## RESEARCH

# A systematic dissection of the epigenomic heterogeneity of lung adenocarcinoma reveals two different subclasses with distinct prognosis and core regulatory networks

Chongze Yuan[1,2,3†], Haojie Chen[4†], Shiqi Tu[4†], Hsin-Yi Huang[5†], Yunjian Pan[1,2,3], Xiuqi Gui[4], Muyu Kuang[1,2,3], Xuxia Shen[3,6], Qiang Zheng[3,6], Yang Zhang[1,2,3], Chao Cheng[1,2,3], Hui Hong[1,2,3], Xiaoting Tao[1,2,3], Yizhou Peng[1,2,3], Xingxin Yao[1,2,3], Feilong Meng[7*], Hongbin Ji[5,8*], Zhen Shao[4*] and Yihua Sun[1,2,3*]

* Correspondence: feilong.meng@
sibcb.ac.cn; hbji@sibcb.ac.cn;
shaozhen@picb.ac.cn;
sun_yihua76@hotmail.com
†Chongze Yuan, Haojie Chen, Shiqi
Tu and Hsin-Yi Huang contributed
equally to this work.
[7]State Key Laboratory of Molecular
Biology, Shanghai Institute of
Biochemistry and Cell Biology,
Center for Excellence in Molecular
Cell Science, Chinese Academy of
Sciences, Shanghai 200031, China
[5]State Key Laboratory of Cell
Biology, Shanghai Institute of
Biochemistry and Cell Biology,
Center for Excellence in Molecular
Cell Science, Chinese Academy of
Sciences, Shanghai 200031, China
[4]CAS Key Laboratory of
Computational Biology, Shanghai
Institute of Nutrition and Health,
University of Chinese Academy of
Sciences, Chinese Academy of
Sciences, Shanghai 200031, China
[1]Department of Thoracic Surgery
and State Key Laboratory of Genetic
Engineering, Fudan University
Shanghai Cancer Center, Shanghai
200032, China
Full list of author information is
available at the end of the article

## Abstract

**Background:** Lung adenocarcinoma (LUAD) is a highly malignant and heterogeneous tumor that involves various oncogenic genetic alterations. Epigenetic processes play important roles in lung cancer development. However, the variation in enhancer and super-enhancer landscapes of LUAD patients remains largely unknown. To provide an in-depth understanding of the epigenomic heterogeneity of LUAD, we investigate the H3K27ac histone modification profiles of tumors and adjacent normal lung tissues from 42 LUAD patients and explore the role of epigenetic alterations in LUAD progression.

**Results:** A high intertumoral epigenetic heterogeneity is observed across the LUAD H3K27ac profiles. We quantitatively model the intertumoral variability of H3K27ac levels at proximal gene promoters and distal enhancers and propose a new epigenetic classification of LUAD patients. Our classification defines two LUAD subgroups which are highly related to histological subtypes. Group II patients have significantly worse prognosis than group I, which is further confirmed in the public TCGA-LUAD cohort. Differential RNA-seq analysis between group I and group II groups reveals that those genes upregulated in group II group tend to promote cell proliferation and induce cell de-differentiation. We construct the gene co-expression networks and identify group-specific core regulators. Most of these core regulators are linked with group-specific regulatory elements, such as super-enhancers. We further show that CLU is regulated by 3 group I-specific core regulators and works as a novel tumor suppressor in LUAD.

(Continued on next page)

(Continued from previous page)

**Conclusions:** Our study systematically characterizes the epigenetic alterations during LUAD progression and provides a new classification model that is helpful for predicting patient prognosis.

**Keywords:** Lung adenocarcinoma, Classification model, Epigenome, Super-enhancers, Core regulators

## Background

Lung cancer is a malignant tumor with highest mortality among all cancers worldwide and could be further classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [1]. Lung cancer exhibits high heterogeneity that enables adaptability, limits therapeutic success, and remains incompletely understood. Lung adenocarcinoma (LUAD), which accounts for most cases of NSCLC, can be classified into several histologic subtypes based on morphological characters [2]. These subtypes are frequently associated with patient prognosis, e.g., the micropapillary and solid predominant subtypes are predictors of poor prognosis [3, 4]. The histologic classification is most widely used in clinical practice but is also limited since it is a subjective judgment by pathologists. Other classification methods like radiological classification can also predict prognosis [5, 6], and with the development of targeted therapy and precision medicine, molecular classification is becoming more and more important [7].

Targeted therapies focusing on *EGFR* mutation, *ALK*-fusion, and *ROS1*-fusion have dramatically improved LUAD patient survival [8–10]. Although LUAD is one of the most heavily mutated cancers, most mutations, such as *TP53*, *KRAS,* and *STK11,* are not pharmaceutically targetable [11]. *EGFR* mutation is one of the most common driver genes, but can only explain about 14% of Caucasian LUAD patients and 50% of East-Asian patients [7, 12]. As a consequence, many patients are lacking therapeutic targets and need to be further characterized. Another important molecular classification is based on transcriptomic signatures. Transcriptomic classification not only defines distinct LUAD subclasses, but also reveals their relation with prognosis, critical biological features, and potential oncogenes [7, 13–15].

The study of cancer epigenetics has radically altered our views in cancer pathogenesis, providing new insights in biomarker development for risk assessment, early detection, and therapeutic stratification [16–19]. DNA methylation profiling of tumor tissues divides LUAD into distinct subsets: significantly altered CpG island methylator phenotype high (CIMP-H(igh)) group, normal-like CIMP-L(ow) group, and intermediate methylation group. Among them, DNA hypermethylation of several key genes, such like *CDKN2A*, *GATA2,* and *WIF1*, are observed in CIMP-H tumors [7]. On the other hand, histone modifications are also found to play important roles in LUAD [20–23]. The histone modification status is closely connected with chromatin configuration and *cis* element activity. However, it is also variable in tumors as it can be easily influenced by abnormal genomic rearrangements, mutations, and histone modification enzymes [24]. Recently, the Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenomics Consortiums have extensively characterized human regulatory landscape across a wide range of cell lines and tissues, expanding our understanding of cancer regulatory abnormalities [25, 26]. Although histone modifications are widely recognized as key

epigenetic regulators, rare study has been performed to profile and understand the dynamic histone modification patterns in primary tumor tissues of LUAD patients.

H3K27ac is closely linked with gene transcriptional activation and acts as a major histone mark of active regulatory elements, especially for enhancers and super-enhancers (SEs). In our previous study, we investigated H3K27ac landscape in two Chinese patient-derived LUAD cell lines and revealed SE-associated gene *RAI14* as a novel biomarker [27]. Here, we aim to further explore the dynamic H3K27ac landscape in LUAD tumors. We performed high-resolution chromatin immunoprecipitation sequencing (ChIP-seq) of H3K27ac in the tumors and normal lung tissues of 42 LUAD patients and uncovered a high intertumoral epigenetic heterogeneity. We classified the LUAD patients into two major subgroups based on the observed epigenetic heterogeneity and found they have distinct prognosis and transcriptomic features. Through an extensive co-expression network analysis, we defined the core transcriptional and epigenetic regulators for each subgroup, and identified *CLU* as a novel tumor suppressor from the downstream target genes of these core regulators in LUAD. Taken together, our study expands the understanding of LUAD complexity by a systematic analysis of epigenetic and transcriptomic signatures, providing important supplement to current histologic and molecular classifications.

## Results

### The epigenetic landscape differed between normal lung and LUAD tissues

To explore the epigenetic landscape in LUAD, we assembled a cohort of clinical samples from patients who received surgery in Fudan University Shanghai Cancer Center (FUSCC). The histological subtype and tumor cell content of each LUAD sample were confirmed by two independent pathologists. H3K27ac high-throughput chromatin immunoprecipitation sequencing (ChIP-seq) was then successfully performed in 42 paired LUAD and adjacent normal lung samples. Among the 42 LUAD samples, 1 sample was minimally invasive adenocarcinoma (MIA) and 41 samples were invasive adenocarcinomas (IACs) containing different subtypes (Additional file 2: Table S1).

We next processed the ChIP-seq data for each sample as previously described before further analysis [28, 29]. Briefly, after data quality control, reads alignment, peak calling, and filtering, in total we identified ~ 100 thousand H3K27ac peaks for LUAD samples and ~ 60 thousand peaks for normal samples, mostly located at intergenic and intronic regions (Additional file 1: Figures S1A and S1C). Interestingly, the LUAD and normal lung tissues can be well separated by their H3K27ac landscape, indicating that epigenetic alteration was one of the major differences and might play important roles in LUAD tumorigenesis (Fig. 1a, Additional file 1: Figure S1B). We grouped the H3K27ac ChIP-seq profiles of LUAD and normal lung samples, and then used the MAnorm2 model to compare these two groups of profiles [29]. By this method, we detected a huge number of differential H3K27ac sites, including 4784 tumor-specific and 7645 normal-specific ones (Fig. 1b, c). We found that genes regulated by the normal-specific sites were enriched for basic cellular functions, such as actin cytoskeleton organization and GTPase activity whereas the tumor-specific sites were significantly associated with a number of genes known to be dysregulated in cancers (Additional file 1: Figure S1D), indicating that aberrant epigenetic modifications directly contribute to transcriptional dysregulation in LUAD.

**Fig. 1** (See legend on next page.)

(See figure on previous page.)
**Fig. 1** Distinct H3K27ac profiles in LUAD and normal lung tissues. **a** Unsupervised hierarchical clustering of H3K27ac profiles for tumor and normal lung tissues of LUAD patients based on pairwise Pearson correlation coefficients (PCCs). **b** An MA plot of differential H3K27ac-modified sites between tumor and normal tissues, "M" represented log 2 fold change and "A" represented average log 2 signal intensities, sites with |M value| ≥ 1 and adjust *p*-value ≤ 0.05 defined as differential sites. **c** A heatmap of the H3K27ac signal in differential sites identified in **b**. The H3K27ac signal is represented as row-normalized z-scores. **d** Differential H3K27ac enrichment in super-enhancers (SE) between tumor and normal tissues. Each row represents an SE with a different enrichment between two tissues. SE scores are represented as row-normalized z-scores. Important differential SE-associated genes shown in the right. **e** Ranked plot for tumor-specific SE-associated TFs. IHC staining validated TFs are indicated with lines. **f** IHC staining results of 4 tumor samples showed tumor-specific SE-associated *SOX9* were highly expressed in tumor. **g** The functional enrichment of tumor-specific (left) and normal-specific (right) SE-associated genes. **h** Track plots of the H3K27ac signal distribution in tumor (top) and normal (bottom) samples across the *SOX9* (tumor-specific super-enhancer associated genes), *CAV1-CAV2* (normal-specific super-enhancer associated genes), and *MET* (other super-enhancer associated genes) loci. "SOX9-SE" represented this super-enhancer associated with *SOX9*. "CAV2&CAV1-SE" represented this super-enhancer associated with *CAV2* and *CAV1*. "MET-SE" represented this super-enhancer associated with *MET*. Heatmap of log2 fold change indicates the H3K27ac signal differences between tumor and normal tissues. **i** An example of super-enhancer hijacking. Number of junction reads from RNA-seq supported *EML4* and *ALK* gene fusion showed in the top left panel. Model of super-enhancer hijacking through chromosome translocation showed in the top right panel. Track plots of the H3K27ac signal distribution and gene expression in fusion and non-fusion samples across the *EML4* and *ALK* loci (bottom)

We and others have previously investigated the H3K27ac landscape in cell lines instead of tissue samples to explore LUAD epigenetic alterations [27, 30]. Strikingly, we found that the H3K27ac profiles of LUAD cell lines, including the widely used A549 cell line and two derived from Chinese patients, were quite different from those of primary tumor tissues (Additional file 1: Figure S1E). This indicated that observation obtained from cell lines may not accurately represent the in vivo situation and directly profiling the epigenetic landscape in primary tissues is of irreplaceable value to cancer studies. Taken together, these data uncovered the dramatic difference of the H3K27ac landscape between LUAD and normal lung tissues, which might be of functional importance.

