



Integrative pharmacogenomics revealed three subtypes with different immune landscapes and specific therapeutic responses in lung adenocarcinoma



Xiaoyong Ge^{a,b,c,1}, Zaoqu Liu^{a,b,c,*,1}, Siyuan Weng^{a,b,c}, Hui Xu^{a,b,c}, Yuyuan Zhang^a, Long Liu^d, Qin Dang^e, Chunguang Guo^f, Richard Beatson^g, Jinhai Deng^h, Xinwei Han^{a,b,c,*}

^a Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China

^b Interventional Institute of Zhengzhou University, Zhengzhou, Henan 450052, China

^c Interventional Treatment and Clinical Research Center of Henan Province, Zhengzhou, Henan 450052, China

^d Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China

^e Department of Colorectal Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China

^f Department of Endovascular Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China

^g King's College London, School of Cancer and Pharmaceutical Sciences, Guy's Cancer Centre, London, UK

^h Richard Dumbleby Laboratory of Cancer Research, School of Cancer and Pharmaceutical Sciences, King's College London, London, UK

ARTICLE INFO

Article history:

Received 4 March 2022

Received in revised form 30 June 2022

Accepted 30 June 2022

Available online 2 July 2022

Keywords:

Lung adenocarcinoma
Pharmacogenomics
Molecular subtypes
Immune landscapes
Precision medication
Therapeutic responses

ABSTRACT

Background: Pharmacogenomics is crucial for individualized drug therapy and plays an increasingly vital role in precision medicine decision-making. However, pharmacogenomics-based molecular subtypes and their potential clinical significance remain primarily unexplored in lung adenocarcinoma (LUAD).

Methods: A total of 2065 samples were recruited from eight independent cohorts. Pharmacogenomics data were generated from the profiling of relative inhibition simultaneously in mixtures (PRISM) and the genomics of drug sensitivity in cancer (GDSC) databases. Multiple bioinformatics approaches were performed to identify pharmacogenomics-based subtypes and find subtype-specific properties.

Results: Three reproducible molecular subtypes were found, which were independent prognostic factors and highly associated with stage, survival status, and accepted molecular subtypes. Pharmacogenomics-based subtypes had distinct molecular characteristics: S-I was inflammatory, proliferative, and immune-evasion; S-II was proliferative and genetics-driven; S-III was metabolic and methylation-driven. Finally, our study provided subtype-guided personalized treatment strategies: Immune checkpoint blockers (ICBs), doxorubicin, tipifarnib, AZ628, and AZD6244 were for S-I; Cisplatin, camptothecin, roscovitine, and A.443654 were for S-II; Docetaxel, paclitaxel, vinorelbine, and BIBW2992 were for S-III.

Conclusion: We provided a novel molecular classification strategy and revealed three pharmacogenomics-based subtypes for LUAD patients, which uncovered potential subtype-related and patient-specific therapeutic strategies.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations: LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; TRU, terminal respiratory unit; PI, proximal inflammatory; PP, proximal proliferative; ICBs, immune checkpoint blockers; PRISM, profiling of relative inhibition simultaneously in mixtures; CTRP, cancer therapeutics response portal; GDSC, genomics of drug sensitivity in cancer; CCLE, cancer cell line encyclopedia; TME, tumor microenvironment; DRGs, drug response-associated genes; NMF, non-negative matrix factorization; IGP, in-group proportion; SubMap, subclass mapping analysis; ssGSEA, single-sample gene set enrichment analysis; TMB, tumor mutation burden.

* Corresponding authors at: Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China.

E-mail addresses: liuzaoqu@163.com (Z. Liu), fchanxw@zzu.edu.cn (X. Han).

¹ These authors contributed equally.

1. Introduction

Non-small cell lung cancer (NSCLC) is a malignant tumor with high incidence and cancer-related mortality, of which lung adenocarcinoma (LUAD) is the most prevalent pathological subclass [1]. Many FDA-approved drugs, such as chemotherapy agents, targeted agents, and immune checkpoint blockers (ICBs), are available for LUAD patients. Platinum-based chemotherapy is the primary first-line treatment for NSCLC patients without driver mutations. Targeted therapy, such as epidermal growth factor receptor (EGFR)

<https://doi.org/10.1016/j.csbj.2022.06.064>

2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

tyrosine kinase inhibitors, is the standard treatment for lung cancer patients with driver mutations. In recent years, ICB (i.e., PD-1/PDL-1 blocker) has been a novel therapeutic strategy for patients with unresectable NSCLC. However, LUAD patients have a dismal survival prognosis despite multiple therapeutic options, with a 5-year relative survival rate of only 21% [2]. The poor survival rate could be attributed to different individual therapeutic responses. Tumor heterogeneity, such as different tumor microenvironment (TME) patterns or cancer cell types, is a significant cause of low response rates and drug resistance [3]. Therefore, LUAD patients are urged to develop a precise classification strategy.

LUAD molecular subtypes are currently based on genomic, transcriptomic, and epigenetic alteration and their combination [4]. The most accepted molecular subtypes were described in 2014 with 230 LUAD samples from the cancer genome atlas (TCGA) cohort, namely terminal respiratory unit (TRU), proximal inflammatory (PI), and proximal proliferative (PP) [5]. Although distinct molecular subtypes have different biological behaviors and prognoses, they have limitations in guiding clinical management decision. Therefore, we need to develop a novel pharmacogenomics-based classifier to achieve precision medication for different populations.

Cancer cell lines are the most commonly used models for defining drug efficacy. The large-scale high-throughput sequencing technique permits cell lines genome-wide analysis of drug response. Thus, we have access to a lot of pharmacogenomics data. Some available pharmacogenomics resources include the profiling of relative inhibition simultaneously in mixtures (PRISM) [6], the cancer therapeutics response portal (CTRP) [7], and the genomics of drug sensitivity in cancer (GDSC) [8]. The cell models of PRISM and CTRP are from the cancer cell line encyclopedia (CCLE). Some drugs overlap and are FDA-approved for LUAD patients. Thus, we can utilize multiple pharmacogenomics databases and obtain multiple drug response data.

This study revealed the three distinct LUAD molecular subtypes based on pharmacogenomics with different prognoses and TME patterns. Our integrative pharmacogenomics-based classification uncovered potential subtype-related and patient-specific therapeutic strategies.

2. Materials and methods

2.1. Data source and processing

RNA expression and clinical data of TCGA-LUAD samples ($n = 497$) were retrieved from UCSC Xena (<https://xenabrowser.net/datapages/>). Expression data in the form of fragments per kilobase of transcripts of million mapped reads (FPKM) were converted to trans per million (TPM) form and then \log_2 transformed for further analysis. Other LUAD datasets (GSE30219 ($n = 85$), GSE31210 ($n = 226$), GSE41271 ($n = 182$), GSE42127 ($n = 133$), GSE50081 ($n = 127$), GSE68465 ($n = 442$), and GSE72094 ($n = 398$)) were obtained from gene expression omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). The RNA microarray data of GEO datasets were normalized between arrays and then \log_2 transformed.

2.2. Drug response data collection and processing

Drug response data of LUAD cell lines were obtained from PRISM [6] and GDSC [9]. Nine FDA-approved and commonly used drugs for the treatment of LUAD were included in this study: gefitinib, erlotinib, afatinib, crizotinib, cisplatin, docetaxel, etoposide, paclitaxel, and vinorelbine. The therapeutic response data of nine drugs all appeared in the above two pharmacogenomics databases. For each drug, cell lines were divided into groups of sensitive, par-

tial response, or resistance by utilizing the mean \pm 0.5 standard deviations (SD) of the IC50, \log_{10} (IC50), EC50, or \log_{10} (EC50) values [10,11]. Cell lines with an IC50, \log_{10} (IC50), EC50, or \log_{10} (EC50) value more significant than the mean + 0.5 SD were defined as resistant to the drug. Cell lines with an IC50, \log_{10} (IC50), EC50, or \log_{10} (EC50) value less than the mean - 0.5 SD were defined as sensitive to the drug, and those with an IC50, \log_{10} (IC50), EC50, or \log_{10} (EC50) value between the mean + 0.5 SD and the mean - 0.5 SD were defined as having a partial response to the drug. This categorization corresponds to the RECIST 1.1 system (i.e., complete response, partial response, and stable disease/disease progression) in evaluating chemotherapeutic response in solid tumors [12]. The corresponding expression data of LUAD cell lines from PRISM were obtained from the CCLE (<https://sites.broadinstitute.org/ccle/datasets>) [13], and the expression data of LUAD cell lines from GDSC were collected from the GDSC1000 resource. CCLE expression data were acquired by RNA-seq-based measurements and transformed into $\log_2(\text{TPM} + 1)$. GDSC1000 was expression array data from Affymetrix Human Genome U219 array platform and normalized by robust multi-array average (RMA) algorithm. Considering the different methods of expression data generation, each gene expression was transformed into Z-score across samples in CCLE and GDSC1000 cell lines.

