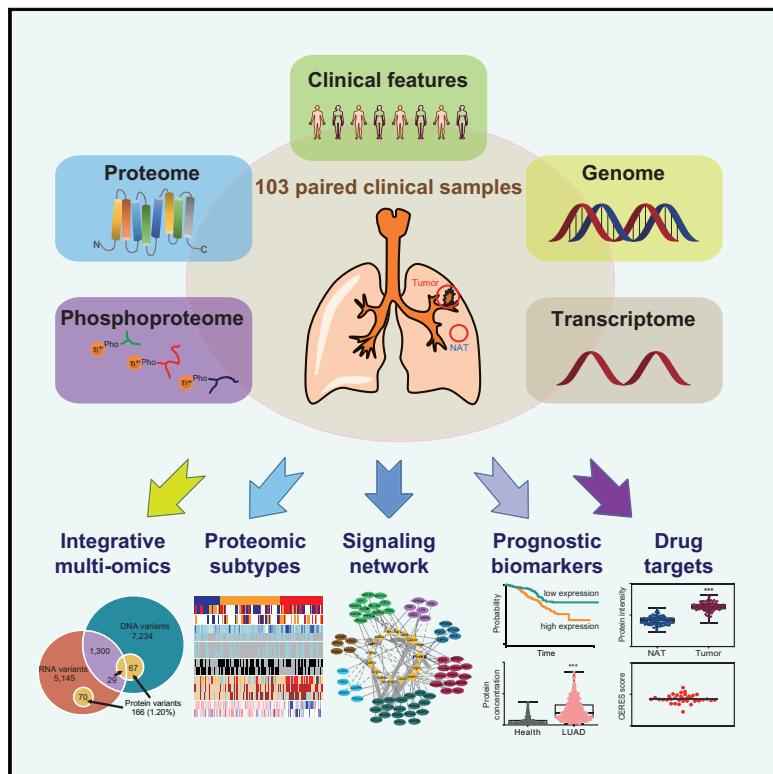


Integrative Proteomic Characterization of Human Lung Adenocarcinoma

Graphical Abstract



Authors

Jun-Yu Xu, Chunchao Zhang,
Xiang Wang, ..., Fuchu He, Ting Xiao,
Minjia Tan

Correspondence

wangyi@ncpsb.org.cn (Y.W.),
jing.li@sjtu.edu.cn (J.L.),
hefc@nic.bmi.ac.cn (F.H.),
xiaot@cicams.ac.cn (T.X.),
mjtan@simm.ac.cn (M.T.)

In Brief

Integrative proteomic characterization of lung adenocarcinoma in 103 Chinese patients identifies three subtypes related to clinical and molecular features and nominates potential prognostic biomarkers and drug targets.

Highlights

- Discovery of prognosis-associated proteins and pathways at early stage of LUAD
- Proteomics revealed three subtypes related to clinical and molecular features
- Identification of subtype-specific kinases and cancer-associated phosphoproteins
- Identification of potential prognostic biomarkers and drug targets in LUAD



Resource

Integrative Proteomic Characterization of Human Lung Adenocarcinoma

Jun-Yu Xu,^{1,11} Chunchao Zhang,^{2,11} Xiang Wang,^{3,11} Linhui Zhai,^{1,11} Yiming Ma,^{4,11} Yousheng Mao,⁵ Kun Qian,⁶ Changqing Sun,⁷ Zhiwei Liu,^{1,8} Shangwen Jiang,¹ Minghui Wang,³ Lin Feng,³ Lei Zhao,¹ Ping Liu,¹ Bo Wang,⁴ Xin Zhao,⁶ Hui Xie,⁷ Xiaoyun Yang,⁴ Liyuan Zhao,⁴ Yafei Chang,⁴ Jingya Jia,⁴ Xijun Wang,³ Yimin Zhang,⁴ Yaru Wang,³ Yikun Yang,⁵ Zhixiang Wu,¹ Longhai Yang,⁵ Bin Liu,¹ Teng Zhao,⁶ Shengguo Ren,⁹ Aihua Sun,² Yang Zhao,² Wantao Ying,² Fei Wang,⁹ Guangshun Wang,¹⁰ Yi Zhang,⁶ Shujun Cheng,³ Jun Qin,² Xiaohong Qian,² Yi Wang,^{2,*} Jing Li,^{4,*} Fuchu He,^{2,*} Ting Xiao,^{3,*} and Minjia Tan^{1,8,12,*}

¹State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

²State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

³State Key Laboratory of Molecular Oncology, Department of Etiology and Carcinogenesis, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

⁴Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

⁵Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

⁶Department of Thoracic Surgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China

⁷Department of Oncology, Tianjin Baodi Hospital, Tianjin 301800, China

⁸University of the Chinese Academy of Sciences, Beijing, China

⁹Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science and Technology, Fudan University, Shanghai 200433, China

¹⁰Department of Oncology, Baodi Clinical College of Tianjin Medical University, Tianjin 301800, China

¹¹These authors contributed equally

¹²Lead Contact

*Correspondence: wangyi@ncpsb.org.cn (Y.W.), jing.li@sjtu.edu.cn (J.L.), hefc@nic.bmi.ac.cn (F.H.), xiaot@cicams.ac.cn (T.X.), mjtanj@mimm.ac.cn (M.T.)

<https://doi.org/10.1016/j.cell.2020.05.043>

SUMMARY

Genomic studies of lung adenocarcinoma (LUAD) have advanced our understanding of the disease's biology and accelerated targeted therapy. However, the proteomic characteristics of LUAD remain poorly understood. We carried out a comprehensive proteomics analysis of 103 cases of LUAD in Chinese patients. Integrative analysis of proteome, phosphoproteome, transcriptome, and whole-exome sequencing data revealed cancer-associated characteristics, such as tumor-associated protein variants, distinct proteomics features, and clinical outcomes in patients at an early stage or with *EGFR* and *TP53* mutations. Proteome-based stratification of LUAD revealed three subtypes (S-I, S-II, and S-III) related to different clinical and molecular features. Further, we nominated potential drug targets and validated the plasma protein level of HSP 90 β as a potential prognostic biomarker for LUAD in an independent cohort. Our integrative proteomics analysis enables a more comprehensive understanding of the molecular landscape of LUAD and offers an opportunity for more precise diagnosis and treatment.

INTRODUCTION

Lung cancer accounts for the highest cancer incidence and mortality worldwide (Bray et al., 2018). The 5-year survival rate is less than 20% (Herbst et al., 2018). Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small cell lung cancer (NSCLC), which accounts for about 40% of lung malignancies (Chen et al., 2014). The most common risk factor for

LUAD is smoking (Gould et al., 2013). Other risk factors include long-term exposure to radon, occupational exposure to carcinogens, and outdoor air pollution (Vineis et al., 2005). In recent years, diagnoses of non-smoking patients have increased drastically, highlighting the importance of non-smoking risk factors in LUAD development (Molina et al., 2008). In the past decade, large-scale genomic studies carried out in Western countries have revealed driver genes of LUAD; among them, the most



common somatic mutations are *TP53*, *KRAS*, *KEAP1*, *STK11*, and *EGFR* (Campbell et al., 2016; Collisson et al., 2014; Devarakonda et al., 2015). In China and other Asian countries, female never-smokers with an *EGFR* mutation account for a large percentage (40%–50%) of LUAD cases (Chen et al., 2020; Nahar et al., 2018; Sun et al., 2007; Wang et al., 2018b). Therapeutic options for LUAD include surgery, radiation, chemotherapy, targeted therapy, and immunotherapy or a combination of these treatments (Herbst et al., 2018). Recent advancements in targeted therapies against several oncogenic drivers and immunotherapies targeting immune checkpoints have achieved remarkable success (Hughes et al., 2016; Lin and Shaw, 2016; Peters et al., 2018; Thomas et al., 2015). Inhibitors against *EGFR* and *BRAF* V600E mutations, *ALK* and *ROS1* rearrangements, as well as immune checkpoint inhibitor antibodies against PD-1 or PD-L1 have been approved for precision treatment of LUAD (Hirsch et al., 2017; Vargas and Harris, 2016). Other emerging therapies targeting *MET* amplification, *RET* rearrangement, and *ERBB2* (*HER2*) mutations have also shown promising potential benefits (Hirsch et al., 2017).

Despite this progress, there is still a large percentage of LUAD without available targeted therapeutic options, either because of the lack of known genetic mutations in the key oncogenic signaling pathways or because of the difficulty of targeting oncogenic mutations (e.g., *KRAS* mutations) (Dang et al., 2017; Papke and Der, 2017). Furthermore, intrinsic and acquired resistance to targeted therapies is usually observed in lung cancer patients (Rotow and Bivona, 2017; Tan et al., 2015). Because proteins are the functional executors of the cell, an in-depth characterization of the proteome and signal transduction (phosphoproteome) of LUAD will lay a foundation for comprehensively understanding the molecular mechanisms of the disease and for developing new therapeutic approaches.

In this study, we performed an integrated omics analyses using proteomics, phosphoproteomics, and genomics data collected from 103 LUAD tumors and their paired non-cancerous adjacent tissues (NATs). Our data provided a resource to explore the relationships between genetic variation and transcriptional or translational regulations. Further analysis of the proteomics data uncovered LUAD-related molecular characteristics and their associated clinical outcomes in subgroups with different genotypes.

RESULTS

Molecular Profiling of Chinese LUAD

The present study collected a total of 103 primary LUAD samples with paired NATs from treatment-naïve Chinese patients (Figure S1A; Table S1). A schematic of the experimental design is shown in Figure 1A. A CNHPP (Chinese Human Proteome Project) mass spectrometry (MS)-based label-free quantification strategy (Ge et al., 2018; Jiang et al., 2019) was adopted in the proteomics study. A phosphoproteomics analysis was conducted on 79 paired samples using a TiO₂ enrichment strategy. Whole-exome sequencing (WES) was carried out on all paired samples to detect any possible genetic variants in the cancer genome. In addition, mRNA sequencing was carried out on 51 tumors and 49 paired NATs. Our study therefore provides a

comprehensive summary of LUAD at the multi-omics level (Figure S1B). WES data led to identification of 17,917 genetic variation events (Table S2). Similar to a recent genomics study of Chinese NSCLC patients, our study revealed a unique mutational landscape of Chinese LUAD patients different from those of the Caucasian patients reported by The Cancer Genome Atlas (TCGA) consortium (Collisson et al., 2014), suggesting genetic diversity of cancer genome between races (Wang et al., 2018b). We observed significantly higher *EGFR* (50%) mutations than in the TCGA study (*EGFR*, 14%; Figure 1B; Table S2). In contrast, oncogenic *KRAS* mutations, a main effector of the receptor tyrosine kinase (RTK) pathway, were found less frequently (6%) than in the TCGA cohort (33%) (Figure 1B; Table S2). The mutation rate of *TP53* (51%) was similar to that of the TCGA cohort (46%) (Figure 1B). RNA sequencing (RNA-seq) analysis identified 16,188 genes with fragments per kilobase of transcript per million fragments mapped (FPKM) of more than 1, providing an opportunity to explore the relationship between the transcriptome and full proteome (Table S3). For the proteomics data analysis, a Spearman's correlation coefficient was calculated for all quality control runs of HeLa cell samples (Figure S1C). The average correlation coefficient among the control samples was 0.93, demonstrating the consistent stability of the MS platform. The correlations of the 103 tumor samples were between 0.62–0.96 (mean = 0.85) (Figure S1D). The tumor proteome (orange) and NAT proteome (blue) exhibited a unimodal distribution and passed the proteomics quality control procedure (Figure S1D). Proteomics measurement of all patient samples resulted in a total of 11,119 protein groups at a 1% false discovery rate (FDR) at the protein and peptide levels (Figure 1C; Table S4). 11,091 and 10,793 protein groups were identified in tumors and NATs, respectively. On average, the LUAD proteome had 6,682 protein groups per sample, ranging from a minimum of 4,266 in NATs to a maximum of 8,655 in tumors (Figure S1E; Table S4). Furthermore, a total of 22,564 phosphosites corresponding to 5,277 phosphoproteins were identified with a confident site localization score (probability, >0.75) (Figure 1D; Figure S1F; Table S5). Our study has so far established a comprehensive landscape of Chinese LUAD at the genomics, transcriptomics, proteomics, and phosphoproteomics levels.

Integrated Multi-omics Analyses of Chinese LUAD

Multi-dimensional omics data provided an excellent chance to explore the relationships between the transcriptome and proteome of LUAD. After appropriate sample quality control (QC) and normalization procedures, we performed principal-component analyses (PCAs) of RNA-seq (15,824 genes) and proteomics data (8,252 proteins) in 49 common paired samples, and phosphoproteomics data (2,130 phosphoproteins) in 79 paired samples (Table S4). All datasets could separate tumors and NATs, with the best separation observed with the proteomics analysis (Figures S2A and S2B).

Analyses of mRNA and protein expressions revealed a medium correlation of 0.28 (Pearson correlation coefficient) in tumors and a weaker correlation of 0.07 in NATs (Figures S2C and S2D). Enrichment analyses of the common genes of RNA-seq and proteomics data identified Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that had relatively strong

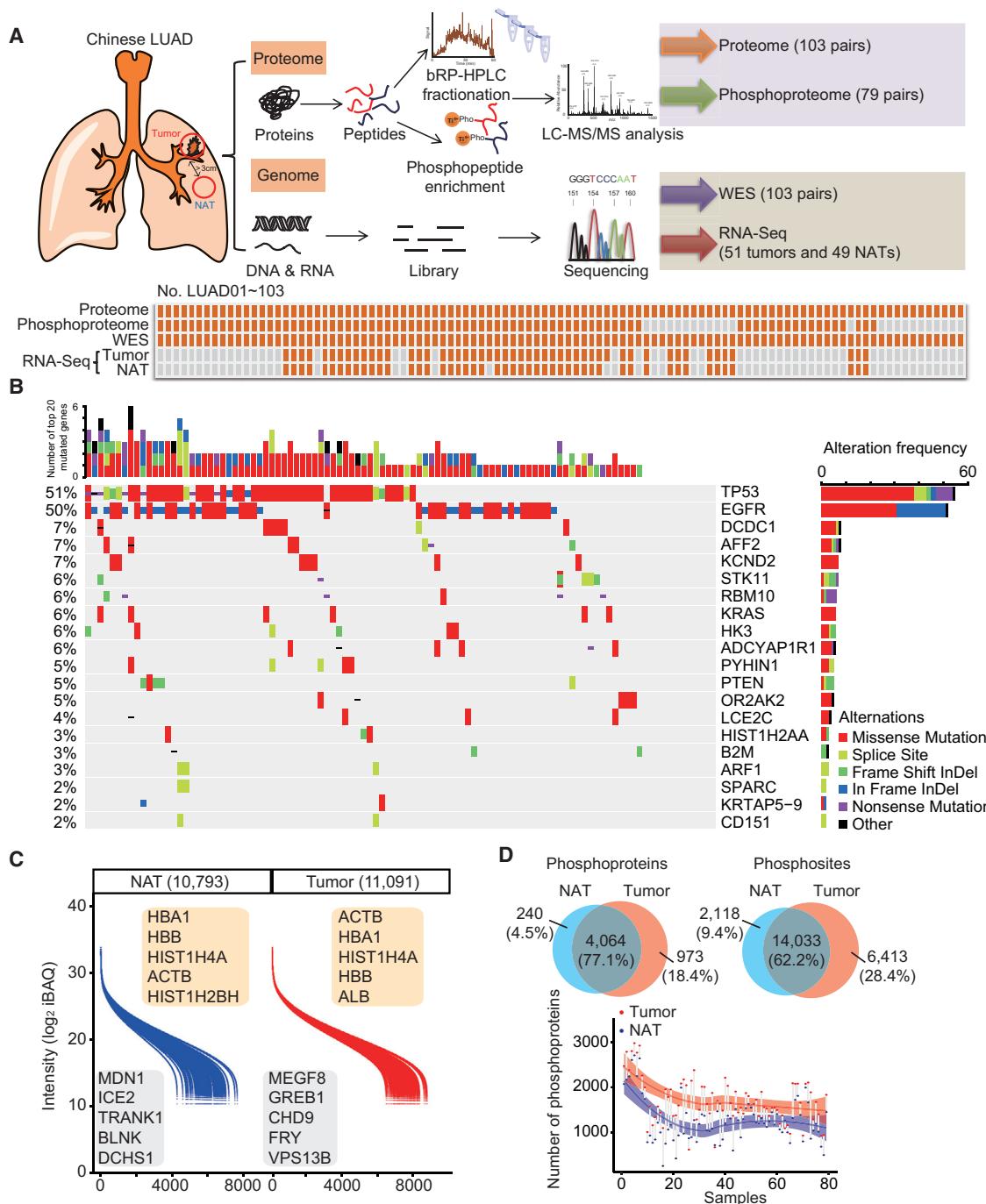


Figure 1. Multi-omics Landscape of LUAD Samples

- (A) Overview of the experimental design and the number of samples for proteomics, phosphoproteomics, WES, and RNA-seq analyses.
- (B) The genomic profiles. Top: mutation counts of the top 20 mutated genes in each patient. Bottom: the top 20 mutated genes and their occurrence in 103 patients. Mutation types and their frequencies are demonstrated by a bar plot in the right panel.
- (C) Overview of the proteomics profile of LUAD patients. Shown are the dynamics of protein abundances identified in tumors (red) and NATs (blue). Proteins were quantified as a normalized intensity-based absolute quantification (iBAQ) value and \log_2 transformed. The highest- and lowest-abundance proteins are shown in the box.
- (D) Overview of the phosphoproteomics profile of LUAD patients. Shown are the numbers of phosphoproteins and phosphosites identified in 79 patients (top) and pairwise comparison of phosphoproteins identified in 79 patients (bottom).

See also Figure S1 and Tables S1, S2, S3, S4, and S5.

or weak mRNA-protein correlations (Table S3). Genes with strong correlations were mainly involved in valine, leucine, and isoleucine degradation and citrate cycle (Figures S2C and S2D), whereas genes with weak or negative correlations were enriched in spliceosome and ribosome pathways (Figures S2C and S2D), which is mostly consistent in tumors and NATs. The discordance of transcriptomics and proteomics suggests that proteomics data have unique oncogenic features that cannot be obtained from genomics data.

We then compared prognostic power (log rank test, Benjamini-Hochberg (BH) $p < 0.01$) of a single gene in two datasets (transcriptomics and proteomics) with 51 common tumor samples. Fifty-two genes were significantly associated with clinical outcomes in both datasets (Figure S2E; Table S3). Among them, four (PRC1, SYDE1, TRIM33, and ZBTB7B) showed opposite associations in two datasets. Additionally, 561 genes with prognostic power were only observed in RNA-seq data, whereas 352 were only found in proteomics data. Although 5.0% of genes and their corresponding proteins showed a similar prognosis, 94.6% of them showed prognostic power at the mRNA or protein level, suggesting complementary prognostic information provided by mRNA and protein data.

We performed whole exome-based somatic copy number alteration (CNA) analyses and examined the regulatory effects of 23,080 CNAs on mRNA, protein, and phosphoprotein expression. As demonstrated in Figure 2A and Figure S2F, CNAs can positively or negatively affect their expressions in a *cis* or *trans* mode: 17,179 CNAs (74.4%) affect 7,686 mRNAs, 10,530 CNAs (45.6%) affect 2,634 proteins, and 4,330 CNAs (18.8%) affect 1,284 phosphoproteins (Spearman correlation analysis, FDR < 0.05) (Figure 2A; Figure S2F; Table S2). Interestingly, *cis*-regulatory effects of CNA (diagonal lines in Figure 2A) on mRNAs and proteins were more prominent than on phosphoproteins (Figure S2F). Strong *trans*-regulatory effects of CNAs were found on all three molecules (vertical stripes in Figure 2A and Figure S2F). The most frequent amplifications were on chromosomes 7p, 7q, 1q, and 20q, and deletions were on 13q, 9p, 9q, and 15q. Among the 20 well-defined driver genes of LUAD (Bailey et al., 2018), the CNAs of four (*EGFR*, *RIT1*, *CDKN2A*, and *STK11*) have a significant *cis* effect on mRNA expression (Table S2). Seven drivers exhibit a *trans* effect on protein expression, whereas only *EGFR* exhibits a *cis* effect on proteins and phosphoproteins (Figure S2G). *EGFR* and *RIT1* had the strongest *cis*- or *trans*-regulatory effects with their targets.

