Journal of Translational Medicine

RESEARCH

Open Access

Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes

Xiaofei Wang†, Shuang Yao†, Zengtuan Xiao, Jialin Gong, Zuo Liu, Baoai Han and Zhenfa Zhang*

Abstract

Background: Given that abnormal autophagy is involved in the pathogenesis of cancers, we sought to explore the potential value of autophagy-associated genes in lung adenocarcinoma (LUAD).

Methods: RNA sequencing and clinical data on tumour and normal samples were acquired from The Cancer Genome Atlas (TCGA) database and randomly assigned to training and testing groups. Differentially expressed autophagy-associated genes (AAGs) were screened. Within the training group, Cox regression and Lasso regression analyses were conducted to screen five prognostic AAGs, which were used to develop a model. Kaplan–Meier (KM) and receiver operating characteristic (ROC) curves were plotted to determine the performance of the model in both groups. Immunohistochemistry was used to demonstrate the differential expression of AAGs in tumour and normal tissues at the protein level. Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were utilized to further elucidate the roles of AAGs in LUAD.

Results: The data from the TCGA database included 497 tumour and 54 normal samples, within which 30 differentially expressed AAGs were screened. Using Cox regression and Lasso regression analyses for the training group, 5 prognostic AAGs were identified and the prognostic model was constructed. Patients with low risk had better overall survival (OS) in the training group (3-year OS, 73.0% vs 48.0%; 5-year OS, 45.0% vs 33.8%; P = 1.305E−04) and in the testing group (3-year OS, 66.8% vs 41.2%; 5-year OS, 31.7% vs 25.8%; P = 1.027E−03). The areas under the ROC curves (AUC) were significant for both the training and testing groups (3-year AUC, 0.810 vs 0.894; 5-year AUC, 0.792 vs 0.749).

Conclusions: We developed a survival model for LUAD and validated the performance of the model, which may provide superior outcomes for the patients.

Keywords: LUAD, Autophagy-associated genes, Survival, The Cancer Genome Atlas

Background

Lung cancer has the highest morbidity and mortality rates worldwide and is therefore a constant threat to human life [1]. Lung adenocarcinoma (LUAD) is a prevalent pathological subtype of lung cancer, accounting for nearly 45% of lung cancer. Despite advances in the global medical industry and changes in health awareness, the outcomes of patients with lung cancer remain poor, in part because almost 80% of the patients are at an advanced stage when diagnosed; another reason may be that the current TNM (tumour size/lymph nodes/distant metastasis) staging system is not always accurate for postoperative tumour staging, and therefore necessary adjuvant treatments may not be applied [2, 3]. Therefore, it is necessary to explore alternative methods for diagnosis and accurate postoperative tumour staging.

*Correspondence: zhangzhenfa@tmu.edu.cn
†Xiaofei Wang and Shuang Yao contributed equally to this work
Department of Lung Cancer Surgery, Tianjin Medical University Cancer Institute and Hospital, Huanhu West Rd, Tianjin, China

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Autophagy is considered a vital catabolic process within eukaryotic cells, allowing lysosomes to degrade damaged, senescent, or nonfunctional proteins and organelles [4, 5]. Early studies have reported that autophagy is involved in many pathophysiological processes such as immune responses, inflammation, neurodegenerative diseases, tumourigenesis and cancer progression [6, 7]. Early in 1976, JS et al. first reported that cellular autophagocytosis progressed in cervical cancer cells in the absence of serum and amino acids [8]. Later studies showed that autophagy may play a part in degrading and recycling components of nonfunctional organelles to supply the demands of tumour progression [9, 10]. Nassour et al. demonstrated that autophagy was vital for tumour suppression, and the absence of autophagy was necessary for the initiation of tumours [11]. Therefore, autophagy may not only be involved in the inhibition of cancer but may also be related to the development and advancement of tumours [12–14].

Over the last decade, scholars have performed many studies to explore the role of autophagy in LUAD. Some studies have concluded that downregulating autophagy indirectly enhances the efficacy of the LUAD suppressors [15–17]; conversely, high-level autophagy was proven to promote tumourigenesis of LUAD in other studies [18–22]. Some results have provided evidence that the upregulation of autophagy is correlated with cisplatin or docetaxel resistance in LUAD [23–25]. Wang et al. found that autophagy impacted the low-dose hyper-radiosensitivity of LUAD [26].

Given these contradictory results, we sought to explore the potential value of autophagy in LUAD by integrating the entire set of autophagy-associated genes (AAGs) and the corresponding gene expression with clinical data acquired from The Cancer Genome Atlas (TCGA) portal. First, 30 AAGs that were differentially expressed in tumour and non-tumour tissues were screened and randomly divided into training and testing groups. We then performed Cox regression and Lasso regression analyses within the training group to identify the AAGs associated with remarkable overall survival (OS) in LUAD patients, and the prognostic model was constructed. To validate the accuracy of the model, the Kaplan–Meier (KM) estimator and the receiver operating characteristic (ROC) curve were applied. In addition, we investigated the results of Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses to further elucidate the role of AAGs in LUAD.

Materials and methods

Data source and pre-processing

The entire set of 232 AAGs was collected from the human autophagy portal (http://www.autophagy.lu/index.html), which is an online database that provides a complete set of human genes related to autophagy as described in the literature. RNA sequencing and clinical data consisting of 497 LUAD and 54 non-tumour tissues was downloaded from the TCGA data portal. The ensemble gene IDs were then converted to gene symbols using the online database GENCODE (https://www.gencodegenes.org/human/releases.html), a project for referencing human genome annotation. Finally, the expression data of the AAGs were extracted.

Screening of differentially expressed AAGs in LUAD

The expression data of 232 AAGs comprising 497 LUAD and 54 non-tumour samples were processed using the mean function, and the mean expression values were normalized by log2 transformation. The 30 AAGs that were differentially expressed between the tumour and normal samples were then identified using the Wilcoxon signed-rank test in R (version 3.6.1, https://www.r-project.org/) with a threshold of $|\log(\text{fold change})| > 1$ and a false discovery rate (FDR) < 0.05. Next, we integrated the expression data of the 30 AAGs with the corresponding clinical information. Finally, the data were randomly divided into training and testing groups for subsequent validation.

The expression data of the 30 AAGs within the training group were then analysed using univariate Cox regression analysis to obtain the AAGs that were significantly related to survival ($P < 0.05$). The least absolute shrinkage and selection operator (Lasso) regression selectively enters variables into the model to obtain improved performance parameters and to control the complexity of the model through a series of parameters to avoid overfitting [27]. Therefore, we employed a Lasso regression analysis to remove highly correlated survival-related AAGs.

Construction of the prognostic model

We performed a multivariate Cox regression analysis using both forward and backward selection to identify the 5 prognostic AAGs and their coefficients, on which we constructed the prognostic model. Every LUAD patient in both training and testing groups received an individual risk score.

The calculation of the risk score based on the AAG model was conducted as follows: $Risk\ score = \sum_{i=1}^{n} \nu_i \times c_i$ (the ν_i is the expression value of gene i, c_i represents the regression coefficient of gene

i in the multivariate Cox regression analysis, and ***n*** represents the number of independent indicators).

Validating the performance of the prognostic model in training and testing groups

Based on the individual risk scores, all patients were separated into one of two groups (low/high score) by the median risk scores. The Kaplan–Meier (K-M) survival curve was plotted to evaluate the differences in overall survival between the two groups using the log-rank test to assign statistical significance. In addition, we generated receiver operating characteristic (ROC) curves to determine the accuracy of the prognostic model.

