A LncRNA panel within EpCAM-specific exosomes for noninvasive early diagnosing non-small cell lung cancer

Linyuan Liu[1†], Danlei Li[1†], Amei Zhuo[1†], Jiachun Lu[1,2,3], Jianjun Zou[1,4], Guitian Huang[5], Zhaoting Hu[6], Zili Zhang[1], Yibin Deng[2*] and Lei Yang[1,2,3*]

Abstract

Background Plasma tumor-associated exosomes represent a promising source for cancer biomarkers; however, the role of long non-coding RNAs (lncRNAs) within these exosomes is not well-defined in non-small cell lung cancer (NSCLC).

Methods We identified a panel of NSCLC-specific lncRNAs within plasma EpCAM-specific exosomes (Epexo) through a comparative analysis of lncRNA profiles between plasma Epexo and lung tissues. The panel's diagnostic value was firstly evaluated in a retrospective cohort of 210 NSCLC patients and 245 healthy controls, and validated in a prospective cohort of 192 patients with pulmonary nodules (nodule size < 3 cm in diameter). The evaluation utilized the area under the ROC curve (AUC) based on a random forest model. For precision, repeat testing was conducted with 31 randomly selected samples. Additionally, 39 paired tissue-plasma samples were employed to assess the concordance of lncRNA expression between tissue and plasma within the same individuals.

Results The panel, including linc01125, HNF1A-AS1, MIR100HG, linc01160, and ZNRF3-AS1, demonstrated superior capability in distinguishing early-stage NSCLC patients from controls, achieving AUC values of 0.805 and 0.856 in the discovery and validation set, respectively. The panel also showed potential for differentiating adenocarcinoma and squamous cell carcinoma. Repeat sample testing showed a consistency of 90.3% for this panel. The expression levels of MIR100HG and HNF1A-AS1 showed significant correlations between plasma Epexo and cancerous tissues.

Conclusions The identified lncRNA panel, consisting of linc01125, HNF1A-AS1, MIR100HG, linc01160, and ZNRF3-AS1, presents a promising diagnostic tool for NSCLC.

Clinical trial number not applicable.

Keywords EpCAM-specific exosomes, Long non-coding RNA, Non-small cell lung cancer, Pulmonary nodule

---

†Linyuan Liu, Danlei Li and Amei Zhuo contributed equally to this work.

*Correspondence:
Yibin Deng
dengyb75@163.com
Lei Yang
leiyang@gzhmu.edu.cn

Full list of author information is available at the end of the article

Liu *et al. Respiratory Research*   (2025) 26:144

Page 2 of 16

Background

Lung cancer has been the leading cause of cancer-related death for years [1, 2]. Despite significant advancements in the treatment of lung cancer in the past decade, the five-year survival rate remains low at only 25.4%, according to the Surveillance, Epidemiology, and End Results Program (SEER: https://seer.cancer.gov/). One of the most signific ant challenges in improving the survival rate of lung cancer is the low rate of early detection. Approximately 53% of lung cancer cases are diagnosed with metastasis, and an additional 21% have already spread to regional lymph nodes at the time of diagnosis. The five-year survival rates for these cases are only 8.2% and 34.8%, respectively (https://seer.cancer.gov/statfacts/html/lungb.html). Low-dose spiral computed tomography (LDCT) has been widely acclaimed in recent years as an effective screening method for early detection of lung cancer. Multiple randomized controlled trials have demonstrated its ability to reduce mortality due to lung cancer by approximately 20.4% in high-risk populations [3–6]. However, LDCT still presents challenges such as a high rate of false positive results, over-diagnosis, and the accompanying economic burden and psychological impact of follow-up for pulmonary nodules (PN) [6]. Therefore, the exploration of alternative early diagnostic technologies for lung cancer is still necessary.

Exosomes are a type of small vesicles ranging in diameter from 30 nm to 150 nm. They contain various cell-derived substances such as DNA, RNA, proteins, and small molecule metabolites, which play a crucial role in facilitating cell-cell communication. Exosomes are abundantly found in a variety of body fluids, and their double-layered membrane efficiently safeguards the components originating from parental cells, shielding them from degradation within the fluid and facilitating their detection. So far, a significant number of studies have demonstrated the diagnostic potential of components within plasma exosomes for non-small cell lung cancer (NSCLC) [7]. Unfortunately, different studies have reported different diagnostic molecules within exosomes, highlighting the lack of stability and reliability in the identified diagnostic biomarkers thus far. One possible reason for this phenomenon is the complex cellular origin of plasma exosomes. Indeed, for tumor-related diseases, there is a pressing need to further understand the role of tumor-associated exosomes (TAEx) in order to unravel the aforementioned challenges.

TAEx are exosomes that bear tumor molecular markers on their membrane. EpCAM is commonly expressed in epithelial tissues and its dysregulation has been linked to various cancers. EpCAM-specific exosomes (Epexo) could represent a significant type of TAEx, as they may be detectable only in the plasma of cancer patients [8, 9]. Individual studies have currently shown that Epexo contents have an impressive diagnostic value for NSCLC, including high accuracy and stability [9, 10]. For instance, miR-486 and miR-21 in plasma Epexo have high sensitivity and specificity for early lung cancer diagnosis [10], and these two microRNAs are also the only two consistently reported NSCLC diagnostic biomarkers in plasma exosomes [11–13].

Long non-coding RNAs (lncRNAs) have emerged as crucial regulators in cancer progression and are increasingly recognized for their potential in cancer diagnosis and prognosis. Recent studies have unveiled their pivotal role in regulating gene expression at various levels, including chromatin modification, transcription, and post-transcription [14, 15]. In cancer, lncRNAs can function as oncogenes or tumor suppressors. For instance, LINC00672 is reported to contributing p53 protein-mediated gene suppression and promotes endometrial cancer chemosensitivity [16]; the lncRNA MALAT1 is known to promote metastasis in lung cancer by affecting epithelial-mesenchymal transition (EMT) and has been implicated in poor patient prognosis [16]. Moreover, lncRNAs are detectable in exosomes, making them promising non-invasive biomarkers for early cancer diagnosis. However, there are also significant challenges regarding the stability and reliability of detecting lncRNAs in exosomes, as mentioned previously.

Since the role of lncRNAs in Epexo in lung cancer diagnosis remains unknown, in this study, we conducted a pioneering analysis of the diagnostic value of lncRNAs in plasma Epexo of patients affected by NSCLC. We discovered that these lncRNAs show significant potential for both the early diagnosis and molecular classification of NSCLC, highlighting their innovative role in enhancing diagnostic accuracy and patient management.