### Super-enhancers control LUAD driver oncogenes

Previous studies have revealed that large clusters of enhancers, named super-enhancers (SEs), have prominent roles in determining cell identity, tumorigenesis, and chronic disease [28]. Compared to typical enhancers (TEs), SEs elicit stronger effects and predominantly exert a transactivation function to induce continuous and high expression of target genes [31]. We thus wanted to identify the SEs in our dataset and determine the role of SEs in LUAD tumorigenesis. To do so, we first separately identified SEs for each H3K27ac ChIP-seq sample using ROSE, a software specifically developed for this purpose [31, 32]. By this means, we identified ~ 800 and ~ 700 SEs for each tumor and normal sample, respectively, leaving the other distal H3K27ac sites as TEs (Additional file 1: Figure S1F). The SEs only accounted for a minor portion of the total enhancer domains (median, 5.58%) but for the majority of H3K27ac signals (median, 33.76%) (Additional file 1: Figure S1G). We merged the SEs identified from all tumors and normal samples and obtained 2,893 SEs in total. We further mapped the differential H3K27ac sites detected between LUAD and normal lung samples to these SEs and thus defined 205 tumor-specific and 324 normal-specific SEs (Fig. 1d; Additional file 1:

Yuan *et al. Genome Biology* (2021) 22:156

Page 6 of 26

Figure S2A - S2C). Interestingly, > 50% of the 11,574 differential H3K27ac sites at distal regions were located within SEs (Additional file 1: Figure S2D). This finding emphasizes that SE abnormalities might play an important role in LUAD development.

Generally, SEs and TEs regulate spatially closed genes, but they can also contact distant gene transcription start sites (TSSs) by forming long-range enhancer-promoter interactions [33]. While the most widely used method for annotating the target gene of each regulatory element is to map it to gene with the nearest TSS, we used a more reliable gene annotation method here which considered the correlation between the normalized H3K27ac intensities of each SE and those at gene promoters within 500 kb of the SE across all tumors and normal samples (Additional file 1: Figure S3A - 3C showing the SE at *MAX-FUT8* gene locus as an example). In this way, we found that a number of well-known oncogenes, including *MYC* and *SOX9*, were annotated as targets of tumor-specific SEs, while several tumor suppressor genes in LUAD, such as *DLC1* and *RB1*, were linked to normal-specific SEs (Fig. 1d; Additional file 3: Table S2, Additional file 4: Table S3). Among tumor-specific SE-associated genes, lots of genes were important transcriptional factors (TFs). Then, we ranked all the TFs and found *SOX9* was ranked as top one tumor-specific SE-associated TF (Fig. 1e). We performed immuno-histochemistry (IHC) staining of tumor-specific SE-associated transcription factor *SOX9* on patient tissue samples and found that higher level of SOX9 was observed in tumors than adjacent normal lung tissues (Fig. 1f). We also performed IHC staining of other top-ranked TFs including *SIX1*, *RUNX1*, and *SOX4* and again found higher expression level of these TFs in tumor tissues (Additional file 1: Figure S3E). The tumor-specific SE-associated genes were enriched in important pathways, such as chordate embryonic development (*RUNX2, FOXA1*), regulation of cell adhesion (*RUNX1, SOX2* and *VEGFA*), and epithelial cell differentiation (*SOX9, ELF3*) (Fig. 1g). Some of these pathways are related to cancer stem cell (CSC) or LUAD cancer cell migration and invasion, including chordate embryonic development, cell fate specification, regulation of cell adhesion, and collagen catabolic process [34, 35]. In particular, the activity of *SOX9* was known to be associated with the primitive transcriptional programs spanning stem cell-like to regenerative pulmonary epithelial progenitor states during metastasis in LUAD, consistent with a lot of epithelial differentiation and embryonic related pathways enriched in tumor-specific super-enhancer associated genes [34]. Genes associated with normal-specific SEs were enriched in cell growth regulation (*DLC1, RB1, GATA6*) and other basic cellular functions, such as cell morphogenesis, GTPase-mediated signal transduction and regulation of phosphate metabolic process (Fig. 1g). For example, *CAV1* and *CAV2* were marker genes of lung alveolar cells [36], and the H3K27ac levels at their nearby SEs were significantly upregulated in normal lung tissues compared to LUAD (Fig. 1h). Thus, in both normal lung and LUAD samples, SEs regulated key cell identity genes, and the SE abnormality contributed to tumorigenesis.

Previous studies have revealed that genomic rearrangement is highly associated with SE abnormality in cancers, and thus we focused on the most frequent fused oncogene *ALK* in LUAD [37]. *ALK*-fusion is a powerful driver mutant and important therapeutic target that occurs in ~ 5% of LUAD patients, especially in advanced stage patients [8]. The malignant behavior caused by *ALK*-fusion can be strongly inhibited by receptor tyrosine kinase inhibitors (TKIs), thus dramatically improved patients' prognosis [38]. Three of the 42 patients involved in our study were detected with *ALK-EML4* gene

fusion, which hijacked (or translocated) the SE located upstream of *EML4* (Fig. 1i). It was suggested that the fusion protein can activate several canonical signaling pathways, including RAS/MEK/ERK pathway and PI3K/AKT cascades [39], and our results indicated that the hijacked SE probably maintained aberrant expression of *ALK-EML4*. Similar SE hijacking was also found in one patient with *ROS1-SLC34A2* fusion (Additional file 1: Figure S3F). These data suggested that SE abnormality could also directly get involved in LUAD oncogenesis through SE hijacking of driver oncogenes, and thus, the hijacked SE might be a potential target for LUAD therapy, especially for TKIs-resistant patients [32].

## LUAD is composed of two different epigenetic states

Our data thus far highlighted the key epigenetic differences between normal lung and LUAD tissues. However, a high epigenetic heterogeneity was also observed among LUAD tissues from different patients (Fig. 1a). We next quantitatively evaluated the intertumoral variations of H3K27ac signals normalized by MAnorm2 on genome scale. In order to remove the mean-variance dependence, we fitted a mean-variance curve and selected those peaks with significant deviations above the curve [29]. By this method, we identified 4615 hyper-variable H3K27ac peaks (HVPs) among the tumor samples (Fig. 2a). We downloaded the susceptible SNPs of lung cancer identified by a previous genome-wide association study (GWAS) [40] and found they were highly enriched in the HVPs of LUAD samples (Additional file 1: Figure S4B). This finding suggests that the fluctuations of H3K27ac levels at these HVPs might provide valuable insights in understanding the epigenetic heterogeneity of LUAD. We further noticed that the vast majority of the tumor HVPs were not identified as HVPs across normal samples (Fig. 2b; Additional file 1: Figure S4C). It indicated that most of the intertumoral epigenetic heterogeneity observed at the tumor HVPs could hardly be explained by the epigenetic variations already existed in the patients' normal lung tissue and it predominantly emerged during tumorigenesis. Meanwhile, there was little overlap between the tumor HVPs and the peaks upregulated in tumor samples compared to normal samples (previous identified tumor-specific peaks) showing consistent H3K27ac changes between tumors and normal tissues (Fig. 2b), suggesting that a very large fraction of them may be peaks specific to a subset of tumor samples. Thus, further investigation on the observed intertumoral epigenetic heterogeneity was needed and might provide insights into LUAD progression.

We next performed principal component analysis (PCA) on the normalized H3K27ac intensities of tumor HVPs across all tumor ChIP-seq samples. Surprisingly, we found that, for each sample, its score on the first principle component (PC1) was significantly associated with the lymph node invasion status of corresponding patient (Fig. 2c). Hierarchical clustering of the tumor samples' PC1 scores clustered them (and thus the corresponding patients) into two subgroups, namely group I and group II (GI and GII, Fig. 2d). Of note, LUAD can be classified into several subtypes which are important predictors of prognosis by its predominant histologic component [41–43]. Based on that, we have divided the 42 LUAD patients in our cohort into three pathological subgroups: low-risk, median-risk, and high-risk (Additional file 2: Table S1). Here, after a comprehensive comparison of the pathological subtype, tumor size, and lymph node
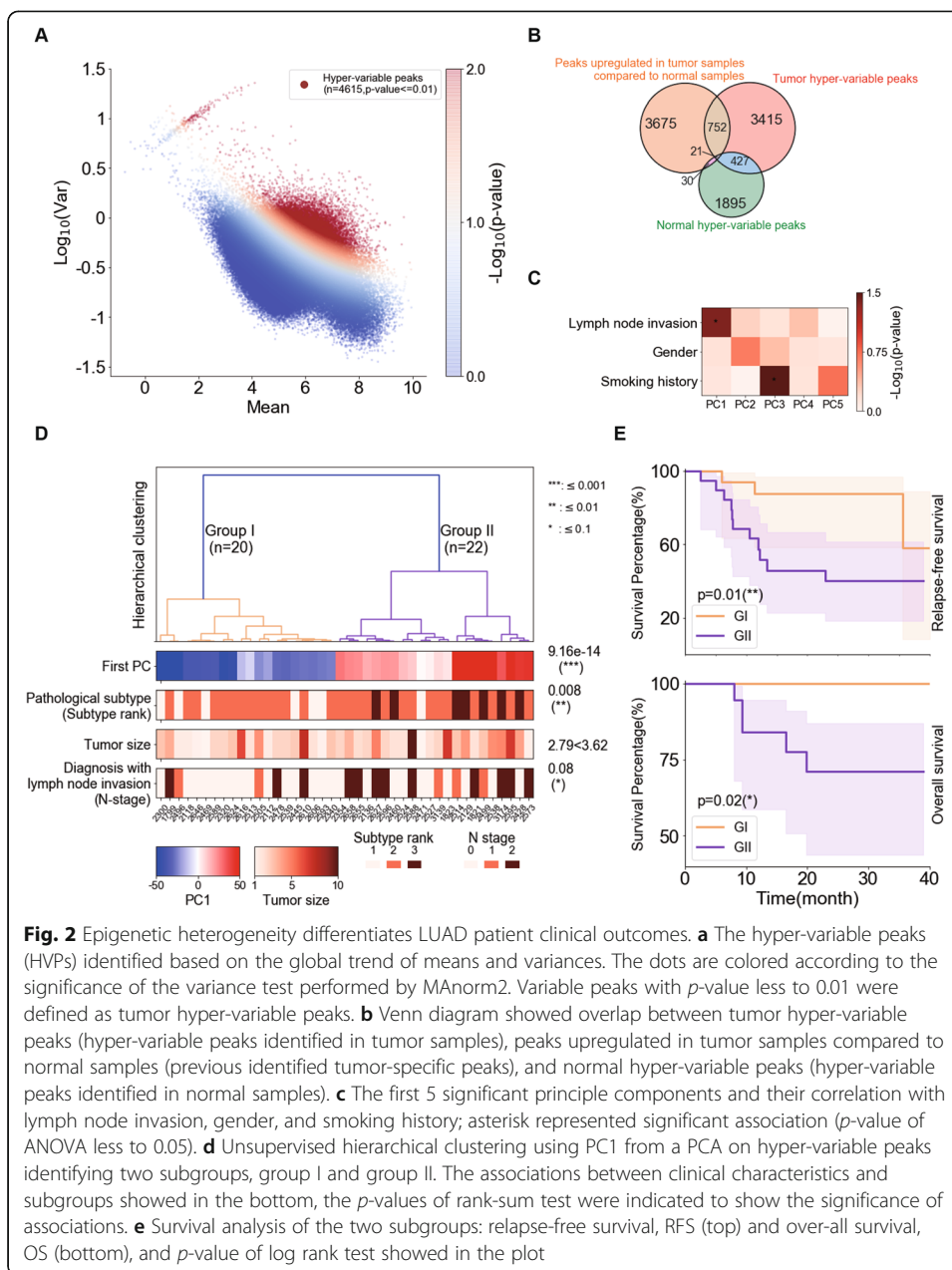
**Fig. 2** Epigenetic heterogeneity differentiates LUAD patient clinical outcomes. **a** The hyper-variable peaks (HVPs) identified based on the global trend of means and variances. The dots are colored according to the significance of the variance test performed by MAnorm2. Variable peaks with *p*-value less to 0.01 were defined as tumor hyper-variable peaks. **b** Venn diagram showed overlap between tumor hyper-variable peaks (hyper-variable peaks identified in tumor samples), peaks upregulated in tumor samples compared to normal samples (previous identified tumor-specific peaks), and normal hyper-variable peaks (hyper-variable peaks identified in normal samples). **c** The first 5 significant principle components and their correlation with lymph node invasion, gender, and smoking history; asterisk represented significant association (*p*-value of ANOVA less to 0.05). **d** Unsupervised hierarchical clustering using PC1 from a PCA on hyper-variable peaks identifying two subgroups, group I and group II. The associations between clinical characteristics and subgroups showed in the bottom, the *p*-values of rank-sum test were indicated to show the significance of associations. **e** Survival analysis of the two subgroups: relapse-free survival, RFS (top) and over-all survival, OS (bottom), and *p*-value of log rank test showed in the plot

