Drug Resistance in NSCLC is Associated with Tumor Micro-environment

Yizhong Ke
University of Shanghai for Science and Technology https://orcid.org/0000-0001-5937-2448
Pinzheng Huang
University of Shanghai for Science and Technology
Lixi Li
University of Shanghai for Science and Technology
Mingming Jin
Shanghai University of Medicine and Health Sciences
Gang Huang (huangg@sumhs.edu.cn)
Shanghai University of Medicine and Health Sciences

Research

Keywords: Single-cell RNA sequencing, tumor microenvironment, non-small cell lung cancer, resistance

Posted Date: September 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-882588/v1

Abstract

Introduction

Tumor cell resistance to chemotherapy is the most critical factor that influences the prognosis of cancer patients. It is generally believed that drug resistance is caused by genetic alterations in tumor cells; however, the relationship between drug resistance and the tumor microenvironment (TME) has not been adequately studied.

Methods

Herein, we successfully identified drug resistance and sensitivity clusters using single-cell transcriptome sequencing data from GSE149383 and established a proportional hazards model to find genes that affected prognosis.

Results

The results showed that marker genes between resistant and sensitive clusters were significantly associated with the TME; additionally, the model showed good reliability. Furthermore, we used bulk RNA-seq data to analyze the expression of CD24 and CYP1B1, which revealed little difference in the levels of the two genes in normal and tumor tissues but a significant difference in their expression between drug-resistant and -sensitive cells.

Conclusion

In conclusion, our study demonstrated a link between drug resistance and the TME, and we found that CD24 and CYP1B1 may be key regulators of drug resistance development in tumor cells via altering the TME.

1. Introduction

Lung cancer is the most common cancer diagnosis worldwide. According to recent projections, new lung cancer cases will account for 12.42% of all cancers diagnoses in 2021; lung cancer is also predicted to be responsible for 21.67% of all cancer-related deaths, making it the malignancy with the highest mortality rate. Approximately 80% of lung cancers are histologically defined as non-small cell lung cancer (NSCLC) [1, 2]. Previous genomics studies have identified several high-frequency genetic variants in NSCLC such as EGFR and KRAS mutations that regulate or participate in the occurrence and development of NSCLC, which has led to the development of a series of EGFR Tyrosine kinase inhibitors (TKIs) including gefitinib and erlotinib [3–5]. Although targeted therapies show significant efficacy and lower toxicity than chemotherapy, most NSCLC patients will develop resistance to targeted drugs, with a median time to disease progression of approximately 12 months [6–8].

Although there is now considerable evidence that the tumor microenvironment (TME) is closely related to tumor growth [9, 10], there have been few studies on the relationship between the TME and drug resistance. The TME provides the necessary conditions for tumor survival and includes tumor cells and other cell types, such as immune cells, mesenchymal cells, and fibroblasts [11]. Previously, it was believed that genetic alterations in tumor cells were solely responsible for drug resistance; however, there is now evidence that the TME plays a vital role in the development of resistance to anticancer treatments [12, 13].

To analyze the relationship between the TME and drug resistance, we obtained data for PC9 cells treated with erlotinib from the Gene Expression Omnibus (GEO) database and performed single-cell transcriptome sequencing (ScRNA-seq) analysis. ScRNA-seq analysis is an effective method of genomic analysis that analyzes the heterogeneity of tumor cells [14]. First, we used Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) algorithms to perform dimensionality reduction and clustering of ScRNA-seq data to determine possible drug-resistant and drug-sensitive clusters and to determine marker genes (MGs) between the two clusters. Searching each MG as a keyword in PubMed revealed that some MGs were involved in immunity and inflammation, in addition to the large number of MGs related to drug resistance. Second, we uniformly clustered each sample using the expression pattern of MGs and found significant differences in immune infiltration across categories. Finally, we constructed a proportional hazards regression model for the MGs, which identified CD24 and CYP1B1 as prognostic genes (PGs) for NSCLC.

We were surprised to find that levels of these two PGs differed little between tumor and nontumor tissue, but CD24 and CYP1B1 expression did differ significantly between the drug-resistant and -sensitive groups. Additionally, the drug sensitivity analysis among PGs showed that certain PGs might be associated with resistance to a large number of drugs. Combined with other reports, we hypothesize that both CD24 and CYP1B1 might contribute to drug resistance through pathways related to immunity and inflammation. The flow chart of this study is shown in Fig. 1.

2. Results

2.1 Two clusters of heterogeneous drug-resistant cells were distinguished using scRNA-seq data

The source of these scRNA-seq data was GSE149383. It has been previously reported that the PC9 cells used in these experiments had developed drug resistance. The number of PC9 cells dropped to the lowest level 4 d after the addition of erlotinib. After 11 d, the number of cells was restored to the level of day 1 (Fig. 2A). Therefore, we selected cells from day 0 (D0), D4, and D11 for analysis. In total, 4574 cells were included. After standardization and quality control (QC), 3848 competent cells remained (Fig. 2B). The results showed that after QC there was no apparent correlation between sequencing depth and the number of mitochondria, while the correlation coefficient between sequencing depth and the number of genes reached 0.94 (Fig. 2C). PCA is a standard data analysis method that is often used to reduce the

dimensionality of high-dimensional data and also can be used to extract the main feature components of data. The results of PCA showed that the data from D0, D4, and D11 did not form significant clusters (Fig. 2D). Thus, we next retrieved the top 15 PCs with their P-values, among which 13 PCs with P-values < 0.05 were selected for subsequent analysis (Fig. 2E).

Next, we clustered the cells using the UMAP algorithm and divided the three groups of cells into five clusters (Fig. 2F). The cells were also color-coded by the number of treatment days (Fig. 2G). We hypothesized that even the D11 cell population contained some drug-sensitive cells, and the clustering results proved our conjecture. We observed that D0, D4, and D11 cells overlapped in some areas, forming clusters 0, 1, and 3. This suggested that cells in these clusters did not (completely) develop resistance. Therefore, subsequent analysis focused on clusters 2 and 4, which contained 553 and 158 cells, respectively. We consider clusters 2 and 4 to represent drug-sensitive and -resistant cells, respectively.

2.2 Identification and analysis of MGs between clusters 2 and 4

There were 64 MGs between clusters 2 and 4, including 28 upregulated genes and 36 downregulated genes. When we selected the top four upregulated MGs and the top four downregulated MGs, we were surprised to find that these genes were almost exclusively expressed between the two clusters. For example, TOP2A was primarily expressed in cluster 2, while SERPINE1 was expressed in cluster 4 (Fig. 3A, B). This laterally verified our hypothesis that clusters 2 and 4 represent drug-sensitive and -resistant cells, respectively. Furthermore, this supposition was also supported by bubble plots, which showed that the MGs were not only differentially expressed between clusters 2 and 4, but were also minimally expressed in other clusters (Fig. 3C). Searching for these MGs in PubMed as keywords revealed that some genes were related to immunity and inflammation in addition to most of the genes being related to drug resistance (Fig. 3D). In addition, Gene Ontology enrichment results showed that the downregulated MGs (mainly expressed in cluster 4) were associated with immunity, such as leukocyte migration and negative regulation of cell migration, while the upregulated MGs (mainly expressed in cluster 2) were associated with cell cycle, such as organelle fission and chromosome segregation (Fig. 3E). Kyoto Encyclopedia of Genes and Genomes pathway enrichment results suggested that the downregulated MGs were enriched in the P53 signaling pathway, while the upregulated MGs were enriched in cell cycle-related pathways. Together, these findings suggest that entirely different pathways are involved in the regulation of the MGs between clusters 2 and 4 (Fig. 3F).

