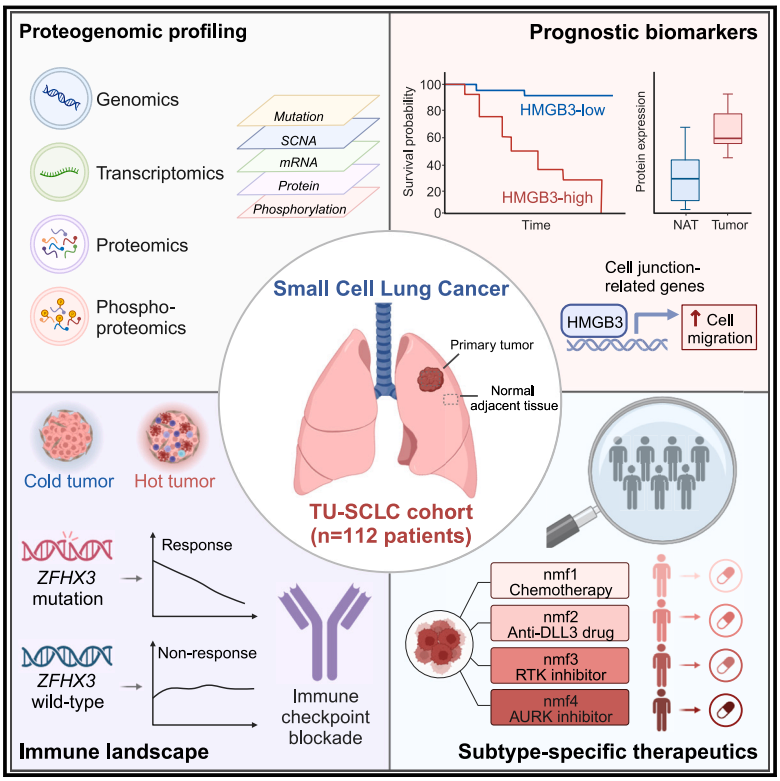
Cell

Resource

Proteogenomic characterization of small cell lung cancer identifies biological insights and subtype-specific therapeutic strategies

Graphical abstract



Highlights

- Integrated multi-omics reveals functional impacts of genomic alteration
- HMGB3 promotes cell migration and is a prognostic indicator
- *ZFHX3* mutation is associated with higher immune response
- Four multi-omics subtypes inform precision patient treatment

Authors

Qian Liu, Jing Zhang, Chenchen Guo, ..., Hongbin Ji, Hu Zhou, Peng Zhang

Correspondence

dgao@sibcb.ac.cn (D.G.), hbji@sibcb.ac.cn (H.J.), zhouhu@simm.ac.cn (H.Z.), zhangpeng1121@tongji.edu.cn (P.Z.)

In brief

Comprehensive proteogenomic characterization of small cell lung cancer using paired tumor and normal adjacent tissues provides insights into tumorigenesis, potential prognostic biomarkers, and immune landscape; identifies therapeutic strategies for four subtypes; and creates a resource for further research.

Liu et al., 2024, Cell *187*, 184–203
January 4, 2024
https://doi.org/10.1016/j.cell.2023.12.004

CellPress

CellPress

Cell

Resource

Proteogenomic characterization of small cell lung cancer identifies biological insights and subtype-specific therapeutic strategies

Qian Liu,[1,2,18] Jing Zhang,[1,18] Chenchen Guo,[3,18] Mengcheng Wang,[3,4,18] Chenfei Wang,[5,6,18] Yilv Yan,[1] Liangdong Sun,[1] Di Wang,[1] Lele Zhang,[7] Huansha Yu,[1] Likun Hou,[8] Chunyan Wu,[8] Yuming Zhu,[1] Gening Jiang,[1] Hongwen Zhu,[2] Yanting Zhou,[2] Shanhua Fang,[2] Tengfei Zhang,[3,4] Liang Hu,[3] Junqiang Li,[9] Yansheng Liu,[10] Hui Zhang,[11] Bing Zhang,[12] Li Ding,[13] Ana I. Robles,[14] Henry Rodriguez,[14] Daming Gao,[3,4,15,*] Hongbin Ji,[3,4,15,17,*] Hu Zhou,[2,4,16,*] and Peng Zhang[1,19,*]

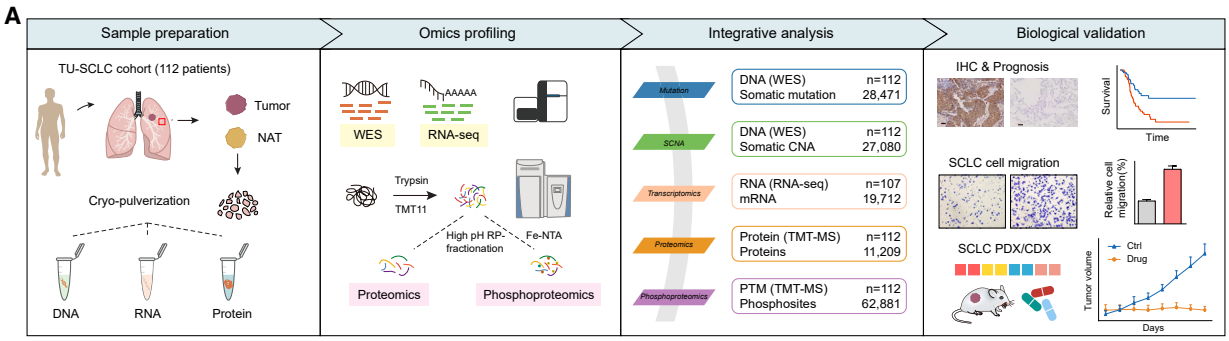
[1]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai 200433, China
[2]Department of Analytical Chemistry, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China
[3]State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, CAS Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China
[5]Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Department of Orthopedics, Tongji Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[6]Frontier Science Center for Stem Cells, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[7]Central Laboratory, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai 200433, China
[8]Department of Pathology, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai 200433, China
[9]D1 Medical Technology, Shanghai 201800, China
[10]Cancer Biology Institute, Yale University School of Medicine, West Haven, CT 06516, USA
[11]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[12]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA
[13]Department of Medicine, McDonnell Genome Institute, Washington University, St. Louis, MO 63108, USA
[14]Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Rockville, MD 20850, USA
[15]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[16]School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[17]School of Life Science and Technology, Shanghai Tech University, Shanghai 200120, China
[18]These authors contributed equally
[19]Lead contact
*Correspondence: dgao@sibcb.ac.cn (D.G.), hbji@sibcb.ac.cn (H.J.), zhouhu@simm.ac.cn (H.Z.), zhangpeng1121@tongji.edu.cn (P.Z.)
https://doi.org/10.1016/j.cell.2023.12.004

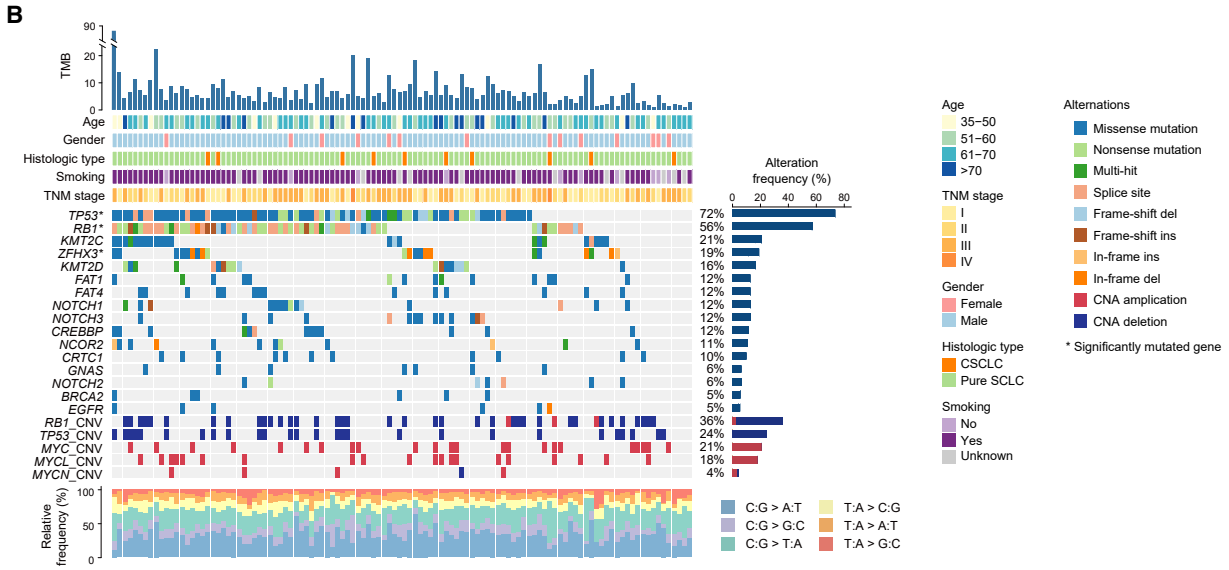
SUMMARY

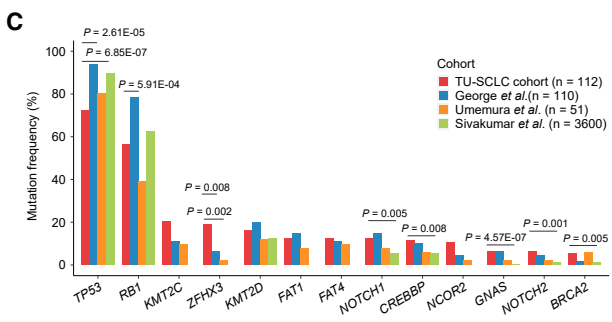
We performed comprehensive proteogenomic characterization of small cell lung cancer (SCLC) using paired tumors and adjacent lung tissues from 112 treatment-naive patients who underwent surgical resection. Integrated multi-omics analysis illustrated cancer biology downstream of genetic aberrations and highlighted oncogenic roles of *FAT1* mutation, *RB1* deletion, and chromosome 5q loss. Two prognostic biomarkers, HMGB3 and CASP10, were identified. Overexpression of HMGB3 promoted SCLC cell migration via transcriptional regulation of cell junction-related genes. Immune landscape characterization revealed an association between *ZFHX3* mutation and high immune infiltration and underscored a potential immunosuppressive role of elevated DNA damage response activity via inhibition of the cGAS-STING pathway. Multi-omics clustering identified four subtypes with subtype-specific therapeutic vulnerabilities. Cell line and patient-derived xenograft-based drug tests validated the specific therapeutic responses predicted by multi-omics subtyping. This study provides a valuable resource as well as insights to better understand SCLC biology and improve clinical practice.

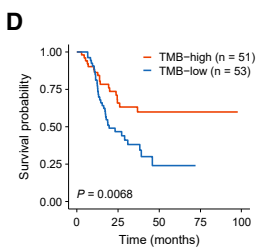
INTRODUCTION

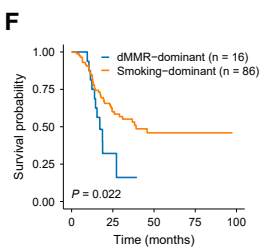
Lung cancer is the leading cause of cancer mortality worldwide.[1] Small cell lung cancer (SCLC) accounts for 15% of all lung cancers and is the most malignant and deadliest subtype.[2] The overall 5-year survival rate of SCLC is only 5%, making it the sixth leading cause of cancer-related death.[3] SCLC is an aggressive neuroendocrine (NE) carcinoma characterized by rapid

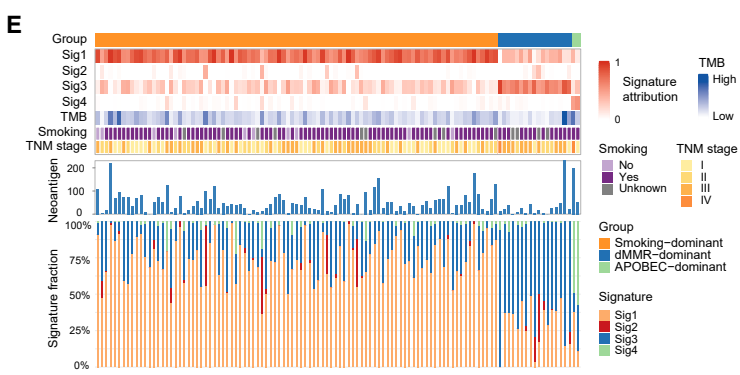


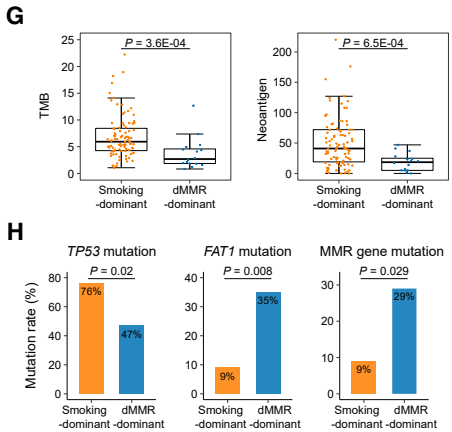












(legend on next page)

proliferation, strong metastatic proclivity, and striking therapeutic resistance, which contribute to its extremely poor prognosis.[4] SCLC is in stark contrast with non-small cell lung cancer (NSCLC) in clinicopathology, biology, and treatment options.[5–8] Although targeted therapies and immunotherapies have tremendously improved the outcomes of NSCLC patients for molecularly defined subtypes, progress in the treatment of SCLC has been slow. Patient survival has not significantly improved over the last few decades, and SCLC still remains outside the realm of precision medicine.

Previous genomic studies revealed that inactivation of *TP53* and *RB1* occurs frequently in SCLC,[9–11] but mutational profiling did not seem to establish molecular subtypes or identify actionable therapeutic targets, likely because some gene alterations identified at the DNA level do not result in changes in gene/protein expression. The consensus definition of SCLC subtypes based on a transcription factor expression profile has been reported previously,[12,13] but its direct impact on subtype-based treatment is limited. A major barrier in the SCLC field is the scarcity of tumor samples available for detailed molecular characterization, especially at the protein level. Therefore, there is an urgent need for the collection of high-quality specimens to better understand the molecular basis of this cancer and accelerate both clinical and basic research.

Here, we performed an integrated genomic, transcriptomic, proteomic, and phosphoproteomic analysis of 112 Chinese SCLC specimens following the guidelines established by the Clinical Proteomic Tumor Analysis Consortium (CPTAC).[14] Our comprehensive proteogenomic study explored the functional consequences of genomic aberrations, cataloged SCLC-associated molecular characteristics, identified actionable prognostic biomarkers, and provided a better understanding of patient stratification with personalized therapeutic strategies. The underlying datasets also furnish an extensive resource to support further research in SCLC.

RESULTS

Overview of SCLC proteogenomic profiling in the Chinese population

To characterize the proteogenomic landscape of SCLC in China, 112 treatment-naive primary SCLC tumors and paired normal adjacent tissues (NATs) (Tongji University [TU]-SCLC cohort) from surgical resection were collected prospectively under standardized protocols. The clinicopathological characteristics are summarized in Table S1. All samples were cryo-pulverized and aliquoted for molecular profiling using whole-exome sequencing (WES), RNA sequencing (RNA-seq), and isobaric tandem mass tag (TMT) labeling-based global proteomics and phosphoproteomics (Figure 1A; Table S1). In total, 28,471 nonsynonymous somatic mutations and 27,080 somatic copy-number alterations (SCNAs) were identified. TMT-based proteomic analysis identified 11,209 proteins, and phosphoproteomic analysis identified 62,881 confidently localized phosphosites from 9,373 phosphoproteins. High data reproducibility and technical quality were demonstrated across the entire TMT sets (Figures S1A–S1F). Examination of gene-wise mRNA-protein correlation indicated a strikingly higher median value for tumors when compared with NATs (R = 0.5 vs. 0.07, Spearman's correlation) (Figure S1G), similar to a previous report on lung adenocarcinoma (LUAD).[5] Furthermore, tumors displayed a higher sample-wise correlation (0.31) than NATs (0.21) (Figure S1H).

Somatic alterations and mutational profiles in SCLC

The most frequently mutated genes in TU-SCLC cohort were *TP53* (72%) and *RB1* (56%) (Figure 1B). Other alterations with mutation frequency greater than 10% were found in genes encoding chromatin-modifying enzymes (*KMT2C*, *KMT2D*, *CREBBP*), protocadherin genes (*FAT1*, *FAT4*), transcriptional regulators (*ZFHX3*, *NCOR2*, *CRTC1*), and *NOTCH* family genes (*NOTCH1/2/3*). With the integrative analyses of copy-number variation, we found only 9 (8%) of the tumors showed no alterations in *TP53* and *RB1* at the genomic level (Figures 1B and S1I), supporting the essential role of TP53 and RB1 inactivation in SCLC tumorigenesis. Despite the dominant role of TP53 and RB1 inactivation in SCLC, the mutation rate of *TP53* in TU-SCLC cohort was relatively lower than that observed in cohorts predominantly from the Western populations,[9,11] yet it was comparable to that in a Japanese cohort[10] (Figure 1C). The mutation rate of *RB1* in TU-SCLC cohort was lower than that in the George et al. cohort,[9] but showed no significant difference with the other two cohorts. Additionally, we observed a high *ZFHX3* mutation frequency (19%) in TU-SCLC cohort (Figure 1C).

TU-SCLC had a median tumor mutation burden (TMB) of 5.45 nonsynonymous mutations per million base pairs. TMB-high and TMB-low were determined based on this median cutoff value, and the mutation rates of *TP53*, *KMT2C*, *FAT4*, and *GNAS* in TMB-high group were significantly higher than those in TMB-low group (Figure S1J). TMB-high patients tended to have a better overall survival (OS) (Figure 1D). This result is consistent with

Figure 1. Genomic landscape of TU-SCLC cohort

(A) Experimental workflow of SCLC proteogenomic profiling.

(B) Genetic profile and associated clinicopathologic features of all the 112 SCLC patients. MutSigCV-based significantly mutated genes ($q < 0.05$) are annotated by asterisk.

(C) Comparison of mutational frequencies between TU-SCLC cohort and other published SCLC cohorts (Fisher's exact test).

(D) Kaplan-Meier curves for overall survival based on TMB (log rank test).

(E) Clustering of SCLC tumors based on proportions of mutation signatures. Each cluster group is named according to the dominant mutation signature (that is, smoking, dMMR, and APOBEC).

(F) Kaplan-Meier curves for overall survival of smoking-dominant and dMMR-dominant patients (log rank test).

(G) Comparison of TMB and predicted neoantigen between smoking-dominant and dMMR-dominant tumors (Wilcoxon ranked-sum test).

(H) Comparison of mutation rates for *TP53*, *FAT1*, and MMR-related genes (*MSH2*, *MSH6*, and *PMS2*) between smoking-dominant and dMMR-dominant tumors (Wilcoxon ranked-sum test).

See also Figure S1 and Table S1.

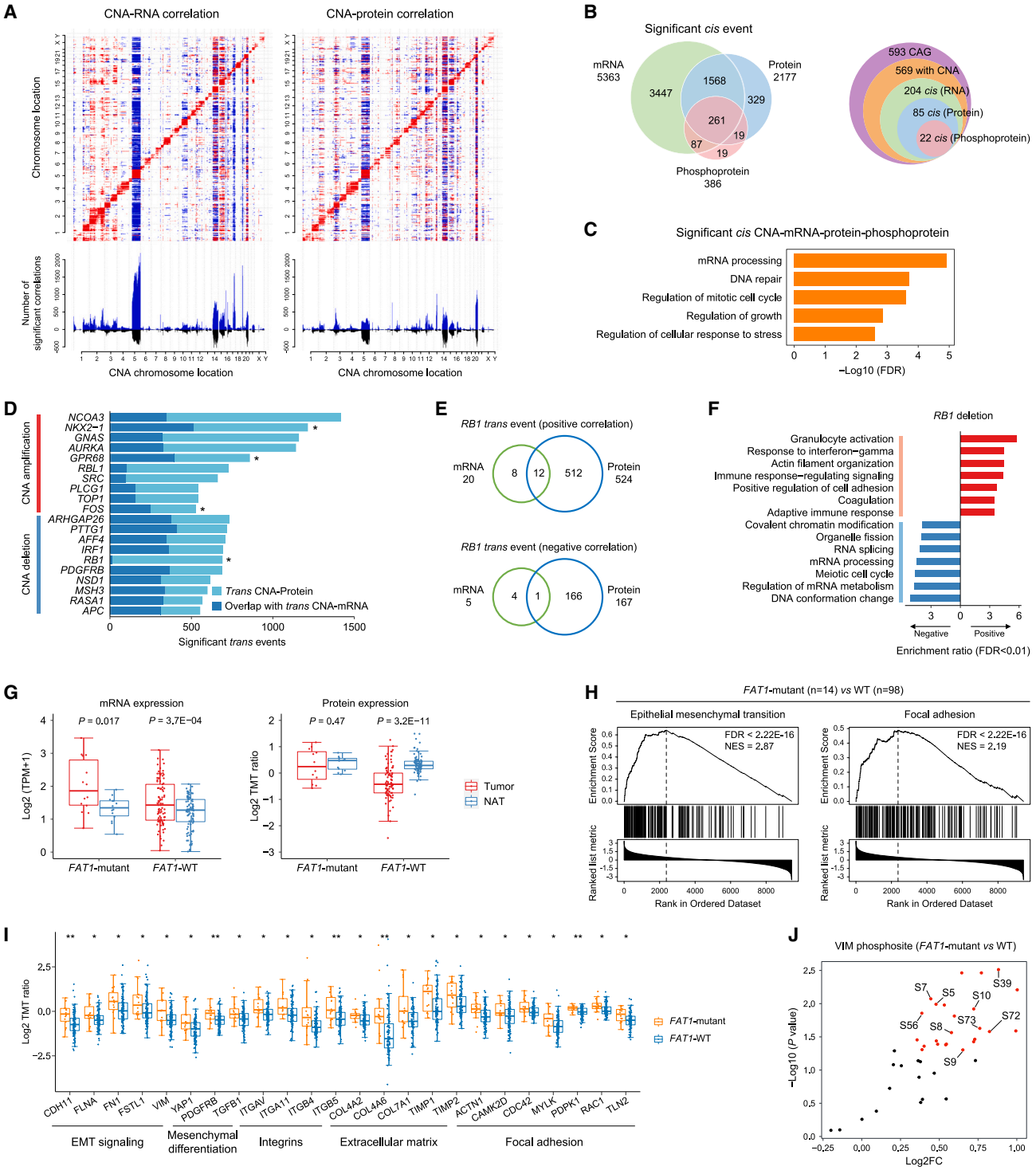


Figure 2. Impacts of genetic aberrations on proteogenomic features

(A) Correlations of CNAs to mRNA (left) and protein (right) abundance. Significant positive (red) and negative (blue) correlations (false discovery rate [FDR] < 0.05, Spearman's correlation) are indicated. The number of significant *cis* and *trans* events is shown at the bottom, blue bars representing associations specific to mRNA or protein and black bars representing associations common to both mRNA and protein.

(B) Venn diagrams depicting the *cis*-cascading effects of SCNAs (CAGs, cancer-associated genes).

(C) Enriched gene ontology (GO) biological processes (BP) for the overlapped genes of *cis* CNA-mRNA, CNA-protein, and CNA-phosphoprotein as well as differentially expressed at protein level in tumors compared with NATs.

(legend continued on next page)

previous report,[15] suggesting the potential prognostic value of TMB in SCLC.

We identified four mutational signatures (Sig1–4) based on the frequencies of mutated trinucleotide sequence motifs (Figure S1K). Tobacco mutagens exposure was best matched with Sig1 and Sig2, defective DNA mismatch repairs (dMMR) were associated with Sig3, and Sig4 bored the signature of APOBEC cytidine deaminases. According to the mutational signature contributions, 93 (83%) patients were classified as smoking-dominant, 17 (15%) were dMMR-dominant, and only 2 (2%) were APOBEC-dominant (Figure 1E). Importantly, dMMR-dominant patients showed a worse OS (Figure 1F), whereas smoking-dominant tumors had significantly higher TMB and neoantigen load (Figure 1G). We also observed less *TP53* mutations, along with more *FAT1* and mismatch repair (MMR) pathway gene (*MSH2*, *MSH6*, *PMS3*) mutations in dMMR-dominant tumors when compared with smoking-dominant ones (Figure 1H). Overall, our results reinforce the well-recognized association of tobacco carcinogens with SCLC tumorigenesis and shed new light on the role of DNA repair defects in SCLC development.

Investigating the impact of genomic aberrations by integrated analyses

SCNA analysis identified amplifications in chromosomes 1p, 1q, 3q, 5p, 8q, 14p, 14q, 18p, and 18q, along with deletions in chromosomes 3p, 4p, 4q, 5q, 13q, 15q, 16q, 17p, and 17q (Figures S2A and S2B; Table S2), as previously described in SCLC.[9,16,17] We examined the impact of SCNAs on mRNA, protein, and phosphoprotein abundance in both *cis*- and *trans*-effects (Figure 2A). A total of 5,363, 2,177, and 386 significant positive *cis*-correlations were observed for mRNA, protein, and phosphoprotein, respectively (Figure 2B). Among the 261 significant *cis*-effects overlapping across all three omics, 223 proteins showed differential abundance between tumors and NATs (Table S2), which were enriched in mRNA processing, DNA repair, and regulation of mitotic cell cycle processes (Figure 2C). We found a similar attenuation trend based on a reference list of 593 cancer-associated genes (CAGs) (Figure 2B). CNA regions with *trans*-associations were observed on chromosome 5q (Figure 2A). Most of the *trans*-effects were negatively correlated with 5q and showed an obvious attenuation from mRNA to protein expression (Figure S2C). These 5q-anticorrelated proteins were converged on multiple biological processes, including mitotic cell cycle phase transition (Figure S2D). Our multi-omics data suggested that 5q loss was associated with high expression of genes involved in DNA replication, DNA repair, and cell cycle progression (Figure S2E). Thus, loss of 5q may promote SCLC tumorigenesis via CNA *trans*-effects.