2.3. Identification of drug response-associated genes

The limma package was applied in PRISM and GDSC databases to screen out differentially expressed genes between sensitive and resistant groups for each drug. $P < 0.05$ and $|\log_2\text{fold change}|$ ($|\log_2\text{FC}|$) > 0.5 served as the cutoff criteria [14]. Differentially expressed genes for each drug that were both up-regulated or down-regulated in two independent datasets were selected. The combined up-regulated and down-regulated genes were identified as drug response-associated genes (DRGs) for each drug. Finally, we combined the DRGs of nine drugs and removed the duplicates as the DRGs of LUAD (Fig. 1A).

2.4. Identification and validation of pharmacogenomics-based subtypes

The DRGs expression matrix of the TCGA-LUAD samples was used to identify the pharmacogenomics-based subtypes performing the non-negative matrix factorization (NMF) consensus cluster method implemented in the R package NMF v.0.23.0. The nsNMF algorithm [15] was performed using 100 iterations for the rank (between 2 and 7 clusters) survey and 500 iterations for the clustering runs. The value of k for which the cophenetic coefficient starts decreasing is chosen as the optimal number of clusters [16]. The silhouette width profiles were assessed, and samples with negative silhouette width were excluded to identify subtype-specific characteristics [17].

Using the centroids of the TCGA-LUAD cohort, the in-group proportion (IGP) algorithm [18] implemented in the R package clusterRepro was conducted to classify LUAD samples of the seven GEO cohorts. Because the expression data generation methods are different in TCGA and GEO datasets, specifically, one based on RNA-seq and the other based on the microarray. Utilizing the *scale* method contained in the R, the gene expression values were normalized to Z-score before the IGP analysis. In addition, subclass mapping analysis (SubMap, Gene Pattern, <https://cloud.genepattern.org/gp/>) [19] was performed to assess the genetic similarity in DRGs expression profiles between subgroups from independent TCGA and GEO cohorts.

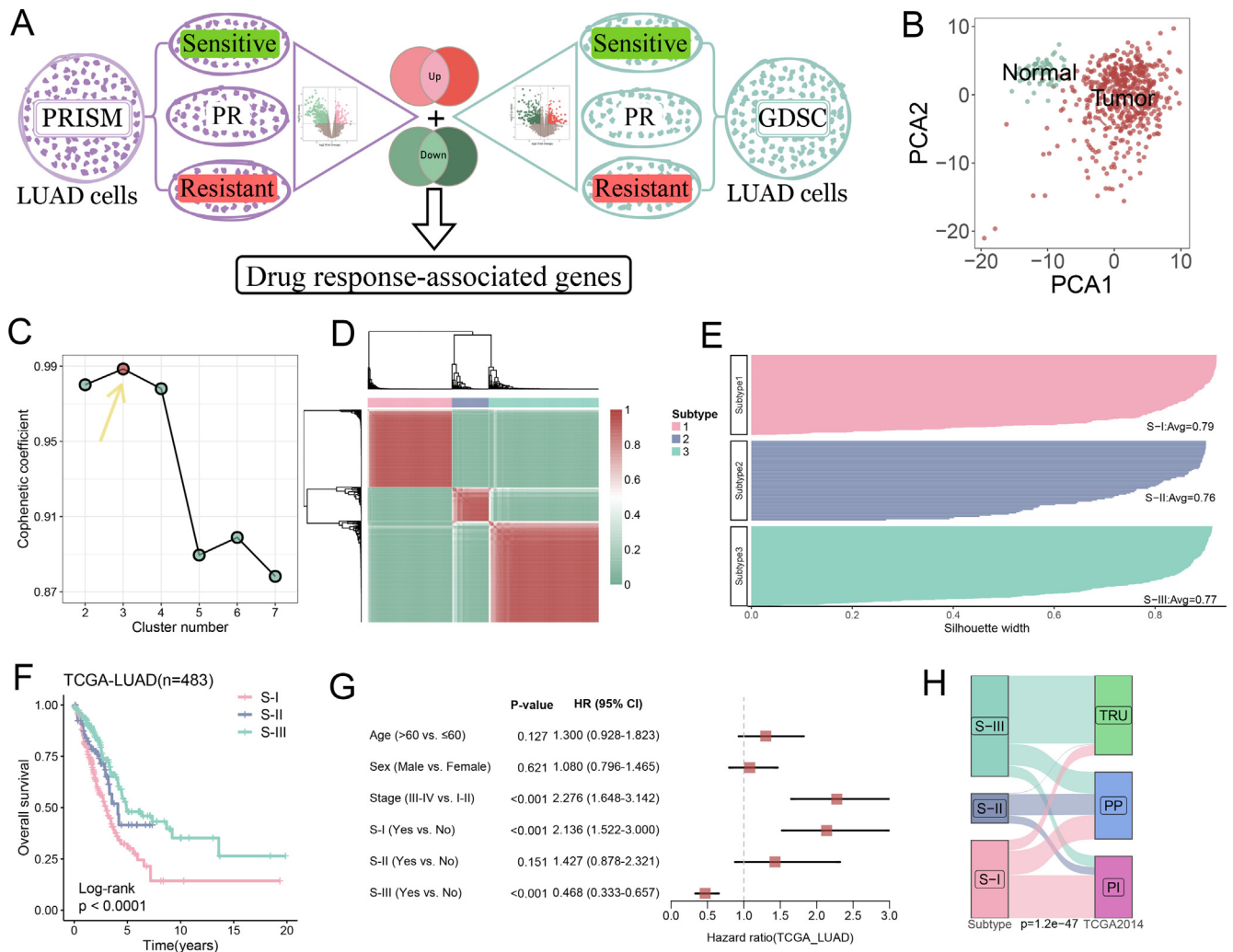


Fig. 1. Identification of pharmacogenomic-based subtypes.(A) Flowchart for screening drug response-associated genes. (B) The principal component analysis (PCA) algorithm displayed the two-dimension spatial distribution of tumor and normal samples. (C) The first rank (K = 3) for which the cophenetic coefficient starts decreasing is generally defined as the optimal rank. (D) The consensus map of NMF clustering results in the TCGA-LUAD cohort. (E) The silhouette statistic of three pharmacogenomic-based subtypes. (F) Kaplan-Meier curves of overall survival according to pharmacogenomic-based subtypes. (G) The multivariate Cox regression analysis in TCGA-LUAD cohort. (H) The links between pharmacogenomic-based subtypes and acceptable subtypes (TCGA 2014).

2.5. Associations of pharmacogenomics-based subtypes with clinical features and classical subtypes

Kaplan-Meier (K-M) analysis was performed to examine differences in survival among the three pharmacogenomics-based subtypes in the TCGA and GEO cohorts. We performed multivariate Cox regression analysis to determine whether pharmacogenomics-based subtypes in the TCGA and GEO cohorts were independent prognostic factors. The LUAD consensus molecular subtypes [13], namely TCGA 2014 (TRU, PP, and PI), were assigned to each LUAD patient of TCGA and GEO cohorts using the nearest centroid predictor. Sankey plot was used to demonstrate the links between pharmacogenomics-based subtypes and TCGA 2014. The pie chart showed the relationship between pharmacogenomics-based subtypes and clinical features.

2.6. Molecular characterization of pharmacogenomics-based subtypes

To understand the biological pathways of pharmacogenomics-based subtypes, quantitative set analysis for gene expression (QuSAGE, R package qusage v2.26.0) was conducted to compare each subtype to all others, leveraging hallmark gene sets from

molecular signature database (MSigDb) to identify pathway activity within each subtype. QuSAGE quantified pathway activity with a complete probability density function. Hallmark pathways with activity scores (adjusted $p < 0.01$) were represented as a heatmap.

To further confirm the biological properties of pharmacogenomics-based subtypes, we again contrasted each subtype to all others. We conducted pathway enrichment analysis using gene sets from the Reactome database implemented through the R package ReactomePA v1.36.0. For each subtype, pathways of adjusted $p < 0.01$ and normalized enrichment score (NES) > 0 were chosen as subtype-specific biological pathways.

Finally, we conducted a single-sample gene set enrichment analysis (ssGSEA) of gene sets for representative pathways to further refine the above discriminatory molecular characterization. Heatmap was used to represent the Z-score value of ssGSEA score for each sample of pharmacogenomics-based subtypes. In addition, we combined these metabolism-related pathways and downloaded inflammatory and proliferative signatures from the CancerSEA website (<http://biocc.hrbmu.edu.cn/CancerSEA/>). Scores of inflammatory, proliferative, and metabolic signatures were calculated by ssGSEA. The difference analysis of cancer signatures among subtypes was presented as boxplots for TCGA and GEO cohorts.