2.3 The MG-based classification of NSCLC patients was related to survival and clinical characteristics

Unsupervised consensus clustering of MG expression was performed on 176 NSCLC samples from GSE42127. The results of cumulative distribution function (CDF) curves showed that the optimal number of clusters was two ($P < 0.05$; Fig. 4A, B). Clusters 1 (C1) and 2 (C2) included 123 and 53 NSCLC samples, respectively. Combining their survival data, we found that the survival time of C2 was much lower than

that of C1 (Fig. 4C). To analyze the survival difference between C1 and C2, we performed correlation analysis using other clinical data in GSE42127, such as age, gender, and medication history. There were significant differences in medication history and gender between C1 and C2 ($P < 0.05$). There were more female patients (52.85%) and a greater proportion received chemotherapy (45.97%) in C1 compared with C2 (22.64% female patients and 33.96% received chemotherapy). We hypothesized that the reason for shorter OS in C2 may be the large percentage of patients who did not receive chemotherapy and the higher number of patients with a history of smoking (males) (Fig. 4D). Overall, these results suggested that classifications based on MGs can effectively differentiate the survival outcomes of patients and may be related to medication history.

2.4 Analysis of the C1 and C2 TME

Our previous results demonstrated that many MGs were related to immunity and inflammation. Additionally, the clustering results based on MGs showed that C1 and C2 may be related to exposure to chemotherapy. Therefore, we hypothesized that the TME of C1 and C2 may also be different due to chemotherapy.

In the TME, immune cells and stromal cells are the two most critical non-tumor cells, and their relative proportions directly determine tumor purity. Therefore, we calculated the scores for immune cells and stromal cells and added the two to get the “Estimatscore,” which could be used to determine tumor purity. The results showed that C1 had a significantly increased ratio of immune cells and stromal cells compared with C2 ($P < 0.001$). Thus, more immune and stromal cells and fewer tumor cells were present in C1 compared with C2 (Fig. 5A–D). Moreover, the number of immune infiltrating cells in the C1 and C2 groups confirmed the difference in the immune TME between the two groups. The numbers of resting memory CD4 + T cells, macrophage M0 cells, and resting mast cells were significantly different between the two groups ($P < 0.001$; Fig. 5E,F). In summary, there was a distinct difference TME between C1 and C2, which may be caused by chemotherapy.

2.5 The proportional hazards model based on MGs and development of a clinical prognostic nomogram

First, univariate Cox analysis was performed using expression data of MGs in GSE42127, and then least absolute contraction sum selection operator (LASSO) and multivariate Cox analysis were performed. These results were used to construct a proportional hazards model consisting of two risk genes: Risk score = (Fig. 6A). The model accurately classified the 176 samples in GSE42127 into high- and low-risk groups using the median value of the risk score as the cutoff value. The receiver operating characteristic (ROC) curve of the model showed that the 5-year area under curve (AUC) reached 0.705, indicating that the model had good predictive value (Fig. 6B). Additionally, Kaplan–Meier survival analysis showed that the overall survival (OS) rate of the high-risk group was significantly lower than that of the low-risk group (log-rank test, $P = 0.003427$; Fig. 6C). We visualized the relationship between risk scores and patient status, including risk score, survival time, survival status, and expression of PGs in the high- and low-risk

groups for each patient (Fig. 6D). The results showed that there were significantly more deaths in the high-risk group than in the low-risk group and that PGs were upregulated in the high-risk group. Finally, the clinical data in GSE42127 was used to construct a prognostic nomogram that could predict OS for each patient. Age, gender, stage, histological characteristics, and medication history were included in the prediction model (Fig. 6E). The calibration chart shows that the predicted and actual risks for 1, 3, and 5 years were almost identical according to the model (Fig. 6F). This suggests that this model has considerable reliability for providing individualized predictions for NSCLC patients.

Next, GSEA was performed to identify the regulatory pathways associated with the identified MGs. The results showed that MGs of the high-risk group mainly regulated primary immunodeficiency, T-cell receptor signaling pathway, P53 signaling pathway, and B-cell receptor signaling pathway. In contrast, MGs in the low-risk group regulated glycine, serine, and threonine metabolism, histidine metabolism, drug metabolism, cytochrome P450, and cytochrome P450 metabolism of xenobiotics. Overall, the high-risk group was primarily enriched for immune-related pathways, while the low-risk group was enriched for metabolism-related pathways (Fig. 6G).

2.6 Differential expression of CD24 and CYP1B1 between drug-resistant and -sensitive cells

GSE116959 included 57 NSCLC and 11 normal samples. We found no significant difference in the expression of PGs between tumor and normal tissue samples, suggesting that PGs are not proto-oncogenes and do not promote cancer cell proliferation (Fig. 7A). Therefore, we hypothesized that PGs increased patient risk by generating drug resistance or altering the TME. To test this speculation, we analyzed differences in PG expression between the drug-resistant and -sensitive groups.

GSE80344 included 12 erlotinib-resistant NSCLC cell lines and four sensitive NSCLC cell lines. The results revealed that PG expression differed significantly between the resistant and sensitive groups ($P < 0.05$; Fig. 7B). Additionally, we calculated the drug sensitivity of PGs with CellMiner using expression profile and drug data. Two drugs with minimal P-values were retained for each gene, and upregulation of PGs was found to produce resistance to three of them: methotrexate, tamoxifen, and pipamperone ($Cor < 0$; Fig. 7C,D).

3. Discussion

Many previous studies have shown that drug resistance is achieved through genetic alterations in tumor cells [15–17]. However, the development of drug resistance is a complex process, which we believe includes changes in the TME that further promote the proliferation of tumor cells, creating a vicious feed-forward cycle. Extensive tumor heterogeneity is also considered a factor in the generation of drug resistance [18, 19]. Previous studies that used macro-level sequencing of individual genes could not fully explain tumor heterogeneity. However, scRNA-seq can greatly facilitate cancer research by sequencing each cell at the individual level [20].

Intratumor heterogeneity refers to the different characteristics of individual tumor cells within the same overall lesion [21]. Our results showed that PC9 cells, even under the same culture conditions, generate different clusters. Among these five identified clusters, clusters 2 and 4 were considered drug-sensitive and -resistant clusters, respectively, because the other three clusters were concentrated in the overlapping region of D0, D4, and D11 cells. These clusters may represent cells in some intermediate stage of the transition from drug-sensitive to -resistant. MGs were defined as differentially expressed genes between clusters 2 and 4, but we unexpectedly found that MGs were minimally expressed in the other clusters. This confirmed our speculation that clusters 2 and 4 represented drug-sensitive and -resistant clusters, respectively.

Chien et al. [22] genetically compared cisplatin-resistant cells with cisplatin-sensitive cells and found that collagen expression was higher in the resistant group than in the sensitive group. This suggests that chemotherapy can alter TME by directly altering tumor DNA. The results of our PubMed search using MGs as keywords showed that in addition to many MGs being related to drug resistance, a considerable number of MGs were associated with keywords such as immunity and inflammation. Bulk RNA-seq data of GSE42127 was then analyzed to explore the expression pattern of MGs. Unsupervised consensus clustering based on the expression of MGs divided 176 NSCLC samples into the groups C1 and C2. The ESTIMATE algorithm is a standard means of calculating tumor purity, and many studies have demonstrated its reliability [23–25]. We used the ESTIMATE algorithm to calculate the immune cell score, stroma cell score, and tumor purity for both groups. C1 samples had lower tumor purity compared with C2. Combined with the analysis of clinical characteristics, this patient stratification could be because there were more male patients (smokers) in C2 and, more importantly, only 22.64% of patients in C2 received chemotherapy compared with 45.97% in C1. This resulted in a higher predicted survival rate in C1 than C2. Additionally, there were more resting memory CD4 + T cells and resting mast cells in the TME of C1, while there were more macrophage M0 cells in C2. Resting Memory CD4 + T cells can promote the expression of CTLA-4 and PD-1, which are costimulatory molecules that play a critical negative regulatory role in T cell activation; thus, maintaining balanced T cell activation. A series of monoclonal antibodies targeting CTLA-4 and PD-1 are available in the clinic [26–28].