Materials and methods
Clinical specimens

Figure 1 illustrates the research design of the study. Initially, five cases of NSCLC and five pneumonia controls were randomly selected for RNA sequencing (RNA-seq) to identify differentially expressed lncRNAs in NSCLC-derived Epexo. The identified lncRNAs were then compared with previously published lung cancer lncRNA expression profiles to find lung cancer-specific exosomal lncRNAs [17]. A retrospective study was conducted from January 2016 to October 2023, involving 210 pathologically confirmed NSCLC patients and 245 healthy individuals, serving as a discovery set to evaluate the diagnostic significance of identified Epexo lncRNAs. Healthy participants were screened using chest computed tomography (CT) or low-dose CT (LDCT) scans. To enhance the study's reliability, a matching design based on age (± 5 years) and gender was implemented. Additionally, a cohort of 192 patients with PN, ranging in size from
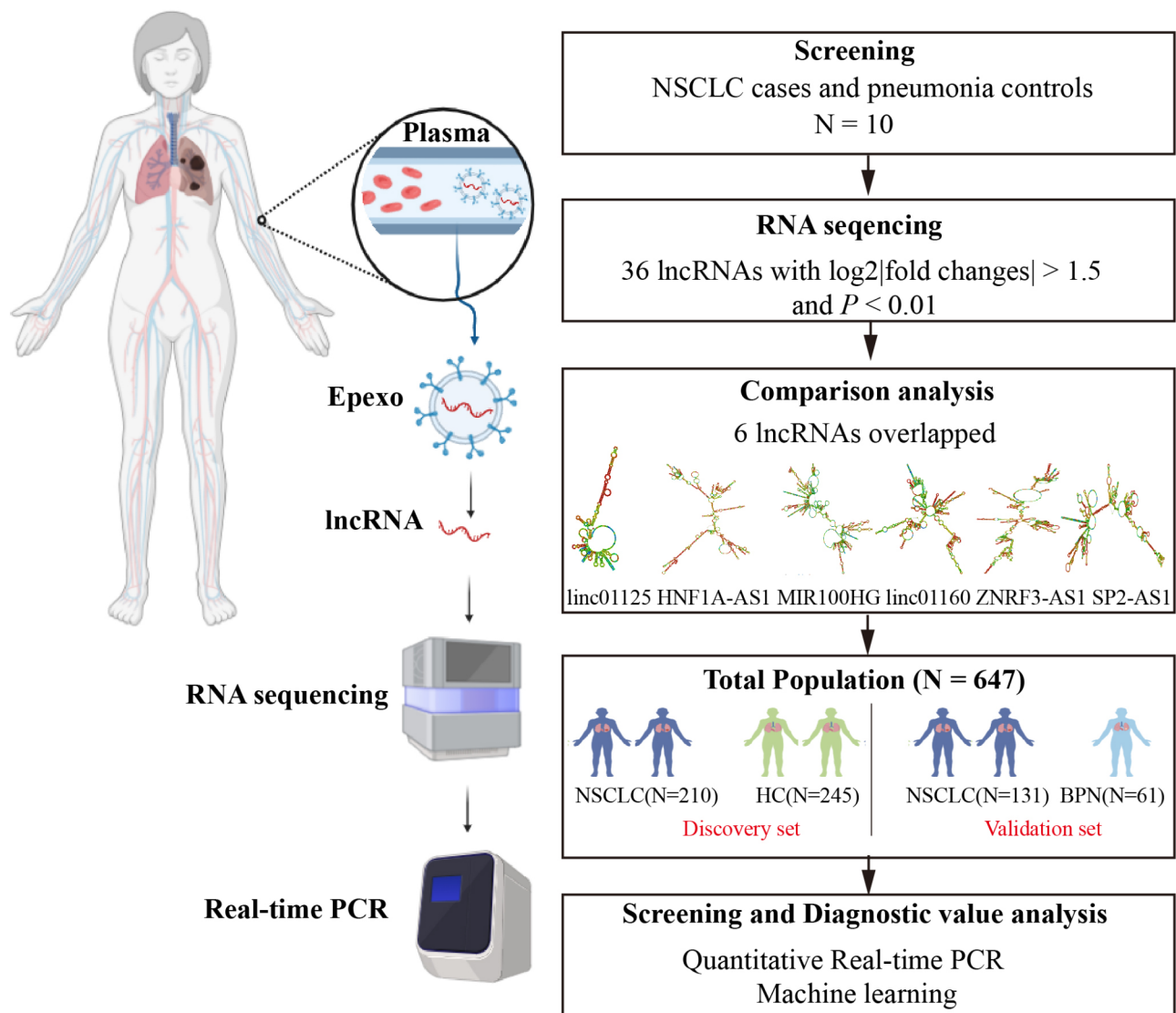
Fig. 1 A flowchart of research design. Epexo: EpCAM-specific exosomes; NSCLC: non-small cell lung cancer; MPN: malignant pulmonary nodule; BPN: benign pulmonary nodule; The secondary structures of the target lncRNAs were predicted using the RNAfold website. (http://rna.tbi.univie.ac.at/cgi-bin /RNAWebSuite/RNAfold.cgi)

4 mm to 3 cm, were enrolled as a validation set to prospectively assess the potential diagnostic value of Epexo lncRNAs for early-stage NSCLC detection. Of these, 131 were confirmed malignant through surgical intervention, while the others were diagnosed as benign or stable nodules over a four-year follow-up. All participants provided informed consent for their plasma samples and clinical data. A cohort of 31 patients also underwent secondary blood sample collection to reevaluate the targeted lncRNAs, enhancing the reliability of the diagnostic tool. The study was approved by the Institutional Review Board of Guangzhou Medical University, and demographic details of the participants are summarized in Table 1.

Plasma total exosomes and Epexo isolation and characterization

Exosome isolation and characterization substantially adhere to Minimal Information for Studies of Extracellular Vesicles (MISEV) 2023 guidelines. Plasma was obtained by immediate centrifugation of 5 ml of heparinized blood, and the plasma was subsequently stored at -80 °C. Total exosomes from human plasma were isolated using the SBI ExoQuick exosome precipitation solution (System Biosciences, Palo Alto, CA) as per a previously described method [18]. Briefly, Plasma samples underwent centrifugation at 3,000×g for 15 min at 4 °C to eliminate cellular debris. Subsequently, the clarified biofluid was treated with the appropriate volume of ExoQuick. The ExoQuick/biofluid mixture was centrifuged

Liu *et al. Respiratory Research*   (2025) 26:144

Page 4 of 16

Table 1 Frequency distributions of demographics in studied subjects

Variables	NSCLC case-control study			PN case-control study		
	Case, *n* (%)	Control, *n* (%)	P^a	Malignant, *n* (%)	Benign, *n* (%)	P^a
No. of subjects	210	245		131	61	
Gender						
Male	156(74.3)	173(70.6)	0.383	65(49.6)	36(59.0)	0.225
Female	54(25.7)	72(29.4)		66(50.4)	25(41.0)	
Age (years)						
>60	109(51.9)	121(49.4)	0.592	70(53.4)	16(26.2)	<0.001
≤60	101(48.1)	124(50.6)		61(46.6)	45(73.8)	
Smoking status						
Ever	105(50.0)	88(35.9)	0.002	37(28.2)	15(24.6)	0.596
Never	105(50.0)	157(64.1)		94(71.8)	46(75.4)	
Stages						
LS	32(15.3)					
RDS	158(75.2)					
Undetermined	20(9.5)					
Histological type						
LUSC	97(46.2)			6(4.6)		
LUAD	81(38.6)			113(86.3)		
Other/unknown	32(15.2)			12(9.1)		
Maximum diameter						
<10 mm				45(34.4)	48(78.7)	<0.001
10–20 mm				57(43.5)	12(19.7)	
>20–30 mm				29(22.1)	1(1.6)	

^{a}*P* value from a Chi-square test

at 3,000×g for 10 min, followed by removal of the supernatant. The resulting pellet was resuspended. The isolated exosomes were then purified and stored at -80 ℃ until required. Epexo were isolated using magnetic beads that target the EpCAM. Capture beads were dispensed into each 12×75 mm Polystyrene Round Bottom tube (cytometer tube), and 45μL of ExoStep Incubation Buffer along with 50μL of the sample were added. A portion of the supernatant (20μL) was retained from the tubes. Subsequently, 100μL of detached solution was introduced, followed by a 2-hour incubation at 37 ℃. After incubation, the mixture was centrifuged at 2,500×g for 5 min to retrieve the supernatant containing the released exosomes detached from the antibody-bead complex, which were then prepared for downstream analysis.

The morphology and grain size of exosomes, including Epexo and EpCAM negative exosomes, were visualized and measured using the Hitachi HT-7700 transmission electron microscopy (TEM, Tokyo, Japan) and the NanoFCM N30E particle size analyzer (Nottingham, UK). The analysis was completed by a technology company (Keyida, Guangzhou, China). Furthermore, the protein expression of EpCAM and CD63 was determined using western blot analysis with antibodies against EpCAM (ab223582, Abcam) and CD63 (ab134045, Abcam). The western blot protocol was described previously [19].

LncRNA identification by RNA-seq

We combined Epexo and EpCAM-negative exosomes in an 8:2 ratio to isolate exosomal RNA using Trizol reagent and was first quantified and assessed for quality. RNA sequencing libraries were prepared using a standardized protocol, including end trimming, fragmentation, cDNA synthesis, adapter ligation and PCR amplification. Quality-verified libraries were sequenced on the Illumina HiSeq 2500 platform. The raw data was processed using fastp to eliminate low-quality bases, adapters, and reads containing excessive unknown bases, resulting in clean reads. After removing ribosomal RNA, the reads depleted of rRNA were aligned to the reference genome. For lncRNAs, the HISAT2 software (version 2.2.1.0) was used to align the reads and the processed reads were quantified to estimate the expression levels of lncRNAs. To identify differentially expressed lncRNAs between NSCLC patients and pneumonia controls, a statistical analysis was performed using the DESeq2 in R. In this analysis, lncRNAs with Fragments Per Kilobase of transcript per Million mapped reads greater than 0.5 in 4 or more samples were retained for analysis. The criteria for significance was set based on log2|fold changes| > 1.5 and *P* value < 0.01. The RNA-seq data for this study (HRA011042) are available on the China National Center for Bioinformation website (https://ngdc.cncb.ac.cn/). All relevant data within the scope of the paper are publicly available.

Liu *et al. Respiratory Research*   (2025) 26:144

Page 5 of 16

Real-Time quantitative PCR of LncRNAs

To ensure consistency between the exosomes derived from cases and controls for potential future clinical testing, we combined Epexo and EpCAM-negative exosomes in an 8:2 ratio. Exosomal RNA was subsequently extracted using Trizol reagent, and complementary DNAs (cDNAs) were synthesized according to the manufacturer's instructions using the PrimeScript reverse transcriptase reagent kit. Quantitative real-time PCR was performed on the ABI 7500 real-time PCR system utilizing the TB Green Master Mix (Takara, Beijing, China). Beta-actin was employed as the internal control, and each experiment was conducted at least twice to ensure reproducibility. The primer sequences are provided in Supplementary Table 1. The ΔCT method (i.e., CT_{target} - $CT_{control}$) was utilized to assess the relative expression levels of the target lncRNAs. Importantly, the health status of the samples was unknown to the testing personnel throughout the study.