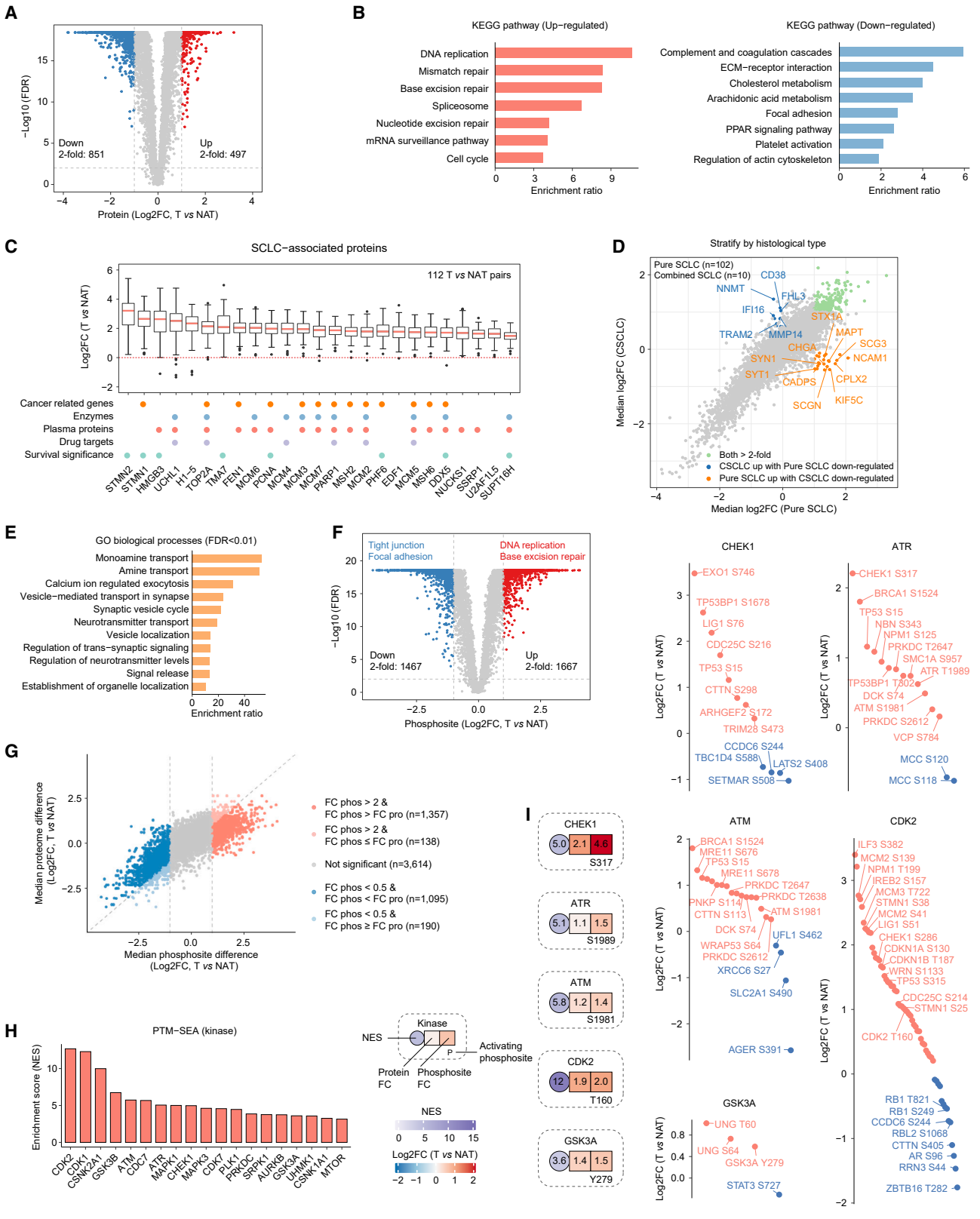
We focused on identifying CAGs with the top 10 protein-level *trans* events (Figure 2D; Table S2). Among them, *NKX2-1* (also known as thyroid transcription factor 1 [*TTF-1*]) has long been considered as a marker of lung cancer, and several CNA genes including *GNAS*, *AURKA*, *SRC*, and *TOP1* are widely reported as proto-oncogenes or drug targets in SCLC. Notably, tumor suppressor *RB1* showed copy-number deletion *trans*-effects mainly at protein level but barely at the mRNA level. Among the 691 significant *trans*-affected proteins, 524 showed positive correlation with *RB1* deletion and were enriched in pathways including granulocyte activation, response to interferon-gamma, actin filament organization, and cell adhesion, whereas the other 167 negative-correlated proteins were mainly involved in DNA conformation change, mRNA processing, and RNA splicing pathways (Figures 2E and 2F). Further *trans*-effects at phosphoprotein level confirmed these pathways affected by *RB1* deletion (Figure S2F). Additionally, we found that *RB1* deletion significantly correlated with worse survival ($p = 0.0021$; Figure S2G). These data demonstrate that protein-level *trans*-effects strongly reflect the tumorigenic impact of CNAs.

Next, we examined the impacts of somatic mutations on their cognate gene products and identified three genes (*TP53*, *FAT1*, and *GNAS*) with significant *cis*-effect (Figures S2H and S2I). Intriguingly, FAT1, which is generally considered as a tumor suppressor,[18,19] displayed increased mRNA and protein expression in *FAT1*-mutant tumors (Figure S2H). However, FAT1 protein expression was significantly downregulated in *FAT1*-wild-type (WT) tumors when compared with NATs, despite of elevated mRNA abundance (Figure 2G). Proteomic-based gene set enrichment analysis (GSEA) linked *FAT1* mutation with the upregulation of epithelial mesenchymal transition (EMT) and focal adhesion pathways (Figure 2H). Proteins involved in EMT, mesenchymal differentiation, integrins, extracellular matrix (ECM), and focal adhesion pathways were upregulated in *FAT1*-mutant tumors (Figure 2I). Moreover, our phosphoproteomic data demonstrated an increase of 24 VIM phosphosites in *FAT1*-mutant tumors (Figure 2J), and these elevated phosphosites were predominantly involved in cytoskeletal reorganization and cell motility processes as summarized in PhosphoSitePlus database.[20] Hence, our proteogenomic data indicate that FAT1 loss of function may promote an EMT phenotype in SCLC, which merits further validation.

Tumor-NAT comparisons reveal SCLC-associated proteomic events

Principal-component analysis (PCA) showed clear discrepancy between tumors and NATs at mRNA, protein, and phosphorylation levels and revealed greater heterogeneity of tumors relative to NATs (Figure S3A). Herein, we focused on the abundance

(D) Barplot showing the total number of *trans*-correlated events at mRNA and protein levels. The top 10 protein-level *trans* events for CAGs are highlighted. SCNAs that are significantly associated with patient survival (log rank test, $p < 0.05$) are annotated by asterisk.
(E) Venn diagrams depicting mRNA/proteins with positive (top) or negative (bottom) *trans* CNA-mRNA and CNA-protein correlations with *RB1* deletion.
(F) Enriched KEGG pathways for proteins with positive or negative *trans* CNA-protein correlations with *RB1* deletion.
(G) Comparison of mRNA and protein abundance between tumors and NATs for *FAT1*-mutant and *FAT1*-WT groups separately (Wilcoxon signed-rank test).
(H) GSEA plots for epithelial mesenchymal transition (EMT) and focal adhesion pathways in *FAT1*-mutant vs. *FAT1*-WT comparisons.
(I) Comparison of protein abundance between *FAT1*-mutant and *FAT1*-WT tumors for annotated pathways (*$p < 0.05$, **$p < 0.01$, Wilcoxon ranked-sum test).
(J) Comparison of VIM phophosites abundance between *FAT1*-mutant and *FAT1*-WT tumors (Wilcoxon ranked-sum test).
See also Figure S2 and Table S2.



(legend on next page)

of proteins and phosphosites with no missing values and performed differential analysis between tumors and NATs (Figures 3A and 3F; Table S3). Enrichment analysis of differentially expressed proteins revealed upregulation of pathways such as DNA replication, spliceosome, cell cycle, and DNA repair, as well as downregulation of complement and coagulation cascades, ECM-receptor interaction, focal adhesion, cholesterol metabolism, and arachidonic acid metabolism pathways in tumors (Figure 3B). After controlling for stromal and immune content, 138 out of 497 upregulated proteins remained significantly upregulated by more than 2-fold in SCLC. Among these, we focused on 25 proteins with over 2-fold elevated expression in over 90% of all tumor-NAT pairs, which were defined as SCLC-associated proteins (Figure 3C). These proteins including microtubule dynamics-related proteins STMN1/2, DNA damage repair proteins TMA7 and PCNA, and transcriptional regulators (HMGB3, PHF6, DDX5, and SUPT16H) were negatively correlated with patient survival (Figure S3B). Experimentally, knockdown of STMN1 as well as TMA7 significantly reduced cell proliferation in H345 cell lines, suggesting their oncogenic role in SCLC (Figures S3C and S3D). Most proteins changed similarly in tumor-NAT comparisons for both pure SCLC and combined SCLC (CSCLC) (Figure 3D). However, 22 proteins, mainly involved in NE-related processes, were increased specifically in pure SCLC tumors (Figure 3E), which was consistent with higher NE scores of pure SCLC (Figure S3E).

Furthermore, we observed 1,667 phosphosites with more than 2-fold upregulation in tumors when compared with NATs, among which 1,357 (81%) had greater changes in phosphosite abundance than in corresponding protein abundance (Figure 3G). We next inferred kinase activity based on the phosphorylation levels of its substrates and the corresponding kinase activating sites. Phosphosite-specific signature enrichment analysis (PTM-SEA)[21] identified 19 kinases with increased activity in SCLC (Figure 3H), among which 5 kinases displayed elevated expression of their activating sites, including CHEK1, ATR, ATM, CDK2, and GSK3A (Figure 3I). Moreover, these kinases had greater changes for tumor-NAT comparisons in phosphosite abundance than in corresponding protein abundance (Figure S3F). Specifically, CHEK1 is known to be critical for G2/M checkpoint in response to genotoxic stress.[22] We found CHEK1 protein and its S317 phosphosite level were increased in tumors, accompanied by elevated expression of substrates involved in cell cycle arrest and DNA damage response (DDR), such as CDC25C S216, TP53BP1 S1678, and TP53 S15 (Figure 3J). Furthermore, high levels of CHEK1 and the S317 phosphorylation were significantly associated with worse patient survival (Figures S3G and S3H).

The proteomic data identified a total of 8 cancer-testis (CT) antigens highly expressed in tumors (STAR Methods; Figure S3I). Unlike predicted neoantigens that were enriched in TMB-high samples, CT antigens were widely distributed independent of TMB. Moreover, compared with other CPTAC studies like LUAD[5] and colon cancer,[23] the percentage of CT antigens presence was relatively higher in SCLC. These proteomics-identified CT antigens may serve as targets for potential cancer vaccines.

Functional analyses of identified proteins associated with survival

We performed supervised analysis to identify prognostic biomarkers and revealed 16 differentially expressed proteins that were associated with patient survival (Figures 4A and S4A; Table S4). Among them, HMGB3 showed elevated expression in tumors and was correlated with worse survival, whereas CASP10 showed decreased expression and was associated with a better prognosis. These results were further validated by immunohistochemistry (IHC) assay on samples from the data producing TU-SCLC cohort and another independent SCLC cohort (n = 111) (Figures 4B–4D and S4B–S4D; Table S4), confirming their potential prognostic value in SCLC.

HMGB3 belongs to high-mobility group superfamily, which binds to nucleosomes and participates in DNA replication, recombination, repair, and transcription.[24] High expression of HMGB3 is observed in many cancers, and correlated with drug resistance and poor patient survival.[25,26] However, its expression or function in SCLC has not been reported. Since HMGB3 is upregulated in SCLC, we speculated that HMGB3 may function as an oncogenic driver in SCLC. Indeed, overexpression of HMGB3 in H345 cells increased cell migration, whereas HMGB3 knockdown led to reduced cell migration (Figures 4E and 4F).

As HMGB3 is reported as a DNA-binding protein that regulates gene transcription, we overexpressed HMGB3 in H345 cells and

Figure 3. Proteogenomic alterations associated with tumorigenesis and prognosis

(A) Volcano plot depicting differentially expressed proteins between tumors and paired NATs (Wilcoxon signed-rank test, Benjamini-Hochberg [BH] adjusted).

(B) Representative KEGG pathways for 2-fold increased and decreased proteins.

(C) Boxplot showing $\log_2$ fold change between tumors and paired NATs for SCLC-associated proteins annotated with potential clinical utilities by the Human Protein Atlas.

(D) Scatterplot depicting comparison of protein changes between tumors and paired NATs in two SCLC histological types (CSCLC, combined SCLC).

(E) Enriched GO terms for the proteins upregulated in pure SCLC but downregulated in CSCLC between tumors and paired NATs.

(F) Volcano plot depicting differentially expressed phosphosites between tumors and paired NATs (Wilcoxon signed-rank test, BH adjusted). Representative pathways are annotated.

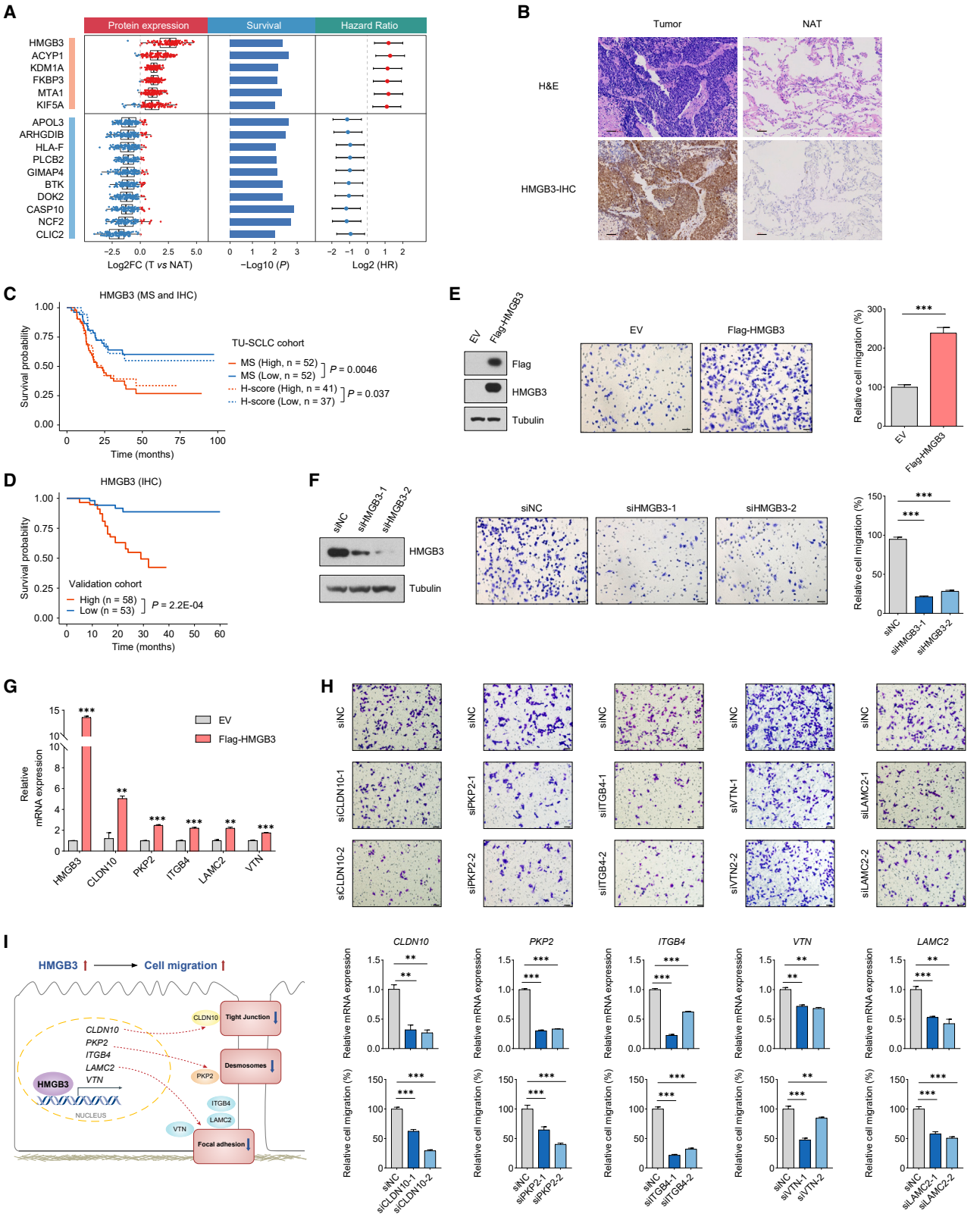
(G) Scatterplot depicting comparison of abundance changes between phosphosites and their corresponding proteins (BH adjusted $p < 0.01$, Wilcoxon signed-rank test).

(H) Significant (FDR < 0.05) kinase signatures between tumors and paired NATs assessed by PTM signature enrichment analysis (PTM-SEA).

(I) Diagram showing kinases with increased activity inferred from the phosphorylation levels of its substrates and the corresponding kinase activating sites.

(J) Dotplot showing the median $\log_2$ fold change of kinase substrates phosphorylation. Pink and blue colors indicate upregulated or downregulated phosphosites between tumors and paired NATs, respectively.

See also Figure S3 and Table S3.



(legend on next page)

conducted chromatin immunoprecipitation sequencing (ChIP-seq) analysis to identify potential HMGB3 target genes. Enrichment analysis revealed that genes bound by HMGB3 were involved in pathways controlling cell migration phenotypes, such as adherens junction, tight junction, and focal adhesion (Figures S4E and S4F). Among these, we focused on 5 genes previously reported to promote cancer cell migration,[27–31] including *CLDN10*, *PKP2, ITGB4*, *VTN*, and *LAMC2* (Figure S4G). Consistent with the ChIP-seq results, HMGB3 overexpression led to increased expression of these genes at mRNA level (Figure 4G). Thus, we hypothesized that HMGB3 might regulate the transcription of *CLDN10*, *PKP2, ITGB4*, *VTN*, and *LAMC2*, by which HMGB3 contributes to the high mobility of SCLC cells. As expected, knockdown of *CLDN10*, *PKP2, ITGB4*, *VTN*, or *LAMC2* significantly reduced cell migration in both parental and HMGB3-overexpressed H345 cells (Figures 4H and S4H). Our results indicate that HMGB3 is a robust prognostic biomarker and functions as an oncogenic factor in SCLC via transcriptional regulation of cell junction-related genes (Figure 4I).

Immune landscape of TU-SCLC cohort

ESTIMATE immune score[32] was used to evaluate the levels of immune cell infiltration, and our results suggested that immune cell infiltration was significantly associated with gender, tobacco smoking, TNM stages, and patient survival (Figures S5A and S5B; Table S5). Unsupervised clustering based on xCell-derived cell type enrichment scores[33] for both tumors and NATs identified three immune clusters including hot-tumor-enriched, cold-tumor-enriched, and NAT-enriched subtypes (Figure 5A). Most tumors belonged to immune-cold subtype and were associated with worse prognosis (Figure 5B). Intriguingly, immune-cold tumors exhibited higher NE scores (Figure 5C), and there was a significant inverse correlation between immune scores and NE scores ($R = -0.59$, Spearman's correlation). This further suggests a linkage between immune-cold tumors and the NE phenotype.[34,35] Additionally, immune-hot subtype was characterized by stronger signatures of both cytotoxic immune cells ($CD8^+$ T cells, natural killer cells, activated dendritic cells [DCs], and M1 macrophages) and immune inhibitory cells (regulatory T cells and M2 macrophages) and showed upregulation of multiple immune-related pathways including interferon-gamma response, antigen processing and presentation, and natural killer cell-mediated cytotoxicity (Figure 5A). T cell markers (CD3D, CD8A), MHC molecules, stimulatory and inhibitory immune modulator protein signatures, cytolytic activity (CYT score; STAR Methods),[36] as well as immune checkpoint molecules (CTLA4, PD-L1, PD-1, and IDO1) expression were also upregulated in immune-hot when compared with immune-cold tumors (Figure 5A). Thus, we speculated that immunotherapies may benefit this SCLC population of immune-hot subtype.

Notably, we observed a significant enrichment of *ZFHX3* mutation in immune-hot tumors (Figure 5D). Tumors harboring *ZFHX3* mutation presented higher immune scores and TMB when compared with *ZFHX3*-WT tumors (Figures 5E and 5F), suggesting an association of *ZFHX3* mutation with elevated tumor immunogenicity. Patients with *ZFHX3* mutation appeared to have better survival than *ZFHX3*-WT patients (Figure S5C). Proteins significantly upregulated in *ZFHX3*-mutant samples were enriched in immune pathways, including response to interferon-gamma and antigen processing and presentation (Figure S5D). Consistently, MHC molecules, stimulatory and inhibitory immune modulators, as well as xCell-derived $CD4^+$, $CD8^+$ T cell, DC, and M1 macrophage signatures were upregulated in *ZFHX3*-mutant tumors (Figures 5G and 5H), whereas *ZFHX3*-mutant-associated NATs showed no statistical difference in these immune cells. Thus, we linked *ZFHX3* mutation to immune activation behavior and speculated *ZFHX3* mutation as a potential predictive biomarker for immunotherapy. In support of this hypothesis, we collected 12 pre-treatment samples from SCLC patients received neoadjuvant PD-1/PD-L1 blockade combined with chemotherapy in our center, involving 2 ongoing phase II trials (NCT04539977 and NCT04542369). Major pathologic response (MPR) was used to assess pathologic response of neoadjuvant immunotherapy efficacy, which is defined as having no more than 10% residual tumors after therapy. Among them, 5 patients were categorized as MPR, and 7 patients were categorized as non-MPR. WES results suggested that 3 patients had *ZFHX3* mutation, and it is worth noting that all the 3 patients (100%) with *ZFHX3* mutation belonged to be MPR, whereas only 2 WT (22%) patients were MPR (Fisher's exact test, $p = 0.045$) (Figures 5I and 5J). Similarly, the number of residual tumor cells in the *ZFHX3*-mutant group was significantly lower than the *ZFHX3*-WT group among these 12 clinical trial patients (Figure 5K). These results support our hypothesis and indicate that patients with *ZFHX3* mutation might benefit more from immunotherapy.

Figure 4. Identification and validation of proteomic prognostic biomarkers

(A) Diagram representing 16 candidate prognostic proteins.

(B) Representative hematoxylin and eosin (H&E) and immunohistochemical (IHC) staining images for HMGB3 on tumors and paired NATs (scale bars, 60 μm).

(C) Kaplan-Meier curves for overall survival based on HMGB3 proteomic abundance or immunostaining scores (log rank test).

(D) Kaplan-Meier curves for overall survival based on HMGB3 immunostaining scores in an independent SCLC cohort (n = 111) (log rank test).

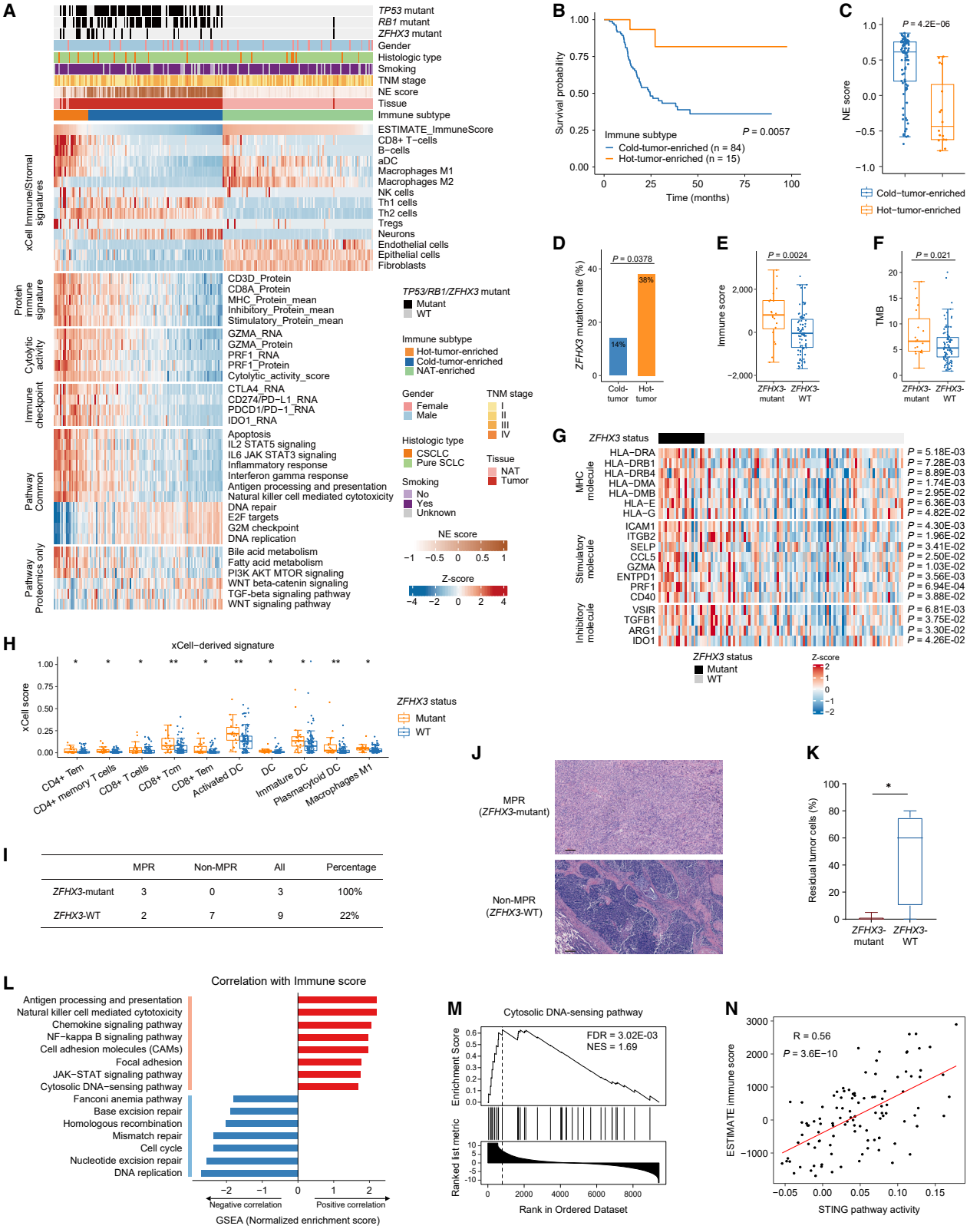
(E and F) The impacts of HMGB3 overexpression (E) or knockdown (F) on SCLC cell migration. Ectopically expressing FLAG-tag empty vector or HMGB3 and HMGB3 knockdown by siRNA in H345 cells were generated and validated by western blot. Transwell migration assays of indicated cells were measured. Data are represented as mean ± SEM (Student's t test), ***$p < 0.001$. (scale bars, 50 μm.)

(G) Real-time qPCR validation of representative HMGB3 target genes in H345 cells overexpressing HMGB3 compared with vector control. Data are shown as mean ± SEM (Student's t test), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(H) The impacts of CLDN10, PKP2, ITGB4, VTN, and LAMC2 knockdown on SCLC cell migration. siRNA transfection was performed in H345 cells, and the depletion efficiency was validated by real-time qPCR for each gene. Transwell migration assays of indicated cells were measured. Data are represented as mean ± SEM (Student's t test), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. (scale bars, 50 μm.)

(I) Schematic work model representing the role of HMGB3 in SCLC cell migration.

See also Figure S4 and Table S4.



(legend on next page)

To further explore potential determinants of immune-cold tumors, we performed correlation analysis using the immune scores and proteomic data. As expected, immune-related pathways showed positive correlations, whereas DNA replication, cell cycle, and DNA damage repair pathways were negatively correlated with the immune scores (Figure 5L). Since the deficiency of DDR in cancer cells is critical for tumor immunogenicity,[37,38] and SCLC cells are especially dependent on high DDR activity for survival because of extremely high genomic instability and replication stress, we thus speculated that elevated activity of DDR may lead to an immunosuppressive response in SCLC. We found negative associations between specific DNA repair pathway scores[39] and immune scores (Figure S5E). DNA repair proteins (PARP1, MSH2, MSH6, XRCC1, XRCC5, XRCC6) and DNA damage checkpoints (CHEK1, WEE1, AURKA/B) were also increased in immune-cold tumors at protein abundance as well as phosphorylation levels (Figures S5F and S5G). In addition, the cytosolic DNA-sensing pathway displayed strongly positive correlation with immune scores (Figures 5L and 5M), indicating that the cGAS-STING pathway, a critical innate immune pathway activated by cytoplasmic double-strand DNA, might be associated with immune infiltration. Indeed, STING pathway activity and the expression of downstream proinflammatory chemokines CCL5 and CXCL10 were significantly correlated with immune features (Figures 5N and S5H). The cGAS-STING pathway-related proteins were upregulated in immune-hot tumors (Figure S5I), and their higher abundance was associated with better survival (Figure S5J). Moreover, cGAS, STING1, CCL5, and CXCL10 protein expression showed significantly negative correlation with DDR activity, and STING1 protein expression was inversely correlated with the expression of DNA repair proteins and DNA damage checkpoints (Figure S5K). Together, our proteogenomic data indicate that elevated DDR activity may contribute to immune suppression by attenuating the activation of the cGAS-STING pathway in SCLC.