Additionally, Liu et al. [29] found that resting memory CD4 + T cells and resting mast cells were negatively correlated with prognosis in all cancers; thus, our findings suggest that chemotherapy altered the TME in the C1 group such that continued chemotherapy may no longer be effective. Tumor-associated macrophages are the most abundant immune cells in the TME [30]. While macrophage M1 cells are involved in inflammatory and antitumor responses, macrophage M2 cells are an immunomodulatory cell type formed by peripheral monocytes in the TME that not only prevent T lymphocytes from attacking tumor cells but also secrete cytokines to nourish tumor cells and promote tumor angiogenesis, leading to tumor cell dissemination and metastasis [31, 32]. Thus, macrophages in most advanced tumors have a pro-cancer effect, as evidenced by the M0 macrophage content in the untreated C2 population. In conclusion, the classification results constructed from drug resistance-associated MGs showed a

potential link between drug resistance and TME that promoted tumor cell proliferation and drug resistance via increasing the numbers of resting memory CD4 + T cells and resting mast cells.

The proportional hazards model uses a mathematical model to fit the relationship between the survival distribution and the influencing factors by evaluating the degree of influence of the influencing factors on the distribution of the survival function; such models are widely used in clinical analyses [33, 34]. Our study showed that PGs affected the prognosis of NSCLC, so we established a risk score formula based on two risk genes. The AUC showed good prediction accuracy of the model at 1, 3, and 5 years. Additionally, we constructed a clinical nomogram to predict the prognosis of NSCLC using risk scores and clinical parameters. This scoring system will help doctors predict OS based on a patient's histological and clinical parameters.

We next analyzed the expression of PGs in different cohorts. Unexpectedly, we found little difference in the expression of PGs between NSCLC patients and the normal population, but there was a significant difference in PG expression between the erlotinib-resistant and erlotinib-sensitive groups. This suggested that PGs may not be proto-oncogenes but alter genes that control the TME by regulating drug resistance pathways, which nonetheless affects the prognosis of NSCLC patients. GSEA results also confirmed this, as the identified genes were concentrated in immune-related pathways in the high-risk group and metabolism-related pathways in the low-risk group. Therefore, we speculate that with continued drug use, MGs alter the TME by regulating metabolic pathways and subsequently promote tumor growth. CD24 is known to interact with sialic-acid-binding Ig-like lectin 10 (Siglec-10) on the surface of innate immune cells, and their interaction promotes immune escape [35, 36]. However, no studies have correlated CD24 with drug resistance. CYP1B1 has been reported to be associated with drug resistance. Mu et al. [37] found that miR-27b enhanced drug sensitivity by activating p53-dependent apoptosis and reducing CYP1B1-mediated drug detoxification.

In conclusion, this study demonstrated a potential link between drug resistance in tumor cells and the TME. Furthermore, we identified several PGs associated with drug resistance. Our drug sensitivity analysis of PGs suggested that PG expression may make cancer cells resistant to a wide range of drugs. However, this study had some limitations as we used a published dataset but did not have our own validation cohort. Therefore, further validation of these results is needed.

4. Conclusion

In conclusion, we used scRNA-seq data to demonstrate that drug resistance is closely linked to the TME and that drug-resistant cells may achieve continued proliferation by altering metabolism along with altering the TME. Additionally, PGs identified based on risk proportional regression models may be key regulators of drug resistance. Our study provides a potential approach for future tumor therapy by cutting off the pathways associated with drug resistance and making the TME unsuitable for tumor cell survival.

5. Methods

5.1 Data acquisition and pre-processing

ScRNA-seq data from GSE149383 [38], which included sequencing data from 4574 cells, were included in the screening, and we performed QC on the cells using the Seurat package in R4.0.1 to screen out low-quality cells with the following QC criteria: 1) > 3 cells with genes were detected; 2) cells showed < 5% of expressed genes from mitochondrial DNA; 3) cells with > 50 detected genes [39]. After QC, 3848 cells remained. PCA was then used to analyze gene expression among eligible cells, and 13 PCs with P-values < 0.05 were retained for subsequent downscaling. The UMAP algorithm was used to cluster the cells, and the LIMMA package was used to calculate MGs between selected clusters with cutoff criteria of P-value < 0.05; and |log2[fold change (FC)]| >2.

Bulk RNA-seq data were obtained from GSE42127, which included 176 NSCLC cases with sequencing data [40, 41]. The data used to visualize the risk associated with gene expression were obtained from GSE80344 and GSE116959. GSE80344 included eight groups of drug-resistant cells and four groups of sensitive cells [42]; GSE116959 included 57 NSCLC samples and 11 normal samples [43].

5.2 Patient classification based on MGs

Unsupervised consensus clustering of the NSCLC patients in GSE42127 was performed using the ConsensusClusterPlus package, and the best clustering results were determined by a combination of the consensus heat map and the area under the CDF curve. Survival rates for different classifications were then evaluated using the Survival and Survminer packages. Finally, we compared differences in the clinical characteristics between C1 and C2, with P-values < 0.05 considered statistically significant.

5.3 Immune cell infiltration

Immune cell and stromal cell scores were calculated for C1 and C2 using the ESTIMATE algorithm, with expression data normalized for all samples. The relative number of immune cell counts for each sample was output by the CIBERSORT algorithm, and these data were used to calculate immune infiltration in C1 and C2.

5.4 Generating the prognostic risk score model based on MGs

For the GSE42127 cohort, univariate Cox–LASSO–multivariate Cox regression analysis was first used to screen for prognostic genes, with P-values < 0.05 considered statistically significant. Then a risk score model based on critical prognosis-related MGs was developed to predict the prognosis of NSCLC. Referring to previous reports, we used the formula: risk score = Exp gene1 × β1 + Exp gene2 × β2+.....+Exp gene(n)×β(n), where "Exp" represents the expression level of the corresponding gene, and "β" represents the regression coefficient calculated by multivariate Cox analysis [44]. In total, 176 groups of NSCLC data were divided into high-risk and low-risk groups, followed by Kaplan–Meier survival analysis to estimate OS in the two groups; a two-sided log-rank test was used to assess survival differences. Finally, time-

dependent ROC curve analysis was used to assess the predictive accuracy of the prognostic model based on MGs.

Subsequently, the identified independent prognostic parameters including gender, age, medication history, and histological characteristics, were used to develop prognostic column line plots. The performance of the nomogram was then quantitatively assessed using calibration plots [45].

5.5 Gene set enrichment analysis (GSEA) and drug sensitivity analysis

GSEA (http://software.broadinstitute.org/gsea/index.jsp) was used to identify pathways regulated by the MGs [46]. Adjusted P-values were used to correct for false-positive results using the Benjamini–Hochberg false discovery rate (FDR) procedure, with FDR values < 0.05 considered statistically significant [47].

The Cellminer database (https://discover.nci.nih.gov/cellminer/home.do) includes approximately 100,000 compounds and natural products, and was used to screen the drug affinity of different genes. The sensitivity of genes to drugs was evaluated using linear regression analysis, with P-values < 0.05 considered statistically significant.

Abbreviations

TME
Tumor microenvironment;
NSCLC
Non-small cell lung cancer;
TKIs
Tyrosine kinase inhibitors;
GEO
Gene Expression Omnibus;
ScRNA-seq
Single-cell transcriptome sequencing;
PCA
Principal Component Analysis;
UMAP
Uniform Manifold Approximation and Projection;
MGs
Marker genes;
PGs
Prognostic genes;
CDF
Cumulative distribution function;

AUC
Area under curve;
QC
Quality control;
LASSO
Least absolute contraction sum selection operator;
OS
Overall survival;
ROC
Receiver operating characteristic;
GSEA
Gene set enrichment analysis;
FDR
False discovery rate;

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The datasets analysed during the current study are available in the GEO repository, (https://www.ncbi.nlm.nih.gov/geo/).

