

Identification and Evaluation of Circulating Small Extracellular Vesicle Micrnas as Diagnostic Biomarkers for Patients with Indeterminate Pulmonary Nodules

Di Zheng

Tongji University Affiliated Shanghai Pulmonary Hospital

Yuming Zhu

Tongji University Affiliated Shanghai Pulmonary Hospital

Jiyang Zhang

3D Medicines Inc.

Wei Zhang

3D Medicines Inc.

Huizhen Wang

3D Medicines Inc.

Hao Chen

3D Medicines Inc.

Chunyan Wu

Tongji University Affiliated Shanghai Pulmonary Hospital

Jian Ni

Tongji University Affiliated Shanghai Pulmonary Hospital

Xiaoya Xu

3D Medicines Inc.

Baoning Nian

3D Medicines Inc.

Sheng Chen

3D Medicines Inc.

Beibei Wang

3D Medicines Inc.

Xiaofang Li

3D Medicines Inc.

Yanan Zhang

3D Medicines Inc.

Jiatao Zhang

Guangdong Provincial People's Hospital

Wenzhao Zhong

Guangdong Provincial People's Hospital

Lei Xiong

3D Medicines Inc.

Fugen Li

3D Medicines Inc.

Dadong Zhang

3D Medicines Inc.

Jianfang Xu

Tongji University Affiliated Shanghai Pulmonary Hospital

Gening Jiang (✉ geningjiang@tongji.edu.cn)

Tongji University Affiliated Shanghai Pulmonary Hospital <https://orcid.org/0000-0002-1351-3409>

Research Article

Keywords: indeterminate pulmonary nodule, low-dose computed tomography, small extracellular vesicle, microRNA, small RNA sequencing

Posted Date: December 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1157761/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Nanobiotechnology on April 2nd, 2022. See the published version at <https://doi.org/10.1186/s12951-022-01366-0>.

Abstract

Background: The identification of indeterminate pulmonary nodules (IPNs) following a low-dose computed tomography is a major challenge for early diagnosis of lung cancer. The inadequate assessment of IPNs' malignancy risk results in a large number of unnecessary surgeries or an increased risk of cancer metastases. However, limited studies on non-invasive diagnosis of IPNs have been reported.

Methods: In this study, we identified and evaluated the diagnostic value of circulating sEV miRNAs in patients with IPNs that had been newly detected using LDCT scanning and were scheduled for surgery. Out of 459 recruited patients, 109 eligible patients with IPNs were enrolled in the training cohort (n = 47) and the test cohort (n = 62). An external cohort (n=99) was used for validation. MiRNAs were extracted from plasma sEVs, and assessed using Small RNA sequencing. 490 lung adenocarcinoma samples and follow-up data were used to investigate the role of miRNAs in overall survival.

Results: A circulating sEV miRNA (CirsEV-miR) model was constructed from five differentially expressed miRNAs (DEMs), showing 0.920 AUC in the training cohort (n = 47), and further identified in the test cohort (n = 62) and in an external validation cohort (n = 99). Among five DEMs of the CirsEV-miR model, miR-101-3p and miR-150-5p were significantly associated with better overall survival (p = 0.0001 and p = 0.0069). The CirsEV-miR scores were calculated, which significantly correlated with IPNs diameters (p < 0.05), and were able to discriminate between benign and malignant PNs (diameter \leq 1 cm). The expression patterns of sEV miRNAs in the benign, adenocarcinoma in situ/minimally invasive adenocarcinoma, and invasive adenocarcinoma subgroups were found to gradually change with the increase in aggressiveness for the first time. Among all DEMs of the three subgroups, five miRNAs (miR-30c-5p, miR-30e-5p, miR-500a-3p, miR-125a-5p, and miR-99a-5p) were also significantly associated with overall survival of lung adenocarcinoma patients.

Conclusions: Our results indicate that the CirsEV-miR model could help distinguish between benign and malignant PNs, providing insights into the feasibility of circulating sEV miRNAs in diagnostic biomarker development.

Trial registration: Chinese Clinical Trials, ChiCTR1800019877. Registered 05 December 2018, <https://www.chictr.org.cn/showproj.aspx?proj=31346>.

Background

Lung cancer is the most frequent malignancy with the highest morbidity and mortality worldwide¹. In the U.S. National Lung Cancer Screening Test (NLST) trial², low-dose computed tomography (LDCT) has been proven to be a highly sensitive method to detect early stage lung cancer in patients with a smoking history. LDCT scanning has dramatically improved the ability to detect pulmonary nodules (PNs); however, a total of 96.4% of false positive screening results mandates improvements in the management of indeterminate pulmonary nodules (IPNs), which are difficult to diagnose. The inadequate assessment

of IPNs' malignancy risk is a major cause of misdiagnosis and mistreatment. Benign PNs require no surgical resection, however, the rates of benign PNs undergoing surgical resection have been reported to range from 1–30% in suspected lung cancer cases³. The invasive thoracic procedure of unnecessary operations performed in patients with benign PNs results in unnecessary costs, societal burden, and morbidity and mortality risk to the patient, with no therapeutic benefit^{4,5}. Hence, a non-invasive auxiliary diagnostic test for LDCT to improve its ability of distinguishing between benign and malignant PNs among IPNs is urgently needed.

Extensive efforts using blood biomarkers, such as DNA, RNA, and proteins to distinguish malignant from benign PNs have yielded novel insights into lung cancer diagnosis⁶. Small extracellular vesicles (sEVs), secreted by a variety of cells into the blood, contain bioactive molecules such as proteins, lipids, and nucleic acids that can mirror the cellular origin and the physiological state, and these molecules are attractive potential biomarkers representing the “fingerprint” or “signature” of the donor cell⁷. Moreover, the membranous structure of the sEVs protects the luminal contents, avoiding degradation by extracellular enzymes. The remarkable stability and activity strengthen the potential of circulating sEVs to be reservoirs for biomarker development^{8,9}. MicroRNAs (miRNAs) have been found to be the most abundant species among plasma-derived sEVs RNAs¹⁰, and are also notably stable under different storage conditions¹¹. Lung cancer cells secrete more sEVs into blood than normal tissue cells¹², and miRNAs derived from sEVs of lung cancer patients have also been found to be significantly different from those of healthy people¹³. Some studies have focused on serum/plasma sEV miRNAs as potential biomarkers of lung cancer¹⁴. However, most of these studies aimed to discriminate lung cancer patients from healthy controls, but without detailed distinction between benign and malignant PNs among IPNs, which are quite difficult to distinguish for early diagnosis of lung cancer in clinical practice.

In the present study, we assessed the expression levels of circulating sEV miRNAs using small RNA sequencing to detect the differences between patients with benign PNs and patients with malignant PNs (early- to mid-stage lung adenocarcinoma). We then developed a CirsEV-miR model to differentiate between benign and malignant PNs in a training cohort (n = 47) and further confirmed the model in a test cohort (n = 62) and in an external validation cohort (n = 99). Furthermore, we calculated CirsEV-miR scores and for the first time explored the relationship between CirsEV-miR scores and clinical characteristics, including the diameter of IPNs and the PN aggressiveness in the benign, adenocarcinoma in situ (AIS)/minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma. We also found that the circulating sEV miRNA signature was able to discriminate between benign and malignant PNs with the diameter ≤ 1 cm, which are otherwise difficult to distinguish in clinical practice. In total, we analyzed circulating sEV miRNAs of 208 patients with IPNs, providing the largest sample size among the available studies. Additionally, 490 lung adenocarcinoma samples and follow-up data from The Cancer Genome Atlas (TCGA) were used to investigate the role of miRNAs in overall survival. The development of such non-invasive diagnostic test, *i.e.*, the CirsEV-miR model, may complement the highly sensitive but insufficiently specific LDCT and be integrated into the diagnostic algorithm to achieve higher diagnostic accuracy for patients with IPNs.

Results

Participants and clinical characteristics

The workflow of our study is illustrated in Figure 1. To explore circulating sEV miRNAs as diagnostic biomarkers in patients with lung IPNs, a total of 199 and 260 patients were recruited in the training phase and the test phase, respectively. After the elimination of hemolysis, non-lung adenocarcinoma (non-LUAD), failure of sequencing library construction, and limitation of the ratio of benign and malignant PNs, a total of 109 patients were included in two independent cohorts (**Additional file 1: Figure S1**). The training cohort consisted of 47 patients with IPNs, including 17 benign PNs (benign group) and 30 malignant PNs (malignant group) identified using pathological diagnosis; the test cohort consisted of 62 patients with IPNs, including 24 benign PNs and 38 malignant PNs. In addition, an external validation cohort (n = 99), including 20 patients with benign PNs and 79 patients with malignant PNs (Figure 1 and Table 1), was used to validate the model. More than 90% of LUAD patients (92.6%, 63/68) were in the early stage (no lymph node metastasis), and 33.0% (36/109) of patients had IPNs with a diameter less than 1 cm (Table 1). The demographic and clinicopathologic characteristics of the patients are shown in Table 1. Benign PNs included cases of atypical adenomatous hyperplasia (AAH), fibrosis, granulomas, hamartoma, organizing pneumonias (OPs), cyst, and other benign subtypes, while malignant PNs included AIS, MIA, and invasive adenocarcinoma. The pathology subtype composition of the three cohorts is shown in (**Additional file 1: Figure S2**). The representative imaging features of benign and malignant PNs are shown in (**Additional file 1: Figure S3**).

Table 1
The demographic and clinicopathologic characteristics of the patients in three cohorts.