To explore DNA or RNA variants and their protein products, patient-derived databases containing somatic mutations (from WES), novel alternative splicing (from RNA-seq), and fusion proteins (from RNA-seq) were created for the MS database, seeking to confirm these novel gene events at the protein level (Wang and Zhang, 2013). Of the 13,845 (8,630 from WES and 6,544 from RNA-seq) nonsynonymous single-nucleotide variants (SNVs) and 1,365 small indels (insertions or deletions), only 166 (1.2%) proteomic variants (corresponding to 171 mutant peptides) were validated by MS, and 120 of these variants have not yet been annotated in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Figure 2B; Table S6). In addition, 37 variants were indels from 29 proteins (Table S6). On the other hand, 0.32% of novel splicing events (1,366 of 428,320) and

1.59% of fusion transcripts (3 of 189) were validated by MS (Figure 2C; Table S6).

These analyses showed that the transcriptome, proteome, and phosphoproteome possess unique features and, when integrated appropriately, can bring new insights and opportunities to find novel functionally important gene products.

Proteomics Features in Tumor Tissues Compared with NATs

Among the 11,119 proteins identified by liquid chromatography-tandem MS (LC-MS/MS), 3,913 were quantified in all 103 patients, and 5,753 were present in more than 90% of patients (Figure S3A; Table S4). PCA analysis revealed a clear boundary between the tumor and NAT proteomes, indicating an abnormal proteomic landscape during the development and progression of LUAD (Figure 3A). A total of 3,355 proteins were differentially expressed between tumors and NATs (Wilcoxon signed-rank test, BH $p < 0.05$, ratio of tumor over non-cancerous adjacent tissue (T/N) > 2 or $< 1/2$). Among them, 2,749 were upregulated and 606 were downregulated in tumors (Table S4). Gene set enrichment analysis (GSEA) demonstrated that tumor-enriched proteins were significantly enriched in pathways including the proteasome, glycolysis or gluconeogenesis, ribosome, ubiquitin-mediated proteolysis, nucleotide excision repair, and cell cycle (Figure S3B), whereas proteins enriched in NATs were mainly involved in pathways related to lung physiology functions (lung ventilation and immune function) (Clarke et al., 1997; Polio et al., 2019; Wittekindt, 2017), including extracellular matrix (ECM) receptor interaction, focal adhesion, tight junction, and ABC transporters (Figure S3B). The representative differentially expressed proteins in these pathways are shown in Figure 3B. Some of them were well-correlated with the clinical outcomes (Figure 3C). Notably, seven previously reported lung signature proteins (AGER, SFTPA2, CACNA2D2, LAMP3, SCGB1A1, SFTPB, and SFTPC) (Uhlén et al., 2015) had lost expression in tumors (Figure 3D). In addition, patients with higher expression of several lung-enriched proteins in tumors appeared to have better prognostic outcomes (log rank test, $p < 0.05$) (Figure 3D).

GSEA of tumor and NAT transcriptomes ($n = 49$) allowed comparison of enriched pathways revealed in their corresponding proteome analyses. Several cancer hallmark pathways involved in metabolism, proliferation, epigenetics, and protein homeostasis were enriched in RNA-seq and proteomics data (FDR < 0.05 ; Figure S3C). Notably, enrichment of metabolic pathways (upregulated in tumors) and adhesion-related pathways (downregulated in tumors) was much more significant in the proteomics data, whereas enrichment of proliferation-related pathways was more significant in the RNA-seq data (Figure S3C).

Proteomics Characteristics of Early-Stage LUAD

Although tumor node metastasis (TNM) stage I LUAD patients who receive surgical resection have a high recovery rate, up to 20%~30% of early-stage patients have a poor prognosis (Zhang et al., 2013). Proteomics investigation of stage I patients (Table S1) with a good ($n = 36$, disease-free survival [DFS] time > 3 years, DFS good (DG) group) or poor ($n = 15$, DFS time < 3 years, DFS poor (DP) group) prognosis (Figure S4A) could provide clues to understanding the difference in clinical outcomes. GSEA

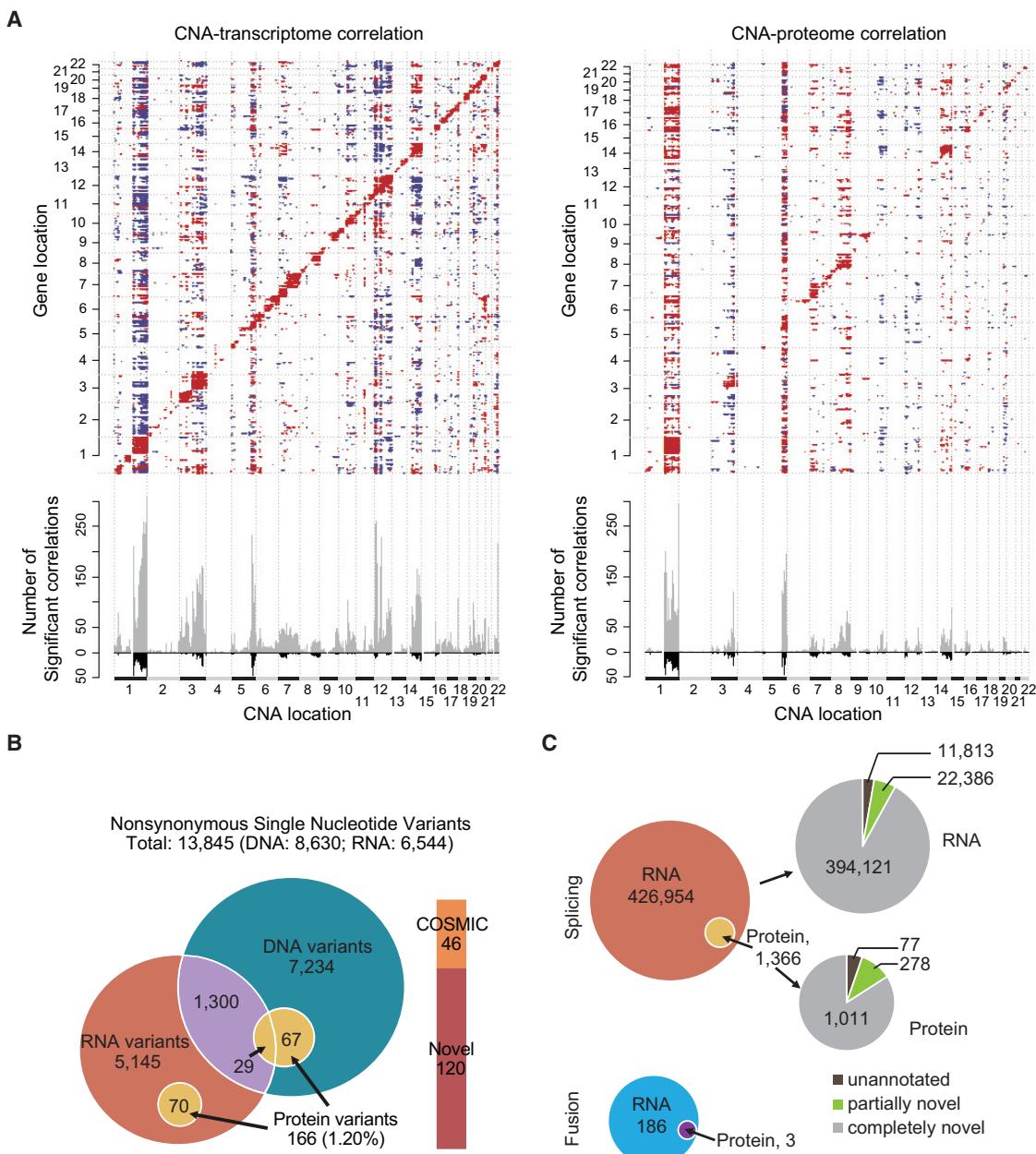


Figure 2. Integrative Omics Analyses of LUAD Samples

(A) Functional effects of CNAs on mRNA and proteins. Top panels: correlation of CNA to mRNA and protein abundance. Positive and negative correlations are indicated in red and blue, respectively. Genes were ordered by chromosomal location on the x and y axes. Diagonal lines indicate *cis* effects of CNA on mRNA or proteins. Bottom panels: number of mRNAs or proteins that were significantly associated with a specific CNA. Gray bars indicate correlations specific to mRNA or proteins, and black bars indicate correlations with both mRNA and proteins.

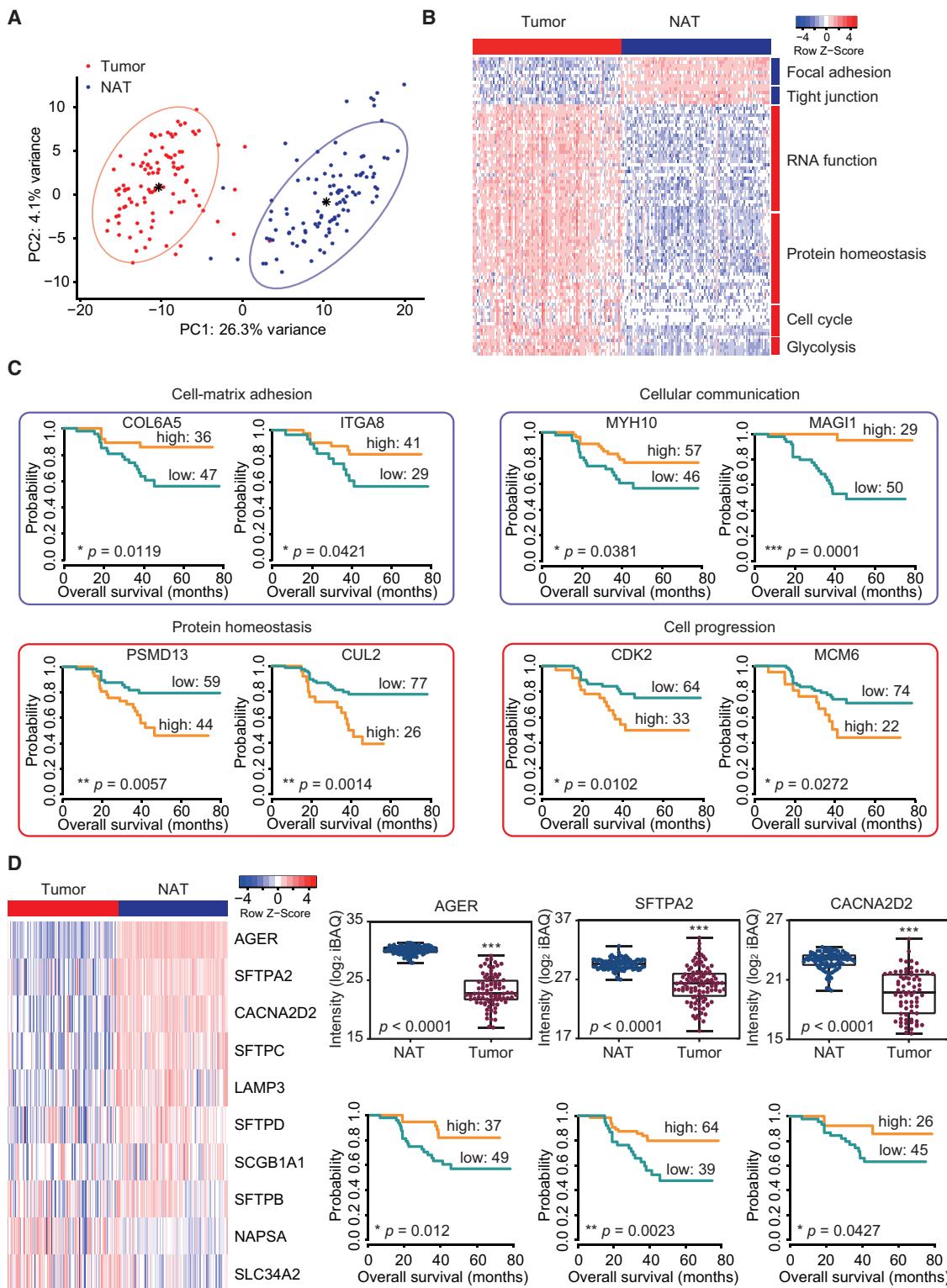
(B) Overlap of SNVs at DNA, mRNA, and protein levels. Blue: DNA variants (from WES); red: RNA variants (from RNA-seq); yellow: protein variants (from LC-MS/MS).

(C) Novel RNA splicing and fusion transcripts validated by LC-MS/MS. Red, novel RNA splicing (from RNA-seq); yellow, novel alternative splice variants (from LC-MS/MS); light blue, fusion transcripts; purple, fusion proteins. Three types of alternative splicing events were explored.

See also Figure S2 and Tables S2, S3, and S6.

Showed that pathways related to energy metabolism, such as fatty acid metabolism and oxidative phosphorylation, were enriched in the DG group (Figure 4A; Figure S4B). On the contrary, epithelial-mesenchymal transition (EMT) as well as other inflam-

matory and oncogenic signaling pathways, such as tumor necrosis factor alpha (TNF- α), mTOR, and MYC, were enriched in the DP group (Figure 4A; Figure S4B). The LUAD histological marker TTF-1 (NKX2-1; Wilcoxon rank-sum test, $p = 0.0287$, fold

**Figure 3. Proteomics Features of Tumors and NATs in LUAD**

(A) PCA of 8,252 proteins in 103 paired LUAD patients. Red, tumors; blue, NATs.

(B) Differentially expressed proteins in tumors and NATs and their associated biological pathways.

(C) Representative proteins from one of the four biological pathways and their association with prognosis (p value from log rank test).

(legend continued on next page)

change = 1.9) was expressed higher in the DG group (Figure S4C). MAOA and ACSS1 in fatty acid metabolism and ANK1 in heme metabolism were also expressed higher in the DG group (Wilcoxon rank-sum test, $p < 0.05$ and fold change > 2) and were positively associated with good prognosis, whereas THBS1 and PLOD2 in EMT process and IGFBP3 in glycolysis were expressed higher in the DP group ($p < 0.05$, fold change > 2) and were negatively associated with a good prognosis (Figure 4B; Figure S4D). To investigate the proteins that may account for the poor prognosis, we evaluated the genetic vulnerabilities of the upregulated proteins in the DP group using data from the Cancer Dependency Map Project (DepMap). Average gene essentiality scores (CRISPR-Cas9 gene knockout scores [CERES]) that reflect gene dependence were calculated in 36 cell lines with LUAD origin, and the essential genes below a threshold of a -0.6 CERES score (Meyers et al., 2017) were retained. TTF2 showed a low average CERES score (-0.74) (Figure S4E).

Proteomics Characteristics of LUADs with EGFR and TP53 Mutations

To investigate how mutation of a single driver gene affects the proteome and the related pathways, we examined significantly altered proteins in patients with or without LUAD driver mutations (Wilcoxon rank-sum test, BH $p < 0.05$, fold change > 1.5 , identified in more than 75% of patient samples). In total, we identified four driver genes that are significantly positively or negatively correlated with 245 protein expressions (Table S4). The majority of these proteins were correlated with *EGFR* or *TP53* mutation, the two predominant drivers in our cohort. Previous studies have focused on the prognostic prediction of *EGFR* and *TP53* mutation in LUAD (Jiao et al., 2018; Labb   et al., 2017; Shepherd et al., 2017; Yu et al., 2018). To further establish a connection between genetic alterations and their downstream pathways, we explored the proteomics profiles in patients with or without mutations of *EGFR* and/or *TP53*. Several oncogenic pathways, including DNA replication, mismatch repair, and the spliceosome, changed noticeably in patients with *TP53* mutations (Figure S4F). Proteins in patients with *EGFR* mutations were elevated in pathways such as the spliceosome and propanoate metabolism (Figure S4F). TTF-1 (NKX2-1) and Napsin-A (NAPSA), two histological markers of LUAD (Travis et al., 2011), showed higher expressions in tumor samples with *EGFR* mutation (TTF-1, $p = 0.0011$; NAPSA, $p = 0.0086$) (Figure 4C). In contrast, NAPSA showed lower expression in tumor samples with *TP53* mutation ($p = 0.0266$), but no obvious difference was observed for TTF-1 in patients with *TP53* mutation (Figure 4C). Poor prognostic outcomes were observed in patients with co-mutation of *EGFR* and *TP53* ($n = 27$) (Figure 4D). These patients had downregulated metabolic pathways, including oxidative phosphorylation, fatty acid metabolism, and propanoate metabolism, and several upregulated oncogenic pathways, including DNA replication, ubiquitin-mediated proteolysis, proteasome, mismatch repair, and cell cycle (Figure 4D). 19DEL (exon 19 deletion) and L858R are two common mutations of *EGFR*. A closer as-

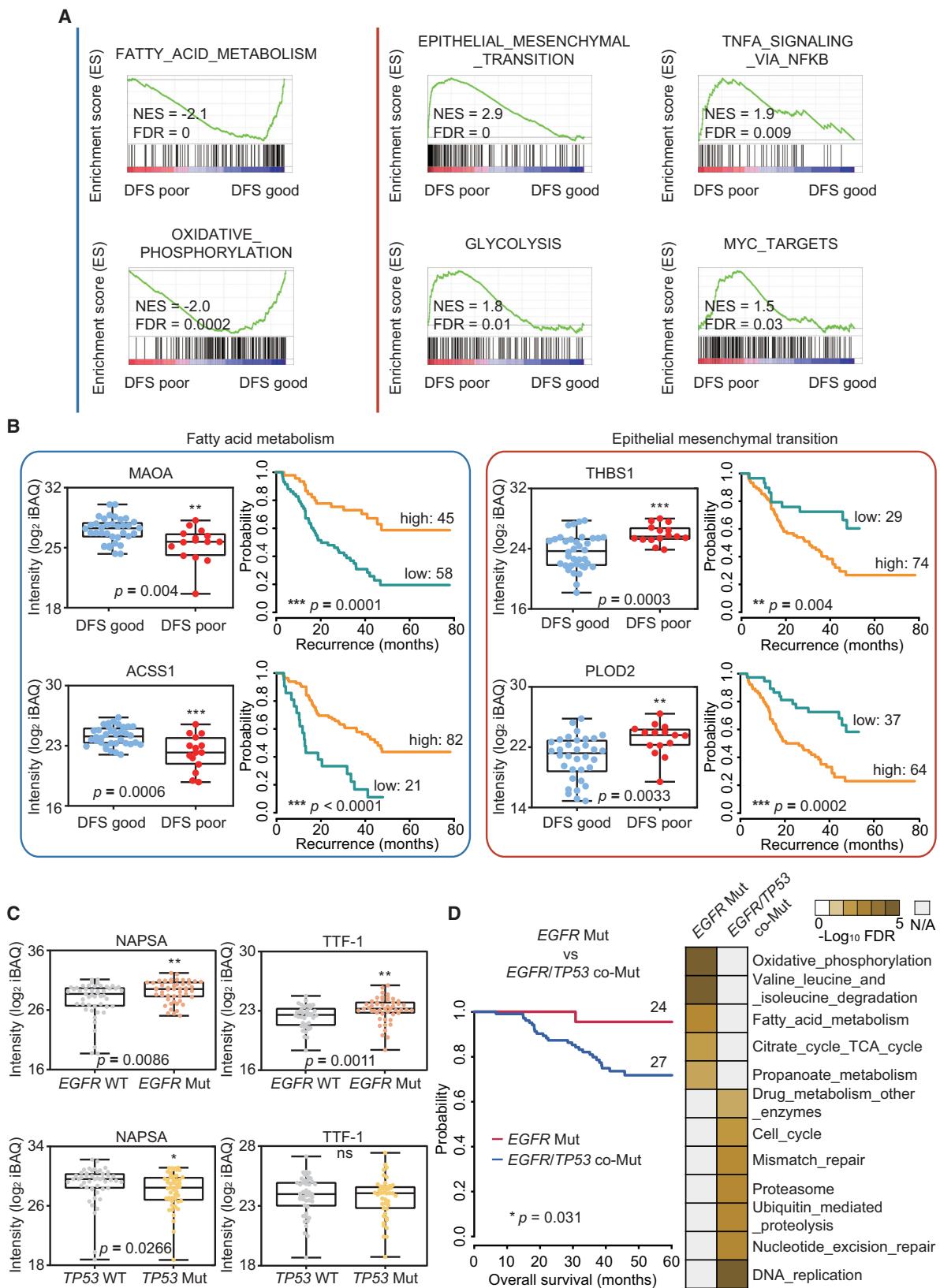
sociation between cell replication and nuclear replication events with *EGFR*-19DEL and micro-environment and protein biosynthesis with *EGFR*-L858R was observed (Figure S4G). Thus, our study provides new insights into the proteomics characteristics of patients with different mutational statuses of *EGFR* and *TP53*.