Exploration of the expression of AAGs at the protein level

The Human Protein Atlas is an interactive web-based database (https://www.proteinatlas.org) that contains the RNA and protein expression profiles of more than ninety percent of the putative protein-encoding genes and includes more than 13 million high-resolution images [28]. The immunohistochemical results of the five prognostic AAGs were explored using this database to verify their differential expression in tumour and normal tissues.

Enrichment analysis of AAGs

To explore the potential tumour-related molecular mechanisms of AAGs, GO functional annotation and KEGG pathway enrichment analyses were performed in R using the packages DOSE, Cluster Profiler, ggplot2, GO plot, etc. with both the p-value and q-value set at 0.05. The outcomes were visually illustrated in multidimensional views.

Statistical analysis

All statistical analyses and graphics were performed using the R 3.6.1 (https://www.r-project.org/) and Perl language packages. Cox regression analyses were utilized to screen the AAGs related to survival. A Lasso regression analysis was used to eliminate highly correlated AAGs and prevent overfitting of the model. The Kaplan–Meier curve was plotted to display the differences in overall survival between the two groups and the log-rank test was performed to determine the significance of the differences. The ROC curve and the corresponding area under the curve (AUC) were used to evaluate the performance of the model. Statistical significance was defined as $P < 0.05$.

Results

Differentially expressed AAGs in lung adenocarcinoma (LUAD)

We analysed the expression of 232 AAGs in 497 LUAD and 54 non-tumour tissues using the Wilcoxon signed-rank test in R, and 30 AAGs were eventually identified using the criteria of $|\log2FC| > 1$ and $FDR < 0.05$, including 12 downregulated genes (NRG3, DLC1, NLRC4, HSPB8, DAPK2, PPP1R15A, FOS, NRG1, PRKCQ, CCL2, GRID1, MAP1LC3C) and 18 upregulated genes (HSPA5, ERBB2, PARP1, P4HB, IKBKE, BNIP3, ATIC, IFNG, VMP1, ITGB4, EIF4EBP1, PTK6, GAPDH, ATG9B, ERO1A, TMEM74, CDKN2A, BIRC5) (Fig. 1). The ggpubr package in R was utilized to exhibit the

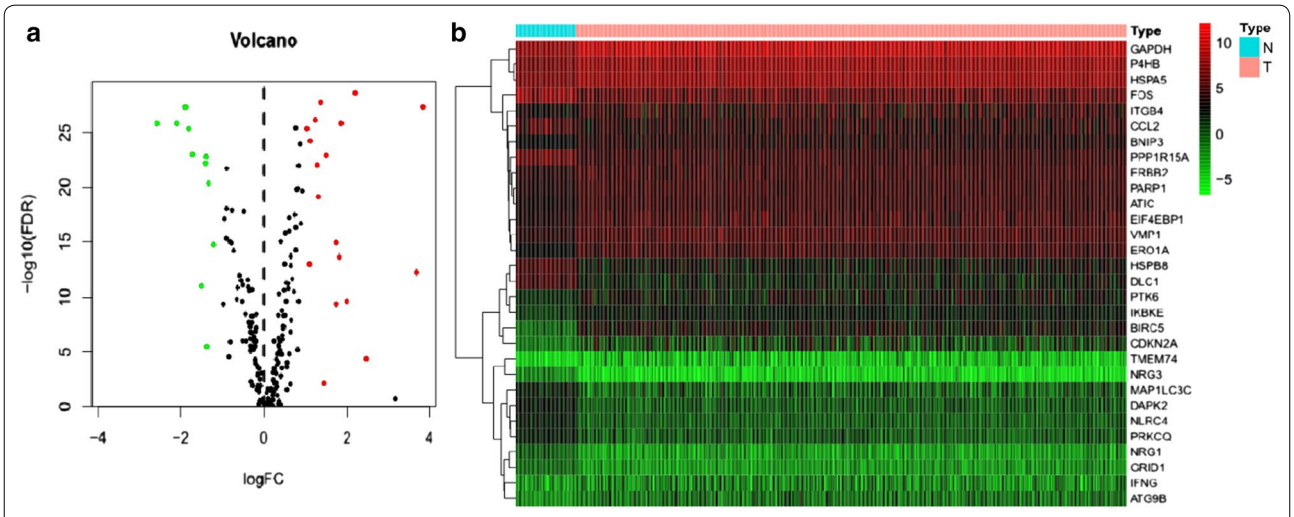


Fig. 1 Differentially expressed autophagy-associated genes (AAGs) in lung adenocarcinoma (LUAD) and non-tumour samples. **a** The volcano map of 232 AAGs. The red dots indicate genes with high expression and the green dots represent genes with low expression. **b** Hierarchical clustering distribution of differentially expressed AAGs in normal and tumour samples

expression patterns of the 30 AAGs in tumour and normal samples. The red box plots above the gene names represent tumour samples and the green box plots represent normal samples (Fig. 2).

Survival-related AAGs and the prognostic model

We conducted a univariate Cox regression analysis and identified 6 AAGs (GAPDH, ERO1A, NLRC4, ITGB4, ATG9B, and CDKN2A) that were significantly related to LUAD survival in the training group. Of the 6 survival-related AAGs, 4 genes (GAPDH, ERO1A, ITGB4, and CDKN2A) were considered risk factors (all $P<0.05$; HRs, 1.0007–1.0175) and that their overexpression may reduce survival; overexpression of the remaining two genes (NLRC4 and ATG9B) (all $P<0.05$; HRs, 0.6913 and 0.7382, respectively) may improve the survival of patients. The Lasso regression analysis was then applied to exclude genes that may be highly correlated with other genes (Fig. 3). The 6 survival-related AAGs were then submitted to a multivariate Cox proportional hazards model, resulting in 5 candidate genes (ITGB4, NLRC4, ATG9B, CDKN2A, and ERO1A) that may serve as significant predictors of the prognosis (Table 1). Based on the 5 candidate AAGs, the formula for the risk score of every LUAD patient was constructed: *risk score* = (expression value of ITGB4 * 0.0063) + (-expression value of NLRC4 * 0.354) + (-expression value of ATG9B * 0.3956) + (expression value of CDKN2A * 0.0202) + (expression value of ERO1A * 0.0122).

Validation of the model performance

To validate the accuracy of the model, we plotted the KM survival curve to evaluate the difference in LUAD survival in both the training and testing groups. In the training group, the median overall survival of low-risk patients was 4.11 years, whereas the survival of high-risk patients was 2.86 years. In comparison, in the testing group, the median overall survival of patients with low-risk scores was 3.9 years, and the survival of patients with high-risk scores was 2.34 years. Low-risk patients exhibited higher survival than high-risk patients in the training group (3-year OS, 73.0% vs 48.0%; 5-year OS, 45.0% vs 33.8%; P = 1.305E−04) and in the testing group (3-year OS, 66.8% vs 41.2%; 5-year OS, 31.7% vs 25.8%; P = 1.027E−03) (Fig. 4a, b). Furthermore, we constructed the ROC curve to assess the accuracy of the model, and the areas under the ROC curves in the training and testing groups were both significant (3-year AUC, 0.810 vs 0.894; 5-year AUC, 0.792 vs 0.749) (Fig. 4c, d).

