



OPEN ACCESS

EDITED BY

Vivek P. Chavda,
L. M. College of Pharmacy, India

REVIEWED BY

Yan Zou,
Henan University, China
Nasima Ahmed,
Dibrugarh University, India
Humzah Postwala,
L. M. College of Pharmacy, India

*CORRESPONDENCE

Jie Li

✉ lijie@kmmu.edu.cn
ChunLing Bai
✉ baichunling@kmmu.edu.cn
Jianhong Hou
✉ hjh@kust.edu.cn
Yang Chen
✉ 13836168124@163.com

[†]These authors have contributed equally to this work

RECEIVED 10 January 2025

ACCEPTED 23 May 2025

PUBLISHED 04 June 2025

CITATION

Li J, Pan X, Bi Y, Huang J, Peng J, Bai C, Hou J and Chen Y (2025) Comprehensive multi-omics analysis of tandem duplicator phenotypes in non-small cell lung cancer. *Front. Med.* 12:1556840. doi: 10.3389/fmed.2025.1556840

COPYRIGHT

© 2025 Li, Pan, Bi, Huang, Peng, Bai, Hou and Chen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive multi-omics analysis of tandem duplicator phenotypes in non-small cell lung cancer

Jie Li^{1*†}, Xinyan Pan^{2†}, Yanting Bi^{1†}, Jian Huang², Jun Peng², ChunLing Bai^{3*}, Jianhong Hou^{2*} and Yang Chen^{2*}

¹Academy of Biomedical Engineering, Kunming Medical University, Kunming, China, ²Department of Pathology, The First People's Hospital of Yunnan Province, Kunming, China, ³Human Resources Office, Kunming Medical University, Kunming, China

Introduction: The Tandem Duplicator Phenotype (TDP) is a prominent genomic metric prevalent in non-small cell lung cancer (NSCLC).

Methods: Multi-omics samples including DNA copy number, transcriptomics, somatic single-nucleotide variations, clinical data, and cell-line drug sensitivity from TCGA and CCLE are delved into TDP characteristics.

Results: Our thorough analysis reveals that patients with smaller span sizes TDP have a more intricate genome, moderate pathology, enhanced therapeutic response, favorable prognosis, a more active immune system, and higher drug sensitivity. In contrast, those with larger span sizes TDP exhibited opposite trends.

Discussion: The integrated analysis underscores that an in-depth molecular assessment of TDP can offer invaluable insights into its role in cancers. Classification strategy of TDP could recognize the chemotherapy sensitivity of different NSCLC patients.

KEYWORDS

tandem duplicator phenotype, non-small cell lung cancer, copy number variance, clinical pathology, chemotherapy

Introduction

As high-throughput biological techniques advance, whole-genome sequencing (WGS) and single nucleotide polymorphism (SNP) datasets now allow for a more profound exploration of instability patterns in human genome rearrangements and the identification of cancer-associated copy number variations (1–6). These recurrent patterns lead to genomic instability and further promote tumorigenesis (7–9). Menghi et al. (10, 11) identified a specific structural variation pattern known as the “tandem duplicator phenotype (TDP).” TDP is characterized by tandem duplication events or segmental duplications where chromosome segments, overlapping with one or more genes, can duplicate, positioning adjacently in various configurations. These duplications can encompass all or part of a gene's regulatory elements, suggesting that TDP could serve as instability-based biomarkers in the human genome (12, 13).

TDP, a unique instability configuration, has been identified in various cancers (14, 15). For instance, research shows that TDP, particularly those with head-to-tail somatic segmental tandem duplications (TDs), are predominantly present in non-small lung cell cancer, breast cancer, ovarian cancer, and endometrial cancers (10, 11, 16–19). Zhou et al. (16) also detected internal tandem duplications in NSCLC linked to focal adhesion kinase anomalies. These duplications augmented the autophosphorylation proteins level and heightened sensitivity to FAK inhibitors. However, most

previous NSCLC studies introduced genomic fingerprints of NSCLC, active immune landscape appears in NSCLC patients with high mutation burden, these multiple somatic mutation gene are associated with DNA repair (20–24). Several studies have demonstrated that *EGFR* kinase domain duplication typically results from an in-frame tandem duplication of exons 18–25, leading to constitutive activation of *EGFR* signaling via the formation of an intramolecular dimer in non-small cell lung cancer (NSCLC). Similarly, *BRAF* rearrangements, including kinase domain duplications, have been identified at a frequency of approximately 4.3%. These findings emphasize the clinical relevance of *EGFR* and *BRAF* kinase domain duplications in NSCLC and highlight the potential of targeted therapies tailored to these recurrent molecular alterations (25–27), the impact of varying TD lengths on tumor progression is yet to be fully understood. Thus, comprehensive investigations considering TDP effects on CNV, clinical variations, transcripts, mutation characteristics, and drug sensitivity are warranted (28–30).

In this study, we developed TDP models for NSCLC, encompassing both lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), by curating existing SNP data from the TCGA and CCLE cohorts (31–33). Our goal is to deeply characterize the molecular attributes of TDP and methodically examine their functions. We also present a comprehensive analysis of clinical variations and drug sensitivity related to TDP. This suggests the prospective utility of TDP groups as predictive genomic biomarkers in clinical scenarios.

Methods

Data collection

To comprehensively characterize somatic tandem duplications (TDs) in NSCLC, including LUAD and LUSC, we amassed diverse genomic datasets. These encompass masked copy number variation (CNV) data, transcriptomics, somatic mutations, and drug sensitivity metrics from the Cancer Genome Atlas (TCGA) and the Cancer Cell Line Encyclopedia (CCLE) (31–33). We derived the masked CNV data from the Affymetrix Genome-Wide Human SNP Array 6.0, utilizing it to identify TDs. To further delineate the distinct features of TDP groups, we utilized transcript and somatic mutation datasets from TCGA. Pertinent clinical data, including relapse events and timelines for TCGA patients, were sourced from the UCSC Cancer Genomic Browser.¹

TDP classification

Utilizing the masked CNV data, we set criteria to define tandem duplication segments based on three rules. Firstly, the amplification CN segment exceeded 100 bp in length; Secondly, the segment exhibited an increase relative to adjacent segments, with a log₂ CN ratio of at least 0.3; Thirdly, differences between consecutive neighboring segments maintained a log₂ CN ratio of 0.3 or less. Following these criteria, we determined a TDP score for each tumor

sample, derived from the count and chromosomal distribution of somatic tandem duplications (TDs). As per this metric,

$$\text{Raw TDP score} = -\frac{\sum_i |\text{Obs}_i - \text{exp}_i|}{\text{TD}},$$

where TD represented the total number of TDs in a sample, Obs_i represented the observed number of TDs of chromosome i , exp_i represented the expected number of TDs of chromosome i across all analysis samples, and i represented chromosome 1 to 22 and chromosome X.

Samples with raw TDP scores below the mean value of the first mode in the Gaussian finite distribution, or those with fewer than 20 TDs, were categorized into the non-TDP group. In contrast, samples with TDP scores exceeding the mean value of the second mode in the Gaussian finite distribution were identified as part of the TDP group, in alignment with the approach taken in the previous study by Menghi et al. (10).

In the TDP group, we analyzed the span size of the density distribution for all detected TDs (34, 35). Notably, the distribution peaks in each TDP group predominantly clustered at recurrent and distinct span-size intervals. To discern the pattern of these distribution peaks, we applied the “mclust” R package for Gaussian fitting distribution (25, 36–38). This helped determine the optimal number of mixture components, utilizing default values for the initialization of the iterative process. In each cohort, by marking thresholds at intersections between consecutive Gaussian curves, we distinguished span size ranges across distinct, non-overlapping intervals. Previous study used TDP group 1 (tens of base pairs), TDP group 2 (hundreds of base pairs), TDP group 3 (> 1.7 mb base pairs) to classify TDP class (19). We define the similar criterion in our TDP group in LUAD and LUSC. This process led to the definition of six unique TDP groups based on different intervals. Of these, TDP group 1, TDP group 1/2 mix, and TDP group 1/3 mix were categorized as small span size groups (SSG).

Clinical information collection and prognostic analysis

We subsequently compared the clinical information across various TDP groups. For the TCGA LUAD cohort, we collected data on clinical stage, distal metastasis stage, and therapy response for analysis. In the TCGA LUSC cohort, we examined the clinical stage, lymph node stage, and therapy response. Additionally, we employed the Kaplan–Meier survival analysis and the log-rank test to compare recurrence-free survival among the different TDP groups.

Analysis of gene differential expression

We selected two TDP groups, TDP group 2 and TDP group 2/3 mix, which exhibited notably different clinical information and had a larger patient count for the analysis of transcript differences. Within the TCGA LUAD cohort, we examined both TDP group 2 and TDP group 2/3 mix patients. For the TCGA LUSC cohort, we focused on SSG and TDP group 2/3 mix patients. We employed the R package

¹ <https://xenabrowser.net/datapages/>

“limma” to determine differentially expressed genes in the FPKM files, and the “quantile” function from “limma” was used for normalization. R package “limma” used empirical bayes to smooth of standard errors, (shrinks standard errors that are much larger or smaller than those from other genes towards the average standard error), and then used t-test to compare the difference between two groups. In the LUAD datasets, differentially expressed genes were identified with criteria of an adjusted p -value <0.01 (BH multiple hypotheses testing) and an absolute fold change ≥ 2 . For the LUSC datasets, the criteria were a p -value <0.001 (39). To further explore the KEGG pathways within distinct TDP groups, we used the R package “clusterProfiler” to identify relevant KEGG pathways with a p -value threshold of <0.05 (40–42).

Somatic copy-number analysis

We utilized the Genomic Identification of Significant Targets in Cancer (GISTIC 2.0) to pinpoint target genes within focal somatic copy-number variants (SCNVs) and to identify significant amplification regions containing genes across the six distinct TDP groups (43). For this analysis, we adopted the hg19 reference genome. The SNP annotation markers file was sourced from the UCSC Genome Browser.² Within GISTIC 2.0, the default settings for determining gene amplification or deletion were set with a threshold of 0.1. We used a q -value cutoff of ≤ 0.05 , and the confidence level for identifying driver aberrations was established at 90%.

TD-impacted cancer genes analysis

From the genes identified with amplifications by GISTIC 2.0, we retained only those classified as cancer genes for further analysis. These cancer genes were categorized into tumor suppressor genes (TSGs) and oncogenes (OGs) as proposed by Davoli et al. (44). We selected 300 TSGs and 250 OGs for evaluation, consistent with a prior study. TD-impacted cancer genes were recognized if their genomic locations overlapped with one or more TDs and met any of the following criteria, (i) duplication (DUP): the TD spanned the entire gene body, resulting in gene duplication; (ii) double transection (DT): both TD breakpoints were located within the gene body, leading to a disruption of gene integrity or (iii) single transection (ST): only one TD breakpoint was situated within the gene body, causing an effective gene copy number neutral rearrangement. We then assessed the TD-impacted cancer genes within the distinct TDP groups of each cohort.

Somatic single-nucleotide variation analysis

We utilized the MutSigCV2.0 algorithm to pinpoint genes that were significantly and recurrently mutated in various TDP groups (45), setting a p -value threshold of <0.05 . The types of somatic

mutations in these genes including synonymous, missense, frame insertion/deletion, frame shift, nonsense, splice sites, and other non-synonymous variants were visualized across distinct TDP groups using the “pheatmap” R package.

Drug sensitivity analyses

Based on the somatic copy-number profiles sourced from CCLE, we also categorized NSCLC cell lines into six distinct TDP groups (31). Subsequently, we assessed the variations in drug sensitivity across these TDP groups. We compared IC50 values between the groups when subjected to treatment with 24 different anticancer drugs.

Results

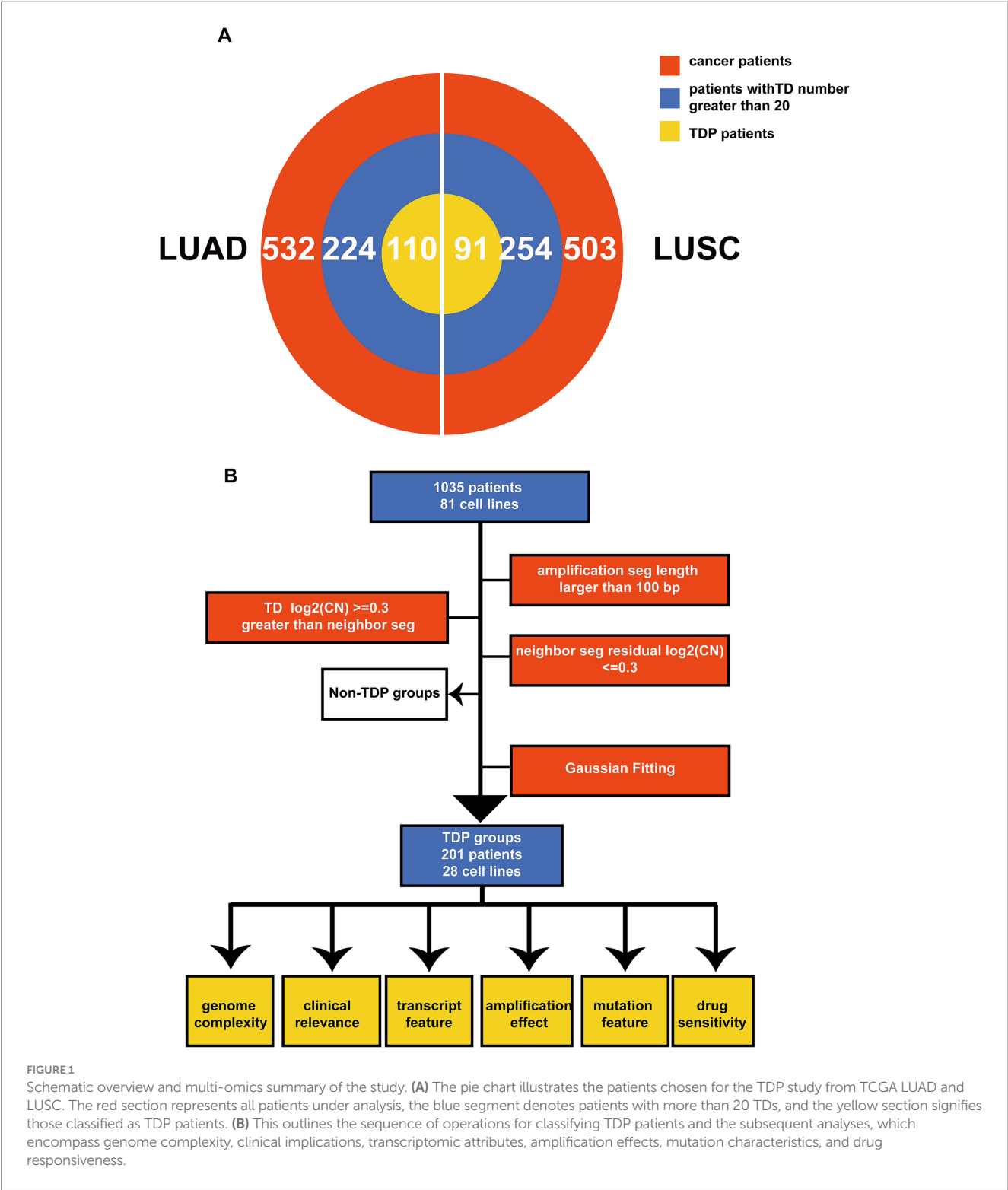
TDP patient identification and subtypes classification

The tandem duplicator phenotype (TDP) is prevalent in NSCLC, encompassing both LUAD and LUSC. To comprehensively delineate somatic TDP, we sourced and analyzed data from the Affymetrix Genome-Wide Human SNP Array 6.0. This dataset consisted of 1,035 samples and 81 cell lines, which include 532 masked CNV segment data from TCGA LUAD, 503 from TCGA LUSC, and 81 from CCLE NSCLC (Figures 1A,B) (31–33). Through the preprocessing of the SNP array data (Figure 1B), we identified appropriate TDs based on the following criteria. Firstly, we selected amplification segments with lengths exceeding 100 bp. Secondly, the log-transformed CNV ratios of somatic tandem duplications (TDs) surpassed 0.3 when compared to adjacent segments. Thirdly, the difference in log-transformed CNV ratios between two neighboring segments was less than 0.3.

For detection of the number and chromosomal distribution of TDs in cancer patients' genomes, we used the raw TDP score to characterize this feature (described in methods) (Figures 2A,B). We removed the samples whose TD counts were less than 20 (Figure 1A), and the remaining samples' raw TDP scores exceeded the mean value of the second mode according to previous study. Subsequently, 110 TDP patients from TCGA LUAD cohort; 91 TDP patients from TCGA LUSC cohort; 28 TDP cell lines from CCLE NSCLC cohort were recognized (Figure 1B).

The “Mclust” function from “Mclust” R package was applied to the TD span size distribution of TDP patients using default parameters. The TD class is defined by the mode in the Gaussian finite distribution. LUAD has three modes, resulting in three TD classes, while LUSC has four modes, leading to four TD classes according to the distribution. We identified three primary TDP groups: LUAD's first mode and LUSC's first and second modes as TDP group 1, LUAD's second mode and LUSC's third mode as TDP group 2, and LUAD's third mode and LUSC's fourth mode as TDP group 3; tumors with bimodal distributions were labeled as TDP group mixes (1/2, 1/3, or 2/3), and notably, TDP group 1, TDP group 1/2 mix, and TDP group 1/3 mix were classified as SSG (Figures 2C,D). The sample counts for TDP groups 1, 2, 3, as well as 1/2, 1/3, and 2/3 mixtures are calculated in TCGA LUAD and LUSC patients (Table 1).

² <http://genome.ucsc.edu/>



Distinct model distribution patterns often correlate with varying genome complexities. Thus, we computed the genome complexities for the three primary TDP groups and non-TDP groups from the aforementioned datasets. It emerged that the non-TDP group exhibited significantly fewer amplification and deletion regions compared to the TDP groups, with TDP group 1 and TDP group 2 exhibiting the most pronounced genome complexity (Figures 2E,F).

Clinical association and prognosis difference in distinct TDP groups

Genomic instability is a recognized hallmark of cancer based on previous research. However, the relationship between TDP and cancer remains elusive. To bridge this gap, we explored the association between patient clinical data and TDP groups and evaluated the prognosis outcomes across these groups.

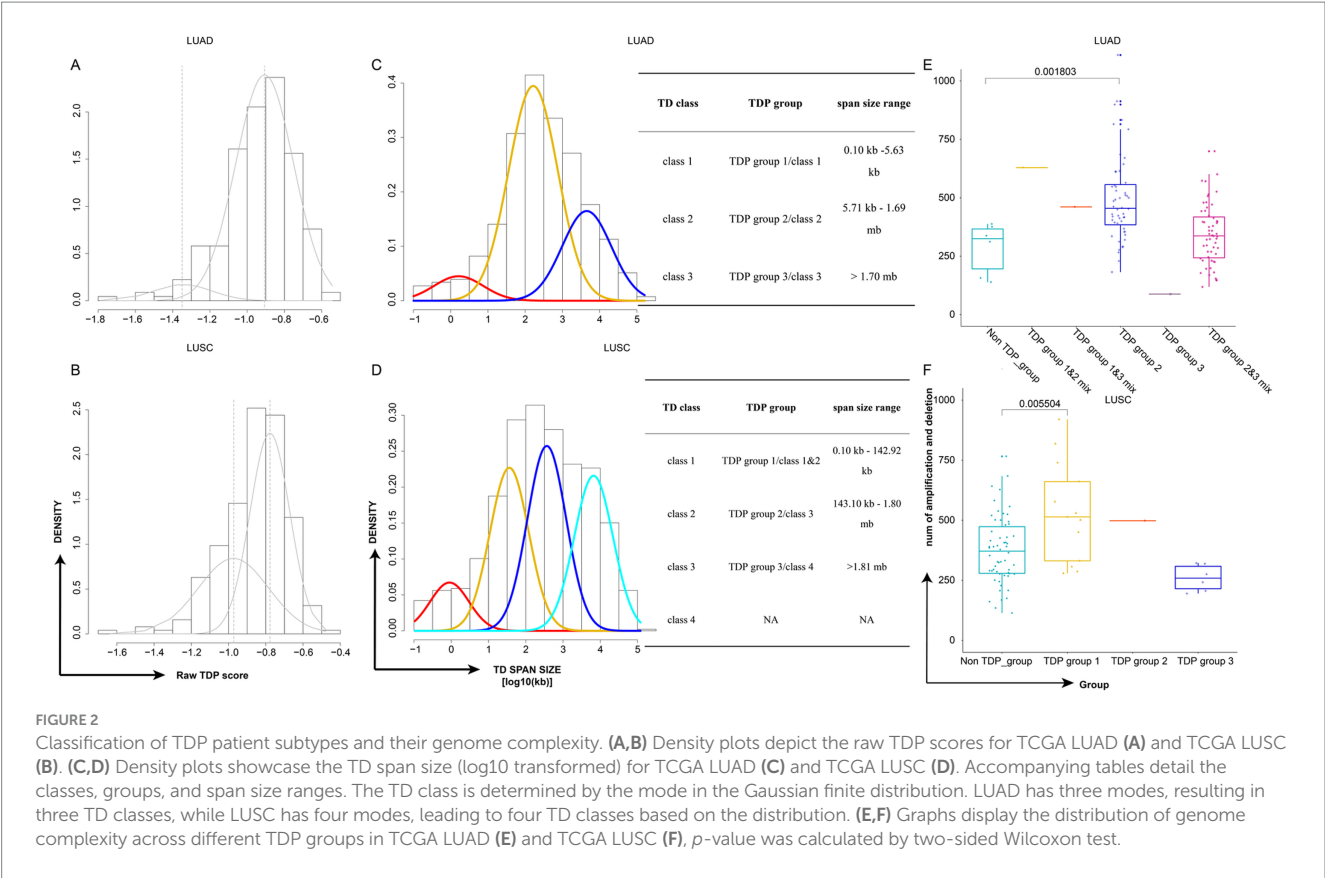


TABLE 1 Number of TDP groups in TCGA LUAD and LUSC.

TDP classification	LUAD	LUSC
TDP group 1	0	11
TDP group 1/2 mix	1	31
TDP group 1/3 mix	1	16
TDP group 2	52	1
TDP group 3	1	6
TDP group 2/3 mix	55	24

For TCGA LUAD patients, our findings indicated that patients in the TDP group 1/2 mix and TDP group 1/3 mix had a relatively moderate clinical stage (Figure 3A). They displayed no distal metastasis (Figure 3B), and notably, the TDP group 1/2 mix patients showed a complete therapy response (Figure 3C). On the other hand, a section of the patients in TDP group 2 and TDP group 2/3 mix exhibited severe clinical stages, manifested metastasis, and had an unsatisfactory therapy response (Figures 3A–C). In terms of relapse-free survival time, patients from TDP group 1/2 mix and TDP group 1/3 mix significantly had better outcome than those from TDP group 2 and TDP group 2/3 mix, with a log-rank test *p*-value of 0.0327 (Figure 3D).

For TCGA LUSC patients, those in TDP group 1 predominantly exhibited clinical stages of I and II (Figure 3E). The majority showed no lymph node metastasis and had a full therapy response (Figures 3E,G). However, patients from other groups, in contrast, had a relatively severe clinical stage, demonstrated lymph node metastasis,

and responded poorly to therapy (Figures 3E–G). Remarkably, TDP group 1 patients exhibited significantly improved outcomes concerning relapse-free survival time, with a log-rank test *p*-value of 0.0398 (Figure 3H).

Drawing conclusions from the clinical associations and prognostic analyses across different TDP groups, it's evident that patients in TDP group 1 and SSG displayed moderate pathology, superior therapy response, and a more favorable prognosis compared to their counterparts in the TDP group 2/3 mix.

Smaller span size group suggested active immune system

To delve deeper into why different TDP groups exhibit varied pathology, we compared transcript differences between SSG patients (with moderate pathology) and TDP group 2/3 mix patients (with severe pathology) in LUAD. We identified 50 genes with elevated expression in SSG patients (Figure 4A). Enriched KEGG pathways related to cell proliferation and migration included “PI3K-Akt signaling pathway,” “JAK-STAT signaling pathway,” “Focal adhesion,” and “ECM-receptor interaction.” Notably, immune pathways like “IL-17 signaling pathway,” “Cytokine-cytokine receptor interaction,” and “Chemokine signaling pathway” were also enriched (Figure 4B). The activation of these immune pathways suggests that SSG patients possess an active immune response, potentially contributing to their moderate pathology. Conversely, six genes showed heightened expression in TDP group 2/3 mix patients (Figure 4A), with cell proliferation and migration pathways such as “PI3K-Akt signaling

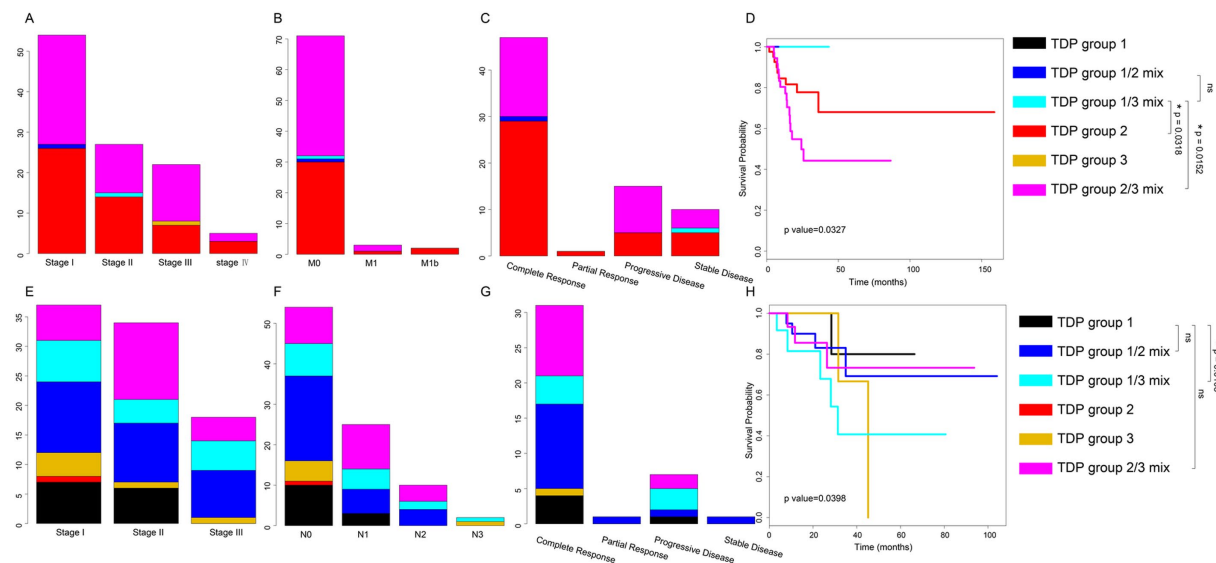


FIGURE 3

Distribution of clinical data and prognosis across various TDP groups. (A–C) Graphs illustrate the distribution of clinical stage (A), distal metastasis (B), and therapy response (C) among TDP group patients within the TCGA LUAD dataset. (D) Kaplan–Meier survival curves and log-rank tests were executed for the TCGA LUAD dataset. TDP group 1/3 mix patients have better prognosis outcome than TDP group 2 patients (p -value = 0.0318), TDP group 1/3 mix patients have better prognosis outcome than TDP group 2/3 patients (p -value = 0.0152). (E–G) Charts depict the distribution of clinical stage (E), lymph node stage (F), and therapy response (G) for patients in different TDP groups within the TCGA LUSC dataset. (H) Kaplan–Meier survival curves and log-rank tests were conducted for the TCGA LUSC dataset. TDP group 1 patients have better prognosis outcome than TDP group 1/3 patients (p -value = 0.0108), TDP group 1 patients have better prognosis outcome than TDP group 3 patients (p -value = 0.0215).

pathway,” “MAPK signaling pathway,” “Ras signaling pathway,” “Rap1 signaling pathway,” and “TNF signaling pathway” being enriched (Figure 4C), possibly driving the severe pathology observed.

In TCGA LUSC patients, we discerned transcript differences between TDP group 1 and TDP group 2/3 mix patients. Twenty-one genes were found to have increased expression in TDP group 1 patients (Figure 4D). The “IL-17 signaling pathway” was similarly enriched as observed in LUAD SSG patients, accompanied by activation of numerous metabolic pathways (Figure 4E). In contrast, seven genes exhibited heightened expression in TDP group 2/3 mix patients (Figure 4D), with the enrichment of pathways like “Wnt signaling pathway,” “Glycolysis/Gluconeogenesis,” and “Galactose metabolism” (Figure 4F).

In summary, the activation of immune pathways in SSG and TDP group 1 patients may underpin their moderate pathology. Conversely, the prominence of cell proliferation and migration pathways in TDP group 2/3 mix patients could be responsible for their severe pathology.

Cancer gene amplification effect in distinct TDP groups

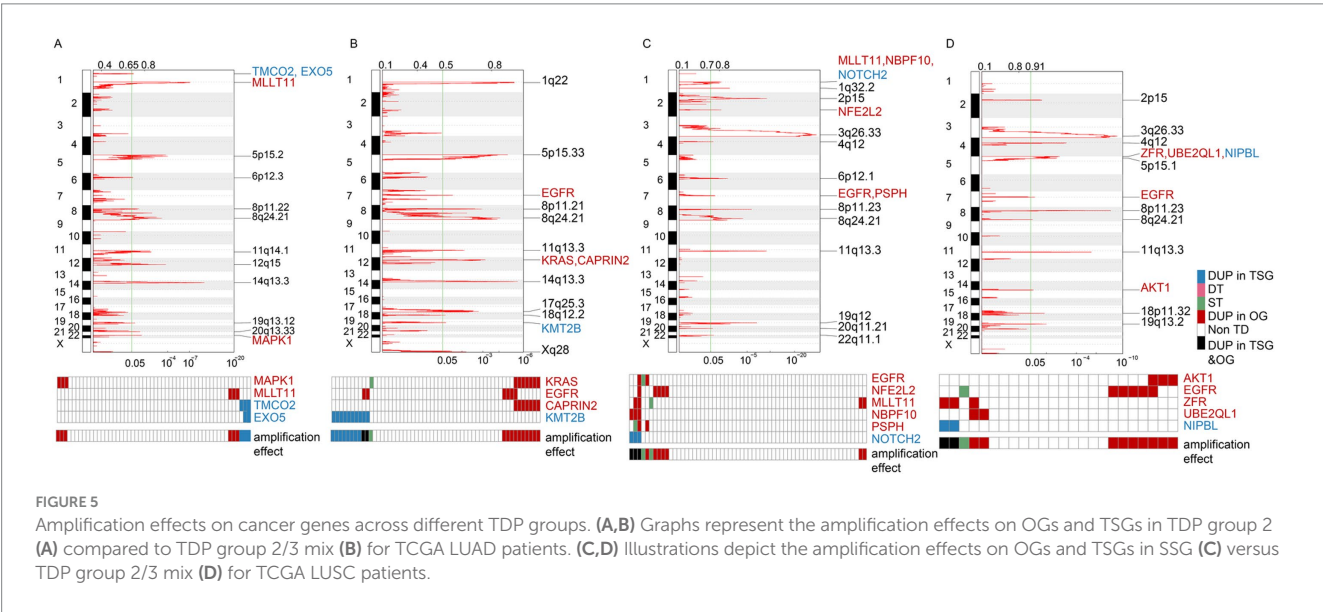
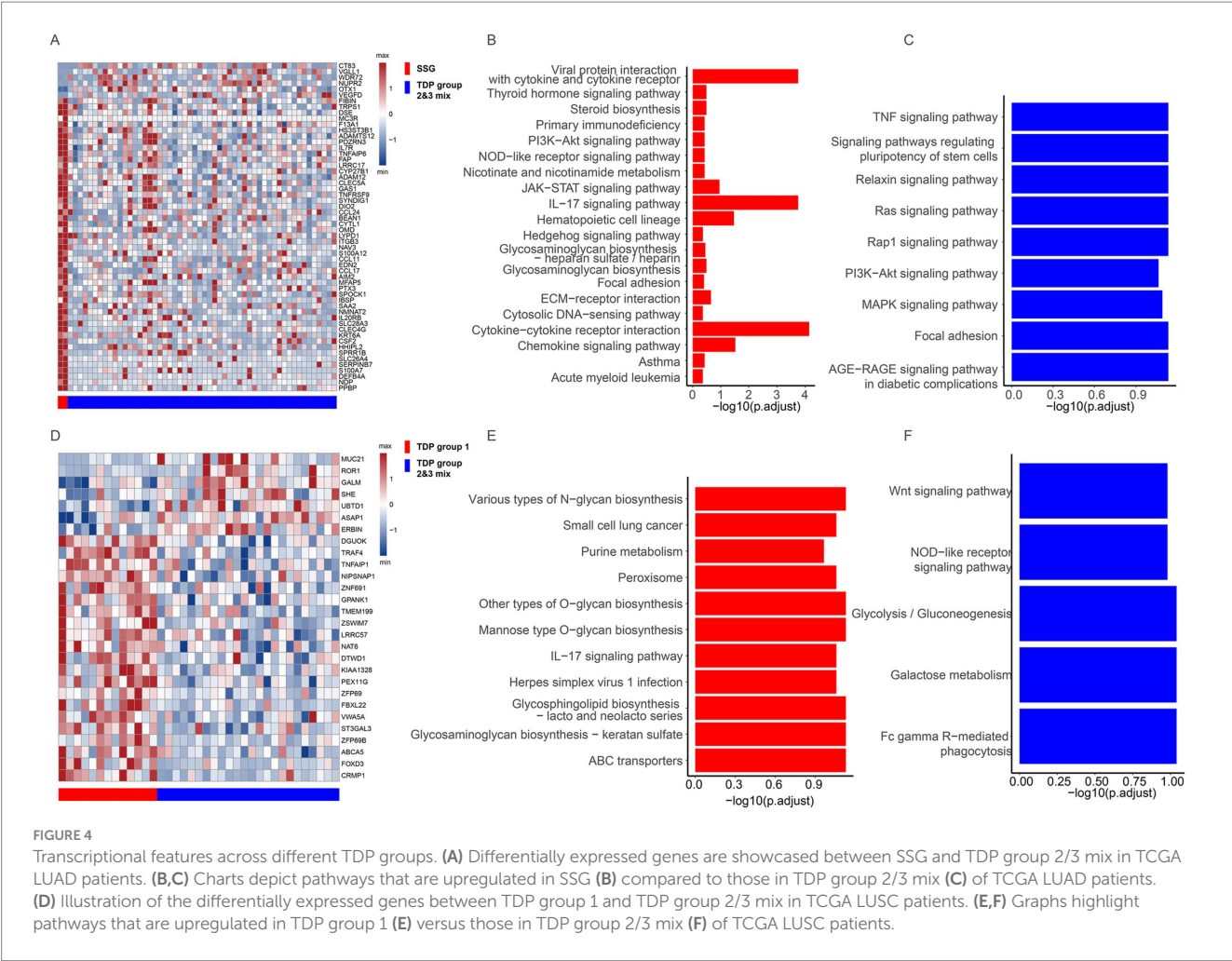
Following our exploration into TDP groups, we further investigated how TDs influence oncogenes (OGs) and tumor suppressor genes (TSGs). For this purpose, we relied on 300 TSGs and 250 OGs classified by Davoli et al. (44) for an in-depth analysis, TD target genes were pinpointed using GISTIC 2.0 (43). As delineated in the methods, TDs can impact the integrity of a gene’s body in three main ways: (1) duplication (DUP) leading to gene replication, (2) double transection (DT) causing gene disruptions, and (3) single

transection (ST) that results in an effective gene copy-number neutral rearrangement.

For the sake of robust results and considering sample size, our focus turned to TDP groups with substantial numbers—specifically, TDP group 2 and TDP group 2/3 mix in LUAD, along with SSG and TDP group 2/3 mix in LUSC—to examine the effects of cancer gene amplification. Amplification regions were identified by GISTIC 2.0, and genes from these regions with a p -value below 0.05 were selected for further scrutiny.

Our findings highlighted that in LUAD TDP group 2 patients, OGs like “MAPK1” and “MLLT11” along with TSGs such as “TMCO2” and “EXO5” were subject to the DUP effect. Specifically, the DUP effect for OGs was observed in 6 out of 52 TDP group 2 patients, and for TSGs, this effect was noted in 4 out of the same 52 patients (Figure 5A). For the LUAD TDP group 2/3 mix, OGs like “KRAS,” “EGFR,” and “CAPRIN2” and TSGs including “KMT2B” manifested the DUP effect. Here, the DUP effect was present in 10 of 55 patients for OGs and 8 of 55 for TSGs (Figure 5B). These amplifications, particularly in OGs, could potentially underlie the severe pathology.

For the TCGA LUSC dataset, OGs such as “EGFR,” “NFE2L2,” “MLLT11,” “NBPF10,” and “PSPH” as well as TSGs like “NOTCH2” exhibited the DUP effect in SSG patients, with the OGs’ DUP effect recorded in 7 out of 60 SSG patients (Figure 5C). In the TDP group 2/3 mix patients, OGs including “AKT1,” “EGFR,” “ZFR,” and “UBE2QL1” and TSGs like “NIPBL” demonstrated the DUP effect. Here, the OGs’ DUP effect was evident in 9 out of the 24 TDP group 2/3 mix patients (Figure 5D). This higher proportion of affected patients underscores the link between severe pathology and the TDP group 2/3 mix.



Insights into mutation feature in distinct TDP groups

We further examined the mutation characteristics of OGs and TSGs across various TDP groups. Among TCGA LUAD patients,

commonly occurring mutations of OG “KRAS” and TSGs such as “TP53,” “KEAP1,” “RB1,” and “SMARCA4” were identified in both TDP group 2 and TDP group 2/3 mix patients. Additionally, each group presented with its unique set of mutated genes (Figures 6A,B). Previous research by Ambrogio et al. (46) highlighted that “KRAS”

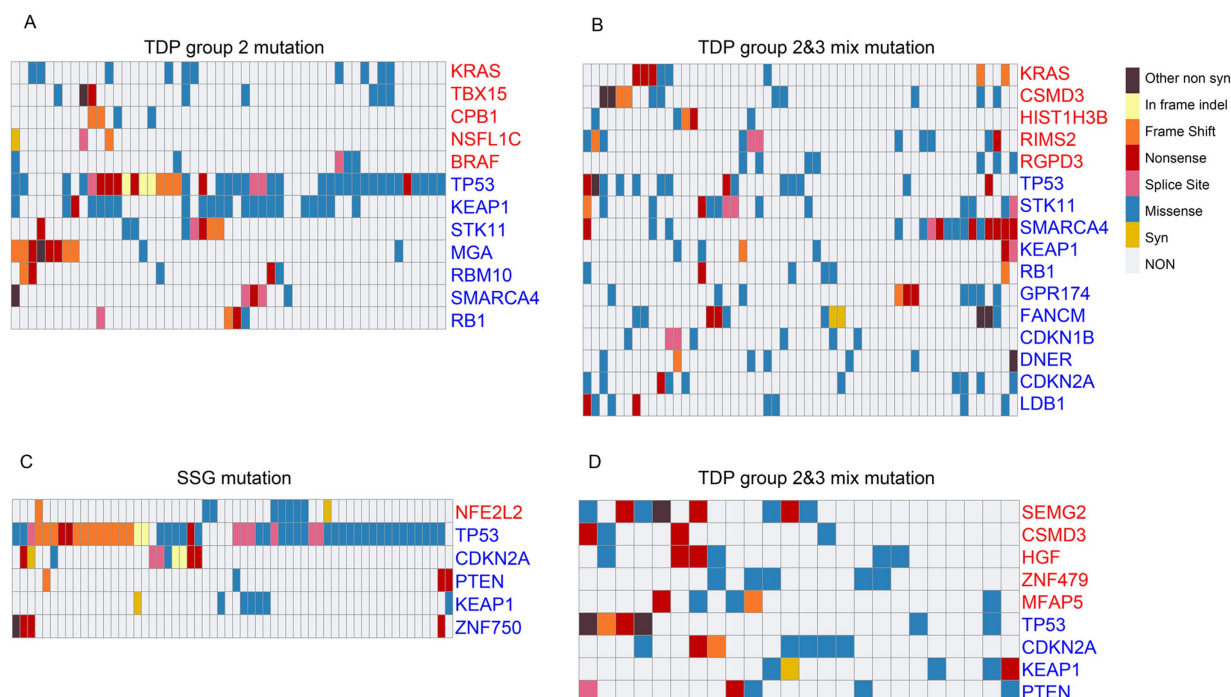


FIGURE 6

Mutation characteristics of cancer genes across different TDP groups. (A,B) The mutation patterns of OGs and TSGs are displayed for TDP group 2 (A) in comparison to TDP group 2/3 mix (B) in TCGA LUAD patients. (C,D) Charts showcase the mutation characteristics of OGs and TSGs for SSG (C) versus TDP group 2/3 mix (D) in TCGA LUSC patients.

dimerization enhances the survival of both human and murine KRAS mutant LUAD tumor cells possessing wild-type KRAS, explaining the observed resistance to MEK inhibition. Further, Romero et al. (47) suggested a link between Keap1 loss and KRAS-driven lung cancer.

For TCGA LUSC patients, recurring mutations in TSGs like “TP53,” “CDKN2A,” “PTEN,” and “KEAP1” were observed in both SSG and TDP group 2/3 mix patients. Specifically, mutations in “NFE2L2” and “ZNF750” were detected in SSG patients, while “SEMG2,” “CSMD3,” “HGF,” “ZNF479,” and “MFAP5” mutations were noted in the TDP group 2/3 mix patients (Figures 6C,D). A study by Zhang et al. (48) identified novel genetic disruptions in squamous cell lung carcinoma, pointing out that “CSMD3” is associated with a high frequency of single nucleotide variants.

Smaller span size TDP groups have greater drug sensitivity

To delve deeper into the varied drug sensitivities across different TDP groups, we sourced pharmacological profiles and segmented copy-number profiles for 24 anticancer drugs, encompassing 504 cell lines from the Cancer Cell Line Encyclopedia³ (31).

For the CCLE LUAD cohort, we segmented cell lines into distinct TDP groups using criteria consistent with the TCGA cohort. Upon calculating the TDP score (Supplementary Figures 1A,B) and evaluating the Gaussian fitting distribution (Supplementary Figures 1C,D),

we classified cell lines into TDP group 1 and TDP group 1/2 mix. The former comprised six samples, while the latter included twenty-two. Aligning TDP groups with drug therapy concentration data, we noted that following Lapatinib treatment (targeting the “EGFR” gene), the IC₅₀ for TDP group 1/2 mix cell lines was considerably higher than for TDP group 1 cell lines, yielding a *p*-value of 0.0107 (one-sided Wilcoxon test) (Supplementary Figure 1E). A parallel outcome emerged when LUAD cell lines were treated with Panobinostat (targeting the “HDAC” gene), where the *p*-value was 0.0318 (Supplementary Figure 1F).

A comparable analysis was executed for CCLE LUSC cell lines treated with AZD6244 (targeting the “MEK” gene). This resulted in the identification of four TDP group 1 cell lines and seventeen TDP group 1/2 mix cell lines. The IC₅₀ of the TDP group 1/2 mix cell lines was markedly higher than that of the TDP group 1 cell lines, with a *p*-value of 0.0333 (one-sided Wilcoxon test) (Supplementary Figure 1G).

Conclusively, TDP groups with smaller span sizes exhibited lower IC₅₀ values, signifying that their drug sensitivity surpassed that of groups with larger span sizes.

Discussion

Recurrent genomic rearrangements, such as duplications, translocations, inversions, and deletions, have become a focal point in current research, specifically in the context of TDP. This article explores of TDP, identifying six unique TDP groups based on TD span size, each with distinct characteristics. Notably, TDP group 1 and SSG stand out as exemplars of genomic complexity, while the non-TDP group exhibits the least complexity. Referring to Macintyre's et al. (49) research, we draw connections between high genomic complexities,

³ <https://portals.broadinstitute.org/ccle/data>

particularly CN signatures 1, 7, and 8, and favorable prognoses, with CN signature 7 indicating a tandem duplicator phenotype.

As expected, both TDP group 1 and SSG patients experience significantly improved recurrence-free survival, characterized by a moderate clinical stage, primarily “stage I” and “stage II” absence of distal metastases, moderate lymph node stages, and positive therapy response. In contrast, a substantial portion of TDP group 2/3 mix patients presents severe pathology, evident metastasis, advanced lymph node stages, and less responsive therapy outcomes. We propose that the length of TDP may serve as a potential biomarker for predicting lung cancer recurrence. Considering the clinical relevance of TDP in oncogenes such as EGFR and BRAF, targeted sequencing probes could be developed to detect these alterations more effectively, thereby enabling a more accurate assessment of patient prognosis.

Examining the clinical pathology at the transcriptomic level across different TDP groups reveals that SSG and TDP group 1 patients activate numerous immune-associated pathways, potentially curbing tumor cell growth and spread. Conversely, the TDP group 2/3 mix shows an enrichment of pathways driving cell growth and mobility, likely contributing to their severe pathology.

Analyzing gene amplification in the TDP group 2/3 mix highlights OG duplication, particularly emphasizing the frequent occurrence of EGFR in TCGA LUAD and LUSC cohorts. EGFR amplification, recognized for spurring cell proliferation, may hinder targeted treatments, especially in tumors with EGFR-Kinase domain duplications (26, 50). Notably, TDP group 2/3 mix patients exhibit a worse prognosis, possibly due to the amplification of EGFR wild-type alleles.

On the mutation front, recurrent “KRAS” mutations are identified in both TDP group 2 and TDP group 2/3 mix of TCGA LUAD, suggesting potential resistance and adverse pathology. Similarly, the “CSMD3” mutation, a driver of LUSC, is prevalent in the TDP group 2/3 mix, contributing to severe pathology.

Neo-adjuvant studies underscore the responsiveness of TDP group 1 cell lines to treatments, with lower IC50 values than TDP group 1/2 mix cell lines. These differences suggest that the TDP tumor classification is a potent predictor of treatment response, irrespective of tumor type. Importantly, SSG and TDP group 1, prevalent in both LUAD and LUSC, share similar clinical, prognostic, transcriptional, mutational, and biological features. It should be noted that the drug susceptibility analysis based on CCLE data included only six LUAD cell lines, which limits the statistical power and raises the possibility of false-positive associations. Therefore, the results should be interpreted as preliminary and hypothesis-generating. Further validation in larger datasets and experimental models is essential to confirm these findings. Thus, classifying TDP tumors into six distinct categories based on molecular determinants holds clinical significance, potentially guiding adjuvant chemotherapy benefits.

In summary, our findings significantly advance the classification of TDP tumor genomic complexity, delineating the molecular features of different TDP groups, elucidating the underpinnings of diverse pathologies and prognoses, and highlighting the clinical implications of distinct TDP groups in adjuvant chemotherapy. Despite these promising findings, several limitations should be acknowledged. First, the analyses were primarily based on retrospective datasets, which may introduce bias. Second, the drug susceptibility analysis was limited by the small number of LUAD cell lines in the CCLE dataset,

which may affect the robustness of the results. Third, functional and mechanistic validation of the TDP subtypes and their association with drug response is lacking. Future studies should address these limitations through larger, prospective cohorts and experimental validation.

Conclusion

We conducted a systematic analysis of the molecular characteristics of TDP, establishing a link between TDP, clinical pathology, therapeutic response, and drug sensitivity. Our findings indicate that patients with a smaller span size exhibit a better therapeutic response and a more active immune system. This insight lays the groundwork for a deeper exploration of the roles TDP plays in cancer, the deep exploration in TDP by using single cell genome data with single cell transcript data might improve the explanation of difference between heterogenous TDP cell types. Moreover, TDP classification would help identify patients who could benefit from chemotherapy.

Code availability

The code of this manuscript was deposited in Github, the website is <https://github.com/jacklee2thu/Comprehensive-multi-omics-analysis-of-tandem-duplicator-phenotypes-in-non-small-cell-lung-cancer>.

Data availability statement

The datasets generated and/or analyzed during the current study are available in the UCSC Cancer Genomic Browser repository and CCLE repository. Pertinent masked CNV data, gene expression data (FPKM value), somatic mutation data, clinical data, including relapse events and timelines for TCGA LUAD patients, were sourced from the UCSC Cancer Genomic Browser [https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Lung%20Adenocarcinoma%20\(LUAD\)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Lung%20Adenocarcinoma%20(LUAD)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443). These data for TCGA LUSC patients were deposited at [https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Lung%20Squamous%20Cell%20Carcinoma%20\(LUSC\)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Lung%20Squamous%20Cell%20Carcinoma%20(LUSC)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443). The cell lines data and drug resistance data were source from [https://xenabrowser.net/datapages/?cohort=Cancer%20Cell%20Line%20Encyclopedia%20\(CCLE\)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=Cancer%20Cell%20Line%20Encyclopedia%20(CCLE)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443).

Author contributions

JL: Writing – original draft, Writing – review & editing. XP: Writing – original draft. YB: Writing – original draft. JHu: Writing – review & editing. JP: Writing – review & editing. CB: Writing – original draft, Writing – review & editing. JHo: Writing – original draft, Writing – review & editing. YC: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China (No. 82404091), the Yunnan Fundamental Research Projects (202501CF070023), the Health Commission Foundation of Yunnan Province (2023-KHRCBZ-B15), the National Natural Science Foundation of China (No. 82302957), Major Science and Technology Project of Yunnan Province (202402AA310016), Yunnan Province Key Basic Research Science and Technology Plan Project (202201AS070009), Yunnan Province Xingdian Famous Doctor Plan Project (XDYC-MY-2022-0029).

Conflict of interest

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

References

- Eisfeldt J, Pettersson M, Vezzi F, Wincent J, Källér M, Gruselius J, et al. Comprehensive structural variation genome map of individuals carrying complex chromosomal rearrangements. *PLoS Genet.* (2019) 15:e1007858. doi: 10.1371/journal.pgen.1007858
- Pros E, Saigi M, Alameda D, Gomez-Mariano G, Martinez-Delgado B, Albuquerque-Bejar JJ, et al. Genome-wide profiling of non-smoking-related lung cancer cells reveals common RB1 rearrangements associated with histopathologic transformation in EGFR-mutant tumors. *Ann Oncol.* (2020) 31:274–82. doi: 10.1016/j.annonc.2019.09.001
- Koch L. Cancer genetics: a 3D view of genome rearrangements. *Nat Rev Genet.* (2017) 18:456. doi: 10.1038/nrg.2017.56
- Xu C, Xia P, Li J, Lewis KB, Ciombor KK, Wang L, et al. Discovery and validation of a 10-gene predictive signature for response to adjuvant chemotherapy in stage II and III colon cancer. *Cell Rep Med.* (2024) 5:101661. doi: 10.1016/j.xcrm.2024.101661
- Zhu X, Li Y, Gu N. Application of artificial intelligence in the exploration and optimization of biomedical nanomaterials. *Nano Biomed Eng.* (2023) 15:342–53. doi: 10.26599/NBE.2023.9290035
- Tade RS, Jain SN, Satyavijay JT, Shah PN, Bari TD, Patil TM, et al. Artificial intelligence in the paradigm shift of pharmaceutical sciences: a review. *Nano Biomed Eng.* (2024) 16:64–77. doi: 10.26599/NBE.2023.9290043
- Soca-Chafre G, Montiel-Davalos A, Rosa-Velazquez IA, Caro-Sanchez CHS, Pena-Nieves A, Arrieta O. Multiple molecular targets associated with genomic instability in lung cancer. *Int J Genomics.* (2019) 2019:9584504. doi: 10.1155/2019/9584504
- Li J, Chen J, Sun X, Yang H, Sun K, Liu C, et al. Uncovering chromatin accessibility and cancer diagnostic potential via cell-free DNA utilization. *Sci Bull.* (2024) 69:2987–92. doi: 10.1016/j.scib.2024.04.013
- Zhang X, Li J, Lan X, Li J. Cell-free DNA-associated multi-feature applications in cancer diagnosis and treatment. *Clin Transl Discov.* (2024) 4:e280. doi: 10.1002/ctd.280
- Menghi F, Inaki K, Woo X, Kumar PA, Grzeda KR, Malhotra A, et al. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc Natl Acad Sci USA.* (2016) 113:E2373–82. doi: 10.1073/pnas.1520010113
- Menghi F, Liu ET. Reply to Watkins et al.: Whole-genome sequencing-based identification of diverse tandem duplicator phenotypes in human cancers. *Proc Natl Acad Sci USA.* (2016) 113:E5259–60. doi: 10.1073/pnas.1610624113
- Catchen JM, Braasch I, Postlethwait JH. Conserved synteny and the zebrafish genome. *Methods Cell Biol.* (2011) 104:259–85. doi: 10.1016/B978-0-12-374814-0.00015-X
- Baik CS, Wu D, Smith C, Martins RG, Pritchard CC. Durable response to tyrosine kinase inhibitor therapy in a lung cancer patient harboring epidermal growth factor receptor tandem kinase domain duplication. *J Thorac Oncol.* (2015) 10:e97–9. doi: 10.1097/JTO.0000000000000586
- Choi EY, Patel K, Haddad MR, Yi L, Holmes C, Goldstein DS, et al. Tandem duplication of exons 1–7 neither impairs Atp7a expression nor causes a Menkes disease phenotype. *JIMD Rep.* (2015) 20:57–63. doi: 10.1007/8904_2014_391

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1556840/full#supplementary-material>

- Bailey EJ, Duffield AS, Greenblatt SM, Aplan PD, Small D. Effect of FLT3 ligand on survival and disease phenotype in murine models harboring a FLT3 internal tandem duplication mutation. *Comp Med.* (2013) 63:218–26. doi: 10.30802/AALAS-CM-13-000014
- Zhou B, Wang GZ, Wen ZS, Zhou YC, Huang YC, Chen Y, et al. Somatic mutations and splicing variants of focal adhesion kinase in non-small cell lung cancer. *J Natl Cancer Inst.* (2018) 110:195–204. doi: 10.1093/jnci/djx157
- Glodzik D, Morganello S, Davies H, Simpson PT, Li Y, Zou X, et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat Genet.* (2017) 49:341–8. doi: 10.1038/ng.3771
- Popova T, Manie E, Boeva V, Battistella A, Goundiam O, Smith NK, et al. Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. *Cancer Res.* (2016) 76:1882–91. doi: 10.1158/0008-5472.CAN-15-2128
- Menghi F, Barthel FP, Yadav V, Tang M, Ji B, Tang Z, et al. The tandem duplicator phenotype is a prevalent genome-wide cancer configuration driven by distinct gene mutations. *Cancer Cell.* (2018) 34:197–210.e5. doi: 10.1016/j.ccell.2018.06.008
- Wang C, Yin R, Dai J, Gu Y, Cui S, Ma H, et al. Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC patients. *Nat Commun.* (2018) 9:2054. doi: 10.1038/s41467-018-04492-2
- Garsed DW, Pandey A, Fereday S, Kennedy CJ, Takahashi K, Alsop K, et al. The genomic and immune landscape of long-term survivors of high-grade serous ovarian cancer. *Nat Genet.* (2022) 54:1853–64. doi: 10.1038/s41588-022-01230-9
- Ricciuti B, Wang X, Alessi JV, Rizvi H, Mahadevan NR, Li YY, et al. Association of high tumor mutation burden in non-small cell lung cancers with increased immune infiltration and improved clinical outcomes of PD-L1 blockade across PD-L1 expression levels. *JAMA Oncol.* (2022) 8:1160–8. doi: 10.1001/jamaoncol.2022.1981
- Wang S, Yuan P, Mao B, Li N, Ying J, Tao X, et al. Genomic features and tumor immune microenvironment alteration in NSCLC treated with neoadjuvant PD-1 blockade. *NPJ Precis Oncol.* (2022) 6:2. doi: 10.1038/s41698-021-00244-6
- Chen Y, Zheng Q, Wang H, Tang P, Deng L, Li P, et al. Integrating transcriptomics and proteomics to analyze the immune microenvironment of cytomegalovirus associated ulcerative colitis and identify relevant biomarkers. *BioData Min.* (2024) 17:26. doi: 10.1186/s13040-024-00382-0
- Du Z, Brown BP, Kim S, Ferguson D, Pavlick DC, Jayakumaran G, et al. Structure-function analysis of oncogenic EGFR Kinase Domain Duplication reveals insights into activation and a potential approach for therapeutic targeting. *Nat Commun.* (2021) 12:1382. doi: 10.1038/s41467-021-21613-6
- Wang J, Li X, Xue X, Ou Q, Wu X, Liang Y, et al. Clinical outcomes of EGFR kinase domain duplication to targeted therapies in NSCLC. *Int J Cancer.* (2019) 144:2677–82. doi: 10.1002/ijc.31895
- Sheikine Y, Pavlick D, Klempner SJ, Trabucco SE, Chung JH, Rosenzweig M, et al. BRAF in lung cancers: analysis of patient cases reveals recurrent BRAF mutations, fusions, kinase duplications, and concurrent alterations. *JCO Precis Oncol.* (2018) 2:2. doi: 10.1200/PO.17.00172

28. Li J, Lan X. Perspective on new cell-free DNA technologies for early cancer detection. *Cancer Biol Med.* (2023) 21:139–43. doi: 10.20892/j.issn.2095-3941.2023.0159
29. Li J, Sun X, Yang H, Chen J, Bu Z, Ji J, et al. Integrated analysis toolkit for dissecting whole-genome-wide features of cell-free DNA. *Clin Transl Med.* (2023) 13:e1212. doi: 10.1002/ctm2.1212
30. Li J, Xu M, Peng J, Wang J, Zhao Y, Wu W, et al. Novel technologies in cfDNA analysis and potential utility in clinic. *Chin J Cancer Res.* (2021) 33:708–18. doi: 10.21147/j.issn.1000-9604.2021.06.07
31. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature.* (2019) 569:503–8. doi: 10.1038/s41586-019-1186-3
32. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* (2014) 511:543–50. doi: 10.1038/nature13385
33. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* (2012) 489:519–25. doi: 10.1038/nature11404
34. Wang X, Chen H, Jiang R, Hong X, Peng J, Chen W, et al. Interleukin-17 activates and synergizes with the notch signaling pathway in the progression of pancreatic ductal adenocarcinoma. *Cancer Lett.* (2021) 508:1–12. doi: 10.1016/j.canlet.2021.03.003
35. Wang S, Xia P, Zhang L, Yu L, Liu H, Meng Q, et al. Systematical Identification of Breast Cancer-Related Circular RNA Modules for Deciphering circRNA Functions Based on the Non-Negative Matrix Factorization Algorithm. *Int J Mol Sci.* (2019) 20:919. doi: 10.3390/ijms20040919
36. Xu C, Ping Y, Zhao H, Ning S, Xia P, Wang W, et al. LncNetP, a systematical lncRNA prioritization approach based on ceRNA and disease phenotype association assumptions. *Oncotarget.* (2017) 8:114603–12. doi: 10.18632/oncotarget.23059
37. Pang H, Jiang Y, Li J, Wang Y, Nie M, Xiao N, et al. Aberrant Nad⁺ metabolism underlies Zika virus-induced microcephaly. *Nat Metab.* (2021) 3:1109–24. doi: 10.1038/s42255-021-00437-0
38. Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, et al. Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. *Int J Cancer.* (2018) 143:2150–60. doi: 10.1002/ijc.31573
39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
40. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
41. Xu C, Qi R, Ping Y, Li J, Zhao H, Wang L, et al. Systemically identifying and prioritizing risk lncRNAs through integration of pan-cancer phenotype associations. *Oncotarget.* (2017) 8:12041–51. doi: 10.18632/oncotarget.14510
42. Chen J, Sun Y, Li J, Lyu M, Yuan L, Sun J, et al. In-depth metaproteomics analysis of tongue coating for gastric cancer: a multicenter diagnostic research study. *Microbiome.* (2024) 12:6. doi: 10.1186/s40168-023-01730-8
43. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi 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. doi: 10.1186/gb-2011-12-4-r41
44. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell.* (2013) 155:948–62. doi: 10.1016/j.cell.2013.10.011
45. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* (2013) 499:214–8. doi: 10.1038/nature12213
46. Ambrogio C, Kohler J, Zhou ZW, Wang H, Paranal R, Li J, et al. KRAS dimerization impacts MEK inhibitor sensitivity and oncogenic activity of mutant KRAS. *Cell.* (2018) 172:857–868.e15. doi: 10.1016/j.cell.2017.12.020
47. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat Med.* (2017) 23:1362–8. doi: 10.1038/nm.4407
48. Zhang P, Kang B, Xie G, Li S, Gu Y, Shen Y, et al. Genomic sequencing and editing revealed the Grm8 signaling pathway as potential therapeutic targets of squamous cell lung cancer. *Cancer Lett.* (2019) 442:53–67. doi: 10.1016/j.canlet.2018.10.035
49. Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, Eldridge M, et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet.* (2018) 50:1262–70. doi: 10.1038/s41588-018-0179-8
50. Nukaga S, Yasuda H, Tsuchihara K, Hamamoto J, Masuzawa K, Kawada I, et al. Amplification of EGFR wild-type alleles in non-small cell lung cancer cells confers acquired resistance to mutation-selective EGFR tyrosine kinase inhibitors. *Cancer Res.* (2017) 77:2078–89. doi: 10.1158/0008-5472.CAN-16-2359