomic, epigenomic and transcriptomic landscape and uncover three molecular subtypes with distinct prognoses of early-stage poorly differentiated LUAD. Our study not only provides important insight into the relatively specific biology of this subset of LUAD but also reveals its molecular heterogeneity, which can potentially facilitate their precise treatment and postoperative monitoring.

### Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12943-025-02333-7.

Supplementary Material 1. Supplementary Material 2.

#### Acknowledgements

We thank all the participants and family members for participating in this study.

### Authors' contributions

BL, WT, XTZ, LDX, SY and NW contributed to the conception and design of the study. BL, WT, XTZ, YRL, LDX, XY, ELZ and YSH acquired and analyzed the data. MH, YGZ, XRC, YQW and TYG provided patient materials and scientific discussion. BL, WT, XTZ and NW drafted the initial version of the Stage 1 protocol. All authors contributed to revisions of the Stage 1 protocol, preparation of the Stage 2 manuscript. All authors reviewed and approved the submitted version.

#### Funding

This study was supported by the Capital's funds for health improvement and research (2024–1 - 1023), the National Natural Science Foundation of China (No. 82303583, 82373082), the Science Foundation of Peking University Cancer Hospital (2022–11), the National Key R&D Program of China (No. 2022YFC2406804), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501702), the National Ten-thousand Talent Program, and the First-Class Discipline Team of Kunming Medical University (2024XKTDYS07).

#### Data availability

The raw data can be obtained from the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/ Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number HRA009302 (https://ngdc.cncb.ac.cn/gsa). To protect patient privacy, data access can be obtained through a request to the data access committee. Access to the data will be restricted to noncommercial entities.

# Declarations

#### Ethics approval and consent to participate

This study was approved by the Ethics Committee of Peking University Cancer Hospital & Institute (Institutional Review Board No. 2024KT65).

#### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Thoracic Surgery II, Peking University Cancer Hospital & Institute, Beijing 100142, China. <sup>2</sup>Genecast Biotechnology Co, Ltd, Wuxi, Jiangsu, China. <sup>3</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Pathology, Peking University Cancer Hospital & Institute, Beijing, China. <sup>4</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, China. <sup>5</sup>State Key Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, China. <sup>5</sup>State Key Laboratory of Molecular Oncology, Frontiers Science Center for Cancer Integrative Omics, Department of Thoracic Surgery II, Beijing Key Laboratory of Carcinogenesis and Translational Research, Peking University Cancer Hospital & Institute, Beijing 100142, China. <sup>6</sup>Yunnan Cancer Hospital, The Third Affiliated Hospital of Kunming Medical University, Peking University Cancer Hospital Yunnan, Yunnan, China.

### Received: 22 February 2025 Accepted: 12 April 2025 Published online: 01 May 2025

#### References

- Moreira AL, Ocampo PSS, Xia Y, Zhong H, Russell PA, Minami Y, Cooper WA, Yoshida A, Bubendorf L, Papotti M, et al. A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the study of lung cancer pathology committee. J Thorac Oncol. 2020;15:1599–610.
- Ruan Y, Cao W, Han J, Yang A, Xu J, Zhang T. Prognostic impact of the newly revised IASLC proposed grading system for invasive lung adenocarcinoma: a systematic review and meta-analysis. World J Surg Oncol. 2024;22:302.
- Xu L, Su H, Hou L, Wang F, Xie H, She Y, Gao J, Zhao S, Dai C, Xie D, et al. The IASLC proposed grading system accurately predicts prognosis and mediastinal nodal metastasis in patients with clinical stage I lung adenocarcinoma. Am J Surg Pathol. 2022;46:1633–41.
- Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50.
- Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, Alvarez JJS, Lu B, Lim JQ, Takano A, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet. 2020;52:177–86.
- Fukui T, Shaykhiev R, Agosto-Perez F, Mezey JG, Downey RJ, Travis WD, Crystal RG. Lung adenocarcinoma subtypes based on expression of human airway basal cell genes. Eur Respir J. 2013;42:1332–44.
- Burdett NL, Willis MO, Alsop K, Hunt AL, Pandey A, Hamilton PT, Abulez T, Liu X, Hoang T, Craig S, et al. Multiomic analysis of homologous recombination-deficient end-stage high-grade serous ovarian cancer. Nat Genet. 2023;55:437–50.
- Davidson NR, Barnard ME, Hippen AA, Campbell A, Johnson CE, Way GP, Dalley BK, Berchuck A, Salas LA, Peres LC, et al: Molecular Subtypes of High-Grade Serous Ovarian Cancer across Racial Groups and Gene Expression Platforms. Cancer Epidemiol Biomarkers Prev 2024, 33:1114-1125.
- Linazi G, Maimaiti A, Abulaiti Z, Shi H, Zhou Z, Aisa MY, Kang Y, Abulimiti A, Dilimulati X, Zhang T, et al. Prognostic value of anoikis-related genes revealed using multi-omics analysis and machine learning based on lower-grade glioma features and tumor immune microenvironment. Heliyon. 2024;10:e36989.