Multi-omics subtypes with distinct biological features

We applied non-negative matrix factorization (NMF)-based unsupervised clustering by integrating mRNA, protein, and phosphorylation data from 107 SCLC tumors and grouped them into four clusters (Figures 6A and S6A; Table S6). Significant associations between the multi-omics subtypes and clinical parameters were presented in Figure S6B. The nmf1 subtype was primarily enriched in cell cycle, DNA damage, chromatin organization, and epigenetic regulation pathways, suggesting a subgroup with a high proliferation rate. The nmf2 subtype had fewer signature proteins or phosphosites but showed the highest TMB (Figure 6B). The nmf3 subtype was associated with ECM-receptor interaction, ECM organization, and focal adhesion pathway at protein level. Remarkably, phosphoproteomic rather than transcriptomic or proteomic data showed increased activity of signaling by receptor tyrosine kinases (RTKs) in the nmf3 subtype. Finally, the nmf4 subtype was distinguished by high expression of MYC targets and enrichment of RNA metabolism pathways, as well as characterized by higher chromosome instability (CIN), stemness score, and mRNA-protein correlation (Figures 6A and 6B). Similarly, subtype-specific pathway enrichment analysis also demonstrated distinct molecular features among the four subtypes (Figure S6C).

Each subtype showed distinct profiles of CNA (Figure S6D). Deletion of *TP53*, *RB1*, and *FAT1* was significantly enriched in the nmf1 and nmf2 subtypes, and *ZFHX3* deletion was mostly restricted to the nmf1 subtype. By contrast, the oncogene *AURKA* and *GNAS* amplification were preferentially associated with the nmf4 subtype (Figure 6C).

We examined the patterns of transcription factors ANPY (ASCL1, NEUROD1, POU2F3, and YAP1) expression across the four subtypes based on mRNA expression and found clear separation (Figures 6D and 6E). The nmf1 subtype showed high expression of ASCL1, NEUROD1, or both, consistent with previous studies that many SCLC tumors co-express ASCL1 and NEUROD1.[40–42] High expression of POU2F3 was mainly restricted to the nmf4 subtype and was mutually exclusive with ASCL1 and NEUROD1. The expression of YAP1 was very low compared with other three transcription factors, even though YAP1 expression was relatively higher in the nmf3 and nmf4 subtypes. In addition, we observed the highest NE scores in the nmf1 subtype (Figure 6F), accompanied by high protein expression of NE markers SYP, CHGA, NCAM1, and INSM1 (Figure 6D). The nmf3 subtype displayed the

Figure 5. Immune landscape of TU-SCLC cohort

(A) Heatmap illustrating three immune clusters based on xCell-derived cell signatures. Cell type compositions, mRNA and protein abundance of key immune modulators, activity scores, and pathways that were differentially regulated in immune-hot and immune-cold subtypes are illustrated.

(B) Kaplan-Meier curves for overall survival based on immune subtypes (log rank test).

(C) Comparison of NE scores between cold-tumor-enriched and hot-tumor-enriched subtypes (Wilcoxon ranked-sum test).

(D) Comparison of *ZFHX3* mutation rate between immune-cold and immune-hot tumors (Fisher's exact test).

(E and F) Comparison of immune scores (E) and TMB (F) between *ZFHX3*-mutant and *ZFHX3*-WT tumors (Wilcoxon ranked-sum test).

(G) Heatmap illustrating differentially expressed proteins between *ZFHX3*-mutant and *ZFHX3*-WT tumors involved in MHC molecules, stimulatory and inhibitory immune modulators (Wilcoxon ranked-sum test).

(H) Comparison of xCell-derived immune cell signatures between *ZFHX3*-mutant and *ZFHX3*-WT tumors (*$p < 0.05$, **$p < 0.01$, Wilcoxon ranked-sum test). Tem, effector memory T cells; Tcm, central memory T cells; DC, dendritic cell.

(I) The association between *ZFHX3* mutation and immunotherapy response observed in ongoing clinical trials.

(J) Representative H&E images for MPR in patient with *ZFHX3* mutation, and non-MPR in patient without *ZFHX3* mutation who received neoadjuvant immunotherapy combined with chemotherapy (scale bars, 100 μm).

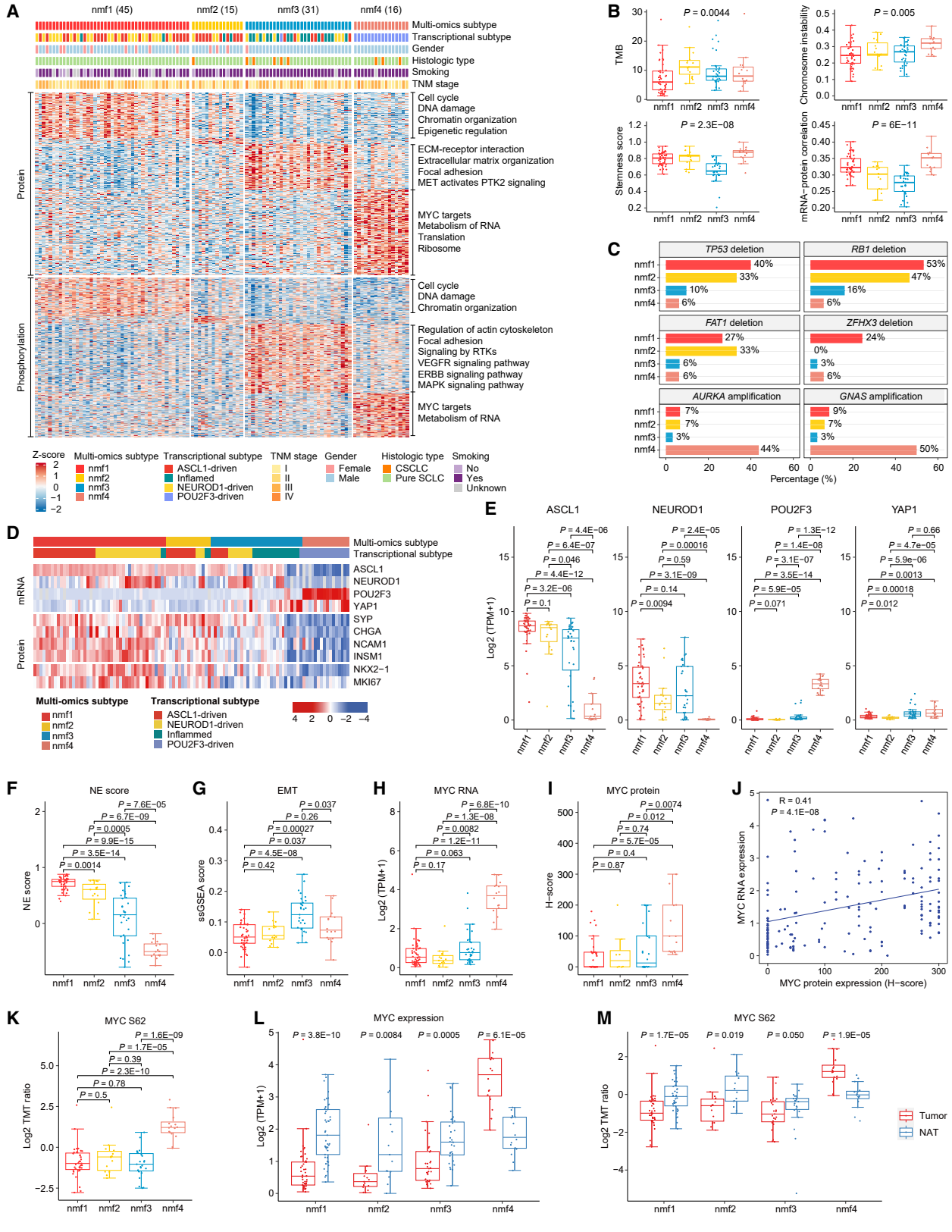
(K) Comparison of residual tumor cells between *ZFHX3*-mutant and *ZFHX3*-WT groups from clinical trials patients (*$p < 0.05$, Wilcoxon ranked-sum test).

(L) Barplot showing normalized enrichment scores for the top KEGG pathways correlated (red) or anticorrelated (blue) with immune scores.

(M) GSEA plot for cytosolic DNA-sensing pathway correlated with immune scores.

(N) Scatterplot showing Spearman's correlation of immune scores and ssGSEA-derived STING pathway activity.

See also Figure S5 and Table S5.



(legend on next page)

highest EMT score (Figure 6G), which was associated with SCLC metastasis and chemoresistance.[43,44] Our proteomic data also showed that the nmf3 subtype expressed the highest levels of mesenchymal markers, including VIM, ZEB2, and AXL, while exhibiting the lowest level of epithelial marker EPCAM (Figure S6E).

The amplification of *MYC* showed no statistically significant difference among these subtypes (Figure S6F), whereas MYC mRNA expression displayed significant upregulation in the nmf4 subtype (nmf4 vs. [nmf1 or nmf2 or nmf3], fold change [FC] > 7.5; Figure 6H). Further, IHC staining for MYC protein in SCLC tumors and NATs confirmed these differences (Figures 6I, 6J, and S6G). Phosphoproteomics data showed that MYC S62, S347, S348, and S161 phosphosites were also upregulated in the nmf4 subtype (Figures 6K, S6F, and S6H). Among them, MYC S62 is the most studied phosphosites, which can promote stabilization of MYC and enhance MYC activity.[45–48] MYC S347 and S348 have been identified to promote cell growth.[49] These results suggested an active oncogenic role of MYC in the nmf4 subtype. When compared with NAT samples, MYC expression was only upregulated in the nmf4 subtype tumors at RNA, protein, and phosphorylation levels, while downregulated in other subtypes (Figures 6L, 6M, S6G, and S6I). Thus, our findings implicate that MYC expression rather than *MYC* gene amplification could better indicate MYC status and may be a more valuable marker to define a subset of SCLC with low NE properties and high-POU2F3 expression.

Additionally, xCell-based immune cell infiltration and proteogenomic-derived immune modulator signatures revealed a differential pattern of immune features across the multi-omics subtypes (Figure S6J). The nmf3 and nmf4 subtypes had higher immune scores (Figure S6K). Immune-hot tumors in the nmf3 subtype included most of the transcriptional inflamed subtype, and other immune-hot tumors were mainly allocated to the nmf4 subtype. This is consistent with our previous study[50] and further supports that the immune infiltration subtype seems to correspond to a combination of POU2F3-driven and inflamed subtypes. Together, these data provide a comprehensive characterization of molecular heterogeneity among SCLC tumors and suggest the necessity of subtype-specific treatment strategies.

Subtype-specific therapeutic strategies in SCLC

Based on the molecular features of multi-omics subtypes, we explored subtype-specific therapeutic vulnerabilities of SCLC. The nmf1 subtype was associated with high proliferation, E2F activity, as well as replication stress (Figure S7A). A recent study suggests that concurrent inhibition of ATR and TOP1 results in durable SCLC regression, especially in patients with high replication stress and NE differentiation.[51] Using the differentially expressed genes between responding and non-responding tumors from this study,[51] we detected a strong response score toward ATR and TOP1 inhibition in the nmf1 subtype (Figure S7B). Consistently, the response score was predicted to be highest in the nmf1 subtype from another two SCLC datasets[9,52] (Figure S7B; STAR Methods). Moreover, we found that the nmf2 subtype showed the highest protein level of the inhibitory NOTCH ligand delta-like protein 3 (DLL3) and the lowest mRNA level of the NOTCH transcriptional target REST (Figure S7C). Similarly, high DLL3 expression was observed in the nmf2 subtype from the George et al. dataset[9] (Figure S7D). Since DLL3 has been proposed as a promising target in SCLC and its expression significantly correlates with high therapeutic efficacy of DLL3-targeted drugs,[53–55] it remains very likely that the nmf2 subtype might benefit from therapeutics targeting DLL3. Remarkably, our phosphoproteomic data showed that the activity of RTK signaling was notably upregulated in the nmf3 subtype (Figure 6A). Anlotinib, a multi-target tyrosine kinase inhibitor known for its anti-angiogenesis and anti-tumor activity, has been proven to improve progression-free survival (PFS) and OS in SCLC patients.[56] We proposed that targeting RTKs may be a potential strategy for the nmf3 subtype. The nmf4 subtype showed high MYC expression and the enrichment of MYC-related pathways, further implicating a potential opportunity for targeting AURKA/B.[57,58]

To validate these proposed subtype-specific therapeutic strategies, we analyzed a total of 16 SCLC patient-derived xenograft (PDX) and 6 human SCLC cell line-derived xenograft (CDX) models through RNA-seq, proteomics, and phosphoproteomics characterization among which 10 tumors were concordant with nmf1, 3 with nmf2, 5 with nmf3, and 4 with nmf4 (Figure S7E; Table S7; STAR Methods). Since we have no DLL3-targeting reagents available, we focused on the validation of the nmf1, nmf3, and nmf4 subtypes. We found that the nmf1 tumor models held great response to ATR and TOP1 inhibition (Figure S7F). *In vivo* results further showed that the nmf1 subtype (H69 CDX model) was susceptible to etoposide + cisplatin (E/P) treatment (Figure S7G). This is consistent with our previous findings in other

Figure 6. Multi-omics classification of TU-SCLC cohort and the corresponding biological features

(A) Integrative multi-omics classification of SCLC tumors into four NMF-derived subtypes. Heatmap showing the signature proteins and phosphosites for each subtype. Representative pathways are also annotated.

(B) Comparison of TMB, chromosome instability (CIN), stemness scores, and mRNA-protein correlation across multi-omics subtypes (Kruskal-Wallis test).

(C) The percentages of specific genes with CNA for each subtype ($p < 0.05$, Fisher's exact test).

(D) Heatmap showing relative mRNA expression of ANPY, protein expression of neuroendocrine markers, thyroid transcription factor 1 (TTF-1/NKX2-1), and proliferation index marker Ki-67/MKI67 across multi-omics subtypes. A, ASCL1; N, NEUROD1; P, POU2F3; Y, YAP1.

(E) Comparison of ANPY expression across multi-omics subtypes (Wilcoxon ranked-sum test).

(F and G) Comparison of NE scores (F) and EMT scores (G) inferred from proteogenomic data across subtypes (Wilcoxon ranked-sum test).

(H and I) Comparison of MYC mRNA abundance (H) and protein abundance by IHC assays (I) across subtypes (Wilcoxon ranked-sum test).

(J) Scatterplot showing Spearman's correlation of MYC mRNA with protein abundance (H scores).

(K) Comparison of MYC S62 abundance across subtypes (Wilcoxon ranked-sum test).

(L and M) Comparison of MYC mRNA (L) and MYC S62 (M) phosphosite abundance between tumors and paired NATs in each subtype, respectively (Wilcoxon signed-rank test).

See also Figure S6 and Table S6.

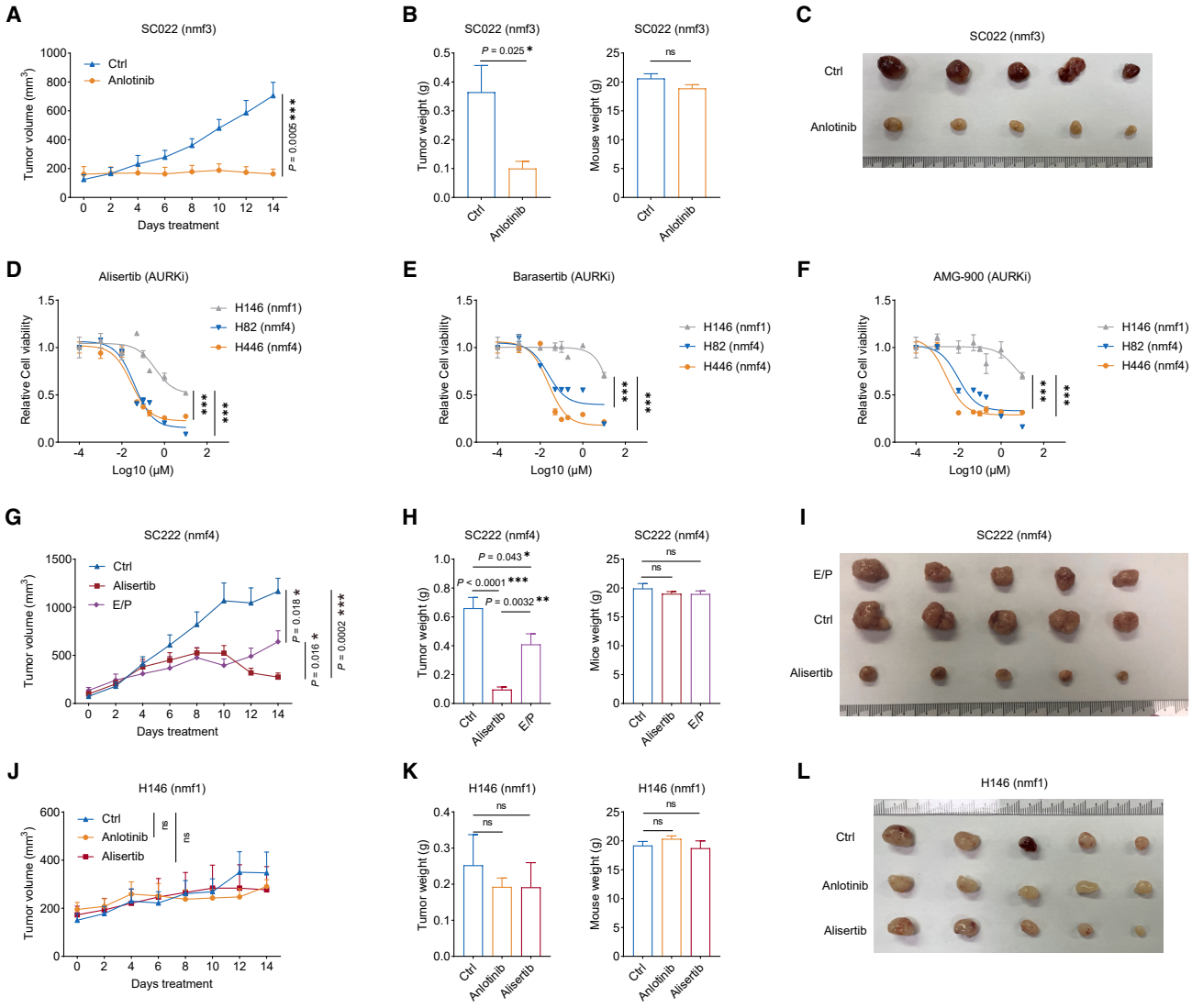


Figure 7. Subtype-specific therapeutic strategies in SCLC

(A) Tumor growth in SC022 tumor-bearing mice treated with anlotinib or vehicle. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(B) Statistical analyses of tumor weights (left) and mouse weights (right) from SC022 PDX models treated with anlotinib or vehicle. Data are represented as mean ± SEM (Student's t test).

(C) The tumors from SC022 PDX models treated with anlotinib or vehicle.

(D–F) Relative viability of H146, H82, and H446 cells after 72 h of treatment with increasing concentrations of AURK inhibitor alisertib (D), barasertib (E), or AMG-900 (F). Data are represented as mean ± SEM (n = 3 replicates, Student's t test, ***p < 0.001).

(G) Tumor growth in SC222 tumor-bearing mice treated with E/P, alisertib, or vehicle. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(H) Statistical analyses of tumor weights (left) and mouse weights (right) from SC222 PDX models treated with E/P, alisertib, or vehicle. Data are represented as mean ± SEM (Student's t test).

(I) The tumors from SC222 PDX models treated with E/P, alisertib, or vehicle.

(J) Tumor growth in H146 tumor-bearing mice treated with anlotinib, alisertib, or vehicle. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(K) Statistical analyses of tumor weights (left) and mouse weights (right) from H146 CDX models treated with anlotinib, alisertib, or vehicle. Data are represented as mean ± SEM (Student's t test).

(L) The tumors from H146 CDX models treated with anlotinib, alisertib, or vehicle.

See also Figure S7 and Table S7.

nmf1 tumors including SC234, SC007, SC185 PDX models, and H146 CDX model.[59]

As for the nmf3 subtype, we examined the therapeutic efficacy of the RTK inhibitor anlotinib in multiple nmf3 PDX/CDX models including SC022, SC224, and DMS114. Strong inhibition of tumor growth was observed in all three nmf3 tumor models (Figures 7A–7C and S7H–S7K), whereas no obvious inhibition was observed in the nmf1 tumor models (H146 and

SC234) (Figures 7J–7L and S7O–S7Q). As for the nmf4 subtype, we tested the effect of AURK inhibitors (alisertib, barasertib, and AMG-900) on nmf4 SCLC cell lines (H82 and H446) as well as PDX/CDX models (SC222 and H446). We found that the nmf4 cell lines were more responsive to alisertib, barasertib, and AMG-900 treatment when compared with the nmf1 subtype (H146) (Figures 7D–7F). Moreover, alisertib treatment significantly inhibited tumor growth in the nmf4 PDX/CDX models (Figures 7G–7I and S7L–S7N) while barely eliciting any inhibition in the nmf1 tumor models (H146 and SC234) (Figures 7J–7L and S7O–S7Q). These data highlight the importance of subtype-specific targeting in SCLC and support the utility of multi-omics subtypes in guiding treatment selection.

DISCUSSION

A significant challenge in the SCLC field is the scarcity of tumor specimens available for research studies.[2] Most of our knowledge about SCLC has been obtained from studies using model-centric approaches, leading to a gap between basic research and patient outcomes.[60] Therefore, there is a great need for the collection of SCLC patient samples to understand the molecular basis of this disease and accelerate both clinical and basic research. Genomic analyses of SCLC have broadened our understanding of the molecular events of this aggressive malignancy.[9,11,16,17,61] In this study, we reported a comprehensive proteogenomic characterization of 112 treatment-naive and resectable SCLC tumors and paired NATs. We found that TP53 and RB1 inactivation were the primary genetic alterations, with only 9 patients (8%) showing no DNA-level alterations. Interestingly, the somatic mutation rate of *TP53* was lower when compared with cohorts predominantly from Western populations,[9,11] but comparable with Asian cohorts from Japanese.[10,62,63] A real-world study also indicates that patients of African genetic ancestry exhibited fewer *TP53* or *RB1* alterations compared with European patients.[11] These findings suggest that differences such as ethnicity or other clinical parameters could influence the mutational profile of SCLC. Interestingly, recent discoveries demonstrated that *TP53*/*RB1*-WT tumors represented a distinct genetic subtype of SCLC associated with human papilloma virus (HPV) infection.[11] Although no HPV infection was detected in TU-SCLC cohort, likely due to the limited patient number, we believed that other mechanisms beyond genomic alteration may inactivate TP53 and RB1.

Paired tumor-NAT analyses enabled us to identify SCLC-associated proteins, activated kinases, and CT antigens. We identified and validated HMGB3 and CASP10 as putative prognostic biomarkers. CASP10 is a member of the caspase family and could suppress tumorigenesis by affecting apoptotic, metabolic, or epigenetic reprogramming pathways.[64,65] We found lower levels of CASP10 expression in SCLC tumors and its association with worse patient outcome. Further studies are warranted to investigate its contribution to SCLC tumorigenesis. HMGB3 has been reported as a highly expressed protein in various cancer types,[25,26] which may drive cancer metastasis and growth. Previous studies found that downregulation of HMGB3 expression can inhibit migration and invasion in various cancer cells,[66–68] but the underlying mechanisms have not been revealed. We found that HMGB3 could promote SCLC cell migration via modulating the transcription of cell junction-related genes. Interestingly, our data also indicated that HMGB3 is associated with the PI3K-Akt and MAPK signaling pathways through a transcription-dependent manner, which may provide more clues to understanding the role of HMGB3 in promoting SCLC tumorigenesis. The connection of HMGB3 to the activation of PI3K-Akt and MAPK signaling pathways has also been reported in colorectal carcinoma[69] and glioblastoma,[70] respectively. Therefore, these results indicate that HMGB3 may play an important and complex role in promoting SCLC, in addition to serving as a robust prognostic biomarker for SCLC patients.

Immunotherapy has been regarded as a promising treatment for SCLC, but its benefits are still limited. Recent clinical trials suggested that predictive biomarkers and combination therapies are required to achieve better clinical outcomes.[71,72] Proteogenomic analysis allowed us to characterize the immune landscape of SCLC and revealed three immune clusters. We linked *ZFHX3* mutation to immune-hot behavior and clinically demonstrated that *ZFHX3* mutation could be a potential predictive biomarker for SCLC patients receiving immunotherapy. Interestingly, an exploratory analysis focused on NSCLC immunotherapy cohort also indicated *ZFHX3* mutation was an independent predictive biomarker.[73] Moreover, we found that elevated DDR activity may promote immune suppression via attenuating the activation of the cGAS-STING pathway and extended recent reports demonstrating strong synergy between DDR inhibitors and immune checkpoint blockade (ICB) by STING-mediated T cell activation in the murine SCLC model.[74,75]

Unsupervised clustering divided SCLC tumors into four subtypes with biological differences and various therapeutic vulnerabilities. The nmf1 subtype was associated with high proliferation rate, replication stress, and NE differentiation. This suggests a potential response to drugs that exacerbate genome instability, and this hypothesis was validated by E/P-based chemotherapy treatment. The nmf2 subtype showed a high abundance of DLL3 expression, suggesting potential response to anti-DLL3 therapies, such as the T cell engager (TCE) molecules tarlatamab.[53,76] The nmf3 subtype was characterized by high EMT status and elevated RTK signaling and thus may benefit from RTK inhibitor treatment. As expected, the experimental results showed that anlotinib was more effective in the nmf3 subtype CDX/PDX models than those subtyped to nmf1. By contrast, the nmf4 subtype displayed a non-NE subtype with high expression of MYC and POU2F3, which may be sensitive to AURK inhibitors.[58] Indeed, we found the nmf4 subtype tumors were more responsive to alisertib than those subtyped to the nmf1 subtype both in *in vitro* cell cultures and *in vivo* PDX/CDX models. Therefore, these results indicate that the multi-omics subtyping and their respective characteristics may be valuable in guiding treatment selection of SCLC.

Altogether, this work extends the understanding of SCLC cancer biology that could not be captured by genomic analysis and generates subtypes that may guide precision therapeutics. We hope that these findings will contribute to more effective SCLC clinical treatment in addition to providing a valuable resource for basic and clinical researchers.

Limitations of the study

The objectives of this study were to comprehensively characterize SCLC using multi-omics analysis and provide proteogenomic resources. For this purpose, collected samples were treatment-naive and surgically resected. Consequently, there are inherent limitations of this study. (1) The present cohort comprises resectable samples, which limits the exploration to progressive and metastatic tumors, even though most patients were diagnosed with metastatic disease.[77] Moreover, analysis of treatment-naive samples lacks treatment data from therapeutic drugs and limits the investigation to therapy-related proteogenomic features. Focused analyses of expanded cohorts in the future will contribute to providing critical insights into metastatic biology and treatment perturbations that may shed light on mechanisms of response and resistance to therapy. (2) Proteogenomic analysis of this study is deployed using bulk tumors and NAT tissues, which improved the depth and internal concordance of molecular analysis but sacrificed intra-tumor heterogeneity information in cellularity and tumor microenvironment. Emerging methods geared to more spatially resolved proteogenomics or using single-cell genomics and proteomics will be useful adjuncts.[78–80] (3) Although some of the key findings from proteogenomic analysis were validated using cell lines and CDX/PDX models, many molecular alterations are hypothesis-generating. Therefore, a wider community effort will be required to validate biological conclusions and treatment predictions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Clinical Specimens
  - Cell Lines
  - Animals
- METHOD DETAILS
  - Specimen Acquisition and Preparation
  - Genomics and Transcriptomics Profiling Experiments
  - Proteomics and Phosphoproteomics Profiling Experiments
  - Genomic and Transcriptomic Data Processing
  - Proteomics and Phosphoproteomics Data Processing
  - Data Quality Control
  - Integrated Analysis
  - Immunohistochemistry Analysis
  - Functional Experiments
  - Drug Response Experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.12.004.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (grants 82125001, 81972172, 81925029, 82230098, 32221002, 81790253, 81872312, 82011540007, 31621003, and 82030083), the National Key Research and Development Program of China (grants 2022YFA1302902, 2020YFA0509001, 2020YFA0803203, 2019YFA0802102, 2022YFA1103900, and 2020YFA0803300), the Innovation Program of Shanghai Municipal Education Commission (grant 2023ZKZD33), the foundation of Shanghai Pulmonary Hospital (grant FKCX2304), the Shanghai Young Excellent Academic Leader Program (grant 20XD1424900), the Shanghai Municipal Science and Technology Major Project, the CAS project for young scientists in basic research (grant YSBR-014), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB19020201), the Basic Frontier Scientific Research Program of Chinese Academy of Science (grant ZDBS-LY-SM006), the International Cooperation Project of Chinese Academy of Sciences (grant 153D31KYSB20190035), the Innovative Research Team of High-Level Local Universities in Shanghai (grant SSMU-ZLCX20180500), and the China Postdoctoral Science Foundation (grants 2023M732650, 2022M713138, and 2022T150646). We would like to thank the Institutional Technology Service Center of the Shanghai Institute of Materia Medica for all the technical support. This work was done under the auspices of the US National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). CPTAC collaborates with international organizations/institutions to accelerate the understanding of the molecular basis of cancer through the application of proteogenomics, standards development, and publicly available datasets. Cartoons in graphical abstract were created with biorender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, P.Z., H. Zhou, H.J., and D.G.; methodology, Q.L., J.Z., C.G., M.W., C. Wang, Y.Y., L.S., and H. Zhu; validation, L. Hou and C. Wu; formal analysis, Q.L., J.Z., C. Wang, Y.Y., L.S., L.Z., H.Y., and Y. Zhou; investigation, Q.L., J.Z., C.G., M.W., T.Z., L. Hu, and S.F.; resources, D.W., Y. Zhu, G.J., J.L., P.Z., H. Zhou, H.J., and D.G.; data curation, Q.L. and J.Z.; writing – original draft, Q.L., J.Z., C.G., and M.W.; writing – review & editing, Y.L., H. Zhang, B.Z., L.D., A.I.R., H.R., D.G., H.J., H. Zhou, and P.Z.; visualization, Q.L., J.Z., C.G., and M.W.; supervision, P.Z., H. Zhou, H.J., and D.G.; funding acquisition, P.Z., H. Zhou, H.J., D.G., Q.L., and C.G.

DECLARATION OF INTERESTS

J.L. is an employee of D1 Medical Technology.

Received: December 19, 2022
Revised: September 25, 2023
Accepted: December 1, 2023
Published: January 4, 2024

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *71*, 209–249.
2. Rudin, C.M., Brambilla, E., Faivre-Finn, C., and Sage, J. (2021). Small-cell lung cancer. Nat. Rev. Dis. Primers *7*, 3.
3. Sabari, J.K., Lok, B.H., Laird, J.H., Poirier, J.T., and Rudin, C.M. (2017). Unravelling the biology of SCLC: implications for therapy. Nat. Rev. Clin. Oncol. *14*, 549–561.
4. Megyesfalvi, Z., Gay, C.M., Popper, H., Pirker, R., Ostoros, G., Heeke, S., Lang, C., Hoetzenecker, K., Schwendenwein, A., Boettiger, K., et al. (2023). Clinical insights into small cell lung cancer: tumor heterogeneity, diagnosis, therapy, and future directions. CA Cancer J. Clin. *73*, 620–652.

5. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell *182*, 200–225.e35.

6. Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H., Chen, H.W., Chang, G.C., Chang, Y.L., Wu, C.T., et al. (2020). Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. Cell *182*, 226–244.e17.

7. Xu, J.Y., Zhang, C., Wang, X., Zhai, L., Ma, Y., Mao, Y., Qian, K., Sun, C., Liu, Z., Jiang, S., et al. (2020). Integrative proteomic characterization of human lung adenocarcinoma. Cell *182*, 245–261.e17.

8. Satpathy, S., Krug, K., Jean Beltran, P.M., Savage, S.R., Petralia, F., Kumar-Sinha, C., Dou, Y., Reva, B., Kane, M.H., Avanessian, S.C., et al. (2021). A proteogenomic portrait of lung squamous cell carcinoma. Cell *184*, 4348–4371.e40.

9. George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

10. Umemura, S., Mimaki, S., Makinoshima, H., Tada, S., Ishii, G., Ohmatsu, H., Niho, S., Yoh, K., Matsumoto, S., Takahashi, A., et al. (2014). Therapeutic priority of the PI3K/AKT/mTOR pathway in small cell lung cancers as revealed by a comprehensive genomic analysis. J. Thorac. Oncol. *9*, 1324–1331.

11. Sivakumar, S., Moore, J.A., Montesion, M., Sharaf, R., Lin, D.I., Colón, C.I., Fleishmann, Z., Ebot, E.M., Newberg, J.Y., Mills, J.M., et al. (2023). Integrative analysis of a large real-world cohort of small cell lung cancer identifies distinct genetic subtypes and insights into histologic transformation. Cancer Discov. *13*, 1572–1591.

12. Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V., Johnson, J.E., Lehman, J.M., MacPherson, D., et al. (2019). Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat. Rev. Cancer *19*, 289–297.

13. Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y., Fujimoto, J., Solis, L.M., Lu, W., et al. (2021). Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell *39*, 346–360.e7.

14. Rodriguez, H., Zenklusen, J.C., Staudt, L.M., Doroshow, J.H., and Lowy, D.R. (2021). The next horizon in precision oncology: proteogenomics to inform cancer diagnosis and treatment. Cell *184*, 1661–1670.

15. Park, S., Lee, H., Lee, B., Lee, S.H., Sun, J.M., Park, W.Y., Ahn, J.S., Ahn, M.J., and Park, K. (2019). DNA damage response and repair pathway alteration and its association with tumor mutation burden and platinum-based chemotherapy in SCLC. J. Thorac. Oncol. *14*, 1640–1650.

16. Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., et al. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat. Genet. *44*, 1111–1116.

17. Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., et al. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. *44*, 1104–1110.

18. Morris, L.G., Kaufman, A.M., Gong, Y., Ramaswami, D., Walsh, L.A., Turcan, Ş., Eng, S., Kannan, K., Zou, Y., Peng, L., et al. (2013). Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat. Genet. *45*, 253–261.

19. Pastushenko, I., Mauri, F., Song, Y., de Cock, F., Meeusen, B., Swedlund, B., Impens, F., Van Haver, D., Opitz, M., Thery, M., et al. (2021). Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. Nature *589*, 448–455.

20. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

21. Krug, K., Mertins, P., Zhang, B., Hornbeck, P., Raju, R., Ahmad, R., Szucs, M., Mundt, F., Forestier, D., Jane-Valbuena, J., et al. (2019). A curated resource for phosphosite-specific signature analysis. Mol. Cell. Proteomics *18*, 576–593.

22. Sen, T., Gay, C.M., and Byers, L.A. (2018). Targeting DNA damage repair in small cell lung cancer and the biomarker landscape. Transl. Lung Cancer Res. *7*, 50–68.

23. Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell *177*, 1035–1049.e19.

24. Nemeth, M.J., Kirby, M.R., and Bodine, D.M. (2006). Hmgb3 regulates the balance between hematopoietic stem cell self-renewal and differentiation. Proc. Natl. Acad. Sci. USA *103*, 13783–13788.

25. Mukherjee, A., Huynh, V., Gaines, K., Reh, W.A., and Vasquez, K.M. (2019). Targeting the high-mobility group box 3 protein sensitizes chemoresistant ovarian cancer cells to cisplatin. Cancer Res. *79*, 3185–3191.

26. Wen, B., Wei, Y.T., and Zhao, K. (2021). The role of high mobility group protein B3 (HMGB3) in tumor proliferation and drug resistance. Mol. Cell. Biochem. *476*, 1729–1739.

27. Sun, H., Cui, C., Xiao, F., Wang, H., Xu, J., Shi, X., Yang, Y., Zhang, Q., Zheng, X., Yang, X., et al. (2015). miR-486 regulates metastasis and chemosensitivity in hepatocellular carcinoma by targeting CLDN10 and CITRON. Hepatol. Res. *45*, 1312–1322.

28. He, X.L., Xiao, Q., Zhou, Z.P., and Hui, C.Y. (2021). PPM1D accelerates proliferation and metastasis of osteosarcoma by activating PKP2. Eur. Rev. Med. Pharmacol. Sci. *25*, 78–85.

29. Liu, Z., Sun, T., Piao, C., Zhang, Z., and Kong, C. (2022). METTL14-mediated N(6)-methyladenosine modification of ITGB4 mRNA inhibits metastasis of clear cell renal cell carcinoma. Cell Commun. Signal. *20*, 36.

30. Wei, F., Wu, Y., Tang, L., He, Y., Shi, L., Xiong, F., Gong, Z., Guo, C., Li, X., Liao, Q., et al. (2018). BPIFB1 (LPLUNC1) inhibits migration and invasion of nasopharyngeal carcinoma by interacting with VTN and VIM. Br. J. Cancer *118*, 233–247.

31. Liang, Y., Chen, X., Wu, Y., Li, J., Zhang, S., Wang, K., Guan, X., Yang, K., and Bai, Y. (2018). LncRNA CASC9 promotes esophageal squamous cell carcinoma metastasis through upregulating LAMC2 expression by interacting with the CREB-binding protein. Cell Death Differ. *25*, 1980–1995.

32. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

33. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

34. Dora, D., Rivard, C., Yu, H., Bunn, P., Suda, K., Ren, S., Lueke Pickard, S., Laszlo, V., Harko, T., Megyesfalvi, Z., et al. (2020). Neuroendocrine subtypes of small cell lung cancer differ in terms of immune microenvironment and checkpoint molecule distribution. Mol. Oncol. *14*, 1947–1965.

35. Cai, L., Liu, H., Huang, F., Fujimoto, J., Girard, L., Chen, J., Li, Y., Zhang, Y.A., Deb, D., Stastny, V., et al. (2021). Cell-autonomous immune gene expression is repressed in pulmonary neuroendocrine cells and small cell lung cancer. Commun. Biol. *4*, 314.

36. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

37. Chabanon, R.M., Rouanne, M., Lord, C.J., Soria, J.C., Pasero, P., and Postel-Vinay, S. (2021). Targeting the DNA damage response in immuno-oncology: developments and opportunities. Nat. Rev. Cancer *21*, 701–717.

38. Xiang, Z., Liu, H., and Hu, Y. (2023). DNA damage repair and cancer immunotherapy. Genome Instab. Dis. *4*, 210–226.

39. Anurag, M., Punturi, N., Hoog, J., Bainbridge, M.N., Ellis, M.J., and Haricharan, S. (2018). Comprehensive profiling of DNA repair defects in breast cancer identifies a novel class of endocrine therapy resistance drivers. Clin. Cancer Res. *24*, 4887–4899.

40. Baine, M.K., Hsieh, M.S., Lai, W.V., Egger, J.V., Jungbluth, A.A., Daneshbod, Y., Beras, A., Spencer, R., Lopardo, J., Bodd, F., et al. (2020). SCLC subtypes defined by ASCL1, NEUROD1, POU2F3, and YAP1: A comprehensive immunohistochemical and histopathologic characterization. J. Thorac. Oncol. *15*, 1823–1835.

41. Caeser, R., Egger, J.V., Chavan, S., Socci, N.D., Jones, C.B., Kombak, F.E., Asher, M., Roehrl, M.H., Shah, N.S., Allaj, V., et al. (2022). Genomic and transcriptomic analysis of a library of small cell lung cancer patient-derived xenografts. Nat. Commun. *13*, 2144.

42. Tlemsani, C., Pongor, L., Elloumi, F., Girard, L., Huffman, K.E., Roper, N., Varma, S., Luna, A., Rajapakse, V.N., Sebastian, R., et al. (2020). SCLC-CellMiner: A resource for small cell lung cancer cell line genomics and pharmacology based on genomic signatures. Cell Rep. *33*, 108296.

43. Allison Stewart, C., Tong, P., Cardnell, R.J., Sen, T., Li, L., Gay, C.M., Masrorpour, F., Fan, Y., Bara, R.O., Feng, Y., et al. (2017). Dynamic variations in epithelial-to-mesenchymal transition (EMT), ATM, and SLFN11 govern response to PARP inhibitors and cisplatin in small cell lung cancer. Oncotarget *8*, 28575–28587.

44. Cañadas, I., Rojo, F., Taus, Á., Arpí, O., Arumí-Uría, M., Pijuan, L., Menéndez, S., Zazo, S., Dómine, M., Salido, M., et al. (2014). Targeting epithelial-to-mesenchymal transition with Met inhibitors reverts chemoresistance in small cell lung cancer. Clin. Cancer Res. *20*, 938–950.

45. Dhanasekaran, R., Deutzmann, A., Mahauad-Fernandez, W.D., Hansen, A.S., Gouw, A.M., and Felsher, D.W. (2022). The MYC oncogene - the grand orchestrator of cancer growth and immune evasion. Nat. Rev. Clin. Oncol. *19*, 23–36.

46. Blake, D.R., Vaseva, A.V., Hodge, R.G., Kline, M.P., Gilbert, T.S.K., Tyagi, V., Huang, D., Whiten, G.C., Larson, J.E., Wang, X., et al. (2019). Application of a MYC degradation screen identifies sensitivity to CDK9 inhibitors in KRAS-mutant pancreatic cancer. Sci. Signal. *12*, eaav7259.

47. Farrington, C.C., Yuan, E., Mazhar, S., Izadmehr, S., Hurst, L., Allen-Petersen, B.L., Janghorban, M., Chung, E., Wolczanski, G., Galsky, M., et al. (2020). Protein phosphatase 2A activation as a therapeutic strategy for managing MYC-driven cancers. J. Biol. Chem. *295*, 757–770.

48. Dauch, D., Rudalska, R., Cossa, G., Nault, J.-C., Kang, T.-W., Wuestefeld, T., Hohmeyer, A., Imbeaud, S., Yevsa, T., and Hoenicke, L. (2016). A MYC–aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. Nat. Med. *22*, 744–753.

49. Wasylishen, A.R., Chan-Seng-Yue, M., Bros, C., Dingar, D., Tu, W.B., Kalkat, M., Chan, P.K., Mullen, P.J., Huang, L., Meyer, N., et al. (2013). MYC phosphorylation at novel regulatory regions suppresses transforming activity. Cancer Res. *73*, 6504–6515.

50. Chen, Y., Fang, Z., Tang, Y., Jin, Y., Guo, C., Hu, L., Xu, Y., Ma, X., Gao, J., Xie, M., et al. (2021). Integrative analysis of multi-omics data reveals the heterogeneity and signatures of immune therapy for small cell lung cancer. Clin. Transl. Med. *11*, e620.

51. Thomas, A., Takahashi, N., Rajapakse, V.N., Zhang, X., Sun, Y., Ceribelli, M., Wilson, K.M., Zhang, Y., Beck, E., Sciuto, L., et al. (2021). Therapeutic targeting of ATR yields durable regressions in small cell lung cancers with high replication stress. Cancer Cell *39*, 566–579.e7.

52. Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., Ogle, C., Reinhart, R., Selby, M., Connelly, J., et al. (2016). Small cell lung cancer screen of oncology drugs, investigational agents, and gene and microRNA expression. J. Natl. Cancer Inst. *108*, djw122.

53. Rudin, C.M., Reck, M., Johnson, M.L., Blackhall, F., Hann, C.L., Yang, J.C., Bailis, J.M., Bebb, G., Goldrick, A., Umejiego, J., and Paz-Ares, L. (2023). Emerging therapies targeting the delta-like ligand 3 (DLL3) in small cell lung cancer. J. Hematol. Oncol. *16*, 66.

54. Rudin, C.M., Pietanza, M.C., Bauer, T.M., Ready, N., Morgensztern, D., Glisson, B.S., Byers, L.A., Johnson, M.L., Burris, H.A., 3rd, Robert, F., et al. (2017). Rovalpituzumab tesirine, a DLL3-targeted antibody-drug conjugate, in recurrent small-cell lung cancer: a first-in-human, first-in-class, open-label, phase 1 study. Lancet Oncol. *18*, 42–51.

55. Paz-Ares, L., Champiat, S., Lai, W.V., Izumi, H., Govindan, R., Boyer, M., Hummel, H.D., Borghaei, H., Johnson, M.L., Steeghs, N., et al. (2023). Tarlatamab, a first-in-class DLL3-targeted bispecific T-cell engager, in recurrent small-cell lung cancer: an open-label, phase I study. J. Clin. Oncol. *41*, 2893–2903.

56. Cheng, Y., Wang, Q., Li, K., Shi, J., Liu, Y., Wu, L., Han, B., Chen, G., He, J., Wang, J., et al. (2021). Anlotinib vs placebo as third- or further-line treatment for patients with small cell lung cancer: a randomised, double-blind, placebo-controlled Phase 2 study. Br. J. Cancer *125*, 366–371.

57. Cardnell, R.J., Li, L., Sen, T., Bara, R., Tong, P., Fujimoto, J., Ireland, A.S., Guthrie, M.R., Bheddah, S., Banerjee, U., et al. (2017). Protein expression of TTF1 and cMYC define distinct molecular subgroups of small cell lung cancer with unique vulnerabilities to aurora kinase inhibition, DLL3 targeting, and other targeted therapies. Oncotarget *8*, 73419–73432.

58. Mollaoglu, G., Guthrie, M.R., Böhm, S., Brägelmann, J., Can, I., Ballieu, P.M., Marx, A., George, J., Heinen, C., Chalishazar, M.D., et al. (2017). MYC drives progression of small cell lung cancer to a variant neuroendocrine subtype with vulnerability to Aurora kinase inhibition. Cancer Cell *31*, 270–285.

59. Guo, C., Wan, R., He, Y., Lin, S.H., Cao, J., Qiu, Y., Zhang, T., Zhao, Q., Niu, Y., Jin, Y., et al. (2022). Therapeutic targeting of the mevalonate-geranylgeranyl diphosphate pathway with statins overcomes chemotherapy resistance in small cell lung cancer. Nat. Cancer *3*, 614–628.

60. Thomas, A., Desai, P., and Takahashi, N. (2022). Translational research: A patient-centered approach to bridge the valley of death. Cancer Cell *40*, 565–568.

61. Zhang, J., Zhang, L., Luo, J., Ge, T., Fan, P., Sun, L., Hou, L., Li, J., Yu, H., Wu, C., et al. (2021). Comprehensive genomic profiling of combined small cell lung cancer. Transl. Lung Cancer Res. *10*, 636–650.

62. Yokouchi, H., Nishihara, H., Harada, T., Yamazaki, S., Kikuchi, H., Oizumi, S., Uramoto, H., Tanaka, F., Harada, M., Akie, K., et al. (2020). Detection of somatic TP53 mutation in surgically resected small-cell lung cancer by targeted exome sequencing: association with longer relapse-free survival. Heliyon *6*, e04439.

63. Udagawa, H., Umemura, S., Murakami, I., Mimaki, S., Makinoshima, H., Ishii, G., Miyoshi, T., Kirita, K., Matsumoto, S., Yoh, K., et al. (2018). Genetic profiling-based prognostic prediction of patients with advanced small-cell lung cancer in large scale analysis. Lung Cancer *126*, 182–188.

64. Mohr, A., Deedigan, L., Jencz, S., Mehrabadi, Y., Houlden, L., Albarenque, S.M., and Zwacka, R.M. (2018). Caspase-10: a molecular switch from cell-autonomous apoptosis to communal cell death in response to chemotherapeutic drug treatment. Cell Death Differ. *25*, 340–352.

65. Kumari, R., Deshmukh, R.S., and Das, S. (2019). Caspase-10 inhibits ATP-citrate lyase-mediated metabolic and epigenetic reprogramming to suppress tumorigenesis. Nat. Commun. *10*, 4255.

66. Li, M., Cai, Y., Zhao, H., Xu, Z., Sun, Q., Luo, M., Gu, L., Meng, M., Han, X., and Sun, H. (2015). Overexpression of HMGB3 protein promotes cell proliferation, migration and is associated with poor prognosis in urinary bladder cancer patients. Tumour Biol. *36*, 4785–4792.

67. Tang, H.R., Luo, X.Q., Xu, G., Wang, Y., Feng, Z.J., Xu, H., Shi, Y.W., Zhang, Q., Wu, L.G., Xue, C.Q., et al. (2012). High mobility group-box 3 overexpression is associated with poor prognosis of resected gastric adenocarcinoma. World J. Gastroenterol. *18*, 7319–7326.

68. Song, N., Wang, B., Feng, G., Duan, L., Yuan, S., Jia, W., and Liu, Y. (2019). Knockdown of high mobility group box 3 impairs cell viability and colony formation but increases apoptosis in A549 human non-small cell lung cancer cells. Oncol. Lett. *17*, 2937–2945.

69. Gu, M., Jiang, Z., Li, H., Peng, J., Chen, X., and Tang, M. (2021). MiR-93/HMGB3 regulatory axis exerts tumor suppressive effects in colorectal carcinoma cells. Exp. Mol. Pathol. *120*, 104635.

70. Liu, J., Wang, L., and Li, X. (2018). HMGB3 promotes the proliferation and metastasis of glioblastoma and is negatively regulated by miR-200b-3p and miR-200c-3p. Cell Biochem. Funct. *36*, 357–365.

71. Poirier, J.T., George, J., Owonikoko, T.K., Berns, A., Brambilla, E., Byers, L.A., Carbone, D., Chen, H.J., Christensen, C.L., Dive, C., et al. (2020). New approaches to SCLC therapy: from the laboratory to the clinic. J. Thorac. Oncol. *15*, 520–540.

72. Iams, W.T., Porter, J., and Horn, L. (2020). Immunotherapeutic approaches for small-cell lung cancer. Nat. Rev. Clin. Oncol. *17*, 300–312.

73. Zhang, J., Zhou, N., Lin, A., Luo, P., Chen, X., Deng, H., Kang, S., Guo, L., Zhu, W., and Zhang, J. (2021). ZFHX3 mutation as a protective biomarker for immune checkpoint blockade in non-small cell lung cancer. Cancer Immunol. Immunother. *70*, 137–151.

74. Sen, T., Rodriguez, B.L., Chen, L., Corte, C.M.D., Morikawa, N., Fujimoto, J., Cristea, S., Nguyen, T., Diao, L., Li, L., et al. (2019). Targeting DNA damage response promotes antitumor immunity through STING-mediated T-cell activation in small cell lung cancer. Cancer Discov. *9*, 646–661.

75. Taniguchi, H., Caeser, R., Chavan, S.S., Zhan, Y.A., Chow, A., Manoj, P., Uddin, F., Kitai, H., Qu, R., Hayatt, O., et al. (2022). WEE1 inhibition enhances the antitumor immune response to PD-L1 blockade by the concomitant activation of STING and STAT1 pathways in SCLC. Cell Rep. *39*, 110814.

76. Tendler, S., and Rudin, C.M. (2023). Tarlatamab: new star on the horizon for small-cell lung cancer? J. Clin. Oncol. *41*, 2877–2880.

77. Ko, J., Winslow, M.M., and Sage, J. (2021). Mechanisms of small cell lung cancer metastasis. EMBO Mol. Med. *13*, e13122.

78. Satpathy, S., Jaehnig, E.J., Krug, K., Kim, B.J., Saltzman, A.B., Chan, D.W., Holloway, K.R., Anurag, M., Huang, C., Singh, P., et al. (2020). Microscaled proteogenomic methods for precision oncology. Nat. Commun. *11*, 532.

79. Hunt, A.L., Bateman, N.W., Barakat, W., Makohon-Moore, S., Hood, B.L., Conrads, K.A., Zhou, M., Calvert, V., Pierobon, M., Loffredo, J., et al. (2021). Extensive three-dimensional intratumor proteomic heterogeneity revealed by multiregion sampling in high-grade serous ovarian tumor specimens. iScience *24*, 102757.

80. Chan, J.M., Quintanal-Villalonga, Á., Gao, V.R., Xie, Y., Allaj, V., Chaudhary, O., Masilionis, I., Egger, J., Chow, A., Walle, T., et al. (2021). Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. Cancer Cell *39*, 1479–1496.e18.

81. Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjatic, S., et al. (2009). CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res. *37*, D816–D819.

82. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

83. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

84. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

85. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

86. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

87. Kim, J., Mouw, K.W., Polak, P., Braunstein, L.Z., Kamburov, A., Kwiatkowski, D.J., Rosenberg, J.E., Van Allen, E.M., D'Andrea, A., and Getz, G. (2016). Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat. Genet. *48*, 600–606.

88. Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316.

89. Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. *8*, 33.

90. Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Ruiter, J., Lolkema, M.P., et al. (2015). CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol. *16*, 49.

91. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.

92. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

93. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

94. Yoo, S., Huang, T., Campbell, J.D., Lee, E., Tu, Z., Geraci, M.W., Powell, C.A., Schadt, E.E., Spira, A., and Zhu, J. (2014). MODMatcher: multi-omics data matcher for integrative genomic analysis. PLoS Comput. Biol. *10*, e1003790.

95. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

96. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205.

97. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

98. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

99. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

100. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

101. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics *16*, 284–287.

102. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest package: tests in linear mixed effects models. J. Stat. Software *82*, 1–26.

103. Kochura, Y., Stirenko, S., Alienin, O., Novotarskiy, M., and Gordienko, Y. (2018). Performance Analysis of Open Source Machine Learning Frameworks for Various Parameters in Single-Threaded and Multi-Threaded Modes, N. Shakhovska and V. Stepashko, eds. (Springer International Publishing), pp. 243–256.

104. Dong, L., Lu, D., Chen, R., Lin, Y., Zhu, H., Zhang, Z., Cai, S., Cui, P., Song, G., Rao, D., et al. (2022). Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma. Cancer Cell *40*, 70–87.e15.

105. Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., et al. (2019). Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. Cell *179*, 561–577.e22.

106. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

107. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

108. Liu, Z., Li, M., Jiang, Z., and Wang, X. (2018). A comprehensive immunologic portrait of triple-negative breast cancer. Transl. Oncol. *11*, 311–329.

109. Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. USA *101*, 4164–4169.

110. Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

111. Tan, V.Y., and Févotte, C. (2013). Automatic relevance determination in nonnegative matrix factorization with the β-divergence. IEEE Trans. Pattern Anal. Mach. Intell. *35*, 1592–1605.

112. Kasar, S., Kim, J., Improgo, R., Tiao, G., Polak, P., Haradhvala, N., Lawrence, M.S., Kiezun, A., Fernandes, S.M., Bahl, S., et al. (2015). Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat. Commun. *6*, 8866.

113. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. *47*, D941–D947.

114. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering signatures of mutational processes operative in human cancer. Cell Rep. *3*, 246–259.

115. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756.

116. Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

117. Staedtler, F., Hartmann, N., Letzkus, M., Bongiovanni, S., Scherer, A., Marc, P., Johnson, K.J., and Schumacher, M.M. (2013). Robust and tissue-independent gender-specific transcript biomarkers. Biomarkers *18*, 436–445.

118. Wang, J., Ma, Z., Carr, S.A., Mertins, P., Zhang, H., Zhang, Z., Chan, D.W., Ellis, M.J., Townsend, R.R., Smith, R.D., et al. (2017). Proteome profiling outperforms transcriptome profiling for coexpression based gene function prediction. Mol. Cell. Proteomics *16*, 121–134.

119. Bailey, M.H., Tokheim, C., Porta-Pardo, E., SenGupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. Cell *173*, 371–385.e18.

120. Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

121. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

122. Cao, L., Huang, C., Cui Zhou, D., Hu, Y., Lih, T.M., Savage, S.R., Krug, K., Clark, D.J., Schnaubelt, M., Chen, L., et al. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell *184*, 5031–5052.e26.

123. Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S., et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome atlas. Cell Rep. *23*, 239–254.e6.

124. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

125. Huang, C., Chen, L., Savage, S.R., Eguez, R.V., Dou, Y., Li, Y., da Veiga Leprevost, F., Jaehnig, E.J., Lei, J.T., Wen, B., et al. (2021). Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell *39*, 361–379.e16.

126. Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al. (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. Cell *183*, 1436–1456.e31.

127. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway KnowledgeBase. Nucleic Acids Res. *48*, D498–D503.

128. Zhang, W., Girard, L., Zhang, Y.A., Haruki, T., Papari-Zareei, M., Stastny, V., Ghayee, H.K., Pacak, K., Oliver, T.G., Minna, J.D., and Gazdar, A.F. (2018). Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. Transl. Lung Cancer Res. *7*, 32–49.

129. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.e15.

130. Daily, K., Ho Sui, S.J., Schriml, L.M., Dexheimer, P.J., Salomonis, N., Schroll, R., Bush, S., Keddache, M., Mayhew, C., Lotia, S., et al. (2017). Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. Sci. Data *4*, 170030.

131. Salomonis, N., Dexheimer, P.J., Omberg, L., Schroll, R., Bush, S., Huo, J., Schriml, L., Ho Sui, S., Keddache, M., Mayhew, C., et al. (2016). Integrated genomic analysis of diverse induced pluripotent stem cells from the progenitor cell biology consortium. Stem Cell Rep. *7*, 110–125.

132. Sokolov, A., Paull, E.O., and Stuart, J.M. (2016). One-class detection of cell states in tumor subtypes. Pac. Symp. Biocomput. *21*, 405–416.

133. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71.

134. Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X., Noushmehr, H., Colaprico, A., and Papaleo, E. (2019). New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLOS Comput. Biol. *15*, e1006701.

135. Ellis, M.J., Suman, V.J., Hoog, J., Goncalves, R., Sanati, S., Creighton, C.J., DeSchryver, K., Crouch, E., Brink, A., Watson, M., et al. (2017). Ki67 proliferation index as a tool for chemotherapy decisions during and after neoadjuvant aromatase inhibitor treatment of breast cancer: results from the American College of Surgeons Oncology Group Z1031 trial (alliance). J. Clin. Oncol. *35*, 1061–1069.

136. Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., et al. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell *13*, 1977–2000.

137. Candel, A., Parmar, V., LeDell, E., and Arora, A. (2016). Deep learning with H_2O. http://h2o.ai/resources.

138. Yoshida, A., Tsuta, K., Wakai, S., Arai, Y., Asamura, H., Shibata, T., Furuta, K., Kohno, T., and Kushima, R. (2014). Immunohistochemical detection of ROS1 is useful for identifying ROS1 rearrangements in lung cancers. Mod. Pathol. *27*, 711–720.

139. Boehm, J.S., Hession, M.T., Bulmer, S.E., and Hahn, W.C. (2005). Transformation of human and murine fibroblasts without viral oncoproteins. Mol. Cell. Biol. *25*, 6464–6474.

140. Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol. Cell. Proteomics *10*, M110.003699.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-HA	Proteintech	Cat# 51064-2-AP; RRID: AB_11042321
Rabbit polyclonal anti-Flag	Sigma	Cat# F7425; RRID: AB_439687
Mouse monoclonal anti-Flag	Sigma	Cat# F3165; RRID: AB_259529
Rabbit polyclonal anti-HMGB3	ABclonal	Cat# A15064; RRID: AB_2761946
Mouse monoclonal anti-Tubulin	Santa Cruz Biotechnology	Cat# SC23948; RRID: AB_628410
Mouse monoclonal anti-MYC	Santa Cruz Biotechnology	Cat# SC-40; RRID: AB_627268
Rabbit monoclonal anti-HMGB3	Abcam	Cat# ab75782; RRID: AB_1310317
Rabbit monoclonal anti-CASP10	Abcam	Cat# ab32155; RRID: AB_2069188
Anti-Rabbit Immunoglobulins/HRP	Dako	Cat# P0217; RRID: AB_2728719
Anti-Mouse Immunoglobulins/HRP	Dako	Cat# P0260; RRID: AB_2636929
Biological samples		
Paired tumor, normal adjacent lung tissues from a cohort of 112 SCLC patients	Shanghai Pulmonary Hospital, Tongji University	This paper
FFPE of 78 paired tumor and normal adjacent lung tissues	Shanghai Pulmonary Hospital, Tongji University	This paper
FFPE of an independent cohort of 111 SCLC patients	Shanghai Pulmonary Hospital, Tongji University	This paper
Biopsy samples of 12 SCLC patients from ongoing phase II trials (NCT04539977 and NCT04542369)	Shanghai Pulmonary Hospital, Tongji University	This paper
Patient-derived xenografts (PDXs)	Guo et al.[59]	https://www.nature.com/articles/s43018-022-00358-1
Chemicals, peptides, and recombinant proteins		
LC-MS grade water	Fisher chemical	Cat# W6-4
Urea	Sigma	Cat# U1250
Sodium dodecyl sulfate	Sigma	Cat# L4509
Formic acid	Tokyo Chemical Industry	Cat# F0654
Tris	VWR Life Science	Cat# 0497
Dithiothretiol	Sigma	Cat# D9163
Iodoacetamide	Sigma	Cat# I1149
LC-MS grade acetonitrile	Fisher chemical	Cat# A955-4
Sequencing-grade modified trypsin	Promega	Cat# V5111
Trifluoroacetic acid	Sigma	Cat# 302031
50% Hydroxylamine	Thermo Fisher Scientific	Cat# 90115
LC-MS grade methanol	Merck	Cat# 1.06007
Ammonium formate	Fluka	Cat# 70221
Triethylammonium bicarbonate	Sigma	Cat# T7408
Ammonium hydroxide	Sigma	Cat# 30501
Reversed-phase tC18 SepPak	Waters	Cat# WAT036790
ReproSil-Pur, 120 Å, C18-AQ, 1.9-μm resin	Dr. Maisch	Cat# r119.aq.
Polyethylenimine	Sigma	Cat# 408727
Polybrene	Sigma	Cat# TR-1003
Puromycin	Beyotime Biotechnology	Cat# ST551
Lipo3000 Transfection Reagent	Thermo Fisher Scientific	Cat# L3000001
Anlotinib	Selleck	Cat# S8726

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Alisertib	Selleck	Cat# S1133
Barasertib	Selleck	Cat# S1147
AMG-900	Selleck	Cat# S2719
Critical commercial assays		
QIAamp Fast DNA tissue kit	QIAGEN	Cat# 51404
SureSelect Human All ExonV6 Kit	Agilent Technologies	Cat# 5190-8865
RNAlater Reagent	Invitrogen	Cat# AM7020
TRIzol Reagent	Invitrogen	Cat# 15596026
NEBNext® UltraTM RNA Library Prep Kit	NEB	Cat# E7530L
BCA Protein Assay Kit	Beyotime Biotechnology	Cat# P0012S
TMT 10-plex Isobaric Label Reagent	Thermo Fisher Scientific	Cat# 90111
TMT 11-131C Label Reagent	Thermo Fisher Scientific	Cat# A34807
High-Select Fe-NTA kit	Thermo Fisher Scientific	Cat# A32992
truChIP® Chromatin Shearing Kit	Covaris	Cat# 520127
EZ-ChIP™ Chromatin Immunoprecipitation Kit	Millipore	Cat# 17-371
NEBNext® Ultra™ DNA Library Prep Kit	NEB	Cat# E7645
Deposited data		
Proteomics data of TU-SCLC cohort	This paper	OMIX database: OMIX002489; https://ngdc.cncb.ac.cn/omix
WES and RNA-seq data of TU-SCLC cohort	This paper	GSA database: HRA003419; http://bigd.big.ac.cn/gsa-human
GENCODE (version 22)	N/A	https://www.gencodegenes.org
GRCh38/hg38	NCBI	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
SwissProt human protein database (version 2018.08, 20,431 entries)	Uniprot	https://www.uniprot.org/
PhosphoSitePlus	Hornbeck et al.[20]	https://www.phosphosite.org
CT Antigen database	Almeida et al.[81]	http://www.cta.lncc.br
Human Protein Atlas	Uhlén et al.[82]	https://www.proteinatlas.org
MSigDB v7.1 gene sets	Liberzon et al.[83]	https://www.gsea-msigdb.org
PTMsigDB version 1.9.0	Krug et al.[21]	https://github.com/broadinstitute/ssGSEA2.0/tree/master/db/ptmsigdb
Mutational Signatures (version 3.2)	COSMIC	https://cancer.sanger.ac.uk/signatures/
COSMIC onco-driver and suppressor gene data	COSMIC	https://cancer.sanger.ac.uk/census
HGNC database	HGNC	https://www.genenames.org
Experimental models: Cell lines		
Human: 293T	Prof. Daming Gao (Shanghai Institute of Biochemistry and Cell Biology)	N/A
Human: H146, male origin	ATCC	Cat# HTB-173
Human: DMS114, male origin	ATCC	Cat# CRL-2066
Human: H345, male origin	ATCC	Cat# HTB-180
Human: H69, male origin	ATCC	Cat# HTB-119
Human: H82, male origin	Dr. Zhe Liu (Tianjin Medical University)	N/A
Human: H446, male origin	Dr. Lei Huang (Shanghai Jiao Tong University)	N/A
Human: H526, male origin	Dr. Lei Huang (Shanghai Jiao Tong University)	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
BALB/c nude mouse, male	BK Company, Shanghai	N/A
C.B-17 SCID mouse, male	SLAC Company, Shanghai	N/A
Oligonucleotides		
siRNA targeting sequence: siSTMN1-1 sense: 5'- UGGAGGAAAUUCAGAAGAA-3' siSTMN1-2 sense: 5'- AGGUGAAAGAACUGGAGAA-3'	This paper	N/A
siRNA targeting sequence: siTMA7-1 sense: 5'-GAAGAAACUCGAGGAGCUA-3' siTMA7-2 sense: 5'-GAAUUAAGAAAUCUGGCAA-3'	This paper	N/A
siRNA targeting sequence: siHMGB3-1 sense: 5'-AGAAGAAGAAGGAUCCUAA-3' siHMGB3-2 sense: 5'-GACUAUAAGUCGAAAGGAA-3'	This paper	N/A
siRNA targeting sequence: siITGB4-1 sense: 5'-GGAAAGAGCUGCAGGUGAA-3' siITGB4-2 sense: 5'-CCACAGAGCUGGUGCCCUA-3'	This paper	N/A
siRNA targeting sequence: siCLDN10-1 sense: 5'-CGAUAAAGCCAAAGCUAAA-3' siCLDN10-2 sense: 5'-GUAUCAUGGUGGAGAAGAU-3'	This paper	N/A
siRNA targeting sequence: siVTN-1 sense: 5-GCUAUGAACUGGACGAAAA-3' siVTN-2 sense: 5-GCUUCAACGUGGACAAGAA-3'	This paper	N/A
siRNA targeting sequence: siPKP2-1 sense: 5-CCCAGAAGUCCGUGGAAGA-3' siPKP2-2 sense: 5-CCAGAGACUUGGAGACUAA-3'	This paper	N/A
siRNA targeting sequence: siLAMC2-1 sense: 5'-CGAAAUGGGUCUCCUGCAA-3' siLAMC2-2 sense: 5'-CGGAGGAGGUGGUGUGCAA-3'	This paper	N/A
Primers for GAPDH, HMGB3, ITGB4, CLDN10, VTN, PKP2, LAMC2; see method details section "Real-time quantitative PCR"	This paper	N/A
Recombinant DNA		
pLEX-MCS-CMV-puro	Thermo Scientific Open Biosystems	Cat# OHS4735
Software and algorithms		
MaxQuant (version 1.6.5.0)	Tyanova et al.[84]	http://www.maxquant.org
GATK (version 4.0.6.0)	Broad Institute	https://software.broadinstitute.org/gatk/
BWA (version 0.7.15)	Li and Durbin[85]	http://bio-bwa.sourceforge.net/
Picard (version 2.9.0)	GitHub	http://broadinstitute.github.io/picard/
MutSigCV (version 1.4)	Lawrence et al.[86]	https://www.genepattern.org/modules/docs/MutSigCV
SignatureAnalyzer	Kim et al.[87]	https://github.com/getzlab/SignatureAnalyzer
OptiType (version 1.2.1)	Szolek et al.[88]	https://github.com/FRED-2/OptiType
NetMHCpan (version 3.0)	Nielsen and Andreatta[89]	http://www.cbs.dtu.dk/services/NetMHCpan-3.0/
CopywriteR (version 2.6.1.2)	Kuilman et al.[90]	https://github.com/PeeperLab/CopywriteR
GISTIC2.0 (version 2.0.23)	Mermel et al.[91]	https://github.com/broadinstitute/gistic2
STAR (version 2.7.6a)	Dobin et al.[92]	https://github.com/alexdobin/STAR
Salmon (version 1.3.0)	Patro et al.[93]	https://combine-lab.github.io/salmon/
OmicsEV	Bing Zhang Lab	https://github.com/bzhanglab/OmicsEV
MODMatcher (version 0.1.0)	Yoo et al.[94]	https://github.com/integrativenetworkbiology/Modmatcher

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
diptest R Package (version 0.75-7)	N/A	https://cran.r-project.org/web/packages/diptest/index.html
multiOmicsViz R package (version 1.20.0)	N/A	http://www.bioconductor.org/packages/release/bioc/html/multiOmicsViz.html
ESTIMATE (version 1.0.13)	Yoshihara et al.[32]	https://bioinformatics.mdanderson.org/public-software/estimate/
xCell	Aran et al.[33]	https://xcell.ucsf.edu/
NMF R-package (version 0.23.0)	Gaujoux and Seoighe[95]	https://cran.r-project.org/web/packages/NMF/index.html
PTM-SEA	Krug et al.[21]	https://github.com/broadinstitute/ssGSEA2.0
WebGestalt	Liao et al.[96]	http://www.webgestalt.org/
GSVA R package (version 1.42.0)	Hänzelmann et al.[97]	https://bioconductor.org/packages/release/bioc/html/GSVA.html
GSEA software (version 4.1.0)	Subramanian et al.[98]	https://www.gsea-msigdb.org/gsea/index.jsp
DreamAI	Pei Wang Lab	https://github.com/WangLab-MSSM/DreamAI
factoextra R package (version 1.0.7)	N/A	https://rpkgs.datanovia.com/factoextra/index.html
Bowtie2 (version 2.4.2)	Langmead and Salzberg[99]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MACS2 (version 2.2.7)	Zhang et al.[100]	https://pypi.org/project/MACS2/
Homer (version 4.11.1)	N/A	http://homer.ucsd.edu/homer/ngs/annotation.html
clusterProfiler R package (version 4.2.0)	Yu et al.[101]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
lmerTest R package (version 3.1-3)	Kuznetsova et al.[102]	https://cran.r-project.org/web/packages/lmerTest/index.html
h2o R package (version 3.36.0.3)	Kochura et al.[103]	https://www.h2o.ai

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and regents should be directed to and will be fulfilled by the lead contact, Peng Zhang (zhangpeng1121@tongji.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The WES and RNA-seq raw data have been deposited in the Genome Sequence Archive (GSA) database in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA database: HRA003419, http://bigd.big.ac.cn/gsa-human). The proteomic data have been deposited in the OMIX database, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (OMIX database: OMIX002489, https://ngdc.cncb.ac.cn/omix).

Sample annotation, processed, and normalized data files were provided in Table S1. Software and code used in this study are referenced in their corresponding STAR Methods sections and also the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical Specimens

All SCLC tumor and normal adjacent tissue (NAT) samples for this study were prospectively collected from Shanghai Pulmonary Hospital (Tongji University, Shanghai, China) from April 2012 to June 2019. This study was approved by the Institutional Review Board of Shanghai Pulmonary Hospital and written informed consents of all participating patients were obtained. The clinicopathological

parameters were collected and summarized in Table S1, including gender, age, smoking status, tumor site, histologic type, tumor size, date of surgical resection, TNM staging (AJCC cancer staging system 8th edition) and VALSG (the Veterans Administration Lung Study Group) staging and so on. This cohort included 98 males and 14 females with the median age at diagnosis of 62 (range 38-81). In total, 35, 30 and 47 patients were classified as TNM stages I, II and III-IV, respectively. Most patients had been diagnosed with limited-stage, except one with extensive-stage. Histologically, 91% (n = 102) of the tumors were pure SCLC, and 9% (n = 10) were combined SCLC, which had an additional component of non-small cell lung cancer. Overall survival (OS) was defined as the interval between surgery and death.

Cell Lines

Human SCLC cells (H345, H69, H146 and DMS114) were purchased from American Type Culture Collection (ATCC). H82 cells were generously provided by Dr. Zhe Liu at Tianjin Medical University; H446 and H526 cells were generously provided by Dr. Lei Huang at Shanghai Jiao Tong University. Human embryonic kidney cells (HEK293T) were kept by Prof. Daming Gao's lab. HEK293T cells were cultured in DMEM with 10% fetal bovine serum (FBS), penicillin and streptomycin. All the human SCLC cells were cultured in RPMI 1640 medium with 10% FBS, penicillin and streptomycin.

Animals

Six-week-old female BALB/c nude mice were purchased from BK Company (Shanghai, China) and used for establishment of human SCLC cell line xenografts. Six-week-old female C.B-17 SCID mice were purchased from SLAC Company (Shanghai, China) and used for establishment of SCLC PDX models. All animal experiments were conducted following a protocol approved by the Institutional Animal Care and Use Committee of the Shanghai Institute of Biological Sciences, Chinese Academy of Sciences.

Mice were maintained in specific pathogen-free facilities and housed in single-sex cages at 23 ± 3 °C, 40-70% humidity under a 12-h light/12-h dark photoperiod with the lights on at 7:00 a.m. The length and width of the tumors (in millimeters) were measured every other day with calipers. Tumor volume was calculated using the formula $(A \times B^2)/2$, where A and B were the long and short dimensions, respectively. Mice were blindly randomized into different groups for treatment studies.

METHOD DETAILS

Specimen Acquisition and Preparation

Following Clinical Proteomic Tumor Analysis Consortium (CPTAC) guideline, all the tissues were treatment-naïve and surgically resected. No prior anticancer treatments, including chemotherapy, radiotherapy or immunotherapy therapy were exposed previously. Paired histologically-NATs were collected from the same patient at tumor resection. Tissues were obtained and stored in liquid nitrogen less than 30 min after resection. For accurate tumor cellularity analysis, hematoxylin and eosin (H&E) staining of the middle section of each tissue was done. Histologic sections from each case were reviewed by two board-certified pathologists to confirm the assigned pathology.

The proteogenomic workflow of our SCLC samples was shown in Figure 1A. SCLC tumor specimens and NATs were cryo-pulverized using the CryoPrepTM CP02 (Covaris), and then divided into three parts. The first part was snap-frozen in liquid nitrogen and stored in -80°C for DNA extraction and whole exome sequencing (WES); the second part was stored in 1 mL RNAlater Reagent (Invitrogen) at -80°C for RNA extraction and RNA sequencing (RNA-seq); the remaining samples were treated with SDS lysis buffer (4% SDS, 0.1 M Tris-HCl, pH 7.6) and kept in -80°C for the following proteomic and phosphoproteomic analyses. According to Clinical Proteomic Tumor Analysis Consortium (CPTAC) clinical sample collection procedures as reported previously,[104,105] 112 high-quality tumors and paired NATs were chosen for WES, proteomic and phosphoproteomic analyses; among which 107 paired samples with qualified RNA were used for RNA-seq.

Genomics and Transcriptomics Profiling Experiments

DNA extraction and whole exome sequencing (WES)

Genomic DNA was extracted from tumors and NATs using QIAamp Fast DNA tissue kit (QIAGEN) according to the manufacturer's protocol. Total DNA was quantified by the Qubit 2.0 Flurometer (Life Technologies) and NanoDrop 2000 (Thermo Fisher Scientific) and the integrity was assessed by TapeStation (Agilent Technologies). For Illumina sequencing library construction, the genomic DNA was fragmented to an average size of 180-280 bp using a Covaris focused-ultrasonicator. Then, WES libraries were prepared and captured using the Agilent SureSelect Human All Exon V6 kit (Agilent Technologies) following the manufacturer's instructions. The DNA library with 150 bp paired-end reads was sequenced on the Illumina Novaseq 6000 platform.

RNA extraction and RNA sequencing

Total RNA was extracted and purified from fresh frozen tissues using the TRIzol reagent (Invitrogen). All RNA analytes were assayed for RNA concentration and purity using Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies) and NanoPhotometer® spectrophotometer (IMPLEN). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies). Samples with RNA integrity number (RIN) > 6.0 were considered high quality and used to prepare the transcriptome library. Total RNA-seq library construction was performed from the RNA samples using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB) following manufacturer's recommendations and index codes were added to attribute sequences to each

sample. Quality control was performed at every step, and the libraries were quantified using the Agilent Bioanalyzer 2100 system (Agilent Technologies). Then, paired-end libraries were sequenced on the Illumina Novaseq 6000 platform and 150 bp paired-end reads were generated. Totally, 107 paired tumor and NAT samples passed the RNA quality control and were used for RNA-seq.

Proteomics and Phosphoproteomics Profiling Experiments

Protein extraction and tryptic digestion

Approximately 25-100 mg of each cryo-pulverized SCLC tumor or NAT sample was resuspended in 500 μL lysis buffer (4% SDS, 0.1 M Tris-HCl, pH 7.6) and sonicated at 20% amplitude for 2 min (5 s on, 5 s off) using Ultrasonic Homogenizer (JY92-IIDN, NingBo Scientz Biotechnology). The proteins were then denatured and reduced at 95°C for 5 min. Lysates were centrifuged at 12,000 g for 10 min to remove the insoluble debris and protein concentrations of the clarified lysates were measured by BCA assay (Beyotime Biotechnology).

Protein lysates were reduced with 0.1 M dithiothreitol (Sigma) for 30 min at 56°C and then an equal amount of proteins for each sample were digested by filter-aided sample preparation (FASP) procedure as previously reported.[104–106] Briefly, proteins were mixed with 200 μL UA buffer (8 M urea in 0.1 M Tris-HCl, pH 8.5) in 10 kDa centrifugal filter tubes (Pall Corporation) and centrifuged at 12,000 g at 22°C. All the following centrifugation steps were performed applying the same conditions, washed twice with 200 μL UA buffer, alkylated with 50 mM iodoacetamide (Sigma) in 200 μL UA buffer for 30 min in the dark at room temperature, washed thrice with 100 μL UA buffer again and finally washed thrice with 100 μL 50 mM triethyl ammonium bicarbonate (TEAB, Sigma). Subsequently, proteins were subjected to proteolytic digestion with sequencing grade modified trypsin (Promega) at 1:50 enzyme-to-substrate ratio in 50 mM TEAB for 18 hours at 37°C. The digests were collected by centrifugation, and the filter device was rinsed with 50 μL 50 mM TEAB and centrifuged again. The concentration of tryptic peptides was determined by BCA assay and 400 μg peptides for each sample were dried using Speed-Vac apparatus (Thermo Fisher Scientific).

Preparation of reference tissue samples

The proteomic and phosphoproteomic analyses of SCLC samples were structured as TMT 11-plex experiments. To facilitate quantitative comparison between all samples in each TMT batch experiment, an 'internal reference' mixed sample was included in each TMT set. Therefore, 40 pairs of tumors and NATs were randomly selected and mixed in equal protein amount, yielding an internal reference represented in this study. The resulting pooled reference material was divided into 1.2 mg aliquots and was also digested by FASP procedure as described above.

TMT 11-plex labeling of peptides

112 tumor samples with paired NATs were distributed among 23 sets of TMT 11-plex experiments, with 10 individual samples occupying the first 10 channels of each set and the 11th channel being reserved for the internal reference sample. For each TMT experiment, five tumor samples were labeled with 126, 127C, 128C, 129C and 130C; paired NAT samples were labeled with 127N, 128N, 129N, 130N and 131N; and the mixed samples were labeled with channel 131C as internal reference.

For each TMT labeling experiment, dried peptides (400 μg) from each sample were dissolved in 200 μL 100 mM TEAB and two sets of TMT reagents (0.8 mg) dissolved in 82 μL (41 μL × 2) anhydrous acetonitrile were added. After 1-hour incubation at room temperature, 16 μL 5% hydroxylamine was added to quench the labeling reaction for 15 min at room temperature. The labeled peptides were pooled, dried down via Speed-Vac, and subsequently desalted on a reversed phase tC18 SepPak column (Waters). The same lot of TMT reagents were used for all samples (TC261829 for TMT 10-plex and TF266762 for TMT 11-131C).