metastasis status of patients in two subgroups, we observed a weak difference in tumor size and a significant difference in lymph node invasion between GI and GII patients (Fig. 2d). More importantly, we found that the tumor samples' PC1 scores were strongly associated with the patient pathological subtypes and could largely distinguish the more aggressive tumors (GII) from the less aggressive ones (GI) (Fig. 2d). High-risk pathological subgroup composed of micropapillary and solid predominant LUAD tumors was only contained in GII patients (Additional file 2: Table S1). In contrast, GI contained more low-risk pathological subgroup patients (Fig. 2d, Additional file 2: Table S1). Survival analysis of these two groups of patients further supported that GII patients had a significantly poorer prognosis than GI patients (Fig. 2e). In summary, by

Yuan *et al. Genome Biology* (2021) 22:156

Page 9 of 26

systematically dissecting the inter-tumor epigenetic heterogeneity, we were able to classify the LUAD samples into two subgroups that correlated with clinical outcomes.

In addition to clinical features, we also compared the gene mutation patterns between GI and GII patients. *EGFR* mutation occurred in > 50% of Asian LUAD patients and was the most frequent driver mutation in our cohort (27 of all 42 patients). However, we found no difference in the *EGFR* mutation frequency between GI and GII (Additional file 1: Figure S4E). Our previous studies revealed that patients with *ALK*-fusion are more likely to be in advanced stages, and the *EML4-ALK* variant 3 usually indicates a poor prognosis [44]. Consistently, all 3 *ALK*-fusion positive patients were found in GII and had a solid predominant subtype (Additional file 1: Figure S4E). In conclusion, based on the intertumoral heterogeneity of H3K27ac profiles, we successfully classified LUAD into two distinct subgroups with different pathological subtypes and clinical outcomes.

## Epigenomic and transcriptomic changes underlie distinct pathways and malignancy between two LUAD subclasses

Our data suggested that patients in GI had better prognosis than patients in GII, as evidenced by the significant prolonged relapse-free survival (RFS) and over-all survival (OS) time (Fig. 2e). To investigate the reasons underlying these differences in survival, we performed RNA sequencing (RNA-seq) of the tumor samples and determined the differentially expressed gene (DEG) signatures for each group. In total, we identified 2501 significant DEGs between GII and GI (Fig. 3a). Functional enrichment analysis revealed that GII-upregulated genes were enriched in the cell cycle and DNA replication pathways, both of which were signatures of highly malignant tumors. Besides, *FOXM1* pathway and *E2F* pathway which were both enriched in GII have been proven to be related to LUAD invasion and metastasis [45, 46]. Conversely, a number of genes involved in maintenance of normal cell functions and metabolic processes were significantly downregulated from GI to GII (Fig. 3b). Consistent with GI-specific genes signatures, several independent studies have shown that *KRAS*-driven murine LUAD preferentially arises from alveolar type II (AT2) cells, supported by the IHC staining of alveolar markers in early-stage tumors [47]. Surfactant metabolism-related pathways were usually highly activated in AT2 cells but significantly decreased in GII compared to GI, indicating that GI might somehow represent early-stage tumors. Next, we moved to test the epigenomic differences between GI and GII. After a quantitative comparison of the H3K27ac profiles between GI and GII tumors using MAnorm2, we defined 17, 713 GII-specific and 13,408 GI-specific H3K27ac peaks, which then led to the identification of 194 GII-specific and 437 GI-specific SEs. We further found that these SE were highly correlated with the genes differentially expressed between GI and GII (Fig. 3c, d; Additional file 1: Figures S6 and S7A). We chose two subgroup-specific SE-associated genes, *RUNX2* and *NKX2-1*(thyroid transcription factor 1, also known as *TTF1*), for IHC staining to assess their protein levels in GI and GII tumors. The LUAD maker *NKX2-1* was reported to suppress LUAD progression, whereas *RUNX2* was reported as key regulator to promote cancer progression [47, 48]. In our study, higher RUNX2 expression was detected in GII tumors compared to GI tumors, whereas NKX2-1 seemed to be specifically expressed in GI tumors (Additional file 1: Figure S7C and D).
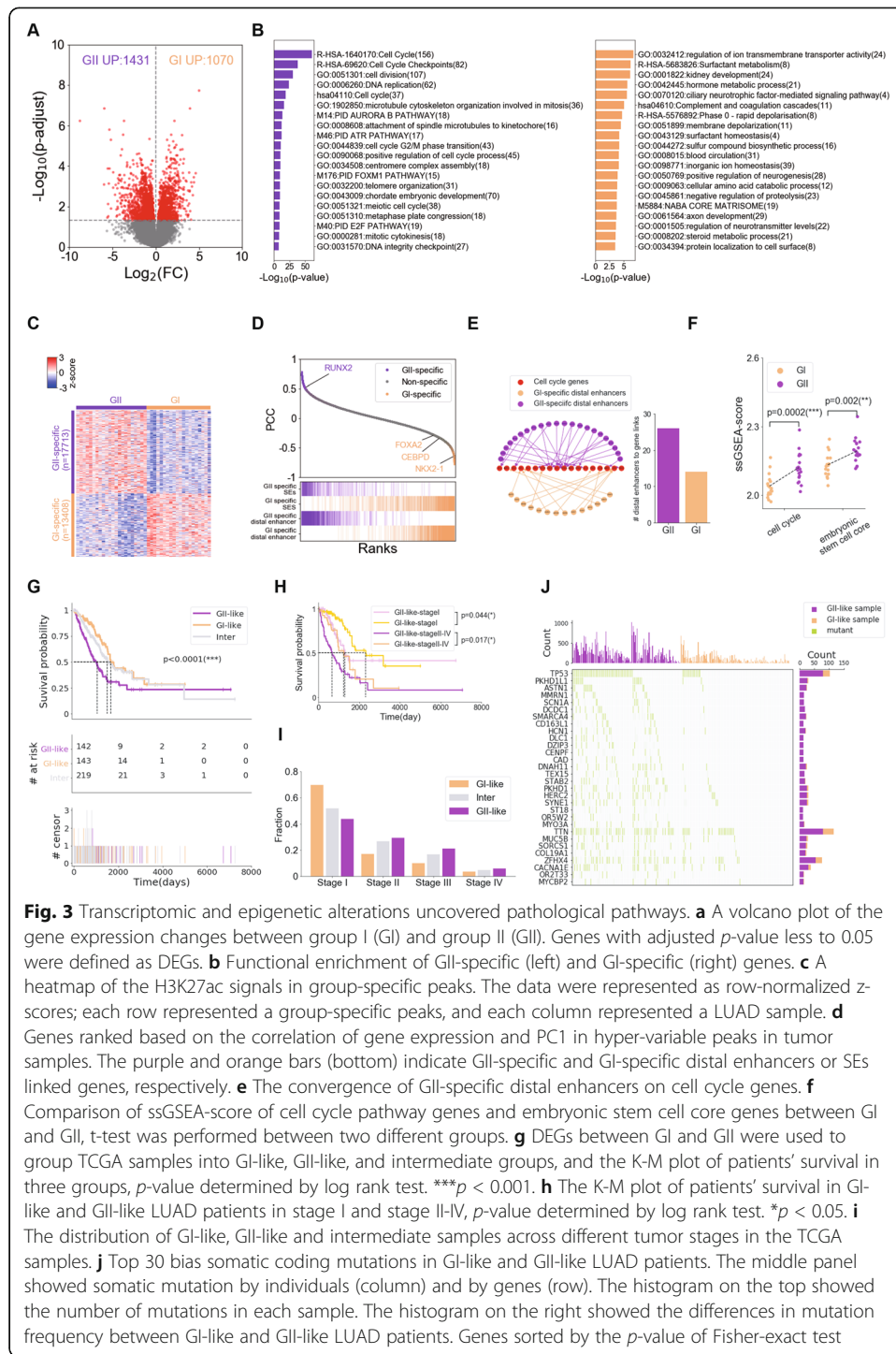
**Fig. 3** Transcriptomic and epigenetic alterations uncovered pathological pathways. **a** A volcano plot of the gene expression changes between group I (GI) and group II (GII). Genes with adjusted *p*-value less to 0.05 were defined as DEGs. **b** Functional enrichment of GII-specific (left) and GI-specific (right) genes. **c** A heatmap of the H3K27ac signals in group-specific peaks. The data were represented as row-normalized z-scores; each row represented a group-specific peaks, and each column represented a LUAD sample. **d** Genes ranked based on the correlation of gene expression and PC1 in hyper-variable peaks in tumor samples. The purple and orange bars (bottom) indicate GII-specific and GI-specific distal enhancers or SEs linked genes, respectively. **e** The convergence of GII-specific distal enhancers on cell cycle genes. **f** Comparison of ssGSEA-score of cell cycle pathway genes and embryonic stem cell core genes between GI and GII, t-test was performed between two different groups. **g** DEGs between GI and GII were used to group TCGA samples into GI-like, GII-like, and intermediate groups, and the K-M plot of patients' survival in three groups, *p*-value determined by log rank test. ****p* < 0.001. **h** The K-M plot of patients' survival in GI-like and GII-like LUAD patients in stage I and stage II-IV, *p*-value determined by log rank test. **p* < 0.05. **i** The distribution of GI-like, GII-like and intermediate samples across different tumor stages in the TCGA samples. **j** Top 30 bias somatic coding mutations in GI-like and GII-like LUAD patients. The middle panel showed somatic mutation by individuals (column) and by genes (row). The histogram on the top showed the number of mutations in each sample. The histogram on the right showed the differences in mutation frequency between GI-like and GII-like LUAD patients. Genes sorted by the *p*-value of Fisher-exact test

To further investigate the significant pathways changed with epigenetic alterations, we ranked genes by the correlation of gene expression with PC1 scores across all tumor samples and performed gene set enrichment analysis (GSEA). The analysis confirmed the positive enrichment of gene sets related to cell cycle and DNA repair. On the contrary, the negative enrichment of genes set was associated with immune-cell signaling related pathway, such as asthma, complement, and coagulation cascades in aggressive

tumors (Additional file 1: Figure S7B). These results indicated the strong connection between epigenetic alterations with transcriptional changes during LUAD progression. Accordingly, we found that more GII-specific enhancers were associated with cell cycle genes than GI-specific enhancers, indicating that epigenetic alterations might lead to group-specific transcription signatures (Fig. 3e; Additional file 1: Figure S8A).