Characteristic	Training cohort (N= 47)	Testing cohort (N= 62)	External cohort (N= 99)
Categories - No. (%)			
Benign	17 (36.2)	24 (38.7)	20 (20.2)
Malignant	30 (63.8)	38 (61.3)	79 (79.8)
Age, mean (SD)	58.7 (12.5)	58.1 (10.7)	57.9 (11.9)
Age, No. (%)			
< 60	24 (51.1)	28 (45.2)	49 (49.5)
≥ 60	23 (48.9)	34 (54.8)	50 (50.5)
Gender, No. (%)			
Female	20 (42.6)	37 (59.7)	55 (55.6)
Male	27 (57.4)	25 (40.3)	44 (44.4)
Smoking status, No. (%)			NA
Yes	12 (25.5)	6 (9.7)	
No	35 (74.5)	56 (90.3)	
Pathology, No. (%)			
AIS	3 (6.4)	6 (9.7)	10
MIA	10 (21.3)	11 (17.7)	5
Invasive	17 (36.2)	21 (33.9)	66
AAH	2 (4.3)	2 (3.2)	0
Fibrosis	2 (4.3)	4 (6.5)	0
Granulomas	4 (8.5)	3 (4.8)	0
Hamartoma	2 (4.3)	3 (4.8)	3
OP	4 (8.5)	4 (6.5)	0
Other benign subtypes	3 (6.4)	8 (12.9)	15
Nodule diameter (cm), No. (%)			

AIS, Adenocarcinoma in situ, MIA, minimally invasive adenocarcinoma; AAH, atypical adenomatous hyperplasia, OP, organizing pneumonias; NA, not available.

Characteristic	Training cohort (N = 47)	Testing cohort (N = 62)	External cohort (N = 99)
≤1	12 (25.5)	24 (38.7)	21 (21.2)
>1	35 (74.5)	38 (61.3)	78 (78.8)
Malignant stages, No. (%)			NA
0	3 (10.0)	6 (15.8)	
IA1	12 (40.0)	13 (34.2)	
IA2	5 (16.7)	11 (28.9)	
IA3	6 (20.0)	4 (10.5)	
IB	1 (3.3)	0	
IIA	2 (6.7)	0	
IIB	0	1 (2.6)	
IIIA	1 (3.3)	3 (7.9)	
AIS, Adenocarcinoma in situ, MIA, minimally invasive adenocarcinoma; AAH, atypical adenomatous hyperplasia, OP, organizing pneumonias; NA, not available.			

Circulating sEV characterization

Circulating sEVs were successfully isolated from patient plasma samples and characterized using western blot (WB) analysis, nanoparticle tracking analysis (NTA), and transmission electron microscopy (TEM). In accordance with the Minimal Information for Studies of Extracellular Vesicles (MISEV) 2018¹⁵, several protein markers were evaluated using WB in eight representative sEV samples from the patients with benign and malignant PNs. The expression levels of TSG101, CD63, CD9, and Syntenin were detected in the eight sEV samples, while a negative marker, Calnexin, was absent in all eight sEV samples (Figure 2a). Furthermore, the majority of the isolated sEVs were around 100 nm in diameter, which is the typical size of sEVs (Figure 2b). The TEM result from a representative sample showed that the isolated sEVs were cup-shaped (Figure 2c), which is the typical morphology of sEVs. Furthermore, we also detected sEV transmembrane proteins (CD63, CD81, CD9) using ExoView platform. As shown in Figure 2d, CD63, CD81, and CD9 were all detected in plasma sEVs of patients with benign or malignant PNs.

Construction and validation of a circulating sEV miRNA (CirsEV-miR) model for IPN diagnosis

We then extracted the RNA from sEVs and performed small RNA sequencing. Raw data of small RNA sequencing were filtered and normalized. Unsupervised hierarchical clustering based on the expression

levels of circulating sEV miRNAs categorized the samples in a similar pattern as the clinicopathologic classifications did, both in the training cohort (Figure 3a) and in the test cohort (**Additional file 1: Figure S4a**), thereby suggesting the feasibility of predicting clinicopathologic classifications using circulating sEV miRNA expression. Six DEMs between benign and malignant PNs were identified in the training cohort (Figure 3b), and these DEMs were selected to construct a model for distinguishing benign PNs from malignant PNs. To verify the potential of miRNAs as a suitable classifier, receiver operating characteristic (ROC) analysis was performed, and the area under curve (AUC) was calculated (Figure 3c, **Additional file 1: Table S1, Additional file 1: Table S2**). Let-7b-3p exhibited an AUC of 0.875, which was the best performance of an individual miRNA. Next, we integrated two or more miRNAs to further improve the performance. Integration of let-7b-3p and miR-125b-5p led to a slightly improved AUC of 0.886; integrating three miRNAs achieved an AUC of 0.892 (let-7b-3p, miR-125b-5p, and miR-197-3p); and integrating four miRNAs achieved an AUC of 0.882 (let-7b-3p, miR-125b-5p, miR-197-3p, and miR-150-5p). Integration of five miRNAs achieved an even higher AUC of 0.904 (let-7b-3p, miR-125b-5p, miR-197-3p, miR-150-5p, and miR-3168), while the AUC of the integration of all six miRNAs dropped to 0.794. We then used LASSO-penalized regression to develop a classifier of five miRNAs (let-7b-3p, miR-125b-5p, miR-150-5p, miR-101-3p, and miR-3168), called the CirsEV-miR model, which exhibited an AUC of 0.920 in the training cohort (Figure 3d, **Additional file 1: Table S2, Additional file 1: Table S3**), which was the highest AUC among all classifiers. The sensitivity reached 0.900, and the specificity was 0.882. Based on DEMs, CirsEV-miR scores were also generated through LASSO analysis. CirsEV-miR scores of malignant PNs were significantly higher than those of benign PNs in the training cohort (Figure 3e, $p < 0.0001$), suggesting that CirsEV-miR scores increase with malignancy. The CirsEV-miR model was further confirmed in the test cohort, showing an AUC of 0.763 (**Additional file 1: Figure S4b**) and significant differences in CirsEV-miR scores between benign and malignant IPNs (**Additional file 1: Figure S4c**, $p=0.0004$). The CirsEV-miR model was also validated in the external cohort consisting of 20 patients with benign PNs and 79 patients with malignant PNs; the AUC in that cohort was 0.781 (**Additional file 1: Figure S4d**). These results revealed that the CirsEV-miR model may be a new promising approach to assist in the differential diagnosis of lung IPNs.

Biological function enrichment and overall survival analysis of five DEMs in the CirsEV-miR model

To evaluate the potential functions of the five miRNAs used in the CirsEV-miR model, we performed Gene Ontology (GO) and pathway analysis of their target genes (**Additional file 1: Table S4**). We found that 2238 mRNAs were targeted by the five miRNAs through the analysis of the miRNA target database. The target genes were enriched in biological processes (BPs), such as cellular nitrogen compound metabolic process, biosynthetic process, and cellular protein modification processes (Figure 4a), and in molecular functions (MFs), such as ion binding, enzyme binding, and RNA binding (Figure 4b). In addition, the target genes were significantly enriched in pathways related to tumorigenesis and progression processes, such as MAPK, TGF β , Hippo, p53 signaling pathways, cell cycle, and adherence junction processes (Figure 4c, 4d). Moreover, 490 lung adenocarcinoma samples and their follow-up data from the TCGA showed that patients with high expression of miR-101-3p or miR-150-5p in tumor tissue samples had better overall

survival (Figure 4e, 4f, $p = 0.0001$; $p = 0.0069$). However, the other three miRNAs, let-7b-3p, miR-125b-5p, and miR-3168, were not significantly associated with overall survival (data not shown). The functional enrichment analysis indicated that five miRNAs were involved in tumorigenesis and progression of lung cancer.

The signature of circulating sEV miRNAs discriminated between benign and malignant PNs with diameter ≤ 1 cm.

The size of IPNs is a key factor that is associated with their malignant potential and patients' long term survival^{16,17}. In clinical practice, IPNs with a diameter > 1 cm are thought to need resection, whereas IPNs with a diameter ≤ 1 cm require repeated scans and a long-term follow-up¹⁸. Thus, we also analyzed the correlation between the nodule size and risk score of the CirsEV-miR model. All 109 eligible samples were recategorized into two subgroups: the IPNs with diameter ≤ 1 cm and IPNs with diameter > 1 cm, and re-ranked according to the CirsEV-miR scores. The CirsEV-miR scores were significantly higher in the subgroup of IPNs with diameter > 1 cm (Figure 5a, $p = 0.0170$), suggesting that the CirsEV-miR scores remarkably increase with the diameter of IPNs. Moreover, among the patients with small IPNs (diameter ≤ 1 cm), malignant PNs also exhibited higher CirsEV-miR scores than benign PNs (Figure 5b, $p = 0.0088$). The AUC of the CirsEV-miR model in patients with small IPNs (diameter ≤ 1 cm) was 0.767 in our cohort (**Additional file 1: Figure S5a**), and 0.721 in the external validation cohort (**Additional file 1: Figure S5b**). Moreover, we achieved the diagnostic specificity of 91.7% for the IPNs with diameter ≤ 1 cm in our cohort, and specificity of 75% in the external validation cohort (Figure 5c). Among the IPNs with diameter ≤ 1 cm, benign and malignant subgroups showed different miRNA expression patterns (Figure 5d). Twelve DEMs between benign and malignant PNs with diameter ≤ 1 cm were identified (Figure 5e), and eight of them were involved in lung cancer tumorigenesis and progression, suggesting that even small IPNs were different from benign PNs. These results suggest that even in IPNs with diameter ≤ 1 cm, the circulating sEV miRNA signature can separate benign PNs from malignant PNs.

