MDPI

Article

Prognostic Ability of Enhancer RNAs in Metastasis of Non-Small Cell Lung Cancer

Jun Liu [1,2], **Jingyi Jia** [1,2,3,4], **Siqiao Wang** [2], **Junfang Zhang** [2], **Shuyuan Xian** [2], **Zixuan Zheng** [2], **Lin Deng** [5], **Yonghong Feng** [2,3,4,*], **Yuan Zhang** [6,*], **Jie Zhang** [1,2,*]

1   Department of Anesthesiology, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China; yirong007@126.com (J.L.); 2031276@tongji.edu.cn (J.J.)
2   School of Medicine, Tongji University, Shanghai 200092, China; 1753111@tongji.edu.cn (S.W.); zhangjunfang@tongji.edu.cn (J.Z.); 1833173@tongji.edu.cn (S.X.); 1653144@tongji.edu.cn (Z.Z.)
3   Shanghai Key Laboratory of Tuberculosis, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China
4   Shanghai Clinical Research Center for Infectious Diseases (tuberculosis), Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China
5   Normal College, Qingdao University, Qingdao 266071, China; bss613@126.com
6   Department of Pulmonary and Critical Care Medicine, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China
*   Correspondence: feng_yonghong@tongji.edu.cn (Y.F.); zhs0905@126.com (Y.Z.); jiezhang@tongji.edu.cn (J.Z.)

Abstract: (1) Background: Non-small cell lung cancer (NSCLC) is the most common lung cancer. Enhancer RNA (eRNA) has potential utility in the diagnosis, prognosis and treatment of cancer, but the role of eRNAs in NSCLC metastasis is not clear; (2) Methods: Differentially expressed transcription factors (DETFs), enhancer RNAs (DEEs), and target genes (DETGs) between primary NSCLC and metastatic NSCLC were identified. Prognostic DEEs (PDEEs) were screened by Cox regression analyses and a predicting model for metastatic NSCLC was constructed. We identified DEE interactions with DETFs, DETGs, reverse phase protein arrays (RPPA) protein chips, immunocytes, and pathways to construct a regulation network using Pearson correlation. Finally, the mechanisms and clinical significance were explained using multi-dimensional validation unambiguously; (3) Results: A total of 255 DEEs were identified, and 24 PDEEs were selected into the multivariate Cox regression model (AUC = 0.699). Additionally, the NSCLC metastasis-specific regulation network was constructed, and six key PDEEs were defined (ANXA8L1, CASTOR2, CYP4B1, GTF2H2C, PSMF1 and TNS4); (4) Conclusions: This study focused on the exploration of the prognostic value of eRNAs in the metastasis of NSCLC. Finally, six eRNAs were identified as potential markers for the prediction of metastasis of NSCLC.

Keywords: Non-small cell lung cancer; metastasis; enhancer RNAs; regulatory network; prognosis

1. Introduction

Lung cancer is the malignant tumor with the highest incidence and mortality rates in recent years. Approximately 1.6 million people die per annum as a result and lung cancer death rates worldwide are estimated to be higher by the World Health Organization (WHO) [1,2]. Non-small cell lung cancer (NSCLC) is the most common lung cancer, consisting of over 80% of cases [3]. NSCLC is mainly treated with surgery, radiotherapy, or chemotherapy [4]. However, only a few patients with early-stage NSCLC can be treated by surgery, which makes the 5-year survival rate for stage IA NSCLC patients reach up to 70% [5]. Patients with more advanced NSCLC are usually treated with chemotherapy or radiotherapy, which means the 5-year survival rate decreases to around 23%. Besides, significant limitations still exist in new treatments, such as immunological and targeted

therapies [6]. Therefore, a more comprehensive understanding of the molecular mechanism of progression and metastasis is critical to improve the prognosis of patients with NSCLC.

Enhancers are short genomic regions [7,8] that can modulate gene expression by interacting with promoters [9]. Enhancer RNAs (eRNAs), produced during the transcription process of enhancers, are of functional importance. Their expression levels correlate with enhancer activity [10,11]. In human cells, quite a lot of eRNAs have been found, many of which play critical roles in the meditation of the activation of target genes by transcriptional circuitry [12]. eRNA promotes transcription by regulating the chromatin accessibility near target gene promoters and binding to target gene promoters [13,14], or forming eRNA-proteins complexes that promote enhancer-promoter loops [15]. Many studies reported the role of eRNAs in cancers. For instance, the activation of ESR1 can broadly induce the increase of eRNA transcription in breast cancer [16], indicating that eRNAs are associated with activating oncogenes or oncogenic signaling pathways. Moreover, KLK3e is an androgen-induced eRNA regulating the gene KLK3, it can regulate AR-dependent gene expression in prostate cancer by scaffolding the androgen receptor (AR)-associated protein complex [17], which means that in some cases, tumorigenesis can be promoted directly by oncogene-induced eRNAs. Importantly, tissue and individual patient specificity were found in the expression of eRNAs, indicating eRNAs' potential clinical utility in diagnosis, prognosis and therapy for cancers [18–20].

However, the eRNA regulation mechanisms underlying NSCLC metastasis have not been elucidated, and eRNA targeted anti-cancer agents that can improve the prognosis of NSCLC patients are still insufficient. Here, differentially expressed transcription factors (DETFs), eRNAs (DEEs), and target genes (DETGs) between primary and metastatic NSCLC patients were determined. Prognostic DEEs (PDEEs) were further identified using univariate and multivariate Cox regression analyses, based on which a metastatic NSCLC-specific prognosis prediction model was constructed. Significant immune cells and immune-related pathways were identified using cell type identification by estimating the relative subsets of RNA transcripts (CIBERSORT) [21] and single sample gene set enrichment (ssGSEA) algorithms [22], respectively. Hallmark pathways correlated with key DEEs were quantified using gene set variation analysis (GSVA) [23]. Importantly, six PDEEs (ANXA8L1, CASTOR2, CYP4B1, GTF2H2C, PSMF1 and TNS4), along with DETFs, DETGs, immune cells, immune-related pathways, and hallmark pathways were integrated into co-expression analysis to construct a regulation network. The results above were demonstrated by multi-dimensional validation to explain the mechanisms and clinical significance unambiguously.

2. Results

2.1. DEG Identification and Functional Enrichment Analysis

The mechanism by which eRNAs promote the transcription of target genes was shown in Figure S1 A. An analysis process of this study was shown in Figure S1B. All clinical baseline information for primary NSCLC samples was summarized in Table 1. The differential expression patterns of 1648 DEGs (506 upregulated DEGs and 1142 downregulated DEGs) between primary NSCLC and metastatic NSCLC samples were illustrated by the heatmap plot (Figure 1A) and volcano plot (Figure 1B). The GO terms, including biological processes (BPs), cellular components (CCs), and molecular functions (MFs) where the DEGs were mostly enriched were the regulation of peptidase activity, extracellular matrix and enzyme inhibitor activity, respectively (Figure 1C). Moreover, amyotrophic lateral sclerosis was the most significant KEGG pathway in which the DEGs were mostly enriched (Figure 1D).

In addition, we can see from the heatmap plot (Figure 2A) and volcano plot (Figure 2B) that 255 DEEs (59 upregulated DEEs and 196 downregulated DEEs) were identified from 5100 eRNAs between primary NSCLC and metastatic NSCLC samples.

Table 1. Clinical baseline information of 829 primary NSCLC patients.

Characteristics	Total Patients (N = 829)
Age, years	
• Mean ± SD	66.30 ± 9.35
• Median(Range)	68 (33–87)
Gender	
• Female	301 (36.31%)
• Male	528 (63.69%)
Stages	
• Stage i	415 (50.06%)
• Stage ii	242 (29.19%)
• Stage iii	144 (17.37%)
• Stage iv	28 (3.38%)
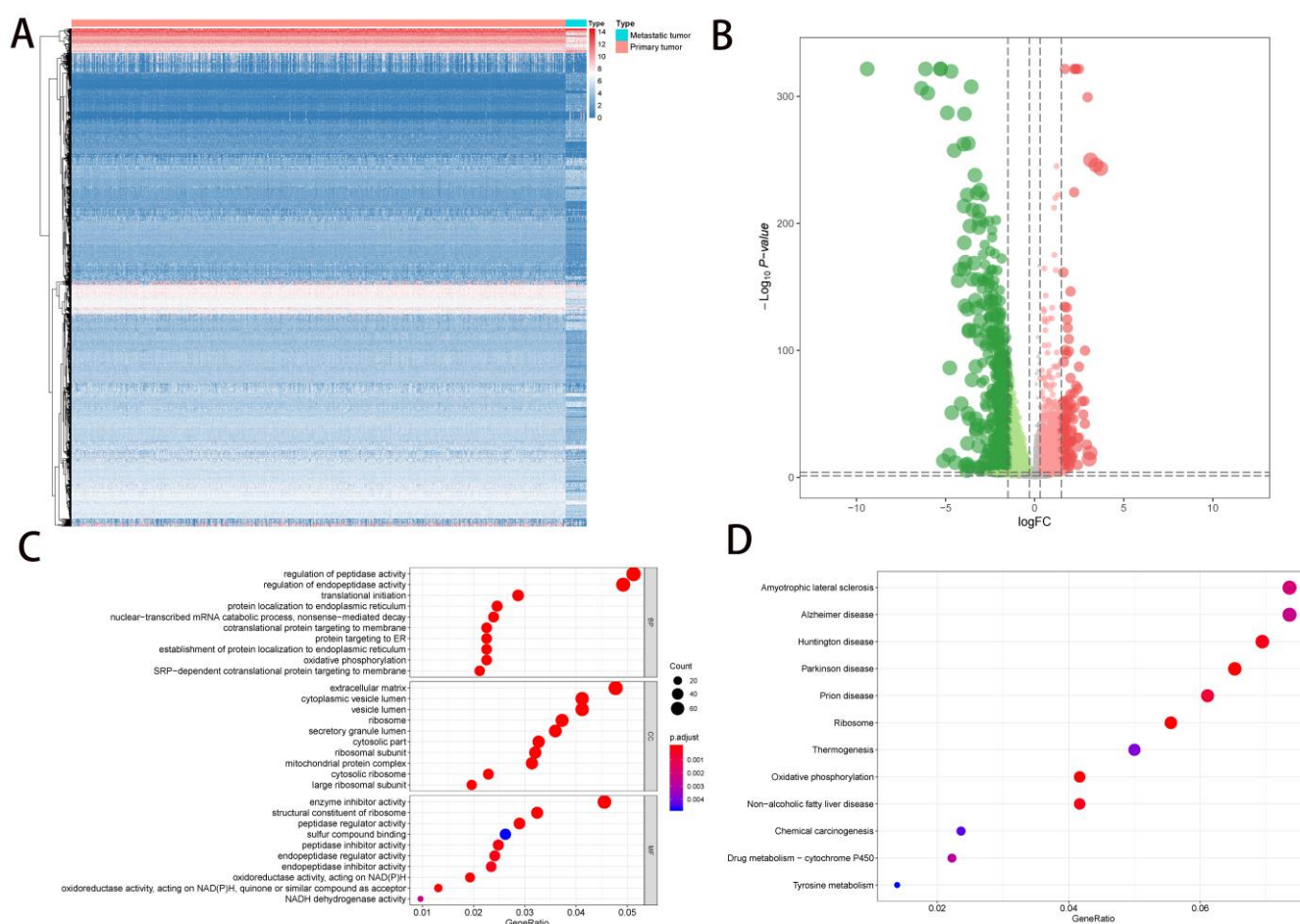
Figure 1. The analysis of DEGs. (**A**,**B**) DEGs were defined between primary NSCLC and metastatic NSCLC. A total of 1648 DEGs (506 upregulated DEGs and 1142 downregulated DEGs) were defined between primary NSCLC and metastatic NSCLC. In the heatmap (**A**), red color represented primary NSCLC and blue color represented metastatic NSCLC. In the volcano plot (**B**), red dot represented upregulated DEGs and green dot represented downregulated DEGs, the two dashed horizontal lines mark the positions of *p*-value = 0.05 and *p*-value = 0.0001, respectively, and the 4 vertical dotted lines mark the positions of log2 foldchange (logFC) = −1.5, logFC = −0.3, logFC = 0.3 and logFC = 1.5, respectively. (**C**,**D**) GO and KEGG pathway enrichment analysis. DEGs enriched in regulation of peptidase activity, extracellular matrix enzyme inhibitor activity and amyotrophic lateral sclerosis significantly. NSCLC, Non-small cell lung cancer; SD, Standard deviation. DEGs, differential