Proteomic Subtypes of LUAD and Their Association with Clinical Outcomes

Consensus clustering identified three proteomic subtypes based on the 1,567 most variable proteins (proteins with the top 25% standard deviations) that were present in more than 70% of tumor samples (Figure S5A). They were designated S-I (environment and metabolism high [EM-H], $n = 21$), S-II (mixed type, $n = 48$), and S-III (proliferation and proteasome [PP], $n = 34$) with distinct molecular and clinical features (Figures 5A–5E; Table S4). To examine whether different cell types are associated with the proteomics clusters, tumor purity and non-tumor cells, including stroma, immune cells, normal tissue, and necrotic cells, were assessed by hematoxylin and eosin (H&E) staining. Neither tumor purity nor non-tumor cells showed an association with the proteomics clusters in a Kruskal-Wallis test (Figures S5B and S5C) or a linear regression model (Table S4). Notably, patients in S-I had the best clinical outcomes, whereas patients in S-III had the poorest prognosis (log rank test, $p < 0.05$) (Figure 5B; Table S4). Further analysis within stage I patients among the three clusters showed that proteomics subtype was significantly correlated with survival (log rank test, $p < 0.05$), indicating the prognostic power of proteomics clustering (Figure S5D). Consistent with current clinical knowledge, the early-stage (Fisher's exact test, $p = 0.006$) and lepidic pathological subtypes ($p = 0.006$) were enriched in patients in S-I, whereas patients in S-III had a lower degree of differentiation ($p = 0.0001$), more solid pathological subtypes ($p = 0.0002$), and a higher level of tumor stage ($p = 0.006$) (Figure 5C). Genomic information showed that S-II had the highest mutation frequency of *EGFR* (Fisher's exact test, $p = 0.011$), whereas S-III had the highest tumor mutational burden (Kruskal-Wallis test, $p = 0.049$). The two histological markers TTF-1 and NAPSA had the lowest expression in S-III patients (Kruskal-Wallis test; TTF-1, $p = 0.003$; NAPSA, $p < 0.0001$) (Figure 5C).

Mutational signature analysis of WES data identified nine main signatures in 103 patients. Signature 1 (aging-related) was a dominant signature identified in 82 patients (Figure 5D; Table S2). The other eight signatures were signature 6 (defective mismatch repair, $n = 46$), signature 15 (defective mismatch repair, $n = 43$), signature 24 (exposure to aflatoxin, $n = 41$), signature 3 (failure of DNA double-strand break repair, $n = 35$), signature 7 (ultraviolet light exposure, $n = 34$), signature 2 (APOBEC, $n = 30$), signature 4 (smoking, $n = 29$), and signature 13 (APOBEC, $n = 27$) (Figure 5D). Five signatures were significantly associated with smoking status, gender, status of *EGFR*, or *KRAS* mutation (Figure S5E; Wilcoxon rank-sum test, $p < 0.05$). The most dominant signature, signature 1, was mainly observed in non-smokers ($p = 0.0019$) and females ($p = 0.0039$). The smoking

(D) A list of lung signature proteins that were differentially expressed in tumors and NATs (p value from Wilcoxon signed-rank test) (left). Three proteins (AGER, SFTPA2, and CACNA2D2) were significantly associated with prognosis (p value from log rank test) (right). See also Figure S3 and Table S4.



(legend on next page)

signature (signature 4) was mostly found in smokers ($p < 0.0001$), males ($p < 0.0001$), patients with *KRAS* mutation ($p = 0.0022$) or *EGFR* wild type ($p = 0.0002$). These results are consistent with the current understanding of LUAD. Among the three subtypes, S-III has the lowest level of signature 1 (Kruskal-Wallis test, $p = 0.0357$; Figure S5F).

GSEA showed that S-I has higher environment and energy metabolism, including ECM receptor interaction, fatty acid metabolism, and amino acid metabolism (Figure 5E); some of them were ventilation- and immune response-related pulmonary function pathways. S-III is more related to protein homeostasis and tumor proliferation, including the ribosome, proteasome, proteolysis, DNA replication, and complement and coagulation cascades (Figure 5E). S-II has the intermediate features of S-I and S-III. The significantly altered proteins in the representative pathways of the three subtypes had roles in focal adhesion and the cell cycle (Figure 5F; Table S4). Several lung signature proteins were highly expressed in S-I but lost in S-III (Kruskal-Wallis test, $p < 0.05$; Figure S5G). Together, it suggests that S-I is more related to normal tissue, whereas S-III has more malignant characteristics. As a complementary approach, weighted gene co-expression network analysis (WGCNA) was performed and identified seven co-expression modules (Figure S5H). These modules were significantly associated with the proteomic subtypes, *TP53* mutation status, differentiation, pathology, and TNM stages (Figure S5H).

We also conducted clustering analyses on the tumor phosphoproteome ($n = 79$, consensus clustering) and transcriptome ($n = 51$, non-negative matrix factorization [NMF]), and identified three subtypes in each dataset (Figure 5A). Generally, a moderate or even poor concordance among the transcriptomics, proteomics, and phosphoproteomic subtypes was revealed (38.0% between proteomics and phosphoproteomic subtypes and 41.2% between proteomics and transcriptomics subtypes). Phosphoproteomic subtypes were associated with overall survival ($p < 0.05$, log rank test). However, transcriptomics subtypes had no association with prognosis, possibly because of the smaller sample size ($n = 51$) (Figure S5I). One of the phosphoproteomic subtypes with poor overall survival (pS-III) agreed well with proteomics subtype S-III. Proteomics subtype S-I had relatively good concordance with its corresponding mRNA-based subtype (tS-I).

Cancer immunotherapy has achieved a remarkable success in certain groups of NSCLC patients (Herbst et al., 2018). No significant difference was observed in the three proteomics clusters based on the immunohistochemistry (IHC) scoring of CD8 T cells (Fisher's exact test, $p = 0.4474$; Table S4). The result

was consistent with analysis of CD8A expression in the three proteomic subtypes (Kruskal-Wallis test, $p = 0.7358$; Figure S5J). Further analyses of 770 immune-related proteins defined by the nCounter PanCancer Immune Profiling Panel (Cesano, 2015) identified 136 differentially expressed proteins across the proteomic subtypes (Kruskal-Wallis test, BH $p < 0.05$; Table S4). Notably, several human leukocyte antigen (HLA) proteins from major histocompatibility complex (MHC) class II, which are normally found on professional antigen-presenting cells (APCs) (Axelrod et al., 2019), had the lowest expression in S-III, whereas the HLA proteins under MHC class I had no obvious difference in the three subtypes (Figure 5G; Figure S5K). To evaluate whether MHC-II was expressed in cancer cells or in tumor-infiltrating lymphocytes (TILs), we examined HLA-DR expression in 103 tumor tissues by IHC. The presence of HLA-DR was confirmed in cancer cells and APCs (Figure S5L). No significant difference in HLA-DR across the three proteomic subtypes was found in TILs, based on the IHC scoring of HLA-DR (Fisher's exact test, $p = 0.7176$), whereas the lowest expression of HLA-DR in cancer cells was found in S-III (Fisher's exact test, $p = 0.0035$) (Table S4). In contrast, proteins in the complement and coagulation cascades were significantly higher in S-III than in the other two subtypes (Figure 5G).

Phosphoproteomics Features and Their Clinical Relevance

Among 1,244 phosphoproteins quantified in at least 50% of patient samples, 474 (38%) were significantly increased and 183 (15%) were significantly decreased in tumors compared with paired NATs (Wilcoxon signed-rank test, BH $p < 0.01$; Figure 6A). Of these, 97 and 44 phosphoproteins had a more than 2-fold increase or decrease, respectively (Table S5). The three most differentially expressed phosphoproteins are CAV1, SDPR, and CAV2 in NATs and HSP90AB1, GFPT1, and CLIC6 in tumors (Figure S6A). GSEA indicated that phosphoproteins in tumors were enriched in oncogenicity-associated pathways such as mTORC1 signaling and glycolysis, whereas those in NATs were enriched in physiological functions like myogenesis and apical junctions (Figure S6B). Kinase substrate enrichment analysis (KSEA) of the phosphoproteome between tumors and NATs identified multiple kinases, including CSNK2A1, MAPKAPK2, CDK1 or 2, and SGK1, that were activated in tumors (Figure S6C; STAR Methods). Comparison of the phosphoproteome among the three proteomic subtypes revealed subtype-specific activated kinases, including AKT1 for S-I, PRKCE for S-II, and AURKB, PRKCA, and MAPK14 for S-III (Figure 6B). Association analysis of differentially expressed phosphoproteins (97 with

Figure 4. Proteomics Characteristics of TNM Stage I LUAD and LUADs with EGFR/TP53 Mutation

- (A) GSEA (H, hallmark gene sets) of TNM stage I LUAD patients revealed pathways associated with a good ($n = 36$, DFS time > 3 years) or poor ($n = 15$, DFS time < 3 years) prognosis (p value from log rank test).
- (B) Proteins differentially expressed in the good or poor prognosis groups in stage I patients (p value from Wilcoxon rank-sum test) and their association with clinical outcomes in all patients (p value from log rank test).
- (C) Expression of two markers (TTF-1 and NAPSA) for LUAD pathological diagnosis in *EGFR/TP53* mutant and wild-type LUADs (p value from Wilcoxon rank-sum test).
- (D) Survival analysis of patients with *EGFR* mutation versus co-mutation of *EGFR* and *TP53* (p value from log-rank test). Enriched pathways revealed by GSEA (C2: curated gene sets, KEGG subset of canonical pathways) in patients with *EGFR* mutation or co-mutation of *EGFR* and *TP53*.
- See also Figure S4 and Tables S1 and S4.

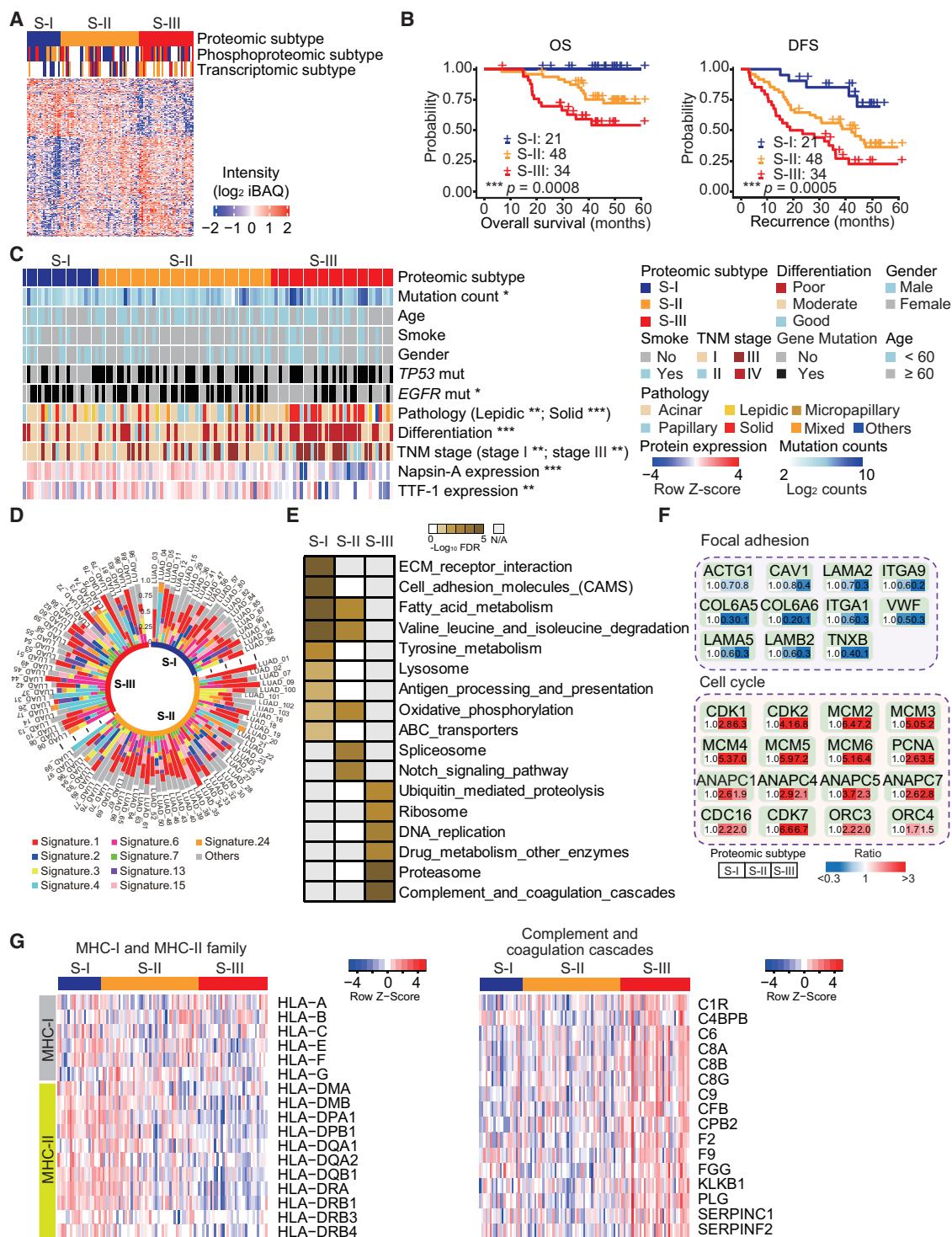


Figure 5. Molecular Subtyping of LUAD and Associations with Clinical Outcomes

(A) The relative abundance of 842 proteins in the three proteomic subtypes: S-I (environment and metabolism high [EM-H]), S-II (mixed), and S-III (proliferation and proteasome [PP]). Above the heatmap: comparison of proteomics ($n = 103$), transcriptomics ($n = 51$), and phosphoproteomic subtypes ($n = 79$).
(B) The association of three proteomic subtypes with clinical outcomes in 103 LUAD patients (p value from log rank test). Left panel: overall survival (OS). Right panel: disease-free survival (DFS).
(C) The association of proteomic subtypes with 11 variables. Fisher's exact test was used for categorical variables: age, gender, smoking status, status of EGFR/TP53 mutation, pathological subtypes, differentiation, and TNM stage; Kruskal-Wallis test was used for continuous variables: protein expression of TTF-1 and NAPSA and mutation counts.

(legend continued on next page)

T/N > 2, 44 with T/N < 0.5) revealed 48 prognosis-associated phosphoproteins (log rank test, $p < 0.05$) (Table S5). Among them, differential phosphorylation of CTNNB1, SPTBN1, and SNTB1 was found across the three proteomic subtypes (Kruskal-Wallis H test, FDR < 0.05) and was correlated with prognosis (Figure 6C). Moreover, there were high correlations between changes in the proteome and phosphoproteome (T/N ratios, Pearson's $r = 0.67$, $p = 2.2 \times 10^{-16}$; Figure S6D). Among them, the changes in phosphorylation of 19 altered protein (T/N ratio > 2) were greater than the changes in their corresponding protein abundance (Table S5). We defined them as cancer-associated phosphoproteins (Figure S6D, red dots). Further investigation into the differentially changed phosphosites showed that elevated substrates involved in focal adhesion and reduced substrates involved in RNA transport were observed in S-I (Figure 6D). Consistently, S23 phosphorylation in CAV2 and S2 phosphorylation in EIF2S2 were correlated with overall survival (log rank test, $p < 0.05$) (Figure 6D).

To pinpoint the effects of these cancer-associated phosphoproteins on the oncogenic signaling pathways (Sanchez-Vega et al., 2018), we built a phospho-regulatory network by calculating the correlation with the expression levels of the protein members in these oncogenic pathways. As shown in Figure 6E, multiple proteins may be regulated by cancer-associated phosphorylation signals (Spearman correlation coefficient, $r \geq 0.4$ or ≤ -0.4 , FDR < 0.05). Collectively, our phosphoproteomics data provide additional insight into the molecular features of LUAD.

Potential Prognostic Biomarkers and Drug Targets in LUAD

In clinical practice, plasma or serum is the most widely used specimen for biomarker discovery because proteins in the circulatory system likely reflect disease pathophysiology. Our dataset provides a good opportunity to identify potential prognostic biomarkers of LUAD. We hypothesized that the most highly and commonly expressed proteins with prognostic power in tumor could be circulated and detected in blood and could be used as potential biomarkers. With the four criteria described (STAR Methods), 27 proteins, including HSP 90 β (HSP90AB1), met the criteria (Table S7).

Because a potential association between plasma HSP 90 β and pathological grade and clinical stage in lung cancer has been observed previously (Rong et al., 2014), we further investigated whether HSP 90 β could be a prognostic biomarker for LUAD. Our proteomic results showed that HSP 90 β was expressed much higher in tumors than in NATs (Wilcoxon signed-rank test, $p < 0.05$, T/N ratio = 4.6) (Figure 7A), and the results were further validated by western blotting (Figure S7A). High expression of HSP 90 β was associated with a poor prognosis (log rank test, $p < 0.05$) (Figure 7A) and was also observed in patients with co-mutation of EGFR and TP53 but not in those with EGFR mutation alone (Wilcoxon rank-sum test, $p < 0.05$) (Fig-

ure S7B). In addition, the lowest protein expression was observed in S-I, with the best prognosis (Wilcoxon rank sum test, $p < 0.05$) (Figure S7C). Notably, two phosphosites of HSP 90 β (S226 and S255; Figure S7D) were higher in tumors than in NATs (Wilcoxon signed-rank test, $p < 0.05$; pS226, T/N ratio = 2.6; pS255, T/N ratio = 4.9) (Figures S7E–S7G) and were correlated with overall survival (log-rank test, $p < 0.05$) (Figures S7E and S7F). To confirm the potential utility of HSP 90 β as a prognostic biomarker of LUAD, we examined the plasma protein level of HSP 90 β using ELISA in an independent cohort of 705 LUAD patients and 282 healthy controls. The result confirmed the higher levels of HSP 90 β in LUAD patients (Wilcoxon rank-sum test, $p < 0.05$) (Figure 7B; Table S1). Analysis of 499 LUAD patients indicated that HSP 90 β was negatively correlated with prognosis (log-rank test, $p < 0.05$) (Figure 7C). The efficacy of tanesipimycin, a small-molecule inhibitor of HSP90 (Schenk et al., 2013), was significantly higher in LUAD cell lines (median half maximal inhibitory concentration (IC_{50}) = 225 nM, $n = 66$) and lung squamous cell carcinoma (LUSC) cell lines (median IC_{50} = 154 nM, $n = 15$) than in SCLC cell lines (median IC_{50} = 3 μ M, $n = 52$) in the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al., 2013; Figure 7D). Together, these data confirmed HSP 90 β as a potential biomarker of LUAD.

Recently, precision medicine has achieved great success in targeting cancers by selectively disrupting cancer driver pathways. To search for potential drug targets in LUAD, we considered druggability, prognostic power, gene dependency, and T/N ratio (STAR Methods) and identified 11 candidates (Table S7), including CARS, MCM2, GAPDH, GMPS, CTPS1, TPI1, COASY, NUP107, IMPDH2, PCYT1A, and DICER1. Two (IMPDH2 and GAPDH) are US Food and Drug Administration (FDA)-approved drug targets (Figure 7E and 7F). Four (GAPDH, IMPDH2, MCM2, and CARS) were associated with a poor prognosis (log rank test, $p < 0.01$) in 51 patients based on RNA-seq data. Similar approach applied to LUAD with EGFR mutations led to the identification of 10 additional potential drug targets (STAR Methods), including PSMB2, CARS, MCM5, MCM2, IARS, POLD1, CHD4, CTPS1, GMPS, and RARS (Table S7).

Identification of cancer-related variant proteins may provide insights into new targetable mutations. Integrative omics analyses (STAR Methods) identified 29 protein variants (corresponding to 31 sites) that may be targetable mutations (Table S7), with 25 mutations having a meta-score higher than 0.57 (a cutoff score to define the pathogenesis of deleterious single-amino-acid polymorphisms; STAR Methods). The efficacy and toxicity of these proposed druggable candidates or mutations should be evaluated further.

DISCUSSION

Despite the remarkable progress in precision oncology by genomic profiling of LUAD, the existing strategy targeting

(D) Distributions of nine main COSMIC signatures in the proteomic subtypes across 103 LUAD patients.

(E) GSEA (C2: curated gene sets, KEGG subset of canonical pathways) revealed the pathways that were significantly enriched in the proteomic subtypes.

(F) Proteins in two pathways (focal adhesion and cell cycle) that were differentially expressed in the three proteomic subtypes.

(G) Expression of MHC class I and II proteins (left) and proteins of complement-related pathways (right) in the three proteomic subtypes.

See also Figure S5 and Tables S2 and S4.