In addition, we ranked the all of the LUAD patients by their risk scores to analyse the survival distribution. From the scatterplot, we could identify the survival status of patients with different risk scores; the mortality rate of patients rose with the increase in risk score. The heat

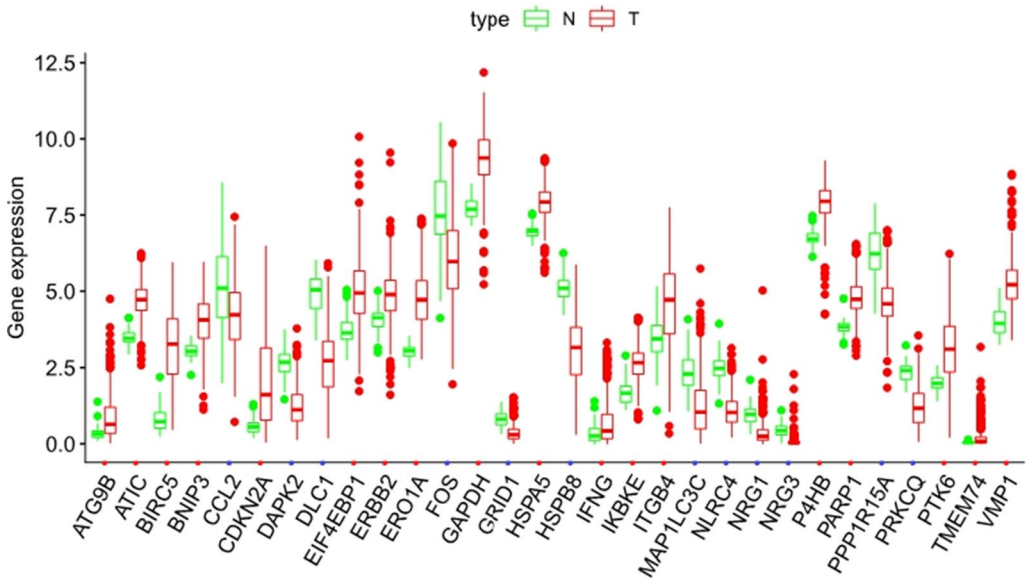


Fig. 2 Boxplots of the expression levels of 30 autophagy-associated genes (AAGs) in tumour and normal tissues. The red box plots above the corresponding gene name represent the expression in tumour samples, whereas the green box plots represent the expression in normal samples; the red dots on the X-axis indicate genes with high levels of expression and the blue dots indicate genes with low levels of expression. (Difference analysis by Wilcoxon signed-rank test and all false discovery rate (FDR) < 0.05)

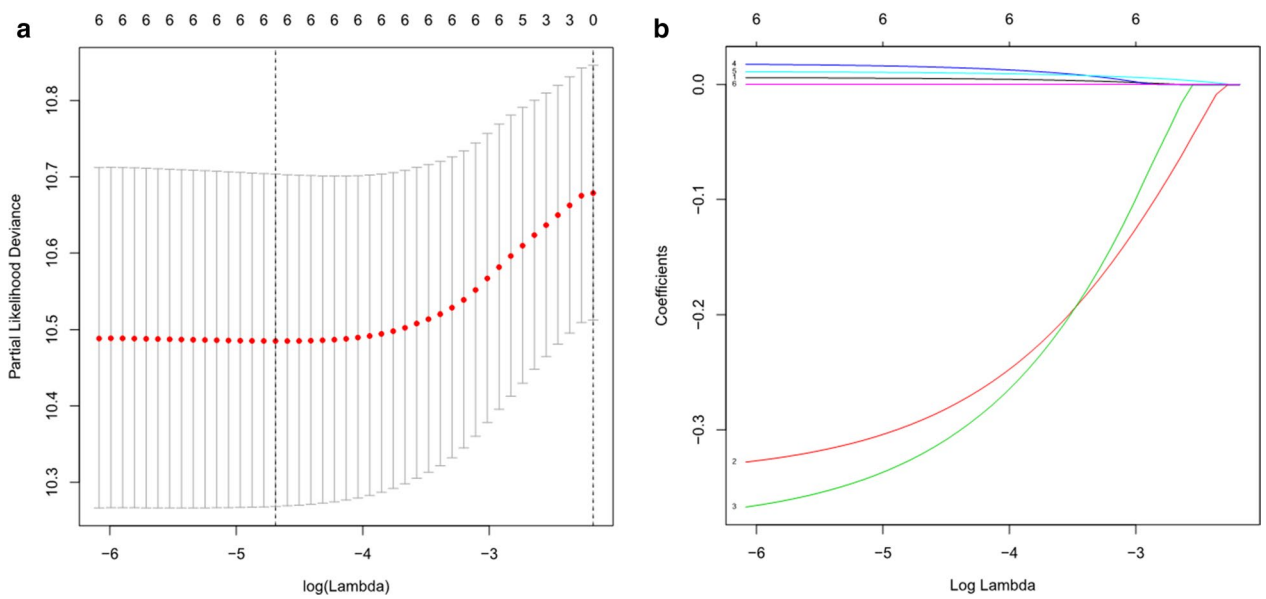


Fig. 3 Screening of the optimal AAGs used for the final construction of the predictive model using a Lasso regression. **a** Screening of optimal parameter (lambda) at which the vertical lines were drawn. **b** Lasso coefficient profiles of the six AAGs with non-zero coefficients determined by the optimal lambda

Table 1 Univariate and multivariate cox regression analyses of OS in lung adenocarcinoma patients

Genes	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P	HR (95% CI)	P	Coef
ITGB4	1.0061 (1.0005–1.0117)	0.0328	1.0064 (1.0005–1.0123)	0.0338	0.0063
NLRC4	0.6913 (0.5023–0.9513)	0.0234	0.7019 (0.5020–0.9813)	0.0384	−0.354
ATG9B	0.7382 (0.5583–0.9761)	0.0332	0.6733 (0.5007–0.9054)	0.0088	−0.3955
CDKN2A	1.0175 (1.0000–1.0353)	0.0495	1.0204 (1.0025–1.3868)	0.0252	0.0202
ERO1A	1.0133 (1.0056–1.0212)	0.0008	1.0123 (1.0040–1.0206)	0.0035	0.0122
GAPDH	1.0007 (1.0003–1.0010)	0.0004			

HR hazard ratio, *OS* overall survival, *Coef* regression coefficient of genes in the multivariate Cox regression analysis

maps illustrate the expression of AAGs with the rising risk scores of patients (Fig. 5a–f).

Differential expression of prognostic AAGs at the protein level

We used immunohistochemistry to compare the expression of the 5 prognostic AAGs (ITGB4, NLRC4, ATG9B, CDKN2A, and ERO1A) in LUAD with their expression in normal tissues (Fig. 6a–d). As expected, the levels of protein expression of the three high-risk genes (ITGB4, CDKN2A, and ERO1A) were demonstrably higher in tumour tissues with more intense antibody staining and a greater proportion of stained cells. In contrast, NLRC4, a protective gene, stained fewer cells with weaker intensity in LUAD. The results were compatible with our findings of AAGs in LUAD; there were no data for another protective gene, ATG9B, in the Human Protein Atlas database.