expressed genes; NSCLC, Non-small cell lung cancer; GO, Gene Oncology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
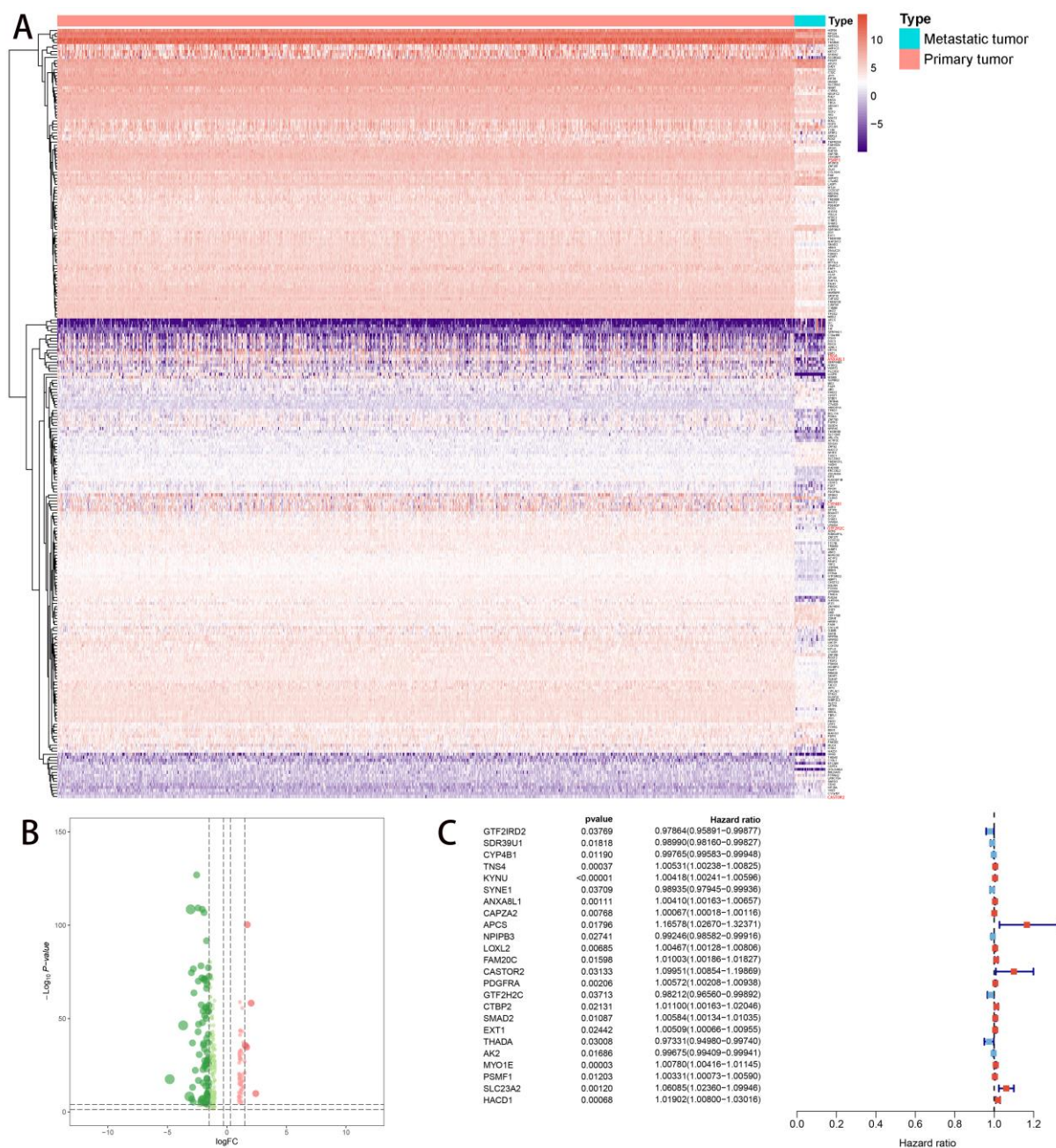
Figure 2. The DEEs identification and DEEs univariate Cox regression analysis. (**A**,**B**) The identification of DEEs. A total of 255 DEEs (59 upregulated DEEs and 196 downregulated DEEs) were identified between primary NSCLC and metastatic NSCLC. In the heatmap (**A**), red color represented primary NSCLC and blue color represented metastatic NSCLC. In the volcano plot (**B**), red dot represented upregulated DEEs and green dot represented downregulated DEEs, the two dashed horizontal lines mark the positions of *p*-value = 0.05 and *p*-value = 0.0001, respectively, and the four vertical dotted lines mark the positions of log2 foldchange (logFC) = −1.5, logFC = −0.3, logFC = 0.3 and logFC = 1.5, respectively. (**C**) DEEs univariate Cox regression analysis. Twenty-four eRNAs were identified as PDEEs (*p* < 0.05). DEEs, differentially expressed eRNAs; NSCLC, Non-small cell lung cancer; PDEEs, prognostic differentially expressed eRNAs; eRNAs, enhancer RNAs.

2.2. Multivariate Prognostic Model Construction and Independent Prognostic Factors Identification

Twenty-four PDEEs were identified by univariate Cox regression analysis ($p < 0.05$) (Figure 2C) and integrated into the multivariate prognostic model. Then, the efficiency of the model was evaluated by an ROC curve (AUC = 0.699) (Figure 3D).

For the evaluation of the independent prognostic value of RS, the formula mentioned in the methods was applied to calculate the RS for each NSCLC patient, and the distribution of NSCLC patients with low and high RS was shown by the risk scatter plot (Figure 3A) and risk line plot (Figure 3B).

Besides, the survival probability of the high-risk group and low-risk group was shown by the Kaplan–Meier survival curve, which indicated a lower survival probability in the high-risk group ($p < 0.001$) (Figure 3C). Furthermore, the distribution of the high-risk group and low-risk group within the principal components (PCs) (PC1, PC2 and PC3) was shown in Figure 3E. Finally, RS was identified as an independent prognostic factor in the univariate Cox regression (hazard ratios (HR) = 76.734, 95% confidence interval (CI) (30.391–193.746), $p < 0.001$) (Figure 3F) and multivariate Cox regression (HR = 1.372, 95% CI (1.271–1.482), $p < 0.001$) models (Figure 3G), which were adjusted by age, gender, and stage. Then, boxplots showed the LncRNA expression of 24 PDEEs in each TNM staging and stage of NSCLC (Figure 4A–D).
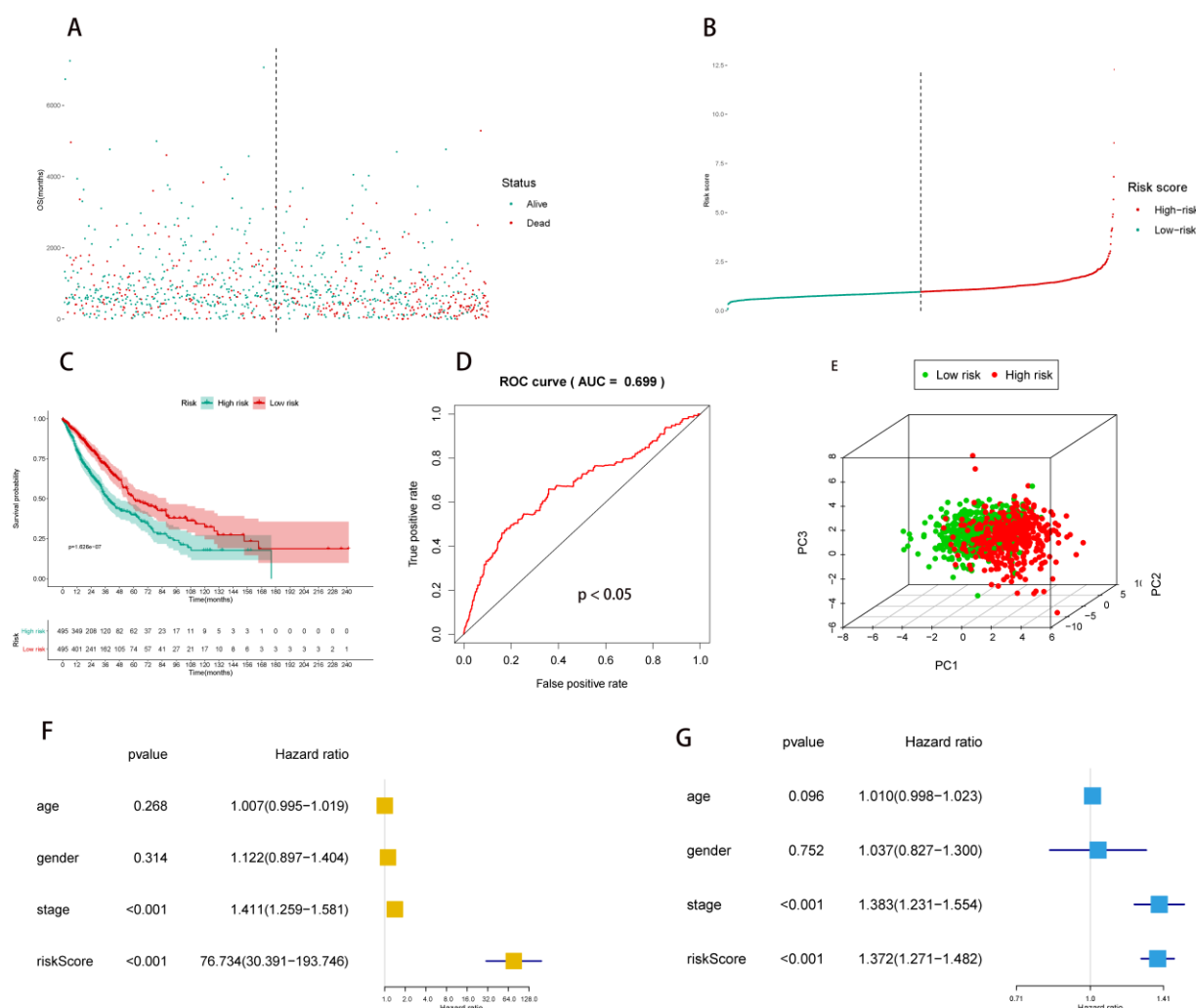
Figure 3. Independent prognosis analysis. RS was calculated by 24 PDEEs integration multivariate Cox model. (**A**,**B**) The scatter plot and line plot illustrated the distribution of RS among all NSCLC patients. (**C**) The ROC curve indicated that the multivariate Cox regression model was of good predictive power (AUC = 0.699). (**D**) The Kaplan–Meier curve showed that RS had prognostic value for NSCLC patients ($p < 0.05$), red line and green line represented low-risk group and high-risk group, respectively. (**E**) The distribution of low-risk and high-risk groups in PC1, PC2 and PC3, red plots represented high-risk groups and green plots represented low risk groups, the proportion of variance for PC1, PC2 and PC3 were 0.141, 0.087 and 0.074, respectively, the cumulative proportion was 0.302. (**F**,**G**) The univariate (HR = 76.734, 95% CI (30.391–193.746), $p < 0.001$) and multivariate (HR = 1.372, 95% CI (1.271–1.482), $p < 0.001$) Cox regression analysis for RS defined RS as an independent prognostic factor for NSCLC patients. The multivariate Cox regression model was corrected by age, gender and stage. RS, risk score; PDEEs, prognostic differentially expressed eRNAs; eRNAs, enhancer RNAs; ROC, receiver operator characteristic; AUC, area under curve; NSCLC, non-small cell lung cancer; HR, hazard ratios.
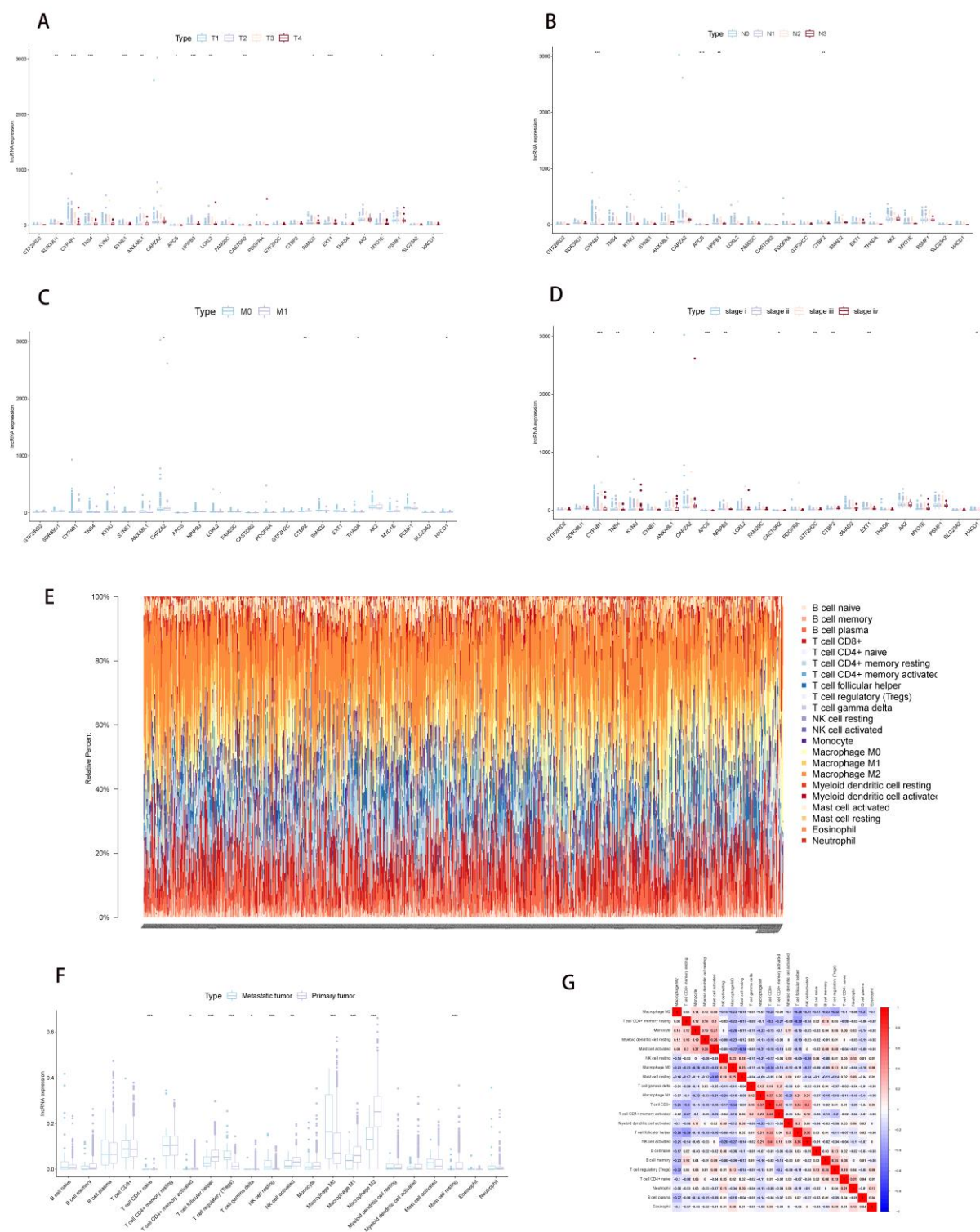
2.3. Correlation Analysis of PDEEs and Immune Cells

In all samples, the composition of 22 immune cells was estimated by the CIBERSORT algorithm (Figure 4E), Macrophage M2, T cell CD4+ memory resting and Macrophage M1 have a higher proportion. Figure 4F, G illustrated the differential infiltration degree of the immune cells between the primary NSCLC and metastatic NSCLC and co-expression analysis of 22 immune cells, respectively. As is shown in Figure 4F, the infiltration degree of T cell CD4+ naive, T cell CD4+ memory activated, T cell follicular helper, T cell gamma delta, NK cell activated, Macrophage M2, Macrophage M1 and Mast cell resting in metastatic NSCLC is significantly lower than that in primary NSCLC, but the infiltration degree of T cell regulatory (Tregs) and NK cell resting is significantly higher in metastatic NSCLC. Figure 4G showed that T cell CD8+ is positively correlated with NK cell activated and T cell CD4+ memory activated significantly, but Mast cell activated and Mast cell resting, T cell CD4+ memory resting and T cell follicular helper showed significant negative correlations. Furthermore, co-expression analysis between PDEEs and 22 immune cells indicated a significant regulatory relationship among 16 immune cells and 15 PDEEs with a |correlation coefficient| > 0.15 and $p < 0.05$.
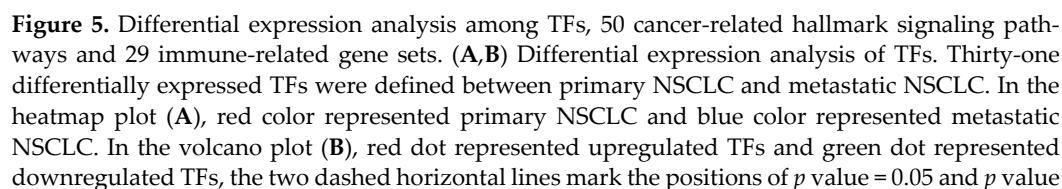
2.4. Correlation Analysis of PDEEs, DETFs, Immune-Related Gene Sets, and Hallmark Pathways

The expression levels of 30 DETFs were illustrated in a heatmap plot (Figure 5A) and volcano (Figure 5B) plot, 29 immune gene sets were shown in a heatmap plot (Figure 5F). Besides, the heatmap plot (Figure 5C), volcano plot (Figure 5D), together with the bar plot (Figure 5E) showed the differentially expressed gene sets of hallmarks of cancer in GSVA, HALLMARK_PANCREAS_BETA_CELLS and HALLMARK_KRAS_SIGNALING_DN are highly expressed in metastatic NSCLC. Afterward, correlation analysis was performed based on PDEEs with DETFs, immune-related gene sets, and hallmark pathways, respectively.

