

Comprehensive Analysis of Prognostic Value and Immune Infiltration of ATX-associated Genes in Non-small Cell Lung Cancer

Xiaoxia Wen

Chengdu University of Traditional Chinese Medicine

Yao Xiao

Sichuan Cancer Hospital & Institute, University of Electronic Science and Technology of China

Ping Leng (✉ 596353806@qq.com)

Chengdu University of Traditional Chinese Medicine

Huaichao Luo

Sichuan Cancer Hospital & Institute, University of Electronic Science and Technology of China

Research Article

Keywords: NSCLC, bioinformatics analysis, biomarker, prognosis

Posted Date: December 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1121362/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Lung cancer is one of the most commonly diagnosed cancer and the leading cause of cancer-related death in the world. AT α -alveolar type II cells, AT α are a key structure of the distal lung epithelium and have a secretory function that is essential to maintain normal lung homeostasis. AT α cells dedifferentiate into a cell stem-like state, which can continuous differentiation, proliferation, repair and damage, and helps initiate and maintain tumor progression. However, the potential mechanistic value of AT α -associated genes as a clinical biomarker and therapeutic target of NSCLC has not been fully elucidated.

Methods: We used the Gene Expression Profile Interaction Analysis (GEPIA) and Oncomine database to explore the expression of AT α -associated genes (*AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) in NSCLC patients. Then we use the Kaplan Meier plotter and the GEPIA website to evaluate the prognosis of survival impact of differential expression of these genes. Finally, we analyzed the correlation between eight AT α -associated genes and infiltration of immune cells using the TIMER website.

Results: The expression levels of *AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC* were remarkably reduced in lung cancer tissues, and also observably related to clinical cancer stages. Low mRNA expression of *AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC* were associated with short overall survival (OS) in NSCLC patients and the low expression of *CLDN18*, *FOXA2*, *NKX2-1*, *PGC*, *SFTPB*, *SFTPC*, *SFTPD* were significantly related to a reduced progression-free survival (FP), and low *CLDN18*, *FOXA2* and *SFTPD* mRNA expression led to a short post-progression survival (PPS). Moreover, the functions of the differentially expressed eight AT α -associated genes were primarily related to lung development, regulation of epithelial to mesenchymal transition, late endosome, antibacterial humoral response. Finally, the expression of *AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC* in LUAD and LUSC patients were significantly correlated with the infiltration of diverse immune cells, including six types of CD4⁺ T cells, macrophages, neutrophils, B cells, CD8⁺ T cells, and dendritic cells.

Conclusion: Our study provided strong evidence of the values of AT α -associated genes (*AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) as clinical biomarkers and therapeutic targets in NSCLC and might provide some new inspirations to assist in the design of new immunotherapies.

Background

Lung cancer is one of the most commonly diagnosed cancer and the leading cause of cancer-related death in the world [1][2][3]. Non-small cell lung cancer (NSCLC) is one of the most main types of lung cancer approximately 85%, mainly including lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) [2][4][5][6]. Studies demonstrated that NSCLC patients with an average five-year survival rate of 15% [1][7]. This poor survival rate is attributable to many factors, including most patients were at an advanced stage at the time of diagnosis, currently available therapies are limited [8][9]. Within the last decade, with the improvement of treatment technology and equipment and the emergence of the

era of precision radiotherapy, the diagnosis and treatment of lung cancer have been improved to a certain extent[8][10][11]. However, despite advances in treatment, the overall prognosis for NSCLC has not yet improved significantly.

The alveolar cells are mainly composed of AT I alveolar type I cells, AT II and AT III alveolar type II cells, AT IV [12][13]. Among them, AT II is a key structure of the distal lung epithelium and has a secretory function that is essential to maintain normal lung homeostasis[14]. In recent years, there is currently substantial evidence showing that AT II and AT II-associated genes abnormal expression is significantly related to the occurrence and development of some diseases including cancer [15], . AT II is essential for normal lung function. One of the pathological features of the idiopathic pulmonary fibrosis (IPF) lung is the senescence of AT II [15][16][17]. AT II is also involved in the occurrence and development of COPD through the upregulated expression of many anti-or pro-inflammatory genes, including genes encoding oxygenase 2 (HO-2) and inducible nitric oxidase (iNOS)[15]. Importantly, several studies have also shown that AT II plays a crucial role in the oncogenesis of lung cancer[7][18]. Ashley detected the AT II-associated genes abnormal expression in the lung cancer cells in lung cancer tissues including aquaporin 4 (AQP4), surfactant pulmonary associated protein B (SFTPB), surfactant pulmonary associated protein C (SFTPC), surfactant pulmonary associated protein D (SFTPD), claudin 18 (CLDN18), forkhead box A2 (FOXA2), NKX homeobox-1 gene (NKX2-1) which also known as thyroid transcription factor-1 (TTF-1) and pepsinogen C (PGC) by single-cell RNA sequencing[19]. However, the potential values of these AT II cell-related genes as clinical biomarkers and therapeutic targets in NSCLC have not been fully clarified. Therefore, in the present study, we performed an in-depth and comprehensive analysis of the potential values of AT II-associated genes in mainly type lung cancer of NSCLC, including LUAD and LUSC based on multiple large bioinformatics databases. It aims to provide clinicians with additional information to assess and adjust the diagnostic methods and treatment options of NSCLC patients. Finally, the overall prognosis for NSCLC patients can be improved.

Results

Differential Expression of AT II-associated genes in Patients With NSCLC

We first explored the expression levels of 8 AT II-associated genes in lung cancer tissues and normal paracarcinoma tissues using the Oncomine database, the mRNA expression levels of *AQP4*, *CLDN18*, *FOXA2*, *NKX2-1*, *PGC* and *SFTPB*, *SFTPC*, *SFTPD* were all remarkably reduced in lung cancer tissues in multiple datasets (Figure 1). Furthermore, we used the GEPIA dataset to compare the mRNA expressions of the AT II-associated genes in both 483 LUAD and 347 normal tissues and 486 LUSC, and 338 normal tissues. Our results indicated that *AQP4*, *CLDN18*, *PGC* and *SFTPB*, *SFTPC*, *SFTPD* were low expression in LUAD tissues, and the *AQP4*, *CLDN18*, *FOXA2*, *NKX2-1*, *PGC* and *SFTPB*, *SFTPC*, *SFTPD* were lower expression in the LUSC tissues (Figure 2). We also contrast the relative expression levels of eight AT II-associated genes

in LUAD and LUSC tissues. The results reveal the highest expression of gene in LUAD and LUSC is *SFTPB* (Figure 3).

According to the above results, we considered that transcriptional expressions of *AQP4*, *CLDN18*, *FOXA2*, *NKX2-1*, *PGC*, *SFTPB*, *SFTPC*, and *SFTPD* were low-expression in patients with NSCLC.

Correlation Between mRNA Expression of Different AT₁-associated genes and Tumor Stages of NSCLC Patients

Lung cancer is divided into four stages according to the disease progression. As the condition develops, the patient's physiology and physical condition will also constantly change. Therefore, we assessed the correlation between the expression of AT₁-associated genes and the patients' pathological cancer stages of LUAD and LUSC patients by using GEPIA. We found that there is a significant correlation between the expression of all eight AT₁-associated genes and pathological stage of NSCLC: *AQP4* (P = 1.81e-06), *CLDN18* (P = 4.64e-06), *FOXA2* (P = 0.000128), *NKX2-1* (P = 0.000756), *PGC* (P = 3.08e-07), *SFTPB* (P = 3.33e-07), *SFTPC* (P = 1.4e-08) and *SFTPD* (P = 1.54e-07), and NSCLC patients who were in more advanced cancer stages were all almost inclined to express higher mRNA expression of AT₁-associated genes. (Figure 4). These data suggested that the 8 AT₁-associated might play a significant role in the tumorigenesis and progression of NSCLC.