- Deng C, Zheng Q, Zhang Y, Jin Y, Shen X, Nie X, Fu F, Ma X, Ma Z, Wen Z, et al. Validation of the novel international association for the study of lung cancer grading system for invasive pulmonary adenocarcinoma and association with common driver mutations. J Thorac Oncol. 2021;16:1684–93.
- Caso R, Sanchez-Vega F, Tan KS, Mastrogiacomo B, Zhou J, Jones GD, Nguyen B, Schultz N, Connolly JG, Brandt WS, et al. The underlying tumor genomics of predominant histologic subtypes in lung adenocarcinoma. J Thorac Oncol. 2020;15:1844–56.
- Dong ZY, Zhang C, Li YF, Su J, Xie Z, Liu SY, Yan LX, Chen ZH, Yang XN, Lin JT, et al. Genetic and immune profiles of solid predominant lung adenocarcinoma reveal potential immunotherapeutic strategies. J Thorac Oncol. 2018;13:85–96.
- Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT: The Eighth Edition Lung Cancer Stage Classification. Chest 2017;151:193–203.
- Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, et al. The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10:1243–60.
- Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012;72:100–11.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
- 17. Aldana R, Freed D. Data processing and germline variant calling with the Sentieon Pipeline. Methods Mol Biol. 2022;2493:1–19.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.
- 21. Mroz EA, Tward AD, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. PLoS Med. 2015;12:e1001786.
- Arora A, Shen R, Seshan VE. FACETS: fraction and allele-specific copy number estimates from tumor sequencing. Methods Mol Biol. 2022;2493:89–105.
- Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12:R41.
- Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell. 2018;33:676-689.e673.
- Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from Targeted DNA sequencing. PLoS Comput Biol. 2016;12:e1004873.
- Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, Quan W Jr, Sangal A, Khemka V, Waypa J, et al. Tumor cell-free DNA copy number instability predicts therapeutic response to immunotherapy. Clin Cancer Res. 2017;23:5074–81.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
- Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat. 2017;38:788–97.
- McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, Birkbak NJ, Veeriah S, Van Loo P, Herrero J, Swanton C. Allele-Specific HLA loss and immune escape in lung cancer evolution. Cell. 2017;171:1259-1271.e1211.
- Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetM-HCpan-4.0: Improved Peptide-MHC Class I Interaction predictions

integrating eluted ligand and peptide binding affinity data. J Immunol. 2017;199:3360–8.

- Chen K, Yang A, Carbone DP, Kanu N, Liu K, Wang R, Nie Y, Shen H, Bai J, Wu L, et al. Spatiotemporal genomic analysis reveals distinct molecular features in recurrent stage I non-small cell lung cancers. Cell Rep. 2022;40:111047.
- Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
- Scire J, Huisman JS, Grosu A, Angst DC, Lison A, Li J, Maathuis MH, Bonhoeffer S, Stadler T. estimateR: an R package to estimate and monitor the effective reproductive number. BMC Bioinformatics. 2023;24:310.
- Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest. 2017;127:2930–40.
- Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18:248–62.
- Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z, et al. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. 2019;47:D900-d908.
- 41. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
- 42. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.
- Bie F, Wang Z, Li Y, Guo W, Hong Y, Han T, Lv F, Yang S, Li S, Li X, et al. Multimodal analysis of cell-free DNA whole-methylome sequencing for cancer detection and localization. Nat Commun. 2023;14:6042.
- 44. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, Gydush G, Reed SC, Rotem D, Rhoades J, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun. 2017;8:1324.
- 45. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, O'Shaughnessy J, Kinzler KW, Parmigiani G, Vogelstein B, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med. 2012;4:162ra154.
- Feng H, Conneely KN, Wu H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res. 2014;42:e69.
- Yu G, Wang LG, He QY. ChlPseeker: an R/Bioconductor package for ChlP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3.
- Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin. 2018;11:37.
- Xu T, Le TD, Liu L, Su N, Wang R, Sun B, Colaprico A, Bontempi G, Li J. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics. 2017;33:3131–3.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
- Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, Jensen JB, Strandgaard T, Nordentoft I, Christensen E, et al. An integrated multi-omics analysis identifies prognostic molecular

- 52. Ando M, Saito Y, Xu G, Bui NQ, Medetgul-Ernar K, Pu M, Fisch K, Ren S, Sakai A, Fukusumi T, et al. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. Nat Commun. 2019;10:2188.
- Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S, et al. Integrative proteomic characterization of human lung adenocarcinoma. Cell. 2020;182:245-261.e217.
- Hudson KM, Klimczak LJ, Sterling JF, Burkholder AB, Kazanov MD, Saini N, Mieczkowski PA, Gordenin DA. Glycidamide-induced hypermutation in yeast single-stranded DNA reveals a ubiquitous clock-like mutational motif in humans. Nucleic Acids Res. 2023;51:9075–100.
- Kim Y, Lee B, Shim JH, Lee SH, Park WY, Choi YL, Sun JM, Ahn JS, Ahn MJ, Park K. Concurrent genetic alterations predict the progression to target therapy in EGFR-Mutated Advanced NSCLC. J Thorac Oncol. 2019;14:193–202.
- Duan J, Xu J, Wang Z, Bai H, Cheng Y, An T, Gao H, Wang K, Zhou Q, Hu Y, et al. Refined stratification based on baseline concomitant mutations and longitudinal circulating tumor DNA monitoring in advanced egfrmutant lung adenocarcinoma under gefitinib treatment. J Thorac Oncol. 2020;15:1857–70.
- Takamochi K, Oh S, Matsunaga T, Suzuki K. Prognostic impacts of EGFR mutation status and subtype in patients with surgically resected lung adenocarcinoma. J Thorac Cardiovasc Surg. 2017;154:1768-1774.e1761.
- Zheng J, Deng Y, Huang B, Chen X. Prognostic implications of STK11 with different mutation status and its relationship with tumor-infiltrating immune cells in non-small cell lung cancer. Front Immunol. 2024;15:1387896.
- He Y, Chen X, Liu H, Xiao H, Kwapong WR, Mei J. Matrix-remodeling associated 5 as a novel tissue biomarker predicts poor prognosis in non-small cell lung cancers. Cancer Biomark. 2015;15:645–51.
- Li M, Shi M, Hu C, Chen B, Li S. MALAT1 modulated FOXP3 ubiquitination then affected GINS1 transcription and drived NSCLC proliferation. Oncogene. 2021;40:3870–84.
- Chen Y, Zhou Y, Han F, Zhao Y, Tu M, Wang Y, Huang C, Fan S, Chen P, Yao X, et al. A novel miR-1291-ERRα-CPT1C axis modulates tumor cell proliferation, metabolism and tumorigenesis. Theranostics. 2020;10:7193–210.
- 62. Chen YJ, Roumeliotis TI, Chang YH, Chen CT, Han CL, Lin MH, Chen HW, Chang GC, Chang YL, Wu CT, et al. Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. Cell. 2020;182:226-244.e217.
- 63. Yu Y, Lin D, Li A, Chen Y, Ou Q, Hu H, Yao H. Association of immune checkpoint inhibitor therapy with survival in patients with cancers with MUC16 variants. JAMA Netw Open. 2020;3:e205837.
- 64. Li J, Liu B, Ye Q, Xiao X, Yan S, Guan W, He L, Wang C, Yu Z, Tai Z, et al. Comprehensive genomic analysis of primary malignant melanoma of the esophagus reveals similar genetic patterns compared with epitheliumassociated melanomas. Mod Pathol. 2022;35:1596–608.
- 65. Yin G, Liu L, Yu T, Yu L, Feng M, Zhou C, Wang X, Teng G, Ma Z, Zhou W, et al. Genomic and transcriptomic analysis of breast cancer identifies novel signatures associated with response to neoadjuvant chemotherapy. Genome Med. 2024;16:11.
- 66. Wang X, Yu X, Krauthammer M, Hugo W, Duan C, Kanetsky PA, Teer JK, Thompson ZJ, Kalos D, Tsai KY, et al. The Association of MUC16 Mutation with Tumor Mutation Burden and Its Prognostic Implications in Cutaneous Melanoma. Cancer Epidemiol Biomarkers Prev. 2020;29:1792–9.
- 67. Duan H, Ren J, Wei S, Yang Z, Li C, Wang Z, Li M, Wei Z, Liu Y, Wang X, et al. Integrated analyses of multi-omic data derived from paired primary lung cancer and brain metastasis reveal the metabolic vulnerability as a novel therapeutic target. Genome Med. 2024;16:138.
- Wang C, Liang H, Lin C, Li F, Xie G, Qiao S, Shi X, Deng J, Zhao X, Wu K, Zhang X: molecular subtyping and prognostic assessment based on tumor mutation burden in patients with lung adenocarcinomas. Int J Mol Sci 2019;20.
- Su C, Wang X, Zhou J, Zhao J, Zhao F, Zhao G, Xu X, Zou X, Zhu B, Jia Q. Titin mutation in circulatory tumor DNA is associated with efficacy to immune checkpoint blockade in advanced non-small cell lung cancer. Transl Lung Cancer Res. 2021;10:1256–65.
- Jia Q, Wang J, He N, He J, Zhu B: Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. JCI Insight 2019;4.