2.7. Immune landscapes of pharmacogenomics-based subtypes

The TIMER, MCPcounter, CIBERSORT, and ssGSEA algorithms assessed immune cellular components among pharmacogenomics-based subtypes. The differences in the immune response under different algorithms were uncovered using the heatmap. Correlation analyses between three subtypes and immune cell ssGSEA score were performed using Pearson correlation implemented in the R package *ggcor* v 0.9.8.1. ESTIMATE algorithms were performed to compare immune score, stromal score, ESTIMATE score, and tumor purity among pharmacogenomics-based subtypes. T cell receptor diversity (Shannon Entropy), tumor neoantigen number, immune costimulatory gene, immune coinhibitory gene, and antigen presentation gene were retrieved from previous literature [20]. Boxplots were used to visualize the differences among pharmacogenomics-based subtypes.

2.8. Molecular mechanisms of pharmacogenomics-based subtypes

Maf file of the mutation data was downloaded from the GDC website (<https://portal.gdc.cancer.gov/>). The top 20 genes with the highest mutation frequency were selected for subsequent analysis. Fisher's exact test compared the frequency of mutated genes among pharmacogenomics-based subtypes. The waterfall and frequency diagram were used to show the genes with significantly different mutation frequencies among three subtypes. We calculated the number of tumor mutations per megabase and obtained each sample's tumor mutation burden (TMB) value.

Copy number analysis data of GISTIC_2.0 level were collected from FireBrowse (<http://firebrowse.org/>). Top 20 chromosome segments with gain or loss frequency were selected for subsequent analysis. Fisher's exact test compared the frequency of gain or loss segment among subtypes. The copy number alteration burden (gain or loss) was obtained by calculating the total number of genes with copy number alteration at the focal and arm levels.

DNA methylation profile was accessed from the UCSC database. MethylMix algorithm [21] was conducted to identify DNA methylation-driven genes for each subtype. Genes whose both methylation and expression levels were different (adjusted $p < 0.05$) were selected for subsequent analysis. Boxplots were used to visualize these genes' methylation and normalized expression levels differences among subtypes. Correlation analysis of DNA methylation and mRNA expression levels was examined and pictured for each subtype. Besides, we further surveyed the correlation between mRNA expression levels and inflammatory, proliferative, and metabolic signatures scores.

2.9. Pharmacotherapy prediction for pharmacogenomics-based subtypes

To assess the sensitivity to ICBs for LUAD patients, we utilized SubMap analysis to evaluate the gene expression profiles similarity between the three subtypes and melanoma patients receiving CTLA-4/PD-1 blockers with different immunotherapy responses. Adjusted $p < 0.05$ was considered a significant sensitivity or insensitivity to ICBs.

As described above, the acquisition of DRGs was derived from gene differential expression analysis between sensitive and resistant cell lines. Thus, the expression level of up-regulated or down-regulated DRGs in each drug may predict the sensitivity of LUAD patients to this drug. Therefore, we conducted the ssGSEA analysis of up-regulated or down-regulated DRGs for each drug. Heatmap was used to represent the median Z-score value of ssGSEA score for each subtype. Furthermore, we queried the con-

nectivity map (CMap) database (<https://clue.io/>) to check the sensitivity of these drugs.

Besides, we further explored potential agents targeting the molecular pathways or genes related to the pharmacogenomics-based subtypes by utilizing the CMap database. It can predict drugs based on gene expression signatures and reveal the mode of action (MoA) of agents targeting corresponding molecular pathways. The differentially expressed genes between each subtype and others were employed to query the CMap database. The most significantly highly expressed genes of each subtype were considered potential targets of compounds. The connectivity scores of agents were calculated. Agents with connectivity scores < -95 were deemed potential therapeutic drugs for each subtype [22]. In addition, we applied the pRRophetic package to examine the therapeutic sensitivity of the above potential drug or molecular targeting, adjudicated by the half-maximal inhibitory concentration (IC₅₀) of LUAD patients.

2.10. Statistical analysis

All statistical analyses were performed utilizing the R statistical environment (R version 4.1.1). The log-rank test was used to compare the survival distributions among the three groups. The Mann-Whitney *U* test compared categorical variables and non-normally distributed variables between two groups. The Kruskal-Wallis test compared differences in more than two groups.

3. Results

3.1. Identification of drug response-associated genes

Sensitive, partially sensitive, and resistant cell lines of nine drugs were identified from the PRISM and GDSC databases, respectively. Based on differential expression analysis (see methods and Fig. 1A), a total of 195 drug response-associated genes (DRGs) were identified. PCA analysis confirmed that the 195 DRGs effectively stratified tumor and normal samples (Fig. 1B).

3.2. Identification of pharmacogenomics-based subtypes

The optimal clustering number of the NMF algorithm was three clusters (Fig. 1C). Fig. 1D showed the clustering heatmap for three subtypes, namely S-I, S-II, and S-III. The mean silhouette widths of S-I, S-II, and S-III were 0.79, 0.76, and 0.77, respectively, indicating that the samples of each subtype had an excellent internal consistency (Fig. 1E). K-M analysis displayed significant survival differences among three subtypes (Fig. 1F, $p < 0.0001$). S-I had the worst prognosis, S-III had the most excellent prognosis, and S-II had an intermediate prognosis. S-I and S-III are independent prognostic factors, which are poor and favorable prognoses, respectively (Fig. 1G). The TCGA 2014 subtypes revealed TRU had a favorable prognosis, while PI and PP had poor prognoses [5]. The comparison showed a strong correlation between pharmacogenomics-based subtypes and TCGA 2014 subtypes (Fig. 1H, $p = 1.2e-47$). S-I was similar to PI, and S-III was akin to TRU.

3.3. Validation of pharmacogenomics-based subtypes in seven cohorts

To investigate the reproducibility and utility of pharmacogenomics-based subtypes, we performed the IGP algorithm to capture the molecular subtypes in seven independent cohorts, including GSE30219, GSE31210, GSE41271, GSE42127, GSE50081, GSE68465, and GSE72094. The proportions of S-I, S-II, and S-III in seven cohorts were almost similar (Fig. S1A). The SubMap analysis revealed the significant consistency between sub-

types from independent TCGA and GEO cohorts. (Fig. S1B-H, $p < 0.05$). K-M analysis showed significantly poor survival with S-I and favorable survival with S-v, similar to the TCGA cohort outcomes (Fig. 2A-G). The multivariate Cox regression analysis findings for each GEO cohort were the same as those of the TCGA cohort (Fig. 2H-N). The connection between subtypes of each GEO cohort and TCGA 2014 subtypes was also consistent with those of the TCGA cohort (Fig. 2O-U). These results indicated that our novel pharmacogenomics-based subtypes were reliable and generalized.

3.4. Associations of pharmacogenomics-based subtypes with clinical features

The pie chart depicted the correlation between clinical features and pharmacogenomics-based subtypes in eight cohorts (Fig. S2A-H). Age distribution was associated with subtypes in three cohorts (TCGA, $p = 0.021$; GSE50081, $p = 0.025$; GSE72094, $p = 0.011$), while the distribution pattern was inconsistent. Gender distribution was associated with subtypes only in the TCGA cohort ($p = 0.018$). The stage or grade distribution was subtype-related in five cohorts, and advanced or high-grade patients were predominantly subtypes with poor prognosis (TCGA, $p = 0.0027$; GSE30219, $p = 0.031$; GSE31210, $p = 2e-05$; GSE41271, $p = 0.033$; GSE68465, $p = 1e-19$). Except for the GSE41271 cohort, the distribution of survival status was correlated with subtypes (TCGA, $p = 7.2e-05$; GSE30219, $p = 0.0056$; GSE31210, $p = 0.00046$; GSE42127, $p = 0.006$; GSE50081, $p = 0.0017$; GSE68465, $p = 7.6e-05$; GSE72094, $p = 6e-04$), and the distribution pattern was consistent with the prognosis of subtypes. In summary, patients with advanced stage and poor survival were mainly distributed in S-I, while patients with early stage and good survival were mainly distributed in S-III.

3.5. Molecular characterization of pharmacogenomics-based subtypes

The MSigDb hallmark and Reactome database were used to investigate subtype-specific biological properties deriving pharmacogenomics-based subtypes (Fig. 3A-B). Inflammatory pathways (i.e., complement, inflammatory response, interferon signaling, signaling by interleukins, and neutrophil degranulation) were enriched only in S-I. Proliferation pathways (i.e., MYC targets v1, MYC targets v2, MTORC1 signaling, G2M checkpoint, E2F targets, telomere maintenance, DNA replication, and regulation of mitotic cell cycle) were enriched in S-I and S-II. Metabolic pathways (i.e., bile acid metabolism, fatty acid metabolism, surfactant metabolism, heme metabolism, and phospholipid metabolism) were enriched only in S-III.

We further summarized the molecular pathways and calculated the ssGSEA score for each sample of pharmacogenomics-based subtypes (Fig. 3C). Thus, we defined the molecular characteristics of S-I as inflammation and proliferation, S-II as proliferation, and S-III as metabolism. We again compared the differences of inflammation, proliferation, and metabolism signature among three subtypes (Fig. 3D). S-I showed high inflammation and proliferation ($p < 0.0001$), S-II showed high proliferation ($p < 0.0001$), and S-III showed high metabolism ($p < 0.0001$). The findings of subtype-specific molecular features were validated in seven independent

cohorts, demonstrating the concordance of subtype-specific biological behaviors from different cohorts (Fig. 3E-K).

3.6. Immune landscapes of pharmacogenomics-based subtypes

Since the three subtypes have significantly different enrichment of inflammation-related pathways, we further explored the immune landscapes of pharmacogenomics-based subtypes. We utilized four algorithms to demonstrate the immune infiltrating status of the three subtypes (Fig. 4A, Fig. S3A-D). S-I belonged to the immune-hot tumor. S-II belonged to the immune-cold tumor. S-III belonged to the moderate immune infiltrating tumor.

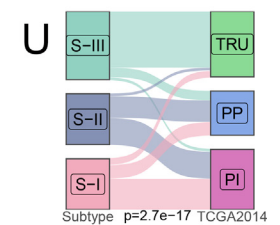
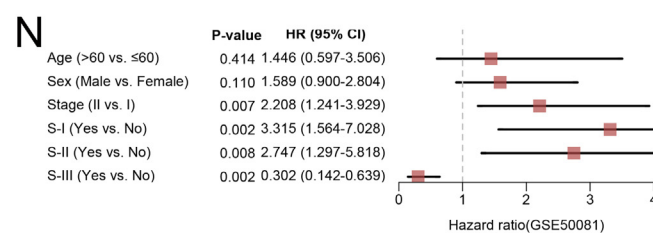
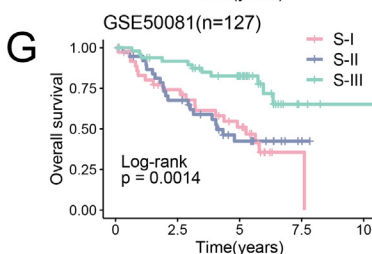
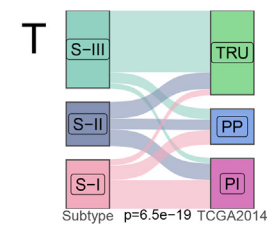
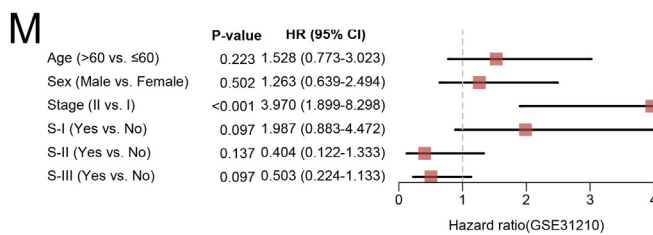
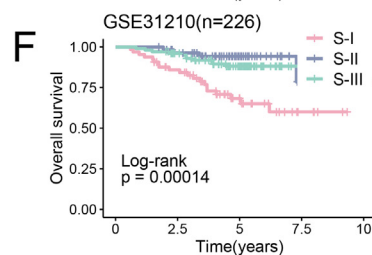
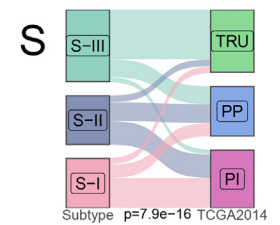
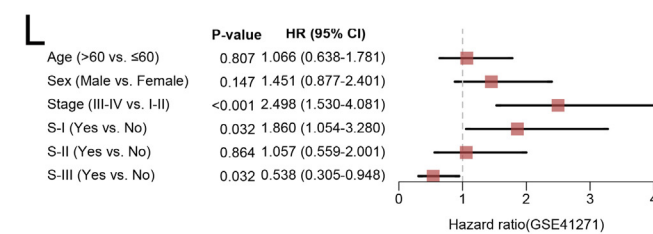
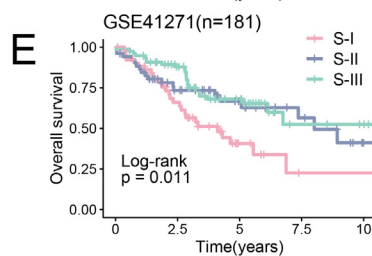
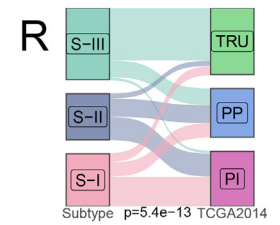
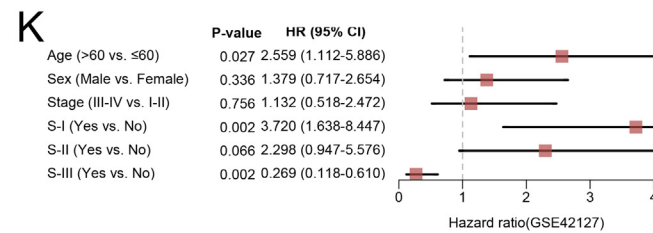
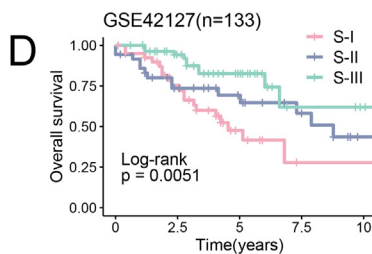
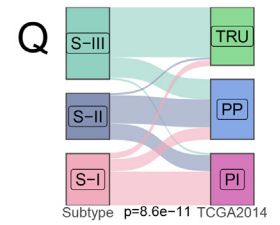
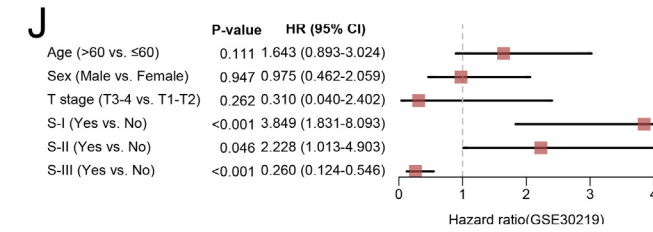
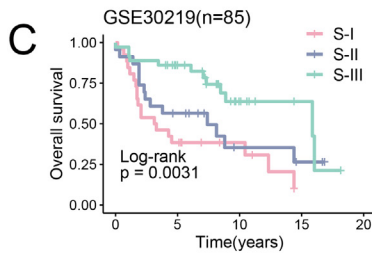
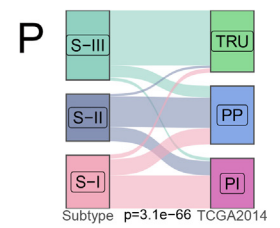
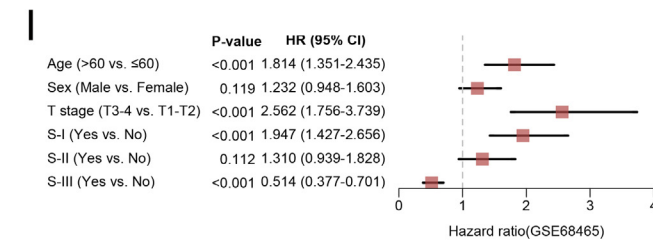
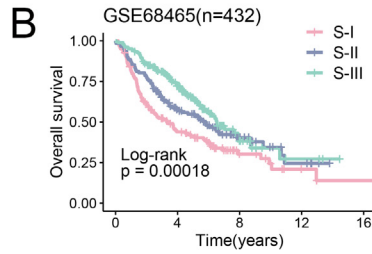
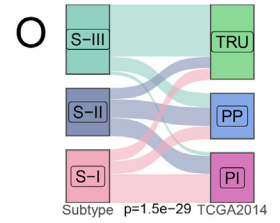
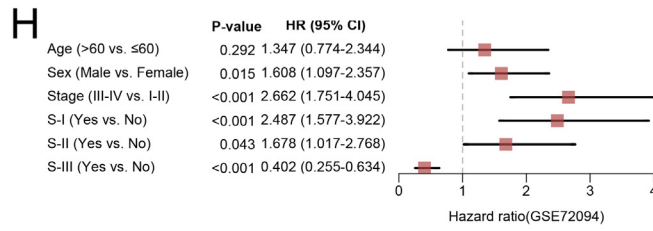
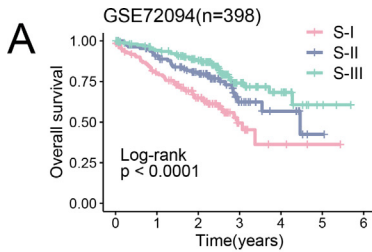
We further analyzed the correlation between each subtype and immune cell components (Fig. 4B). S-I was markedly positively correlated with T cell exhaustion, regulatory T cell, type-1 T helper cell, type-2 T helper cell activated CD4 T cell, central memory CD4 T cell, gamma delta T cell, natural killer T cell, effector memory CD8 T cell, central memory CD8 T cell, CD56bright natural killer cell, memory B cell, and neutrophil. S-II was markedly negatively correlated with T follicular helper cell, type-1 T helper cell, type-2 T helper cell, central memory CD4 T cell, central memory CD8 T cell, effector memory CD8 T cell, activated dendritic cell, CD56bright natural killer cell, eosinophil, and mast cell. There were fewer immune cell components associated with S-III only two (including eosinophil and mast cell) were positively correlated, and three (including T cell exhaustion, type-2 T helper cell, activated CD4 T cell, natural killer T cell, memory B cell) were negatively correlated. In general, S-I was immune evasion, S-II was immune desert, S-III was immune activation.

In line with these results, S-I presented the highest immune score (Fig. 4C), stromal score (Fig. 4D), and ESTIMATE score (Fig. 4E), while S-II with the highest tumor purity (Fig. 4F). Surprisingly, the number of neoantigens was significantly highest in S-II (Fig. 4G), and there was no difference in T cell receptor diversity among the three subtypes (Fig. S3E). Furthermore, expression levels of most costimulator (Fig. 4H), coinhibitor (Fig. 4I), and antigen presentation (Fig. 4J) genes were the highest in S-I, the lowest in S-II, and fluctuated in S-III. Altogether, these analyses confirmed pharmacogenomics-based subtypes corresponded to distinct TME patterns.

3.7. Molecular mechanisms of pharmacogenomics-based subtypes

We explored the underlying molecular mechanisms of pharmacogenomics-based subtypes at gene mutation, copy number alteration, and DNA methylation levels, respectively. First, we assayed the mutation landscapes of three subtypes. Eleven of the top 20 mutated genes were significantly different among the three subtypes. Most of those genes were frequently mutated in S-II, less regularly in S-III (Fig. 5A-B). Interestingly, the mutation frequency of TP53 was almost the same between S-I and S-II, and the loss of TP53 function leads to proliferative features of cancers. In addition, S-II exhibited the highest TMB (Fig. 5C), accounting for the mechanism by which S-II had the highest number of tumor neoantigens. Second, copy number alteration was detected in three subtypes (Fig. 5D). Most of the chromosome fragments were gained or lost significantly in S-II. In S-I,

Fig. 2. Validation of pharmacogenomic-based subtypes in seven cohorts.(A-G) Kaplan-Meier curves demonstrating survival differences of pharmacogenomic-based subtypes in the GSE72094 (A), GSE68465 (B), GSE30219 (C), GSE42127 (D), GSE41271 (E), GSE31210 (F), GSE50081 (G) cohorts. (H-N) The multivariate Cox regression analysis in the GSE72094 (H), GSE68465 (I), GSE30219 (J), GSE42127 (K), GSE41271 (L), GSE31210 (M), GSE50081 (N) cohorts. (O-U) The links between pharmacogenomic-based subtypes and acceptable subtypes (TCGA2014) in the GSE72094 (O), GSE68465 (P), GSE30219 (Q), GSE42127 (R), GSE41271 (S), GSE31210 (T), GSE50081 (U) cohorts.



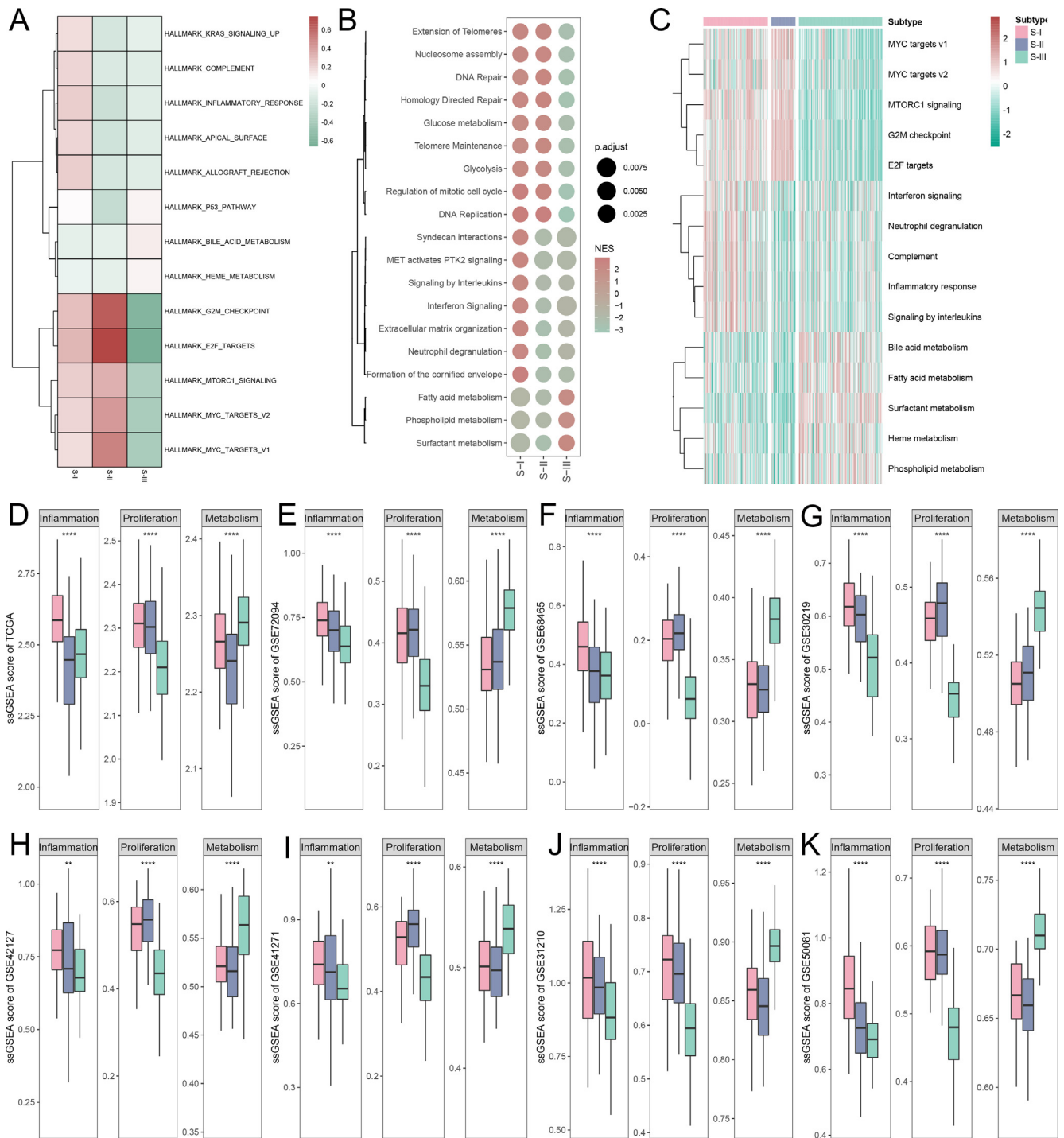


Fig. 3. Molecular characterization of pharmacogenomic-based subtypes. (A) Heatmap representing MSigDb hallmark gene set QuSAGE activity scores for each subtype compared with all others. The higher the score, the higher the pathway activity. (B) Dot plot depicting normalized enrichment score (NES) of Reactome gene set. (C) Heatmap depicting representative pathways ssGSEA scores for each patient from the three subtypes. (D–K) Boxplots representing difference of inflammatory, proliferative, and metabolic signatures among subtypes in TCGA-LUAD (D), GSE72094 (E), GSE68465 (F), GSE30219 (G), GSE42127 (H), GSE41271 (I), GSE31210 (J), GSE50081 (K) cohorts. P values are shown as *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

chromosome fragments 7p11.2, 7p11.2, and 7q31.2 were gained considerably, and chromosome fragments 9p21.3 and 9p23 were lost significantly. In S-III, chromosome fragments 16p11.2 were gained especially, while chromosome fragments 6q14.3 were lost significantly. Furthermore, the burden of copy number alterations (gain or loss) at arm and focal levels were highest in S-II (Fig. 5F–I). Finally, we identified DNA methylation-driven genes that dif-

fered significantly among the three subtypes (Fig. 5J–L). In addition, we further analyzed the correlation between the expression of these genes and inflammation, proliferation, and metabolic signatures (Fig. 5M). Most DNA methylation-driven genes showed significant positive correlation with metabolic signatures. In summary, all these results suggested that S-II was genetics-driven and S-III is methylation-driven.