Prognostic Features of AT₁-associated genes in Patients With Lung Cancer

To analyze the prognostic values of AT₁-associated genes in NSCLC patients, we assessed the correlation between these genes and prognosis using Kaplan-Meier plotter (Table 1). The survival of patients including overall survival (OS), progression-free survival (FP), and post-progression survival (PPS) (p < 0.05) (Figure 5). The results revealed that the decreased *AQP4* (HR = 0.74, p = 0.00024), *CLDN18* (HR = 0.76, p = 1.9e-05), *FOXA2* (HR = 0.63, p = 1.6e-12), *NKX2-1* (HR = 0.67, p = 4.9e-10), *PGC* (HR = 0.69, p = 1e-08), *SFTPB* (HR = 0.67, p = 6.3e-10), *SFTPC* (HR = 0.81, p = 0.0014) and *SFTPD* (HR = 0.66, p = 1.6e-10) mRNA levels were expressively associated with low OS. Besides, the low expression of *CLDN18* (HR = 0.72, p = 0.00091), *FOXA2* (HR = 0.68, p = 6.7e-05), *NKX2-1* (HR = 0.81, p = 0.031), *PGC* (HR = 0.7, p = 0.00024), *SFTPB* (HR = 0.82, p = 0.048), *SFTPC* (HR = 0.82, p = 0.04) and *SFTPD* (HR = 0.68, p = 6.2e-05) were significantly related to a reduced FP. Finally, low *CLDN18* (HR = 0.98, p = 0.032), *FOXA2* (HR = 0.74, p = 0.021) and *SFTPD* (HR = 0.96, p = 0.021) mRNA expression apparently led to a short PPS. However, no significant difference was found between the AT₁-associated genes and Disease-free survival (DFS) in NSCLC patients. (Table 1).

Table 1
The Relationship Between the Expression Level of ATII-associated genes and Prognosis in Patients with NSCLC

ATII-associated genes	Kaplan-Meier Plotter (Logrank p)			GEPIA (Logrank p)
	OS	FP	PPS	DFS
AQP4	0.00024	0.056	0.4	0.74
CLDN18	1.90E-05	0.00091	0.032	0.69
FOXA2	1.60E-12	6.70E-05	0.021	0.81
NKX2-1	4.90E-10	0.031	0.12	0.13
PGC	1.00E-08	0.00024	0.59	0.22
SFTP-B	6.30E-10	0.048	0.38	0.42
SFTP-C	0.0014	0.04	0.11	0.29
SFTP-D	1.60E-10	6.20E-05	0.021	0.4

Genetic Alteration, Expression and Protein/Gene Interaction Analyses of ATII-associated genes in Patients With NSCLC

Epigenetic alteration plays a vital role in early malignancies, so a comprehensive analysis of the molecular characteristics of ATII-associated genes was further performed on the LUAD and LUSC samples, respectively. We used the cBioPortal online tool to analyze the ATII-associated genes alterations for LUAD (TCGA, Pan-Cancer Atlas) and LUSC (TCGA, Pan-Cancer Atlas). As a group, two or more alterations were detected in different subtypes of NSCLC, and the 8 ATII-associated genes were varied in 273 samples out of 1053 patients with NSCLC (26%) (Figure 6A). Moreover, the mutation rates of *AQP4*, *CLDN18*, *FOXA2*, *NKX2-1*, *PGC* and *SFTP-B*, *SFTP-C*, *SFTP-D* were 3, 5, 2.4, 9, 2.8, 1.8, 5, and 1.1% of the investigated lung cancer samples, respectively (Figure 6A).

Moreover, a PPI network analysis of ATII-related genes was conducted with STRING. The results in Figure 6B exposed that the (deleted in malignant brain tumors) *DMBT1* gene which is a candidate tumor suppressor gene recently discovered in recent years was closely connected with ATII-associated genes alterations (Figure 6B). Besides, some genes that play an important role in immune response regulation, blood cell proliferation, defense mechanisms, and acute phase response genes are also significantly connected with ATII-associated genes alterations. including Microfibril-associated glycoprotein 4 (*MFAP4*), Pulmonary surfactant-associated protein A1 (*SFTPA1*) (Figure 6B). The GeneMANIA results also revealed that the functions of the differentially expressed ATII-associated genes and their associated molecules (such as, Leucine-rich repeat kinase 2 (*LRRK2*), lysosomal-associated membrane protein 3 (*LAMP3*), Cathepsin E (CTSE), ATP-binding cassette transporter A3 (*ABCA3*), forkhead box F1 (*FOXF1*),

and Napsin A (*NAPSA*) were primarily related to lung development, late endosome, aspartic-type peptidase activity (Figure 6C).

Immune Cell Infiltration of AT α -associated genes in Patients With NSCLC

Immune cell level is associated with the proliferation and progression of the cancer cell. In this study, to verify AT α -associated genes have been involved in cancer-related inflammation and the infiltration of immune cells, thus affecting the clinical outcome of NSCLC patients, we use the TIMER database to provide a comprehensive analysis of the correlation between eight AT α -associated genes and immune cell infiltration (Figure 7, Figure 8). As the result shows, the expression levels of *SFTPC* and *CLDN18* and the infiltration of neutrophils, B cells, macrophages, CD4+ T cells, dendritic cells, and CD8+ T cells have a positive correlation both in LUSC and LUAD ($P < 0.05$). In LUSC, all AT α -associated genes (including *AQP4*, *FOXA2*, *NKX2-1*, *PGC*, *SFTPB*, *SFTPD*, *CLDN18*, *SFTPC*) were positively associated with the infiltration of six immune cell types (neutrophils, B cells, macrophages, CD4+ T cells, dendritic cells, and CD8+ T cells; all $P < 0.05$) and these genes also were positively related to the infiltration of B cells in LUAD ($P < 0.05$). Besides, the expression levels of *AQP4*, *FOXA2*, *NKX2-1*, *SFTPB*, *CLDN18*, *SFTPC* had a positive relation to the infiltration of CD8+T cells ($P < 0.05$) and the expression of *NKX2-1*, *CLDN18*, *SFTPD*, *SFTPC* had a positive relation to the infiltration of CD4+T cells ($P < 0.05$).

Discussion

The occurrence of lung cancer is a multistep process. For example, LUAD has thought to progress always from atypical adenomatous hyperplasia (AAH) to adenocarcinoma in situ (AIS)[20] and before the development of LUSC, we can observe pre-invasive lesions in the airways[21]. Distinct molecular events and other malignant phenotypes make normal lung cells gain or lose some functions leading to deregulation of key genetic signals involved in cell proliferation, differentiation, apoptosis, migration, invasion[22][23]. The study showed that AT α cells can dedifferentiate into a cell stem-like state, which can continuous differentiation, proliferation, repair and damage. Therefore, AT α is suspected to be the cell of origin in oncogene-driven lung cancers and can help maintain tumor progression[19].

In recent years, these 8 AT α -associated genes have been confirmed to play key roles in growth and development, multiple diseases (including several cancers). For example, *FOXA2* has been proved that plays crucial roles during lung morphogenesis, surfactant protein production, goblet cell differentiation and mucin expression[24]. Besides, Liu experimentally found that the histone demethylase PHF8 can drive neuroendocrine prostate cancer (NEPC) development by epigenetically upregulation of *FOXA2*[25]. Thyroid transcription factor 1 (TTF-1 or *NKX2-1*) has long been known as an important development regulator of driving the brain, lungs, and thyroid, maturation and morphogenesis[26][27]. Studies have demonstrated that *NKX2-1* gene mutations related to compensated congenital hypothyroidism and unexplained respiratory distress due to lung hypoplasia in neonates[28]. *NKX2-1* amplification and overexpression also be proved that contribute to lung cancer cell proliferation rates and survival[29].

Interestingly, some researchers found an opposite phenomenon that *NKX2-1* can constrain lung adenocarcinoma in part by repressing the embryonically restricted chromatin regulator *Hmga2*[30]. Thus, the oncogenic and inhibitory function of *NKX2-1* in the same tumor type confirms its role as a bifunctional lineage factor. Aquaporins (AQPs) are water channel proteins that can capable of selectively transporting water and other small solutes across cells[31][32]. In the lung, AQPs were supposed to facilitating fluid transport in alveolar space, airway humidification, pleural fluid absorption, and submucosal gland secretion. *AQP4* is one of a member of the aquaporin family which was first discovered in 1994[32][33]. The change of *AQP4* expression is associated with many central nervous system (CNS) diseases including epilepsy, edema, stroke, and glioblastoma[34]. Besides, in breast cancer, thyroid carcinoma (undifferentiated) and stomach cancer, *AQP4* is low expression [35][36][37][38]. On the contrary, studies found that *AQP4* is a high expression in lung cancer and is involved in the invasion of lung cancer cells[39][40]. Surfactant proteins (SP) are involved in surfactant function and innate immunity in the human lung. In cystic fibrosis (CF), the genetic contribution of the surfactant protein genes, *SFTPB*, *SFTPC*, and *SFTPD* have been proved. Besides, a study has shown that major genetic mutations with childhood intermittent lung disease (ILD) also occur in surfactant genes, including *SFTPA1*, *SFTPA2*, *SFTPB*, *SFTPC*, *ABCA3* and *NKX2-1*. Finally, *CLDN18* is required for intercellular connectivity and has been reported to be involved in cell migration and metastasis, making it an oncogene in various cancer types, including pancreatic, esophageal, ovarian, and lung cancer.

In this study, we first systematically analyzed the expression of eight AT β -associated genes (*AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) in lung cancer. The expression levels of *AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC* in lung cancer tissues were lower than those in normal tissues. Additionally, we also verified that differential expression of AT β -associated genes (*AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) was observably related to clinical cancer stages in NSCLC patients. These results indicate that all of these eight AT β -associated genes function as an oncogene and might take a significant part in the tumorigenesis and progression of NSCLC. Besides, all of these eight AT β -associated genes were found to be notably related to OS in lung cancer patients. Low mRNA expression was associated with short OS in lung cancer patients. The remaining seven genes (*SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) except *AQP4*, were significantly associated with FP and lower mRNA expression was related to shorter FP while low *CLDN18*, *FOXA2* and *SFTPD* mRNA expression apparently led to a short PPS. All these results indicate AT β -associated genes might be a risk factor for survivals of NSCLC patients and could be potential prognostic biomarkers. In addition, our study showed that the expression of AT β -associated genes might be significantly correlated with and the infiltration of six immune cell types. The tumor microenvironment (TME) is complex and continuously evolving and could affect tumor progression and recurrence[41]. Immune cells are important constituents of the tumor stroma and critically take part in this process[42]. This result also suggest that AT β -associated genes may also reflect the immune status besides the disease prognosis.

We analyzed AT β -associated genes comprehensively in NSCLC based on their expression, mutation,, survival analysis, and infiltration of immune cell. Undeniably, our study also had some limitations. All of our results were acquired from public databases so that our results need to be validated. The potential

mechanisms and molecules of eight AT β -associated genes also should be further explored in the progression of NSCLC.

Conclusion

In conclusion, this work provided strong evidence of the values of AT β -associated genes (*AQP4*, *SFTPB*, *SFTPC*, *SFTPD*, *CLDN18*, *FOXA2*, *NKX2-1* and *PGC*) as clinical biomarkers and therapeutic targets in NSCLC. In the future we hope the results could make these eight AT β -associated genes were expected to become new prognostic biomarkers in NSCLC and provide some new inspirations to assist in the design of new immunotherapies.

Methods

Oncomine

Oncomine database is a publicly accessible online cancer microarray database. (www.oncomine.org), which provides a genome-wide expression analysis[43]. In this study, it was utilized to analyze the transcription levels of AT β -associated genes in NSCLC tissues and their corresponding adjacent normal control samples. A p -value < 0.05, a fold change of 2, and a gene rank in the top 10% were set as the significance thresholds. Student's t -test was used to analyze.

Gene Expression Profiling Interactive Analysis (GEPIA)

GEPIA (<http://gepia.cancer-pku.cn/index.html>) is a newly developed interactive web server for analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from the TCGA and Genotype-tissue Expression dataset[44]. GEPIA offers customizable functions such as tumor/normal differential expression analysis, profiling according to cancer types or pathological stages, patient survival analysis, similar gene detection, correlation analysis, and dimensionality reduction analysis. In this study, we performed the pathological type and stage analysis of eight AT β -associated genes using the "LUAD" and "LUSC" datasets. The Student's t -test was used to generate a p -value and the p -value cutoff was 0.01.

Kaplan-Meier Plotter

Kaplan–Meier Plotter (<https://kmplot.com/analysis/>) is a useful prognostic biomarker assessment tool that can assess the effect of 54 k genes on survival in 21 cancer types[45]. In this study, LUAD and LUSC patients were split into high and low expression groups based on median values of AT β -associated genes expression and analyze the prognostic value of the AT β -associated genes in LUAD and LUSC regarding OS (overall survival), FP (first progression), and PPS (post-progression survival). The hazard ratio with 95% confidence intervals and log-rank P value was calculated and the statically significant difference was considered when a p -value is <0.05.

cBioPortal

cBioPortal (www.cbioportal.org) is a comprehensive web resource that could visualize and analyze multidimensional cancer genomics data[46][47]. In this study, we analyze the AT α -associated genes alterations for LUAD (TCGA, PanCancer Atlas) and LUSC (TCGA, PanCancer Atlas), which contained mutations, Structural variants copy-number alterations.

STRING

STRING (<https://string-db.org/>) is a database of known and predicted protein–protein interactions (PPI) [48]. In this study, we conducted associations among the PPI network of AT α -associated genes to explore the role of AT α -related genes co-expressed genes with STRING

GeneMANIA

GeneMANIA (<http://www.genemania.org>) is a useful website that can find information on protein-protein, protein-DNA and genetic interactions, pathways, reactions, gene and protein expression data, protein domains and phenotypic screening profiles[49]. In this study, we used it to weights that indicates the predictive value of AT α -associated genes.

Timer

Timer web server (<https://cistrome.shinyapps.io/timer/>) is a comprehensive resource for systematical analysis of the infiltration of different immune cells and their clinical impact across diverse cancer types[50]. In this study, we use the “Gene module” and “Survival module” to explore the correlation of eight AT α -associated gene levels and the immune cell infiltration, the clinical outcome in LUAD and LUSC, respectively.

Abbreviations

AT α : alveolar type II cells

NSCLC: Non-small cell lung cancer

LUAD: lung adenocarcinoma

LUSC: lung squamous cell carcinoma

AT I: alveolar type I cells

IPF: idiopathic pulmonary fibrosis

HO-2: oxygenase 2

iNOS: nducible nitric oxidase

AQP4: aquaporin 4

SFTPB: surfactant pulmonary associated protein B

SFTPC: surfactant pulmonary associated protein C

SFTPD: surfactant pulmonary associated protein D

CLDN18: claudin 18

FOXA2: forkhead box A2

NKX2-1: NKX homeobox-1

TTF-1: thyroid transcription factor-1

PGC: pepsinogen C

OS: overall survival,

FP: progression-free survival

DFS: disease-free survival

PPS: post-progression survival

MFAP4: microfibril-associated glycoprotein 4

SFTPA1: pulmonary surfactant-associated protein A1

LRRK2: leucine-rich repeat kinase 2

LAMP3: lysosomal-associated membrane protein 3

CTSE: cathepsin E

ABCA3: ATP-binding cassette transporter A3

FOXF1: forkhead box F1,

NAPSA: napsin A

AIS: adenocarcinoma in situ

NEPC: neuroendocrine prostate cancer

AQPs: aquaporins

CNS: central nervous system

SP: surfactant proteins

CF: cystic fibrosis

ILD: childhood intermittent lung disease

TME: the tumor microenvironment

Declarations

Ethics approval and con sent to participate: Not applicable.

Availability of data and materials: Not applicable.

Author Contributions: Huaichao Luo, Ping Leng P conceived of the studies. Xiaoxia Wen and Yao Xiao complete statistical analysis of data, Xiaoxia Wen prepared and wrote the manuscript. All authors read and approved the final manuscript.

Funding: There no funding has been involved, with no further customization needed.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, et al. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics*. 2019;9:2056–70.
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553:446–54.
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941–53.
4. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the Notion of Non-Small Cell Lung Cancer. *Trends in Molecular Medicine*. 2019;25:585–94.

5. Roointan A, Ahmad Mir T, Ibrahim Wani S, Mati-ur-Rehman, Hussain KK, Ahmed B, et al. Early detection of lung cancer biomarkers through biosensor technology: A review. *Journal of Pharmaceutical and Biomedical Analysis*. 2019;164:93–103.
6. Didkowska J, Wojciechowska U, Mańczuk M, Łobaszewski J. Lung cancer epidemiology: contemporary and future challenges worldwide. *Ann Transl Med*. 2016;4:150–150.
7. Lin C, Song H, Huang C, Yao E, Gacayan R, Xu S-M, et al. Alveolar Type II Cells Possess the Capability of Initiating Lung Tumor Development. *Morrissey E, editor. PLoS ONE*. 2012;7:e53817.
8. Sears CR, Mazzone PJ. Biomarkers in Lung Cancer. *Clinics in Chest Medicine*. 2020;41:115–27.
9. Clausen MM, Langer SW. Improving the prognosis for lung cancer patients. *Acta Oncologica*. 2019;58:1077–8.
10. Inage T, Nakajima T, Yoshino I, Yasufuku K. Early Lung Cancer Detection. *Clinics in Chest Medicine*. 2018;39:45–55.
11. Toumazis I, Bastani M, Han SS, Plevritis SK. Risk-Based lung cancer screening: A systematic review. *Lung Cancer*. 2020;147:154–86.
12. Aspal M, Zemans RL. Mechanisms of ATII-to-ATI Cell Differentiation during Lung Regeneration. *IJMS*. 2020;21:3188.
13. Mason RJ. Biology of alveolar type II cells. *Respirology*. 2006;11:S12–5.
14. Zhao L, Yee M, O'Reilly MA. Transdifferentiation of alveolar epithelial type II to type I cells is controlled by opposing TGF- β and BMP signaling. *American Journal of Physiology-Lung Cellular and Molecular Physiology*. 2013;305:L409–18.
15. Ruaro B, Salton F, Braga L, Wade B, Confalonieri P, Volpe MC, et al. The History and Mystery of Alveolar Epithelial Type II Cells: Focus on Their Physiologic and Pathologic Role in Lung. *IJMS*. 2021;22:2566.
16. Rana T, Jiang C, Liu G, Miyata T, Antony V, Thannickal VJ, et al. PAI-1 Regulation of TGF- β 1-induced Alveolar Type II Cell Senescence, SASP Secretion, and SASP-mediated Activation of Alveolar Macrophages. *Am J Respir Cell Mol Biol*. 2020;62:319–30.
17. Selman M, Pardo A. Revealing the Pathogenic and Aging-related Mechanisms of the Enigmatic Idiopathic Pulmonary Fibrosis. An Integral Model. *Am J Respir Crit Care Med*. 2014;189:1161–72.
18. Shaurova T, Zhang L, Goodrich DW, Hershberger PA. Understanding Lineage Plasticity as a Path to Targeted Therapy Failure in EGFR-Mutant Non-small Cell Lung Cancer. *Front Genet*. 2020;11:281.
19. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell*. 2020;182:1232-1251.e22.
20. Inamura K. Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *Int J Mol Sci*. 2018;19:E1259.
21. Pennycuik A. Immune surveillance in clinical regression of pre-invasive squamous cell lung cancer. :31.

22. Marino FZ, Bianco R, Accardo M, Ronchi A, Cozzolino I, Morgillo F, et al. Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications. *Int J Med Sci.* 2019;16:981–9.
23. de Sousa VML, Carvalho L. Heterogeneity in Lung Cancer. *Pathobiology.* 2018;85:96–107.
24. Choi W, Choe S, Lau GW. Inactivation of *FOXA2* by Respiratory Bacterial Pathogens and Dysregulation of Pulmonary Mucus Homeostasis. *Front Immunol.* 2020;11:515.
25. Liu Q, Pang J, Wang L, Huang Z, Xu J, Yang X, et al. Histone demethylase PHF8 drives neuroendocrine prostate cancer progression by epigenetically upregulating *FOXA2*. *J Pathol.* 2021;253:106–18.
26. Termsarasab P. Chorea: CONTINUUM: Lifelong Learning in Neurology. 2019;25:1001–35.
27. Peall KJ, Lumsden D, Kneen R, Madhu R, Peake D, Gibbon F, et al. Benign hereditary chorea related to *NKX2.1*: expansion of the genotypic and phenotypic spectrum. *Dev Med Child Neurol.* 2014;56:642–8.
28. Kostopoulou E. Genetics of primary congenital hypothyroidism—a review. 2021;12.
29. Kwei KA, Kim YH, Girard L, Kao J, Pacyna-Gengelbach M, Salari K, et al. Genomic profiling identifies *TITF1* as a lineage-specific oncogene amplified in lung cancer. *Oncogene.* 2008;27:3635–40.
30. Winslow MM. Suppression of lung adenocarcinoma progression by *NKX2-1*. :7.
31. IJMS Editorial Office. Erratum: Mangiatordi, G.F., et al. Human Aquaporin-4 and Molecular Modeling: Historical Perspective and View to the Future. *Int. J. Mol. Sci.* 2016, 17, 1119. *IJMS.* 2016;17:1720.
32. He D, Zhang A, Li Y, Cai G, Li Y, Guo S. Autoimmune aquaporin-4 induced damage beyond the central nervous system. *Multiple Sclerosis and Related Disorders.* 2017;18:41–6.
33. Hasegawa H, Ma T, Skach W, Matthay MA, Verkman AS. Molecular cloning of a mercurial-insensitive water channel expressed in selected water-transporting tissues. *J Biol Chem.* 1994;269:5497–500.
34. Vandebroek A, Yasui M. Regulation of *AQP4* in the Central Nervous System. *IJMS.* 2020;21:1603.
35. Papadopoulos MC, Saadoun S. Key roles of aquaporins in tumor biology. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 2015;1848:2576–83.
36. Niu D, Kondo T, Nakazawa T, Kawasaki T, Yamane T, Mochizuki K, et al. Differential Expression of Aquaporins and Its Diagnostic Utility in Thyroid Cancer. Fusco A, editor. *PLoS ONE.* 2012;7:e40770.
37. Shi Z, Zhang T, Luo L, Zhao H, Cheng J, Xiang J, et al. Aquaporins in human breast cancer: identification and involvement in carcinogenesis of breast cancer. *J Surg Oncol.* 2012;106:267–72.
38. Xu H, Zhang Y, Wei W, Shen L, Wu W. Differential expression of aquaporin-4 in human gastric normal and cancer tissues. *Gastroenterol Clin Biol.* 2009;33:72–6.
39. Song Y, Wang L, Wang J, Bai C. Aquaporins in Respiratory System. In: Yang B, editor. *Aquaporins [Internet].* Dordrecht: Springer Netherlands; 2017 [cited 2021 Oct 10]. p. 115–22. Available from: http://link.springer.com/10.1007/978-94-024-1057-0_7
40. Xie Y, Wen X, Jiang Z, Fu HQ, Han H, Dai L. Aquaporin 1 and aquaporin 4 are involved in invasion of lung cancer cells. *Clin Lab.* 2012;58:75–80.

41. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 2019;79:4557–66.
42. Lei X, Lei Y, Li J-K, Du W-X, Li R-G, Yang J, et al. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Letters.* 2020;470:126–33.
43. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia.* 2004;6:1–6.
44. 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:W98–102.
45. Nagy Á, Lánckzy A, Menyhárt O, Gyórfy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep.* 2018;8:9227.
46. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery.* 2012;2:401–4.
47. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling.* 2013;6:pl1–pl1.
48. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research.* 2019;47:D607–13.
49. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research.* 2010;38:W214–20.
50. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017;77:e108–10.

Figures

Analysis Type by Cancer	Cancer vs. Normal													
	AQP4	CLDN18	FOXA2	NKX2-1	PGC	SFTPB	SFTPC	SFTPD						
Bladder Cancer														
Brain and CNS Cancer						1	1	1						
Breast Cancer				3										
Cervical Cancer								2						
Colorectal Cancer			6	1	1	2								
Esophageal Cancer		3	5		3		1							
Gastric Cancer	7		3	1		4		1						
Head and Neck Cancer						1								
Kidney Cancer				4	2	3	1	1						
Leukemia						1	1							
Liver Cancer	1													
Lung Cancer	10	14	4	5	11	1	9	12						
Lymphoma				1			1							
Melanoma							1							
Myeloma														
Other Cancer			1			1								
Ovarian Cancer								1						
Pancreatic Cancer		1					1							
Prostate Cancer	2			1										
Sarcoma	1		1	1		1	1							
Significant Unique Analyses	21	4	18	12	2	9	4	16	6	17	1	17	1	15
Total Unique Analyses	400	406	397	444	411	425	439	401						



Figure 1

mRNA expression of AT α -associated genes in different cancer types (Oncomine). The graphic demonstrates the numbers of datasets with statistically significant alterations in the mRNA expression of the target gene: upregulated (red) and downregulated (blue). The following criteria were used: p-value: 0.05, fold change: 2, gene rank: 10%, data type: mRNA, analysis type: cancer vs. normal tissue. As shown in the green frame, transcriptional levels of AQP4, SFTPB, SFTPC, SFTPD, CLDN18, FOXA2, NKX2-1 and PGC were significantly reduced in lung cancer tissues vs. normal tissues.

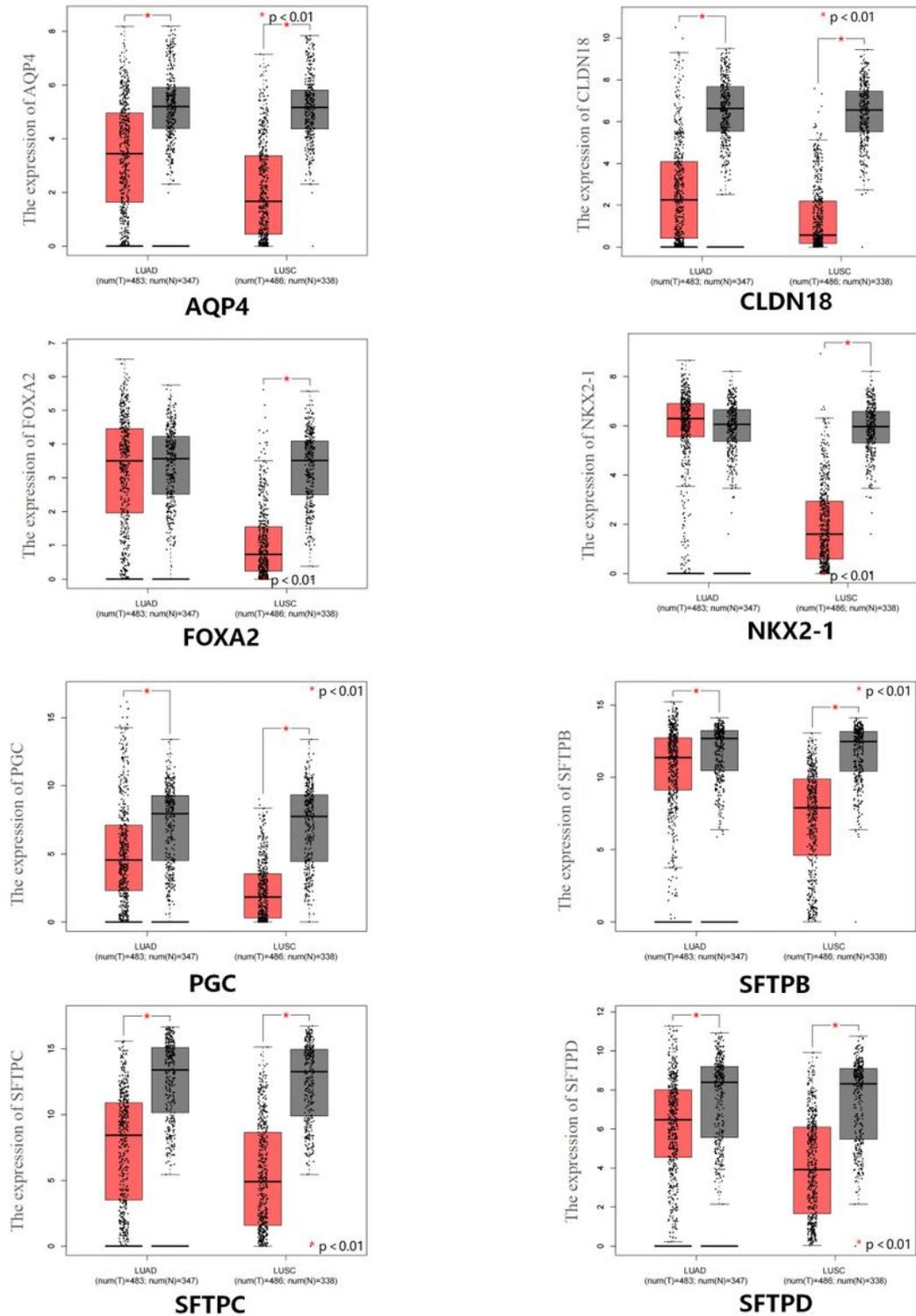


Figure 2

The mRNA expressions of the AT β -associated genes in LUAD and normal tissues, LUSC and normal tissues (GEPIA). The results indicated that AQP4, CLDN18, PGC and SFTPB, SFTPC, SFTPD were lower in LUAD tissues than in normal tissue, and AQP4, CLDN18, FOXA2, NKX2-1, PGC and SFTPB, SFTPC, SFTPD were lower in the LUSC tissues than the normal tissues. * p < 0.01

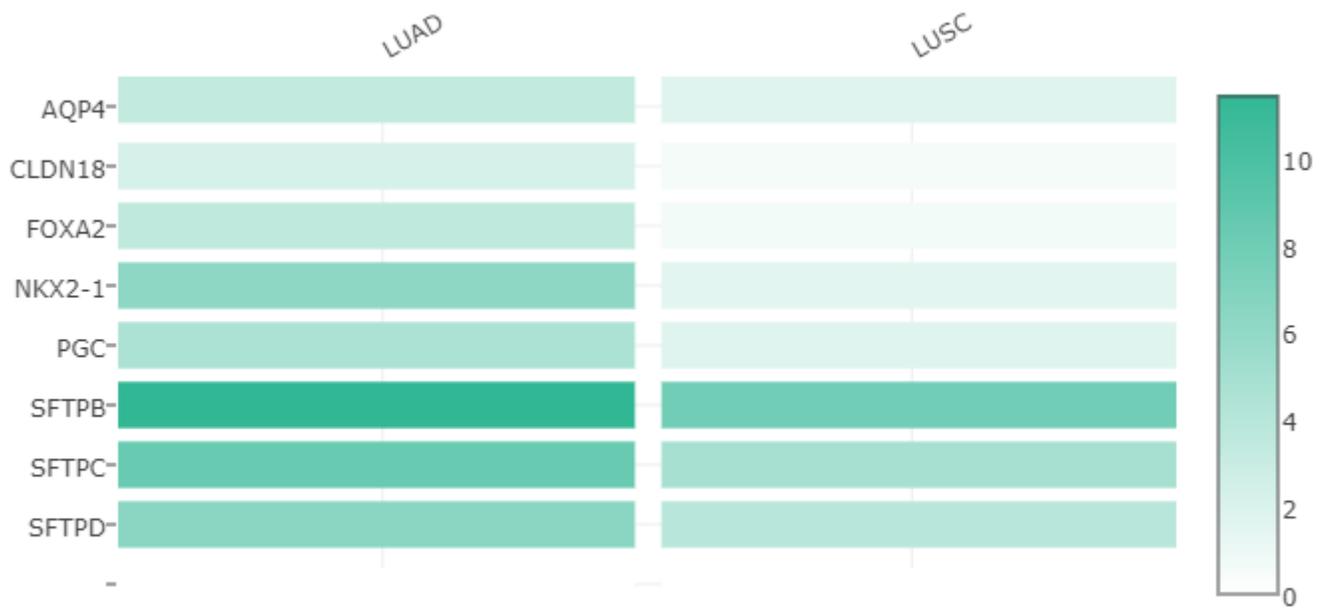


Figure 3

The relative expression level of ATX-associated genes in LUAD and LUSC (GEPIA). The darker the color of the bar, the higher the relative expression. The result evaluated that SFTPB was the highest expression in both LUAD and LUSC

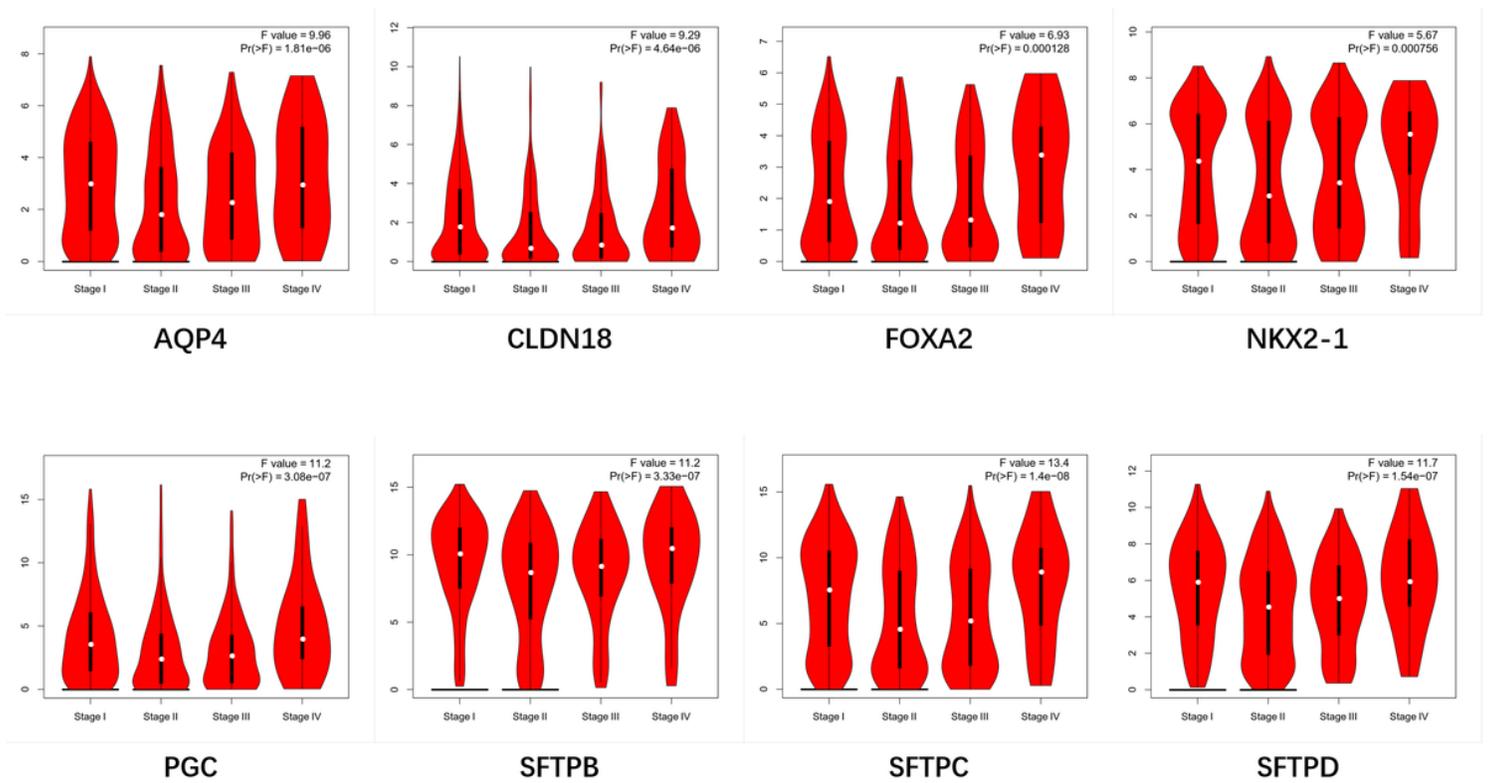


Figure 4

Correlation between AT-associated genes (AQP4, CLDN18, FOXA2, NKX2-1, PGC and SFTPB, SFTPC, SFTPD) expression and tumor stage in NSCLC patients (GEPIA). The mRNA expressions of AQP4, CLDN18, FOXA2, NKX2-1, PGC and SFTPB, SFTPC, SFTPD were distinctly related to patients' individual cancer stages.

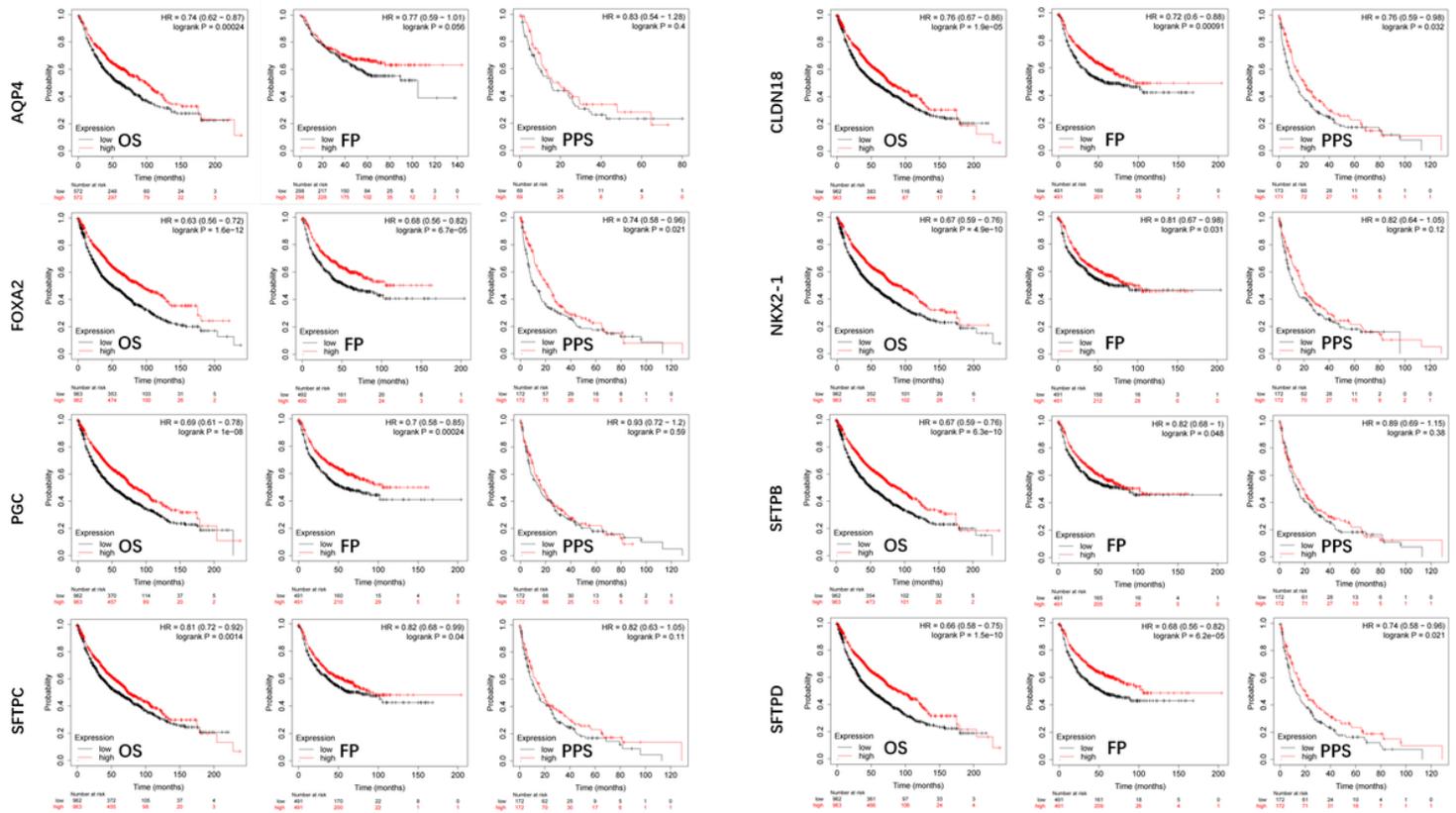


Figure 5

Prognostic value of ATX-associated genes in LUAD and LUSC (Kaplan-Meier plotter). Low mRNA expression of AQP4, SFTPB, SFTPC, SFTPD, CLDN18, FOXA2, NKX2-1 and PGC were associated with short overall survival (OS) while the low expression of CLDN18, FOXA2, NKX2-1, PGC, SFTPB, SFTPC, SFTPD were significantly related to a reduced FP, and low CLDN18, FOXA2 and SFTPD mRNA expression apparently led to a short PPS in NSCLC patients.

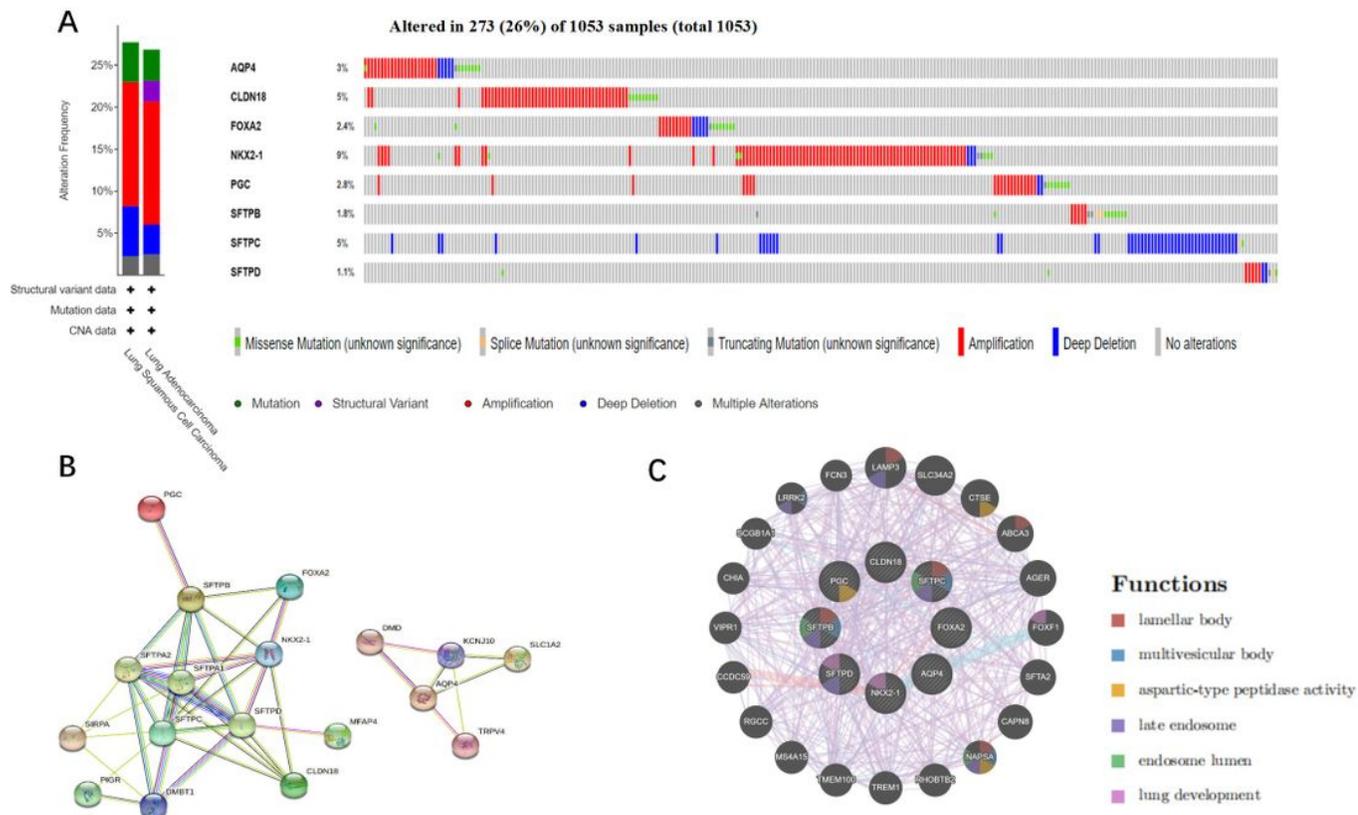


Figure 6

ATX-associated genes mutation and expression analyses in NSCLC (cBioPortal and STRING). (A) Summary of alterations in different expressed ATX-associated genes in LUAD and LUSC. (B, C) Protein-protein interaction network of different expressed ATX-associated genes.

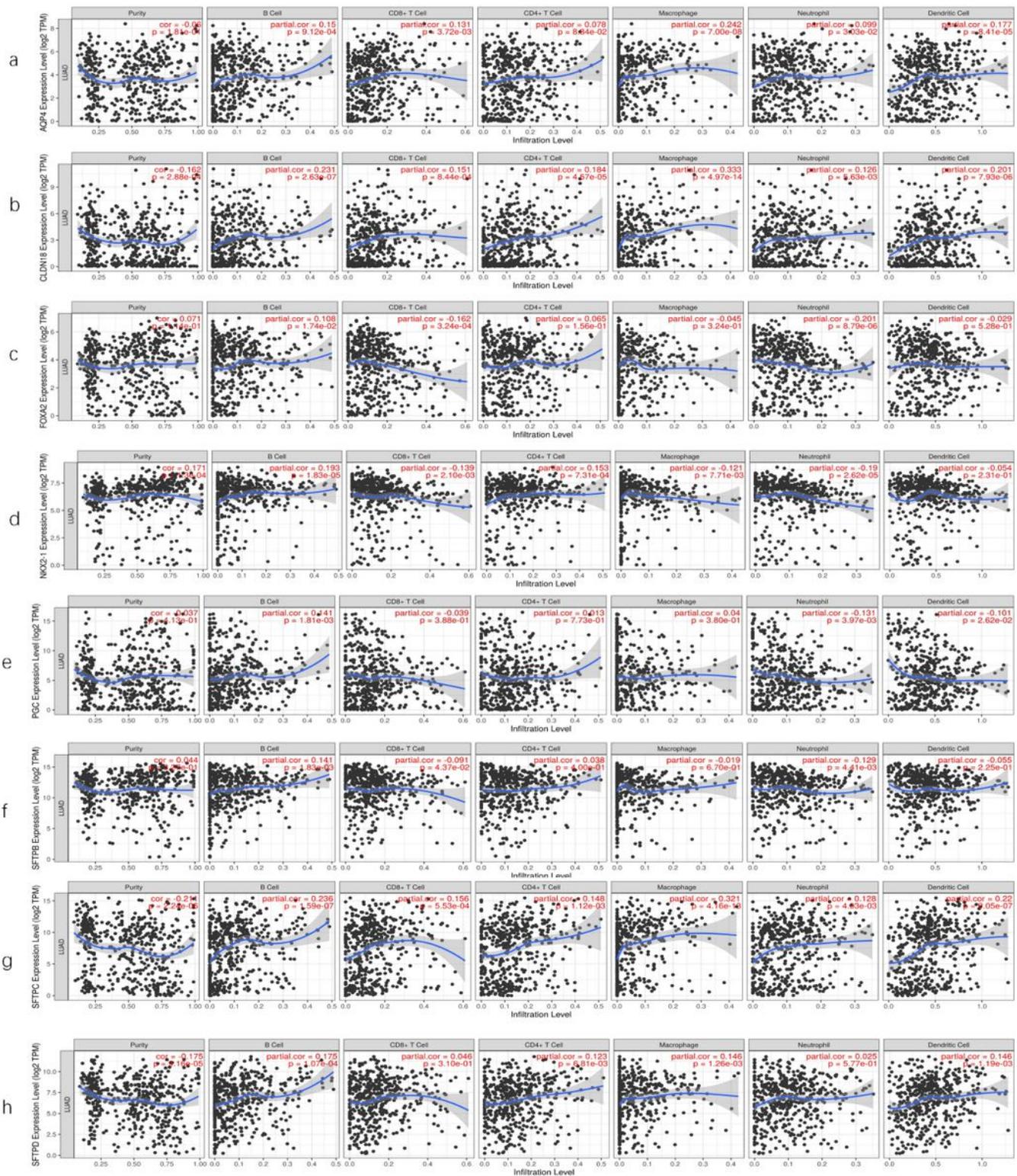


Figure 7

Correlations between differentially expressed AT-associated genes and immune cell infiltration (TIMER). Correlations between the abundance of immune cells and the expression of AQP4, CLDN18, FOXA2, NKX2-1, PGC and SFTPB, SFTPC, SFTPD in LUAD

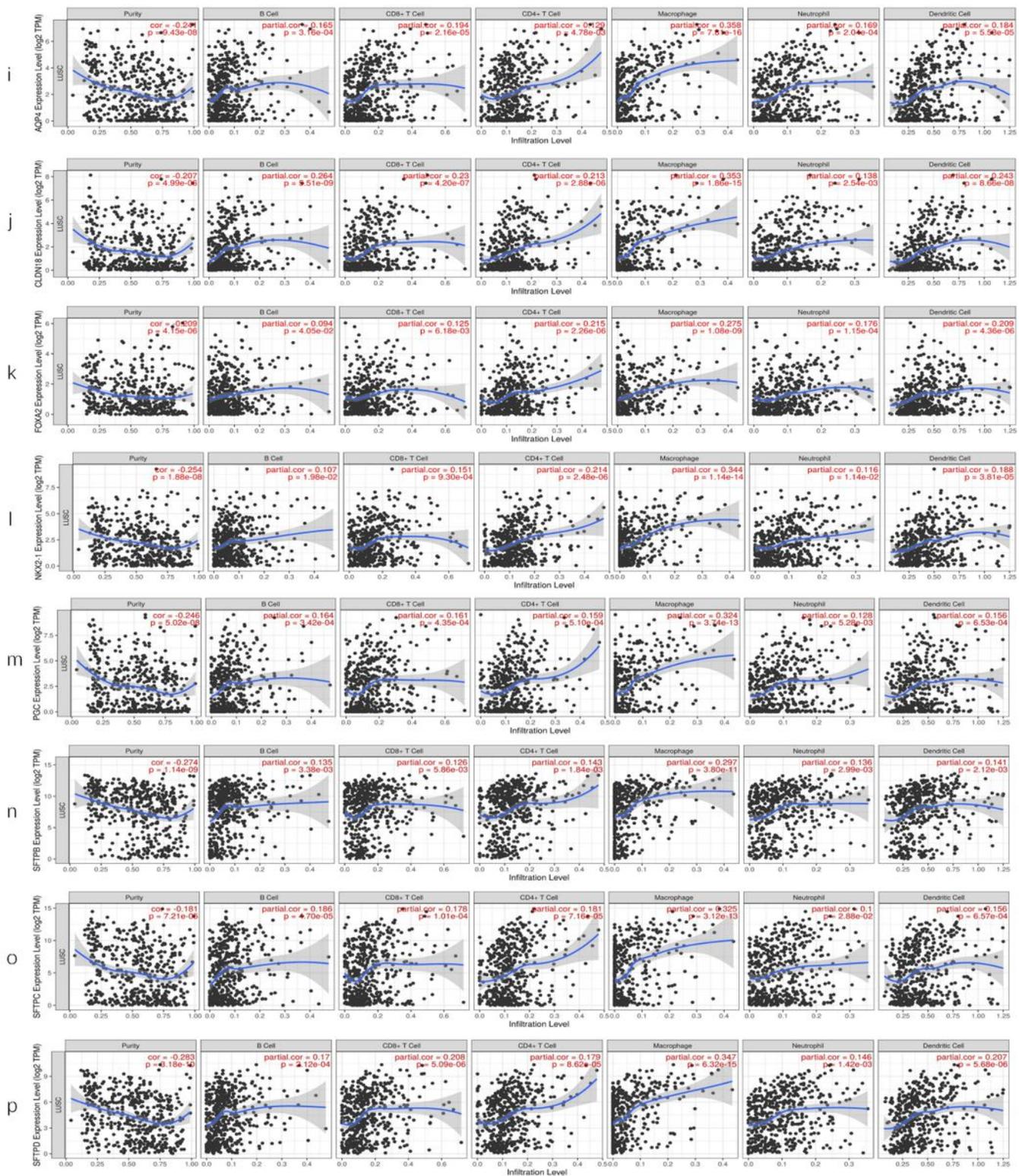


Figure 8

Correlations between differentially expressed ATX-associated genes and immune cell infiltration (TIMER). Correlations between the abundance of immune cells and the expression of AQP4, CLDN18, FOXA2, NKX2-1, PGC and SFTPB, SFTPC, SFTPD in LUSC