Circulating sEV miRNA analysis reveals that AIS/MIA can alleviate the differentiation between benign and malignant PNs

AIS and MIA represent early-stage lung cancer, and both AIS and MIA patients possess superior prognoses to invasive adenocarcinoma^{19,20}. Therefore, we speculated that the possible reason for the relatively low sensitivity of the current model was the misprediction of the clinicopathologically equivocal AIS/MIA samples. To test this hypothesis, 109 eligible samples were categorized into three subgroups: benign (41, 37.6%), AIS/MIA (30, 27.5%), and invasive adenocarcinoma (38, 34.9%). The proportions of AIS/MIA samples in the training and the test cohorts were 27.7% and 27.4%, respectively (Table 1). For all 109 samples, we calculated the CirsEV-miR scores and ranked each subgroup. The CirsEV-miR scores in

the benign, AIS/MIA and invasive adenocarcinoma subgroups gradually changed with the increase in aggressiveness (Figure 6a). We also tested the miRNA expression levels; we found that the expression pattern of the AIS/MIA subgroup was intermediate between benign and invasive adenocarcinoma subgroups (Figure 6b). The expression levels of the five DEMs in the AIS/MIA subgroup were located between the benign and the invasive adenocarcinoma subgroups (Figure 6c). We further compared all DEMs between the benign, AIS/MIA, and invasive adenocarcinoma subgroups (Figure 6d, 6e, **Additional file 1: Table S5, Additional file 1: Table S6**). We only found two overlapping upregulated DEMs (let-7b-3p and miR-125b-5p), which were also elements of our CirsEV-miR model. Thirteen miRNAs (six upregulated and seven downregulated) with repeated emergence were found in Benign_vs_AIS/MIA and Benign_vs_Invasive adenocarcinoma subgroups, showing some similarity in sEV miRNA profiles of AIS/MIA and invasive adenocarcinoma. Meanwhile, three upregulated miRNAs and four downregulated miRNAs were found in Benign_vs_AIS/MIA alone, while nine miRNAs showed altered expression in invasive adenocarcinoma compared with AIS/MIA (three upregulated and six downregulated), suggesting difference in sEV-miRNA profiles between AIS/MIA and invasive adenocarcinoma. Among all DEMs, five miRNAs (miR-30c-5p, miR-30e-5p, miR-500a-3p, miR-125a-5p, and miR-99a-5p) were significantly associated with overall survival from the TCGA database (Figure 6f, $p = 0.0008$; $p = 0.0090$; $p = 0.0110$; $p = 0.0310$, $p = 0.0007$). Taken together, the results of the circulating sEV miRNA analysis demonstrated that the presence of AIS/MIA could affect the differentiation between benign and malignant PNs, and that AIS/MIA presented the intermediate molecular features of circulating sEV miRNAs between benign and malignant PNs.

Discussion

With the development of LDCT scanning for early screening and early diagnosis of lung cancer, the number of detected IPNs has dramatically increased. The inadequate assessment of IPNs' malignancy risk is a major cause of a large number of unnecessary surgeries or an increased risk of cancer metastases; hence, the accurate diagnosis of IPNs becomes one of the major challenges in clinic. However, a non-invasive diagnostic test to assist LDCT to distinguish benign PNs from malignant PNs in IPN diagnosis is not yet available. In this study, we included 208 patients with IPNs, which is the largest sample size among the available studies; and we developed, tested, and validated the CirsEV-miR model to discriminate between benign and malignant PNs using small RNA sequencing of circulating sEV miRNAs. We also calculated a CirsEV-miR score from the model, and we found that this score remarkably increased with the diameter of IPNs, and then gradually changed with the increase in aggressiveness. Moreover, we found that the molecular features of circulating sEV miRNAs of AIS/MIA were between benign and invasive adenocarcinomas, meanwhile, AIS/MIA also possessed unique circulating sEV miRNA profiles different from the other two subgroups. Furthermore, 490 lung cancer samples and follow-up data from TCGA were used to investigate the role of miRNAs in overall survival. In summary, the CirsEV-miR model developed using circulating sEV miRNAs could serve as a non-invasive auxiliary test to help physicians to manage the IPNs for early-stage lung cancer diagnosis. The differentially expressed

circulating sEV miRNAs identified in this study might further serve as potential therapeutic targets in future investigations.

This study focused on IPNs and enrolled the patients with highly suspicious malignant PNs undergoing surgery (404/459, confirmed by pathology diagnosis after surgery) from the routine work-ups in a tertiary hospital. Unlike previous studies that focused on early lung cancer screening in low-prevalence populations, our study recruited patients with a high prevalence, and we used benign PNs as controls. The discrimination between malignant and benign PNs was more difficult than that between the nodule population and non-nodule controls²¹. Benign PNs in this study contained a variety of pathological subtypes without selection, reflecting real-world clinical situations. The benign PNs in the external validation cohort contained different pathological subtypes from those in the training and test cohorts, yet yielding no significant impairment in the model performance, which indicates that our CirsEV-miR model may have broad clinical application potential.

The diameter is an important parameter in assessing the malignant potential of an IPN. The possibility of malignancy positively correlates with the diameter. Lung cancer probability was low in IPNs smaller than 5 mm, intermediate in IPNs with a diameter of 5–10 mm, and high in IPNs larger than 10 mm¹⁸. Clinically, IPNs > 10 mm would be considered to need resection, while IPNs ≤ 10 mm would require repeated scans. We found that CirsEV-miR score was significantly higher in IPNs > 10 mm than in IPNs ≤ 10 mm. Even in IPNs ≤ 10 mm, CirsEV-miR score of malignant PNs was significantly higher than that of benign PNs (Figure 5a, 5b), suggesting that malignancy may increase with IPNs diameter. Our CirsEV-miR model also had good performance in IPNs with a diameter of less than 1 cm, exhibiting AUC of 0.767 and specificity of 91.7% in our cohorts, and AUC of 0.721 and specificity of 75.0% in the external validation cohort (Figure 5c, **Additional file 1: Figure S5**). Taken together, these findings indicate that our model has the ability to distinguish between benign and malignant PNs, even in IPNs ≤ 10 mm, thereby expanding the applications widely.

According to IASLC/ATS/ERS and 2015 WHO classification of lung tumors, AIS is a tumor smaller than 3 cm with no invasion, and MIA denotes tumor smaller than 3 cm with invasion less than 0.5 cm²². Both AIS and MIA patients achieve 100% DFS with complete resection²³. It is believed that LUAD develops stepwise from benign AAH to AIS, to MIA, and eventually to overt invasive adenocarcinoma²⁴. However, most studies have focused on histopathological subtypes, while little is known about the genomic evolution from preinvasive to invasive adenocarcinoma of LUAD. In 2019, Zhang et al. reported that mutations in *EGFR*, *ERBB2*, *NRAS*, and *BRAF*, as well as genomic intratumor heterogeneity and immunoediting, are common and early phenomena harbored by AIS, MIA, and IA, and that mutations in *TP53*, as well as cell mobility and metastasis-related genes, may be later events during neoplastic progression²⁵. Our results also demonstrated that the expression profiles of circulating sEV miRNAs in patients with AIS/MIA were intermediate between the patients with benign PNs and the patients with invasive adenocarcinoma (Figure 6b), supporting the stepwise development hypothesis. Yet, we found seven expression-altered miRNAs unique to AIS/MIA compared with benign PNs, and nine miRNAs unique

to AIS/MIA compared with invasive adenocarcinoma (Figure 6d, 6e, **Additional file 1: Table S5, Additional file 1: Table S6**). Among them, miR-30e-5p, a well-known tumor suppressor that suppresses tumorigenesis via the Sirt1/JAK/STAT3 signaling pathway²⁶, was downregulated in both AIS/MIA and invasive adenocarcinoma compared with benign PNs. However, miR-25-3p and miR-128-3p, which have been reported to promote metastasis, exhibited higher expression in AIS/MIA than in invasive adenocarcinoma^{27,28}. These findings suggest that AIS/MIA exhibit unique features different from benign PNs or invasive adenocarcinoma. It is also proposed that due to a lack of signals in AIS/MIA specimens, the difficulty in distinguishing benign PNs from AIS/MIA IPNs is an intrinsic feature²⁹. This study is the first report that revealed the circulating sEV miRNA signatures of benign, AIS/MIA, and invasive adenocarcinoma, suggesting that the presence of AIS/MIA could impact the differentiation between benign and malignant PNs.

miRNAs bind to the 3'-untranslated region (UTR) of target mRNA, resulting in post-translational gene silencing either by mRNA degeneration or by inhibition of translation. Thousands of miRNAs have been linked to various human diseases, including cancers³⁰. In this study, we identified differentially expressed miRNAs between benign and malignant PNs and constructed the diagnostic model composed of five DEMs using LASSO-penalized regression. Functional enrichment analysis revealed that the target genes of the five miRNAs of our diagnostic model were involved in many cancer-related pathways, such as MAPK, TGF β , Hippo, p53 signaling pathway, focal adhesion, and cell cycle (Figure 4c, 4d). We also found sEV-miRNAs enriched in immune system processes and metabolic processes, in accordance with the previous report that intratumor heterogeneity and immunoediting are early phenomena in PNs²⁵. Moreover, we found seven miRNAs (miR-101-3p, miR-150-5p, miR-30c-5p, miR-30e-5p, miR-500a-3p, miR-125a-5p, and miR-99a-5p) that were associated with better overall survival based on the TCGA databank (Figure 4e, 4f, 6f). Interestingly, these miRNAs have all been reported to function as tumor suppressors³¹⁻³⁶, although some of them were upregulated in tumor tissues^{33,35}, and in circulating sEVs of patients with malignant PNs in our study, indicating that further research is required to clarify the molecular mechanism underlying benign and malignant PNs.

Several limitations of this study are worth mentioning. First, this was a single-center case control study that was performed in a high prevalence population. Thus, the performance of the CirsEV-miR model needs to be explored further in a community where a low prevalence is expected. Second, a prospective validating study was not conducted. Third, the study lacks *in vitro* experiments to verify the miRNA expression level in lung adenocarcinoma tissues. To overcome the above limitations, a prospective multicenter clinical trial in a larger population-based setting is ongoing, and the expression profiles and biological functions of the identified miRNAs need to be validated *in vitro* in lung adenocarcinoma tissues or in lung adenocarcinoma cell lines.