We noticed that a number of lineage-specific TFs were upregulated in GI group and associated with GI-specific SEs, such as *NKX2-1* (LUAD marker), *FOXA2* (regulating surfactant protein production), and *CEBPD* (AT2 cell marker) (Fig. 3d). Loss of lineage identity during tumor development was shown in a recent study [47]. Thus, we used gene single sample set enrichment analysis (ssGSEA) to evaluate the activities of these two important pathways related to stem cell characteristics in GI and GII (Fig. 3f), and then performed GSEA analysis to verify the differential activation of stem cell related pathways between GI and GII (Additional file 1: Figure S8C and D). Compared with GI, the GII-upregulated genes were enriched in processes associated with rapid cell proliferation and a de-differentiated state (Additional file 1: Figure S8C and D). These results suggested that GI represented as more differentiated state while GII displayed a more stem cell-like phenotype. However due to the lack of temporal relationship between samples, we could not perform a trajectory analysis to check whether GI and GII corresponded to early and late-stage during tumor progression.

In order to investigate the epigenomic and transcriptomic changes in the continuum progression from differentiated state to de-differentiated state, we further divided GII tumor samples into GII.1 and GII.2 based on previous hierarchical clustering result and re-analyzed the epigenetic and transcriptomic changes among these three subgroups (Additional file 1: Figure S7E and F). GII.1 and GI had great differences in epigenome but not in transcriptome (Additional file 1: Figure S7G, third column), compared to the differential analysis between GII.2 and GI (Additional file 1: Figure S7G, first column). The epigenetic alterations in GII.1 compared to GI seemed like prior to the transcriptional changes between them during the progression. These results indicated that epigenetic regulators might be involved in the epigenome remodeling in GII.1 during the progression. Taken together, epigenetic heterogeneity led to the discovery of two LUAD subgroups and then different driver genes and biological pathways enriched in each LUAD subgroup, which can greatly help to explain the differences in clinical outcomes among patients.

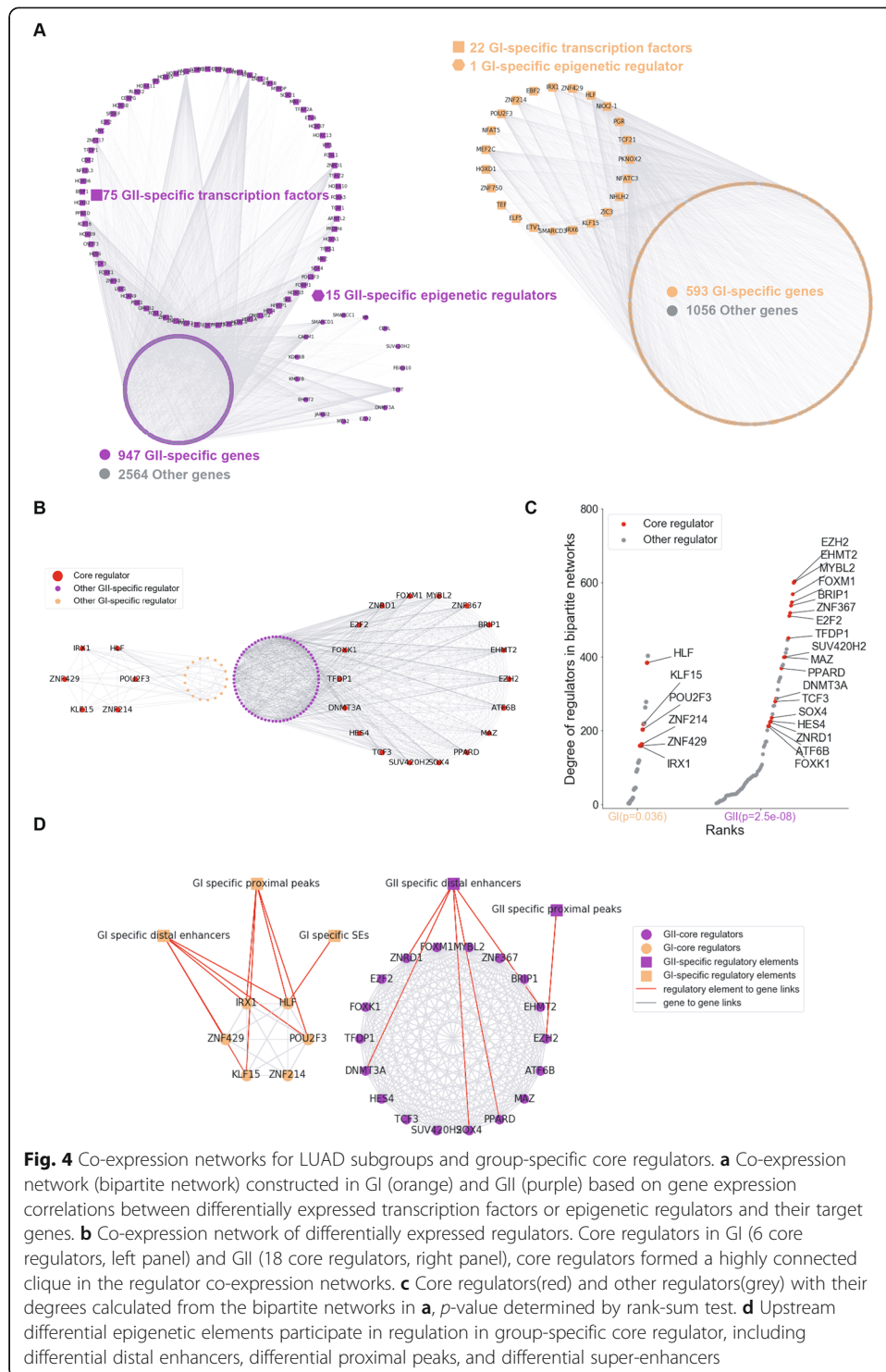### DEGs between GI and GII tumors can predict LUAD prognosis

To verify the utility of our LUAD classification model, we incorporated the transcriptomic data of 504 LUAD patients in TCGA database from diverse populations and tried to map each patient to GI or GII based on transcriptome similarity. We first calculated the ssGSEA scores of GI-specific and GII-specific genes in the RNA-seq profile of each TCGA-LUAD patient' tumor sample, which were called the GI and GII scores, respectively. We designated the TCGA patients with high GI score and low GII score samples as GI-like, and the patients with low GI score and high GII score samples as GII-like, leaving the remainder as the intermediate group (intergroup). In total, we defined 143 GI-like, 142 GII-like, and 219 intergroup patients. Consistently, GI-like patients had significantly better OS than GII-like patients (Fig. 3g). We also noticed that the

majority of GI-like patients were labeled as stage I patients by TCGA, while most GII-like patients were at later stages (II-IV, Fig. 3i). In particular, even for the stage I patients, the GII-like ones showed clearly worse prognosis compared to the GI-like ones (Fig. 3h). These findings indicated that our LUAD classification model could be potentially used as a reference to predict tumor progression and risk.

Additionally, genomic mutation analysis of TCGA-LUAD datasets revealed more gene mutations in GII-like patients than in GI-like patients (Fig. 3j). Especially, some tumor suppressor genes (TSGs) including *DLC1* and *SYNE1*, which have been found to be downregulated in our GII tumor samples compared to GI samples, were predominantly mutated in GII-like TCGA patients (Fig. 3j). This result indicated that loss of TSG function caused by gene mutation and downregulation can directly contribute to progression in tumorigenesis. Taken together, these results supported our new classification model and showed pronounced transcriptomic changes in these two epigenetic subgroups that correlate with the OS rate. The GI-GII score derivative from our classification model could be used to identify high-risk patients, especially in stage I, and, therefore, constituted an important supplement to current TNM staging system.

**Subgroup-specific core regulators determined transcription characteristics in GI and GII**

In order to explore key regulators driving the transcriptional differences between two subgroups, a co-expression network was constructed based on the differential expressed epigenetic regulators and transcription factors between GI and GII, which were considered regulators here, as well as their target genes identified by co-expression analysis. We found that the network formed two different sub-networks, which correspond to the two subgroups and were named as GI-specific network and GII-specific network, respectively (Fig. 4a). More importantly, GI-specific network contained 539 (50.4%) GI-specific genes, and GII-specific networks covered 947 (66.2%) GII-specific genes (Additional file 5: Table S4, Additional file 6: Table S5). These results indicated that the transcriptional characteristics of GI and GII were driven by different regulatory modules. We also found that the degrees of the transcription and epigenetic regulators showed a power law distribution (Additional file 1: Figure S9A), indicated that a small set of them had very high connectivity in the network, and thus might play important roles in driving the transcription characteristics of GI and GII. Then, we constructed the co-expression network of regulators and identified the core regulators of GI and GII, which not only had high connectivity with each other but also with other regulators (see methods). By this means, we identified 6 core regulators for GI and 18 core regulators for GII (Fig. 4b, c). In addition, 9 of the 18 GII-specific core regulators were annotated to cell cycle pathway (Additional file 1: Figure S9B), such as *MYBL2*, *FOXM1*, previously known to be related to cancer abnormal cell cycle [49]. In contrast, GI-specific core regulators were assigned to complement and coagulation cascades and metabolic related pathways (Additional file 1: Figure S9B), such as *POU2F3*, which has been identified as a good prognosis marker of LUAD. The better prognosis of high *POU2F3* expression patients was confirmed by The Human Protein Atlas [50], with a 46% 5-year survival rate compared to 34% 5-year survival rate of low *POU2F3* expression patients. The shift of core regulators might happen at the initial stage and drives the transition from GI to GII. We observed an upregulation of GII-specific core

**Fig. 4** Co-expression networks for LUAD subgroups and group-specific core regulators. **a** Co-expression network (bipartite network) constructed in GI (orange) and GII (purple) based on gene expression correlations between differentially expressed transcription factors or epigenetic regulators and their target genes. **b** Co-expression network of differentially expressed regulators. Core regulators in GI (6 core regulators, left panel) and GII (18 core regulators, right panel), core regulators formed a highly connected clique in the regulator co-expression networks. **c** Core regulators(red) and other regulators(grey) with their degrees calculated from the bipartite networks in **a**, *p*-value determined by rank-sum test. **d** Upstream differential epigenetic elements participate in regulation in group-specific core regulator, including differential distal enhancers, differential proximal peaks, and differential super-enhancers

regulators *EZH2* and *EHMT2* in GII.1 compared to GI, and they were further upregulated in GII.2 (Additional file 1: Figure S9C). We also detected a progressive downregulation of GI-specific core regulators *HLF* and *IRX1* (Additional file 1: Figure S9C). Compared with GI samples, GII.1 upregulated genes were highly enriched in epigenetic regulation of gene expression, consistent with the function of GII-specific core regulators

(Additional file 1: Figure S9D). Moreover, by integrating with our previously defined links from super-enhancers, distal enhancers and proximal H3K27ac peaks to gene s[51], a large fraction of the GI and GII-specific core regulators were found to be directly assigned to at least one differential epigenetic element (Fig. 4d). In summary, the transcriptional characteristics of the LUAD subgroups identified here were driven by a number of subgroup-specific core regulators, which potentially influenced the biological pathways related to clinical prognosis and contributed to LUAD progression. Meanwhile, subgroup-specific epigenetic elements acted as upstream factors driving the differential expression of these core regulators.