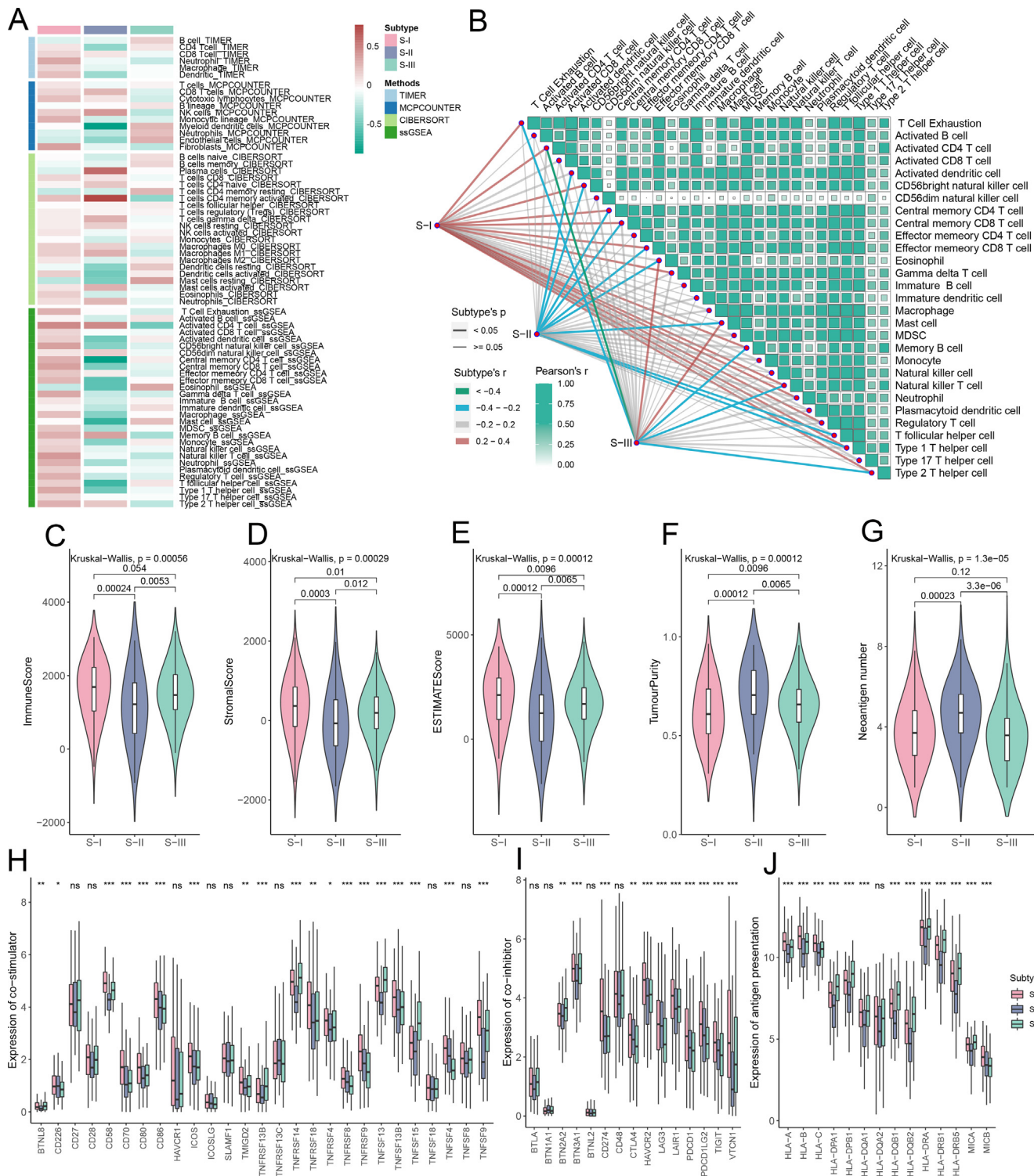


Fig. 4. Immune landscapes of pharmacogenomic-based subtypes. (A) Heatmap for immune responses based on TIMER, MCPcounter, CIBERSORT, and ssGSEA algorithms among three subtypes. The heatmap is colored by the mean score of each subtype. (B) Correlations between three subtypes and immune cellular components. (C–G) Violin plot of the immune score (C), stromal score (D), ESTIMATE score (E), tumor purity (F), and tumor neoantigen number (G) in three subtypes. (H–J) Boxplots representing different expression levels of immune co-stimulatory genes (H), immune co-inhibitory genes (I), and Antigen presentation genes (J) among three subtypes. P values are shown as *P < 0.05; **P < 0.01; ***P < 0.001.

3.8. Pharmacotherapy prediction for pharmacogenomics-based subtypes

First, we applied SubMap analysis to predict immunotherapy responses of pharmacogenomics-based subtypes. SubMap uncov-

ered that S-I was genetically similar to melanoma tumors responding to PD-1 (P = 0.03) and CTLA4 (P = 0.02) blockers, suggesting that ICBs had potential responses to S-I (Fig. 6A).

Second, we investigated the sensitivity of the above nine drugs in the pharmacogenomics-based subtypes based on the ssGSEA

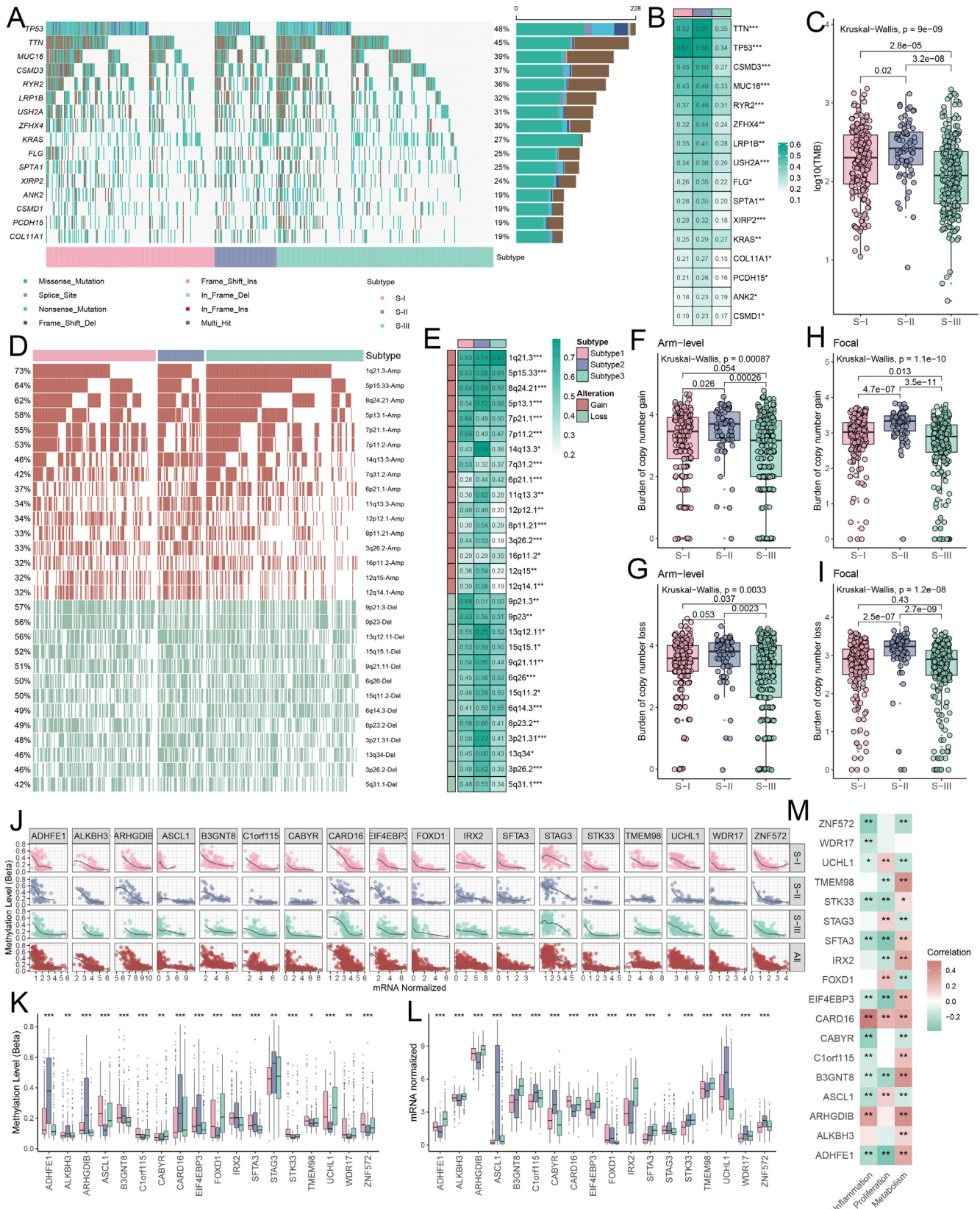


Fig. 5. Mutation, Copy number alteration, and DNA methylation landscapes of pharmacogenomic-based subtypes. (A) Waterfall plot of genes with significantly different mutations among three subtypes. (B) The mutation frequency of significantly different mutated genes among three subtypes. P values are shown as *P < 0.05; **P < 0.01; ***P < 0.001. (C) Boxplot representing difference analysis of TMB among three subtypes. (D) Waterfall plot of segments with significantly different alterations (gain and loss) among three subtypes. (E) The copy number alteration frequency of significantly different segments among three subtypes. P values are shown as *P < 0.05; **P < 0.01; ***P < 0.001. (F) The burden of copy number gain at arm (F) and focal levels (H), (G, I) The burden of copy number loss at arm (G) and focal levels (I). (J) Correlation analysis between DNA methylation and mRNA expression levels for methylation-driven genes. (K, L) Boxplot representing methylation (K) and mRNA expression (L) levels for methylation-driven genes. P values are shown as *P < 0.05; **P < 0.01; ***P < 0.001. (M) Heatmap represents the correlation between mRNA expression levels and scores of inflammatory, proliferative, and metabolic signatures scores. P values are shown as *P < 0.05; **P < 0.01.