Conclusion

In summary, our study profiled circulating sEV miRNAs in patients with IPNs and provided a diagnostic model, the CirsEV-miR model, based on the measurement of circulating sEV miRNAs to distinguish benign PNs from malignant PNs to assist LDCT scanning for early-stage lung cancer diagnosis. Five circulating sEV miRNAs (let-7b-3p, miR-101-3p, miR-125b-5p, miR-150-5p, and miR-3168) were revealed, and the CirsEV-miR model consisting of these miRNAs was established and validated in 208 patients with IPNs, which is the largest sample size so far. Moreover, CirsEV-miR could discriminate between benign and malignant PNs with diameter ≤ 1 cm, which are mostly difficult to distinguish in clinical setting. We primarily revealed that the molecular feature of AIS/MIA was intermediate of benign and invasive PNs, while exhibiting its own characteristics. These results suggest that circulating sEV miRNAs as diagnostic biomarkers could be integrated with LDCT scan to obtain further evaluation work-ups for IPNs.

Methods

Patient enrollment and study design

This study was approved by the Institutional Review Board in Shanghai Pulmonary Hospital affiliated with Tongji University (K18-199Y) and registered at the Chinese Clinical Trial Registry (<http://www.chictr.org.cn/>) with registration number ChiCTR1800019877. All patients were from Shanghai Pulmonary Hospital affiliated with Tongji University and had signed written consents for their blood samples and clinical information to be used in this study.

Patients with IPNs detected using LDCT scanning, who subsequently underwent surgical resections and diagnosed with LUAD (malignant PNs) and various benign PNs, were enrolled in this study. 199 patients with IPNs were recruited in the training phase from April to May 2019, including 20 patients with benign PNs and 179 with malignant PNs diagnosed using pathological examination. 260 patients with IPNs were recruited in the test phase from September to October 2019, including 35 patients with benign PNs and 225 patients with malignant PNs. Plasma samples had been prospectively collected in a vacutainer with anticoagulant (REF367863; Becton Dickinson, Franklin Lakes, NJ, USA) prior to surgical operation. After the elimination of pathology samples from patients with non-LUAD ($n = 65$), serious hemolysis (above or equal to grade 5, $n = 107$), failure to meet the ratio of benign and malignant PNs set at 1:2 ($n = 155$), and failure in construction of sequencing libraries ($n = 23$), finally, the training cohort consisted of 47 patients with IPNs (17 benign PNs and 30 malignant PNs), and the test cohort consisted of 62 patients with IPNs (24 benign PNs and 38 malignant PNs). In addition, an external cohort consisting of 20 patients with benign PNs and 79 patients with malignant PNs was used for validation. In this case-control study, no healthy individuals were included, and a variety of benign PNs without selection served as controls. The whole study design and inclusion/exclusion criteria are depicted in Figure 1 and **(Additional file 1: Figure S1)**.

The LUAD consisted of three pathological subtypes, namely, AIS, MIA, and invasive adenocarcinoma. Nine patients were diagnosed with AIS, which is technically not a malignant disease. Considering the

perspective of the pathological progression of lung adenocarcinoma, we still classified AIS as a type of malignant nodule. The pathological subtypes of benign PNs were without selection and comprised more than 10 subtypes. The pathological information of all of the samples was obtained from surgically resected tissue sections in accordance with the 2015 WHO Histological Classification of Lung Cancer²². The pathological diagnosis of each patient was confirmed by two pathologists. The tumor–node–metastasis (TNM) stage was determined in accordance with the 8th edition International Association for the Study of Lung Cancer (IASLC) lung cancer staging system³⁷. The pathological subtypes of our training and test cohorts, and those of the external validation cohort, are shown in (**Additional file 2: Figure S2**).

The accuracy of a diagnostic test is usually measured by its sensitivity and specificity³⁸. In this study, sensitivity represented the model's ability to correctly identify individuals with malignant PNs, and specificity represented the model's ability to correctly identify individuals with benign PNs.

Plasma isolation and sEVs isolation

Blood samples were collected from patients in 10-mL vacutainer tubes containing an anticoagulant of K2EDTA (REF367863; Becton Dickinson, Franklin Lakes, NJ, USA), mixed by gently inverting several times, stored with the tubes placed upright, and then transported on ice within 1 hour after collection. To harvest the plasma, the samples were centrifuged at $1600 \times g$ for 10 min at 4°C , after which the hemolysis level was determined and recorded. Samples with hemolysis grade of no more than 4 were used²⁰. The collected supernatant was centrifuged again at $16,000 \times g$ for 15 min at 4°C , and then the 1 mL supernatant was transferred into a fresh 1.5-mL tube and stored at -80°C prior to use.

For the sEV isolation from plasma, a polyethylene glycol-based 3D Medicine isolation reagent²¹ (L3525; 3DMed, Shanghai, China) was used. This isolation reagent has been modified and improved based on the work of Rider et al.²² and has been registered to the National Medical Products Administration as a Class I medical device (#HMXB20190091), specifically for the isolation of sEVs in the clinical setting. The plasma samples were centrifuged at $12,000 \times g$ for 10 min at 4°C after a static water bath incubation at 37°C for 5 min. The supernatant was transferred to a $0.45\text{-}\mu\text{m}$ tube filter (CLS8163-100EA; Costar, Corning, NY, USA), followed by transfer to a $0.22\text{-}\mu\text{m}$ tube filter (CLS8161-100EA; Costar) and then centrifuged at $12,000 \times g$ for 5 min at 4°C . The filtered supernatant was transferred to a fresh 1.5-mL tube. One-quarter volume of an isolation reagent (L3525) was added to the supernatant; gently inverted and incubated for 30 min at 4°C and then centrifuged at $4700 \times g$ for 30 min at 4°C . Finally, the supernatant was removed and the pellets containing the total sEVs were re-suspended in 0.2 mL phosphate-buffered saline (PBS).

Western blot analysis

The isolated sEVs were lysed in 200 μL lysis buffer (P0013B, Beyotime, Shanghai, China); next, the proteins were extracted using an isolation reagent (N3525, 3DMed, Shanghai, China). The protein concentration of the sEVs was measured using a Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific,

USA). 20 µg of total protein was resolved on a 12% SDS-PAGE gel, electrotransferred onto a PVDF membrane (Millipore, USA). The membranes were blocked in 5% non-fat milk for 60 min, and incubated with anti-CD9 antibody (diluted 1:500; cat. no. ab92726; Abcam, Cambridge, UK), anti-CD63 antibody (1:2000, ab216130; Abcam, Cambridge, UK), anti-Syntenin antibody (diluted 1:500; cat. no. ab19903; Abcam, Cambridge, UK), anti-TSG101 polyclonal antibody (diluted 1:500; cat. no. abs115706; Absin Bioscience Inc., Shanghai, China), and anti-Calnexin antibody (diluted 1:1000; cat. no. 2679; Cell Signaling Technology, Danvers, MA, USA) primary antibodies overnight at 4°C. Horseradish peroxidase-conjugated goat anti-rabbit IgG and goat anti-mouse IgG antibodies (Beyotime Biotechnology, China) were used as secondary antibodies. Antibody binding was detected using an enhanced chemiluminescence system according to the manufacturer's protocol (Tanon-5200 Multi; Tanon Science & Technology Co. Ltd., Shanghai, China).

Nanoparticle tracking analysis (NTA)

Nanosight NS 300 system (NanoSight Technology, Malvern, UK) was used to characterize the number and size of EVs. Isolated sEVs were resuspended in PBS at a concentration of 5 µg/mL and were further diluted 100- to 1000-fold, to achieve between 20 and 100 objects/frame. Samples were manually injected into the sample chamber at ambient temperature. Each sample was configured using a 488-nm laser and a high-sensitivity scientific complementary metal-oxide semiconductor camera, and the measurements were performed in triplicate at a camera setting of 13 with an acquisition time of 30 seconds and a detection threshold setting of 7. At least 200 completed tracks were analyzed and obtained per video. Finally, the NTA analytical software (version 2.3) was used to analyze the nanoparticle tracking data of the sEV samples in this study.

Transmission electron microscopy (TEM)

For TEM analysis, plasma sEVs were suspended in PBS prior to fixing in 4% paraformaldehyde and transferred to the carbon-coated electron microscopy grids. They were washed with PBS twice, and the third time with PBS containing glycine (50 mM), each for 3 min; then, they were incubated with PBS containing BSA (0.5%) for 10 min. Finally, the grids were stained with 2% uranyl acetate. After the staining, TEM (H-7650, Hitachi High-Technologies, Japan) was used to analyze the morphology of sEVs.

ExoView analysis

Plasma sEVs were detected using ExoView chips (NanoView Biosciences, Brighton, MA) printed with antibodies against CD63, CD81, CD9, and mouse IgG1 as a negative control. 35 µL samples were dropped onto the chip and incubated for 16 hours. After washing, chips were incubated with a fluorescence antibody cocktail of anti-CD9 (CF® 488), anti-CD81 (CF® 555), and anti-CD63 (CF® 647) for 1 hour at room temperature. Chips were then imaged in the ExoView R100 Scanner (NanoView Biosciences, Brighton, MA). Data were analysed using NanoViewer Software (NanoView Biosciences, Brighton, MA).

RNA isolation from sEVs

RNA was extracted from sEVs using the miRNeasy Serum/Plasma Kit (217184; QIAGEN, Shanghai, China) in accordance with the manufacturer's protocol. The miRNA quality, yield, and distribution were analyzed using the Agilent 2100 Bioanalyzer with Small RNA Chips (5067-1548; Agilent, Savage, MD, USA).