### Screening of genes regulated by GI-specific core regulators and in vivo validation reveals CLU as a potential tumor suppressor

After constructing a co-expression network dominated by the core regulators comprised of two subgroup-specific sub-networks, we asked whether the function of downstream genes regulated by these core regulators were directly linked with subgroup biological features. Making use of a previously compiled list of 49 TSGs whose expression was consistently downregulated in 11 different cancer types [52], we found that the GI core regulators target genes enriched more TSGs than GII core regulators target genes and the expression of these TSGs predominantly showed a positively correlation with GI-specific core regulators but a negatively correlation with GII-specific core regulators (Additional file 1: Figure S10A). Group-specific distal enhancers or SEs could control TSGs such as *DLC1* and *MAPK10* directly or indirectly via group-specific core regulators, which might help to explain the downregulation of these TSGs and tumor progression in GII (Fig. 5a). We then screened the genes directed linked with GI-specific core regulators to discover new candidate TSGs that can inhibit LUAD progression. As the top 3 GI core regulators were also regulated by SEs, we firstly searched the downstream genes co-regulated by *HLF*, *KLF15,* and *POU2F3* (Fig. 5b). A total of 38 protein-coding genes were identified (Additional file 7: Table S6). Interestingly, there were multiple well-known TSGs, including *MAPK10*, a well-studied TSG in esophageal cancer, cervical cancer, and breast cancer (Additional file 1: Figure S10B) [53–55]. To screen for progression associated TSGs, we further selected the candidate genes whose expression was only downregulated from GI-like to GII-like TCGA-LUAD tumor samples, but not from normal to GI-like ones. Three GI-specific genes were filtered out and we noticed *CLU*, one potential LUAD TSG which was included in our previous TSGs screening study [56]. During LUAD progression, *CLU* was significantly downregulated from GI-like to GII-like samples. However, there was no significant difference between normal and GI-like samples (Fig. 5c). The epigenetic analysis also showed that *CLU* was directly regulated by GI-specific enhancers (Fig. 5d). These results suggested that *CLU* might act as a TSG to inhibit LUAD progression. Our previous study showed that *CLU* knockout mediated by CRISPR/Cas9 technique in Kras$^{G12D}$-based genetically engineered mouse model (GEMM) promoted LUAD malignant progression [56]. We further found that ectopic *CLU* expression in CRL-5803 and PC9 cells significantly inhibited cell proliferation and suppressed colony formation in soft agar (Fig. 5e–g). Conversely, *CLU* knockdown in CRL-5872 cells accelerated cell proliferation and promoted colony formation in soft agar (Fig. 5h–j). We also assessed
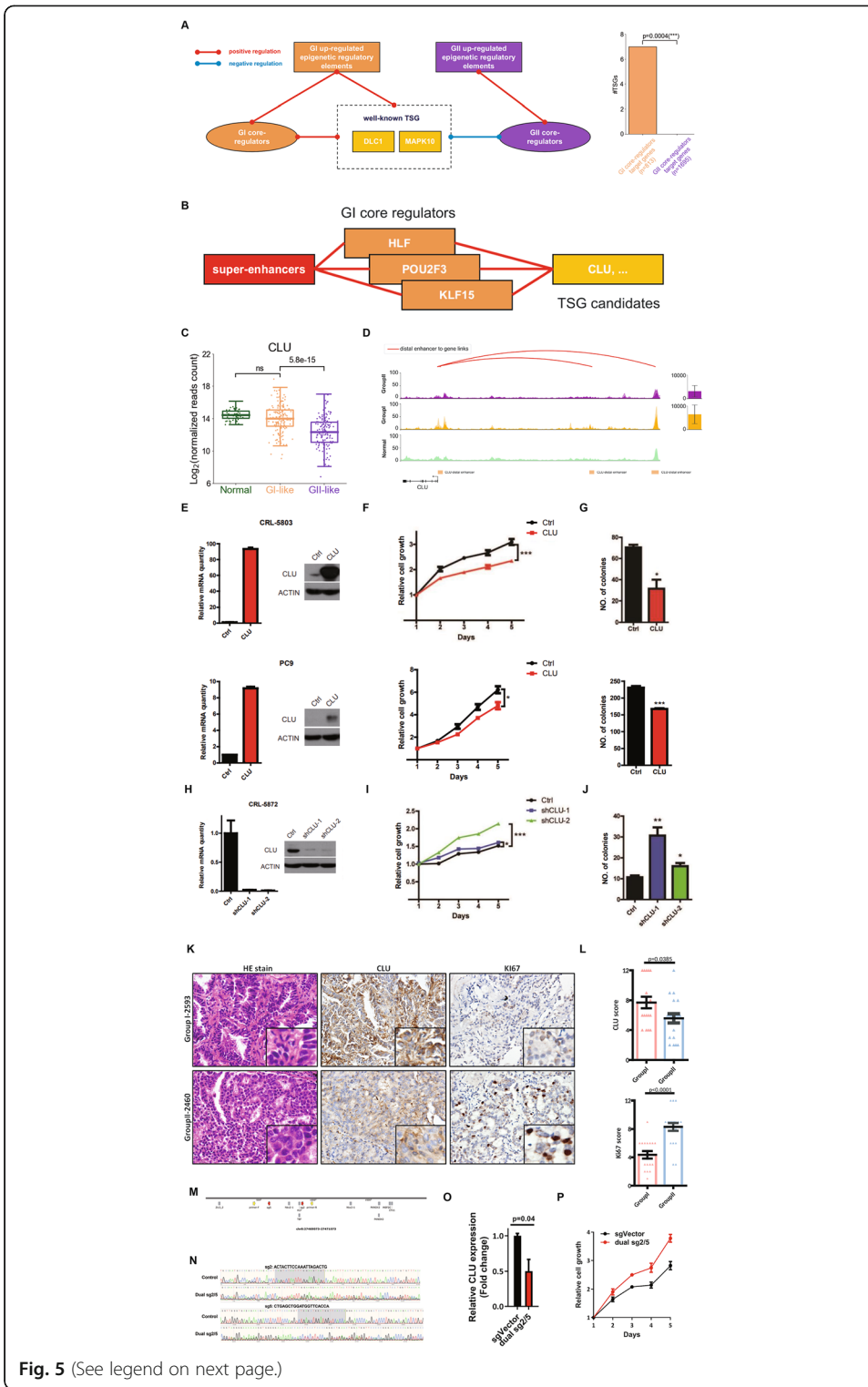
**Fig. 5** (See legend on next page.)

(See figure on previous page.)

**Fig. 5** In vivo and in vitro validation of potential function of identified TSGs. **a** A schematic diagram illustrated how core regulators and epigenetic regulatory elements (including H3K27ac-marked distal enhancers, promoters, and super-enhancers) regulated well-known TSGs. Bar plot in the right indicated that there were more active TSGs in GI core regulator target genes than GII core regulator target genes. The *p*-value of Fisher-exact test showed in the plot. **b** The strategy used for selecting TSG candidates. Three super-enhancer-associated GI core regulators co-regulated genes were selected as TSG candidates. **c** Gene expression level of *CLU* in GI-like, GII-like, and normal samples in TCGA-LUAD cohort. The *p*-value of t-test showed in the plot. ns, not significant. **d** Track plots revealed *CLU* gene expression was regulated by GI upregulated distal enhancers. The *CLU* gene expression in GI and GII samples was shown on the right. **e** Real-time PCR quantification and Western blot detection of *CLU* in CRL-5803 (upper panel) and PC9 (lower panel) cells with or without *CLU* ectopic expression. **f** Cell proliferation assay in CRL-5803 (upper panel) and PC9 (lower panel) cells with or without *CLU* overexpression. **g** Soft agar colony formation assay in CRL-5803 (upper panel) and PC9 (lower panel) cells with or without *CLU* overexpression. **h** Real-time PCR quantification and Western blot detection of *CLU* in CRL-5872 cells with or without CLU knockdown. **i** Cell proliferation assay in CRL-5872 cells with or without *CLU* knockdown. **j** Soft agar colony formation assay in CRL-5872 cells with or without *CLU* knockdown. **k** Representative photos of HE and IHC staining of CLU and Ki-67 in GI and GII samples. **l** Statistical analyses of CLU and Ki-67 IHC scores in group I and group II samples. Data were shown as mean with S.E.M. Statistical analyses was calculated by two-tailed, unpaired *t*-test. **m** The TF binding motifs on the enhancer region. Grey boxes indicate TFs; red arrows, predicted sgRNAs; and yellow arrows, PCR primers for testing knockout efficiency. **n** Sequencing of the PCR products by reverse (R) primers to validate dual gRNA knockout efficiency. **o** *CLU* expression of the indicated sgRNAs by real-time PCR quantification. **p** Cell proliferation analysis of the indicated sgRNAs

the protein levels of CLU as well as Ki-67, a proliferation marker, in our cohort using IHC staining. Compared with GI tumors, the CLU level was significantly downregulated in GII tumors. In contrast, the Ki-67 level was conversely upregulated in GII tumors (Fig. 5k, l). We here identified *CLU* as a new TSG in LUAD which was related to tumor progression.

We then further investigated the regulatory relationship between *CLU* and associated enhancers. Using dual gRNA CRISPR system [57], we successfully performed knockout of a GI-specific enhancer, which contains the binding motif of GI core regulator *HLF*, in CRL-5872 cells (Fig. 5m, n). Consistent with our *CLU* knockdown results, *CLU* enhancer knockout significantly downregulated the expression of *CLU* and accelerated cell proliferation (Fig. 5o, p). Taken together, these data support that GI tumors display relatively high expression of TSGs, which may be directly regulated by GI-specific enhancers to suppress malignant progression and contribute to better patient prognosis.

## Discussion

In this study, we described the active H3K27ac landscape of 42 LUAD samples and paired normal lung tissues. Our data added new knowledge on pathogenic LUAD mechanisms at the epigenetic level and proposed a classification model based on inter-tumoral H3K27ac heterogeneity. Although genomic classification is mostly related to the current target therapy of LUAD, a majority of patients are lacking known therapeutic targets and need to be further characterized. The transcriptomic classification is also widely used. Various classification models in LUAD cohorts revealing morphologic features, gene expression signatures, and gene mutation landscapes in each subgroup were reported [15, 58, 59]. Although these subclassifications vary considerably from study to study, most of them can effectively predict patient prognosis and provide new insight on LUAD progression [60]. Recently, proteogenomics studies revealed LUAD molecular signatures and therapeutic vulnerabilities, which provided new insights of

LUAD classification models [61, 62]. Other classification criterions like CpG island methylation were also investigated but still needed more comprehensive study [63].

Compared with previous methods, our H3K27ac-based classification provides new insights of epigenetic alterations in LUAD and investigates the transcriptomic features, co-expression network and core regulators which maintain the biological behavior in each subclass. Our epigenetic classification is highly associated with tumor histological features defined by previous reports and effectively predicts patient prognosis. Furthermore, our GI-GII score evaluating method can also effectively predict the prognosis of patients in the same TNM stage, thus could work as critical supplements to current TNM staging system in clinical application [7]. Plenty of biomarkers of LUAD have been found by our and others' previous studies including many oncogenes, but few of them are valuable therapeutic targets as most biomarkers lacking effective and specific inhibitors, thus hindering the application of these findings. The relevant epigenetic elements and upstream core regulators, which are identified in our study, seem to be alternative targets to inhibit the expression of oncogenes. *BRD4* and *CDK7* are key factors for the function of epigenetic elements such as enhancers, and their inhibitors, JQ1 and THZ1, globally inhibit the expression of associated genes and suppressed lung cancer cell proliferation [32, 64]. Recently, the FDA has approved tazemetostat, the first inhibitor of *EZH2* (a top core regulator in GII), for treating epithelioid sarcoma, indicating *EZH2* as a potential target in LUAD [65].

In our study, a transformation process might occur from GI to GII during tumor progression, including the loss of expression of the LUAD differentiation marker and gradually inactivation of TSGs. This potential transformation process might explain the invasive behavior and worse prognosis in GII. Similar transformation process has been reported by Lindsay et al. in mouse model [47], but the transformation process from GI to GII seems very complicated. The identification and intervention of early changes are important because the prognosis of lung cancer is completely different between early-stage and late-stage patients [66]. The existence of a GII.1 subgroup hints that epigenetic alterations might act as pioneers in lung cancer progression, consistent with a recent study [47]. However, further studies by more extensive investigations in larger cohorts, direct tracer study, and pioneer gene screening are necessary to confirm this hypothesis.

In this study, we have identified *CLU* as a potential LUAD TSG. Previous study has proposed controversial functions of *CLU* in tumorigenesis, e.g., *CLU* works as a proto-oncogene in prostate cancer and colorectal cancers whereas as a TSG in lung cancer [67–71]. In a functional screening of the TSGs using CRISPR/Cas9 knockout platform in *Kras*[G12D]-based GEMM, we have previously identified *CLU* as an important tumor suppressor. Interestingly, we here find that *CLU* is simultaneously regulated by 3 top SE-associated GI core regulators, highlighting the potential function of *CLU* in LUAD malignant progression. Our in vitro study further supports the tumor-suppressive function of *CLU*. Future efforts will be interesting to look into in-depth link between *CLU* and SE-associated core regulators.

There were some limitations of this study. Previous studies reported several forms of histone modification, mainly involving the methylation or acetylation on K4, K9, K27, K36, and K79 on histone3 [24]. Due to the limitation of tissue samples, our study only focuses on H3K27ac, a hallmark of active of enhancer and promoter. Of course, other

histone modifications such as H3K27me3 level will be interesting to explore in future in context of H3K27ac dynamic changes. We have used the transcriptomic profiles of a total 504 TCGA-LUAD patients without the H3K27ac profiling to validate our classification. More comprehensive study involving multiple histone modifications like H3K27me3 will be interesting to investigate in larger cohorts. Moreover, the traditional method of ChIP-seq to detect epigenetic changes requires high quality and quantity of tumor samples. However, the amount of available tissue is highly restricted especially in early-stage tumors and biopsy samples, which is also a main restriction to cohort size in such studies [72]. Future efforts will be interesting to take advantage of some newly developed techniques including CUT&Tag which requires only a small amount of samples or even single cells [73].

## Conclusions

We here describe the epigenetic alterations during LUAD tumorigenesis and provide a new classification model to predict patients' prognosis. Through comprehensively analyses of epigenomic and transcriptomic features, we have constructed the co-expression networks that determine subgroup-specific biological characteristics. We also reveal epigenetic modifications, especially super-enhancers, and core regulators in regulating tumor progression, which potentially serve as novel therapeutic targets of LUAD. Loss of function of various TSGs is prominent in GII group, which might promote the LUAD progression. Our data further identify *CLU* as an important TSG in inhibiting LUAD progression.

## Methods

### Sample collection

This study was approved by the Ethics Committee of Fudan University Shanghai Cancer Center. A consent form was signed by every patient who received surgical resection for LUAD or by his/her legal representative before surgery. The tumor samples and adjacent normal samples (> 2 cm from the tumor) were collected immediately after resection of the tumor and stored in liquid nitrogen before subsequent analyses. Tumor tissues were subjected to a pathological review to confirm the histology and the tumor cell content of each tumor tissue. Forty-two patients were included in this study.

### ChIP-seq in tissues

H3K27ac ChIP-seq was performed using an anti-H3K27ac antibody (abcam, ab4729). In brief, frozen LUAD and normal tissues were submersed in RPMI-1640 medium (Corning, 10-040-CVR), cut into small pieces, and homogenized on ice with a Dounce homogenizer. The tissue homogenate was filtered with a 70-μm cell strainer and fixed in 1% formaldehyde at room temperature for 10 min. The tissue pellet was collected through centrifugation, washed twice with cold PBS, and then incubated with lysis buffer on ice for 20 min. The tissue lysate was sonicated, and DNA was sheared to an average size of 100–300 bp. Approximately 10% of the total tissue lysate was collected as input, while the other portion was first precleared with uncoupled Protein-A Dynabeads (Novex, 10002D) and incubated with H3K27ac-coupled Protein-A Dynabeads for 6 h in a cold room. Then, the Dynabeads were collected using a magnet track, and

chromatin was eluted. The immunoprecipitant (IP) and input chromatin were incubated at 65 °C overnight and treated with RNase A and proteinase K, and then the DNA was purified (Tiangen, DP214-03). The DNA was then processed for NGS library construction and sequenced.

### Peak calling and classification of LUAD patients

H3K27ac ChIP-seq reads were mapped to the human reference genome hg19/GRCh37 using Bowtie v1.2.2. Only uniquely mapped reads were retained. Duplicated reads were removed before downstream analysis. H3K27ac peaks were called by using MACS v1.3.7, and the 30,000 most significant peaks associated with each sample were retained. All H3K27ac peaks were merged into a consensus list of peaks, and reads finally within these peaks were counted by using MAnorm2-utils [29]. Differential H3K27ac sites between tumor and normal samples (or between GII and GI) were identified using MAnorm2 [29]. Those sites with adjusted $p$-value below 0.05 and fold change greater than 2 were defined as differential ones.

To perform a classification of tumor ChIP-seq samples, we first identified peaks that were associated with hyper-variable signals across all the tumor samples by using MAnorm2 ($p$-value $\leq$ 0.01) [29]. These peaks were defined as tumor hyper-variable peaks (HVPs). Peaks in sex chromosomes were removed for downstream analysis, and principle component analysis (PCA) was performed on the tumor HVPs. Permutation analysis was used to determine significant principal components (PCs). Among the significant PCs, which were the first 5 PCs, only PC1 showed significant association with clinical information of patients (e.g., ANOVA test for lymph node invasion gave a $p$-value of 0.04). We therefore used only PC1 for hierarchically clustering the tumor samples. ConsensusClusterPlus algorithm was used to determine the optimal number of clusters [74], and we observed that k = 2 led to clearly more stable clustering results compared to k = 3 (Additional file 1: Figure S5A). Considering the limited cohort size, we have chosen k = 2 and have classified the tumor samples into group I (GI) and group II (GII).

### Peak saturation analysis

To better understand whether our epigenetic profiling adequately captured the epigenome landscape of paired primary LUADs and normal samples, we used the peaks of interest (not in the black list from https://www.encodeproject.org/files/ENCFF001 TDO/) in each sample to calculate the number of discrete peaks and the number of novel peaks gained with increasing sample number. This was performed across 100 permutations of the tumor and normal samples (Additional file 1: Figure S1A).

### Super-enhancer analysis

ROSE was employed to identify super-enhancers (SEs) in each sample. A 12.5-kb stitching distance was used to connect the proximal cluster of H3K27ac peaks into continuous enhancer regions, and regions around the TSS within 2.5 kb were excluded. We found that tumor-specific peaks and normal-specific peaks were mutually exclusive, and with increasing SE size (the length of SE), the fraction of differential H3K27ac sites was reduced. Thus, we used Fisher's exact test to define differential SEs based on the relative enrichment of differential H3K27ac sites compared to the background

(Additional file 1: Figure S2). SEs were mapped to genes based on the correlation between the SE score and H3K27ac signal intensities of the promoter (peaks within 2.5 kb of the TSS) or the gene expression level within ± 500 kb using a adjusted *p*-value of 0.01 and 0.05, respectively. Any SE without a linked gene was linked to the most highly correlated or closest gene. For each SE, SE score is the average normalized signal intensities of enhancers within this SE.

### Identification of distal enhancer-linked genes

To assign distal enhancers to genes, we used a correlation-based method. All possible interactions between the distal enhancers and TSSs within 500 kb were identified. For each interaction, we calculated the Pearson's correlation coefficient (PCC) between H3K27ac signal intensities and gene expression levels (Additional file 1: Figure S6A). And we constructed a null distribution of non-specific correlations using random pairs of distal enhancers and genes (not within ± 500kb or not in the same chromosome) to determine the significance of PCC (Additional file 1: Figure S6B). Then, we calculated the mean and standard deviation of these non-specific correlations. We used a normal distribution with the previously calculated mean and standard deviation to compute the *p*-value for each correlation and adjusted for multiple hypotheses using the Benjamini-Hochberg procedure (false discovery rate, FDR). Then, all correlations with an FDR below 0.05 were defined as distal enhancers linked genes.

### Pathway enrichment analysis

To identify the epigenetic signatures, GSEA was performed based on the PCC of PC1 and gene expression using the gseapy Python package, which implements GSEA on a preranked gene list, and the Molecular Signatures Database (MsigDB). For the functional characterization of differential enhancers-associated genes, DEGs, and differential SE-associated genes, we used DAVID [75] and Metascape [76] to identify pathways that were significantly represented in the gene list from our dataset.

### RNA-seq and data analysis

Tissue samples were homogenated in lysis buffer and RNA was extracted as user manual (AllPrep DNA/RNA Mini Kit, Cat.no 80204, Qiagen). RNA samples were then processed to library construction. RNA and library DNA quantity were evaluated with Qubit 3.0 and suitable kit. RNA and library size distributions were measured with Aglient Bioanalyzer 2100 or QIAxcel system. High-throughput sequencing was performed with Illumina Hi-seq X10 system. RNA-seq reads were aligned to hg19 human genome and transcript annotation (GENCODE). Gene raw read counts were calculated via feature counts in R package Subread. Differentially expressed genes (DEGs) analysis was perform using DESeq2 [77]. DESeq2 was run with the formula "~ group + gender", where group has levels "Group I" and "Group II" and "gender" has levels "male" and "female". Genes with adjusted *p*-value less than 0.05 defined as DEGs.