Competing interests: The authors declare that they have no competing interests.

Funding: This research was funded by National Natural Science Foundation of China (NO.81830052) and the Shanghaihai Key Laboratory of Molecular Imaging.

Authors' contributions: Yizhong Ke and Gang Huang designed the study; Yizhong Ke searched the data; Yizhong Ke, Pinzheng Huang, and Lixi Li performed the data analysis; Yizhong Ke completed the first draft, and Mingming Jin proofread the first draft. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

[1] Agustoni F, Suda K, Yu H, Ren SX, Rivard CJ, Ellison K, et al. EGFR-directed monoclonal antibodies in combination with chemotherapy for treatment of non-small-cell lung cancer: an updated review of clinical trials and new perspectives in biomarkers analysis. Cancer Treatment Reviews. 2019;72:15-27.

[2] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. Ca-a Cancer Journal for Clinicians. 2021;71(1):7-33.

[3] Mok TS, Wu YL, F.A.C.S, Thongprasert S, Yang CH, Chu DT, et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. The New England Journal of Medicine.2009;361:947-57.

[4] Shepherd FA, Pereira JR, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in Previously Treated Non–Small-Cell Lung Cancer. The New England Journal of Medicine. 2005;353:123-32.

[5] Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncology. 2012;13(3):239-46.

[6] Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. The New England Journal of Medicine. 2009;361(10):958-67.

[7] Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nature Reviews Cancer. 2013;13(10):714-26.

[8] Obenauf AC, Zou YL, Ji AL, Vanharanta S, Shu WP, Shi HB, et al. Therapy-induced tumour secretomes promote resistance and tumour progression. Nature. 2015;520(7547):368-72.

[9] Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng WG, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nature Medicine. 2015;21(8):938-45.

[10] Zhan HX, Zhou B, Cheng YG, Xu JW, Wang L, Zhang GY, et al. Crosstalk between stromal cells and cancer cells in pancreatic cancer: New insights into stromal biology. Cancer Letters. 2017;392:83-93.

[11] Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell. 2010;141(1):52-67.

[12] Eckstein N, Servan K, Hildebrandt B, Pölitz A, von Jonquières G, Wolf-Kümmeth S, et al. Hyperactivation of the insulin-like growth factor receptor I signaling pathway is an essential event for cisplatin resistance of ovarian cancer cells. Cancer Res. 2009;69(7):2996-3003.

[13] Meads MB, Hazlehurst LA, Dalton WS. The Bone Marrow Microenvironment as a Tumor Sanctuary and Contributor to Drug Resistance. Clinical Cancer Research. 2008;14(9):2519-26.

[14] Andrews TS, Hemberg M. Identifying cell populations with scRNASeq. Molecular Aspects of Medicine. 2018;59:114-22.

[15] Niederst MJ, Engelman JA. Bypass Mechanisms of Resistance to Receptor Tyrosine Kinase Inhibition in Lung Cancer. Science Signaling. 2013;6(294).

[16] Rotow J, Bivona TG. Understanding and targeting resistance mechanisms in NSCLC. Nature Reviews Cancer. 2017;17(11):637-58.

[17] Harrison PT, Vyse S, Huang PH. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. Seminars in Cancer Biology. 2020;61:167-79.

[18] Lupia M, Cavallaro U. Ovarian cancer stem cells: still an elusive entity? Mol Cancer. 2017;16(1):64.

[19] Steg AD, Bevis KS, Katre AA, Ziebarth A, Dobbin ZC, Alvarez RD, et al. Stem Cell Pathways Contribute to Clinical Chemoresistance in Ovarian Cancer. Clinical Cancer Research. 2012;18(3):869-81.

[20] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377-82.

[21] Ling B, Huang ZL, Huang SY, Qian L, Li GL, Tang QL. Microenvironment Analysis of Prognosis and Molecular Signature of Immune-Related Genes in Lung Adenocarcinoma. Oncology Research. 2021;28(6):561-78.

[22] Chien J, Kuang R, Landen C, Shridhar V. Platinum-sensitive recurrence in ovarian cancer: the role of tumor microenvironment. Front Oncol. 2013;3:251.

[23] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013;4:2612.

[24] Jia D, Li SL, Li DL, Xue HP, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging-Us. 2018;10(4):592-605.

[25] Cao R, Yuan L, Ma B, Wang G, Tian Y. Tumour microenvironment (TME) characterization identified prognosis and immunotherapy response in muscle-invasive bladder cancer (MIBC). Cancer Immunol Immunother. 2021;70(1):1-18.

[26] Cutolo M, Sulli A, Paolino S, Pizzorni C. CTLA-4 blockade in the treatment of rheumatoid arthritis: an update. Expert Review of Clinical Immunology. 2016;12(4):417-25.

[27] Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature. 2017;541(7637):321-30.

[28] Das S, Johnson DB. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. Journal for Immunotherapy of Cancer. 2019;7(1):306.

[29] Liu R, Yang F, Yin JY, Liu YZ, Zhang W, Zhou HH. Influence of Tumor Immune Infiltration on Immune Checkpoint Inhibitor Therapeutic Efficacy: A Computational Retrospective Study. Frontiers in Immunology. 2021;12.

[30] Qi YY, Chang Y, Wang ZW, Chen LL, Kong YY, Zhang PP, et al. Tumor-associated macrophages expressing galectin-9 identify immunoevasive subtype muscle-invasive bladder cancer with poor

prognosis but favorable adjuvant chemotherapeutic response. Cancer Immunology Immunotherapy. 2019;68(12):2067-80.

[31] Kumar AT, Knops A, Swendseid B, Martinez-Outschoom U, Harshyne L, Philp N, et al. Prognostic Significance of Tumor-Associated Macrophage Content in Head and Neck Squamous Cell Carcinoma: A Meta-Analysis. Front Oncol. 2019;23;9:656.

[32] Bercovici N, MV Guérin, Trautmann A, Donnadieu E. The remarkable plasticity of macrophages: a chance to fight cancer. Frontiers in Immunology.2019;10:1563.

[33] Fibrinogen Studies Collaboration. Measures to assess the prognostic ability of the stratified Cox proportional hazards model. Stat Med. 2009;28(3):389-411.

[34] Zhang HH, Lu W. Adaptive Lasso for Cox's proportional hazards model. Biometrika.2007;94(3):691-703.

[35] Barkal AA, Brewer RE, Markovic M, Kowarsky M, Barkal SA, Zaro BW, et al. CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature. 2019;572(7769):392-6.

[36] Chen GY, Tang J, Zheng P, Liu Y. CD24 and Siglec-10 selectively repress tissue damage-induced immune responses. Science. 2009;323(5922):1722-5.

[37] Mu WJ, Hu CB, Zhang HB, Qu ZQ, Cen J, Qiu ZX, et al. miR-27b synergizes with anticancer drugs via p53 activation and CYP1B1 suppression. Cell Research. 2015;25(4):477-95.

[38] Aissa AF, Islam A, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nature Communications. 2021;12(1).

[39] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology. 2018;36(5):411-20.

[40] Hight SK, Mootz A, Kollipara RK, McMillan E, Yenerall P, Otaki Y, et al. An in vivo functional genomics screen of nuclear receptors and their co-regulators identifies FOXA1 as an essential gene in lung tumorigenesis. Neoplasia. 2020;22(8):294-310.

[41] Tang H, Xiao GH, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-Gene Set Predicts Survival Benefits from Adjuvant Chemotherapy in Non-Small Cell Lung Cancer Patients. Clinical Cancer Research. 2013;19(6):1577-86.

[42] Leon LM, Gautier M, Allan R, Ilie M, Nottet N, Pons N, et al. The nuclear hypoxia-regulated NLUCAT1 long non-coding RNA contributes to an aggressive phenotype in lung adenocarcinoma through regulation of oxidative stress. Oncogene. 2019;38(46):7146-65.