Small RNA libraries preparation and sequencing

To prepare and construct the small RNA sequencing libraries, a NEB Next Multiplex Small RNA Library Prep Set for Illumina (E7300L; New England Biolabs, Ipswich, MA, USA) was used in accordance with the manufacturer's protocol. Briefly, the reverse transcription primer was hybridized after 3' adaptor ligation of 100 ng RNA per sample, following 5' adaptor ligation. A total of 18 PCR cycles were performed with Illumina feasible barcode primers after the first strand cDNA synthesis. The prepared libraries were resolved on NucleoSpin Gel and PCR Clean-up (740609.50; MACHEREY-NAGEL, Germany) and recovered in 30 μ L DNase- and RNase-free water. The DNA quality, yield, and distribution were analyzed using the LabChip® GX Touch™ HT Nucleic Acid Analyzer with DNA High Sensitivity Reagent Kit (CLS760672; PerkinElmer, Waltham, MA, USA) and the DNA Extended Range LabChip (CLS138948; PerkinElmer). A total of 20–25 libraries were pooled into a single sequencing lane and sequenced using an Illumina HiSeq PE150 analyzer.

Bioinformatics analysis of small RNA sequencing data

The 3' adaptors of reads were cleaved using a custom program. Subsequently, the reads were aligned to the human genome hg19 assembly (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/>) using BWA 0.7.12³⁹. The annotations were generated from Gencode v25⁴⁰ and miRBase v21⁴¹ for statistical analysis and to determine expression levels. The miRNA expressions were determined by counting the number of reads mapped to the regions annotated by mature miRNAs. The miRNA mapped by at least two reads in each of the samples and with length less than 30 nt was saved for miRNA expression analysis. The miRNA expression analysis was performed using the voom function in the limma package⁴², with normalization by Trimmed Mean of M-values (TMM) via the edgeR package, and the miRNA expression level was converted to log₂-counts-per-million (logCPM)⁴³. The Empirical Bayes algorithm implemented in ComBat was applied to the training and the test cohort data sets adjusted for batch effects^{44,45}.

Statistical analysis

The samples in the training and test cohorts in this study and the samples in the external validation cohort from another study were analyzed⁴⁶. The diagnostic model was constructed using least absolute shrinkage and selection operator (LASSO) in the training cohort. The test cohort and the external cohort were used to test and validate the diagnostic model. We selected the differentially expressed sEV-miRNAs (DEMs) determined according to the stringent statistical threshold (Student's t-test P-value \leq 0.05, 1.5-fold change, and the mean expression CPM \geq 50) between the benign and malignant PNs. Based on DEMs, the risk scores were generated using LASSO analysis, and the best parameters of the model constructed using LASSO were ultimately selected using 10-fold cross-validation.

Statistical analysis was performed using the statistical programming language R (version 3.6). The dendextend package⁴⁷ in R was used to perform average linkage hierarchical clustering of genes and cases. The heatmap was constructed using the ComplexHeatmap package⁴⁸ in R/Bioconductor. The biological processes of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment of the experimentally validated targets of miRNAs were examined using mirPath v.3, which provided the Expression Analysis Systematic Explorer (EASE) score and false-discovery rates using the Fisher's exact tests and unbiased empirical distributions⁴⁹. The Kaplan–Meier plot analysis of the TCGA data was performed using OncoLnc⁵⁰.

Abbreviations

AAH: atypical adenomatous hyperplasia; AIS: adenocarcinoma in situ; AUC: area under curve; BPs: biological functions; CPM: counts-per-million; DEMs: differentially expressed miRNAs; GO: gene ontology; IPNs: indeterminate pulmonary nodules; ISEV: International Society for Extracellular Vesicles; LASSO: least absolute shrinkage and selection operator; LDCT: low-dose computed tomography; LUAD: lung adenocarcinoma; MFs: molecular functions; MIA: minimally invasive adenocarcinoma; miRNAs: microRNAs; MISEV: Minimal Information for Studies of Extracellular Vesicles; NTA: nanoparticle tracking analysis; OPs: organizing pneumonias; OS: overall survival; PNs: pulmonary nodules; ROC: receiver operating characteristic; sEV: small extracellular vesicles; TCGA: The Cancer Genome Atlas; TEM: transmission electron microscopy; UTR: untranslated region; WB: western blot.

Declarations

Ethics approval and consent to participate

All patients enrolled signed informed consent. This study was approved by institutional review board (K18-199Y).

Consent for publication

Not applicable.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

Authors affiliated with 3D Medicines Inc. are current or former employees. No other disclosures were reported.

Funding

This work was supported by Grant No. 16CR2016A from the Shanghai Metropolitan Key Program.

Authors' contributions

Conceived and designed the experiments: GN-J, D-Z, DD-Z, JF-X. Performed the experiments: YM-Z, CY-W, J-N, XX-X, BN-N, S-C, YN-Z, XF-L. Analyzed the data: JY-Z, HZ-W, W-Z, H-C, CY-W, BB-W, DD-Z, FG-L. Contributed reagents/materials/analysis: D-Z, L-X, JT-Z, WZ-Z, JF-X, GN-J. All authors read and approved the final manuscript.

Acknowledgements

We appreciate the support and participation of the physicians and patients in this study. We would like to thank LetPub (www.letpub.com.cn) for English language editing.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249, doi:10.3322/caac.21660 (2021).
2. National Lung Screening Trial Research, T. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395–409, doi:10.1056/NEJMoa1102873 (2011).
3. Grogan, E. L. *et al.* Thoracic operations for pulmonary nodules are frequently not futile in patients with benign disease. *J Thorac Oncol* **6**, 1720–1725, doi:10.1097/JTO.0b013e318226b48a (2011).
4. Black, W. C. & Baron, J. A. CT screening for lung cancer: spiraling into confusion? *JAMA* **297**, 995–997, doi:10.1001/jama.297.9.995 (2007).
5. Wilson, D. O. *et al.* The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med* **178**, 956–961, doi:10.1164/rccm.200802-3360C (2008).

6. Tao, R. *et al.* Liquid biopsies to distinguish malignant from benign pulmonary nodules. *Thorac Cancer* **12**, 1647–1655, doi:10.1111/1759-7714.13982 (2021).
7. Zhang, Y., Liu, Y., Liu, H. & Tang, W. H. Exosomes: biogenesis, biologic function and clinical potential. *Cell Biosci* **9**, 19, doi:10.1186/s13578-019-0282-2 (2019).
8. Kalra, H. *et al.* Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. *Proteomics* **13**, 3354–3364, doi:10.1002/pmic.201300282 (2013).
9. Boukouris, S. & Mathivanan, S. Exosomes in bodily fluids are a highly stable resource of disease biomarkers. *Proteomics Clin Appl* **9**, 358–367, doi:10.1002/prca.201400114 (2015).
10. Huang, X. *et al.* Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics* **14**, 319, doi:10.1186/1471-2164-14-319 (2013).
11. Ge, Q. *et al.* miRNA in plasma exosome is stable under different storage conditions. *Molecules* **19**, 1568–1575, doi:10.3390/molecules19021568 (2014).
12. Rabinowits, G., Gercel-Taylor, C., Day, J. M., Taylor, D. D. & Kloecker, G. H. Exosomal microRNA: a diagnostic marker for lung cancer. *Clin Lung Cancer* **10**, 42–46, doi:10.3816/CLC.2009.n.006 (2009).
13. Jin, X. *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 Res* **23**, 5311–5319, doi:10.1158/1078-0432.CCR-17-0577 (2017).
14. Smolarz, M. & Widlak, P. Serum Exosomes and Their miRNA Load-A Potential Biomarker of Lung Cancer. *Cancers (Basel)* **13**, doi:10.3390/cancers13061373 (2021).
15. They, C. *et al.* Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **7**, 1535750, doi:10.1080/20013078.2018.1535750 (2018).
16. McWilliams, A. *et al.* Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* **369**, 910–919, doi:10.1056/NEJMoa1214726 (2013).
17. Snoeckx, A. *et al.* Evaluation of the solitary pulmonary nodule: size matters, but do not ignore the power of morphology. *Insights Imaging* **9**, 73–86, doi:10.1007/s13244-017-0581-2 (2018).
18. Horeweg, N. *et al.* Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* **15**, 1332–1341, doi:10.1016/S1470-2045(14)70389-4 (2014).
19. Nakagiri, T. *et al.* Evaluation of the new IASLC/ATS/ERS proposed classification of adenocarcinoma based on lepidic pattern in patients with pathological stage IA pulmonary adenocarcinoma. *Gen Thorac Cardiovasc Surg* **62**, 671–677, doi:10.1007/s11748-014-0429-3 (2014).
20. Boland, J. M. *et al.* Adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive pulmonary adenocarcinoma—analysis of interobserver agreement, survival, radiographic characteristics, and gross pathology in 296 nodules. *Hum Pathol* **51**, 41–50, doi:10.1016/j.humpath.2015.12.010 (2016).