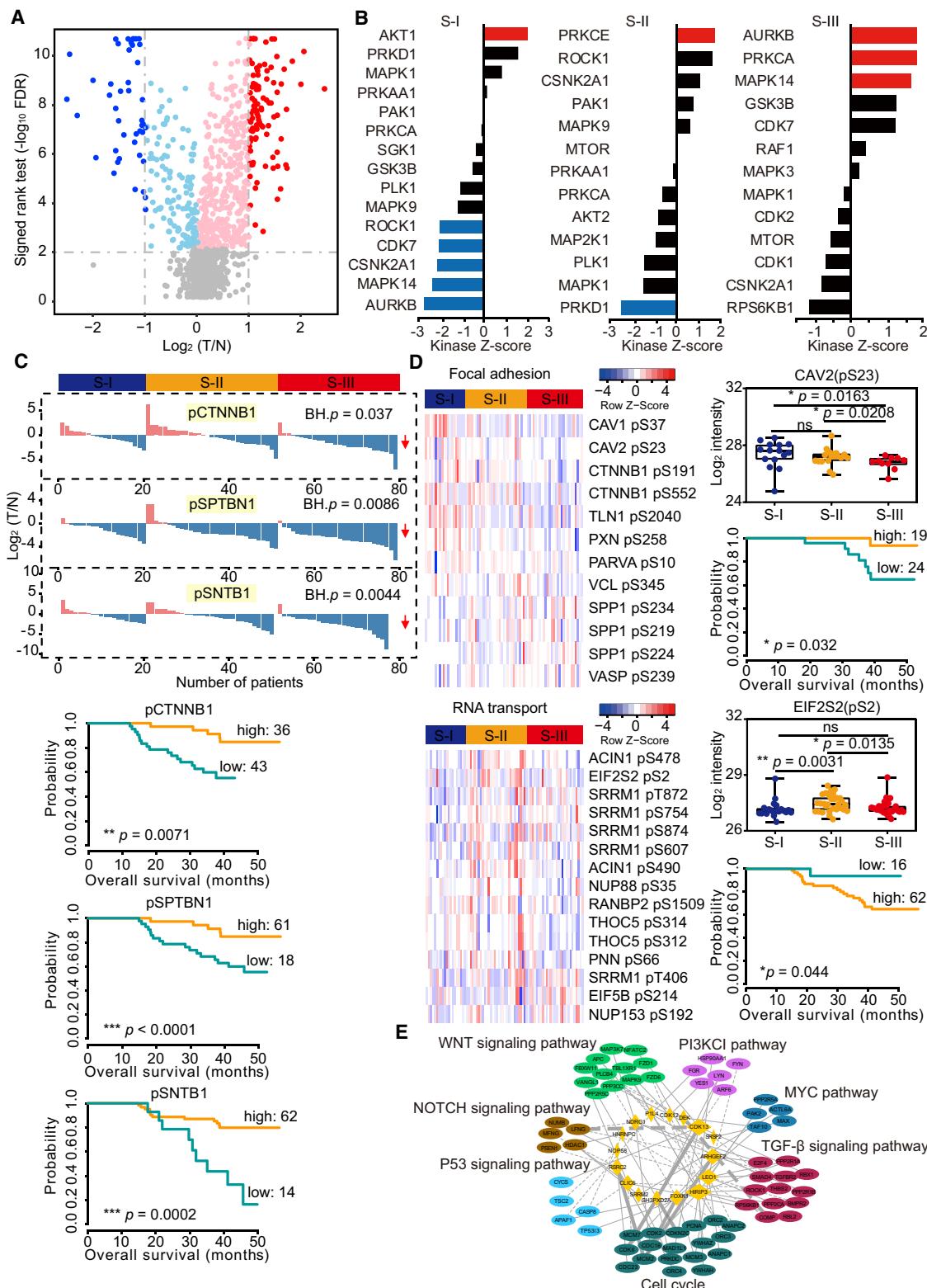


Figure 6. Phosphoproteomics Profiles of LUAD and Their Association with Clinical Outcomes

(A) Expression of phosphoproteins in tumors or NATs. Dark red and dark blue colors, upregulated and downregulated phosphoproteins in tumors, respectively (FDR < 0.01 from Wilcoxon signed-rank test and fold change > 2).

(legend continued on next page)

genetic abnormalities has limitations. A more comprehensive understanding of LUAD based on proteomics can fill the gap between genome abnormalities and oncogenic protein machinery. Compared with the NAT proteome, LUAD-related cancer pathways were identified. Of note, some lung signature proteins were less expressed in tumors than in NATs. These proteins are expressed specifically in pneumocytes, necessary for surfactant maintenance and normal respiration or epithelial regeneration after oxidant-induced injury (Lindskog et al., 2014). Loss of these proteins in tumors may suggest cellular atypia because LUAD starts in epithelial cells. Our proteomics analysis of stage I LUAD revealed that EMT as well as inflammatory and oncogenic signaling pathways (TNF- α , mTOR, and MYC) likely lead to poor prognosis. The phosphoproteomics analysis further revealed activated signaling pathways and kinases in LUAD. Our proteomic analysis of Chinese LUAD patients has added another omics layer and provides complementary insights into LUAD beyond the current genomic understanding.

A data-driven method identified three new proteomic subtypes that have distinct molecular features connecting the clinical, pathological, and prognostic features of LUAD. Among the most striking proteomics features, cellular environment and energy metabolism are associated with a good prognosis, whereas cell proliferation and proteome stability are associated with a poor prognosis. S-III presented the worst prognosis, highest tumor mutation burden (TMB), and lowest levels of MHC class II, which was consistent with previous findings that high TMB was associated with poor clinical outcomes patients of East Asian ancestry (Chen et al., 2020; Owada-Ozaki et al., 2018), whereas positive expression of MHCs suggested a good prognosis in various types of cancers (He et al., 2017; Simpson et al., 2010; Takeuchi et al., 2019). A possible explanation is that the frequent occurrence of allele-specific MHC loss in NSCLC is associated with a high subclonal neoantigen burden, and could serve as an immune escape mechanism subject to strong selection pressure in tumor evolution (McGranahan et al., 2017).

Largely consistent with recent genomic studies (Chen et al., 2020; Wang et al., 2018b), our genomic analysis showed that genetic mutations in *EGFR* and *TP53* are predominant in Chinese LUAD patients. In agreement with previous research (Cardnell et al., 2015; Lee et al., 2012), two diagnostic biomarkers, TTF-1 (NKX2-1) and NAPSA, were positively correlated with *EGFR* mutation (Figure 4C). TTF-1 regulates RTK ROR1 activity to maintain EGFR signalling in LUAD via two possibly distinct mechanisms: (1) ROR1 kinase activity-dependent c-Src-mediated signaling and (2) ROR1 kinase activity-independent sustainment of EGFR-ERBB3-phosphatidylinositol 3-kinase (PI3K) signaling

(B) Evaluation of kinase activities by KSEA in tumors across the three proteomic subtypes.

(C) Identification of three differentially expressed phosphoproteins in tumors and NATs across the three proteomic subtypes (top panel) and the association of three phosphoproteins with OS (bottom panel, p value from log-rank test).

(D) A list of phosphosites involved in the two pathways and their abundance in the three proteomic subtypes and the association of two phosphosites with OS (p value from log-rank test).

(E) Phospho-regulatory network in LUAD. Shown is the network linking cancer-related phosphoproteins (yellow diamonds) and proteins (ellipses) in oncogenic signaling pathways. The size of the diamonds indicates the degree of correlation. The ellipses in different colors represent the proteins in different oncogenic signaling pathways. Solid and dashed lines indicate positive or negative correlations between phosphoproteins and proteins, respectively. Bold lines indicate higher correlations ($r > 0.7$ or $r < -0.7$).

See also Figure S6 and Table S5.

(Yamaguchi et al., 2012). Similarly, we observed higher mRNA expression of ROR1 in *EGFR* mutant patients (Wilcoxon rank-sum test, $p = 0.0213$). However, the exact mechanism of how *EGFR* mutation regulates TTF-1 expression remains to be investigated. Compared with the TCGA study in patients from Western countries (Collisson et al., 2014), the main differences are the higher frequency of *EGFR* mutations (50% versus 14%) and the lower frequency of *KRAS* mutations (6% versus 33%) in our cohort, indicating a race-biased genetic profile. It is worth mentioning, in contrast to a recent large-cohort study in LUAD patients of Western countries who showed no prognostic effect of *TP53/EGFR* co-mutation (Shepherd et al., 2017), that patients in our cohort with *TP53/EGFR* co-mutations often had poorer prognoses than those who had the *EGFR* mutation alone. Taken together, our analyses suggest that race is one of the key factors that should be evaluated in precision oncology in LUAD.

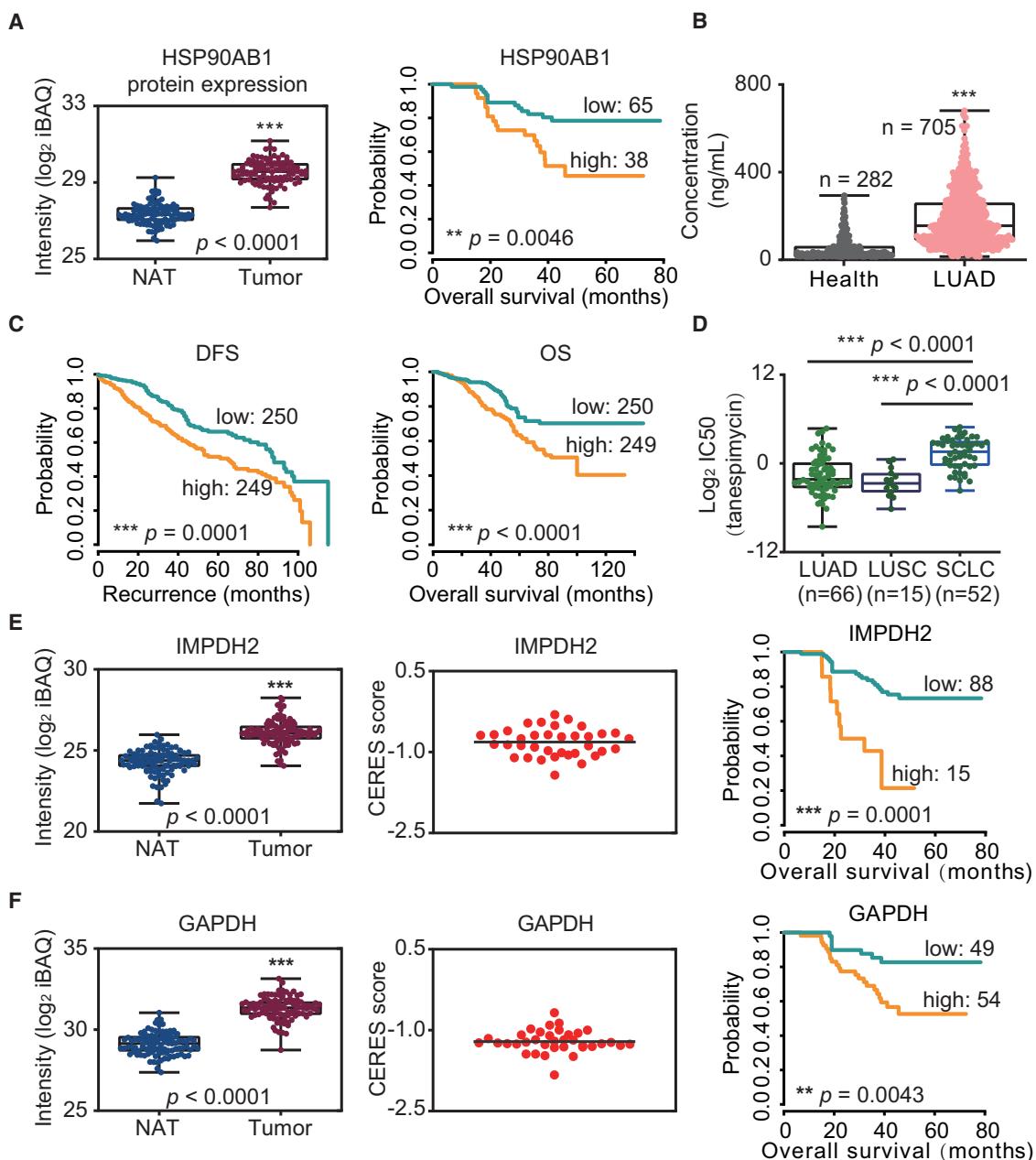
The proteomics data confirmed HSP 90 β as a prognostic marker of LUAD and identified a set of proteins as potential drug targets. Many of them are known to be associated with cancer development. Notably, IMPDH2 is a target of the immunosuppressant mycophenolic acid. Recent studies have shown that IMPDH2 plays important roles in promoting several types of cancer, including small-cell lung cancer (Huang et al., 2018) and glioblastoma (Kofuji et al., 2019). Targeting GAPDH, a key enzyme in the Warburg effect in cancers, has shown potential therapeutic efficacies in several types of cancers (Liberti et al., 2017; Yun et al., 2015). PSMB2, a potential target for *EGFR* mutant cancer, is the target of bortezomib and carfilzomib, two proteasome inhibitors for treatment of multiple myeloma. Previous studies showed that D148E mutation in APEX1, a DNA repair gene, could decrease the risk of cutaneous melanoma (Li et al., 2006) and was more sensitive to APEX1 inhibitor in a melanoma cancer cell line (Mohammed et al., 2011). These results suggest potential therapeutic benefits of these candidates in LUAD.

In summary, multi-omics integrative analysis is a valuable and powerful tool that provides a complementary and more comprehensive understanding of LUAD and offers an opportunity to expedite translation of basic research to more precise diagnosis and treatment in the clinic.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY

**Figure 7. Potential Prognostic Biomarkers or Drug Targets**

(A) Protein expression of HSP 90 β in tumors or NATs (p value from Wilcoxon signed-rank test) and associations with OS (p value from log-rank test).
 (B) Measurement of the plasma protein level of HSP 90 β in an independent cohort of 705 LUAD patients and 282 healthy individuals (p value from Wilcoxon rank-sum test).
 (C) Survival analyses of LUAD patients with high or low levels of plasma HSP 90 β . Left panel, DFS; right panel, OS (p value from log rank test).
 (D) Drug sensitivity of tanespimycin in LUAD, lung squamous cell carcinoma (LUSC), and small-cell lung cancer (SCLC) cell lines (from the GDSC database).
 (E and F) IMPDH2 (E) and GAPDH (F) as potential drug targets. Left panel: protein abundance in tumors and NATs. Center panel: CERES scores of IMPDH2 and GAPDH in 36 cell lines in the DepMap database (the black line illustrates the mean CERES score). Right panel: association of protein abundance with OS in 103 patients.

See also Figure S7 and Tables S1 and S7.

- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Patient samples of LUADs
 - Independent cohort

- **METHOD DETAILS**
 - Proteomic and phosphoproteomic analysis
 - Whole-exome sequencing

- RNA-Seq
 - Enzyme-linked immunosorbent assays (ELISA)
 - Western blot assay and antibodies
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Quality control and assessment of LC-MS/MS data
 - Differential protein analysis
 - Pathway enrichment analysis
 - Consensus clustering analyses
 - Clustering analysis of transcriptomic data
 - WGCNA analysis
 - Correlation between subtype and clinical features
 - Phosphoproteomic data analysis
 - mRNA-protein correlation in tumors and paired NATs
 - Survival analysis
 - Potential prognostic biomarkers or drug targets
 - **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.05.043>.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program of China (973 Program) (2014CBA02004), the Natural Science Foundation of China (81821005, 91753203, 81872888, 31901036, 31871329, and 31700724), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA12050406), the CAMS Innovation Fund for Medical Sciences (CIFMS), China (2016-I2M-1-001), the National Key Research and Development Program of China (2017YFC0908401), the Shanghai Sailing Program (19YF1457700), the Innovative Research Team of High-Level Local Universities in Shanghai (SSMU-ZDCX20181202), the K. C. Wong Education Foundation, and the SA-SIBS Scholarship Program. We thank the pathologists, Dr. Huiqin Guo and Dr. Huan Zhao, for their professional pathological opinions.

AUTHOR CONTRIBUTIONS

M.T., T.X., F.H., J.L., and Yi Wang conceived and supervised the project. T.X., S.C., Yi Zhang, G.W., C.S., T.Z., X.Z., K.Q., Y. Mao, Y.Y., L.Y., Xijun Wang, Yaru Wang, and M.W. coordinated the acquisition, distribution, and quality evaluation of LUAD tumors and adjacent tissues. J.-Y.X., L. Zhai, S.J., H.X., P.L., Z.W., and B.L. directed and performed analyses and quality control of MS data. T.X., Xiang Wang, and M.W. performed ELISA. C.Z., J.-Y.X., Y. Ma, J.L., Z.L., Lei Zhao, B.W., X.Y., Yimin Zhang, Liyuan Zhao, Y.C., S.R., J.J., Y. Zhao, W.Y., L.F., and F.W. performed proteomics and genomics data analyses. C.Z., J.-Y.X., M.T., L. Zhai, J.L., F.H., A.S., J.Q., X.Q., and Y.W. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 29, 2019

Revised: April 8, 2020

Accepted: May 21, 2020

Published: July 9, 2020

REFERENCES

- Axelrod, M.L., Cook, R.S., Johnson, D.B., and Balko, J.M. (2019). Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. *Clin. Cancer Res.* 25, 2392–2402.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendt, M.C., Kim, J., Reardon, B., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 174, 1034–1035.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424.
- Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al.; Cancer Genome Atlas Research Network (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* 48, 607–616.
- Cardnell, R.J.G., Behrens, C., Diao, L., Fan, Y., Tang, X., Tong, P., Minna, J.D., Mills, G.B., Heymach, J.V., Wistuba, I.I., et al. (2015). An Integrated Molecular Analysis of Lung Adenocarcinomas Identifies Potential Therapeutic Targets among TTF1-Negative Tumors, Including DNA Repair Proteins and Nrf2. *Clin. Cancer Res.* 21, 3480–3491.
- Casado, P., Rodriguez-Prados, J.-C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S., and Cutillas, P.R. (2013). Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* 6, rs6.
- Cesano, A. (2015). nCounter® PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). *J. Immunother. Cancer* 3, 42.
- Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F., and Wong, K.-K. (2014). Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat. Rev. Cancer* 14, 535–546.
- Chen, J., Yang, H., Teo, A.S.M., Amer, L.B., Sherbaf, F.G., Tan, C.Q., Alvarez, J.J.S., Lu, B., Lim, J.Q., Takano, A., et al. (2020). Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 52, 177–186.
- Clarke, M.R., Landreneau, R.J., Finkelstein, S.D., Wu, T.T., Ohori, P., and Yousem, S.A. (1997). Extracellular matrix expression in metastasizing and nonmetastasizing adenocarcinomas of the lung. *Hum. Pathol.* 28, 54–59.
- Collisson, E.A., Campbell, J.D., Brooks, A.N., Berger, A.H., Lee, W., Chmielecki, J., Beer, D.G., Cope, L., Creighton, C.J., Danilova, L., et al.; Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Dang, C.V., Reddy, E.P., Shokat, K.M., and Soucek, L. (2017). Drugging the ‘undruggable’ cancer targets. *Nat. Rev. Cancer* 17, 502–508.
- Devarakonda, S., Morgensztern, D., and Govindan, R. (2015). Genomic alterations in lung adenocarcinoma. *Lancet Oncol.* 16, e342–e351.
- Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. (2011). Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.* 39, D261–D267.
- Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11, 367.
- Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., et al. (2018). A proteomic landscape of diffuse-type gastric cancer. *Nat. Commun.* 9, 1012.
- Gould, M.K., James, F., Iannettoni, M.D., Lynch, W.R., Midthun, D.E., Naidich, D.P., and Ost, D.E.; American College of Chest Physicians (2013). Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines. *Chest* 132, 108S–130S.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.
- He, Y., Rozeboom, L., Rivard, C.J., Ellison, K., Dziadziuszko, R., Yu, H., Zhou, C., and Hirsch, F.R. (2017). MHC class II expression in lung cancer. *Lung Cancer* 112, 75–80.