Finally, several elements aforementioned were considered to have significant correlations which were extracted for the construction of a regulation network. Specifically, this regulation network consisted of pairwise interactions between 24 DETFs and 15 PDEEs with a |correlation coefficient| > 0.20 and $p < 0.05$, pairwise interactions between 24 immune-related gene sets and 13 PDEEs with a |correlation coefficient| > 0.20 and $p < 0.05$, and pairwise interactions between 34 hallmark pathways and 16 PDEEs with a |correlation coefficient| > 0.25 and $p < 0.05$.

Figure 4. Clinical correlation analysis and immune infiltration analysis. (**A–D**) Boxplots showed the LncRNA expression of 24 PDEEs in each TNM staging and stage of NSCLC. (**E**) CIBERSORT algorithm estimated the composition of 22 immune cells in entire NSCLC samples. (**F**) The immune cells that were differentially expressed between primary NSCLC and metastatic NSCLC, purple color and blue color represented primary NSCLC and metastatic NSCLC, respectively. (**G**) Co-expression analysis in entire 22 immune cells. NSCLC, Non-small cell lung cancer.

Figure 5. Differential expression analysis among TFs, 50 cancer-related hallmark signaling pathways and 29 immune-related gene sets. (**A**,**B**) Differential expression analysis of TFs. Thirty-one differentially expressed TFs were defined between primary NSCLC and metastatic NSCLC. In the heatmap plot (**A**), red color represented primary NSCLC and blue color represented metastatic NSCLC. In the volcano plot (**B**), red dot represented upregulated TFs and green dot represented downregulated TFs, the two dashed horizontal lines mark the positions of *p* value = 0.05 and *p* value

= 0.0001, respectively, and the four vertical dotted lines mark the positions of log2 foldchange (logFC) = −1.5, logFC = −0.3, logFC = 0.3 and logFC = 1.5, respectively. (**C–E**) Differential expression analysis of 50 cancer-related hallmark signaling pathways. In the heatmap (**C**), red color represented primary NSCLC and blue color represented metastatic NSCLC. In the volcano plot (**D**), red dot represented the upregulated cancer-related hallmark signaling pathways and green dot represented the downregulated cancer-related hallmark signaling pathways, the two dashed horizontal lines mark the positions of th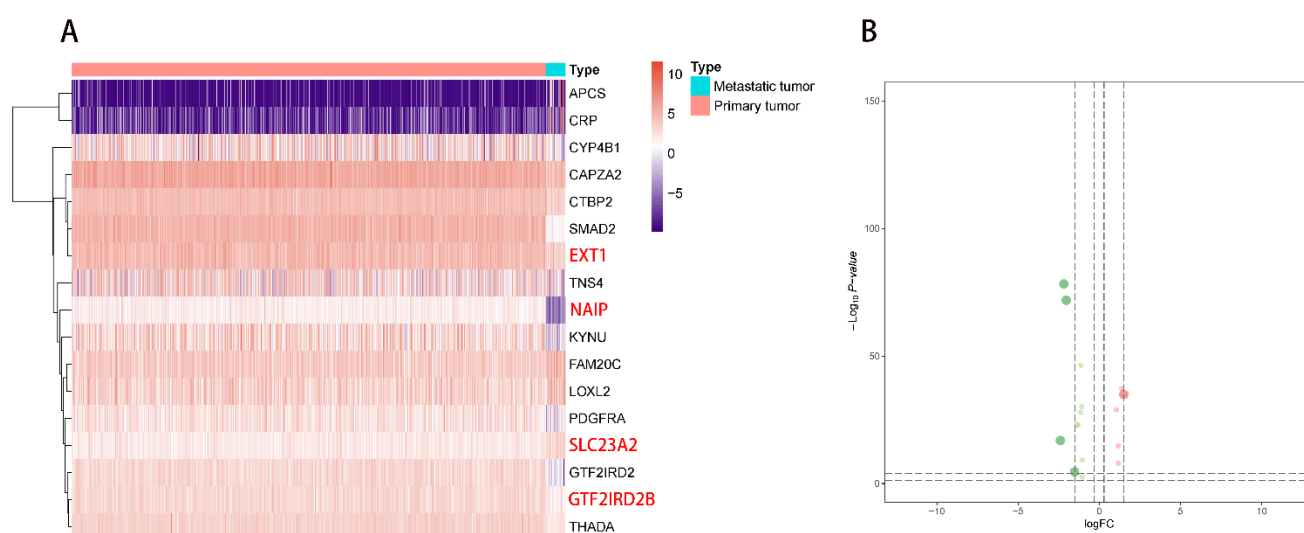e p-value = 0.05 and p-value = 0.0001, respectively, and the four vertical dotted lines mark the positions of log2 foldchange (logFC) = −1.5, logFC = −0.3, logFC = 0.3 and logFC = 1.5, respectively. Bar plot (**E**) showed the t-value of the GSVA score among all cancer-related hallmark signaling pathways. (**F**) Differential expression analysis for 29 immune-related gene sets. Red color represented primary NSCLC and blue color represented metastatic NSCLC. TFs, transcription factors; NSCLC, non-small cell lung cancer. 2.6 The Construction of NSCLC metastasis-specific eRNA regulation network.

2.5. Correlation Analysis of PDEEs, DETGs, and RPPA Protein Chips

The heatmap plot (Figure 6A) and volcano (Figure 6B) plot showed significantly differential expression patterns for 17 DETGs. Afterward, correlation analysis indicated strong correlations between 17 DETGs which were selected for the construction of a regulation network based on a |correlation coefficient| > 0.20 and p-value < 0.05.


Figure 6. Differential expression analysis of target genes of eRNAs. (**A**,**B**) Differential expression analysis of target genes of eRNAs. Seventeen differentially expressed target genes of eRNAs were defined between primary NSCLC and metastatic NSCLC. In the heatmap plot (**A**), red color represented primary NSCLC and blue color represented metastatic NSCLC. In the volcano plot (**B**), red dot represented the upregulated part and green dot represented the downregulated part in target genes of eRNAs, the two dashed horizontal lines mark the positions of p-value = 0.05 and p-value = 0.0001, respectively, and the four vertical dotted lines mark the positions of log2 foldchange (logFC) = −1.5, logFC = −0.3, logFC = 0.3 and logFC = 1.5, respectively. eRNAs, enhancer RNAs; NSCLC, non-small cell lung cancer.

Besides, 38 RPPA protein chips and 11 PDEEs were identified to have significant correlations with the criterion of a |correlation coefficient| > 0.25 and p-value < 0.05.

2.6. The Construction of NSCLC Metastasis-Specific eRNA Regulation Network

Twenty-eight DEGs were eventually integrated to construct a regulatory network, and the expression levels of them between primary NSCLC samples and metastatic NSCLC samples were shown in the heatmap plot (Figure 7A). Specifically, 6 PDEEs, 17 DETFs, 5 DETGs, 33 RPPA protein chips, 23 hallmark pathways, 13 immune cells, and 21 immune-related gene sets were selected to construct the NSCLC metastasis-specific eRNA regulation network (Figure 7B). Figure 7C showed the results of co-expression analysis among 6 PDEEs, 17 DETFs, 5 DETGs, 33 RPPA protein chips, 23 hallmark pathways, 13 immune cells, and 21 immune-related gene sets. Finally, ANXA8L1 mainly positively regulated TP63 (DETF, r = 0.345, *p* < 0.001), EXT1 (DETG, r = 0.361, *p* < 0.001), PAI1 (RPPA protein chip, r = 0.368, *p* < 0.001), EGFR (RPPA protein chip, r = 0.258, *p* < 0.001), HALLMARK_P53_PATHWAY (hallmark signaling pathway, r = 0.398, *p* < 0.001), and Mast cell resting (immune cell, r = 0.214, *p* < 0.001) in NSCLC metastasis. CASTOR2 mainly positively regulated H2AFX (DETF, r = 0.451, *p* < 0.001), SLC23A2 (DETG, r = 0.316, *p* < 0.001), TFRC (RPPA protein chip, r = 0.318, *p* < 0.001), HALLMARK_G2M_CHECKPOINT (hallmark signaling pathway, r = 0.336, *p* < 0.001) and NK cell resting (immune cell, r = 0.204, *p* < 0.001) in NSCLC metastasis. CYP4B1 mainly positively regulated FOS (DETF, r = 0.302, *p* < 0.001), GTF2IRD2B (DETG, r = 0.250, *p* < 0.001), NAPSINA (RPPA protein chips, r = 0.365, *p* < 0.001), HALLMARK_FATTY_ACID_METABOLISM (hallmark signaling pathway, r = 0.365, *p* < 0.001), Mast cell activated (immune cell, r = 0.359, *p* < 0.001) and Type_II_IFN_Reponse (immune-related gene set, r = 0.269, *p* < 0.001) in NSCLC metastasis. GTF2H2C mainly positively regulated NAIP (DETG, r = 0.469, *p* < 0.001), TTF1 (RPPA protein chip, r = 0.412, *p* < 0.001), HALLMARK_PROTEIN_SECRETION (hallmark signaling pathway, r = 0.323, *p* < 0.001), Macrophage M2 (immune cell, r = 0.179, *p* < 0.001) and iDCs (immune-related gene sets, r = 0.231, *p* < 0.001) in NSCLC metastasis. PSMF1 mainly positively regulated TP63 (DETF, r = 0.320, *p* < 0.001), EXT1 (DETG, r = 0.276, *p* < 0.001), TFRC (RPPA protein chip, r = 0.251, *p* < 0.001), HALLMARK_MYC_TARGETS_V1 (hallmark signaling pathway, r = 0.344, *p* < 0.001) and Mast cell resting (immune cell, r = 0.152, *p* < 0.001) in NSCLC metastasis. TNS4 mainly positively regulated TP63 (DETF, r = 0.289, *p* < 0.001), EXT1 (DETG, r = 0.310, *p* < 0.001), CD49B (RPPA protein chip, r = 0.332, *p* < 0.001), HALLMARK_P53_PATHWAY (hallmark signaling pathway, r = 0.320, *p* < 0.001) and Mast cell resting (immune cell, r = 0.200, *p* < 0.001) in NSCLC metastasis.

Because eRNA-related transcriptional regulatory changes were implicated in the pathological processes of NSCLC, and traditional long-term treatment with drugs may result in refractoriness and unsatisfactory outcomes, it is urgent to find potential inhibitors which target NSCLC-related PDEEs, DETGs, and DETFs. Therefore, the small-molecule bioactive inhibitors for DEGs within the regulatory network in this study were identified based on the CMap database. The heatmap plot (Figure 7D) showed the statistically significant small-molecule bioactive inhibitors in more than 10 types of cancers. The results indicated that irinotecan (enrichment score = 0.996, *p* value < 0.001) may be the best small-molecule bioactive inhibitor that may inhibit NSCLC metastasis by suppressing the expression of key DEGs in this study.
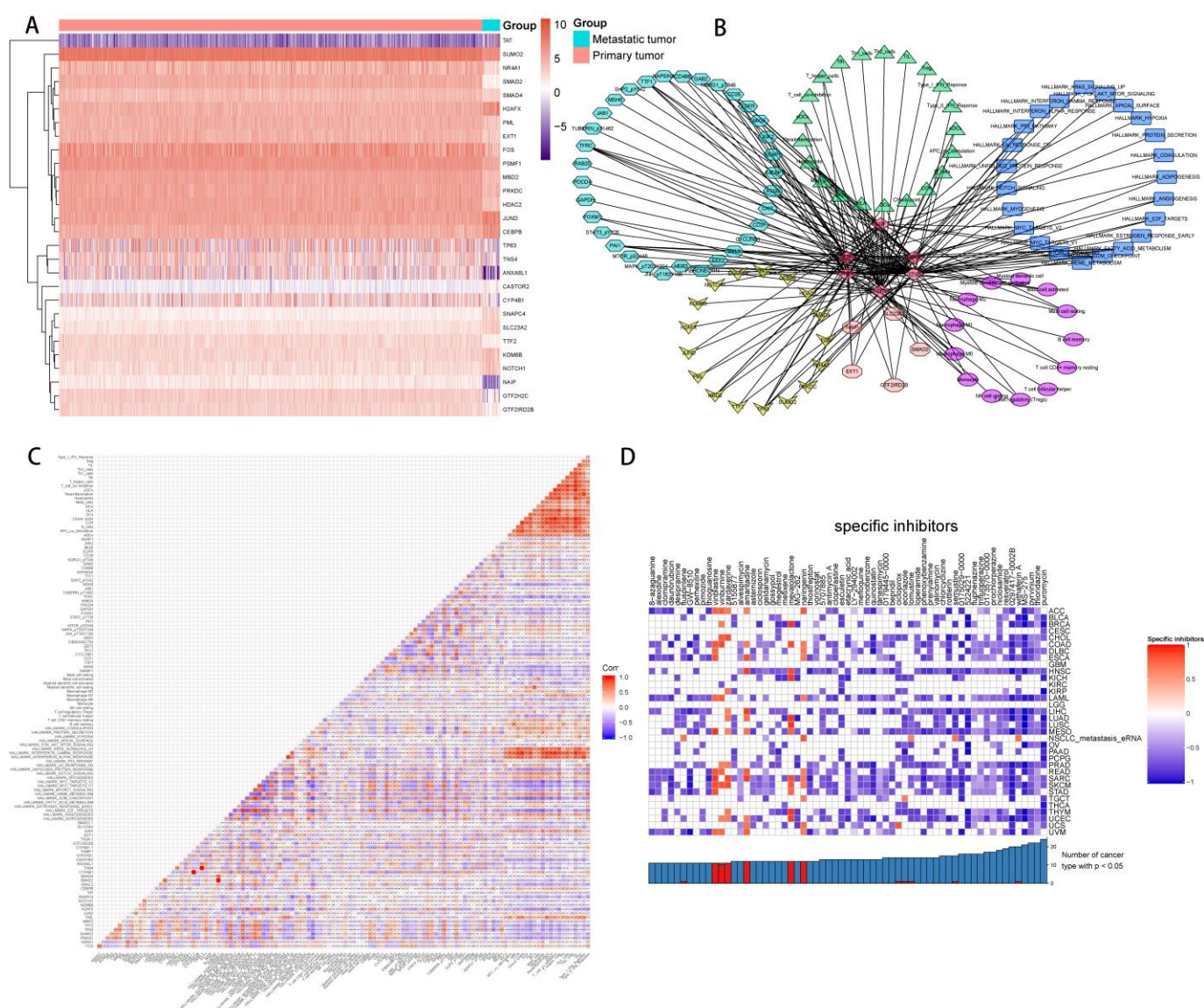
Figure 7. The metastasis-specific regulation network construction of NSCLC and Cmap analysis. (**A**) A total of 28 DEGs were extracted in the construction of the regulatory network. Red color and blue color represented primary and metastatic NSCLC. (**B**) Construction of the NSCLC metastasis-specific regulation network. Six PDEEs, 17 DETFs, 5 DETGs, 33 RPPA protein chips, 23 hallmark signaling pathways, 13 immune cells, and 21 immune-related gene sets in total were selected for the construction of the regulation network. In the network, PDEEs were represented by red rhombus in the center, immune-related gene sets were represented using green triangles, hallmark signaling pathways were represented by dark blue rectangles, immune cells were represented by purple ellipses, DETGs were represented by pink octagons, DETFs were represented by yellow concave quadrilaterals, and RPPA protein chips were represented by light blue hexagons. (**C**) Co-expression analysis among all elements selected in the regulatory network. (**D**) Cmap analysis. Heatmap plot of Cmap analysis showed significant small-molecule bioactive inhibitors in more than 10 types of cancer. Irinotecan (enrichment score = 0.996 *p*-value < 0.001) may be the potential small-molecule bioactive inhibitor for key PDEEs in NSCLC metastasis. DEGs, differentially expressed genes; NSCLC, non-small cell lung cancer; PDEEs, prognostic differentially expressed eRNAs; Cmap, connectivity map; eRNAs, enhancer RNAs.

2.7. Analysis of Single-Cell RNA-Seq Transcriptomes

Data of the single-cell RNA sequencing (scRNA-seq) from 24 NSCLC samples (GSE153935) were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153935, accessed on 8 June 2022) for validation of the subcellular locations of six key PDEEs. Fifteen cell clusters and seven cell types (Alveolar, B, NK/T, Endothelial, Epithelial, Fibroblast and Myeloid) were identified by t-distributed Stochastic Neighbor Embedding (t-SNE) analysis (Figure 8A). The heatmap showed the genes that were up- or down-regulated in the 15 clusters (Figure 8B). Figure 8C showed the expression of the major genes of the seven cell types in each cell type. The Cleveland plot (Upper part of Figure 8D) showed the expression of canonical markers in seven cell types and the bar plot (Lower part of Figure 8D) showed the distribution of seven cell types in 24 NSCLC samples, which validated the accuracy of our cell type annotations.

Expressions of five key PDEEs (ANXA8L1, CYP4B1, GTF2H2C, PSMF1 and TNS4) (Figure 9A), three key TFs (TP63, H2AFX and FOS) and four key DETGs (EXT1, GTF2IRD2B, NAIP and SLC23A2) (Figure S23) in seven cell types were displayed in feature plots. The UMAP plot displayed the cell cycle of cells above (Figure 9B), which showed that cells from cluster 1 were mainly in G1 phase and cells from cluster 4 were mainly in S phase. Pairs of ligand and receptor among the clusters above were displayed by the ligand-receptor plot (Figure 9C). All the results above suggested that in NSCLC, ANXA8L1 and CYP4B1 are mainly expressed in Alveolar cells, TNS4 is mainly expressed in epithelial cells, GTF2H2C is mainly expressed in NK/T cells and PSMF1 is significantly expressed in the seven cell types above.
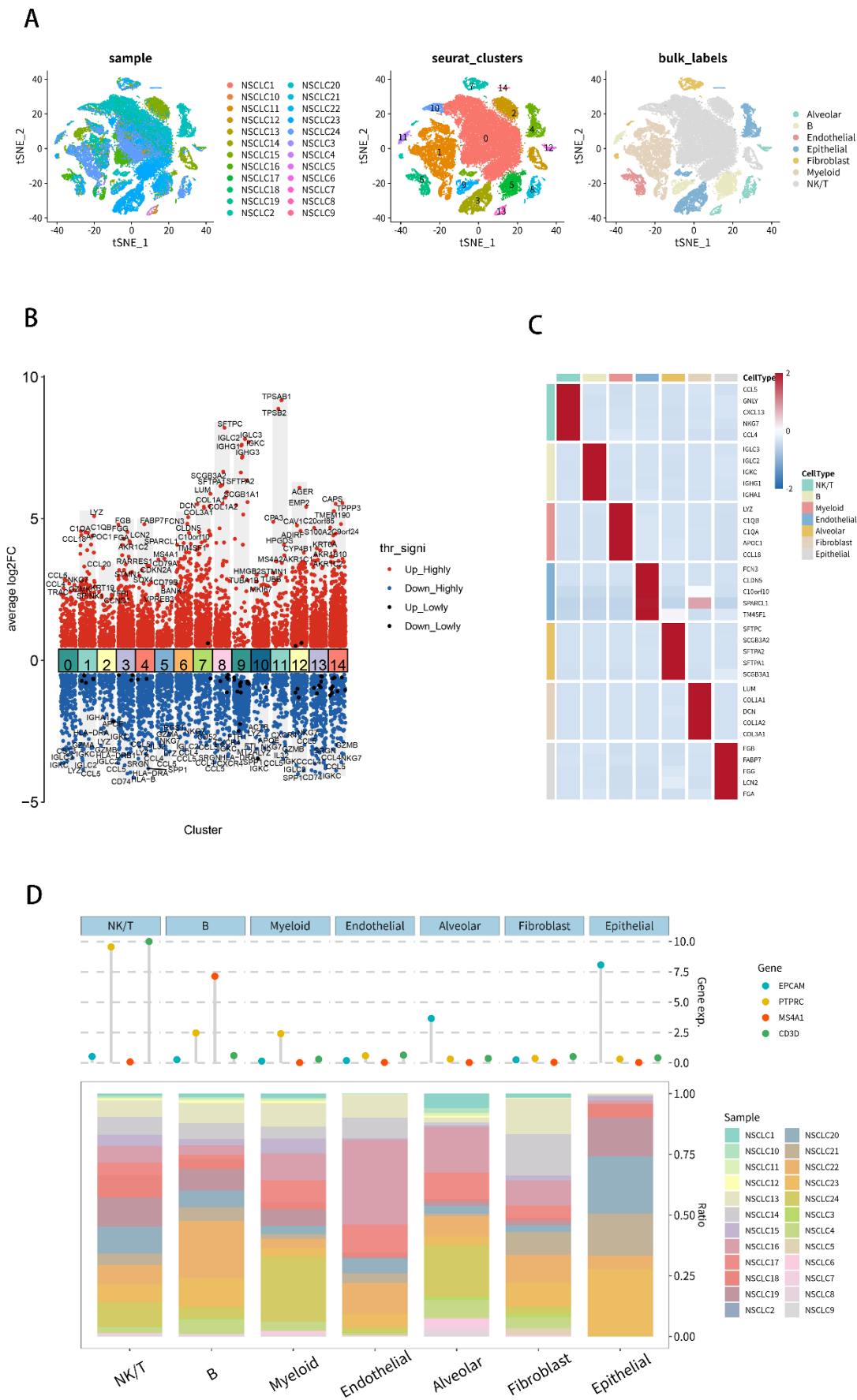
Figure 8. Analysis of scRNA-seq transcriptome for NSCLC to verify the distribution of key bi-omarkers in various cell types. (**A**) Distribution of 24 NSCLC samples, 15 cell clusters and seven cell types identified by t-SNE analysis. (**B**) Heatmap plot of genes that were up- or down-regulated in the 15 clusters. (**C**) Heatmap plot for the expression of the major genes of the seven cell types in each cell type. (**D**) Cleveland plot (Upper) showed the expression of canonical markers in seven cell types and the bar plot (Lower) showed the distribution of seven cell types in 24 NSCLC samples. scRNA-seq, single-cell RNA sequencing.
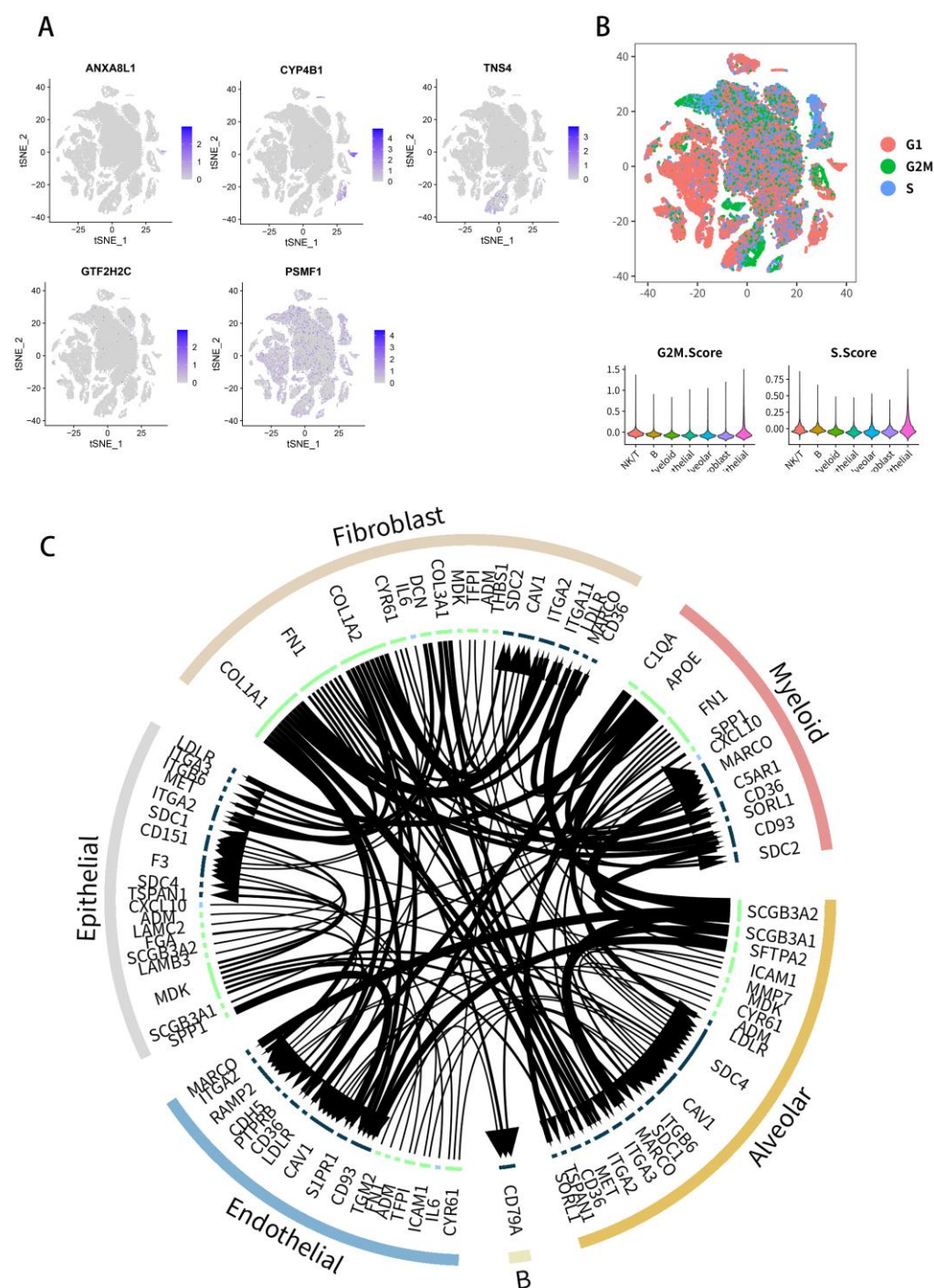
Figure 9. Distribution of five key PDEEs in seven cell types in NSCLC. (**A**) Expressions of five key PDEEs (ANXA8L1, CYP4B1, GTF2H2C, PSMF1, and TNS4) in each cluster. (**B**) Cell cycle of each cell. (**C**) ligand-receptor plot for pairs of ligand and receptor among the seven cell types. NSCLC, Non-small cell lung cancer; PDEEs, prognostic differentially expressed eRNAs.

2.8. Multidimensional validation

Multiple databases were utilized for reducing the bias caused by different platforms and to improve the reliability of our results. The key role of PDEEs including ANXA8L1, CASTOR2, CYP4B1, GTF2H2C, PSMF1, and TNS4 in pathogenesis and metastasis of NSCLC were validated based on GEPIA (Figure S2), UCSC xena (Figure S3), The Human Protein Atlas (Figure S4), cBioPortal (Figure S5), UALCAN (Figure S6, S7) and OncoLnc (Figure S8). In GEPIA and UALCAN databases, it was shown that CASTOR2, GTF2H2C, PSMF1 and TNS4 were highly-expressed in Lung Adenocarcinoma (LUAD), whereas ANXABL1 and CYP4B1 were lowly-expressed in the TCGA LUAD cohort (Figures S2 and S6); ANXA8L1, CASTOR2, PSMF1 and TNS4 were highly-expressed in the Lung Squamous Cell Carcinoma (LUSC) cohort, while CYP4B1 and GTF2H2C were lowly-expressed in the TCGA LUSC cohort (Figures S2 and S7). CYP4B1 (GEPIA, Figure S2; UALCAN, Figure S6; OncoLnc, Figure S8), TNS4 (GEPIA, Figure S2; OncoLnc, Figure S8) were significantly associated with the prognosis of LUAD. CASTOR2 (GEPIA, Figure S2), CYP4B1 (GEPIA, Figure S2; OncoLnc, Figure S8) and PSMF1 (GEPIA, Figure S2; UALCAN, Figure S6) were significantly related to the prognosis of LUSC. Besides, TNS4 (UCSC xena, Figure S3) was significantly associated with the prognosis of pan NSCLC (all $p < 0.05$).

The clinical characteristics, drug responses, and target genes of six key PDEEs were validated by the eRic database (https://hanlab.uth.edu/eRic/, accessed on 6 August 2021) [19]. The results indicated that ANXA8L1 (cholangiocarcinoma), PSMF1 (stomach adenocarcinoma) and TNS4 (head and neck squamous cell carcinoma, lung squamous cell carcinoma, rectum adenocarcinoma, stomach adenocarcinoma) were highly expressed in these cancers. CYP4B1 was lowly expressed in lung adenocarcinoma (Figure S9). ANXA8L1 (kidney renal papillary cell carcinoma) and CYP4B1 (lung adenocarcinoma and bladder urothelial carcinoma) were highly expressed at the early stage of these cancers, but PSMF1 was highly expressed in stage III stomach adenocarcinoma (Figure S10). CYP4B1 was lowly expressed in high-grade bladder urothelial carcinoma (Figure S11). The expression levels of CYP4B1 (lung adenocarcinoma, bladder Urothelial Carcinoma, bladder urothelial carcinoma, and breast invasive carcinoma) showed a significant relationship with different subtypes of these cancers (Figure S12). Furthermore, ANXA8L1 (kidney renal papillary cell carcinoma), CYP4B1 (lung adenocarcinoma) and TNS4 (stomach adenocarcinoma) showed a significant relation with survival in these cancers (Figure S13). Moreover, smoking showed a significant relation with the expression of CYP4B1 in lung adenocarcinoma (Figure S14) (all FDR $p < 0.05$). Table S1 showed the information of target genes and the drug responses of six key PDEEs.

The accessibility in chromatin of key PDEEs was validated by the data from ATAC-seq downloaded from the TCGA database and the results in LUAD and LUSC were shown in Figures S15 and S16, and Figures S17 and S18, respectively. Twenty-four chromosome open regions of genic, intergenic, intron, exon, upstream, downstream and distal intergenic were shown in Figures S15A and S17A, and Figures S15B and S17B, respectively (open regions were shown as green peaks). The distance between the open regions of all chromosomes and the regions of gene transcription were calculated and shown in Figures S15C and S17C, and Figures S15D and S17D. Figure S15E and 15F showed that in the open region, genes were mainly enriched for skeletal system development, centrosome, cell adhesion molecule binding and MAPK signaling pathway in LUAD, and Figures S17E and S17F showed that the open region genes were mainly enriched for the positive regulation of nervous system development, centrosome, cell adhesion molecule binding and MAPK signaling pathway in LUSC. Then, six key PDEEs were detected to be accessible in chromatin (Figures S16 and S18).

The Chip-seq validation was performed by the Cistrome database to determine whether PDEEs are bound to DETFs binding sites, PDEEs and DETFs (ANXA8L1-TP63, CASTOR2-H2AFX, CYP4B1-FOS, GTF2H2C-TP63, PSMF1- TP63 and TNS4-TP63) with the highest correlation were selected for further analysis. Finally, six key PDEEs were determined to be bound to the DETFs binding sites (Figure S19).

3. Discussion

NSCLC is the main type of lung cancer [24], whose mortality ranks top 1 all over the world [25]. In general, the prognosis of patients with NSCLC metastasis is not good [6], it is vital to explore the potential biological mechanisms and biomarkers for prognostic and therapeutic targets related to NSCLC metastasis. Multiple eRNAs are implicated in tumorigenesis and metastasis, which were significant therapeutic targets in metastatic NSCLC.

In our study, 24 PDEEs were identified and based on them, the risk score was defined by the multivariate Cox regression model (AUC = 0.699). Furthermore, the risk score was identified as the significant predictive factor for the prognosis of NSCLC patients by Univariate and multivariate Cox regression analysis adjusted by age, gender and stage. In addition, for the exploration of potential mechanisms of NSCLC metastasis, correlation analysis was utilized to construct the NSCLC metastasis-specific regulation network, including key PDEEs, DETGs, DETFs, RPPA protein chips, hallmark signaling pathways, immune-related gene sets, and immune cells. Finally, six key PDEEs including ANXA8L1, CASTOR2, CYP4B1, GTF2H2C, PSMF1 and TNS4 were identified.

ANXA8L1 (annexin A8 like 1) encoded one of the top 10 antigens identified by the majority of serological tests for pemphigus vulgaris patients [26]. High expression of ANXA8L1 was detected to be associated with poor prognosis in cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) [27], but there were few studies on the relationship between ANXA8L1 and lung cancer.

CASTOR2 (cytosolic arginine sensor for mechanistic target of rapamycin complex 1 (mTORC1) subunit 2) is a subtype of CASTOR, a kind of protein readily detectable in vertebrates [28]. CASTOR2 was found to be related to the regulation of mTORC1 [28–30], a central growth controller that integrates diverse environmental inputs to coordinate anabolic and catabolic processes in cells [31], cancer cells might promote growth transformation and tumorigenesis by manipulating CASTOR2, which was reported in Kaposi sarcoma [32].

As a superfamily of enzymes related to phase I drug metabolism, Cytochrome P450 (CYP) was involved in multiple biological processes, such as maintaining calcium homeostasis, fatty acid metabolism, and steroid and cholesterol biosynthesis [33]. CYP4B1 (cytochrome P450 family 4 subfamily B member 1) is an extrahepatic form of cytochrome P450 predominantly and responsible for the bioactivation of multiple protoxins with tissue-specific toxicological effects [34]. CYP4B1 was found to be associated with a variety of cancers [35] and Czerwinski et al. discovered that CYP4B1 from normal and neoplastic lung tissues, compared with normal tissues, had mRNA levels in tumor tissues that were reduced by 2.4 times [36].

GTF2H2C (GTF2H2 family member C) was a subtype of general transcription factor IIH subunit 2-like, related to transcription by RNA polymerase II and DNA nucleotide excision repair [37]. The expression level of transcription factor IIH was identified to be significantly reduced in alveolar macrophages of idiopathic pulmonary fibrosis patients [38]. In fetal lung and placenta, altered methylation may occur in GTF2H2C to repair the DNA damage caused by exposure to smoking [37]. However, the relationship between GTF2H2C and cancer remained unclear.

PSMF1 (proteasome inhibitor subunit 1), the proteasome inhibitor PI31 subunit, was able to bind to the outer rings of the 20S proteasome directly or compete for 20S binding with the activating particles to inhibit the proteasome activities [39–41]. High expression of PSMF1 was shown to be associated with better survival of NSCLC patients, hence PSMF1 was considered as an underlying suppresser gene in NSCLC [42], which demonstrated the accuracy and clinical practicality of our hypothesis.

TNS4 (Tensin 4) participated in the cell movement, which was induced by MET and was related to the GPCR signaling pathway. High expression of TNS4 was reported to be associated with poor prognosis in gastric cancer and esophageal squamous cell carcinoma patients [43,44]. Furthermore, the differential expression and abnormal methylation of

TNS4 were identified in LUAD patients. It was found that high expression of TNS4 leads to poor prognosis, and TNS4 may be involved in the mechanisms of DNA methylation in LUAD, which means it may be a potential marker for the prognosis of LUAD patients. [45].

Six key PDEEs mainly positively regulated TP63, H2AFX, FOS (TFs), EXT1, SLC23A2, GTF2IRD2B, NAIP (DETGs), PAI1, TFRC, NAPSINA, TTF1, CD49B (RPPA protein chips), P53 pathway, G2M checkpoint, fatty acid metabolism, protein secretion, MYC targets V1 (hallmark signaling pathway), Mast cell resting, NK cell resting, Mast cell activated, Macrophage M2 (immune cell), Type_II_IFN_Reponse, and iDCs (immune-related gene sets).

TP63 (tumor protein p63) was associated with the proliferation, migration, colony formation, and invasion of certain squamous cell carcinomas (SCCs) [46,47]. The expression of H2AFX (H2A histone family, member X) histone was promoted by USP22 (Ubiquitin-specific protease 22), which was involved in the occurrence and progression of LUAD [48]. FOS (Fos proto-oncogene, AP-1 transcription factor subunit) can induce the abnormal proliferation of lung cancer cells [48].

EXT1 (exostosin glycosyltransferase 1) methylation can regulate gene expression and activate the WNT pathway, which affected the proliferation and migration of NSCLC and predicted a poor prognosis [49]. SLC23A2 (solute carrier family 23 member 2) was known to be required for sodium-dependent transporters of vitamin C [50], and several types of cancer were reported to be linked with a deficiency in vitamin C [51]. GTF2IRD2B (GTF2I repeat domain containing 2B) was involved in chromatin structure modification and gene expression regulation [52], mutations in GTF2IRD2B may cause disorders of gene expression regulation and contribute to carcinogenesis [53]. NAIP (NLR family apoptosis inhibitory protein), is a major anti-apoptotic protein and is targeted by miR-1 and miR-145, which induces cell death and contributes to the development of cancer[54].

PAI1 (phosphoribosylanthranilate isomerase 1) promoted glycolytic metabolism [55], and the migration and chemotaxis of cancer cells relies on the energy obtained via enhanced glycolysis primarily [56,57]. EGFR (Epidermal growth factor receptor) plays an important role in the regulation of the proliferation, differentiation, survival and motility of the tumor cells and was found to be highly expressed in over 60% of NSCLCs [58], which indicates that it has the potential to promote NSCLC metastasis. TFRC (transferrin receptor) promoted the proliferation and metastasis of cancer cells by upregulating the expression of AXIN2, which accelerated the development of cancer [59]. NapsinA (napsin A aspartic peptidase) and TTF1 (thyroid transcription factor-1) were reported as the specific clinical diagnosis indexes and prognostic markers for LUAD [60–62]. CD49B was shown to be the cell-surface marker for the enrichment of a subpopulation of leiomyoma cells that possess stem/progenitor cell properties [63]. The functions of RPPA and their regulatory relationship with PDEE were shown in Table S2.

The transcription factor p53 participated in the mechanism of the cell cycle and was reported as an important tumor suppressor [64]. The G2/M checkpoint checked cell size and DNA damage in mitosis of multiple organisms. Disorder of mitotic entry can often cause oncogenesis or cell death [65]. Deregulated anabolism and catabolism of fatty acids metabolic were identified as metabolic regulators that support cancer cell growth [66]. Protein secretion signaling was reported to be associated with multiple biological processes, such as cancer cell migration and invasion [67]. MYC targets v1 signaling and might be related to the higher rates of cell proliferation, resulting in increased aggressiveness of the tumor and worse survival [68]. The tumor microenvironment (TME) was significantly associated with the pathogenesis of lung cancer [69,70]. The proportion of resting NK cells and mast cells resting in lung cancer tumor tissues was reported to be lower than in normal tissues [71]. Mast cells and their activation might result in tumor cytotoxicity and tumor angiogenesis [72]. M2 Macrophages promote the growth, invasion, metastasis, and angiogenesis of cancer cells, and are regarded as one of the main tumor-infiltrating immune cells [73]. IFN-γ (type II IFN) played an important role in anti-tumor responses [74]. IFN-γ receptor impairment or IFN-γ-mediated signal disruption may cause

insensitivity to IFN-γ, which may promote the development and progress of tumors [75]. Dendritic cells (DCs) were associated with the regulation of the immune response and played key roles in inducing anti-tumor activity [76], a decrease in immature dendritic cells (iDC) and impaired migration ability may cause cancer [77].

Irinotecan is a topoisomerase inhibitor that causes cytotoxic protein-linked DNA breaks [78], which is a kind of p53-dependent small-molecule bioactive inhibitor [79]. Linked to the results of Cmap, irinotecan may exert an inhibitory effect through the P53 pathway.

Five key eRNAs (ANXA8L1, CYP4B1, GTF2H2C, PSMF1, and TNS4) were found to be expressed in Alveolar, B, NK/T, Endothelial, Epithelial, Fibroblast and Myeloid cells in single cell sequencing analysis and the analysis of the scRNA-seq transcriptomes. The tumor-microenvironment was detected to play an important role in each step of NSCLC metastasis, and the synergistic mechanisms of tumor cells and the microenvironment may provide biomarkers or potential therapeutic targets for cancers [80], which suggests that eRNAs may regulate NSCLC metastasis by interacting with the tumor microenvironment.

The ATAC-seq validation showed that six key PDEEs were chromatin accessible in the NSCLC metastasis state, which indicated that these eRNAs were related to NSCLC metastasis. In addition, through Chip-seq validation analysis, these key PDEEs were found to bind to the binding sites of enhancers and DETFs in this study. Therefore, our study verified that these key eRNAs regulated target genes and mediated NSCLC metastasis by recruiting TFs, which may be implicated in the metastasis of NSCLC.

This is the first study that explored the roles that eRNAs played in NSCLC metastasis, however, several limitations still remained in our study. Firstly, since the samples were all from America, selection bias was inevitable, and the applicability of the prediction model to other countries was uncertain. Secondly, our results were completely based on public databases and have not been verified by our own population studies and experiments. Therefore, multidimensional validation was performed to validate our hypothesis based on various online databases in the multi-omics dimension. Moreover, further cell and animal experimental validation and clinical trials were warranted in future studies, demonstrating correlations between identified key PDEEs and other multi-omics biomarkers, and validating the clinical relevance of our key findings regarding novel eRNA regulatory mechanisms from multiple dimensions.

4. Materials and Methods

4.1. Data Acquisition

RNA-seq data from 1011 primary NSCLC, clinical information from 829 primary NSCLC, and 258 protein chips were obtained from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov, accessed on 3 August 2020) [81]. RNA-seq data from 42 metastatic NSCLC were obtained from the MET500 database (https://met500.path.med.umich.edu/, accessed on 4 August 2020) [82], 318 TFs were obtained from the Cistrome database (http://cistrome.org, accessed on 14 August 2020) [83]. Immune gene expression profiles were downloaded from the ImmPort database (https://www.immport.org/, accessed on 19 February 2020). Fifty cancer-related hallmark pathways and 29 immune-related pathways were obtained from the Molecular Signatures Database (MSigDB, Version 7.4) (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp, accessed on 10 September 2020) [84].

4.2. The eRNA Expression Data

The list of eRNAs, which was normalized and annotated by Ensemble ID and the corresponding target gene list was available in the eRic (enhancer RNA in cancers) database (https://hanlab.uth.edu/eRic/, accessed on 6 June 2020) [19] in addition to the Chromatin Immunoprecipitation Sequencing (CHIP-seq) results containing acetylated histone

H3 lysine 27 (H3K27ac) [85]. In addition, the official gene symbol of each eRNA was iden-
tified by the CHIP seeker package based on the location in the hg38 genome [86].

4.3. Differential Expression Analysis

Differentially expressed genes (DEGs) between primary NSCLC and metastatic
NSCLC were identified by the limma and edgeR algorithm [87,88] with the criteria of a
False Discovery Rate (FDR) *p*-value < 0.05 and |log2 Fold Change (FC)| > 1.0; DEEs,
DETGs, and DETFs were all identified. To determine the most enriched pathways, the
Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Oncology (GO) enrichment
analysis was applied[89].

4.4. Multivariate Risk-Prediction Model Construction and Independent Prognostic Factors Identification

Prognostic DEEs (PDEEs) were screened by univariate Cox regression analysis from
identified DEEs. Furthermore, the Least Absolute Shrinkage and Selection Operator
(LASSO) regression was used to determine the independent variables with great signifi-
cance, reducing the over-fitting phenomenon [90]. Afterward, we integrated all PDEEs
into a multivariate Cox regression model, and the receiver operator characteristic (ROC)
curve was used to evaluate the predictive power of the model. The multivariate Cox re-
gression model formula was applied to calculate the risk score (RS) for each NSCLC sam-
ple as follows:

$$RS = \beta_1 \times NSCLC_1 + \beta_2 \times NSCLC_2 + \beta_3 \times NSCLC_3 + \ldots\ldots + \beta_n \times NSCLC_n$$

In this formula, "n" represented the number of PDEEs in the multivariate model; "β"
represented the coefficient of corresponding PDEEs. Then, the median of the RS was used
to divide all the NSCLC samples into two groups, including the high-risk group and the
low-risk group. Additionally, the survival conditions of the high-risk group and the low-
risk group were described by Kaplan–Meier survival analysis. The independent prognos-
tic value of RS was evaluated by univariate and multivariate Cox regression analyses, and
the multivariate Cox regression model was adjusted by clinical variables, such as age,
gender, and stage.

4.5. Identification of PDEE-Related Immune Cells and Immune-Reltaed Gene Sets

The composition of 22 types of immune cells in all samples was evaluated by the
CIBERSORT algorithm. To determine the correlations between PDEE signature and im-
mune cell infiltration in NSCLC tissues, PDEE expression matrix data were uploaded to
CIBERSORT [21]. Infiltrating immune cells were all extracted for subsequent analysis.
Moreover, nonparametric tests were performed to determine associations between the im-
mune cells and different clinical phenotypes.

Additionally, a single-sample gene set enrichment analysis (ssGSEA) was carried out
to quantify 29 immune-related gene sets in all samples[22]. Further, Pearson correlation
analysis was performed to identify significant correlations between PDEEs and immune
cells/immune-related gene sets, where a *p*-value < 0.05 was considered statistically signif-
icant.

4.6. Identification of Downstream Hallmark Pathways

Gene Set Variation Analysis (GSVA) was utilized to explore potential downstream
hallmark pathways of PDEEs. Absolute quantification of 50 hallmark pathways was cal-
culated to determine the differentially expressed hallmark pathways between primary
NSCLC samples and metastatic NSCLC samples using the GSVA package [23].

4.7. Construction of Metastasis-Specific eRNA Regulation Network for NSCLC

Key PDEEs along with DETGs, DETFs, RPPA proteins chips, 50 hallmark pathways, 29 immune gene sets, 22 immune cells, were integrated into Pearson correlation analysis to construct a NSCLC metastasis-specific eRNA regulation network with the criterion of |correlation coefficient| > 0.20 and *p* value < 0.05 for PDEEs and DETGs; |correlation coefficient| > 0.20 and *p* value < 0.05 for PDEEs and DETFs; |correlation coefficient| > 0.25 and *p* value < 0.05 for PDEEs and RPPA protein chips; |correlation coefficient| > 0.25 and *p* value < 0.05 for PDEEs and hallmark pathways; |correlation coefficient| > 0.20 and *p* value < 0.05 for PDEEs and immune-related gene sets, and |correlation coefficient| > 0.15 and *p* value < 0.05 for PDEEs and immune cells.

Furthermore, the connectivity map (Cmap) database was utilized to identify bioactive small molecule inhibitors with the potential as target drugs for DEGs of the NSCLC metastasis-specific regulation network (https://portals.broadinstitute.org/cmap/, accessed on 25 September 2020) [91,92]. DEGs, including PDEEs, DETGs, and DETFs were used as input data in the Cmap analysis, results of stemness-related DEG analysis in pan-cancer were also integrated into the Cmap analysis [91]. Information on targeting inhibitors could be acquired from the mechanism of actions (MoA) (http://clue.io/, accessed on 25 September 2020) including human cell lines' transcriptional responses to perturbagens, structural formulas, and protein targets. Therefore, based on MoA, inhibitors that may target NSCLC metastasis-related DEGs in this study were all identified.

4.8. Analysis of scRNA-Seq Transcriptomes

scRNA-seq data from 24 NSCLC samples obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153935, accessed on 8 June 2022) were applied to analyze the cellular localization of key PDEEs. The analysis of the integrated data was carried out by the R toolkit Seurat [93]. Genes that were expressed in over 200 single cells and cells with 1500 to 100,000 gene transcripts were selected for further analysis. Moreover, the "vst" method, "FindMarkers" and "FindConservedMarkers" functions were utilized to identify the variable genes, and the marker genes for each cell type were defined. Then, the variable genes above were used to carry out the principal component analysis (PCA) to reduce data dimensionality. The top 15 PCs were selected for further analysis, including Unified Manifold Approximation and Dimensionality Reduction Projection (UMAP) analysis and cluster analysis. The cell clusters, which were based on key PCs, were identified by t-SNE (resolution = 0.50) [94] and absolute values of FDR < 0.05 and log2 (FC) > 0.5 were the criterion for DEG in each cell cluster. Annotation for every cluster was performed by the CellMarker database and the singleR method [95,96]. Visualization of cell cycle stages was performed by the "CellCycleScoring" function and markers of phases. Finally, in different cell types, the pairs of receptor and ligand were identified by the "iTALK" package[97] and visualization of intercellular communication was performed by the "edgebundleR" package (https://github.com/garthtarr/edgebundleR, accessed on 8 June 2022).

4.9. Multidimensional Validation

For balancing false positive results and reducing information bias, multi-database external verification was performed based on various online databases including Gene Expression Profiling Interactive Analysis (GEPIA) [98], UCSC xena [99], The Human Protein Atlas [100], cBioPortal [101,102], UALCAN [103] and OncoLnc [104]. In addition, Single Cell Expression Atlas was adopted to identify eRNA expression at the cellular level (https://www.ebi.ac.uk/gxa/sc/experiments, accessed on 30 August 2021) [105]. The assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) data was obtained from the TCGA database to explore the accessibility of eRNAs in chromatin [106]. Chromatin immunoprecipitation sequencing (Chip-seq) data of eRNAs were collected from the Cistrome database to identify specific binding relationships between key

PDEEs and DETFs in this study (http://cistrome.org/, accessed on 30 August 2021) [107–117].

4.10. Statistics Analysis

All statistical analyses were performed by R version 3.6.1 (Institute for Statistics and Mathematics, Vienna, Austria) (Package: e1071, parallel, preprocessCore, sva, limma, edgeR, ggplot2, survminer, survival, rms, randomForest, pROC, glmnet, pheatmap, timeROC, vioplot, corrplot, ConsensusClusterPlus, forestplot, survivalROC, beeswarm, edgeR, chromVAR, Biostrings, BSgenome.Hsapiens.UCSC.hg38, ChIPseeker, TxDb.Hsapiens.UCSC.hg38.knownGene, clusterProfiler, org.Hs.eg.db, ggplot2, karyoploteR, limma, pheatmap, GSVA, limma, GSEABase, stringr, GEOquery, dplyr, limma, ComplexHeatmap, RColorBrewer, clusterProfiler, tibble, ggplot2, cowplot, ggcorrplot, xlsx, tidyverse, GEOquery, plyr, circlize, ComplexHeatmap, TCGAbiolinks, SummarizedExperiment, dplyr, tidyverse, fgsea, ggplot2, ImmuLncRNA, iTALK, edgebundleR). A two-sided p-value < 0.05 was considered statistically significant, and the utilization of the Pearson/Spearman correlation coefficient depended on the results of the normality tests. The machine diagram was drawn on the biorender website (https://biorender.com/, accessed on 2 June 2022).

5. Conclusions

This study verified that eRNAs played significant roles in NSCLC metastasis. Moreover, six key PDEEs (ANXA8L1, CASTOR2, CYP4B1, GTF2H2C, PSMF1, and TNS4) were identified as potential markers to predict the prognosis of NSCLC and provide references for the treatment of metastatic NSCLC, which occupied the central position in the NSCLC metastasis-specific regulation network.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/molecules27134108/s1, Table S1: The target genes and drug response information of eRNAs; Figure S1 A: Mechanism by which eRNAs promote transcription of target genes; Figure S1 B: flow chart of our study; Figure S2: GEPIA database validation; Figure S3: UCSC Xena database validation; Figure S4: the Human Protein Atlas database validation; Figure S5: cBioPortal database validation; Figure S6: UALCAN database validation in LUAD; Figure S7: UALCAN database validation in LUSC; Figure S8: OncoLnc database validation; Figure S9: eRic database validation (a); Figure S10: eRic database validation (b); Figure S11: eRic database validation (c); Figure S12: eRic database validation (d); Figure S13: eRic database validation (e); Figure S14: eRic database validation (f); Figure S15: ATAC-seq validation for LUAD (a); Figure S16: ATAC-seq validation for LUAD (b); Figure S17: ATAC-seq validation for LUSC (a); Figure S18: ATAC-seq validation for LUSC (b); Figure S19: Chip-seq validation; Figure S20: Scatter plot of the expression of ANXA8L1 and RPPA|EGFR; Figure S21: Scatter plot of the expression of ANXA8L1 and GTF2H2C|KEAP1; Figure S22: Scatter plot of the expression of CASTOR2 and HALLMARK_KRAS_SIGNALING_UP; Figure S23: Feature plots of the expressions of 3 key TFs (TP63, H2AFX and FOS) and 4 key DETGs (EXT1, GTF2IRD2B, NAIP and SLC23A2) in 7 cell types.

Author Contributions: Conceptualization, J.L. and J.J.; methodology, J.L.; software, S.X., J.L. and J.J.; validation, J.J.; data curation, Z.Z.; writing—original draft preparation, J.J.; writing—review and editing, S.W.; visualization, L.D.; supervision, J.Z.; project administration, J.Z. and Y.F.; funding acquisition, Y.Z., J.L., J.J. have contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable, only public data was used in the study.

Informed Consent Statement: Not applicable, only public data was used in the study.

Abbreviations

WHO: World Health Organization; NSCLC, Non-small cell lung cancer; LUAD, lung adeno-carcinoma; LUSC, lung squamous cell carcinoma; eRNAs, enhancer RNAs; TCGA, The Cancer Genome Atlas; eRic, enhancer RNA in cancers; H3K27ac, containing acetylated histone H3 lysine 27; DEGs, differential expressed genes; DEEs, differential expressed eRNAs; PDEEs, Prognostic DEEs; AUC, area under the curve; ROC, receiver operator characteristic; TFs, transcription factors; GSVA, gene set variation analysis; ssGSEA, single sample gene set enrichment analysis; FC, Fold Change; FDR, False Discovery Rate; GO, Gene Oncology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BPs, biological processes; CCs, cellular components; MFs, molecular functions; RS, risk score; Cmap, connectivity map; GEPIA, Gene Expression Profiling Interactive Analysis; scRNA-seq, single-cell RNA sequencing; t-SNE, t-distributed Stochastic Neighbor Embedding; ATAC-seq, assay for transposase-accessible chromatin with high-throughput sequencing; Chip-seq, Chromatin immunoprecipitation sequencing; Hi-C, high-throughput chromatin conformation capture; ANXA8L1, annexin A8 like 1; CASTOR2, cytosolic arginine sensor for mTORC1 subunit 2; CYP4B1, cytochrome P450 family 4 subfamily B member 1; GTF2H2C, GTF2H2 family member C; PSMF1, proteasome inhibitor subunit 1; TNS4, Tensin 4; TP63, tumor protein p63; squamous cell carcinomas, SCCs; H2AFX, H2A histone family, member X USP22, Ubiquitin-specific protease 22; FOS, Fos proto-oncogene, AP-1 transcription factor subunit; EXT1, exostosin glycosyltransferase 1; SLC23A2, solute carrier family 23 member 2; GTF2IRD2B, GTF2I repeat domain containing 2B; NAIP, NLR family apoptosis inhibitory protein; PAI1, phosphoribosylanthranilate isomerase 1; TFRC, transferrin receptor; NapsinA, napsin A aspartic peptidase; TTF1, thyroid transcription factor-1; TME, tumor microenvironment; type II IFN, type II interferon; DCs, Dendritic cells; iDCs, immature dendritic cells.

References

1. Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. Global cancer statistics, 2012. CA: a cancer journal for clinicians 2015, 65, 87–108, doi:10.3322/caac.21262.
2. Dela Cruz, C.S.; Tanoue, L.T.; Matthay, R.A. Lung cancer: epidemiology, etiology, and prevention. Clinics in chest medicine 2011, 32, 605–644, doi:10.1016/j.ccm.2011.09.001.
3. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clinic proceedings 2008, 83, 584–594, doi:10.4065/83.5.584.
4. Duma, N.; Santana-Davila, R.; Molina, J.R. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clinic proceedings 2019, 94, 1623–1640, doi:10.1016/j.mayocp.2019.01.013.
5. Strand, T.E.; Rostad, H.; Møller, B.; Norstein, J. Survival after resection for primary lung cancer: a population based study of 3211 resected patients. Thorax 2006, 61, 710–715, doi:10.1136/thx.2005.056481.
6. Miller, K.D.; Nogueira, L.; Mariotto, A.B.; Rowland, J.H.; Yabroff, K.R.; Alfano, C.M.; Jemal, A.; Kramer, J.L.; Siegel, R.L. Cancer treatment and survivorship statistics, 2019. CA: a cancer journal for clinicians 2019, 69, 363–385, doi:10.3322/caac.21565.
7. Pennacchio, L.A.; Bickmore, W.; Dean, A.; Nobrega, M.A.; Bejerano, G. Enhancers: five essential questions. Nature reviews. Genetics 2013, 14, 288–295, doi:10.1038/nrg3458.
8. Mossing, M.C.; Record, M.T., Jr. Upstream operators enhance repression of the lac promoter. Science (New York, N.Y.) 1986, 233, 889–892, doi:10.1126/science.3090685.
9. Blackwood, E.M.; Kadonaga, J.T. Going the distance: a current view of enhancer action. Science (New York, N.Y.) 1998, 281, 60–63, doi:10.1126/science.281.5373.60.
10. Bulger, M.; Groudine, M. Enhancers: the abundance and function of regulatory sequences beyond promoters. Developmental biology 2010, 339, 250–257, doi:10.1016/j.ydbio.2009.11.035.
11. Kim, T.K.; Hemberg, M.; Gray, J.M.; Costa, A.M.; Bear, D.M.; Wu, J.; Harmin, D.A.; Laptewicz, M.; Barbara-Haley, K.; Kuersten, S.; et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010, 465, 182–187, doi:10.1038/nature09033.

12.   Li, W.; Notani, D.; Rosenfeld, M.G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nature reviews. Genetics 2016, 17, 207–223, doi:10.1038/nrg.2016.4.

13.   Mousavi, K.; Zare, H.; Dell'orso, S.; Grontved, L.; Gutierrez-Cruz, G.; Derfoul, A.; Hager, G.L.; Sartorelli, V. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Molecular cell 2013, 51, 606–617, doi:10.1016/j.molcel.2013.07.022.

14.   Maruyama, A.; Mimura, J.; Itoh, K. Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding. Nucleic acids research 2014, 42, 13599–13614, doi:10.1093/nar/gku1169.

15.   Li, W.; Notani, D.; Ma, Q.; Tanasa, B.; Nunez, E.; Chen, A.Y.; Merkurjev, D.; Zhang, J.; Ohgi, K.; Song, X.; et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 2013, 498, 516–520, doi:10.1038/nature12210.

16.   Li, W.; Hu, Y.; Oh, S.; Ma, Q.; Merkurjev, D.; Song, X.; Zhou, X.; Liu, Z.; Tanasa, B.; He, X.; et al. Condensin I and II Complexes License Full Estrogen Receptor α-Dependent Enhancer Activation. Molecular cell 2015, 59, 188–202, doi:10.1016/j.molcel.2015.06.002.

17.   Hsieh, C.L.; Fei, T.; Chen, Y.; Li, T.; Gao, Y.; Wang, X.; Sun, T.; Sweeney, C.J.; Lee, G.S.; Chen, S.; et al. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proceedings of the National Academy of Sciences of the United States of America 2014, 111, 7319–7324, doi:10.1073/pnas.1324151111.

18.   Lee, J.H.; Xiong, F.; Li, W. Enhancer RNAs in cancer: regulation, mechanisms and therapeutic potential. RNA biology 2020, 17, 1550–1559, doi:10.1080/15476286.2020.1712895.

19.   Zhang, Z.; Lee, J.H.; Ruan, H.; Ye, Y.; Krakowiak, J.; Hu, Q.; Xiang, Y.; Gong, J.; Zhou, B.; Wang, L.; et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. Nature communications 2019, 10, 4562, doi:10.1038/s41467-019-12543-5.

20.   Léveillé, N.; Melo, C.A.; Agami, R. Enhancer-associated RNAs as therapeutic targets. Expert opinion on biological therapy 2015, 15, 723–734, doi:10.1517/14712598.2015.1029452.

21.   Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015, 12, 453–457, doi:10.1038/nmeth.3337.

22.   Huang, R.; Li, Z.; Zhu, X.; Yan, P.; Song, D.; Yin, H.; Hu, P.; Lin, R.; Wu, S.; Meng, T.; et al. Collagen Type III Alpha 1 chain regulated by GATA-Binding Protein 6 affects Type II IFN response and propanoate metabolism in the recurrence of lower grade glioma. Journal of cellular and molecular medicine 2020, 24, 10803–10815, doi:10.1111/jcmm.15705.

23.   Hänzelmann, S.; Castelo, R.; Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 2013, 14, 7, doi:10.1186/1471-2105-14-7.

24.   Houston, K.A.; Henley, S.J.; Li, J.; White, M.C.; Richards, T.B. Patterns in lung cancer incidence rates and trends by histologic type in the United States, 2004-2009. Lung cancer (Amsterdam, Netherlands) 2014, 86, 22–28, doi:10.1016/j.lungcan.2014.08.001.

25.   Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. CA: a cancer journal for clinicians 2018, 68, 7–30, doi:10.3322/caac.21442.

26.   Kalantari-Dehaghi, M.; Anhalt, G.J.; Camilleri, M.J.; Chernyavsky, A.I.; Chun, S.; Felgner, P.L.; Jasinskas, A.; Leiferman, K.M.; Liang, L.; Marchenko, S.; et al. Pemphigus vulgaris autoantibody profiling by proteomic technique. PloS one 2013, 8, e57587, doi:10.1371/journal.pone.0057587.

27.   Ding, H.; Xiong, X.X.; Fan, G.L.; Yi, Y.X.; Chen, Y.R.; Wang, J.T.; Zhang, W. The New Biomarker for Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC) Based on Public Database Mining. BioMed research international 2020, 2020, 5478574, doi:10.1155/2020/5478574.

28.   Chantranupong, L.; Scaria, S.M.; Saxton, R.A.; Gygi, M.P.; Shen, K.; Wyant, G.A.; Wang, T.; Harper, J.W.; Gygi, S.P.; Sabatini, D.M. The CASTOR Proteins Are Arginine Sensors for the mTORC1 Pathway. Cell 2016, 165, 153–164, doi:10.1016/j.cell.2016.02.035.

29.   Xia, J.; Wang, R.; Zhang, T.; Ding, J. Structural insight into the arginine-binding specificity of CASTOR1 in amino acid-dependent mTORC1 signaling. Cell discovery 2016, 2, 16035, doi:10.1038/celldisc.2016.35.

30.   Saxton, R.A.; Chantranupong, L.; Knockenhauer, K.E.; Schwartz, T.U.; Sabatini, D.M. Mechanism of arginine sensing by CASTOR1 upstream of mTORC1. Nature 2016, 536, 229–233, doi:10.1038/nature19079.

31.   Saxton, R.A.; Sabatini, D.M. mTOR Signaling in Growth, Metabolism, and Disease. Cell 2017, 168, 960–976, doi:10.1016/j.cell.2017.02.004.

32.   Li, T.; Ju, E.; Gao, S.J. Kaposi sarcoma-associated herpesvirus miRNAs suppress CASTOR1-mediated mTORC1 inhibition to promote tumorigenesis. The Journal of clinical investigation 2019, 129, 3310–3323, doi:10.1172/jci127166.

33.   Adhikari, A.; De, M.J.A.J.P.P. De M. Association of cytochrome P450 with cancer induced by Betel quid (BQ): a review. Am. J. Pharmacol. Pharmacother. 2015, 2, 81–94.

34.   Baer, B.R.; Rettie, A.E. CYP4B1: an enigmatic P450 at the interface between xenobiotic and endobiotic metabolism. Drug metabolism reviews 2006, 38, 451–476, doi:10.1080/03602530600688503.

35.   Lim, S.; Alshagga, M.; Ong, C.E.; Chieng, J.Y.; Pan, Y. Cytochrome P450 4B1 (CYP4B1) as a target in cancer treatment. Human & experimental toxicology 2020, 39, 785–796, doi:10.1177/0960327120905959.

36.   Czerwinski, M.; McLemore, T.L.; Gelboin, H.V.; Gonzalez, F.J.J.C.R. Quantification of CYP2B7, CYP4B1, and CYPOR messenger RNAs in normal human lung and lung tumors. Cancer Res. 1994, 54, 1085–1091.

37. Chhabra, D.; Sharma, S.; Kho, A.T.; Gaedigk, R.; Vyhlidal, C.A.; Leeder, J.S.; Morrow, J.; Carey, V.J.; Weiss, S.T.; Tantisira, K.G.J.E. Fetal lung and placental methylation is associated with in utero nicotine exposure. Epigenetics 2014, 9, 1473–1484.

38. Ren, P.; Rosas, I.O.; Macdonald, S.D.; Wu, H.P.; Billings, E.M.; Gochuico, B.R. Impairment of alveolar macrophage transcription in idiopathic pulmonary fibrosis. American journal of respiratory and critical care medicine 2007, 175, 1151–1157, doi:10.1164/rccm.200607-958OC.

39. McCutchen-Maloney, S.L.; Matsuda, K.; Shimbara, N.; Binns, D.D.; Tanaka, K.; Slaughter, C.A.; DeMartino, G.N. cDNA cloning, expression, and functional characterization of PI31, a proline-rich inhibitor of the proteasome. The Journal of biological chemistry 2000, 275, 18557–18565, doi:10.1074/jbc.M001697200.

40. Zaiss, D.M.; Standera, S.; Holzhütter, H.; Kloetzel, P.; Sijts, A.J. The proteasome inhibitor PI31 competes with PA28 for binding to 20S proteasomes. FEBS letters 1999, 457, 333–338, doi:10.1016/s0014-5793(99)01072-8.

41. Cho-Park, P.F.; Steller, H. Proteasome regulation by ADP-ribosylation. Cell 2013, 153, 614–627, doi:10.1016/j.cell.2013.03.040.

42. Yang, S.; Tang, D.; Zhao, Y.C.; Liu, H.; Luo, S.; Stinchcombe, T.E.; Glass, C.; Su, L.; Shen, S.; Christiani, D.C.; et al. Potentially functional variants of ERAP1, PSMF1 and NCF2 in the MHC-I-related pathway predict non-small cell lung cancer survival. Cancer immunology, immunotherapy : CII 2021, 70, 2819–2833, doi:10.1007/s00262-021-02877-9.

43. Sakashita, K.; Mimori, K.; Tanaka, F.; Kamohara, Y.; Inoue, H.; Sawada, T.; Hirakawa, K.; Mori, M. Prognostic relevance of Tensin4 expression in human gastric cancer. Annals of surgical oncology 2008, 15, 2606–2613, doi:10.1245/s10434-008-9989-8.

44. Shi, Z.Z.; Wang, W.J.; Chen, Y.X.; Fan, Z.W.; Xie, X.F.; Yang, L.Y.; Chang, C.; Cai, Y.; Hao, J.J.; Wang, M.R.; et al. The miR-1224-5p/TNS4/EGFR axis inhibits tumour progression in oesophageal squamous cell carcinoma. Cell death & disease 2020, 11, 597, doi:10.1038/s41419-020-02801-6.

45. Wang, X.; Shi, D.; Zhao, D.; Hu, D. Aberrant Methylation and Differential Expression of SLC2A1, TNS4, GAPDH, ATP8A2, and CASZ1 Are Associated with the Prognosis of Lung Adenocarcinoma. BioMed research international 2020, 2020, 1807089, doi:10.1155/2020/1807089.

46. Xie, J.J.; Jiang, Y.Y.; Jiang, Y.; Li, C.Q.; Lim, M.C.; An, O.; Mayakonda, A.; Ding, L.W.; Long, L.; Sun, C.; et al. Super-Enhancer-Driven Long Non-Coding RNA LINC01503, Regulated by TP63, Is Over-Expressed and Oncogenic in Squamous Cell Carcinoma. Gastroenterology 2018, 154, 2137-2151.e2131, doi:10.1053/j.gastro.2018.02.018.

47. Kandoth, C.; Schultz, N.; Cherniack, A.D.; Akbani, R.; Liu, Y.; Shen, H.; Robertson, A.G.; Pashtan, I.; Shen, R.; Benz, C.C.; et al. Integrated genomic characterization of endometrial carcinoma. Nature 2013, 497, 67–73, doi:10.1038/nature12113.

48. Han, B.; Sun, Y.; Yang, D.; Zhang, H.; Mo, S.; Chen, X.; Lu, H.; Mao, X.; Hu, J. USP22 promotes development of lung adenocarcinoma through ubiquitination and immunosuppression. Aging 2020, 12, 6990–7005, doi:10.18632/aging.103056.

49. Kong, W.; Chen, Y.; Zhao, Z.; Zhang, L.; Lin, X.; Luo, X.; Wang, S.; Song, Z.; Lin, X.; Lai, G.; et al. EXT1 methylation promotes proliferation and migration and predicts the clinical outcome of non-small cell lung carcinoma via WNT signalling pathway. Journal of cellular and molecular medicine 2021, 25, 2609–2620, doi:10.1111/jcmm.16277.

50. Takanaga, H.; Mackenzie, B.; Hediger, M.A. Sodium-dependent ascorbic acid transporter family SLC23. Pflugers Archiv : European journal of physiology 2004, 447, 677–682, doi:10.1007/s00424-003-1104-1.

51. Gillberg, L.; Ørskov, A.D.; Liu, M.; Harsløf, L.B.S.; Jones, P.A.; Grønbæk, K. Vitamin C - A new player in regulation of the cancer epigenome. Seminars in cancer biology 2018, 51, 59–67, doi:10.1016/j.semcancer.2017.11.001.

52. Gunbin, K.V.; Ruvinsky, A. Evolution of general transcription factors. Journal of molecular evolution 2013, 76, 28–47, doi:10.1007/s00239-012-9535-y.

53. Candia, J.; Bayarsaikhan, E.; Tandon, M.; Budhu, A.; Forgues, M.; Tovuu, L.O.; Tudev, U.; Lack, J.; Chao, A.; Chinburen, J.; et al. The genomic landscape of Mongolian hepatocellular carcinoma. Nature communications 2020, 11, 4383, doi:10.1038/s41467-020-18186-1.

54. Xu, X.; Wu, X.; Jiang, Q.; Sun, Y.; Liu, H.; Chen, R.; Wu, S. Downregulation of microRNA-1 and microRNA-145 contributes synergistically to the development of colon cancer. International journal of molecular medicine 2015, 36, 1630–1638, doi:10.3892/ijmm.2015.2364.

55. Humphries, B.A.; Buschhaus, J.M.; Chen, Y.C.; Haley, H.R.; Qyli, T.; Chiang, B.; Shen, N.; Rajendran, S.; Cutter, A.; Cheng, Y.H.; et al. Plasminogen Activator Inhibitor 1 (PAI1) Promotes Actin Cytoskeleton Reorganization and Glycolytic Metabolism in Triple-Negative Breast Cancer. Molecular cancer research : MCR 2019, 17, 1142–1154, doi:10.1158/1541-7786.Mcr-18-0836.

56. Beckner, M.E.; Stracke, M.L.; Liotta, L.A.; Schiffmann, E. Glycolysis as primary energy source in tumor cell chemotaxis. Journal of the National Cancer Institute 1990, 82, 1836–1840, doi:10.1093/jnci/82.23.1836.

57. Shiraishi, T.; Verdone, J.E.; Huang, J.; Kahlert, U.D.; Hernandez, J.R.; Torga, G.; Zarif, J.C.; Epstein, T.; Gatenby, R.; McCartney, A.; et al. Glycolysis is the primary bioenergetic pathway for cell motility and cytoskeletal remodeling in human prostate and breast cancer cells. Oncotarget 2015, 6, 130–143, doi:10.18632/oncotarget.2766.

58. da Cunha Santos, G.; Shepherd, F.A.; Tsao, M.S. EGFR mutations and lung cancer. Annual review of pathology 2011, 6, 49–69, doi:10.1146/annurev-pathol-011110-130206.

59. Huang, Y.; Huang, J.; Huang, Y.; Gan, L.; Long, L.; Pu, A.; Xie, R. TFRC promotes epithelial ovarian cancer cell proliferation and metastasis via up-regulation of AXIN2 expression. American journal of cancer research 2020, 10, 131–147.

60. Jiang, Y.; Zhao, J.; Zhang, Y.; Li, K.; Li, T.; Chen, X.; Zhao, S.; Zhao, S.; Liu, K.; Dong, Z. Establishment of lung cancer patient-derived xenograft models and primary cell lines for lung cancer study. Journal of translational medicine 2018, 16, 138, doi:10.1186/s12967-018-1516-5.

61. Yatabe, Y.; Dacic, S.; Borczuk, A.C.; Warth, A.; Russell, P.A.; Lantuejoul, S.; Beasley, M.B.; Thunnissen, E.; Pelosi, G.; Rekhtman, N.; et al. Best Practices Recommendations for Diagnostic Immunohistochemistry in Lung Cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 2019, 14, 377–407, doi:10.1016/j.jtho.2018.12.005.

62. Yu, H.; Li, L.; Liu, D.; Li, W.M. [Expression of TTF-1, NapsinA, P63, CK5/6 in Lung Cancer and Its Diagnostic Values for Histological Classification]. Sichuan da xue xue bao. Yi xue ban = Journal of Sichuan University. Medical science edition 2017, 48, 336–341.

63. Yin, P.; Ono, M.; Moravek, M.B.; Coon, J.S.t.; Navarro, A.; Monsivais, D.; Dyson, M.T.; Druschitz, S.A.; Malpani, S.S.; Serna, V.A.; et al. Human uterine leiomyoma stem/progenitor cells expressing CD34 and CD49b initiate tumors in vivo. The Journal of clinical endocrinology and metabolism 2015, 100, E601-606, doi:10.1210/jc.2014-2134.

64. Lane, D.; Levine, A. p53 Research: The Past Thirty Years and the Next Thirty Years. Cold Spring Harbor Perspect. Biol. 2010, 2, 10, doi:10.1101/cshperspect.a000893.

65. Yasutis, K.M.; Kozminski, K.G. Cell cycle checkpoint regulators reach a zillion. Cell cycle (Georgetown, Tex.) 2013, 12, 1501–1509, doi:10.4161/cc.24637.

66. Li, Z.; Zhang, H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cellular and molecular life sciences : CMLS 2016, 73, 377–392, doi:10.1007/s00018-015-2070-4.

67. Liu, C.; Qu, L.; Shou, C. Role and Characterization of Synuclein-γ Unconventional Protein Secretion in Cancer Cells. Methods in molecular biology (Clifton, N.J.) 2016, 1459, 215–227, doi:10.1007/978-1-4939-3804-9_15.

68. Schulze, A.; Oshi, M.; Endo, I.; Takabe, K. MYC Targets Scores Are Associated with Cancer Aggressiveness and Poor Survival in ER-Positive Primary and Metastatic Breast Cancer. International journal of molecular sciences 2020, 21, 8127, doi:10.3390/ijms21218127.

69. Korniluk, A.; Koper, O.; Kemona, H.; Dymicka-Piekarska, V. From inflammation to cancer. Irish journal of medical science 2017, 186, 57–62, doi:10.1007/s11845-016-1464-0.

70. Candido, J.; Hagemann, T. Cancer-related inflammation. Journal of clinical immunology 2013, 33 Suppl 1, S79-84, doi:10.1007/s10875-012-9847-0.

71. Zhong, R.; Chen, D.; Cao, S.; Li, J.; Han, B.; Zhong, H. Immune cell infiltration features and related marker genes in lung cancer based on single-cell RNA-seq. Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico 2021, 23, 405–417, doi:10.1007/s12094-020-02435-2.

72. Virk, H.; Arthur, G.; Bradding, P. Mast cells and their activation in lung disease. Translational research : the journal of laboratory and clinical medicine 2016, 174, 60–76, doi:10.1016/j.trsl.2016.01.005.

73. Xu, F.; Cui, W.Q.; Wei, Y.; Cui, J.; Qiu, J.; Hu, L.L.; Gong, W.Y.; Dong, J.C.; Liu, B.J. Astragaloside IV inhibits lung cancer progression and metastasis by modulating macrophage polarization through AMPK signaling. Journal of experimental & clinical cancer research : CR 2018, 37, 207, doi:10.1186/s13046-018-0878-0.

74. Gao, Y.; Yang, J.; Cai, Y.; Fu, S.; Zhang, N.; Fu, X.; Li, L. IFN-γ-mediated inhibition of lung cancer correlates with PD-L1 expression and is regulated by PI3K-AKT signaling. International journal of cancer 2018, 143, 931–943, doi:10.1002/ijc.31357.

75. Mandai, M.; Hamanishi, J.; Abiko, K.; Matsumura, N.; Baba, T.; Konishi, I. Dual Faces of IFNγ in Cancer Progression: A Role of PD-L1 Induction in the Determination of Pro- and Antitumor Immunity. Clinical cancer research : an official journal of the American Association for Cancer Research 2016, 22, 2329–2334, doi:10.1158/1078-0432.Ccr-16-0224.

76. Banchereau, J.; Steinman, R.M. Dendritic cells and the control of immunity. Nature 1998, 392, 245–252, doi:10.1038/32588.

77. Obleukhova, I.; Kiryishina, N.; Falaleeva, S.; Lopatnikova, J.; Kurilin, V.; Kozlov, V.; Vitsin, A.; Cherkasov, A.; Kulikova, E.; Sennikov, S. Use of antigen-primed dendritic cells for inducing antitumor immune responses in vitro in patients with non-small cell lung cancer. Oncology letters 2018, 15, 1297–1306, doi:10.3892/ol.2017.7403.

78. Bailly, C. Irinotecan: 25 years of cancer treatment. Pharmacological research 2019, 148, 104398, doi:10.1016/j.phrs.2019.104398.

79. Carlsen, L.; Schorl, C.; Huntington, K.; Hernandez-Borrero, L.; Jhaveri, A.; Zhang, S.; Zhou, L.; El-Deiry, W.S. Pan-drug and drug-specific mechanisms of 5-FU, irinotecan (CPT-11), oxaliplatin, and cisplatin identified by comparison of transcriptomic and cytokine responses of colorectal cancer cells. Oncotarget 2021, 12, 2006–2021, doi:10.18632/oncotarget.28075.

80. Wood, S.L.; Pernemalm, M.; Crosbie, P.A.; Whetton, A.D. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. Cancer treatment reviews 2014, 40, 558–566, doi:10.1016/j.ctrv.2013.10.001.

81. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. Nature genetics 2013, 45, 1113–1120, doi:10.1038/ng.2764.

82. Robinson, D.R.; Wu, Y.M.; Lonigro, R.J.; Vats, P.; Cobain, E.; Everett, J.; Cao, X.; Rabban, E.; Kumar-Sinha, C.; Raymond, V.; et al. Integrative clinical genomics of metastatic cancer. Nature 2017, 548, 297–303, doi:10.1038/nature23306.

83. Liu, T.; Ortiz, J.A.; Taing, L.; Meyer, C.A.; Lee, B.; Zhang, Y.; Shin, H.; Wong, S.S.; Ma, J.; Lei, Y.; et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome biology 2011, 12, R83, doi:10.1186/gb-2011-12-8-r83.

84. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdóttir, H.; Tamayo, P.; Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. Bioinformatics (Oxford, England) 2011, 27, 1739–1740, doi:10.1093/bioinformatics/btr260.

85. Lai, F.; Gardini, A.; Zhang, A.; Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. Nature 2015, 525, 399–403, doi:10.1038/nature14906.

86. Yu, G.; Wang, L.G.; He, Q.Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics (Oxford, England) 2015, 31, 2382–2383, doi:10.1093/bioinformatics/btv145.

87. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015, 43, e47, doi:10.1093/nar/gkv007.

88. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) 2010, 26, 139–140, doi:10.1093/bioinformatics/btp616.

89. Meng, T.; Huang, R.; Zeng, Z.; Huang, Z.; Yin, H.; Jiao, C.; Yan, P.; Hu, P.; Zhu, X.; Li, Z.; et al. Identification of Prognostic and Metastatic Alternative Splicing Signatures in Kidney Renal Clear Cell Carcinoma. Frontiers in bioengineering and biotechnology 2019, 7, 270, doi:10.3389/fbioe.2019.00270.

90. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.D.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 2017, 171, 1437–+.

91. Malta, T.M.; Sokolov, A.; Gentles, A.J.; Burzykowski, T.; Poisson, L.; Weinstein, J.N.; Kamińska, B.; Huelsken, J.; Omberg, L.; Gevaert, O.; et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 2018, 173, 338-354.e315, doi:10.1016/j.cell.2018.03.034.

92. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science (New York, N.Y.) 2006, 313, 1929–1935, doi:10.1126/science.1132939.

93. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology 2018, 36, 411–420, doi:10.1038/nbt.4096.

94. Linderman, G.C.; Rachh, M.; Hoskins, J.G.; Steinerberger, S.; Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. Nature methods 2019, 16, 243–245, doi:10.1038/s41592-018-0308-4.

95. Aran, D.; Looney, A.P.; Liu, L.; Wu, E.; Fong, V.; Hsu, A.; Chak, S.; Naikawadi, R.P.; Wolters, P.J.; Abate, A.R.; et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature immunology 2019, 20, 163–172, doi:10.1038/s41590-018-0276-y.

96. Zhang, X.; Lan, Y.; Xu, J.; Quan, F.; Zhao, E.; Deng, C.; Luo, T.; Xu, L.; Liao, G.; Yan, M.; et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic acids research 2019, 47, D721-d728, doi:10.1093/nar/gky900.

97. Wang, Y.; Wang, R.; Zhang, S.; Song, S.; Jiang, C.; Han, G.; Wang, M.; Ajani, J.; Futreal, A.; Wang, L. iTALK: an R Package to Characterize and Illustrate Intercellular Communication. 2019, 507871, doi:10.1101/507871 %J bioRxiv. https://doi.org/10.1101/507871.

98. Tang, Z.; Li, C.; Kang, B.; Gao, G.; Li, C.; Zhang, Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research 2017, 45, W98-w102, doi:10.1093/nar/gkx247.

99. Ellrott, K.; Bailey, M.H.; Saksena, G.; Covington, K.R.; Kandoth, C.; Stewart, C.; Hess, J.; Ma, S.; Chiotti, K.E.; McLellan, M.; et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell systems 2018, 6, 271-281.e277, doi:10.1016/j.cels.2018.03.002.

100. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. Science (New York, N.Y.) 2015, 347, 1260419, doi:10.1126/science.1260419.

101. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 2013, 6, pl1, doi:10.1126/scisignal.2004088.

102. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery 2012, 2, 401–404, doi:10.1158/2159-8290.Cd-12-0095.

103. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia (New York, N.Y.) 2017, 19, 649–658, doi:10.1016/j.neo.2017.05.002.

104. Anaya, J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. PeerJ Computer Science 2016, 2, e67.

105. Papatheodorou, I.; Moreno, P.; Manning, J.; Fuentes, A.M.; George, N.; Fexova, S.; Fonseca, N.A.; Füllgrabe, A.; Green, M.; Huang, N.; et al. Expression Atlas update: from tissues to single cells. Nucleic acids research 2020, 48, D77-d83, doi:10.1093/nar/gkz947.

106. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current protocols in molecular biology 2015, 109, 21.29.21–21.29.29, doi:10.1002/0471142727.mb2129s109.

107. Mei, S.; Qin, Q.; Wu, Q.; Sun, H.; Zheng, R.; Zang, C.; Zhu, M.; Wu, J.; Shi, X.; Taing, L.; et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic acids research 2017, 45, D658-d662, doi:10.1093/nar/gkw983.

108. Park, P.J. ChIP-seq: advantages and challenges of a maturing technology. Nature reviews. Genetics 2009, 10, 669–680, doi:10.1038/nrg2641.

109. Yamane, A.; Robbiani, D.F.; Resch, W.; Bothmer, A.; Nakahashi, H.; Oliveira, T.; Rommel, P.C.; Brown, E.J.; Nussenzweig, A.; Nussenzweig, M.C.; et al. RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. Cell reports 2013, 3, 138–147, doi:10.1016/j.celrep.2012.12.006.

110. Periyasamy, M.; Patel, H.; Lai, C.F.; Nguyen, V.T.M.; Nevedomskaya, E.; Harrod, A.; Russell, R.; Remenyi, J.; Ochocka, A.M.; Thomas, R.S.; et al. APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. Cell reports 2015, 13, 108–121, doi:10.1016/j.celrep.2015.08.066.

111. Capell, B.C.; Drake, A.M.; Zhu, J.; Shah, P.P.; Dou, Z.; Dorsey, J.; Simola, D.F.; Donahue, G.; Sammons, M.; Rai, T.S.; et al. MLL1 is essential for the senescence-associated secretory phenotype. Genes & development 2016, 30, 321–336, doi:10.1101/gad.271882.115.

112. Sastre-Perona, A.; Hoang-Phou, S.; Leitner, M.C.; Okuniewska, M.; Meehan, S.; Schober, M. De Novo PITX1 Expression Controls Bi-Stable Transcriptional Circuits to Govern Self-Renewal and Differentiation in Squamous Cell Carcinoma. Cell stem cell 2019, 24, 390-404.e398, doi:10.1016/j.stem.2019.01.003.

113. Saladi, S.V.; Ross, K.; Karaayvaz, M.; Tata, P.R.; Mou, H.; Rajagopal, J.; Ramaswamy, S.; Ellisen, L.W. ACTL6A Is Co-Amplified with p63 in Squamous Cell Carcinoma to Drive YAP Activation, Regenerative Proliferation, and Poor Prognosis. Cancer cell 2017, 31, 35–49, doi:10.1016/j.ccell.2016.12.001.

114. Pope, B.D.; Ryba, T.; Dileep, V.; Yue, F.; Wu, W.; Denas, O.; Vera, D.L.; Wang, Y.; Hansen, R.S.; Canfield, T.K.; et al. Topologically associating domains are stable units of replication-timing regulation. Nature 2014, 515, 402–405, doi:10.1038/nature13986.

115. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012, 489, 57–74, doi:10.1038/nature11247.

116. Morrow, J.J.; Bayles, I.; Funnell, A.P.W.; Miller, T.E.; Saiakhova, A.; Lizardo, M.M.; Bartels, C.F.; Kapteijn, M.Y.; Hung, S.; Mendoza, A.; et al. Positively selected enhancer elements endow osteosarcoma cells with metastatic competence. Nature medicine 2018, 24, 176–185, doi:10.1038/nm.4475.

117. Watanabe, H.; Ma, Q.; Peng, S.; Adelmant, G.; Swain, D.; Song, W.; Fox, C.; Francis, J.M.; Pedamallu, C.S.; DeLuca, D.S.; et al. SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. The Journal of clinical investigation 2014, 124, 1636–1645, doi:10.1172/jci71545.