21. Cazzoli, R. *et al.* microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. *J Thorac Oncol* **8**, 1156–1162, doi:10.1097/JTO.0b013e318299ac32 (2013).
22. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* **10**, 1243–1260, doi:10.1097/JTO.0000000000000630 (2015).
23. Behera, M. *et al.* Lung Adenocarcinoma Staging Using the 2011 IASLC/ATS/ERS Classification: A Pooled Analysis of Adenocarcinoma In Situ and Minimally Invasive Adenocarcinoma. *Clin Lung Cancer* **17**, e57–e64, doi:10.1016/j.clc.2016.03.009 (2016).
24. Inamura, K. Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *Int J Mol Sci* **19**, doi:10.3390/ijms19041259 (2018).
25. Zhang, C. *et al.* Genomic Landscape and Immune Microenvironment Features of Preinvasive and Early Invasive Lung Adenocarcinoma. *J Thorac Oncol* **14**, 1912–1923, doi:10.1016/j.jtho.2019.07.031 (2019).
26. Xu, G. *et al.* MicroRNA-30e-5p suppresses non-small cell lung cancer tumorigenesis by regulating USP22-mediated Sirt1/JAK/STAT3 signaling. *Exp Cell Res* **362**, 268–278, doi:10.1016/j.yexcr.2017.11.027 (2018).
27. Zhong, J. *et al.* miR-21-5p promotes lung adenocarcinoma progression partially through targeting SET/TAF-I α . *Life Sci* **231**, 116539, doi:10.1016/j.lfs.2019.06.014 (2019).
28. Cai, J. *et al.* Simultaneous overactivation of Wnt/beta-catenin and TGF β signalling by miR-128-3p confers chemoresistance-associated metastasis in NSCLC. *Nat Commun* **8**, 15870, doi:10.1038/ncomms15870 (2017).
29. Liang, W. *et al.* Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics* **9**, 2056–2070, doi:10.7150/thno.28119 (2019).
30. Lee, Y. S. & Dutta, A. MicroRNAs in cancer. *Annu Rev Pathol* **4**, 199–227, doi:10.1146/annurev.pathol.4.110807.092222 (2009).
31. Zhang, X. *et al.* MiR-101-3p inhibits the growth and metastasis of non-small cell lung cancer through blocking PI3K/AKT signal pathway by targeting MALAT-1. *Biomed Pharmacother* **93**, 1065–1073, doi:10.1016/j.biopha.2017.07.005 (2017).
32. Dai, F. Q. *et al.* miR-150-5p Inhibits Non-Small-Cell Lung Cancer Metastasis and Recurrence by Targeting HMGA2 and beta-Catenin Signaling. *Mol Ther Nucleic Acids* **16**, 675–685, doi:10.1016/j.omtn.2019.04.017 (2019).
33. Chan, L. W. *et al.* MiR-30 Family Potentially Targeting PI3K-SIAH2 Predicted Interaction Network Represents a Novel Putative Theranostic Panel in Non-small Cell Lung Cancer. *Front Genet* **8**, 8, doi:10.3389/fgene.2017.00008 (2017).

34. Liao, X. H., Xie, Z. & Guan, C. N. MiRNA-500a-3p inhibits cell proliferation and invasion by targeting lymphocyte antigen 6 complex locus K (LY6K) in human non-small cell lung cancer. *Neoplasma* **65**, 673–682, doi:10.4149/neo_2018_170516N355 (2018).
35. Wu, S. *et al.* Genetic variations in miR-125 family and the survival of non-small cell lung cancer in Chinese population. *Cancer Med* **8**, 2636–2645, doi:10.1002/cam4.2073 (2019).
36. Mizuno, K. *et al.* FAM64A: A Novel Oncogenic Target of Lung Adenocarcinoma Regulated by Both Strands of miR-99a (miR-99a-5p and miR-99a-3p). *Cells* **9**, doi:10.3390/cells9092083 (2020).
37. Goldstraw, P. *et al.* The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* **11**, 39–51, doi:10.1016/j.jtho.2015.09.009 (2016).
38. Akobeng, A. K. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* **96**, 338–341, doi:10.1111/j.1651-2227.2006.00180.x (2007).
39. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, doi:10.1093/bioinformatics/btp698 (2010).
40. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760–1774, doi:10.1101/gr.135350.111 (2012).
41. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-158, doi:10.1093/nar/gkm952 (2008).
42. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
43. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, doi:10.1093/bioinformatics/btp616 (2010).
44. Lazar, C. *et al.* Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform* **14**, 469–490, doi:10.1093/bib/bbs037 (2013).
45. Taminau, J. *et al.* Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinformatics* **13**, 335, doi:10.1186/1471-2105-13-335 (2012).
46. Zhang, J. T. *et al.* Plasma extracellular vesicle microRNAs for pulmonary ground-glass nodules. *J Extracell Vesicles* **8**, 1663666, doi:10.1080/20013078.2019.1663666 (2019).
47. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720, doi:10.1093/bioinformatics/btv428 (2015).
48. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849, doi:10.1093/bioinformatics/btw313 (2016).
49. Vlachos, I. S. *et al.* DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res* **43**, W460-466, doi:10.1093/nar/gkv403 (2015).

Figures

Figure 1

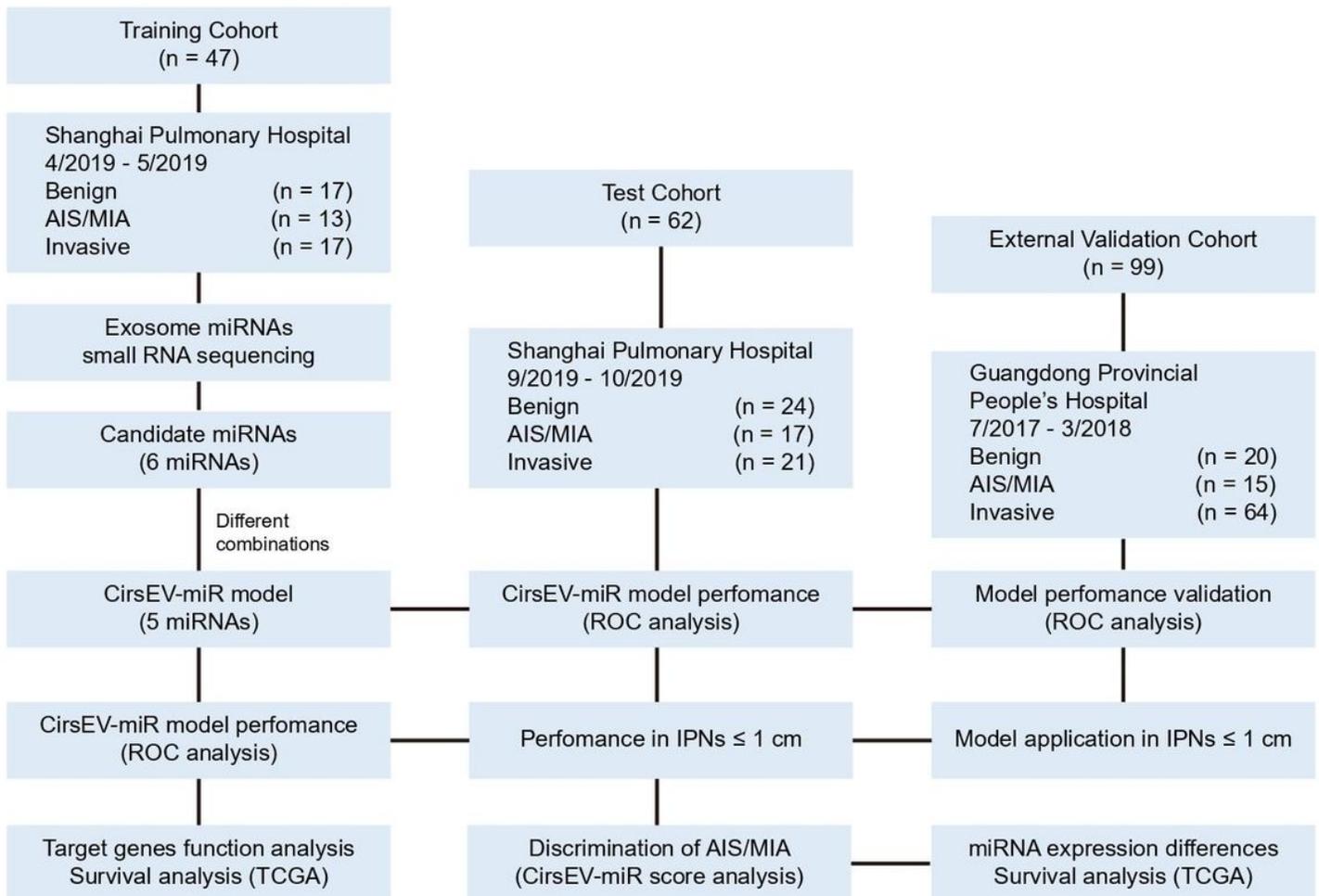


Figure 1

Overall study design and patients in the training, test, and external validation cohorts.

AIS, adenocarcinoma in situ. MIA, minimally invasive carcinoma. ROC, receiver operating characteristic.

Figure 2

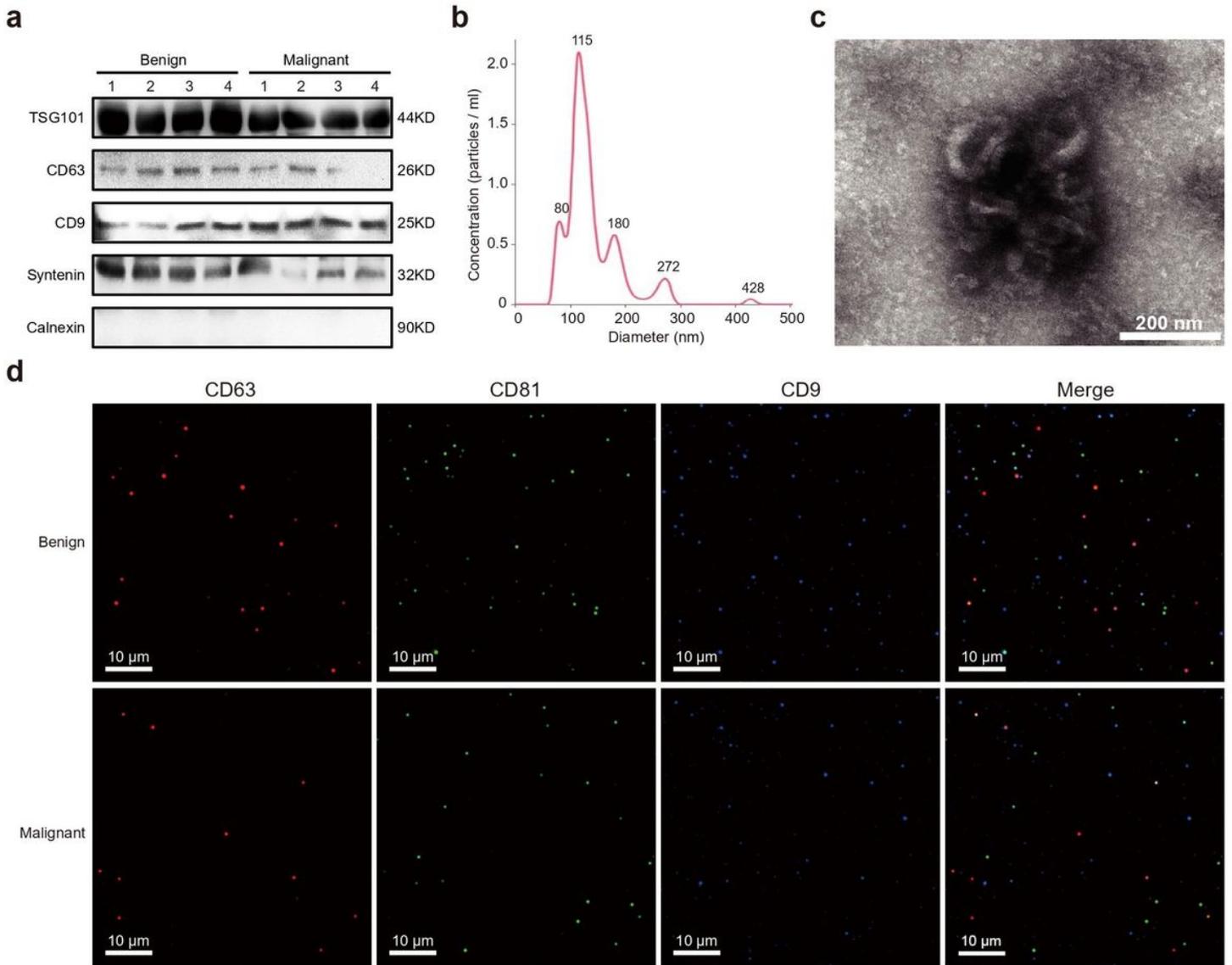


Figure 2

Characterization of circulating small extracellular vesicles (sEVs).

(a) Western blotting identified sEV proteins of eight representative samples, including the sEV positive markers TSG101, CD63, CD9, Systemin, and the negative marker Calnexin. **(b)** Nanoparticle tracking analysis (NTA) results from representative sEV samples. **(c)** Transmission electron microscopic (TEM) images of representative sEVs. **(d)** Representative spots and immunofluorescence staining of CD63 (red), CD81 (green), and CD9 (blue).

Figure 3

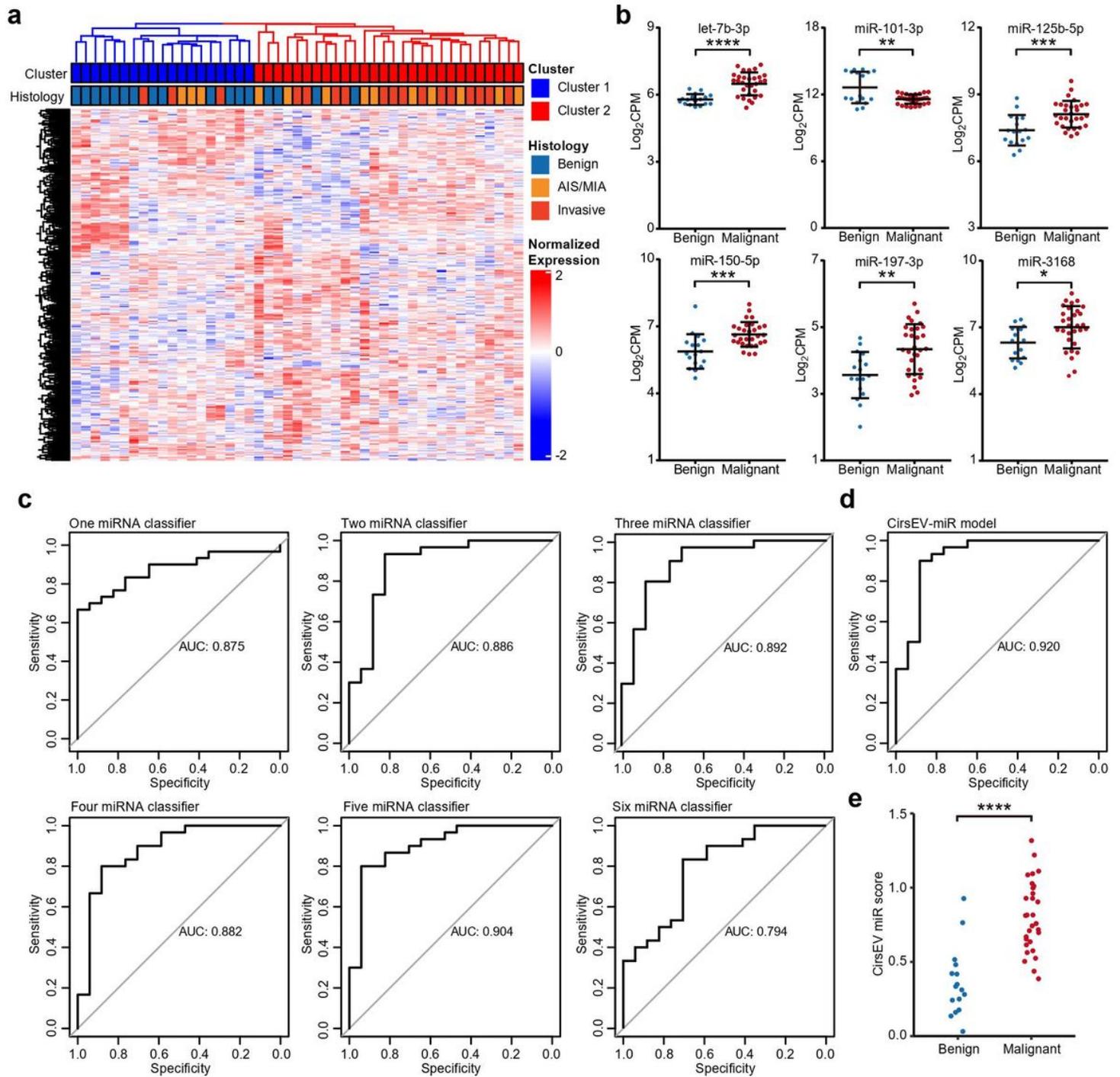


Figure 3

Classifier development process for IPNs.

(a) Heatmap of sEV-miRNA expression in the training cohort by unsupervised hierarchical clustering. (b) Expression levels of six identified DEMs in the training cohort. (c) ROCs of classifiers constructed with different combinations of the six DEMs. Only ROCs with the best diagnostic powers of different miRNA combinations are shown. (d) ROC curve of the CirsEV-miR model in the training cohort. (e) CirsEV-miR scores of benign and malignant PNs in the training cohort.

Figure 4

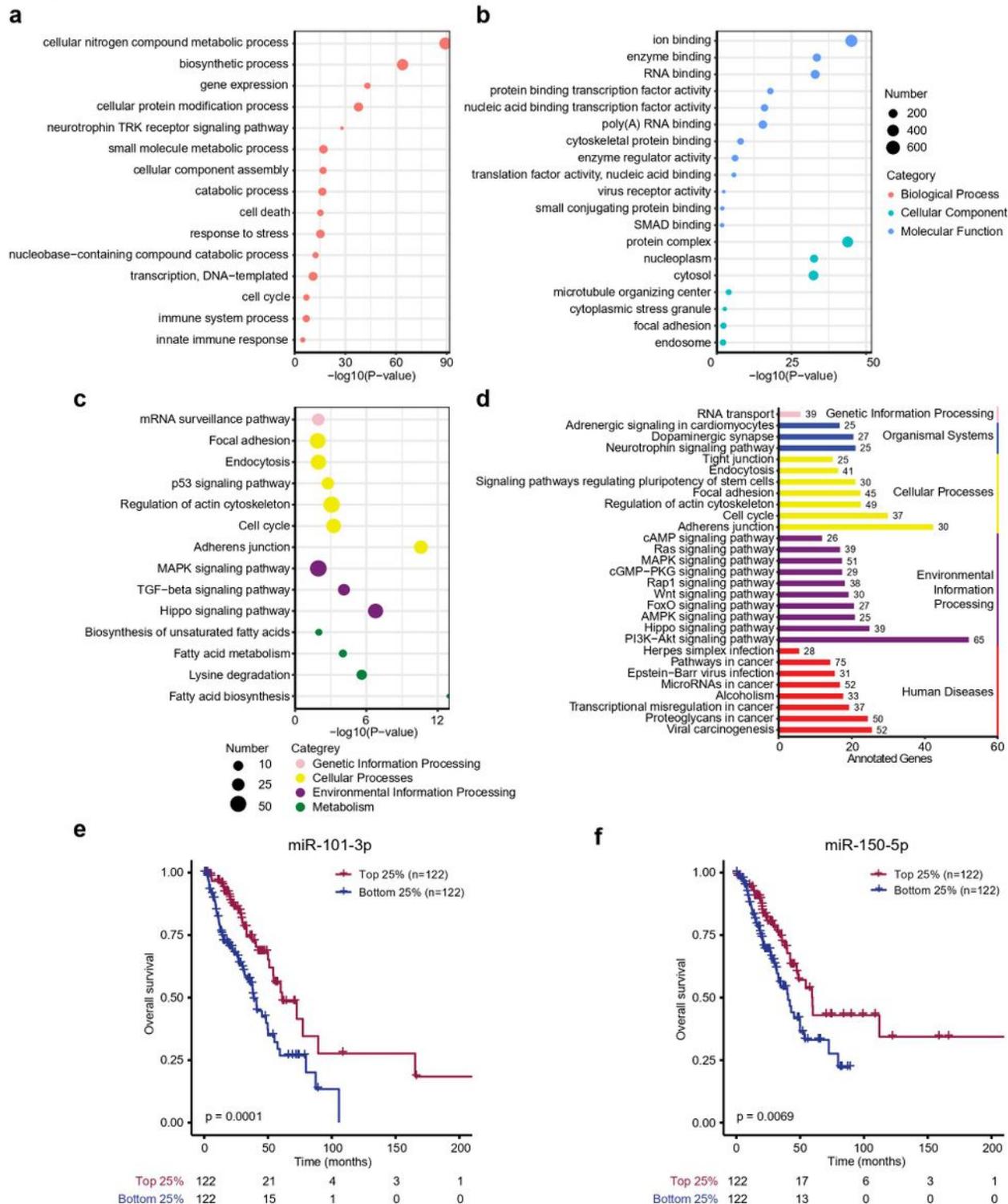


Figure 4

Target analysis and survival analysis of the five DEMs used for CirsEV-miR model development.