GO and KEGG analyses of AAGs

To evaluate the molecular mechanisms of AAGs in LUAD, GO functional annotation and KEGG pathway enrichment analyses were conducted (Table 2). GO analysis consists of three categories: biological processes (BP), cellular components (CC) and molecular function (MF). We found that the most significant GO enriched terms involved in autophagy were the intrinsic apoptotic signalling pathway, cellular response to unfolded/topologically incorrect proteins and neuron death (BP); autophagosome and endoplasmic reticulum-Golgi intermediate compartment (CC); and

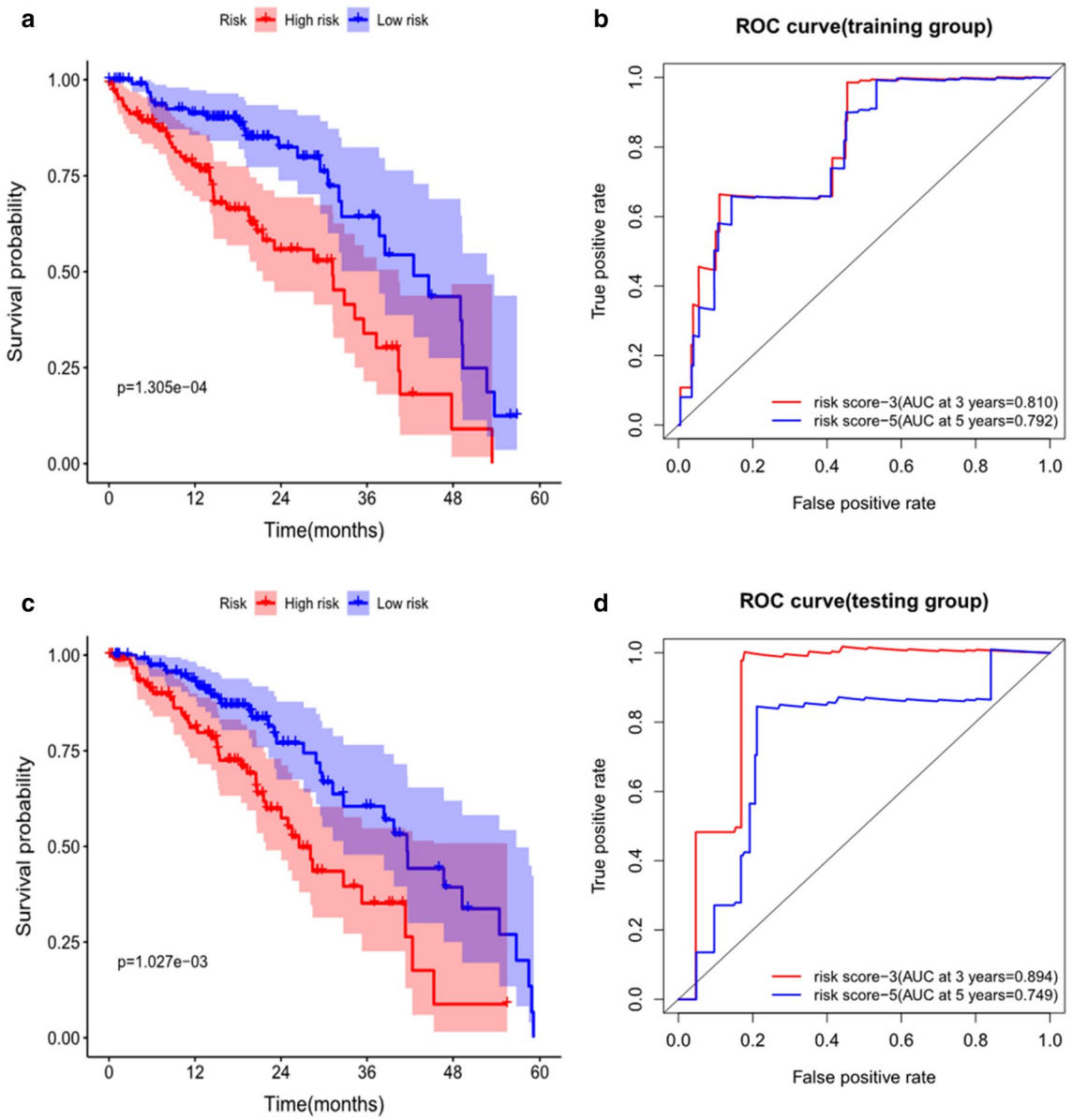


Fig. 4 **a** K-M curve of the high-risk (red) and low-risk (blue) LUAD patients in the training group. **b** The 3-year (red) and 5-year (blue) ROC curves in the training group of LUAD patients. **c** K–M curve of the high-risk (red) and low-risk (blue) LUAD patients in the testing group. **d** The 3-year (red) and 5-year (blue) ROC curves in the testing group of LUAD patients

protein phosphatase binding, chemorepellent activity, receptor activator activity, R-SMAD binding and NAD binding (MF) (Fig. 7a, b). In the KEGG enrichment analysis, the AAGs were primarily correlated with pathways related to autophagy-animal, the ErbB signalling pathway, bladder cancer, the HIF-1 signalling pathway, platinum drug resistance, proteins processed in the endoplasmic reticulum, EGFR tyrosine kinase inhibitor resistance, PD-L1 expression and the PD-1 checkpoint pathway in cancer (Fig. 8a, b).

Discussion

Adenocarcinoma is the most prevalent histological subtype of lung cancer and has been broadly explored for distinct genetic drivers and diverse prognostic factors. However, LUAD patients still experience high mortality due to undetected pathogenesis [29–31]. Many researchers believe that the existing guidelines and definitions of lung cancer may result in different clinical decisions for preoperative and postoperative patients. Zhang et al. found that dissection of the 4th lymph