### TCGA sample classification based on RNA-seq

GI- and GII-specific genes were identified using DEseq2, and ssGSEA was used to calculate the sample wise gene set enrichment score of GI- and GII-specific genes in each

Yuan *et al. Genome Biology*     (2021) 22:156

Page 21 of 26

TCGA sample (called the GI score and GII score, respectively). We used 1/3 and 2/3 quantiles of the GI score and GII score across all samples to divide the TCGA samples into three groups, high GI score and low GII score samples as GI-like group, low GI score and high GII score samples as GII-like group, and the remainder as the intergroup. Then, the R package survminer was used to perform a survival analysis of these groups. Gene expression matrix of TCGA-LUAD was downloaded via TCGA-biolinks [78].

### Co-expression networks construction

We identified transcription factors(TFs) and epigenetic regulators based on functional annotation databases [74, 79] and selected those were DEGs as regulators, we used these regulators as key nodes to expand the networks based on significant positive correlation in gene expression (FDR ≤ 0.01) of these regulators and other genes. Co-expression network visualized using Python package networkx. To identify core regulators among these differentially expressed regulators, which not only had high connectivity with each other but also with other regulators, the co-expression network of differentially expressed regulators was constructed based on significant gene expression correlation (FDR ≤ 0.01). Core regulator clique was defined as regulator clique with the highest clique score in the regulators co-expression network and regulators in this clique were defined as core regulators. Two important definitions: "regulator score" was defined as the number of cliques that in which regulator participated. "clique score" for a clique was defined as the sum of "regulator score" of all regulators in that clique. We first calculated the "regulator score" of each regulator in the regulators co-expression network, then we calculated the "clique score" of each clique in this network, finally selected regulators from the clique with highest "clique score" as core regulators.

### Cell culture and functional assay

The HEK-293T cell and human NSCLC cell lines including PC9, CRL-5803, and CRL-5872 were purchased from the American Type Culture Collection and cultured in DMEM supplemented with 8% FBS. All cell lines used in this study were mycoplasma-free. The HEK-293T cells were used for lentivirus production for ectopic *CLU* expression and *CLU* knockdown as previously described [56]. MTT assay was performed following the manufacture manual. Briefly, 2000 cells were seeded on 96-well plates, and the viability of cells was measured daily for 5 days. All experiments were performed in triplicate. The CRL-5803 and PC9 cells were virally infected for ectopic *CLU* expression and the CRL-5872 for sh*CLU* knockdown. These cell lines together with the parental cells were then used for MTT assay and the colony formation in soft agar. For soft agar assay, 5000 cells were mixed with 0.2% agar containing growth medium and layered onto 1% agar containing growth medium in 6-well plates. Medium were changed every 3 days and the colonies were stained with 0.005% crystal violet and counted in 3 weeks.

### Plasmid construction

For gene expression in cell lines, the coding sequences of CLU were PCR amplified and cloned into pCDH-puro vector. For dual gRNA CRISPR screening, we insert another

Yuan *et al. Genome Biology* (2021) 22:156

Page 22 of 26

U6-filler-TA into pSECC, thus two gRNAs can be cloned into one vector. Briefly, the first gRNA, sg2, was cloned into BsmBI sites of pSECC vector. For the second gRNA, sg5, the U6-filler-TA sequence was amplified from pSECC and inserted into the vector pCDNA3.1 between EcoRI and XhoI which we then called pCDNA3.1-linker. The sg5 was cloned into BsmBI sites of pCDNA3.1-linker and then the U6-sg5-TA was cloned into pSECC-sg2 using ECoRI and XhoI. All primers used are listed as follows: For q-RT PCR analysis: qF-CLU: CCAATCAGGGAAGTAAGTACGTC, qR-CLU: CTTGCGCTCTTCGTTTGT TTT. For PCR sequencing: primer-F: CCATTCAGAACTAGGTTCTGACC, primer-R: GTGGCCTCTGTGTGCTTGTCT. Dual sgRNA target sites: sg2: ACTACTTCCAAAT TAGACTG, sg5: CTGAGCTGGATGGTTCACCA. ShRNA target sites: shCLU-1: AAC CAGAGCTCGCCCTTCTAC, shCLU-2: AGCAGCTGAACGAGCAGTTTA.

## Quantitative real-time PCR

Total RNA was extracted using TRIzol reagent (Invitrogen) and retrotranscribed into first strand cDNA using PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa). The cDNAs were subjected to real-time PCR with gene-specific primers using SYBR-Green Master PCR mix (Roche). Actin was served as internal controls.

## Immunohistochemistry (IHC)

IHC was performed as described previously [80]. Paraffin-embedded clinical tissues were incubated with the following antibodies: anti-SOX9 (abcam, ab185966; 1:1000 dilution), anti-RUNX2 (abcam, ab192256; 1:1000 dilution), anti-TTF1 (abcam, ab76013; 1:250 dilution), anti-Ki67 (Proteintech, 27309-1-AP; 1:5000 dilution), anti-SOX4 (abcam, ab243041, 1:1000 dilution), anti-RUNX1 (abcam, ab240639, 1:2000 dilution), anti-SIX1 (abcam, ab252224, a:100 dilution), and anti-Clusterin (abcam, ab92548; 1:200 dilution). The immunostaining was reviewed and scored blindly. The scoring system for grading expression level was reported previously [81]. The score of each sample was multiplied by the grading of intensity and staining area.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-021-02376-1.

---

**Additional file 1: Supplementary figures.** All figures explain the content of this study but were not listed as figures in the article.

**Additional file 2: Table S1.** Sample clinical information.

**Additional file 3: Table S2.** Normal tissue specific SEs and associated genes.

**Additional file 4: Table S3.** Tumor tissue specific SEs and associated genes.

**Additional file 5: Table S4.** The co-expression network constructed based on the GI specific regulators.

**Additional file 6: Table S5.** The co-expression network constructed based on the GII specific regulators.

**Additional file 7: Table S6.** 38 protein-coding genes regulated by top 3 GI regulators.

**Additional file 8: Table S7.** Hyper variable peaks identified in tumor samples.

**Additional file 9: Table S8.** Hyper variable peaks identified in normal samples.

**Additional file 10: Table S9.** Enrichment results of normal specific SEs associated genes by metascape.

**Additional file 11: Table S10.** Enrichment results of tumor specific SEs associated genes by metascape.

**Additional file 12.** Review history.

---

**Review history**
The review history is available as Additional file 12.

**Peer review information**
Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Experimental methods**
All the experimental methods in this study comply the Helsinki declaration.

**Authors' contributions**
Conception and design: CY, FM, ZS, and YS. Development of methodology: CY, HC, ST, HHu, FM, ZS, and HJ. Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): CY, HC, HHu, YPa, XG, MK, XS, QZ, CC, HHo, XT, YPe, and XY. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, bioinformatic analysis): CY, HC, ST, and HHu. Writing, review, and/or revision of the manuscript: CY, HC, ST, HHu, HJ, ZS, and YS. The authors read and approved the final manuscript.

**Availability of data and materials**
All the raw sequencing data of ChIP-seq and RNA-seq generated in this study have been deposited into the European Genome-phenome Archive (EGAD00001007066, https://ega-archive.org/datasets/EGAD00001007066) [82]. The transcriptomic data, survival data, and genomic mutation data of TCGA-LUAD here is in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

## Declarations

**Ethics approval and consent to participate**
This study was approved by the Committee for Ethical Review of Research (Fudan University Shanghai Cancer Center Institutional Review Board No. 090977-1). Informed consents of all patients for donating their samples to the tissue bank of Fudan University Shanghai Cancer Center were obtained from patients themselves or their relatives. Informed consent of all patients for publication was obtained from patients themselves or their relatives.

**Consent for publication**
Not applicable.

**Competing interests**
No potential conflicts of interest were disclosed.

**Author details**
[1]Department of Thoracic Surgery and State Key Laboratory of Genetic Engineering, Fudan University Shanghai Cancer Center, Shanghai 200032, China. [2]Institute of Thoracic Oncology, Fudan University, Shanghai 200032, China. [3]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China. [4]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [5]State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China. [6]Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. [7]State Key Laboratory of Molecular Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China. [8]School of Life Science and Technology, Shanghai Tech University, Shanghai 200120, China.

## References
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7–30.
2. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol. 2011;6:244–85.
3. Tsutsumida H, Nomoto M, Goto M, Kitajima S, Kubota I, Hirotsu Y, et al. A micropapillary pattern is predictive of a poor prognosis in lung adenocarcinoma, and reduced surfactant apoprotein A expression in the micropapillary pattern is an excellent indicator of a poor prognosis. Mod Pathol. 2007;20(6):638–47. https://doi.org/10.1038/modpathol.3800780.

4.    Zhang Y, Wang R, Cai D, Li Y, Pan Y, Hu H, et al. A comprehensive investigation of molecular features and prognosis of lung adenocarcinoma with micropapillary component. J Thorac Oncol. 2014;9(12):1772–8. https://doi.org/10.1097/JTO.0000000000000341.

5.    Ye T, Deng L, Wang S, Xiang J, Zhang Y, Hu H, et al. Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype. J Thorac Oncol. 2019;14(4):617–27. https://doi.org/10.1016/j.jtho.2018.12.030.

6.    Fu F, Zhang Y, Wen Z, Zheng D, Gao Z, Han H, et al. Distinct prognostic factors in patients with stage I non-small cell lung cancer with radiologic part-solid or solid lesions. J Thorac Oncol. 2019;14(12):2133–42. https://doi.org/10.1016/j.jtho.2019.08.002.

7.    Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543–50. https://doi.org/10.1038/nature13385.

8.    Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561–6. https://doi.org/10.1038/nature05945.

9.    Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012;13(3):239–46. https://doi.org/10.1016/S1470-2045(11)70393-X.

10.   Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. N Engl J Med. 2020;382(1):41–50. https://doi.org/10.1056/NEJMoa1913662.

11.   Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010;466(7308):869–73. https://doi.org/10.1038/nature09208.

12.   Shi Y, Li J, Zhang S, Wang M, Yang S, Li N, et al. Molecular epidemiology of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology - Mainland China subset analysis of the PIONEER study. PLoS One. 2015;10(11):e0143515. https://doi.org/10.1371/journal.pone.0143515.

13.   Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A. 2001;98(24):13790–5. https://doi.org/10.1073/pnas.191502998.

14.   Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol. 2006;24(31):5079–90. https://doi.org/10.1200/JCO.2005.05.1748.

15.   Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002;8(8):816–24. https://doi.org/10.1038/nm733.

16.   Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12–27. https://doi.org/10.1016/j.cell.2012.06.013.

17.   Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005;352(10):997–1003. https://doi.org/10.1056/NEJMoa043331.

18.   Lu Z, Zou J, Li S, Topper MJ, Tao Y, Zhang H, et al. Epigenetic therapy inhibits metastases by disrupting premetastatic niches. Nature. 2020;579(7798):284–90. https://doi.org/10.1038/s41586-020-2054-x.

19.   Li F, Huang Q, Luster TA, Hu H, Zhang H, Ng WL, et al. In vivo epigenetic CRISPR screen identifies Asf1a as an immunotherapeutic target in Kras-mutant lung adenocarcinoma. Cancer Discov. 2020;10(2):270–87. https://doi.org/10.1158/2159-8290.CD-19-0780.