A bubble plot of enriched GOs ((Biological process, **a**), and (Cellular component and Molecular Function, **b**)) of target genes of the five DEMs. (**c**) A bubble plot of enriched KEGG pathways of target genes of the five DEMs. (**d**) A bar plot showing the number of DEM target genes in each of the KEGG pathways (target

genes > 20). **(e, f)** Overall survival analysis of miR-101-3p **(e)** and miR-150-5p **(f)** from the TCGA database.

Figure 5

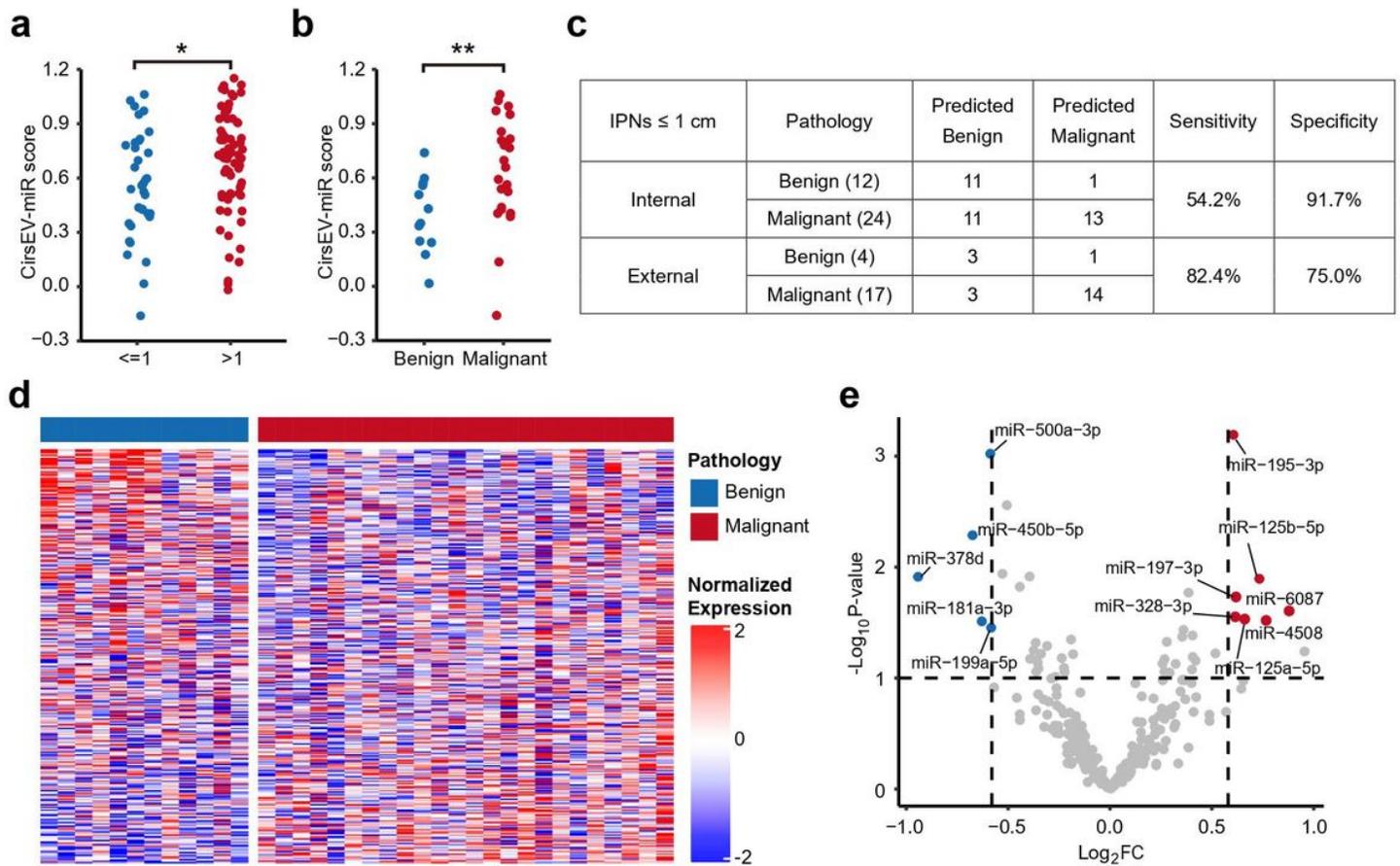


Figure 5

The CirsEV-miR scores increased with the increase in the diameter of IPNs.

(a) CirsEV-miR scores of IPNs ≤ 1 cm and > 1 cm. **(b)** The CirsEV-miR scores of benign and malignant subgroups of IPNs ≤ 1 cm. **(c)** Model sensitivity and specificity of IPNs ≤ 1 cm. **(d)** Gene expression heatmap of benign and malignant subgroups of IPNs ≤ 1 cm. **(e)** Differentially expressed miRNAs between the benign and malignant subgroups of IPNs ≤ 1 cm.

Figure 6

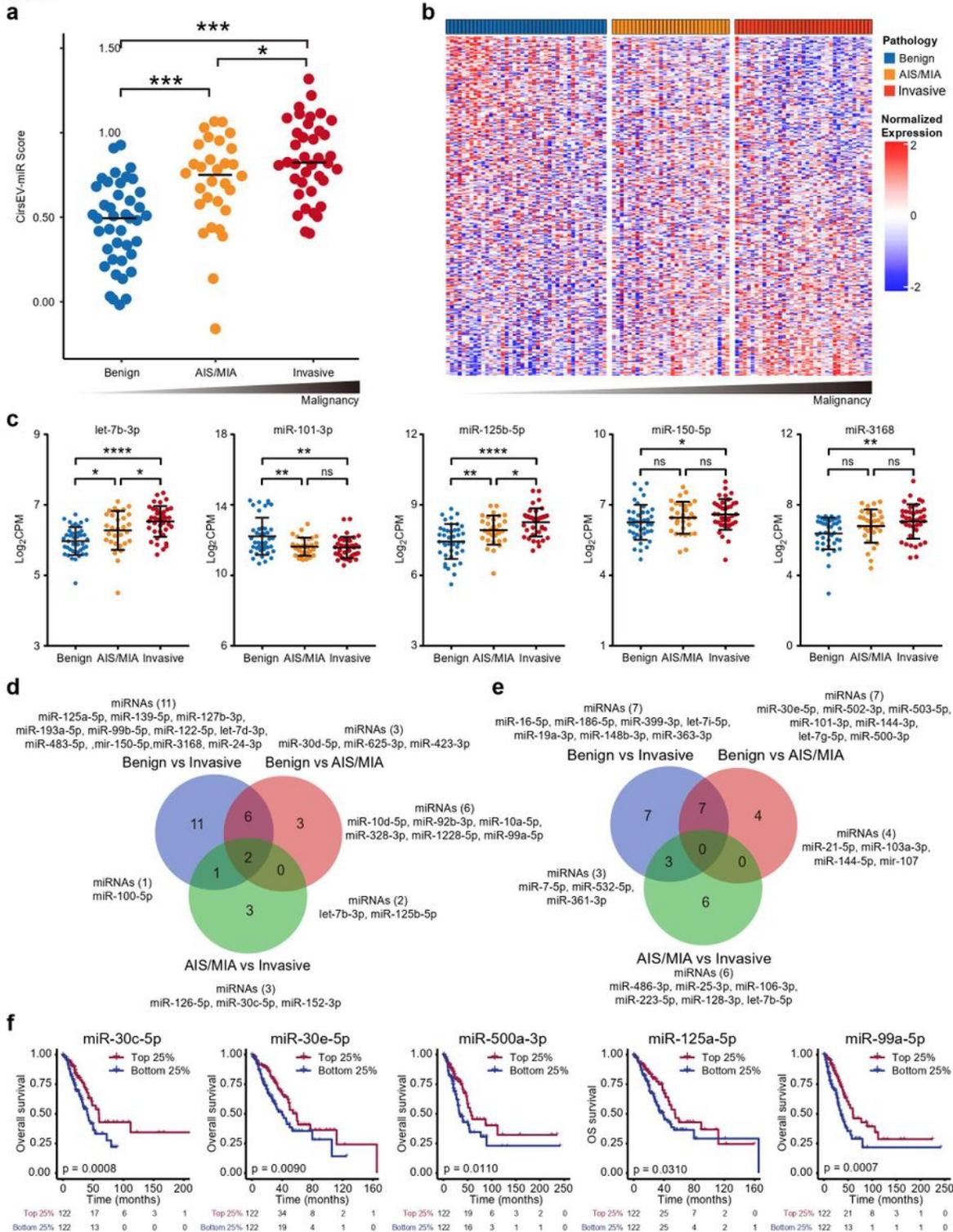


Figure 6

AIS/MIA are intermediate between benign and malignant PNs.

(a) The CirsEV-miR scores of the benign, AIS/MIA, and invasive adenocarcinoma subgroups categorized according to the pathologies. (b) Gene expression heatmap of all genes in the benign, AIS/MIA and invasive adenocarcinoma subgroups. (c) The expression levels of the five DEMs in the benign, AIS/MIA,

and invasive adenocarcinoma subgroups. **(d, e)** Venn diagrams of upregulated **(d)** or downregulated **(e)** genes of the benign, AIS/MIA, and IA subgroups. **(f)** Overall survival analysis of miR-30c-5p, miR-30e-5p, miR-500a-3p, miR-125a-5p, and miR-99a-5p from the TCGA database. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; ns, not significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf](#)