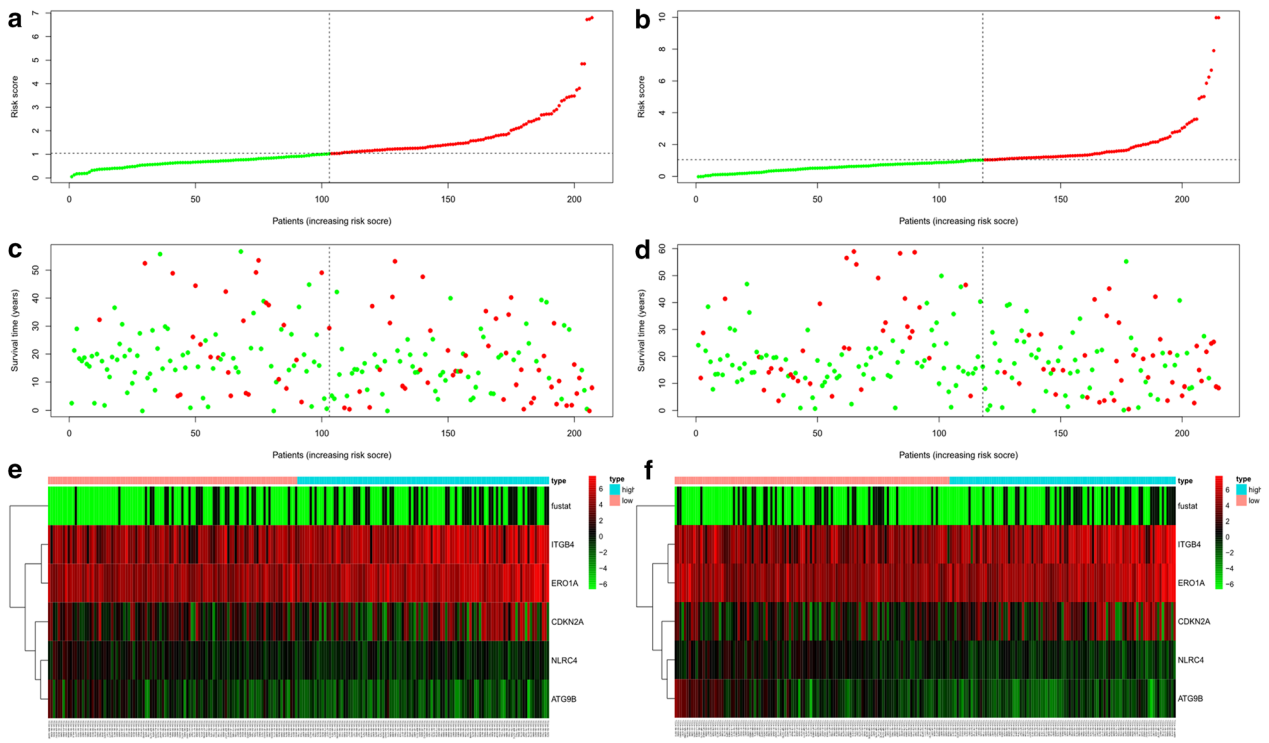


Fig. 5 **a** Risk score distribution of LUAD patients with different risks (low, green; high, red) in the *training group*. **b** Risk score distribution of LUAD patients with different risks (low, green; high, red) in the *testing group*. **c** Scatterplots of LUAD patients with different survival status in *training group*. **d** Scatterplots of LUAD patients with different survival status in *testing group*. **e** Expression of risk genes in LUAD patients with different risks (low, pink; high, blue) in the *training group*. **f** Expression of risk genes in LUAD patients with different risks (low, pink; high, blue) in the *testing group*

nodes was related to a better prognosis, although this was not recommended by the International Association for the Study of Lung Cancer (IASLC) [32]. There are also scholars who believe that the present staging guide is insufficient for predicting individual-level overall survival because many early-stage patients may experience a later relapse [33, 34]. Moreover, Valeria et al. suggested abandoning the concept of non-small cell lung cancer because a large body of experimental evidence suggests that LUAD and lung squamous cell carcinoma appear to be distinct tumours at the molecular, pathological and clinical levels [35]. Therefore, academics have placed increasing emphasis on the use of precision medicine in lung cancer [36–38]. It is necessary to explore methods to consolidate the current staging system and to improve the prognosis for lung cancer patients. Over the last decade, breakthroughs in microarrays and genome sequencing have promoted the discovery of prognostic biomarkers, which have greatly increased the accurate classification of diseases and improved individual treatment. Many studies have demonstrated that genomic data, particularly multigene signatures, demonstrate superior performance in prognosis analysis compared with the current staging system [39–42].

To our knowledge, this is the first study to combine the entire set of AAGs with LUAD and explore as well as validate the potential value of AAGs in LUAD. First, we selected 30 differentially expressed AAGs from 497 tumour samples and 54 normal samples. We then randomly divided the data into training and testing groups. Using Lasso regression and Cox survival analyses, we constructed a risk model based on five prognostic AAGs (ITGB4, NLRC4, ATG9B, CDKN2A, and ERO1A). Using the model, every LUAD patient was assigned a risk score. The differences in survival between patients with low and high scores were significant in both the training group and the testing group. The ROC curves and AUCs indicated that models of the training and testing groups both performed well. In addition, we performed immunohistochemistry that further proved the significant roles of AAGs in LUAD. Furthermore, GO and KEGG enrichment analyses of the differentially expressed AAGs were conducted to explore the underlying molecular mechanisms. The results of GO functional annotation revealed that the AAGs were primarily enriched in the intrinsic

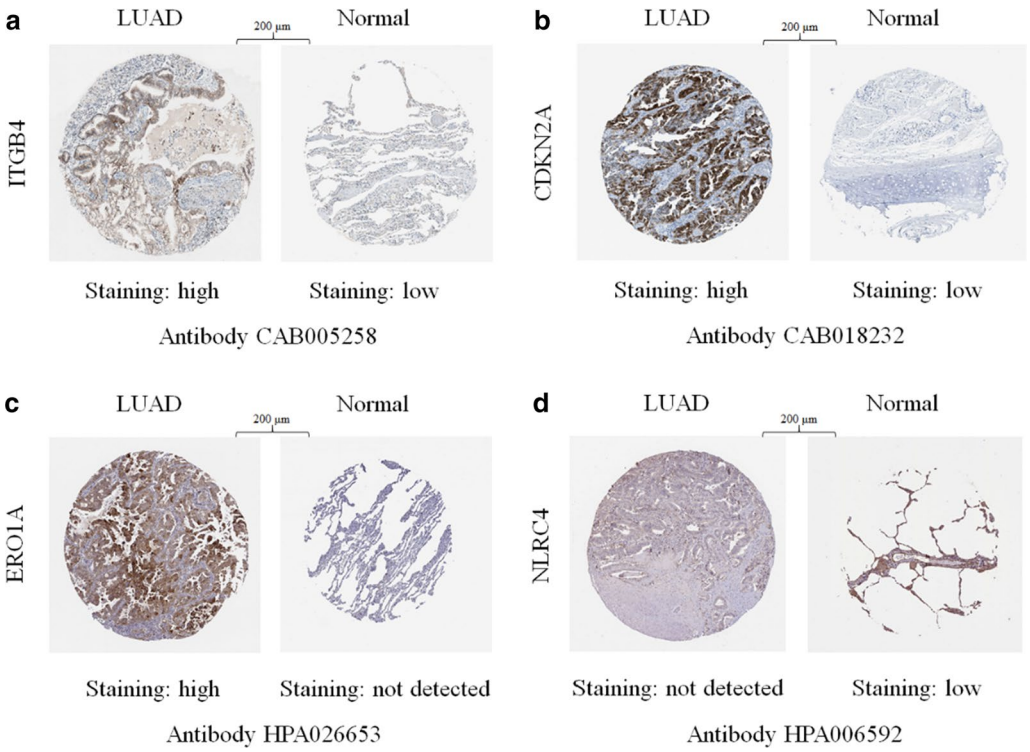


Fig. 6 Immunohistochemistry (IHC) results showing protein levels of autophagy-associated genes in LUAD and normal tissues, **a** IHC results of ITGB4 in LUAD (staining: high; intensity: strong; quantity: 75–25%; location: cytoplasmic/membranous) and in normal tissue (staining: low; intensity: weak; quantity: 75–25%; location: cytoplasmic/membranous). **b** IHC results of CDKN2A in LUAD (staining: high; intensity: strong; quantity: > 75%; location: cytoplasmic/membranous/nuclear) and in normal tissue (staining: low; intensity: moderate; quantity: < 25%; location: cytoplasmic/membranous). **c** IHC results of ERO1A in LUAD (staining: high; intensity: strong; quantity: 75–25%; location: cytoplasmic/membranous) and in normal tissue (staining: not detected; intensity: negative; quantity: none; location: N/A). **d** IHC results of NLRC4 in LUAD (staining: not detected; intensity: negative; quantity: none; location: N/A) and in normal tissue (staining: low; intensity: moderate; quantity: < 25%; location: cytoplasmic/membranous)