- Herbst, R.S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553, 446–454.
- Hirsch, F.R., Varella-Garcia, M., Bunn, P.A., Jr., Di Maria, M.V., Veve, R., Bremmes, R.M., Barón, A.E., Zeng, C., and Franklin, W.A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J. Clin. Oncol.* 21, 3798–3807.
- Hirsch, F.R., Scagliotti, G.V., Mulshine, J.L., Kwon, R., Jr., Curran, W.J., Jr., Wu, Y.L., and Paz-Ares, L. (2017). Lung cancer: current therapies and new targeted treatments. *Lancet* 389, 299–311.
- Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E., and Zhang, B. (2004). PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 4, 1551–1561.
- Huang, F., Ni, M., Chalishazar, M.D., Huffman, K.E., Kim, J., Cai, L., Shi, X., Cai, F., Zacharias, L.G., Ireland, A.S., et al. (2018). Inosine Monophosphate Dehydrogenase Dependence in a Subset of Small Cell Lung Cancers. *Cell Metab.* 28, 369–382.e5.
- Hughes, P.E., Caenepeel, S., and Wu, L.C. (2016). Targeted Therapy and Checkpoint Immunotherapy Combinations for the Treatment of Cancer. *Trends Immunol.* 37, 462–476.
- Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., et al.; Chinese Human Proteome Project (CNHPP) Consortium (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 567, 257–261.
- Jiao, X.D., Qin, B.D., You, P., Cai, J., and Zang, Y.S. (2018). The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base. *Lung Cancer* 123, 70–75.
- Kim, D., and Salzberg, S.L. (2011). TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* 12, R72.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kofuji, S., Hirayama, A., Eberhardt, A.O., Kawaguchi, R., Sugiura, Y., Sampteeran, O., Ikeda, Y., Warren, M., Sakamoto, N., Kitahara, S., et al. (2019). IMP dehydrogenase-2 drives aberrant nucleolar activity and promotes tumorigenesis in glioblastoma. *Nat. Cell Biol.* 21, 1003–1014.
- Labbé, C., Cabanero, M., Korpany, G.J., Tomasini, P., Doherty, M.K., Mascaux, C., Jao, K., Pitcher, B., Wang, R., Pintilie, M., et al. (2017). Prognostic and predictive effects of TP53 co-mutation in patients with EGFR-mutated non-small cell lung cancer (NSCLC). *Lung Cancer* 111, 23–29.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lee, J.G., Kim, S., and Shim, H.S. (2012). Napsin A is an independent prognostic factor in surgically resected adenocarcinoma of the lung. *Lung Cancer* 77, 156–161.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, C., Liu, Z., Wang, L.E., Strom, S.S., Lee, J.E., Gershenwald, J.E., Ross, M.I., Mansfield, P.F., Cormier, J.N., Prieto, V.G., et al. (2006). Genetic variants of the ADPRT, XRCC1 and APE1 genes and risk of cutaneous melanoma. *Carcinogenesis* 27, 1894–1901.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liberti, M.V., Dai, Z., Wardell, S.E., Baccile, J.A., Liu, X., Gao, X., Baldi, R., Mehrmohamadi, M., Johnson, M.O., Madhukar, N.S., et al. (2017). A Predictive Model for Selective Targeting of the Warburg Effect through GAPDH Inhibition with a Natural Product. *Cell Metab.* 26, 648–659.e8.
- Lin, J.J., and Shaw, A.T. (2016). Resisting Resistance: Targeted Therapies in Lung Cancer. *Trends Cancer* 2, 350–364.
- Lindskog, C., Fagerberg, L., Hallström, B., Edlund, K., Hellwig, B., Rahnenführer, J., Kampf, C., Uhlén, M., Pontén, F., and Micke, P. (2014). The lung-specific proteome defined by integration of transcriptomics and antibody-based profiling. *FASEB J.* 28, 5184–5196.
- McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., and Swanton, C.; TRACERx Consortium (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171, 1259–1271.e11.
- Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* 49, 1779–1784.
- Mohammed, M.Z., Vyjayanti, V.N., Laughton, C.A., Dekker, L.V., Fischer, P.M., Wilson, D.M., 3rd, Abbotts, R., Shah, S., Patel, P.M., Hickson, I.D., and Madhusudan, S. (2011). Development and evaluation of human AP endonuclease inhibitors in melanoma and glioma cell lines. *Br. J. Cancer* 104, 653–663.
- Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., and Adjei, A.A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* 83, 584–594.
- Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* 52, 91–118.
- Nahar, R., Zhai, W., Zhang, T., Takano, A., Khng, A.J., Lee, Y.Y., Liu, X., Lim, C.H., Koh, T.P.T., Aung, Z.W., et al. (2018). Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. *Nat. Commun.* 9, 216.
- Okuda-Ozaki, Y., Muto, S., Takagi, H., Inoue, T., Watanabe, Y., Fukuhara, M., Yamaura, T., Okabe, N., Matsumura, Y., Hasegawa, T., et al. (2018). Prognostic Impact of Tumor Mutation Burden in Patients With Completely Resected Non-Small Cell Lung Cancer: Brief Report. *J. Thorac. Oncol.* 13, 1217–1221.
- Papke, B., and Der, C.J. (2017). Drugging RAS: Know the enemy. *Science* 355, 1158–1163.
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667.
- Peters, S., Kerr, K.M., and Stahel, R. (2018). PD-1 blockade in advanced NSCLC: A focus on pembrolizumab. *Cancer Treat. Rev.* 62, 39–49.
- Polio, S.R., Stasiak, S.E., Jamieson, R.R., Balestrini, J.L., Krishnan, R., and Parameswaran, H. (2019). Extracellular matrix stiffness regulates human airway smooth muscle contraction by altering the cell-cell coupling. *Sci. Rep.* 9, 9564.
- Reuben, A., Zhang, J., Chiou, S.H., Gittelman, R.M., Li, J., Lee, W.C., Fujimoto, J., Behrens, C., Liu, X., Wang, F., et al. (2020). Comprehensive T cell repertoire characterization of non-small cell lung cancer. *Nat. Commun.* 11, 603.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Rong, B., Zhao, C., Liu, H., Ming, Z., Cai, X., Gao, W., and Yang, S. (2014). Identification and verification of Hsp90-beta as a potential serum biomarker for lung cancer. *Am. J. Cancer Res.* 4, 874–885.
- Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 17, 31.
- Rotow, J., and Bivona, T.G. (2017). Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* 17, 637–658.
- Ruggles, K.V., Tang, Z., Wang, X., Grover, H., Askenazi, M., Teubl, J., Cao, S., McLellan, M.D., Clouser, K.R., Tabb, D.L., et al. (2016). An Analysis of the

- Sensitivity of Proteogenomic Mapping of Somatic Mutations and Novel Splicing Events in Cancer. *Mol. Cell. Proteomics* 15, 1060–1071.
- Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al.; Cancer Genome Atlas Research Network (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 173, 321–337.e10.
- Schenk, E., Hendrickson, A.E.W., Northfelt, D., Toft, D.O., Ames, M.M., Meneffe, M., Satele, D., Qin, R., and Erlichman, C. (2013). Phase I study of tanespimycin in combination with bortezomib in patients with advanced solid malignancies. *Invest. New Drugs* 31, 1251–1256.
- Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* 23, 1391–1398.
- Seckinger, A., Meissner, T., Moreaux, J., Depeweg, D., Hillengass, J., Hose, K., Rème, T., Rösen-Wolff, A., Jauch, A., Schnettler, R., et al. (2012). Clinical and prognostic role of annexin A2 in multiple myeloma. *Blood* 120, 1087–1094.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Shepherd, F.A., Lucas, B., Le Teuff, G., Hainaut, P., Jänne, P.A., Pignon, J.P., Le Chevalier, T., Seymour, L., Douillard, J.Y., Graziano, S., et al.; LACE-Bio Collaborative Group (2017). Pooled Analysis of the Prognostic and Predictive Effects of TP53 Comutation Status Combined With KRAS or EGFR Mutation in Early-Stage Resected Non-Small-Cell Lung Cancer in Four Trials of Adjuvant Chemotherapy. *J. Clin. Oncol.* 35, 2018–2027.
- Simpson, J.A., Al-Attar, A., Watson, N.F., Scholefield, J.H., Ilyas, M., and Durant, L.G. (2010). Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer. *Gut* 59, 926–933.
- 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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Sun, S., Schiller, J.H., and Gazdar, A.F. (2007). Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer* 7, 778–790.
- Takeuchi, M., Miyoshi, H., Asano, N., Yoshida, N., Yamada, K., Yanagida, E., Moritsubo, M., Nakata, M., Umeno, T., Suzuki, T., et al. (2019). Human leukocyte antigen class II expression is a good prognostic factor in adult T-cell leukemia/lymphoma. *Haematologica* 104, 1626–1632.
- Tan, C.S., Gilligan, D., and Pacey, S. (2015). Treatment approaches for EGFR-inhibitor-resistant patients with non-small-cell lung cancer. *Lancet Oncol.* 16, e447–e459.
- Thomas, A., Liu, S.V., Subramanian, D.S., and Giaccone, G. (2015). Refining the treatment of NSCLC according to histological and molecular subtypes. *Nat. Rev. Clin. Oncol.* 12, 511–526.
- Travis, W.D., Brambilla, E., Noguchi, M., Nicholson, A.G., Geisinger, K.R., Yatabe, Y., Beer, D.G., Powell, C.A., Riely, G.J., Van Schil, P.E., et al. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* 6, 244–285.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindsjög, C., Oksvold, P., Mardinoğlu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33.
- Vargas, A.J., and Harris, C.C. (2016). Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer* 16, 525–537.
- Vineis, P., Airoldi, L., Veglia, F., Olgiati, L., Pastorelli, R., Autrup, H., Dunning, A., Garte, S., Gormally, E., Hainaut, P., et al. (2005). Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ* 330, 277–280.
- Wang, X., and Zhang, B. (2013). customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* 29, 3235–3237.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.
- Wang, B., Li, J., Cheng, X., Zhou, Q., Yang, J., Zhang, M., Chen, H., and Li, J. (2018a). NIPS, a 3D network-integrated predictor of deleterious protein SAPs, and its application in cancer prognosis. *Sci. Rep.* 8, 6021.
- Wang, C., Yin, R., Dai, J., Gu, Y., Cui, S., Ma, H., Zhang, Z., Huang, J., Qin, N., Jiang, T., et al. (2018b). Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC patients. *Nat. Commun.* 9, 2054.
- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573.
- Wiredja, D.D., Koyutürk, M., and Chance, M.R. (2017). The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* 33, 3489–3491.
- Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.
- Wittekindt, O.H. (2017). Tight junctions in pulmonary epithelia during lung inflammation. *Pflügers Arch.* 469, 135–147.
- Yamaguchi, T., Yanagisawa, K., Sugiyama, R., Hosono, Y., Shimada, Y., Arima, C., Kato, S., Tomida, S., Suzuki, M., Osada, H., and Takahashi, T. (2012). NKX2-1/TITF1/TTF-1-Induced ROR1 is required to sustain EGFR survival signaling in lung adenocarcinoma. *Cancer Cell* 21, 348–361.
- Yang, C.-Y., Chang, C.-H., Yu, Y.-L., Lin, T.-C.E., Lee, S.-A., Yen, C.-C., Yang, J.-M., Lai, J.-M., Hong, Y.-R., Tseng, T.-L., et al. (2008). PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. *Bioinformatics* 24, i14–i20.
- Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961.
- Yoshida, A., Tsuta, K., Wakai, S., Arai, Y., Asamura, H., Shibata, T., Furuta, K., Kohno, T., and Fukushima, R. (2014). Immunohistochemical detection of ROS1 is useful for identifying ROS1 rearrangements in lung cancers. *Mod. Pathol.* 27, 711–720.
- Yu, H.A., Suzawa, K., Jordan, E., Zehir, A., Ni, A., Kim, R., Kris, M.G., Hellmann, M.D., Li, B.T., Somwar, R., et al. (2018). Concurrent alterations in EGFR-mutant lung cancers associated with resistance to EGFR kinase inhibitors and characterization of MTOR as a mediator of resistance. *Clin. Cancer Res.* 24, 3108–3118.
- Yun, J., Mullarky, E., Lu, C., Bosch, K.N., Kavalier, A., Rivera, K., Roper, J., Chio, I.I., Giannopoulou, E.G., Rago, C., et al. (2015). Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. *Science* 350, 1391–1396.
- Zhang, J., Wu, J., Tan, Q., Zhu, L., and Gao, W. (2013). Why do pathological stage IA lung adenocarcinomas vary from prognosis?: a clinicopathologic study of 176 patients with pathological stage IA lung adenocarcinoma based on the IASLC/ATS/ERS classification. *J. Thorac. Oncol.* 8, 1196–1202.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-HSP90AB1	ABGENT	Cat# AP7867D; RRID: AB_2121404
Rabbit polyclonal anti-Phospho-HSP90AB1-S254	Abclonal	Cat# AP0203; RRID: AB_2771182
Mouse monoclonal anti-β-actin antibody	Proteintech	Cat# 66009-1-Ig; RRID: AB_2687938
Anti-Rabbit IgG, HRP-linked Antibody	Cell Signaling Technology	Cat# 7074P2; RRID: AB_2099233
Rabbit Anti-CD8 antibody (SP16)	ZSGB Bio	Cat# ZA-0508
Rabbit Anti-HLA-DR antibody	Abcam	Cat# ab92511; RRID: AB_10563656
Goat Anti-Mouse IgG, HRP Conjugated antibody	CWBio	Cat# CW0102; RRID: AB_2736997
Biological Samples		
Paired tumor and paired non-cancerous adjacent tissues from a cohort of 103 lung adenocarcinoma patients	National Cancer Center and Cancer Hospital of Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Xuanwu Hospital, and Tianjin Baodi Hospital	This paper
Paired tumor and paired non-cancerous adjacent tissues from a cohort of 20 lung adenocarcinoma patients for western blot analysis	National Cancer Center and Cancer Hospital of Chinese Academy of Medical Sciences & Peking Union Medical College	This paper
Peripheral blood samples of 282 healthy individuals and 705 lung adenocarcinoma patients	National Cancer Center and Cancer Hospital of Chinese Academy of Medical Sciences & Peking Union Medical College	This paper
Chemicals, Peptides, and Recombinant Proteins		
PMSF	Sigma	Cat#: P-7626
Dithiothreitol	Sigma	Cat#: 43815-1G
Iodoacetamide	Amresco	Cat#: M216-30G
Formic acid	Sigma	Cat#: 56302-50ML
Trifluoroacetic acid	Sigma	Cat#: T6508-100ML
RNAlater stabilization solution	Invitrogen	Cat#: AM7021
Protease inhibitor cocktail	Roche	Cat#: 04693159001
Phosphatase inhibitor cocktail	Roche	Cat#: 04906837001
Trypsin	Hualishi Scientific	Cat#: HLS TRY001C
Ammonium bicarbonate	Sigma	Cat#: A6141-500G
Cysteine	Sigma	Cat#: C7352-25G
Acetonitrile	Fisher	Cat#: A955-4
C18 resin	Dikma Technologies	Cat#: 85252
SepPak C18 cartridges	Waters	Cat#: WAT054960
Xbridge C18 column	Waters	Cat#: 186003576
Titanium dioxide (TiO ₂) beads	GL Sciences Inc	Cat#: 5020-75000
30 kDa Microcon filter	Millipore	Cat#: UFC503096
Critical Commercial Assays		
QIAamp DNA mini kit	QIAGEN	Cat# 51304
BCA protein assay kit	Beyotime Biotechnology	Cat#: P0011
DNA assay kit	Life Technologies	Cat# Q10212
Agilent SureSelect Human All Exon V6 kit	Agilent Technologies	Cat# 5190-8863

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TRIzol reagent	Invitrogen	Cat# 15596026
UltraTM RNA library prep kit	NEB	Cat# E7530L
HiSeq PE Cluster Kit v4-cBot-HS	illumina	Cat# PE-401-4001
RIPA lysis buffer	Applygen	Cat# C1053
ECL detection kit	Applygen	Cat# P1030
Polyvinylidene difluoride membrane	Millipore	Cat# ISEQ0010
ELISA kit for heat shock Protein 90 kDa Alpha B1	USCN	Cat# SED522Hu
Deposited Data		
Proteomics data	This paper	iProx: IPX0001804000
RNA-Seq data	This paper	GSE140343
DepMap portal	NA	https://depmap.org/portal/
GDSC database	Yang et al., 2013	https://www.cancerrxgene.org
Software and Algorithms		
Trim_Galore script (version 0.4.4)	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
BWA MEM (version 0.7.12-r1039)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
Samtools (version 1.5)	Li et al., 2009	http://samtools.sourceforge.net/
picard (version 22)	GitHub	http://broadinstitute.github.io/picard/
GATK (version 3.7)	Broad Institute	https://gatk.broadinstitute.org/hc/en-us
Mutect2 (version 2)	Broad Institute	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Oncotator (version 1.9.2.0)	Broad Institute	https://software.broadinstitute.org/cancer/cga/oncotator
MutSigCV tool	Broad Institute	https://software.broadinstitute.org/cancer/cga/mutsig
FastQC (version 0.11.7)	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
hisat2 software (version 2.0.4)	Pertea et al., 2016	https://ccb.jhu.edu/software/hisat2/index.shtml
StringTie software (version 1.3.4d)	Pertea et al., 2016	https://ccb.jhu.edu/software/stringtie/#install
MaxQuant (version 1.6.5.0)	Cox and Mann, 2008	http://www.coxdocs.org/doku.php?id=maxquant:start
limma R package (version 3.40.4)	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
ConsensusClusterPlus R package (version 1.48.0)	Wilkerson and Hayes, 2010	RRID: SCR_016954
Survminer R package (version 0.2.4)	NA	https://cran.r-project.org/web/packages/survminer/index.html
diptest R Package (version 0.75-7)	NA	https://cran.r-project.org/web/packages/diptest/index.html
Cytoscape (version 3.6.1)	Shannon et al., 2003	NA
ComplexHeatmap R package (version 2.0.0)	Gu et al., 2016	http://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
pyQUILTS	Ruggles et al., 2016	https://github.com/ekawaler/pyQUILTS
TopHat (version 2.1.0)	Kim et al., 2013	http://ccb.jhu.edu/software/tophat
TopHat-Fusion (version 0.1.0)	Kim and Salzberg, 2011	http://tophat-fusion.sourceforge.net/
Bowtie1 (version 1.2.3)	Langmead et al., 2009	http://bowtie-bio.sourceforge.net

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie2 (version 2.2.9)	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
deconstructSigs (version 1.8.0)	Rosenthal et al., 2016	https://cran.r-project.org
ANNOVAR (version 2017-07-17)	Wang et al., 2010	http://www.openbioinformatics.org/annovar/annovar_download.html
NMF R package (version 0.22.0)	Gaujoux and Seoighe, 2010	https://cran.r-project.org/web/packages/NMF/index.html
X!tandem (version 2017.02.01.4)	The Global Proteome Machine Organization	https://www.thegpm.org/tandem/
NIPS	Wang et al., 2018a	http://lilab.life.sjtu.edu.cn:8080/nips/
WGCNA R package	Langfelder and Horvath, 2008	https://cran.r-project.org/web/packages/WGCNA/
ImageJ software (version 1.51j)	National Institutes of Health	https://imagej.nih.gov/ij/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Minjia Tan (mjtan@simm.ac.cn). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Patient samples of LUADs*****Sample collection***

The LUAD samples used for this study were collected from three hospitals in China including National Cancer Center and Cancer Hospital of Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Xuanwu Hospital, and Tianjin Baodi Hospital. Patients were randomly selected from March 2010 to May 2016 upon their first visit and did not undergo any anti-cancer treatments prior to surgery. Primary tumor tissues and paired non-cancerous adjacent tissues (> 3 cm apart from tumor edge) were surgically resected and transferred to sterile freezing vials. A total of 103 paired specimens were collected with the clinical information including gender, age, tumor size, histological subtype, degree of differentiation, TNM staging (AJCC cancer staging system 8th edition), tumor purity, smoke status, date of surgical resection, status of cancer recurrence, status of survival, and drug information after surgery. Two very early stage pathological subtypes of LUAD, namely adenocarcinoma-*in situ* (AIS) and minimally invasive adenocarcinomas, were excluded since the tumor size of these two subtypes was usually too small to conduct multi-omic experiments. All patient samples were obtained with the hospital's approval of the Research Ethics Committee with written informed consent provided by all participants.

Sample preparation

A single site of primary tumor and paired non-cancerous adjacent tissues was collected. Collected specimens were divided into three parts: the first part was snap frozen in liquid nitrogen and then stored in a -80°C refrigerator until being used for proteomics and phosphoproteomics as well as DNA extraction; the second part was processed with RNAlater Stabilization Solution (Invitrogen, catalog No: AM7021, Carlsbad, CA) and stored at -80°C until being used for the RNA extraction; the third part was treated with formalin for HE staining and used for histological evaluation. Each sample was assigned a new research ID and the patient's name or medical record number used during hospitalization was de-identified.

Histopathology and immune cell infiltration

Formalin-fixed, paraffin-embedded specimens were prepared and provided by each hospital. HE-stained slides were examined and evaluated independently by two experienced pathologists and information regarding tumor histological subtype, degree of differentiation, TNM staging, and tumor purity were provided. All samples used in this study had tumor purity ≥ 50%. In addition, the non-tumor cell populations, including stroma, immune cells, normal tissue, and necrosis, were evaluated. Immunohistochemical (IHC) staining of CD8 and HLA-DR were performed to evaluate the immune cell infiltration. The IHC results were evaluated independently by two pathologists. All the information was included in [Table S1](#).

Immunohistochemistry and scoring of IHC

Tumor infiltrations were validated through IHC staining of CD8 and HLA-DR. Formalin-fixed, paraffin-embedded tumor tissues were sliced into 4 µm slides for IHC staining. Slides were stained with CD8 antibody (1:50, ZSGB Bio, catalog No: ZA-0508) and HLA-DR antibody (1:800, Abcam, catalog No: ab92511). Slides were processed using an automated Leica Bond staining system according to the manufacturer's protocol.

For scoring the IHC image, Histoscore (H-score) was calculated by multiplying the proportion of positive cells in the sample (0–100%) by the average intensity of the positive staining (1+, 2+, or 3+) to obtain a score ranging between 1 and 300 as previously described (Hirsch et al., 2003; Reuben et al., 2020; Yoshida et al., 2014). CD8 was classified into CD8 high (≥ 150), medium ($50 \leq H\text{-score} < 150$) or low (< 50) whereas HLA-DR were classified into high (≥ 150) or low (< 150) on the basis of individual biomarker distribution (Table S1).

Independent cohort

To measure plasma HSP 90 β , peripheral blood samples of 282 healthy individuals (182 males and 100 females with the mean age of 55) and 705 diagnosed LUAD patients (332 males and 373 females with mean age at 60) who underwent surgery were collected at the National Cancer Center and Cancer Hospital of the Chinese Academy of Medical Sciences & Peking Union Medical College between July 2007 and August 2013. Patients were excluded if they have been treated with radiotherapy or chemotherapy or suffering from other cancers at the same time. All patients provided written informed consent before surgery, and the treatments were performed in accordance with current ethical guidance. Peripheral blood samples were collected prior to surgery by venipuncture and preserved in EDTA-coated tubes. Samples were centrifuged at 4°C for 10 minutes at 1,000 g to separate plasma from blood cells. Supernatants were collected, divided into aliquots, and stored at –80°C until use. Of the 705 patients, 499 had the post-hospital follow-up visit. The median follow-up time was 48 months. Overall survival (OS) and disease-free survival (DFS) were given to evaluate the clinical outcomes.

To validate the phosphorylation of HSP 90 β S255 by immunoblotting, 20 additional lung adenocarcinoma tumors and paired non-cancerous adjacent tissues were collected at Cancer Hospital of Chinese Academy of Medical Sciences & Peking Union Medical College. Specimens were collected and prepared in the same manner as those of the 103 LUAD patients.

METHOD DETAILS

Proteomic and phosphoproteomic analysis

Protein extraction and tryptic digestion

Collected samples were washed five times with phosphate buffer saline (PBS) buffer to remove blood and debris. Samples were lysed in lysis buffer (8 M urea, 100 mM ammonium bicarbonate, pH 8.0) supplemented with protease inhibitors and phosphatase inhibitors (Roche) for 20 min on ice. Samples were then sonicated for 2 min (3 s on and 3 s off) on ice and centrifuged at 21,000 g for 10 min to remove the tissue debris. The supernatants were collected, and the protein concentration was measured by using BCA protein assay. Extracted proteins were reduced in 5 mM dithiothreitol at 56°C for 30 min and then alkylated in 15 mM iodoacetamide at room temperature for 30 min in darkness. Reaction was quenched with 30 mM cysteine at room temperature for additional 30 min. Protein samples underwent trypsin digestion (enzyme-to-substrate ratio of 1:50 at 37°C for 16 hours) followed by desalting through SepPak C18 cartridges (Waters, MA) and vacuum-dried by Speed Vac.

Peptide pre-fractionation by high-pH HPLC

Tryptic peptides were fractioned by using high-pH HPLC to reduce sample complexity. Briefly, tryptic peptides were dissolved in buffer A (2% acetonitrile (ACN), pH 9.5) and then loaded on an Xbridge C18 column (Waters, MA, 4.6 mm × 100 mm, 130Å, 3.5 µm) and eluted with a 60 min gradient from 0 to 95% buffer B (98% ACN, pH 9.5) at a flow rate of 0.6 mL/min. Aliquots were combined into 10 fractions before MS analysis.

The enrichment of phosphorylated peptides

For phosphoproteomic analysis, peptides were extracted after trypsin digestion using a FASP method (Wiśniewski et al., 2009). Briefly, samples were transferred into a 30 kD Microcon filter (Millipore) and centrifuged at 14,000 g at 15°C for 20 min. The precipitate in the filter was washed twice by adding 300 µL washing buffer (8 M urea in 100 mM Tris, pH 8.0) into the filter and centrifuged at 14,000 g at 15°C for 20 min. The precipitate was resuspended in 300 µL 100 mM NH₄HCO₃. Trypsin with a protein-to-enzyme ratio of 50:1 (w/w) was added into the filter. Proteins were digested at 37°C for 16 h. After tryptic digestion, peptides were collected by centrifugation at 14,000 g at 15°C for 20 min and then dried in the SpeedVac. After tryptic digestion, the resulting peptides were enriched with titanium dioxide (TiO₂) beads (GL Sciences Inc., Japan) incubated in 800 µL incubation solution (70% ACN, 5% trifluoroacetic acid, 8% lactic acid) with a peptide-to-bead ratio of 1:2.5. Beads were collected by centrifugation at 2,000 g for 2 min. The incubation reaction was conducted for three times. Bound peptides were washed with incubation solution for five times. An addition washing procedure was carried out with wash buffer 1 (30% ACN, 0.5% trifluoroacetic acid) for one time and then wash buffer 2 (80% ACN, 0.5% trifluoroacetic acid) for two times to further remove the unphosphorylated peptides. The remaining peptides on beads were eluted with elution buffer (15% ammonia, 80% ACN), resolved on a C18 stage tip, and then fractionated into three fractions for LC-MS/MS analysis.

Nano-LC-MS/MS

Peptide samples were analyzed on an EASY-nLC 1000 LC system (Thermo Fisher Scientific, Waltham, MA) coupled with an Orbitrap Fusion mass spectrometry (Thermo Fisher Scientific, Waltham, MA). Peptides were re-dissolved in mobile phase A (2% ACN and 0.1% formic acid) and directly loaded onto an 18 cm long, home-made C18 nano-capillary analytical column (75 µm inner diameter) packed with C18 resin (particle size 3 µm, pore size 100 Å, Dikma Technologies Inc., Lake Forest, CA).

For the proteome profiling samples, peptides were separated onto the analytical column with a 60 min gradient (buffer A: 0.1 % Formic acid in water; buffer B: 0.1 % Formic acid in 90 % ACN) at a constant flow rate of 350 nL/min (0-50 min, 5 to 45 % of buffer B; 50-55 min, 45% to 80% buffer B; 55-60 min, 80% of buffer B). The eluted peptides were ionized under 2 kiloVolts and introduced into mass spectrometry. Mass spectrometry was operated under a data-dependent acquisition mode. For the MS1 full scan, ions with m/z ranging from 350 to 1300 were acquired by Orbitrap mass analyzer at a high resolution of 120,000. The automatic gain control (AGC) was set as 5×10^5 . The maximal ion injection time was 50 ms. MS2 acquisition was performed in a top-speed mode and the duty cycle time was 3 s. Precursor ions were selected and fragmented with higher energy collision dissociation (HCD) with normalized collision energy of 32%. Fragment ions were analyzed by ion trap mass analyzer with AGC at 7000. The maximal ion injection time of MS2 was 35 ms and the dynamic exclusion was 60 s.

For the phosphoproteomic samples, peptide separation was achieved with a 90 min gradient (0-50 min, 2 to 18% of buffer B; 50-80 min, 18% to 45% of buffer B; 80-85 min, 45% to 80% of buffer B; 85-90 min, 80% of buffer B). The eluted phosphopeptides were ionized and detected by an Orbitrap Fusion mass spectrometry. Mass range was 300-1400 m/z for MS1. The maximal ion injection time for MS2 was 50 ms and the dynamic exclusion was set to 30 s. All other parameters were set as the same as those used for the proteome profiling samples.

MS database searching

MS raw files generated by LC-MS/MS were searched against the UniProt human proteome database (version 2019-03-07, 20,404 sequences) using MaxQuant (version 1.6.5.0) software enabled with Andromeda search engine. Protease was Trypsin/P. Up to 2 missed cleavages were allowed. Carbamidomethyl (C) was considered as a fixed modification. For the proteome profiling data, variable modifications were oxidation (M) and acetylation (Protein N-term). For the phosphoproteomic data, variable modifications were oxidation (M), acetylation (Protein N-term) and phospho (S/T/Y). The cutoff of false discovery rate (FDR) by using a target-decoy strategy was 1% for both proteins and peptides.

Construction of patient-specific database

To create patient-specific database containing somatic mutations (nonsynonymous single nucleotide variations and small INDELS derived from WES and RNA-Seq data), an approach for searching customized protein sequence database was adopted for each patient sample based on the somatic variants from patient's matched RNA-Seq and WES data ([Wang and Zhang, 2013](#)).

To create patient-specific database containing novel alternative splice variants (derived from RNA-Seq data) and fusion proteins (derived from RNA-Seq data), pyQUILTS ([Ruggles et al., 2016](#)), an open-source proteogenomic tool, was used to translate novel transcripts *in silico* into junction-specific peptides using human UniProt proteome database (version 2019-03-07, 20,404 sequences) and GRCh38.p12 as the genome reference. An in-house python script was developed to filter out any variant peptides that were present in the human UniProt proteome database, or had a peptide length less than five amino acids, or had less than ten supporting reads spanning the novel junction site. Alternative splice variants were grouped into three classes: unannotated, partially novel, and completely novel ([Ruggles et al., 2016](#)).

Identification of variant peptides

MS/MS data were searched by X!Tandem (version 2017.2.14) against the patient-derived protein databases with decoy sequences. For variant peptide identification, 0.1 % separated FDR at PSM level was used so that there is no obvious bias between the detected variants and the wide-type peptides in the searching score distributions. ANNOVAR and COSMIC databases were used for variant annotation. The deleterious risk of single amino acid variations was evaluated by using NIPS ([Wang et al., 2018a](#)).

Whole-exome sequencing

DNA extraction

Total DNA from lung cancer tissues and matched non-cancerous adjacent tissues were extracted using QIAamp DNA mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. DNA degradation and contamination were monitored on 1% agarose gels. DNA concentration was measured by Qubit® DNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA). A total amount of 0.6 µg genomic DNA per sample was used as input for DNA sequencing. Sequencing libraries were generated by using Agilent SureSelect Human All Exon V6 kit (Agilent Technologies, CA, USA) following manufacturer's recommendations and barcodes were added to each sample. DNA fragments were sequenced on an Illumina HiSeq X ten sequencing system.

Alignment

The exome sequencing data were used to identify somatic variations that were found in tumor samples but not in matched adjacent tissues. The Trim_Galore script (version 0.4.4) was employed to perform adaptor trimming and low-quality reads filtering. Clean reads were aligned via BWA MEM (version 0.7.12-r1039) against the human reference genome hs37d5 (based on GRCh37 assembly with human virus sequences) with default parameters. Primitive alignment results were sorted and indexed through Samtools (version 1.5); they were then marked and duplicate reads were removed by running picard (version 22). We ran GATK (version 3.7) to perform local realignment and recalibration following the tool's best practice and the output list was prepared for downstream analysis.

Variant detection and annotations

We used Mutect2 in GATK3.7, a variant caller, to detect possible SNVs and small indels (insertion/deletion) in the tumor genome. Mutect2 was run in default setting by taking tumor and matched adjacent tissues as input and only mutations called in targeted area were evaluated. Human SNP database (dbSNPv132) and COSMICv81 coding and noncoding mutation data were also provided

as reference inputs of Mutect2. Passed mutations were annotated by somatic mutation annotation software Oncotator (version 1.9.2.0).

Somatic copy number alteration

Exome-based somatic copy number alteration (SCNA) was called by following somatic CNV calling pipeline in GATK's (GATK 4) Best Practice. The results of this pipeline, segment files of every 1000 were input in GISTIC2, to identify significantly amplified or deleted regions across all samples, which could be accumulated driving regions. To exclude false positives as much as possible, relatively stringent cut off thresholds were used with parameters: -ta 0.5 -tb 0.5 -brlen 0.5 -conf 0.9. Other parameters were the same as default values. For those genes with SCNA, Spearman correlation coefficients were calculated between CNAs and the abundances of mRNA, protein, or phosphoprotein (FDR < 0.05).

Calling significantly mutated genes

MutSigCV tool was used to identify significantly mutated genes by taking mutations of all samples as input and comparing them against background mutation rate. Genes mutated more often than expected by chance were regarded as candidates of driver genes. The first 20 most significantly mutated genes and their co-mutation status were analyzed. The software ran with default parameters and background mutation rate files were offered by the tool package.

Mutational signature analysis

Mutational signature analysis in 103 LUAD patients were performed by using the deconstructSigs approach (Rosenthal et al., 2016) and its R package (deconstructSigs v1.8.0) with default parameters. Thirty COSMIC cancer signatures were considered and their contributions (weights) in each patient were normalized between 0 and 1, and signatures with a weight below 0.06 were filtered out.

RNA-Seq

RNA extraction

RNA was extracted from tissues by using TRIzol reagent kit (Ambion, Invitrogen, USA) according to the reagent protocols. The concentration and RNA integrity were then determined by using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) and an Agilent 2100 Bioanalyzer (Agilent, CA, USA). RNA samples exhibiting an RNA integrity number (RIN) greater than 6.0 were included in the study. For library preparation of RNA sequencing, a total amount of 2 µg RNA per sample was used as the input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and RNase H. Second strand cDNA synthesis was subsequently performed using buffer, dNTPs, DNA polymerase I and RNase H. The library fragments were purified with QiaQuick PCR kits and eluted with EB buffer, followed by terminal repair, A-tailing and adaptor addition. The aimed products were retrieved and PCR was performed, then the library was completed. The RNA concentration of the library was measured using Qubit® RNA Assay Kit in Qubit® 3.0 and then it was diluted to 1 ng/µL. Insert size was assessed using the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), and qualified insert size was accurately quantified using StepOnePlusTM Real-Time PCR System (Library valid concentration > 10 nM). The clustering of the index-coded samples was performed on a cBot cluster generation system using HiSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the libraries were sequenced on an Illumina platform and 150 bp paired-end reads were generated.

RNA-Seq data analysis

RNA-Seq reads were adaptor trimmed and the data quality was assessed with the FastQC (version 0.11.7) software before any data filtering criteria was applied. Reads were mapped onto the human reference genome (GRCh38.p12 assembly) by using HISAT2 software (v2.0.4). The mapped reads were assembled into transcripts or genes by using StringTie software (v1.3.4d) and the genome annotation file (hg38_ucsc.annotated.gtf). For quantification purpose, the relative abundance of the transcript/gene was measured by a normalized metrics, FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Transcripts with an FPKM score above one were retained, resulting in a total of 16,188 gene IDs. All known exons in the annotated file were 100% covered.

Calling mutated genes from RNA-Seq data

For the variant discovery from RNA-seq data, we followed the GATK Best Practice Variant Detection protocol on RNASeq (<https://gatkforums.broadinstitute.org/dsde/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail>). For 49 paired tumors and NATs, only the SNVs, INDELs, and junctions obtained in the tumor tissues exclusively were retained. The resulting VCF file was further filtered by dbSNP.

Fusion genes and novel alternative splicing events

To search potential fusion genes in tumors and adjacent tissues, reference genome (GRCh38.p12) was indexed with the NGS software - Bowtie1 (version 1.2.3) (Langmead et al., 2009). RNA-Seq reads were aligned to the indexed genome with reads aligner, TopHat (version 2.1.0) (Kim et al., 2013), to identify exon-exon junctions with the fusion search option and other parameters (-fusion-search-keep-fasta-order-no-coverage-search -r 100-mate-std-dev 250-max-intron-length 100000-fusion-min-dist 100000-fusion-anchor-length 20-fusion-ignore-chromosomes chrM). TopHat-Fusion (version 0.1.0) (Kim and Salzberg, 2011) was then employed to identify novel fusion transcripts (tophat-fusion-post:-num-fusion-reads 1-num-fusion-pairs 2-num-fusion-both 5). Only candidates in the output file with at least ten junction reads (JunctionReadCount) were considered and retained.

To search novel alternative splicing transcripts, reference genome (GRCh38.p12) was indexed with Bowtie2 (version 2.2.9) ([Langmead and Salzberg, 2012](#)). RNA-Seq raw files were then processed with reads aligner TopHat (version 2.1.0) to identify exon-exon splicing junction sites with–max-multihits (-g 1),–prefilter-multihits ([Van der Auwera et al., 2013](#)),–transcriptome-max-hits (-x 1) options. The identified novel alternative splicing transcripts with at least ten junction reads were considered and retained.

Enzyme-linked immunosorbent assays (ELISA)

Plasma HSP 90 β concentrations were assessed by ELISA kits (USCN, China) according to the manufacturer's instructions. Briefly, 100 μ L of diluted plasma was added into wells on an anti- HSP 90 β microplate and incubated at 37°C for 2 hours. Then, 100 μ L of the biotinylated HSP90 β detector antibody was added to each well and incubated at 37°C for 1 hour. After 3 washes, 100 μ L of the conjugate was added to each well and incubated at 37°C for 1 hour. After 5 washes, samples were measured immediately using a microplate reader (Bio-Rad Laboratory, Hercules, CA, USA) with absorbance at 450 nm.

Western blot assay and antibodies

To extract the proteins from the tissues, 20 paired LUAD tumor tissues and paired non-cancerous adjacent tissues stored in a –80°C refrigerator were thawed, and then cut into pieces and lysed on ice using RIPA lysis buffer (Applygen, China) supplemented with PMSF and Protease Inhibitor Cocktail and then quantified using the BCA Protein Assay Kit. Equal amounts of the cell protein lysates (20 μ g) were separated by 10% SDS-PAGE and transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore, Darmstadt, Germany). After being blocked with 10% nonfat dry milk in phosphate-buffered saline and 0.1% Tween 20 solution, membranes were incubated overnight at 4°C with the following primary antibodies: HSP90AB1 (ABGENT, China, 1:1000), Phospho-HSP90AB1-S254 (Abclonal, China, 1:1000) and β -actin (Proteintech, China, 1:5000). The membranes were then incubated with the appropriate secondary antibody for 1 hour at room temperature. The specific bands were detected using an ECL detection kit (Applygen, China) and captured on an ImageQuant LAS 4000 mini system (GE Healthcare, NJ, USA). Relative expression was determined by normalizing with β -actin using the ImageJ software (Version 1.51j, National Institutes of Health, MD, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality control and assessment of LC-MS/MS data

A label-free quantification algorithm, iBAQ, was used for protein quantitation. Values of iBAQ calculated by MaxQuant (version 1.6.5.0) software were quantile normalized, and log2 transformed if necessary. The density plot of the normalized intensities of the proteins identified in each sample showed that all samples passed the quality control with an expected unimodal distribution (dip statistic test). PCA analyses of both proteomic and phosphoproteomic profiles showed no significant batch effects.

Differential protein analysis

Wilcoxon signed-rank test was used to examine whether proteins were differentially expressed between the tumors and matched NATs. Wilcoxon rank-sum test was used to examine whether proteins were differentially expressed between patients with good and poor prognosis, or patients with different mutation status. Upregulated or downregulated proteins in tumors were defined as proteins differentially expressed in tumors compared with their matched adjacent tissues ($T/N > 2$ or $< 1/2$, Wilcoxon signed-rank test, Benjamini-Hochberg adjusted $p < 0.05$). Wilcoxon rank-sum test ($p < 0.05$, fold change > 2) was used to identify differentially expressed proteins in stage I patients with poor ($n = 15$, DFS time less than 3 years) or good prognosis ($n = 36$, DFS time more than 3 years).

Pathway enrichment analysis

Gene set enrichment analysis (GSEA) was used for pathway enrichment analysis ([Subramanian et al., 2005](#)). GSEA evaluates and determines whether *a priori* defined sets of genes show statistically significant, cumulative changes in gene expression that are correlated with a specific phenotype. Samples grouped according to the paired T/N group, mutation status, or the clinical outcomes of patients were subjected to GSEA, respectively. Proteins with less than 50% of missing values were selected, and missing values were then imputed with the minimum value of the proteomic data as GSEA required input data with no missing values. Molecular Signatures Database (MSigDB) of hallmark gene sets (H), curated gene sets (H2) and GO gene sets (C5) were used for enrichment analysis. An FDR value of 0.05 was used as a cutoff. The enrichment score (ES) in GSEA was calculated by first ranking the proteins from the most to least significant with respect to the two phenotypes (i.e., patients with good and poor prognosis), the entire ranked list was then used to assess how the proteins of each gene set were distributed across the ranked list.