Gene function annotation

To demonstrate the potential impact of target lncRNAs, we downloaded RNA-seq raw counts from TCGA-LUSC and TCGA-LUAD tumor tissues. A co-expression analysis was carried out between the target lncRNAs and the whole transcriptomes using the Spearman rank correlation test. Subsequently, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and gene-set enrichment analysis (GSEA) were conducted using the "clusterProfiler" package in R and GSEA 4.1.0 software, respectively.

Statistical analysis

Tidymodels, a collection of R packages tailored for modeling and machine learning tasks, was harnessed to construct a predictive random forest model. The model was trained using the expression levels of six Epexo lncRNAs (serving as predictors) and diagnostic outcomes from 210 NSCLC cases and 245 normal controls. The dataset was randomly partitioned into training ($n = 340$) and testing ($n = 115$) subsets for model development and evaluation. All possible combinations of predictors were examined. The developed model underwent validation to evaluate its diagnostic efficacy, utilizing receiver operating characteristic (ROC) curve analysis and calculation of the area under the curve (AUC). The 95% confidence interval (95% CI) for the AUC was estimated employing the Delong model, providing a rigorous assessment of the model's performance and reliability [20]. Additionally, the logit model was employed to estimate the AUC for the diagnostic model developed within the PN group. Similarly, the predictive random forest model was utilized to determine the accuracy of all possible combinations of Epexo lncRNAs for different NSCLC pathological types

and M staging. In this section, patients with lung adenocarcinoma (LUAD) and squamous carcinoma (LUSC) were split randomly into training ($n = 124$) and testing ($n = 54$) groups for model development and evaluation. Differences in target lncRNA expressions between cases and controls were evaluated using unpaired t-tests with Welch's correction. All statistical analyses were conducted using GraphPad Prism (version 9.0) or R (v4.2.3). A P value of less than 0.05 was deemed statistically significant.

Results

Characteristic and identification of Epexo

Total exosomes has been characterized in previous studies using transmission electron microscopy (TEM), particle size analysis, and protein marker analysis [18, 19]. Here, as derived from total exosomes, Epexo presence was exclusively identified in plasma sourced from NSCLC cases, absent in healthy controls, as delineated through TEM (Fig. 2a). This observation aligns with findings reported in extant literatures [8, 9]. However, particle size analysis was capable of detecting diameter data in Epexo from both cases and controls, which might be attributed to its higher detection sensitivity compared to TEM. Interestingly, the mean diameter of Epexo substantially exceeded that of the EpCAM-negative counterparts (Fig. 2b). In concordance, Epexo demonstrated positive protein expression of EpCAM, a feature not detected in the EpCAM-negative entities (Fig. 2c). Intriguingly, divergent from preceding research which cited an absence of EpCAM protein expression in exosomes from healthy controls [9], this investigation uncovered positive expression thereof in plasma exosomes derived from subjects harboring either malignant or benign PN (Fig. 2d).

Identification of differentially expressed LncRNAs in exosomes between cases and controls

Utilizing RNA-seq analysis, a comprehensive study identified 2935 lncRNAs in the Epexo of all NSCLC patients and 5082 lncRNAs in all pneumonia controls. Among these, 2744 lncRNAs were commonly expressed in both cases and controls. Applying predefined criteria, 15 down-regulated and 21 up-regulated lncRNAs were singled out in the plasma Epexo of NSCLC cases (Fig. 3a and b). Subsequently, a selection of six overlapped lncRNAs underwent further scrutiny after alignment with previously published profiles of dysregulated lncRNA in lung cancer tissues (Fig. 3c) [17]. These lncRNAs demonstrated discriminatory potential in distinguishing NSCLC cases from pneumonia controls (Fig. 3d). Upon examination in the discovery set, it was found that five lncRNAs, specifically linc01125 ($P < 0.001$), linc01160 ($P = 0.005$), SP2-AS1 ($P < 0.001$), MIR100HG ($P = 0.015$), and HNF1A ($P = 0.035$), exhibited significant differences among the

Liu *et al. Respiratory Research*   (2025) 26:144

Page 6 of 16
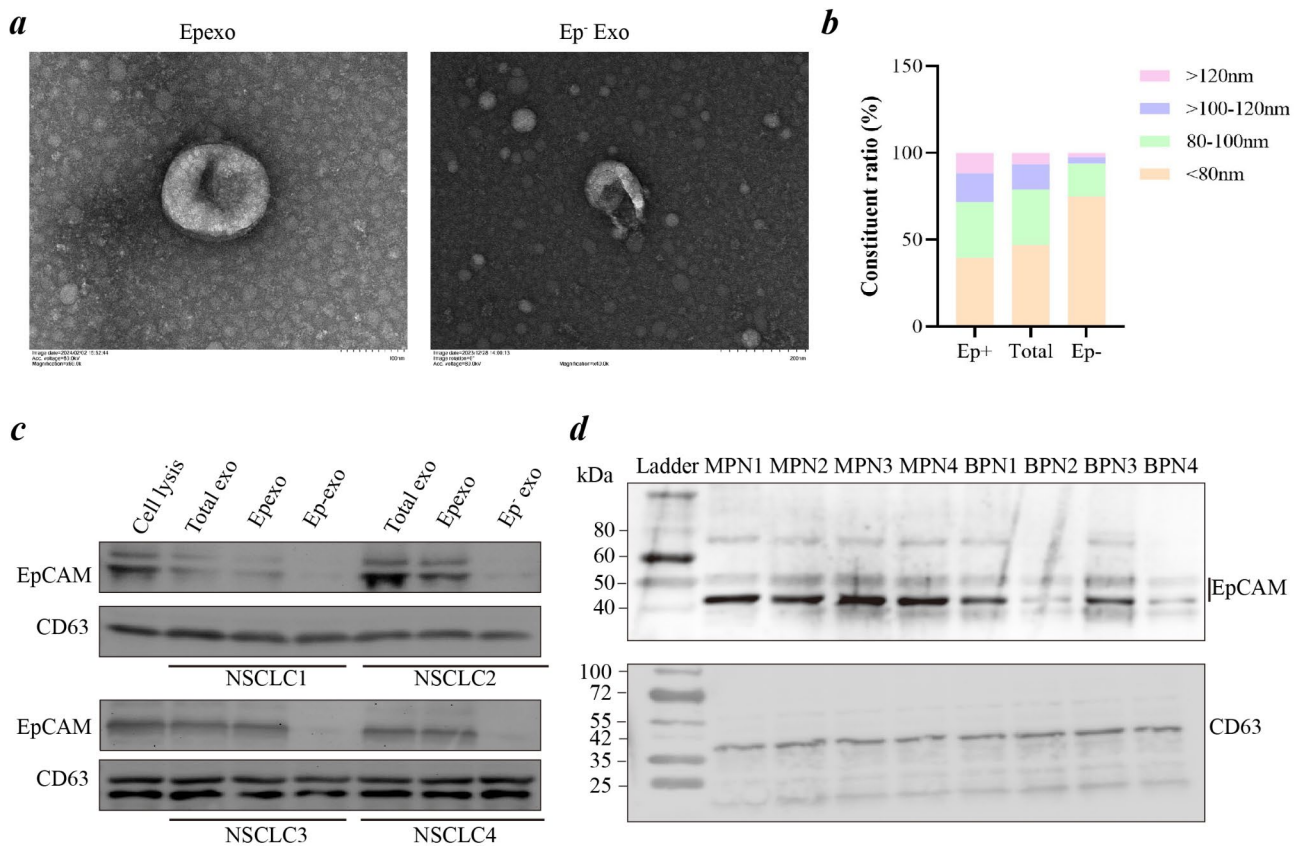
Fig. 2 Comprehensive characterization of plasma-derived EpCAM-specific Exosomes (Epexo). (**a**) Depicts representative transmission electron microscopy images showcasing the morphological attributes of exosomes with EpCAM expression (Epexo) contrasted against those without (Ep⁻ Exo). (**b**) Illustrates the size distribution profile for exosomes, as determined through nanoparticle tracking analysis, highlighting uniform particulate dimensions. (**c**) Presents representative western blot analyses, depicting the protein expression levels of EpCAM and CD63 across different samples, including a positive control-cell lysis, total exosome fractions, and sub-populations of Epexo and Ep⁻ Exo, isolated from four individuals diagnosed with NSCLC. (**d**) Quantifies and compares the expression levels of EpCAM and CD63 protein within Epexo obtained from patients exhibiting malignant PN (MPN) versus those with benign PN (BPN)
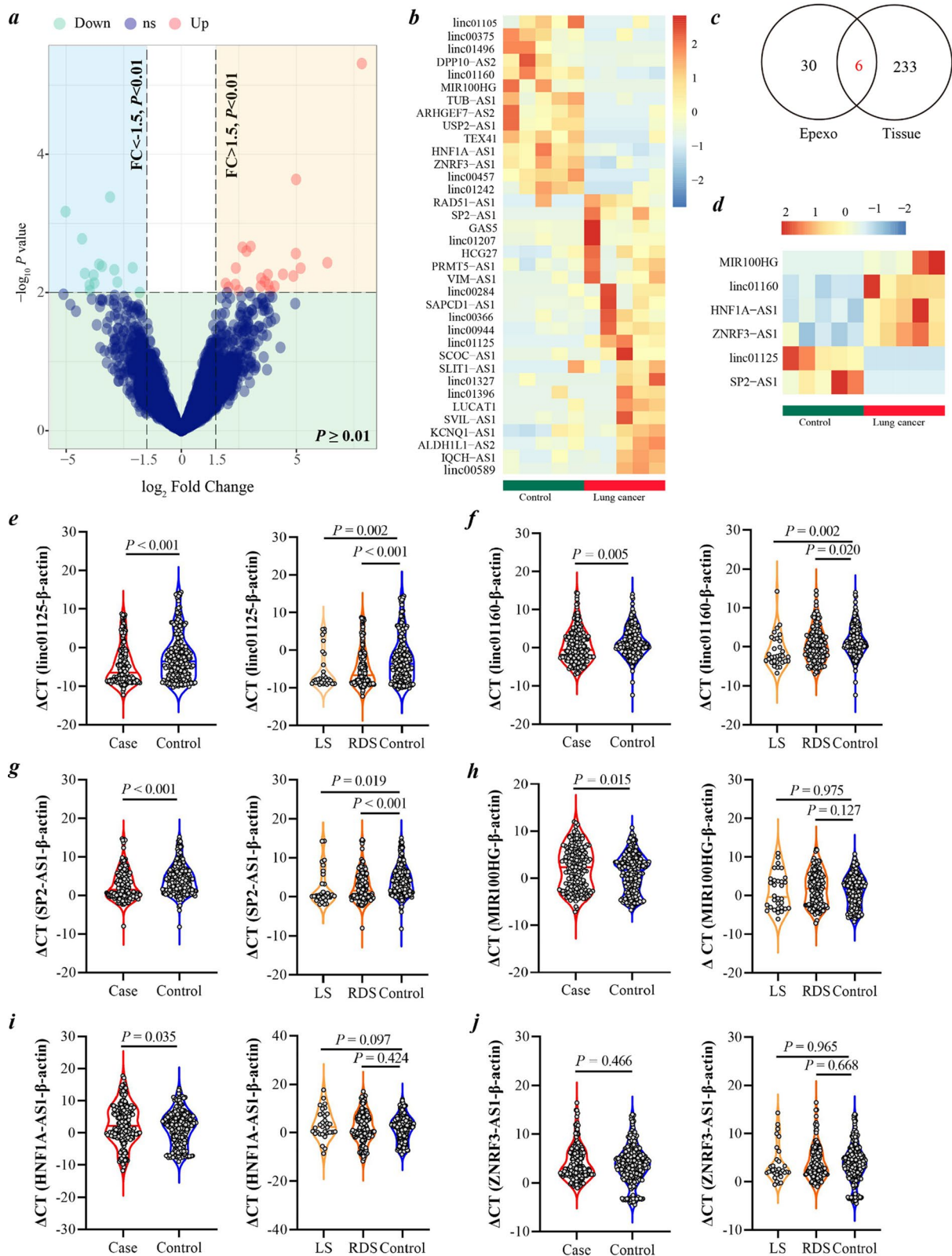
groups. Notably, ZNRF3-AS1 ($P = 0.466$) did not show significant differences. Furthermore, the observed significant disparities persist for linc01125, linc01160, and SP2-AS1 both in patients with localized stage (LS) and those with regional or distant stage (RDS) when contrasted with the control subjects (Fig. 3e-j).

Construction of NSCLC diagnostic model through LncRNAs combination in plasma Epexo

Utilizing a predictive random forest model, the optimal performance of lncRNA combinations in Epexo for the diagnosis of NSCLC was evaluated in the discovery set. Among sixty-three distinct combinations, a quintet lncRNA panel, comprising linc00125, linc01160, MIR100HG, HNF1A-AS1, and ZNRF3-AS1, manifested the most superior AUC value (Fig. 4a). In contrast to the AUC values procured from each individual lncRNA analysis (Fig. 4b)-0.649 (95% CI = 0.549–0.750) for linc01125, 0.667 (95% CI = 0.566–0.768) for HNF1A-AS1, 0.593 (95% CI = 0.484–0.702) for SP2-AS1, 0.556 (95%

CI = 0.451–0.662) for MIR100HG, 0.565 (95% CI = 0.459–0.670) for linc01160, and 0.636 (95% CI = 0.533–0.740) for ZNRF3-AS1-this composite model achieved a markedly elevated AUC of 0.928 (95% CI = 0.883–0.974; Fig. 4c). Furthermore, the precision-recall curve nearly reached the optimal top-right position, denoting significant precision across various levels of recall (Fig. 4d). Intriguingly, the lncRNA ensemble demonstrated a remarkable AUC value (0.805, 95% CI = 0.559-1.000) for early-stage NSCLC detection in distinguishing patients at LS from healthy control subjects (Fig. 4e). Furthermore, despite the fact that resampling ($n = 31$) and subsequent retesting only identified significant correlations for linc01125 (Fig. 4f) and linc01160 (Fig. 4g), while failing to do so for the other lncRNAs (Figs. 4h-j), the diagnostic lncRNA panel exhibited a notable level of coherence (concordance rate = 90.3%; Fig. 4k).

To be consistent, significant differences were observed in the expression levels of Epexo linc01125 ($P = 0.012$; Fig. 5a), SP2-AS1 ($P = 0.035$; Fig. 5b), MIR100HG

Fig. 3 (See legend on next page.)

Liu *et al. Respiratory Research*   (2025) 26:144

Page 8 of 16

(See figure on previous page.)

Fig. 3 Profiling Novel Non-Small Cell Lung Cancer (NSCLC)-Specific Long Non-Coding RNAs (lncRNAs) in plasma-derived Epexo. (**a**) Depiction of a volcano plot comparing lncRNA expression profiles in plasma-derived Epexo between NSCLC cases and pneumonia controls, with 1.5-fold up- and down-regulated lncRNAs (with $P < 0.01$) in NSCLC samples highlighted by light green and pink dots, respectively. (**b**) Heat map illustrating the altered expression levels of lncRNAs in plasma-derived Epexo between NSCLC cases and pneumonia controls. (**c**) Venn diagram displaying six dysregulated lncRNAs shared between plasma-derived Epexo and cancerous tissues from NSCLC patients. (**d**) Heat map visualizing the six common dysregulated lncRNAs in plasma-derived Epexo between NSCLC cases and pneumonia controls. (**e-j**) Expression patterns of linc01125 (**e**), linc01160 (**f**), SP2-AS1 (**g**), MIR100HG (**h**), HNF1A-AS1 (**i**), and ZNRF3-AS1 (**j**) in plasma-derived Epexo from a cohort of 210 NSCLC patients, including patients with local stage (LS) and regional or distant stage (RDS), as well as 245 healthy individuals (control). Statistical analysis was conducted using unpaired t-tests with Welch's correction. Each dot in the violin plots represents the expression value of the respective lncRNA, with the solid line in the center indicating the median expression level

($P = 0.040$; Fig. 5c), and ZNRF3-AS1 ($P = 0.009$; Fig. 5d), whereas no statistically significant differences were found for HNF1A-AS1 and linc01160 (Fig. 5e, f) between malignant PN patients and benign PN controls in the validation set. The corresponding AUC values for distinguishing malignant patients from benign controls were 0.619 (95%CI = 0.528–0.711), 0.576 (95%CI = 0.483–0.670),0.607 (95%CI = 0.519–0.695), 0.625 (95%CI = 0.534–0.717), 0.539 (95%CI = 0.449–0.629), and 0.559 (95%CI = 0.465–0.654), respectively. The diagnostic lncRNA panel, comprising Epexo linc01125, linc01160, MIR100HG, HNF1A-AS1, and ZNRF3-AS1, achieved an AUC of 0.716 (95% CI = 0.637–0.794; Fig. 5g). Upon integration of the nodule size into the panel, the AUC was elevated to 0.856 (95%CI = 0.800-0.913; Fig. 5g). The precision-recall curve neared the optimal top-right position (Fig. 5h). Moreover, the diagnostic accuracy of the model, incorporating the optimized lncRNA combination and nodule size, remained remarkable for differentiating maligant patients from benign controls, with AUCs of 0.854 (95% CI = 0.781–0.927) for nodules ≤ 10 mm (Fig. 5i).The findings demonstrate that a panel consisting of five lncRNAs within plasma Epexo serves as a potent diagnostic instrument for NSCLC early detection.

Utilization of Epexo LncRNAs as a classifier for discrimination of pathological types

A significant disparity in the expression levels of all six examined lncRNAs in plasma Epexo was observed between patients with LUAD and those with LUSC (Fig. 6a). This finding prompted an investigation into the potential of these lncRNAs to serve as classifiers for distinguishing between the pathological subtypes of NSCLC.Following an evaluation of various lncRNA combinations, two specific quintet panels emerged as highly accurate classifiers. One panel included SP2-AS1 and ZNRF3-AS1, while the other featured linc01125, HNF1A-AS1, and SP2-AS1. Both panels demonstrated exceptional accuracy (0.722, $P < 0.001$; Fig. 6b) and maintained consistent true performance indices in the testing samples (Fig. 6c, d).

Differential expression of Epexo LncRNAs across TNM stages in NSCLC patients

Upon comparing the expression levels of Epexo lncRNAs across TNM stages, significant differences were noted in the expression of linc01125 ($P = 0.007$), SP2-AS1 ($P = 0.007$), linc01160 ($P = 0.016$), and ZNRF3-AS1 ($P = 0.037$) among NSCLC patients with and without distant metastasis (Fig. 6e). However, no significant effects were observed for all Epexo lncRNAs across different T or N stages (Figure S1).

Correlation between LncRNA expression in plasma Epexo and cancer tissues

Upon evaluating expression levels of the aforementioned six lncRNAs in 39 NSCLC tissue samples, our investigation uncovered significant inverse correlations for both MIR100HG ($r = -0.401$, $P = 0.011$; Fig. 7a) and HNF1A-AS1 ($r = -0.327$, $P = 0.042$; Fig. 7b) across plasma-derived Epexo and cancerous tissues. Nonetheless, while negative associations were likewise observed for the remaining lncRNAs across these biological specimens, such correlations failed to achieve statistical significance (Fig. 7c-f).

Bioinformatics analysis for selected LncRNAs and their targets

In the aforementioned diagnostic framework, the majority of lncRNAs have been substantiated to exert critical functions in the pathogenesis and progression of neoplasms, inclusive of lung carcinoma. However, the comprehensive elucidation of the biological roles played by linc01160 and ZNRF3-AS1 remain insufficiently explored. Upon querying the GEPIA database (http://gepia.cancer-pku.cn/index.html), a high expression level of linc01160 was found to correlate with a favorable overall survival of NSCLC patients (Fig. 7g). The top 200 genes exhibiting significant correlations with linc01160, ranked by correlation coefficients, were enriched in the Hippo signaling pathway (Fig. 7h). Also, a high expression level of ZNRF3-AS1 was found to correlate with a poor disease-free survival of NSCLC patients (Fig. 7i). The top 200 genes exhibiting significant correlation with ZNRF3-AS1, were enriched in the Peroxisome and Carbon metabolism (Fig. 7j).

Discussion

Despite numerous studies highlighting the significant diagnostic value of plasma exosome components for tumors, including lung cancer, inconsistencies in the molecules reported across different studies have introduced challenges in the translational application of exosome diagnostics. TAEx in plasma offer distinct advantages over total exosomes as diagnostic tools. These TAEx are more enriched with biomarkers that are highly relevant to the presence, type, and progression of tumors [21–23]. This targeted approach can provide more precise diagnostic information, enhancing the detection and monitoring of cancers such as lung cancer. Moreover, the selectivity of TAEx can reduce background noise, thereby improving diagnostic specificity and sensitivity. In this study, through comprehensive analysis, a panel of Epexo lncRNAs was identified in plasma, including linc01125, HNF1A-AS1, MIR100HG, linc01160, and ZNRF3-AS1. This panel demonstrated strong diagnostic potential for the early detection of NSCLC. Notably, the identified lncRNAs showed a distinctive ability to differentiate between LUAD and LUSC, effectively distinguishing patients from controls while also enabling precise pathological diagnosis.

The current understanding of the specific cargoes contained within TAEx in plasma is still emerging. Recent studies have highlighted the potential of certain biomarkers, notably microRNA miR-21 and mRNA TTF-1, in plasma exosomes bearing EGFR or PD-L1, for differentiating NSCLC patients from normal controls [24]. While the Epexo 5-lncRNA panel demonstrates comparable diagnostic accuracy to established biomarkers like miR-21 (AUC: 0.88–0.95) and TTF-1 mRNA (AUC: 0.87–0.88) in distinguishing NSCLC from healthy controls (AUC: 0.92), it offers three distinct clinical advantages: dual diagnostic utility for both NSCLC detection and benign/malignant PN differentiation (AUC: 0.856, even in nodules ≤ 10 mm), addressing a critical unmet need in small nodule management; subtype discrimination of NSCLC pathological subtypes, surpassing single-marker assays; and enhanced methodological rigor, including systematic reliability testing demonstrating < 5% inter-batch variability, thereby ensuring reproducibility.

The lncRNAs linc01125, HNF1A-AS1, MIR100HG, linc01160, and ZNRF3-AS1 were deliberately chosen as constituents of the diagnostic lncRNA panel for NSCLC. The pivotal roles of linc01125 as a tumor suppressor, and the oncogenic functions exhibited by HNF1A-AS1 and MIR100HG have been well-documented in previously published studies [18, 25–30]. Particularly in the context of lung cancer, linc01125 orchestrates inhibitory effects on cancer progression and metastasis by up-regulating tumor necrosis factor alpha-induced protein 3 expression through sequestration of miR-19b-3p [18]. Conversely,

HNF1A-AS1 facilitates cellular proliferation, invasion, and radiotherapy resistance by orchestrating the modulation of several microRNAs [31–33]. MIR100HG has been implicated in enhancing metastatic potential in NSCLC through the activation of glycolytic pathways [30]. The functional implications of linc01160 and ZNRF3-AS1 in lung cancer remain largely unexplored. Existing literature posits that LINC01160 is upregulated in nasopharyngeal carcinoma, leading to the promotion of a malignant cell phenotype [34]. On the other hand, ZNRF3-AS1's potential association with ivermectin sensitivity in ovarian cancer has been suggested [35]. Through rigorous bioinformatic analyses, it has been revealed that decreased expression of linc01160 and elevated levels of ZNRF3-AS1 are significantly correlated with adverse clinical outcomes in NSCLC. Furthermore, the targets of these lncRNAs were found to be enriched in cancer-related pathways such as the Hippo signaling pathway, Peroxisome, and Carbon metabolism. These compelling findings underscore the plausible utility of these lncRNAs as promising candidates for diagnostic biomarker development in the context of NSCLC.

An unexpected observation gleaned from our investigation unveils an overarching negative correlation in the expression levels of lncRNAs between plasma Epexo and primary cancer tissues. Intriguingly, only the associations of MIR100HG and HNF1A-AS1 exhibited statistical significance. This implies that while the expression of lncRNAs in plasma Epexo may be under the influence of the originating cancer cells, their abundance is more likely subject to modulation through mechanisms such as transport.

Despite the insights provided by the current study, several notable limitations must be acknowledged. Firstly, the reliance on a limited cohort of only five paired samples for lncRNA expression profiling raises the possibility of missing key lncRNAs that may possess significant diagnostic relevance. Secondly, the lack of random participant recruitment, primarily due to resource constraints, may introduce an element of selection bias, potentially compromising the robustness of our findings. Lastly, our current cohort lacks ethnic diversity, necessitating validation in heterogeneous populations. Additionally, the cost-effectiveness and clinical scalability of the 5-lncRNA panel require optimization, particularly regarding exosome isolation complexity and the development of simplified, cost-effective assays.

In conclusion, our research has established a reliable and robust 5-lncRNA panel in plasma Epexo for early detection of NSCLC. Furthermore, this panel exhibits the potential to classify different histological types of NSCLC. To enhance the practical utility of the 5-lncRNA panel, additional validation in a larger, randomly recruited cohort is essential.
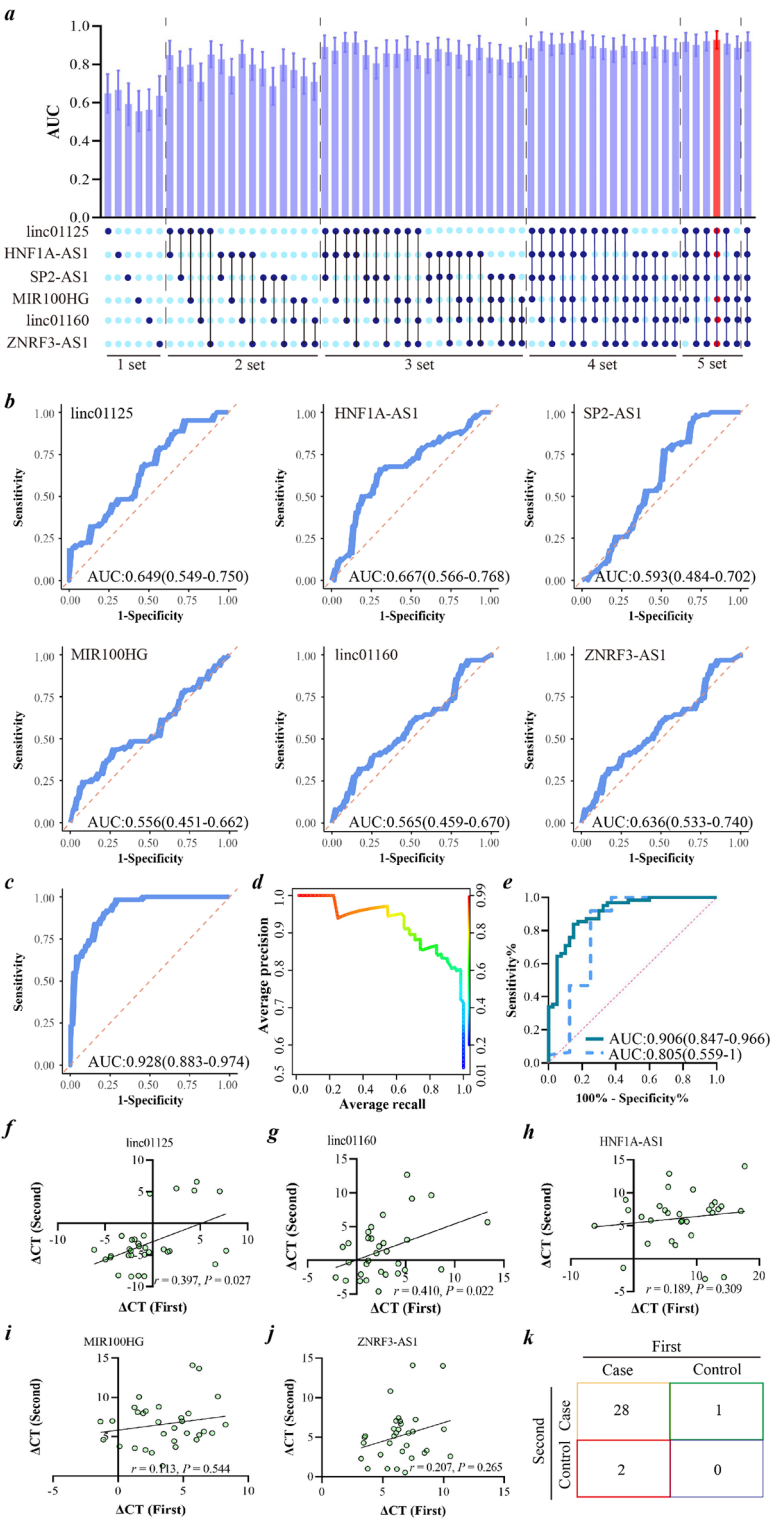
Fig. 4 Assessment of diagnostic efficacy using selected lncRNAs and their combinations on NSCLC. (**a**) Area under the Receiver Operating Characteristics (ROC) curve (AUC) values for six selected lncRNAs and all possible combinations in distinguishing NSCLC patients from healthy controls. The optimal diagnostic combination, identified based on maximizing the AUC values, is highlighted in red. (**b**) Individual ROC curves for each lncRNA. (**c-d**) ROC curve (**c**) and precision-recall curve (**d**) for the best diagnostic combination. (**e**) Subgroup ROC curves for the most effective diagnostic combination in differentiating patients with local stage (LS) or regional or distant stage (RDS) from healthy controls. (**f-j**) Scatter plots illustrating the correlation between the results of repeated measurements of linc01125 (**f**), linc01160 (**g**), HNF1A-AS1 (**h**), MIR100HG (**i**), and ZNRF3-AS1. (**j**) Expression levels in plasma-derived Epexo from 31 NSCLC patients. Spearman rank correlation test was used to calculate the correlation coefficient (*r*) and *P* value. (**k**) Comparative analysis of model diagnostic outcomes from two sets of repeated testing in 31 NSCLC patients
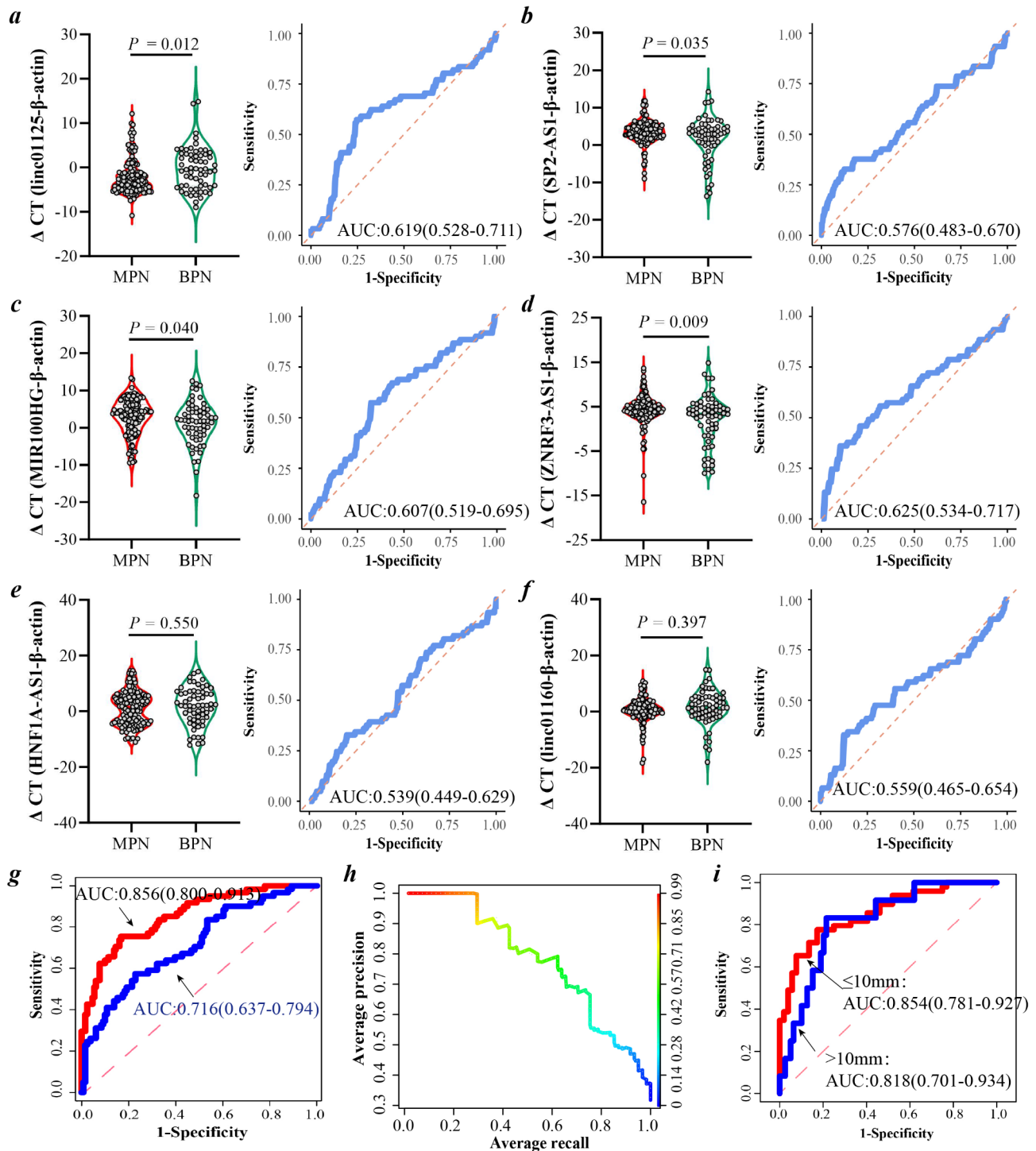
Fig. 5 Evaluation of Discriminatory Efficacy Using Selected lncRNAs and Their Combinations on Malignant Pulmonary Nodules (MPN) and Benign Pulmonary Nodules (BPN).**(a-f)**Depiction of the expression profiles of linc01125 **(a)**, SP2-AS1 **(b)**, **MIR100HG (c)**, ZNRF3-AS1 **(d)**, HNF1A-AS1 **(e)**, and linc01160 **(f)** in plasma-derived Epexo from patients diagnosed with MPN and BPN. Corresponding ROC curves for these lncRNAs are displayed alongside the bar chart illustrating their expression levels. The statistical significance was determined using unpaired t-tests with Welch's correction. **(g)**ROC curves illustrating the diagnostic performance of the best combination with (red line) or without consideration of nodule size (blue line). **(h)**Precision-recall curve demonstrating the diagnostic accuracy of the best combination in conjunction with nodule size. **(i)**Subgroup ROC curves showcasing the effectiveness of the optimal diagnostic combination in distinguishing MLN patients from BLN controls, categorized based on nodule size (≤10mm and > 10mm).