apoptotic signalling pathway, cellular response to topologically incorrect proteins and the PERK-mediated unfolded protein response, which is consistent with the conclusion of previous studies that autophagy is a physiological process that eliminates misfolded proteins and damaged organelles in response to cellular stress [12, 43]. In the KEGG pathway analysis, AAGs were primarily enriched in the ErbB, IL-17 and HIF-1 signalling pathways. EGFR (ErbB1) is not unknown to us; in 2004, there was a major discovery that treatment with the EGFR-TKI (epidermal growth factor receptor-tyrosine kinase inhibitor) gefitinib caused tumour regression in some patients with NSCLC [44], and the third-generation EGFR-TKI axitinib confers greater survival benefits to patients, particularly those with the T790M mutation [45]. In addition to EGFR (ErbB1), the proteins HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4) compose the ErbB family of transmembrane receptor tyrosine kinases (RTKs), which is one of the most broadly explored therapeutic targets in human malignancies [46]. Jutten et al. found that autophagy activity influenced the expression of EGFR and the resistance to EGFR-targeting therapies could be reduced by downregulating autophagy [47, 48]; IL-17 (interleukin-17), as a signature proinflammatory cytokine of the CD4+ T helper 17 (Th17) cells [49], was shown to participate in the formation and advancement of various tumours [50] and was widely distributed in the tumour microenvironment, where it has twin roles in tumourigenesis and tumour suppression [51]. Previous studies have indicated that the formation of lung cancer is closely related to local dysbiosis and inflammation mediated by Th17 cells [52, 53]. The 2019 Nobel Prize in physiology or medicine was awarded to Professor William G. Kaelin Jr., Sir Peter J. Ratcliffe and Gregg L. Semenza for their contributions to elucidating the mechanisms by which cells sense and adapt to the availability of oxygen. They found that HIF-1 (hypoxia-inducible factor-1) regulates more than 4000 targeted genes, some of which can increase oxygen transport for angiogenesis and erythropoiesis. Another study reported that under emergent

Table 2 GO and KEGG pathway enrichment analysis of AAGs in lung adenocarcinoma

Terms	Pathway ID	Pathway description	GeneID	Count	FDR
BP	GO:0050873	Brown fat cell differentiation	BNIP3/ERO1A	2	0.04082
	GO:0010660	Regulation of muscle cell apoptotic process	BNIP3/IFNG/CDKN2A	3	0.02240
	GO:0097193	Intrinsic apoptotic signaling pathway	DAPK2/PPP1R15A/PARP1/P4HB/IKBKE/BNIP3/ERO1A	7	0.00011
	GO:0001953	Negative regulation of cell–matrix adhesion	DLC1/CDKN2A	2	0.04082
	GO:0052548	Regulation of endopeptidase activity	DLC1/NLRC4/GAPDH/BIRC5	4	0.04616
	GO:0052547	Regulation of peptidase activity	DLC1/NLRC4/GAPDH/BIRC5	4	0.04990
	GO:0034599	Cellular response to oxidative stress	FOS/PARP1/P4HB/BNIP3/ERO1A	5	0.00916
	GO:0061919	Process utilizing autophagic mechanism	HSPB8/DAPK2/MAP1LC3C/BNIP3/IFNG/VMP1/ITGB4/GAPDH/ATG9B/TMEM74	10	0.00000
	GO:0016236	Macroautophagy	HSPB8/MAP1LC3C/BNIP3/VMP1/GAPDH/ATG9B/TMEM74	7	0.00011
	GO:0035967	Cellular response to topologically incorrect protein	HSPB8/PPP1R15A/CCL2/HSPA5/ERO1A	5	0.00066
	GO:0006986	Response to unfolded protein	HSPB8/PPP1R15A/CCL2/HSPA5/ERO1A	5	0.00118
	GO:0001558	Regulation of cell growth	NRG3/PRKCQ/ERBB2/CDKN2A	4	0.04988
	GO:0006575	Cellular modified amino acid metabolic process	P4HB/ATIC/ERO1A	3	0.04632
	GO:0071456	Cellulaar response to hypoxia	P4HB/BNIP3/EIF4EBP1/ERO1A	4	0.01895
	GO:0019471	4-hydroxyproline metabolic process	P4HB/ERO1A	2	0.00966
	GO:0030968	Endoplasmic reticulum unfolded protein response	PPP1R15A/CCL2/HSPA5/ERO1A	4	0.00411
	GO:0034976	Response to endoplasmic reticulum stress	PPP1R15A/CCL2/HSPA5/P4HB/ERO1A	5	0.00551
	GO:0070059	Intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress	PPP1R15A/ERO1A	2	0.04990
CC	GO:0005776	Autophagosome	DAPK2/MAP1LC3C/VMP1/ATG9B/TMEM74	5	0.00001
	GO:0000421	Autophagosome membrane	MAP1LC3C/VMP1/ATG9B/TMEM74	4	0.00001
MF	GO:0019887	Protein kinase regulator activity	NRG3/NRG1/CDKN2A	3	0.03024
KEGG Pathways	hsa04140	Autophagy-animal	VMP1/DAPK2/BNIP3/ATG9B/PRKCQ	5	0.00940
	hsa04012	ErbB signaling pathway	NRG1/ERBB2/NLRC4/EIF4EBP1	4	0.01011
	hsa04657	IL-17 signaling pathway	CCL2/IFNG/ITGB4/FOS	4	0.01011
	hsa05219	Bladder cancer	DAPK2/CDKN2A/ERBB2	3	0.01028
	hsa04066	HIF-1 signaling pathway	IFNG/ERO1A/GAPDH/EIF4EBP1	4	0.01071
	hsa05323	Rheumatoid arthritis	CCL2/ITGB4/FOS	3	0.03726
	hsa01524	Platinum drug resistance	BIRC5/CDKN2A/ERBB2	3	0.03236
	hsa04141	Protein processing in endoplasmic reticulum	ERO1A/P4HB/PPP1R15A/HSPA5	4	0.03236
	hsa01521	EGFR tyrosine kinase inhibitor resistance	NRG1/ERO1A/EIF4EBP1	3	0.03236
	hsa05132	Salmonella infection	NLRC4/IFNG/FOS	3	0.03236
	hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer	IFNG/FOS/CDKN2A	3	0.03726
	hsa01522	Endocrine resistance	CDKN2A/FOS/ERBB2	3	0.03988
	hsa05142	Chagas disease (American trypanosomiasis)	CCL2/IFNG/FOS	3	0.04085
	hsa04660	T cell receptor signaling pathway	IFNG/FOS/NLRC4	3	0.04085
	hsa04659	Th17 cell differentiation	IFNG/PRKCQ/NLRC4	3	0.04147

GO gene ontology, *KEGG* Kyoto encyclopedia of genes and genomes, *AAG* autophagy-associated genes, *FDR* false discovery rate

oxygen fluctuations, autophagy can be harmful and can lead to cell death [54]. Moreover, Bellot et al. reported that various mechanisms such as autophagy activation enabled tumour cells to adjust to hypoxia [55]. Therefore, the regulation of HIF-1 may represent an important breakthrough in tumour therapy, as angiogenesis and erythropoiesis play crucial roles in the occurrence and development of cancer.

It is apparent that autophagy plays many roles in tumourigenesis and development, which is consistent