[43] Ruden DM, Chen L, Possidente D, Possidente B, Rasouli P, Wang L, et al. Genetical toxicogenomics in Drosophila identifies master-modulatory loci that are regulated by developmental exposure to lead. Neurotoxicology. 2009;30(6):898-914.

[44] Wang ZH, Guo XP, Gao L, Wang Y, Ma WB, Xing B. Glioblastoma cell differentiation trajectory predicts the immunotherapy response and overall survival of patients. Aging-Us. 2020;12(18):18297-321.

[45] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.

[46] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545-50.

[47] Thissen D, Steinberg L, Kuang D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. Journal of Educational & Behavioral Statistics.2002;27(1):77-83.

Figures

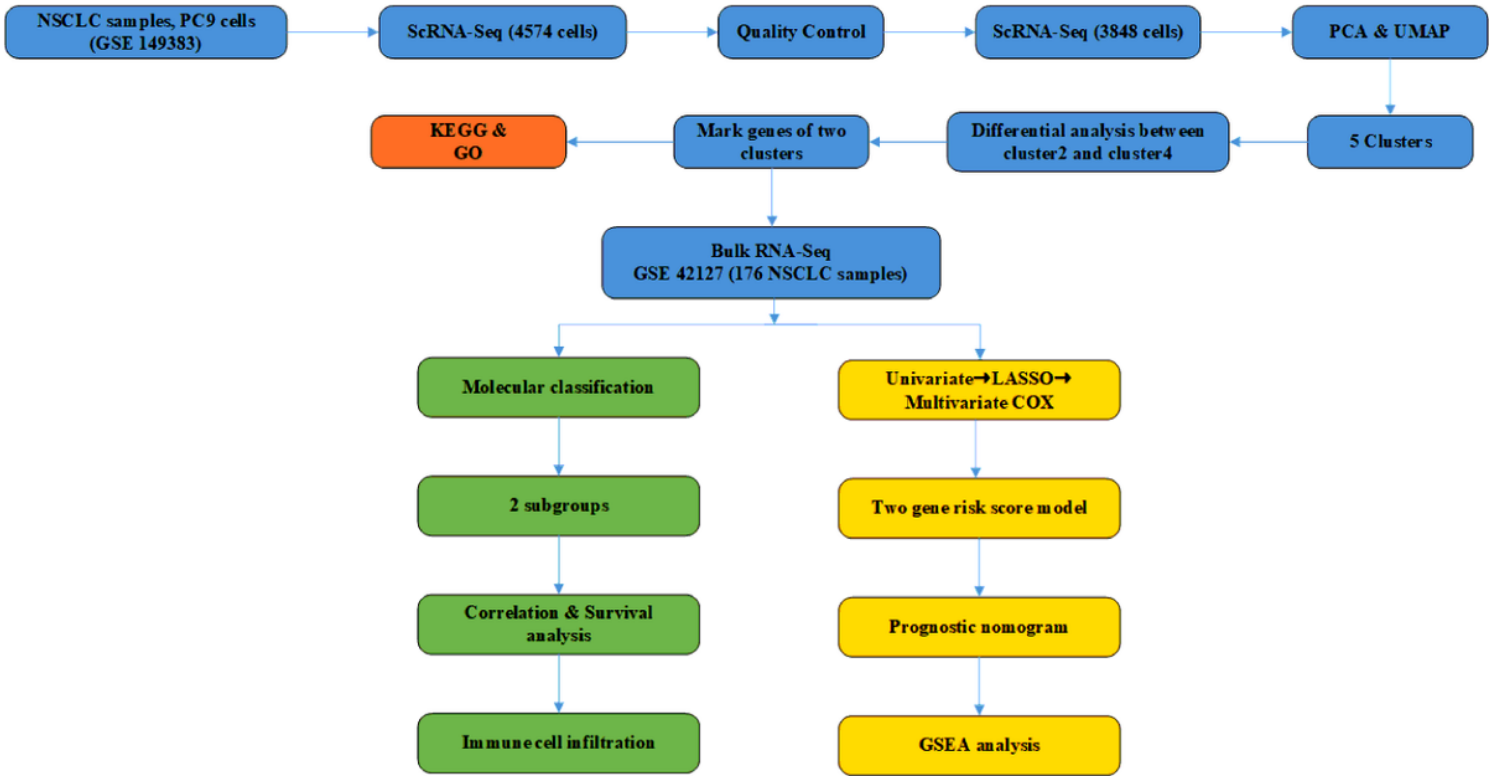

Figure 1

Flow chart of the study design.