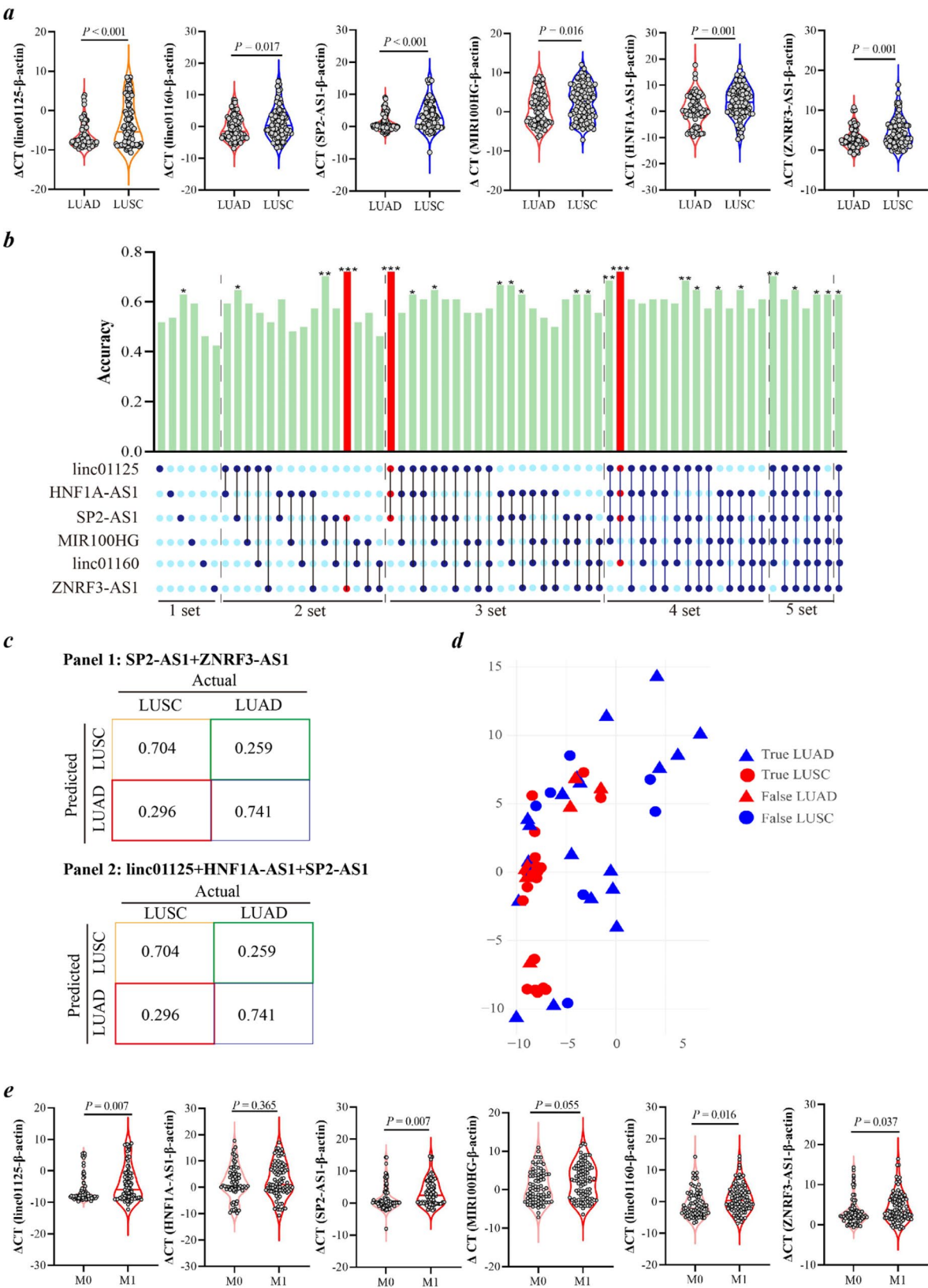
Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Epexo lncRNAs as a Classifier for Discrimination of Pathological Types and M Staging. (a) Depiction of the expression profiles of six selected lncRNAs in plasma-derived Epexo from patients diagnosed with lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC). **(b)**Accuracy, represented as AUC, of the six selected lncRNAs and all possible combinations in distinguishing between LUAD and LUSC patients. (*$P<0.05$, **$P<0.01$, and ***$P<0.001$). **(c)**Comparative analysis of model diagnostic outcomes and actual outcomes using two classifiers: one comprising SP2-AS1 and ZNRF3-AS1, and the other comprising linc01125, HNF1A-AS1, and SP2-AS1. **(d)**Scatter plot displaying the distribution consistency between model diagnosis and actual diagnosis. Blue triangles and red circles represent true LUAD and LUSC, indicating correct diagnostic outcomes by the model, while red triangles and blue circles represent false LUAD and LUSC, implying erroneous diagnostic outcomes by the model. **(e)**Expression patterns of the six selected lncRNAs in plasma-derived Epexo among patients with positive (M0) and negative (M1) distant metastasis. *P* value was obtained using unpaired t-tests with Welch's correction.
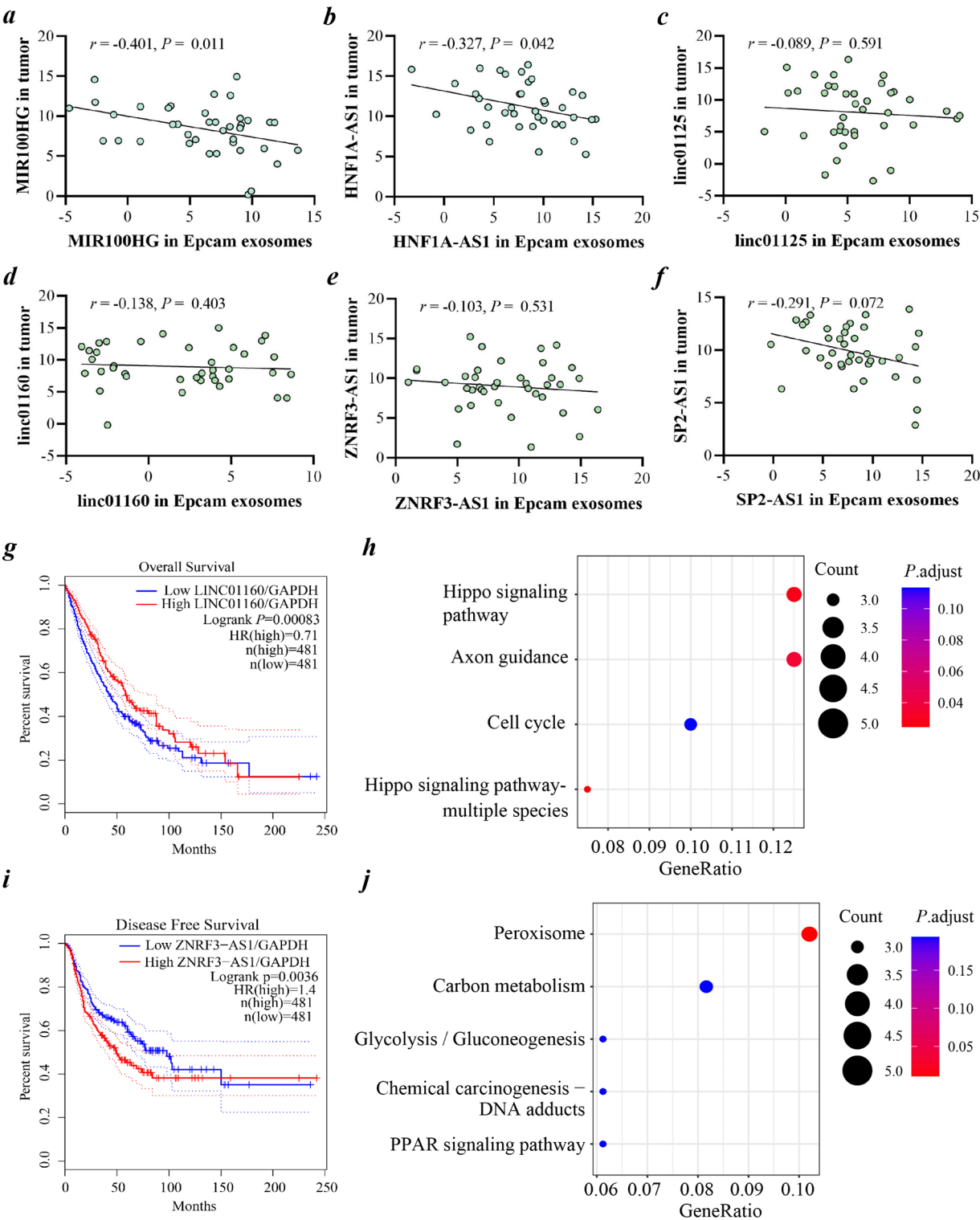
Fig. 7 Correlation of selected lncRNAs in plasma-derived Epexo and cancerous tissues, and functional evaluation. (**a-f**) Scatter plots illustrating the correlation between MIR100HG (**a**), HNF1A-AS1 (**b**), linc01125 (**c**), linc01160 (**d**), ZNRF3-AS1 (**e**), and SP2-AS1 (**f**) expression levels in plasma-derived Epexo and cancerous tissues from 39 NSCLC patients. Spearman rank correlation test was conducted to determine the correlation coefficient (*r*) and associated *P* value. (**g**) Survival curve depicting the impact of linc01160 on NSCLC overall survival, as analyzed using the GEPIA database. (**h**) KEGG analysis performed on the top 200 genes displaying a significant correlation with linc01160. (**i**) Survival curve illustrating the effect of ZNRF3-AS1 on NSCLC disease-free survival, based on analysis from the GEPIA database. (**j**) KEGG analysis conducted on the top 200 genes demonstrating a significant correlation with ZNRF3-AS1

Liu *et al. Respiratory Research*   (2025) 26:144

Page 15 of 16

Abbreviations

NSCLC   Non-small cell lung cancer
Epexo   EpCAM-specific exosomes
LDCT   Low-dose computed tomography
lncRNA   Long non-coding RNA
PN   Pulmonary nodules
AUC   Area under the ROC curve

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-025-03220-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

LYL: Conceptualization, Data curation, Formal analysis. DLL: Data curation, Investigation, Writing - review & editing. AMZ: Data curation, validation, investigation. JCL: Methodology, Resources. JJZ: Investigation, Resources. GTH: Investigation, Resources. ZTH: Investigation, Resources. ZLZ: validation, Writing - review & editing. YBD: Resources, validation, Writing - review & editing. LY: Conceptualization, Methodology, Resources, Writing - original draft, Project administration, Funding acquisition.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

All subjects provided informed consent for the use of their plasma samples and clinical information. This study was approved by the institutional review board of Guangzhou Medical University.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹The Key Laboratory of Advanced Interdisciplinary Studies, The First Affiliated Hospital, Guangzhou Medical University, Guangzhou 510120, China
²Key Laboratory of Clinical Molecular Diagnosis and Research for High Incidence Diseases in Western Guangxi Universities, The Affiliated Hospital of Youjiang Medical University for Nationalities, Baise 533000, China
³School of Public health, Guangzhou Medical University, Guangzhou 511436, China
⁴Department of Oncology, Guangzhou Chest Hospital, Guangzhou 510095, China
⁵Physical Examination Center, Affiliated First People's Hospital, Guangzhou Medical university, Guangzhou 511468, China
⁶Department of Health Management Center, The Third Affiliated Hospital, Southern Medical University, Guangzhou 510630, China

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71:209–49.
2. Global Burden of Disease, Cancer C, Fitzmaurice C, Abate D, et al. Global, regional, and National cancer incidence, mortality, years of life lost, years lived with disability, and Disability-Adjusted life-Years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. JAMA Oncol. 2019;5:1749–68.
3. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med. 2020;382:503–13.
4. Li N, Tan F, Chen W, et al. One-off low-dose CT for lung cancer screening in China: a multicentre, population-based, prospective cohort study. Lancet Respiratory Med. 2022;10:378–91.
5. Becker N, Motsch E, Trotter A, et al. Lung cancer mortality reduction by LDCT screening-Results from the randomized German LUSI trial. Int J Cancer. 2020;146:1503–13.
6. Jonas DE, Reuland DS, Reddy SM, et al. Screening for lung cancer with Low-Dose computed tomography: updated evidence report and systematic review for the US preventive services task force. JAMA. 2021;325:971–87.
7. Li MY, Liu LZ, Dong M. Progress on pivotal role and application of exosome in lung cancer carcinogenesis, diagnosis, therapy and prognosis. Mol Cancer. 2021;20:22.
8. Yoh KE, Lowe CJ, Mahajan S, et al. Enrichment of circulating tumor-derived extracellular vesicles from human plasma. J Immunol Methods. 2021;490:112936.
9. Shen X, Yang Y, Chen Y, et al. Evaluation of EpCAM-specific Exosomal LncRNAs as potential diagnostic biomarkers for lung cancer using droplet digital PCR. J Mol Med. 2022;100:87–100.
10. Jin X, Chen Y, Chen H, et al. Evaluation of Tumor-Derived Exosomal MiRNA as potential diagnostic biomarkers for Early-Stage Non-Small cell lung cancer using Next-Generation sequencing. Clin cancer Research: Official J Am Association Cancer Res. 2017;23:5311–9.
11. Yang G, Wang T, Qu X, et al. Exosomal miR-21/Let-7a ratio distinguishes non-small cell lung cancer from benign pulmonary diseases. Asia-Pac J Clin Oncol. 2020;16:280–6.
12. Liu C, Kannisto E, Yu G, et al. Non-invasive detection of Exosomal MicroRNAs via tethered cationic lipoplex nanoparticles (tCLN) biochip for lung cancer early detection. Front Genet. 2020;11:258.
13. Dejima H, Iinuma H, Kanaoka R, et al. Exosomal MicroRNA in plasma as a non-invasive biomarker for the recurrence of non-small cell lung cancer. Oncol Lett. 2017;13:1256–63.
14. Chen LL, Kim VN. Small and long non-coding RNAs: past, present, and future. Cell. 2024;187:6451–85.
15. Coan M, Haefliger S, Ounzain S, et al. Targeting and engineering long non-coding RNAs for cancer therapy. Nat Rev Genet. 2024;25:578–95.
16. Li W, Li H, Zhang L, et al. Long non-coding RNA LINC00672 contributes to p53 protein-mediated gene suppression and promotes endometrial cancer chemosensitivity. J Biol Chem. 2017;292:5801–11.
17. Gong W, Yang L, Wang Y et al. Analysis of Survival-Related LncRNA landscape identifies A role for LINC01537 in energy metabolism and lung cancer progression. Int J Mol Sci. 2019; 20:3713.
18. Xian J, Zeng Y, Chen S, et al. Discovery of a novel linc01125 isoform in serum exosomes as a promising biomarker for NSCLC diagnosis and survival assessment. Carcinogenesis. 2021;42:831–41.
19. Xian J, Su W, Liu L, et al. Identification of three circular RNA cargoes in serum exosomes as diagnostic biomarkers of Non-Small-Cell lung cancer in the Chinese population. J Mol Diagnostics: JMD. 2020;22:1096–108.
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–45.
21. Liu X, Wu F, Pan W, et al. Tumor-associated exosomes in cancer progression and therapeutic targets. MedComm. 2024;5:e709.
22. Iaccino E, Mimmi S, Dattilo V, et al. Monitoring multiple myeloma by idiotype-specific peptide binders of tumor-derived exosomes. Mol Cancer. 2017;16:159.
23. Castillo J, Bernard V, San Lucas FA, et al. Surfaceome profiling enables isolation of cancer-specific Exosomal cargo in liquid biopsies from pancreatic cancer patients. Annals Oncology: Official J Eur Soc Med Oncol. 2018;29:223–9.

24. Yang Y, Kannisto E, Yu G, et al. An Immuno-Biochip selectively captures Tumor-Derived exosomes and detects Exosomal RNAs for cancer diagnosis. ACS Appl Mater Interfaces. 2018;10:43375–86.

25. Wan W, Hou Y, Wang K, et al. The LXR-623-induced long non-coding RNA LINC01125 suppresses the proliferation of breast cancer cells via PTEN/AKT/p53 signaling pathway. Cell Death Dis. 2019;10:248.

26. He T, Xia H, Chen B, et al. m6A writer METTL3-Mediated LncRNA LINC01125 prevents the malignancy of papillary thyroid cancer. Crit Rev Immunol. 2023;43:43–53.

27. Zhang Y, Shi J, Luo J, et al. Regulatory mechanisms and potential medical applications of HNF1A-AS1 in cancers. Am J Translational Res. 2022;14:4154–68.

28. Lin Z, Wu Y, Xu Y, et al. Mesenchymal stem cell-derived exosomes in cancer therapy resistance: recent advances and therapeutic potential. Mol Cancer. 2022;21:179.

29. Lu Y, Zhao X, Liu Q, et al. LncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/beta-catenin signaling. Nat Med. 2017;23:1331–41.

30. Shi L, Li B, Zhang Y, et al. Exosomal LncRNA Mir100hg derived from cancer stem cells enhance Glycolysis and promote metastasis of lung adenocarcinoma through mircroRNA-15a-5p/31-5p. Cell Communication Signaling: CCS. 2023;21:248.

31. Liu L, Chen Y, Li Q, et al. LncRNA HNF1A-AS1 modulates non-small cell lung cancer progression by targeting miR-149-5p/Cdk6. J Cell Biochem. 2019;120:18736–50.

32. Zhang G, An X, Zhao H, et al. Long non-coding RNA HNF1A-AS1 promotes cell proliferation and invasion via regulating miR-17-5p in non-small cell lung cancer. Volume 98. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie; 2018. pp. 594–9.

33. Wang Z, Liu L, Du Y, et al. The HNF1A-AS1/miR-92a-3p axis affects the radio-sensitivity of non-small cell lung cancer by competitively regulating the JNK pathway. Cell Biol Toxicol. 2021;37:715–29.

34. Ai J, Tan G, Wang T, et al. Transcription factor STAT1 promotes the proliferation, migration and invasion of nasopharyngeal carcinoma cells by upregulating LINC01160. Future Oncol. 2021;17:57–69.

35. Li N, Zhan X. Anti-parasite drug Ivermectin can suppress ovarian cancer by regulating lncRNA-EIF4A3-mRNA axes. EPMA J. 2020;11:289–309.

Publisher's note