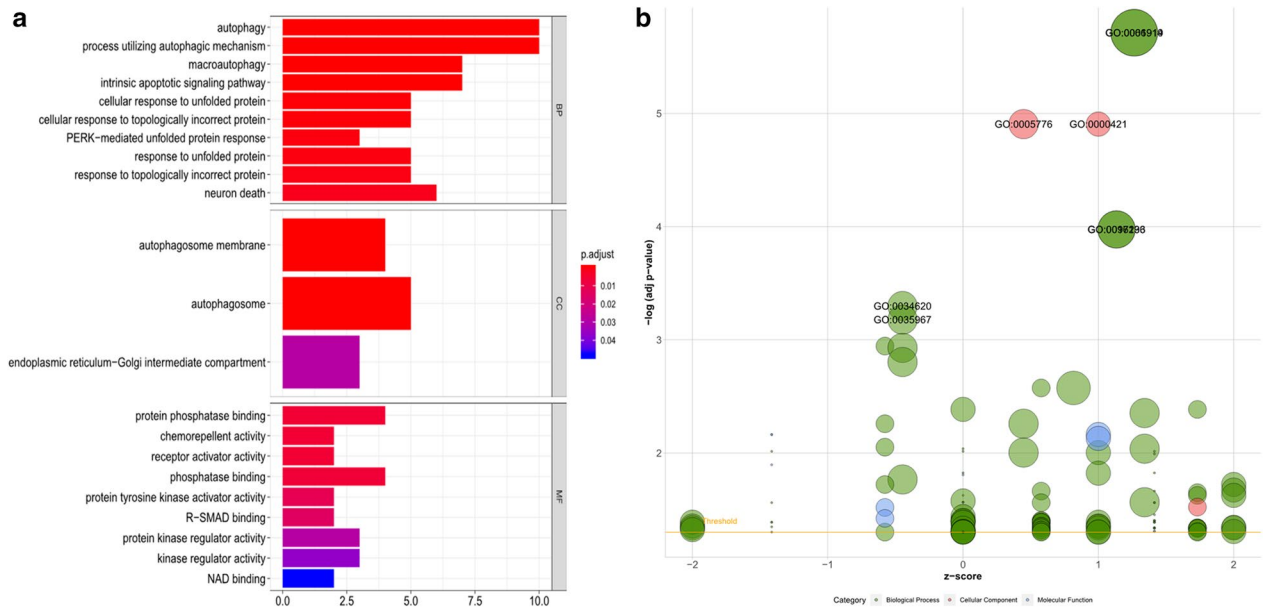
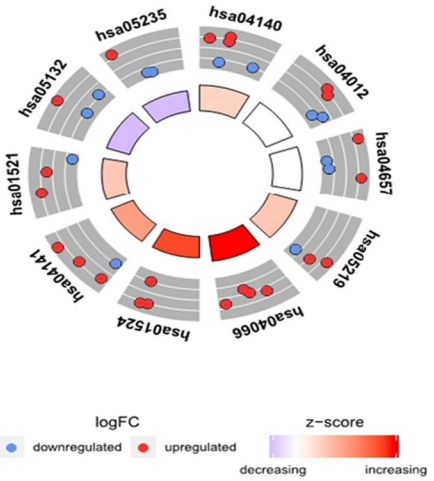


Fig. 7 Results of Gene Ontology (GO) functional annotation analysis. **a** Bar chart of significant terms. The change in colour from blue to red represents the increase in the adjusted P-value, and the length of the bar indicates the number of gene enrichment terms. **b** Bubble plot of enriched GO terms. The Z-score is plotted on the x-axis and the -log(adjusted p-value) is plotted on the y-axis; green represents a biological process, red represents cellular components and blue represents molecular function. The size of the circles is proportional to the number of genes enriched in the term



ID	Description
hsa04140	Autophagy − animal
hsa04012	ErbB signaling pathway
hsa04657	IL−17 signaling pathway
hsa05219	Bladder cancer
hsa04066	HIF−1 signaling pathway
hsa01524	Platinum drug resistance
hsa04141	Protein processing in endoplasmic reticulum
hsa01521	EGFR tyrosine kinase inhibitor resistance
hsa05132	Salmonella infection
hsa05235	PD−L1 expression and PD−1 checkpoint pathway in cancer

Fig. 8 Results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analyses of autophagy-associated genes

with the association between autophagy genes and LUAD in our study. However, some limitations are worth noting. As a retrospective study, this research has an inherent bias; although we validated the model using training/testing sets and immunohistochemistry, additional validation of prognostic designs should be conducted in vitro, in vivo and in clinical trials; moreover, the biological processes and molecular mechanisms of the 5 AAGs should be further evaluated to accelerate their clinical application in LUAD.

Conclusions

In this study, we provided insights into the roles of autophagy genes in LUAD and constructed a promising model, which could provide a reference to determine whether postoperative/preoperative patients are at high

risk. These patients are then more likely to receive more comprehensive neoadjuvant/adjuvant therapy with an improved prognosis.

Abbreviations
LUAD: Lung adenocarcinoma; TCGA: The Cancer Genome Atlas (database); ROC: Receiver operating characteristic; AUC: The area under the ROC curve; TNM: Tumour size/lymph nodes/distance metastasis, a tumour staging system used in oncology and constructed by the American Joint Committee on Cancer and the Union for International Cancer Control; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False discovery rate; IASLC: International Association for the Study of Lung Cancer.

Acknowledgements
Not applicable.

Authors' contributions
ZZ and XW contributed to the conception and design of the study; all authors participated in data acquisition, interpretation and statistical analysis; XW and SY were responsible for writing the manuscript, and all authors agreed to be accountable for all aspects of the work. All authors read and approved the final manuscript.

Funding
None.

Availability of data and materials
All data used in this study were acquired from The Cancer Genome Atlas (TCGA) portal.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 2 December 2019 Accepted: 27 March 2020
Published online: 01 April 2020