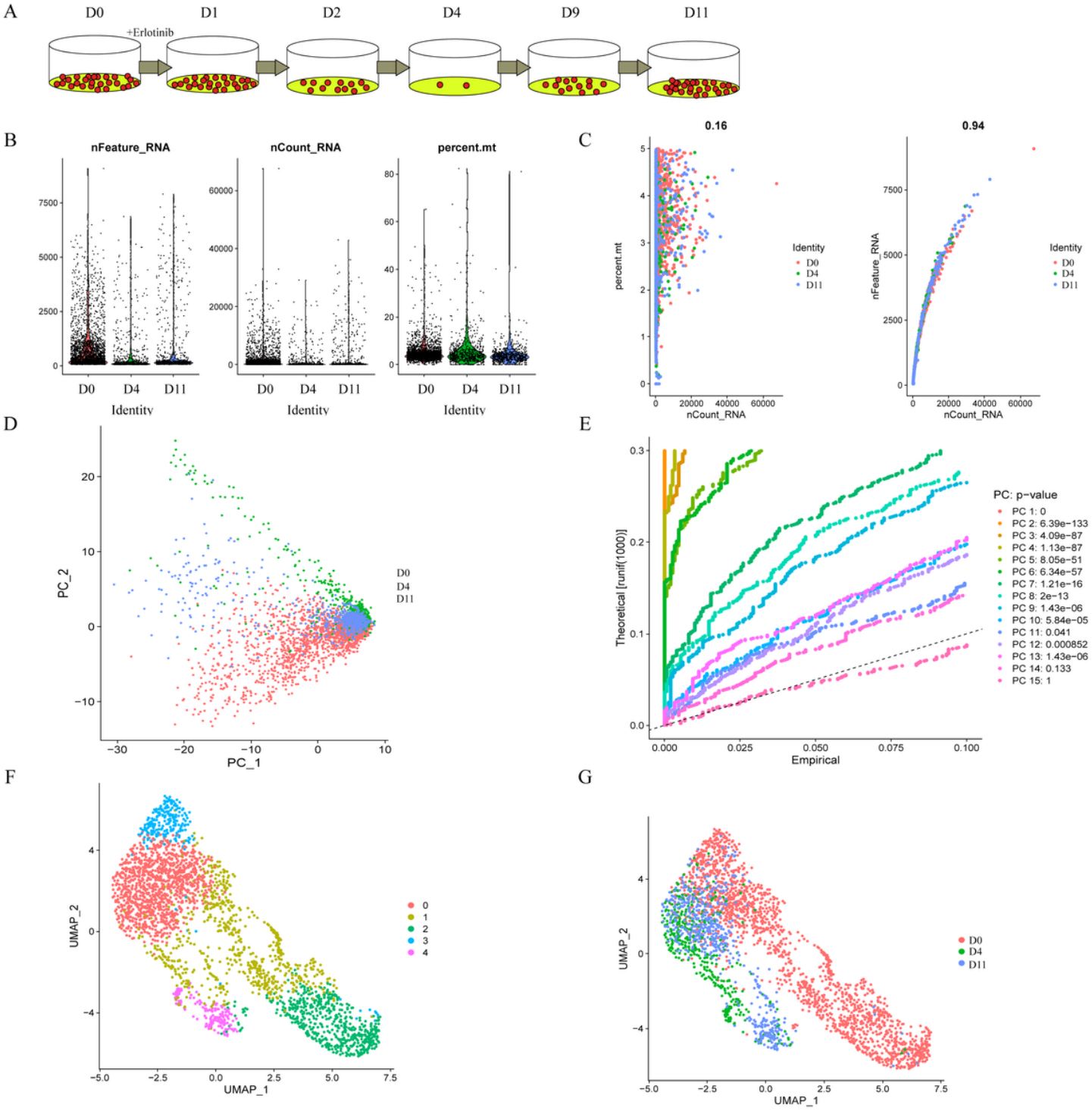

Figure 2

The identification of five clusters based on single cell (sc)RNA-seq data revealed abundant heterogeneity in NSCLC. A: Schematic diagram of a contiguous scRNA-seq sample set. D0 is untreated cells. D1–D11 represent the duration of drug treatment. Cells were titrated on day 0 (D0), D1, D2, D4, D9, and D11 of treatment. B: In total, 4574 cells from D0, D4, and D11 were selected for quality control, of which 3848 cells met the criteria. C: Pearson's correlation coefficients of sequencing depth with mitochondrial number and gene number. D: Principal component analysis (PCA) showed no significant separation of cells in the

three groups. E: PCA identified 15 PCs and their P-values. F: The Uniform Manifold Approximation and Projection algorithm was applied to 13 PCs and successfully divided the cells into 5 clusters. G: Each cell is color-coded according to the number of days of erlotinib treatment.

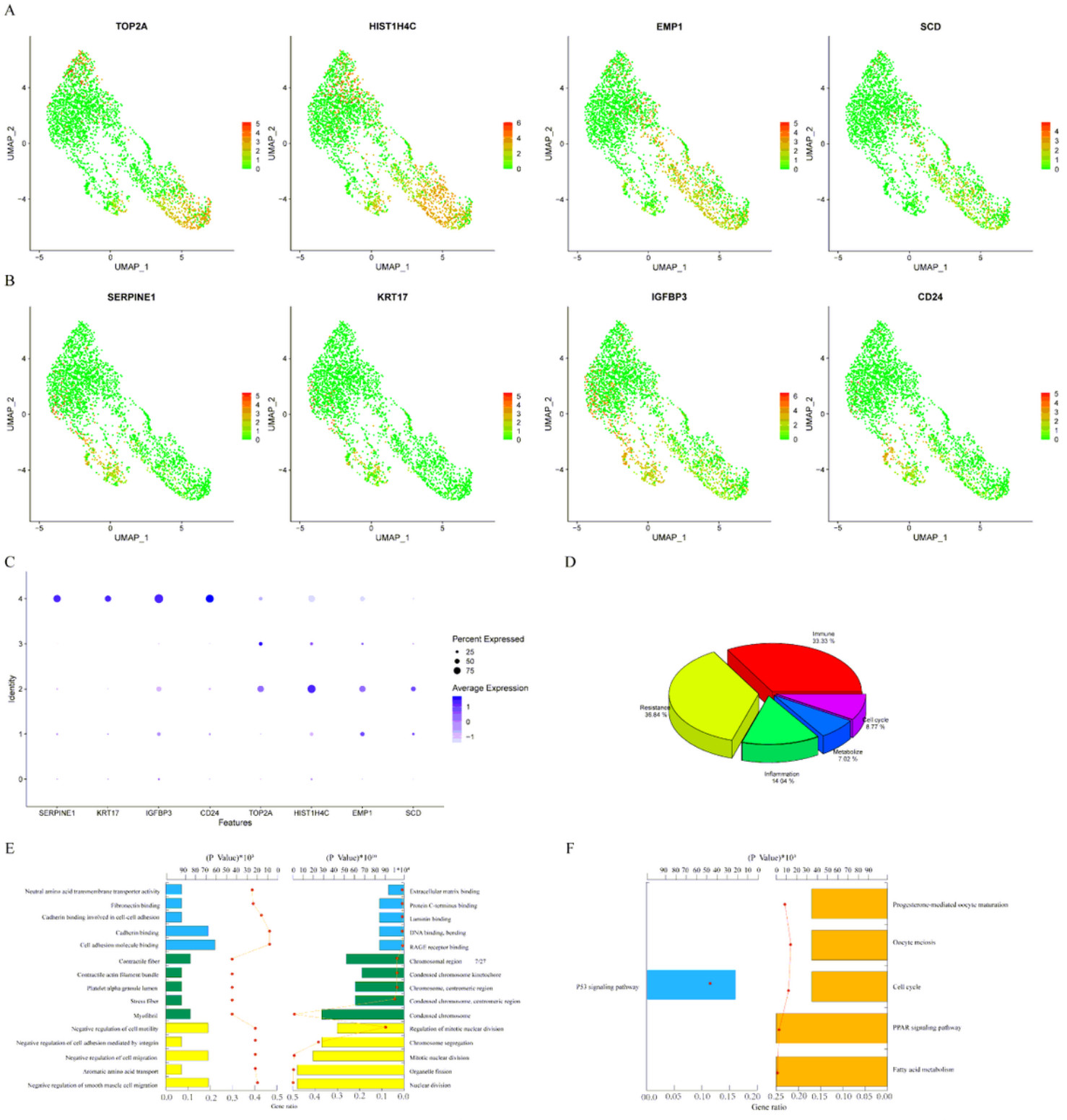

Figure 3

Expression profiles of marker genes (MGs) and related pathways. A: Four MGs that were downregulated in the sample. B: Four MGs that were upregulated in the sample. C: Bubble plots of the expression levels

of eight MGs in five clusters. D: PubMed results from searching MGs as the keywords; 36.84% of MGs were associated with drug resistance, 33.33% with immunity, and 14.04% with inflammation. E: Gene Ontology enrichment results of MGs; down- and up-regulated MGs are shown on the left and right, respectively; yellow, green, and blue represent biological process, cellular component, and molecular function, respectively. F: Kyoto Encyclopedia of Genes and Genomes pathway enrichment results of MGs; down- and up-regulated MGs are shown on the left and right, respectively.

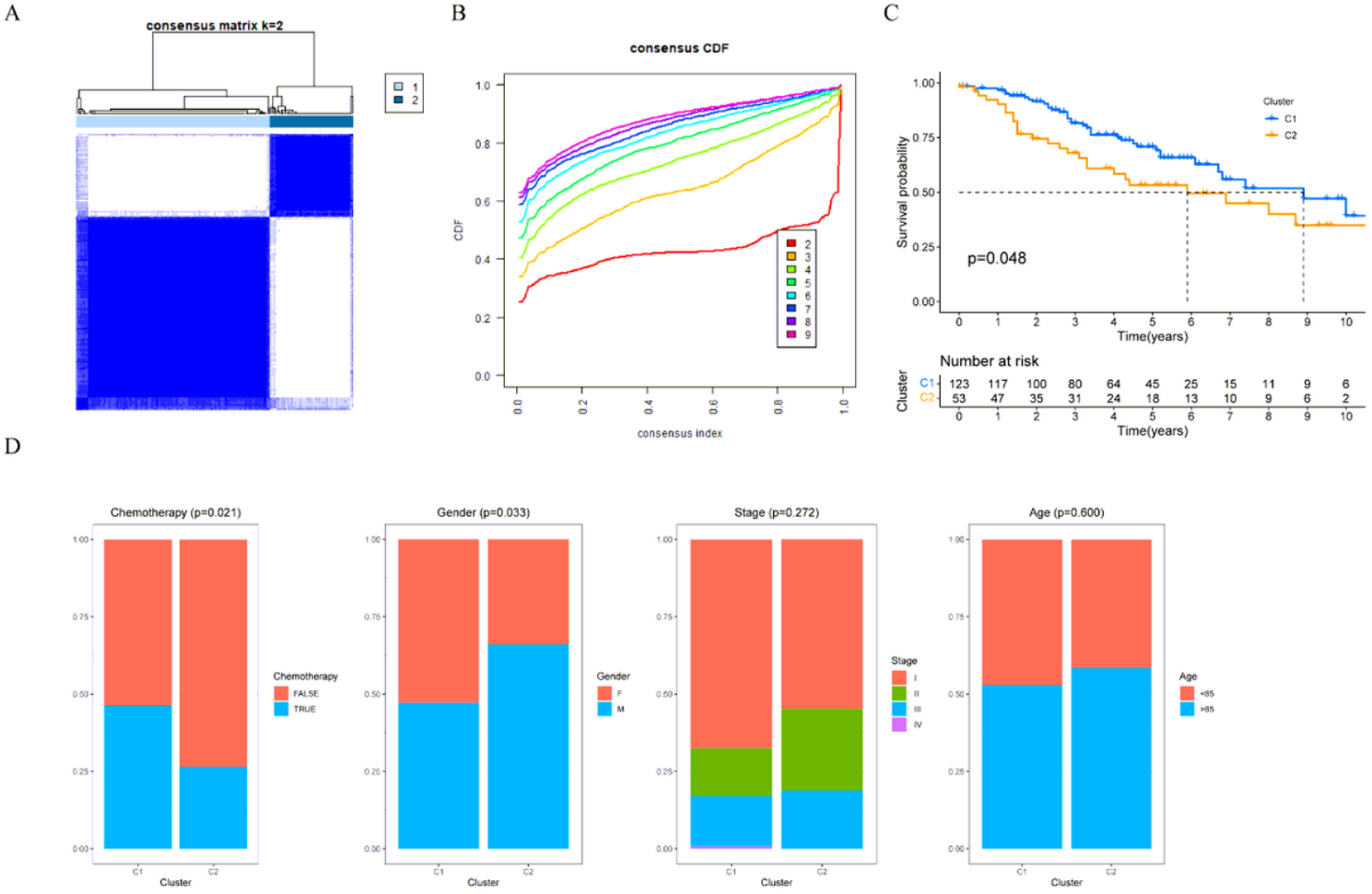

Figure 4

Classification and clinical correlation analysis of NSCLC patients based on MGs. A: Consensus clustering matrix at K = 2. B: Cumulative distribution function curves of consensus clustering (K = 2–9). C: Overall survival (OS) of NSCLC patients was decreased in C2 compared with C1 (P = 0.048). D: Correlation between subgroups and chemotherapy history, gender, stage, and age, which showed that subgroups were correlated with chemotherapy history (P = 0.021) and gender (P = 0.033).

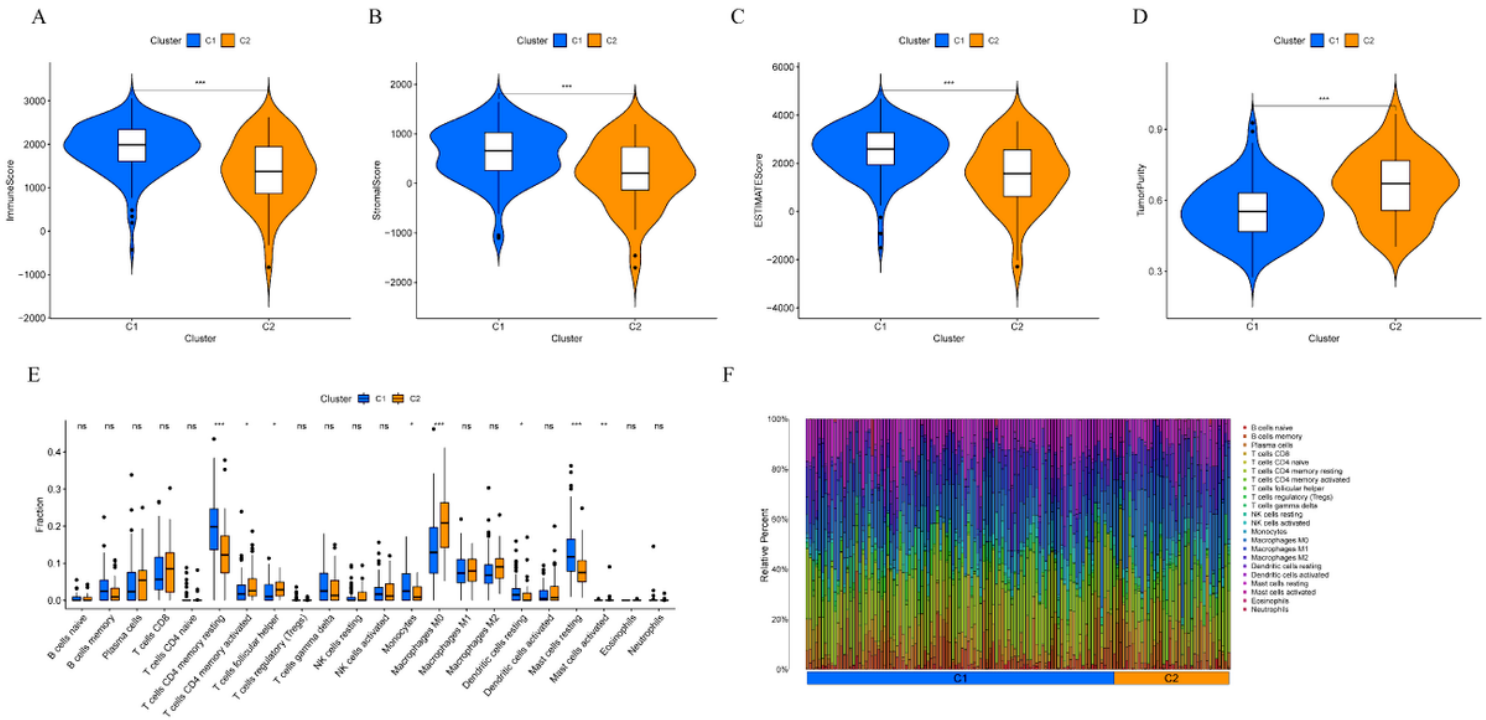

Figure 5

ESTIMATE scores and immune infiltration in C1 and C2. A–D: Immune cell scores, stromal cell scores, ESTIMATE scores (immune cell score + stromal cell score), and tumor purity of C1 and C2. C1 had higher ESTIMATE scores and lower tumor purity. E: The numbers of various immune cells in the TME of C1 and C2. F: Heat map of the percentages of various immune cells in the TME of C1 and C2. *P < 0.05, **P < 0.01, ***P < 0.001.