- Li J, Xiong S, He P, Liang P, Li C, Zhong R, Cai X, Xie Z, Liu J, Cheng B, et al. Spatial whole exome sequencing reveals the genetic features of highly-aggressive components in lung adenocarcinoma. Neoplasia. 2024;54:101013.
- Li L, Li J. Correlation of tumor mutational burden with prognosis and immune infiltration in lung adenocarcinoma. Front Oncol. 2023;13:1128785.
- 73. Lakhani AA, Thompson SL, Sheltzer JM. Aneuploidy in human cancer: new tools and perspectives. Trends Genet. 2023;39:968–80.
- Ilves I, Petojevic T, Pesavento JJ, Botchan MR. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell. 2010;37:247–58.
- Kingsley G, Skagia A, Passaretti P, Fernandez-Cuesta C, Reynolds-Winczura A, Koscielniak K, Gambus A. DONSON facilitates Cdc45 and GINS chromatin association and is essential for DNA replication initiation. Nucleic Acids Res. 2023;51:9748–63.
- Liang J, Yao N, Deng B, Li J, Jiang Y, Liu T, Hu Y, Cao M, Hong J. GINS1 promotes ZEB1-mediated epithelial-mesenchymal transition and tumor metastasis via β-catenin signaling in hepatocellular carcinoma. J Cell Physiol. 2024;239:e31237.
- Fadó R, Zagmutt S, Herrero L, Muley H, Rodríguez-Rodríguez R, Bi H, Serra D, Casals N. To be or not to be a fat burner, that is the question for cpt1c in cancer cells. Cell Death Dis. 2023;14:57.
- Li J, Zheng W, Wu J, Zhang J, Lv B, Li W, Liu J, Zhang X, Huang T, Luo Z. CPT1C-mediated fatty acid oxidation facilitates colorectal cancer cell proliferation and metastasis. Acta Biochim Biophys Sin (Shanghai). 2023;55:1301–9.
- Zhao H, Cheng X, Yan L, Mi F, Wang W, Hu Y, Liu X, Fan Y, Min Q, Wang Y, et al. APC/C-regulated CPT1C promotes tumor progression by upregulating the energy supply and accelerating the G1/S transition. Cell Commun Signal. 2024;22:283.
- Muley H, Valencia K, Casas J, Moreno B, Botella L, Lecanda F, Fadó R, Casals N: Cpt1c Downregulation Causes Plasma Membrane Remodelling and Anthracycline Resistance in Breast Cancer. Int J Mol Sci 2023, 24.
- Mollinedo F. Neutrophil Degranulation, Plasticity, and Cancer Metastasis. Trends Immunol. 2019;40:228–42.
- Shi RY, Zhou N, Xuan L, Jiang ZH, Xia J, Zhu JM, Chen KM, Zhou GL, Yu GP, Zhang J, et al. Trafficking circuit of CD8(+) T cells between the intestine and bone marrow governs antitumour immunity. Nat Cell Biol. 2024;26:1346–58.
- Lin X, Kang K, Chen P, Zeng Z, Li G, Xiong W, Yi M, Xiang B. Regulatory mechanisms of PD-1/PD-L1 in cancers. Mol Cancer. 2024;23:108.
- Cui L, Liu R, Han S, Zhang C, Wang B, Ruan Y, Yu X, Li Y, Yao Y, Guan X, et al. Targeting Arachidonic Acid Metabolism Enhances Immunotherapy Efficacy in ARID1A-Deficient Colorectal Cancer. Cancer Res. 2025;85:925–41.
- Wu W, Zhou Q, Masubuchi T, Shi X, Li H, Xu X, Huang M, Meng L, He X, Zhu H, et al. Multiple Signaling Roles of CD3ε and Its Application in CAR-T Cell Therapy. Cell. 2020;182:855-871.e823.
- Velasco Cárdenas RM, Brandl SM, Meléndez AV, Schlaak AE, Buschky A, Peters T, Beier F, Serrels B, Taromi S, Raute K, et al. Harnessing CD3 diversity to optimize CAR T cells. Nat Immunol. 2023;24:2135–49.
- Lavy M, Gauttier V, Poirier N, Barillé-Nion S, Blanquart C. Specialized proresolving mediators mitigate cancer-related inflammation: role of tumorassociated macrophages and therapeutic opportunities. Front Immunol. 2021;12:702785.
- Serhan CN. Pro-resolving lipid mediators are leads for resolution physiology. Nature. 2014;510:92–101.

# **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.