Peptides fractionation by high-pH reversed-phase liquid chromatography

To deduce sample complexity and increase the depth of protein identification, high-pH reverse phase liquid chromatography (RPLC) was used for peptides fractionation. For each TMT set, about 4.4 mg TMT 11-plex labeled peptides were fractionated using a 4.6 mm × 250 mm Waters XBridge BEH300 C18 column with 3.5 μm size beads (Waters). Peptides were separated using an Agilent 1100 HPLC instrument via high-pH reversed-phase liquid chromatography with solvent A (10 mM ammonium formate, pH 10) and a non-liner increasing concentration of solvent B (90% ACN, 10 mM ammonium formate, pH 10) at a flow rate of 0.7 mL/min. The 110-min separation gradient was set as follows: 1%-5% B in 2 min; 5%-25% B in 35 min; 25%-40% B in 43 min; 40%-55% B in 6 min; 55%-95% B in 3 min; 95% B for 4 min; 95%-1% B in 1 min; 1% B for 16 min. Peptides were separated and collected every minute for a total of 96 fractions from 3 min to 99 min, and were subsequently combined into 24 fractions by a stepwise concatenation strategy. 5% of each of the 24 fractions was allocated and dried down via Speed-Vac for global proteome analysis. The remaining 95% sample was then utilized for phosphopeptides enrichment as described below.

Phosphopeptides enrichment by Fe-IMAC

High-Select Fe-NTA kit (Thermo Fisher Scientific) was used for phosphopeptides enrichment according to the manufacturer's instructions. In brief, fractionated peptides were reconstituted in 200 μL 80% ACN/0.1% trifluoroacetic acid (TFA) and incubated with 50 μL Fe^{3+}-NTA agarose beads for 20 min at room temperature. Then, the mixture was transferred into the filter tip (Axygen, TF-200-L-R-S) and clarified peptide flow-throughs with unbound peptides were collected by centrifugation. After successive washes with 200 μL 80% ACN/0.1% TFA three times and 200 μL H_2O three times, the bound phosphopeptides were eluted twice with 200 μL 50% ACN/5% $NH_3 \cdot H_2O$ and dried down via Speed-Vac. All centrifugation steps above were conducted at 50 g at room temperature.

Benchmark sample preparation

Longitudinal quality control of mass spectrometry performance was tested by periodic analysis of full process replicates of a benchmark sample. The benchmark sample was prepared from five pairs of hepatocellular carcinoma tissues and their adjacent non-tumor

liver tissues as previously reported.[105] After protein extraction, tryptic digestion, TMT 10-plex labeling (tumor tissues labeled with 126, 127C, 128C, 129C and 130C; adjacent non-tumor liver tissues labeled with 127N, 128N, 129N, 130N and 131) and peptides fractionation as described above, 1 μg and diluted samples of 100 ng from each fraction were analyzed before every three sets of SCLC proteomic and phosphoproteomic samples on a Q Exactive HF mass spectrometer, respectively.

LC-MS/MS analysis

For proteomic analysis, the fractionated peptides were re-suspended in 2%ACN/0.1% formic acid (FA) and about 1 μg of each fraction was separated using an in-house packed 20 cm × 75 μm internal diameter C18 column (1.9 μm ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH, Germany) on a nanoflow Easy nLC 1200 UHPLC system (Thermo Fisher Scientific). The column was heated to 50°C using a home-made column heater. The flow rate was set at 300 nL/min. Buffer A and B were 0.1% FA in H_2O and 0.1% FA in 80% acetonitrile, respectively. The 120-min separation gradient was set as follows: 2%-5% B in 1 min; 5%-35% B in 93 min; 35%-50% B in 15 min; 50%-100% B in 3 min; 100% B in 8 min. Samples were analyzed with a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) equipped with a nanoflow ionization source. Data-dependent acquisition was performed using Xcalibur software (version 4.0.27.19) in positive ion mode at a spray voltage of 2,300 V. The MS1 spectra was measured with a resolution of 120,000 @ m/z 200, an AGC target of 3e6, a maximum injection time (IT) of 50 ms and a mass range of 350 to 1,700 m/z. The data-dependent mode cycle was set to trigger MS2 scan on up to the top 20 most abundant precursors per cycle at an MS2 resolution of 60,000 @ m/z 200, an AGC target of 1e5, a maximum injection time of 120 ms, an isolation window of 1.0 m/z, an HCD (high collision dissociation) collision energy of 32, and a fixed first mass of 105.0 m/z. The dynamic exclusion time was set as 40 s and precursor ions with charge 1, 7, 8 and > 8 were excluded for MS2 analysis. The 24 benchmark fractions were analyzed using a 90-min gradient as follows: 2%-6% B in 1 min; 6%-32% B in 64 min; 32%-50% B in 13 min; 50%-70% B in 3 min; 70%-100% B in 1 min and 100% B in 8 min. The parameters of MS were set the same as SCLC proteomic samples.

For phosphoproteomic analysis, the enriched phosphopeptides were re-suspended in 2%ACN/0.1% formic acid and half of each fraction was analyzed using the same instrumentation and MS parameters as the global SCLC proteomic analysis, except for a 70-min LC gradient (2%-5% B in 1 min; 5%-32% B in 48 min; 32%-45% B in 10 min; 45%-100% B in 3 min and 100% B in 8 min). And the 24 benchmark fractions for monitoring SCLC phosphoproteomic analysis were analyzed using a 70-min gradient as follows: 5%-8% B in 4 min; 8%-35% B in 46 min; 35%-50% B in 10 min; 50%-100% B in 2 min and 100% B in 8 min. The parameters of MS were set the same as described above.

Genomic and Transcriptomic Data Processing

Somatic variant calling

To detect single nucleotide variant (SNV) and small insertion/deletion (INDEL), the Genome Analysis Toolkit (GATK, version 4.0.6.0) best practice guideline was followed. After excluding low-quality reads, qualified paired-end WES sequencing reads were aligned to human reference genome (hg38) with BWA MEM (version 0.7.15).[85] The resulting BAM files were further processed with Picard tools (version 2.9.0, http://broadinstitute.github.io/picard/) to remove PCR duplicates. Subsequently, recalibration and INDEL realignment was implemented using GATK modules IndelRealigner and BaseRecalibrator. Cross-sample contamination was assessed with GATK module CalculateContamination tool with a 5% threshold. SNVs and INDELs were detected using MuTect2 tools[107] embedded in GATK from SCLC tumor and paired non-tumor samples, then filtered out short tandem repeat region downloaded from UCSC table browser, and finally annotated using Funcotator from GATK.

Analysis of significantly mutated genes

Significance of mutated genes were evaluated by MutSigCV (version 1.4)[86] in GenePattern, and mutated genes were deemed significant above the background mutation rate if q value was less than 0.05.

Mutation frequency in TU-SCLC and previous studies

Mutation frequencies for three previous SCLC studies were obtained from George et al.,[9] Sivakumar et al.,[11] and Umemura et al.[10] cohort. The frequencies of mutated genes in Figure 1B were compared from TU-SCLC cohort using Fisher's exact test.

Tumor mutational burden (TMB)

For each patient, we calculated the TMB score as previously described.[108] Briefly, TMB = total number of truncating mutations × 1.5 + total number of non-truncating mutations × 1.0. Truncating mutations included nonsense, frame-shift deletion, frame-shift insertion, and splice-site, while non-truncating mutations included missense, in-frame deletion, in-frame insertion, and nonstop. Silent mutations were excluded from these analyses since they did not result in an amino acid change. Truncating mutations were given a higher weight considering their higher deleterious effects on gene function than non-truncating mutations. Based on the median value of TMB scores, we classified the patients into TMB-high and TMB-low group.

Correlation of somatic mutation and clinical parameters

To evaluate the correlation between somatic mutation genes and clinical parameters, we first separated SCLC patients into younger (≤ 62 years, median age in TU-SCLC cohort) and older group or TMB-high and TMB-low group. Then Fisher's exact test was employed to determine whether the mutation was enriched in a certain group as shown in Figure S1J.

Mutational signature analysis

Mutational signatures in 112 SCLC tumors were explored by non-negative matrix factorization (NMF) approach.[109,110] The 96 mutational contexts generated by somatic SNVs based on six base substitutions (C > A, C > T, C > G, T > A, T > G, and T > C) were referred as input data to infer their contributions to detected mutations. SignatureAnalyzer[87] was utilized to evaluate the Bayesian variant of

the NMF approach and discern the optimal number of signature profiles.[87,111,112] Then, each profile of signatures was compared with mutation patterns of validated cancer signatures reported from COSMIC database,[113] and Cosine similarity analysis[114,115] was used to find the best match. The value of Cosine similarity ranged from 0 to 1, which indicated maximal dissimilarity to maximal similarity.

Neoantigen prediction

To predict neoantigen, OptiType (version 1.2.1)[88] was used to identify the human leukocyte antigen (HLA) class I genotype (HLA-A, HLA-B, HLA-C) in each specimen. NetMHCpan (version 3.0)[89] was applied to identify MHC ligands and predict the binding affinity of peptides. Finally, peptides with robust binding affinity ($IC_{50} \leq 500$ nM) were referred as predicted neoantigen, while those with weak binding affinity (binding affinity parameters > 1,000 nM) were excluded.

Somatic copy-number alteration (SCNA) analysis

To estimate SCNAs in each tumor, paired BAM files from SCLC tumor and paired non-tumor samples were processed by the R package CopywriteR (version 2.6.1.2).[90] With default parameters, circular binary segmentation (CBS) algorithm[116] implemented in the CopywriteR package was used for copy number segmentation. Segment-level ratios were calculated and $\log_2$ transformed.

Then, Genomic Identification of Significant Targets in Cancer (GISTIC2.0, version 2.0.23) algorithm[91] was used to identify significantly amplified and deleted arm-level and focal-level SCNA events, with q value less than 0.25 considered significant. A $\log_2$ ratio cut-off of $\pm$ 0.8 was used to define SCNA amplification and deletion. Default value was used for other parameters.

Chromosomal instability index (CIN)

The CIN score reflects the overall copy-number alterations across the whole genome. From the SCNA segment results, we used a weighted-sum approach to summarize the chromosome instability for each sample as previously described.[23] Specifically, the absolute $\log_2$ ratios of all segments within a chromosome were weighted by the segment length and summed to infer the instability score for the chromosome. Then, the genome-wide chromosome instability index was calculated by summing the instability score of all 22 autosomes.

RNA quantification

After removal of adaptors and low-quality reads, RNA-seq cleaned sequence data were aligned to human reference sequence hg38 using STAR (version 2.7.6a).[92] The resulting BAM files were analyzed for gene expression using Salmon (version 1.3.0)[93] against a transcriptomic reference GENCODE (version 22, https://www.gencodegenes.org). Then, the resulting transcripts per million (TPM) values of protein-coding genes were $\log_2$ transformed and data rows of redundant gene symbols were aggregated by calculating the average expression values.

Proteomics and Phosphoproteomics Data Processing

Database searching of MS data

All MS raw files were searched against the human Swiss-Prot database containing 20,431 sequences (downloaded in August, 2018) using MaxQuant (version 1.6.5.0).[84] TMT 11-plex (SCLC tumor and NAT samples) and TMT 10-plex (benchmark samples and PDX/CDX samples) based MS2 reporter ion quantification was chosen with reporter mass tolerance set as 0.003 Da. The purities of TMT labeling channels were corrected according to the kit LOT number. The PIF (precursor intensity fraction) filter value was set at 0.5. Enzyme digestion specificity was set to Trypsin and maximum two missed cleavages were allowed. Carbamidomethyl cysteine was set as fixed modification. Oxidized methionine, protein N-term acetylation, lysine acetylation, asparagine and glutamine (NQ) deamidation were set as variable modifications. For phosphorylation data analysis, phospho (STY) was also chosen as a variable modification. The tolerances of first search and main search for peptides were set at 20 ppm and 4.5 ppm, respectively. A cut-off of 1% FDR was applied at the peptide, protein, and site level. A minimum of 7 amino acids was required for peptide identification. For phosphosite localization, the localization probability > 0.75 was considered as confident phosphosite.

Data normalization

Global protein and phosphosite abundance from 224 samples were measured in 23 sets of TMT 11-plex labeling experiment, which identified 11,209 proteins at gene-level (unique peptides $\geq$ 2) and 62,881 confident phophosites (localization probability > 0.75). In each set of TMT experiments, relative protein or phosphosite abundance was calculated as the ratio of sample abundance to internal reference sample using the reporter ion intensities. Then the relative abundances were $\log_2$ transformed and normalized using median centering method, resulting these proteins and phosphosites having a log TMT ratio value centered at zero. Among the data, 7,010 proteins and 7,388 phosphosites were quantified across all the samples. DreamAI ensemble algorithm (https://github.com/WangLab-MSSM/DreamAI) was applied to impute missing values by DreamAI R package. Only those proteins and phosphosites with a missing rate less than 50% were imputed. After imputation, 9,559 proteins and 26,979 phosphosites were used for downstream analyses.

Data Quality Control

Sample labeling checking

Sample labeling check is a critical data quality control step before integrated multi-omics analysis. In this study, we checked tissue annotations (tumor or NAT), gender annotations (male or female), and sample matching among RNA-seq, proteomics and phosphoproteomics data. We performed principal component analysis (PCA) using factoextra R package (version 1.0.7) independently on the RNA-seq, global proteome and phosphoproteome data respectively. As expected, tumor and NAT samples were clearly separated (Figure S3A), suggesting that the tissue annotation was consistent with given information. Then, gender-specific marker gene

expression (XIST, DDX3Y, RPS4Y1)[117] was used to evaluated the gender annotation. Sample matching was assessed by a pairwise alignment procedure as previously described[94] by comparing RNA-seq and proteomics, proteomics and phosphoproteomics, and RNA-seq and phosphoproteomics data. This analysis was performed using MODMatcher R package (version 0.1.0) (https://github.com/integrativenetworkbiology/Modmatcher).

Batch effect assessment and proteomics quality control

During the LC-MS/MS analysis of global proteomics and phosphoproteomics, we established a quality control procedure by periodic analysis of full process replicates of a benchmark sample to assess the instrument performance. The benchmark sample was prepared from five pairs of hepatocellular carcinoma tissues and their adjacent non-tumor liver tissues as previously reported.[105] We assessed the Pearson correlation of $\log_2$ (TMT126/TMT127N) ratios across the longitudinal benchmark samples and observed high reproducibility across the TMT plexes measurement (Figure S1C). In addition, we calculated the numbers of peptides, proteins and phosphosites identified in each TMT set (Figure S1A). PCA for both proteomics and phosphoproteomics data showed no obvious batch effect across the 23 TMT batches (Figure S1B). The density plots of $\log_2$ TMT ratios for proteins and phosphosites identified in each sample showed that all samples conformed to an expected unimodal distribution evaluated by dip statistic test ($P > 0.05$, diptest R package, version 0.75-7) (Figure S1D). Finally, protein complex correlation analysis and co-expression network-based gene function prediction for KEGG pathways were performed using RNA-seq and proteomics data as previously described[118] by OmicsEV (https://github.com/bzhanglab/OmicsEV) (Figures S1E and S1F).

Integrated Analysis

mRNA-protein correlation analysis

A total of 9,428 genes or proteins with less than 50% missing values were used to measure gene-wise and sample-wise mRNA and protein correlations (Figures S1G and S1H). Spearman's correlation coefficient and corresponding *P* value were calculated for each mRNA-protein pair across tumors and NATs separately and for each individual sample by *cor.test* function in R (Table S1). Further, adjusted *P* value was calculated using the Benjamini-Hochberg (BH) procedure and a cut-off of 0.01 was determined as statistical significance. In order to explore the biological functions enriched for genes with low and high protein-RNA correlations, Spearman's correlation coefficient was used as the ranking metric for gene set enrichment analysis (GSEA) using the GSEA software (version 4.1.0).[98] Molecular Signatures Database (MSigDB, c2.cp.kegg.v7.1.symbols.gmt)[83] was used for enrichment analysis.

Defining cancer-associated genes

Cancer-associated genes (CAGs) were compiled from genes defined by Bailey et al.[119] and cancer-associated genes listed in Mertins et al.[120] and adapted from Vogelstein et al.[121] as previously described.[5] A total of 593 genes were defined as CAGs.

SCNA-driven cis and trans effects

The *cis* and *trans* regulation of SCNAs on mRNA, protein and phosphoprotein expression were determined using Spearman's correlation and visualized by multiOmicsViz R package. *P* values were calculated to assess the statistical significance of the correlation values and corrected for multiple testing using Benjamini-Hochberg procedure. As a result, FDR < 0.05 was considered as statistically significant positive or negative correlations for CNA-mRNA, CNA-protein and CNA-phosphoprotein correlations. In addition, 261 genes which had significant *cis*-effects at mRNA, protein and phosphoprotein levels were filtered to keep the ones with significant differential protein expression between tumors and NATs (BH adjusted *P* value < 0.01, Wilcoxon signed-rank test). This resulted in a number of 223 SCNAs (Table S2) and were further used to perform GO biological processes enrichment analysis. To identify genes with most significant trans-effects at protein level, three filter criteria were used as follows: 1) belongs to CAGs; 2) amplification or deletion frequency > 10% in the TU-SCLC cohort; 3) the number of significant protein-level *trans* events of the SCNA ranked for top 10 (Spearman's correlation, adjusted *P* value < 0.05) (Table S2).

Mutation effects on RNA, proteome and phosphoproteome

We examined the effects of somatic mutation genes as shown in Figure 1B on the RNA, protein and phosphosite levels. Briefly, samples were classified into mutated and wild-type (WT) groups, and Wilcoxon ranked-sum test was used to identify differentially expressed RNA, proteins and phosphosites. For *FAT1* mutation, signed -$\log_{10}$ *P* value was used as the pre-ranked list for GSEA analysis using WebGestalt[96] (http://www.webgestalt.org/).

Tumor-NAT differential expression analysis

Differential expression analysis was performed for tumors and paired NATs using the Wilcoxon signed-rank test. *P* values were adjusted using the Benjamini-Hochberg method. Each feature was required to be non-missing values. Proteins or phosphosites (collapsed into gene-level) differentially expressed between tumors and NATs (BH adjusted *P* value < 0.01, $\log_2$FC > 1 or < -1) were further used for over-representation analysis by WebGestalt.[96]

A linear mixed model implanted in R package lmerTest[102] was used to correct for stromal and immune content for 107 patients with an ESTIMATE score for both tumors and NATs as previously described.[122] Specifically, the sample type and z-scored ESTIMATE score were used as fixed effects and the patient was used as random effect. Benjamini-Hochberg adjusted *P* value < 0.01 was considered as significant. Furthermore, we filtered proteins with at least 2-fold upregulated in more than 90% tumor-NAT pairs, and defined as SCLC-associated proteins. Then, proteins of specific type or function including cancer related genes, enzymes, plasma proteins, drug targets (FDA approved drug targets and potential drug targets) were annotated by the Human Protein Atlas[82] (HPA, https://www.proteinatlas.org).

Kinase activity prediction

Activated kinases were filtered using two methods. Firstly, we performed phosphosite-specific signature enrichment analysis (PTM-SEA)[21] to infer kinase activity from the phosphorylation of its substrates using the PTM signatures database (PTMsigDB, version 1.9.0) (https://github.com/broadinstitute/ssGSEA2.0/tree/master/db/ptmsigdb). Phosphosites with less than 50% missing values were applied to Wilcoxon signed-rank test between tumors and NATs. *P* values were then log-transformed and signed according to the fold change. The resulting signed -$\log_{10}$ *P* value for each site was used as the ranking metric for PTM-SEA. The parameters were used as follows: statistic – area.under.RES; output.score.type – NES; correl.type – rank; weight – 1; nperm – 1,000; min.overlap – 5. *P* value for each signature was derived from 1,000 random permutations and further adjusted by BH method. Signature of kinases with adjusted *P* value < 0.05 and NES score > 0 were considered to show increased activity in SCLC tumors. In addition, activated kinases were inferred from significantly increased phosphorylation of their sites annotated as "enzyme activity induced" from PhosphoSitePlus database[20] (https://www.phosphosite.org). Kinases with a median phosphorylation fold change greater than the median protein fold change were considered as significantly activated.

Cancer/testis antigen prediction

Cancer-testis (CT) antigens were downloaded from CTdatabase (http://www.cta.lncc.br).[81] The CT antigens that overlapped with SCLC proteomics data were selected for further analysis and those overexpressed at least 2-fold in tumors compared with NATs in more than 10% of samples were displayed in Figure S3I.

Prognostic biomarker analysis for SCLC

To identify potential protein prognostic biomarkers, four filter criteria were used as follows (also shown in Figure S4A): 1) the proteins should be quantified in all samples; 2) the correlation coefficient between mRNA and protein expression should be > 0.7; 3) the candidate proteins should be differentially expressed between tumors and NATs with adjusted *P* value < 0.01 (Wilcoxon signed-rank test, BH adjusted) and fold change > 2 at both mRNA and protein levels; 4) we stratified the TU-SCLC patients into two groups based on the median expression of each protein. Kaplan-Meier curve with log rank test was used to visualize the survival difference and Cox proportional hazard model was used to evaluate the hazard ratio (HR) for each protein. The candidate proteins should significantly correlate with the overall survival (log rank *P* value < 0.01, and HR (high/low) > 2 for upregulated or < 0.5 for downregulated proteins).

Immune subtype identification and downstream analysis

The abundances of 64 different cell types in 107 SCLC tumors and paired NAT samples were computed via xCell[33] (https://xcell.ucsf.edu/) using the RNA-seq data (Table S5). Then unsupervised clustering was implemented on these cell signatures using the NMF R package (version 0.23.0). We chose $k = 3$ based on the maximal cophenetic correlation coefficients using 50 iterations, and subsequently repeated the NMF analysis using 200 iterations. Furthermore, ESTIMATE algorithm[32] was used to derive the overall immune score and stromal score for each sample based on RNA-seq data. Protein-based immune signatures for MHC proteins, stimulatory and inhibitory immune modulators were calculated as the mean value of the protein expression for each gene set. Immune cytolytic activity (CYT) scores were obtained by calculating the geometric mean of two key cytolytic effectors, granzyme A (GZMA) and perforin (PRF1) mRNA expression (TPM) per sample as previously reported.[36]

Immune scores correlation analysis

To identify the potential drivers of immunosuppression, ESTIMATE immune scores were correlated with proteomics data using Spearman's correlation analysis. Then, WebGestalt[96] was used to perform GSEA for KEGG pathways using the signed -$\log_{10}$ *P* values. In order to further evaluate the correlation between immune scores and specific DNA repair pathways, we used a DDR gene set containing unique proteins from a specific pathway as previously reported.[39] Damage sensor scores were calculated based on the protein expression of 9 DNA damage response genes.[123] Gene set based scores were the mean protein expression of all genes in that set (Figure S5E).

Inferred DNA repair, replication stress, STING pathway, EMT, and E2F activity scores

All scores were inferred by single sample gene set enrichment analysis (ssGSEA)[124] method from GSVA package.[97] The Hallmark DNA_repair, replication_stress, epithelial_mesenchymal_transition, E2F_targets gene sets were used to calculate the DNA repair scores, replication stress scores, EMT scores and E2F activity scores using the proteomics data, respectively. STING pathway activity scores were inferred from Reactome 'STING mediated induction of host immune response' gene set based on mRNA expression.

Unsupervised multi-omics clustering

As previously described,[5,8,125,126] non-negative matrix factorization (NMF)[95] was used to perform unsupervised clustering of tumor samples based on mRNA, protein, and phosphosite expression data. Briefly, we required all features quantified in more than 50% samples and the remaining missing values were imputed as described above. All data tables were then concatenated and features with the lowest standard deviation (bottom 5th percentile) across all samples were removed for subsequent analysis. Each row in the data matrix was further scaled and standardized such that all features from different data types were represented as z-scores. Since NMF requires a non-negative input matrix, we converted the z-scores in the data matrix into a non-negative matrix as follows: 1) create one data matrix with all negative numbers zeroed; 2) create another data matrix with all positive numbers zeroed and the signs of all negative numbers removed; 3) concatenate both matrices resulting in a data matrix twice as large as the original, but containing only positive values and zeros and hence appropriate for NMF.

The resulting matrix was then subjected to NMF analysis via the NMF R-package (version 0.23.0) and using the 'brunet' factorization method.[109] A range of clusters between $k = 2$ - 8 were tested to determine the optimal factorization rank *k* (number of clusters). 50

iterations were set for each *k*. We calculated the cophenetic correlation coefficients and chose the *k* with maximal cophenetic correlation (Figure S6A). After determining the optimal factorization rank *k*, the NMF analysis was repeated using 1,000 iterations to achieve stability and a robust result.

Comparison of the multi-omics subtyping with previous transcriptional subtyping

Transcriptional subtype signature genes were obtained from previously reported NMF-derived gene list.[13] NMF analysis was performed to define transcriptional subtypes based on the mRNA expression (TPM values) for overlapped genes in TU-SCLC cohort, and comparison results were evaluated with our multi-omics subtypes by Fisher's exact test.

Subtype signature protein and phosphosite identifications

For each of the multi-omics subtype, the subtype signature protein and phosphosite were defined as those significantly more abundant in that subtype compared with all the other three subtypes (Student's t test). The resulting subtype signature protein and phosphosite (collapsed into gene-level) were used to perform over-representation analysis to identify enriched GO biological processes, KEGG and Reactome[127] pathway analysis via WebGestalt.[96]

Multi-omics single-sample gene set enrichment analysis

To further analyze biological characteristics of multi-omics subtypes, we performed ssGSEA[124] to identify the pathway alterations among the four SCLC subtypes. To achieve pathway enrichment scores for each sample, R package GSVA[97] and Hallmark pathway gene sets (MSigDB database v.7.1) were used with a least 10 overlapping genes. A Wilcoxon ranked-sum test was applied subsequently to define pathways differentially expressed between one subtype to others. *P* values were adjusted via the Benjamini-Hochberg procedure.

Neuroendocrine (NE) score inference

SCLC neuroendocrine (NE) scores were calculated based on a gene set of 25 neuroendocrine and 25 non-neuroendocrine genes established by a reported study.[128] For each sample, the NE score was calculated with the following formula: NE score = (correl NE - correl non-NE)/2, where correl NE or non-NE was the Pearson correlation between expression of the 50 genes in our SCLC samples and the expression of these genes in the NE or non-NE cell line group from Zhang et al.[128] The resulting NE score ranged from -1 to +1, which indicates non-NE phenotype to NE phenotype. RNA-seq and global proteomics data-based NE scores were highly correlated (Pearson correlation coefficient higher than 0.96).

Stemness score inference

We used stemness score to assess the degree of oncogenic dedifferentiation as previously described.[129] Specifically, the pluripotent stem cell samples (ESC and iPSC) were downloaded from the Progenitor Cell Biology Consortium (PCBC) dataset[130,131] and preprocessed. Then, one-class logistic regression (OCLR) machine-learning algorithm[132] was used to build a predictive model on the PCBC dataset. mRNA expression values for 107 SCLC tumor samples and 107 NAT samples were applied to calculate the stemness index. The function of TCGAanalyze Stemness deriving from TCGAbiolinks R package[133] was used and previous published workflow[134] was followed, with ''stemSig'' argument set as PCBC_stemSig.

Proliferation index inference

Proliferation index was defined as multi-gene proliferation scores (MGPS) and was calculated as described previously.[135] Specifically, MGPS were the mean expression of gene-wise z-scores for mRNA expression for all cell cycle-regulated genes identified by Whitfield et al. in each sample.[136]

Assessment of therapy response

To predict the therapy response of ATR inhibitor (ATRi) combined with TOP1 inhibitor (TOP1i), ATRi and TOP1i response score was inferred based on a gene set of 225 differentially expressed genes between responding and non-responding tumors from an SCLC clinical trial as previously described.[51] The score was calculated by the median expression of upregulated genes – the median expression of downregulated genes using the RNA-seq data.

Development and application of a signature-based classifier for multi-omics subtype

Subtype signature genes, proteins and phosphosites were defined as described above. Then, we performed stratified random sampling for subtype with a ratio of 7:3. We divided our SCLC dataset into a training dataset of 74 samples and a test dataset of 33 samples. Deep learning models could extract useful information from original data to a large extent, and show high performance in processing complex data. Here we used the R platform to develop an H2O-based deep learning model.[137] To construct a fully connected neural network, five hidden layers, in which 100, 200, 100, 200 and 100 nodes were allocated respectively. When training the model, a Tanh function was used as an activation function for the neural network, and a training procedure was repeated during 500 epochs. Using the expression information of the identified signature list, a fully connected neural network prediction classifier was constructed using the H2O[103] (https://www.h2o.ai) deep learning platform (version 3.36.0.3).

Survival analysis

We used Kaplan-Meier analysis to explore survival differences associated with TMB status, mutation signature dominant groups, SCNA status, immune subtypes, and somatic mutation status. For continuous variables, including mRNA, protein and phosphosite expression, maximally selected rank statistics (maxstat) implemented in survimer R package (version 0.4.9) was performed to determine the optimal cutpoint for selected samples. For protein biomarkers selection and validation, samples were stratified based on the median protein abundance by MS data or immunohistochemistry scores (H-scores). Kaplan-Meier plots, log rank tests and Cox proportional hazards regression for statistical significance were implemented via the survminer (version 0.4.9) and survival (version 3.2-13) packages in R. Patients died for operation-associated complication were excluded for survival analysis.

Immunohistochemistry Analysis

Apart from the use for the multi-omics analysis, the rest SCLC tumor and NAT samples were applied to construct tissue microarray (TMA). Briefly, paired SCLC tumor and NAT samples were paraffin-embedded, and then inspected by H&E staining. Representative areas were marked on the paraffin-embedded tissues and 2.0-mm-diameter cylinder was used for the construction of TMA. Finally, 78 patients with complete clinical prognosis information were applied for survival analysis. Immunohistochemistry (IHC) analysis was applied in TMA to validate the expression of MYC, HMGB3 and CASP10 expression. After deparaffinization, slides were prepared for antigen retrieval and blockade of endogenous peroxidase. Then, slides were stained with MYC antibody (1:50, Santa Cruz Biotechnology, SC-40), HMGB3 antibody (1:200, Abcam, ab75782) and CASP10 antibody (1:50, Abcam, ab32155), respectively. Subsequently, slides were incubated with secondary antibody and finally covered by cover slips. Immunostained slides were scored using the H-score method as previously described,[138] H-score = intensity of positive staining × percentage of positively-stained cells. H-scores range from 0 to 300 and the median value was termed as the cutoff to classify the groups. In addition, to validate the prognostic value of HMGB3 and CASP10, another independent cohort consisting of 111 SCLC tumors were collected from Shanghai Pulmonary Hospital (Tongji University, Shanghai, China) from December 2014 to July 2019 and were used for IHC assays in the same way. These patients received surgical resection and received no prior anticancer treatments.

Functional Experiments

Plasmid construct

3×Flag-HMGB3 plasmid was constructed by cloning coding sequence of human HMGB3 into the lentiviral vector pLEX-MCS-CMV-puro.

Lentivirus production and infection

Lentiviral packaging and subsequent generation of stable cell lines by infection were done as previously described.[139] Briefly, HEK293T cells were co-transfected with lentiviral vectors and the packaging plasmids psPAX2 and pMD2.G using polyethylenimine (Sigma). Lentivirus-containing medium was collected after 48 hours and supplemented with 20 mg/mL Polybrene (Sigma). H345 cells were infected by mixing the lentivirus-containing medium and the cell culture medium for 6 hours. Virally infected cells were selected by puromycin (Beyotime Biotechnology) for 96 hours.

RNA interference

The siRNAs were synthesized by Biotend Company. All siRNA transfections were performed with Lipo3000 Transfection Reagent (Thermo Fisher Scientific) at 50 nM final concentration according to the manufacturer's protocol. The siRNA transfected cells were harvested for western blot or qPCR assays 48 hours after transfection. Oligonucleotide sequences are shown in the key resources table.

Western blot analysis

Cells were harvested with EBC lysis buffer (50 mM Tris HCl, pH 8.0, 120 mM NaCl, 0.5% Nonidet P-40) supplemented with protease inhibitors (Selleck) and phosphatase inhibitors (Selleck). 30 μg of total proteins were separated by SDS-PAGE gel and blotted with primary antibodies. The details about primary antibodies were listed below with vendor and catalog numbers: anti-HMGB3 (1:3000; ABclonal, A15064), anti-Tubulin (1:10000; Santa Cruz Biotechnology, SC23948), anti-FLAG (1:3000; Sigma, F7425). Peroxidase-labeled anti-mouse (1:5000; DAKO, P0217) or anti-rabbit (1:5000; DAKO, P0260) IgG secondary antibody was used. The western blot gel image was obtained with an Minichemi 610 chemiluminescent imager (Sagecreation, Beijing, China).

Real-time quantitative PCR

Total RNA was extracted from cells using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. Total RNA was reversely transcribed into first-strand cDNA using the PrimeScript™RT reagent Kit with gDNA Eraser (Takara). The cDNAs were then used for real-time PCR (qPCR) on a Roche LightCycler 96 Real-Time quantitative PCR System (Roche) using TB Green® Premix Ex Taq™ II (Takara). GAPDH was served as an internal control. The relative quantification of gene expression was analyzed by the $2^{-\Delta\Delta Ct}$ method. The primers used for qPCR analyses are as follows:

GAPDH Forward: 5'-GGAGCGAGATCCCTCCAAAAT-3', Reverse: 5'-GGCTGTTGTCATACTTCTCATGG-3'
STMN1 Forward: 5'-TCAGCCCTCGGTCAAAAGAAT-3', Reverse: 5'-TTCTCGTGCTCTCGTTTCTCA-3'
TMA7 Forward: 5'-AAGGTGGCAAGAAGAAGCCA-3', Reverse: 5'-CTTCCCCGCGGCCTTC-3'
HMGB3 Forward: 5'-CCCAGAGGTCCCTGTCAATTT-3', Reverse: 5'-CGATCATAGCGCACTTTATCTGC-3'
ITGB4 Forward: 5'-GCAGCTTCCAAATCACAGAGG-3', Reverse: 5'-CCAGATCATCGGACATGGAGTT-3'
CLDN10 Forward: 5'-GCATGTAGAGGACTTATGATCGC-3', Reverse: 5'-TCCGACTTTGGTACACTTCATTC-3'
VTN Forward: 5'-TGACCAAGAGTCATGCAAGGG-3', Reverse: 5'-ACTCAGCCGTATAGTCTGTGC-3'
PKP2 Forward: 5'-GTGGGCAACGGAAATCTTCAC-3', Reverse: 5'-CCAGCCTTTAGCATGTCATAGG-3'
LAMC2 Forward: 5'-GACAAACTGGTAATGGATTCCGC-3', Reverse: 5'-TTCTCTGTGCCGGTAAAAGCC-3'

Cell proliferation Assay

For cell proliferation assay, H345 cells (1×10^3 cells) were seeded in 96-well plates. CCK-8 solution (Beyotime Biotechnology, C0039) at the final concentration of 10% was added to the wells, and absorbance at 450 nm was measured 2 hours after incubation to represent the relative cell numbers.

Transwell migration assay

For H345 cell migration, 5 × 10^4 cells were plated onto Transwell filters with 8-mm pores, a 24-well plate chamber insert (Corning). The top of the insert was supplemented with serum-free medium, while the bottom was supplemented with RPMI 1640 with 10% FBS. Cells were incubated for 24 hours and washed 2 times with PBS before fixed with methanol for 30 min. Then cells were stained with crystal violet for 30 min and washed 2 times with PBS. Cells at the top of the insert were scraped with a cotton swab while the positively stained cells at the bottom of the insert were photographed and examined under the microscope. Five visual fields were randomly chosen to calculate the number of migrated cells, each assay was repeated at least three times independently.

Chromatin immunoprecipitation sequencing (ChIP–seq)

For ChIP-seq, 4 × 10^7 H345 cells were harvested, cross-linked, and fragmented using the truChIP® Chromatin Shearing Kit (Covaris) according to the manufacturer's instructions to achieve a DNA shear length of approximately 200 bp. Solubilized chromatin was incubated with the appropriate antibody (anti-FLAG; Sigma, F3165) and eluted using the EZ-ChIP™ Chromatin Immunoprecipitation Kit (Millipore). Then, libraries were prepared using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) and assessed by the Agilent 2200 TapeStation (Agilent Technologies, USA) and Qubit (Thermo Fisher Scientific, USA). Sequencing was performed using an Illumina platform with 150 bp paired-end reads.

After pre-processing with Trimmomatic tools (version 0.36) and quality control by FastQC software, the sequencing reads were aligned to a reference genome (UCSC hg38) using Bowtie2 (version 2.4.2).[99] MACS2 (version 2.2.7)[100] was employed to perform peak calling and Homer software (version 4.11.1) was used to annotate the Peaks. GO and KEGG pathway enrichment analysis were performed using the R package clusterProfiler (version 4.2.0).[101]

Drug Response Experiments

Cell line-based drug response test

Cells were seeded in 96-well plates at 2,500 cells per well and cultured for 24 hours, followed by different concentrations of alisertib, barasertib and AMG-900 treatment for 72 hours. Cell viability was assessed using 60 μL of Cell Titer Glo reagent (Promega), and luminescence was determined using a Bio Tek Synergy NEO.

Xenograft tumorigenesis assay

For the establishment of human SCLC cell line xenografts, nude mice (BK Co.) were subcutaneously injected with 1 × 10^6 cells (H146, H82, H446, H69, H526, DMS114). For the establishment of SCLC PDX models, transbronchoscopic lung biopsies from 16 SCLC patients were collected and embedded in matrigel for injection into the right flank of severe combined immune deficiency (SCID) mice (SLAC Co.). Details were performed as previously described.[59]

Multi-omics profiling of tumor tissues from xenograft mouse models

Tumor tissues from 16 patient-derived tumor xenograft (PDX) and 6 SCLC cell line-derived xenograft (CDX) mouse models were used for multi-omics profiling. The method of RNA extraction, RNA sequencing was exactly the same as that of the mentioned above. For proteomic and phosphoproteomic analysis, approximately 20-50 mg tumor tissues were resuspended in 300 μL SDT lysis buffer (4% SDS, 0.1 M Tris-HCl, 0.1 M DTT, pH 7.6), sonicated at 20% amplitude for 2 min (5 s on, 5 s off), denatured and reduced at 95°C for 5 min. Then, lysates were centrifuged at 12,000 g for 10 min and protein concentration of the supernatant was determined using tryptophan-based fluorescence quantification method.[140] 16 samples with relative high protein concentration were selected and mixed in equal protein amount, yielding an internal reference sample for PDX/CDX TMT-based proteomic analysis. Three TMT 10-plex (Thermo Fisher Scientific) sets were conducted to label these individual mouse tumor tissue and mixed samples. The methods of tryptic digestion, TMT 10-plex labeling, high-pH RPLC peptides fractionation and phosphopeptides enrichment were the same as SCLC patient samples.

For proteomic analysis, LC-MS/MS analysis was performed with a nanoflow Easy nLC 1200 UHPLC (Thermo Fisher Scientific) coupled to a Q Exactive HF-X (Thermo Fisher Scientific) over a 120-min separation gradient at a flow rate of 300 nL/min (2%-5% B in 1 min; 5%-32% B in 94 min; 32%-45% B in 15 min; 45%-65% B in 3 min; 65%-100% B in 1 min; 100% B in 6 min). The RP chromatographic column and the solvent A/B was the same as above. MS parameters were as follows: MS1 resolution – 120,000, mass range – 350 – 1,700 m/z, AGC target – 3e6, maximum IT – 50 ms, charge state exclude – 1,7,8,>8, dynamic exclusion – 40 s, top 20 ions selected for MS2; MS2: resolution – 45,000, HCD energy – 32, isolation window – 1.0 m/z, AGC target – 1e5, maximum IT – 120 ms.

For phosphosproteomic analysis, LC-MS/MS analysis was performed with a nanoflow Easy nLC 1200 UHPLC (Thermo Fisher Scientific) coupled to a Q Exactive HF (Thermo Fisher Scientific) over a 90-min separation gradient at a flow rate of 300 nL/min (2%-25% B in 70 min; 25%-32% B in 10 min; 32%-100% B in 2 min; 100% B in 8 min). The parameters of Q Exactive HF mass spectrometer were set the same as described above.

Cell line and xenograft model-based drug response test

4 PDX models (SC022, SC222, SC224, SC234) and 4 SCLC cell line xenografts (H69, DMS114, H446, H146) were used for evaluation of specific compounds or drugs. When tumor volume had reached 60-200 mm^3, the mice were randomized into different treatment groups. For nmf1 SCLC models (SC234, H146), mice were orally administered with anlotinib (6 mg/kg) or alisertib (30 mg/kg) for consecutive 14 days respectively. For nmf3 SCLC models (SC022, SC224, DMS114), mice were orally administered with anlotinib (6 mg/kg) for consecutive 14 days respectively. For nmf4 SCLC models (SC222, H446), mice were orally administered with alisertib (30 mg/kg) for consecutive 14 days respectively. For evaluation of the effects of E/P in H69 (nmf1) and SC222 (nmf4) models, mice

were i.p. injection with cisplatin (DDP, 6 mg kg^{-1} d^{-1} intraperitoneally) at day 1 and etoposide (VP16, 10 mg kg^{-1} d^{-1}) at days 1-3. One week was considered as one cycle. Nude mice were generally given two or more cycles of E/P treatment.

QUANTIFICATION AND STATISTICAL ANALYSIS

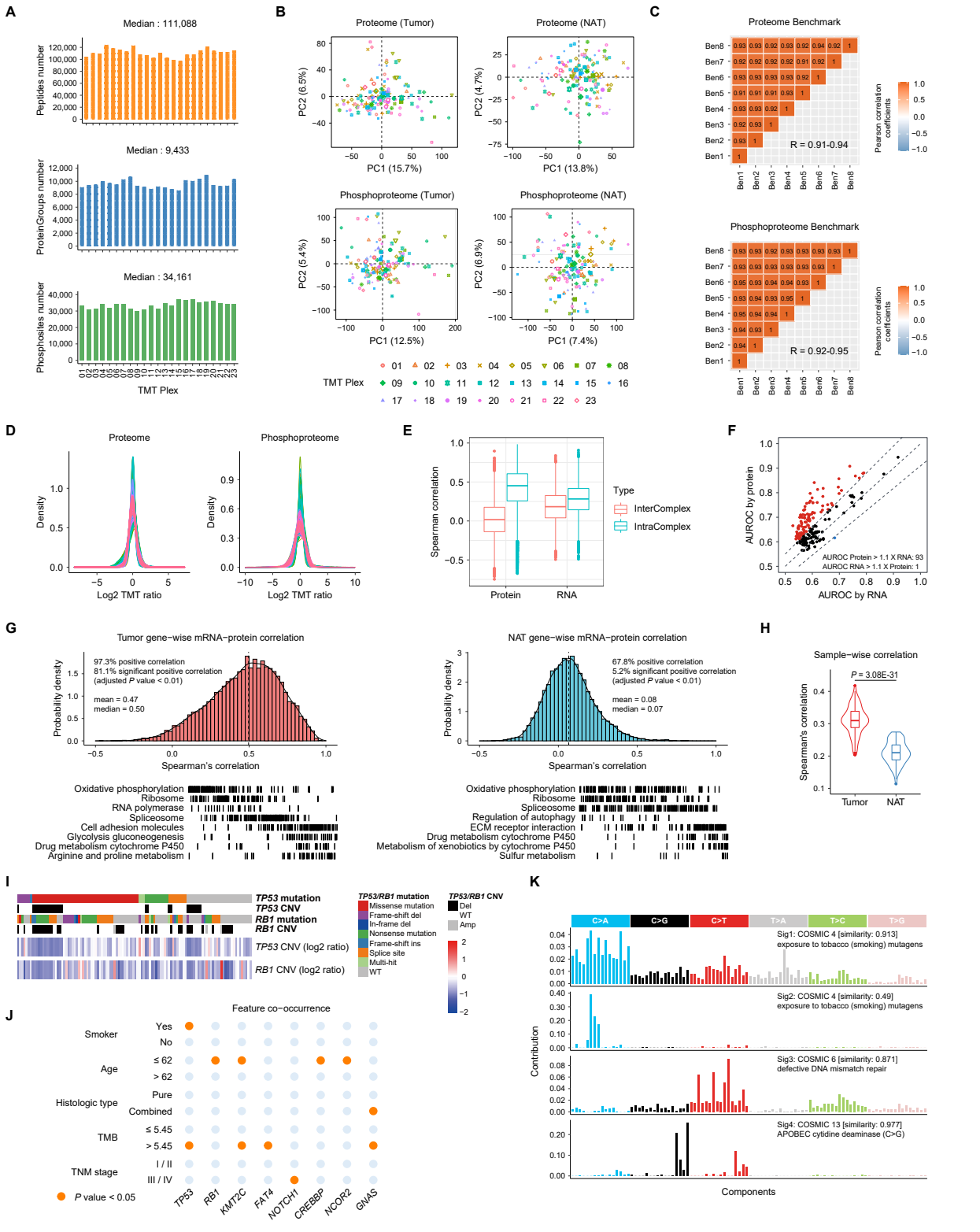
Quantification methods and statistical analysis methods for single-omics and multi-omics analyses were mainly described and referenced in the respective method details subsections.

Additionally, standard statistical tests were used to analyze the data, including but not limited to Wilcoxon test, Fisher's exact test, Chi-square test, Kruskal-Wallis test, log rank test and Cox proportional hazards regression model. For categorical variables versus categorical variables, Fisher's exact test or Chi-square test was used; for categorical variables versus continuous variables, Student's t test, Wilcoxon test or Kruskal-Wallis test was used; and for continuous variables versus continuous variables, Pearson or Spearman's correlation was used. All statistical tests were two-sided, and statistical significance was considered when P value < 0.05 unless otherwise indicated. Multiple comparisons were adjusted by the Benjamini-Hochberg procedure. Kaplan-Meier plots (log rank test) were used to describe overall survival. Variables associated with overall survival were identified using univariate Cox proportional hazards regression models. Significant factors in univariate analysis were further subjected to a multivariate Cox regression analysis. Statistical analyses were performed using R unless explained otherwise.

For functional experiments, each was repeated at least three times independently, and the results were shown as mean ± standard error of the mean (SEM). The statistical significance of differences was determined by Student's t test using GraphPad Prism (version 8). Details can be found in results and Figure Legends.

CellPress

Supplemental figures



(legend on next page)

Figure S1. Data quality assessments and mutational profiles, related to Figure 1

(A) Barplots showing the numbers of identified peptides, proteins, and phosphosites across the 23 TMT sets.

(B) Principal-component analysis (PCA) plots of proteome and phosphoproteome separately for tumor and NAT samples. Samples analyzed in different TMT sets are shown with different colors and shapes.

(C) Pearson correlation plots of longitudinal benchmark samples showing robust and accurate quantification reproducibility across several months of data acquisition time by mass spectrometry. (Pearson's correlation coefficients, proteome R = 0.91–0.94, phosphoproteome: R = 0.92–0.95.) The benchmark samples were interposed every three TMT 11 sets in proteome and phosphoproteome platform.

(D) Density plots of TMT protein/internal reference ratios and phosphosite/internal reference ratios showing no obvious protein degradation in tumor and NAT samples. (Hartigans' dip test, $p > 0.05$.)

(E) Boxplots showing quality assessment of proteomic and transcriptomic data based on protein complex correlation analysis.

(F) Scatterplots comparing the gene function prediction accuracy using co-expression networks based on RNA-seq and TMT proteomics data. Network-based gene function prediction was performed using the random walk-based network propagation algorithm for each KEGG pathway. Prediction performance was evaluated using 5-fold cross validation and quantified based on the area under the receiver operating characteristic curve (AUROC). Red and blue indicate pathways with >10% difference.

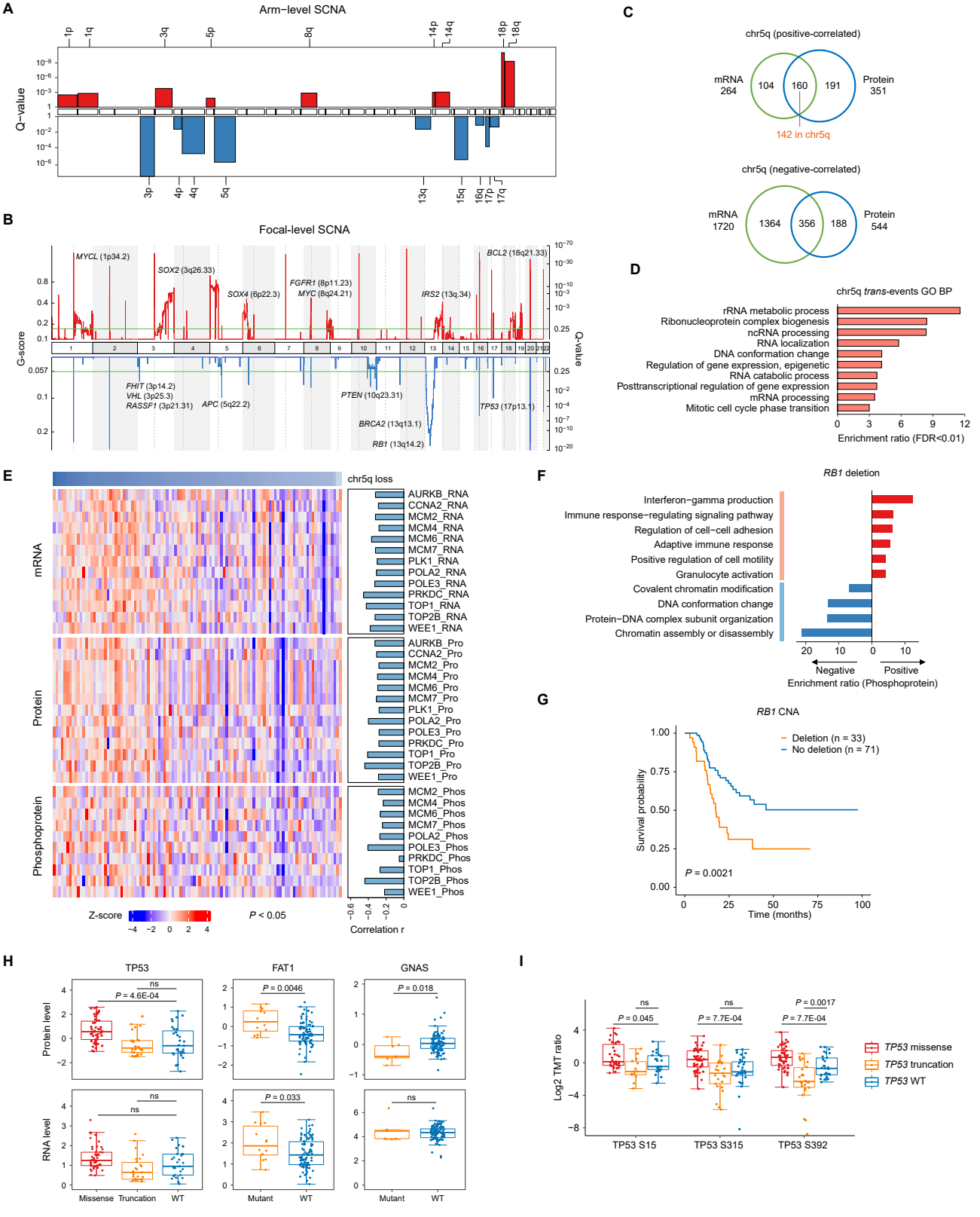
(G) Histograms of gene-wise correlation between mRNA and protein expression in SCLC tumors (top left) and NATs (top right). Different GSEA enrichment KEGG pathways representing different levels of correlation are annotated (bottom).

(H) Violin plot showing the comparison of tumors and NATs sample-wise correlation between mRNA and protein expression (Wilcoxon ranked-sum test).

(I) *TP53* and *RB1* somatic mutation and copy-number variation in SCLC.

(J) Correlation between somatic mutations and clinical features in TU-SCLC cohort, significant correlations were highlighted in orange (Fisher's exact test). Patients were grouped based on the median age of 62 years and the median tumor mutation burden (TMB) value of 5.45.

(K) Trinucleotide motif frequency plots and enriched mutational signatures identified in TU-SCLC cohort. Four mutational signatures were inferred, matching previously described mutational signatures associated with exposure to tobacco (smoking) mutagenes (cos sim = 0.913 and 0.49 with COSMIC 4), defective DNA mismatch repair (cos sim = 0.871 with COSMIC 6) and APOBEC cytidine deaminse (C>G) (cos sim = 0.977 with COSMIC 13).



(legend on next page)

Figure S2. Proteogenomic consequences of genetic aberrations, related to Figure 2

(A) Chromosome arm-level SCNAs in TU-SCLC cohort.

(B) Focal-level SCNAs with selected genes labeled in the significantly peak regions.

(C) Venn diagrams depicting mRNA/proteins with positive (upper) or negative (lower) *trans* CNA-mRNA and CNA-protein correlations with chromosome 5q loss.

(D) Enriched GO biological processes terms for proteins with negative *trans* CNA-protein correlations with chromosome 5q loss.