Consensus clustering analyses

Prior to clustering analysis, proteins that were expressed in more than 70% of patient samples were selected ($n = 6,267$). Standard deviations were calculated for each protein and top 25% most variably expressed proteins ($n = 1,567$) were subjected to subsequent statistical analyses. To identify new proteomic subtypes of LUAD, consensus clustering (R package ConsensusClusterPlus v.1.48.0) ([Monti et al., 2003](#); [Wilkerson and Hayes, 2010](#)), a reconciling clustering algorithm with resampling technique, was conducted on 1,567 proteins. Clustering algorithm was k -means using Euclidean distance. A total number of subsampling was 1,000. Proportion

of samples selected was 80% in each resampling. Number of clusters considered was 3 to 6. A consensus matrix with $k = 3$ appeared to have the clearest cut between clusters and showed significant association with the patients' survival. The same strategy was applied on the phosphoproteomic data.

Clustering analysis of transcriptomic data

RNA-Seq of 51 tumor samples was used for subtyping. Non-negative matrix factorization (NMF) ([Gaujoux and Seoighe, 2010](#)), an unsupervised clustering algorithm, and the R-package (NMF, Version: 0.22.0) were adopted. Prior to the clustering analysis, a list of genes with an FPKM value found in more than 50% of samples were selected. Genes with top 15% standard variations ($n = 2,312$) were retained and subjected to the clustering analysis (number of runs = 400).

WGCNA analysis

To identify differentially co-expressed gene modules, WGCNA ([Langfelder and Horvath, 2008](#)) was applied to the proteins ($n = 1,567$) that have been used for proteomic subtyping. KEGG enrichment analysis was used for the functional annotation of the identified modules ($n = 7$). The eigengenes of each module were used to measure the association between a module and clinical information.

Correlation between subtype and clinical features

In the measurements of correlations between proteomic subtypes and clinical features, Fisher's exact test was performed on categorical variables, including driver gene mutations, gender, age group, smoke status, differentiation, TNM stage, and histological types. Kruskal-Wallis test was performed on continuous variables including, mutation counts, TTF-1 and Napsin-A expression analyses.

Phosphoproteomic data analysis

For the phosphoproteomic data, the intensities of the phosphopeptides of 79 patient samples were extracted from the MaxQuant software. The confidently identified phosphopeptides (1% FDR and $\geq 75\%$ localization probability) were applied with the mean-centering correction to adjust for sample-specific biases.

The correlation network between the cancer-related phosphoproteins and the protein expression in oncogenic signaling pathways was calculated with pair-wised Spearman correlation coefficients ($r \geq 0.4$, $p < 0.05$ with BH adjustment). The network was built, visualized and generated by the software Cytoscape (version 3.6.1).

Gene Set Enrichment Analysis (GSEA) was applied to find enriched pathways in phosphoproteins. Phosphoproteins with less than 50% of missing values were selected, and missing values were then imputed with the minimum value of the phosphoproteomic data.

KSEA algorithm was used to estimate the kinase activities by using the ratios of all identified phosphosites between tumor and NAT. Kinase-Substrate Enrichment Analysis (KSEA) estimates changes in a kinase's activity by measuring and averaging the amounts of its identified substrates instead of a single substrate, which enhances the signal-to-noise ratio from inherently noisy phosphoproteomics data ([Casado et al., 2013](#); [Wiredja et al., 2017](#)). If the same phosphorylation motif was shared by multiple kinases, it was used for estimating the activities of all known kinases. The use of all curated substrate sequences of a particular kinase minimizes the overlapping effects from other kinases and thus improves the precise measurement of kinase activities. The information of kinase-substrate relationships was obtained from publicly available databases including PhosphoSite ([Hornbeck et al., 2004](#)), Phospho.ELM ([Dinkel et al., 2011](#)), and PhosphoPOINT ([Yang et al., 2008](#)). The information of substrate motifs was obtained either from the literatures ([Schwartz and Gygi, 2005](#)) or from an analysis of KSEA dataset with Motif-X ([Casado et al., 2013](#)).

mRNA-protein correlation in tumors and paired NATs

The Pearson correlation coefficients of genes/proteins were calculated for those that were detected in more than 50% of the tumor tissues (4,879 genes in 51 samples) or paired non-cancerous adjacent tissues (3,762 genes in 49 samples) in both RNA-Seq and MS data. ConsensusPathDB KEGG enrichment analysis (<http://cpdb.molgen.mpg.de/CPDB>) were performed using the over-representation method and chose the full list of 20,027 IDs as background genes identified by either RNA-Seq or LC-MS/MS. Among the top 20 most significant calls (ranked by q values with cutoff of 0.05), pathways with good or poor mRNA-protein correlations were analyzed.

Survival analysis

Kaplan-Meier survival curves (log-rank test) were used for overall survival (OS) or disease-free survival (DFS) of the proteomic subtypes and patients with *EGFR* mutation versus co-mutation of *EGFR* and *TP53*.

Prior to the log-rank test of a given mRNA, protein, phosphoprotein, or phosphosite, survminer 0.2.4 R package with maxstat (maximally selected rank statistics) (<http://r-addict.com/2016/11/21/Optimal-Cutpoint-maxstat.html>) was used to determine the optimal cutpoint for the selected samples according to the previous study ([Seckinger et al., 2012](#)). DFS or OS curves were then calculated (Kaplan-Meier analysis, log-rank test) based on the optimal cutpoint.

Potential prognostic biomarkers or drug targets**Nominating prognostic biomarkers of LUAD**

The following four criteria were used to identify potential plasma biomarkers with prognostic power: 1) The candidate proteins were expressed in all of the 103 tumor samples; 2) The abundances of candidate biomarkers were among the top 1,000 protein list; 3) The candidates were expressed at least 4-fold higher in tumors than the adjacent tissues (Wilcoxon signed-rank test, $p < 0.05$); and 4) The high expression of candidates were negatively correlated with the overall survival (Kaplan-Meier analysis, log-rank test, $p < 0.01$).

Screening potential drug targets of LUAD

The following four criteria were used to screen the potential drug targets in LUAD patients: 1) The candidates should be an FDA-approved drug target or potential drug target in the Human Protein Atlas database (<https://www.proteinatlas.org/humanproteome/proteinclasses>); 2) Overexpression of a candidate was associated with poor prognosis (Kaplan-Meier log-rank test, $p < 0.01$) in 103 patients. 3) Average gene dependency score (CERES) of a candidate in 36 available LUAD cell lines in the DepMap database was below -0.6 (A cutoff score defined the gene dependency in a given cancer cell line) and 4) A candidate drug target was expressed higher in tumor (T/N ratio > 2 -fold) in all 103 patients (Wilcoxon signed-rank test, $p < 0.05$).

Screening potential drug targets of EGFR mutant LUAD

Similar four criteria were used to screen the potential drug targets in *EGFR* mutant LUAD patients with some modifications in three criteria: 1) The prognostic power of a candidate was observed in the 51 *EGFR* mutant patients; 2) The average CERES scores were calculated in the 9 *EGFR* mutant cell lines in the DepMap database; 3) The candidate drug target was expressed higher in tumor (T/N ratio > 2 -fold) in the 51 *EGFR* mutant patients (Wilcoxon signed-rank test, $p < 0.05$).

Screening potential targetable mutations of LUAD

For screening the new targetable mutations, 166 novel or COSMIC-annotated protein variants mapped to 151 genes were identified by integrative analyses of genomic, transcriptomic, and proteomic data. Among them, 35 genes (corresponding to 39 mutated peptides) were classified as FDA-approved drug targets or potential drug targets in the Human Protein Atlas database. These variant peptides were further validated by manual inspection of the MS/MS spectra. Finally, 31 mutant sites (corresponding to 29 proteins) were identified as potential targetable mutations. NIPS meta-scores (a 3D network-integrated predictor of single amino acid polymorphism (SAP) in mRNA and proteomic data) we established previously ([Wang et al., 2018a](#)) were used to estimate the pathogenesis of deleterious SAPs.

DATA AND CODE AVAILABILITY

Raw proteomics data have been uploaded to the iProx Consortium with the subproject ID IPX0001804000 <https://www.iprox.org/page/PSV023.html?url=15717295820259eyj>.

Raw RNA-Seq data have been deposited into the GEO dataset with an accession number: GSE140343. As publicly sharing of the raw genomic data is restricted by the regulation of the Human Genetic Resources Administration of China, detailed results of whole exome sequencing were included in [Table S2](#).

Supplemental Figures

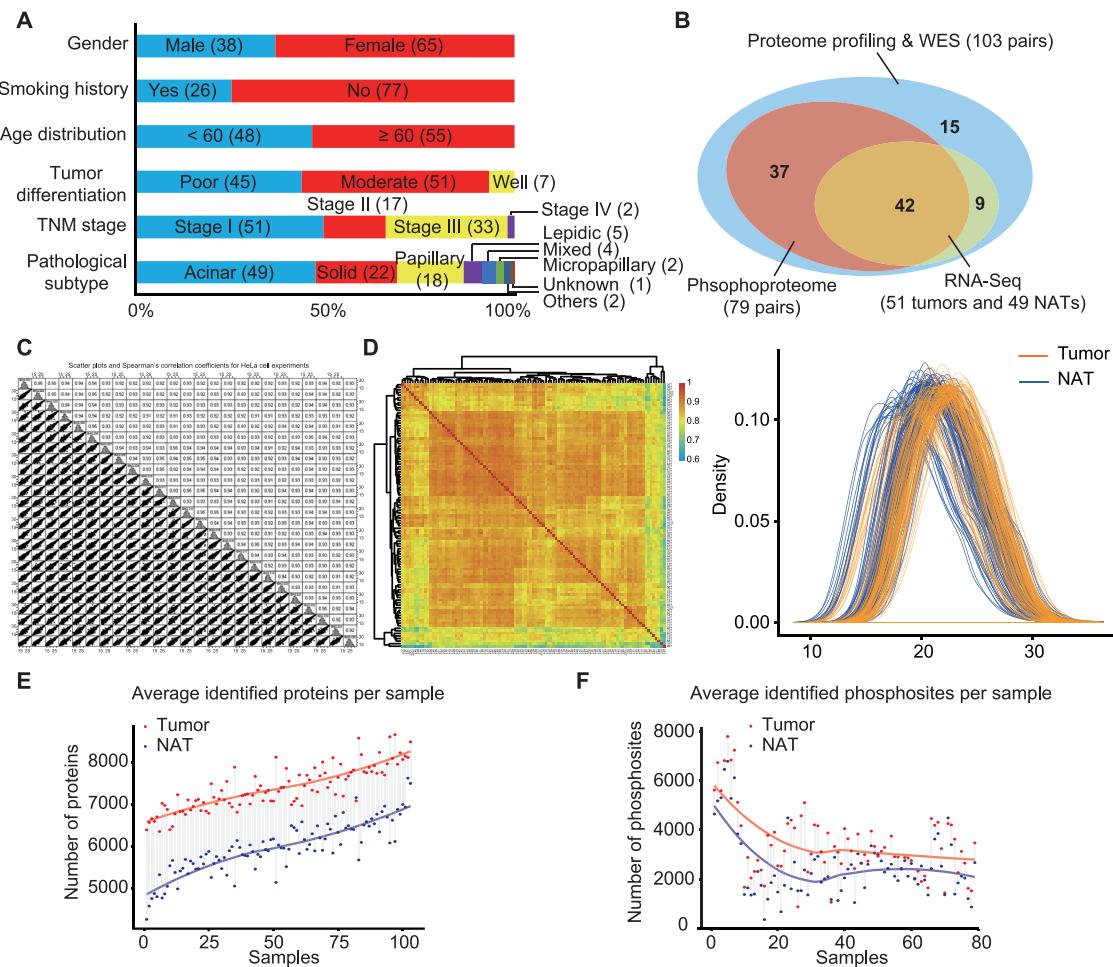


Figure S1. Multi-omics Landscape of LUAD in Chinese Patients, Related to Figure 1

- (A) Summary of clinical characteristics of 103 LUAD patients. Clinical features are gender, smoking history, age distribution, tumor differentiation, TNM stage and subtypes.
- (B) Venn diagram summary of the number of patients/samples used in proteomics, phosphoproteomics, whole-exome sequencing, and RNA-Seq experiments. A total of 103 pairs of tumors and NATs were used for proteome profiling and whole exome sequencing (blue circle). 79 pairs of tumors and NATs were used for phosphoproteome profiling (red circle). 51 tumors and 49 NATs were used for RNA-Seq (yellow circle).
- (C) Correlation analysis of 25 HeLa cell samples as MS quality control to evaluate the robustness of label-free quantification. Top-right half panel: pairwise calculation of Spearman's correlation coefficients between 25 samples; bottom-left half panel: pairwise comparison of 25 samples by scatterplots. An average correlation coefficient of 0.93 was observed.
- (D) Correlation matrix of 103 tumor proteomes (Spearman's correlation coefficients). Red color indicates high correlation whereas blue color indicates low correlation (left). Distribution of protein abundances in tumors (orange) and NATs (blue) by a density plot. A unimodal distribution (dip test) was observed. All of samples passed proteomic quality control (right).
- (E) Number of proteins identified in 103 tumors (red dots) and NATs (blue dots).
- (F) Number of phosphosites identified in 79 tumors (red dots) and NATs (blue dots).

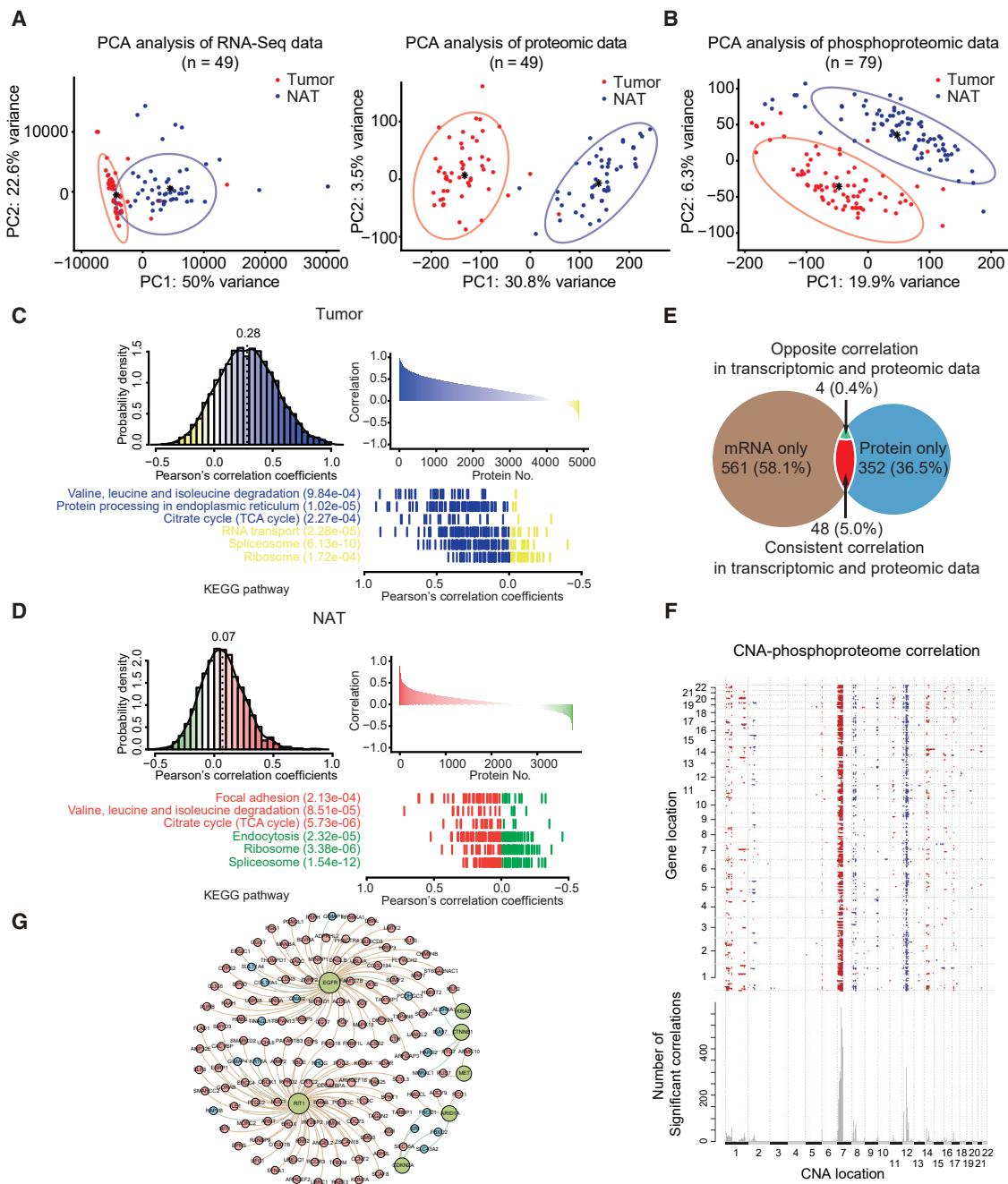


Figure S2. Integrative Analyses of Genomics, Transcriptomics, Proteomics, and Phosphoproteomics Data in LUAD Samples, Related to Figure 2

(A) Principal component analysis (PCA) of RNA-Seq (15,824 genes) and proteomic (8,252 proteins) data in 49 common patient samples. Red dots: tumors; blue dots: NATs.

(B) Principal component analysis (PCA) of phosphoproteomics data (2,130 phosphoproteins) in 79 patient samples. Red dots: tumors, blue dots: NATs.

(C) mRNA-protein correlation in tumors. Blue: pathways in which positively correlated genes were involved; yellow: pathways in which negatively correlated genes were involved.

(D) mRNA-protein correlation in NATs. Red: pathways in which positively correlated genes were involved; green: pathways in which negatively correlated genes were involved.

(E) Overlap of genes with prognostic power based on RNA-Seq or proteomic data (log-rank test, BH p < 0.01).

(F) Functional impacts of CNA on phosphoprotein. Top panel: Positive and negative correlations (BH p < 0.05, Spearman's correlation coefficient) were indicated by red and blue colors, respectively. Bottom panel: Number of phosphoproteins that were significantly associated with a specific CNA.

(G) Significant *cis*- or *trans*-effects of CNAs of driver genes on protein abundances. The network contains seven LUAD driver genes (green) with CNAs and their regulatory targets, where nodes and edges in pink represent positive correlations and those in blue represent negative correlations.

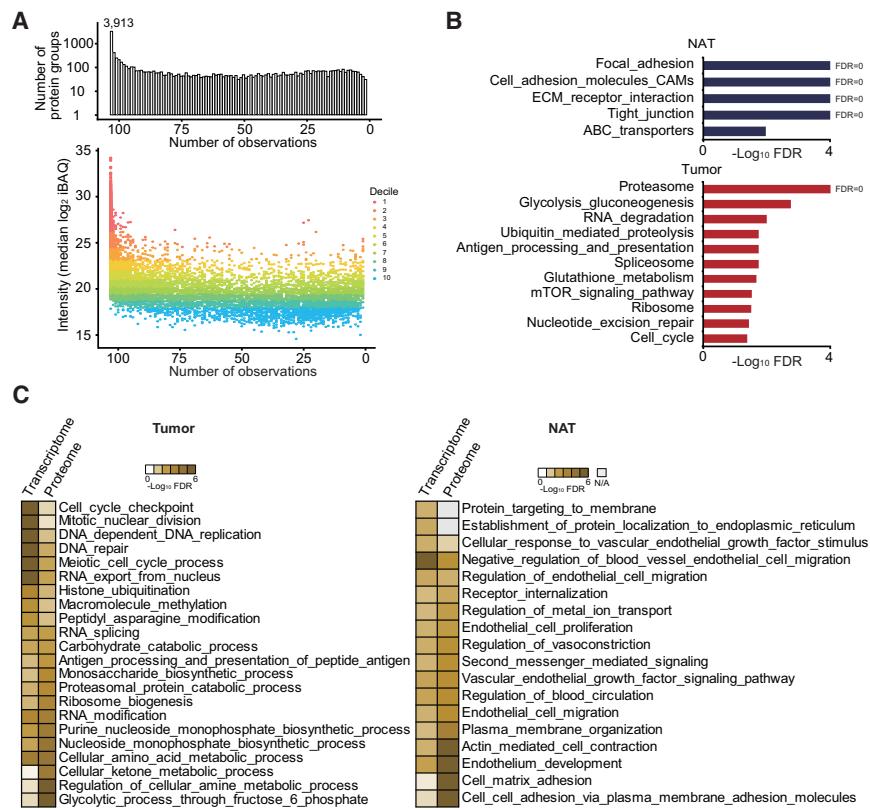


Figure S3. Proteomics Features of Tumors and NATs in LUAD, Related to Figure 3

(A) Distribution of protein abundance presented as median intensity in 103 patient samples. Bar plot on the top shows the number of proteins quantified in each patient sample.