20.   Van Den Broeck A, Brambilla E, Moro-Sibilot D, Lantuejoul S, Brambilla C, Eymin B, et al. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. Clin Cancer Res. 2008;14(22):7237–45. https://doi.org/10.1158/1078-0432.CCR-08-0869.

21.   Alam H, Tang M, Maitituoheti M, Dhar SS, Kumar M, Han CY, et al. KMT2D deficiency impairs super-enhancers to confer a glycolytic vulnerability in lung cancer. Cancer Cell. 2020;37:599–617 e597.

22.   Sun RC, Dukhande VV, Zhou Z, Young LEA, Emanuelle S, Brainson CF, et al. Nuclear glycogenolysis modulates histone acetylation in human non-small cell lung cancers. Cell Metab. 2019;30(5):903–16 e907. https://doi.org/10.1016/j.cmet.2019.08.014.

23.   Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov. 2011;1(7):598–607. https://doi.org/10.1158/2159-8290.CD-11-0214.

24.   Audia JE, Campbell RM. Histone modifications and cancer. Cold Spring Harb Perspect Biol. 2016;8:a019521.

25.   Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75–82. https://doi.org/10.1038/nature11232.

26.   Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.

27.   Yuan C, Hu H, Kuang M, Chen Z, Tao X, Fang S, et al. Super enhancer associated RAI14 is a new potential biomarker in lung adenocarcinoma. Oncotarget. 2017;8(62):105251–61. https://doi.org/10.18632/oncotarget.22165.

28.   Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934–47. https://doi.org/10.1016/j.cell.2013.09.053.

29.   Tu S, Li M, Chen H, Tan F, Xu J, Waxman DJ, et al. MAnorm2 for quantitatively comparing groups of ChIP-seq samples. Genome Res. 2021;31(1):131–45. https://doi.org/10.1101/gr.262675.120.

30.   Suzuki A, Makinoshima H, Wakaguri H, Esumi H, Sugano S, Kohno T, et al. Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. Nucleic Acids Res. 2014;42(22):13557–72. https://doi.org/10.1093/nar/gku885.

31.   Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307–19. https://doi.org/10.1016/j.cell.2013.03.035.

32.   Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320–34. https://doi.org/10.1016/j.cell.2013.03.036.

33.   Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016;351(6280):1454–8. https://doi.org/10.1126/science.aad9024.

Yuan *et al. Genome Biology*      (2021) 22:156

Page 25 of 26

34. Laughney A, Hu J, Campbell N, Bakhoum S, Setty M, Lavallée V, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat Med. 2020;26(2):259–69. https://doi.org/10.1038/s41591-019-0750-6.

35. Ni KW, Sun GZ. The identification of key biomarkers in patients with lung adenocarcinoma based on bioinformatics. Math Biosci Eng. 2019;16(6):7671–87. https://doi.org/10.3934/mbe.2019384.

36. Marudamuthu AS, Bhandary YP, Fan L, Radhakrishnan V, MacKenzie B, Maier E, et al. Caveolin-1-derived peptide limits development of pulmonary fibrosis. Sci Transl Med. 2019;11(522):eaat2848. https://doi.org/10.1126/scitranslmed.aat2848.

37. Northcott PA, Lee C, Zichner T, Stutz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature. 2014;511:428–34.

38. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. N Engl J Med. 2014;371(23):2167–77. https://doi.org/10.1056/NEJMoa1408440.

39. Shaw AT, Solomon B. Targeting anaplastic lymphoma kinase in lung cancer. Clin Cancer Res. 2011;17(8):2081–6. https://doi.org/10.1158/1078-0432.CCR-10-1591.

40. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet. 2017; 49(7):1126–32. https://doi.org/10.1038/ng.3892.

41. Witt C. European respiratory society/american thoracic society/international association for the study of lung cancer international multidisciplinary classification of lung adenocarcinoma: state of the art. J Thorac Oncol. 2011;6:1451.

42. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: a clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. J Thorac Oncol. 2011;6(9):1496–504. https://doi.org/10.1097/JTO.0b013e318221f701.

43. Zhang Y, Li J, Wang R, Li Y, Pan Y, Cai D, et al. The prognostic and predictive value of solid subtype in invasive lung adenocarcinoma. Sci Rep. 2014;4:7163.

44. Zheng D, Wang R, Zhang Y, Pan Y, Cheng X, Cheng C, et al. Prevalence and clinicopathological characteristics of ALK fusion subtypes in lung adenocarcinomas from Chinese populations. J Cancer Res Clin Oncol. 2016;142(4):833–43. https://doi.org/10.1007/s00432-015-2081-4.

45. Wei P, Zhang N, Wang Y, Li D, Wang L, Sun X, et al. FOXM1 promotes lung adenocarcinoma invasion and metastasis by upregulating SNAIL. Int J Biol Sci. 2015;11(2):186–98. https://doi.org/10.7150/ijbs.10634.

46. Chen L, Yu JH, Lu ZH, Zhang W. E2F2 induction in related to cell proliferation and poor prognosis in non-small cell lung carcinoma. Int J Clin Exp Pathol. 2015;8:10545–54.

47. LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. Cancer Cell. 2020;38(2):212–28 e213. https://doi.org/10.1016/j.ccell.2020.06.006.

48. Winslow MM, Dayton TL, Verhaak RG, Kim-Kiselak C, Snyder EL, Feldser DM, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. Nature. 2011;473:101–4.

49. Ahmed F. Integrated network analysis reveals FOXM1 and MYBL2 as key regulators of cell proliferation in non-small cell lung cancer. Front Oncol. 2019;9:1011. https://doi.org/10.3389/fonc.2019.01011.

50. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357(6352):eaan2507. https://doi.org/10.1126/science.aan2507.

51. Hnisz D, Weintraub A, Day D, Valton A, Bak R, Li C, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science (New Yorkg, NY). 2016;351:1454–8.

52. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2016;44(D1):D1023–31. https://doi.org/10.1093/nar/gkv1268.

53. Sun R, Xiang T, Tang J, Peng W, Luo J, Li L, et al. 19q13 KRAB zinc-finger protein ZNF471 activates MAPK10/JNK3 signaling but is frequently silenced by promoter CpG methylation in esophageal cancer. Theranostics. 2020;10:2243–59.

54. Zhang L, Li H, Yuan M, Li M, Zhang S. Cervical cancer cells-secreted exosomal microRNA-221-3p promotes invasion, migration and angiogenesis of microvascular endothelial cells in cervical cancer by down-regulating MAPK10 expression. Cancer Manag Res. 2019;11:10307–19. https://doi.org/10.2147/CMAR.S221527.

55. Xie Y, Liu Y, Fan X, Zhang L, Li Q, Li S, et al. MicroRNA-21 promotes progression of breast cancer via inhibition of mitogen-activated protein kinase10 (MAPK10). Biosci Rep. 2019:BSR20181000. https://doi.org/10.1042/BSR20181000. Epub ahead of print.

56. Wu Q, Tian Y, Zhang J, Tong X, Huang H, Li S, et al. In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. Proc Natl Acad Sci U S A. 2018;115(17):E3978–86. https://doi.org/10.1073/pnas.1716589115.

57. Hong H, Yao S, Zhang Y, Ye Y, Li C, Hu L, et al. In vivo miRNA knockout screening identifies miR-190b as a novel tumor suppressor. PLoS Genet. 2020;16(11):e1009168. https://doi.org/10.1371/journal.pgen.1009168.

58. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A. 2001;98(24):13784–9. https://doi.org/10.1073/pnas.241500798.

59. Tomida S, Koshikawa K, Yatabe Y, Harano T, Ogura N, Mitsudomi T, et al. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. Oncogene. 2004;23:5360–70.

60. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol. 2006;24:1679–88.

61. Chen YJ, Roumeliotis TI, Chang YH, Chen CT, Han CL, Lin MH, et al. Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. Cell. 2020;182(1):226–44 e217. https://doi.org/10.1016/j.cell.2020.06.012.

62. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell. 2020;182(1):200–25 e235. https://doi.org/10.1016/j.cell.2020.06.013.

63. Medina PP, Romero OA, Kohno T, Montuenga LM, Pio R, Yokota J, et al. Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. Hum Mutat. 2008;29(5):617–22. https://doi.org/10.1002/humu.20730.
64. Christensen CL, Kwiatkowski N, Abraham BJ, Carretero J, Al-Shahrour F, Zhang T, et al. Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. Cancer Cell. 2014;26(6):909–22. https://doi.org/10.1016/j.ccell.2014.10.019.
65. First EZH2 Inhibitor Approved-for Rare Sarcoma. Cancer Discov. 2020;10(3):333–34. https://doi.org/10.1158/2159-8290.CD-NB2020-006. Epub 2020 Feb 10.
66. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the rorthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol. 2016;11:39–51.
67. Panico F, Rizzi F, Fabbri LM, Bettuzzi S, Luppi F. Clusterin (CLU) and lung cancer. Adv Cancer Res. 2009;105:63–76. https://doi.org/10.1016/S0065-230X(09)05004-0.
68. Takeuchi A, Shiota M, Beraldi E, Thaper D, Takahara K, Ibuki N, et al. Insulin-like growth factor-I induces CLU expression through Twist1 to promote prostate cancer growth. Mol Cell Endocrinol. 2014;384:117–25.
69. Mazzarelli P, Pucci S, Spagnoli LG. CLU and colon cancer. The dual face of CLU: from normal to malignant phenotype. Adv Cancer Res. 2009;105:45–61. https://doi.org/10.1016/S0065-230X(09)05003-9.
70. Albert JM, Gonzalez A, Massion PP, Chen H, Olson SJ, Shyr Y, et al. Cytoplasmic clusterin expression is associated with longer survival in patients with resected non small cell lung cancer. Cancer Epidemiol Biomark Prev. 2007;16(9):1845–51. https://doi.org/10.1158/1055-9965.EPI-07-0146.
71. Kevans D, Foley J, Tenniswood M, Sheahan K, Hyland J, O'Donoghue D, et al. High clusterin expression correlates with a poor outcome in stage II colorectal cancers. Cancer Epidemiol Biomark Prev. 2009;18(2):393–9. https://doi.org/10.1158/1055-9965.EPI-08-0302.
72. Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods. 2009;48:240–8.
73. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019;10:1930.
74. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3. https://doi.org/10.1093/bioinformatics/btq170.
75. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
76. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. https://doi.org/10.1038/s41467-019-09234-6.
77. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
78. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71. https://doi.org/10.1093/nar/gkv1507.
79. Singh Nanda J, Kumar R, Raghava GP. dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes. Sci Rep. 2016;6(1):19340. https://doi.org/10.1038/srep19340.
80. Li F, Han X, Li F, Wang R, Wang H, Gao Y, et al. LKB1 inactivation elicits a redox imbalance to modulate non-small cell lung cancer plasticity and therapeutic response. Cancer Cell. 2015;27:698–711.
81. Huang H, Zhang W, Pan Y, Gao Y, Deng L, Li F, et al. YAP suppresses lung squamous cell carcinoma progression via deregulation of the DNp63-GPX2 axis and ROS accumulation. Cancer Res. 2017;77(21):5769–81. https://doi.org/10.1158/0008-5472.CAN-17-0449.
82. Sun Y, Yuan C: A systematic dissection of the epigenomic heterogeneity of lung adenocarcinoma reveals two different subclasses with distinct prognosis and core regulatory networks. European Genome-Phenome Archive, EGAD00001007066. 2021. https://ega-archive.org/datasets/EGAD00001007066.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.