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Moon Y, Park JK, Lee KY. The effect of resection margin distance and invasive component size on recurrence after sublobar resection in patients with small (≤2 Cm) lung adenocarcinoma. World J Surg. 2020;44(3):990–997.
3. Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e142S–65S.
4. Wang SS, Chen G, Li SH, Pang JS, Cai KT, Yan HB, Huang ZG, He RQ. Identification and validation of an individualized autophagy-clinical prognostic index in bladder cancer patients. Onco Targets Ther. 2019;12:3695–712.
5. Towers CG, Thorburn A. Therapeutic targeting of autophagy. EBioMedicine. 2016;14:15–23.
6. Vidal RL, Hetz C. Crosstalk between the UPR and autophagy pathway contributes to handling cellular stress in neurodegenerative disease. Autophagy. 2012;8(6):970–2.
7. Cadwell K. Crosstalk between autophagy and inflammatory signalling pathways: balancing defence and homeostasis. Nat Rev Immunol. 2016;16(11):661.
8. Mitchener JS, Shelburne JD, Bradford WD, Hawkins HK. Cellular autophagocytosis induced by deprivation of serum and amino acids in HeLa cells. Am J Pathol. 1976;83(3):485–92.
9. Janku F, McConkey DJ, Hong DS, Kurzrock R. Autophagy as a target for anticancer therapy. Nat Rev Clin Oncol. 2011;8(9):528–539.
10. White E. Deconvoluting the context-dependent role for autophagy in cancer. Nat Rev Cancer. 2012;12(6):401–10.
11. Nassour J, Radford R, Correia A, Fuste JM, Schoell B, Jauch A, Shaw RJ, Karlseder J. Autophagic cell death restricts chromosomal instability during replicative crisis. Nature. 2019;565(7741):659–63.
12. Yun CW, Lee SH. The roles of autophagy in cancer. Int J Mol Sci. 2018;19(11):3466.
13. Rosenfeldt MT, Ryan KM. The multiple roles of autophagy in cancer. Carcinogenesis. 2011;32(7):955–63.
14. Gewirtz DA. The four faces of autophagy: implications for cancer therapy. Cancer Res. 2014;74(3):647–51.
15. Fan J, Zhang X, Wang S, Chen W, Li Y, Zeng X, Wang Y, Luan J, Li L, Wang Z, et al. Regulating autophagy facilitated therapeutic efficacy of the sonic Hedgehog pathway inhibition on lung adenocarcinoma through GLI2 suppression and ROS production. Cell Death Dis. 2019;10(9):626.
16. Zheng Z, Yang J, Zhao D, Gao D, Yan X, Yao Z, Liu Z, Ma Z. Downregulated adaptor protein p66(Shc) mitigates autophagy process by low nutrient and enhances apoptotic resistance in human lung adenocarcinoma A549 cells. FEBS J. 2013;280(18):4522–30.
17. Hu X, Shi S, Wang H, Yu X, Wang Q, Jiang S, Ju D, Ye L, Feng M. Blocking autophagy improves the anti-tumor activity of afatinib in lung adenocarcinoma with activating EGFR mutations in vitro and in vivo. Sci Rep. 2017;7(1):4559.
18. Liu F, Gao S, Yang Y, Zhao X, Fan Y, Ma W, Yang D, Yang A, Yu Y. Curcumin induced autophagy anticancer effects on human lung adenocarcinoma cell line A549. Oncol Lett. 2017;14(3):2775–82.
19. Park JH, Lee YR, So HS, Lee KK, Lee SY, Moon SR, Jo HJ, Lee S, Jeong K, Kwon KB, et al. The role of autophagy induced by pemetrexed in lung adenocarcinoma cells. Oncol Rep. 2014;31(5):2365–70.
20. Li YY, Lam SK, Zheng CY, Ho JC. The effect of tumor microenvironment on autophagy and sensitivity to targeted therapy in egfr-mutated lung adenocarcinoma. J Cancer. 2015;6(4):382–6.
21. Xiao K, Jiang J, Guan C, Dong C, Wang G, Bai L, Sun J, Hu C, Bai C. Curcumin induces autophagy via activating the AMPK signaling pathway in lung adenocarcinoma cells. J Pharmacol Sci. 2013;123(2):102–9.
22. Gurpinar E, Grizzle WE, Shacka JJ, Mader BJ, Li N, Piazza NA, Russo S, Keeton AB, Piazza GA. A novel sulindac derivative inhibits lung adenocarcinoma cell growth through suppression of Akt/mTOR signaling and induction of autophagy. Mol Cancer Ther. 2013;12(5):663–74.
23. Ren JH, He WS, Nong L, Zhu QY, Hu K, Zhang RG, Huang LL, Zhu F, Wu G. Acquired cisplatin resistance in human lung adenocarcinoma cells is associated with enhanced autophagy. Cancer Biother Radiopharm. 2010;25(1):75–80.
24. Wu T, Wang MC, Jing L, Liu ZY, Guo H, Liu Y, Bai YY, Cheng YZ, Nan KJ, Liang X. Autophagy facilitates lung adenocarcinoma resistance to cisplatin treatment by activation of AMPK/mTOR signaling pathway. Drug Des Devel Ther. 2015;9:6421–31.
25. Pan B, Chen D, Huang J, Wang R, Feng B, Song H, Chen L. HMGB1-mediated autophagy promotes docetaxel resistance in human lung adenocarcinoma. Mol Cancer. 2014;13:165.
26. Wang Q, Xiao Z, Lin Z, Zhou J, Chen W, Jie W, Cao X, Yin Z, Cheng J. Autophagy influences the low-dose hyper-radiosensitivity of human lung adenocarcinoma cells by regulating MLH1. Int J Radiat Biol. 2017;93(6):600–6.
27. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med. 2007;26(30):5512–28.
28. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
29. Zhao Y, Wang R, Shen X, Pan Y, Cheng C, Li Y, Shen L, Zhang Y, Li H, Zheng D, et al. Minor components of micropapillary and solid subtypes in lung adenocarcinoma are predictors of lymph node metastasis and poor prognosis. Ann Surg Oncol. 2016;23(6):2099–105.
30. Fujimoto J, Nunomura-Nakamura S, Liu Y, Lang W, McDowell T, Jakubek Y, Ezzeddine D, Kapere Ochieng J, Petersen J, Davies G, et al. Development

of Kras mutant lung adenocarcinoma in mice with knockout of the airway lineage-specific gene Gprc5a. Int J Cancer. 2017;141(8):1589–99.
31. Li X, Shi Y, Yin Z, Xue X, Zhou B. An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. J Transl Med. 2014;12:159.
32. De Ruysscher DKM, Decaluwe H. 4L lymph node involvement in left-sided lung cancer: unique or not? J Clin Oncol. 2018;36(29):2907–8.
33. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, Postmus PE, Rusch V, Sobin L. The IASLC lung cancer staging project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. J Thorac Oncol. 2007;2(8):706–14.
34. He R, Zuo S. A robust 8-gene prognostic signature for early-stage non-small cell lung cancer. Front Oncol. 2019;9:693.
35. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. Trends Mol Med. 2019;25(7):585–94.
36. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, Mak R, Aerts HJ. Exploratory study to identify radiomics classifiers for lung cancer histology. Front Oncol. 2016;6:71.
37. Inamura K. Update on immunohistochemistry for the diagnosis of lung cancer. Cancers. 2018;10(3):72.
38. Shah DR, Masters GA. Precision medicine in lung cancer treatment. Surg Oncol Clin N Am. 2020;29(1):15–21.
39. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu CQ, Strumpf D, Johnston MR, Darling G, Keshavjee S, Waddell TK, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol. 2007;25(35):5562–9.
40. Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, Girard L, Erickson HS, Roth J, Heymach JV, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. Clin Cancer Res. 2011;17(17):5705–14.
41. Wu M, Li X, Zhang T, Liu Z, Zhao Y. Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer. Front Oncol. 2019;9:996.
42. Zhang G, Fan E, Yue G, Zhong Q, Shuai Y, Wu M, Feng G, Chen Q, Gou X. Five genes as a novel signature for predicting the prognosis of patients with laryngeal cancer. J Cell Biochem. 2019. https://doi.org/10.1002/jcb.29535.
43. Russell RC, Yuan HX, Guan KL. Autophagy regulation by nutrient signaling. Cell Res. 2014;24(1):42–57.
44. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500.
45. Stirrups R. Osimertinib improves progression-free survival in NSCLC. Lancet Oncol. 2018;19(1):e10.
46. de Bono JS, Rowinsky EK. The ErbB receptor family: a therapeutic target for cancer. Trends Mol Med. 2002;8(4 Suppl):S19–26.
47. Kolch W, Pitt A. Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. Nat Rev Cancer. 2010;10(9):618–29.
48. Jutten B, Rouschop K. EGFR signaling and autophagy dependence for growth, survival, and therapy resistance. Cell Cycle. 2014;13(1):42–51.
49. Chang SH, Dong C. A novel heterodimeric cytokine consisting of IL-17 and IL-17F regulates inflammatory responses. Cell Res. 2007;17(5):435–40.
50. Zhao J, Chen X, Herjan T, Li X. The role of interleukin-17 in tumor development and progression. J Exp Med. 2019.
51. Murugaiyan G, Saha B. Protumor vs antitumor functions of IL-17. J Immunol. 2009;183(7):4169–75.
52. Chang SH, Mirabolfathinejad SG, Katta H, Cumpian AM, Gong L, Caetano MS, Moghaddam SJ, Dong C. T helper 17 cells play a critical pathogenic role in lung cancer. Proc Natl Acad Sci USA. 2014;111(15):5664–9.
53. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, Ameh S, Sandel D, Liang XS, Mazzilli S, et al. Commensal microbiota promote lung cancer development via γδ t cells. Cell. 2019;176(5):998–1013.e1016.
54. Fang Y, Tan J, Zhang Q. Signaling pathways and mechanisms of hypoxia-induced autophagy in the animal cells. Cell Biol Int. 2015;39(8):891–8.
55. Bellot G, Garcia-Medina R, Gounon P, Chiche J, Roux D, Pouysségur J, Mazure NM. Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains. Mol Cell Biol. 2009;29(10):2570–81.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC