



## Article

# Cancer Essential Genes Stratified Lung Adenocarcinoma Patients with Distinct Survival Outcomes and Identified a Subgroup from the Terminal Respiratory Unit Type with Different Proliferative Signatures in Multiple Cohorts

Kuo-Hao Ho <sup>1,2</sup>, Tzu-Wen Huang <sup>1,3</sup> , Ann-Jeng Liu <sup>4</sup>, Chwen-Ming Shih <sup>1,2,\*</sup> and Ku-Chung Chen <sup>1,2,\*</sup> 

<sup>1</sup> Graduate Institute of Medical Sciences, College of Medicine, Taipei Medical University, Taipei 11031, Taiwan; d119106001@tmu.edu.tw (K.-H.H.); tw.huang@tmu.edu.tw (T.-W.H.)

<sup>2</sup> Department of Biochemistry and Molecular Cell Biology, School of Medicine, College of Medicine, Taipei Medical University, Taipei 11031, Taiwan

<sup>3</sup> Department of Microbiology and Immunology, School of Medicine, College of Medicine, Taipei Medical University, Taipei 11031, Taiwan

<sup>4</sup> Department of Neurosurgery, Taipei City Hospital Ren-Ai Branch, Taipei 10629, Taiwan; DAB90@tpech.gov.tw

\* Correspondence: cmsih@tmu.edu.tw (C.-M.S.); kuchung@tmu.edu.tw (K.-C.C.)



**Citation:** Ho, K.-H.; Huang, T.-W.; Liu, A.-J.; Shih, C.-M.; Chen, K.-C. Cancer Essential Genes Stratified Lung Adenocarcinoma Patients with Distinct Survival Outcomes and Identified a Subgroup from the Terminal Respiratory Unit Type with Different Proliferative Signatures in Multiple Cohorts. *Cancers* **2021**, *13*, 2128. <https://doi.org/10.3390/cancers13092128>

Academic Editor: Rory Johnson

Received: 30 March 2021

Accepted: 26 April 2021

Published: 28 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Several genes are essential for tumor growth and predict poor prognoses of patients, however, most of their roles in lung adenocarcinoma (LUAD) are unclear. In addition, a good classification strategy for cancer patients would be a useful tool for future personalized medicine. In LUAD, the existing subtype classification, including the terminal respiratory unit (TRU), proximal-inflammatory (PI), and proximal-proliferative (PP) subtypes, is mainly based on genes with variant expression levels across patients without considering the oncogenetic roles of those genes. Thus, the LUAD essential genes were identified and used to stratify patients into distinct survival outcomes, TP53 mutation statuses, E2F target activities, and tumor mutation burdens. Moreover, TRU-type patients could be further divided into clinically and molecularly different subgroups based on our classifier. Integration of existing subtypes with our classification strategy provides a more comprehensive understanding of the heterogeneity of LUAD, and can guide us to identifying potential targets for future personalized medicine.

**Abstract:** *Background:* Heterogeneous features of lung adenocarcinoma (LUAD) are used to stratify patients into terminal respiratory unit (TRU), proximal-proliferative (PP), and proximal-inflammatory (PI) subtypes. A more-accurate subtype classification would be helpful for future personalized medicine. However, these stratifications are based on genes with variant expression levels without considering their tumor-promoting roles. We attempted to identify cancer essential genes for LUAD stratification and their clinical and biological differences. *Methods:* Essential genes in LUAD were identified using genome-scale CRISPR screening of RNA sequencing data from Project Achilles and The Cancer Genome Atlas (TCGA). Patients were stratified using consensus clustering. Survival outcomes, genomic alterations, signaling activities, and immune profiles within clusters were investigated using other independent cohorts. *Findings:* Thirty-six genes were identified as essential to LUAD, and there were used for stratification. Essential gene-classified clusters exhibited distinct survival rates and proliferation signatures across six cohorts. The cluster with the worst prognosis exhibited TP53 mutations, high E2F target activities, and high tumor mutation burdens, and harbored tumors vulnerable to topoisomerase I and poly(ADP ribose) polymerase inhibitors. TRU-type patients could be divided into clinically and molecularly different subgroups based on these essential genes. *Conclusions:* Our study showed that essential genes to LUAD not only defined patients with different survival rates, but also refined preexisting subtypes.

**Keywords:** lung adenocarcinoma (LUAD); terminal respiratory unit (TRU); TP53 mutation; E2F

## 1. Introduction

Lung cancer, a leading cause of cancer-associated mortality, can be categorized into small-cell lung carcinoma (SCLC) and non-SCLC (NSCLC). Lung adenocarcinoma (LUAD) is the major subtype of NSCLC which accounts for around 40% of all lung cancer patients [1]. Although various drugs for treating LUAD have been extensively investigated, survival rates of LUAD patients have still not dramatically improved [2]. Therapeutic strategies for LUAD are based on histopathological features or tumors presenting with targetable genomic alterations like *epidermal growth factor receptor 1 (EGFR1)* mutations [3], or translocation of *ALK*, *RET*, or *ROS1* [4–6]. However, these factors still do not completely capture the highly heterogeneous features of LUAD. Thus, it is necessary to characterize this complex disease using more-sophisticated approaches.

With the development of high-throughput sequencing technology, several studies have tried to define LUAD based on transcriptome profiles. Initially, LUAD was stratified into bronchoid, magnoid, and squamoid subtypes with significant clinical differences such as stage-specific survival [7]. Through large-scale multiomics investigations of LUAD by The Cancer Genome Atlas (TCGA), these subtypes were renamed terminal respiratory unit (TRU), proximal-proliferative (PP), and proximal-inflammatory (PI) subtypes [8]. Among these types, TRU has favorable prognoses and harbors tumors presenting *EGFR* mutations. The PP and PI subtypes have poorer survival outcomes. The former harbors *KRAS* mutations and inactivation of *STK11*, while the latter has co-mutations of *NF1* and *TP53*. However, stratification of these subtypes is mainly based on gene candidates with highly variant expressions across tumor samples without considering the roles of these candidates in tumor malignancy. Thus, we attempted to identify a subset of cancer essential genes to reclassify LUAD patients.

Project Achilles uses a genome-scale CRISPR-Cas9 tool to individually knock out each gene, thereby identifying candidates which are critical for cancer survival [9]. Taking advantage of Project Achilles and RNA sequencing (RNA-Seq) data from LUAD patients, we were able to pinpoint essential genes responsible for LUAD malignancy. These essential genes were used to classify LUAD patients into different molecular types. Clinical differences of these molecular types in multiple cohorts were investigated. Additionally, a new subset of patients with distinct prognoses in the TRU subtype was identified using our classification. These findings may allow us to refine the preexisting subtype classification of LUAD, and also guide us in identifying tumors that may be vulnerable to specific treatments.

## 2. Materials and Methods

### 2.1. Retrieving LUAD Patient Data and Identifying Essential LUAD Genes

RNA Seq data and clinical information of TCGA LUAD patients ( $n = 511$ ) were retrieved from UCSC Xena (<https://xena.ucsc.edu/>; 20 July 2019). Other LUAD datasets were obtained from Gene Expression Omnibus (GEO) datasets. Expression data from TCGA and GSE140343 ( $n = 51$ ) were normalized as fragments per kilobase of transcripts of million mapped reads (FPKM) and then log<sub>2</sub> transformed. The GSE68465 ( $n = 432$ ), GSE72094 ( $n = 398$ ), GSE50081 ( $n = 127$ ), and GSE31210 ( $n = 226$ ) datasets consisted of microarray data, and their expressions were normalized by robust multichip averaging with log<sub>2</sub> transformation. Genome-wide CRISPR screening of LUAD cells was downloaded from the DepMap portal (<https://depmap.org/portal/download/>; 20 July 2019). Dependency scores for around 17,000 candidate genes were calculated using the CERES algorithm [9]. Candidate genes were defined as essential genes with a CERES score of  $< -1$  across 75% of LUAD cell lines ( $n = 31$ ). Using the limma package, differentially expressed gene (DEG) analyses were conducted between tumor and paired normal tissues from TCGA and GSE140343 RNA Seq data. Candidates with a false discovery rate (FDR) of  $< 0.01$  and a multiple of change of  $> 1$  were considered to be significantly upregulated in tumor tissues.

## 2.2. Classification of LUAD Subtypes and Clinical Feature Analysis

To identify RNA expression subtypes of LUAD, i.e., TRU, PP, and PI, in a previous study [8], the subtypes were assigned to each LUAD tumor using the nearest centroid predictor. A Pearson correlation analysis was performed to correlate expression profiles of each tumor with the nearest centroid predictor. The subtype of each patient was predicted based on the maximum correlation coefficient. Essential genes we identified were used to perform an unsupervised consensus clustering with TCGA LUAD data, and a partitioning around medoids (PAM) clustering algorithm was used. One thousand permutations with a 0.95 random fraction of essential genes in each iteration were repeated to perform the clustering analyses. Judging from the delta area plot, the optimal cluster was selected according to whether no appreciable increase was present. Expressions of essential genes in TCGA were median centered, and they were used to develop the nearest centroid classifier that could predict essential gene-classified clusters. The classifier was applied to the GEO and GDSC datasets to predict subtypes of LUAD. Expressions of essential genes are shown in a heatmap. To compare 5-year overall survival in each subtype, a log-rank test was performed. A multivariate Cox regression was conducted to investigate whether the essential gene-classified cluster was an independent prognostic factor considering tumor stages. To evaluate the relation between essential-classified clusters and tumor stage, a logistic regression analysis was performed. The dependent variable was tumor stage, and we divided it into binary variable. The patients in TCGA, GSE140343, GSE68645 and GSE72094 were categorized into high stage (stage III and stage IV) and low stage (stage II and stage I) groups. For the low stage patients in GSE50081 and GSE31210, they were divided into stage II and stage I. According to previous studies [10–12], four pathological subtypes with distinct survival have been identified including lepidic predominant non-mucinous adenocarcinomas (ADCs), acinar/papillary predominant non-mucinous ADCs, micropapillary/solid predominant non-mucinous ADCs, and invasive mucinous ADCs. A Fisher exact test was performed to compare the enrichment for a certain essential gene-stratified cluster in a given pathological type.

## 2.3. Comparisons of Pathway and Transcription Factor (TF) Activities in LUAD Patients

To compare signaling pathways that differed within essential gene-classified clusters, a single-sample gene set enrichment analysis (ssGSEA) was performed to evaluate the degree of activation of Hallmark pathways [13]. TF activity was calculated according to Garcia-Alonso et al. [14]. Briefly, a group of high-confidence human TFs and their target genes were defined based on TF-binding site predictions, text mining-derived TF-target interactions, and chromatin immunoprecipitation coupled with high-throughput (ChIP-X) data. Transcriptome profiles of these TF targets were used to infer TF activities of each patient by performing analytical rank-based enrichment analyses (aREAs). The Kruskal-Wallis test with post-hoc Dunn's test was conducted to identify the top signaling pathways or TFs that significantly differed within essential gene-stratified clusters ( $p < 10^{-3}$ ). A multivariate linear regression was conducted to evaluate the association between ssGSEA-inferred E2F targets activity and essential gene-classified clusters considering tumor stage as a covariate.

## 2.4. Copy Number Variations and Mutation Analyses

Copy number variations and genomic mutation data of TCGA LUAD were downloaded from UCSC Xena (<https://xena.ucsc.edu/>; 20 July 2019). Copy number segment data were analyzed using GISTIC 2.0 to identify significant focal copy number alterations in LUAD patients. Mutation data were generated by the Multi-Center Mutation Calling in Multiple Cancers (MC3) project. Genes were divided into binary calls as either a non-silent mutation or wild type (WT). Fisher's exact test was carried out to explore changes in genomic mutations and copy numbers that were significantly enriched in specific essential gene-classified clusters ( $p < 10^{-3}$ ).

### 2.5. Drug Discovery in Essential Genes-Stratified Clusters

To uncover the drug candidates that exhibited different efficacy in essential gene-classified clusters, we queried the genomics of drug sensitivity in cancer (GDSC) database (<https://www.cancerrxgene.org>; 16 January 2020). This database contains nearly 1000 genetically characterized human cancer cell lines treated with an array of anti-cancer therapeutics (367 compounds). Among these cell lines, 46 of them belongs to LUAD. We applied our KNN prediction to classify these cell lines into our essential genes-stratified clusters and compared the differences of the area under the receiver operator characteristics curve (AUC) drug responses among these clusters. Only 257 compounds which were tested on more than 75% of the LUAD cell lines were selected to analyze. The Kruskal-Wallis test with a  $p$  value  $< 0.05$  was considered as significant difference.

### 2.6. Evaluating the Tumor Mutation Burden (TMB) and Immune Infiltration in Essential Gene-Classified Clusters

The TMB was derived from the sum of gene-coding errors, and base substitution insertions or deletions divided by the length of the human exon (38 MB) [15]. The degree of immune infiltration was measured using the method described in ESTIMATE. Briefly, an ssGSEA using RNA expression signatures related to immune cells was conducted to infer immune infiltration in tumors [16]. This method was applied to GEO and TCGA LUAD data to estimate the degree of immune infiltration. The TMB and immune infiltration were compared in essential gene-stratified clusters by the Kruskal-Wallis test with post-hoc Dunn's test.

## 3. Results

### 3.1. Identification of Essential Genes for Promoting LUAD Malignancy

An analysis pipeline was designed to identify and characterize subtypes, clinical features, and molecular profiles of LUAD patients in the present study (Figure 1A). To pinpoint crucial gene candidates responsible for LUAD malignancy, we investigated genome-wide CRISPR-based loss-of-function screens derived from DepMap. In total, 693 genes were found to be crucial for maintaining survival in 31 LUAD cell lines (Table S1). To identify which candidates among these 693 genes were aberrantly expressed in tumor tissues, DEG analyses were carried out to compare tumor tissues with paired normal tissues in TCGA (number of pairs = 59) and GSE140343 (number of pairs = 49) RNA Seq data. Thirty-six of 693 essential genes were significantly upregulated in tumor tissues (with a multiple of change of  $>1$  and an FDR of  $<0.01$ ) (Figure 1B, Tables S2–S4). Gene expression correlation analyses indicated that these 36 candidate genes were strongly associated with each other in both TCGA (median Pearson's  $R = 0.625$ ) and GSE140343 (median Pearson's  $R = 0.621$ ) (Figure 1C and Figure S1). This finding implied that cancer essential genes were coordinately modulated by common regulators. Functionally, pathway analyses revealed that these 36 LUAD essential genes were enriched in cell proliferative signaling pathways including DNA replication checkpoint, transcription involving the  $G_1/S$  transition, and DNA replication initiation (Figure 1D).

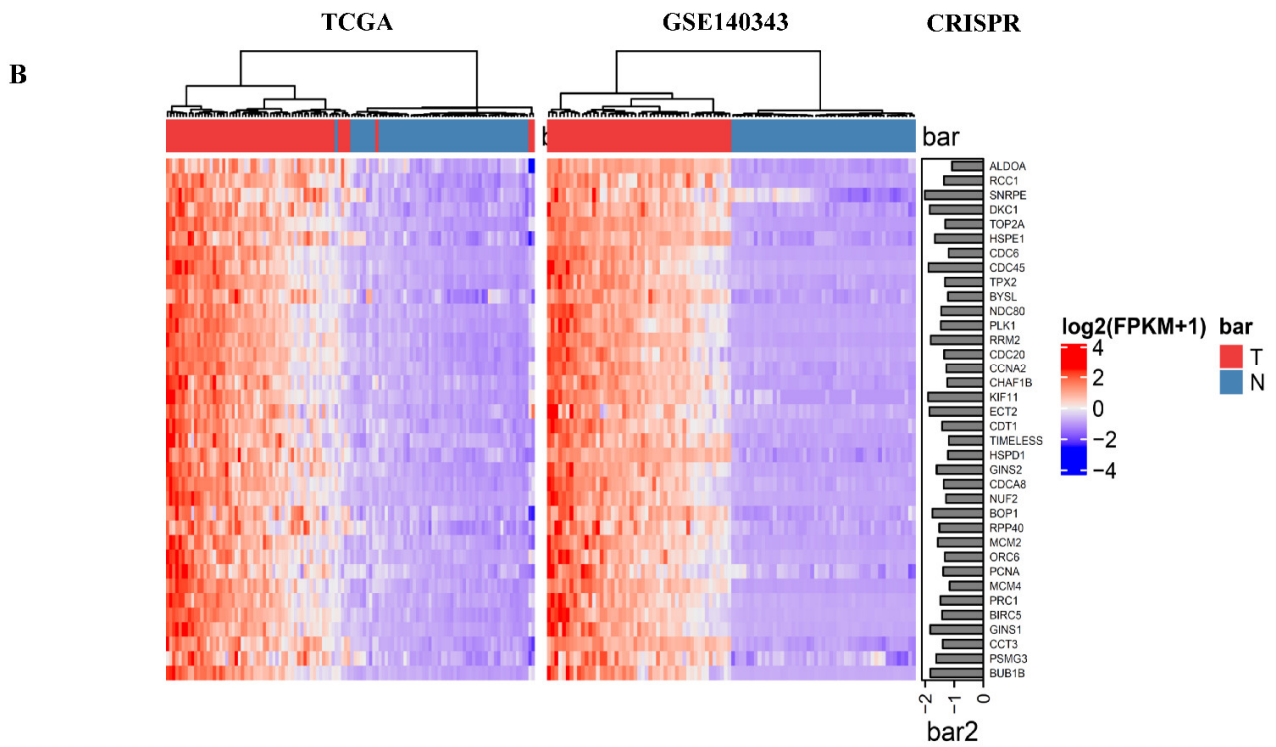
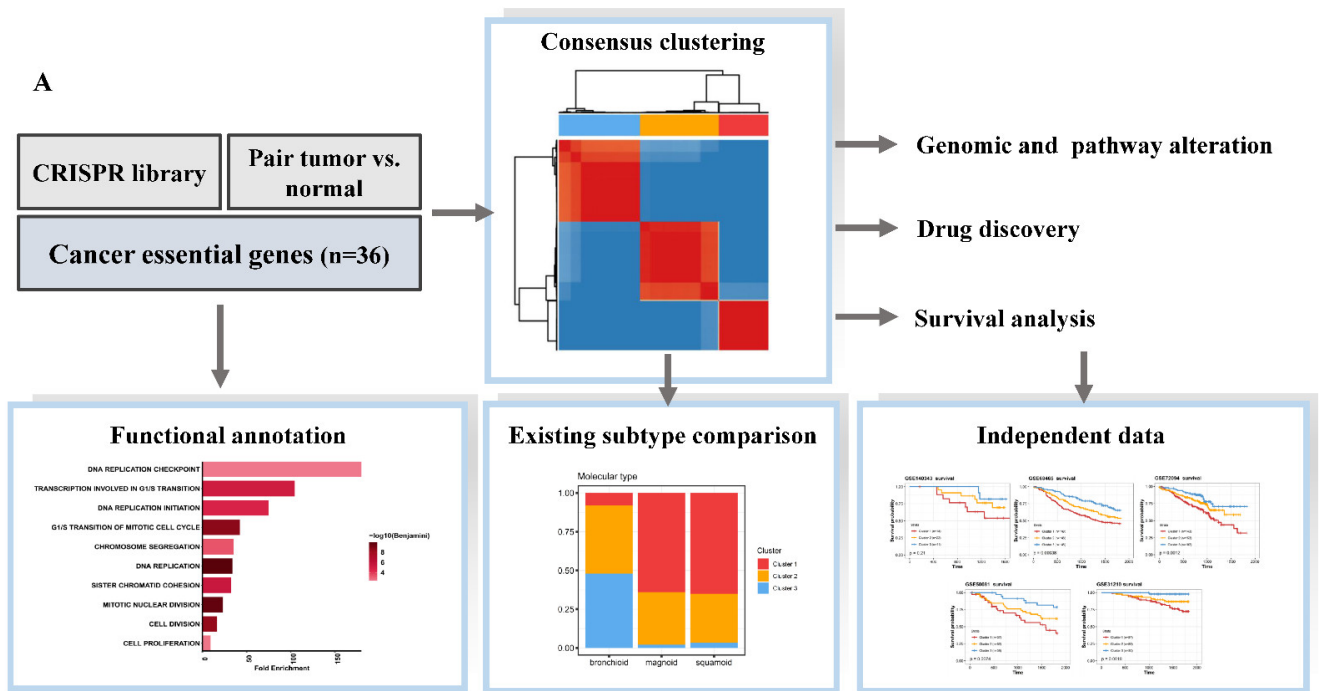
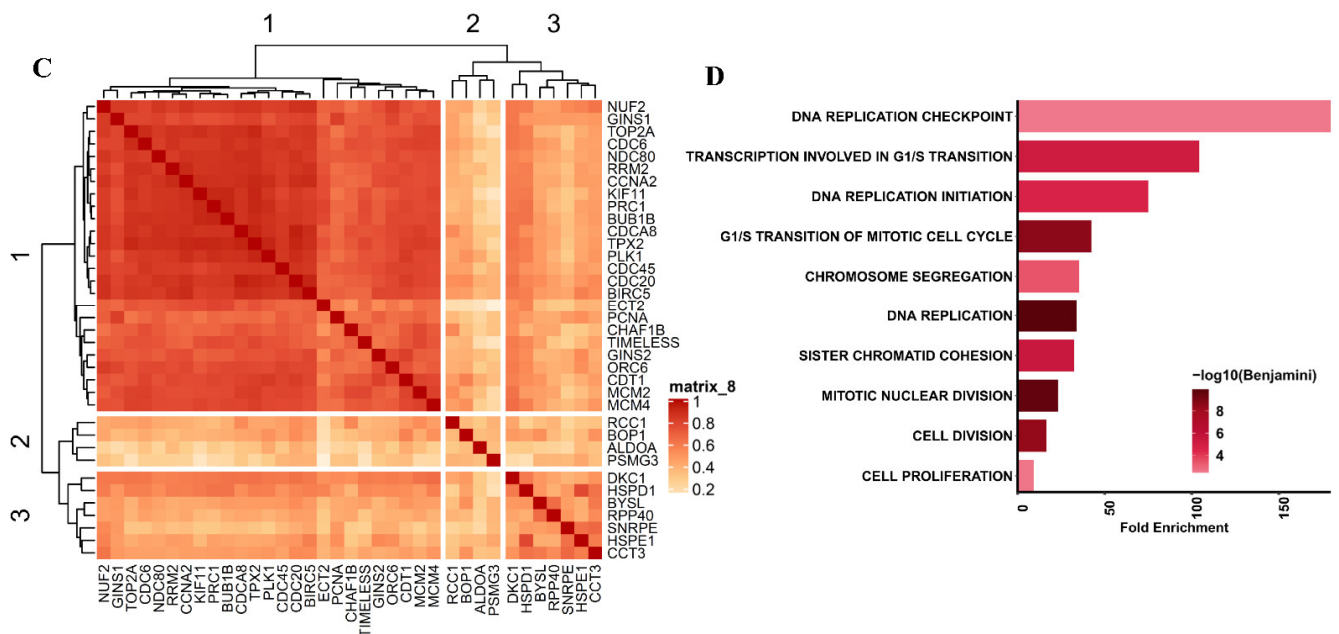


Figure 1. Cont.





**Figure 1.** Identification of cancer essential genes in lung adenocarcinoma (LUAD). (A) Flowchart demonstrating our investigation of the clinical importance and molecular associations among LUAD essential gene-stratified clusters. (B) Heatmap showing significantly upregulated genes in LUAD tumor tissues compared to paired normal tissues using RNA sequencing data of TCGA and GSE140343. The horizontal bar plot represents the median CERES score in LUAD cells from Project Achilles. (C) Associations of cancer essential genes from TCGA data are shown as a correlation heatmap. (D) Horizontal bar plot demonstrating the top essential gene-enriched signaling pathways using pathway enrichment analyses.

### 3.2. Essential Genes Stratified LUAD Patients with Different Prognoses and E2F Signaling Activities

Unsupervised consensus clustering was performed to classify TCGA LUAD patients into three robust clusters using the 36 identified essential genes (Figure 2A,B). A log-rank test demonstrated that these three clusters exhibited significantly different prognoses (log-rank test  $p = 0.0039$ ) (Figure 2C). Clusters 1, 2, and 3 respectively exhibited the worst prognosis, a moderate survival time, and the most favorable prognosis. The associations between tumor stages and essential genes-stratified clusters were analyzed by conducting logistic regression. The cluster 1 patients were significantly associated with higher tumor stage comparing with cluster 3 (log of the odds ratio = 0.99,  $p$  value < 0.01) (Table S5). Besides tumor stages, the associations between pathological types and clusters were also analyzed. We identified that cluster 3, cluster 2, and cluster 1 patients were respectively enriched in lepidic (odds ratio = 32.86,  $p$  value < 0.01), acinar/papillary (odds ratio = 1.96,  $p$  value = 0.02), and micropapillary/solid ADCs (odds ratio = 5.35,  $p$  value < 0.01) (Figure S2). A multivariate Cox regression analysis indicated that the essential gene-stratified subtype was a prognostic factor independent of the lung cancer stage in TCGA LUAD patients (cluster 1 vs. cluster 3; hazard ratio (HR) = 1.6,  $p = 0.039$ ) (Figure 2D). Expression levels of essential LUAD genes subsequently increased in the order of clusters 3, 2, and 1 (Figure 2E). To identify central regulators associated with distinct expression profiles of LUAD essential genes, we compared genomic alterations among these three clusters. Genomic mutation and copy number variation analyses revealed that the *TP53* mutation and chromosome 3q26.2 amplification were significantly enriched in cluster 1 (Figure 2F,G). The ssGSEA identified that E2F targets, the G<sub>2</sub>M checkpoint, and mitotic signaling were significantly upregulated in cluster 1 (Figure 2H, upper panel). Additionally, an aREA was performed to infer TF activities in LUAD patients. Activities of multiple E2F TFs, including E2F2, E2F3, and E2F4, were significantly higher in cluster 1 (Figure 2H, lower panel). A multivariate linear regression analysis demonstrated that the association between E2F targets and essential gene-classified clusters was independent of tumor stage (Table S6). Combining

findings of the pathway and transcription factor analyses, these clusters exhibited distinct E2F activities. Thus, these three clusters were denoted as having high (cluster 1), medium (cluster 2), and low (cluster 3) E2F activities.

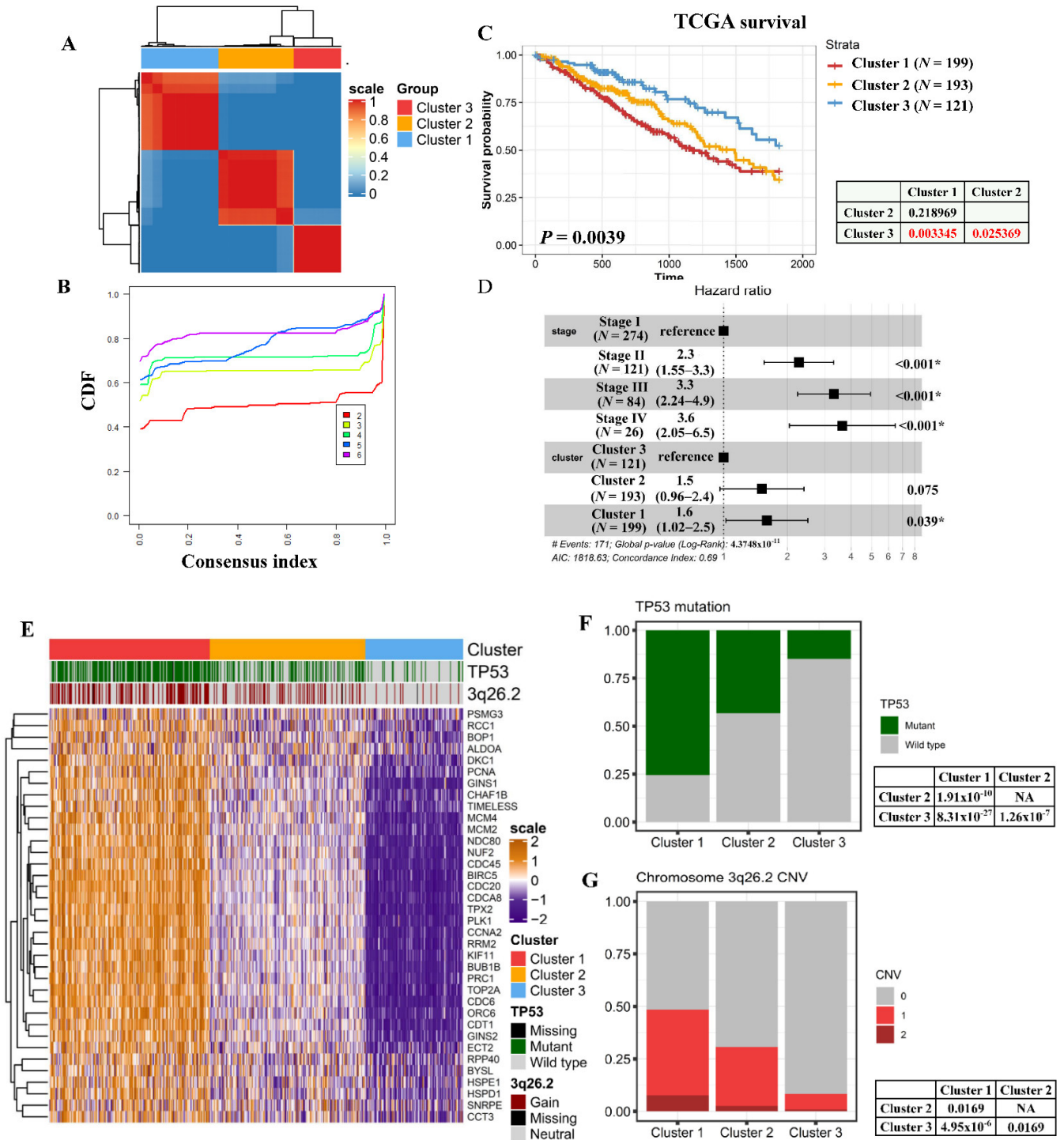
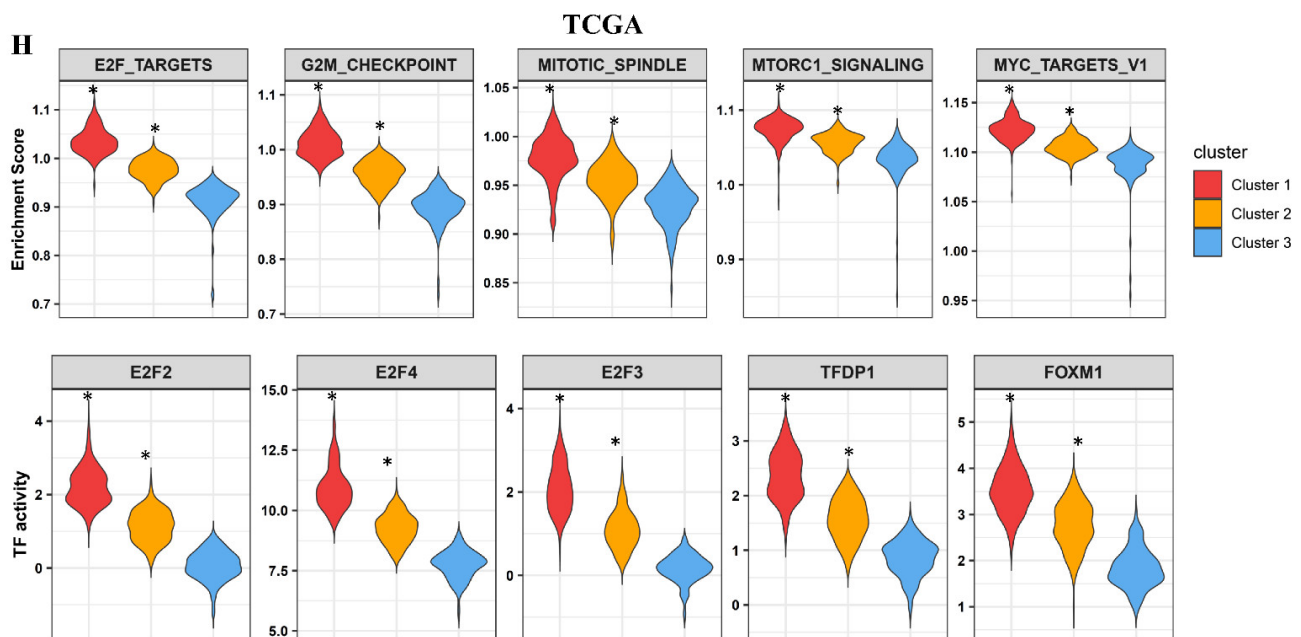


Figure 2. Cont.

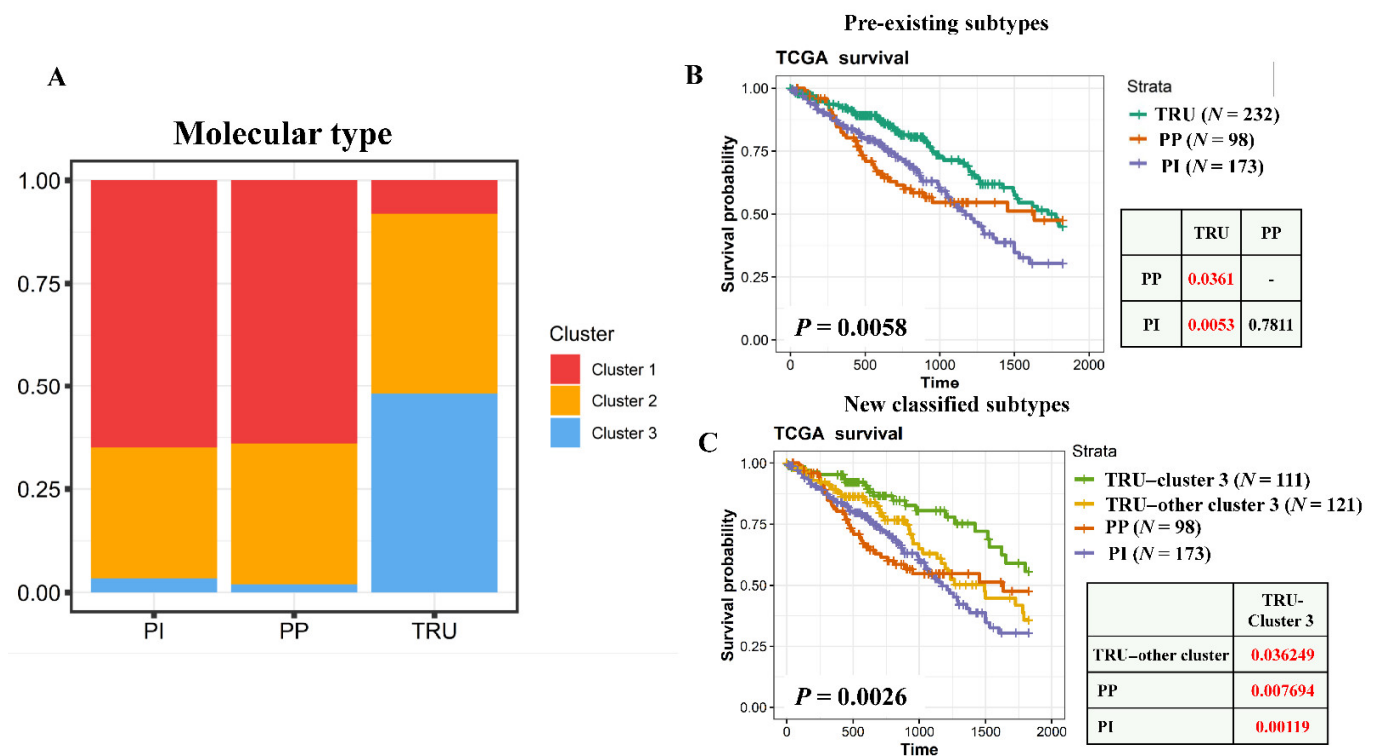


**Figure 2.** Essential gene-stratified clusters exhibit distinct survival and molecular profiles. (A) Similarity matrix of TCGA patients derived from consensus clustering assays shown as a heatmap. (B) Cumulative distribution function (CDF) plot used to decide optimal cluster numbers. (C) Survival differences within each cluster compared using a log-rank test and demonstrated as a Kaplan-Meier plot. (D) Forest plot indicating hazard ratios of essential gene-stratified clusters considering tumor stages. (E) Heatmap indicating expressions of essential genes within clusters. The status of the *TP53* mutation and chromosome 3q26.2 alterations are demonstrated as annotation plots. The frequency of *TP53* mutations (F) and chromosome 3q26.2 alterations (G) within each cluster are demonstrated. (H) Violin plots demonstrating the top signaling and transcription factor (TF) activities that significantly differed within essential gene-classified clusters. \* means  $p < 0.05$ .

### 3.3. Essential Genes Identify a New Subgroup from TRU-Type Patients with a Favorable Prognosis

A previous study divided LUAD patients into different molecular subtypes including TRU, PP, and PI based on 506 candidate gene expressions [8]. The TRU type had favorable prognoses, but PP and PI type patients had poor survival rates. The three essential gene-classified clusters were used to compare existing molecular subtypes (Figure 3A). Cluster 1 and 2 patients were highly enriched in PI (65.0% in cluster 1, 31.6% in cluster 2, and 3.4% in cluster 3) and PP (64.0% in cluster 1, 34% in cluster 2, and 2% in cluster 3). Cluster 3 patients were significantly enriched in TRU (8.1% in cluster 1, 43.8% in cluster 2, and 48.1% in cluster 3). Although both cluster 3 and TRU-type patients had favorable prognoses, nearly half (51.9%) of TRU patients contained cluster 1 and 2 types, which are poor survival groups. Therefore, we wondered whether essential gene-classified clusters could further subdivide TRU-type patients into distinct survival groups. By performing a log-rank test, TRU belonging to the cluster 3 subtype exhibited more-favorable prognoses compared to the other molecular types (Figure 3B). No survival differences were found among TRU without cluster 3, and PP and PI subtype patients. Additionally, we found that patients with lepidic pathological type were significantly enriched in TRU-cluster 3 (odds ratio = 36.86,  $p$  value < 0.001) (Figure S2). From these results, essential genes provide an additional characterization from preexisting molecular types of LUAD.





**Figure 3.** Essential gene-stratified clusters identified a subset from terminal respiratory unit (TRU) tumors with distinct proliferative signatures and prognoses. (A) Bar plot showing the frequencies of essential gene-classified clusters in previously defined molecular types. Survival differences within previously defined molecular types (B) and essential gene-classified clusters (C) are demonstrated as Kaplan-Meier plots using log-rank tests.

### 3.4. Validation of the Survival Significance of Essential Gene-Stratified Clusters in Multiple Cohorts

To validate the prominent roles of essential gene-classified clusters, TCGA LUAD patients were used as a training set to build a prediction model using the nearest centroid classifier. Because 14 of the 36 essential genes overlapped with candidates that stratified preexisting molecular types, only 22 genes remained for establishing the classification model (Figure S3A). In addition, 13 of the 36 essential genes directly belonged to E2F targets (Figure S3B). LUAD patients derived from GSE140343 ( $n = 51$ ), GSE68465 ( $n = 432$ ), GSE72094 ( $n = 398$ ), and low-stage LUAD patients derived from GSE50081 ( $n = 127$ ) and GSE31210 ( $n = 226$ ) were divided into three clusters based on the classification model. Expressions of essential genes and E2F target signaling were obviously activated in the order of clusters 3, 2, and 1 (Figures S4 and S5). Logistic regression analysis demonstrated that cluster 1 patients were significantly associated with higher tumor stage comparing with cluster 3 in GSE50081 (log of odds ratio = 1.16,  $p$  value = 0.04) and GSE31210 (log of odds ratio = 2.05,  $p$  value < 0.01) (Table S5). Additionally, GSE68465 contained the tumor grade information, and we identified that cluster 1 (log of the odd ratio = 2.16,  $p$  value < 0.01) and cluster 2 (log of the odd ratio = 2.89,  $p$  value < 0.01) patients associated with higher tumor grade compared with cluster 3 patients (Table S6). Multivariate linear regression analysis showed that the clusters were correlated with E2F target activity independent of tumor stages in all the cohorts (Table S7). Log-rank survival analyses demonstrated that cluster 1 patients had the significantly worst prognoses, while cluster 3 patients had favorable survival rates in the GSE68465, GSE72094, GSE50081, and GSE31210 datasets (Figure 4A). Although survival times did not significantly differ across essential gene-classified clusters in the GSE140343 dataset, their survival trends still followed a similar pattern as the other cohorts. The insignificant survival differences in the GSE140343 dataset might have been due to its small sample size ( $n = 51$ ). Additionally, a multivariate Cox regression confirmed

that essential gene-classified clusters acted as an independent prognostic factor considering tumor stages as covariates in the GSE68465, GSE72094, GSE50081, and GSE31210 datasets (Figure 4B). Similar to TCGA data, around half of TRU patients in these cohorts belonged to clusters 1 and 2 which exhibited high E2F signaling and proliferative signatures (Figure 4C). A survival analysis was also used to identify that TRU patients with cluster 3 genetic features exhibited better prognoses among different molecular types in three (GSE31210, GSE68465, and GSE50081) of the five datasets. These results suggested that essential LUAD genes were prominent molecular predictors that could identify LUAD patients with distinct proliferative signatures and prognoses. Importantly, this signature could be further used to refine preexisting RNA expression subtypes of LUAD.

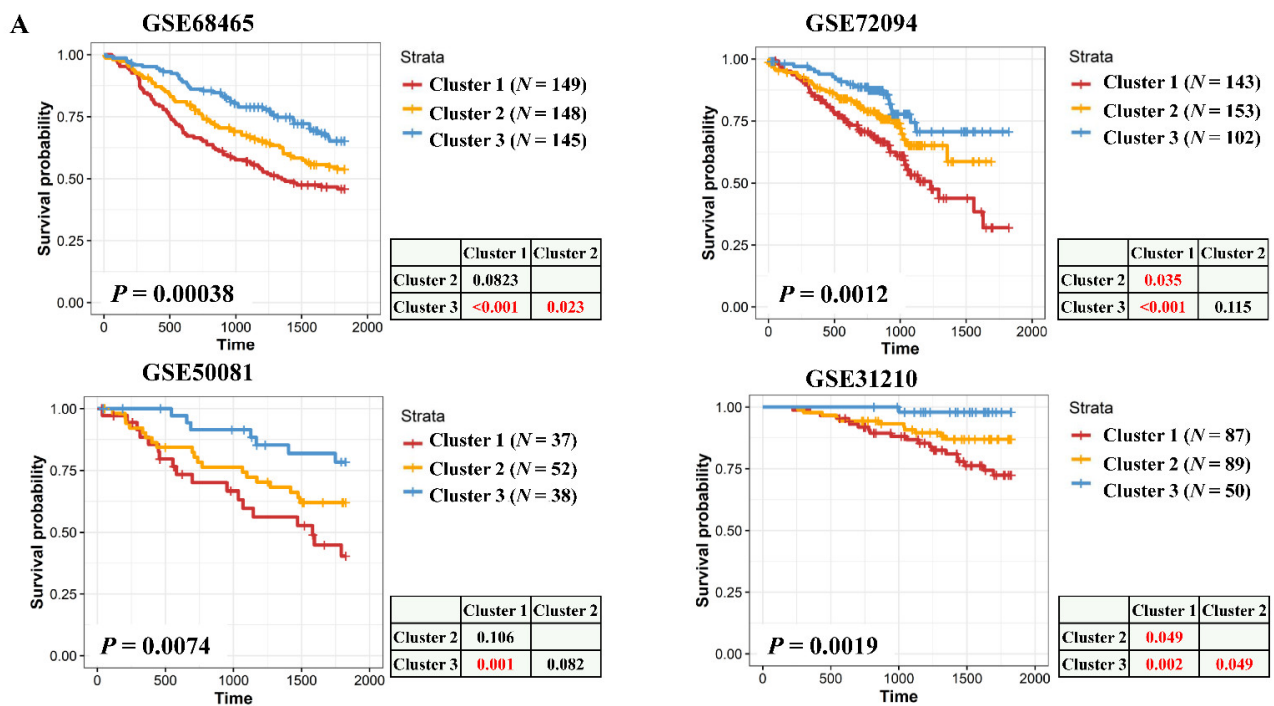


Figure 4. Cont.

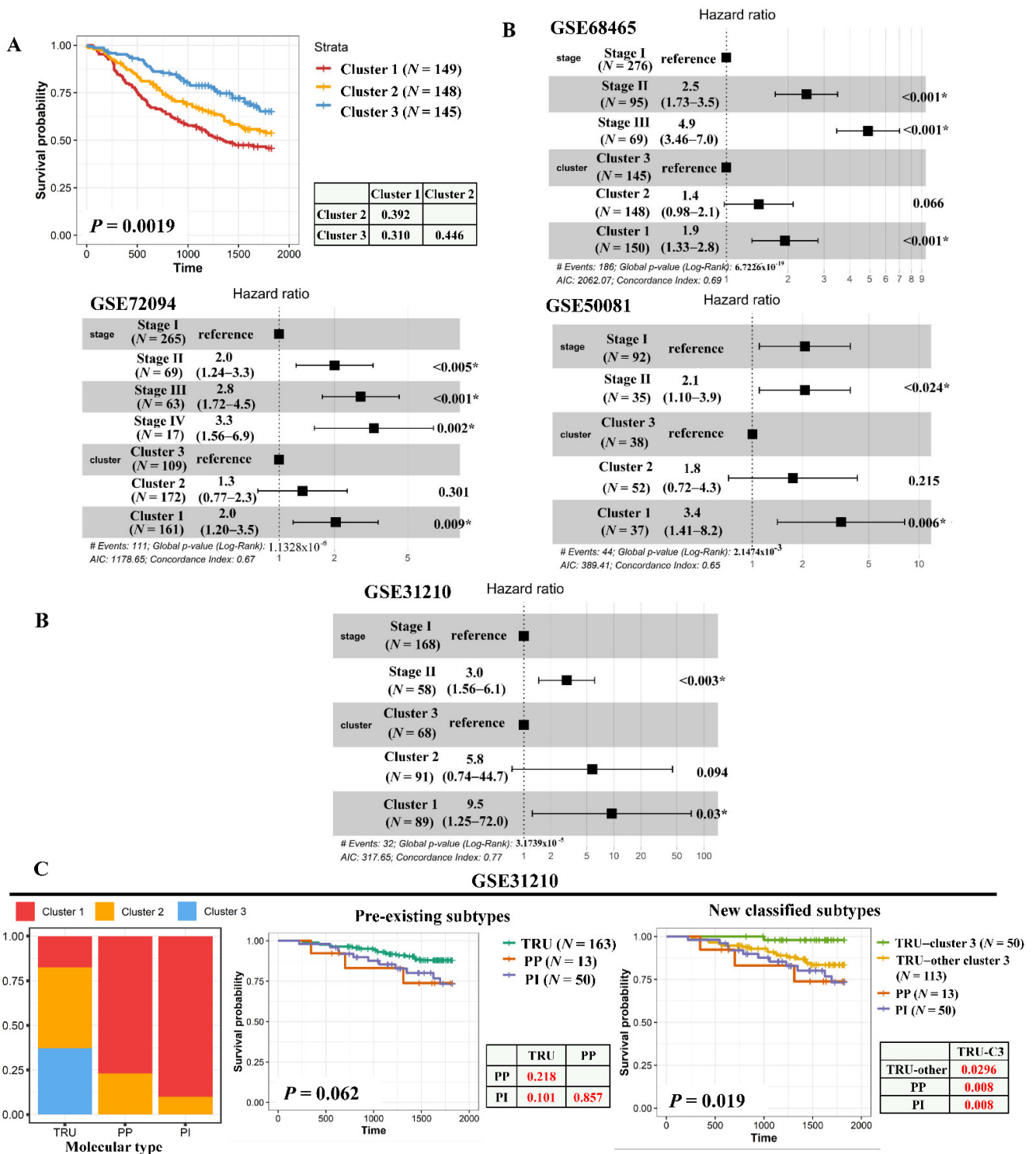
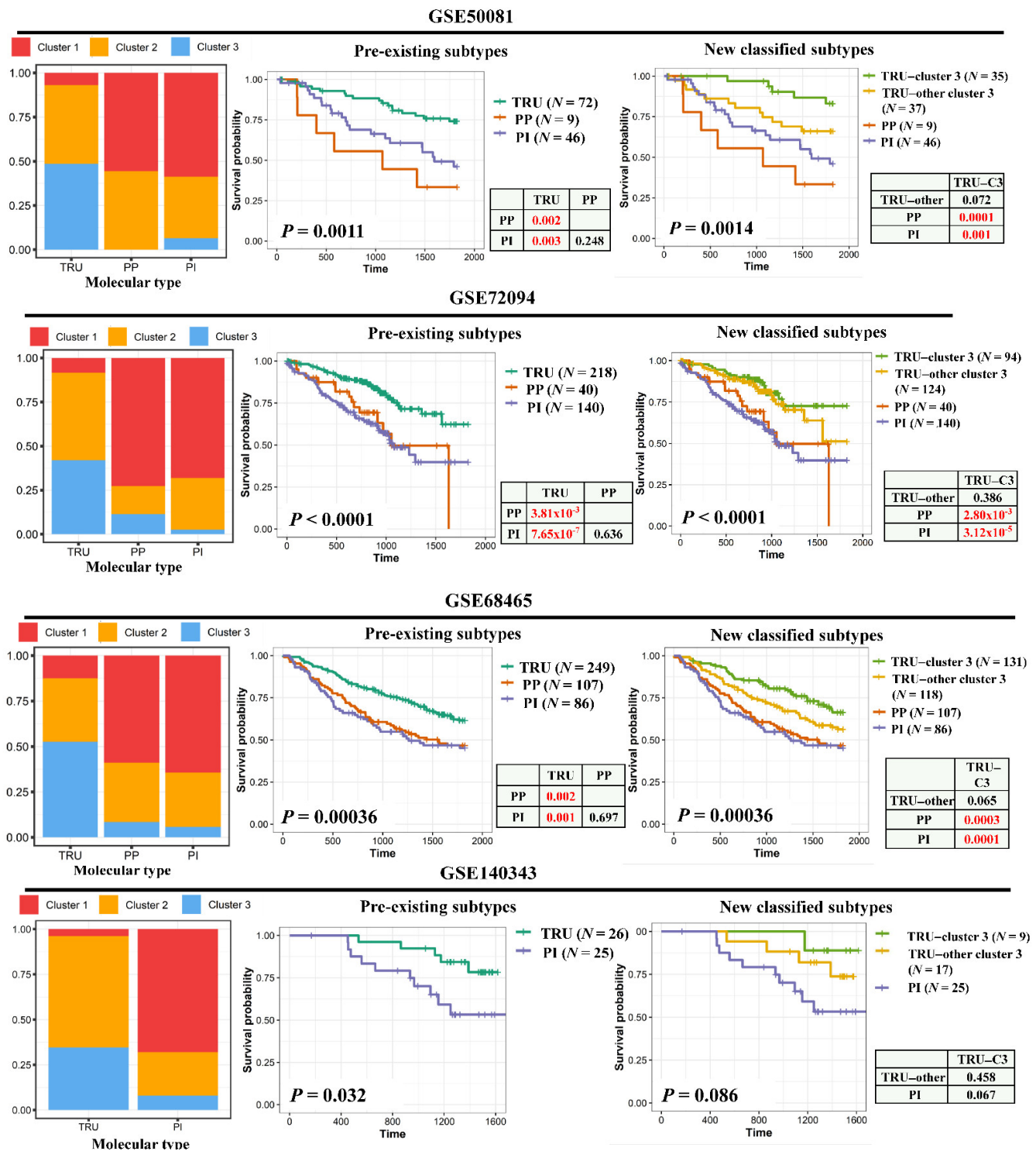


Figure 4. Cont.



**Figure 4.** Validation of survival and molecular differences of essential gene-stratified clusters of multiple GEO data. (A) Kaplan-Meier plots demonstrating survival differences of essential gene-classified clusters in the GSE140343, GSE68465, GSE72094, GSE50081, and GSE31210 datasets. Survival differences were evaluated using log-rank tests. (B) Forest plot using multivariate Cox regression analyses showing essential gene-classified clusters as an independent prognostic factor in the GSE68465, GSE72094, GSE50081, and GSE31320 datasets. (C) Distributions of essential gene-classified clusters in previously reported expression subtypes shown in stack bar plots. Kaplan-Meier plots demonstrated survival rates of different molecular types with or without considering cancer essential gene-classified clusters. \* means  $p < 0.05$ .



### 3.5. Potential Drug Discovery and Immune Environment Characterization of Essential Gene-Classified Clusters

To identify drugs that exhibited distinct efficacy in essential gene-classified clusters, the prediction model was applied to the GDSC database (Figure 5A). Areas under the receiver operator characteristics curve (AUC) of drug responses within clusters were compared. Totally, the GDSC database contains 367 compounds. Only 257 compounds which were tested on more than 75% of the LUAD cell lines were used for analyses (Tables S8 and S9). AUCs of SN-38, a topoisomerase I inhibitor, and talazoparib, a poly(ADP ribose) polymerase (PARP) inhibitor, were significantly lower in cluster 1 LUAD cells (Figure 5B). This suggested that cluster 1 tumors with highly proliferative signatures were vulnerable to drugs targeting DNA-replication mechanisms. Understanding the immune microenvironment and mutation burden in tumors could guide us in identifying tumors sensitive to immunotherapies. Thus, we compared differences in these factors within essential gene-classified clusters in TCGA and GEO datasets. The ESTIMATE-derived immune infiltration score did not prominently differ within these clusters (Figure 5C and Figure S6A), but the TMB subsequently decreased in the order of clusters 1, 2, and 3 (Figure 5D). Previously reported expression subtypes exhibited distinct immune profiles with the TRU and PI subtypes demonstrating higher immune infiltration compared to the PP subtype. Activities of E2F target signaling were lower in the TRU subtype (Figure 5E and Figure S5B). Further, categorization of these previously reported subtypes with essential gene-stratified clusters led to distinct E2F signaling (Figure 5F and Figure S6B). Herein, a combination of our essential gene-stratified clusters and previously reported expression subtypes more comprehensively captured proliferative and immune profiles of LUAD (Figure 5G). The PI type presented high immune infiltration and high proliferative features; the PP type exhibited low immune infiltration and high proliferative features; the TRU type was subdivided into high immune infiltration/high proliferative and high immune infiltration/low proliferative groups with distinct prognoses. In conclusion, E2F signaling-related essential genes could identify highly immune infiltrative TRU patients with distinct proliferative signatures and prognoses.

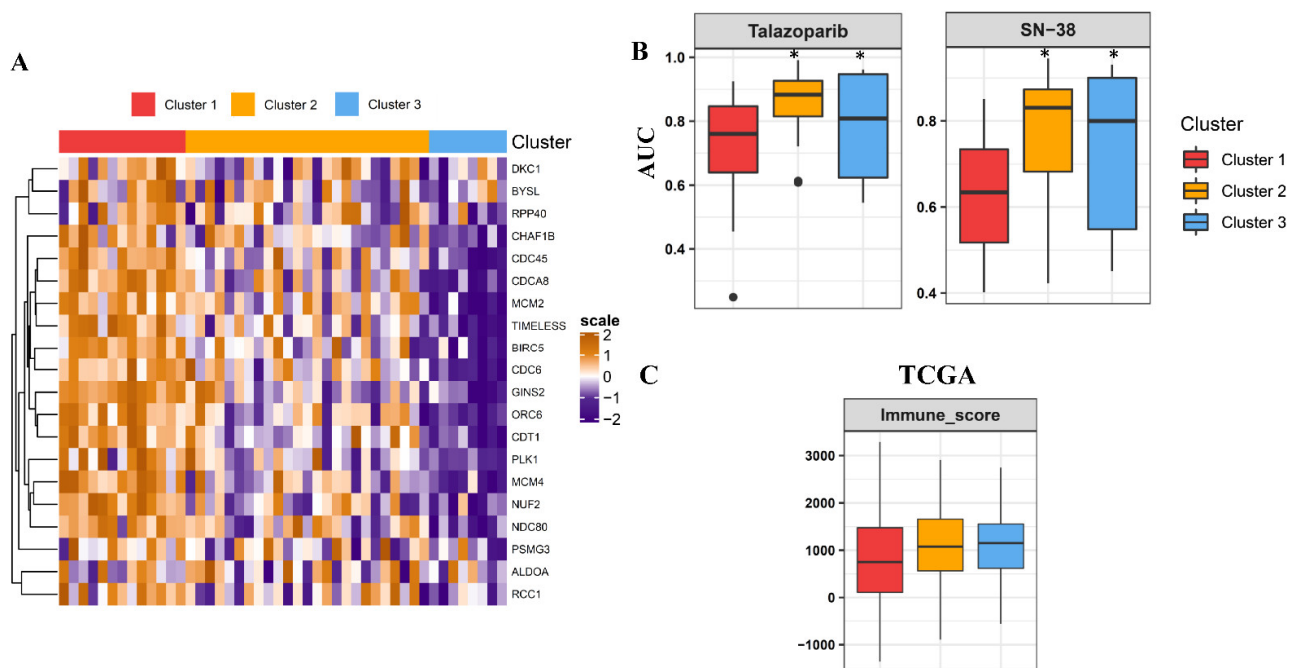
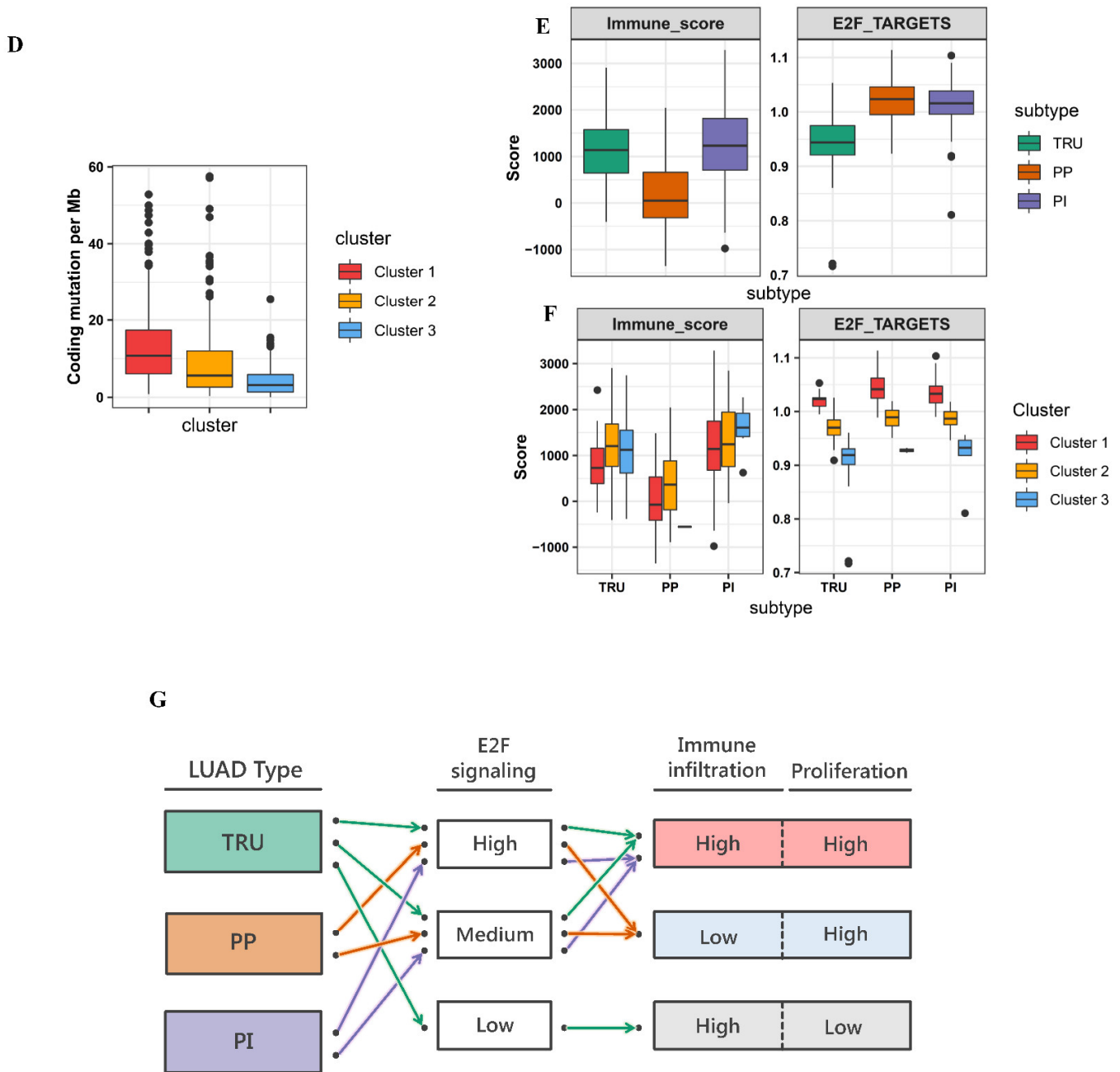


Figure 5. Cont.



**Figure 5.** Drug discovery and immune characterization in essential gene-classified clusters of lung adenocarcinoma (LUAD). (A) Heatmap showing essential gene-stratified-clusters exhibiting distinct expressions in GDSC lung cancer cells. (B) Drug response area under the receiver operator characteristics curve (AUC) of SN-38 and talazoparib within essential gene-classified clusters demonstrated as a boxplot. ESTIMATE-derived immune infiltration (C) and the tumor mutation burden (D) were compared within each cluster and are shown as boxplots. Boxplot of immune infiltration and E2F signaling within previously defined subgroups (E) and essential gene-classified clusters (F). (G) Integration of previously defined molecular types and essential gene-stratified clusters categorizing LUAD patients into distinct immune infiltration and proliferative signature groups. \* means  $p < 0.05$ .

#### 4. Discussion

Inter-patient diversity in LUAD indicates the importance of identifying genetically different subgroups with distinct survival and druggable targets. By combining Project Achilles and LUAD patient RNA Seq data, 36 cancer essential genes involved in cell proliferation pathways were nominated. Clinically, these essential gene signatures stratified LUAD patients into different survival groups in multiple cohorts. Molecularly, *TP53* mutations

and chromosome 3q26.2 amplification were enriched in patients with the worst prognoses as identified by essential genes. Additionally, E2F targets and E2F transcription activities were activated in the group with the worst prognoses. GDSC drug analyses identified that the high E2F-signaling group was sensitive to cell proliferation inhibitors including talazoparib and SN-38. Intriguingly, essential gene-classified clusters further identified a group of TRU patients with favorable prognoses and low proliferative signatures.

In our pathway analysis, we identified distinct E2F activities within essential gene-classified clusters, and 13 of 36 essential genes directly belonged to E2F targets. Specifically, we identified that E2F2/3/4 were activated in cluster 1 patients. The E2F family consists of eight members (E2F1~8), and they are TFs that are responsible for inducing G<sub>1</sub>/S and G<sub>2</sub>/M phase transitions [17–19]. It was indicated that E2F2/3 exhibit higher expressions in LUAD tumors compared to healthy lung tissues [20]. E2F2/4 were correlated with LUAD stages and negatively associated with relapse-free survival [20]. Functionally, E2F2 was identified as a direct microRNA (miR)-99a target and is involved in miR-99a- suppression of lung cancer stemness and the epithelial-to-mesenchymal transition [21]. Suppression of E2F3 was shown to synergize the cytotoxicity of paclitaxel in lung cancer cells [22]. However, few studies reported the function of E2F4 in LUAD. Hence, future studies to achieve a more detailed understanding of E2F4 are still needed. Interestingly, in the present study, we identified that E2F2, E2F3, and E2F4 were not essential genes in lung cancer cells from Project Achilles. This result suggests that E2F members have redundant roles, and knocking-out one of the E2F members might not sufficiently reduce essential gene expressions. Still, future experimental studies are needed to verify this speculation.

In the copy number variation analyses, amplification of chromosome 3q26.2 was enriched in cluster 1 patients (48%). The *protein kinase C iota (PRKCI)* gene, located in chromosome 3q26.2, was indicated to phosphorylate the cancer stemness regulator, SOX2 [23]. PRKCI/SOX2 signaling promotes the Hedgehog pathway and sustains lung squamous cell cancer stemness [23]. However, few studies have investigated the roles of chromosome 3q26.2 amplification and PRKCI function in LUAD. Gene mutation analyses revealed that non-silent somatic mutations of *TP53* were highly enriched in cluster 1 cancer patients. *TP53* is a well-known tumor suppressor, and it is frequently mutated in various cancers including LUAD. *TP53* functions as a cell cycle suppressor by promoting G<sub>1</sub>/S and G<sub>2</sub>/M arrest [24]. Thus, the loss of function of *TP53* leads to uncontrolled proliferative features of lung cancers. Moreover, one study reported that *TP53* suppresses transcription activities of E2F proteins [25]. A pan-cancer analysis of genomic profiles of patients indicated that E2F signaling is activated in *TP53*-mutant tumors [26]. Those previous results echo our findings that activation of E2F signaling accompanies *TP53* mutations. However, 24% of LUAD patients in cluster 1 harbored tumors with WT *TP53*. No significant differences in E2F activities between WT and mutant *TP53* were found in cluster 1 (Wilcox  $p = 0.1794$ , data not shown). These findings might be explained by other non-mutational mechanisms suppressing *TP53*'s functions, since MDM2, MDM4, and PPM1D were identified to function as negative regulators of *TP53* [27–29]. It is worth noting that *MDM2* gene expression was significantly higher with WT *TP53* compared to the mutant type (Wilcox  $p = 0.03829$ , data not shown) in cluster 1. Taken together, these findings imply that upstream genomic mutations might not comprehensively capture clinically distinct patients, since alternative signaling might compensate for their roles. In contrast, cancer essential genes which are predominately involved in direct processes of the cell cycle can more accurately reflect clinical differences in LUAD patients.

Immune infiltration and the TMB are crucial factors determining the efficacy of immunotherapy. Our data demonstrated significant differences in the mutation burden within essential gene-classified clusters. Cluster 1 patients had the highest mutation burden. Because *TP53* maintains genomic stability [30], loss of function of *TP53* accompanied by activation of cell proliferative signatures might lead to genomic instability and the high TMB in cluster 1 patients. No significant differences in immune cell infiltration within essential gene-classified clusters were identified, since these essential genes were derived

from cancer cell line knockout experiments without considering the tumor microenvironment. In contrast, previously reported expression subtypes were demonstrated to possess distinct immune infiltration levels [31], but they did not exhibit prominent differences in E2F signaling activities.

In our analysis, we identified that essential genes expression exhibited an increased trend in an order of cluster 3, 2 and 1 judging from the heatmap from Figure 2E, Figures S4 and S5. Further, a strong association among these essential genes implicated that it is feasible to reduce the genetic signature and develop a simplifier prediction model. In the future, a functional study is needed to better characterize the roles of essential genes in LUAD. This could further guide us to categorize those genes into functional distinct subsets. Then, we could select more representative genes from those subsets to refine our predictive model.

## 5. Conclusions

An integration of our classification with previously reported subtypes provides a more thorough understanding of immune and proliferative profiles of LUAD. The TRU subtype with high immune infiltration could be further divided into low and high proliferative groups based on our identified essential genes. Consequently, these findings can guide us in identifying subgroups of LUAD tumors that may be vulnerable to immunotherapy or cell proliferation inhibitors in the future.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13092128/s1>, Figure S1. Associations of cancer essential genes from GSE140343, Figure S2. Association between pathological type and essential gene-stratified cluster, Figure S3. List of 22 selected essential genes and 13 E2F target genes, Figure S4 Validation of expression patterns in essential gene-classified clusters, Figure S5 Validation of associated signaling in essential gene-classified clusters, Figure S6 Immune infiltration and E2F signaling profiles, Table S1: 31 LUAD cell lines, Table S2: 36 essential genes in TCGA, Table S3: 36 essential genes in GSE140343, Table S4: The CERES Score of 36 essential genes, Table S5: Association between essential gene-classified clusters and tumor stages, Table S6: Association between essential gene-classified clusters and tumor grade, Table S7: Association between essential gene-classified clusters and E2F targets activity after adjusting tumor stages, Table S8: The LUAD cell lines derived from GDSC database, Table S9: Compounds for drug discovery analysis.

**Author Contributions:** K.-H.H.: Conceptualization, Methodology, Writing—Original Draft; T.-W.H.: Validation, Data Curation; A.-J.L.: Visualization, Funding acquisition; C.-M.S.: Formal analysis, Data Curation, Supervision; K.-C.C.: Writing—Review & Editing, Supervision, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was sponsored by the Ministry of Science and Technology, Taiwan (grant no. MOST 109-2320-B-038-014 to Ku-Chung Chen), Taipei City Government (grant no. 11001-62-022 to Ann-Jeng Liu), and Taipei City Hospital Ren-Ai Branch (grant no. TPCH-110-07 to Ann-Jeng Liu). The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy concerns.

**Conflicts of Interest:** The authors declare that there are no conflict of interest associated with this work.

## References

1. Siegel, R.; Naishadham, D.; Jemal, A. Cancer statistics. *CA Cancer J. Clin.* **2012**, *62*, 10–29. [[CrossRef](#)]
2. Herbst, R.S.; Heymach, J.V.; Lippman, S.M. Lung cancer. *N. Engl. J. Med.* **2008**, *359*, 1367–1380. [[CrossRef](#)]



3. Paez, J.G.; Jänne, P.A.; Lee, J.C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F.J.; Lindeman, N.; Boggon, T.J.; et al. EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* **2004**, *304*, 1497–1500. [[CrossRef](#)]
4. Kwak, E.L.; Bang, Y.J.; Camidge, D.R.; Shaw, A.T.; Solomon, B.; Maki, R.G.; Ou, S.H.; Dezube, B.J.; Jänne, P.A.; Costa, D.B.; et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **2010**, *363*, 1693–1703. [[CrossRef](#)]
5. Bergethon, K.; Shaw, A.T.; Ou, S.H.; Katayama, R.; Lovly, C.M.; McDonald, N.T.; Massion, P.P.; Siwak-Tapp, C.; Gonzalez, A.; Fang, R.; et al. ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **2012**, *30*, 863–870. [[CrossRef](#)]
6. Drilon, A.; Wang, L.; Hasanovic, A.; Suehara, Y.; Lipson, D.; Stephens, P.; Ross, J.; Miller, V.; Ginsberg, M.; Zakowski, M.F.; et al. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov.* **2013**, *3*, 630–635. [[CrossRef](#)] [[PubMed](#)]
7. Hayes, D.N.; Monti, S.; Parmigiani, G.; Gilks, C.B.; Naoki, K.; Bhattacharjee, A.; Socinski, M.A.; Perou, C.; Meyerson, M. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **2006**, *24*, 5079–5090. [[CrossRef](#)]
8. Collisson, E.A.; Campbell, J.D.; Brooks, A.N.; Berger, A.H.; Lee, W.; Chmielecki, J.; Beer, D.G.; Cope, L.; Creighton, C.J.; Danilova, L.; et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550.
9. Meyers, R.M.; Bryan, J.G.; McFarland, J.M.; Weir, B.A.; Sizemore, A.E.; Xu, H.; Dharia, N.V.; Montgomery, P.G.; Cowley, G.S.; Pantel, S.; et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **2017**, *49*, 1779–1784. [[CrossRef](#)] [[PubMed](#)]
10. Lee, H.Y.; Cha, M.J.; Lee, K.S.; Lee, H.Y.; Kwon, O.J.; Choi, J.Y.; Kim, H.K.; Choi, Y.S.; Kim, J.; Shim, Y.M. Prognosis in Resected Invasive Mucinous Adenocarcinomas of the Lung: Related Factors and Comparison with Resected Nonmucinous Adenocarcinomas. *J. Thorac. Oncol.* **2016**, *11*, 1064–1073. [[CrossRef](#)] [[PubMed](#)]
11. Cha, M.J.; Lee, H.Y.; Lee, K.S.; Jeong, J.Y.; Han, J.; Shim, Y.M.; Hwang, H.S. Micropapillary and solid subtypes of invasive lung adenocarcinoma: Clinical predictors of histopathology and outcome. *J. Thorac. Cardiovasc. Surg.* **2014**, *147*, 921–928. [[CrossRef](#)] [[PubMed](#)]
12. Travis, W.D.; Brambilla, E.; Riely, G.J. New pathologic classification of lung cancer: Relevance for clinical practice and clinical trials. *J. Clin. Oncol.* **2013**, *31*, 992–1001. [[CrossRef](#)] [[PubMed](#)]
13. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)]
14. Garcia-Alonso, L.; Iorio, F.; Matchan, A.; Fonseca, N.; Jaaks, P.; Peat, G.; Pignatelli, M.; Falcone, F.; Benes, C.H.; Dunham, I.; et al. Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. *Cancer Res.* **2018**, *78*, 769–780. [[CrossRef](#)] [[PubMed](#)]
15. Bai, R.; Lv, Z.; Xu, D.; Cui, J. Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. *Biomark. Res.* **2020**, *8*, 34. [[CrossRef](#)]
16. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Treviño, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612. [[CrossRef](#)]
17. Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **1998**, *12*, 2245–2262. [[CrossRef](#)]
18. Polager, S.; Kalma, Y.; Berkovich, E.; Ginsberg, D. E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. *Oncogene* **2002**, *21*, 437–446. [[CrossRef](#)]
19. Ishida, S.; Huang, E.; Zuzan, H.; Spang, R.; Leone, G.; West, M.; Nevins, J.R. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol. Cell Biol.* **2001**, *21*, 4684–4699. [[CrossRef](#)]
20. Sun, C.C.; Zhou, Q.; Hu, W.; Li, S.J.; Zhang, F.; Chen, Z.L.; Li, G.; Bi, Z.Y.; Bi, Y.Y.; Gong, F.Y.; et al. Transcriptional E2F1/2/5/8 as potential targets and transcriptional E2F3/6/7 as new biomarkers for the prognosis of human lung carcinoma. *Aging* **2018**, *10*, 973–987. [[CrossRef](#)]
21. Feliciano, A.; Garcia-Mayea, Y.; Jubierre, L.; Mir, C.; Hummel, M.; Castellvi, J.; Hernández-Losa, J.; Paciucci, R.; Sansano, I.; Sun, Y.; et al. miR-99a reveals two novel oncogenic proteins E2F2 and EMR2 and represses stemness in lung cancer. *Cell Death Dis.* **2017**, *8*, e3141–e3153. [[CrossRef](#)]
22. Kurtyka, C.A.; Chen, L.; Cress, W.D. E2F inhibition synergizes with paclitaxel in lung cancer cell lines. *PLoS ONE* **2014**, *9*, e96357–e96366. [[CrossRef](#)]
23. Justilien, V.; Walsh, M.P.; Ali, S.A.; Thompson, E.A.; Murray, N.R.; Fields, A.P. The PRKCI and SOX2 oncogenes are coamplified and cooperate to activate Hedgehog signaling in lung squamous cell carcinoma. *Cancer Cell* **2014**, *25*, 139–151. [[CrossRef](#)] [[PubMed](#)]
24. Bargonetti, J.; Manfredi, J.J. Multiple roles of the tumor suppressor p53. *Curr. Opin. Oncol.* **2002**, *14*, 86–91. [[CrossRef](#)] [[PubMed](#)]
25. Vaishnav, Y.N.; Pant, V. Differential regulation of E2F transcription factors by p53 tumor suppressor protein. *DNA Cell Biol.* **1999**, *18*, 911–922. [[CrossRef](#)]
26. Donehower, L.A.; Soussi, T.; Korkut, A.; Liu, Y.; Schultz, A.; Cardenas, M.; Li, X.; Babur, O.; Hsu, T.K.; Lichtarge, O.; et al. Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Rep.* **2019**, *28*, 1370–1384. [[CrossRef](#)]
27. Soussi, T.; Kroemer, G. MDM2-TP53 Crossregulation: An Underestimated Target to Promote Loss of TP53 Function and Cell Survival. *Trends Cancer* **2018**, *4*, 602–605. [[CrossRef](#)] [[PubMed](#)]

28. Wasylshen, A.R.; Lozano, G. Attenuating the p53 Pathway in Human Cancers: Many Means to the Same End. *Cold Spring Harb. Perspect. Med.* **2016**, *6*, a026211–a026233. [[CrossRef](#)] [[PubMed](#)]
29. Lu, X.; Nguyen, T.A.; Moon, S.H.; Darlington, Y.; Sommer, M.; Donehower, L.A. The type 2C phosphatase Wip1: An oncogenic regulator of tumor suppressor and DNA damage response pathways. *Cancer Metastasis Rev.* **2008**, *27*, 123–135. [[CrossRef](#)] [[PubMed](#)]
30. Stracquadanio, G.; Wang, X.; Wallace, M.D.; Grawenda, A.M.; Zhang, P.; Hewitt, J.; Zeron-Medina, J.; Castro-Giner, F.; Tomlinson, I.P.; Goding, C.R.; et al. The importance of p53 pathway genetics in inherited and somatic cancer genomes. *Nat. Rev. Cancer* **2016**, *16*, 251–265. [[CrossRef](#)] [[PubMed](#)]
31. Faruki, H.; Mayhew, G.M.; Serody, J.S.; Hayes, D.N.; Perou, C.M.; Lai-Goldman, M. Lung Adenocarcinoma and Squamous Cell Carcinoma Gene Expression Subtypes Demonstrate Significant Differences in Tumor Immune Landscape. *J. Thorac. Oncol.* **2017**, *12*, 943–953. [[CrossRef](#)] [[PubMed](#)]