(B) GSEA (C2: curated gene sets, KEGG subset of canonical pathways) revealed pathways that were significantly enriched in tumors and their matched NATs.

(C) GSEA (C5: GO gene sets, biological process) analyses of the RNA-Seq and proteomic data of 49 common patients revealed pathways that were significantly altered (FDR < 0.05) in tumors and NATs.

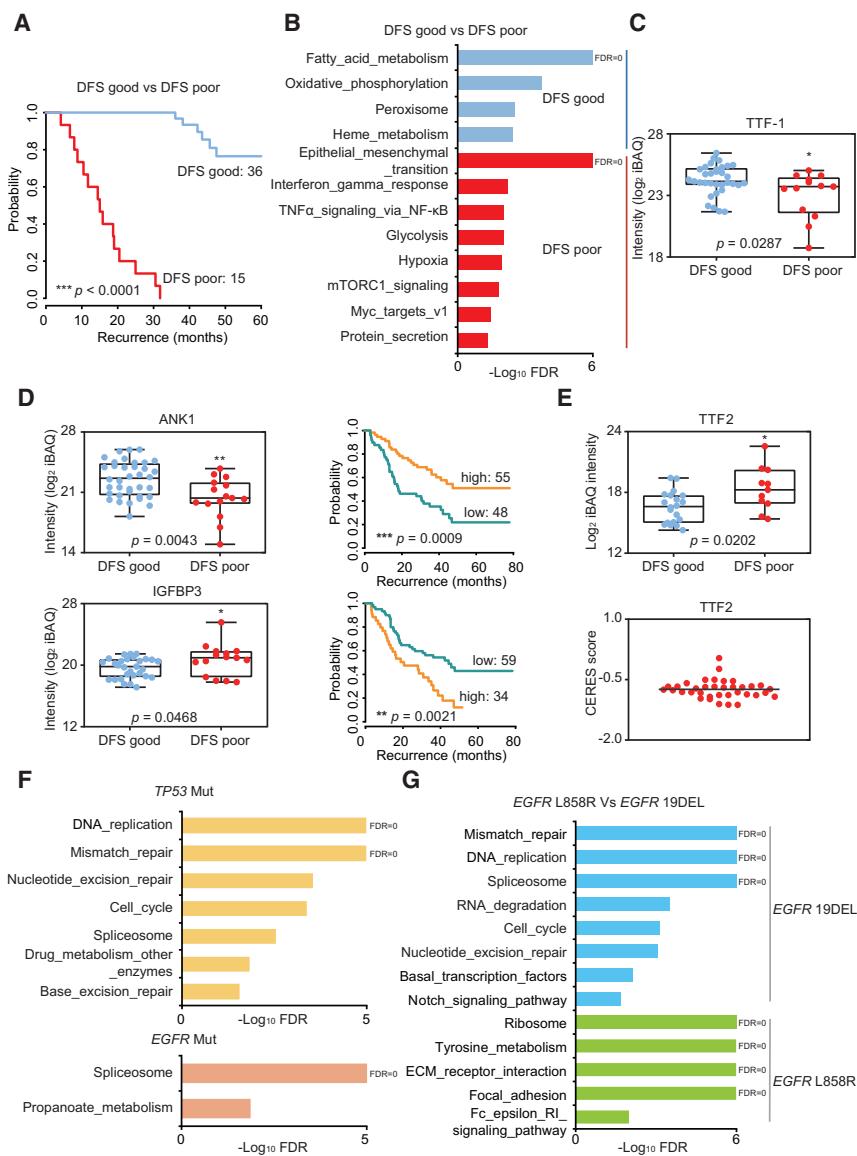


Figure S4. Proteomics Characteristics of TNM Stage I LUADs or LUADs with EGFR/TP53 Mutations, Related to Figure 4

- (A) Disease free survival analysis of patients in early stage LUAD with good or poor prognosis (p -value from log-rank test).
- (B) GSEA (H: hallmark gene sets) analysis of stage I LUAD patients revealed the pathways associated with good ($n = 36$, DFS time > 3 years) or poor ($n = 15$, DFS time < 3 years) prognosis.
- (C) Protein abundance of TTF-1 in DFS good and poor groups (p -value from Wilcoxon rank-sum test).
- (D) Protein abundances of ANK1 and IGFBP3 in DFS good and poor groups (p -value from Wilcoxon rank-sum test) and their associations with prognosis (p -value from log-rank test).
- (E) Protein abundance of TTF2 in DFS good and poor groups (p -value from Wilcoxon rank-sum test). CERES scores of TTF2 in 36 cell lines of LUAD in DepMap database (The black line illustrates the mean CERES score).
- (F) GSEA (C2: curated gene sets, KEGG subset of canonical pathways) revealed the enriched pathways in TP53 mutant patients in comparison with TP53 wild-type patients (top). GSEA revealed the enriched pathways in EGFR mutant patients in comparison with EGFR wild-type patients (bottom).
- (G) GSEA (C2: curated gene sets, KEGG subset of canonical pathways) revealed the enriched pathways in patients with EGFR 19DEL or L858R mutation.

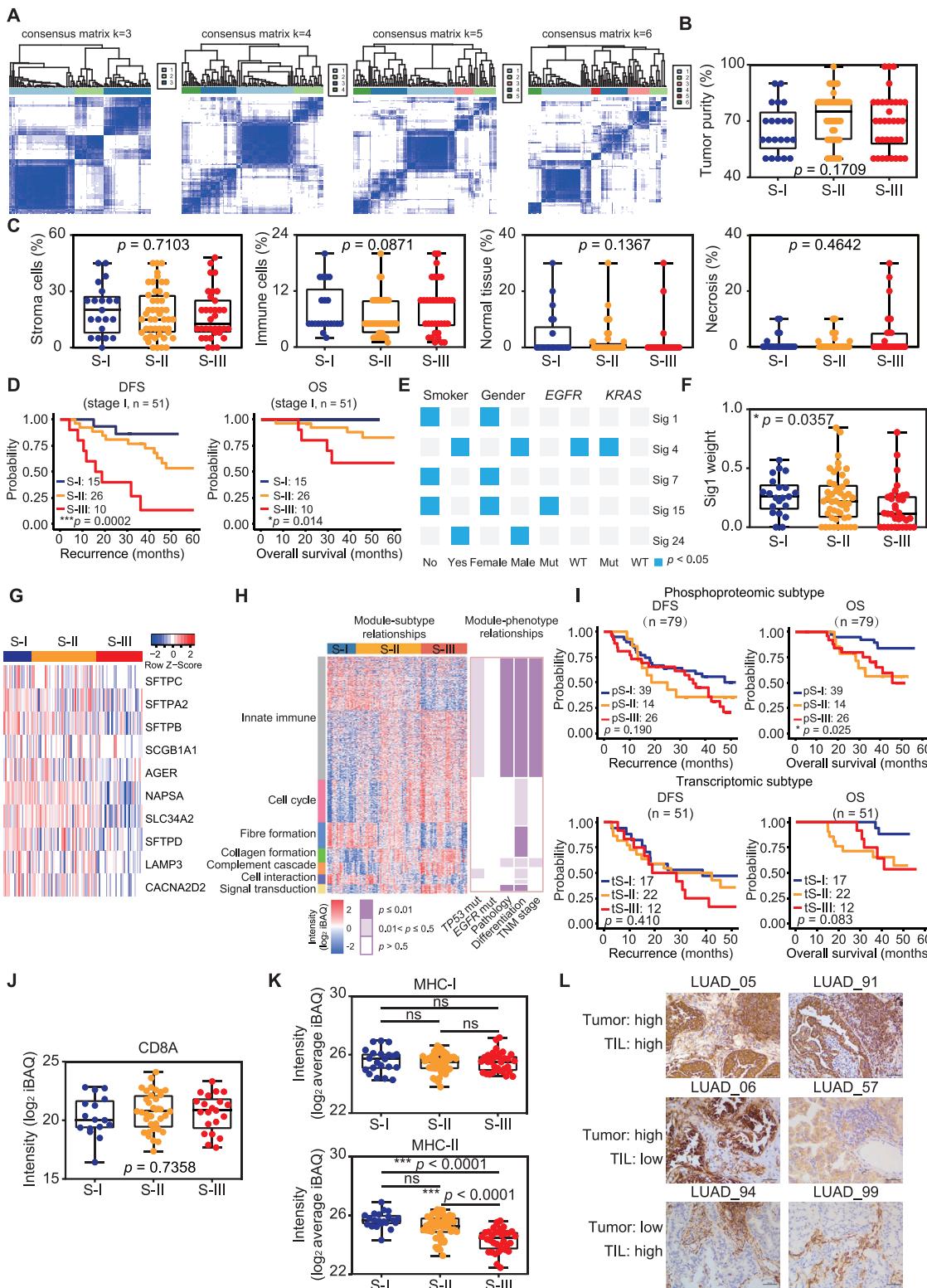


Figure S5. Molecular Subtyping of LUAD and Correlation with Clinical Outcomes, Related to Figure 5

(A) Consensus matrices of identified clusters (k = 3 to 6).

(B) Tumor purity in the three proteomic subtypes. (p-value from Kruskal-Wallis test).

(legend continued on next page)

-
- (C) Statistics of hematoxylin-eosin (HE) staining on non-tumor cell populations (immune cells, stroma cells, normal tissue and necrosis) in the three proteomic subtypes. (*p-value* from Kruskal-Wallis test).
- (D) Disease-free survival and overall survival of stage I patients (S-I, n = 15; S-II, n = 26; S-III, n = 10) among the three clusters (*p-value* from log-rank test).
- (E) Association of mutational signatures with clinical features or mutation status of two driver genes (*EGFR* and *KRAS*) (*p-value* from Wilcoxon rank-sum test).
- (F) Weights of COSMIC Signature 1 (*p-value* from Kruskal-Wallis test) in the three proteomic subtypes.
- (G) Abundances of ten lung signature proteins (identified in more than 50% of patients) in the three proteomic subtypes.
- (H) Seven modules (innate immune, cell cycle, fiber formation, collagen formation, complement cascade, cell interaction, and signal transduction) identified by WGCNA and their correlations with the three proteomic subtypes (left), and clinical variables (right).
- (I) Association of phosphoproteomic or transcriptomic subtypes with clinical outcomes. Top panels: overall survival (OS) and disease-free survival (DFS) analyses of the three phosphoproteomic subtypes (pS-I-III) in 79 patients; Bottom panels: overall survival (OS) and disease-free survival (DFS) analyses of the three transcriptomic subtypes (tS-I-III) in 51 patients (*p-value* from log-rank test).
- (J) Protein abundance of CD8A (*p-value* from Kruskal-Wallis test) in the three proteomic subtypes.
- (K) Average protein abundances of MHC-I and MHC-II (*p-value* from Wilcoxon rank-sum test) in the three proteomic subtypes.
- (L) Representative HLA-DR expression in cancer cells or in TILs of 103 patients. IHC analyses of HLA-DR in 103 tumors. FFPE sections were stained with HLA-DR antibody to determine MHC-II abundance in tumor cells or TILs. Three patterns were identified. Top panels: high HLA-DR in both tumor cells and TILs. Middle panels: high HLA-DR in tumor cells and low HLA-DR in TILs. Bottom panels: low HLA-DR in tumor cells and high HLA-DR in TILs. The scale bar indicates 50 μ m.

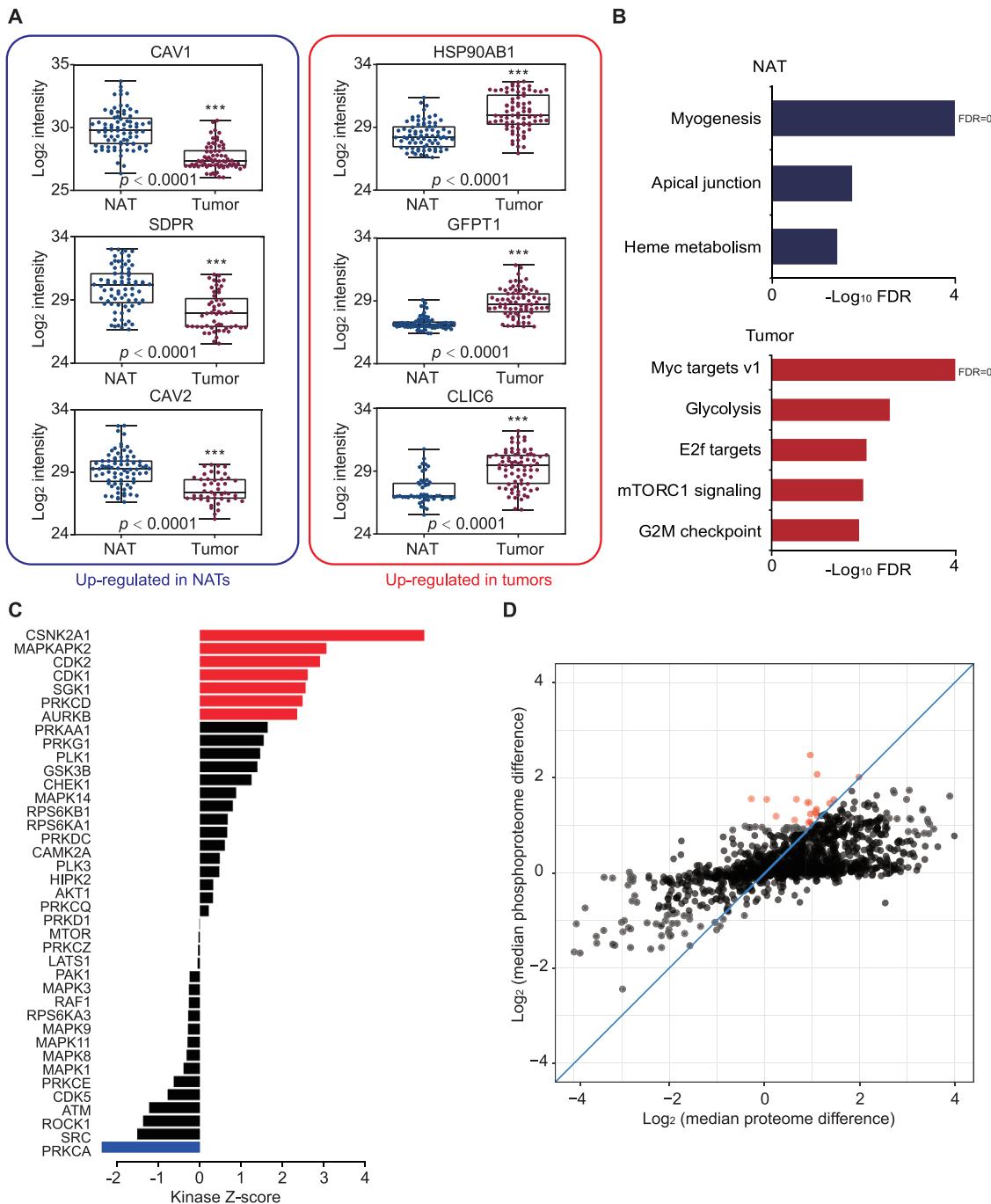


Figure S6. Phosphoproteomics Analysis of 79 LUADs, Related to Figure 6

(A) The top 3 upregulated phosphoproteins in either NATs (left panels) or tumors (right panels) (p -value from Wilcoxon signed-rank test).

(B) GSEA (H: hallmark gene sets) of differently altered phosphoproteins revealed pathways enriched in tumors and NATs.

(C) KSEA analyses of kinase activities in tumors and NATs.

(D) Fold-changes of proteins and phosphoproteins, and their correlations in tumors and NATs (Pearson's $r = 0.67$, $p = 2.2 \times 10^{-16}$). Red dots (defined as cancer-related phosphoproteins): phosphoproteins are greater than 2-fold changes in tumor than in NATs, and changes of phosphoprotein abundance are greater than changes of their corresponding protein abundance.

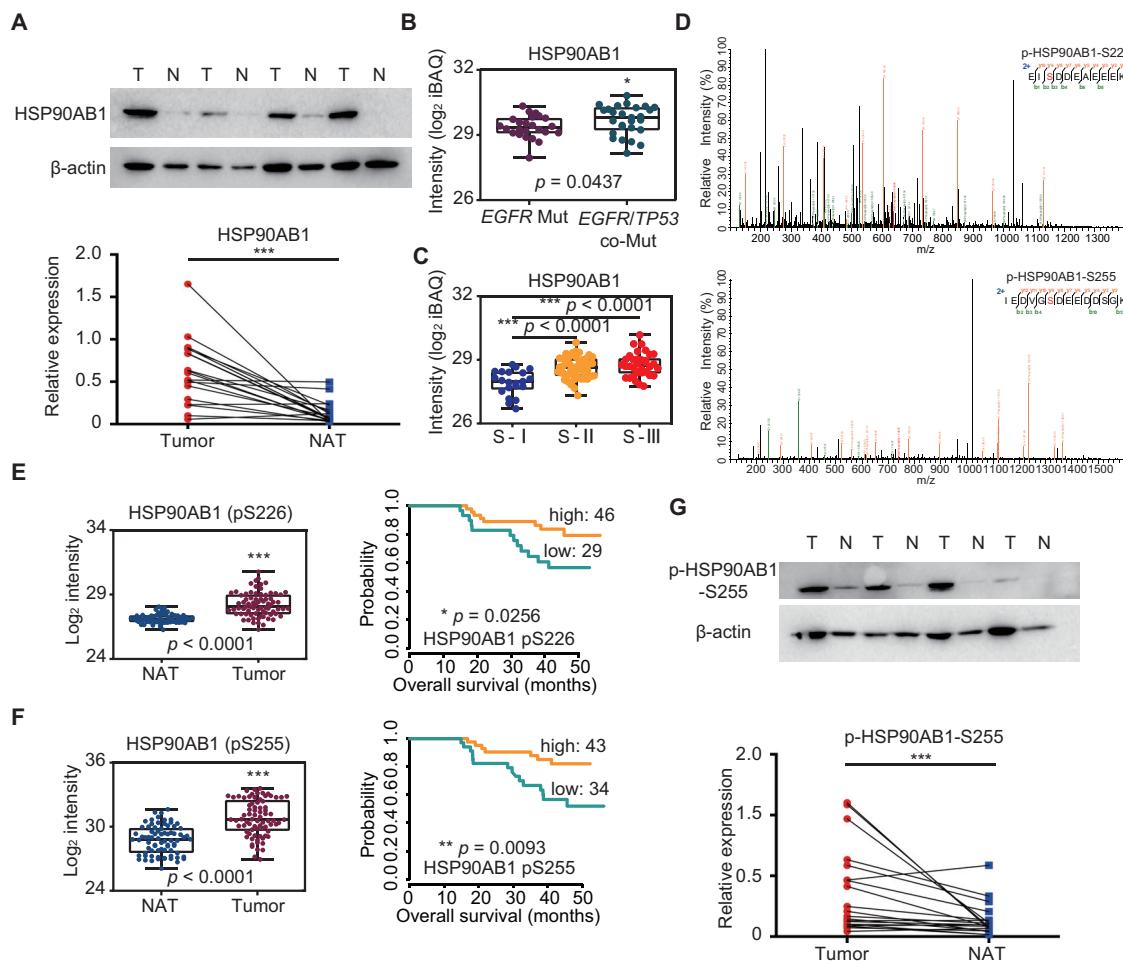


Figure S7. Screening Potential Prognostic Biomarkers and Drug Targets, Related to Figure 7

- (A) Representative western blot analysis of HSP 90 β in 20 independent LUAD patient samples collected with tumor (T) and their matched NATs (N) (top panel). Relative expression of HSP 90 β in 20 tumors and NATs (bottom panel) (*p*-value from Wilcoxon signed-rank test).
- (B) Protein abundance of HSP 90 β in 103 tumors with EGFR mutation or co-mutation of EGFR and TP53 (*p*-value from Wilcoxon rank-sum test).
- (C) Protein abundance of HSP 90 β in 103 tumors across the three proteomic subtypes (*p*-value from Wilcoxon rank-sum test).
- (D) MS/MS spectra of HSP 90 β S226 and S255 phosphorylation.
- (E) Phosphorylation level of HSP 90 β S226 in tumors or NATs (*p*-value from Wilcoxon signed-rank test), and their associations with clinical outcomes (*p*-value from log-rank test).
- (F) Phosphorylation level of HSP 90 β S255 in tumors or NATs (*p*-value from Wilcoxon signed-rank test), and their associations with clinical outcomes (*p*-value from log-rank test).
- (G) Representative western blot analysis of HSP 90 β S255 phosphorylation (corresponding antibody: phospho-HSP90AB1-S254) in 20 independent LUAD patient samples collected with tumor (T) and their matched NATs (N) (top panel). Relative expression of HSP 90 β S255 phosphorylation in 20 tumors and NATs (bottom panel) (*p*-value from Wilcoxon signed-rank test).