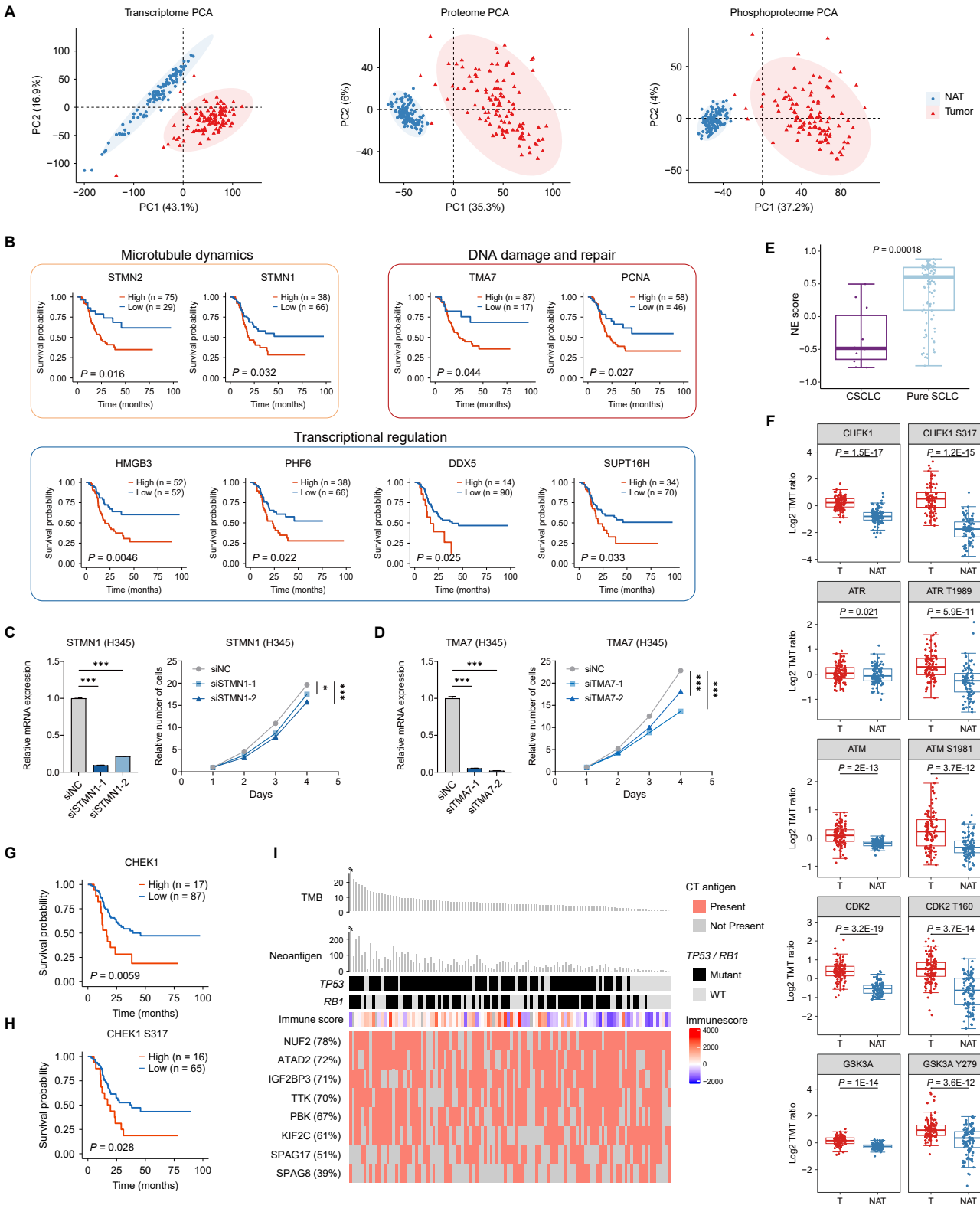
(E) Heatmap of chromosome 5q copy-number loss and mRNA/protein/phosphoprotein abundance of genes involved in DNA synthesis, DNA replication and repair, and cell cycle progression pathways. The right panel shows the correlation between CNAs in chromosome 5q and mRNA/protein/phosphoprotein expression (Spearman's correlation, $p < 0.01$).

(F) Enriched KEGG pathway terms for phosphoproteins with positive or negative *trans* CNA-phosphoprotein correlations with *RB1* deletion.

(G) Kaplan-Meier curves for overall survival of patients with or without *RB1* copy-number deletion (log rank test).

(H) Boxplots showing comparisons of mRNA and protein abundance for TP53, FAT1, and GNAS across tumor samples stratified by different mutation status (Wilcoxon ranked-sum test, ns, not significant).

(I) Boxplots showing comparisons of TP53 S15, S315, and S392 phosphosite abundance between samples with *TP53* missense/truncating mutations and WT samples (Wilcoxon ranked-sum test, ns, not significant).



(legend on next page)

Figure S3. Proteomic alterations in tumors and paired NATs, related to Figure 3

(A) Principal-component analysis (PCA) plots of mRNA, protein, and phosphosite in tumors (pink triangles) and paired NATs (blue circles).

(B) Kaplan-Meier curves for overall survival based on proteomic abundance of SCLC-associated proteins (log rank test). Proteins were classified by their function as microtubule dynamics, DNA damage and repair, and transcriptional regulation. Patients were stratified by the optimal cutpoint using maximally selected rank statistics (maxstat) on protein abundance. Numbers in parentheses represent the sample sizes for the involved groups.

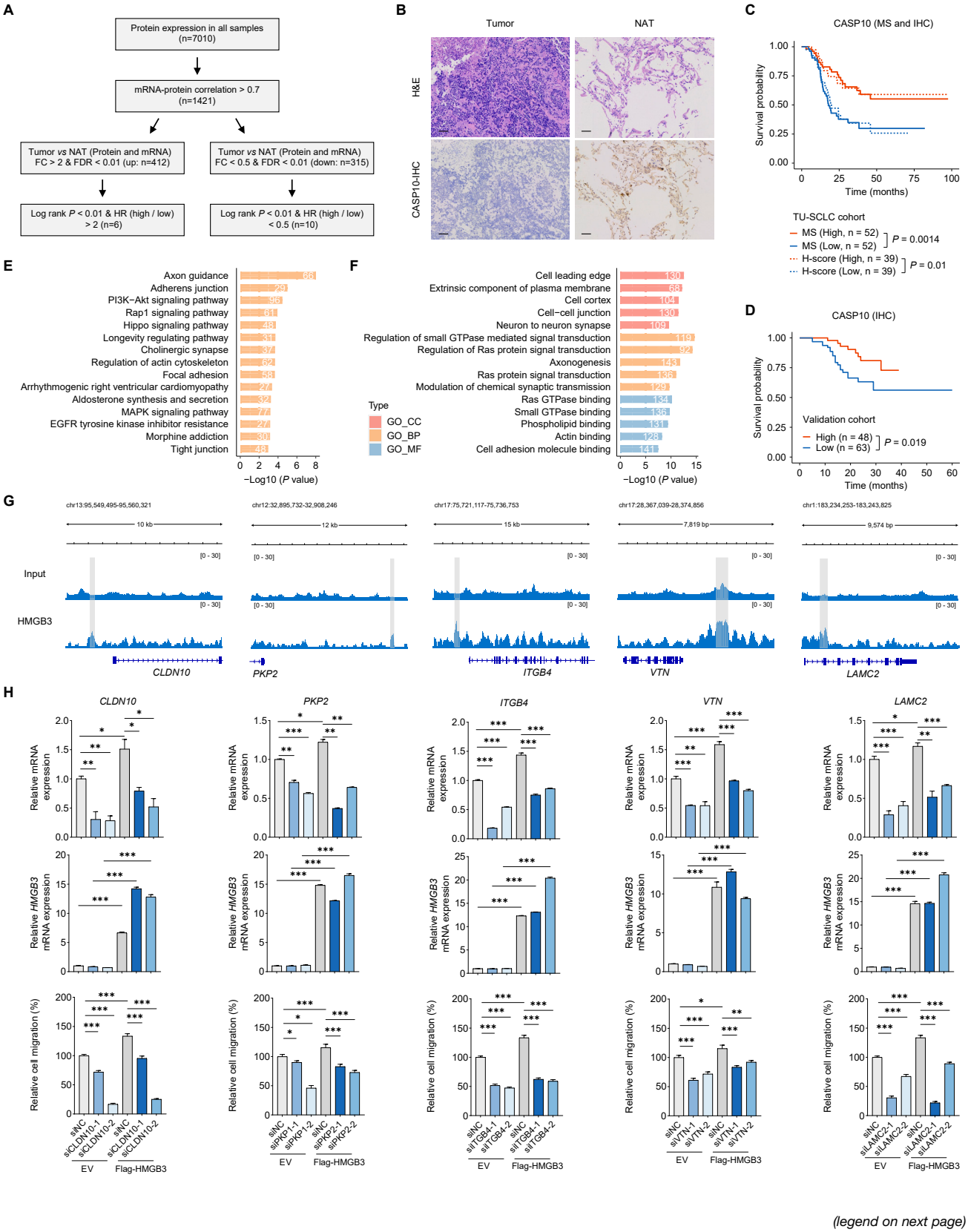
(C and D) The impacts of STMN1 (C) and TMA7 (D) gene knockdown on H345 cell growth. Proliferation of indicated cells was measured by CCK8 method. The efficiency of siRNA gene knockdown was validated by real-time qPCR and was displayed on the left of each panel. Data were represented as mean ± SEM (Student's t test), *$p < 0.05$, ***$p < 0.001$.

(E) Comparison of NE scores between CSCLC and pure SCLC tumors (Wilcoxon ranked-sum test).

(F) Boxplots showing the comparison of kinase protein or its activating site phosphorylation abundance between tumors and paired NATs (Wilcoxon signed-rank test).

(G and H) Kaplan-Meier curves for overall survival based on CHEK1 protein (G) and CHEK1 S317 phosphosite abundance (H) (log rank test). Patients were stratified by the optimal cutpoint using maxstat on protein or phosphosite abundance. Numbers in parentheses represent the sample sizes for the involved groups.

(I) Heatmap representing cancer testes (CT) antigens that were overexpressed at least 2-fold in tumors compared with paired NATs in more than 10% of samples. The percentage was indicated in the brackets. The upper barplots show TMB and predicted neoantigen counts in each sample. Samples are ordered by decreasing TMB.



(legend on next page)

Figure S4. Identification and validation of proteomic prognostic biomarkers, related to Figure 4

(A) Workflow for selecting potential SCLC prognostic proteins. FC, fold change; HR, hazard ratio.

(B) Representative H&E and IHC staining images for CASP10 on tumors and paired NATs (scale bars, 60 μm).

(C) Kaplan-Meier curves for overall survival based on CASP10 proteomic abundance or immunostaining scores (log rank test).

(D) Kaplan-Meier curves for overall survival based on CASP10 immunostaining scores in an independent SCLC cohort (n = 111) (log rank test).

(E and F) Representative KEGG (E) and GO (F) enrichment results for genes bound by HMGB3 from the ChIP-seq assay.

(G) ChIP-seq in H345 cells revealed binding peaks for HMGB3 on *CLDN10*, *PKP2*, *ITGB4*, *VTN*, and *LAMC2* genes.

(H) The impacts of CLDN10, PKP2, ITGB4, VTN, and LAMC2 knockdown on HMGB3-overexpressed or parental H345 cell migration. The efficiency of siRNA gene knockdown and HMGB3 overexpression was validated by real-time qPCR. Transwell migration assays of indicated cells were measured. Data were represented as mean ± SEM (Student's t test), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

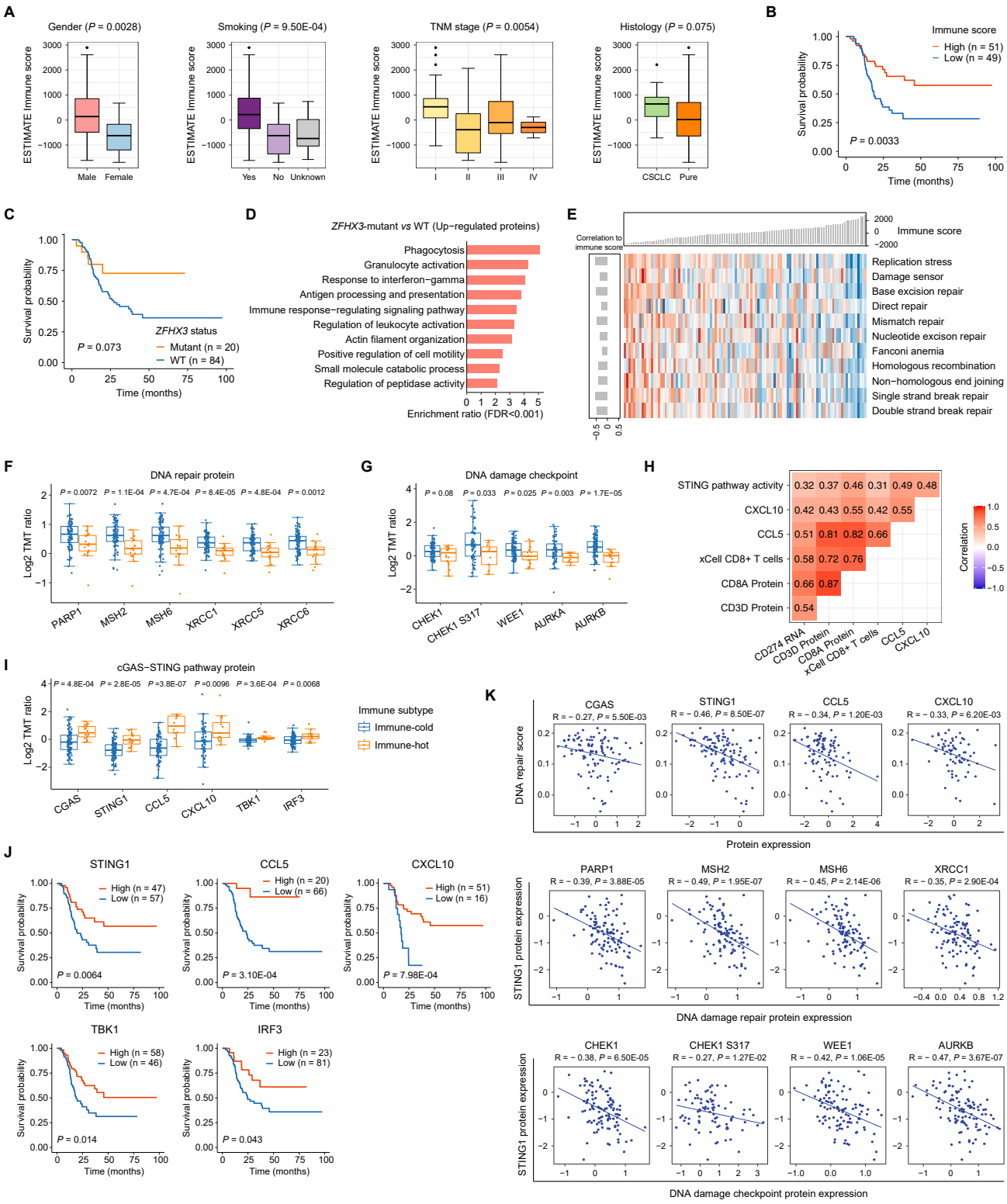


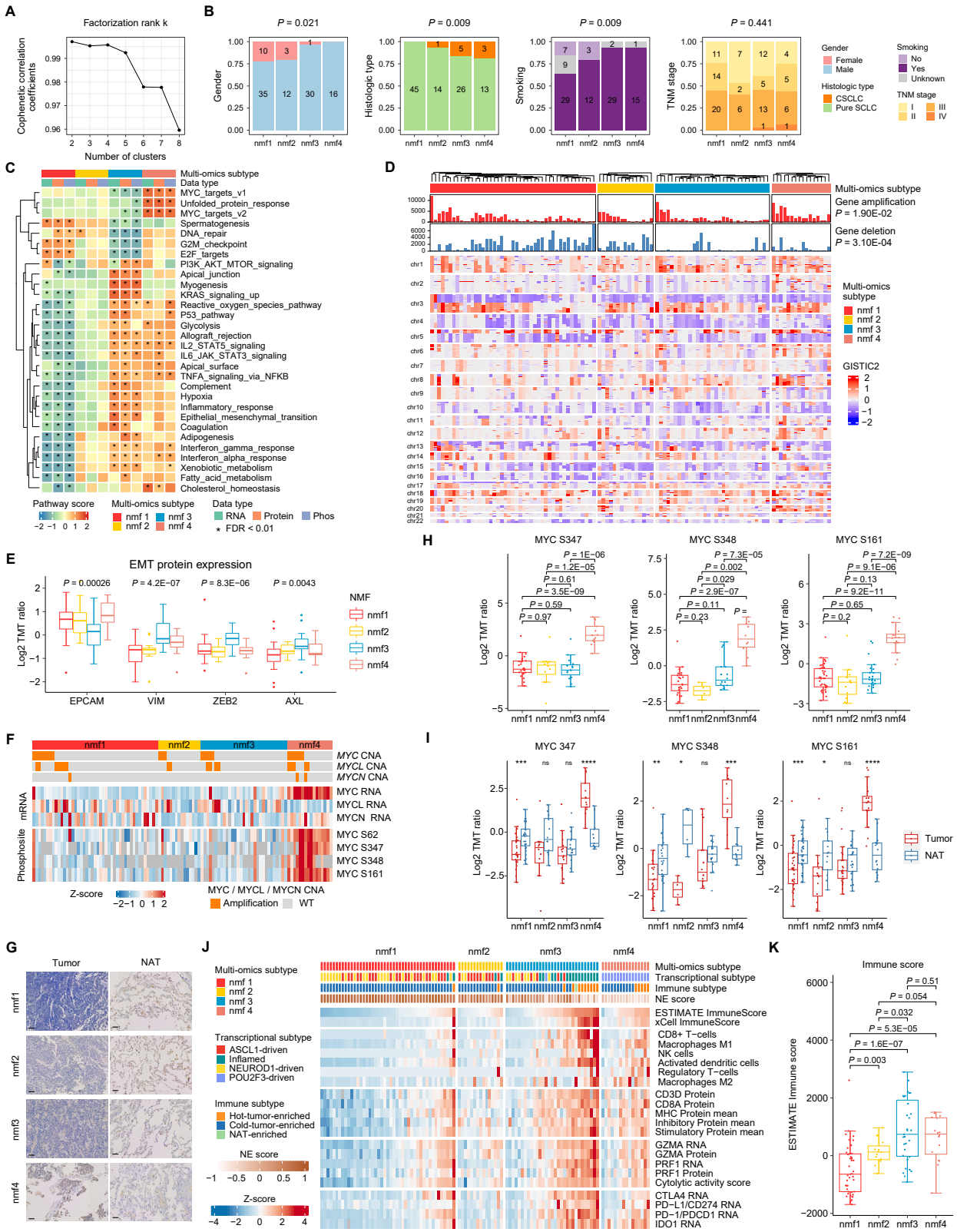
Figure S5. Immune landscape of TU-SCLC cohort, related to Figure 5

(A) Boxplot showing comparison of ESTIMATE immune scores across different clinical parameters (Wilcoxon ranked-sum test for gender and histology, Kruskal-Wallis test for smoking history and TNM stage).

(B) Kaplan-Meier curves for overall survival based on ESTIMATE immune scores (log rank test).

(legend continued on next page)

(C) Kaplan-Meier curves for overall survival of patients with or without *ZFHX3* mutation (log rank test).
(D) Enriched KEGG pathway terms for upregulated proteins in *ZFHX3*-mutant when compared with *ZFHX3*-WT tumors.
(E) Heatmap showing Spearman's correlations between immune scores and proteomics-derived signatures for DNA damage response (DDR) gene sets. Samples were ordered by immune scores as shown in the upper barplot.
(F and G) Boxplots showing comparison of protein or phosphosite expression between immune-cold and immune-hot tumors for DNA repair protein (G) and DNA damage checkpoints (H) (Wilcoxon ranked-sum test).
(H) Spearman's correlation among STING pathway activity and immune-related features.
(I) Boxplots showing comparison of protein expression between immune-cold and immune-hot tumors for cGAS-STING pathway proteins (Wilcoxon ranked-sum test).
(J) Kaplan-Meier curves for overall survival based on proteomic abundance of cGAS-STING pathway proteins (log rank test). Patients were stratified by the optimal cutpoint using maximally selected rank statistics (maxstat) on protein abundance. Numbers in parentheses represent the sample sizes for the involved groups.
(K) Scatterplot comparing DNA repair score and cGAS-STING pathway protein and the protein expression of STING1 with DNA repair protein and DNA damage checkpoints (Spearman's correlation).



(legend on next page)

Figure S6. Characterization of the integrated multi-omics subtypes, related to Figure 6

(A) Cophenetic correlation coefficients for a range of factorization ranks. The maximal cophenetic correlation coefficient was observed for rank k = 4.

(B) Sample distribution across different clinical parameters in each subtype (Fisher's exact test).

(C) Enrichment of cancer hallmark gene sets in multi-omics subtypes detected by single sample gene set enrichment analysis (ssGSEA).

(D) Heatmap depicting all somatic copy-number aberrations (SCNAs) detected in TU-SCLC cohort. Genes are ordered by chromosomal location and samples are organized by multi-omics subtype and separately clustered for each subtype.

(E) Boxplots showing mesenchymal and epithelial proteins across multi-omics subtypes (Kruskal-Wallis test).

(F) Heatmap showing MYC family genes CNA and relative mRNA and phosphosite expression across multi-omics subtypes.

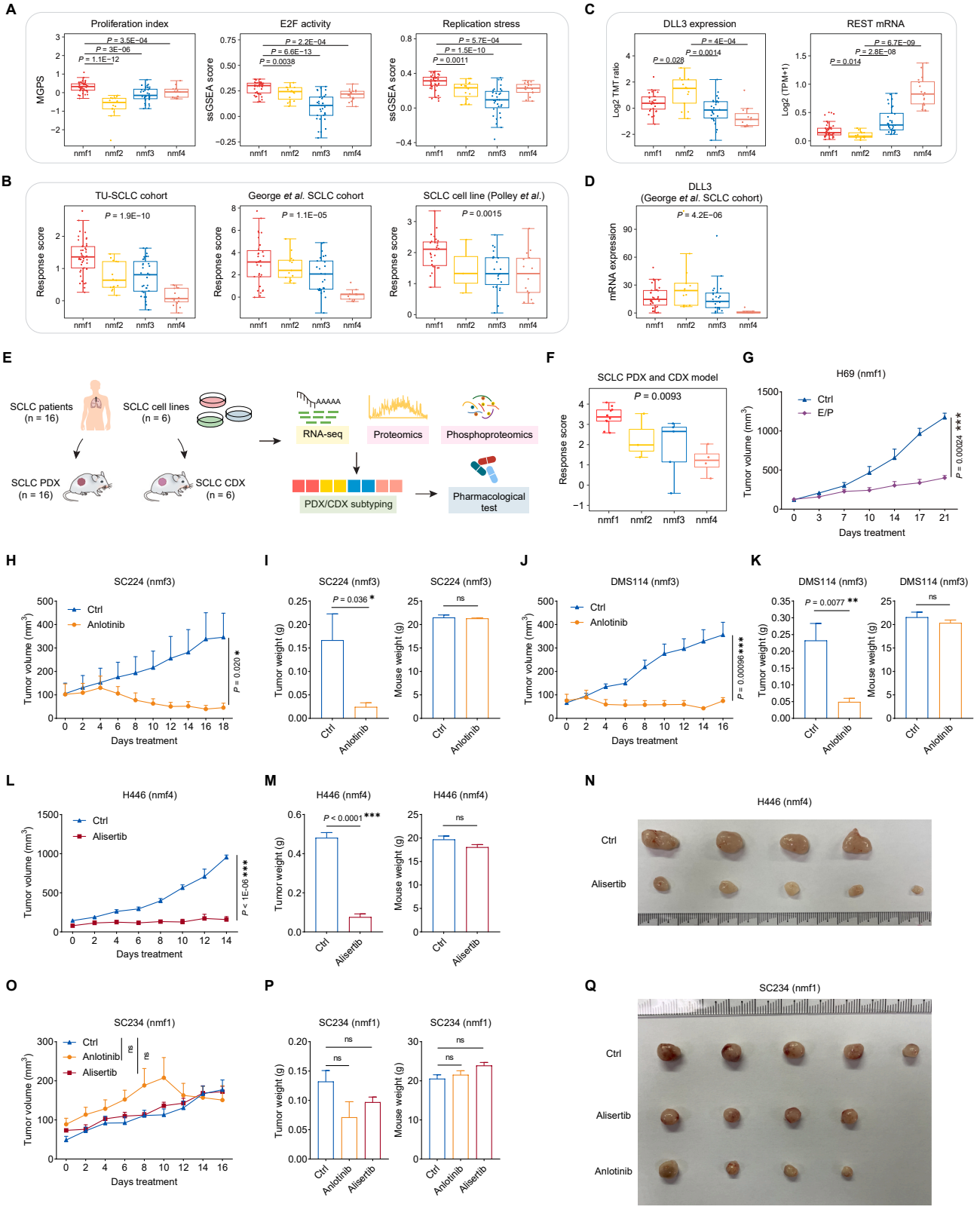
(G) Representative IHC staining images for MYC protein on tumor and NAT samples in each subtype (scale bars, 60 μm).

(H) Boxplot showing comparison of MYC S347, MYC S348, and MYC S161 phosphosite abundance across subtypes (Wilcoxon ranked-sum test).

(I) Boxplots showing MYC S347, MYC S348, and MYC S161 phosphosite expression between tumors and paired NATs in each subtype, respectively (Wilcoxon signed-rank test).

(J) Heatmap showing the wide range of expression levels for immune-related features across multi-omics subtypes. Within each subtype, samples were ordered by increasing ESTIMATE immune scores.

(K) Boxplot showing comparison of ESTIMATE immune scores across multi-omics subtypes (Wilcoxon ranked-sum test).



(legend on next page)

Figure S7. Subtype-specific therapeutic strategies in SCLC, related to Figure 7

(A) Comparisons of molecular features inferred from proteogenomic data across multi-omics subtypes (Wilcoxon ranked-sum test). Proliferation index was inferred from mRNA-based multi-gene proliferation score (MGPS).[135] E2F activity and replication stress activity were inferred by ssGSEA.

(B) Boxplots showing comparisons of response score toward ATR and TOP1 inhibition across subtypes in TU-SCLC cohort, George et al. SCLC cohort,[9] and SCLC cell lines[52] (Kruskal-Wallis test).

(C) Boxplot showing comparison of DLL3 protein expression and REST mRNA expression across multi-omics subtypes (Wilcoxon ranked-sum test).

(D) Comparisons of DLL3 mRNA expression across subtypes from George et al. SCLC cohort[9] (Kruskal-Wallis test).

(E) Experimental workflow of multi-omics profiling of tumor tissues from xenograft mouse models.

(F) Boxplot showing comparison of response score toward ATR and TOP1 inhibition across subtypes in SCLC PDX/CDX tumor models (Kruskal-Wallis test).

(G) Tumor growth in H69 tumor-bearing mice treated with E/P (etoposide + cisplatin) or vehicle. Tumor volumes were monitored every other day by caliper. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(H) Tumor growth in SC224 tumor-bearing mice treated with anlotinib or vehicle. Tumor volumes were monitored every other day by caliper. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(I) Statistical analyses of tumor weights (left) and mouse weights (right) from SC224 PDX models treated with anlotinib or vehicle. Data are represented as mean ± SEM (Student's t test).

(J) Tumor growth in DMS114 tumor-bearing mice treated with anlotinib or vehicle. Tumor volumes were monitored every other day by caliper. Data are represented as mean ± SEM (n = 5 mice per group, Student's t test).

(K) Statistical analyses of tumor weights (left) and mouse weights (right) from DMS114 CDX models treated with anlotinib or vehicle. Data are represented as mean ± SEM (Student's t test).

(L) Tumor growth in H446 tumor-bearing mice treated with alisertib or vehicle. Tumor volumes were monitored every other day by caliper. Data are represented as mean ± SEM ($n \geq 4$ mice per group, Student's t test).

(M) Statistical analyses of tumor weights (left) and mouse weights (right) from H446 CDX models treated with alisertib or vehicle. Data are represented as mean ± SEM (Student's t test).

(N) The tumors from H446 CDX models treated with alisertib or vehicle.

(O) Tumor growth in SC234 tumor-bearing mice treated with anlotinib, alisertib, or vehicle. Tumor volumes were monitored every other day by caliper. Data are represented as mean ± SEM ($n \geq 4$ mice per group, Student's t test).

(P) Statistical analyses of tumor weights (left) and mouse weights (right) from SC234 PDX models treated with anlotinib, alisertib, or vehicle. Data are represented as mean ± SEM (Student's t test).

(Q) The tumors from SC234 PDX models treated with anlotinib, alisertib, or vehicle.