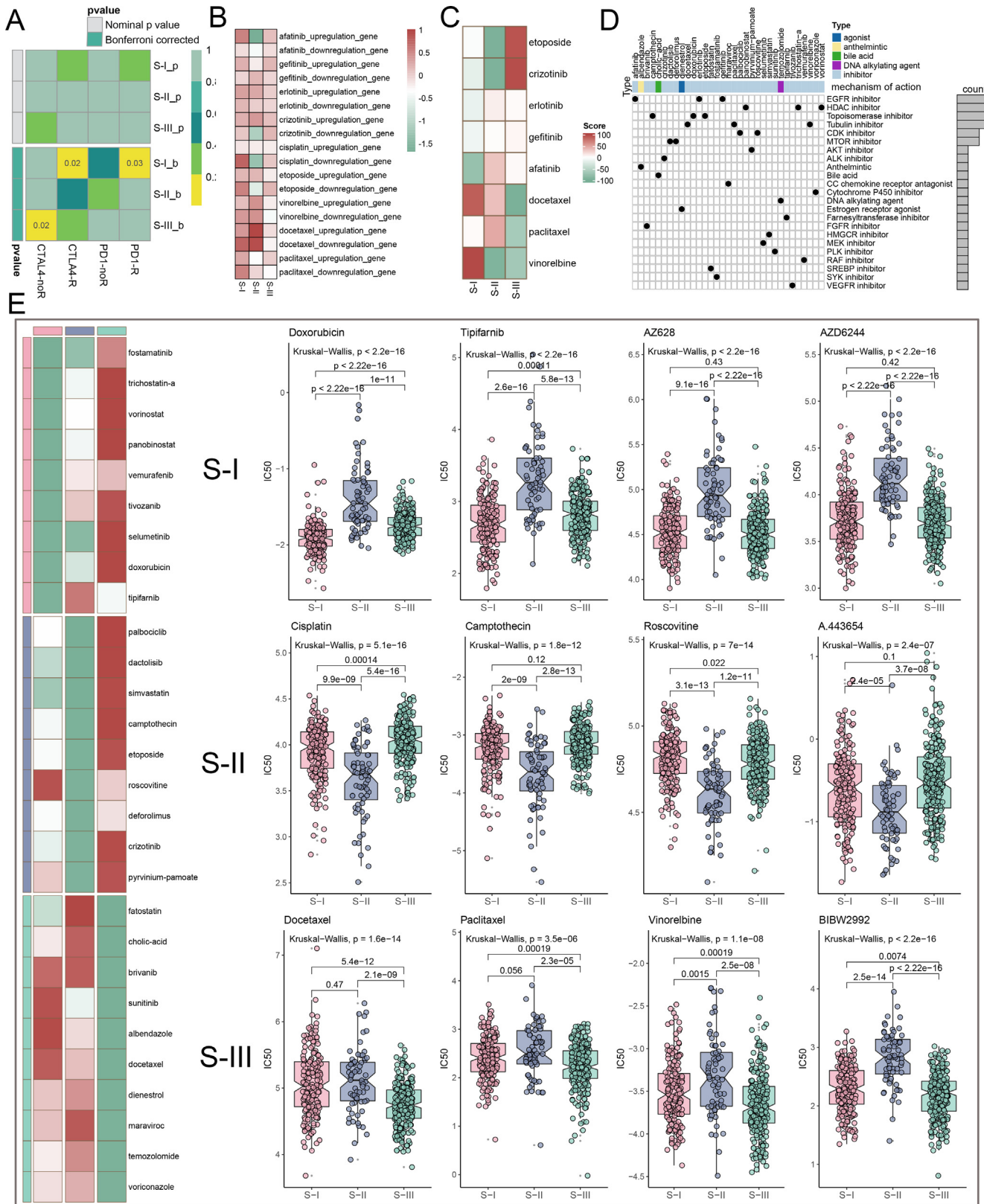


Fig. 6. Pharmacotherapy prediction for pharmacogenomic-based subtypes. (A) Submap analysis of the three subtypes and melanoma patients receiving anti-CTLA-4/PD-1 treatment with different immunotherapy responses. (B) Heatmap representing the ssGSEA median z-score value of up-regulated or down-regulated DRGs for each subtype. (C) Heatmap representing the connection score of CMap analysis for eight drugs. Drugs with Lower scores suggest better sensitivity for patients. (D) Heatmap showing each compound (perturbagen) from the CMap that shares mechanisms of action (rows) and sorted by descending number the compound with shared mechanisms of action. (E) Prediction of subtype-specific therapeutics by integrating CMAP database and pRRophetic package.

scores of up-regulated or down-regulated DRGs and the CMap database (Fig. 6B–C). Integrating the ssGSEA and CMap analysis results, we discovered that afatinib was more sensitive to S-I, etoposide, and crizotinib was more sensitive to S-II, and docetaxel, paclitaxel, and vinorelbine was more sensitive to S-III. Although ssGSEA analysis suggested that cisplatin was more sensitive in S-II, its sensitivity could not be verified in the CMap database due to the lack of cisplatin data.

Finally, we displayed potential agents targeting the molecular pathways and genes by utilizing the CMap database (Fig. 6D). In addition, we applied the pRRophetic package to examine the therapeutic sensitivity of the above potential drug or molecular targeting and screened subtype-specific drugs (Fig. 6E). Notably, pRRophetic package analysis suggested that cisplatin had the highest sensitivity in S-II, confirming the ssGSEA results. The three studies predicted that the sensitivity of the docetaxel, paclitaxel, and vinorelbine drugs in S-III were consistent.

Altogether, ICBs, doxorubicin (topoisomerase inhibitor), tipifarnib (farnesyltransferase inhibitor), AZ628 (or vemurafenib, RAF inhibitor), and AZD6244 (or selumetinib, MEK inhibitor) were potential treatments for S-I. Cisplatin, camptothecin (topoisomerase inhibitor), roscovitine (CDK inhibitor), and A.443654 (or pyrvinium-pamoate, AKT inhibitor) were potential treatments for S-II. Docetaxel (tubulin inhibitor), paclitaxel (tubulin inhibitor), vinorelbine (tubulin inhibitor), and BIBW2992 (TKI inhibitor) were potential treatments for S-III.

4. Discussion

In this study, we integrated pharmacogenomics data of LUAD cell lines and found three well-defined molecular subtypes in the TCGA-LUAD cohort: S-I (inflammatory and proliferative), S-II (proliferative), and S-III (metabolic). Our pharmacogenomics-based classification had significantly different survival implication and was independent prognostic factor. Besides, our pharmacogenomics-based subtypes were highly associated with the classical subtypes defined by previous reports. These pharmacogenomics-based subtypes were validated in seven independent GEO cohorts. The prognosis, association with previously reported subtypes, and subtype-specific molecular behaviors in all validation cohorts were highly consistent with the TCGA cohort. These results demonstrated the robustness and reproducibility of pharmacogenomics-based subtypes.

The three pharmacogenomics-based subtypes had different immune landscapes. S-I was an immuno-hot tumor with high immune infiltration but in the state of immune evasion. Although tumor tissue of S-I was heavily infiltrated with cytotoxic T cells (CD8⁺ T cells) and antigen-presenting cells (dendritic cells), the overall antitumor effect of the body's immune system was poor due to the low number of neoantigens, weak antigen-recognition ability (abundant T helper cells), and T cell exhaustion [23]. Unfortunately, the stimulation of inflammatory cells can promote the growth, invasion, and metastasis of cancer tissue [24]. This also explained the reason for S-I with the worst prognosis from the TME perspective. S-II was an immune-cold tumor with low immune infiltration and high tumor purity. Although tumor neoantigens were the most numerous, the body's immune system cannot kill cancer cells due to the lack of various immune cells in the TME. Of course, without stimulating inflammatory cells, S-II survived better than S-I. S-III was moderate immune infiltration. The body's immune system can play an anti-tumor effect on cancer cells, and S-III had the best prognosis. In short, subtype-specific TME patterns were vital factors affecting patient survival because of the essential influence of TME in tumor progression [25].

TMB was previously reported to be an ICBs biomarker [26]. Nevertheless, our outcomes suggested that S-II with high TMB did not respond to ICBs, indicating the limitations of prior TMB-based predictive biomarkers. We must examine all elements that contribute to the immunotherapy process in an integrative and global manner, rather than just one element. Besides, even tumors with low TMB have been shown to provide high-quality neoantigens that elicit antitumor T cell responses [27]. Consequently, only S-I responded well to ICBs due to enhanced tumor antigen recognition.

Herein, we found the molecular mechanisms of cancer cells were crucial factors driving S-II and S-v. S-II was genetics-driven, and S-III is methylation-driven. The mutation analysis identified TP53 as the primary mutant gene causing proliferation signature. TP53, as a tumor suppressor gene, is frequently mutated in various cancers, including LUAD. Wild-type TP53 suppressed cell cycle via enhancing multiple cell cycle checkpoints (including G2M checkpoint) arrest [28]. Thus, the mutant TP53 lost its normal function, leading to uncontrolled proliferation features of cancer cells. Copy number alteration analyses revealed that chromosomes 14q13.3 and 11q13.3 were amplified significantly in S-II. Amplification or overexpression of TTF1 (thyroid transcription factor 1) on chromosome 14q13.3 can enhance lung cancer cell proliferation via inducing the expression of the ROR1 (receptor tyrosine kinase-like orphan receptor 1) [29,30]. Overexpression of CCND1 (Cyclin D1) promoted NSCLC proliferation and progression through regulating the cell cycle [31]. Finally, we examined subtype-specific DNA methylation-driven genes. We found that except for WDR17, expression profiles of other methylation-driven genes were strongly correlated with metabolic signature. Upregulated ADHFE1 (alcohol dehydrogenase iron containing 1) promoted metabolic reprogramming of breast cancer [32]. SFTA3 (surfactant-associated protein 3), belonging to the family of lung surfactant proteins, is involved in surfactant metabolism [33].

This study aims to screen subtype-specific therapeutic agents and achieve precision medicine for cancer patients. The positive response rate of anti-PD-1 therapy for NSCLC patients was only 20% [34]. The response rate range to targeted treatment is 50 to 80% for patients with EGFR, ALK, ROS1, and BRAF mutations [35]. Although chemotherapy has been the traditional treatment for the management of NSCLC, individual patients respond differently to chemotherapy and have different survival rates. Therefore, it is urgent to develop a novel classification strategy that accurately identify the patients responding to drugs. In contrast to most previous reports studying precision medicine approaches from the cancer hallmark perspective [36], we integrated pharmacogenomics data of LUAD cell lines and proposed the pharmacogenomics-based system. We integrated CMap and pRRophetic analysis to identify subtype-specific sensitive drugs: ICBs, doxorubicin, tipifarnib, AZ628, and AZD6244 were for S-I; Cisplatin, camptothecin, roscovitine, and A.443654 were for S-II; Docetaxel, paclitaxel, vinorelbine, and BIBW2992 were for S-III.

This study has some limitations. DRGs were obtained from pharmacogenomics data of lung cancer cell lines. The influence of the tumor microenvironment on drug response cannot be considered. Nevertheless, there are distinct TME patterns among our pharmacogenomics-based subtypes. This study needs further conduct randomised controlled studies in homogenous cohorts based on three subtypes. Besides, in this study, intra-tumor molecular heterogeneity was not considered, which needs to be further explored from a single-cell level. In addition, the three subtypes may not exist independently. The relationship among the three subtypes needs further study, which contributes to revealing the mechanism of drug resistance.

5. Conclusion

We found three reproducible and pharmacogenomics-based subtypes with different prognoses. The three subtypes had distinct molecular characteristics: S-I was inflammatory, proliferative, and immune-evasion, S-II was proliferative and genetics-driven, and S-III was metabolic and methylation-driven. Finally, our study provided subtype-guided personalized treatment strategies by applying integrative pharmacogenomics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.064>.

References

- [1] Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers* 2015;1:15009.
- [2] Siegel RL, Miller KD, Fuchs HE, Jemal A. *Cancer Statistics, 2021*. *CA Cancer J Clin* 2021;71(1):7–33.
- [3] Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018;15(2):81–94.
- [4] Tlemsani C, Pécuchet N, Gruber A, Laurendeau I, Danel C, Riquet M, et al. NF1 mutations identify molecular and clinical subtypes of lung adenocarcinomas. *Cancer Med* 2019;8(9):4330–7.
- [5] Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–50.
- [6] Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* 2020;1(2):235–48.
- [7] Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* 2015;5(11):1210–23.
- [8] Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;166(3):740–54.
- [9] Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery cancer cells. *Nucleic Acids Res*. 2013;41(Database issue):D955–61.
- [10] Ma Y, Ding Z, Qian Y, Shi X, Castranova V, Harner EJ, et al. Predicting cancer drug response by proteomic profiling. *Clin Cancer Res* 2006;12(15):4583–9.
- [11] Ye Q, Putila J, Raese R, Dong C, Qian Y, Dowlati A, et al. Identification of prognostic and chemopredictive microRNAs for non-small-cell lung cancer by integrating SEER-medicare data. *Int J Mol Sci* 2021;22(14).
- [12] Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekas S, et al. RECIST 1.1-Update and clarification: From the RECIST committee. *Eur J Cancer* 2016;62:132–7.
- [13] Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald 3rd ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019;569(7757):503–8.
- [14] Deng ZM, Hu W, Dai FF, Yuan MQ, Hu M, Cheng YX. Immune-related genes to construct a novel prognostic model of breast cancer: a chemosensitivity-based study. *Front Immunol* 2021;12:734745.
- [15] Pascual-Montano A, Carazo JM, Kochi K, Lehmann D, Pascual-Marqui RD. Nonsmooth nonnegative matrix factorization (nsNMF). *IEEE Trans Pattern Anal Mach Intell* 2006;28(3):403–15.
- [16] Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 2004;101(12):4164–9.
- [17] Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschlegler S, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 2013;19(5):619–25.
- [18] Kapp AV, Tibshirani R. Are clusters found in one dataset present in another dataset? *Biostatistics* 2007;8(1):9–31.
- [19] Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS ONE* 2007;2(11):e1195.
- [20] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* 2018;48(4):812–30.
- [21] Cedoz P-L, Prunello M, Brennan K, Gevaert O. MethylMix 2.0: an R package for identifying DNA methylation genes. *Bioinformatics* 2018;34(17):3044–6.
- [22] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437–52.e17.
- [23] Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 2013;39(1):1–10.
- [24] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144(5):646–74.
- [25] Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* 2013;501(7467):346–54.
- [26] Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (New York, NY)*. 2015;348(6230):124–8.
- [27] Balachandran VP, Łuksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* 2017;551(7681):512–6.
- [28] Giono LE, Manfredi JJ. The p53 tumor suppressor participates in multiple cell cycle checkpoints. *J Cell Physiol* 2006;209(1):13–20.
- [29] Kwei KA, Kim YH, Girard L, Kao J, Pacyna-Gengelbach M, Salari K, et al. Genomic profiling identifies TTF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 2008;27(25):3635–40.
- [30] Yamaguchi T, Yanagisawa K, Sugiyama R, Hosono Y, Shimada Y, Arima C, et al. NKX2-1/TTF1/TTF-1-induced ROR1 is required to sustain EGFR survival signaling in lung adenocarcinoma. *Cancer Cell* 2012;21(3):348–61.
- [31] Zhang Z, Cui Z, Xie Z, Li C, Xu C, Guo X, et al. Deubiquitinase USP5 promotes non-small cell lung cancer cell proliferation by stabilizing cyclin D1. *Transl Lung Cancer Res* 2021;10(10):3995–4011.
- [32] Mishra P, Tang W, Putluri V, Dorsey TH, Jin F, Wang F, et al. ADHFE1 is a breast cancer oncogene and induces metabolic reprogramming. *J Clin Investig* 2018;128(1):323–40.
- [33] Schicht M, Rausch F, Finotto S, Mathews M, Mattil A, Schubert M, et al. SFTA3, a novel protein of the lung: three-dimensional structure, characterisation and immune activation. *Eur Respiratory J* 2014;44(2):447–56.
- [34] Garon EB, Rizvi NA, Hui R, Leighi N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *New England J Med* 2015;372(21):2018–28.
- [35] Chen R, Manochakian R, James L, Azzouqa AG, Shi H, Zhang Y, et al. Emerging therapeutic agents for advanced non-small cell lung cancer. *J Hematol Oncol* 2020;13(1):58.
- [36] Hijazo-Pechero S, Alay A, Marín R, Vilariño N, Muñoz-Pinedo C, Villanueva A, et al. Gene expression profiling as a potential tool for precision oncology in non-small cell lung cancer. *Cancers* 2021;13(19).