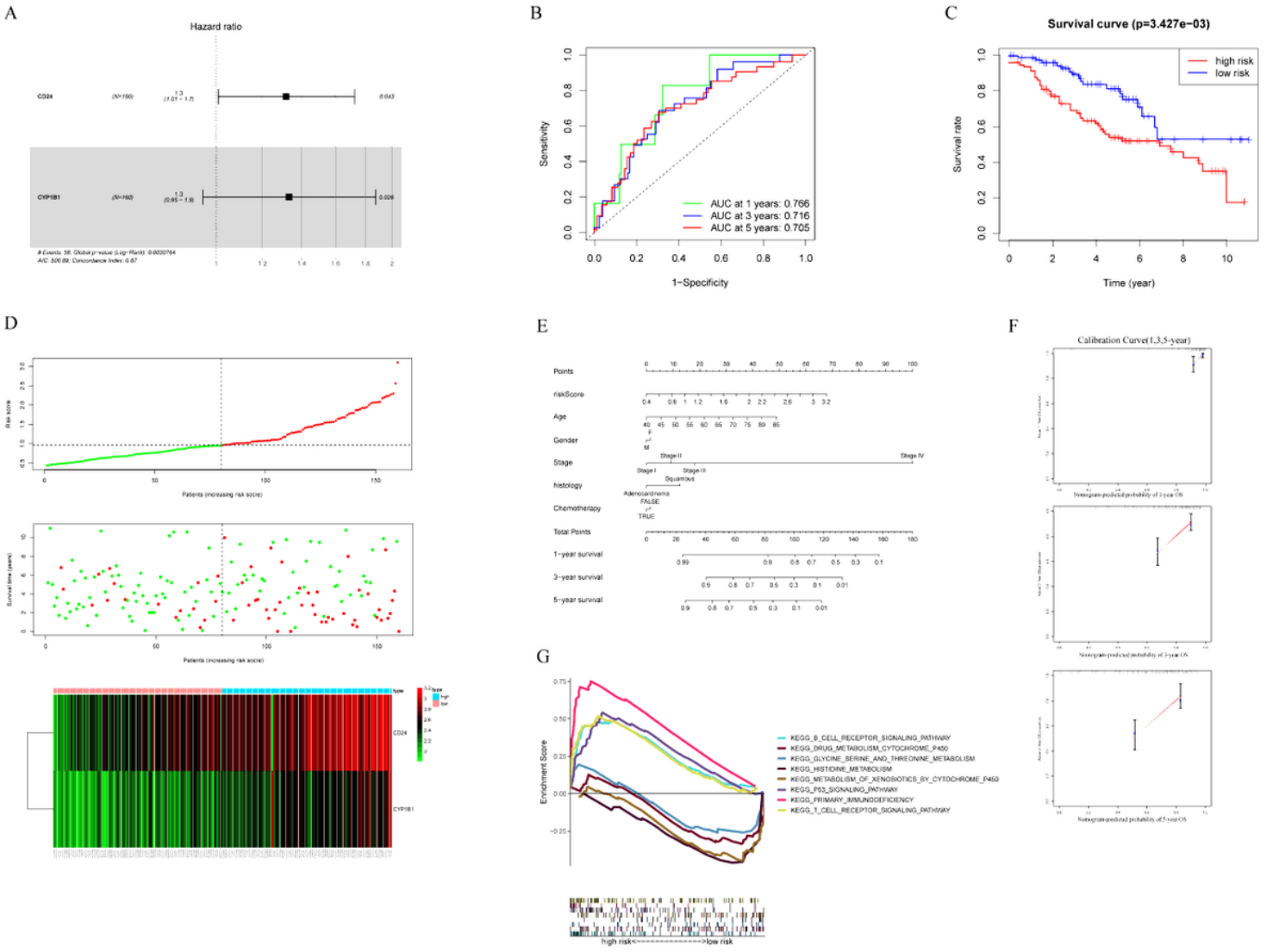

Figure 6

Construction of proportional hazards models and predictive prognostic nomograms based on MGs. A: Forest plot of risk for prognostic genes (PGs). B: Receiver operating characteristic curves showing area under the curve values at 1, 3, and 5 years. C: Kaplan–Meier survival analysis used to assess OS in high- and low-risk patients. OS was significantly worse in the high-risk group than in the low-risk group. D: Risk score analysis was used to calculate the PG signature and to classify patients into the high- or low-risk groups using the median risk score as the cutoff value. Upper panel: risk score curves for the PG signature; middle panel: patient survival status and time by risk score distribution; lower panel: heat map of the expression levels of two PGs in NSCLC samples. From green to red indicates gene expression levels from low to high. E: Nomogram for predicting the clinical prognosis of NSCLC patients, which included age, gender, stage, histological features, chemotherapy history, and PGs. F: Calibration plots of the predictive nomograms for 1-, 3-, and 5-year OS. G: Gene set enrichment analysis for the high- and low-risk groups.

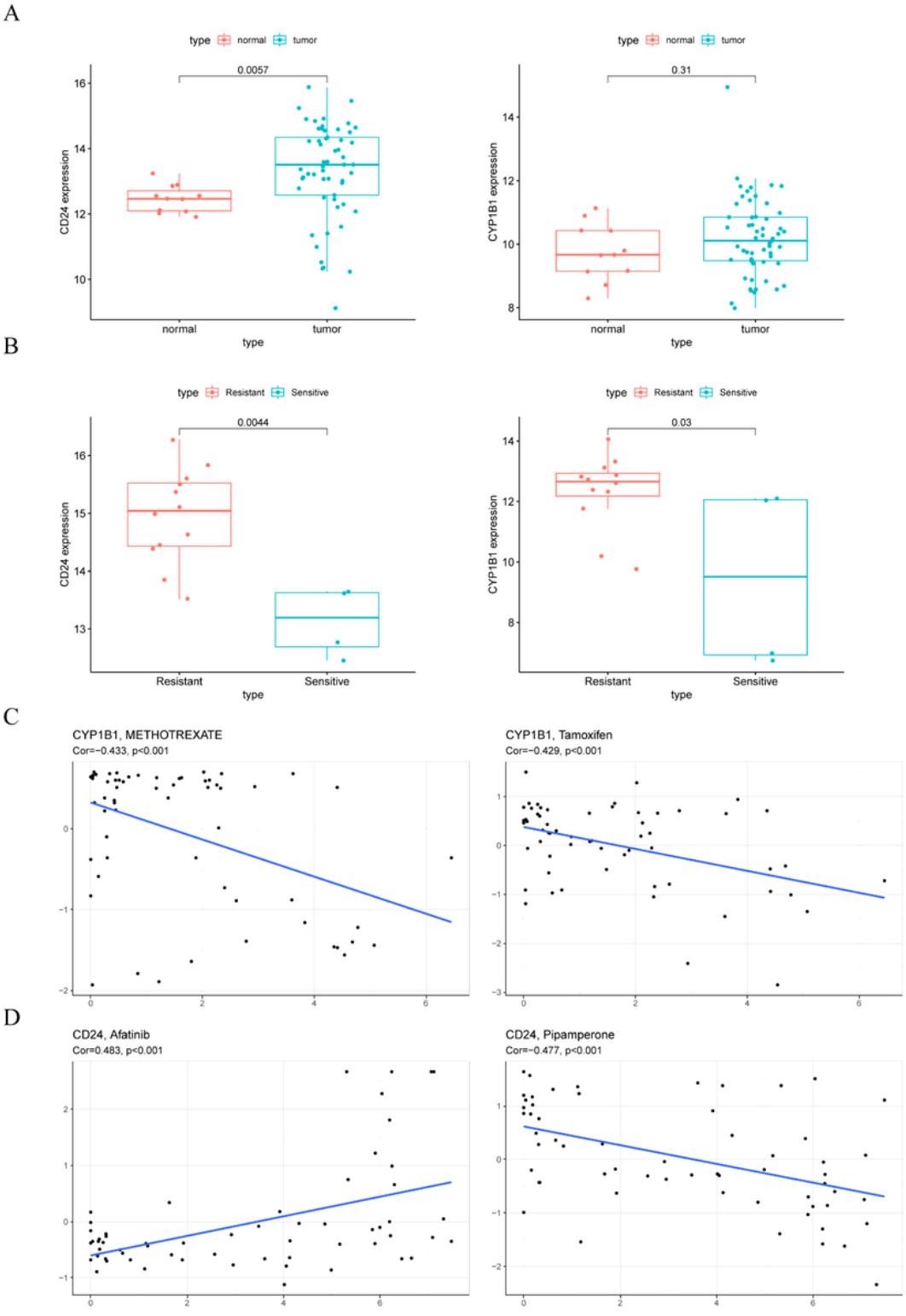

Figure 7

Expression of PGs in different samples and drug sensitivity analysis. A: There was almost no difference in the expression of PGs between the tumor and normal groups. B: Significant differences in the expression of PGs were found in the drug-resistant and -sensitive groups. C: The degree of tolerance for the four drugs most associated with PGs. PG expression caused cells to become resistant to three of the drugs.